# Quantitative Analysis of the Low Molecular Weight Serum Proteome Using $^{18}O$ Stable Isotope Labeling in a Lung Tumor Xenograft Mouse Model

Brian L. Hood, David A. Lucas, Grace Kim, King C. Chan, Josip Blonder, Haleem J. Issaq, Timothy D. Veenstra, and Thomas P. Conrads

Laboratory of Proteomics and Analytical Technologies, National Cancer Institute at Frederick, SAIC-Frederick, Inc., Frederick, Maryland, USA

Ingrid Pollet and Aly Karsan

British Columbia Cancer Research Center, Vancouver, British Columbia, Canada

With advancements in the analytical technologies and methodologies in proteomics, there is great interest in biomarker discovery in biofluids such as serum and plasma. Current hypotheses suggest that the low molecular weight (LMW) serum proteome possesses an archive of clipped and cleaved protein fragments that may provide insight into disease development. Though these biofluids represent attractive samples from which new and more accurate disease biomarkers may be found, the intrinsic person-to-person variability in these samples complicates their discovery. Mice are one of the most extensively used animal models for studying human disease because they represent a highly controllable experimental model system. In this study, the LMW serum proteome was compared between xenografted tumor-bearing mice and control mice by differential labeling utilizing trypsin-mediated incorporation of the stable isotope of oxygen, $^{18}O$. The digestates were combined, fractionated by strong cation exchange chromatography, and analyzed by nanoflow reversed-phase liquid chromatography coupled online with tandem mass spectrometry, resulting in the identification of 6003 proteins identified by at least a single, fully tryptic peptide. Almost 1650 proteins were identified and quantitated by two or more fully tryptic peptides. The methodology adopted in this work provides the means for future quantitative measurements in comparative animal models of disease and in human disease cohorts. (J Am Soc Mass Spectrom 2005, 16, 1221–1230) © 2005 American Society for Mass Spectrometry

Biomarker investigations are increasingly exploiting the developing technologies and methodologies of proteomics with the hope of discovering better indicators of the onset or progression of diseases. Many investigations aim to discover biomarkers in various biofluids (i.e., serum, plasma, and cerebrospinal fluid) as these samples are easily obtainable, contain high concentrations of proteins, and perfuse key tissues, undoubtedly endowing these fluids with protein species indicative of diseased states [1–4]. Hence, detection of changes in the abundance or characteristics of different proteins in biofluids may provide the means for disease detection, enabling earlier intervention in disease progression.

Much of the promise of proteomics for identification of biomarkers in complex biological fluids can be traced to seminal studies performed in John Yates' laboratory. This laboratory was the first to develop a high-throughput method for large-scale proteome analysis, which combined multidimensional liquid chromatography, tandem mass spectrometry (MS/MS), and database searching by the SEQUEST algorithm [5]. This method became popular under the acronym MudPIT (i.e., multidimensional protein identification technology). MudPIT, or variations thereof, has been used in countless studies to characterize complex proteome samples, with the typical goal of obtaining as much possible information concerning the sample of interest. This technology has ushered in an era

Published online June 24, 2005

Address reprint requests to Dr. T. D. Veenstra, Laboratory of Proteomics and Analytical Technologies, National Cancer Institute at Frederick, SAIC-Frederick, Inc., P.O. Box B, Frederick, MD 21702, USA. E-mail: veenstra@ncifcrf.govE-mail: conrads@ncifcrf.gov

or Dr. T. P. Conrads, Mass Spectrometry Center, Laboratory of Proteomics and Analytical Technologies, National Cancer Institute at Frederick, SAIC-Frederick, Inc., P.O. Box B, Frederick, MD 21702, USA.


1044-0305/05/$30.00
doi:10.1016/j.jasms.2005.02.005

Received November 24, 2004
Revised February 2, 2005
Accepted February 2, 2005

where upwards of 4000 proteins can be characterized in a single sample [6, 7] and enables proteomics with the technology to live up to its expectations of comprehensive coverage of cellular proteins.

Perhaps the largest challenge posed in characterizing the serum proteome arises from its inherently wide dynamic range of protein concentration. Indeed, proteins in serum have been estimated to be present across >9 orders of magnitude in concentration, in which only 22 species comprise almost 99% of the total protein mass [8]. Logically, depletion of one or more of the high abundance, high molecular weight proteins would assist the detection and identification of low abundance species. In an effort to decrease this dynamic range of protein concentration, numerous affinity purification and extraction methods have been developed to enrich the smaller, less-abundant proteins and peptides in complex biofluids [9, 10]. Recently, a simple methodology was developed to deplete serum of high molecular weight proteins by size partitioning through the use of centrifugal ultrafilters [2]. The result was recovery of what was coined the low molecular weight (LMW) serum proteome. Microcapillary reversed-phase liquid chromatography ($\mu$RPLC) coupled online with tandem mass spectrometry (MS/MS) enabled the identification of greater than 880 peptides in this fraction, corresponding to ~340 proteins, which demonstrated the efficacy of this method for the removal of large abundant proteins and the enrichment of the LMW serum proteome.

The work described herein expands upon the LMW serum proteome enrichment methodology previously developed in our laboratory to enable quantitative measurements of peptide relative abundances between serum samples utilizing trypsin-mediated $^{18}$O-labeling. Specifically a quantitative analysis of the LMW serum proteome from a murine model of human lung carcinoma utilizing $^{18}$O stable isotope labeling and MS has resulted in the identification and quantitation of 6003 proteins from 8992 unique tryptic peptides. This analysis establishes the basis for high throughput quantitative analysis of the LMW serum proteome for disease biomarker discovery.

## Experimental

### Materials

Ammonium bicarbonate ($NH_4HCO_3$), ammonium formate ($NH_4HCO_2$), formic acid (HCOOH), trifluoroacetic acid (TFA), and dithiothreitol (DTT) were purchased from Sigma (St. Louis, MO). Porcine sequencing grade modified trypsin was obtained from Promega (Madison, WI). High performance liquid chromatography (HPLC)-grade acetonitrile ($CH_3CN$) was obtained from EMD Chemicals Inc. (Gibbstown, NJ). All buffers and reagents were used as supplied from the manufacturer and prepared in double distilled water using a NANOPure Diamond water system (Barnstead International, Dubuque, IA).

