6-1-2006

Sampling and Analytical Strategies for Biomarker Discovery Using Mass Spectrometry

Thomas P. Conrads

Brian L. Hood

Timothy D. Veenstra

Sampling and analytical strategies for biomarker discovery using mass spectrometry

Thomas P. Conrads, Brian L. Hood, and Timothy D. Veenstra

BioTechniques 40:799-805 (June 2006)
doi 10.2144/000112196

There is an often unspoken truth behind the course of scientific investigation that involves not what is necessarily academically worthy of study, but rather what is scientifically worthy in the eyes of funding agencies. The perception of worthy research is, as cost is driven in the simplest sense in economics, often driven by demand. Presently, the demand for novel diagnostic and therapeutic protein biomarkers that possess high sensitivity and specificity is placing major impact on the field of proteomics. The focal discovery technology that is being relied on is mass spectrometry (MS), whereas the challenge of biomarker discovery often lies not in the application of MS but in the underlying proteome sampling and bioinformatic processing strategies. Although biomarker discovery research has been historically technology-driven, it is clear from the meager success in generating validated biomarkers that increasing attention must be placed at the pre-analytic stage, such as sample retrieval and preparation. As diseases vary, so do the combinations of sampling and sample analyses necessary to discover novel biomarkers. In this review, we highlight different strategies used toward biomarker discovery and discuss them in terms of their reliance on technology and methodology.

INTRODUCTION

Protein biomarker discovery requires multidisciplinary strategies that incorporate intelligent sample collection, processing, data acquisition, and analysis (see Reference 2). Protein identification has historically been accomplished by Edman degradation. Although an arguably powerful technology, it suffers both in terms of throughput and scope, as proteins can only be sequenced consecutively and only after appreciable purification. Furthermore, even if these criteria are fulfilled, this technology falls short, in that it provides no facile translation to routine protein assay development. Improvements in biomarker assays have all relied on the development and use of high affinity reagents (i.e., antibodies). The development of modern separations coupled with advanced mass spectrometry (MS) has ushered in a new paradigm in the throughput and scope with which proteins can be identified and characterized. These proteomic-scale capabilities now enable thousands of proteins to be identified from complex mixtures (1). Indeed, no other technology parallels the capability of MS in the identification of proteins from complex proteomic samples, such as serum, plasma, urine, or cerebrospinal fluid (CSF). It is for these reasons that conventional wisdom suggests that through application of these new tools, novel and specific disease biomarkers will be identified. Accordingly, the recent past has seen an exponential rise in data acquisition (i.e., MS) and data analysis (i.e., computer hardware and software) capabilities, while it may be argued that commensurate advances in sample collection and processing have lagged. The increasing power of MS and bioinformatic tools often results in experimental designs that are overly dependent on technology and suffer from lack of imaginative sample preparation. As shown in Figure 1, there is often an inverse relationship between the complexity of sample preparation and the amount of data acquired or the sophistication of the bioinformatic analysis. Simply put, minimal sample preparation prior to MS analysis will require more data acquisition and more sophisticated bioinformatic analysis. There is, however, a direct correlation between the amount of data acquired and the sophistication of the bioinformatic analysis.

ANALYTICAL SAMPLING OF BIOFLUIDS

In the simplest sense, the goal of protein biomarker discovery is to identify a protein or panel of proteins that distinguish patients afflicted with a particular disease from healthy individuals (2). While the premise seems simple enough, achieving this goal has not been a trivial pursuit. Ideally, such a biomarker or biomarkers would be assayable in biological samples obtained through minimal invasion. Biofluids such as urine, serum, and plasma readily fulfill such criteria and are routinely collected during physical examinations. Unfortunately, these biofluids represent an extremely difficult matrix to characterize, even by the most advanced MS technologies. These difficulties are clear if one considers the physiological and analytical

National Cancer Institute, Frederick, MD, USA

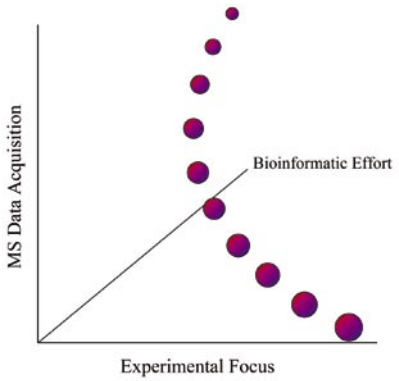

Figure 1. Graphical representation of the dependency of the mass spectral data acquisition and bioinformatic effort based on the level of experimental focus in proteomic investigations. As illustrated, increased experimental focus results in a corresponding decrease in both mass spectrometry (MS) data acquisition and bioinformatic effort.

