Journal of Chromatography B, 877 (2009) 1222–1228

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Review

# Electrophoresis and liquid chromatography/tandem mass spectrometry in disease biomarker discovery☆

Haleem J. Issaq*, Josip Blonder

*Laboratory of Proteomics and Analytical Technologies, Advanced Technology Program, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD 21702, United States*

ARTICLE INFO

*Article history:*
Received 13 August 2008
Accepted 10 December 2008
Available online 24 December 2008

*Keywords:*
Disease biomarkers
HPLC–MS/MS
Proteomics
Metabolomics

ABSTRACT

The search for disease markers is not new; however, with the emergence of new technologies such as nano-HPLC and electrospray ionization and time of flight mass spectrometry, the search has intensified considerably. Genomic, proteomic and metabolomic technologies are being used to search for novel disease markers. In this manuscript emphasis will be on different HPLC and MS methods that are used to search for metabolites and proteins that can be used for the discovery of novel, sensitive and specific disease biomarkers. Definitions of terms such as sensitivity, specificity, and protein profiles will be given. Methods used for effective fractionation, separation and quantitation of proteins and peptides using HPLC/MS will be discussed and examples are presented. A brief discussion of electrophoretic procedures used for protein fractionation and biomarker discovery is also included.



**Contents**



## 1. Introduction

A biomarker is a substance that is objectively measured that indicates the presence of an abnormal condition within a patient and allows disease progression and/or therapeutic response to be monitored. A biomarker can be gene- (e.g. single nucleotide polymorphism), protein- (e.g. prostate-specific antigen), or metabolite-based (e.g. glucose, cholesterol, etc.) that has been shown to correlate with the characteristics of a specific disease [1]. An ideal diagnostic biomarker test should be fast, easy to perform and interpret and allow the early detection of a disease, preferably with 100% sensitivity and specificity (no false positive or false negative), and the disease progression and/or therapeutic response to be monitored. To have the greatest impact, a biomarker should be present within an easily obtainable sample, such as urine or blood so that it can be monitored through a non-invasive testing. The test must be capable of screening thousands of samples in a high-throughput manner at affordable cost using accessible equipment.

In the past decade tremendous efforts have been made in the search for novel disease biomarkers using genomic, proteomic

☆ This paper is part of the special issue "Quantitative Analysis of Biomarkers by LC–MS/MS", J. Cummings, R.D. Unwin and T. Veenstra (Guest Editors).

* Corresponding author. Fax: +1 301 846 6037.
*E-mail address:* issaqh@mail.ncifcrf.gov (H.J. Issaq).

1570-0232/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2008.12.028

and metabolomic technologies [2–8]. The search for biomarkers in tissue and biofluids is a worthwhile endeavor because of the tremendous benefits of early detection of a disease when the chances of survival are the highest. The major reason for the high rate of mortality from a disease such as cancer can be attributed to the lack of biomarkers amenable for early screening with high sensitivity and specificity. There is an urgent need for the discovery of disease biomarkers, and this area of research should be given the highest priority.

Proteomic- and metabolomic-based approaches have the potential of leveraging novel discoveries into diagnostic tests that will allow better understanding of disease pathogenesis and progression. The search for a molecule(s) that can be used as biomarker is to search for proteins in a proteome, or metabolites in a metabolome, that show a different characteristic (e.g. abundance, modification, etc.) between control and diseased specimen. Unfortunately, most of the recently identified protein or metabolite biomarkers have failed to replace existing clinical procedures because verification and validation phases of biomarkers are costly, labor intensive and require more time than just the discovery of a biomarker, and they suffer from low diagnostic sensitivity and specificity [9,10]. In spite of the fact that proteome and metabolome profiling is still far from demonstrating its application in clinical diagnosis, preliminary studies clearly indicate its significant potential.

Separation science, chromatography and electrophoresis, combined with mass spectrometry (MS), are essential technologies in the search for proteins and metabolites that can be used as biomarkers. However, the complexity of the proteome and metabolome (number of compounds, concentration dynamic range and chemical diversity) make the search equivalent to looking for a needle in a hay stack. The large concentration dynamic range of a serum proteome (i.e. $\sim 10^{12}$) or complexity of the metabolome (hundreds to thousands of compounds) exceeds the capabilities of current analytical methodologies. This complexity requires the development of special fractionation and separation procedures in order to simplify the mixtures and allow their efficient detection when using MS [11].

While different chromatographic and electrophoretic approaches have been developed for analysis of the metabolome, many more have been developed for proteome analysis. Scientists have used electrophoretic and chromatographic technologies, separately and in combination, both off-line and on-line for the fractionation and separation of protein digests. These combinations include two-dimensional high-performance liquid chromatography [12–14]. In spite of all these efforts, it is still not possible to completely separate, detect, and identify the entire proteome or metabolome.

The complexity and concentration dynamic ranges are major problems for the analysis of serum or plasma proteins; arguably the most interrogated specimen used for biomarker discovery studies. A handful of highly abundant proteins, such as albumin, make up over 90% of serum protein content and can mask the detection of low-abundant proteins. To remove these proteins, affinity chromatography or membrane filters have been used to deplete the high abundance proteins when proteomics is used for biomarker discovery. Depletion of these proteins, however, may cause the loss of proteins that are complexed (generally non-specifically) to highly abundant proteins resulting in the potential loss of beneficial information [15].

