A rational approach for discovering and validating cancer markers in very small samples using mass spectrometry and ELISA microarrays

Richard C. Zangar[a,*], Susan M. Varnum[a], Chandice Y. Covington[b] and Richard D. Smith[a]
[a]*Pacific Northwest National Laboratory, Richland, WA, USA*
[b]*University of California School of Nursing, Los Angeles, CA, USA*

Abstract. Identifying useful markers of cancer can be problematic due to limited amounts of sample. Some samples such as nipple aspirate fluid (NAF) or early-stage tumors are inherently small. Other samples such as serum are collected in larger volumes but archives of these samples are very valuable and only small amounts of each sample may be available for a single study. Also, given the diverse nature of cancer and the inherent variability in individual protein levels, it seems likely that the best approach to screen for cancer will be to determine the profile of a battery of proteins. As a result, a major challenge in identifying protein markers of disease is the ability to screen many proteins using very small amounts of sample. In this review, we outline some technological advances in proteomics that greatly advance this capability. Specifically, we propose a strategy for identifying markers of breast cancer in NAF that utilizes mass spectrometry (MS) to simultaneously screen hundreds or thousands of proteins in each sample. The best potential markers identified by the MS analysis can then be extensively characterized using an ELISA microarray assay. Because the microarray analysis is quantitative and large numbers of samples can be efficiently analyzed, this approach offers the ability to rapidly assess a battery of selected proteins in a manner that is directly relevant to traditional clinical assays.

Keywords: Proteomics, protein microarray, biomarker, mass spectrometry

1. Introduction

At present, the best way for most individuals to reduce their overall risk of developing cancer may be changes in lifestyle or diet. Beyond these steps, modern medicine can do little to reduce the risk of cancer in most people. For individuals who develop cancer, the best way to improve survival and minimize the adverse effects of therapeutic intervention is early detection of the disease. In spite of considerable research to improve our ability to detect cancer in its early stages, only modest gains have been made in this area. In the case of breast cancer, mammography and self-examination have helped to reduce overall mortality rates. However, these examination methods are limited by the size and density of the tumor and are subject to high levels of false positives and negatives. Cervical cancer can be detected by microscopic examination of cells obtained in a Pap test. Widespread use of this test has reduced the mortality from cervical cancer dramatically [1]. Both the cervical and breast cancer tests rely on a physical or visual examination of either tissue or cells.

In addition to physical examinations, it is also possible to detect cancer by measuring changes in protein levels or other molecular markers of disease. The use-

*Corresponding author: Richard Zangar, Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, WA 99352, USA. Tel.: +1 509 376 8596; Fax: +1 509 376 6767; E-mail: richard.zangar@pnl.gov.

fulness of molecular markers of cancer are primarily limited by the ability to define a concentration cut-off that accurately distinguishes between individuals with and without cancer. The basic measure of this accuracy is the true positive (sensitivity) and false positive rates (specificity). It is not unusual for concentrations of individual protein markers to vary by two orders of magnitude even between healthy individuals. This high level of inter-individual variability makes it much more unlikely a single protein biomarker can accurately detect disease. Rather, it is becoming increasingly clear that the accurate identification of early-stage cancer will likely result from an analysis of a battery or profile of markers.

Serum levels of prostate-selective antigen (PSA[1]) are widely used as a screen for prostate cancer, and this assay currently represents the most successful use of a protein marker for the early detection of cancer. Although PSA levels alone are commonly insufficient to determine the presence or absence of prostate cancer, they provide a simple screening method to identify individuals that either are at high risk or already have this disease. Markers for other cancers have shown promise but have yet to be demonstrated as useful for cancer detection. An example of this is CA125, which is elevated in woman diagnosed with ovarian cancer [2]. Initial studies suggested that this marker had great promise in early detection. However, because levels can be increased during menstruation, pregnancy and other benign conditions, the clinical utility of CA125 may be limited to use in monitoring disease recurrence and tumor burden after surgery [3,4].

In spite of the limited success of past studies for cancer biomarkers, recent advances in the ability to efficiently screen proteins in complex biological samples have renewed optimism about our ability to find markers of disease. Within a single sample, proteomic methods that employ MS technology can now simultaneously analyze hundreds or thousands of proteins that may potentially function as disease markers. This lack of bias in the experimental design encourages the identification of new markers of disease. These proteomic analyses are expected to result in many new candi-

date marker proteins being identified in the near future. Even so, this technology is presently limited by the high cost of the instrumentation and sample throughput. In addition, MS analyses presently only provide a relative measure of protein levels, although there are significant efforts in progress aimed at making such measurements much more quantitative. At present, however, the MS data is fundamentally different from traditional clinical protein assays, which have typically required a truly quantitative value that can be reproducibly measured in different laboratories.

Because of the anticipated increase in new candidate markers and the expectation that analyzing multiple proteins will improve the accuracy of cancer detection, it seems likely there will be a growing need to rapidly screen multiple markers in the future. Enzyme-linked immunosorbent assay (ELISA) microarrays are one emerging technology that can perform this task. We feel that this approach will prove particularly valuable since the resulting data are directly comparable to traditional clinical analyses. Once the initial assay is developed, the antibody microarrays offer a relatively inexpensive system for rapidly measuring trace levels of multiple proteins in very small samples. Therefore, we consider the MS technology to be primarily suited for the discovery of new biomarkers (Table 1). The ELISA microarray technology offers an efficient way to validate sets of potential marker proteins identified by MS, and this approach can also serve as a discovery tool for a targeted set of proteins such as cytokines.

In this review, we discuss our efforts to develop and integrate MS proteomics and ELISA microarray analyses into a system for screening biological samples for disease markers. In particular, we will focus on the analysis of nipple aspirate fluid (NAF), a very small volume fluid that is likely to be an ideal source of markers for breast cancer.

2. Nipple aspirate fluid (NAF)

A liquid secretion is present in the ducts and sinuses of a woman's breast prior to pregnancy and after lactation has ceased. These intraductal fluids are continuously secreted, concentrated and absorbed. Components in this fluid are thought to be either actively absorbed from plasma by breast cells and then transported intracellularly to the ducts, or synthesized within the breast epithelium and directly released into the duct [5, 6].

---

[1]Abbreviations: AMT, accurate mass and time tag; cLC, capillary liquid chromatography; DREAMS, Dynamic Range Enhancement Applied to Mass Spectrometry; ELISA, enzyme linked immunosorbent assay; FTICR, Fourier Transform Ion Cyclotron Resonance; i.d., internal diameter; LCQ, liquid chromatography classic ion trap; MMA, mass measurement accuracy; MS, mass spectrometry; NAF, nipple aspirate fluid; nanoESI, nano-electrospray ionization; PMT, potential mass and time tag; PSA, prostate-selective antigen.

Table 1
Complimentary strengths and weaknesses on the MS proteomics and ELISA microarray analyses for the discovery and validation of cancer biomarkers

	Advantages	Limitations	Applications
MS Analysis	100's-1000's of proteins Unbiased analysis	Quantitation Clinical usefulness Throughput	Discovery
ELISA Microarray	High-throughput Relatively inexpensive Quantitative Clinically relevant results	Biased sample set Limited # of proteins (<50/microarray?)	Discovery & Validation

With the assistance of a non-invasive aspiration device, this ductal fluid can be extracted through the nipple and is referred to as NAF. In comparison to serum, NAF should offer a superior fluid for detecting breast cancer since NAF proteins are specifically from breast tissue. NAF collects from the epithelial cells lining the ductal system of the breast, the same cells that are the source of ~85% of breast cancers [7]. Therefore, it is not surprising that NAF has been found to be a rich source of cancer markers [8]. For example, median concentrations of carcinoembryonic antigen (CEA) in NAF were reported to be 1100 ng/ml, compared to normal serum levels of less than 6 ng/ml [9]. As breast tissue is the major source of prostate-selective antigen (PSA) in women, it is also not surprising that the reported mean levels of PSA in NAF are ~10,000 times greater than mean serum levels in healthy woman [10]. In addition, PSA levels in NAF samples from healthy women are inversely correlated with breast cancer risk and these changes may precede detectable mammographic abnormalities [11,12]. In women with breast cancer, PSA levels in NAF averaged 0.2 ± 0.03 ng/ml compared to 63 ± 10 ng/ml in NAF from women without breast cancer [13]. Therefore, the evidence suggests that NAF is a highly concentrated source of protein markers for the early detection of breast cancer.