challenges in discovering, for example, a tumor-specific protein biomarker in serum. Assume that a population of tumor cells secretes an aberrant protein into the circulatory system. The blood, collected from a vein at the inner elbow and from which the serum sample is prepared, is derived from a 7.5-L circulatory system that encompasses approximately 100,000 km of veins, arteries, and capillaries. While the local concentration of the biomarker may be high in the microenvironment of the tumor, its travels take it through thousands of kilometers of biological highways until it reaches the point of extraction (i.e., inner elbow). This journey will have many confounding effects that present several analytical challenges to its facile detection, the most obvious of which is dilution. The high concentration of the biomarker within the vicinity of the tumor will be dramatically decreased as it moves within the circulatory system. Since the activity level of proteases in blood is high, the biomarker may also be digested into a variety of different fragments prior to collection. Therefore the primary sequence, and potentially the functional significance, of the biomarker can radically change between the point of entry into the circulation system and collection.

Probably the most challenging aspect of discovering biomarkers in biofluids such as serum and plasma is their proteomic complexity (3). While the exact number of proteins is not known, it has been estimated that has many as 10^5–10^6 different species may be constituents in the blood proteome. Adding to that complexity, the blood proteome is in a persistent state of flux, with materials being exchanged with healthy cells and proteins being released from necrotic and apoptotic cells. Furthermore, this complexity is not simply a function of the sheer numbers of proteins, but also the dynamic range of their concentrations estimated conservatively at minimally ten orders of magnitude (4).

With these factors to consider, obviously the ability to comprehensively characterize these biofluid proteomes is a critical first step in the hope of identifying clinically significant biomarkers. In the recent past, there have been a number of studies utilizing different separation technologies followed by MS protein identification to characterize these proteomes (5). The two primary separation methods for obtaining comprehensive proteome coverage have been two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and multidimensional liquid chromatography (LC) (i.e., strong cation exchange LC followed by reversed-phase LC). Research conducted at Large Scale Biology Corporation (LSBC; Vacaville, CA, USA) has been instrumental in highlighting the utility of 2D-PAGE/MS analyses in biofluid proteomics. Studies conducted at LSBC demonstrated the resolution of over 1400 and 3700 distinct protein spots on 2D-PAGE gels from urine and serum samples, respectively, resulting in the respective identification of 150 and 325 unique proteins (6,7). Investigations using multidimensional LC coupled to MS analysis have typically resulted in the identification of 1500–2000 proteins in serum or plasma and 100–250 proteins in urine (8,9). Although it might be argued that 2D-PAGE results in fewer protein identifications, the upshot is that it provides quantitative information in comparative analyses through the comparison of stained protein spot intensities and information related to protein isoforms and posttranslational modifications. Multidimensional fractionation allows more protein identifications but suffers from less rigorous quantitative comparisons that are based typically on the number of peptides observed and/or their raw ion current for a given protein. While the numbers of proteins identified using either method do not approach the expected complexity of the given biofluid, they do provide orders of magnitude more coverage and information than available prior to the proteome era and warrant the use of these technologies in biomarker discovery.

In many ways, biomarker discovery using such technologies is a case of the proverbial needle-in-the-haystack expedition without knowing what the haystack or the needle looks like. The hope is that the data reveal a handful of proteins that distinguish disease and fulfill the necessary criteria to move to a validation stage using expanded clinical cohorts. There has been a tremendous amount of resources, both in terms of time and money, devoted to global biofluid analysis for biomarker discovery. While there have been many potentially useful biomarkers put forth in the literature, there have been none that have been validated to meet Food and Drug Administration (FDA) requirements. Does this mean the leveraging of MS and bioinformatics to analyze complex biofluid samples has been a waste of time? Absolutely not! The analytical advances made in the development of instrumentation with greater acquisition speed, sensitivity, dynamic range, and resolution, as well as bioinformatic tools that more accurately identify proteins, have allowed a greater understanding of the proteome content of many biofluids. We are only now beginning to understand changes in the proteomes common to many disease conditions and in a sense recognizing just what the haystack looks like.

SOURCES OF CLINICAL SAMPLES

An ideal biomarker would be found in a biological specimen that can be obtained noninvasively and, in the case of cancer, be as directly proximal to the

underlying tumorigenesis as possible to endow high levels of specificity and sensitivity to the disease. Biofluids such as serum, plasma, urine, and CSF have been the predominant choice for biomarker discovery. One obvious reason is their accessibility; acquiring tumor biopsies from patients is invasive, costly, and impractical. Another reason arises from a forward thinking vision for development of routine clinical assays. Unfortunately, the sheer complexity of these biofluids prevents direct association of a protein identified from a proteomic study to its site of origination. Fortunately, recent developments in utilizing new techniques that enable sampling of material closer to the site of interest may provide more directed routes to the identification of biomarkers suitable for clinical validation.