## 2. Definitions

Before we discuss biomarker discovery, a number of important definitions should be clarified. In this section, for the benefit of readers who are not familiar with the terminology, we will define a few terms that will be used throughout this manuscript.

### 2.1. Sensitivity

The percentage of positive samples is identified by a model as true positive. Sensitivity decreases with an increase in false negatives. It can be defined by the following equation:

$$\text{Sensitivity} = \frac{\#\text{of true positives}}{(\#\text{ of true positives} + \#\text{of false negatives})} \times 100$$

### 2.2. Specificity

The percentage of negative samples is identified by a model as true negative. Specificity decreases with an increase in false positives. Specificity is defined by the following equation:

$$\text{Specificity} = \frac{\#\text{of true negatives}}{(\#\text{ of true negatives} + \#\text{of false positives})} \times 100$$

### 2.3. Proteomic patterns

Alterations in proteins abundance, structure, or function, act as useful indicators of pathological abnormalities prior to development of clinical symptoms and as such are often useful diagnostic and prognostic biomarkers. It is for this reason that recent hypotheses suggest that detection of panels of biomarkers may provide higher sensitivities and specificities for disease diagnosis than is afforded with single markers. Proteomic patterns as a clinical test for disease diagnosis is a concept that was first introduced by Petricoin et al. [16]. Proteomic pattern analysis relies on comparison of differences in relative abundance of a number of polypeptides/proteins [mass-to-charge ratio ($m/z$) and intensity] within the mass spectrum of two sample sets. For a review of proteomic patterns and their potential in disease diagnosis see reference [17].

## 3. Sample fractionation for effective analysis

The strategy for the discovery of protein or metabolite biomarkers requires the identification and quantitation of as many entities in a proteome or metabolome as possible. The complexity of the proteome and the metabolome requires the development of efficient fractionation, concentration and separation procedures in the hope of detecting the largest possible number of compounds using MS. Analytical chemists have made many attempts to develop methods that possess sufficient resolution to separate large numbers of compounds (proteins and metabolites), as well as be sensitive enough to detect those compounds present in low abundance. Single dimension separations are not capable of resolving all the compounds in the proteome or metabolome. Therefore, the focus has been on developing multi-dimensional separation methods where the first dimension is a fractionation procedure followed by separation and MS analysis. The procedures used are mostly two-dimensional (2D) liquid chromatography and 2D gel electrophoresis. In spite of the fact that 2D fractionation procedures can resolve thousands of proteins and hundreds of metabolites, no method presently exists that can separate, detect and quantify *all* the proteins within a given proteome, or *all* the metabolites in a metabolome.

Since the physicochemical nature of the metabolome and proteome differ, they require different fractionation and separation procedures. The proteome is made up of thousands of polypeptides/proteins that possess the same amino acid building blocks; however, they vary in size and cover a wide range of chemical properties (e.g. acidity, basicity, hydrophobicity, hydrophilicity, etc.). Unlike the proteome, the metabolome contains compounds that

are structurally unrelated and have different chemical and physical properties. Compounds such as sugars, lipids, steroid hormones, hydrocarbons, etc., require different procedures for extraction, separation and detection. A comprehensive separation of all the metabolites in a metabolome may require multiple solvent systems for extraction followed by different separation methods, reversed- and normal-phases in addition to ion exchange and gas chromatography (GC). Also, MS analysis should be carried out using positive and negative ionization modes, since the ionization potential of metabolites differ as well.

Fractionation of a serum, urine or tissue proteome, prior to separation and MS analysis can be carried out at the protein or peptide level after extraction of proteins from the sample. Protein properties make them amenable to different electrophoretic and chromatographic fractionation approaches. Fractionation of the metabolome requires the use of different solvent systems for extraction followed by a chromatographic or CE separation and MS for identification. Most electrophoretic and chromatographic fractionation procedures published to date employ 2D separations, which results in a large number of fractions that require MS analysis. While 2D-PAGE is the most popular electrophoretic procedure for the separation of proteins, SCX/RP HPLC on-line with MS is one of the most popular chromatographic methods. Researchers have attempted to use more than 2D fractionation to increase the number of proteins identified; however, increasing the number of dimensions causes a large increase in the time and cost required to conduct the experiment, and may lead to sample loss [18–21].

### 3.1. Electrophoretic fractionation

Electrophoretic fractionation of a proteome has been mostly carried out by gel electrophoresis although there have been a few attempts at using liquid-phase IEF. Gel electrophoretic fractionation of a proteome is carried out at the protein level according to their isoelectric point (charge) by isoelectric focusing (IEF) in a pH gradient or size by SDS-PAGE. Due to pH gradient instability and irreproducibility, Bjellqvist et al. introduced immobilized pH gradients (IPG) for IEF [22]. In IPGs the pH gradient is generated by a limited number of chemicals. The carrier ampholytes are attached to acrylamide molecules and cast into the gels to form stable, reproducible and fixed pH gradients capable of focusing acidic and basic proteins on a single gel prepared with broad pH gradients. Studies have shown that using narrow-range IPG strips allowed a larger number of proteins to be separated than had been possible with standard 2D-PAGE [23]. IEF using IPGs is an excellent first step for the fractionation of complex mixture of proteins.