### Mouse Tumor Xenograft Model

Mice were xenografted subcutaneously with $2 \times 10^6$ Lewis lung carcinoma cells in 100 $\mu$l of medium or with medium alone. Thirteen days following tumor implantation or sham injection, blood was obtained by cardiac puncture from 5 tumor-bearing mice and 5 control (medium injected) mice, placed into Microtainer tubes (Becton Dickinson, Franklin Lakes, NJ), allowed to clot, and then centrifuged at $2500 \times g$ for 10 min. Serum was aliquoted and stored at $-80$ °C until further use. Tumors ranged in size from 250 to 300 mm at the time of serum collection.

### Low Molecular Weight Fractionation of Pooled Mouse Serum Samples

Serum samples from each group of mice were pooled and 500 $\mu$l of each sample was diluted 20-fold with 50 mM $NH_4HCO_3$, pH 8.4, 20% $CH_3CN$ and incubated on ice for 90 min with occasional gentle mixing. The low molecular weight proteins were obtained by centrifugal ultrafiltration using Centriplus ultrafilters (Milipore, Billerica, MA) at $750 \times g$ at 4 °C until approximately 80% (8 mL) of the sample filtered through the membrane. The filtrates were removed and stored at $-20$ °C until further use. The retentates (approximately 2 mL) were further diluted to 10 mL with 50 mM $NH_4HCO_3$, pH 8.4, 20% $CH_3CN$ and incubated on ice for an additional 2 h with occasional mixing. These samples were fractionated as above overnight at $500 \times g$ at 4 °C. Filtrates from both fractionations were pooled, lyophilized and stored at $-80$ °C.

### Tryptic Digestion

Total protein in each LMW serum sample was quantified using the BCA protein assay (Pierce, Rockford, IL) and removal of the high molecular weight proteins was confirmed by SDS-PAGE analysis of the pooled filtrates and retentates. Equivalent amounts of each sample (approximately 300 $\mu$g) were diluted to 500 $\mu$L with 25 mM $NH_4HCO_3$, pH 8.4. Freshly prepared 1 M DTT was added to a final concentration of 10 mM and the samples were boiled for 10 min. Trypsin was added at an enzyme to protein ratio of 1:50 and each sample was incubated for 16 h at 37 °C after which a second aliquot of trypsin was added (enzyme to protein ratio of 1:50) and the samples were incubated an additional 8 h at 37 °C.

### Trypsin-Mediated $^{18}$O Labeling of Mouse LMW Serum

The tryptically digested LMW serum proteins were lyophilized and resuspended in 145 $\mu$L of either $^{16}$O (control) or $^{18}$O (lung carcinoma) water and 40 $\mu$L of methanol (20% (vol/vol) final concentration). Fifteen $\mu$L of trypsin, resuspended in the appropriately labeled water, was added to each sample at an enzyme to

protein ratio of 1:50 and the samples were incubated for 16 h at 37 °C. Concentrated TFA was added to a final concentration of 0.4% (vol/vol) and the mixtures were boiled for 10 min. The samples were pooled and lyophilized prior to strong cation exchange (SCX) fractionation.

### *Strong Cation Exchange Fractionation*

Strong cation exchange fractionation was performed using a HP 1090 LC system (Agilent Technologies, Palo Alto, CA) equipped with a polysulfoethyl A column (4.6 × 200 mm, 5 μm, 300 Å pore size, PolyLC, Inc., Columbia, MD). The isotopically labeled mouse LMW serum protein pool (~600 μg) was reconstituted in 10 mL 25% acetonitrile/0.1% formic acid and loaded onto the column and washed with 2% mobile phase B for 12 min at a flow rate of 1.5 mL/min. The following $NH_4HCO_3/CH_3CN$ multi-step gradient was used to elute the peptides from the column at a flow rate of 1 mL/min: 2% mobile phase B (25% $CH_3CN$, 0.5 M $NH_4HCO_3$, pH 3.0) for 3 min, followed by 23% B in 70 min, then to 60% B in 15 min, and finally to 100% B in 1 min and maintained at 100% B for 7 min. Mobile phase A was 25% $CH_3CN$. Peptide separation was monitored by fluorescence (280 nm excitation/350 nm emission). Fractions were collected every minute, lyophilized and reconstituted in 30 μL of 0.1% (vol/vol) TFA prior to nanoflow (nano) reversed-phase (RP) liquid chromatography (LC)-tandem mass spectrometry (MS/MS) analysis.

### *Nanoflow Reversed-Phase Liquid Chromatography Tandem Mass Spectrometry*

Nanoflow RPLC analyses were performed using an Agilent 1100 nanoLC system (Agilent Technologies) coupled online to a linear ion trap (LIT) mass spectrometer (LTQ, Thermo Electron, San Jose, CA) with the nanoelectrospray interface supplied by the manufacturer. Separations were performed using 75 μm i.d. × 360 μm o.d. × 10 cm long fused silica capillary columns (Polymicro Technologies, Phoenix, AZ) that were slurry-packed in-house with 3 μm, 300 Å pore size C-18 silica-bonded stationary phase (Vydac, Hysperia, CA). After injecting 7 μL of sample, the column was washed for 30 min with 98% mobile phase A (0.1% formic acid in water) at a flow rate of 500 nL/min. Peptides were eluted using a linear gradient of 2% mobile phase B (0.1% formic acid in acetonitrile) to 40% mobile phase B in 110 min, then to 98% B in an additional 30 min, all at a constant flow rate of 200 nL/min.

