Proteomics in human cancer research. Proteomics Clin Appl

ARTICLE *in* PROTEOMICS. CLINICAL APPLICATIONS · JANUARY 2007

Impact Factor: 2.96 · DOI: 10.1002/prca.200600369 · Source: PubMed

4 AUTHORS, INCLUDING:

Stella B Somiari
Windber Research Institute
38 PUBLICATIONS **2,710** CITATIONS
SEE PROFILE

Malgorzata Czyz
Medical University of Lódz
64 PUBLICATIONS **934** CITATIONS
SEE PROFILE

Richard Idem Somiari
ITSI - Biosciences, Johnstown, PA USA
47 PUBLICATIONS **783** CITATIONS
SEE PROFILE

DOI 10.1002/prca.200600369 *Proteomics Clin. App.* 2007, *1*, 4–17

REVIEW

Proteomics in human cancer research

*Elzbieta Pastwa[1], Stella B. Somiari[2], Malgorzata Czyz[3] and Richard I. Somiari[4]**

[1] Molecular Genetics Department, Medical University of Lodz, Lodz, Poland
[2] Windber Research Institute, Windber, PA, USA
[3] Department of Molecular Biology of Cancer, Medical University of Lodz, Lodz, Poland
[4] Functional Genomics & Proteomics, ITSI-Biosciences, Johnstown, PA, USA

Received: May 16, 2006
Revised: August 27, 2006
Accepted: August 28, 2006

Proteomics is now widely employed in the study of cancer. Many laboratories are applying the rapidly emerging technologies to elucidate the underlying mechanisms associated with cancer development, progression, and severity in addition to developing drugs and identifying patients who will benefit most from molecular targeted compounds. Various proteomic approaches are now available for protein separation and identification, and for characterization of the function and structure of candidate proteins. In spite of significant challenges that still exist, proteomics has rapidly expanded to include the discovery of novel biomarkers for early detection, diagnosis and prognostication (clinical application), and for the identification of novel drug targets (pharmaceutical application). To achieve these goals, several innovative technologies including 2-D-difference gel electrophoresis, SELDI, multidimensional protein identification technology, isotope-coded affinity tag, solid-state and suspension protein array technologies, X-ray crystallography, NMR spectroscopy, and computational methods such as comparative and *de novo* structure prediction and molecular dynamics simulation have evolved, and are being used in different combinations. This review provides an overview of the field of proteomics and discusses the key proteomic technologies available to researchers. It also describes some of the important challenges and highlights the current pharmaceutical and clinical applications of proteomics in human cancer research.

Keywords:
Cancer / Proteome

1 Introduction

Proteomics typically involves the large-scale isolation, identification, and characterization of the protein complement of the genome under a defined state. This discipline has been evolving and developing since the last 20 years [1]. Proteomics now involves the more complex studies of the total set and structure of proteins, their PTMs, and their interactions, thereby opening up a new postgenomic era and opportunities in cancer research.

The central dogma of molecular biology is based on the transcription of genomic DNA into mRNA and translation of mRNA into proteins, the functional units of genes. Cancer development is caused by the accumulation of DNA changes in genes. Unfortunately, genetic abnormalities in tumor cells, which are measured *via* DNA and RNA analysis, do not accurately portray the situation at the protein levels. Cancer is increasingly recognized as a "proteomic disease", but until now only a small percentage of proteins, which are the major players in the cellular processes, have been well described.

Correspondence: Dr. Elzbieta Pastwa, Molecular Genetics Department, Medical University of Lodz, Mazowiecka 6/8, 92-215 Lodz, Poland
E-mail: epastwa@csk.umed.lodz.pl
Fax: +48-42-678-42-77

Abbreviations: ATP, adenosine triphosphate; **2-D DIGE,** 2-D difference gel electrophoresis; **ICAT,** isotope-coded affinity tag; **IDCA,** infiltrating ductal carcinoma; **LCM,** laser capture microdissection; **MD,** molecular dynamic; **MudPIT,** multidimensional protein identification technology; **SAT,** suspension array technology

* Additional corresponding author: Dr. Richard Idem Somiari; E-mail: richard@itsi.us

The ultimate goal of proteomics is to separate and elucidate the structure of all proteins (structural proteomics), their function (functional proteomics), expression profiles (protein profiling), and finally, the detection of protein-based cancer biomarkers that can be targeted for therapy (drug targets/drug development).

New technologies and strategies have facilitated the expansion and categorization of proteomics into three key areas: protein expression profiling, functional proteomics, and structural proteomics (Fig. 1). Protein expression profiling is the most popular, commonly practiced, and most expensive proteomics strategy. It aims to isolate, identify, and define the expression pattern of all differentially expressed proteins as a function of different physiological states, *e.g.*, cancer *versus* normal-state or early-stage cancer *versus* late-stage cancer. Both 2-DE and nongel-based technologies in combination with MS and database search engines are used for protein expression profiling.

Whereas profiling proteomics provides a list of proteins that are differentially expressed, functional proteomics, a specialized form of proteomics, aims to directly elucidate the protein's role in the cancerous process. Functional proteomics, therefore, involves the comprehensive study of proteins and their biological functions and interactions. Human genes produce a lot of proteins *via* their PTMs and mutations. Each protein can participate in or mediate up to 5–10 interactions in a cell. The knowledge of all protein interactions, PTMs, subcellular localization, and protein function is still a big challenge. To meet this challenge, several methods for protein separation and characterization including the affinity-based protein purification approaches like the SELDI-TOF and protein microarrays are being used.

Structural proteomics is a very important proteomic field that aims to determine the tertiary structures of proteins and their structure–function relationships. The precise 3-D structures of proteins help to find their "Achilles heels", *i.e.*, where drugs might turn their activity "on" or "off". The estimation of the structure–function relationship of proteins is important for studies on protein–protein interactions and elucidation of metabolic pathways in cells, tissues, and organisms. Protein structures are identified by experimental approaches such as X-ray crystallography or NMR spectroscopy, and computational methods such as comparative and *de novo* structure prediction or molecular dynamic (MD) simulations. These techniques also involve biochemistry, bioinformatics, and molecular biology.

2 Proteomics strategies

Proteomics mainly consists of three steps: (i) sample preparation, (ii) protein separation, and (iii) protein identification and annotation [1, 2]. The typical workflow for proteomics analysis is shown in Fig. 2. The standard technologies used for protein separation, *e.g.*, 2-D PAGE or HPLC and protein identification, *e.g.*, MS have been around for many years. New technological developments have led to the emergence of the second generation of proteomic techniques typified by large-scale, high-throughput analysis made possible by the extensive use of robotic workstations for sample preparation, spot picking, and in-gel digestion of proteins into peptides, protein identification by PMF, and automated database search and mining tools.

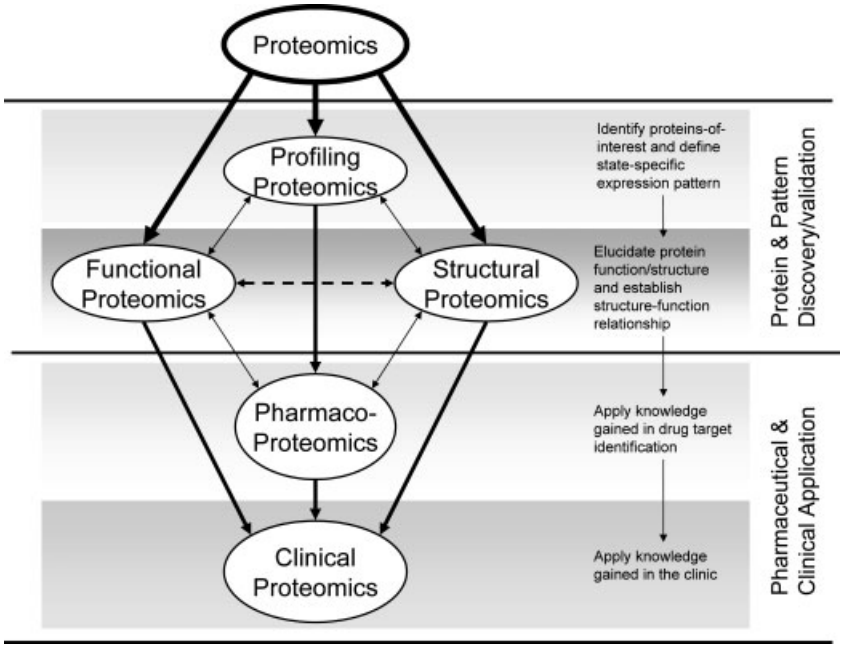

Figure 1. Illustration of the key categories of proteomics and their application. The hallmark of proteomics is the identification of biological markers that can be used for early cancer detection, prognostication and monitoring of treatment, and identification of novel drug targets.