Protein fractionation by IEF is not limited to gel electrophoresis [24]. In liquid-phase-based IEF the proteins in a sample are mixed with the desired pH range carrier ampholyte buffer in a focusing cell. Application of an electric potential to the focusing cell causes the proteins to migrate to a position in the established pH gradient equivalent to their respective p*I*. If a protein diffuses away from this pH region, its net charge will change and the resulting electrophoretic forces will influence its migration back to its p*I* point. The net result is the "focusing" of proteins into sharp bands at their p*I* values. The pH gradients are established by using carrier ampholytes. The advantage of liquid-phase-based IEF is that the proteins are separated in a solution and can be eluted in different vials for analysis. For example, the fractionation of proteins using Rotofor from BioRad results in 20 fractions. The disadvantage is that if concentrated protein solution is used, protein precipitation will take place. Liquid-phase-based IEF is faster than gel electrophoresis and is easier to handle; however, it does not have the resolving power of gel electrophoresis, and cannot be used in a 2D-format.

### 3.2. Chromatographic fractionation

Fractionation is the first step in the comprehensive analysis of the proteome which is mostly done by gel electrophoresis and liquid chromatography using different column chemistries (reversed-phase, ion exchange, affinity, size exclusion). Most of the published work in this area is carried out by fractionation of the proteome at the protein or at the peptide level after protein digestion. It is easier to do the fractionation of peptides rather than proteins because peptides are more soluble and easier to handle. For reasons discussed earlier, extraction of the metabolites from serum or urine for global metabolome analysis is more difficult than proteins from the proteome. For example, extraction of the metabolites from serum requires, first, the precipitation of proteins; then, because of the diversity of the metabolites, extraction of the metabolites by different solvent systems. Want et al. [25] evaluated 14 different solvents including methanol, ethanol, acetonitrile and acetone as well as different combinations of methanol and acetonitrile or ethanol in different ratios to select the best solvent system for the comprehensive extraction of metabolites from serum. Methanol was proven to be the best solvent not only for extraction of metabolites but also for precipitation of serum proteins. Targeted analysis of the proteome and the metabolome simplifies the analysis considerably because the analysis is limited to a specific group of compounds to which an analytical procedure is devised. A review of the role of separation science in metabolomics was recently published [26].

Many approaches, using two or more orthogonal HPLC separation procedures, for example, IEX/RP, Affinity/IEX/RP, have been tried to separate complex peptide mixtures such as cell lysate digest. A review by Issaq et al. [4,27] discussed the different multidimensional approaches for the extensive separation of peptides.

## 4. Protein markers

Alterations in proteins abundance, structure, or function, act as useful indicators of pathological abnormalities prior to development of clinical symptoms and as such are often useful diagnostic and prognostic biomarkers. The discovery of protein markers is based on quantitatively measuring the same protein in two different samples reproducibly and accurately. For the sake of reproducibility, it is preferred that both control and diseased specimen be treated differently then analyzed, in the same experiment. The two most common forms of gel electrophoresis are 2D-PAGE and 2D-DIGE. In both methods, proteins are resolved in two sequential steps, according to their charge in the first dimension and by their mass in the second dimension. Good resolution of proteins is obtainable; however, the technique is not amenable to automation or direct coupling to the mass spectrometer. In 2D-PAGE aliquots from two samples are spotted and analyzed on two separate gel plates. After separation staining is carried out to reveal protein expression. The protein levels are compared, spots of interest are excised, the proteins within the spot are enzymatically digested, and the resultant peptides are extracted and analyzed using MS. Ünlü et al. [28] introduced 2D-DIGE to improve the reproducibility and reliability issues encountered with 2D-PAGE. Proteins from two samples are labeled, each with a different cyanine fluorescent dye that has a different excitation and emission wavelengths. The two samples are mixed and spotted on the same plate. The same protein labeled with any of the dyes will migrate to the same position on the gel because the dyes do not affect the size or the isoelectric point of the protein. The proteins are visualized by illuminating the gel with the excitation wavelengths of each of the dyes. Proteins of interest are excised and treated as mentioned earlier. This method allows comparative samples to be run together on the same gel, while 2D-PAGE requires each sample to be run on separate gels and

the staining intensity of coincident spots to be compared between gels. For reviews on 2D electrophoresis the reader should consult two recent reviews [29,30].

Gel-free 2D LC, unlike 2D SDS-PAGE, allows the direct coupling of the LC system to the mass spectrometer so that the effluent from the column (separated species) can be detected and characterized by MS/MS. While 2D-PAGE and LC/MS are used to fractionate complex mixtures of proteins, they can also be used to quantify the relative abundances of proteins within two different samples, one healthy and one diseased. The difference between the two methods of comparisons is that the relative abundances of proteins in the samples are compared before MS analysis (2D-PAGE), while in LC/MS the comparison is carried out after the MS data have been acquired. Relative quantitation of proteins in a bottom-up LC/MS-based proteomics quantitation is carried out by a variety of stable isotope labeling techniques that rely on post-digestion and quantitation of differentially labeled peptides/proteins were recently reviewed by Fenselau [31]. For a general review of mass spectrometry-based proteomic quantitation the reader should consult the review by Ong and Mann [32].