The LIT mass spectrometer was operated in a data dependent MS/MS mode in which each full MS scan was followed by five MS/MS scans where the five most abundant peptide molecular ions are dynamically selected for collision-induced dissociation (CID) using a normalized collision energy of 35%. Dynamic exclusion was utilized to minimize redundant acquisition of peptides previously selected for MS/MS. The heated capillary temperature and electrospray voltage were set at 160 °C and 1.5 kV, respectively.

### *Bioinformatic Analysis*

Tandem mass spectra were searched against the UniProt mouse proteomic database (05-12-04 release) from the European Bioinformatics Institute (http://www.ebi.ac.uk/) with SEQUEST operating on an 18 node Beowulf cluster. Peptides were searched using tryptic cleavage constraints and a dynamic 4 Da modification on the C-terminus. For a peptide to be considered legitimately identified, it had to achieve stringent charge state and proteolytic cleavage-dependent cross correlation ($X_{corr}$) scores of 1.9 for $[M + H]^{1+}$, 2.2 for $[M + 2H]^{2+}$, 3.1 for $[M + 3H]^{3+}$, and a minimum delta correlation ($\Delta C_n$) of 0.08.

## Results and Discussion

It is predicted that serum is likely to contain an abundant archive of pathophysiological information because of its constant perfusion through tissues [11]. A number of published and ongoing proteomic investigations have been, or are being, conducted of serum that will usher in a new paradigm of understanding of this complex and vital clinical sample. An interesting feature of many of the diagnostic studies that have employed various proteomic strategies is that, consistently, proteins (and/or their fragments) that are considered diagnostic for the disease of interest possess low molecular weights, suggesting that the LMW serum proteome may contain an unexplored archive of histopathological information and provide useful biomarkers for disease detection. Indeed, LMW serum proteins, peptides, and other small components have been associated with pathological conditions such as cancer [12], diabetes [13], cardiovascular, and infectious diseases [14]. A simple method for the enrichment of the LMW components in serum has previously been developed in our laboratory and used in conjunction with multidimensional chromatographic fractionation and MS analysis to better characterize this unexplored fraction [2]. We have expanded this methodology in the present work to incorporate trypsin-mediated $^{18}O$ isotope labeling to enable quantitative measurements of the LMW serum proteome from a mouse model of human lung carcinoma. The experimental design (Figure 1) relies on the ultrafiltration of two (or more) serum samples, individually, to recover the LMW proteomes to be quantitatively compared. In the present case, each of the recovered LMW proteome samples were first digested to completion with trypsin and lyophilized to dryness. The control LMW proteome sample was resuspended in a buffered solution prepared in $H_2{}^{16}O$ and the lung carcinoma LMW proteome sample was resuspended in the same buffered solution