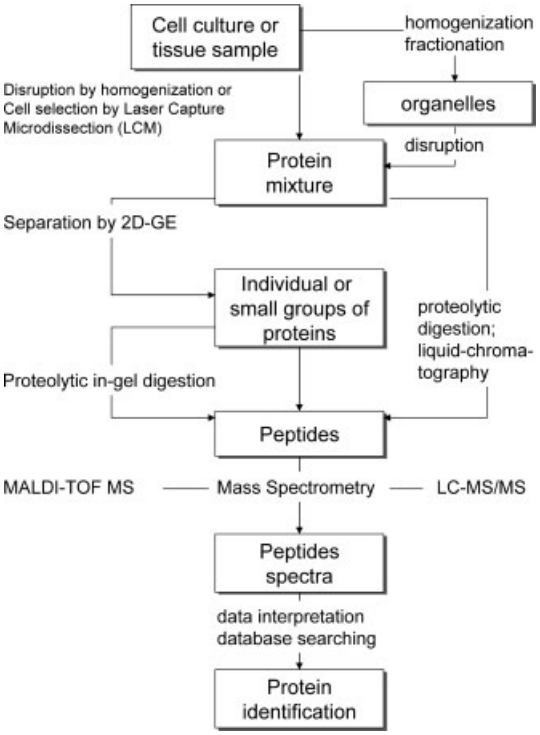

Figure 2. Schematic representation of the standard proteomics workflow process, which starts from sample preparation to protein identification by MS.

2.1 Sample preparation

A biological or clinical sample will typically be a complex mixture of thousands of proteins that differ greatly in their physicochemical and biological properties, and so must be preprocessed prior to analysis. Sample preparation, therefore, is a very crucial step for any large-scale proteomics study. The type of specimens that are currently interrogated by proteomics includes cultured cell lines and tissue (tumor biopsies, serum, plasma, cerebrospinal fluid saliva, and urine). Although cultured cell lines are relatively easier to work with, tissue samples are more informative because they more closely reflect the *in vivo* situation. Unfortunately, tissue specimens are more highly heterogeneous, due to the presence of various cell types under different states. Additionally, because tissue is also more likely to be affected by different sample collection, handling, and storage procedures, it poses an analytical challenge when analyzed in different laboratories and the data generated may not always be a true reflection of the overall functional state of each cell type within the tissue. The use of standard operating procedures for sample collection, processing, and storage and standardization of key analytical steps could minimize sample-to-sample variability, whereas the problem of tissue heterogeneity can be partially solved by the use of the laser capture microdissection (LCM) technology. The LCM which is a device that permits the selection of homogenous tumor populations from normal cells and *vice versa* is now widely applied in cancer research, and it has also been successfully used in many comparative proteomic studies [3–6].

Prefractionation of a protein mixture is usually essential to reduce sample complexity and selectively concentrate low-abundant proteins, for example, in serum and plasma. Subcellular fractionation can provide information about the biological function and localization of proteins, including proteins of very low abundance. A three-stage fractionation strategy comprising of nuclear pellet, vesicular membranes, and cytoplasmic components can be achieved by gradient centrifugation or the use of solvent-based protein fraction enrichment kits that only require bench-top centrifuges. However, it is a challenging task to obtain high reproducibility and resolution of subcellular compartments. Centrifugation can be supplemented with affinity-mediated isolation of organelles using antibodies against cytoplasmic domains of transmembrane proteins present on the surface of organelles. Using this method, significant enrichment could be achieved but new problems are introduced as well. First of all, the composition of proteins in organelles is always changing due to transportation of proteins from and to organelles by vesicles and tubules. Some molecules present in cytoplasm can temporarily interact with organelles to carry out defined functions. Moreover, the fraction of interest often consists of only a small proportion of the starting material, so the requirement for significantly large pool of any material becomes problematic. Despite these limitations, several successful studies employing a qualitative proteomic approach to study organelles and large cellular structures have been described [7, 8]. Protein solubilization is also a strategic point before 2-D PAGE. Soluble proteins can mask membrane-bound proteins, which are nearly insoluble even in harsh chemicals. Acidic proteins can mask basic proteins. The use of proper reagents, *e.g.*, Triton X-100 and DTT for sample preparation helps resolve some of the problems associated with protein solubilization [9].

2.2 Protein separation

Many chromatographic and electrophoresis-based technologies are available for protein separation, but only a few, particularly 2-D PAGE and LC, are routinely used as a step in the proteomic workflow process. The 2-D PAGE is still the benchmark and the most widely used and reported protein separation method. 2-D PAGE has been used for over 20 years to separate complex protein mixtures into individual and small groups of proteins [10, 11]. This protein separation method currently offers the best resolution and coverage amongst all the protein separation methods. In 2-D PAGE, proteins are first separated according to pH (p*I*) and then according to size (molecular weight) [12]. The protein spots are visualized by staining the electrophoresis gels with Coomasie, silver, or a fluorescent dye. Fluorescent dyes such as SYPRO Ruby are becoming popular [13], but they can only be used by those laboratories that can afford the high cost. They

are expensive because you have to employ expensive laser-based excitation scanner systems due to their ruthenium-based dye. SYPRO Ruby has high sensitivity (1–2 ng) and shows a linear dynamic range over three orders of magnitude, whereas Coomasie possesses limited sensitivity and silver, although very sensitive, has poor dynamic range and reproducibility [14].

The availability of IPG strips and high-resolution precast gels has significantly increased reproducibility of 2-D PAGE [15] and reduced the drudgery of performing 2-D PAGE. The major advantage of 2-D PAGE is the possibility for simultaneous separation, visualization, and capture of thousands of unknown proteins. To date, the 2-D PAGE is the only protein separation strategy that can effectively separate, display, and serve as a fraction collector for thousands of proteins simultaneously. Gels can be run first in a broad-range of pH (*e.g.*, 3–10 or 3–11), and then in a narrow p*I* range of, *e.g.*, pH 4–5 to zoom into an area of interest. The number of separations will increase with the number of gels for narrow pH ($\leq$1 U) and the number of different proteomes under investigation. Separation in narrow p*I* can greatly increase resolution and give access to low-abundant proteins that would otherwise be undetected due to saturation by highly expressed proteins [16]. However, the narrow-range IPGs have some drawbacks. When more unfractionated sample is applied, the abundant proteins which possess p*I*s either higher or lower than the actual pH gradient used, will migrate toward each electrode, resulting in streaking. To solve this problem, IPGs may need to be redesigned to allow higher loading capacity and samples may need buffering or prefractionation to remove abundant proteins. Additionally, the use of multicompartment electrolyzers for isoelectrical fractionation of samples prior to narrow-range electrophoresis can be a remedy for this disadvantage [7]. Alkaline IPGs (pH$>$8) are also commercially available and are of special interest, since many important classes of proteins such as DNA-binding proteins (BPs) and integral membrane proteins are predicted to have basic p*I*s. However, to obtain the basic range of IEF, acrylamide matrices stable at alkaline pH and resistant to hydrolysis are needed. There are already some promising results in this respect. Acrylamide N-substituted derivatives such as dimethylacrylamide, found to be more resistant to hydrolysis, are now available [17]. The absence of a suitable acrylamide buffer above pH ~ 10.5 is, however, still a major problem. It is hoped that this will become available in the future.

2.3 Protein identification

Protein identification is an important unit operation in the proteomic workflow process, and the most important step prior to assigning a functional role to candidate proteins. Typically, the proteins are separated by chromatography or electrophoresis and digested with trypsin before assignment of identity by MS, using different types of sources and mass analyzers. Electrophoretic separation, particularly by 2-D PAGE, prior to mass spectrometric analysis is the most widely practiced, although online LC-based separation of peptides and direct injection into mass spectrometers are becoming popular because of the speed, improved resolution, and data-dependent capabilities that have been built into the newer generation LC-MS instruments optimized for proteomics. When 2-D PAGE is used, the proteins separated by electrophoresis are excised from the gels and digested prior to mass spectrometric analysis by PMF (Fig. 2). PMF involves matching constituent peptide fragment masses obtained by MS (peptide masses) to the theoretical peptide masses generated from a protein database. The premise of using PMF for protein identification is that every unique protein will have a unique set of peptides and hence unique peptide masses. Because varying amounts of information can be assembled from each fragment spectrum, the data obtained needs to be interpreted carefully. PMF has been widely and successfully adopted in proteome analysis for microorganisms and yeast [18, 19] and for identification of proteins isolated from disease tissue [13]. MALDI-TOF-MS is the most common high-throughput and rapid method for protein identification by PMF. MS/MS techniques which provide amino acid sequence data are becoming increasingly popular as the primary technique for unambiguous identification of proteins. Additionally, MS/MS allows the fragmentation of individual peptide ions, and after correct interpretation, can provide information about primary protein structure [20]. MS/MS techniques are relatively more advanced and expensive compared to MALDI-TOF-MS techniques. They include technologies such as ESI-MS/MS, ion-traps, quadropole-TOFs, and more recently MALDIs with TOF-TOF capabilities, and the even more advanced linear ion-traps with fourier transformed MS backends (LTQ-FT-MS) instruments. In both PMF and MS/MS approaches, the more the number of peptide matches, the greater the degree of confidence. The strength of these techniques includes high sensitivity and specificity combined with a high-throughput.

The large volume of data generated from proteomics (state-specific expression pattern, protein identity, cellular location, modifications) needs to be converted into their biological meaning. Bioinformatics software's utilizing advanced algorithms are useful tools for studying the involvement of the identified proteins in signaling pathways, important protein–protein interactions, and other biological processes in cancer and normal samples.

