Serum and Plasma Proteomics

Haleem J. Issaq, Zhen Xiao, and Timothy D. Veenstra*

Laboratory of Proteomics and Analytical Technologies, Advanced Technology Program, SAIC-Frederick, Inc., NCI-Frederick, P.O. Box B, Frederick, Maryland 21702

Received March 28, 2007

Contents

1. Introduction 3601
2. Why Study the Proteome of Blood? 3601
   2.1. What's the Difference between Serum and Plasma? 3602
   2.2. Collection, Handling, and Storage of Serum and Plasma 3603
3. Characterization of the Serum Proteome 3603
   3.1. Mass Spectrometry 3604
   3.2. The Dynamic Range Problem Related to Protein Concentration 3604
   3.3. Fractionation Methods for Serum Proteomics 3605
      3.3.1. Two-Dimensional Polyacrylamide Gel Electrophoresis of Serum and Plasma 3606
   3.4. Solution-Based Methods for Characterizing the Serum Proteome 3608
      3.4.1. Affinity Chromatography 3608
      3.4.2. Isoelectric Focusing 3609
      3.4.3. Ion Exchange Chromatography 3609
      3.4.4. Reversed-Phase Liquid Chromatography 3609
      3.4.5. Multidimensional Protein Identification Technology (MudPIT) 3610
      3.4.6. Capillary Electrophoresis 3611
4. Comparative Analysis of Serum Proteomes 3611
   4.1. Two-Dimensional Polyacrylamide Gel Electrophoresis 3612
      4.1.1 Two-Dimensional Differential In-Gel Electrophoresis 3612
   4.2. Stable-Isotope Labeling for Quantitative Proteomics 3613
   4.3. Subtractive Proteomics 3614
   4.4. Proteomic Profiling 3615
5. Automation 3617
6. Conclusions 3617
7. Acknowledgments 3618
8. References 3618

1. Introduction

There is a saying, "Eyes are the windows to the soul". In much the same vein, blood, since it flows through the entire body, can be considered the window to the physical condition of the patient. Virtually everyone who has been to visit their doctor has given blood that is tested to detect any present disease conditions and determine the functional state of the body's organs. A variety of different characteristics of blood are often measured depending on the physical condition of the patient (Figure 1). The fact that blood is comprised of several different types of cells and compounds such as salts and proteins makes blood tests very useful diagnostic tools. For example, the physician may request a white blood cell count if an infection is suspected or a Basic Metabolite Profile (i.e., measurement of sodium, potassium, chloride, calcium, bicarbonate, blood urea nitrogen, creatinine, and glucose levels)[1] may be needed to determine the present state of renal function. Depending on the required test, blood can be separated into different components. A simple blood smear can be prepared on a microscope slide to examine and count red and white blood cells, as well as platelets. The liquid portion of blood is referred to as plasma.[2] When blood is permitted to clot after it has been drawn from the patient, the blood cells and some of the proteins precipitate. Centrifugation of this sample is used to separate these solids, leaving behind the serum portion of blood.[3]

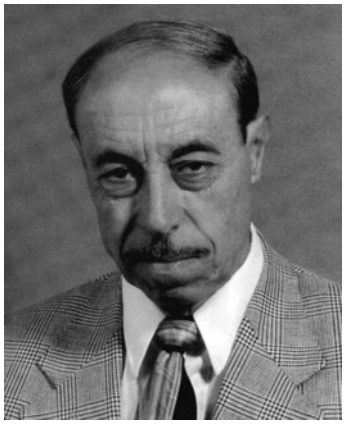
While numerous diagnostic tests for diseases ranging from cancer to diabetes are presently conducted using blood samples, the general consensus is that the archive of information within this biofluid has only begun to be understood. It is this hope that has driven the efforts to develop technologies necessary to mine this information within blood. These developments have impacted not only the instrumentation used to acquire the data but also how the samples are prepared and how the data is analyzed before and after data acquisition, respectively. This review will focus on the advances made in characterizing the serum and plasma proteomes, beginning with sample preparation methods, proceeding through fractionation methods, and, finally, ending with mass spectrometry (MS) analysis. While protein arrays are becoming increasingly important, MS remains the analytical technique of choice for serum and plasma profiling. For information on the use of protein arrays, the reader is directed to several interesting reviews.[4]

2. Why Study the Proteome of Blood?

The human body possesses over 60 000 miles of veins, arteries, and capillaries. Approximately five liters of blood travels continuously through the body by way of the circulation system. Blood carries oxygen and nutrients to cells and transports carbon dioxide and waste products excreted from cells.[5] It is estimated that no cell is more than four cell units removed from the circulation system. No other biofluid has intimacy with the body like blood has, and therefore, it is not surprising that it possesses such a richness of information concerning the overall pathophysiology of the patient. Unlike specific cell types, however, blood does not

* To whom correspondence should be addressed. Telephone: 301-846-7286. Fax: 301-846-6037. E-mail: veenstra@ncifcrf.gov.

10.1021/cr068287r CCC: $65.00
Published on Web 07/18/2007



Haleem J. Issaq received his B.S. degree from the American College of Istanbul, Turkey, a M.S. degree from the Technion, Israel Institute of Technology, and a Ph.D. in analytical chemistry from Georgetown University in 1972. He is a principal scientist and head of the Separation Technologies and Assay Development Group within the Laboratory of Proteomics and Analytical Technologies, SAIC-Frederick, Inc., located within the National Cancer Institute at Frederick, Maryland. His current research interests include multidimensional separation of complex protein/peptide mixtures using liquid-phase-based separations (LC/LC, IEF/LC, and LC/CE), fast separation of estrogens, and use of HPLC/MS metabolomics for disease biomarker detection. He is the author/coauthor of over 250 scientific publications and the editor of *A Century of Separation Science*. He is the recipient of numerous awards, including Maryland Chemist of the Year.

Dr. Zhen Xiao obtained her B.S. degree in biochemistry with distinction from Zhongshan University, Guangzhou, China, and a Ph.D. degree in Biomedical Sciences from Eastern Virginia Medical School, Norfolk, Virginia. After completing her postdoctoral training in the Laboratory of Proteomics and Analytical Technologies, SAIC-Frederick, Inc., located within the National Cancer Institute at Frederick, Maryland, she was promoted to a scientist position in the same laboratory. Her current research interests include developing and applying mass-spectrometry-based proteomic technologies in the areas of prevention, diagnosis, and treatment of cancer, AIDS, and other diseases. In particular, she has expertise in biomarker discovery and studying extracellular matrix and vesicular proteins. She was a two-time recipient of the prestigious American Association for Cancer Research AFLAC Young Scholar Award, in 2001 and 2003.

Timothy D. Veenstra received his B.Sc. degree in chemistry from Trent University, Peterborough, Ontario, Canada, in 1990 and a Ph.D. in biochemistry from the University of Windsor, Windsor, Ontario, Canada, in 1994. After a three-year postdoctoral fellowship at the Mayo Clinic/Foundation, he studied mass-spectrometry-based proteomics under the guidance of Dr. Richard D. Smith at Pacific Northwest National Laboratories. Dr. Veenstra has been the director of the Laboratory of Proteomics and Analytical Technologies, SAIC-Frederick, Inc., located within the National Cancer Institute at Frederick, Maryland, for the past six years. His current research interests include the use of mass spectrometry to interrogate clinical samples in the search for disease-related biomarkers.

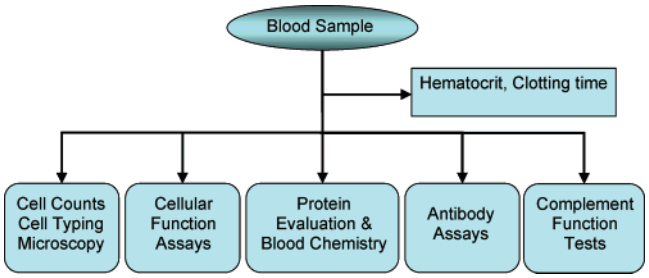


Figure 1. Blood is an extremely valuable fluid for monitoring the well-being of the patient. A variety of different tests are conducted on a blood sample to assist the physician in diagnosing the patient's physical condition.

contain its own genome. Its genome can be considered as a compilation of the organism's foundational genome along with all of the variations (i.e., mutations, single nucleotide polymorphisms, gene duplications, etc.) that are found in particular cells. Since it lacks a specific genome, blood does not have its own transcriptome. Rather, it can potentially contain any portion of a transcript that is transcribed within any cell in the body. Likewise, the proteome of blood can potentially contain portions of any protein found within any cell. A recent study comparing N-linked glycopeptides within cultured cells, solid tissues, and plasma, and other studies that have characterized the serum proteome have clearly substantiated this hypothesis.[6]

2.1. What's the Difference between Serum and Plasma?

There has been an overwhelming interest in trying to decipher valuable information from the proteome content of blood. To be clear, it is not actually blood that is directly analyzed in most proteomic studies; rather, it is the plasma or serum portion of blood. As briefly mentioned above, plasma is the liquid portion of *unclotted* blood that is left behind after all the various cell types are removed. To prepare plasma, blood is withdrawn from the patient using venipuncture in the presence of an anticoagulant and the sample is centrifuged to remove cellular elements. The most commonly used anticoagulants include heparin, ethylenediamine tetraacetic acid (EDTA), or sodium citrate.[2] Heparin prevents coagulation by activating anti-thrombin while both EDTA and sodium citrate prevent coagulation by chelating calcium ions. It is also possible to draw blood through a resin that

removes calcium, thereby preventing the coagulation. Serum is prepared by collecting plasma in the absence of any coagulant. Under these conditions, a fibrin clot forms. This clot is then removed using centrifugation, leaving behind serum.[3]

The process of coagulation makes serum qualitatively different from plasma. The removal of a large portion of the fibrinogen content of plasma in the form of the fibrin clot results in serum having a protein concentration lower than that of plasma. This difference, however, is only on the order of 3−4%.[7] Other proteins are also removed by specific or nonspecific interactions within the fibrin clot. Conventional thinking would surmise that many coagulation factors are also removed in the preparation of serum. Actually, factors IX, X, XI, and VII/VIIa are found within serum.[8] While the primary effect of the coagulation process is the removal of the fibrin clot, platelets, erythrocytes, and leukocytes secrete and increase certain proteins in serum during the same process. One study showed that the levels of platelet-secreted vascular endothelial growth factor (VEGF) are 230 ± 63 and 38 ± 8 pg/mL in the serum and plasma of normal individuals, respectively.[8] In studies of patients suffering from thrombocytosis, in which their platelet count is substantially increased compared to those of matched healthy controls, VEGF levels are also much higher.[9] These results show that serum and plasma VEGF levels are affected by platelets, but more markedly so in serum.

2.2. Collection, Handling, and Storage of Serum and Plasma

One issue that is a constant concern in the proteomic analysis of serum or plasma is the method used for collection, preparation, and storage of the samples. It is universally recognized that sample collection, handling, and storage have a great impact on the sensitivity, selectivity, and reproducibility of any given analysis. Unfortunately, considering its perceived importance as a source of biomarkers, little has been done to address these effects on serum or plasma samples. A recent study by Hsieh et al. showed that using different blood collection tubes affects the observable proteome of serum and plasma.[10] In addition, other factors utilized in the preparation of serum, such as the anticoagulant used, the clotting time allowed, and the length of the time period before centrifugation, had a significant effect on the serum proteome. Other handling parameters, such as overnight fasting, the time and speed of centrifugation, storage conditions (e.g., temperature and time), and repeated freeze/thaw cycles, had only a minor effect on the sample's proteome. Other studies, however, have shown a significant effect of freeze/thaw cycles on the proteome profile of serum/plasma.[11] These comparative studies, however, were conducted by profiling the low molecular weight proteome of serum and plasma using MALDI-TOF/MS. Given the penchant for this technology to detect only high abundance species,[12] more subtle effects on lower abundant proteins would remain undetected. In addition, it is impossible to study the effects of storage beyond a couple of years, as the analytical technology used today in serum/plasma proteome analysis has changed and continues to rapidly change. It is virtually impossible to precisely determine what proteomic changes have occurred in samples that have been stored in repositories over decades of time.

Commercially available blood collection tubes are in widespread use today for clinical sample collection. The two primary effects of collection tubes on the blood proteome are the addition or removal of components. Silicones are frequently used to coat the internal surface of these tubes, and polymers such as polyvinylpyrrolidines or polyethylene glycols may also be added. A study by Drake et al. showed that almost 65% of the tubes tested shed polymeric compounds into the clinical sample.[13] Instead of using serum or plasma, the study was carried out with an aqueous saline solution simulating typical contact times of blood components with the tube from collection to processing, thus avoiding the potential confusion of whether observed peaks represent peaks derived from the tubes or peaks from the serum/plasma components. The study showed that seven out of 11 tubes tested added polymeric components detected as multiple signals in the m/z range 1000−3000. These peaks could potentially complicate and compromise the interpretation of MS spectra in the low molecular weight range, especially when using MALDI or SELDI, in which a broad spectrum of different components is measured in a single analysis. Due to the mass to charge (m/z) values and propensity for such compounds to ionize, polymeric contamination can severely confound the ability to analyze the proteomic components of serum/plasma. Polymers also wreak havoc by plugging and prematurely degrading chromatography columns. In addition to the shedding of components from the tube, adsorption of serum/plasma proteins to the tube may occur, so different protein profiles from the same sample can be obtained depending on the kind of tube used. Actually, significant differences have been found when comparing red-top tubes (glass tubes containing no preservatives or anticoagulants) and tiger-top tubes (also known as serum separator tubes or SST) in different studies.[14]

A few studies have been carried out showing that sampling procedures (i.e., fasting, time sample acquired from patient, etc.) had the greatest effects on proteome profiling, while handling procedures and storage conditions had relatively minor effects.[15] However, everyone agrees that standardized protocols for serum/plasma sampling, handling, and storage are required, since the issue is not about which procedure is better but rather about using standardized procedures to obtain comparable and reproducible results between different laboratories.[16]

3. Characterization of the Serum Proteome

The proteome describes the entire compliment of proteins expressed by a cell at a point in time. Alterations in protein abundance, function, and structure can serve to indicate pathological abnormalities even before clinical symptoms are observed. Therefore, if the early detection of diseases such as cancer is to become a reality, it is vital to identify useful diagnostic and prognostic biomarkers. The methods that have been evaluated for discovering proteins for diagnostic purposes are often based on separation (electrophoresis and chromatography) followed by MS for detection and identification. Since blood circulates throughout the body, early hypotheses suggested that the serum/plasma proteome contains a treasure trove of protein biomarkers. Without having any prior information, discovering these biomarkers requires analytical methods that can be used to examine as much of this biofluid proteome as possible. Therefore, much of the analytical development in the past few years has focused on methods to identifying ever increasing numbers of proteins within serum and plasma.

