Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography B, 815 (2005) 215–225
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Review

Proteomics of breast carcinoma

Richard I. Somiari[a,*], Stella Somiari[b], Stephen Russell[a], Craig D. Shriver[c]

[a] *Functional Genomics and Proteomics, ITSI Biosciences, 938 Mt. Airy Drive, Johnstown, PA 15904, USA*
[b] *Clinical Breast Care Project, Windber Research Institute, Windber, PA, USA*
[c] *General Surgery Services/Clinical Breast Care Project, Walter Reed Army Medical Center, Washington, DC, USA*

Received 31 August 2004; accepted 1 November 2004
Available online 15 December 2004

Abstract

Beast cancer is the most diagnosed cancer in women, accounting for approximately 40,000 deaths annually in the USA. Significant advances have been made in the areas of detection and treatment, but a significant number of breast cancers are detected late. The advent of proteomics provides the hope of discovering novel biological markers that can be used for early detection, disease diagnosis, prognostication and prediction of response to therapy. Several proteomics technologies including 2D-PAGE, 2D-DIGE, ICAT, SELDI-TOF, MudPIT and protein arrays have been used to uncover molecular mechanisms associated with breast carcinoma at the global level, and a number of these technologies, particularly the SELDI-TOF hold promise as a proteomic approach that can be applied at the bedside for discovering protein patterns that distinguish disease and disease-free states with high sensitivity and specificity. Laser microdissection, a method for selection of homogenous cell populations, coupled to 2D-DIGE or MudPIT constitute a new proteomics-based paradigm for detecting disease in pathology specimens and monitoring disease response to therapy. This review describes proteomics technologies, and their application in the proteomic analysis of breast carcinoma.

Keywords: Reviews; Proteomics

Contents

* Corresponding author. Tel.: +1 814 241 4715; fax: +1 814 269 9289.
E-mail address: richard@itsi.us (R.I. Somiari).

1570-0232/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2004.11.012

1. Introduction

Breast cancer still imposes significant healthcare burden on women worldwide. For example, in the United States, women have a one in eight lifetime risk of developing breast cancer, and it is estimated that over 200,000 women will be diagnosed with breast cancer in 2004. The standard clinical and pathological approaches to breast cancer staging are the AJCC criteria of tumor size, axillary lymph node status, and presence or absence of distant metastases. Other validated predictive factors include the assessment of estrogen-receptor (ER), progesterone-receptor (PR), and human epidermal growth factor receptor 2 [1,2]. Wider adoption of mammographic screening has increased the number of breast lesions detected but many breast cancers are still not detected early enough, and patients with nodal involvement or metastatic disease often display significantly different clinical phenotypes and responses to therapy [3]. This underscores the need to identify new biological markers and adopt new strategies for detection and management of breast cancer.

The most commonly reported and evaluated breast cancer biomarkers are either (i) associated with hormone reception, cell cycle regulation, extracellular matrix modification, cell proliferation, tumor suppression, (ii) are oncogenes or proto-oncogenes, (iii) are linked to familial/hereditary breast cancer, or (iv) are involved in critical cancer-associated biological mechanisms, including involvement in pathways and/or co-regulated by the same or closely related proteins [4]. Since the molecular character of each tumor is different, accurate and objective classification of the tumor can only be carried out after screening the breast tumor with multiple markers. But choosing the right combination of markers that will be sufficiently sensitive and selective for all cases is not straightforward. Several biomarkers that have been used for breast cancer diagnosis, prognostication, prediction of response to therapy and overall survival have been described [5]. Amongst the reported markers, the hormonal receptors, particularly, estrogen (ER) and progesterone (PR) are the most widely used, particularly, to identify patients that will benefit from hormonal therapy. A number of the reported biological markers are used to identify metastatic disease or identify patients at high risk for disease progression or recurrence [5]. Unfortunately, no single marker with sufficient "predictive power" across all categories of patients has been reported [6] and our experience supports the general notion that breast cancer prognostication and management is significantly improved by the use of multiple biomarkers [5] that will most likely be identified by proteomic analysis of breast carcinoma. The central dogma of molecular biology is still based on the transcription of genomic DNA into mRNA and translation of mRNA into proteins, the functional unit of genes [7]. But, while DNA and RNA analysis through techniques such as cDNA microarrays, comparative genomic hybridization, loss of heterozygosity (LOH) and single nucleotide polymorphism (SNP) analysis are important in identifying genetic abnormalities and uncovering the molecular dysfunctions existing in tumor cells, the presence of single nucleotide polymorphisms, changes in DNA copy numbers or altered levels of RNA may have little or no effect on the events actually happening at the protein level, or directly be relevant to the biological nature of the disease and response to therapy. The requirement therefore exists for comprehensive protein expression profiling using modern proteomics technologies, especially because cancer is increasingly being recognized as a proteomic disease.

It is now widely recognized that proteomics has the potential to revolutionize disease diagnosis and management [8]. The greatest expectations from proteomics come from pharmaceutical research for new protein targets and validation of detected targets [9], whereas clinical researchers hope that proteomics will facilitate the identification of diagnostic, prognostic and predictive biological markers. The human genome sequence has been completely determined and thousands of genes identified or predicted [10,11]. Although 62 genes or more are possibly associated with the onset, progression and/or severity of breast cancer [5], the specific roles played by the majority of these genes are yet to be clearly elucidated at the protein level, and only a small number have been clinically validated or associated with clinical phenotypes. New developments in functional genomics and proteomics will enable high throughput parallel analysis of thousands of genes in individual patients and amongst populations, and open up the possibility of providing more details at the global level on the molecular mechanisms associated with breast carcinoma. The new sub-discipline of proteomics termed "clinical proteomics", that emphasizes the application of proteomics technologies at the bedside [12] holds significant promise for early detection and management of breast cancer patients.

2. Proteomics technologies

The first requirement of proteomic analysis is the separation of the complex mixtures containing as many as several thousand proteins [13]. A complex protein mixture must be resolved into individual proteins or manageable group of proteins, or digested into peptides before identification by mass spectrometry (Fig. 1). A survey of the proteomics literature indicates that several strategies and enabling technologies are in use for proteomic analysis of normal and diseased specimens and for identification of cancer-specific protein markers [4,14–22]. While many of the reported strategies and techniques hold promise, only a few have been rigorously tested, widely used and/or proven to be reasonably effective for comprehensive proteomics of breast carcinoma. Amongst the well-known large-scale biology methods that have been adopted for proteomics, polyacrylamide gel electrophoresis (PAGE), Isotope Coded Affinity Tags (ICAT) technique, Multidimensional Protein Identification Technology (MudPIT), protein array technology and surface enhanced laser-desorption ionisation-time of flight

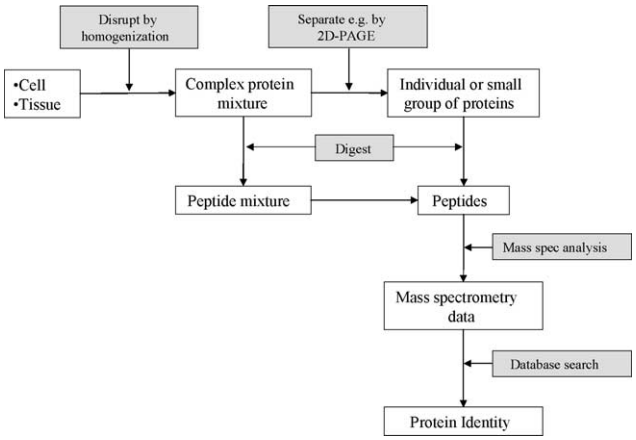

Fig. 1. Generalized schematic representation of the workflow process for proteomic analysis of biological samples.