3 Expression profiling and functional proteomics

3.1 Gel electrophoresis-based proteomics

3.1.1 2-DE

There are many reports involving the separation of proteins from cancer and noncancer cells by gel electrophoresis. Amongst the gel electrophoretic techniques, the 2-D PAGE

is the most popular and widely used. After separation, the protein spots are excised from the polyacrylamide gel and digested into peptides with enzymes, especially trypsin. The mixture of digested peptides is then analyzed by MS and the obtained peptide spectra are used to search databases to establish the protein identity (Fig. 2). The combination of 2-D PAGE and MS has been used to study protein expression in many cancers, including breast biopsies from metastatic and invasive breast cancer patients [21, 22]. Although the standard 2-D PAGE method is the most widely used, it still has some limitations, such as gel-to-gel variations, difficulties in automating this process and in displaying hydrophobic, basic, and less-abundant proteins. It is also difficult to study high-molecular weight proteins (*e.g.*, proteins >400 kDa) *via* the classical 2-D PAGE approach. In addition, the sensitivity is never below the micromolar (10^{-6}) range, unless larger amount of the sample is applied. Unfortunately this is not possible with all clinical samples.

3.1.2 2-D difference gel electrophoresis (DIGE)

The introduction of a variant of the 2-D PAGE method, termed 2-D DIGE has solved some of the inherent problems associated with the standard 2-D PAGE method. 2-D DIGE has been reported to improve the speed, reproducibility, and sensitivity of 2-D PAGE-based proteomic analysis [13, 23–25], but not the overall drudgery associated with the technology and inability to effectively analyze high-molecular weight proteins. The 2-D DIGE concept is based on the covalent labeling of protein extracts with different fluorescent dyes, *e.g.*, cyanine (Cy2, Cy3, or Cy5) dyes or Alexa dyes [13, 25, 26]. Two main protocols for 2-D DIGE are the minimal and saturated labeling and these have been described by Somiari *et al.* [27]. In the minimal labeling procedure, the protein-to-dye ratio is kept high (>95%), so that only the proteins containing a single dye molecule are visualized on the gel [28]. Typically the test and the reference samples are differentially labeled with Cy dyes (Cy 3 and Cy5, respectively), and equal proportions (1:1 ratio) of the two samples are also mixed and labeled with Cy2. The labeled samples are then mixed and coseparated in the same 2-D PAGE process (Fig. 3). The generated gel is scanned at three wavelengths using a fluorescence imager. A comparison of the images generated by scanning of the 2-D DIGE gels at the Cy3, Cy5, and Cy2 wavelengths allows the quantitation of each spot using fully integrated softwares, *e.g.*, DeCyder™ (GE Healthcare). The use of a third dye, *e.g.* Cy2, permits the comparison of multiple gels and performance of biological variation analysis (BVA) [13, 23, 24]. The protein spots are then picked from SYPRO Ruby™ or Deep-Purple™ stained gels, in-gel digested in robotic workstations, and identified by MS [13]. Incorporation of LCM method improves the accuracy of comparative and quantitative proteomics. Determination of quantitative protein expression profiles between the "test" and the "reference" sample is fast and accurate because it is based on the relative fluorescence intensities captured from a single 2-D DIGE gel. The saturated labeling protocols were designed for the analysis of small protein samples, *e.g.*, samples obtained by LCM [3]. The saturation labeling dyes are different from the minimal labeling dyes because they consist of maleimide reactive groups that covalently bind to the cysteine residues of proteins *via* a thioether linkage [27]. The resulting fluorescent signal generated by labeled protein samples and the total number of proteins detectable on a gel

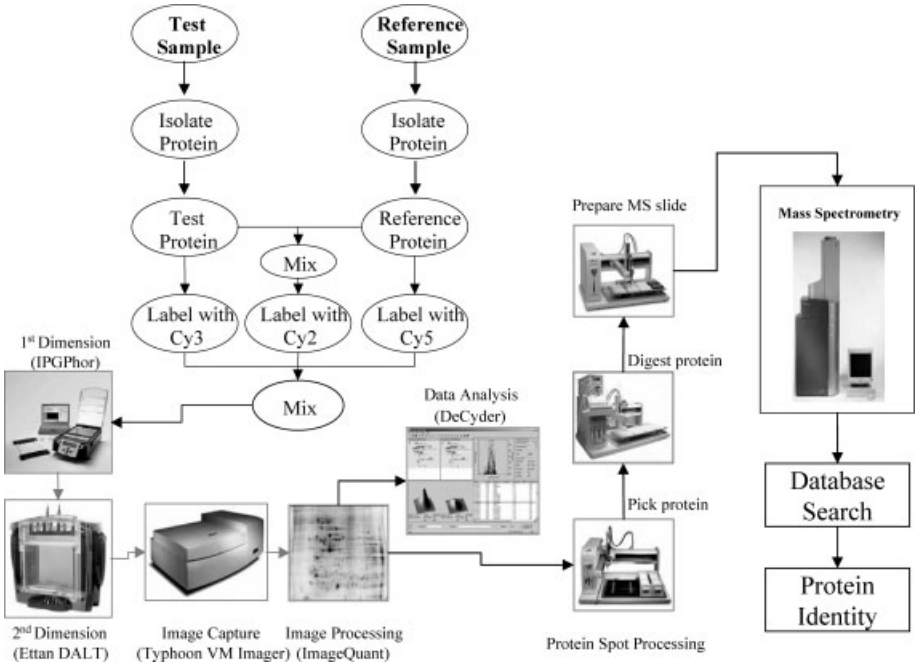

Figure 3. Schematic representation of the 2-D DIGE technology. The 2-D DIGE process involves coseparation of test and reference samples labeled with Cy3 and Cy5, respectively. Aliquots of each sample are mixed and labeled with Cy2. The Cy-labeled samples are mixed and coseparated in the same 2-DE process. After electrophoresis, the gel is scanned at three wavelengths and image generated is analyzed. Candidate protein spots are picked and processed prior to identification by MS. The illustration shows all the equipment and software provided by GE Healthcare for the fully integrated 2-D DIGE process.

is increased [27]. The minimal and saturation labeling protocols have been successfully used in the study of serum and plasma samples [3, 13, 24, 27, 29–32]. The high-throughput proteomic study of human breast infiltrating ductal carcinoma (IDCA) using the 2-D DIGE technology, and a fully automated spot handling workstation has been reported [13]. The 2-D DIGE technology in combination with MS has been found to be a powerful tool for quantifying the differences in protein expression in breast biopsies [13], and between patient-matched esophageal carcinoma cells and normal epithelial cells [23]. A major drawback of separation by 2-DE is the drudgery associated with the process, incomplete recovery of peptide fragments after digestion, and relatively low-throughput.

3.2 Nongel-based proteomics

3.2.1 LC-MS

An alternate protein separation technology upstream of the MS step is 1-D or 2-D LC. The basic liquid chromatograph consists of, the mobile (solvent) phase supply system, the solvent pump and programmer, the sample valve, the separation column (stationary phase), the detector, and a means of displaying and analyzing the results. For proteomic applications, the LC is directly interfaced with an MS to obtain an LC-MS instrument, which now combines the physical separation properties of LC with the mass analysis capabilities of MS. The interface between the liquid phase-end that must continuously deliver liquid and the gas phase-end that must continuously remove all but the gas phase ions is critical. Several interfaces including electrospray ion source or variant such as a nanospray source, fast atom bombardment, and atmospheric pressure chemical ionization interfaces are now available and routinely used in several laboratories (Fig. 4). There are several important differences between the traditional HPLC and the chromatography used in LC-MS, particularly with respect to the internal diameter of column and flow rate. In general, LC-MS instruments utilize columns with internal diameters of 1 mm or less and flow rates that can be as low as 100 nL/min for nanospray instruments.

The LC-MS is now recognized as a powerful technique for many applications which require very high sensitivity and specificity. While the use of the LC-MS has traditionally been oriented toward the specific detection and potential identification of specific chemicals in the presence of other chemicals, there are now many LC-MS instruments that are dedicated for analysis of biological molecules. LC-MS is now routinely applied for protein separation and identification in many proteomics laboratories. A popular adaptation of the LC-MS for proteomics is "bottom-up approach" which generally involves digestion of the proteins (usually with trypsin) into peptides followed by PMF by LC-MS. This "gel-free" approach has many advantages. First, it is relatively fast and semiquantitative; second, it is suitable for searching for low-abundant proteins; and third, smaller amounts of samples, *e.g.*, cancerous cells obtained by LCM can be applied, which is enormously important in proteomic discovery in the early stages of the disease.

LC-MS is now commonly used in pharmacokinetics to study how quickly a drug will be cleared from the hepatic blood flow and organs of the body and in pharmacoproteomics which is the use of proteomics technologies in drug discovery and development. An important advantage of MS is the ability to perform MS/MS-MS. That is, the detector is programmed to select out certain ions for fragmentation. There is, however, a limit in proteome coverage attainable with current online LC-based separations when dealing with a complex genome. Even though the 2-D PAGE-based technologies are still capable of resolving and displaying a greater number of proteins compared to the best HPLC-based techniques available, LC-MS is gaining wide acceptance and fast becoming the platform of choice for laboratories entering the field of proteomics that can afford the instrument.