NAF has been used to investigate breast tissue since first described by Papanicolaou and colleagues [14] and Fleming [15] in the mid 1950s. The literature now includes over 100 references on noninvasively-obtained "nipple aspirate fluids" (not including those on breast cystic fluid or lavaged fluids, which require invasive approaches to obtain). These reports provide a description of the sloughed-off epithelial and foam cells and other components of this intraductal fluid. This area of research was greatly enhanced when Petrakis, King and colleagues [16,17] pioneered the use of NAF for epidemiological studies of the structure and function of breast cells and the biochemistry of NAF, which was collected from over 2000 women who underwent mammography at a breast care center in Northern California.

In those early studies, it was found that women whose NAF samples contained unusual or "atypical" breast epithelial cells had a five-fold increased risk of breast cancer compared to women who did not yield any breast fluid when aspirated. Among premenopausal women the presence of atypical cells was associated with a 15-fold increase in risk compared to non-yielders of NAF [18–24]. Because of these early results, some experts have suggested the mere presence of fluid denotes risk, however the ductal tree is a mucosal surface and fluid can be obtained from virtually all adult women [8, 25]. Since this early work, NAF has been used for studies on lactose, cholesterol, cholesterol epoxides, lipids, lipid peroxides, immunoglobulins, proteins, sodium, potassium, hormones, growth factors, with differences found in compositions between those with breast cancer, benign disease, and controls for some components [26–31].

The reason that NAF can be difficult to obtain is that the nipple ducts are typically occluded by keratin plugs, which maintain the closed microenvironment of the ductal tissue except during late pregnancy and lactation [29]. As a result, in early studies, the aspiration of the fluid was often uncomfortable and/or unattainable. Until the last five years, only about 50% of women (parous and nulliparous) enrolled in NAF studies were able to aspirate and express ductal fluids [29]. Newer methods developed by Covington and colleagues have resulted in greater yields and are well-tolerated by patients [32,33]. This collection method includes use of home preparation prior to the clinic/lab visit, patient-applied heat and massage, and the use of a nipple aspirator. (A detailed protocol is available from Dr. Covington at chandice@ucla.edu). This method has resulted in success rates of 85% and 98% in two separate populations of women [25]. Repeated attempts to aspirate over several days have been reported to increase success rates [8].

In collaboration with NeoMatrix (Irvine, CA), this NAF collection system has undergone clinical trials and will soon be marketed for use (see http://www.

neomatrix.com/halo_sys/works.asp). Another system called Mammary Aspirate Specimen Cytology Test (MASCT), which was originally developed for isolating cells from NAF for cytological analysis, has recently been purchased by Cytyc Corporation. As such, soon it should be possible to routinely obtain NAF samples from women in clinical settings using a commercial apparatus.

In spite of the potential of NAF for detection of breast cancer, it should be noted that there are issues associated with the use of this fluid that have not been adequately resolved. It is important to realize each breast is composed of many isolated ductal systems and that the fluid from each of these ductal systems is expressed through a separate pore in the nipple. It is likely that even when NAF is successfully collected from a single nipple, fluid will not be collected from all of the ducts. Larger tumors may also occlude the distal part of the breast ductal system, preventing aspiration of ductal fluid from this region. As such, the significance of these issues needs to be addressed before NAF can be routinely used in the analysis of breast.

3. Identification of disease markers in minute samples using mass spectrometry

It is not unusual for proteomes to have differences in protein abundance that exceed 9 orders of magnitude. For instance, albumin is present in human serum at a concentration of $\sim$35 mg/ml whereas many cytokines are present at less than 10 pg/ml [34]. As a result, the detection of low-abundance proteins, including disease markers, in complex mixtures represents a challenge for global proteomic analyses. Substantial advances in MS technology have been made in recent years so that it is now possible to analyze minute, complex samples. These technological advances have been recently reviewed [35] and will only be briefly described here. The results of these technological advances is that it is currently feasible to analyze trace amounts of sample by MS to quickly identify the presence of hundreds or thousands of proteins and to make semi-quantitative comparisons of individual protein levels between samples on a proteomic scale.

3.1. An overview of accurate mass and time (AMT) tag proteomics approach

The proteomic approach developed by Smith and coworkers (diagrammed in Fig. 1) appears to be well

suited for the discovery of biomarkers of disease in complex protein mixtures. The first phase of this approach requires the development of a potential mass and time (PMT) tag database of tryptic peptides using combined capillary liquid chromatography (cLC) separations with conventional tandem MS (e.g., using an ion trap MS). The PMT tags are then converted to AMT tags based on the high mass measurement accuracy (MMA) of the Fourier transform ion cyclotron resonance (FTICR) MS and the known elution time of the peptide. The PMT tag database is commonly prepared using samples from diverse individuals in order to ensure that as many proteins as possible are included in this database.

Once developed, a PMT tag database can be used to analyze individual samples with greater throughput and sensitivity. This second phase of this approach relies strictly on the high MMA obtained with the cLC-FTICR MS analysis and known elution time of each peptide: further tandem MS characterization will not be required (Fig. 1). Although this approach requires considerable effort to develop the initial AMT tag database, once completed it is possible to determine the presence of hundreds or thousands of proteins in a sample using a single LC-FTICR run. Because tandem MS is no longer required in this analysis, it is possible to integrate peak intensities for each peptide for the duration of its elution time. As such, peptide peak intensities for isotopically-labeled peptide pairs can be integrated over the full elution time, which results in greater sensitivity and more accurate information on the relative abundances of individual proteins. In the section below, we discuss this approach in more detail and provide some examples of the technological advances that make this system particularly well suited for identifying proteins markers in minute samples.

3.2. Sample fractionation

For a complex protein sample such as serum, there may be dozens of peptides eluting from the cLC column and entering the MS at any one time. In order to select peptides for tandem MS analysis, the instrument does an initial scan to identify peptides that are eluting at any one time. The most abundant peptides are then isolated and fragmented, and the mass of the fragments determined. The mass analysis of the fragmentation pattern is the second or "tandem" step of the MS analysis and these data are used for identifying the parent peptide and associated protein. This selection process is repeated throughout the run, with the exception that
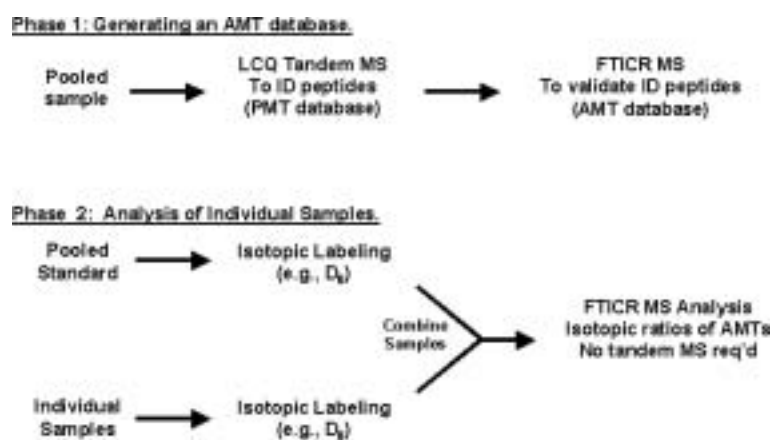
Fig. 1. Schematic overview of the methodology used to generate and validate AMT tags (Phase 1) and then to analyze individual samples using the AMT database using isotopic labeling (Phase 2).