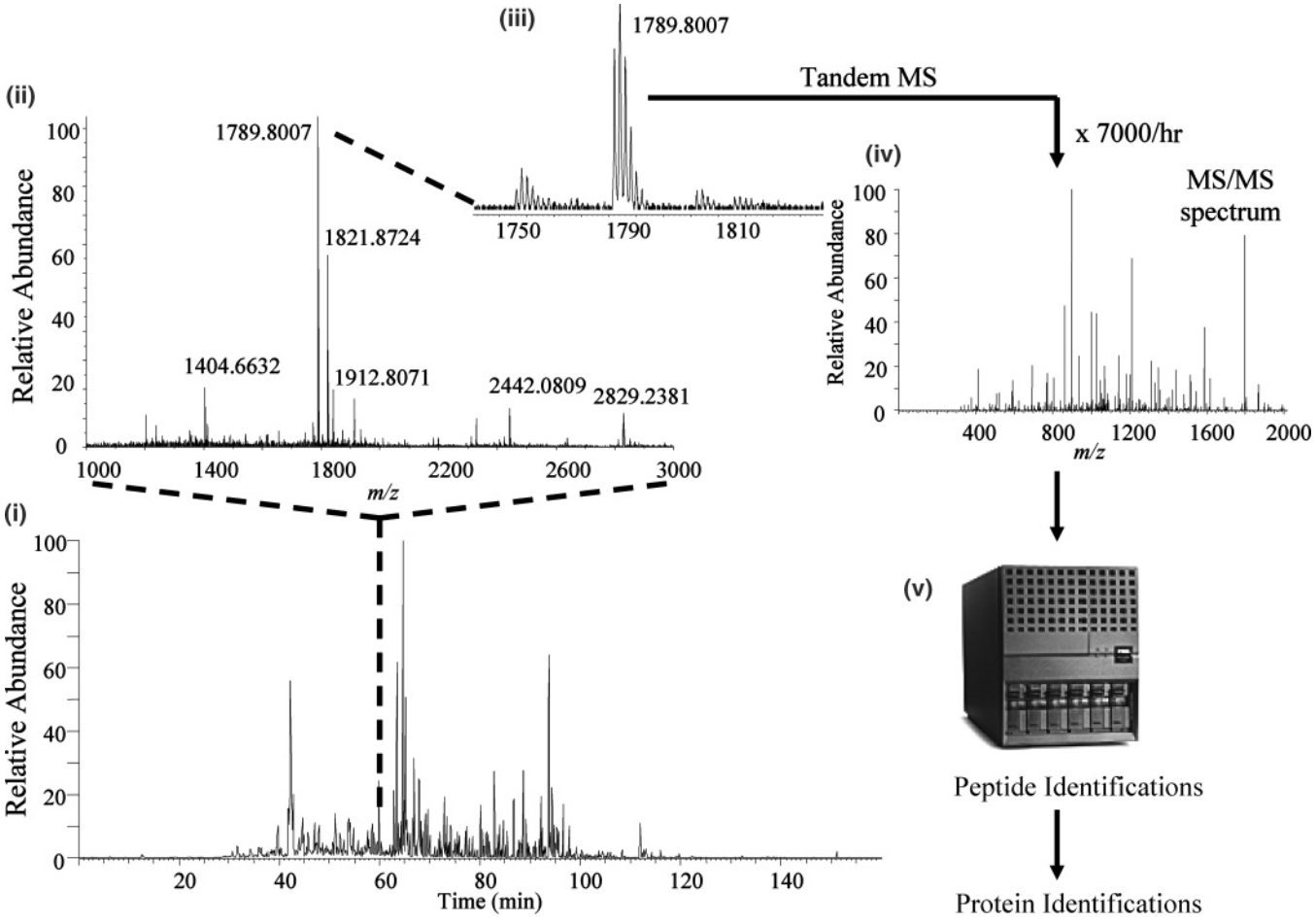


Figure 2. Identification of peptides within complex mixtures using data-dependent tandem mass spectrometry (MS/MS). In data-dependent MS/MS, the mass spectrometer specifically isolates peptide ions based on their signal intensity. Collisional induced dissociation is then used to fragment the peptide, and the mass-to-charge ratios of the resulting pieces are recorded. Software is used to compare the resulting MS/MS spectra against protein sequences within an appropriate database to determine the sequence of the peptide. The peptide sequence is then correlated back to its protein of origin. Modern mass spectrometers are able to repeat this process approximately 7000 times per hour.

3.1. Mass Spectrometry

The ability of the mass spectrometer to rapidly identify proteins is arguably the parameter that makes this instrumentation the driving force in proteomics today. How exactly does a mass spectrometer identify peptides? As shown in Figure 2, peptides are being constantly eluted from a reversed-phase column into the mass spectrometer (i). During this separation, the instrument records the mass-to-charge (*m*/*z*) ratios of the peptides that are eluting at a specific time point (ii). The instrument then selects and isolates the most intense ion observed in the previous scan (iii) and fragments it, in a process referred to as tandem MS, to create a series of sequence ladders (iv). After this fragmentation event, the instrument proceeds to isolate and fragment the next most abundant peptide ion. It does this sequential ion selection and fragmentation for anywhere from 3 to 10 of the most abundant peptide ions (depending on the operator setting). Today's mass spectrometers are able to collect approximately 7000 tandem mass spectra in a single hour. All of these spectra are then analyzed using the appropriate software and protein or genome database to identify the peptides that gave rise to the individual spectra (v). In a typical analysis, 10−20% of the spectra will give a "hit", allowing between 700 and 1400 peptides to be confidently identified and then correlated to their proteins of origin.

3.2. The Dynamic Range Problem Related to Protein Concentration

At first glance, serum and plasma seem to be the ideal clinical samples for MS-based proteomic analysis. They are relatively easily obtained from the patient and have a very high protein concentration (e.g., on the orders of tens of mg/mL). The protein concentration, however, is deceiving. Twenty-two proteins make up approximately 99% of the protein content of serum and plasma (Figure 3).[17] It is estimated that the protein concentrations in these samples span 10 orders of magnitude, and the prevailing thought is that specific disease biomarkers for diagnostic and prognostic purposes are most likely within the very low concentration range. It was recognized early on, particularly in the analysis of serum and plasma, that the high dynamic range of protein concentrations found in these two fluids was going to be problematic for downstream MS analysis.[18] Considering that the dynamic range of a mass spectrometer is on the order of 2 orders of magnitude, it is easy to figure out that a straightforward liquid chromatography (LC)-MS/MS analysis will result in the characterization of only the highest abundance, and probably least interesting, proteins. While strong cation exchange fractionation prior to reversed phase (RP) LC-MS/MS analysis has been shown to increase the ability to identify low abundant proteins in many proteomic

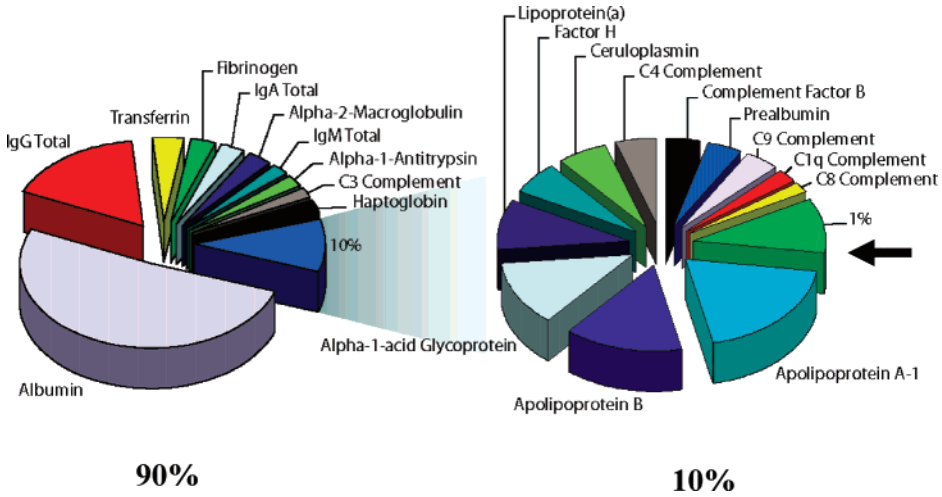


Figure 3. The dynamic range of protein concentrations in human serum. While the overall protein concentration of serum is high, 22 proteins make up 99% of the total protein amount.

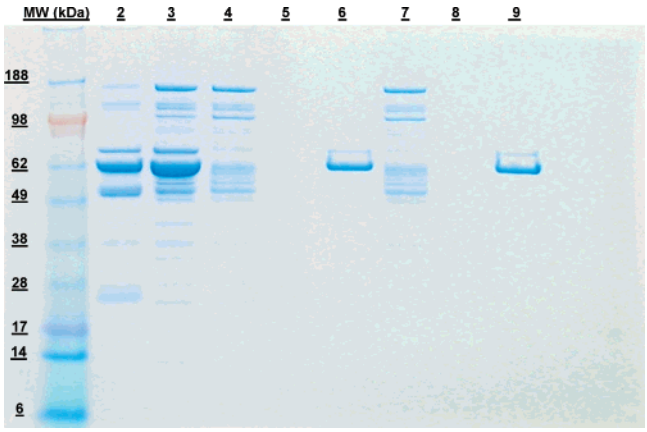


Figure 4. Depletion of serum using a multiple affinity removal system (MARS): lane 2, serum standard; lane 3, raw plasma; lane 4, plasma proteins that flow through a MARS immunodepletion column during the first washing step; lane 5, plasma proteins that flow through a MARS immunodepletion column during the second washing step; lane 6, elution of high abundant proteins that are retained by the MARS immunodepletion column. Lanes 7, 8, and 9 are replicates of lanes 4, 5, and 6, respectively, using a second MARS column.

studies,[19] this strategy alone is not sufficient to gain comprehensive coverage of the low abundance proteins within biofluids.

It was quickly recognized that to effectively characterize serum or plasma was going to require methods to remove the high abundance proteins prior to downstream analysis. One of the earliest approaches used to deplete high abundance proteins was to pass a serum/plasma sample over Cibracon blue, a dye with high affinity for albumin.[20] Albumin, as shown in Figure 3, comprises approximately 50% of the protein content of serum/plasma; therefore, removal of this single protein has an immediate impact on the dynamic range problem related to protein concentration. Recently, Agilent introduced the multiple affinity removal system (MARS) for the immunodepletion of six high abundant proteins (i.e., albumin, IgG, IgA, transferrin, haptoglobin, and alpha-1-antitrypsin) in serum/plasma.[21] The usefulness of this product in removing high abundant proteins is illustrated in Figure 4. Similar products have been developed, including a Top 20 depletion column from Sigma, and the Seppro MIXED12 IgY-based affinity LC column, for the depletion of the twelve highest abundance plasma proteins manufactured by GenWay Biotech Inc.[22] The reproducibility and effectiveness of these products to deplete major proteins in serum/plasma samples have always been a concern. In fact, a recent study published the results of the reproducibility of a MARS column across serum samples from patients with prostate cancer. They found that the depletion of high abundant proteins from all 250 serum samples was complete and reproducible, with a relative standard deviation below 7%, over a six week period.[21] Another study comparing a series of sample preparation methods has also confirmed the effectiveness and robustness of immunoaffinity subtraction methods for simplifying the serum proteome prior to MS analysis.[23] Depletion of high abundant proteins is now considered an essential sample handling step in any serum/plasma study regardless of subsequent analytical strategies. There are always concerns, however, when using affinity-based depletion strategies that potentially important biomarkers will be lost either through the possible "sponge" effect of the high abundant proteins or by the nonspecific binding to the affinity column used. Indeed, studies have shown that proteins remain bound to the targeted high abundance proteins during their depletion.[24] Moreover, major protein depletion alone certainly was not enough to deal with the dynamic range problem.

3.3. Fractionation Methods for Serum Proteomics

Historically, the gold standard for the fractionation of complex proteomes has been two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). The primary characteristics of 2D-PAGE that make it effective for fractionating these very complex mixtures are its resolution and inherent quantitative abilities.[25] Studies utilizing 2D-PAGE have shown the ability to resolve upward of 3000 protein spots.[26] In addition, when comparing different proteomes, the visualized spot intensities provide a direct measure of the relative abundance of specific proteins between samples. Like most analytical methods, however, 2D-PAGE is not without its drawbacks. It is not amenable to proteins having extreme molecular weights, isoelectric points, or hydrophobicity. To overcome some of the drawbacks of 2D-PAGE, researchers have developed solution-based methods that obviate the use of gels altogether. While many different chromatographic and electrophoretic options (see below) are available for conducting solution-based fractionation of

proteomes, the most common approach utilizes a combination of strong cation exchange (SCX) and reversed-phase liquid chromatograph (RPLC). Another critical difference between 2D-PAGE and solution-based fractionation strategies is the form of the proteins that are separated. While 2D-PAGE is used to fractionate intact proteins, solution-based methods generally require the proteome to be proteolytically digested prior to their separation. While there is an increasing interest in top-down proteomic methods in which the proteins are kept intact throughout the analyses,[27] these methods are not yet mature enough to apply to serum/plasma proteomes.

3.3.1. Two-Dimensional Polyacrylamide Gel Electrophoresis of Serum and Plasma

Seventy five years ago, Tiselius introduced the moving boundary method as an analytical tool for studying the electrophoresis of proteins.[28] Using this method, he was able to resolve serum globulin into its ∀, ∃, and (components. Since his pioneering work, electrophoresis methods have been employed for the separation of complex mixtures of proteins with steadily increasing degrees of resolution. This increasing resolution can be attributed to the introduction of acrylamide gels,[29] stacking systems,[30] isoelectric focusing (IEF),[31] and a variety of 2D gel electrophoretic separations. Electrophoretic migration of proteins in gel electrophoresis depends on two important parameters: the electrophoretic mobility of the protein and the resistance of the medium. Currently, the electrophoretic procedures used for protein analysis are primarily one- and two-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoreses (1D and 2D SDS-PAGE), 2D differential gel electrophoresis (DIGE), and capillary electrophoresis (CE).

One-dimensional SDS-PAGE is by far the most commonly used method in biological sciences today to fractionate proteins. To separate complex protein mixtures, such as a serum/plasma proteome, and be able to effectively interrogate many of the species present in the mixture requires the use of 2D electrophoresis. While we tend to think of serum/plasma proteomics as a recent venture, investigators began developing better methods to resolve serum/plasma proteins over a half a century ago. For example, Smithies and Poulik realized in 1956 that "a combination of two electrophoretic processes on a gel at right angles should give a much greater degree of separation than is possible with either separately".[32] These two electrophoretic processes resolve proteins based on their (a) molecular size and (b) free solution mobility on a starch gel. In these early experiments, the first separation dimension was carried out on a 5 mm wide strip of filter paper using a buffer of pH 8.55. The strip containing serum proteins separated according to their free-solution mobilities was then inserted into a 12 cm wide starch gel, and a second dimension of separation was carried out at right angles to the first. This early report of the separation of the serum proteome reported the resolution of more than 15 proteins.[32] Hermans and co-workers described a similar technique to separate serum proteins in which paper electrophoresis was used in the first dimension and cyanogum in the second dimension.[33] Ashton used agar in the first dimension and starch gel in the second dimension.[34]

Raymond and Aurell realized the significant nonlinear effects the gel concentration had on the electrophoretic mobility of proteins.[35] Therefore, they employed two different gel concentrations to separate serum proteins: 5% acrylamide in the first dimension followed by 8% gel in the second dimension. Two years later, Raymond realized that conducting electrophoresis in the flat slab gel format had several advantages over the cylindrical tube format.[36] First, the flat slab provides the maximum surface area for cooling the gel. Second, the resulting patterns are easier to quantify in standard recording densitometers. Third, a large number of samples can be processed in a single gel, facilitating the direct comparison of specimens processed under identical conditions. Fourth, and most importantly for modern proteomics research, the flat slab permits the application of two-dimensional techniques. These insightful statements are as true today as they were in 1964 and are the basis of modern 2D-PAGE. Raymond used acrylamide gel in his 2D experiments, which he named "orthogonal gel electrophoresis or Orthacryl". He realized that using acrylamide gel in two orthogonal directions provides increased resolution and information about the molecular size and shape of related proteins. Raymond also hypothesized that using different pH values in each dimension would lead to better separation of proteins using 2D electrophoresis. Human plasma proteins were resolved by a combination of polyacrylamide discs in the first dimension at pH 9.4 and gradient gel electrophoresis in the second dimension at pH 8.6.[37] In gradient gel electrophoresis, proteins are driven through progressively decreasing pore sizes until they are brought to a near dead stop based on their size. Using this method, Slater was able to resolve about 30 proteins within human plasma.[38]

Subsequent to these studies, human serum was resolved using a soft polyacrylamide gel column followed by electrophoresis in a 2−30% polyacrylamide linear gradient gel slab.[39] In this procedure, serum proteins were separated in the first dimension using a 4.75% gel (2% cross-linkage). The gel column was removed from the glass cylinder and laid on the upper edge of a 2% gradient slab. Electrophoresis was carried out at 20 mA per slab (140−160 V) at 4−10 °C for 22 h. The gel mold was removed and placed in a staining solution of 0.5% amido black for 30 min. The procedure showed the capability of resolving 112 proteins. Many of these proteins were subsequently identified as IgG, IgA, IgM, haptoglobin, ceruloplasmin, transferrin, and albumin. This same study also reported one of the first differential proteome findings, as several post- and pre-albumin components were detected in sera from patients with myeloma, leukemia, and Hodgkin's disease but were not detected in sera of normal subjects.