(SELDI-TOF) technologies are the proteomics strategies that have been most widely used for cancer research. An emerging tool with increasing application in proteomics research is the laser assisted microdissection technology. Laser capture microdissection technology permits selection of a homogenous tumor population from a field of normal appearing cells and vice versa, to improve the accuracy of comparative proteomics studies.

2.1. Polyacrylamide gel electrophoresis (PAGE)

Out of all the protein separation methods in literature, polyacrylamide gel electrophoresis (PAGE) is the most widely used technique for separation of complex protein mixtures prior to protein identification by mass spectrometry. There are two main versions of PAGE, viz. one-dimensional gel electrophoresis (1D-PAGE) and two-dimensional gel electrophoresis (2D-PAGE).

2.1.1. One dimensional gel electrophoresis (1D-PAGE)

1D-PAGE is the single most widely used and least expensive analytical protein separation technique that is reasonably useful for proteomic analysis. In 1D-PAGE, the protein sample is dissolved in a loading buffer that usually contains a reductant (dithiothreitol, or mercaptoethanol) and SDS [23]. Separation, in the presence of an electrical field is mainly based on molecular weight. Separated proteins are visualized by staining the gel with colloidal dyes such as Coomassie brilliant blue or silver stain and excised from the gel for identification by mass spectrometry. The degree of protein resolution by 1D-PAGE is relatively low and a single protein band from a 1D-gel may contain several hundred proteins. Thus, although 1D-PAGE can help reduce protein complexity when dealing with biological specimens like cerebrospinal fluid and lung-lining fluid that contain limited amounts of proteins, it is of little utility in proteomic analysis of biopsies and tumor specimens containing complex protein mixtures because of the limited separation capacity [24].

2.1.2. Two-dimensional gel electrophoresis (2D-PAGE)

Two-dimensional gel electrophoresis (2D-PAGE) is the "benchmark" for large-scale separation of complex protein mixtures and is currently regarded as the most practical and useful separation method for proteomics. Amongst all the proteomics methods currently available, 2D-PAGE is the most reported in breast cancer literature as the technique used for proteomic characterization of breast cancer, probably because it offers the best resolution and is amenable to automation. Varieties of the 2D-PAGE have been used for proteomics analysis of human breast ductal carcinoma in situ [25], infiltrating ductal carcinoma of the breast [4] and for comparison of the protein profiles from fibroadenomas to those from IDCA [26]. In 2D-PAGE, proteins are separated according to two independent physico-chemical parameters. (i) Proteins are first separated by iso-electric focusing (IEF) based on the iso electric point of proteins and then (ii) by PAGE based on the molecular weight of the protein [9,13,23]. Due to new technological developments, a standard spectrum of 2D-PAGE reagents, high resolution pre-cast gels, immobilized pH gradient strips, multiple detection and identification techniques, integrated imaging and bioinformatics tools are now available for proteomic characterization of biological specimens [9,27]. 2D-PAGE and mass spectrometry has been used to study protein expression in breast biopsies [25] and for proteomic phenotyping of metastatic and invasive breast cancer [28].

The recently introduced variety of 2D-PAGE, termed "two dimensional difference gel electrophoresis" [29] has significantly improved the speed, reproducibility and sensitivity of 2D-PAGE based proteomics [4,17,30,31] and permitted the use of this versatile technology in a high throughput research environment [32]. The 2D-DIGE concept involves the covalent labeling of protein extracts with different fluorescent dyes, e.g. cyanine (Cy2, Cy3, or Cy5) dyes [4,17,30,31,33,34] or Alexa dyes [35]. Two approaches of DIGE, termed "minimal" and "saturation" labeling procedures have been described [32]. The earliest and by far the most commonly used DIGE application is the "minimal labeling" procedure, in which the protein-to-dye ratio is deliberately kept high (>95%) so that only the proteins containing a single dye molecule are visualized on the gel [33].

Typically, the "test" protein sample is labeled with Cy3 and the reference sample is labeled with Cy5 [4]. Equal concentrations of the differentially labeled protein samples are mixed and co-separated during the same 2D-PAGE process. The 2D-DIGE gel pattern is then visualized by scanning the gel at two wavelengths using a fluorescence imager (Fig. 2). Since the charge and mass of the fluorescence dyes used in 2D-DIGE are carefully matched, there is minimal dye-induced shift of proteins during 2D-PAGE [4,33]. A comparison of the images generated by scanning of the 2D-DIGE gel at the Cy3 and Cy5 wavelengths allows the quantitation of each spot with the accompanying image analysis software, e.g. the 2D-DIGE DeCyder Software [4,15,30], and