Formalin-Fixed Paraffin-Embedded Tissues

To correlate the biomarker-tumor connection, tumor biopsies have to be performed and other biochemical or imaging methods, such as immunohistochemistry, also need to be performed to validate the site of origination of a given protein biomarker. The extraction of tumor material either surgically or through a needle biopsy, however, is highly invasive, and the costs (both in terms of time and money) of obtaining sizable cohorts of samples from willing donors makes such studies almost impractical for discovery-driven biomarker research. Fortunately, there is a vast archive of tumors in the form of formalin-fixed paraffin-embedded (FFPE) tissues. While these tissues have been routinely used for immunohistochemistry, in situ hybridization studies, and more recently, in array-based genomic and transcriptomic analyses, their utility in discovery-driven proteomics has been essentially nonexistent. The general belief has been that the fixation procedure results in formaldehyde-induced inter- and intramolecular covalent cross-links that renders proteins intractable to downstream fractionation, digestion, and MS analysis (10).

Recently, three studies have been reported that demonstrate the potential of using FFPE samples in discovery-driven proteomics (11–13). While slight overall differences exist, two of these combined enzymatic digestion with liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis of proteins extracted from FFPE samples. In one study, this capability was demonstrated through fixation of a confluent culture of follicular B cell lymphoma cells (SUDHL-4) (11). Lysis buffer was added to 10 sections cut from the fixed sample to extract the protein complement of the cells, which was subsequently digested into peptides using either trypsin or glutamic-C endopeptidase (Glu-C). Analysis of the samples by LC-MS/MS resulted in the identification of 324 proteins from two or more unique peptides. The identification of important signaling proteins such as Raf-B, JAK1, STAT1, and protein kinase C (PKC) was confirmed by immunoblot analysis.

In a clinically relevant application, another study combined the use of laser capture microdissection with proteomic analysis to investigate prostate cancer (PCa) and benign prostatic hyperplasia (BPH) from cells captured from archived FFPE prostate tissue (12). Tryptic peptides were extracted from each of the histologically distinct cells and separately analyzed by LC-MS/MS, as illustrated in Figure 2. In total, 1156 and 702 unique proteins were identified from the PCa and BPH FFPE tissue extracts, respectively. Differences in protein abundance from the PCa and BPH cells were determined by comparing the number of unique peptides identified from a specific protein. A variety of prostate-related proteins, such as prostatic acid phosphatase (PAP), Raf kinase inhibitor protein (RKIP), and most notably prostate-specific antigen (PSA), were identified in both sample types by multiple peptides (Table 1). In contrast, growth differentiation factor 15 (GDF-15) was only identified by four unique peptides in PCa cells. This result is in agreement with a previous study that showed an up-regulation of GDF-15 in PCa cells compared with high-grade prostatic intraepithelial neoplasia (hPIN) cells (14).

These studies also evaluated the effect of FFPE processing versus more commonly used fresh/frozen cells and tissues, by comparing the proteomes from each sample source. Comparison of FFPE and fresh SUDHL-4 cells resulted in the identification of 324 and 512 proteins, respectively (11), while a comparison of 30,000 cells microdissected from FFPE and frozen mouse liver resulted in similar numbers of protein identifications from each sample (approximately 88% as many proteins could be identified from the FFPE sections as from frozen sections) (12). Besides the number of peptides and proteins identified, what was particularly encouraging in the analysis

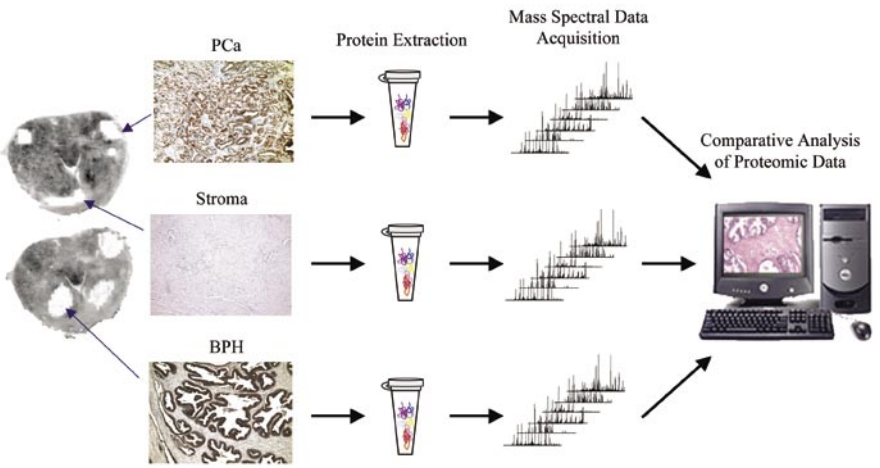

Figure 2. Characterization of formalin-fixed paraffin-embedded (FFPE) prostate tissue by liquid chromatography-tandem mass spectrometry (LC-MS/MS). (A) In this study (Reference 12), cells from regions of prostate cancer (PCa), stroma, and benign prostatic hyperplasia (BPH) were laser capture microdissected from the tissue. Peptides were extracted from each cell type and analyzed by LC-MS/MS.

of the FFPE and frozen tissue was the similarity in the base-peak chromatograms (BPC) and the tandem mass spectra obtained from both samples. The similarities in the BPC and the tandem mass spectra suggest that the population of proteins extracted from each sample type is comparable.