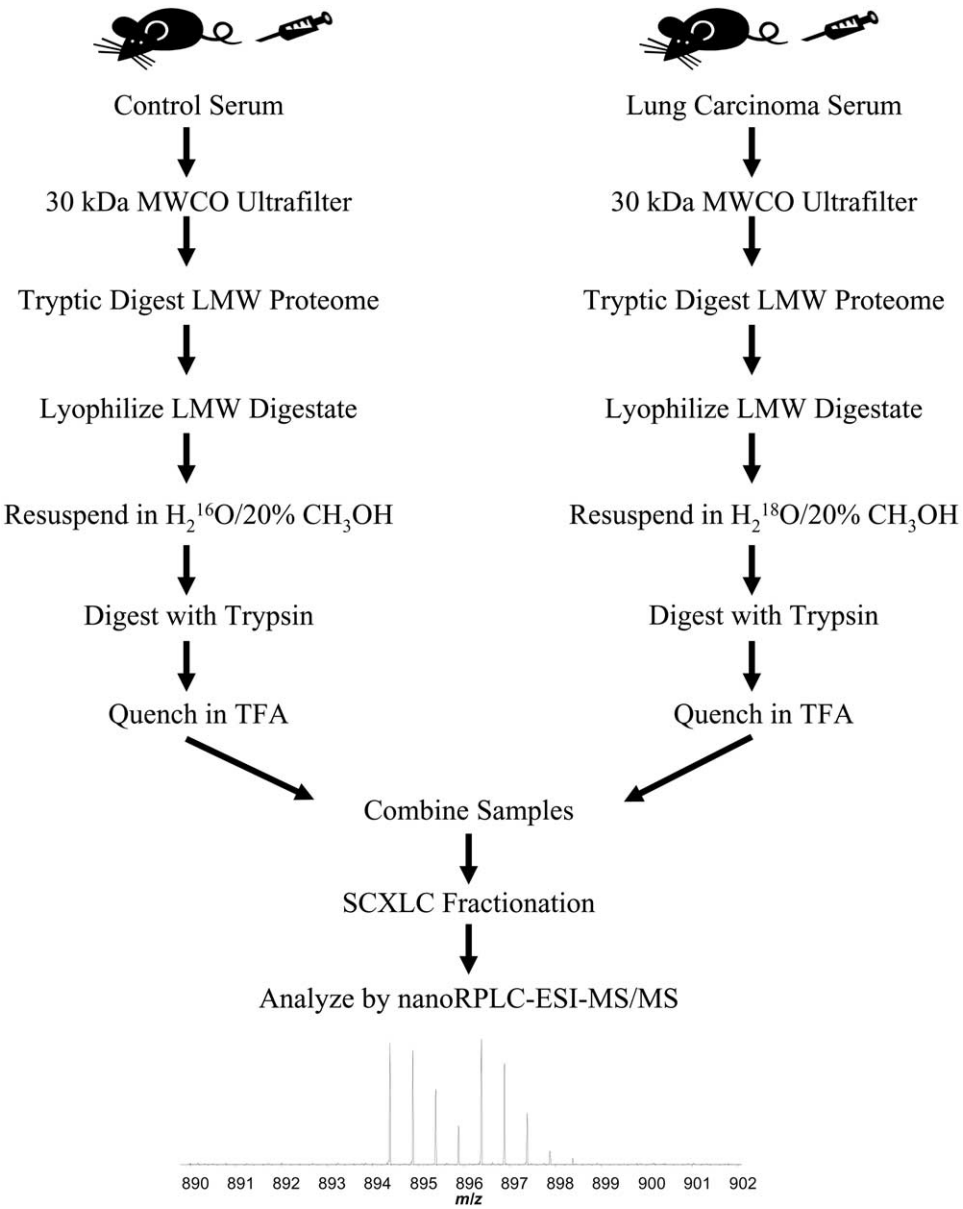


**Figure 1.** Schematic representation of the experimental design utilized for trypsin-mediated $^{18}O$-labeling of the mouse LMW serum proteome. The LMW serum proteome from a pool of control mouse serum and a pool from mice with lung carcinoma were each individually size fractionated after acetonitrile incubation and utilizing molecular weight cutoff ultrafilters. Each of the LMW serum proteomes were fully digested with trypsin and lyophilized. The control LMW digestate was resuspended in buffered methanol and the lung carcinoma LMW digestate was resuspended in buffered methanol prepared in $H_2{}^{18}O$ and separately incubated with trypsin. Each of the LMW samples were quenched in TFA, combined, fractionated by strong cation exchange liquid chromatography, and analyzed by nanoRPLC-MS/MS.

prepared in $H_2{}^{18}O$. The addition of trypsin to these samples serves to mediate the exchange of two equivalents of $^{16}O$ at the carboxy-terminus of each peptide for two equivalents of $^{18}O$ in the sample reconstituted in $H_2{}^{18}O$. Two complete enzyme turnovers in the presence of $H_2{}^{18}O$ results in a 4 Da increase in mass of each tryptic peptide [15–17]. The LMW proteome digestates were combined, fractionated into 96 samples using SCXLC, and each analyzed by nanoRPLC-MS/MS.

An important consideration in the extraction of the LMW serum proteome is the extent to which high abundant components are depleted. As shown in lane 2 of Figure 2, the most highly abundant proteins in serum have molecular weights in excess of 45 kDa. Ultrafiltration of serum through a membrane with a 30 kDa molecular weigh cutoff almost completely depletes these proteins from the sample as shown through coomassie staining of a SDS-PAGE gel (Figure 2, lane 3). Although peptides assigned to high abundance pro-

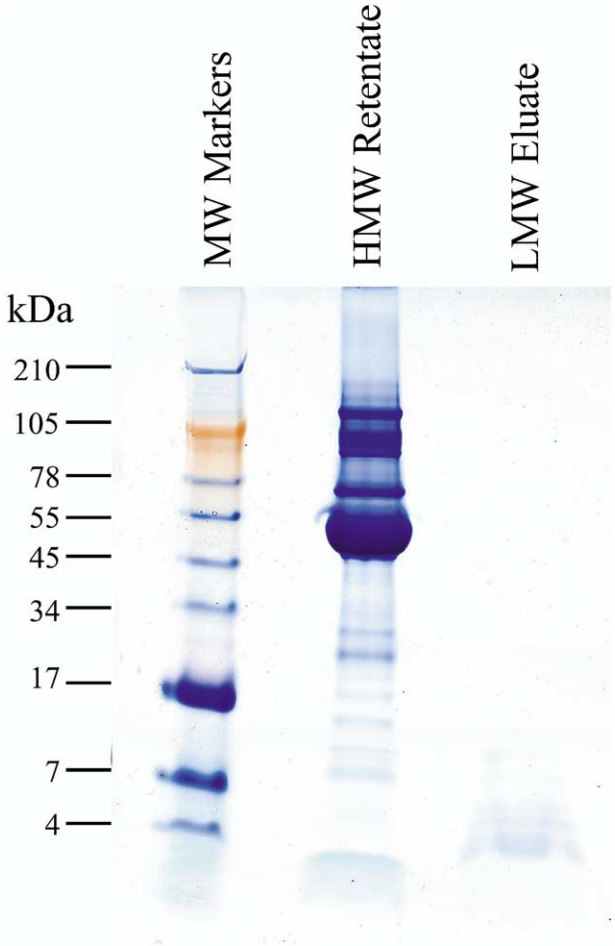


**Figure 2.** SDS-PAGE analysis of high- and low-molecular weight mouse serum purification. Tricine SDS-PAGE of the ultrafiltration of mouse serum through a membrane with a 30 kDa molecular weigh cutoff filter. Lane 2 shows the high-molecular weight (HMW) retentate after filtration while lane 3 shows the low-molecular weight eluate, which results in depletion of the majority of the HMW proteins from the sample as shown through coomassie staining.

teins such as albumin were identified in the analysis, their abundance had been decreased to an extent to which they no longer posed a challenge to the dynamic range measurement capabilities of the combined off-line MudPIT/MS system used in this study, in which SCXLC is performed offline and separate from RPLC. The identification of proteins with masses greater than 30 kDa in the low molecular weight fractions can be attributed to the presence of proteolytic fragments of the native proteins present within serum.

Critical to any study that utilizes a discovery-based approach to find disease-specific biomarkers is the ability to identify a large number of proteins that can be used in a comparative analysis between samples. Presently, the largest databases of identified proteins from either human serum or plasma contain between approximately 1450 [18] to 1700 proteins [19]. The proteins in these studies, however, were identified from searching MS/MS spectra in which multiple different enzyme constraints were allowed and many of the proteins were identified by a single peptide. In this study, almost 9000 fully tryptic peptides originating from 6003 proteins were identified, making this in itself the largest number of proteins identified in serum to date. To assemble the list of peptide and protein identifications, only those proteins identified by two or more tryptic peptides that possess stringent $X_{corr}$ and $\Delta C_n$ scores (as described in the Experimental section) are reported. Indeed, 1647 proteins were identified by at least two unique, fully tryptic peptides, increasing the confidence in those species identified in this manner. Over 500 proteins (i.e., 532) were identified by three or more unique tryptic peptides, already surpassing the number of proteins identified in a previous study in our laboratory examining the LMW serum proteome. This large increase in the number of tryptic peptides and total proteins identified is a result of the resolution of the SCX and RP separations as well as the use of a LIT mass spectrometer, which has a much faster scan time than the three-dimensional ion-trap mass spectrometer for performing MS/MS experiments. This faster duty cycle enables more peptides to be interrogated by CID per unit time. A histogram showing the distribution of the number of proteins identified per unique, fully tryptic peptide is presented in Figure 3. Although it is well accepted that many nontryptic peptides exist in serum, identification of such peptides does not directly contribute to the present quantitative proteomic methodology, based on the use of trypsin-mediated $^{18}O$ labeling, and are therefore excluded from the present analysis.