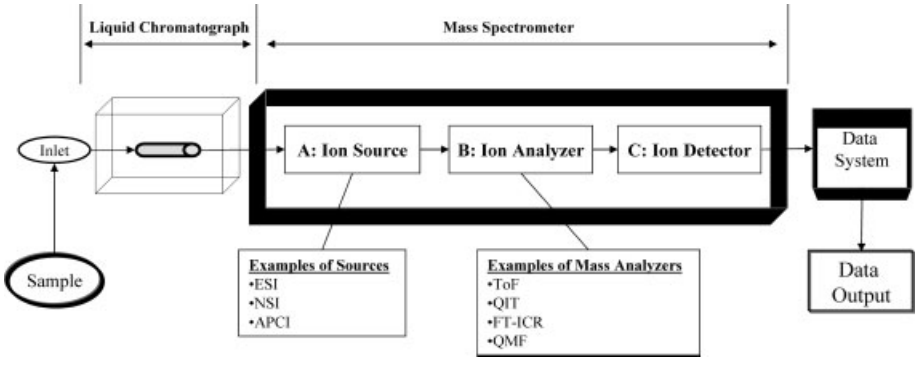

Figure 4. Schematic representation of the LC-MS. The MS section comprises of (A) the source which generates the gas phase ions, (B) the analyzer which sorts the ions, and (C) the detector which detects the selected ions. Examples of the ion sources found in the LC-MS optimized for proteomics include the ESI, nanospray ionization (NSI), and atmospheric pressure chemical ionization (APCI), whereas the examples of analyzers in the instruments include TOF, quadruple IT (QIT), fourier transform ICR (FT-ICR), and quadruple mass filter (QMF).

3.2.2 "Labeled" and "label-free" technologies

Aebersold and coworkers [33] have developed an emerging protein profiling gel-free technology based on isotope tagging of peptides – the isotope-coded affinity tag (ICAT) technology. This strategy involves a single chromatographic step coupled with MS/MS. The ICAT reagent has a thiol-specific reactive group, a linker, and a biotin tag [33, 34]. A linker can contain eight hydrogen atoms (a light version) and eight deuterium isotopic forms (a heavy version). A biotin moiety allows purification of protein mixture *via* affinity chromatography using avidin beads. The light and heavy isotopic tags bind covalently to cysteine moieties of amino acids within proteins. The tags are similar in structure and chemical properties, but are different in mass. When two samples – one treated with the light and the other with the heavy tag – are mixed and purified, the relative intensity of each labeled peptide is measured by MS. The downside of this strategy is the targeting of only cysteine-containing peptides that are accessible to labeling. In consequence, important peptides that do not contain a cysteine residue will not be detected by ICAT. Routinely, about 300–400 proteins *per* sample can be identified by ICAT technology, a number far less than that with the 2-D PAGE technology. The ICAT approach has been used successfully to study protein expression in mammalian cells and human liver cells [35, 36]. Also, it was shown that ICAT was an excellent tool for discovery of proteases making it useful in degradomic research [37]. Recently, a new version of the ICAT method was investigated [38]. The advantage of this new strategy is the use of ^{13}C as the heavy version and ^{12}C as the light version to eliminate the slight delay in retention time of peptides treated with the light tag that occurs with hydrogen and deuterium. This approach leads to about two times more proteins being identified in a single proteomic analysis without losing the accuracy of protein quantification.

The reproducibility and linearity of the new generation MS instruments has led to increasing efforts directed to label free analysis of peptides. This approach which has been termed "differential MS" [39], relies on direct comparison of peptide peak areas among LC-MS runs. The use of the MALDI-TOF as a label-free approach for inhibitor screening has also been described [40]. Briefly, the label-free method specifically finds statistically significant differences in the intensity of both high-abundance and low-abundance ions that account for the variability of measured intensities [39]. Several controlled studies have demonstrated linear responses of peptide ion peaks and showed that mass spectral peak intensities of peptide ions correlate well with protein abundances [41–43] and approximately two-fold differences can be detected [39]. In other reports, RSDs of $<11\%$ were obtained [41, 42] and more than 50% and nearly 90% of peptide ion ratios showed deviations that were less than 10 and 20%, respectively [44].

Recently, the label-free approach was used to analyze plasma from donors diagnosed with invasive breast cancer or benign disease to identify peptide patterns that can be used to classify the donors [45]. Interestingly, artificial neural network models developed with the extracted MS data accurately identified 94% of the cancer cases and 81% of the benign cases [45]. Overall, the label-free quantitative method represents an alternative approach that will allow the global quantitation of differentially expressed proteins without labeling. Because label-free methods are relatively easier and cheaper compared to quantitative approaches requiring stable-isotope labeling, it is expected that this emerging method will find increased utilization and adoption in many laboratories for clinical proteomics and pharmaceutical research.

3.2.3 Multi-dimensional protein identification technology (MudPIT)

Another emerging gel-free proteomic approach is the MudPIT. This technology consists of a multidimensional LC/LC front-end coupled to an MS/MS back-end [46, 47]. A microcapillary column is packed with two independent chromatographic phases: upstream with strong cation exchange resin, and downstream with RP resin, and interfaced with MS. Typically, the complex peptide mixture is loaded onto the biphasic column for simultaneous separation, and peptides are eluted directly from the column into the mass spectrometer to be rapidly analyzed. The above process can be fully automated, thereby satisfying a key requirement of proteomics. Full automation and speed are the most serious drawbacks for all the gel-based proteomics technologies. MudPIT has been used to resolve and identify around 1500 proteins [48], a number that is far below that typically reported for 2-D PAGE-based methods. There is very little information on the application of MudPIT in cancer research. It was reported that this technology has been successfully applied to study cytokinesis proteins, consensus mammalian mediator subunits, and membrane and soluble proteins [49–51]. While the older generation instruments used for MudPIT were not suitable for quantitative proteomics, the introduction of DeCyder MS software (GE Healthcare), and the multidimensional LC (MDLC, GE-Healthcare) which incorporates advanced LC systems and can be coupled to new MS instruments, *e.g.*, the LTQ (Thermo-Finnegan) may position the MudPIT as a key label-free proteomic technology in the near future.

3.2.4 SELDI-TOF technology

SELDI-TOF is a relatively novel affinity-based MS technology, in which proteins (<20 kDa) are selectively absorbed to a chemical or biological surface, and impurities are removed by washing with buffer. The chemical surfaces have hydrophobic, hydrophilic, anionic, or cationic properties, whereas the biological surfaces have antibodies, antigen-binding fragments, DNA, or receptors. High-speed and high-resolution chromatographic separations are achieved by use of several chromatographic arrays and wash condi-

tions [52]. This method is a potentially powerful clinical proteomics tool that can be utilized for the identification of patients at risk for development of cancer, based on the direct analysis of body fluids like serum, plasma, ductal lavage, cerebrospinal fluid, and urine [53–57]. Petricoin *et al.* [58] demonstrated for the first time the potential to use serum proteomic signature for diagnosis of patients in the clinic. SELDI-TOF MS combined with a genetic algorithm has been utilized for the identification of protein peaks in serum that appears to distinguish patients with ovarian cancer from healthy individuals [59]. SELDI-TOF has also been applied to a range of pathological stages of ovarian cancer by other investigators [60] and recently, Pusztai *et al.* [61] used the SELDI-TOF technology to distinguish plasma obtained from patients with breast cancer from healthy individuals. Although there are still questions related to the sensitivity and reproducibility of the SELDI-TOF, and isolation of proteins responsible for the peaks detected by the SELDI-TOF is still challenging, the SELDI-TOF is presently one of the only proteomic tools that has been used to directly analyze clinical samples and stratify patients into "normal" and "cancer" groups. The early detection research network (EDRN) established by the National Cancer Institute (NCI), National Institutes of Health (NIH), USA, reported the cross-laboratory validation of the SELDI-TOF in 2005, making it a platform that holds promise for the screening of body fluids. The strength of the SELDI-TOF approach is that you do not need to know the identity of the protein in order to use the pattern for disease classification. In spite of numerous landmark studies, the usefulness of the SELDI-TOF as a cancer biomarker discovery and screening tool is still debatable. It is apparent however that the technology holds promise and its usefulness will benefit from improved sensitivity and reproducibility as well as standardization of sample collection and preparation protocols.

3.2.5 Protein microarrays

The second affinity-based technology that is becoming important in proteomics is protein microarrays. The idea of protein microarrays is similar to DNA microarrays. In this method, protein chips allow for high-throughput screening of molecular interactions. There are three different approaches. The first one is a direct assay, in which a complex mixture of proteins is immobilized on a solid support, and then some specific proteins from that mixture are visualized using labeled antibodies [62, 63]. The second approach is a sandwich immunoassay, which is opposite to the first method. In this case, antibodies are immobilized on the solid support, and bound proteins are detected using a second labeled antibody [64]. The third method is an antigen capture assay and is similar to the second approach. However, in this case, the bound proteins are detected directly by using chemically labeled complex protein mixtures [65]. Protein microarrays have some drawbacks, which includes cost of producing antibodies, limited availability of antibodies with high specificity and high affinity for the target, and lack of purified recombinant proteins [66].