The above developments form the basis for the modern advancements that have made it possible to fractionate serum/plasma proteins using 2D-PAGE. In 1975, O'Farrell introduced 2-DE for the separation of cellular proteins under denaturing conditions.[40] These conditions allowed hundreds of proteins from cells to be resolved on a single gel. The procedure's principle is inherently simple to understand: separation of the proteins by IEF in the first dimension, followed by molecular mass in the second. Recognizing its utility for proteomics, Anderson and Anderson quickly applied O'Farrell's method for the analysis of human plasma proteins.[41] They were able to resolve approximately 300 distinct spots upon staining, which they surmised were comprised of 75−100 proteins. Manabe et al. employed 2-DE for the separation of plasma proteins without denaturing agents.[42] They used IEF in the first dimension followed by electrophoresis in a 4−21% linear gradient slab gel. No denaturing agent was used throughout the experiment, enabling proteins to retain their native conformation. Al-

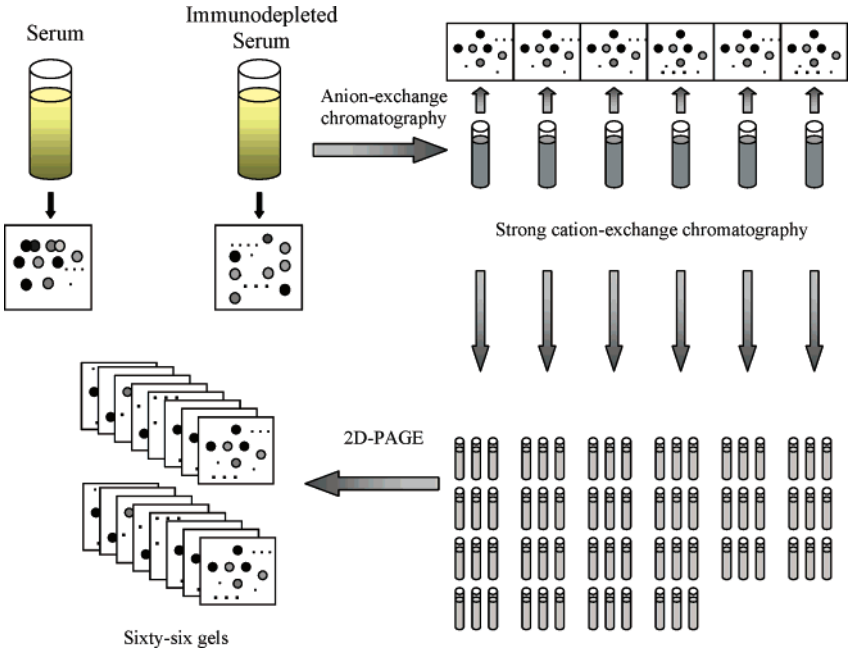


Figure 5. Characterization of the human serum proteome using an immunodepletion/chromatographic/two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) fractionation strategy followed by mass spectrometry (MS) identification. Serum, in which the high abundance proteins had been immunodepleted, was fractionated using anion and strong cation exchange chromatography, resulting in a total of 72 fractions that were separated and visualized on 2D-PAGE gels. Raw and immunodepleted sera were directly separated on two other gels. Analysis of the accumulative 20 000 spots resulted in the identification of 350 unique proteins. Data derived from ref 48.

though more than 230 protein spots could be observed on the gel, the spots were smeared and not very well resolved.

Modern 2D-PAGE has proven to be an extremely efficient method for the separation of complex protein mixtures. As mentioned above, 2D-PAGE is comprised of IEF in the first dimension followed by SDS-PAGE in the second. Since its introduction by Kolin in 1954,[43] IEF, which separates proteins based on their isoelectric points (pI), has undergone several advances. The most prominent advance, and one that has had a great impact on serum/plasma proteomics, was the introduction of immobilized pH gradients (IPGs) by Bjellqvist et al.,[44] which allowed the use of stable and reproducible pH gradients. In IPGs, the carrier ampholytes are attached to acrylamide molecules and cast into the gels to form a fixed pH gradient. Fixing the gradient prevents drift in the gel and also ensures that they can be cast in an efficient and reproducible manner. Using narrow-range IPG strips allowed a larger number of proteins to be separated than were possible with standard 2D-PAGE, because a narrower pH range was spread out over a greater physical distance. This spread allowed proteins with similar pI values to be separated with higher resolution. To illustrate this point, Hoving et al. developed a 2D-PAGE method in which they applied narrow range IPG strips in the first dimension.[45] The IPG strips were typically 1−3 pH units wide and overlap with one another by at least 0.5 pH units. Six IPG strips covering the pH ranges 3.5−5, 4.5−5.5, 5−6, 5.5−6.7, 6.2−8.2, and 7−10 were used. Proteins from a B-lymphoma cell line were applied to each strip and separated by IEF. The strips were then applied to individual SDS-PAGE gels and separated in the second dimension based on molecular weight. The same sample was also run on a single standard 2D-PAGE gel with a single IPG strip with a pH range of 3−10. Using a single gel, the group was able to detect approximately 1500 spots; however, using the six gels run using the narrow range IPG strips, they were able to detect approximately 5000 distinct spots. In addition, while only 0.8 mg of protein could be loaded onto a single gel, the use of narrow range IPG strips allowed 11 mg of total protein to be loaded onto the gels. The increase in resolution was estimated to permit proteins present only at about 300 copies per cell to be detected. Since then, the use of narrow-range IPG strips has undergone significant improvements, as discussed in an excellent review by Righetti et al.[46]

An extremely beneficial advantage of 2D PAGE is the ability to compare expression levels of proteins extracted from two different serum samples using two separate gel plates. For example, proteins extracted from two serum samples (healthy and diseased) are loaded on a separate gel and fractionated. After separation, the proteins are stained, and their abundance levels are compared by the intensity of their staining.[47] There are a variety of staining agents available, both colorimetric and fluorometric. The most popular staining techniques for 2D-PAGE separated proteins are Coomassie blue and silver staining. Other more sensitive stains such as Sypro ruby have been utilized; however, since MS is ultimately used to identify the protein spot, it is often folly to visualize a protein whose abundance is below the detection level of the mass spectrometer.

One of the first large scale studies that showed the ability to identify hundreds of proteins within serum utilized 2D-PAGE fractionation (Figure 5).[48] The serum sample was immunodepleted to remove the most abundant proteins (i.e., albumin, haptoglobins, transferrins, transthyretin, α-1-antitrypsin, α-1-acid glycoprotein, hemopexin, and β-2-macroglobulin). To further reduce the complexity of the sample that would be run on any one gel, the remaining proteins were separated into 74 fractions using sequential anion-exchange and size-exclusion chromatography. Each of these 74 fractions was separated on its own 2D-PAGE gel. After visualization of the proteins using Coomassie staining, approximately 20 000 individual spots could be seen. Removal of redundant spots that were seen across many gels left approximately 3700 unique spots. Analysis of these spots

Table 1. Advantages and Limitations of MS Analysis of Proteins by "Top-Down" and "Bottom-Up" Approaches

top-down	bottom-up
intact proteins	peptides
proteins are fragmented in the MS	protein digests
large peptide fragments	small peptides
applied to single purified protein	applied to complex protein mixtures
sophisticated and expensive instrumentation	simple instrumentation
easier to identify protein modifications	harder to identify protein modifications
solubility problems	easier to solubilize
ESI, ETD, ECD	ESI, MALDI
identifies proteins by MS/MS	identifies peptides by MS/MS
harder to separate different forms of the same protein	easier to separate

using MS resulted in the identification of 1800 of the spots, which correlated to 325 unique proteins. Approximately 39% of the proteins identified were known to be localized within the circulatory system, while 35% represented intracellular proteins that are hypothesized to enter circulation from cells and tissues. Cell surface proteins made up just over 6% of the total number of unique protein identifications. The value of the extensive fractionation was reflected in the number of proteins identified with known serum concentrations less than 10 ng/mL (e.g., interleukin-6, metallothionein II, cathepsins, and various peptide hormones).

From an analytical point of view, this effort provided the largest characterization, in terms of the number of identified proteins, of the serum proteome to date. From a practical viewpoint, however, this study suggested that the comparative analysis of serum/plasma samples from healthy and disease-affected patients using 2D-PAGE was going to be extremely laborious. While improvements in 2D-PAGE are going to occur, it was obvious that other methods to conduct serum/plasma proteomics were going to be needed.

3.4. Solution-Based Methods for Characterizing the Serum Proteome

Investigators soon began looking for ways to analyze the serum/plasma proteome without using 2D-PAGE. While serum/plasma proteomics was relatively new, fortunately chromatography and electrophoresis were not. As with 2D-PAGE developments, however, it was going to require actual experimental evidence to determine which solution-based fractionation methods were best suited to maximize the protein coverage of these complex proteomes.

The first step in the analysis of the blood proteome is to decide if the analysis should be carried out at the intact proteins or the proteins' digest (peptide) level. There are advantages and disadvantages to each, as indicated in Table 1. The primary advantage of initially digesting the proteins into peptides is solubility. Peptides generally have a wider range of solubility than proteins, and even difficult to solubilize proteins (e.g., membrane proteins) will produce a population of peptides in which at least a few are soluble. In a vast majority of proteomic studies of serum/plasma, MS is used as the identification tool. While top-down methods to identify intact proteins using MS are being rapidly developed and implemented, mass spectrometers are still best at identifying peptides rather than intact proteins. Therefore, proteome samples need to be broken down into peptides at some stage prior to MS analysis. Digestion of proteins into peptides, however, increases the complexity of an already complex sample. If 20 000 proteins are present within serum/plasma, digestion of this mixture could result in anywhere between 200 000 and 1 000 000 peptides. These numbers are well beyond the sampling capability of any mass spectrometer. A quick scan of the literature reveals, however, that most solution-based serum/plasma proteome studies (outside of protein profiling, which will be discussed later) are conducted at the peptide level.[23,49]

3.4.1. Affinity Chromatography

An effective method to simplify a proteomic mixture is through affinity chromatography. Affinity chromatography selects for a specific group of proteins with the selection criteria typically based on some unique characteristic of the protein such as its modification status (e.g., phosphorylation, glycosylation, etc.) or class (e.g., kinases). Application of affinity chromatography in proteomic research can be done at the protein or peptide level as shown in Figure 2. Affinity is based on the ability of a biologically active molecule to bind specifically and reversibly to a complimentary molecule, a ligand that is often bound to a solid support. These molecules may include antibodies, metals, lectins, biotin, aptamers, etc. The binding sites of the immobilized substances should be sterically accessible after their coupling to the solid support and should not be deformed by immobilization. In the case of specific proteins, an affinant is attached to the active surface of the column packing material or to the column surface. The sample is injected onto the column, and the protein(s) of interest are captured by the affinant. Proteins that do not possess a complementary binding site for the bound ligand will either pass directly through the column or be eluted using a low stringency washing step. The bound protein(s) is recovered by washing the column with a competitive substrate or a solution that disrupts the interaction between the protein and the affinant (e.g., denaturants). While the use of antibodies directed to a specific protein remains the most popular affinity-based fractionation method, many other affinity methods have been developed in order to isolate a specific class of proteins or peptides. These methods include immobilized metal affinity columns (IMACs) containing nickel, to capture histidine-containing peptides,[50] or gallium, to isolate phosphopeptides.[51] Recently, columns packed with resins containing titanium oxide (TiO_2) or zirconium oxide (ZrO_2) have been used for the affinity capture of phosphopeptides from complex protein digests.[52] Phosphopeptides are captured under acidic conditions on a TiO_2 or ZrO_2 column and then selectively released using an alkaline solution. Since the number of phosphopeptides within a biological mixture is quite large, the enriched mixture must still be fractionated prior to MS analysis. Although different approaches to phosphopeptide enrichment have been used, none of these methods is able capture *all* the phosphopeptides in a proteome, and no phosphoproteome has been mapped to completion. In addition, affinity methods have been devel-

oped to select peptides containing specific types of residues, such as cysteine, tryptophan, or methionine.[53] There are a variety of different lectins that have been used to selectively separate glycoproteins based on the composition of the carbohydrate side chain.[54] Glycoproteins from human serum were identified by a combination of lectin affinity chromatography along with anion exchange and Cu-IMAC selection of tryptic peptides. Glycoproteins were selected using a concanavalin A lectin column and tryptically digested prior to sequential chromatographic selection of acidic and histidine-containing peptides.[55] Unfortunately, affinity methods have not been widely used in the analysis of serum and plasma proteomes. For example, while thousands of phosphopeptides have been identified in a number of different cell types,[56] no comprehensive analysis of serum or plasma has yet been reported. Since it is widely accepted that serum and plasma contain proteins originating from various cells throughout the body, the use of affinity chromatography to capture and analyze modified proteins within the blood proteome will be an important tool for biomarker discovery in the future.

3.4.2. Isoelectric Focusing

Separations using isoelectric focusing (IEF) methods are based on the differences in proteins' isoelectric points. Proteins can carry positive, negative, or zero net charge depending on the pH of the buffer in which they are dissolved. Every protein has a specific pH at which its net charge is zero. Isoelectric focusing can be carried out in a liquid phase or a gel.

In liquid-phase-based IEF, the proteins in a sample are mixed with the desired pH range carrier ampholyte buffer in a focusing cell. Application of an electric potential to the focusing cell causes the proteins to migrate to a position in the established pH gradient equivalent to their respective pI. If a protein diffuses away from this pH region, its net charge will change and the resulting electrophoretic forces will influence its migration back to its pI point. The net result is the "focusing" of proteins into sharp bands at their pI values. The pH gradients are established by using carrier ampholytes. Ampholytes are compounds synthesized with a particular p*K* that when combined form a mixture that will establish a pH gradient. An advantage of liquid-phase IEF is the ability to fractionate a complex mixture of proteins according to their pI values in a nongel medium. The fractions can be collected and further analyzed, if needed, by electrophoresis or chromatography. The disadvantages of liquid-phase IEF are that high concentrations of "neutral" proteins (e.g., when focused at their pI) often precipitate from solution. Additionally, the ampholytes used to establish the pH gradient may interfere with subsequent analysis using techniques such as electrospray ionization mass spectrometry (ESI-MS). Also, highly hydrophobic proteins may be lost in sample preparation or during focusing when the proteins reach their isoelectric point.[57] Several IEF devices that can be used to fractionate a complex mixture of proteins at the preparative level are commercially available, including the Rotofor cell apparatus, a Multicompartment Electrolyzer with isoelectric membranes, recycling IEF, and free flow IEF.[58]

Isoelectric focusing has been used to fractionate human serum prior to analysis via LC-MS.[2] The human serum proteome was analyzed in our laboratory using two different strategies.[59] In an early study, serum proteins were extracted, digested with trypsin, and separated into twenty fractions using ampholyte free IEF with a Rotofor apparatus. The IEF fractions were collected and analyzed by reversed-phase LC-MS.[2] In addition, these twenty peptide fractions were each resolved into seven fractions using SCX for a total of 140 fractions. Analysis of these fractions by LC-MS^2 resulted in the identification of 1444 unique proteins.

Fractionation of proteins/peptides by IEF is also accomplished using gel electrophoresis, which is well suited for the separation of proteins. In a recent study, human plasma proteins were fractionated by IEF using immobilized pH gradient (IPG) strips.[60] The plasma protein mixture was separated on an IPG strip, after which the entire strip was cut into sections. Proteins in each section were digested, and the resulting peptides were analyzed by RP-HPLC followed by electrospray-linear ion-trap MS. A total of 744 distinct proteins were identified from an IPG strip loaded with 300 μg.