An important consideration when identifying proteins in biofluids is that the characterization of the identified species intuitively reflects the source of the proteins. In the analysis of serum, one could expect an enrichment of extracellular proteins compared to other compartmental classes. In this study, 22% of the pro-

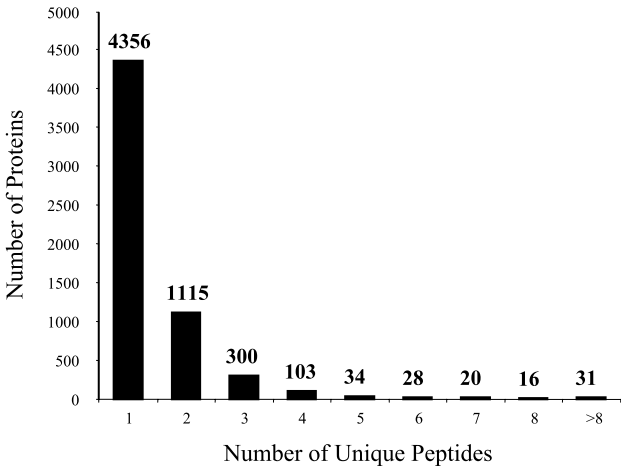


**Figure 3.** Proteins identified per number of unique, fully tryptic peptides. Histogram showing the number of proteins identified in the LMW mouse serum proteome per number of unique, fully tryptic peptides. Of 6003 identified proteins, 1647 were identified by two or more unique peptides.

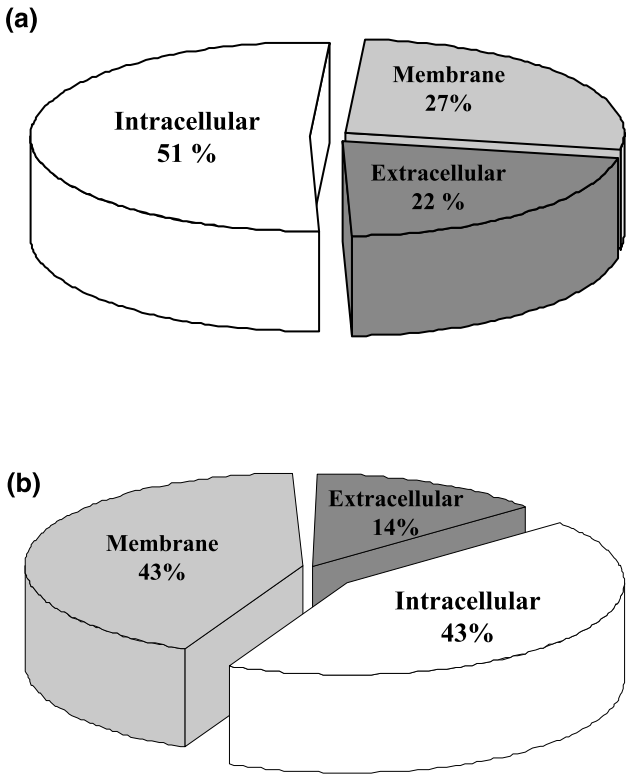


**Figure 4.** Gene ontology classification of cellular location of mouse LMW serum proteins. (**a**) Cellular compartmentalization of low molecular weight mouse serum proteins (~30% mapped by GO) reveal that 22% of these are classified as extracellular. (**b**) Gene ontology classification predicts only 14% of the entire mouse proteome encodes for extracellular proteins.

teins identified by at least two unique tryptic peptides were classified as extracellular (Figure 4a). An additional 27% of the proteins were localized to the membrane, with the remainder (i.e., 51%) originated from intracellular compartments. Unfortunately, only about 30% of the proteins identified in this study by two or more unique tryptic peptides could be linked to gene ontology classifications. Gene ontology of the entire mouse proteome (18,887 out of 27,459 possible proteins) predicts that 14% of the proteins will be extracellular (Figure 4b). Why there was a substantial increase in the percentage of extracellular proteins observed in this study compared to the mouse proteome is not clear. Coincidentally, this same phenomenon was observed in a gene ontology comparison of experimentally identified human serum proteins and the entire human proteome [20].

An important factor when using $^{16}O/^{18}O$ labeling for quantitative measurements is the completeness of enzyme-mediated isotope exchange. Shown in Figure 5a and b are selected mass spectra of differentially labeled LMW serum $^{16}O/^{18}O$ peptides from control mice and those bearing Lewis lung carcinoma, indicating complete $^{18}O$ incorporation of peptides within serum obtained from the mouse lung cancer model. These two sets of peptides were identified as apolipoprotein A1, which has recently been reported as a biomarker for the detection of early stage ovarian cancer [21], and $\alpha$-1-antitrypsin 1-1, which has been observed to have increased levels in lung cancer patients [22], demonstrating the facile ability of the present methodology to provide both quantitation and identification of putative biomarkers in the same high-throughput experiment. We hypothesize that efficient $^{18}O$ incorporation, as indicated by the mass spectra in Figure 5, may be attributed to the observed increase (~20%) in trypsin activity in 20% (vol/vol) $CH_3OH$ as previously demonstrated by a N-benzoyl-L-arginine ethyl ester trypsin activity assay [23].