3.2.6 Suspension array technology (SAT)

The recent introduction of suspended bead-based arrays offers a new approach to multiplexed, real-time, and high-throughput analysis of clinical samples. The commercially available Luminex system (also called BioPlex by BioRad) involves the use of suspensions of specific sets of fluorescent microspheres that are spectrally resolvable [67, 68]. This flow-cytometry-based technology involves optical encoding of micron-sized polymer particles to create suspended microarrays that permits highly multiplexed analysis of complex samples. Each array element can consist of a subpopulation of particles that have distinct optical properties. When coupled with recent innovations for rapid serial analysis of samples, it is believed that molecular analysis with microsphere arrays coated with specific receptors holds significant potential as a general analysis platform for both research and clinical applications [69]. SAT technology can permit the simultaneous analysis of antibodies with specificities for up to 100 different antigens in a single reaction, and the high-throughput screening of up to 1000 sera *per* day [70], thereby satisfying the key proteomic requirement of speed and throughput. A commercially available SAT has been used successfully to study angiogenic factor expression after chemotherapy in an *in vitro* tumor model, [71] and for profiling circulating angiogenic growth factor levels in mice bearing human tumors [72]. A major problem associated with the SAT is the fact that sera may contain antibodies that directly bind to the microspheres, leading to abnormally high nonspecific background. The proportion of such "bead binders" in different serum panels may exceed 5% [70]. Since this problem has been significantly solved by preincubation of sera with PVA, or PVP or partially by preincubation with SeroMap beads provided by Luminex [70], it is expected that this powerful proteomic technology, which has attributes that are lacking in ELISA-based assays, will find increased utilization for biomarker validation and screening in the clinic.

4 Structural and computational proteomics

4.1 Structural proteomics

The complete sequencing of many genomes and the corresponding explosive influx of genomic sequence information have led to the concept of structural proteomics, which is defined as the determination of protein structures on a genome-wide scale [73]. An important bottleneck and the primary obstacle to the full exploitation of the large volume of available genomic information are the structural elucidation

and functional characterization of the gene products. To address this issue, large-scale protein structure determination projects have been initiated in several laboratories [74, 75]. These initiatives which are referred to as "structural proteomics" is initiated because protein 3-D structure is more conserved than DNA sequence and this opens up the possibility of biochemical or biophysical functional characterization *via* structure [76]. An early success of the structural proteomics initiative was the elucidation of the crystal structure of a functionally unannotated protein, MJ0577, from *Methanococcus jannaschii* by Zarembinski *et al.* [77]. The authors reported co-crystallization of the protein with adenosine triphosphate (ATP), and because the crystal structure contained bound ATP, it was postulated that MJ0577 was an ATPase or an ATP-mediated molecular switch. Structural proteomics is also useful in instances when the biochemical function is known. In this regard, knowledge of the 3-D structure of proteins may provide novel insights into the function and mechanism of action of the protein. The availability of structural information allows the direct comparison of proteins with good sequence similarity to the protein of a known structure. Elucidating the structure of proteins can also reveal unique sites for binding, interaction, catalysis, and allosteric regulation. It is believed that the availability of three dimensional protein crystal structures and structures obtained by modeling, in addition to better computing power, will allow the dynamics and function of proteins to be better investigated [76, 78].

X-ray crystallography is still the most powerful method for protein structure determination. First, proteins of interest must be correctly expressed and purified using cloning, expression, and detection tags. The next step is to form crystals of sufficient quality to collect high-resolution data for structure determination. All these steps including expression, crystallization, crystal-mounting alignment in the X-ray beam, and image analysis are now fully automated [79]. These high-throughput capabilities allow the use of small protein sample volumes and performance of thousands of experiments in parallel [80].

The second instrumental method used in structure proteomics does not require crystals, and thus may reduce the data collection time required for structure determination [73]. Moreover, NMR experiments are performed usually in aqueous solutions under conditions similar to the physiological conditions in which the proteins normally function. These features will allow the understanding of the structure–function relationships for a large number of proteins. Biomolecular NMR was found to be useful in the analysis of 25% of proteome of yeast and other organisms, providing its essential role in structural proteomics [81]. However, this technique presents some significant drawbacks, including protein insolubility and aggregation. The amount of time required for the data collection and analysis could also extend to months. This process also still lacks complete automation. Nevertheless, NMR spectroscopy can provide structural and biophysical information that is complementary to crystallographic methods, and these two approaches used together at various laboratories can play synergistic roles in structural proteomics.

4.2 Computational proteomics

The approach used in computational proteomics generally includes comparative structure prediction and *de novo* structure prediction. Both of them are conformational search approaches. Comparative methods are based on the observation that proteins with similar sequences almost always have similar structures. Comparative methods consist of four basic steps: (i) finding a template; (ii) sequence-template alignment; (iii) model building; and (iv) model assessment [82]. High-accuracy comparative models share more than 50% sequence identity to their templates. Comparative structure prediction was used, for example, to construct the homology model of the pore loop domain of potassium channel from *Homo sapiens* based on the crystal structure of the bacterial potassium channel [83, 84]. Some mistakes in this method can come from side-chain packing, shifts or distortions in the core main-chain regions, and improper sequence alignment.

In contrast to comparative methods that are restricted to proteins with at least one known structure, *de novo* structure prediction is free of such a limitation. This strategy is based on the assumption that native form of a protein exists at the global-free energy minimum. Consequently, for a given amino acid sequence, it is possible to find conformational spaces for 3-D structures that are particularly low in free energy. These above conformational search approaches have some disadvantages: slowness, the requirement for human intervention to interpret the results, and the inaccuracy of sequence-structure alignments produced. Despite their drawbacks both comparative and *de novo* structure prediction methods offer some hope to fill the gap between the number of known protein sequences and the number of experimentally determined tertiary structures of proteins.

MD simulations try to characterize particular folded states of proteins in solution without the use of physical force fields and atomic degrees of freedom [85]. Proteins can fold from fully unfolded states to the native state by going through many intermediate states. The tendency in MD simulations is to simplify models to study the mechanism of protein folding *via* these intermediate states, the relationship between the intermediate states, and between the intermediate and native state. There are some successes using this method to study the structures and conformational changes of proteins of interest in cancer, including the RAS proteins and protein kinases [86]. RAS proteins play important roles in signal transduction pathways. Specifically, they regulate cancer-related mechanisms like proliferation, differentiation, and metabolism by acting as molecular switches between the guanosine diphosphate (GDP)-bound inactive form and the guanosine triphosphate (GTP)-bound active forms [86]. The 3-D structure of the RAS-binding do-

main (RBD) of c-Raf-1 (one of three Raf isoforms, c-Raf-1, B-Raf, and A-Raf) has been solved by NMR spectroscopy [87] and using this information, a targeted MD approach has been used to determine possible paths for the switch I and II regions of RAS in going from the GTP-bound form to the GDP-bound form [86]. Protein kinases are enzymes that modify other proteins by chemically adding phosphate groups to them. This results in a fucntional change of the target which manifests as a change in specific activity, cellular localization, and/or protein binding properties. Protein kinases are highly regulated because of the profound effect they have on cells. Dysregulation of kinase activity is associated with many diseases including cancer. Several drugs that inhibit kinases are being developed, and some like Gleevec (imatinib) and Iressa (gefitinib) have already been introduced into the clinic. The ability of kinases to regulate several cellular mechanisms is associated with the range of distinct inactive conformations that exist in several kinase domains [88]. This array of alternative inactive states provides opportunities for the development of selective kinase inhibitors, as exemplified by the success of Gleevec in blocking the activity of break point cluster region–Abelson murine leukemia viral oncogen homolog-1 (BCR–ABl), and its efficacy in the treatment of chronic myelogenous leukemia [88]. By applying crystallographic techniques on the kinase domain of AB1, Levinson *et al.* [88] have been able to reveal a hitherto unappreciated degree of conformational variability in this kinase domain. While several successes utilizing MD simulations have been reported, it is still not feasible in many proteomics laboratories, and there are still limitations associated with this approach because of the transient nature of intermediate states and the necessity of using faster computer speed [89, 90].

5 Clinical proteomics

The hallmark of proteomics is the discovery of diagnostic, prognostic and predictive biological markers (clinical application), and the detection and validation of novel drug targets (pharmaceutical application). Biological markers are physiological measurements or specific biochemicals in the body that are helpful for either disease prediction or treatment. They should have several properties including: (i) sensitivity, selectivity, and a strong association with the disease; (ii) they should be measurable *via* noninvasive procedure; (iii) should be useful in the majority of the population; and (iv) have strong diagnostic, prognostic, and predictive significance [91]. Although it has been reported that a specific biomarker should have a positive predictive value of 100%, a specificity of 99.6%, and a sensitivity of 100% [92], it has not been easy finding a biomarker with a positive predictive value of 100% in all cases. It may also not be possible for all biomarkers to attain this level of accuracy because there will always be some variability and idiopathic differences, resulting in prediction error in real-life situations. Additionally, issues related to sample collection/handling, processing, and storage could affect predictive accuracy of biomarkers. In general, the effort is currently to identify biomarkers that will be sensitive, specific, quantitative, and reproducible when compared to a known standard. Early detection of cancer is critical for effective prevention and treatment and as such, the EDRN is currently employing proteomics technologies in the discovery and evaluation of biomarkers for early cancer detection.