Taken together, these observations suggest that the vast archive of FFPE tissues, a heretofore unexplored clinical sample source, may provide a valuable new avenue for biomarker discovery. This new accessibility of FFPE tissues to proteomic research could catalyze a shift in the manner in which candidate biomarkers are identified for validation in biofluids (e.g., in cancer research). An alternate option to the present paradigm of analyzing biofluids to identify differentially abundant protein biomarkers related to disease may be to first analyze clinical cohorts of FFPE tumor tissues, rendering a subset of proteins proximal to the disease mechanism that can be validated via assay in the appropriate biofluid in expanded clinical cohorts.

This new paradigm has many potential advantages over standard biomarker identification/evaluation protocols. First, potential biomarkers discovered using MS approaches could be validated by immunohistochemistry within a section cut from the same tumor. Second, if the presence of the tumor-related proteins can be validated in a corresponding biofluid, the confidence in linking the origin of the proteins to the diseased tissue is greatly increased than if they were simply identifying the proteomic survey of the biofluid alone. While not discussed in great depth, a major problem in attempting to find disease-specific biomarkers in serum or plasma is that acute phase and inflammatory response proteins are typically up-regulated under many different conditions, adding to the complexity of the biofluid, but having limited utility for diagnostic purposes. Direct analysis of the tumor should also ameliorate the background differences associated with these proteins that are commonly observed in the analysis of serum or plasma from disease-stricken patients.

Table 1. Total Number of Peptides Identified

Protein	PCa	BPH	Stroma
PSA	10	16	1
PEBP	12	7	3
PAP	31	28	0
GDF-15	4	0	0

Comparative analysis of the proteomes showed that proteins such as prostate-specific antigen (PSA), prostatic acid phosphatase (PAP), and phosphatidylethanolamine binding protein (PEBP) were highly represented within both prostate cancer (PCa) and benign prostatic hyperplasia (BPH) cells, based on total peptide count. Growth-differentiation factor 15 (GDF-15), however, was observed exclusively with PCa cells.

Organ Perfusion

A clever strategy for eliminating the complexity of these biofluids and permitting the identification of proteins that may be shed or secreted from the diseased tissue was recently described by Koomen et al. (15). In this study, isolated beating hearts were dissected from rats and immediately perfused for approximately 5 min to remove blood from the organs. The hearts were then subjected to ischemic conditions and reperfused. The effluent from this perfusion was collected in various fractions that were subsequently analyzed by LC-MS/MS for protein identification and sodium dodecyl sulfate PAGE (SDS-PAGE) to verify the removal of background serum/plasma proteins. The SDS-PAGE results showed a dramatic decrease (i.e., at least 90%) in the level of albumin in the perfused samples collected after blood had been removed from the heart. Analysis of the perfusion effluent by LC-MS/MS resulted in the identification of 342 proteins. Approximately one-third of the proteins were cytoplasmic in origin, while another third were common resident serum/plasma proteins. Several biomarkers for myocardial injury such aspartate aminotransferase, creatine kinase MB, cardiac troponin I, cardiac troponin T, glycogen phosphorylase, and heart fatty acid binding protein (FABP) were identified in the effluent. Other potentially valuable markers of myocardial ischemia were also identified such as myotrophin, neuroprotective protein DJ-1, atriopeptigen, which is the precursor for atrionatriuretic peptides, and the AMP phosphoramidate-hydrolyzing enzyme HINT1.

Obviously this technology has some drawbacks in application to human patients, primarily in the ability to acquire samples as well as the cellular heterogeneity of organ systems. Accordingly, the investigators proposed some intriguing applications of this sampling technology particularly in the analysis of human xenograft tumors in encapsulated organs of athymic or immunodeficient mice (15). This technique should allow tumor-specific (human) and host response (mouse) protein biomarkers to be easily discriminated.

Functional Proteomic Assays

While not applicable to every type of study, using sample preparation methods to target a specific functional characteristic can often facilitate the discovery of a biomarker(s). These directed studies rely heavily on sample preparation techniques but do not require excessive MS data acquisition nor the need to burden the bioinformatic analysis to find the desired compounds among a complicated matrix. For instance, Dr. Benjamin Cravatt and coworkers have been leaders in developing activity-based protein profiling (ABPP) for the discovery of functional states specific to tumors (16). This technology uses chemical probes directed toward the active sites of specific classes of enzymes to detect functional changes in entire proteome samples. Recently this group teamed up with Dr. John Yates' group to complement their functional assay capabilities with MS identification of the enzymes that were detected.