The identified peptides were quantified by calculating the relative abundances ($^{18}O/^{16}O$, in this data set) of peptides based on the area of their extracted ion chromatograms (XIC) reconstructed from the molecular ion MS scans. There is no standard for what change in relative abundance, as measured by quantitative proteome analyses, represent statistically significant changes in protein abundance ratios. This lack of a standard is because the determination of isotopomeric peptide ratios depends directly on many experimental factors, such as the equivalence of the initial sample mixing, the relative efficiency of recovery of each of the peptide pairs, the inherent resolution and accuracy of the MS instrumentation, and the number of molecular ion scans across each peptide pair nanoRPLC peak. We sought to determine the threshold above and below which the measured protein abundance ratios were statistically significant within this dataset. This analysis was conducted by plotting the normalized percent of the number of $^{18}O/^{16}O$ labeled proteins identified within binned abundance ratios (e.g., 1.0-1.09, 1.1-1.19, etc.) as determined from integration of each of the respective isotopomeric peptide pair XIC (Figure 6). Theoretically, the distribution of the protein abundance ratios should be a normal distribution, defined as,

$$P = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \quad (1)$$

where $P$ is the normalized number of proteins identified at a given protein abundance ratio $x$, $\mu$ is defined as the experimentally measured mean value of the total population of the protein abundance ratios and $\sigma$ is standard deviation. The z score is defined as,

$$z = \frac{x-\mu}{\sigma} \quad (2)$$

Using non-linear least squares regression method, $\mu$ and $\sigma$ can be obtained from the data series of normalized number of proteins identified and the experimentally determined ratios of the protein abundances. The protein abundance ratio above which a statistically significant increase in abundance should

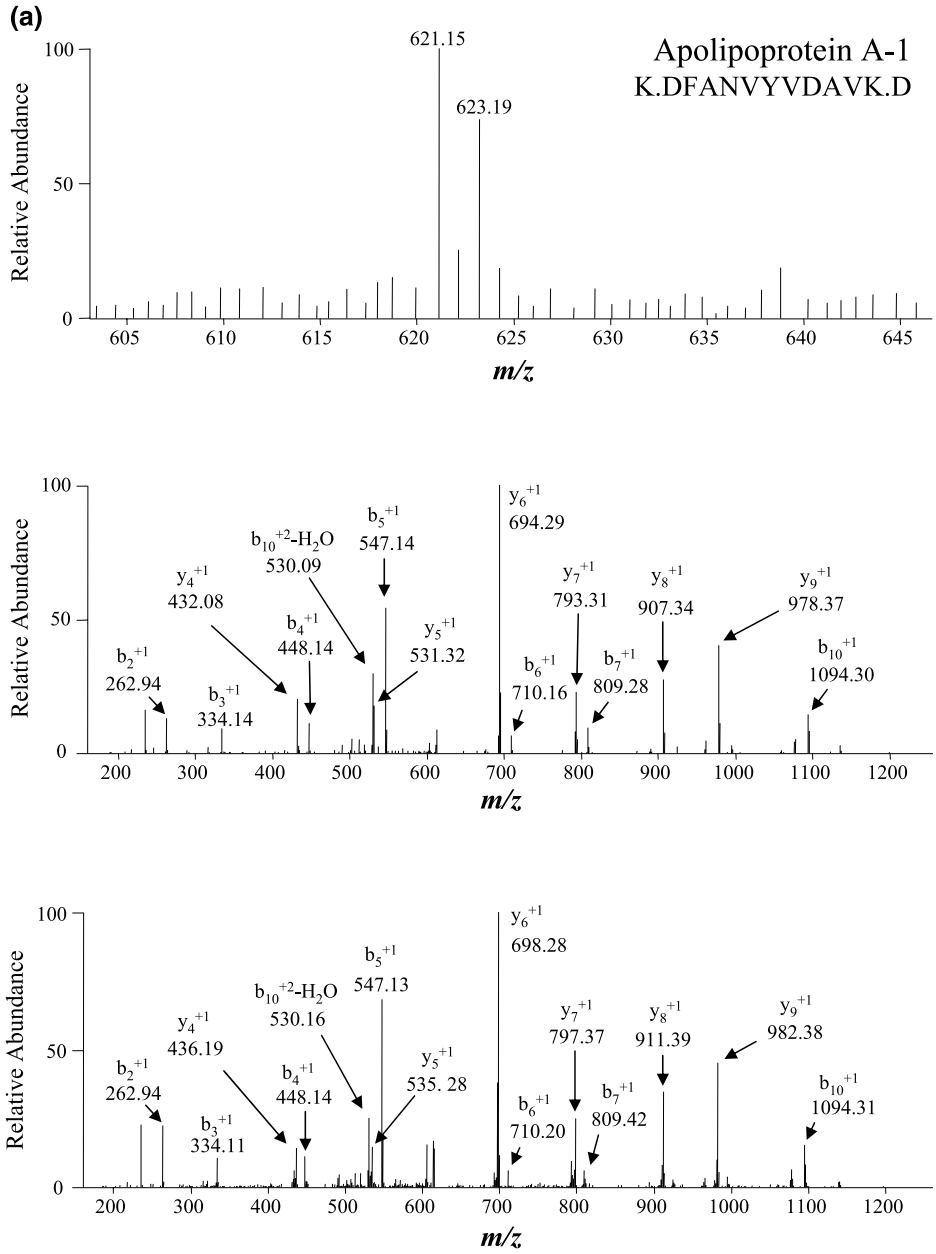


**Figure 5.** Selected mass spectra of differentially labeled LMW serum $^{16}O/^{18}O$ peptides. Full range mass spectra of the $^{16}O/^{18}O$ molecular ion pairs with ($[M + H]^{+1}$) or 2 Da their corresponding MS/MS fragmentation ion series showing ~4 Da ($[M + 2H]^{+2}$) mass shifts in the *y*-type ions for (**a**) apolipoprotein A-1, and (**b**) $\alpha$-1-antitrypsin.

be considered is $2\sigma$ above the measured mean ratio, while the protein abundance ratio below which a statistically significant decrease in abundance should be considered $1/(2\sigma + \mu)$. Using these criteria, the threshold above which a measured protein abundance ratio reflects a statistically significant increase in abundance should be considered to be $\geq$1.708 and the threshold below which a measured abundance ratio reflects a statistically significant decrease in abundance should be considered to $\leq$0.585, for these data. The present data indicate that 211 proteins and 246 proteins are significantly increased and decreased in abundance in serum from mice bearing Lewis lung carcinoma, respectively. Hence, of the proteins identified by two or more unique fully tryptic peptides, approximately 28%, are observed to change in abundance in mice with Lewis lung carcinoma.