Ovarian cancer is one of the most aggressive gynecological conditions and the high mortality associated with it is most often a consequence of delays in diagnosis. In the first SELDI-TOF-based studies in ovarian cancer by Petricoin *et al.* [58], all cases of early-stage ovarian cancers were correctly identified based on their serum proteomic signature. The proteomic pattern represents thousands of proteins and peptides that are largely unidentified. However, other proteomic studies have combined LCM and 2-D PAGE to discover differentially expressed proteins between two ovarian cancer subtypes; invasive epithelial ovarian cancer and low malignant potential ovarian tumors [93]. Three overexpressed proteins in the invasive samples were identified as glyoxalase I, RhoGDI, and FK506 BP. All of these proteins have been involved in oncogenesis affected through tumor cell apoptosis, DNA synthesis, and mutagenesis [94, 95]. Furthermore, Ye *et al.* [96] identified the haptoglobin-α subunit as a possible serum biomarker for ovarian cancer using the SELDI method. Combination of haptoglobin-α subunit with other serum biomarkers such as CA-125 may produce results with increased specificity and sensitivity. In another study, Zhang *et al.* [97] described the identification of three biomarkers that can be used for the detection of early-stage invasive epithelial ovarian cancer using the SELDI-TOF method. An apolipoprotein A1 and a truncated form of transthyretin were down-regulated in cancer, whereas a cleavage fragment of inter-alpha-trypsin inhibitor heavy chain H4 was up-regulated. In an independent validation, the sensitivity and specificity of combined action of the three biomarkers and CA-125 was higher than that of CA-125 alone, demonstrating their potential to improve the detection of early-stage ovarian cancer.

Breast cancer is still the most common disease diagnosed in women. Since the molecular character of each tumor is different, accurate classification of the tumor can only be achieved by the use of multiple biomarkers [91]. Currently, the clinically validated predictive biomarkers of breast cancer include estrogen-receptor, progesterone-receptor, BRCA1, BRCA2, and the overexpression of ErbB-2 [98, 99]. Proteomic analysis enabled the identification of many differentially expressed proteins in cancer. Somiari *et al.* [13] have used 2-D-DIGE technology for proteomic analysis of human breast IDCA. Examples of the proteins differentially expressed in IDCA and the normal breast tissue were carbonic dehydratase, disulfide isomerase, gelsolin, and fibrinogen β [13]. Fibrinogen γ- and β-chain fragments have been identified in various solid tumor types, and particularly fi-

brinogen γ-chain was detected in plasma from tumor patients but not in plasma from controls [100]. It suggests that overexpression of fibrinogen β is correlated with tumor-associated fibrin deposition [100]. Wulfkuhle *et al.* [22] have used 2-D PAGE technology to identify a number of differentially expressed proteins in ductal carcinoma *in situ* (DCIS) of human breast. Interestingly, it has been shown that proteins differentially expressed in DCIS and IDCA overlap [13, 22]. It demonstrates some relationship triggering both kinds of breast carcinomas. 2-D DIGE technology was also successfully applied in breast cells and epithelial cells with ErbB-2 overexpression [101, 102]. Li *et al.* [103] used the SELDI-TOF in combination with bioinformatics tests to identify three serum biomarkers with high sensitivity and specificity for detection of breast cancer. They observed that BC1 protein was down-regulated, whereas BC2 and BC3 proteins were up-regulated in serum samples from patients with benign breast disease. The identities of these potential biomarkers are currently under investigation. Li *et al.* [104], using the same proteomic method, also selected a panel of three proteins PC-1, PC-2, and PC-3 as the serum biomarkers in the early detection of prostate cancer.

Another method employed in the identification of disease-specific protein biomarkers is the study of the proteomes of organelles and large cellular structures. This study combines large-scale proteomic analysis with classical cell biology methods, and this is supplemented by novel centrifugation and affinity-mediated techniques [105, 106]. Proteomic analyses combined with protein localization and protein-knock-down techniques could identify the phenotype, which is produced by disruption of the protein, and elucidate the function of organelle and the individual protein in disease. An example is the silica-coated bead methodology, which was used to isolate plasma membrane of lung endothelial cells, and allowed the identification of Annexin A1 as a novel target for cancer treatment [107].

6 Pharmacoproteomics

Pharmacoproteomics is the use of proteomics technologies in drug discovery and development. Pharmacoproteomics is a more functional representation of patient–patient variation than that provided by genotyping [108]. This approach is being applied to drug discovery, for both disease-related protein target identification and target validation, and to define and identify patients that will most likely benefit from molecular targeted therapeutic compounds [109]. The first goal is usually achieved by global profiling to detect proteins significantly altered in patients with a particular disorder. Chemical proteomic techniques are now used increasingly in drug discovery. This proteomic approach involves the use of small drug-like molecules that bind to the proteins, and then proteins binding to the ligand are viewed as potential drug targets [110–112]. Chemical proteomics can also identify previously unknown protein targets for a known compound. The 2-D DIGE quantitative proteomic method has been used in the identification of maleate dehydrogenase as a primary target for the novel anticancer agent, E7070 [111, 113]. Proteomics-based approaches can also be used to identify the molecular targets of compounds with unknown mechanism of action. For example, the elusive target of LAF-389 was identified by 2-D PAGE approach *via* molecular changes in expression of cellular proteins following treatment of cell culture with this drug [114, 115]. LAF-389 is a synthetic analog of bengamide B possessing antitumor activity. However, its mechanism of action is not well understood.

The second step of drug discovery is target validation. This step provides more specific information about the structure, biological role, and molecular interactions of proteins involved in the disease process. Target validation occurs when a drug-like compound, selected as a novel target for the disease, has efficacy in its target patient population [116, 117]. Determination of the toxicity of potential drugs is investigated during preclinical drug development. Proteomics is now used as a sensitive tool to screen for drug efficacy and toxicity. For example, these technologies were applied to study a broad range of xenobiotics, nephrotoxicity of drugs, neurotoxicity, and cardiac toxicity [118–120]. Transition from the drug target identification to its validation as a *bonafide* drug target using genomic and proteomic technologies still remains a challenge.

7 Conclusions

The term "proteome" was coined to describe the protein complement of the genome in 1994. Since then many technological advances have led to the emergence of the discipline now termed proteomics. Protein profiling, the most popular proteomics technology, can play an important role in the discovery of disease markers and the detection of novel drug targets. Functional proteomics has a wide application in drug validation, whereas structural proteomics is now influencing drug designs.

New technologies such as the 2-D DIGE and ICAT, MudPIT, SAT, SELDI-TOF and X-ray crystallography, or NMR have been developed as enabling proteomic technologies and their advantages and disadvantages have been outlined in a previous review [121]. However, because of the complexity of the proteome, there is currently no single technology that gives a complete coverage of the proteome. Comprehensive coverage of the proteome will require the use of two or more proteomics technologies in tandem, *e.g.*, 2-D DIGE and ICAT, 2-D PAGE and MudPIT, or X-ray crystallography and NMR. An excellent review was released recently describing efforts, which are being undertaken to develop proteomics as a quantitative technology [122].

Improved sample preparation is also critical for accurate proteomic analysis. A new technique, LCM, coupled with 2-D PAGE, 2-D DIGE, or the SELDI-TOF will allow the study

of specific cell-types in tumor and normal samples in more detail. Also, the use of standardized protocols and prefractionation of cell and tissue samples by novel centrifugation techniques and affinity-mediated methods prior to proteomic analysis will increase the sensitivity of protein detection.

Early detection, diagnosis, and prognostication of cancer are some of the fields in which proteomics is beginning to impact cancer research in many ways. Hundreds of clinical samples can be rapidly analyzed using proteomics technologies like SAT, LC-MS, and the SELDI-TOF, and the development of reliable biomarkers, with sensitivity and specificity better than existing standards, remains the main priority. It is recognized that more useful information about cancer can be generated with multiple markers. However, finding the right combination of biomarkers is still a challenge since multiple pathways and proteins are involved in cancer-related mechanisms, and screening for multiple biomarkers is still time consuming and expensive. The demonstration that patients with ovarian and breast cancer can be identified based on the serum/plasma proteomic signature highlights the potential importance of proteomics as a powerful emerging diagnostic tool for early detection and prognostication of cancer, and for identification of the panel of biomarkers that when used together will provide improved sensitivity and specificity, and be adequate for cancer detection and patient stratification.

There has been frustrations and criticism of certain proteomic technologies by some investigators trying to reproduce or validate studies performed in other laboratories. However, there is no clear distinction between problems attributable to the lack of technical reproducibility and external validity of the data obtained. The majority of the controversies arising from the early report that utilized proteomics platforms like the SELDI-TOF can be linked to attempts to compare experiments that should not be compared in the first place, incompatible experimental design, differences in sample collection, storage and preparation procedures, inappropriate use of home-brew algorithms for data analysis, use of publicly available mass spectral data sets that were not meant to be compared, over fitting of data, and/or variability in the quality of proteinchips used. The landmark studies and associated controversies have however been beneficial to the proteomics community because; (i) they fueled an explosion of interest in the use of MS for detection and screening of biomarkers and (ii) underscored the need to establish standards and best practices in the way specimens are collected, stored, processed, and analyzed. To fully move proteomics from the "bench" to the "clinic", researchers will need to control preanalytical, analytical, and postanalytical sources of variation and define the effect of sample storage/processing, sample type, patient selection, and demographic variables on test outcomes. Also, although informatics is beginning to play a crucial role in high-throughput proteomics research, there are still areas that need improvement. This includes the development of proteomic-specific algorithms capable of dealing with the high dimensionality of proteomics data, integrated systems, databases, sample tracking, and user interfaces. Validation of the biomarkers and the choice of an independent validation set and population will also be critical, because only independently validated biomarkers that show reproducibility across multiple laboratories will have the greatest potential for utilization in the clinic.