3.4.3. Ion Exchange Chromatography

One of the most popular methods to fractionate peptides for proteomic analysis is ion exchange chromatography,[61] which separates proteins/peptides according to their net charge. Most published studies dealing with proteome analysis using ion exchange for fractionation of the complex protein mixture have been used to fractionate protein digests rather than proteins because it is easier to work with peptides than with proteins, as mentioned earlier. Since proteins and peptides contain ionizable groups, the net charge is dependent on the composition of the mobile phase. In general, as the pH of a solution increases, deprotonation of the acidic and basic groups on these biomolecules occurs, so that carboxyl groups are converted to carboxylate anions ($R\text{-}COO^-$) and ammonium groups are converted to amino groups ($R\text{-}NH_2$). At a certain pH, the protein/peptide becomes neutral (has no net charge). This value is known as its isoelectric point (pI). At a pH > pI, the protein is negatively charged, and when the pH < pI, the protein is positively charged. The pI is different for different proteins. Separations using ion exchange chromatography are based on charge−charge interactions between the proteins in a sample and the immobilized stationary phase (resin). Separation is based on differences between the overall charges of the proteins or peptides. Ion exchange chromatography is subdivided into cation exchange chromatography, in which positively charged proteins bind to a negatively charged resin, and anion exchange chromatography, in which negatively charged proteins bind to a positively charged resin. Cation and anion exchange chromatography can be subsequently broken down into strong or weak cation and anion exchange. Proteins are selectively eluted from the column by changing the ionic strength of the mobile phase.

3.4.4. Reversed-Phase Liquid Chromatography

No single chromatographic or electrophoretic technique has had a bigger impact on proteomics than reversed-phase liquid chromatography (RPLC).[52] While improvements in MS instrumentation are most often thought of as the driving force, it is unlikely that proteomics would be where it is today without the developments made in coupling RPLC with MS. Not only do RPLC columns provide high-resolution separations, they also utilize solvent conditions that make it compatible with electrospray ionization (ESI) MS detection. The column used in RPLC separation of peptides is packed with bonded silica particles. The functionalities most widely

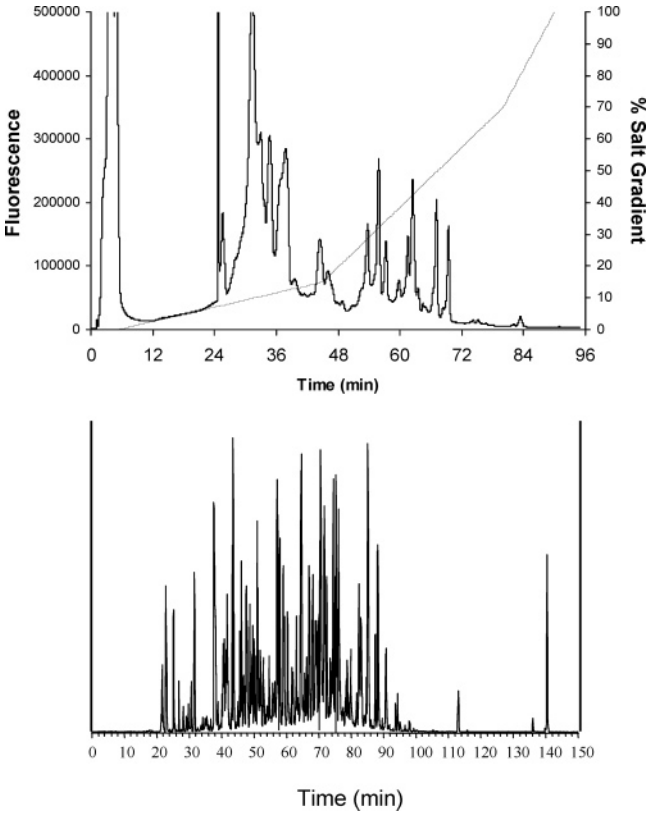


Figure 6. Fractionation of serum using strong cation exchange (SCX) (top) and reversed-phase liquid chromatography (RPLC) (bottom). The higher resolution afforded using RPLC dictates that it follows SCX in a multidimensional separation scheme for the analysis of complex proteome samples. While fractions are collected throughout the entire SCX separation, only those collected 12 min after the start of the separation are subsequently analyzed using RPLC chromatography.

bonded to silica are C_{18}, C_8, C_4, and aromatic phenyl groups. The bonded silica particles vary in size from approximately 1 to 10 μm with a pore size, depending on the size of the peptides, ranging from 90 to 300 Å. RPLC has been used primarily as the second dimension of separation in the analysis of peptides due to its high resolution and ability to be coupled directly on-line with electrospray ionization (ESI) MS analysis. In RPLC, a solute molecule binds to the immobilized hydrophobic molecule and is eluted using a polar solvent (water modified with a polar organic solvent such as acetonitrile, methanol, or tetrahydrofuran). Partitioning of the solutes occurs as a result of hydrophobic−hydrophobic interactions between the hydrophobic moiety of the peptide and that of the stationary phase. Polar solvents with increasing organic modifiers are used to break the hydrophobic bond, and elution of the peptides off the column will occur. The most hydrophobic peptides elute last due to their stronger hydrophobic interaction with the alkyl group. Normally, for the separation of complex mixtures of peptides, a solvent gradient (linear or step) with increasing organic modifier, and an ion pairing agent such as trifluoroacetic or formic acid, is used. The separation of low molecular weight serum protein digests using SCX and C-18 RPLC is shown in Figure 6. Since the resolution of RPLC is much greater than that of SCX, RPLC is always used after SCX in a two-dimensional separation scheme.

3.4.5. Multidimensional Protein Identification Technology (MudPIT)

Ion exchange and RP have been used extensively as the first and second stage chromatographies for the MS analysis of complex proteomes. The popularization of this combination can be directly attributed to the laboratory of John Yates III, who initially utilized this technique in combination with RPLC to develop a multidimensional separation method termed MudPIT.[63] Many of these early developmental studies using this combination of chromatographic procedures revolved around the analysis of the yeast proteome. In these studies, the yeast proteome was first digested into peptides and this mixture was loaded onto a SCX column. A discrete fraction of peptides was displaced from the SCX column directly onto a reversed-phase (RP) column. This population was then eluted from the RP column directly into the mass spectrometer for identification of the individual peptides. This iterative process was repeated twelve times using steps of increasing salt concentrations to elute the peptides from the SCX column to the RP column.[64] To minimize sample losses between the two separation dimensions, the SCX and RP columns were packed at opposing ends of a single capillary column. An advantage of this separation technology is that the entire system is coupled directly on-line with MS, enabling a large number of peptides to be directly identified in a high-throughput manner. These initial studies showed the capability of identifying almost 1500 proteins within the yeast proteome. This combination of SCX and RPLC has become the most popular multidimensional fractionation method used in proteomics today and has been used to analyze a variety of clinically relevant proteome samples including plasma, serum, urine, cerebrospinal fluid, plasma filtrate, and blood ultrafiltrate.[64]

While ion exchange chromatography has been conducted primarily at the peptide level for proteomic studies, it has also been used to fractionate intact proteins. In a study analyzing the mouse serum proteome, ion exchange chromatography was used to first fractionate the sample at the intact protein level.[65] In this procedure depicted in Figure 7, anion and cation exchange columns were connected in series as a first step in the fractionation of the intact proteins. The columns were then decoupled from each other, and proteins eluted separately from each. These fractions were then tryptically digested, and each was separated using SCX chromatography. Final analysis of the peptide mixture was carried out by RP-HPLC/MS. This method resulted in the identification of almost 5000 unique proteins.

In another study, proteins from human plasma were separated as intact proteins, not peptides, into 30 fractions using sequential linear gradient elution ion exchange chromatography.[66] A 1.1 mL sample of human plasma (equivalent to ~20 mg of total protein) was injected onto an SCX column and eluted using a salt gradient. The gradient mobile phase was made up of solvent A (50 mM phosphate buffer at pH 2.5) and solvent B (2 M NaCl in 50 mM phosphate buffer at pH 2.5), with both solvents containing 5% ACN (v/v). Elution of the proteins was carried out sequentially using six different linear gradients as follows: 0% B, 10% B, 20% B, 40% B, 80% B, and 100% B corresponding to 0, 200, 400, 800, 1600, and 2000 mM NaCl buffers. Samples were collected into tubes every 8 min (from 0 to 40 min), resulting in a total of 30 samples. Each protein fraction was then digested into tryptic peptides and analyzed by RP-HPLC MS/MS. A total of 1292 unique proteins were identified.

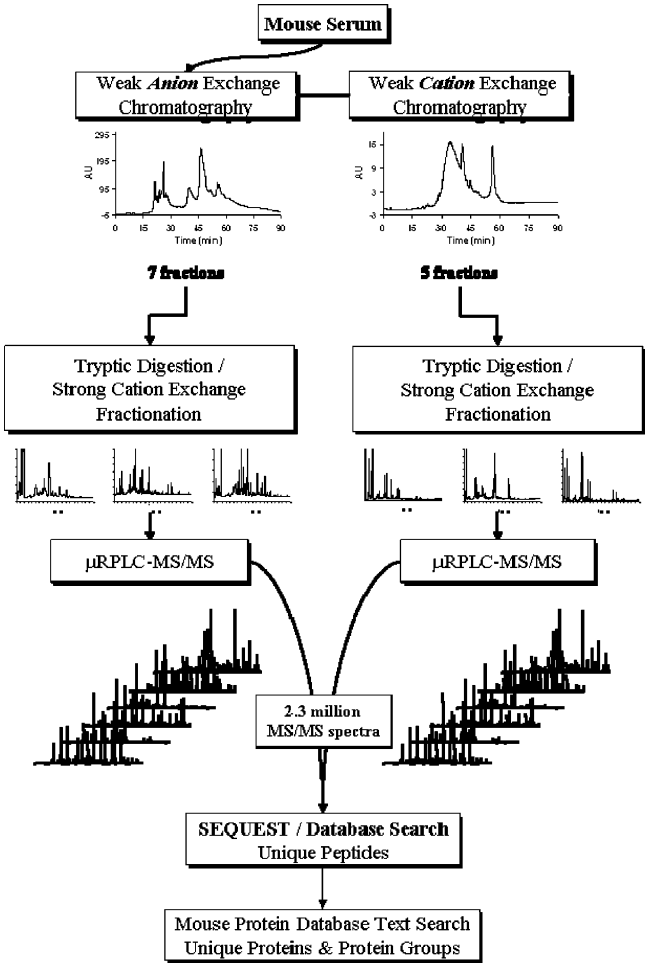


Figure 7. Schematic representation of the experimental design utilized to characterize the mouse serum proteome. Serum was initially fractionated at the protein level using weak anion and cation exchange chromatography columns in series. After separation of the serum, the columns were decoupled from one another and proteins were eluted from each. After tryptically digesting the collected fractions, each was fractionated using strong cation exchange (SCX). The SCX fractions were analyzed by microcapillary reversed-phase liquid chromatography/tandem mass spectrometry.

3.4.6. Capillary Electrophoresis

Capillary electrophoresis (CE) is the general term that encompasses capillary zone electrophoresis (CZE), micellar electrokinetic chromatography (MEKC), capillary isotachophoresis (cITP), and IEF.[67] Capillary electrophoresis, similar to HPLC, is a liquid separation technique. Unlike HPLC, separations in CE are carried out in narrow fused silica capillaries filled with a buffer. The inner surface of the fused silica capillaries is permanently or dynamically coated to prevent the proteins from adsorbing to the inner surface. Capillary electrophoresis is a robust analytical technique that can be used for protein/peptide separation. The advantages of CE over 2D-GE and HPLC are its simplicity and speed of analysis. In addition, it is automated, requires small sample volumes (nanoliters vs microliters for HPLC), and is not limited by the charge of the protein. Separations by CE, as in HPLC and 2D-GE, can be performed as part of a multidimensional fractionation method when analyzing complex proteome mixtures.[68]

Two-dimensional chromatographic methods that employ RP-HPLC in the first dimension and CE in the second dimension are a powerful combination because they combine two high resolving orthogonal techniques with different mechanisms of separation. In comparison to HPLC, CE has many advantages, including higher column efficiency, speed, and simplicity. Since the CE column is constructed from an open tubular fused silica capillary, contamination due to carry-over effects is minimized. Moreover, the time between consecutive CE experiments is much shorter than that of LC-based separations with gradient elution, because the CE capillary column does not require re-equilibration between analyses.

Since narrow capillaries are used for CE separations, the sample size injected is limited and the optical path length for on-column detection is extremely short. These factors serve to affect the detection sensitivity, especially when dilute solutions are used. To overcome such problems, many methods have been developed for on-column sample enrichment for enhancement of detection sensitivity.[69] These include different stacking procedures that are the result of either the manipulation of differences in the electrophoretic mobility of analytes at the boundary of two buffers with differing resistivities or the partitioning of analytes into a stationary or pseudostationary phase. A number of different techniques have been used, including field-amplified sample stacking, large-volume sample stacking, pH-mediated sample stacking, on-column isotachophoresis, and chromatographic preconcentration, whereby a plug of chromatographic material, C-18 or affinity material, is placed at the injection side of the column. Sample stacking for micellar electrokinetic chromatography has been used for on-column sample enrichment.[70] These procedures can enrich the sample by up to 1000-fold.[69,70]

Besides enrichment techniques designed to increase the concentration of the sample that is loaded onto a CE column, more selective devices have also been used to focus on a specific protein or class of proteins. One of the earliest examples of specific protein enrichment was demonstrated in the analysis of human cardiac troponin I (cTnI), a diagnostic marker for myocardial infarction.[71] In this approach, termed precolumn affinity capillary electrophoresis (PA-CE), a packed bed of porous silica to which a monoclonal antibody to cTnI was covalently linked was integrated into the CE column. Injection of serum onto this column, followed by the necessary washing steps to remove nonspecifically bound proteins and other impurities, allowed cTnI levels in the femtomolar range to be detected. Conducting CE-based immunoaffinity studies provided a number of advantages, including the ability to use MS for antigen detection, which eliminates many uncertainties related to antibody cross-reactivity. The CE platform also provides the automation and throughput necessary to assay a large number of serum or plasma samples in a clinical setting.

4. Comparative Analysis of Serum Proteomes

As described above, developments made in sample preparation, chromatographic fractionation, and MS have permitted thousands of proteins to be identified within single studies. If proteomics, however, is to deliver on the promise of novel diagnostic and therapeutic biomarkers for disease states, methods to compare proteins found within serum/plasma samples obtained from different patients are necessary. The primary characteristic that is compared in the analysis of serum/plasma samples is protein relative abundance. While the ultimate goal of routine biomarker discovery has yet to

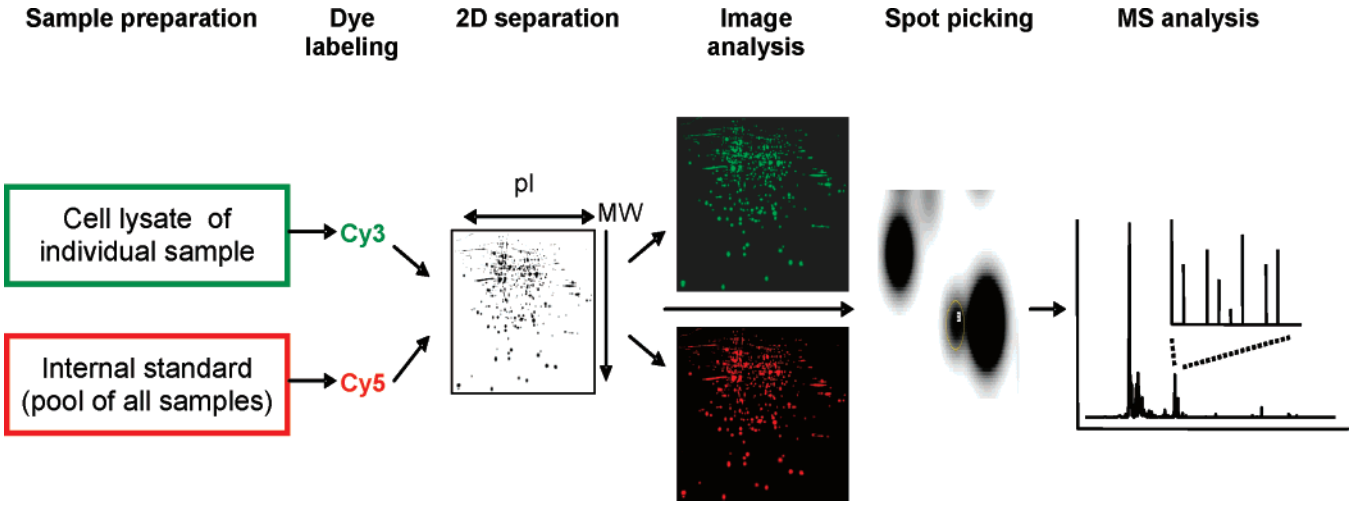


Figure 8. Schematic showing the comparative analysis of proteomes using two-dimensional difference in-gel electrophoresis (2D-DIGE). Proteins showing differential fluorescent intensity (i.e., Cy3 versus Cy5) are extracted from the gel and identified using mass spectrometry.

be achieved, significant developments in the way these complex proteome samples are compared have been made in recent years.