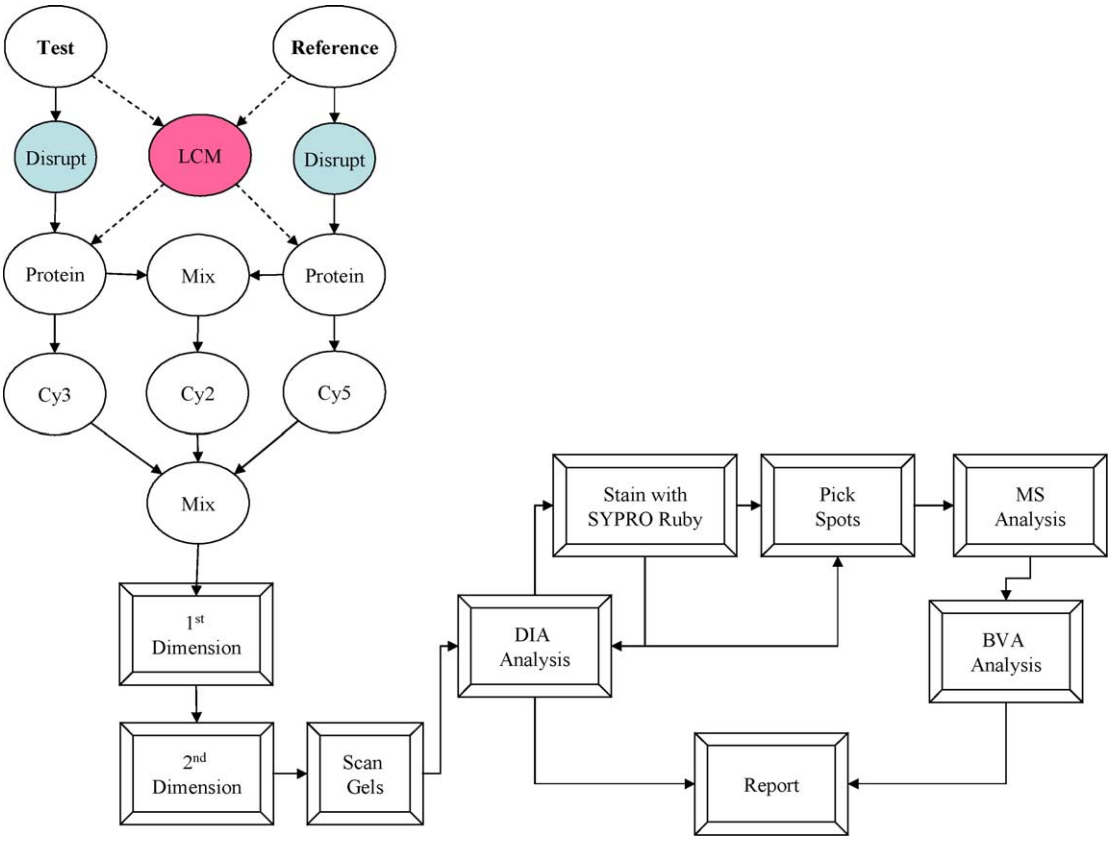

Fig. 2. Generalized workflow process for the GE Healthcare 2D-DIGE technology. Test, reference and a 1:1 mixture of the two samples are labeled with cyanine (Cy2, Cy3 and Cy5), mixed and separated by 2D-PAGE (iso-electric focusing, first dimension; and polyacrylamide gel electrophoresis, second dimension). The gel generated is scanned at three wavelengths and the image generated analyzed with the accompanying difference in gel analysis (DIA) and biological variation analysis (BVA) software's available in DeCyder (GE Healthcare). Candidate protein spots are picked from Sypro Ruby stained gels and identified by mass spectrometry (MS analysis). Incorporation of laser microdissection (LCM) improves the accuracy of comparative and quantitative proteomics.

for making multiple gel comparisons. 2D-DIGE effectively eliminates or drastically reduces gel-to-gel variability associated with standard 2D-PAGE and improves the accuracy of quantitative protein profiling. Determination of protein expression differences between a "test" and a "reference" sample is fast and accurate because it is based on the relative fluorescence intensities captured from a single 2D-DIGE gel [4,30].

The "saturation labeling" procedure is specifically developed for the analysis of scarce protein samples, e.g. samples obtained by laser-microdissection [36]. The "saturation" labeling dyes (CyDye DIGE Fluor Cy3 and CyDye DIGE Fluor Cy5) are different from the "minimal labeling" dyes because of the presence of maleimide reactive groups that covalently bind to the cysteine residues on proteins via a thioether linkage [32]. In the "saturation" labeling protocol, the aim is to label all available cysteine residues thereby increasing the fluorescence signal generated by labeled protein samples and hence the total number of proteins detectable on a gel (Fig. 3). The minimal and saturation labeling protocols have been successfully used to study protein isolated from cells captured from breast tissue by laser microdissection [32,34,36].

2.2. Isotope coded affinity tags (ICAT) technology

Isotope-coded affinity tagging (ICAT) originally developed in the laboratory of Dr. Reudi Aebersold is an emerging protein profiling technology that utilizes stable isotope labeling to perform quantitative analysis of paired protein samples (Fig. 4). Protein samples or peptides are isotope labeled, separated by high performance liquid chromatography and identified by mass spectrometry [37]. While other protein profiling techniques like 2D-DIGE comparatively profile the naturally occurring forms of peptides and proteins, ICAT analysis profiles the relative amounts of cysteine-containing peptides derived from tryptic digests of protein extracts. The isotope tags, e.g. ^{12}C (light) and ^{13}C (heavy) bind covalently to cysteine moieties of amino acids within proteins [38,39]. Like the cyanine dyes used in 2D-DIGE, the isotopic tags used are similar in structure and chemical properties, but are different in mass. In one of the most popular versions of ICAT, the light tag contains eight hydrogen atoms whereas the heavy tag contains eight deuterium isotopic forms. There will therefore be a mass difference of exactly eight mass units between similar proteins labeled with the light and heavy tags [37,39]. Available literature indicate that the ICAT technology can be

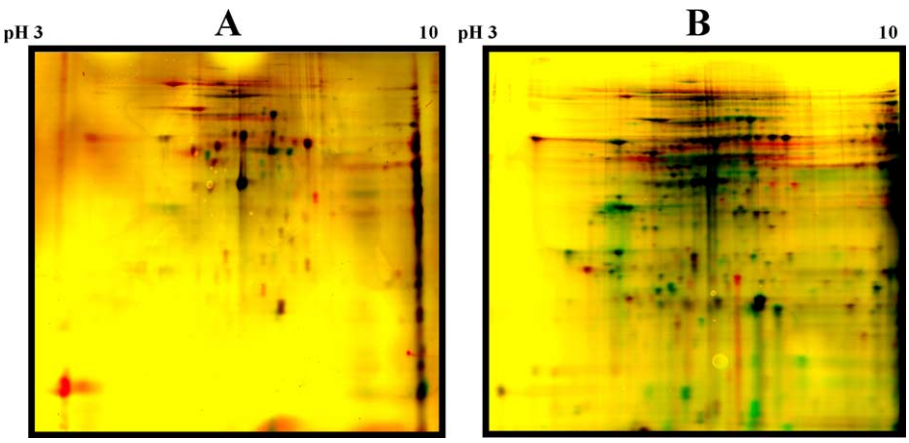

Fig. 3. 2D-DIGE gels showing the fluorescent signal obtained from of 100 μg breast tissue proteins and analyzed using the "minimal" (A) or "saturated" (B) 2D-DIGE protocols. Since more proteins are detectable with the "saturated" labeling dyes, this is the recommended approach for analysis of samples such as ductal lavage where low amounts of protein are expected.

used to routinely identify 300–400 proteins per sample, a number that is far less than what is typically achieved with the 2D-PAGE technology.