In conclusion, it is clear that proteomics is still at its infancy when compared to genomics, and very few biomarkers discovered *via* proteomics have found their way into the clinic. It is apparent that to effectively apply current proteomics technologies for cancer research, biomarker discovery, and drug development, adequate consideration will need to be given to the degree of sensitivity actually needed to detect a clinically useful change. The selection of study populations as well as the development of standard protocols for sample collection, handling, storage, and preparation prior to analysis will be needed. Additionally, informatics tool for data collection and novel algorithms for data analyses and interpretation will need to be developed to streamline and accelerate the discovery of novel biomarkers and proteomic patterns that will be clinically useful.

8 References

[1] Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D. *et al.*, *Biotechnol. Genet. Rev.* 1996, *13*, 19–50.

[2] Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y. *et al.*, *Proc. Natl. Acad. Sci. USA* 2000, *97*, 9390–9395.

[3] Ellsworth, D. L., Russell, S., Deyarmin, B., Sullivan, A. *et al.*, in: Walker, J. (Ed.), *The Proteomics Protocols Handbook*, Humana Press, NJ, USA 2005, pp. 59–66.

[4] Ellsworth, D. L., Shriver, C. D., Ellsworth, R. E., Deyarmin, B. *et al.*, *Biotechniques* 2003, *34*, 42–44, 46.

[5] Shekouh, A. R., Thompson, C. C., Prime, W., Campbell, F. *et al.*, *Proteomics* 2003, *3*, 1988–2001.

[6] Lechpammer, M., Sgroi, D. C., *Expert. Rev. Mol. Diagn.* 2004, *4*, 429–430.

[7] Molloy, M. P., Witzmann, F. A., *Brief Funct. Genomic Proteomic* 2002, *1*, 23–29.

[8] Tribl, F., Gerlach, M., Marcus, K., Asan, E. *et al.*, *Mol. Cell. Proteomics* 2005, *4*, 945–957.

[9] Molloy, M. P., *Anal. Biochem.* 2000, *280*, 1–10.

[10] Klose, J., *Humangenetik* 1975, *26*, 231–243.

[11] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[12] Liebler, D. C., *Introduction to Proteomics: Tools for the New Biology*, Humana Press, Totowa, NJ 2002, pp. 198.

[13] Somiari, R. I., Sullivan, A., Russell, S., Somiari, S. B. *et al.*, *Proteomics* 2003, *3*, 1863–1873.

[14] Berggren, K., Chernokalskaya, E., Steinberg, T. H., Kemper, C. *et al.*, *Electrophoresis* 2000, *21*, 2509–2521.

[15] Poznanovic, S., Schwall, G., Zengerling, H., Cahill, M. A., *Electrophoresis* 2005, *26*, 3185–3190.

[16] Hoving, S., Voshol, H., van Oostrum, J., *Electrophoresis* 2000, *21*, 2617–2621.

[17] Gorg, A., Obermaier, C., Boguth, G., Csordas, A. *et al.*, *Electrophoresis* 1997, *18*, 328–337.

[18] Halden, R. U., Colquhoun, D. R., Wisniewski, E. S., *Appl. Environ. Microbiol.* 2005, *71*, 2442–2451.

[19] Makrantoni, V., Antrobus, R., Botting, C. H., Coote, P. J., *Yeast* 2005, *22*, 401–414.

[20] Mann, M., Wilm, M., *Anal. Chem.* 1994, *66*, 4390–4399.

[21] Bisca, A., DÁmbrosio, C., Scaloni, A., Puglisi, F. *et al.*, *Cancer Lett.* 2004, *204*, 79–86.

[22] Wulfkuhle, J. D., Sgroi, D. C., Krutzsch, H., McLean, K. *et al.*, *Cancer Res.* 2002, *62*, 6740–6749.

[23] Zhou, G., Li, H., DeCamp, D., Chen, S. *et al.*, *Mol. Cell. Proteomics* 2002, *1*, 117–124.

[24] Somiari, R. I., Somiari, S., Russell, S., Shriver, C. D., *J. Chromatogr. B, Analyt. Technol. Biomed. Life Sci.* 2005, *815*, 215–225.

[25] Unlu, M., Morgan, M. E., Minden, J. S., *Electrophoresis* 1997, *18*, 2071–2077.

[26] Von Eggeling, F., Gawriljuk, A., Fiedler, W., Ernst, G. *et al.*, *Int. J. Mol. Med.* 2001, *8*, 373–377.

[27] Somiari, R. I., Russell, S., Somiari, S. B., Sullivan, A. G. *et al.*, in: Walker, J. (Ed.), *The Proteomics Protocols Handbook*, Humana Press, NJ, USA 2005, pp. 223–237.

[28] Tonge, R., Shaw, J., Middleton, B., Rowlinson, R. *et al.*, *Proteomics* 2001, *1*, 377–396.

[29] Russell, S., Sullivan, A. G., Shriver, C., Brzeski, H. *et al.*, *Proceedings of the ASMS* 2002, A020445, June 2nd-6th, Orlando, Florida.

[30] Brzeski, H., Katenhusen, R. A., Sullivan, A. G., Russell, S. *et al.*, *Biotechniques* 2003, *35*, 1128–1132.

[31] Sullivan, A., Brzeski, H., Somiari, R., Shriver, C., in: Walker J. (Ed.) *Preparation of Serum/Plasma Samples for 2D-PAGE in the Proteomics Protocols Handbook*, Humana Press Inc., NJ, USA 2005, pp. 49–54.

[32] Brzeski, H., Sullivan, A., Somiari, R., Shriver, C., in: Walker, J. (Ed.), *The Proteomics Protocols Handbook*, Humana Press, NJ, USA 2005, pp. 119–132.

[33] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F. *et al.*, *Nat. Biotechnol.* 1999, *17*, 994–999.

[34] Aebersold, R., *J. Infect. Dis.* 2003, *187*, S315–S320.

[35] Tian, Q., Stepaniants, S. B., Mao, M., Weng, L. *et al.*, *Mol. Cell. Proteomics* 2004, *3*, 960–969.

[36] Yan, W., Lee, H., Yi, E. C., Reiss, D. *et al.*, *Genome Biol.* 2004, *5*, R54.

[37] Tam, E. M., Morrison, C. J., Wu, Y. I., Stack, M. S. *et al.*, *Proc. Natl. Acad. Sci. USA* 2004, *101*, 6917–6922.

[38] Weiner, M. C., Sachs, J. R., Dayanova, E. G., Yates, N. A., *Anal Chem.* 2004, *76*, 6085–6096.

[39] Yi, E. C., Li, X. J., Cooke, K., Lee, H. *et al.*, *Proteomics* 2005, *5*, 380–387.

[40] Greis, K. D., Zhou, S., Burt, T. M., Carr, A. N. *et al.*, *J. Am. Soc. Mass Spectrom.* 2006, *17*, 815–822.

[41] Bondarenko, P. V., Chelius, D., Shaler, T. A., *Anal Chem.* 2002, *74*, 4741–4749.

[42] Celius, D., Bondarenko, P. V., *J. Proteome Res.* 2002, *1*, 317–323.

[43] Wang, W., Zhou, H., Lin, H., Roy, S. *et al.*, *Anal. Chem.* 2003, *75*, 4818–4826.

[44] Wang, G., Wu, W. W., Zeng, W., Chou, C. L., *J. Proteome Res.* 2006, *5*, 1214–1223.

[45] Ru, Q. C., Zhu, L. A., Silberman, J., Shriver, C. D., *Mol. Cell. Proteomics* 2006, *5*, 1095–1104.

[46] Wolters, D. A., Washburn, M. P., Yates, J. R. III, *Anal. Chem.* 2001, *73*, 5683–5690.

[47] Wu, C. C., MacCoss, M. J., *Curr. Opin. Mol. Ther.* 2002, *4*, 242–250.

[48] Naistat, D. M., Leblanc, R., *J. Environ. Pathol. Toxicol. Oncol.* 2004, *23*, 161–178.

[49] Skop, A. R., Liu, H., Yates, J. III, Meyer, B. J. *et al.*, *Science* 2004, *305*, 61–66.

[50] Sato, S., Tomomori-Sato, C., Parmely, T. J., Florens, L. *et al.*, *Mol. Cell.* 2004, *14*, 685–691.

[51] Wu, C. C., MacCoss, M. J., Howell, K. E., Yates, J. R. III, *Nat. Biotechnol.* 2003, *21*, 532–538.

[52] Merchant, M., Weinberger, S. R., *Electrophoresis* 2000, *21*, 1164–1177.

[53] Cazares, G. W., Jr, Leung, L. H., Nasim, S. M., Adam, B. L. *et al.*, *Prostate Cancer Prostatic. Dis.* 1999, *2*, 264–276.

[54] Wadsworth, J. T., Somers, K. D., Cazares, L. H., Malik, G. *et al.*, *Clin. Cancer Res.* 2004, *10*, 1625–1632.

[55] Hu, Y., Zhang, S., Yu, J., Liu, J. *et al.*, *Breast* 2005, *14*, 250–255.

[56] Yang, S. Y., Xiao, X. Y., Zhang, W. G., Zhang, L. J. *et al.*, *BMC Cancer* 2005, *5*, 83.

[57] Schwegler, E. E., Cazares, L., Steel, L. F., Adam, B. L. *et al.*, *Hepatology* 2005, *41*, 634–642.

[58] Petricoin, E. F., Ardekani, A. M., Hitt, B. A., Levine, P. J. *et al.*, *Lancet* 2002, *359*, 572–577.