peptide masses that were selected for fragmentation are subsequently excluded for a short time period. Peptides with strong ion currents, can be accumulated most rapidly and typically produce the best quality tandem MS spectra that facilitate peptide identification. Identification of lower level peptides can be problematic since these peptides are less likely to be selected for MS/MS analysis and a complex set of peptides is eluted too rapidly for all peptides to be subjected to tandem MS analysis. Even if selected, levels of these peptides may be too low to yield a high quality fragmentation spectrum. Therefore, tandem MS analysis has a limited ability to identify low-abundance peptides due to the fact that these peptides are inherently more challenging to analyze and because of the manufactured instrument bias for the selection of abundant peptides. In order to better identify low abundance peptides, fractionation (e.g., cation exchange) prior to reverse phase cLC-MS analysis is beneficial, since this step effectively increases the concentration of individual peptides in a particular fraction and decreases the overall sample complexity.

3.3. Microcapillary columns and nano-electrospray ionization (nanoESI)

The development of nanoESI allows for the efficient ionization of column effluent flow rates extending as low as ~20 nl/min [35]. In conventional ESI MS analysis, the large majority of the sample is wasted (i.e., is not ionized properly or the ions do not reach the MS detector). However, in nanoESI, a finer droplet size is produced. This improvement allows the nanoESI tip to be placed closer to the MS inlet with the result that a much greater portion of the spray reaches the detector. In addition, because the spray is finer, ionization of the peptides is more efficient. These changes result in greater signal intensity and decreased background noise. The overall result is that although nanospray requires less sample, overall sensitivity is improved even when working with samples that have low concentrations of peptides.

For analysis of small volumes of complex peptide mixtures, it is advantageous to use an LC system that has the lowest feasible flow rates so as to optimize sensitivity. Reducing the column diameter of the LC allows for efficient analysis of small samples, and microcapillary columns with 50 μM i.d. are now routinely used and work with columns having 15 μM i.d. has been shown to be feasible. Sample volumes of 10 μl or less are optimal for these microcapillary separations, allowing for routine analysis of 10 μg or less of digested protein for a typical MS analysis and as little as a few nanograms in more specialized applications (e.g., using 15 μM i.d. capillary columns). The end result of combining μLC separations with nanoESI is the ability to efficiently analyze the proteome of minute biological samples. Furthermore, the combination of μLC separations with multidimensional separations greatly extend the protein identification dynamic range to >8 orders of magnitude in protein relative abundance [36].

One developing technology that offers potential for further improving detection of low level peptides is Dynamic Range Enhancement Applied to Mass Spectrometry (DREAMS) [37]. This technology relies on the fact that in tandem MS analysis, the first MS serves to selectively collect ions. The capacity of this first MS ion trap is limited. Therefore, when one or a few

highly abundant peptides are present, the intensities of low-abundance peptides may be insufficient for detection. What the DREAMS technology does is utilize the first-phase MS instrument to rapidly identify and then selectively eject the highly abundant peptides during the analysis. This is in contrast to conventional tandem MS analysis, where the first MS instrument serves to isolate one peptide prior to fragmentation for tandem MS analysis. The first MS step in DREAMS eliminates the most abundant peptides, a process which has the opposite role of the first step of conventional tandem MS. This step increases the number of low abundance peptides that will accumulate to a point where the signal intensity is above the background noise. An initial application of DREAMS resulted in an ∼50% increase in the number of detectable peptides [38].

3.4. *Utility of FTICR MS and high mass measurement accuracy (MMA) for protein identification*

Although more expensive than conventional MS instrumentation, FTICR MS appears to be particularly well suited for analyzing trace biological samples. This is because FTICR simultaneously provides exceptionally high sensitivity, resolution and MMA [39]. Within a proteome, many tryptic peptides have unique masses making it possible to identify the presence of the parent protein based on the mass of a single peptide even without tandem MS analysis [40]. Therefore, with sufficiently high MMA and knowledge of the complete proteome, measurement of a single peptide mass can be sufficient to identify the parent protein [40]. Computer simulations [40] and application to a microorganism [41] have supported the overall feasibility of this approach. The number of tryptic peptides with unique masses has been calculated for C. elegans using several levels of MMA [40]. (The C. elegans genome encodes for ∼19,000 predicted proteins, a number that likely exceeds what is expressed in a single mammalian tissue and, as such, this analysis of the C. elegans proteome may be relevant to mammalian samples.) About 95% of C. elegans proteins contain at least one tryptic peptide that has a unique mass at 1 ppm resolution [40]. The unique mass designation means that this peptide could be differentiated from every other peptide formed by tryptic digestion of the whole proteome at a MMA of 1 ppm. The ability to discriminate between peptides continues to increase with improved MMA, highlighting the advantage of FTICR compared with conventional mass spectrometers [40].

As discussed above, sample fractionation can greatly extend the concentration range over which proteins can be detected. In the absence of prior fractionation, it has been demonstrated that 10^6 to 10^7 dynamic range is obtainable in a single cLC-FTICR MS analysis [42]. This range of sensitivity is in stark contrast to proteomics analyses based on two-dimensional gels, which likely have a dynamic range of 10^3 or less.

3.5. *Generation of AMT tags and their use for protein identification*

The AMT tag approach to protein identification was originally developed by Smith and coworkers [43–45] and has been adapted by others [46]. Furthermore, by using differential isotopic labeling, this approach provides a rapid method for determining the relative abundance of individual peptides from separate samples. This system relies on a tiered approach to match a particular peptide with the parent protein (Fig. 1). Initially, tandem MS spectra for a tryptic peptide from a complex protein mixture are generated using liquid chromatography interfaced with a classic ion trap (i.e., LCQ tandem MS). These spectra are searched against the appropriate database in SEQUEST, a program that aligns measured tandem MS fragmentation patterns (partial sequence information) with predicted fragmentation patterns of peptides formed by the same protease for all proteins in the database [47]. Based on the fit of the experimental and theoretical tandem MS data, SEQUEST assigns a cross correlation ("Xcorr") value. Therefore, at this first stage of the analysis, the peptide is characterized by the tandem MS analysis and the LC elution time. If the tandem MS analysis produces a SEQUEST cross correlation score above a given level that is dependent upon peptide charge state, the peptide is considered a potential mass and time (PMT) tag. A second analysis of the peptide-containing sample is then undertaken using FTICR-MS and identical chromatographic conditions. To account for cLC column variation over time and between columns, software has been developed that corrects for subtle shifts in elution times by normalizing peak times based on the elution of several well-defined, major peaks [48]. If the peptide has the same mass as predicted by SEQUEST analysis of the tandem MS spectrum and the same LC elution time, then the peptide is considered an accurate mass and time (AMT) tag. The AMT tag can be identified in subsequent FTICR runs based solely on its retention time and the high mass accuracy measurement: further tandem MS analysis is not required.

One of the major advantages of this process is that in order to develop the AMT tag database, samples can be pooled and fractionated. As discussed above, this process serves to increase the relative and absolute concentrations of low-abundance peptides and thereby increase the likelihood of obtaining a quality tandem MS analysis and an accurate identification by SEQUEST. Larger amounts of peptides are required to obtain good tandem MS data than are required for identification of the intact peptide by FTICR MS. This increased sensitivity for the intact peptide is due to the exceptional sensitivity of the FTICR MS and the fact that there is a loss of signal intensity associated with peptide fragmentation during tandem MS.