The ICAT technology has been used successfully to study protein expression in mammalian [40] and liver cells [41] and for quantitative expression proteomics on limited breast tumor cells obtained by laser microdissection [42]. In the later study, a total of 76 proteins were identified and some of the proteins for example, mitochondrial isocitrate dehydrogenase, actin and 14-3-3 protein xi/delta were found to be significantly upregulated in breast tumor cells [42]. Even though the number of proteins identified in a typical ICAT run are far less than those reported for 2D-PAGE, the high throughput, quantitative nature and reproducibility achievable makes ICAT one of the most powerful emerging proteomic technologies that will increasingly be used for high

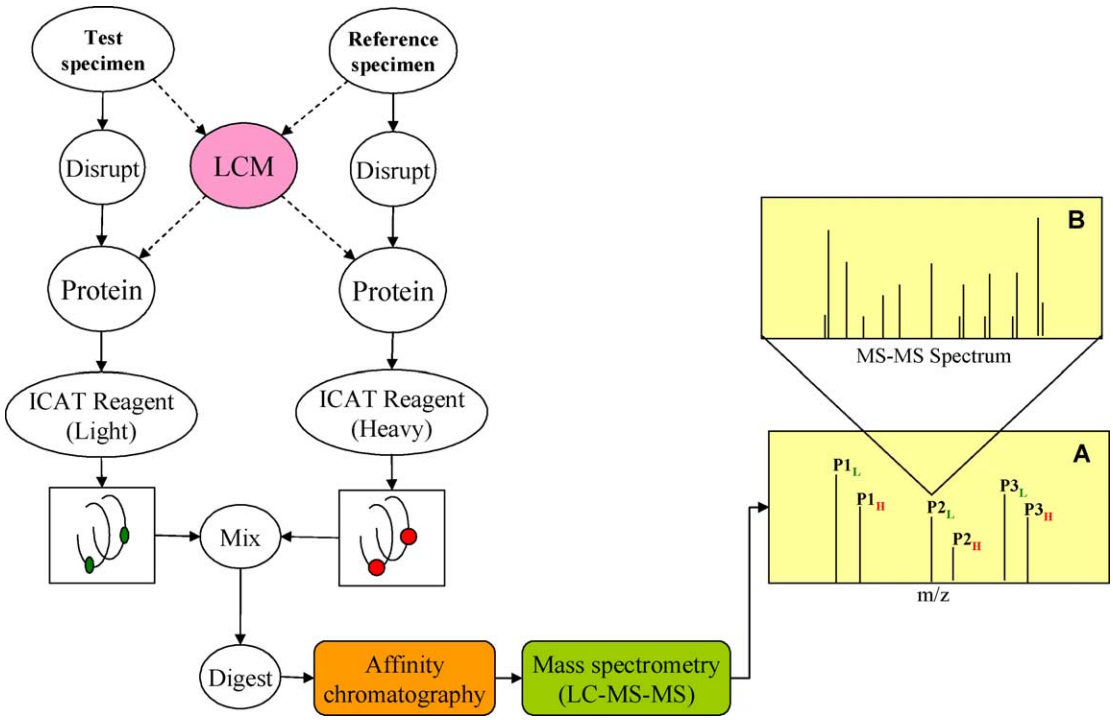

Fig. 4. Flow chat showing the ICAT/MS based proteomic profiling of biological samples. Test and reference samples are labeled with light (L) and heavy (H) ICAT reagents, mixed and digested into peptides using trypsin. The reagents label one or more cysteinyl thiols on the proteins. The mixture is applied to a column of avidin beads to isolate the ICAT-tagged peptides by affinity chromatography before LC–MS–MS by data dependent scanning. The full scan mass spectrum (A) reveals the concentration of the L-labeled and H-labeled proteins in the two samples and the MS–MS spectrum (B) reveals the protein from which the peptide originated. P1–P3 are hypothetical proteins in the mixture analyzed. Incorporation of laser microdissection (LCM) improves the accuracy of comparative and quantitative proteomics of pathology samples.

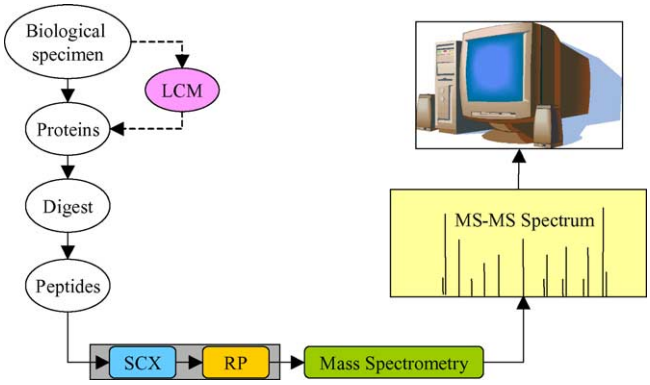

Fig. 5. Schematic representation of the process of multidimensional protein identification technology (MudPIT). Denatured protein complex isolated from biological samples are digested into peptides and subjected to two-dimensional liquid chromatography on strong cation exchange (SCX) column coupled to a reverse-phase (RP) column. Peptides are fragmented by tandem mass spectrometry and proteins in the complex are identified by computational translation of tandem mass spectra to amino acid sequence using genomic sequences.

throughput proteomic characterization of biological specimens. Recently, ICAT was described as an excellent tool for degradomic research, particularly, for discovery of proteases and for exploring proteolytic function in complex and dynamic biological context [43].

2.3. Multidimensional protein identification technology

Multidimensional protein identification technology (MudPIT) is a novel proteomic technique in which two liquid chromatographic steps are interfaced back-to-back in a fused silica capillary to permit two-dimensional high performance liquid chromatography [44]. MudPIT incorporates high-pressure liquid chromatography (HPLC, LC/LC), tandem mass spectrometry (MS/MS) and database-searching algorithms to rapidly analyze complex protein mixtures [45]. Specifically, a column containing a strong cation exchange (SCX) material is coupled to a column containing reversed phase (RP) materials and then to a tandem mass spectrometer. Typically the complex peptide mixture generated from protein lysates are loaded onto the biphasic column for simultaneous separation and analysis by mass spectrometry (Fig. 5).

The biphasic column which is placed in-line with the HPLC system also acts as an ion source for the tandem mass spectrometer thereby reducing dead-volumes and band broadening. Chromatography proceeds in steps with increasing salt concentration to selectively release proteins from the SCX resin unto the reversed phase resin. A reversed phase reagent with increasing hydrophobicity is then slowly introduced to progressively elute peptides from the RP resin into the mass spectrometer where they are ionized and identified based on their time-of-flight. The mass spectrometer data-dependent acquisition isolates peptides as they elute and subjects them to collision induced dissociation and records the fragment ions in a tandem mass spectrum. To identify the peptide and hence the protein, the mass spectra generated are matched to database entries of peptide sequences, by using, e.g. the SEQUEST algorithm to interpret the MS/MS data and identify the peptide sequence from which it was generated.