These techniques are used in realm of biomarker discovery. Typically, control and compared samples are differentially stable isotope tagged using *in vivo* or *in vitro* labeling. Commonly used *in vivo* (metabolic) labeling techniques depend on supplying the cell/organism with nutrients containing stable isotopes, allowing simultaneous differential isotope incorporation into all cellular proteins [33,34]. *In vitro* labeling relies on chemical or enzymatic incorporation of stable isotopes allowing labeling of specimens after cell lysis or tissue homogenization [35–37]. Enzymatic labeling can be efficiently used for quantitative profiling of amount-limited specimens, including proteins isolated from membrane microdomains [38] or for direct labeling of clinically relevant human specimens [39].

There is no clear consensus in the literature for a "best practice" isotope labeling strategy for biomarker discovery. Our opinion is that the choice of isotope labeling technique is highly dependent on experimental design, the scope of a particular analysis and the sample or system being analyzed. Quantitation of peptides/proteins is not limited to isotope labeling as discussed above. An alternative approach to stable isotope-based quantitation of proteins has been developed that does not require differential stable isotope labeling and is thus referred to as label-free method. This approach is attractive because it is simple, economical and easy to apply. However, this approach requires high reproducibility of HPLC separation (peptide retention times) and high sensitivity to detect low abundance peptides. Statistical methods have been developed for peak alignment. For a review of these methods the interested reader should consult ref. [40,41]. Mass spectrometry-based quantitation of proteins is achieved by measuring the peptide levels in a mixture. It has been reported that the peak areas of measured peptides is directly proportional to the concentration of proteins [42].

Few selected examples of the role of proteomics in biomarker discovery are presented here. The first example is an interesting approach because the researchers, using their biological knowledge combined PAGE with LC/MS, searched for discriminating proteins in urines from bladder cancer patients and controls [43]. The authors' hypothesis is that since most bladder cancers originate in the urothelial cells lining the lumen of the organ, these cells will release microparticles into the urine. The objective of the study was to identify potential biomarkers in the urinary microparticles of individuals with bladder cancer. This is interesting because, rather than doing a global proteomic analysis of whole urine a specific target in the urine was isolated. Urine microparticles from five healthy individuals and four individuals with bladder cancer were isolated. Samples were isolated from lipids by PAGE followed by digestion of the proteins with trypsin, and analysis of the resulting peptides by RPLC–MS/MS. The results showed that eight proteins, potential biomarkers, were elevated in the microparticles from individuals with bladder cancer. The authors used four patients and five controls, which is a small number. For these eight proteins to be validated as potential biomarkers for bladder cancer, a larger number of subjects (hundreds) should be studied.

The following example deals with the search for biomarkers for epithelial ovarian cancer which is the most lethal gynecological malignancy. An integrated proteomic and bioinformatics analysis was undertaken in the search of biomarkers for this cancer [44]. Abdominal distention is a common symptom that is the result of the accumulation of ascites fluid in the peritoneal cavity. Ascites fluid represents the local microenvironment that is secreted by ovarian tumors that contains different cell types. It is logical, therefore, to believe that proteomic analysis of ascites may lead to the identification of proteins that can be used as biomarkers of ovarian carcinoma. The proteomic study (SCX followed by LC/MS/MS and Gel fractionation/MS/MS) resulted in the identification of more than 2500 proteins. Comparison with other existing data sets, such as urine and plasma proteome profiles, and available 59 ovarian cancer microarray data sets resulted in 80 proteins that can be used as putative biomarkers, of which 18 proteins were detected in the ovarian cancer ascites of the four patients. Only four patients, which is rather a small number, were used in this study. Also, the resulting number of proteins that are labeled as putative biomarkers is quite large. Many other proteomic studies of disease biomarkers can be found by searching PubMed.

## 5. Metabolomic markers

Analysis of the metabolome gives an indication of cell function. The current standard of care for detecting and monitoring bladder tumors is cystoscopy, which is invasive, painful and costly, and not suitable as a screening test. Issaq et al. [45] demonstrated that metabolite variations can be used to discriminate between urines from healthy individuals and transitional cell carcinoma (TCC) patients. HPLC metabolite separations and MS detection along with statistical data analysis were used to profile metabolite variations between the normal and diseased individuals. Urine samples from 48 healthy volunteers and 41 bladder cancer patients were each analyzed by HPLC/MS. Two different statistical methods were used to analyze the data; principal component analysis (PCA) an unsupervised procedure in which linear combinations of all peak intensities are constructed to produce orthogonal components that maximize the total variance in the samples independent of their group labels, and orthogonal partial least square–discriminate analysis (OPLS–DA) a supervised procedure that constructs a linear combination of all peak intensities, which maximizes the separation between healthy and diseased samples. The results indicated that OPLS–DA gave the best sensitivity and specificity by correctly predicting 48 of 48 healthy and 41 of 41 TCC urines, while PCA correctly predicted 46 of 48 healthy and 40 of 41 TCC urines. This proof-of-concept study indicates that metabolomics using HPLC–MS combined with OPLS–DA has the potential of becoming a non-invasive clinical test for bladder cancer.

In a recent study [46] urinary ribonucleosides were used to differentiate between breast cancer patients and healthy volunteers. The nucleosides were extracted from 113 urine samples from breast cancer patients and 99 control samples. Using affinity chromatography combined with LC/MS and bioinformatics pattern recognition, a sensitivity of 87.67% and a specificity of 89.90% were achieved. The study is based on the concept that modified nucleosides are formed post-transcriptionally in RNA. In cancer disease, the cell turnover is increased, yielding higher concentrations of excreted modified nucleosides. The authors concluded that metabonomics based on

the urinary nucleoside profile significantly improved classification compared with breast cancer biomarkers such as CA15-3.