In one such application, this effort used fluorophosphonate (FP)-based ABPP probes that target serine hydrolases, a diverse class of enzymes that have been implicated in cancer (17). In the first phase of the study, rhodamine-tagged FP-ABPP probes were added to 12 µg protein obtained from homogenized breast tumor tissue sections and then analyzed by SDS-PAGE with fluorescence detection, as shown in Figure 3A. The gel image revealed the profile of changes in serine hydrolase

activity across different breast tumors. While the image demonstrated different activity profiles from different tumors, direct identification of the enzymes linked to these changes was not possible using this analytical platform.

To redress this deficiency, they employed a biotinylated FP-ABPP probe (instead of rhodamine-tag) as shown in Figure 3B. The probe-labeled proteins were enriched using avidin-conjugated beads. After on-bead trypsin digestion, the peptide mixture was analyzed by multidimensional LC-MS/MS. Over 50 serine hydrolases were identified in the tumor proteomes using the biotin-FP ABPP probe. Only a few of the same activities were observed in the control reactions in which the probe was not added to the proteome sample. These enzymes included proteases, lipases, esterases, several proteasome subunits, and at least 15 uncharacterized hydrolases.

For quantitation, the spectral count (e.g., number of peptide identifications by MS/MS) for each hydrolase was averaged for the two different breast cancer samples. Several enzymes were identified that had at least a 3-fold change in activity among the breast cancer specimens. For example, fibroblast activation protein, KIAA1363 and platelet-activating factor acetylhydrolase 2 were elevated in estrogen receptor/progesterone receptor double negative [i.e., ER(-)/PR(-)] tumors compared with either ER(+)/PR(+) tumors or normal breast tissue. Other enzyme activities such as thrombin, dipeptidyl-peptidase IV, and hormone-sensitive lipase, however, were found to be higher in normal breast tissue. While this study focused on the identification of changes in serine hydrolase activity, a number of functional probes are presently available for hypothesis-driven analyses of functional changes in tumor cells.

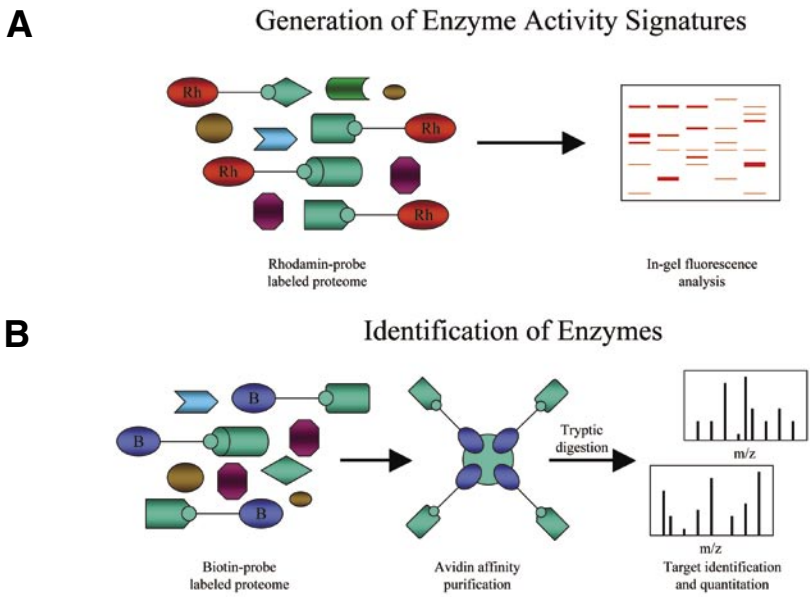

Figure 3. Combination of activity-based protein profiling (ABPP) and mass spectrometry (MS)-based proteomic profiling to identify specific activities within complex human samples. (A) A rhodamine (Rh)-tagged ABPP probe is used to isolate species that contribute a specific enzymatic activity from a proteome. The isolated proteins are then resolved by one-dimensional electrophoresis to provide enzyme activity signatures. (B) In the second phase, the same ABPP probe tagged with biotin (B) is then used in combination with avidin chromatography to extract the proteins giving rise to the specific activity. The extracted proteins are tryptically digested and analyzed by multidimensional liquid chromatography-MS, for identification and to estimate their respective levels. m/z, mass-to-charge ratio.

Cell-Based Assays

Cell-based assays are widely used in biological discovery (18); however, their combination with MS for biomarker discovery has been limited. In general, cell-based assays are used to determine the effect of a known compound on cells in culture or to attempt to find novel compounds within extracts that produce an effect. Unfortunately, there has not been much research showing a combination of cell-based assays and MS to discover biomarkers in biofluids. An excellent example of combining a cell-based assay with multidimensional fractionation and MS analysis was demonstrated by the discovery of a biomarker for interstitial cystitis (IC) (19). Interstitial cystitis (which affects approximately one million Americans) is a chronic, debilitating, bladder disorder that results in the thinning of the bladder epithelial cell lining (20). This disorder is typically diagnosed only by exclusion of other maladies followed by a bladder biopsy under general anesthesia. It was discovered in the 1990s that the urine of patients with IC contained a factor that inhibited bladder epithelial cell growth in culture. This antiproliferative factor (APF) was also found in the supernatant of epithelial cells in culture explanted from the bladders of IC patients, but not from healthy individuals (20). The true molecular nature (e.g., identity) of APF was not discovered, however, until 2004 (19).

