Biomarkers: Mining the Biofluid Proteome*

Timothy D. Veenstra‡§, Thomas P. Conrads‡, Brian L. Hood‡, Anthony M. Avellino¶, Richard G. Ellenbogen¶, and Richard S. Morrison¶

Proteomics has brought with it the hope of identifying novel biomarkers for diseases such as cancer. This hope is built on the ability of proteomic technologies, such as mass spectrometry (MS), to identify hundreds of proteins in complex biofluids such as plasma and serum. There are many factors that make this research very challenging beginning with the lack of standardization of sample collection and continuing through the entire analytical process. Fortunately the advances made in the characterization of biofluids using proteomic techniques have been rapid and suggest that these mainly discovery driven approaches will lead to the development of highly specific platforms for diagnosing diseases and monitoring responses to different treatments in the near future. *Molecular & Cellular Proteomics* 4:409–418, 2005.

Diagnosis represents the first line of defense for any medical condition. There are many different ways that a condition can be recognized including those that manifest visually (i.e. a cut or laceration), those that rely on an indirect diagnosis (i.e. measurement of prostate-specific antigen (PSA) levels), and those that are diagnosed directly (i.e. tumor biopsy). For many, debilitating health conditions such as cancer and heart disease that affect populations worldwide, early diagnosis is a key predictor of successful treatment and/or patient survival. Medical science has discovered many drugs, surgical methods, and other useful therapies to treat patients with such conditions; however, too often diagnoses are made too late for these remedies to have any useful effect on the patient’s survival.

The diagnostic potential of biofluids such as serum and plasma is underscored by the current expenditure of research effort and money on their investigation. An example of the international efforts being applied to characterize plasma and serum is the Plasma Proteome Project (PPP) that is organized by the Human Proteome Organization (HUPO) (1). The PPP incorporates a pilot project to identify the optimal methods of collecting and handling biofluid specimens, as well as standardizing analytical methods such as fractionation, mass spectral analysis, search engines, and databases to identify proteins within plasma and serum. The goal of this project is to not only recognize the best ways of analyzing complex biological specimens, but also to standardize methods and data so that large-scale clinical and epidemiological studies conducted around the world can be properly compared.

Even with the analytical tools that have been recently developed, the discovery of biomarkers in biological fluids is still an enormous challenge. Consider the case of trying to identify a protein that is produced by the presence of a malignant tumor. In a study using mammography to identify breast tumors in Swedish and Canadian women, tumors were identified at an average size of 21 and 16 mm, respectively (2). These dimensions would correspond to tumors with approximate volumes between 4,000 and 9,000 mm³. Considering the average woman occupies a volume of roughly $1.5 \times 10^8$ mm³, the tumor would represent less than 0.006% of the total body volume. Additionally, considering that many modern technologies attempt to identify tumor-specific biomarkers in blood where over 60,000 miles of blood vessels make up the circulatory system and considering that blood samples are routinely obtained from a site that is distinct from the location of the tumor, it is easy to envisage why identifying a protein that is specific to the tumor is so challenging.

The ability to analyze biofluids at the proteome level becomes more critical when one considers the available options. Biofluids such as plasma, serum, urine, and cerebrospinal fluid do not have a corresponding genome or transcriptome that enables gene expression to be measured (3). Biofluids are believed to reflect the ensemble of tissues present within a patient. While it is known that many of the highest abundance proteins in plasma are synthesized in the liver (4), their protein concentration in plasma and mRNA abundance in the liver correlate poorly (5). If this situation exists for highly abundant proteins, the situation would be expected to be worse for low-abundance plasma proteins expressed within smaller tissues. Because DNA- or RNA-based diagnostics are not applicable to biofluids, proteomics (as well as metabolomics) remains one of the few options for identifying biomarkers in biofluids. While the challenge is great, the progress...
made in just the past few years in proteomic analyses of biofluids provides plenty of optimism. In particular, proteomics has greatly increased our knowledge of the protein content of serum, plasma, and cerebrospinal fluid.

There have been two routes that investigators have followed in the proteomic analyses of biofluids with the goal of finding differences in the content of samples obtained from patients with specific disease states compared with healthy, matched controls (6). These approaches can be loosely categorized into biomarker discovery and diagnostic development (Fig. 1). In the biomarker discovery approach, multidimensional fractionation is used in combination with MS/MS analysis to identify proteins that are unique or highly abundant in samples obtained from patients with specific disease states compared with healthy, matched controls. In the diagnostic proteomics pathway, biofluid samples from healthy and disease-affected individuals are applied to protein chips that are modified with a specific chromatographic resin. After a series of washing and binding steps a proteomic pattern of the proteins that are retained is acquired using MS. The source of the biofluid (i.e. disease-affected or healthy patient or unclassifiable) is then determined using bioinformatic algorithms that search for differences in peak intensities between the sample sets. The diagnostic proteomic method does not rely on the actual identification of the protein(s) through which the diagnosis is determined, although identification of the selected peaks can be pursued.
rectly diagnose patients in blind validation studies, it lacks the capability of directly identifying specific proteins within the peaks that indicate the source of the biofluids. In cases where the diagnostic protein peaks have been identified, their relationship to the disease condition is not clear because the proteins comprising the peak are not identified (10).