4.1. Two-Dimensional Polyacrylamide Gel Electrophoresis

Since the fractionation capabilities of 2D-PAGE have been discussed previously in this manuscript, suffice it to say for this section that this technique has also been a workhorse for conducting quantitative comparisons of serum/plasma samples.[72] In this technique, proteomes extracted from different sources (e.g., matched control and disease-affected patients) are individually separated using 2D-PAGE. The gels are stained (usually with Coomassie blue or silver stain) to visualize the resolved proteins. After aligning the spots, the relative intensities of the individual proteins between gels are measured and those that appear to be more abundant in one gel compared to the other are excised. The gel spot is digested using trypsin, and the resultant peptides are analyzed using MS. The protein is identified from the raw MS data by matching either the masses or the tandem MS spectra of the resultant peptides to data obtained from an *in silico* analysis of an appropriate genomic or proteomic database. The staining component inherent to 2D-PAGE analysis provides a direct method by which to visualize changes in protein abundances between complex proteome samples. While the throughput of 2D-PAGE is comparatively slow, it does have the advantage that only those spots that appear differentially abundant need to be analyzed by MS.

A 2D-PAGE approach was recently applied to compare plasma samples obtained from patients with severe acute respiratory syndrome (SARS) and healthy individuals.[73] Twenty-two plasma samples from four different SARS patients were separated by 2D-PAGE using a narrow range immobilized pH gradient (IPG) strip (pH 4−7), and the resulting profiles were compared to 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 to the healthy controls. Many of the proteins up-regulated in plasma from SARS patients can be classified as acute phase proteins (APPs) 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 in the plasma of approximately 36% of SARS patients but in only 10% of patients with fever. This rate of detection is higher than that found in human immunodeficiency virus (HIV) patients, suggesting that peroxiredoxin II may function as a useful serum biomarker for SARS infection.

4.1.1 Two-Dimensional Differential In-Gel Electrophoresis

One of the general difficulties in comparing samples analyzed on different 2D-PAGE gels is the inability to perfectly align protein spots between gels. While many advances in software alignment tools have been made, sometimes they cannot overcome the inherent irreproducibility that is present within almost every analytical technique. To overcome this limitation, 2D differential gel electrophoresis (DIGE) was developed in 1997 by Unlü et al.[74] This approach allows for up to three different proteome samples to be separated on a single 2D-PAGE gel. This coseparation ensures accurate quantitation of the same spots from up to three different samples on the same gel, eliminating any error related to gel misalignment.[75] In a typical 2D-DIGE experiment, as shown in Figure 8, proteins extracted from three different samples, healthy, diseased, and internal control (a pooled sample comprising equal amounts of the proteome extracted from the healthy and diseased samples), are covalently labeled with three different fluorophores, 1-(5-carboxypentyl)-1′-propylindocarbocyanine halide *N*-hydroxysuccinimidyl ester (Cy3), 1-(5-carboxypentyl)-1′-methylindodicarbocyanine halide *N*-hydroxysuccinimidyl ester (Cy5), and 3-(4-carboxymethyl)phenylmethyl-3′-ethyloxacarbocyanine halide *N*-hydroxysuccinimidyl ester (Cy2). The control sample serves as an internal standard, enabling both inter- and intragel matching. The control sample should contain every protein present across all samples in an experiment. This means that every protein in the experiment has a unique signal in the internal standard, which is used for direct quantitative comparisons within each gel and to normalize quantitative abundance values for each protein between gels. Equal concentrations of the differentially labeled proteomes and the control sample are mixed, applied to a single gel, and coseparated in the same 2D-PAGE experiment. Scanning the gel at Cy2, Cy3, and Cy5 excitation wavelengths using a fluorescence imager allows visualization

of the differentially labeled proteins. The images are then merged and analyzed using image analysis software, which enables differences between the abundance levels of proteins to be compared. Proteins of interest are excised off the gel, enzymatically digested, and identified using MS. Since it is performed using a single gel plate, 2D-DIGE is more reproducible and accurate than 2D-PAGE, which requires each proteome sample to be fractionated on a separate gel.[76]

The advantages of using 2D-DIGE over 2D-PAGE for proteomic analysis of the serum and plasma proteomes are as follows: separation of the two protein extracts of interest employing the same gel improves reproducibility; it is more economical in terms of materials, since 50% fewer gels are required; differences in protein expression between two different samples of proteins are easier to compare by DIGE and are more accurately imaged; DIGE requires less time to detect the protein spots because the labeling reaction in DIGE is faster than visualization using staining methods; and it is the method of choice when the absolute protein expression level between two biological samples is the primary target.[77]

One of the first studies utilizing 2D-DIGE for the comparative analysis of serum samples was for the comparison of samples obtained from patients with pancreatic cancer to those obtained from matched controls.[78] In the first stage of the analysis, high abundance proteins (i.e., albumin, immunoglobulins, transferrin, haptoglobin, and antitrypsin) were depleted from the serum samples. Serum samples from three groups (pooled internal standard; cancer; and healthy, matched controls) were labeled with Cy2, Cy3, and Cy5, respectively. Serum samples from three individuals with pancreatic cancer and three individuals without cancer were compared. After analysis of the serum samples using 2D-DIGE, a total of 56 protein spot features were found to be significantly increased and 43 significantly decreased in cancer serum samples. These spot features were excised from the gel, digested with trypsin, and analyzed by MS. The MS analysis resulted in the identification of 24 and 17 unique proteins that were increased and decreased, respectively, in cancer serum samples. Western blot analysis was conducted to confirm the comparative levels of several of these proteins in the pancreatic cancer serum samples. As is critical to every biomarker discovery project, an independent series of serum samples from 20 patients with pancreatic cancer and 14 controls were used to validate the 2D-DIGE/MS results. This independent study confirmed increased levels of apolipoprotein E, α-1-antichymotrypsin, and inter-α-trypsin inhibitor in samples obtained from patients with pancreatic cancer.

4.2. Stable-Isotope Labeling for Quantitative Proteomics

As for the identification of proteins with complex mixtures, investigators have also been developing methods to do quantitative comparisons without having to utilize 2D gels. One of the popular methods is the use of stable-isotope tagging.[79] A popular method of stable-isotope tagging, the use of isotope-coded affinity tags (ICATs),[80] which represents a good model for most of these types of studies, is shown in Figure 9. Proteomes are extracted from two comparative samples and are then labeled with functionally and chemically identical reagents (in this case the ICAT reagents) that differ in their mass (i.e., 9.03 Da) based on their stable-isotope content (i.e., nine carbon-13 atoms in the heavy ICAT reagent in place of carbon-12 atoms in the light version). Once the proteins are differentially labeled, the two proteome samples are combined and digested into tryptic peptides. These peptides are then passed over an avidin column to extract out the stable-isotope tagged peptides. The ICAT reagent is unique in that it has iodoacetamide and biotin groups at opposite ends, resulting in the modification of cysteinyl residues and the ability to reclaim these peptides using avidin chromatography. The biotin portion is then removed from the peptides and they are analyzed through a combination of multidimensional chromatography coupled directly on-line with data-dependent MS/MS. The mass spectrometer is operated in such a way that an MS scan is used to quantitate the relative abundance of the peptide within the different samples and MS/MS is used to identify the peptide in the same experiment. The net result is a list of identified proteins with a measure of their relative abundance between the samples being compared. Other stable-isotope labeling approaches utilizing both chemical modification and metabolic labeling have also been developed.[81] While slightly different than the ICAT method, they all use stable isotopes and ultimately result in the same types of data sets.

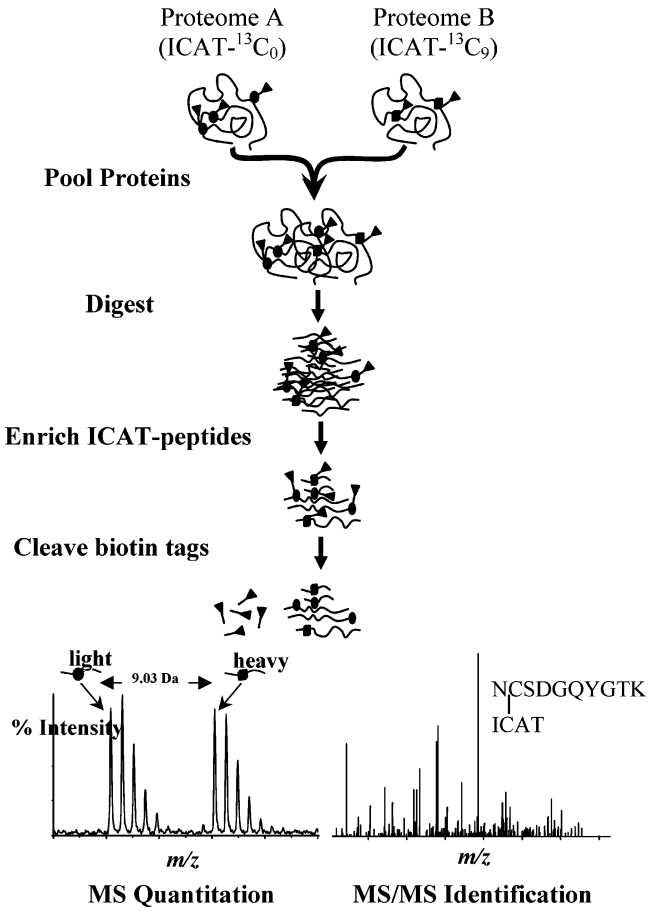


Figure 9. Quantitative proteomics using isotope-coded affinity tags (ICATs). Comparative proteome samples are labeled with chemically identical reagents that differ in their carbon isotope content (i.e., nine carbon-12 atoms for the light reagent and nine carbon-13 atoms for the heavy reagent). After chemically modifying the proteins, the proteomes are combined and digested into tryptic peptides, and the ICAT-labeled peptides are extracted using avidin chromatography. The ICAT-labeled peptides are then analyzed by reversed-phase liquid chromatography coupled directly on-line with a mass spectrometer operating in a data-dependent tandem mass spectrometry (MS/MS) mode. This instrument method enables the relative quantitation of the peptide in the two samples to be measured in the MS mode as well as the identification of the peptides to be determined through data acquired by MS/MS.

While ICAT-labeling has been used extensively in proteomic research, the number of applications to serum/plasma has been limited. In a recent study employing ICAT-labeling, sera obtained from six pediatric patients with severe traumatic brain injury (TBI) (heavy ICAT-labeled) were compared to a pooled sample of sera from healthy adults (light ICAT-labeled).[82] Ninety-five proteins were found to be differentially abundant in the TBI serum samples compared to the pooled control. While most of these proteins are involved in inflammation, innate immunity, and early stress/defense response and therefore are not particularly useful as biomarkers, several low abundant proteins such as Toll receptors, signaling kinases, serine/threonine-protein kinases, transcription factors (serum response factor, golgin 45, myocyte-specific enhancer factor 2B), proteases (pappalysin-2 precursor, MMP-9), and proteins involved in response to oxidative-stress were also identified. Overall, these changes reflect a massive defense response characterized by the recruitment of proteins involved in inflammatory and immune pathways. Several brain-specific proteins such as α-enolase, amyloid β4 precursor, α-spectrin, and cleaved microtubule-associated protein tau, which have been previously detected in serum or CSF from TBI, or other types of brain injury, were found at increased levels in pediatric TBI patients.

Another popular isotope-labeling method that has been applied to serum/plasma proteomics is trypsin-mediated O^{16}/O^{18} labeling.[83] This method relies on the hydrolysis reaction that occurs whenever a peptide bond is cleaved using trypsin. If a proteome is tryptically digested in the presence of water in which the oxygen atom has been substituted with a heavy oxygen isotope (i.e., H_2O^{18}), the two oxygen atoms at the carboxy-terminus of the resultant peptides will be displaced with O^{18} atoms. This stable-isotope labeling method requires minimal sample preparation. To conduct the comparative analysis, two samples are lyophilized and one is resuspended in normal H_2O, while the other is resuspended in H_2O^{18}. Trypsin is added to each, and after the samples are fully digested, trypsin activity is quenched and the samples are relyophilized. Both samples are resuspended in H_2O, and equal aliquots are combined. This single aliquot is then analyzed using RPLC-MS/MS. As with ICAT-labeling, doublets of peaks are observed within the MS chromatogram, separated in this case by 4.02 Da (i.e., the mass difference between two atoms of O^{18} and O^{16}). These doublets represent equivalent peptides observed within the two proteome samples being compared. The relative abundances of the peptides are measured based on the peak areas of the individual peptides within the pair.

An example of O^{16}/O^{18} stable isotope labeling is illustrated in the analysis of serum proteomes from control mice and those bearing human Lewis lung carcinoma.[49c] This study resulted in 1647 proteins, which were identified by at least two tryptic peptides, being quantitated. Two-hundred and eleven and 246 of these proteins were measured as being at an increased and decreased level of abundance, respectively, in the serum of mice bearing Lewis lung carcinoma. Among the proteins found to be up-regulated, many have been implicated in cancer progression. In particular, vascular endothelial growth factor receptor 1 (VEGFR-1) was up-regulated over 7-fold in the mouse lung carcinoma model.

Stable-isotope labeling methods have shown the capability of quantitating thousands of proteins in complex biological samples. Unfortunately, they suffer similar disadvantages to 2D-PAGE. They are low-throughput, requiring days to compare two samples. They are generally limited to comparing two samples; however, the development of iTRAQ has allowed up to four samples to be compared simultaneously.[84] Even this number, however, is far too few to generate the level of confidence required to take the next step and attempt to validate the findings. Techniques that use metabolic stable-isotope labeling, although not impossible,[85] are highly impractical for the study of serum/plasma samples. A unique problem observed with O^{16}/O^{18} labeling is the incomplete exchange of both oxygen atoms at the carboxy-terminus of peptides. Many of the peptides undergo incomplete exchange so that a significant population of peptides with only a single O atom is observed. This incomplete incorporation adds a significant uncertainty to the accuracy of the relative abundance measurements. While they have made a major impact in the analysis of cellular and tissue proteomes, stable-isotope labeling methods, particularly ICAT-labeling, have not been widely used in biomarker discovery. The reasons for this are not readily obvious. It is possible that the domination of serum and plasma by a few high abundant proteins impacts the chemical labeling of lower abundant proteins by the stable-isotope reagents.

4.3. Subtractive Proteomics

One common feature of 2D-PAGE and isotope-labeling quantitative proteomics is their lack of throughput. Since most developments in serum/plasma proteomics are aimed toward discovering novel biomarkers, increasing the throughput is going to be critical. For this reason, investigators worked to determine if the data obtained within the MS and MS/MS data could be used to measure relative protein abundance between samples. These methods do not use gels or stable-isotopes but simply rely on quantifying proteins based on either the number of peptides identified for each species or the intensity of individual peptide peaks within the MS chromatogram. In this method, the proteome is extracted from a series of biological samples and digested into tryptic peptides. The tryptic peptides are analyzed using multidimensional chromatography coupled directly on-line with MS.

The acquired data can be analyzed using two different methods. In one method, the relative abundances of the proteins are measured based on the number of peptides identified for that specific protein in the comparative samples.[86] The validity of this hypothesis can be illustrated in the analysis of a single serum sample. If digested serum is analyzed directly by RPLC-MS/MS, a large number of peptides from albumin will be identified. It is unlikely, however, that even one peptide from a low-abundance protein, such as a chemokine, will be identified. This result is due to the high serum concentration of albumin (i.e., ~60 mg/mL) compared to chemokine proteins (i.e., in the ng/mL range). In a practical example, if nine peptides are identified for cancer antigen-125 (CA-125) in serum sample A and only three peptides are identified in serum sample B, the conclusion is made that CA-125 is three times more abundant in sample A.