To discover the identity of APF, the groups involved used several chromatographic steps and a cell-based assay to track APF through the various collected fractions, scored by measuring [^{3}H]-thymidine incorporation into cultured bladder epithelial cells (an indicator of inhibition of cell proliferation) as illustrated in Figure 4 (19). The fraction possessing the APF activity was graduated through the various chromatographic steps to increase its fold-enrichment, which was subsequently analyzed by LC-MS/MS. After a series of validation steps that included comparison of synthetically generated molecules with structures approximating the isolated molecule, APF was validated as a nine-residue sialoglycopeptide. Subsequent clinical validation demonstrated the unique nature of APF in that the transcript that gives rise to this peptide is solely present in the urine of patients with IC and not in that obtained from healthy-matched controls.

Considering the structure of APF, it is highly unlikely that it could have been identified in a global analysis comparing urine from IC and healthy patients using conventional proteomic strategies. Indeed, most global studies

focus on comparing quantitative changes in unmodified, tryptic peptides derived after proteome digestion. The very nature of APF, namely a glycosylated, nontryptic peptide, would have obviated its identification, particularly when one considers the current guidelines being established in the proteomic community (www.mcponline.org/misc/ParisReport.shtml). The discovery of the structure of APF required a large effort in sample preparation, but in turn required a single LC-MS/MS experiment. The bioinformatic burden on the structure determination was accordingly slight as well, as the tandem mass spectrum was interpreted de novo. Validation of the final structure also required minimal MS and bioinformatics support, but instead relied on techniques such as lectin chromatography and synthetic chemistry. Ultimately the successful identification of this important biomarker for IC depended on the cell-based assay to pinpoint the activity of APF in various chromatography fractions. The entire process is analogous to separating "the needle from the haystack" prior to MS identification.

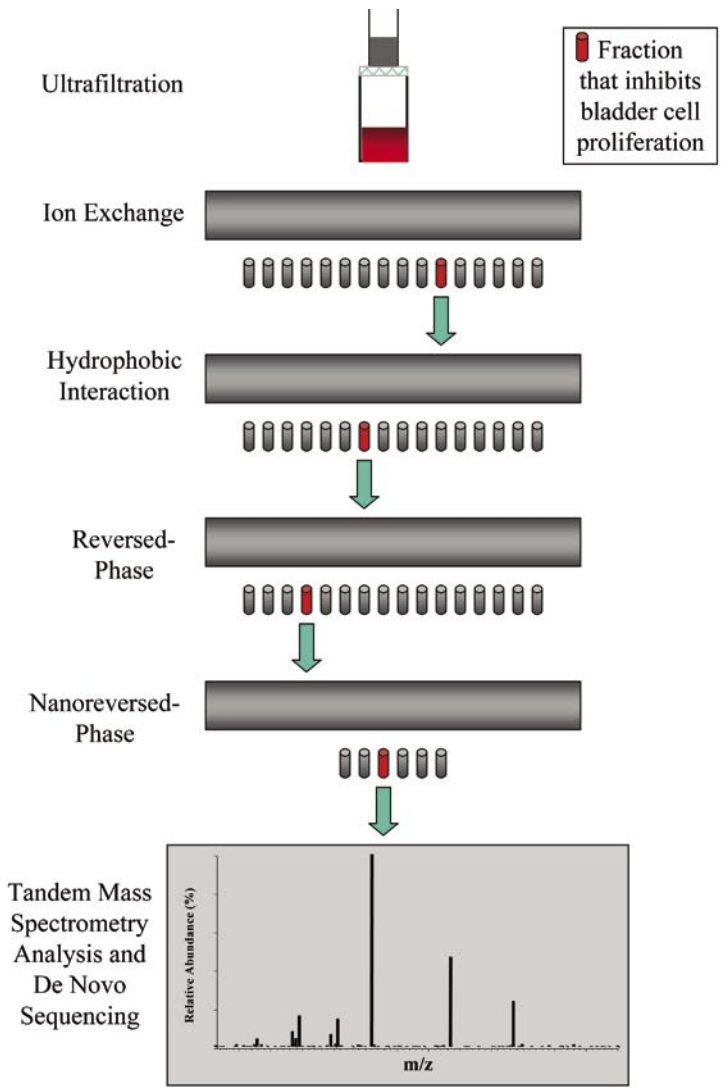

Figure 4. Incorporation of chromatography and a cell-based assay to aid in the identification of the antiproliferative factor (APF) for interstitial cystitis (IC). Fractions were collected during each chromatographic separation and tested for their ability to inhibit bladder epithelial cell proliferation in vitro. Fractions containing APF activity were then promoted to the next chromatographic step until the enrichment of APF was high enough to enable it to be identified by mass spectrometry and de novo sequencing. m/z, mass-to-charge ratio.