DEVELOPMENT OF TECHNOLOGY TO CHARACTERIZE BIOFLUID PROTEOMES

Identification of biomarkers in biofluids most often takes a nonbiased approach in which the aim is to characterize as many possible species as possible. The identities of the proteins identified within various samples are then compared with the hopes of finding a protein (or proteins) that is unique to a sample obtained from patients with a particular disease condition. While this methodology appears to provide a straightforward analytical approach, significant development has been, and continues to be, required to make this technique amenable to the analysis of biofluids, in particular serum and plasma.

Plasma and serum are dual-edged swords when it comes to proteome analysis by MS. On the positive side, they have a very high protein concentration. On the negative side, only 22 proteins account for 99% of their protein content. These proteins, which include albumin, transferrins, immunoglobulins, and complement factors, have all been well characterized as blood proteins. What are of interest are the remaining 1%, which is made up of lower abundance circulatory proteins as well as proteins that are shed or excreted by not only live cells, but also apoptotic and necrotic cells. This division between high- and low-abundance proteins in serum and plasma is believed to span 8–10 orders of magnitude in protein concentration, and this estimate may prove to be very conservative. This range is much greater than the typical two orders of magnitude that constrains most measurements by MS.

Prior to conducting comparative analyses to identify disease-specific biomarkers, it had to be shown that MS-based proteomic technology was capable of providing the coverage required to identify hundreds of proteins in biofluids. Fittingly, one of the first large-scale studies that showed the ability to identify hundreds of proteins within plasma incorporated two-dimensional (2D)-PAGE with MS identification of the visualized protein spots (Fig. 2) (11). Due to the large dynamic range of protein concentrations found in plasma, the entire plasma sample had to be prefractionated prior to 2D-PAGE separation. Plasma was initially immunodepleted to remove the most abundant serum proteins (i.e. albumin, haptoglobins, transferrins, transthyretin, α-1-antitrypsin, α-1-acid glycoprotein, hemopexin, and α-2-macroglobulin) and then fractionated by sequential anion-exchange and size-exclusion chromatography to create a total of 74 unique fractions that were separated on individual 2D-PAGE gels. Visualization of the proteins with Coomassie Brilliant Blue G-250 resulted in ~20,000 spots that were registered by image processing. After the redundant spots were removed, about 3,700 unique protein spots were subsequently selected for MS analysis, of which 1,800 were successfully identified. These 1,800 identifications corresponded to 350 unique proteins. Almost 39% of the proteins identified were known to be localized within the circulatory system, while another 9% were characterized as extracellular matrix proteins or proteins that are secreted into biofluids other than blood. Several proteins, such as IL-6, metallothionein II, cathepsins, and peptide hormones, known to be present in serum at a concentration of less than 10 ng/ml, were identified. While the entire analysis is quite labo-
Biomarkers: Mining the Biofluid Proteome

rious, this strategy has considerable advantages in that it allows quantitative measurements of proteins in comparative samples and provides a visual record of the number of different isoforms present for a particular protein.

The same group that profiled plasma using 2D-PAGE coupled to MS employed a similar strategy to characterize the urinary proteome (12). Urine provides an alternative to blood as a potential source of disease biomarkers, and its proteome is known to change as a result of diseases, particularly those affecting the kidney. In addition, the urinary proteome (and probably to a greater extent, its metabolome) is affected by drug treatments, allowing drug metabolism to be assayed through this complex biofluid. The identification of biomarkers in urine is particularly attractive as its collection is minimally obtrusive to the patient. A prefractionation strategy was devised to separate proteins with molecular masses less than 30 kDa from larger, more-abundant proteins. The high-molecular-mass fraction was then immunodepleted of albumin and immunoglobulin G (IgG) prior to separating all of the fractions via 2D-PAGE. A total of almost 1,400 distinct spots were resolved on the gels. Of these, 420 were identified by either MALDI-TOF peptide mass fingerprinting or LC-ESI MS/MS.

A total of almost 1,400 distinct spots were resolved on the gels. Of these, 420 were identified by either MALDI-TOF peptide mass fingerprinting or LC-ESI MS/MS. These identified spots originated from 150 unique proteins, of which ~50 had been described as classical plasma proteins that are also known to be relatively abundant in urine.