## 6. Current HPLC/MS approaches to biomarker discovery

The analysis of the metabolome and the proteome is tedious, time consuming, and does not possess the throughput required for use as a clinical test. Identification of protein biomarkers from clinically relevant samples including human tissues, and biofluids is further complicated by heterogeneity present in both, human specimens and human population. Importantly, a discussion "tissue vs. fluids", related to the specimens of choice for protein biomarker discovery, gained significant momentum. In terms of biomarker discovery, current results indicate that it is too early to decide what should be a preferential specimen. It appears that both specimens should be used concurrently relying on tissue samples to identify and subsequently track potential biomarker in peripheral serum/plasma [47].

The search for biomarker requires the proteome analysis of hundreds of human serum, urine or tissue samples. It would be advantageous if there were a way to use animal specimen in place of human samples in the search process, then confirm the results using human samples. Pitteri et al. [48] of the Fred Hutchinson Cancer Research Center in Seattle, Washington developed a quantitative proteomic LC/MS approach for the identification of protein changes in plasma samples obtained from a mouse model of breast cancer. They hypothesized that proteins up-regulated in the plasma from mice with tumors may be expressed in human breast cancer cell lines. They hoped to identify a subset of up-regulated proteins in common with proteins expressed in breast cancer cell lines that may represent candidate biomarkers for breast cancer. The authors used a three-dimensional method for fractionating and separating the proteins extracted from plasma. Anion exchange chromatography was used in the first dimension to collect eight separate fractionations. Each fraction was further fractionated using RPLC. These fractions were then treated and digested with trypsin. The resulting peptides in each fraction were analyzed by nano-RPLC/MS–MS. A total of 2310 proteins were identified, of which 133 and 162 proteins were found to be increased or decreased by 1.5-fold or greater, respectively. Those proteins that increased in plasma were compared with a list of 100 reported proteins from proteomic analysis of human breast cancer cell lines. This comparison yielded 49 proteins with increased levels in mouse plasma that were identified in breast cancer cell lines. While it has not yet been reported, it would be of interest to see if any of the 49 common up-regulated proteins could be detected in plasma obtained from breast cancer patients. Differentiating proteins could then be analyzed using specific procedures, such as affinity chromatography/MS. ELISA or antibody high-throughput methods might be a good choice for the analysis of specific differentiating proteins for confirmation purposes.

Separation is an important aspect of the search for disease markers. The HPLC methods used are nano-LC using narrow bore columns, mostly fused silica capillaries packed with different chromatographic materials, connected on-line to a high resolution mass spectrometer. The most widely used columns for biomarker discovery are packed with 3 μm I.D. RP particles that have 300 Å (proteomics) or 100 Å (metabolomics) pore size. Conventional HPLC, not ultrahigh pressure LC (UPLC), have been mostly used for biomarker discovery.

Recently monolithic capillary columns have been used for peptide separation. Monolithic columns consist of a single, rigid or semi-rigid, porous rod that can be organic-based (polymeric) or silica-based. The application of monolithic columns in HPLC and capillary electrochromatography (CEC) have gained momentum in the last decade and have been used for the separation of different groups of compounds, including peptides [49]. The separation of hydrophilic peptides can be resolved in HPLC using a column packed with porous graphitic carbon (PGC), which has a different selectivity than columns packed with RP-derivatized silica and polymer-based bonded phases. In PGC, compounds are retained by a polar retention mechanism [50].

Affinity chromatography, which was first introduced in 1968, is a selective purification, separation, and enrichment technique in which a specific peptide or group of peptides can be enriched from a complex mixture of peptides, for example, cell lysate. The peptide of interest must have a specific property that can be exploited during the affinity procedure. The column is then designed in such a way as to take advantage of the specific property of the peptide(s) of interest, glycopeptides, and phosphopeptides. An organomercurial-agarose (Hg beads) column, which specifically captures cysteine-containing peptides, was used for the enrichment of cysteine-containing peptides from yeast cell lysate [51]. Boronic acid affinity chromatography was used for the enrichment of glycated peptides [52].

## 7. Mass spectrometry in biomarker discovery

Mass spectrometry is arguably the most critical technology in the search for biomarkers in both proteomics and metabolomics. Needless to say that without the recent advances in MS, proteomics and metabolomics would not exist as they do today. The major developments that have had the greatest impact on the ability to use MS for the study of biomolecules such as proteins and peptides are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) [53,54]. In ESI–MS, highly charged droplets dispersed from a capillary in an electric field are evaporated, and the resulting ions are drawn into the inlet of a mass spectrometer where they can be manipulated and eventually detected. Chemistry professor Malcolm Dole of Northwestern University in Evanston, IL first conceived of the technique in the 1960s; however, it was first applied to the analysis of proteins in the early 1980s by Fenn while at Yale University [55]. MALDI–MS was developed in 1987 at the University of Frankfurt, Germany, by Hillenkamp and coworkers [54], and independently by Tanaka et al. at Shimadzu Corp., Kyoto, Japan [56]. In MALDI, sample molecules are laser-desorbed from a solid or liquid matrix containing a highly UV-absorbing substance. Drs. Fenn and Tanaka were awarded the 2002 Nobel Prize in Chemistry for their development of ESI and MALDI and their application to macromolecules. Both ESI–MS and MALDI–MS have made MS increasingly useful for sophisticated biomedical analysis. While MS used to be restricted to the analysis of small organic and inorganic molecules, the development of ESI and MALDI has allowed for such applications as the sequencing and analysis of peptides and proteins, studies of non-covalent complexes, characterization of immunological molecules (i.e. antibodies), and the analysis of intact viruses.