CONCLUSIONS

Proteomics currently exists in a discovery and survey era where MS, and its affiliate technologies, is focused on identification of candidate novel, biofluid-based biomarkers for various disorders. As shown through the variety of examples presented in this review, the challenges are immense, and there is often a trade-off between the amount of effort placed in sample preparation, the amount of MS data acquisition, and the need for bioinformatics analysis. While the investments in these technologies have yet to yield large numbers of validated biomarkers, they have allowed researchers to begin analyzing complex samples in ways not possible as recently as 5 years ago. The past couple of years have been witness to new and imaginative ways to analyze clinical samples that target the site of the disorder and may eventually yield validatable biomarkers with high sensitivity and specificity.

As the analytical technology applied to proteomic analysis of complex biological samples increases in sensitivity, there is an increasing burden placed on the standardization of sample collection and handling. Artifacts arising from nonspecific proteolysis and protein degradation are likely to be quite variable depending on details such as sampling-handling parameters. A recent study showing the ability to classify solid tumors based on the signature peptides produced by the action of exopeptidases postextraction (21), underscores the need for consistent sample handling prior to analysis. Even for samples such as FFPE tissues, the variability introduced by formalin-fixation time and length of storage are not yet well understood. The effects (and potential artifacts) introduced by sample acquisition, storage, and processing will need to be measured in a systematic manner within a well-controlled study.

Although the application of MS for protein biomarker investigations is arguably the most powerful discovery-driven methodology, the ultimate goal is to mobilize such discoveries to routine clinical application for in vitro diagnostic assays. Although these MS-based applications exist, namely the powerful assay

to screen newborns for inborn errors of metabolism (22), there are many hurdles to overcome in the translation of MS technology to the development of routine clinical diagnostic assays of protein biomarkers from complex matrices. In perhaps a more idealized biomarker developmental workflow, high affinity reagents, such as monoclonal antibodies, would be generated against lead biomarkers forthcoming from MS-based discovery-driven investigations. These reagents have a long and rigorously validated history of use in in vitro diagnostic applications and in large part are likely to contribute to a shorter trajectory to clinical adoption of newly discovered biomarkers, especially when the validation and regulatory due diligence that underpin such translations is considered. The production of an antibody with high affinity and specificity for an antigen, however, is never a guarantee. It will not be surprising if MS begins to play a larger role in biomarker validation as the analytical attributes (i.e., sensitivity, resolution, direct detection, etc.) that make it the premier technology for proteome characterization also enable it to interrogate specific species in complex mixtures (23).

ACKNOWLEDGMENTS

This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract no. NO1-CO-12400. By acceptance of this article, the publisher or recipient acknowledges the right of the U.S. Government to retain a non-exclusive, royalty-free license and to any copyright covering the article. 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.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