Although very little information currently exist on the application of MudPIT in cancer research, this technology MudPIT has been used to study cytokinesis proteins [46], consensus mammalian mediator subunits [47], and for concurrent proteomic analysis of both membrane and soluble proteins [48] demonstrating the potential power and versatility of this emerging high throughput proteomic technology. Quantitative MudPIT has also been found to be both intra- and inter-experimentally reproducible at the peptide and protein levels and suitable for identification of low abundant proteins [49]. MudPIT is now widely regarded as the proteomic strategy that may help alleviate some of the problems associated with 2D-PAGE and ICAT based proteomics technologies, because it permits a rapid and simultaneous separation and identification of proteins and peptides in a complex mixture without the need for pre- or post-separation labeling.

2.4. Surface enhanced laser desorption ionization-time of flight technology

Surface enhanced laser desorption ionization-time of flight (SELDI-TOF) is a relatively novel and straightforward proteomic technology that can be used for quantitative analysis of protein mixtures after selectively capturing proteins with unique attributes on activated surfaces. This technology utilizes specialized stainless steel or aluminum-based supports coated with chemical or biological baits to selectively capture proteins based on the intrinsic properties of the proteins. The chemical surfaces that have been described either have hydrophobic, hydrophilic, anionic or cationic properties whereas the biological surfaces can be antibodies, antigen binding fragments, DNA or receptors.

SELDI-TOF is a potentially powerful clinical proteomics tool for identification of patients at risk for development of cancer based on the direct analysis of body fluids like serum, plasma, ductal lavage, cerebro spinal fluid and urine. A high profile and well publicized ovarian cancer study utilized SELDI-TOF to identify protein peaks in serum that distinguished patients with ovarian cancer from those without ovarian cancer [17]. While the study design, sensitivity and specificity reported have generated counter comments, this ovarian cancer study is still a “landmark” study because it demonstrated, for the first time, the potential of utilizing a proteomic approach, specifically the serum proteomic signature, for stratifying patients in the clinic. It has also been reported that the SELDI-TOF is being applied to a range of pathological stages of ovarian cancer [8]. Interestingly, the SELDI-TOF technology has also been recently applied to breast cancer research [50]. In this study, Pusztai and co workers used the SELDI-TOF technique to examine proteomic changes that occur in response to paclitaxel chemotherapy or 5-fluorouracil, doxorubicin and cyclophos-

phamide chemotherapy in plasma of 69 patients with stage III breast carcinoma and 15 normal volunteers. The authors report the identification of a single chemotherapy inducible SELDI-MS peak and five additional peaks that appear to distinguish plasma obtained from patients with breast carcinoma from the plasma obtained from normal healthy volunteers. Considering that the ovarian [17] and breast cancer [50] studies dealt with different cancers and patient cohorts, it is clear that in spite of the criticism and skepticism, the SELDI-TOF is an emerging and potentially powerful proteomic tool that has attributes, e.g. cost and ease of use, that is lacking in other proteomics technologies. In our opinion, the new generation SELDI-TOF's incorporating improved/novel ion sources and new peak identification algorithms, are amongst the reasonably sensitive and affordable proteomic tools currently available that can be used in the clinic for rapid screening of patient serum or plasma to identify protein patterns and signatures associated with breast cancer.

2.5. Protein array technology

Protein arrays are one of the latest proteomic technologies that hold significant promise for molecular and biochemical pathway elucidation via the mapping of all the expressed proteins. This technology, which emulates the DNA microarray technology, can be used to directly measure the levels of proteins in tissues using fluorescence-based imaging techniques [51]. The proteins can be arrayed on solid surfaces, capillary systems or immobilized on beads. The most popular protein arrays are constructed with antibodies or cytokines. Protein array technology is attracting a lot of attention because of the potential of analyzing the levels of hundreds of proteins within a pathway of interest [52,53]. While a number of reports have demonstrated the utility of this technology, the wider application of protein arrays in biomedical research is still limited, partly because of the cost of producing antibodies and the limited availability of antibodies with high specificity and high affinity for the target. Additionally, the difficulty associated with preserving proteins in their biologically active conformation before analysis with protein arrays will further limit the application of this technology as a routine proteomic strategy. Information gathered from recent proteomics meetings indicate that many biotechnology companies are at advanced stages of developing new generation protein arrays using new technologies. One of such developments is the planed production of protein bioarrays using the highly successful CodeLink© (GE-Healthcare) proprietary oligonucleotide microarray slides, that will maintain the spotted antibodies on a three dimensional matrix, thereby increasing the surface area for antibody–antigen interactions.

3. Proteomic applications

Proteomics, defined as the study of the expressed part of the genome, involves the comprehensive display, quantitation and identification of the proteins expressed by the genome of an organism, as a function of normal development, aging, disease or environment. The ultimate goal of proteomics is to characterize protein pathways, networks and signaling events that are relevant in disease. It is now widely recognized that comparative proteomics will play an important role in providing new insights into cancer development and progression as well as in the identification and validation of new protein targets for diagnostic and therapeutic purposes [9,16,54,55].

Unlike the genome, the proteome is at a constant flux due to development, aging, signaling, disease and environmental insults. Thus, the cellular or tissue proteome of most organisms is significantly more challenging to map, when compared to the genome [4]. The current basic application of proteomics can be grouped under four general categories, namely (i) global protein mining, (ii) protein expression profiling, (iii) protein network mapping and (iv) posttranslational modification characterization [23]. Global protein mining is the ultimate brute-force in proteomics and the most challenging, demanding and expensive proteomics application. This proteomic application is particularly important and informative because it could allow the direct analysis and identification of proteins present in a breast cancer specimen rather than inferring the composition of the proteome by transcriptomics (transcriptome profiling). In protein expression profiling, also known as "expression proteomics", the aim is to identify proteins that are characteristic of a clearly defined state such as stage 1 breast cancer or that change as a function of perturbations, e.g. exposure to chemotherapy. Protein expression profiling is typically practiced as a differential analysis by comparing proteins expressed in the "reference" sample, e.g. normal breast tissue to proteins expressed in the "test" sample, e.g. breast tumor biopsy or breast cancer cell lines exposed to cancer drugs. Protein network mapping and post translational modification characterization are highly specialized proteomic applications involving (i) the determination of the condition and the manner in which proteins interact with each other in their natural environment and (ii) the determination of how, the extent and where proteins are modified, respectively. Both applications will increasingly become important in breast cancer research because of the need for a comprehensive characterization of candidate protein targets prior to design and development of antibody based drugs.

The application of a relatively new and sub-discipline of proteomics termed "clinical proteomics" in early detection and diagnosis of cancer has been reported [12]. Clinical proteomics emphasizes the application of proteomic technologies at the bedside, to acquire "serum-based proteomic patterns" characteristic of the blood-proteome of a normal and diseased state. This proteomic approach has been applied successfully to ovarian cancer research, at which, the diagnostic end-point for the detection of ovarian cancer was a proteomic pattern that comprised many individual proteins, none of which could be used in isolation to differentiate between "normal" individuals and ovarian cancer patients. This potentially powerful direct bedside application of proteomics

could affect cancer detection in particular and clinical practice in general because of its potential effect on detection, diagnosis, selection of the line of therapy and assessment of the effect of treatment.