While 2D-PAGE analysis of biofluids was being conducted, investigators were also developing nongel-based (referred to here as solution-based) methods to identify plasma and serum proteins. The earliest study utilized immunoglobulin depletion, multidimensional chromatography, and MS to identify serum proteins (13). Serum that had been depleted of immunoglobulins was digested with trypsin and the resulting peptides fractionated by strong cation exchange (SCX) chromatography. Sixty fractions collected from the SCX column were analyzed by microcapillary reversed-phase HPLC coupled on-line with MS/MS. The net result was the identification of 490 proteins in serum, including PSA, which was believed to be present at a concentration of about 1 pg/ml in the serum sample used in this study.

This study spurred many other solution-based studies aimed at maximizing the coverage of serum or plasma proteins that could be identified by combining peptide fractionation and MS. Two recent studies utilizing SCX chromatography with LC-MS/MS culminated in the identification of ~1,500 proteins in tryptically digested serum (14) or plasma (15) samples that were not previously depleted of high-abundance proteins such as albumin. Other studies have taken approaches aimed at characterizing specific fractions of proteins within serum. A simple methodology was developed to deplete serum of high-molecular-mass proteins by size partitioning through the use of centrifugal ultrafilters (16, 17). This study was based on the hypothesis that the low molecular weight (LMW) region of serum potentially contained an archive of diagnostic information. This hypothesis was predicated on the results of several proteomic profiling studies that identified fragments of large proteins as being diagnostic for diseases such as ovarian cancer (18) and hepatocellular carcinoma (19). These profiling studies typically attribute the generation of these fragments to the proteolytic action of the tumor on high-abundance serum/plasma proteins. An initial study of a tryptic digestate of the LMW serum proteome by microcapillary reversed-phase LC coupled on-line with MS/MS resulted in the identification of 340 unique proteins, with no peptides originating from albumin being observed, demonstrating the efficacy of this technique for removing such high-abundant proteins (16). Indeed, many of the proteins identified in this study originated from proteins of very high molecular mass (>150 kDa) even though a 30-kDa molecular mass cutoff membrane was used to isolate the LMW proteome.

**INVESTIGATION OF SUB-BIOFLUID PROTEOMES**

While many proteomic studies use techniques to deplete serum and plasma of high-abundance proteins prior to either gel- or solution-based analysis, recent studies have shown that these practices may concomitantly remove potentially important diagnostic information. Albumin, for instance, is known to function as a carrier and transporter of proteins within the blood and binds physiologically important species such as hormones, cytokines, and lipoproteins (20, 21). Analogous to subcellular fractionation, Zhou et al. (22) used a “sub-biofluid” fractionation procedure in which a series of antibodies against albumin, IgA, IgG, IgM, apolipoprotein, and transferrins were used to specifically capture these proteins as well as proteins associated with each of these high-abundant serum proteins. A total of 209 unique proteins were found bound to these highly abundant proteins, but of greater interest 12 proteins that are currently utilized as clinical biomarkers were identified (Table I). These proteins include PSA, pregnancy plasma protein A, meninigioma-expressed antigen, and dihydropteronine reductase. While these aforementioned proteins have been identified in previous serum analysis studies (11), other clinically relevant biomarkers such as bone morphogenetic protein 3b, prostate transglutaminase, paraneoplastic antigen MA1, glycosylasparaginase, coagulation factor VII precursor, ryanodine receptor 2, and acid sphingomyelinase-like phosphodiesterase 3a have only been identified in this study bound to high-abundance proteins such as albumin, IgA, and apolipoprotein, but not in other global serum or plasma proteomic analyses. This study was the first to show that a potential archive of diagnostic information may be bound to large, high-abundant circulatory proteins preventing them from being rapidly cleared through the action of the renal system.

Another group used a sub-biofluid fractionation strategy to identify proteins within exosomes in urine (23). Exosomes are membrane vesicles that originate as internal vesicles of cells that are excreted into the extracellular space. Differential cen-
trifugation was used to isolate urinary exosomes from healthy subjects. The proteins extracted from these vesicles were digested into peptides that were subsequently analyzed using LC-MS/MS, resulting in the identification of almost 300 proteins. Proteins known to be involved in a variety of renal and systemic diseases, such as autosomal dominant polycystic kidney disease, Gitelman syndrome, Bartter syndrome, autosomal recessive syndrome of osteopetrosis with renal tubular acidosis, and familial renal hypomagnesemia were identified as components of urinary exosomes. This urinary study, as well as the serum interactome study presented above, suggests that sub-biofluid fractionation strategies may be a valuable tool in the enrichment of low-abundance proteins in blood and urine.

### QUANTITATIVE COMPARISON OF BIOFLUIDS

While quantitative approaches for conducting comparative proteomic analyses of samples such as cell and tissue lysates have been widely applied (24), comparative proteomics of biofluids has not been as widely disseminated.