This approach, known by a variety of terms, including subtractive proteomics, peptide count, etc., is a very attractive method for biomarker discovery because of its inherent simplicity. Except for the depletion of high abundance proteins, it requires minimal sample preparation. Most proteomic laboratories have the capability of identifying

thousands of proteins with serum/plasma, and this quantitative measurement allows an unlimited number of samples to be compared to one another. Like most techniques, however, it also has its disadvantages. Although faster than 2D-PAGE analysis and stable-isotope labeling, it is still relatively low-throughput, requiring several hours just to acquire the raw MS data for each sample. This peptide subtractive method is quantitatively imprecise compared to stable-isotope labeling methods and can only measure differences greater than 3-fold. Low abundance proteins, which are generally only identified through one of two peptides, may not be quantifiable using this method.

A second method of making quantitative measurements from such data sets is to directly compare the peak areas of individual peptides identified in different samples.[87] This type of quantitation requires software to generate selected ion chromatograms of the peptides of interest so that the peak area of each peptide can be determined. Strict analytical care needs to be taken when acquiring data for this type of analysis, as factors such as unequal starting amounts of protein, irreproducible chromatography, etc. can skew the results. Ultimately, since the relative abundances are described based on the protein, abundance ratios are calculated by averaging the peptide peak area ratios for the same protein.

An excellent example of both of these methods combined into one study was published by Richard D. Smith's lab in which peptide peak areas and the number of peptide identifications from 2D-LC-MS/MS analyses were used to garner a quantitative comparison of protein abundances between plasma samples obtained from a human subject prior to (untreated) and 9 h after lipopolysaccharide (LPS) administration (treated).[88] Lipopolysaccharide is an endotoxin released by Gram-negative bacteria that is known to induce inflammatory reactions, such as cytokine production, cell migration, and production of acute-phase proteins. This study sought to quantitate changes in the acute phase plasma proteome in response to LPS administration. The untreated and LPS-treated plasma samples were digested with trypsin, and each sample was fractionated using strong cation exchange chromatography. A total of 50 fractions were collected for each sample, and each of these was analyzed by RPLC-MS/MS. Some of the SCX fractions that had a high peptide content were run twice, resulting in a total of 148 RPLC-MS/MS analyses. Combining both analyses (i.e., treated and nontreated) resulted in a total of 804 unique plasma proteins (not including IgG's) being identified from 5176 unique peptides. Of these, 83% (669 proteins) were identified by at least two unique tryptic peptides.

To determine if the number of peptide identifications for each protein could be used in a quantitative manner, the group plotted the number of peptides identified for 74 specific proteins against their literature documented concentrations in plasma (Figure 10). In general, the correlation was quite good, suggesting peptide hit number is at least semiquantitative. The group also compared the peak areas for peptides that were identified in both samples and used this ratio, along with the number of peptide hits, to identify proteins that were differentially abundant in LPS-treated plasma. As shown in Table 2, eight out of the nine proteins listed for which a protein abundance ratio was determined showed an increase in concentration following LPS administration by both the protein abundance ratios and the ratios of peptide hits. The two computational approaches, however, are generally complementary, as many of the up-regulated proteins were identified in only one of the two methods. This study was one of the first to show that signal intensity and peptide hit count could be used to quantitatively compare protein abundances in biofluids analyzed by LC-MS/MS. Presently, most non-gel-based comparative studies of serum and plasma are conducted using either of these two computational approaches to measure the relative quantitation of proteins in two or more samples.

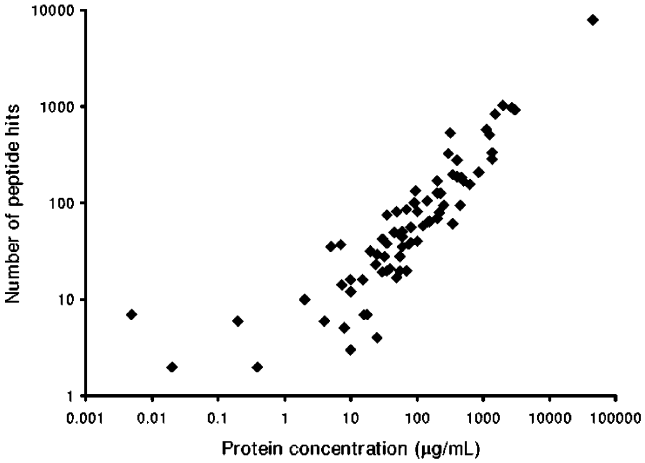


Figure 10. Correlation between the known concentrations of various plasma proteins and the number of peptides identified for each during a multidimensional fractionation/tandem mass spectrometry analysis. Data derived from ref 88.

Table 2. Comparison of the Ratio of Peptide Hits and the Relative Abundance Ratio (Determine by Measuring Peak Areas) for Nine Proteins Observed To Be Up-Regulated in the Comparison of Plasma Taken from a Patient prior to (Untreated) and after Treatment with Lipopolysaccharide[72]

protein	ratio of peptide hits (treated/untreated)	abundance ratio (treated/untreated)
KIAA1009 protein	7	2.9
von Willebrand factor	4.3	1.1
hepatocyte growth factor activator	4	3.8
serum amyloid A2	4	4.3
KIAA1301 protein	4	2.87
lipopolysaccharide binding protein	3.9	2.6
NADH oxidase	2.6	1.95
serum amyloid A1	2.3	5.9
leucine-rich-α-glycoprotein	1.5	2.87

4.4. Proteomic Profiling

The proteomic methods discussed in the previous sections rely on extensive fractionation and MS identification of as many proteins as possible within these complex mixtures. A few years ago, a concept was put forth by a couple of laboratories hypothesizing that a diagnostic fingerprint obtained from a clinical sample could function as a diagnostic or therapeutic indicator, without the need to identify the species that comprise the fingerprint.[89,90] The concept was based on the hypothesis that diseases such as cancer manifest themselves by changing the proteomic content of blood. The hope in proteomic profiling was that the detection of a number of proteins, rather than a single biomarker, may provide higher sensitivities and specificities for diagnosis than those that can be afforded with single markers. This approach offers the advantage that it can quickly discover

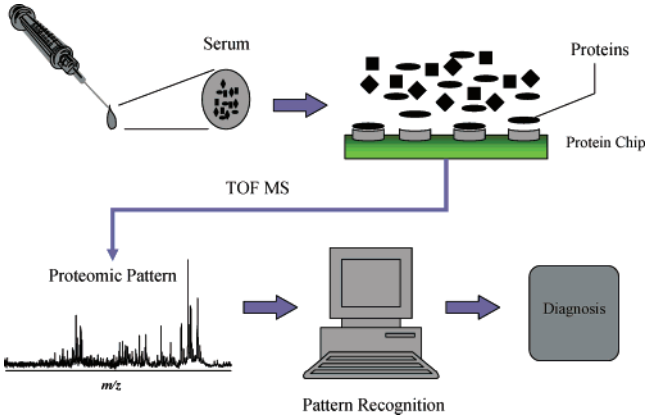


Figure 11. Disease diagnostics using proteomic patterns. The sample drawn from the patient is applied to a protein chip, which is made up of a specific chromatographic surface. After several washing steps and the application of an energy-absorbing molecule, the species that are retained on the surface of the chip are detected using mass spectrometry. The pattern of peaks within the spectrum is analyzed using sophisticated bioinformatic software to diagnose the source of the biological sample.

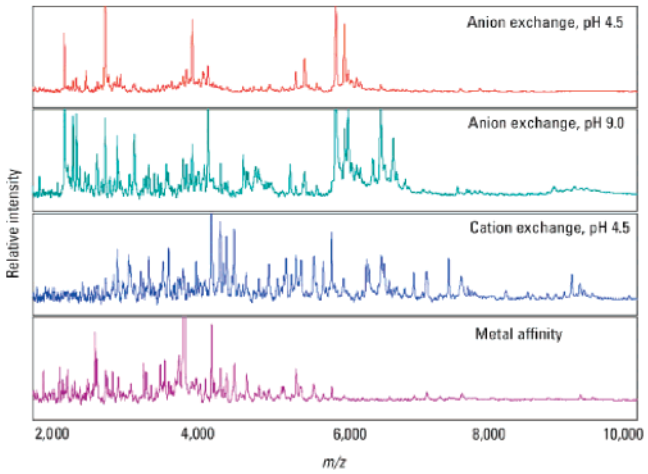


Figure 12. Mass spectrum of human serum obtained using surface enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI TOF/MS) technology employing different chromatographic protein chip surfaces. The top two panels illustrate the effect of pH when using a single chromatographic surface.

proteins as potential biomarkers by examining proteomic profiles, and this method can be transformed to a clinical application to predict susceptibility to certain diseases, obtain an early diagnosis before symptoms emerge, and monitor disease progression and treatment outcome. While proteomic profiling has come under severe criticism in the recent past, its impact has been such that it deserves mentioning within this article.

The first technology utilized to obtain proteomic patterns of serum and plasma samples for diagnostic purposes was surface-enhanced laser desorption ionization (SELDI) time-of-flight mass spectrometry (TOF-MS).[91] This technique combines protein retention chromatography with TOF-MS detection, enabling proteins/peptides to be profiled from complex biological samples, such as cancer cell lines, serum, plasma, urine, nipple aspirate, and tissue extracts with very little sample preparation.[92] The SELDI TOF-MS system is composed of three major components: the protein chip, mass analyzer, and data analysis software.[93] To analyze samples using SELDI TOF-MS (Figure 11), a few (i.e., 1−5) microliters are deposited on the chromatographic surface of the protein chip. Depending on the surface chemistry, proteins are captured on the chromatographic surface by adsorption, partition, electrostatic interaction, or affinity chromatography. After processing the protein chips (i.e., using a series of binding and washing steps), an energy absorbing matrix such as sinapinic acid is deposited to embed the proteins in crystalline form. Finally, the protein chip is analyzed by TOF-MS. The result is a mass spectrum comprised of the *m*/*z* values and intensities of the bound proteins/peptides. While SELDI provides a unique sample preparation platform, its protein detection principle is similar to that of MALDI-MS. The resulting mass spectrum is a function of both the chromatographic chip surface and the experimental conditions used (Figure 12). Based on the spectral output, the analysis software can recognize peaks, compare two or three groups of spectra, and conduct cluster analyses to pinpoint significant protein abundance differences between samples.

The first study that utilized SELDI TOF-MS in cancer research was for the detection of ovarian cancer utilizing serum samples.[89] Serum samples from 50 healthy women and 50 women in different stages of ovarian cancer were used as the training set in the genetic algorithm and clustering bioinformatic software that performed the analysis in two stages: pattern discovery and pattern matching stages. In the pattern discovery stage, the genetic algorithm and clustering tool were used to analyze the SELDI serum mass spectral data obtained from the training set using a hydrophobic chip surface. The *m*/*z* and intensity values were employed to identify a proteomic pattern that discriminated cancer from noncancer cases. In the pattern matching stage, the discriminatory proteomic pattern from the training set was compared with that in the test set of 116 masked serum samples. All 50 blinded ovarian cancer cases and 63 of the 66 nonmalignant cases were correctly identified, yielding 100% sensitivity and 95% specificity. The significantly high sensitivity implied the clinical potential of the proteomic pattern as a diagnostic approach. On the other hand, if the same method was to be applied for cancer screening in the general population, the specificity of the serum patterns would need to be improved to nearly 100% to avoid high false-positive identifications (Rockhill 2002). To achieve higher specificity, a hybrid quadrupole (Qq) TOF-MS instrument equipped with the SELDI protein chip interface was used to acquire high-resolution serum proteomic profiles in ovarian cancer patients. As a result, four optimum discriminatory patterns correctly identified 43 healthy women and 68 ovarian cancer patients, including 18 stage I patients with 100% sensitivity and specificity in the validation test.[94] The study implied that the combination of SELDI and high-resolution mass spectrometry instruments may offer great potential in the field of in vitro diagnostics.

The results presented in the first SELDI TOF/MS manuscript showing the remarkable ability to correctly diagnose serum samples from women with ovarian cancer created a frenzy in the clinical science and proteomic community. This studied spurred a large number of investigations that employed a SELDI TOF-MS-based approach to generate protein profiles of serum for the early detection of cancers such as breast,[95] prostate,[96] cervix,[97] lung,[98] bladder,[99] colon,[100] head and neck,[101] and pancreas.[102] In addition to cancer, SELDI TOF-MS has also been applied to study

proteins related to other maladies, such as Alzheimer's disease,[103] rheumatoid arthritis,[104] and AIDS.[105]

The primary advantage of this proteomic pattern analysis over other MS-based methods is throughput. Hundreds of samples could be processed and analyzed in a single day, allowing large cohorts of clinical samples to be analyzed. The challenges facing proteomic pattern analysis at present are daunting. One of these challenges is the reproducibility of proteomic profiling experiments.[106] It was reported that chip-to-chip coefficients of variation of peak intensities ranged from 10% to 40%.[95] A number of approaches are now routinely incorporated to improve reproducibility, such as including a quality control sample on each chip array and normalizing spectral data. In addition to this inherent irreproducibility is the more disturbing issue of lab-to-lab variability. Laboratories that have conducted studies using the same technological platform to investigate the same cancer condition have found different groups of diagnostic peaks. Another challenge is answering what the bioinformatics analysis of all of these spectra really means. Many different types of algorithms have been brought to bear on proteomic pattern datasets.[107] Unfortunately, the peaks that are selected as diagnostic for a particular disease condition often vary depending on the algorithm selected. Therefore, it is not clear which algorithm is selecting the diagnostically valuable peaks. Finally, the obvious challenge is the identification of the important proteins and peptides that contribute to a diagnostic pattern. While few studies actually proceed to identify the peaks of interest, those that do typically find that they are intact, or fragments of, highly abundant proteins. These proteins are generally acute-phase or inflammatory response proteins and are unlikely to have the necessary specificity to diagnose a particular cancer.

5. Automation

The goal of finding biomarkers in serum and plasma using proteomic technologies such as fractionation and MS is extremely ambitious. Unless specific proteins are targeted, the approaches are largely discovery driven and typically rely on finding a protein(s) that is more abundant in serum or plasma obtained from disease-afflicted individuals than in healthy controls. Unfortunately, there are a number of cumulative factors that make such discovery extremely difficult. First of all, quantitative proteomic results do not have the precision and accuracy found in standardized assays, for instance. This deficiency requires a large number of samples to be analyzed in order to achieve the level of confidence that would be required before even considering taking a potential biomarker into a validation phase. Unfortunately, as can be surmised by the many sections within this review, the throughput associated with proteomic analysis of serum and plasma samples is very low. While tremendous strides have been achieved in the past few years, comprehensive analysis of a single complex proteome still requires several days from start to finish.

One potential solution to this problem is automation. The development of automated approaches for gene sequencing was absolutely critical to the success of the Human Genome Project. Proteomics, however, is not as straightforward as sequencing. The Human Genome Project required continuous sequencing of four bases, while proteomics requires the identification of tens of thousands (to possibly >1 000 000) of peptides with a variety of different physicochemical properties. Present technology requires these peptides to be separated prior to MS analysis. While separation, in the form of chromatography and electrophoresis, can be automated, it still must be done over a time domain that provides the opportunity for adequate resolution of the analytes. The interfacing of separations with MS has also been automated, allowing unattended acquisition of the raw MS data. While MS instrumentation is able to sequence an individual peptide very rapidly, the analysis of a serum or plasma sample will require tens of thousands of sequencing events. It is not difficult to see that the more pressing issue in proteomics today for clinical discovery is not automation but increasing the throughput. Historically, the solution to increasing throughput in proteomics has been to increase the number of analytical platforms recording the necessary data.