A large database of AMT tags can be generated by this process, allowing for the analysis of hundreds or thousands of proteins in a single FTICR run. Furthermore, by using differential isotopic labeling, it is possible to determine the relative levels of individual peptides from separate samples [43–45]. That is, the peak areas of both isotopic variants of a peptide can be determined in a single FTICR run to provide a measure of the relative levels of the parent protein in the original samples. An alternative approach for determining relative protein levels that does not rely on isotopic labeling has recently been proposed [49]. In this approach, the peak area (i.e., the ion current) of each peptide is determined and normalized based on all peak intensities to correct for run-to-run variation. Using this approach, it should be possible to determine differences in individual peptide abundances without prior identification of the parent protein. As such, tandem MS experiments could follow these initial analyses and could be focused on identifying peptides that are known to change in cancer patients. This approach should make these types of studies more efficient.

Although it may be possible to combine the tandem MS analyses and the determination of isotopic peptide ratios in a single run, it seems likely that combining both analyses will sacrifice both sensitivity and quantitation. The reason for this is that quantitative analysis of peptide ion peaks is not possible when the instrument is dedicated to tandem MS analysis. Therefore, one of the real advantages of the above AMT approach for determining relative abundance of peptides is that it physically separates the tandem MS identification from the quantitative analysis.

4. Examples of the proteome studies

4.1. Analysis of Deinococcus radiodurans proteome

The first application of the AMT tag approach is in the bacterium, *D. radiodurans*, which has a predicted complete proteome of 3,187 proteins [50,51]. A combination of cLC-tandem MS and FTICR MS analyses identified with very high confidence 6,997 AMT tags (i.e., tryptic peptides with unique masses at 1 ppm MMA) corresponding to 1,910 proteins from *D. radiodurans* [52]. This represents 61% of the predicted proteins present in this species and is the broadest proteome coverage for an organism achieved to date. The proteins identified included a majority of the predicted proteins in most of the defined functional categories [51], including 88% of the proteins associated with protein synthesis and 78% with transcription. Approximately 32% of the predicted proteins from the *D. radiodurans* database are hypothetical (i.e., have no significant homology to any proteins in public databases at the time of annotation) and 16% are conserved hypothetical (i.e., have limited homology to a protein with unknown function). These analyses identified 48% of the hypothetical proteins and 55% of the conserved hypothetical proteins that were predicted from the genomic sequence. Since the initial report, continuation of *D. radiodurans* proteomic studies has resulted in ∼83% coverage of the *D. radiodurans* proteome [53]. This is the first study to demonstrate such comprehensive proteome coverage and is suggestive of the breadth of proteomic studies now possible. This is particularly true since these initial studies provide a large set of AMT tags for future quantitative expression studies, which can be conducted using only FTICR analysis. Therefore, greater sensitivity and higher throughput can be obtained than would be possible for studies that rely on routine tandem MS measurements.

4.2. Human plasma/serum proteome

A project was recently initiated at the Pacific Northwest National Laboratory to examine the proteome of human serum or plasma. The goal of this project is to develop the capabilities to screen human blood for biomarkers of selected disease states. In an intitial study, select, abundant serum proteins (i.e., immunoglobulins and transferrin) were removed from serum by affinity chromatography. The remaining proteins were digested with trypsin and analyzed by LCQ tandem MS. In order to improve detection of

low-abundance peptides, samples were fractionated by cation exchange chromatography. The result of this analysis was the identification of 490 plasma proteins with reasonable confidence [54], a result that was noted as substantially expanding the known serum proteome in an independent review article [34]. These analyses have been extended to include FTICR-MS analysis. As noted above, the higher MMA of FTICR provides another method to confirm a peptide that was initially identified by tandem MS analysis. The cLC-FTICR analysis in combination with an updated database approximately doubled the number of proteins that could be identified with high confidence in serum (Joel Pounds, personal communication).

More recently, in an effort to streamline proteomic procedures, an analysis of human blood plasma was undertaken without affinity removal of abundant proteins or cation exchange chromatography of peptides [36]. Rather, chromatographic conditions and tandem MS default settings were systematically tested in order to elucidate the advantages of various analytical methods in regards to peptide identification. A total of 99 LCQ runs were performed under various conditions, but only ∼365 μg of non-depleted plasma protein was used in these analyses. Since plasma typically contains about 70 mg protein/ml, this is the equivalent of ∼5 μl of sample. Over 800 proteins with concentration ranges spanning over 8 orders of magnitude were identified. These results suggest that it is possible to develop an extensive PMT tag database without affinity purification and multidimensional chromatography. The elimination of these steps would considerably reduce the mass of protein required and the amount of labor.

4.3. Initial NAF proteome analysis

NAF samples are typically only 10 to 50 μl, and therefore provide a good test case for the ability of nanoESI and MS analyses to work with small volume samples. Prior to our proteomics characterization, very little information was available about the protein characteristics of NAF. One biochemical characterization had identified 10 proteins or associated enzymatic activities [55]. An analysis using two-dimensional gel electrophoresis identified ∼50–100 spots, but most of these spots appeared to represent multiple glycosylation states of a few highly abundant, unidentified proteins [56].

In our study, we initially identified the most abundant proteins in NAF using in-gel digestion and cLC-ion trap tandem MS. These proteins were immunoglobins, poly-
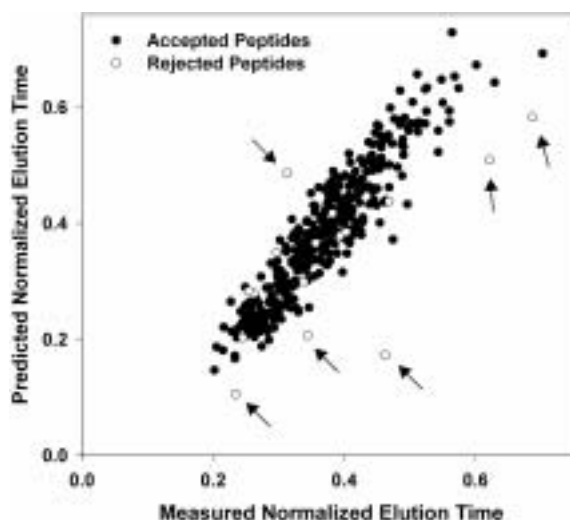

Fig. 2. Comparing measured and predicted peptide elution times as a means of validating tandem MS identifications. Tandem MS analysis and stringent SEQUEST criteria were used to obtain peptide sequence information. The measured elution times for individual peptides are graphed against the elution time predicted by an artificial neural network for the predicted peptide sequence. Closed circles indicate peptides that passed all evaluation criteria, including manual evaluation of the tandem MS spectrum for peptides from proteins with less than three hits. Open circles represent peptides that passed all criteria except manual inspection of the spectrum. Arrows indicate six peptides that are apparently outliers from the rest of the data. Figure from [25], with permission.

immunoglobin receptor, albumin and lactoferrin [25]. Combined, these abundant proteins account for about two-thirds of the total protein content of NAF. In order to increase our ability to detect lower-abundance proteins, the abundant proteins were removed by affinity chromatography. Tryptic digests of the remaining proteins were fractionated by ion exchange chromatography and analyzed by cLC-ion trap tandem MS. These analyses identified 64 proteins in NAF with a very high level of confidence. Fifteen (23%) of these proteins have been previously reported to be altered in the blood or tumor tissue of breast cancer patients [25], a surprisingly high level of potential biomarkers for breast cancer. Therefore, this finding is consistent with the expectation that NAF will be an exceptional fluid for identifying markers of breast cancer.