A 2D-PAGE approach was recently applied to compare plasma samples obtained from patients with severe acute respiratory syndrome (SARS) and healthy individuals (25). Twenty-two plasma samples from four different SARS patients were separated by 2D-PAGE using a narrow-range IPG strip (pH 4–7), and the resulting profiles were compared with those obtained from six healthy plasma samples. Seven proteins were exclusively present in the 22 SARS samples. Eight additional spots were up-regulated in all 22 SARS patients compared with the healthy controls. Many of the proteins up-regulated in plasma from SARS patients can be classified as acute-phase proteins that are produced as a consequence of serial cascades initiated by the SARS-coronavirus infection. Interestingly, the intracellular, antioxidant protein peroxiredoxin II was found to be up-regulated in all of the 22 SARS plasma samples. In a separate validation study, peroxiredoxin II was found to be up-regulated in all of the 22 SARS patients, suggesting that peroxiredoxin II may function as a useful serum biomarker for SARS infection.

A recent study using multidimensional fractionation combined with MS analysis to analyze synovial fluid (SF) was designed to identify candidate protein biomarkers of rheumatoid arthritis (RA) that can predict which patients will develop erosive, disabling disease (26). Synovial fluid is a thick, straw-colored, lubricating compound found in small amounts in joints, bursae, and tendon sheaths (27). In this study, SF was obtained from five patients with erosive RA and five with nonerosive RA. After removal of albumin and IgG, followed by size-exclusion chromatography to enrich for the LMW (i.e. <40 kDa) protein fraction, the SF proteins were tryptically digested and each sample was analyzed by multidimensional chromatography coupled with MS/MS. This procedure generated lists of peptides that were identified in the SF digestate from each of the patients. Potentially interesting biomarkers within the SF from patients with erosive RA were determined by measuring the number of peptides identified from a unique protein compared with the number identified within the SF from the individuals with nonerosive RA, as well as measuring the ion current for each peptide. From these calculations, a total of 33 candidate biomarkers for erosive RA were identified from a total of 418 observed proteins. Among the proteins that were elevated in the SF of patients with erosive RA were C-reactive protein (CRP) and six members of the S100 protein family of calcium-binding proteins. Significantly, levels of CRP, S100A8 (calgranulin A), S100A9 (calgranulin B), and S100A12 (calgranulin C) proteins were also elevated in the serum of patients with erosive disease compared with patients with nonerosive RA or healthy individuals. What set this study apart from many others that use similar technology to identify candidate biomarkers was the subsequent validation of the proteomic results. Pools of serum samples were obtained from a separate group of patients with erosive-RA, nonerosive RA, and healthy, matched controls, and the levels of CRP, S100A8, S100A9, and S100A12 were measured using

### Table I

| Protein                                   | Association | Marker specificity                  |
|-------------------------------------------|-------------|-------------------------------------|
| Glycosylasparaginase                      | Albumin     | I-cell disease                      |
| Paraneoplastic antigen MA1                | IgA         | Paraneoplastic neurological syndrome |
| Meningioma-expressed antigen 6/11         | Albumin     | Meningioma                          |
| Dihydropyrimidine reductase               | Apolipoprotein | Tetrahydrobiopterin deficiency      |
| Coagulation factor VII precursor          | Albumin     | Liver cirrhosis                     |
| ASM-like phosphodiesterase 3a            | Apolipoprotein | Bladder tumors                     |
| Prostate transglutaminase                 | IgA, IgG    | Prostate cancer                     |
| Pregnancy-plasma protein-A               | IgG, IgM    | Down syndrome                       |
| Hsc70-interacting protein                | Albumin     | Cell growth                         |
| Ryanodine receptor 2                     | IgA         | Diabetes                            |
| Bone morphogenetic protein 3b            | Albumin, IgG | Osteoarthropy退          |
| PSA                                       |             | Prostate cancer                     |

...
either multiple reaction monitoring (MRM) MS using isotope-labeled internal standards to quantitate peptides originating from the four proteins of interest or, as in the case of CRP, also by immunoassay. In the measurements obtained by MRM, nonerosive RA samples showed CRP levels ranging from 2- to 12-fold higher than in serum obtained from healthy controls; however, patients with erosive RA showed CRP levels 47- to 142-fold higher than in healthy control samples. While the measurements of CRP by immunoassay did not give the same absolute magnitude of change, they unequivocally confirmed these differences observed by MRM. Similar results were observed for MRM analysis of S100A8, S100A9, and S100A12 using internal isotope-labeled standards. The relative abundance of all three proteins was marginally higher (i.e. 0.7- to 5.8-fold) in three pools of serum samples obtained from individuals with nonerosive RA compared with serum from matched controls. Patients with erosive RA, however, showed substantially higher levels (ranging from 3- to 111-fold) of these proteins than in healthy control samples. While admittedly the study is limited by the sample size that was used and the resource- and labor-intensive nature of the MS approach, nonetheless it demonstrates the efficacy of using global MS analysis of biofluids for the discovery and subsequent validation of clinically relevant sets of disease biomarkers. Now that reasonable targets for erosive-RA diagnosis have been discovered, it is easy to envision using more targeted approaches, such as immunoassays, to evaluate a much larger number of patient samples for further verification of the candidate biomarkers.