Several of the proteins that were found to be upregulated in the serum obtained from the lung carcinoma mouse model have been implicated in cancer progression. Vascular endothelial growth factor receptor 1 (VEGFR-1) was identified to be upregulated over 7-fold in the mouse lung carcinoma model. VEGF is a key

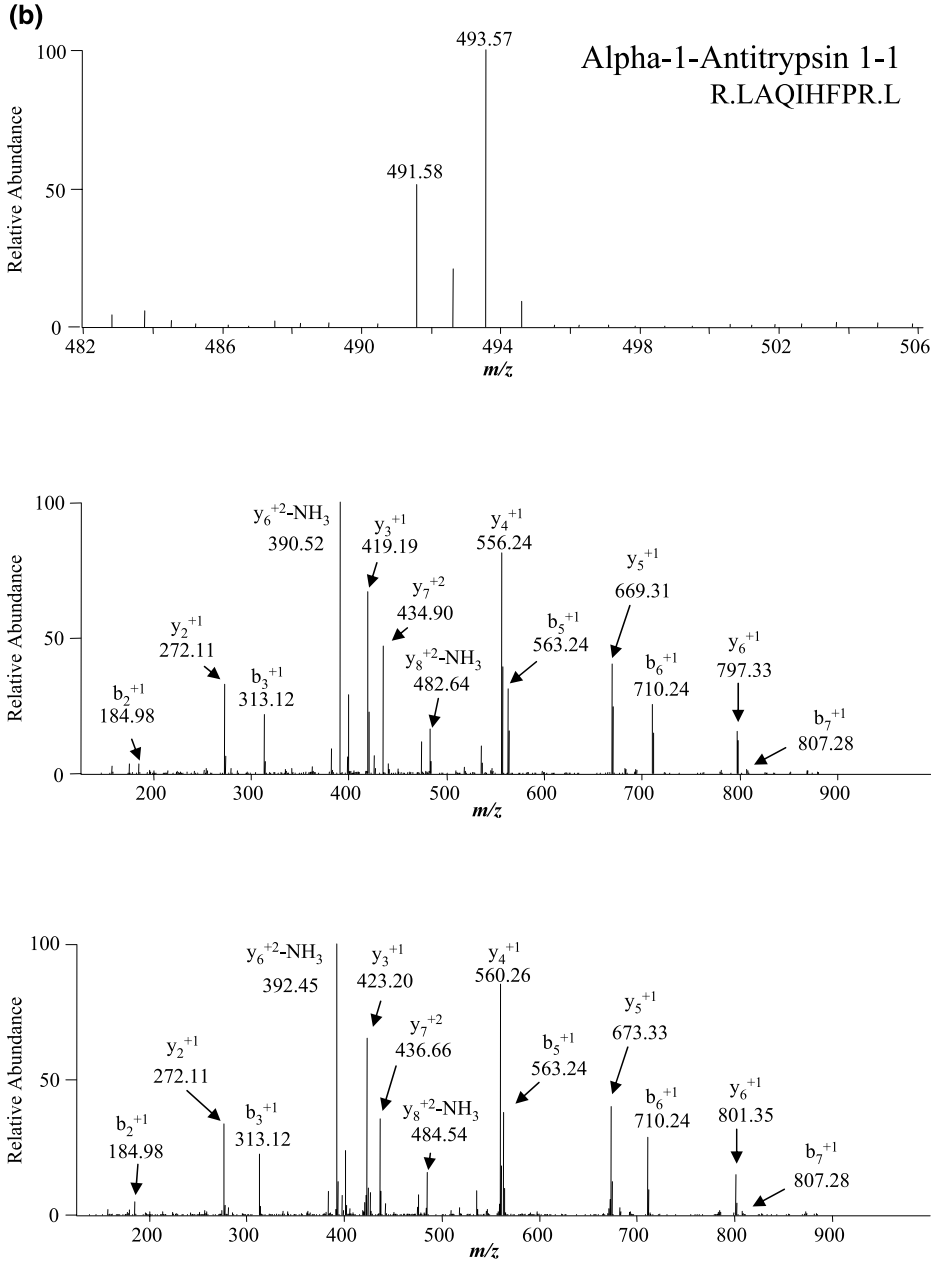


**Figure 5**(continued)

angiogenic factor that is expressed and secreted by nearly all tumors at high levels and stimulates angiogenesis [24]. Members of the VEGF family's angiogenic signals are mainly mediated by activation of the structurally related homologous tyrosine kinase receptors, VEGFR-1 and VEGFR-2. A previous study has shown that blocking VEGFR-1 activity reduces lung metastasis [25]. These results are consistent with the discovery of increased amounts of VEGFR-1 in the serum of the mouse lung cancer model compared to the matched control found in this study. Several of the proteins identified to be differentially abundant in the Lewis lung cancer mouse model are currently being validated with the hope of linking their dysregulation back to the tumor site.