Further important developments in MS include improved sensitivity, resolution, mass accuracy, and scanning speed for tandem MS analysis. MS, unlike other detection methods such as UV/vis and fluorescence, provides a direct, not average, signal for each analyte in the mixture, enabling the surveying of large numbers of analytes in complex mixtures. Other advantages include wider dynamic range, the ability to detect widely different compounds, and its ability to make both, qualitative and quantitative measurements when using the proper experimental design.

In a recent study Nordstrom et al. [57] compared different MS ionization strategies that included ESI, atmospheric pressure chemical ionization (APCI), and MALDI for the identification of metabolites in complex mixture. For this study, extracted serum

**Table 1**
Characteristics of different types of mass analyzers.

| Analyzer | Sensitivity | Resolution | Mass accuracy |
|---|---|---|---|
| Ion-trap | Good | Low | Low |
| Linear ion-trap | Excellent | Low | Low |
| Triple-quadrupole | Good | Good | Good |
| TOF | Good | High | High |
| FTICR | Excellent | High | High |

analyzed using ESI (in both negative and positive ionization modes) resulted in greater than 90% additional unique ions being detected compared to the use of negative ionization mode only. Complementing the ESI analysis with APCI resulted in an approximately 20% increase in unique ions [57]. The results of this study suggest that for true global metabolomics, multiple ionization technologies should be used in order to identify the largest number of metabolites in a metabolome. This is especially true when searching for a biomarker by a strategy that usually involves trying to identify as many components within a mixture as possible.

The majority of published HPLC/MS studies dealing with disease marker discovery utilized mass analyzers of lower resolving power that are not able to resolve overlapping peptides exhibiting similar $m/z$ ratios. Also, potential errors in mass measurement, charge state determination and quantitation cannot be ignored. Recently, MS possessing high resolution capability and high mass measurement accuracy became commercially available (i.e. hybrid linear ion-trap Fourier-transform instruments). A broader use of high resolution and mass accuracy MS instruments should result in enhanced identification of specific and clinically relevant biomarkers [58]. Table 1 gives the overall sensitivity, resolution and mass accuracy of different mass analyzers. As the table indicates FTICR mass analyzers give the highest values and are the preferred analyzers for detecting low abundance molecules.

In contrast to identity-based biomarker discovery that relies on MS to identify potential biomarkers, a significant effort has been invested in MS-derived pattern for biomarker discovery [59]. This strategy relies on detection of protein/peptide peaks that differ in their mass-to-charge ratio ($m/z$) and intensity in patients with cancer compared with healthy individuals. Typically patterns are obtained from human plasma specimens analyzed by MS [60,61]. Raw data are subsequently processed using variety of algorithms and analyzed employing statistical methods, such as principal component analysis (PCA), partial least square-discriminate analysis (PLS–DA), orthogonal PLS–DA and support vector machines (SVM). PCA and OPLS–DA are the two programs that have been used the most. PCA is an unsupervised technique whereby a large data set can be simplified so that components which contribute to the variance can be ranked and compared. It uses linear transformation to define a new coordinate system so that the data set contributing the greatest variance is projected on the first axis (the first principal component), the second greatest variance on the second axis, and so on. PCA reduces dimensionality while it keeps those characteristics that contribute most to the variance. It does this by giving more importance to 'lower order' principal components and ignoring 'higher order' ones, using the assumption that low-order components often contain the "most important" aspects of the data. The data collected from the HPLC–MS chromatograms are subjected to a base-e logarithmic weighing step and scaled using an Applied Biosystems beta version of the MarkerView™ software package prior to performing PCA. OPLS–DA is also a dimensionality reduction method, the same as PCA; however, OPLS–DA is a supervised technique that uses additional information about group or class separation when the new coordinate system is calculated by emphasizing those variables that are responsible for the group separation. Orthogonal PLS–DA (OPLS–DA, SIMCA-P Statistical Program, from Umetrics, Kinnelon, NJ, USA) takes this one step further by capturing variables responsible for group separation in the very first principal component, and other variances in the data set such as intra-group variability are explained by the second and higher principal components. Data obtained by pattern recognition indicate that some proteins/metabolites found in biofluids differ between cancer patients and healthy individuals. The major drawback of this approach is the difficulty in deciding whether the observed differences represent systemic responses to a disease (i.e. immune response) or genuine cancer-specific markers.

## 8. Future outlook

Genomics, proteomics and metabolomics are three important branches of life sciences that are being used for biomarker discovery. Genetic analyses will identify individuals with a predisposition to certain disease, and therefore long-term risk, while proteomic analysis provides the opportunity to detect diseases as they occur. Therefore, direct measurement of protein expression and metabolites is essential to analyze biological processes in normal and disease states.