Another area in serum and plasma proteomics that requires automation is sample preparation. Many of the steps required to prepare a sample for MS analysis are still done manually. Robotic systems have been developed for a number of different areas of proteomics, such as structural proteomics and protein arrays. Probably the most utilized robotic system in MS-based proteomics is the use of workstations that pick spots from 2D-PAGE gels and automatically conduct the in-gel digestion to prepare the protein for MS analysis. Automation will fill the need to standardize sample preparation for clinical samples so that results obtained in different biomarker discovery projects can be more effectively compared. Unfortunately, until many of the issues related to throughput are resolved, there will not be a strong push for improved automated sample preparation methods.

6. Conclusions

Probably no samples have been the subject of more proteomic studies in the last 5 years than serum and plasma. The reason is clear; the discovery of novel biomarkers that are indicative of diseases such as early stage cancer would have a huge impact on public health. If the discovery of biomarkers is used as the sole metric, then it must be concluded that serum and plasma proteomics has been somewhat of a disappointment. Looking closer, however, reveals quite the opposite. As described in this review, it was only 4 years ago that the first manuscript came out describing the identification of over 300 proteins in plasma. State-of-the-art MS-based proteomic laboratories are presently capable of identifying thousands of proteins in serum and plasma samples. The need to garner this level of coverage has spurred the investment of time and money into the development of powerful analytical methods and instrumentation to meet the challenge of discovering useful diagnostic biomarkers. While the development of faster and more sensitive mass spectrometers has obviously had a major impact on the ability to conduct serum and plasma proteomics, improvements in sample preparation methods such as high abundance protein depletion and chromatography have also played key roles. The next major hurdle to overcome if scientists hope to translate all of the information made available by proteomic technologies is to find intelligent solutions to determine which proteins discovered in a comparative analysis have the greatest likelihood of being validated as useful biomarkers.

7. 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 N01-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the United States Government.

8. References

(1) (a) Engle, W. A.; Schreiner, R. L.; Baehner, R. L. *Semin. Perinatol.* **1983**, *7*, 184. (b) http://www.columbia.edu/itc/hs/medical/selective/AdvClinicalPathology/2005 /lecture/DiabeticCareKaplanBW.pdf.
(2) (a) http://en.wikipedia.org/wiki/Plasma. (b) http://www.redcross.org.
(3) (a) http://www.answers.com/topic/blood-serum-1. (b) http://www.greenfacts.org/glossary/abc/blood-serum.htm.
(4) (a) Wingren, C.; Borrebaeck, C. A. *OMICS* **2006**, *10*, 411. (b) Cekaite, L.; Hovig, E.; Sioud, M. *Methods Mol. Biol.* **2007**, *360*, 335. (c) Sharp, V.; Utz, P. J. *Nat. Clin. Rheumatol.* **2007**, *3*, 96. (d) Hu, S.; Loo, J. A.; Wong, D. T. *Proteomics* **2006**, *6*, 6326.
(5) (a) http://en.wikipedia.org/wiki/Blood. (b) http://health.howstuffworks.com/blood.html.
(6) (a) Zhang, H.; Liu, A. Y.; Loriaux, P.; Wollshceid, B.; Zhou, Y.; Watts, J. D.; Aebersold, R. A. *Mol. Cell. Proteomics* **2007**, *6*, 64. (b) Adkins, J. N.; Varnum, S. M.; Auberry, K. J.; Moore, R. J.; Angell, N. H.; Smith, R. D.; Springer, D. L.; Pounds, J. G. *Mol. Cell. Proteomics* **2002**, *1*, 947−955. (c) Chan, K. C.; Lucas, D. A.; Schaefer, C. F.; Xiao, Z.; Janini, G. M.; Buetow, K. H.; Issaq, H. J.; Veenstra, T. D.; Conrads, T. P. *Clin. Prot.* **2004**, *1*, 101.
(7) Lum, G.; Gambino, S. R. *Am. J. Clin. Pathol.* **1974**, *61*, 108.
(8) (a) Kimball, D. B.; Rickles, F. R.; Gockerman, J. P.; Hattler, B. G.; Light, J. A.; Conrad, M. E. *J. Lab. Clin. Med.* **1976**, *87*, 868. (b) Spence, G. M.; Graham, A. N.; Mulholland, K.; Maxwell, P.; McCluggage, W. G.; Sloan, J. M.; McGuigan, J. A. *Int. J. Biol. Markers* **2002**, *17*, 119.
(9) Benoy, I.; Salgado, R.; Colpaert, C.; Weytjens, R.; Vermeulen, P. B.; Dirix, L. Y. *Clin. Breast Cancer* **2002**, *2*, 311.
(10) Hsieh, S. Y.; Chen, R. K.; Pan, Y. H.; Lee, H. L. *Proteomics* **2006**, *6*, 3189.
(11) (a) West-Nielsen, M.; Hogdall, E. V.; Marchiori, E.; Hogdall, C. K.; Schou, C.; Heegaard, N. H. *Anal. Chem.* **2005**, *77*, 5114. (b) Baumann, S.; Ceglarek, U.; Fiedler, G. M.; Lembcke, J.; Leichtle, A.; Thiery. J. *Clin. Chem.* **2005**, *51*, 973.
(12) Verrills, N. M. *Clin. Biochem. Rev.* **2006**, *27*, 99.
(13) Drake, S. K.; Bowen, R. A.; Remaley, A. T.; Hortin, G. L. *Clin. Chem.* **2004**, *50*, 2398.
(14) (a) Luque-Garcia, J. L.; Neubert, T. A. *J. Chromatogr., A* **2007**, *1153*, 259. (b) Villanueva, J.; Philip, J.; Chaparro, C. A.; Li, Y.; Toledo-Crow, R.; DeNoyer, L.; Fleisher, M.; Robbins, R. J.; Tempst, P. *J. Proteome Res.* **2005**, *4*, 1060.
(15) (a) Banks, R. E.; Stanley, A. J.; Cairns, D. A.; Barrett, J. H.; Clarke, P.; Thompson, D.; Selby, P. J. *Clin. Chem.* **2005**, *51*, 1637. (b) Omenn, G. S.; States, D. J.; Adamski, M.; Blackwell, T. W.; Menon, R.; Hermjakob, H.; Apweiler, R.; Haab, B. B.; Simpson, R. J.; Eddes, J. S.; Kapp, E. A.; Moritz, R. L.; Chan, D. W.; Rai, A. J.; Admon, A.; Aebersold, R.; Eng, J.; Hancock, W. S.; Hefta, S. A.; Meyer, H.; Paik, Y. K.; Yoo, J. S.; Ping, P.; Pounds, J.; Adkins, J.; Qian, X.; Wang, R.; Wasinger, V.; Wu, C. Y.; Zhao, X.; Zeng, R.; Archakov, A.; Tsugita, A.; Beer, I.; Pandey, A.; Pisano, M.; Andrews, P.; Tammen, H.; Speicher, D. W.; Hanash, S. M. *Proteomics* **2005**, *5*, 3226. (c) Tammen, H.; Schulte, I.; Hess, R.; Menzel, C.; Kellman, M.; Schulz-Knappe, P. *Proteomics* **2005**, *5*, 3414.
(16) Zolg, W. *Mol. Cell Proteomics* **2006**, *5*, 1720.
(17) (a) www.plasmaproteome.org. (b) Anderson, N. L.; Anderson, N. G. *Mol. Cell. Proteomics* **2002**, *11*, 845. (c) Tirumalai, R. S.; Chan, K. C.; Prieto, D. A.; Issaq, H. J.; Conrads, T. P.; Veenstra, T. D. *Mol. Cell. Proteomics* **2003**, *2*, 1096.
(18) Anderson, L., *J. Physiol.* **2005**, *563*, 23.
(19) (a) Fujii, K.; Nakano, T.; Kawamura, T.; Usui, F.; Bando, Y.; Wang, R.; Nishimura, T. *J. Proteome Res.* **2004**, *3*, 712. (b) Jin, W. H.; Dai, J.; Li, S. J.; Xia, Q. C. *J. Proteome Res.* **2005**, *4*, 613.
(20) Zolotarjova, N.; Martosella, J.; Nicol, G.; Bailey, J.; Boyes, B. E.; Barrett, W. C. *Proteomics* **2005**, *5*, 3304.
(21) Darde, V. M.; Barderas, M. G.; Vivanco, F. *Methods Mol. Biol.* **2007**, *357*, 351.
(22) Gong, Y.; Li, X.; Yang, B.; Ying, W.; Li, D.; Zhang, Y.; Dai, S.; Cai, Y.; Wang, J.; He, F.; Qian, X. *J. Proteome Res.* **2006**, *5*, 1379.
(23) Whiteaker, J. R.; Zhang, H.; Eng, J. K.; Fang, R.; Piening, B. D.; Feng, L. C.; Lorentzen, T. D.; Schoenherr, R. M.; Keane, J. F.; Holzman, T.; Fitzgibbon, M.; Lin, C.; Zhang, H.; Cooke, K.; Liu, T.; Camp, D. G., II; Anderson, L.; Watts, J.; Smith, R. D.; McIntosh, M. W.; Paulovich, A. G. *J. Proteome Res.* **2007**, *6*, 828.
(24) (a) Gundry, R. L.; Fu, Q.; Jelinek, C. A.; Van Eyk, J. E.; Cotter, R. J. *Proteomics Clin. Appl.* **2007**, *1*, 73. (b) Zhou, M.; Lucas, D. A.; Chan, K. C.; Issaq, H. J.; Petricoin, E. F., III; Liotta, L. A.; Veenstra, T. D.; Conrads, T. P. *Electrophoresis* **2004**, *25*, 1289.
(25) Gorg, A.; Weiss, W.; Dunn, M. J. *Proteomics* **2004**, *4*, 3665.
(26) (a) Taoka, M.; Wakamiya, A.; Nakayama, H.; Isobe, T. *Electrophoresis* **2000**, *21*, 1872. (b) Imin, N.; De Jong, F.; Mathesius, U.; van Noorden, G.; Saeed, N. A.; Wang, X. D.; Rose, R. J.; Rolfe, B. G. *Proteomics* **2004**, *4*, 1883.
(27) (a) Kelleher, N. L. *Anal. Chem.* **2004**, *76*, 197A. (b) Whitelegge, J.; Halgand, F.; Souda, P.; Zabrouskov, V. *Expert Rev. Proteomics* **2006**, *3*, 585. (c) Han, X.; Jin, M.; Breuker, K.; McLafferty, F. W. *Science* **2006**, *314*, 109.
(28) Tiselius, A. The Moving Boundary Method of Studying the Electrophoresis of Proteins, Inaugural Dissertation, Almqvist & Wiksells, AB, Upsala, 1930.
(29) Raymond, S.; Weintraub, L. *Science* **1959**, *130*, 711.
(30) Laemmli, U. K. *Nature* **1970**, *227*, 680.
(31) Righetti, P.; Drysdale, J. W. *Biochim. Biophys. Acta* **1971**, *236*, 17.
(32) Smithies, O.; Poulik, M. D. *Nature* **1956**, *177*, 1033.
(33) Hermans, P. E.; McGuckin, W. F.; McKenzie, B. F.; Baird, E. D. *Proc. Staff Meet. Mayo Clin.* **1960**, *35*, 792.
(34) Ashton, G. C. *Nature* **1957**, *180*, 917.
(35) Raymond, S.; Aurell, B. *Science* **1962**, *138*, 152.
(36) Raymond, S. *Ann. N. Y. Acad. Sci.* **1964**, *121*, 350.
(37) Margolis, J.; Kenrick, K. G. *Nature* **1969**, *221*, 1056.
(38) Slater, G. G. *Anal. Biochem.* **1965**, *24*, 215.
(39) Wright, G. L. *Am. J. Clin. Pathol.* **1972**, *57*, 173.
(40) O'Farrell, P. H. *J. Biol. Chem.* **1975**, *250*, 4007.
(41) Anderson, L.; Anderson, N. G. *Proc. Nat. Acad. Sci. U.S.A.* **1977**, *74*, 5421.
(42) Manabe, T.; Tachi, K.; Kojima, K.; Okuyama, T. *J. Biochem.* **1979**, *85*, 649.
(43) Kolin, A. *J. Chem. Phys.* **1954**, *22*, 1628.
(44) Bjellqvist, B.; Ek, K.; Righetti, P. G.; Gianazza, E.; Gorg, A.; Westermeier, R.; Postel, W. *J. Biochem. Biophys. Methods* **1982**, *6*, 317.
(45) Hoving, S.; Voshol, H.; van Ostrum, J. *Electrophoresis* **2000**, *21*, 2617.
(46) Righetti, P. G.; Gianazza, E.; Gelfi, C.; Chiari, M.; Sinha, P. K. *Anal. Chem.* **1989**, *61*, 1602.
(47) (a) Miller, I.; Crawford, J.; Gianazza, E. *Proteomics* **2006**, *6*, 5385. (b) Westermeier, R.; Marouga, R. *Biosci. Rep.* **2005**, *25*, 19. (c) Moritz, B.; Meyer, H. E. *Proteomics* **2003**, *3*, 2208. (d) Miura, K. *Proteomics* **2003**, *3*, 1097. (e) Patton, W. F. *J. Chromatogr., B* **2002**, *771*, 3.
(48) Pieper, R.; Gatlin, C. L.; Makusky, A. J.; Russo, P. S.; Schatz, C. R.; Miller, S. S.; Su, Q.; McGrath, A. M.; Estock, M. A.; Parmar, P. P.; Zhao, M.; Huang, S. T.; Zhou, J.; Wang, F.; Esquer-Blasco, R.; Anderson, N. L.; Taylor, J.; Steiner, S. *Proteomics* **2003**, *3*, 1345.
(49) (a) Chertov, O.; Simpson, J. T.; Biragyn, A.; Conrads, T. P.; Veenstra, T. D.; Fisher, R. J. *Expert Rev. Proteomics* **2005**, *2*, 139. (b) Gu, X.; Deng, C.; Yan, G.; Zhang, X. *J. Proteome Res.* **2006**, *5*, 3186. (c) Hood, B. L.; Lucas, D. A.; Kim, G.; Chan, K. C.; Blonder, J.; Issaq, H. J.; Veenstra, T. D.; Conrads, T. P.; Pollet, I.; Karsan, A. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1221.
(50) (a) Nakagawa, Y.; Yip, T.; Belew, M.; Porath, J. *Anal. Biochem.* **1981**, *75*, 168. (b) Ji, J.; Chakraborty, A.; Geng, M.; Zhang, X.; Amini, A.; Bina, M.; Regnier, F. *J. Chromatogr., B* **2000**, *745*, 197.
(51) (a) Posewitz, M. C.; Cooper, J. *Anal. Chem.* **1999**, *71*, 2883. (b) Bodenmiller, B.; Mueller, L. N.; Mueller, M.; Domon, B.; Aebersold, R. *Nat. Methods* **2007**, *4*, 231.
(52) (a) Pinske, M. W.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. *Anal. Chem.* **2004**, *76*, 3935. (b) Kweon, H. K.; Hakannson, K. *Anal. Chem.* **2006**, *78*, 1743.
(53) (a) Barnard, G.; Bayer, E.; Wilchek, M.; Amir-Zaltsman, Y. *Meth. Enzymol.* **1986**, *133*, 284−288. (b) Coligan, J. E. In *Current Protocols in Protein Science*; Coligan, J. E., Dunn, B., Ploegh, H., Speicher, D., Wingfield, P., Eds.; John Wiley and Sons: New York, 1999; p A.1A.
(54) Schachter, H. *Glycobiology* **1991**, *1*, 453.
(55) Qui, R.; Zhang, X.; Regnier, F. E. *J. Chromatogr., B* **2007**, *845*, 143.
(56) (a) Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2199. (b) Villen, J.; Beausoleil, S. A.; Gerber, S. A.; Gygi, S. P. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1488. (c) Chi, A.; Huttenhower, C.; Geer, L. Y.;

Coon, J. J.; Syka, J. E.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2086.

(57) Harry, J. L.; Wilkins, M. R.; Herbert, B. R.; Packer, N. H.; Gooley, A. A.; Williams, K. L. *Electrophoresis* **2000**, *21*, 1071.