IDENTIFICATION OF BIOMARKERS IN CEREBROSPINAL FLUID

Cerebrospinal fluid (CSF) is a clear liquid produced in the ventricles of the brain that surrounds and protects the central nervous system from trauma. CSF directly contacts the extracellular space of the brain, making it a unique medium by which to detect biochemical changes in the central nervous system. Recent evidence suggests that CSF is also involved in the clearance of toxic molecules from the interstitial space of the brain and changes in CSF physiology may be associated with aging and the development of neurodegenerative disease (28–30). It has been hypothesized that alterations in the protein composition of CSF may reflect abnormalities in protein expression associated with trauma, neurodegenerative disorders, multiple sclerosis, and hydrocephalus (31–37). The comprehensive analysis of CSF might provide: 1) improved diagnostic markers for nervous system diseases, 2) a better understanding of the pathological processes involved in nervous system trauma and disease, and 3) novel therapeutic targets for treating central nervous system pathologies.

A variety of proteomic approaches have recently been used to characterize the total protein composition and the presence of posttranslationally modified proteins in CSF. The most common approach involves the use of 2D-PAGE fractionation followed by MS identification of the resolved protein spots. As with other biofluids, multidimensional fractionation combined with MS/MS has been applied more recently. Lumbar and ventricular CSF has been analyzed from patients expressing a diverse array of neurological diseases. CSF from multiple sclerosis patients was resolved into 430 spots corresponding to 61 distinct proteins (38). Four proteins not previously described in CSF were identified including CRTAC-IB (cartilage acidic protein), tetranection (a plasminogen-binding protein), SPARC-like protein (a calcium-binding cell-signaling glycoprotein), and autotaxin t (a phosphodiesterase). It was not determined if these proteins were causally related to multiple sclerosis. Analysis of CSF from 10 neurologically normal, elderly subjects resulted in the identification of 249 proteins. Interestingly, 38% of the proteins were unique to individual subjects and only 6% were identified in the CSF from all 10 patients. While this might represent considerable subject-to-subject variability, differences in post-mortem delay and cause of death may have influenced the composition of these CSF samples. Alternatively, if larger numbers of proteins were identified it is conceivable that the variability would be minimized.

Differences in the composition of CSF were also observed between Alzheimer’s disease (AD) patients and controls. The levels of eight proteins and their isoforms, including apolipoprotein A1, apolipoprotein E, apolipoprotein J, β-trace, retinol-binding protein, kininogen, α-1 antitrypsin, cell-cycle progression 8 protein, and α-1β glycoprotein were significantly altered in the CSF of AD patients compared with controls (39). Characterization of CSF is also proving useful in the area of neuro-oncology. The N-myc oncoprotein and LMW caldesmon were identified as candidate tumor-related proteins in the CSF from patients with primary brain tumors (40). These proteins are currently being evaluated for their ability to predict patient survival and response to chemotherapy. MS-based protein-mapping efforts are also being used to compare the proteomes obtained from the CSF of controls and low back pain patients to characterize differentially expressed proteins and to elucidate biological markers for idiopathic low back pain (41). These studies demonstrate the diversity of neurological problems that could benefit from the identification of well-defined biological markers that associate with disease onset, progression, length of survival, and response to therapy.

Our own efforts have been directed toward characterizing CSF from patients with pediatric brain tumors. The goal of the investigation is to identify proteins that are associated with tumor growth and linked to the development of hydrocephalus that frequently manifests in these patients. We recently obtained and analyzed CSF samples from the same patient before and after surgical resection for a choroid plexus carcinoma. Multidimensional fractionation combined with MS/MS was used to identify over 1,000 proteins within each of the samples, with almost 400 of these being found within both pre- and postoperative CSF. Not only were a number of...
brain-specific proteins (i.e. amyloid proteins) observed in this study, but numerous low-abundance species such as kallikreins, cytokines, and chemokines were also identified. Several proteins that modulate angiogenesis, including kininogen (42), tetranectin (43), secretoneurin (44), and netrin-1 (45) were identified in the CSF when the tumor was present, but were absent or poorly expressed following tumor resection. These findings demonstrate that the protein content of CSF can potentially be used as a diagnostic for the overall profile of the brain proteome, and a straightforward fractionation strategy combined with MS/MS is entirely capable of characterizing a significant fraction of the CSF proteome. Moreover, characterizing the CSF proteome greatly increases our knowledge of the protein content of this clinically important biofluid.