[59] Rapkiewicz, A. V., Espina, V., Petricoin, E. F. III, Liotta, L. A., *Eur. J. Cancer* 2004, *40*, 2604–2612.

[60] Espina, V., Dettloff, K. A., Cowherd, S., Petricoin, E. F. III *et al.*, *Expert Opin. Biol. Ther.* 2004, *4*, 83–93.

[61] Pusztai, L., Gregory, B. W., Baggerly, K. A., Peng, B. *et al.*, *Cancer* 2004, *100*, 1814–1822.

[62] Paweletz, C. P., Charboneau, L., Bichsel, V. E., Simone, N. L. *et al.*, *Oncogene* 2001, *20*, 1981–1989.

[63] Charboneau, L., Scott, H., Chen, T., Winters, M. *et al.*, *Brief Funct. Genomic Proteomic* 2002, *1*, 305–15.

[64] Schweitzer, B., Roberts, S., Grimwade, B., Shao, W. *et al.*, *Nat. Biotechnol.* 2002, *20*, 359–365.

[65] Haab, B. B., Dunham, M. J., Bron, P. O., *Genome Biol.* 2001, *2*, 0004.1–0004.13.

[66] Baak, J. P., Path, F. R., Hermsen, M. A., Meijer, G. *et al.*, *Eur. J. Cancer* 2003, *39*, 1199–1215.

[67] Fulton, R. J., McDade, R. L., Smith, P. L., Kienker, L. J. *et al.*, *Clin. Chem.* 1997, *43*, 1749–1756.

[68] Krenn, V., Petersen, I., Haupl, T., Koepenik, A. *et al.*, *Pathol. Res. Pract.* 2004, *200*, 95–103.

[69] Nolan, J. P., Sklar, L. A., *Trends Biotechnol.* 2002, *20*, 9–12.

[70] Waterboer, T., Sehr, P., Pawlita, M., *J. Immunol. Methods* 2006, *309*, 200–204.

[71] Keyes, K., Cox, K., Treadway, P., Mann, L. *et al.*, *Cancer Res.* 2002, *62*, 5597–5602.

[72] Keyes, K. A., Mann, L., Cox, K., Treadway, P. *et al.*, *Cancer Chemother. Pharmacol.* 2003, *51*, 321–7.

[73] Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A. *et al.*, *PNAS* 2002, *99*, 1825–1830.

[74] Burley, S. K., Almo, S. C., Bonanno, J. B., Capel, M. *et. al*, *Nature Genetics* 1999, *23*, 151–157.

[75] Dry, S., McCarthy, S., Harris, T., *Nat. Struct. Biol. 2000, Structural Genomics Supplement*, 2000, 946–949.

[76] Wild, D. L., Saqi, M. A. S., *Current Proteomics* 2004, *1*, 59–65.

[77] Zarembinski, T. I., Hung, L. W., Mueller-Dieckmann, H. J., Kim, *Proc. Natl. Acad. Sci. USA* 1998, *95*, 15189–15193.

[78] Maggio, E. T., Ramnarayan, K., *Trends Biotechnol.* 2001, *19*, 266–272.

[79] Muchmore, S. W., Olson, J., Jones, R., Pan, J. *et al.*, *Struct. Fold Des.* 2000, *8*, R243–R246.

[80] Abola, E., Kuhn, P., Earnest, T., Stevens, R. C., *Nat. Struct. Biol.* 2000, *7*, 973–977.

[81] Montelione, G. T., Zheng, D., Huang, Y. J., Gunsalus, K. C. *et al.*, *Nat. Struct. Biol.* 2000, *7*, 982–985.

[82] Liu, H. L., Hsu, J. P., *Proteomics* 2005, *5*, 2056–2068.

[83] Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A. *et al.*, *Science* 1998, *280*, 69–77.

[84] Liu, H. L., Lin, J. C., *Proteins* 2004, *55*, 558–567.

[85] Kovacs, H., Mark, A. E., Johansson, J., van Gunsteren, W. F., *J. Mol. Biol.* 1995, *247*, 808–822.

[86] Ma, J., Karplus, M., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 11905–11910.

[87] Emerson, D. S., Madison, V. S., Palermo, R. E., Waugh, D. S. *et al.*, *Biochemistry* 1995, *34*, 6911–6918.

[88] Levinson, NM, Kunchment, O, Shen, K, Young, M. A. *et al*, *PLoS Biology* 2006, *4*, 0753–0764.

[89] Eaton, W. A., Munoz, V., Hagen, S. J., Jas, G. S. *et al.*, *Annu. Rev. Biophys. Biomol. Struct.* 2000, *29*, 327–359.

[90] Chamberlain, A. K., Marqusee, S., *Adv. Protein Chem.* 2000, *53*, 283–328.

[91] Arciero, C., Somiari, S. B., Shriver, C. D., Brzeski, H. *et al.*, *Int. J. Biol. Markers* 2003, *18*, 241–272.

[92] Jacobs, I. J., Skates, S. J., MacDonald, N., Menon, U. *et al.*, *Lancet* 1999, *353*, 1207–1210.

[93] Jones, M. B., Krutzsch, H., Shu, H., Zhao, Y. *et al.*, *Proteomics* 2002, *2*, 76–84.

[94] Olofsson, B., *Cell Signal* 1999, *11*, 545–554.

[95] Takahashi, T., Irie, R. F., Nishinaka, Y., Hoon, D. S., *Clin. Cancer Res.* 1997, *3*, 1363–1370.

[96] Ye, B., Cramer, D. W., Skates, S. J., Gygi, S. P. *et al.*, *Clin. Cancer Res.* 2003, *9*, 2904–2911.

[97] Zhang, Z., Bast, R. C., Jr., Yu, Y., Li, J. *et al.*, *Cancer Res.* 2004, *64*, 5882–5890.

[98] Leinster, S. J., Downey, H., Gibbs, T. J., *Shared Care for Breast Disease*, Isis Medical Media Ltd., Oxford, UK 2000, p. 125.

[99] Dowsett, M., Cooke, T., Ellis, I., Gullick, W. J. *et al.*, *Eur. J. Cancer* 2000, *36*, 170–176.

[100] Gerner, C., Steinkellner, W., Holzmann, K., Gsur, A. *et al.*, *Thromb. Haemost.* 2001, *85*, 494–501.

[101] Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M. J. *et al.*, *Mol. Cell. Proteomics* 2002, *1*, 91–98.

[102] Chan, H. L., Gharbi, S., Gaffney, P. R., Cramer, R. *et al.*, *Proteomics* 2005, *11*, 2908–2926.

[103] Li, J., Zhang, Z., Rosenzweig, J., Wang, Y. Y. *et al.*, *Clin. Chem.* 2002, *48*, 1296–1304.

[104] Li, J., White, N., Zhang, Z., Rosenzweig, J. *et al.*, *J. Urol.* 2004, *171*, 1782–1787.

[105] Brunet, S., Thibault, P., Gagnon, E., Kearney, P. *et al.*, *Trends Cell. Biol.* 2003, *13*, 629–638.

[106] Yates, J. R., Gilchrist, A., Howell, K. E., Bergeron, J. J. M., *Nature Rev.* 2005, *6*, 702–714.

[107] Oh, P., Li, Y., Yu, J, Durr, E. *et al.*, *Nature* 2004, *429*, 629–635.

[108] Jain, K. K., *Pharmocogenomics* 2004, *5*, 331–336.

[109] Wulfkuhle, J. D., Edmiston, K. H., Liotta, L. A., Petricoin, E. F., *Nat. Clin. Pract. Oncol.*, 2006, *3*, 256–268.

[110] Jeffery, D. A., Bogyo, M., *Curr. Opin. Biotechnol.* 2003, *14*, 87–95.

[111] Oda, Y., Owa, T., Sato, T., Boucher, B. *et al.*, *Anal. Chem.* 2003, *75*, 2159–2165.

[112] Huang, J., Zhu, H., Haggarty, S. J., Spring, D. R. *et al.*, *Proc. Natl. Acad. Sci. USA* 2004, *101*, 16594–16599.

[113] Kopec, K. K., Bozyczko-Coyne, D., Williams, M., *Biochem. Pharmacol.* 2005, *69*, 1133–1139.

[114] Towbin, H., Bair, K. W., DeCaprio, J. A., Eck, M. J. *et al.*, *J. Biol. Chem.* 2003, *278*, 52964–52971.

[115] Fischer, H. P., Heyse, S., *Curr. Opin. Drug Discov. Devel.* 2005, *8*, 334–346.

[116] Williams, M., *Curr. Opin. Pharmacol.* 2003, *3*, 571–577.

[117] Bilello, J. A., *Curr. Mol. Med.* 2005, *5*, 39–52.

[118] Watanabe, T., Lalwani, N. D., Reddy, J. K., *Biochem. J.* 1985, *227*, 767–775.

[119] Arnott, D., O'Connell, K. L., King, K. L., Stults, J. T., *Anal. Biochem.* 1998, *258*, 1–18.

[120] Krapfenbauer, K., Berger, M., Lubec, G., Fountoulakis, M., *Electrophoresis* 2001, *22*, 2086–2091.

[121] Engwegen, J. Y. M. N., Gast, M. W., Schellens, J. H. M., Beijnen, J. H., *Trends Pharmacol. Sci.* 2006, *27*, 251–259.

[122] Ong, S.-E., Mann, M., *Nat. Chem. Biol.* 2005, *1*, 252–262.