1. **Issaq, H.J., K.C. Chan, G.M. Janini, T.P. Conrads, and T.D. Veenstra.** 2005. Multidimensional separation of peptides for effective proteomic analysis. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. *817*:35-47.
2. **Conrads, T.P., B.L. Hood, E.F. Petricoin 3rd, L.A. Liotta, and T.D. Veenstra.** 2005. Cancer proteomics: many technologies, one goal. Expert Rev. Proteomics *2*:693-703.
3. **Hood, B.L., M. Zhou, K.C. Chan, D.A. Lucas, G.J. Kim, H.J. Issaq, T.D. Veenstra, and T.P. Conrads.** 2005. Investigation of the mouse serum proteome. J. Proteome Res. *4*:1561-1568.
4. **Wasinger, V.C., V.L. Locke, M.J. Raftery, M. Larance, D. Rothemund, A. Liew, I. Bate, and M. Guilhaus.** 2005. Two-dimensional liquid chromatography/tandem mass spectrometry analysis of Gradiflow fractionated native human plasma. Proteomics *5*:3397-3401.
5. **Veenstra, T.D., T.P. Conrads, B.L. Hood, A.M. Avellino, E.G. Ellenbogen, and R.S. Morrison.** 2005. Biomarkers: mining the biofluid proteome. Mol. Cell. Proteomics *4*:409-418.
6. **Pieper, R., C.L. Gatlin, A.M. McGrath, A.J. Makusky, M. Mondal, M. Seonarain, E. Field, C.R. Schatz, et al.** 2004. Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots. Proteomics *4*:1159-1174.
7. **Pieper, R., C.L. Gatlin, A.J. Makusky, P.S. Russo, C.R. Schatz, S.S. Miller, Q. Su, A.M. McGrath, et al.** 2003. Human serum proteome: display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins. Proteomics *3*:1345-1364.
8. **Shen, Y., J.M. Jacobs, D.G. Camp 2nd, R. Fang, R.J. Moore, R.D. Smith, W. Xiao, R.W. Davis, and R.G. Tompkins.** 2004. Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome. Anal. Chem. *76*:1134-1144.
9. **Spahr, C.S., M.T. Davis, M.D. McGinley, J.H. Robinson, E.J. Bures, J. Beierle, J. Mort, P.L. Courchesne, et al.** 2001. Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfractionated tryptic digest. Proteomics *1*:93-107.
10. **Ahram, M., M.J. Flaig, J.W. Gillespie, P.H. Duray, W.M. Linehan, D.K. Ornstein, S. Niu, Y. Zhao, et al.** 2003. Evaluation of ethanol-fixed, paraffin-embedded tissues for proteomic applications. Proteomics *3*:413-421.
11. **Crockett, D.K., Z. Lin, C.P. Vaughn, M.S. Lim, and K.S. Elenitoba-Johnson.** 2005. Identification of proteins from formalin-fixed paraffin-embedded cells by LC-MS/MS. Lab. Invest. *85*:1405-1415.
12. **Hood, B.L., M.M. Darfler, T.G. Guiel, B. Furusato, D.A. Lucas, B.R. Ringeisen, I.A. Sesterhenn, T.P. Conrads, et al.** 2005. Proteomic analysis of formalin-fixed prostate cancer tissue. Mol. Cell. Proteomics *4*:1741-1753.
13. **Palmer-Toy, D.E., B. Krastins, D.A. Sarracino, J.B. Nadol, Jr., and S.N. Merchant.** 2005. Efficient method for the proteomic analysis of fixed and embedded tissues. J. Proteome Res. *4*:2404-2411.
14. **Cheung, P.K., B. Woolcock, H. Adomat, M. Sutcliffe, T.C. Bainbridge, E.C. Jones, D. Webber, T. Kinahan, et al.** 2004. Protein profiling of microdissected prostate tissue links growth differentiation factor 15 to prostate carcinogenesis. Cancer Res. *64*:5929-5933.
15. **Koomen, J.M., C.R. Wilson, C.R. Guthrie, M.J. Androlewicz, R. Kobayashi, and H. Taegtmeyer.** 2006. Proteome analysis of isolated perfused organ effluent as a novel model for protein biomarker discovery. J. Proteome Res. *5*:177-182.
16. **Speers, A.E. and B.F. Cravatt.** 2004. Profiling enzyme activities in vivo using click chemistry methods. Chem. Biol. *11*:535-546.
17. **Jessani, N., S. Niessen, B.Q. Wei, M. Nicolau, M. Humphrey, Y. Ji, W. Han, D.Y. Noh, et al.** 2005. A streamlined platform for high-content functional proteomics of primary human specimens. Nat. Methods *2*:691-697.
18. **Indelicato, S.R., S.L. Bradshaw, J.W. Chapman, and S.H. Weiner.** 2005. Evaluation of standard and state of the art analytical technology-bioassays. Dev. Biol. (Basel) *122*:103-114.
19. **Keay, S., Z. Szekely, T.P. Conrads, T.D. Veenstra, J. Barchi, C. Zhang, K. Koch, and C. Michjeda.** 2004. An antiproliferative factor from interstitial cystitis patients is a frizzled 8 protein-related sialoglycopeptide. Proc. Natl. Acad. Sci. USA *101*:11803-11808.
20. **Warren, J.W. and S.K. Keay.** 2002. Interstitial cystitis. Curr. Opin. Urol. *12*:69-74.
21. **Villanueva, J., D.R. Shaffer, J. Philip, C.A. Chaparro, H. Erdjument-Bromage, A.B. Olshen, M. Fleisher, H. Lilja, et al.** 2006. Differential exoprotease activities confer tumor-specific serum peptidome patterns. J. Clin. Invest. *116*:271-284.
22. **Chace, D.H., T.A. Kalas, and E.W. Naylor.** 2003. The application of tandem mass spectrometry to neonatal screening for inherited disorders of intermediary metabolism. Annu. Rev. Genomics Hum. Genet. *3*:17-45.
23. **Kuhn, E., J. Wu, J. Karl, H. Liao, W. Zolg, and B. Guild.** 2004. Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and 13C-labeled peptide standards. Proteomics *4*:1175-1186.

Address correspondence to:

Timothy D. Veenstra
Laboratory of Proteomics and Analytical Technologies
National Cancer Institute at Frederick
SAIC-Frederick, Inc.
P.O. Box B
Frederick, MD 21702, USA
e-mail: veenstra@ncifcrf.gov