IDENTIFICATION AND VALIDATION OF A URINARY BIOMARKER FOR INTERSTITIAL CYSTITIS

While there has been a major emphasis on using technologies such as MS to gather copious amounts of data with the hope of finding validatable biomarkers within a complex milieu of background proteins, the contribution of other techniques that help to target this search should not be ignored. As illustrated in the above studies, any biofluid is tremendously complex. Any technique that can point to the fraction of a biofluid that contains a relevant biomarker is invaluable. An excellent example of combining various techniques to pinpoint, identify, and subsequently validate a relevant biomarker is the discovery of an antiproliferative factor (APF) that is present in urine obtained from patients suffering from interstitial cystitis (IC) (46). While not commonly known, yet as prevalent as multiple sclerosis, IC is a debilitating, chronic bladder disorder that is characterized by thinning and ulceration of the bladder epithelial lining (47). There is presently no definitive diagnostic marker for IC; it is usually identified through the exclusion of other conditions after which point a bladder biopsy is performed to provide a definitive diagnosis.

While APF had been described based on its function approximately 9 years ago, its structure had remained elusive. A critical characteristic of urine obtained from patients with IC was instrumental in the eventual identification of APF. Application of urine from IC patients has an antiproliferative effect on bladder epithelial cells in culture. This cell-based assay allowed the investigators to use a series of chromatographic steps to isolate a specific fraction that contained this antiproliferative activity and hence identify APF. After a specific fraction containing a high concentration of antiproliferative activity was identified, it was analyzed by LC-MS/MS (Fig. 3A). Database analysis of the resultant spectra did not provide an immediate identification of APF; however, four putative structures were determined through de novo sequencing of the MS/MS spectra. Each of these putative structures indicated that APF was a glycosylated nonapeptide. To determine the correct nonapeptide, the proposed amino acid sequences were utilized to conduct a BLAST search against the human nonredundant protein database. Only one of these sequences showed a complete match to a known protein. This sequence had 100% homology to a peptide contained within the sixth transmembrane domain of frizzled-8, a Wnt ligand receptor.

Identification of the structure of the glycosyl moiety was significantly more challenging. The tandem mass spectrum suggested the presence of an O-linked N-acetylgalactosamine (GlcNAc) or N-acetylgalactosamine (GalNAc), followed by an unknown hexose group, culminating with a terminal sialic acid (SA) residue. The SA group was confirmed and the hexose group identified as a galactose (Gal) via lectin binding studies. Unfortunately, these studies could not sufficiently differentiate between a GlcNAc and GalNAc residue. To definitively determine this moiety, two versions of APF were synthesized; containing either a GlcNAc or GalNAc residue. These synthetic glycopeptides were utilized to spike separate aliquots of as isolated APF. These mixtures were separately analyzed by LC-MS/MS. The results showed that synthetic APF modified with GalNAc co-eluted with as isolated APF while the GlcNAc derivative did not. This final experiment confirmed the putative structure of APF as Thr-Val-Pro-Ala-Ala-Val-Val-Ala with a GalNAc-Gal-SA modification O-linked to the N-terminal Thr residue. This proposed structure of APF was further confirmed through demonstrating that a synthetic version of APF elicits identical results to series of biological and analytical tests as compared with as isolated APF. Most importantly, mRNA transcripts that bound a cDNA probe based on the nonapeptide sequence of APF were found to be uniquely present in bladder epithelial cells harvested from IC patients (Fig. 3B). The discovery of the structure of APF now opens up avenues to design affinity probes to test for the presence of this molecule in the urine of patients with IC to provide a definitive, diagnostic for this debilitating condition. The identification of the structure of APF also provides a great example of how concerted use of a variety of analytical techniques aid in biomarker discovery. Obviously not every study will be able to make use of all of the analytical tools described for the identification of APF; however, any method that helps limit the complexity of candidate biomarkers that must be characterized in order to identify a real biomarker is invaluable.

CONCLUSIONS

Proteomics has reached a precarious position in its brief history. Much of the excitement that was initially directed toward the human genome project has now been translated to proteomics and with it the hope of finding more effective biomarkers for human diseases. Fortunately the progress that has been made toward fulfilling this goal is remarkable. A literature search demonstrates the rapid and significant strides made in analyzing biofluids using MS-based proteomics technologies. It was only last year that studies showing the ability to identify hundreds of proteins in samples such as
serum, plasma, urine, and CSF were being published. Groups are presently using this technology to perform comparative analyses for finding potential biomarkers for a variety of different conditions. Obviously only a few of the potential candidates that are identified in these global studies will turn out to be clinically useful biomarkers, but the technological breakthroughs that have been made in the past few years allows the initial discovery of these candidates to be conducted in a more comprehensive manner. So why is proteomics in a precarious position? Simply because it is critical for candidate biomarkers discovered with this technology to find their way into proper clinical trials so that their efficacy as diseasespecific markers can be validated. If the necessary effort is not directed toward investing the resources to translate such discoveries to the clinic, then proteomics may languish as a technology that creates large data repositories of minimal use. The burden will be on proteomic programs to ensure that the quality of data produced in biomarker discovery studies is of sufficient quality that a reasonable percentage of the candidate markers produced have high disease specificity and are actually clinically validated.