4. Protein expression profiling of breast carcinoma

Most proteomics experiments are aimed at identification of proteins that are differentially expressed in normal versus diseased specimens. Majority of the proteomics studies in which breast cancer has been used as a model have utilized breast cancer cell lines and core biopsies, and the focus has been the identification of differentially expressed proteins as a way of (i) defining the molecular and biochemical pathways by which normal cells progress to cancer and/or (ii) uncovering biological markers and therapeutic targets for cancer. The successful performance of proteomics and accurate measurement of altered protein expression depends on the availability of good quality specimens and ability to obtain cell populations enriched for nonmalignant or malignant breast cells. Large scale proteomics of breast carcinoma is now possible because of the establishment of biorepositories dedicated to the collection and banking of normal and diseased tissue suitable for integrated high throughput genomics and proteomics research [56] and availability of reproducible proteomics technologies.

A major obstacle to accurate protein expression profiling is the degree of tissue heterogeneity of breast carcinomas, a phenomenon that can affect the result obtained from comparative proteomics experiments. While several methods have been used to select homogenous cell populations prior to proteomic analysis [57], laser capture microdissection (LCM, PixCell II, Arcturus) or laser microdissection (LMD, *ASLMD*, Leica Microsystems) is the most appropriate for proteomic analysis of breast biopsies and solid tissue. Laser microdissection is extremely important in quantitative and comparative proteomics where tissue heterogeneity may skew the results obtained. The LCM procedure involves the placement of a transparent plastic cap (CapSure^{TM}) containing a thermoplastic membrane over a section of tissue mounted on a histological glass slide, visualizing the tissue section under the microscope and selectively capturing cells unto the membrane on the cap by applying short focused pulses from an infra red laser [58]. The *ASLMD* platform uses laser ablation to cut out selected sections of tissue mounted on a membrane, using UV laser. With the *ASLMD*, the cut section drops by gravity into the cap of a capture tube located beneath the stage [59].

In proteomic experiments, laser microdissection permits the dissection of cancerous tissue and selection of subpopulation of cells, e.g. tumor cells from a field of normal appearing cells. Using LCM or LMD researchers can obtain more accurate representation of cells (tumor and normal appearing cells) and make more accurate comparisons of protein expression in normal and diseased specimens. Laser microdissection has been coupled to 2D-PAGE [25,57,60] and 2D-DIGE [36] for the proteomics characterization of cancerous tissues. Approximately 50,000 cells have been found sufficient for 2D-PAGE separation and visualization [60,25] and our laboratory has successfully used ∼30,000 cells for comparative proteomics by 2D-DIGE [36].

The revolutionary "clinical proteomics" approach also termed "proteomic pattern analysis" [17] could also be applicable for breast cancer, and is currently being applied to a subset of serum samples ($n = \sim 500$) collected by the authors from donors with, and without, breast cancer, as part of the Clinical Breast Care Project (CBCP), biomarker discovery program. Clinical proteomics has shown great promise for early diagnosis of ovarian cancer and is considered suitable for identifying distinguishing protein patterns in serum, plasma, ductal lavage, cerebro spinal fluid or urine of breast cancer patients. Proteomic pattern analysis relies on the pattern of proteins observed and does not rely on the identification of a traceable biological marker [61]. While there are reports on the proteomic analysis of other breast malignancies, the two most studied forms of breast cancer: ductal carcinoma in situ (DCIS) and infiltrating ductal carcinoma (IDCA) are discussed in more detail here.

4.1. Proteomics of ductal carcinoma in situ of the breast

Ductal carcinoma in situ (DCIS) is a heterogeneous disease characterized by noninvasive clonal proliferation of malignant epithelial cells arising from the mammary ducts and terminal ductal units [62]. DCIS represents the earliest detectable cancerous lesion in breast and its incidence is rising because of increased mammographic screening. A diagnosis of DCIS increases the risk of developing IDCA by 8–10-fold [37,63] and it is estimated that 25–50% of DCIS lesions progress to IDCA if left untreated [62,64,65]. DCIS can be treated by total mastectomy, local excision plus adjuvant therapy or local excision alone [66]. A number of high throughput technologies including DNA microarrays and SAGE offer the potential to discover previously unknown alterations in gene expression and identification of biological markers. However, proteomics of breast DCIS has revealed protein expression and modification trends that are distinct from results obtained by nucleic-acid based methods [25].

The proteomic analysis of DCIS has revealed 57 proteins that show differential expression between normal cells and DCIS [25]. Although the differential expression was predominantly due to differences in overall abundance, there was also evidence of posttranslational modification indicating that proteomics is capable of uncovering qualitative and quantitative differences that exist between normal and diseased breast tissue. Proteins such as transgelin and the voltage-dependent anion channel protein (VDAC) showed evidence of posttranslational modification [25]. A subset of proteins including, Annexin V, profiling and HSP 90 showing differential expression on 2D gels have been confirmed by IHC, a further evidence that current proteomic strategies have sufficient sensitivity to detect clinically relevant changes and

therefore can be successfully used to explore protein expression trends and discover novel clinically relevant protein expression portraits previously unconnected to breast cancer. Many of the differentially expressed proteins identified by Wulfkuhle et al. [25] have not been previously associated with breast cancer, underscoring the importance of proteomics as a powerful hypothesis-generating tool in breast cancer research.

4.2. Proteomics of infiltrating ductal carcinoma of the breast

Infiltrating ductal carcinoma (IDCA) of the breast is the most common and potentially aggressive form of breast cancer. In our laboratory, we have looked at protein expression differences in normal and diseased specimens and also compared protein expression changes between primary tumors and lymph nodes. It is believed that such studies will uncover novel biological markers and provide the basis for the development of new methods to detect and treat breast cancer. The proteomic analysis of infiltrating ductal carcinoma (IDCA) of the breast by 2D-DIGE reveals a number of proteins that appear to be differentially expressed between breast IDCA and matching normal tissue [4]. By comparing the test sample (IDCA) to a common reference (normal tissue) it was possible to objectively compare protein expression differences between IDCA samples using the difference in gel analysis (DIA) module of DeCyder^{TM} software [4]. Based on a threshold mode of $\geq$3.0, protein expression differences that ranged from 15.5% in stage-I IDCA to 30.6% in the stage-IIB IDCA were detected [4]. Approximately, 69–85% of the proteins detected were not significantly different between the normal and diseased samples. While 744 candidate protein spots displayed $\geq$3.0 differences in levels between breast IDCA's and the normal breast tissue, the functional role or how majority of these proteins mediate the tumor process are largely unknown. Examples of the proteins detected as differentially expressed were carbonic dehydratase, disulfide isomerase, gelsolin and fibrinogen beta. The protein expression profile and observed trend for a number of the proteins is consistent with information in literature for some of the proteins. For example, fibrinogen gamma-chain and fibrinogen beta-chain fragments have been identified in various solid tumor types at the protein level, and fibrinogen gamma-chain dimmer crosslinked by transglutaminase were detected in plasma from tumor patients but not in plasma from controls [67]. It is suggested that the elevation of β-fibrinogen correlates with tumor-associated fibrin deposition [67]. Carbonic dehydratase, disulfide isomerase, gelsolin and fibrinogen beta are overexpressed in IDCA whereas gelsolin, which is known to bind to β-actin, is less abundant in IDCA compared to normal tissue.