Although the identity-based biomarker discovery plays major role in MS-based biomarker research certain challenges still exist [9]. In addition to the large concentration dynamic range of protein/metabolites in human biofluids the biological variability of protein/metabolites expression in specimens obtained form diverse human population can be significant. Also, certain disease may represent a collection of different sub-phenotypes involving pathological derangement of biologically diverse pathways/proteins. These factors can introduce significant pre-analytical, analytical and subsequent statistical biases. One of the solutions is to collect the samples under stringent and well controlled processing conditions while employing rigorous statistical data processing. At present, MS-based analysis of human specimens for biomarker discovery is still technically challenging. The use of high resolution and high mass measurement accuracy MS, prospective collection of specimens, appropriate sample processing and statically rigorous analysis should result in more effective biomarker discovery.

Proteomic and metabolomic analyses have not resulted, to date, in a biomarker to replace a clinical test; however, the hope is that with increased efforts by talented scientists, development of efficient separation techniques, sensitive MS and NMR instrumentation and high-throughput validation procedures the future outlook is bright. This requires the support and infusion of funds by the federal government, universities and pharmaceutical companies.

## Acknowledgements

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract number N01-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

## References

[1] J.S. MacNiel, Genome Technol. (2004) 24.
[2] C.M. Michener, A.M. Ardekani, E.F. Petricoin 3rd, L.A. Liotta, E.C. Kohn, Cancer Detect. Prev. 26 (2002) 249.
[3] B.A. Merrick, M.E. Bruno, Curr. Opin. Mol. Ther. 6 (2004) 600.
[4] F.B. Collings, V.S. Vaidya, Toxicology 245 (2008) 167.
[5] X. Zhang, D. Wei, Y. Yap, L. Li, S. Guo, F. Chen, Mass Spectrom. Rev. 26 (2007) 403.
[6] D. Ghosh, L.M. Poisson, Genomics (September) (2008) [Epub ahead of print]