* 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 and by Grants NS35533, NS047670, and NS42699 from the National Institutes of Health (to R. S. M.). By acceptance of this article, the publisher or recipient acknowledges the right of the U.S. Government to retain a nonexclusive, 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
13. Adkins, J. N., Varnum, S. M., Auberry, K. J., Moore, R. J., Angell, N. H., Pieper, R., Gatlin, C. L., Makusky, A. J., Russo, P. S., Schatz, C. R., Miller, Zhang, Z., Bast, R. C. Jr., Yu, Y., Li, J., Sokoll, L. J., Rai, A. J., Rosenzweig, J. M., Cameron, B., Wang, Y. Y., Meng, X. Y., Berchuck, A., Van Haften-Day, C., Hacker, N. F., de Bruijn, H. W., van der Zee, A. G., Jacobs, I. J., Fung, E. T., and Chan, D. W. (2004) Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. Cancer Res. 64, 5882–5890

14. Paradis, V., Degos, F., Dargere, D., Pham, N., Belghiti, J., Degott, C., Janeu, J. L., Bezaud, A., Delforge, D., Cubizolles, M., Laurendeau, I., and Bedossa, P. (2004) Identification of a new marker of hepatocellular carcinoma by serum protein profiling of patients with chronic liver diseases. Hepatology 41, 40–47

15. Schüssler, G. C. (2000) The thyroxine-binding proteins. Thyroid 10, 141–149

16. Dea, M. K., Hamilton-Wessler, M., Ader, M., Moore, D., Schaffer, L., Loff- ager, M., Volund, A., and Bergman, R. N. (2002) Albumin binding of acylated insulin (N3N304) does not deter action to stimulate glucose uptake. Diabetes 51, 762–769

17. Zhou, M., Lucas, D. A., Chan, K. C., Issaq, H. J., Petricoin, E. F., 3rd, Liotta, L. A., Veenstra, T. D., and Conrads, T. P. (2004) An investigation into the human serum “interactome.” Electrophoresis 25, 1289–1298

18. Pisitkun, T., Shen, R. F., and Knepper, M. A. (2004) Identification and proteomic profiling of exosomes in human urine. Proc. Natl. Acad. Sci. U. S. A. 101, 13368–13373

19. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. Nature 422, 198–207

20. Chen, J. H., Chang, Y. W., Yao, C. W., Chiu, T. S., Huang, S. C., Chien, K. Y., Chen, A., Chang, F. Y., Wong, C. H., and Chen, Y. J. (2004) Plasma proteome of severe acute respiratory syndrome analyzed by two-dimen-sional gel electrophoresis and mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. 101, 17039–17044

21. Liao, H., Wu, J., Kuhn, E., Chin, W., Chang, B., Jones, M. D., O’Neil, S., Clauser, K. R., Hailer, F., Roubenoff, R., Zolg, W., and Guild, B. C. (2004) Use of mass spectrometry to identify protein biomarkers of disease severity in the synovial fluid and serum of patients with rheuma-toid arthritis. Arthritis Rheum. 50, 3792–3803

22. Swan, A., Amer, H., and Dieppe, P. (2002) The value of synovial fluid assays in the diagnosis of joint disease: A literature survey. Ann. Rheum. Dis. 61, 493–498

23. Silverberg, G. D., Heit, G., Huhn, S., Jaffe, R. A., Chang, S. D., Bronte-Stewart, H., Rubenstein, E., Possin, K., and Saul, T. A. (2001) The cerebrospinal fluid production rate is reduced in dementia of the Alzheimer’s type. Neurology 57, 1763–1766

24. Silverberg, G. D., Levinthal, E., Sullivan, E. V., Bloch, D. A., Chang, S. D., Leverenz, J., Fittman, S., Winn, R., Marciano, F., Saul, T., Huhn, S., Mayo, M., and McGuire, D. (2002) Assessment of low-flow CSF drainage as a treatment for AD: results of a randomized pilot study. Neurology 59, 1139–1145

25. Silverberg, G. D., Mayo, M., Saul, T., Rubenstein, E., and McGuire, D. (2003) Alzheimer’s disease, normal-pressure hydrocephalus, and senescent changes in CSF circulatory physiology: A hypothesis. Lancet Neurol. 2, 506–511