(58) (a) Hanning, K. *Electrophoresis* **1982**, *3*, 235. (b) Bier, M. *Electrophoresis* **1998**, *19*, 1057. (c) Herbert, B.; Righetti, P. G. *Electrophoresis* **2000**, *21*, 3639. (d) Kachman, M. K.; Wang, H.; Schwartz, D. R.; Cho, K. R.; Lubman, D. M. *Anal. Chem.* **2002**, *74*, 1779.

(59) (a) Xiao, Z.; Conrads, T. P.; Lucas, D. A.; Janini, G. M.; Schaefer, C. E.; Buetow, K. H.; Issaq, H. J.; Veenstra, T. D. *Electrophoresis* **2004**, *25*, 128. (b) Chan, K. C.; Lucas, D. A.; Schaefer, C. F.; Xiao, Z.; Janini, G. M.; Buetow, K. H.; Issaq, H. J.; Veenstra, T. D.; Conrads, T. P. *Clin. Proteomics* **2004**, *1*, 101.

(60) Tu, C. J.; Dai, J.; Li, S. J.; Sheng, Q. H.; Deng, W. J.; Xia, Q. C.; Zeng, R. *J. Proteome Res.* **2005**, *4*, 1265.

(61) (a) Jacobs, S. *Methods Biochem. Anal.* **1966**, *14*, 177. (b) Cerrai, E. *Chromatogr. Rev.* **1964**, *6*, 129. (c) Small, H. Half a century in separation science. In *A Century of Separation Scienc*e; Issaq, H. J., Ed.; Marcel Dekker, Inc.: New York, pp 283−292. (d) Dizdaroglu, M. Separation of peptides by high-performance ion-exchange chromatography. In *Handbook of HPLC for the Separation of Amino Acids, Peptides, and Proteins*; CRC Press, Inc.: Boca Raton, FL, pp 23−39.

(62) (a) Nibbering, N. M. *Mass Spectrom. Rev.* **2006**, *25*, 962. (b) Jemal, M.; Xia, Y. Q. *Curr. Drug Metab.* **2006**, *7*, 491. (c) Liu, H.; Lin, D.; Yates, J. R., III. *Biotechniques* **2002**, *32*, 898. (d) Strege, M. A. *J. Chromatogr., B* **1999**, *725*, 67. (e) Hancock, W. S.; Apffel, A.; Chakel, J.; Souders, C.; M'Timkulu, T.; Pungor, E., Jr.; Guzzetta, A. W. *Methods Enzymol.* **1996**, *271*, 403.

(63) (a) Link, A. J.; Carmack, E.; Yates, J. R. *Int. J. Mass Spectrom. Ion Processes* **1997**, *160*, 303. (b) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G.; Morris, D. R.; Garvik, B. M.; Yates, J. R. *Nat. Biotechnol.* **1999**, *17*, 676. (c) Washburn, M. P.; Wolters, D.; Yates, J. R. *Nat. Biotechnol.* **2001**, *19*, 242.

(64) (a) Raida, M.; Schulz-Knape, P.; Heine, G.; Forssmann, W. G. *J. Am. Mass Spectrom.* **1999**, *10*, 45. (b) Richter, R.; Schulz-Knape, P.; Schrader, M.; Standker, L.; Jurgens, M.; Tammen, H.; Forssmann, W. G. *J. Chromatogr., B* **1999**, *726*, 25. (c) Schrader, M.; Jurgens, M.; Hess, R.; Schulz-Knape, P.; Raida, M.; Forssmann, W. G. *J. Chromatogr., A* **1997**, *766*, 139.

(65) Hood, B. L.; Zhou, M.; Chan, K. C.; Lucas, D. A.; Kim, G. J.; Issaq, H. J.; Veenstra, T. D.; Conrads, T. P. *J. Proteome Res.* **2005**, *4*, 1561.

(66) Jin, W. H.; Dai, J.; Li, S. J.; Xia, Q. C.; Zou, H. F.; Zeng, R. *J. Proteome Res.* **2005**, *4*, 613.

(67) Issaq, H. J.; Janini, G. M.; Chan, K. C.; Veenstra, T. D. *J. Chromatogr., A* **2004**, *1053*, 37.

(68) (a) Michels, D. A.; Hu, S.; Schoenherr, R. M.; Eggertson, M. J.; Dovichi, N. J. *Mol. Cell. Proteomics* **2002**, *1*, 69. (b) Kraly, J. R.; Jones, M. R.; Gomez, D. G.; Dickerson, J. A.; Harwood, M. M.; Eggerston, M.; Paulson, T. G.; Sanchez, C. A.; Odze, R.; Feng, Z.; Reid, B. J.; Dovichi, N. J. *Anal. Chem.* **2006**, *78*, 2977.

(69) (a) Guzman, N. A.; Phillips, T. M. *Anal. Chem.* **2005**, *77*, 61A. (b) Guzman, N. A.; Stubbs, R. J. *Electrophoresis* **2001**, *22*, 3602. (c) Bonneil, E.; Waldron, K. C. *J. Chromatogr., A* **1999**, *736*, 273. (d) Stroink, T.; Paarlberg, E.; Waterval, J. C. M.; Buly, A.; Underberg, W. J. M. *Electrophoresis* **2001**, *22*, 2375. (e) Janini, G. M.; Zhou, M.; Yu, L-R.; Blonder, J.; Gignac, M.; Conrads, T. P.; Issaq, H. J.; Veenstra, T. D. *Anal. Chem.* **2003**, *75*, 5984. (f) Osburn, D. M.; Weiss, D. J.; Lunte, C. E. *Electrophoresis* **2000**, *21*, 2768.

(70) Kim, J. B.; Quirino, J. P.; Otsuka, K.; Terabe, S. *J. Chromatogr., A* **2001**, *916*, 123.

(71) Dalluge, J. J.; Sander, L. C. *Anal. Chem.* **1998**, *70*, 5339.

(72) (a) Elrick, M. M.; Walgren, J. L.; Mitchell, M. D.; Thompson, D. C. *Basic Clin. Pharmacol. Toxicol.* **2006**, *98*, 432. (b) Sheta, E. A.; Appel, S. H.; Goldknopf, I. L. *Expert Rev. Proteomics* **2006**, *3*, 45. (c) Pawlik, T. M.; Kuerer, H. M. *Int. J. Fertil. Women's Med.* **2005**, *50*, 212.

(73) Chen, J. H.; Chang, Y. W.; Yao, C. W.; Chiueh, T. S.; Huang, S. C.; Chien, K. Y.; Chen, A.; Chang, F. Y.; Wong, C. H.; Chen, Y. J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17039.

(74) Unlü, M.; Morgan, M. E.; Minden, J. S. *Electrophoresis* **1997**, *18*, 2071.

(75) (a) Herbert, B.; Righetti, P. G. *Electrophoresis* **2000**, *21*, 3639. (b) Righetti, P. G.; Castagna, A.; Antonucci, F.; Piubelli, C.; Cecconi, D.; Campostrini, N.; Antonioli, P.; Astner, H.; Hamdan, *J. Chromatogr., A* **2004**, *1051*, 3.

(76) Lilley, K. S.; Friedman, D. B. *Expert Rev. Proteomics* **2004**, *1*, 401.

(77) (a) Hoom, E. J.; Hoffert, J. D.; Knepper, M. A. *Nephron Physiol.* **2006**, *104*, 61. (b) Marouga, R.; David, S.; Hawkins, E. *Anal. Bioanal. Chem.* **2005**, *382*, 669. (c) Swatton, J. E.; Prabakaran, S.; Karp, N. A.; Lilley, K. S.; Bahn, S. *Mol. Psychiatry* **2004**, *9*, 128.

(78) Yu, K. H.; Rustqi, A. K.; Blair, I. A. *J. Proteome Res.* **2005**, *4*, 1742.

(79) (a) Miyagi, M.; Rao, K. C. *Mass Spectrom. Rev.* **2007**, *26*, 121. (b) Leitner, A.; Lindner, W. *Proteomics* **2006**, *6*, 5418. (c) Ong, S. E.; Mann, M. *Nat. Chem. Biol.* **2005**, *1*, 252. (d) Heck, A. J.; Krijsveld, J. *Expert Rev. Proteomics* **2004**, *1*, 317. (e) Beynon, R. J.; Pratt, J. M. *Mol. Cell. Proteomics* **2005**, *4*, 857.

(80) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994.

(81) (a) Heller, M.; Mattou, H.; Manzel, C.; Yao, X. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 104. (b) Ong, S. E.; Foster, L. J.; Mann, M. *Methods* **2003**, *29*, 124.

(82) Haqqani, A. S.; Hutchison, J. S.; Ward, R.; Stanimirovic, D. B. *J. Neurotrauma* **2007**, *24*, 54.

(83) (a) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 2836. (b) Gruhler, A.; Matthiesen, R. *Methods Mol. Biol.* **2006**, *367*, 139.

(84) Zieske, L. R. *J. Exp. Bot.* **2006**, *57*, 1501.

(85) Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Matthews, D. E.; Yates, J. R., III. *Anal. Chem.* **2004**, *76*, 4951.

(86) (a) Dieguez-Acuna, F. J.; Gerber, S. A.; Kodama, S.; Elias, J. E.; Beausoleil, S. A.; Faustman, D.; Gygi, S. P. *Mol. Cell. Proteomics* **2005**, *4*, 1459. (b) Oh, P.; Li, Y.; Yu, J.; Durr, E.; Krasinska, K. M.; Carver, L. A.; Testa, J. E.; Schnitzer, J. E. *Nature* **2004**, *429*, 629.

(87) Wiener, M. C.; Sachs, J. R.; Deyanova, E. G.; Yates, N. A. *Anal. Chem.* **2004**, *76*, 6085.

(88) Qian, W. J.; Jacobs, J. M.; Camp, D. G., II; Monroe, M. E.; Moore, R. J.; Gritsenko, M. A.; Calvano, S. E.; Lowry, S. F.; Xiao, W.; Moldawer, L. L.; Davis, R. W.; Tompkins, R. G.; Smith, R. D. *Proteomics* **2005**, *5*, 572.

(89) Petricoin, E. F.; Ardekani, A. M.; Hitt, B. A.; Levine, P. J.; Fusaro, V. A.; Steinberg, S. M.; Mills, G. B.; Simone, C.; Fishman, D. A.; Kohn, E. C.; Liotta, L. A. *Lancet* **2002**, *359*, 572.

(90) Qu, Y.; Adam, B. L.; Yasui, Y.; Ward, M. D.; Cazares, L. H.; Schellhammer, P. F.; Feng, Z.; Semmes, O. J.; Wright, G. L., Jr. *Clin. Chem.* **2002**, *48*, 1835.

(91) Kuwata, H.; Yip, T. T.; Yip, C. L.; Tomita, M.; Hutchens, T. W. *Biochem. Biophys. Res. Commun.* **1998**, *245*, 764.

(92) Issaq, H. J.; Conrads, T. P.; Prieto, D. A.; Tirumalai, R.; Veenstra, T. D. *Anal. Chem.* **2003**, *75*, 148A.

(93) (a) James, P. *Biotechniques* **2002**, *4*, 12. (b) Domon, B.; Alving, K.; He, T.; Ryan, T. E.; Patterson, S. D. *Curr. Opin. Mol. Ther.* **2002**, *4*, 577. (c) Pusch, W.; Flocco, M. T.; Leung, S. M.; Thiele, H.; Kostrzewa, M. *Pharmacogenomics* **2003**, *4*, 463. (d) Espina, V.; Dettloff, K. A.; Cowherd, S.; Petricoin, E. F., III; Liotta, L. A. *Expert Opin. Biol. Ther.* **2004**, *4*, 83.

(94) Conrads, T. P.; Fusaro, V. A.; Liotta, L. A.; Petricoin, E. F., III; Veenstra, T. D. *Endocr.-Relat. Cancer* **2004**, *11*, 163.

(95) (a) Wulfkuhle, J. D.; McLean, K. C.; Paweletz, C. P.; Sgroi, D. C.; Trock, B. J.; Steeg, P. S.; Petricoin, E. F., III. *Proteomics* **2001**, *1*, 1205. (b) Li, J.; Zhang, Z.; Rosenzweig, J.; Wang, J. J.; Chan, D. W. *Clin. Chem.* **2002**, *48*, 1296.

(96) (a) Adam, B. L.; Qu, Y.; Davis, J. W.; Ward, M. D.; Clements, M. A.; Cazares, L. H.; Semmes, O. J.; Schellhammer, P. F.; Yasui, Y.; Feng, Z.; Wright, G. L., Jr. *Cancer Res.* **2002**, *62*, 3609. (b) Petricoin, E. F., III; Ornstein, D. K.; Paweletz, C. P.; Ardekani, A.; Hackett, P. S.; Hitt, B. A.; Velassco, A.; Trucco, C.; Wiegand, L.; Wood, K.; Simone, C. B.; Levine, P. J.; Linehan, W. M.; Emmert-Buck, M. R.; Steinberg, S. M.; Kohn, E. C.; Liotta, L. A. *J. Natl. Cancer Inst.* **2002**, *94*, 1576. (c) Banez, L. L.; Prasanna, P.; Sun, L.; Ali, A.; Zou, Z.; Adam, B. L.; McLeod, D. G.; Moul, J. W.; Srivastava, S. *J. Urol.* **2002**, *170*, 442. (d) Li, J.; White, N.; Zhang, Z.; Rosenzweig, J.; Mangold, L. A.; Partin, A. W.; Chan, D. W. *J. Urol.* **2004**, *171*, 1782.

(97) Wong, Y. F.; Cheung, T. H.; Lo, K. W.; Wang, V. W.; Chan, C. S.; Ng, T. B.; Chung, T. K.; Mok, S. C. *Cancer Lett.* **2004**, *211*, 227.

(98) Xiao, X.; Liu, D.; Tabg, Y.; Guo, F.; Xia, L.; Liu, J.; He, D. *Dis. Markers* **2003**, *19*, 33.

(99) Vlahou, A.; Schellhammer, P. F.; Mendrinos, S.; Patel, K.; Kondylis, F. I.; Gong, L.; Nasim, S.; Wright, G. L., Jr. *Am. J. Pathol.* **2001**, *158*, 1491.

(100) Lawrie, L. C.; Curran, S.; McLeod, H. L.; Fothergill, J. E.; Murray, G. L. *Mol. Pathol.* **2001**, *54*, 253.

(101) Wadsworth, J. T.; Somers, K. D.; Cazares, L. H.; Malik, G.; Adam, B. L.; Stack, B. C., Jr.; Wright, G. L., Jr.; Semmes, O. J. *Clin. Cancer Res.* **2004**, *10*, 1625.

(102) Koopmann, J.; Zhang, Z.; White, N.; Rosenzweig, J.; Fedarko, N.; Jagannath, S.; Canto, M. I.; Yeo, C. J.; Chan, D. W.; Goggins, M. *Clin. Cancer Res.* **2004**, *10*, 860.

(103) (a) Carrette, O.; Demalte, I.; Scherl, A.; Yalkinoglu, O.; Corthals, G.; Burkhard, P.; Hochstrasser, D. F.; Sanchez, J. C. *Proteomics* **2003**, *3*, 1486. (b) Lewczuk, P.; Esselmann, H.; Groemer, T. W.; Bibl, M.;

Maler, J. M.; Steinacker, P.; Otto, M.; Kornhuber, J.; Wiltfang, J. *Biol. Psychiatry* **2002**, *55*, 524.
(104) Uchida, T.; Fukawa, A.; Uchida, M.; Fujita, L.; Saito, K. *J. Proteome Res.* **2002**, *1*, 495.
(105) Zhang, L.; Yu, W.; He, T.; Yu, J.; Caffrey, R. E.; Dalmasso, E. A.; Fu, S.; Pham, T.; Mei, J.; Ho, J. J.; Zhang, W.; Lopez, P.; Ho, D. D. *Science* **2002**, *298*, 995.
(106) Baggerly, K. A.; Morris, J. S.; Coombes, K. R. *Bioinformatics* **2004**, *20*, 777.
(107) Yasui, Y.; Pepe, M.; Thompson, M. L.; Adam, B. L.; Wright, G. L., Jr.; Qu, Y.; Potter, J. D.; Winget, M.; Thornquist, M.; Feng, Z. *Biostatistics* **2003**, *4*, 449.

CR068287R