About half of the proteins identified in NAF were typical milk or immune proteins [25], consistent with evidence that the breast ductal cells have a residual secretory function in the absence of active lactation [29]. It is also interesting that approximately two-thirds of the proteins identified in NAF were likely to be glycosylated, as judged by annotated information in the SwissProt database [25]. One function of glycosyla-
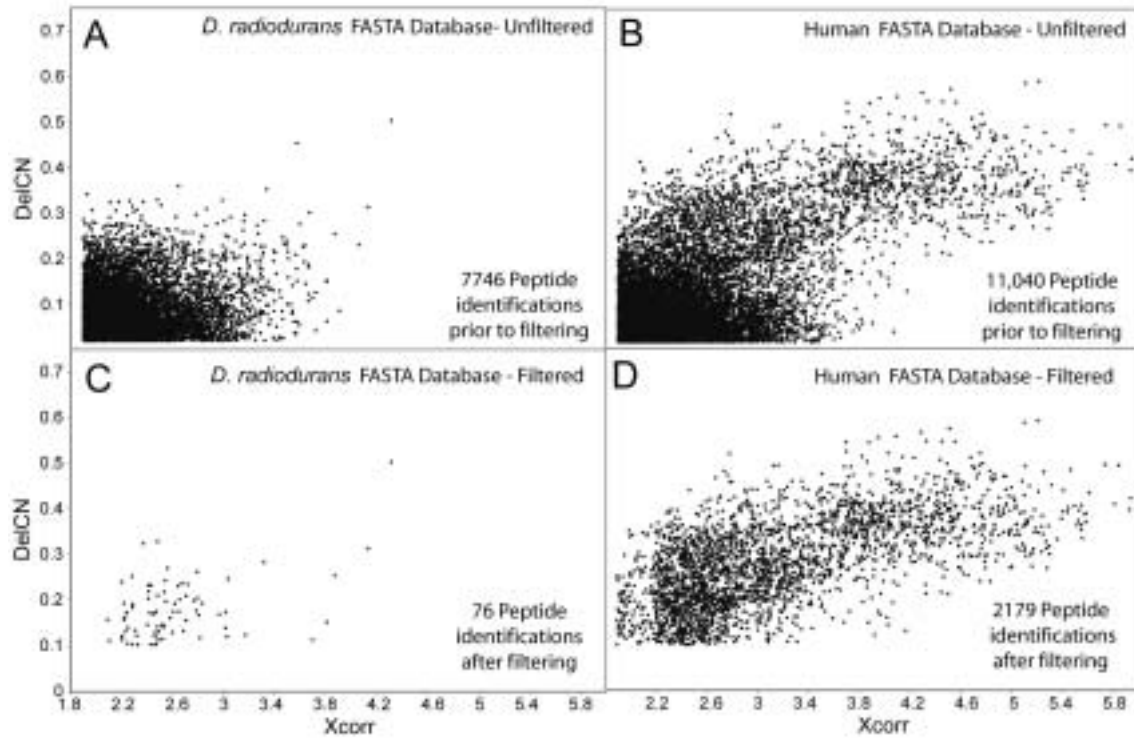
Fig. 3. Effects of low and high stringency SEQUEST peptide filters on the number of peptides identified in an analysis of digested human serum. Tandem mass spectra from serum peptides were compared to databases containing either the whole proteome of the bacterium D. radiodurans or a partial human proteome. Results shown represent the peptides identified before (unfiltered) and after (filtered) the application of stringent identification criteria. Figure from [54], with permission.

tion is to protect a protein against proteolysis, so there has been concern that trypsin would not be able to effectively digest these proteins. However, most proteins are only glycosylated at a few sites and part of the digestion procedure involves denaturing the proteins in 8 M urea, followed by dilution to 2 M urea prior to addition of trypsin. Our experience suggests that once the protein is denatured, glycosylation prevents neither tryptic digestion nor protein identification by tandem MS analysis of non-glycosylated peptides [25].

5. Critical evaluation of proteomic data

One of the greatest challenges to large-scale proteomic analyses is determining protein identifications with a high level of confidence. There are inherent trade-offs between the number of proteins identified and the confidence in the identifications. That is, as the criteria for protein matching using tandem MS data are relaxed, more identifications will be made but at the cost of an increasing number of false positive identifi-

cations. Generally, proteins for which >3 peptides can be identified using a strict criteria are considered to be identified with a very high level of confidence, and are not typically evaluated further. However, when identifying proteins based on smaller numbers of peptides, more care must be taken in order to obtain a reasonable level of confidence in the resulting data.

We have initially applied the criteria developed by Yates and coworkers [57,58], and found this system to work well for initial identification of proteins. This approach initially identifies proteins based on peptide charge state, presence of predicted tryptic cut sites, and a number of parameters provided by SEQUEST analysis of the tandem MS spectrum. For proteins with only a few peptides identified, the tandem MS spectrum should also be "manually" examined as an additional test to eliminate incorrect assignments. After applying these criteria, we then attempted to validate the results by alternative methods. In the case of the peptides from the NAF proteins, we also analyzed the data using a computer program that predicts the cLC elution time of a peptide based on sequence analysis [48]. For
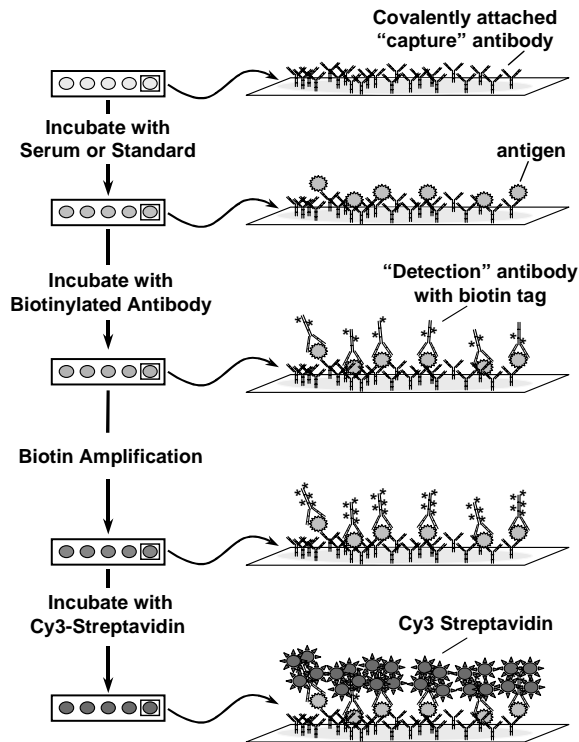
Fig. 4. Schematic diagram of experimental approach for a "sandwich" ELISA microarray analysis. Figure from [59], with permission.

comparative purposes, peptides that passed all evaluation criteria except the manual evaluation of the spectra were also included in this analysis. The results of the predicted and measured normalized elution times are shown in Fig. 2. Of the 412 peptides analyzed, only 6 peptides clearly were outliers from the rest of the data. All six of these peptides belong to the group of 18 peptides (indicated by open circles) that had failed manual evaluation. Therefore, this analysis suggests that the criteria used to evaluate the NAF peptides was sufficiently stringent to allow for a high level of confidence in the identified proteins.

In the analysis of the human serum proteome discussed above, criteria were similar to those used with the NAF samples, except that allowances were made for the presence of endogenous chymotrypsin and elastin proteolytic activity in this fluid [54]. In order to further evaluate the likelihood of false identifications, the human peptide data was compared to a partial proteomic database for humans or the complete database from *Deinococcus radiodurans*. Since manual inspection of the spectra matched with the bacterial database was not undertaken, the comparable data from the human database were used. As can be seen in Fig. 3, in

the absence of stringent criteria to evaluate the data, comparison to the *D. radiodurans* database resulting in 7746 peptide matches compared to 11,040 in the human database. Once stringent data filters were applied, only 76 protein matches were present in the bacterial database compared to 2179 matches in the human database. The number of protein matches from the bacterial database presumably would decrease if the individual spectra had been manually inspected.