26. Blennow, K., and Hampel, H. (2003) CSF markers for incipient Alzheimer’s disease. Lancet Neurol. 2, 605–613

27. Blennow, K. (2004) CSF biomarkers for mild cognitive impairment. J. Intern. Med. 255, 224–234

28. Masliah, E., Yamada, K., Shimazu, N., Seiki, K., Oda, H., Nakau, H., Inui, T., Li, W., Eguchi, N., and Urade, Y. (2003) Lipocalin-type prostaglandin D synthase (beta-trace) in cerebrospinal fluid: A useful marker for the diagnosis of normal pressure hydrocephalus. Neurosci. Res. 47, 455–459

29. Conti, A., Sanchez-Ruiz, Y., Bach, A., Beretta, L., Grandi, E., Beltramo, M., and Alessio, M. (2004) Proteome study of human cerebrospinal fluid following traumatic brain injury indicates fibrinogen degradation products as trauma-associated markers. J. Neurotrauma 21, 854–863

30. Lescuyer, P., Allard, L., Zimmermann-Ivol, C. G., Burgess, J. A., Hughes-Frutiger, S., Burkhard, P. R., Sanchez, J. C., and Hochstrasser, D. F. (2004) Identification of post-mortem cerebrospinal fluid proteins as potential biomarkers of ischemia and neurodegeneration. Proteomics 4, 2234–2241
36. Ramstrom, M., Ivonin, I., Johansson, A., Askmark, H., Markides, K. E., Zubarev, R., Hakansson, P., Aquilonius, S. M., and Bergquist, J. (2004) Cerebrospinal fluid protein patterns in neurodegenerative disease revealed by liquid chromatography-Fourier transform ion cyclotron resonance mass spectrometry. *Proteomics* 4, 4010–4018

37. Zhang, J., Goodlett, D. R., Peskind, E. R., Quinn, J. F., Zhou, Y., Wang, Q., Pan, G., Yi, E., Eng, J., Aebersold, R. H., and Montine, T. J. (2005) Quantitative proteomic analysis of age-related changes in human cerebrospinal fluid. *Neurobiol. Aging* 26, 207–227

38. Hammack, B. N., Fung, K. Y., Hunsucker, S. W., Duncan, M. W., Burgoon, M. P., Owens, G. P., and Gilden, D. H. (2004) Proteomic analysis of multiple sclerosis cerebrospinal fluid. *Mult. Scler.* 10, 245–260

39. Puchades, M., Hansson, S. F., Nilsson, C. L., Andreasen, N., Blennow, K., and Davidsson, P. (2003) Proteomic studies of potential cerebrospinal fluid protein markers for Alzheimer’s disease. *Brain Res. Mol. Brain Res.* 118, 140–146

40. Zheng, P. P., Luider, T. M., Pieters, R., Avezaat, C. J., van den Bent, M. J., Sillevis Smitt, P. A., and Kros, J. M. (2003) Identification of tumor-related proteins by proteomic analysis of cerebrospinal fluid from patients with primary brain tumors. *J. Neuropathol. Exp. Neurol.* 62, 855–862

41. Yuan, X., Russell, T., Wood, G., and Desiderio, D. M. (2002) Analysis of the human lumbar cerebrospinal fluid proteome. *Electrophoresis* 23, 1185–1196

42. Renne, T., Dedio, J., David, G., and Muller-Esterl, W. (2000) High molecular weight kininogen utilizes heparan sulfate proteoglycans for accumulation on endothelial cells. *J. Biol. Chem.* 275, 33688–33696

43. Mogues, T., Etzerodt, M., Hall, C., Engelich, G., Graversen, J. H., and Hartshorn, K. L. (2004) Tetranectin binds to the Kringle 1–4 form of angiostatin and modifies its functional activity. *J. Biomed. Biotechnol.* 2004, 73–78

44. Kirchmair, R., Egger, M., Walter, D. H., Eisterer, W., Niederwanger, A., Woell, E., Nagl, M., Pedrini, M., Murayama, T., Frauscher, S., Hanley, A., Silver, M., Brodmann, M., Sturm, W., Fischer-Colbrie, R., Losordo, D. W., Patsch, J. R., and Schratzberger, P. (2004) Secretoneurin, an angiogenic neuropeptide, induces postnatal vasculogenesis. *Circulation* 110, 1121–1127

45. Park, K. W., Crouse, D., Lee, M., Karnik, S. K., Sorensen, L. K., Murphy, K. J., Kuo, C. J., and Li, D. Y. (2004) The axonal attractant Netrin-1 is an angiogenic factor. *Proc. Natl. Acad. Sci. U. S. A.* 101, 16210–16215

46. Keay, S. K., Szekely, Z., Conrads, T. P., Veenstra, T. D., Barchi, J. J., Jr., Zhang, C. O., Koch, K. R., and Michejda, C. J. (2004) An antiproliferative factor from interstitial cystitis patients is a frizzled 8 protein-related sialoglycopeptide. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11803–11808

47. Warren, J. W., and Keay, S. K. (2002) Interstitial cystitis. *Curr. Opin. Urol.* 12, 69–74