In a separate study in our laboratory, 12 breast tumors and the associated lymph nodes were analyzed by 2D-DIGE with the aim of providing insights into the global pattern of protein expression in primary tumors compared to the associated lymph nodes and adjacent normal appearing tissue. Secondary aims of the study were to determine if the differentially expressed proteins have been reported to play any role in breast cancer development, progression and/or severity and if they are associated with known tumor activation or suppression mechanisms. The breast tumors studied were infiltrating ductal or infiltrating lobular carcinomas and the lymph nodes were either positive or negative. We detected qualitative and quantitative differences in protein expression between tumors (IDCA and ILCA) and the associated lymph nodes with an overall differential protein expression ($\geq$3-fold) ranging from 6.2 to 17% between normal and diseased specimens. Example of the proteins that showed $\geq$3-fold difference in expression between breast carcinomas (IDCA and ILCA) and associated lymph nodes are ubiquitin-activating enzyme E1, transferrin, annexin VI, L-plastin, Glutathione S-transferase A1, and protein disulfide isomerase.

While molecular alterations accompanying DCIS is the earliest detectable form of breast cancer, the changes associated with, or that may trigger the transformation of a DCIS to IDCA are largely unknown. A comparison of the proteins identified as differentially expressed between breast DCIS and normal breast tissue [16] and between breast IDCA and normal breast tissue [4] shows some overlap. For example, L-plastin, Annexin V, and the 78 kDa glucose regulated protein (GRP78) were identified as differentially expressed between breast carcinoma (DCIS and IDCA) and normal breast tissue in both studies. L-Plastin is associated with actin binding, Annexin V is associated with membrane trafficking and linked to cytoskeletal inhibition of PKC and pLAC whereas GRP78 is active in the endoplasmic reticulum and believed to protect secretary proteins, bind peptides and is associated with the MHC [16]. The overlap in proteins differentially expressed in DCIS and IDCA is noteworthy and suggests some relationship in the mechanisms triggering and/or driving both breast carcinomas.

5. Conclusion

Available breast cancer literature reveals the potential power of proteomics as a tool for characterization of molecular dysfunctions associated with breast carcinoma and for uncovering targets for therapeutic interventions. Although new strategies are needed to improve resolution and sensitivity, the technologies currently available have sufficient resolving power (e.g. 2D-PAGE), sensitivity (e.g. 2D-DIGE, ICAT), speed (e.g. MudPIT) and versatility (SELDI-TOF) to allow scientists to begin to perform mining and protein expression analysis of breast carcinoma. Because of the complexity of the proteome and limitations associated with the available proteomic technologies, comprehensive and accurate proteomic analysis of breast carcinoma will require the use of two or more proteomic technologies in tandem, e.g. 2D-PAGE and MuDPIT or 2D-DIGE and ICAT, and incorpo-

ration of noncore proteomic technologies like LCM, to study each sample in detail.

In spite of the current limitations, proteomics is currently the most promising approach that can be used at the global level to reveal (i) relevant tumor associated biological networks, (ii) molecular relationships between different breast carcinomas and (iii) the molecular mechanisms that drive the progression of breast carcinomas, e.g. DCIS to IDCA. In conclusion, the emerging recognition that cancer is a "proteomic disease", and the recent demonstration that ovarian cancer can be detected via a "clinical proteomics" approach has the potential to increase the significance, and application of proteomics in breast cancer research, diagnosis, prognostication, assessment of therapeutic efficacy and/or toxicity.

Acknowledgements

The authors acknowledge the funding received from the United States Department of defence for the Clinical Breast Care Project of Walter Reed Army Medical Center, Washington, DC, USA, Windber Research Institute, Windber, PA, USA and Windber Medical Center, Windber, PA, USA. The support received from the Henry Jackson Foundation for the Advancement of Military Medicine is also acknowledged.