For many proteomic studies, it is important to have very high confidence in each protein identified. For biomarker studies, however, it may be preferable to include greater numbers of protein identifications even if some may be incorrect. The main reason for this is that by utilizing less stringent criteria, more proteins total and proportionally more low-abundance proteins will be included in the data pool. This pool of low-abundance proteins includes many cytokines and other proteins that may be the best disease markers. Since most proteins will not be useful as markers, it is worthwhile to include as many proteins as possible to improve the odds of finding the few that are most likely. Regardless of the level of confidence in the protein identification, if a protein appears to be a good marker of a disease, it will be necessary to confirm this observation by an independent method (e.g., immunoblot or ELISA), as well as by using a different set of samples. As such, low stringency in the initial identification is acceptable if the key protein identifications are independently validated.

6. ELISA microarrays: An ideal tool for targeted, high-throughput, quantitative analysis of protein levels in minute samples

The small NAF sample volumes, which typically range from 10 to 50 μl, have limited the ability to systematically evaluate individual proteins as markers of breast cancer. What is required is a highly sensitive technology capable of simultaneously analyzing multiple proteins in small clinical samples such as NAF. We have developed a "sandwich" ELISA in a microarray format (Fig. 4) that is ideally suited for such uses [59, 60]. This assay is exceptionally sensitive, being able to quantify protein concentrations down to the sub-pg/ml range. Standard curves extend over $\sim$1000-fold range of protein concentration, easily bracketing normal human variations in protein levels. Using this ELISA microarray, we demonstrated that levels of hepatocyte growth factor were elevated in serum from recurrent
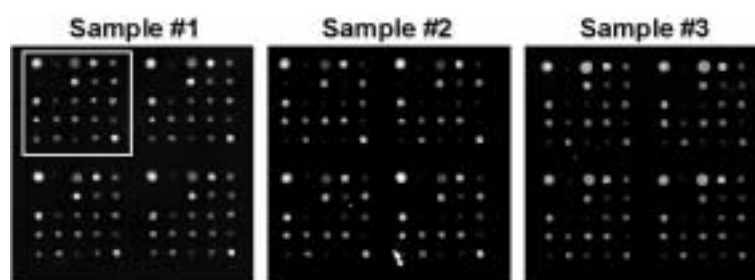
Fig. 5. ELISA microarray analysis of NAF samples from three healthy women. 100 μg of NAF protein was analyzed on each microarray. The quadrant enclosed in the image on the left represents one set of 20 different ELISA assays that are performed in quadruplicate on each slide. The bright spot on the top left of each quadrant is an orientation marker.

breast cancer patients [59], even though hepatocyte growth factor accounted for as little as 10 parts per billion of the total protein content in serum. Results from the hepatocyte growth factor microarray ELISA also correlated well ($r^2 = 0.90$) with a commercial 96-well ELISA assay [59]. Inter-plate reproducibility (coefficients of variation) ranged from 4–13% for five separate ELISA assays simultaneously analyzed on a single plate. Therefore, it is clear that this ELISA microarray is suitable for quantifying trace amounts of proteins in complex protein mixtures.

The primary costs with the ELISA microarrays are associated with developing and validating the individual assays themselves. If commercial antibodies and an antigen are available, it typically costs at least $1000 per assay for these reagents. If antigens and antibodies must be generated, then initial costs will increase substantially. Once the basic assay is developed, however, the per-assay costs are very low, since only trace amounts of antigen and antibodies are used for each analysis and the analysis time on a per sample basis is small.

We have recently developed a multiplex microarray ELISA that contains assays for 20 proteins that have been reported to be altered in sera from cancer patients. We have undertaken a preliminary analysis of NAF samples from 3 individuals using this microarray (Fig. 5). The smallest of the 77 NAF samples we have obtained contained over 200 μg protein, and all but two samples contain over 300 μg protein. Therefore, in preliminary ELISA microarray analyses of NAF, we used 100 μg of NAF protein per analysis, since this mass of protein is sufficient for duplicate microarray analyses of the smallest NAF samples and triplicate analyses for almost all samples. Although this array targeted marker proteins found in serum, most of these proteins were readily detectable in each of the NAF samples. Consistent with other studies (see Section 2),

we estimated that many of the proteins were at least 10-fold more concentrated in NAF than serum. More recently, we have used as little as 2 μg of NAF and found we can get good results with this mass of sample (unpublished results). These data demonstrate that the ELISA microarray approach can simultaneously assay multiple, low-abundance proteins in NAF samples and that sample sizes are not a problem.

7. Summary

NAF is likely to be a concentrated source of protein markers for breast cancer, as this fluid specifically collects proteins from the breast ductal system that is the origin of ~85% of breast cancer cases. This is in contrast to serum, which contains proteins from every tissue in the body. A major challenge in working with NAF is the very small sample sizes, which limits the number and type of assays that can be performed.

Recent advances in proteomics will enhance our ability to search for disease markers in minute samples such as NAF. In particular, it appears that proteome approaches employing MS detection will allow for analysis of low-abundance proteins [61,62]. Excellent sensitivity is expected to be essential for early detection of cancer, since it is unlikely that small tumors will significantly alter levels of abundant proteins. Advances in MS technology now appear to have the required sensitivity for measuring disease markers. For example, recent studies have identified various membrane receptors, cytokines and related proteins that are potential markers for disease in human fluids such as serum and NAF [25,54]. With parallel advances in the ability to determine the relative or absolute levels of proteins by mass spectrometry with extremely high sensitivity, it seems likely that there will be large numbers of potential disease markers identified in the future. How-

ever, it seems unlikely that MS data alone will be able to adequately define the clinical usefulness of disease markers in the near future due to the immature state of the technology, its expense, and present limitations on throughput.

As such, there will be a need to critically evaluate these potential markers for clinical usefulness, using a technology that is quantitative, has high throughput, and can handle minute samples. We believe that ELISA microarrays are ideally suited for this type of analysis. The ELISA microarray is fundamentally the same analysis as the 96-well ELISAs traditionally used in the clinical, but offers the advantage of being able to simultaneously screen many proteins in minute samples. There is also theoretical and experimental evidence that suggests that reducing assay size will improve signal to noise ratios and thereby enhance sensitivity [63]. Although it is time consuming to initially develop and optimize each ELISA microassay, once this has been accomplished, it is a relatively quick and inexpensive to process hundreds or thousands of samples. Overall, we believe that the combined use of advanced mass spectrometry capabilities and ELISA microarrays offers a way to quickly advance the discovery of new and better markers for early detection of cancer and other diseases, particularly when working with small volume samples such as NAF.