[7] H.J. Issaq, O. Nativ, T. Waybright, B. Luke, T.D. Veenstra, E.J. Issaq, A. Kravsto, M. Mullerad, J. Urol. 179 (2008) 2422.
[8] R.J. Simpson, O.K. Bernhard, D.W. Greening, R.L. Moritz, Curr. Opin. Chem. Biol. 12 (2008) 72.
[9] N. Rifai, M.A. Gillette, S.A. Carr, Nat. Biotechnol. 24 (2006) 971.
[10] V. Kulasingam, E.P. Diamandis, Int. J. Cancer 9 (2007) 123.
[11] H.J. Issaq, Z. Xiao, T.D. Veenstra, Chem. Rev. 107 (2007) 3601.
[12] C.R. Evans, J.W. Jorgenson, Anal. Bioanal. Chem. 378 (2004) 1952.
[13] H.J. Issaq, K.C. Chan, G.M. Janini, T.P. Conrads, T.D. Veenstra, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 817 (2005) 35.
[14] A. Motoyama, J.R. Yates 3rd, Anal. Chem. 80 (2008) 187.
[15] M. Zhou, D.A. Lucas, K.C. Chan, H.J. Issaq, E.F. Petricoin, L.A. Liotta, T.D. Veenstra, T.P. Conrads, Electrophoresis 25 (2004) 1289.
[16] E.F. Petricoin, A.M. Ardekani, B.A. Hitt, P.J. Levine, V.A. Fusaro, S.M. Steinberg, G.B. Mills, C. Simone, D.A. Fishman, E.C. Kohn, L.A. Liotta, Lancet 359 (2002) 572.
[17] Z. Xiao, D. Prieto, T.P. Conrads, T.D. Veenstra, H.J. Issaq, Mol. Cell Endocrinol. 230 (2005) 95.
[18] M.K. Dayarathna, W.S. Hancock, M. Hincapie, J. Sep. Sci. 31 (2008) 1156.
[19] Fan, W.W. Zhang, J. Proteomics 71 (2008) 284.
[20] J. Mortosella, N. Zolotarjova, H. Liu, G. Nicol, B.E. Boyes, J. Proteome Res. 4 (2005) 1522.
[21] B.L. Hood, M. Zhou, K.C. Chan, D.A. Lucas, G.J. Kim, H.J. Issaq, T.D. Veenstra, J. Proteome Res. 4 (2005) 1561.
[22] B. Bjellqvist, K. Ek, P.G. Righetti, E. Gianazza, A. Gorg, R. Westermeier, W. Postel, J. Biochem. Biophys. Methods 6 (1982) 317.
[23] S. Hoving, H. Voshol, J. van Ostrum, Electrophoresis 21 (2000) 2617.
[24] Z. Xiao, T.P. Conrads, D.A. Lucas, G.M. Janini, C.F. Schafer, K.H. Buetow, H.J. Issaq, T.D. Veenstra, Electrophoresis 25 (2003) 128.
[25] E.J. Want, G. O'Maille, C.A. Smith, T.R. Brandon, W. Uritboonthai, C. Qin, S.A. Trauger, G. Siuzdak, Anal. Chem. 78 (2006) 743.
[26] H.J. Issaq, E. Abbott, T.D. Veenstra, J. Sep. Sci. 31 (2008) 1936.
[27] H.J. Issaq, K.C. Chan, G.M. Janini, T.P. Conrads, T.D. Veenstra, J. Chromatogr. B 817 (2005) 35.
[28] M. Ünlü, M.E. Morgan, J.S. Minden, Electrophoresis 18 (1997) 2071.
[29] V.-D. Bergh, G. Arckens, Exp. Rev. Proteomics 2 (2005) 343.
[30] H.J. Issaq, T.D. Veenstra, Biotechniques 44 (2008) 697.
[31] C. Fenselau, J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 855 (2007) 14.
[32] S.-E. Ong, M. Mann, Nat. Chem. Biol. 1 (2005) 252.
[33] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 6591.
[34] S.E. Ong, B. Blagoev, I. Kratchmarove, D.G. Kristensen, H. Steen, A. Pandey, M. Mann, Mol. Cell Proteomics 1 (2002) 376.
[35] S.P. Gigy, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994.
[36] P.L. Ross, Y.N. Huang, J.N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, D.J. Pappin, Mol. Cell Proteomics 3 (2004) 1154.
[37] X. Yao, C. Afonso, C. Fenselau, J. Proteome Res. 2 (2003) 147.
[38] J. Blonder, M.L. Hale, K.C. Chan, L.R. Yu, D.A. Lucas, T.P. Conrads, M. Zhou, M.R. Popoff, H.J. Issaq, B.G. Stiles, T.D. Veenstra, J. Proteome Res. 4 (2005) 523.
[39] W.J. Qian, M.E. Monroe, T. Liu, J.M. Jacobs, G.A. Anderson, Y. Shen, R.J. Moore, D.J. Anderson, R. Zhang, S.E. Calvano, S.F. Lowry, W. Xiao, L.L. Moldawer, R.W. Davis, R.G. Tompkins, D.G. Camp 2nd, R.D. Smith, Mol. Cell Proteomics 4 (2005) 700.
[40] A.H. America, J.H. Cordewener, Proteomics 8 (2008) 731.
[41] H. Liu, R.G. Sadygov, J.R. Yates 3rd, Anal. Chem. 76 (2004) 4193.
[42] D. Chelius, P.V. Bondarenko, J. Proteome Res. 1 (2002) 317.
[43] D.M. Smalley, N.E. Sheman, K. Nelson, D. Theodorescu, J. Proteome Res. 7 (2008) 2088.
[44] L. Gortzak-Uzan, A. Ignatchenko, A.I. Evangelou, M. Agochiya, K.A. Brown, P. St.Onge, I. Kireeva, G. Schmitt-Ulms, T.J. Brown, J. Murphy, B. Rosen, P. Shaw, I. Jurisica, T. Kislinger, J. Proteome Res. 7 (2008) 339.
[45] H.J. Issaq, O. Nativ, T. Waybright, B. Luke, T.D. Veenstra, E.J. Issaq, A. Kravstov, M. Mullerad, J. Urol. 179 (2008) 2422.
[46] A. Frickenschmidt, H. Frhlich, D. Bullinger, A. Zell, S. Laufer, C.H. Gleiter, H. Liebich, B. Kamerrer, Biomarkers 13 (2008) 435.
[47] D.J. Johann Jr., J. Blonder, Exp. Rev. Mol. Diagn. 7 (2007) 473.
[48] S.J. Pitteri, V.M. Foca, K.S. Kelly-Spratt, A.E. Kasadra, J. Proteome Res. 7 (2008) 1481.
[49] M. Batycka, N.F. Inglis, K. Cook, A. Adam, D. Frazer-Pitt, D.G. Smith, L. Main, A. Lubben, B.M. Kessler, Rapid Commun. Mass Spectrom. 20 (2006) 2074.
[50] E.T. Chin, D.I. Papac, Anal. Biochem. 273 (1999) 179.
[51] J.M. Raftery, Anal. Chem. 80 (2008) 3334.
[52] A. Frolov, R. Hoffman, Ann. N.Y. Acad. Sci. 1126 (2008) 253.
[53] M. Dole, L.D. Ferguson, R.L. Hines, R.C. Moble, L.D. Ferguson, M.B. Alice, J. Chem. Phys. 49 (1968) 2240.
[54] M. Karas, D. Bachmann, U. Bahr, F. Hillenkamp, Int. J. Mass Spectrom. Ion Proc. 78 (1987) 53.
[55] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, Science 246 (1989) 64.
[56] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida, Rapid Commun. Mass Spectrom. 8 (1988) 151.
[57] A. Nordstrom, E. Want, T. Northen, J. Lehtio, G. Siuzdak, Anal. Chem. 80 (2008) 421.
[58] M. Mann, N.L. Kelleher, Proc. Natl. Acad. Sci. U.S.A. (September) (2008) [Epub ahead of print]
[59] M.A. Gillette, D.R. Mani, S.A. Carr, J. Proteome Res. 4 (2005) 1143.
[60] A. Ardekani, L.A. Liotta, E.F. Petricoin 3rd, Exp. Rev. Mol. Diagn. 2 (2002) 312.
[61] K. Yanagisawa, Y. Shyr, B.J. Xu, P.P. Massion, P.H. Larsen, B.C. White, R. Roberts, M. Edgerton, A. Gonzalez, S. Nadaf, J.H. Moore, R.M. Caprioli, D.P. Carbone, Lancet 362 (2003) 433.