References

[1] M. Dowsett, T. Cooke, I. Ellis, et al., Eur. J. Cancer 36 (2) (2000) 170.
[2] R.M. Merrill, D.L. Weed, Ann. Epidemiol. 11 (8) (2001) 547.
[3] E.T. Liu, Curr. Opin. Genet. Dev. 13 (1) (2003) 97.
[4] R.I. Somiari, A. Sullivan, S. Russell, et al., Proteomics 10 (2) (2003) 1863.
[5] C. Arciero, S. Somiari, C. Shriver, et al., Int. J. Biol. Markers 18 (4) (2003) 241.
[6] S Alberti, Int. J. Biol. Markers 18 (2003) 35.
[7] D. Thieffry, S. Sarkar, Trends Biochem. Sci. 23 (8) (1998) 312.
[8] V. Espina, K.A. Dettloff, S. Cowherd, et al., Expert Opin. Biol. Ther. 4 (1) (2004) 83.
[9] R. Westermeier, T. Naven, Proteomics in Practice: a Laboratory Manual of Protein Analysis, Weiley-VCH, Verlag-GmbH, Weinheim, 2002, p. 316.
[10] J.C. Venter, M.D. Adams, E.W. Myers, et al., Science 291 (5507) (2001) 1304.
[11] F.S. Collins, M. Morgan, A. Patrinos, Science 5617 (2003) 286.
[12] E.F. Petricoin III, M.A. Ardekani, B.A. Hitt, et al., Lancet 359 (2002) 572.
[13] M.J. Dunn, A. Gorg, in: S.R. Pennington, MJ. Dunn (Eds.), Proteomics: from Protein Sequence to Function, Bios Scientific Publishers, Springer Verlag, New York, 2001, p. 43.
[14] L. Liotta, E. Petricoin, Nat. Rev. Genet. 1 (2000) 48.
[15] G. Zhou, H. Li, D. DeCamp, et al., Mol. Cell. Proteomics 1 (2002) 117.
[16] J.D. Wulfkuhle, K.C. Mclean, C.P. Paweletz, et al., Proteomics 1 (2001) 1205.
[17] E.F. Petricoin, K.C. Zoon, E.C. Kohn, J.C. Barrett, L.A. Liotta, Nat. Rev. Drug Discov. 1 (2002) 683.
[18] E.R. Sauter, W. Zhu, X.-J. Fan, et al., Br. J. Cancer 86 (2002) 1440.
[19] H. Hondermarck, A.-S. Vercoutter-Edouart, F. Revillion, et al., Proteomics 1 (2001) 1216.
[20] S. Kennedy, Toxicol. Letts. 120 (2001) 379.
[21] P.H. O'Farrell, J. Biol. Chem. 250 (1975) 4007.
[22] A. Gorg, W. Postel, S. Gunther, Electrophoresis 9 (1988) 531.
[23] D.C. Liebler, Introduction to Proteomics: Tools for the New Biology, Humana Press, Totowa, NJ, 2002, p. 198.
[24] H. Liu, D. Lin, Yates, J.R. III, Biotechniques 32 (2) (2002) 898, 900, 902.
[25] J.D. Wulfkuhle, D.C. Sgroi, H. Krutzsch, et al., Cancer Res. 62 (2002) 6740.
[26] A. Bisca, C. D'Ambrosio, A. Scaloni, et al., Cancer Lett. 204 (2) (2004) 79.
[27] V.E. Bichsel, L.A. Liotta, E.F. Petricoin III, Cancer J. 7 (2001) 69.
[28] Y. Hathout, M.L. Gehrmann, A. Chertov, et al., Cancer Lett. 210 (2) (2004) 245.
[29] M. Unlu, M.E. Morgan, J.S. Minden, Electrophoresis 18 (1997) 2071.
[30] S. Gharbi, P. Gaffney, A. Yang, et al., Mol. Cell. Proteomics 1 (2002) 91.
[31] S. Steiner, F. Witsmann, Electrophoresis 21 (2000) 2099.
[32] R.I. Somiari, S. Russell, S.B. Somiari, et al., Difference gel electrophoresis in a high throughput environment, in: J.M. Walker (Ed.), The Proteomics Protocol Handbook, Humana Press Inc., NJ, USA, 2005, p. 850.
[33] R. Tonge, J. Shaw, B. Middleton, et al., Proteomics 1 (2001) 377.
[34] A.G. Sullivan, S. Russell, C. Arciero, et al., in: Proceedings of the 50th American Society for Mass Spectrometry Annual Meeting, ASMA A020445, 2002.
[35] F. von Eggeling, A. Gawriljuk, W. Fiedler, et al., Int. J. Mol. Med. 8 (2001) 373.
[36] D.L. Ellsworth, S. Russell, B. Deyarmin, et al., in: J.M. Walker (Ed.), The Proteomics Protocol Handbook, Humana Press Inc., NJ, USA, 2005, p. 850.
[37] S.P. Gygi, B. Rist, S.A. Gerber, et al., Nat. Biotechnol. 17 (1999) 994.
[38] S.D. Paterson, Curr. Opin. Biotechnol. 11 (2000) 413.
[39] D.K. Han, J. Eng, H. Zhou, Nat. Biotechnol. 19 (2001) 946.
[40] Q. Tian, S.B. Stepaniants, M. Mao, et al., Mol. Cell Proteomics 3 (10) (2004) 960.
[41] W. Yan, H. Lee, E.C. Yi, et al., Genome Biol. 5 (2) (2004) 54.
[42] L. Zang, D.P. Toy, W.S. Hancock, et al., J. Proteome Res. 3 (2) (2004) 604.
[43] E.M. Tam, C.J. Morrison, Y.I. Wu, et al., Proc. Natl. Acad. Sci. U.S.A. 101 (2) (2004) 6917.
[44] D.A. Wolters, M.P. Washburn, J.R. Yates III, Anal. Chem. 73 (2) (2001) 5683.
[45] C.C. Wu, M.J. MacCoss, Curr. Opin. Mol. Ther. 4 (3) (2002) 242.
[46] A.R. Skop, H. Liu, J.R. Yates III, Science 305 (2) (2004) 61.
[47] S. Sato, C. Tomomori-sato, T.J. Parmely, et al., Mol. Cell. 14 (2) (2004) 685.
[48] C.C. Wu, M.J. MacCoss, K.E. Howell, et al., Nat. Biotechnol. 21 (2) (2003) 532.
[49] M.P. Washburn, R.R. Ulaszek, J.R. Yates III, Anal. Chem. 75 (19) (2003) 5054.
[50] L. Pustai, B.W. Gregory, K.A. Baggerly, et al., Cancer 100 (2) (2004) 1814.
[51] G. MacBeath, S.L Schreiber, Science 289 (2000) 1760.
[52] D. Stoll, M.F. Templin, M. Schrenk, et al., Frontiers Biosci. 7 (2002) c13.
[53] G. Walter, K. Bussow, A. Lueking, et al., Trends Mol. med. 8 (6) (2002) 250.
[54] S. Steiner, C.L. Gatlin, J.J. Lennon, et al., Electrophoresis 21 (2) (2000) 2129.
[55] J. Cellis, M. Kruhoffer, I. Gromova, et al., FEBS Lett. 480 (2000) 2.

[56] S.B. Somiari, R.I. Somiari, J. Hooke, et al., TIBETS 1 (2004) 131–143 (published online at www.tibetsjournal.org).
[57] A.R. Shekouh, C.C. Thompson, W. Prime, et al., Proteomics 3 (2003) 1988.
[58] M.R. Emmert-Buck, R.F. Bonner, P.D. Smith, et al., Science 274 (1996) 998.
[59] J. Ossi, A. Lee, J. Zhou, TIBETS 1 (2004) 118–123 (published online at www.tibetsjournal.org).
[60] R.A. Craven, N. Totty, P. Harnden, et al., Am. J. Pathol. 160 (2002) 815.
[61] T.P. Conrads, T.D. Veenstra, Curr. Drug Targets Immune Endocr. Metabol. Disord. 4 (2) (2004) 41.
[62] K. Mokbel, Int. J. Fertil. Womens Med. 48 (2) (2003) 217.
[63] B. Futcher, G.L. Latter, P. Monardo, et al., Mol. Cell Biol. 19 (1999) 7357.
[64] A.G. Hinnebusch, in: E.W. Jones, J.R. Pringle, JR. Broach (Eds.), The Molecular and Cellular Biology of the Yeast Sacharomyces: Gene Expression, vol. 2, Cold Spring Habour Press, Plainview, NY, 1992, p. 319.
[65] M.C. Costanzo, M.E. Crawford, J.E. Hirschman, et al., Nucleic Acid Res. 30 (2001) 31.
[66] K. Mokbel, Eur. J. Surg. Oncol. 29 (2) (2003) 191.
[67] C. Gerner, W. Steinkellner, K. Holtzmann, et al., Thromb. Haemost. 85 (2001) 494.