Acknowledgements

References

[1] T.J. Herzog, New approaches for the management of cervical cancer, *Gynecol Oncol* **90** (2003), S22–S27.
[2] F. Alexander-Sefre, U. Menon and I.J. Jacobs, Ovarian cancer screening, *Hosp Med* **63** (2002), 210–213.
[3] U. Menon and I. Jacobs, Screening for ovarian cancer, *Best Pract Res Clin Obstet Gynaecol* **16** (2002), 469–482.
[4] A.E. Guppy and G.J. Rustin, CA125 response: can it replace the traditional response criteria in ovarian cancer? *Oncologist* **7** (2002), 437–443.
[5] J. Russo, P. Furmanski, R. Bradley, P. Wells and M.A. Rich, Differentiation of normal human mammary epithelial cells in culture: an ultrastructural study, *Am J Anat* **145** (1976), 57–77.
[6] M.E. Rennison, S.E. Handel, C.J. Wilde and R.D. Burgoyne, Investigation of the role of microtubules in protein secretion from lactating mouse mammary epithelial cells, *J Cell Sci* **102**(Pt 2) (1992), 239–247.
[7] T. Wright and A. McGechan, Breast cancer: new technologies for risk assessment and diagnosis, *Mol Diagn* **7** (2003), 49–55.
[8] E.R. Sauter, E. Ross, M. Daly, A. Klein-Szanto, P.F. Engstrom, A. Sorling, J. Malick and H. Ehya, Nipple aspirate fluid: a promising non-invasive method to identify cellular markers of breast cancer risk, *Br J Cancer* **76** (1997), 494–501.
[9] L. Foretova, J.E. Garber, N.L. Sadowsky, S.J. Verselis, D.M. Joseph, A.F. Andrade, P.G. Gudrais, D. Fairclough and F.P. Li, Carcinoembryonic antigen in breast nipple aspirate fluid, *Cancer Epidemiol Biomarkers Prev* **7** (1998), 195–198.
[10] K.K. Kaulsay, E.H. Ng, C.Y. Ji, G.H. Ho, T.C. Aw and K.O. Lee, Serum IGF-binding protein-6 and prostate specific antigen in breast cancer, *Eur J Endocrinol* **140** (1999), 164–168.
[11] E.R. Sauter, M. Daly, K. Linahan, H. Ehya, P.F. Engstrom, G. Bonney, E.A. Ross, H. Yu and E. Diamandis, Prostate-specific antigen levels in nipple aspirate fluid correlate with breast cancer risk, *Cancer Epidemiol Biomarkers Prev* **5** (1996), 967–970.
[12] E.R. Sauter, J. Babb, M. Daly, P.F. Engstrom, H. Ehya, J. Malick and E. Diamandis, Prostate-specific antigen production in the female breast: association with progesterone, *Cancer Epidemiol Biomarkers Prev* **7** (1998), 315–320.
[13] F. Mannello, M. Malatesta, M. Sebastiani, S. Battistelli and G. Gazzanelli, Molecular forms and ultrastructural localization of prostate-specific antigen in nipple aspirate fluids, *Clin Chem* **45** (1999), 2263–2266.
[14] G.N. Papanicolaou, G.M. Bader, D.G. Holmquist and E.A. Falk, Cytologic evaluation of breast secretions, *Ann N Y Acad Sci* **30** (1956), 1409–1421.
[15] R.M. Fleming, Cytological studies in lesions of the breast: findings in nipple secretions and aspirates from tumors, *South Med J* **48** (1955), 74–78.
[16] E.B. King, D. Barrett, M.C. King and N.L. Petrakis, Cellular composition of the nipple aspirate specimen of breast fluid. I. The benign cells, *Am J Clin Pathol* **64** (1975), 728–738.
[17] E.B. King, D. Barrett and N.L. Petrakis, Cellular composition of the nipple aspirate specimen of breast fluid. II. Abnormal findings, *Am J Clin Pathol* **64** (1975), 739–748.
[18] N.L. Petrakis, L. Mason, R. Lee, B. Sugimoto, S. Pawson and F. Catchpool, Association of race, age, menopausal status, and cerumen type with breast fluid secretion in nonlactating women, as determined by nipple aspiration, *J Natl Cancer Inst* **54** (1975), 829–834.
[19] N.L. Petrakis, C.A. Maack, R.E. Lee and M. Lyon, Mutagenic activity in nipple aspirates of human breast fluid, *Cancer Res* **40** (1980), 188–189.
[20] N.L. Petrakis, M.R. Wrensch, V.L. Ernster, R. Miike, E.B. King and W.H. Goodson, Prognostic-Significance of Atypical Epithelial Hyperplasia in Nipple Aspirates of Breast Fluid, *Lancet* **2** (1987), 505–505.
[21] N.L. Petrakis, M.R. Wrensch, V.L. Ernster, R. Miike, J. Murai, N. Simberg and P.K. Siiteri, Influence of pregnancy and lactation on serum and breast fluid estrogen levels: implications for breast cancer risk, *Int J Cancer* **40** (1987), 587–591.
[22] N.L. Petrakis, M.L. Lim, R. Miike, R.E. Lee, M. Morris, L. Lee and L. Mason, Nipple Aspirate Fluids in Adult Nonlactating Women – Lactose Content, Cationic Na^+, K^+, Na^+/K^+

Ratio, and Coloration, *Breast Cancer Research and Treatment* **13** (1989), 71–78.

[23] N.L. Petrakis, Physiological, Biochemical, and Cytologic Aspects of Nipple Aspirate Fluid, *Breast Cancer Research and Treatment* **8** (1986), 7–19.

[24] O.W. Sartorius, H.S. Smith, P. Morris, D. Benedict and L. Friesen, Cytologic evaluation of breast fluid in the detection of breast disease, *J Natl Cancer Inst* **59** (1977), 1073–1080.

[25] S.M. Varnum, C.C. Covington, R.L. Woodbury, K. Petritis, L.J. Kangas, M.S. Abdullah, J.G. Pounds, R.D. Smith and R.C. Zangar, Proteomic characterization of nipple aspirate fluid: identification of potential biomarkers of breast cancer, *Breast Cancer Res Treat* **80** (2003), 87–97.

[26] N.L. Petrakis, R.E. Lee, R. Miike, M.E. Dupuy and M. Morris, Coloration of breast fluid related to concentration of cholesterol, cholesterol epoxides, estrogen, and lipid peroxides, *Am J Clin Pathol* **89** (1988), 117–120.

[27] N.L. Petrakis, M.L. Lim, R. Miike, R.E. Lee, M. Morris, L. Lee and L. Mason, Nipple Aspirate Fluids in Adult Nonlactating Women – Lactose Content, Cationic Na$^+$, K$^+$, Na$^+$/K$^+$ Ratio, and Coloration, *Breast Cancer Research and Treatment* **13** (1989), 71–78.

[28] F. Vizoso, L.M. Sanchez, I. Diez-Itza, L.M. Luz and C. Lopez-Otin, Factors affecting protein composition of breast secretions from nonlactating women, *Breast Cancer Res Treat* **23** (1992), 251–258.

[29] M.R. Wrensch, N.L. Petrakis, L.D. Gruenke, V.L. Ernster, R. Miike, E.B. King and W.W. Hauck, Factors associated with obtaining nipple aspirate fluid: analysis of 1428 women and literature review, *Breast Cancer Res Treat* **15** (1990), 39–51.

[30] D.P. Rose, B. Berke, L.A. Cohen and H. Lahti, A comparison of serum and breast duct fluid-immunoassayable prolactin and growth hormone with bioassayable lactogenic hormones in healthy women and patients with cystic breast disease, *Cancer* **60** (1987), 2761–2765.

[31] D.P. Rose, Hormones and growth factors in nipple aspirates from normal women and benign breast disease patients, *Cancer Detect Prev* **16** (1992), 43–51.

[32] C. Covington, A. Mitchell-Gieleghem, D. Lawson, I. Eto and C. Grubbs, Presence of carotenoid, an anticarcinogenic marker, in nipple aspirates postlactation, *Adv Exp Med Biol* **501** (2001), 143–152.

[33] L.M. Nantais-Smith, C.Y. Covington, B.A. Nordstrom-Klee, C.J. Grubbs, I. Eto, D.M. Lawson, B.A. Pieper and L.L. Northouse, Differences in plasma and nipple aspirate carotenoid by lactation status, *Nurs Res* **50** (2001), 172–177.

[34] N.L. Anderson and N.G. Anderson, The human plasma proteome: history, character, and diagnostic prospects, *Mol Cell Proteomics* **1** (2002), 845–867.

[35] R.D. Smith, Y. Shen and K. Tang, Ultra-sensitive and quantitative analyses from combined separations-mass spectrometry for the characterization of proteomes, *Accounts Chem Res* **37** (2004), 269–278.

[36] Y. Shen, J.M. Jacobs, D.G. Camp, R. Fang, R.J. Moore and R.D. Smith, High efficiency SCXLC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome, *Anal Chem* **16** (2004), 144–154.

[37] M.E. Belov, G.A. Anderson, N.H. Angell, Y. Shen, N. Tolic, H.R. Udseth and R.D. Smith, Dynamic range expansion applied to mass spectrometry based on data-dependent selective ion ejection in capillary liquid chromatography fourier transform ion cyclotron resonance for enhanced proteome characterization, *Anal Chem* **73** (2001), 5052–5060.

[38] L. Pasa-Tolic, R. Harkewicz, G.A. Anderson, N. Tolic, Y. Shen, R. Zhao, B. Thrall, C. Masselon and R.D. Smith, Increased proteome coverage for quantitative peptide abundance measurements based upon high performance separations and DREAMS FTICR mass spectrometry, *J Am Soc Mass Spectrom* **13** (2002), 954–963.

[39] R.D. Smith, Evolution of ESI-mass spectrometry and Fourier transform ion cyclotron resonance for proteomics and other biological applications, *Internation Journal of Mass Spectrometry* **200** (2000), 509–544.

[40] T.P. Conrads, G.A. Anderson, T.D. Veenstra, L. Pasa-Tolic and R.D. Smith, Utility of accurate mass tags for proteome-wide protein identification, *Anal Chem* **72** (2000), 3349–3354.

[41] M.S. Lipton, L. Pasa-Tolic', G.A. Anderson, D.J. Anderson, D.L. Auberry, J.R. Battista, M.J. Daly, J. Fredrickson, K.K. Hixson, H. Kostandarithes, C. Masselon, L.M. Markillie, R.J. Moore, M.F. Romine, Y. Shen, E. Stritmatter, N. Tolic', H.R. Udseth, A. Venkateswaran, K.K. Wong, R. Zhao and R.D. Smith, Global analysis of the Deinococcus radiodurans proteome by using accurate mass tags, *Proc Natl Acad Sci USA* **99** (2002), 11049–11054.

[42] Y. Shen, C.D. Masselon, N. Tolic, L. Pasa-Tolic, H.K. Hixson, J.M. Jacobs and R.D. Smith, Probing the dynamic range of high resolution RPLC/FTICR MS proteomics measurements, *Anal Chem* (2004), In press.

[43] T.P. Conrads, K. Alving, T.D. Veenstra, M.E. Belov, G.A. Anderson, D.J. Anderson, M.S. Lipton, L. Pasa-Tolic, H.R. Udseth, W.B. Chrisler, B.D. Thrall and R.D. Smith, Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and 15N-metabolic labeling, *Anal Chem* **73** (2001), 2132–2139.

[44] M.A. Moseley, Current trends in differential expression proteomics: isotopically coded tags, *TRENDS in Biotechnology* **19** (2001), S10–S16.

[45] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb and R. Aebersold, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nat Biotechnol* **17** (1999), 994–999.

[46] A.I. Nepomuceno, D.C. Muddiman, H.R. Bergen, J.R. Craighead, M.J. Burke, P.E. Caskey and J.A. Allan, Dual electrospray ionization source for confident generation of accurate mass tags using liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry, *Analytical Chemistry* **75** (2003), 3411–3418.

[47] A. Ducret, O. Van, I, J.K. Eng, J.R. Yates, III and R. Aebersold, High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry, *Protein Sci* **7** (1998), 706–719.

[48] K. Petritis, L.J. Kangas, P.L. Ferguson, G.A. Anderson, L. Pasa-Tolic', M.S. Lipton, K.J. Auberry, E. Stritmatter, Y. Shen, R. Zhao and R.D. Smith, Use of artificial neural networks for the prediction of peptide liquid chromatography elution times in proteome analyses, *Anal Chem* **75** (2003), 1039–1048.

[49] W.X. Wang, H.H. Zhou, H. Lin, S. Roy, T.A. Shaler, L.R. Hill, S. Norton, P. Kumar, M. Anderle and C.H. Becker, Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards, *Analytical Chemistry* **75** (2003), 4818–4826.

[50] O. White, J.A. Eisen, J.F. Heidelberg, E.K. Hickey, J.D. Peterson, R.J. Dodson, D.H. Haft, M.L. Gwinn, W.C. Nelson, D.L. Richardson, K.S. Moffat, H. Qin, L. Jiang, W. Pamphile, M. Crosby, M. Shen, J.J. Vamathevan, P. Lam, L. McDonald, T. Utterback, C. Zalewski, K.S. Makarova, L. Aravind, M.J. Daly and C.M. Fraser, Genome sequence of the radioresistant

bacterium Deinococcus radiodurans R1, *Science* **286** (1999), 1571–1577.

[51] M.D. Ermolaeva, H.G. Khalak, O. White, H.O. Smith and S.L. Salzberg, Prediction of transcription terminators in bacterial genomes, *J Mol Biol* **301** (2000), 27–33.

[52] R.D. Smith, G.A. Anderson, M.S. Lipton, L. Pasa-Tolic, Y. Shen, T.P. Conrads, T.D. Veenstra and H.R. Udseth, An accurate mass tag strategy for quantitative and high-throughput proteome measurements, *Proteomics* **2** (2002), 513–523.

[53] J.S. MacNeil, Making FT/MS matter, *Genome Tech* **34** (2003), 34–40.

[54] J.N. Adkins, S.M. Varnum, K.J. Auberry, R.J. Moore, N.H. Angell, R.D. Smith, D.L. Springer and J.G. Pounds, Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry, *Mol Cell Proteomics* **1** (2002), 947–955.

[55] M. Malatesta, F. Mannello, G. Bianchi, M. Sebastiani and G. Gazzanelli, Biochemical and ultrastructural features of human milk and nipple aspirate fluids, *J Clin Lab Anal* **14** (2000), 330–335.

[56] P. Klein, E. Glaser, L. Grogan, M. Keane, S. Lipkowitz, P. Soballe, L. Brooks, J. Jenkins, S.M. Steinberg, D.M. DeMarini and I. Kirsch, Biomarker assays in nipple aspirate fluid, *Breast J* **7** (2001), 378–387.

[57] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik and J.R. Yates, III, Direct analysis of protein complexes using mass spectrometry, *Nat Biotechnol* **17** (1999), 676–682.

[58] M.P. Washburn, D. Wolters and J.R. Yates, III, Large-scale analysis of the yeast proteome by multidimensional protein identification technology, *Nat Biotechnol* **19** (2001), 242–247.

[59] R.L. Woodbury, S.M. Varnum and R.C. Zangar, Elevated HGF levels in sera from breast cancer patients detected using a protein microarray ELISA, *Journal of Proteome Research* **1** (2002), 233–237.

[60] S.M. Varnum, R.L. Woodbury and R.C. Zangar, A protein microarray ELISA for screening biological fluids, *Methods Mol Biol* (2003), In Press.

[61] W.H. McDonald and J.R. Yates, III, Shotgun proteomics: integrating technologies to answer biological questions, *Curr Opin Mol Ther* **5** (2003), 302–309.

[62] S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon and R. Aebersold, Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology, *Proc Natl Acad Sci USA* **97** (2000), 9390–9395.

[63] R.P. Ekins, Ligand assays: from electrophoresis to miniaturized microarrays, *Clin Chem* **44** (1998), 2015–2030.