## Conclusions

The heart of proteomics represents a discovery science that enables investigators to gather information on hundreds, if not thousands, of proteins in a relatively high-throughput, nonbiased, manner. This basic characteristic has brought great hope that proteomics includes the sets of technologies that will enable the discovery of biomarkers for the diagnosis of early stages of diseases, such as cancer, a reality. Obviously, developments in MS instrumentation have been critical for enabling the broad identification of proteins in complex mixtures such as serum. What is sometimes under-appreciated, however, is the sample processing steps conducted prior to any portion of the sample

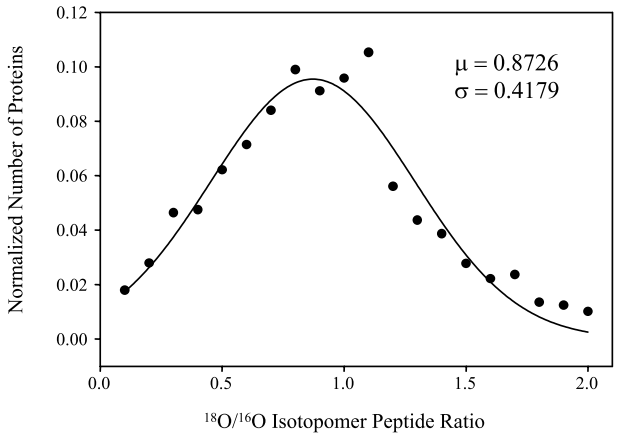


**Figure 6.** Distribution of ratios of $^{18}O/^{16}O$ labeled proteins. Plot of the normalized percent of the number of proteins identified by two or more unique peptides versus their $^{18}O/^{16}O$ ratios. Abundance ratios greater than 1.708 and less than 0.585 are considered statistically significant and result in 211 proteins and 246 proteins that are over- and underexpressed in the LMW mouse serum proteome respectively.

being introduced into the mass spectrometer. In this study, we utilized ultrafiltration to enrich for the LMW component of serum. This targeted fractionation is not only simple, but removes a large proportion of the high abundance proteins that are primarily responsible for making the comprehensive characterization of serum and plasma so challenging. Of equal importance is the use of an off-line MudPIT approach to separate the resultant peptides into manageable fractions for the spectrometer to identify individual peptides. Although many other groups have explored the use of alternative techniques to provide broad proteome characterization, the components of MudPIT (i.e., SCX and $\mu$RPLC) remain the optimal choice for fractionation of complex peptide mixtures. If proteomics is to find effective disease biomarkers without any prior knowledge as to what their identity may be, methods such as LMW fractionation and off-line MudPIT will be crucial to gathering the critical mass of data necessary to make assumptions about the relative abundances of proteins in samples obtained from normal and disease-afflicted individuals that can be validated over larger populations.

Although comparative proteomic analysis of human serum promises the ability to discover new and more effective biomarkers for the early indication of disease and response to therapy, such investigations are complicated because of the intrinsic variability in each serum sample attributed to genetics, lifestyle, and environmental differences amongst people. A vital step in making this distinction has relied on the use of animal models such as the mouse where these biological variables can be carefully controlled to allow increased rigor with which comparisons and pathophysiological conclusions can be made. Mouse models, either transgenic or xenograft, represent an invaluable experimental system for understanding human cancer pathogenesis because, unlike human samples derived from clinical trials, experimental artifacts related to genetic background and environment can be more carefully accounted for, thus minimizing the variability seen within samples acquired from human patients. Along these lines, we have developed a methodology to permit direct quantitative measurement of abundance changes in serum proteins from mouse models of human cancer utilizing trypsin-mediated stable oxygen isotope labeling. In this work, a global quantitative proteomic investigation of the LMW serum proteomes from control mice and those bearing human Lewis lung carcinoma was conducted. This investigation resulted in the identification of more than 6000 proteins; all from peptides bearing full tryptic termini. In an effort to increase the confidence in the population of proteins utilized in the quantitative comparison, we culled from this list only those proteins identified from two or more tryptic peptides, which equates to 1647 proteins. From this list of proteins identified with high confidence, 211 and 246 were quantitatively determined to be present at an increased and decreased level of abundance, respectively, in the serum of mice bearing Lewis lung carcinoma. Current studies are focused on further analysis of this differentially abundant dataset to determine if any of these serum proteins can be directly linked to the presence of the lung carcinoma.

## Acknowledgments

This work was supported by Federal funds from the National Cancer Institute, National Institutes of Health, under contract NO1-CO-12400, and in part by a grant to AK from the National Cancer Institute of Canada. By acceptance of this article, the publisher or recipient acknowledges the right of the United States Government to retain a nonexclusive, royalty-free license and to any copyright covering the article. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the United States Government.

## Supplementary Material

Supplementary material (Tables 1 and 2) can be found in the online version at doi:10.1016/j.jasms.2005.02.005.

## References

1. Adkins, J. N.; Varnum, S. M.; Auberry, K. J.; Moore, R. J.; Angell, N. H.; Smith, R. D.; Springer, D. L.; Pounds, J. G. Toward a Human Blood Serum Proteome: Analysis by Multidimensional Separation Coupled with Mass Spectrometry. *Mol. Cell. Proteom.* **2002,** *1,* 947–955.
2. Tirumalai, R. S.; Chan, K. C.; Prieto, D. A.; Issaq, H. J.; Conrads, T. P.; Veenstra, T. D. Characterization of the Low Molecular Weight Human Serum Proteome. *Mol. Cell. Proteom.* **2003,** *13,* 13–18.
3. Carrette, O.; Demalte, I.; Scherl, A.; Yalkinoglu, O.; Corthals, G.; Burkhard, P.; Hochstrasser, D. F.; Sanchez, J. C. A Panel of Cerebrospinal Fluid Potential Biomarkers for the Diagnosis of Alzheimer's Disease. *Proteomics* **2003,** *3,* 1486–1494.

4. Cepok, S.; Jacobsen, M.; Schock, S.; Omer, B.; Jaekel, S.; Boddeker, I.; Oertel, W. H.; Sommer, N.; Hemmer, B. Patterns of Cerebrospinal Fluid Pathology Correlate with Disease Progression in Multiple Sclerosis. *Brain* **2001,** *124,* 2169–2176.
5. Washburn, M. P.; Wolters, D.; Yates, J. R. Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology. *Nat. Biotechnol.* **2001,** *19,* 242–247.
6. Resing, K. A.; Meyer-Arendt, K.; Mendoza, A. M.; Aveline-Wolf, L. D.; Jonscher, K. R.; Pierce, K. G.; Old, W. M.; Cheung, H. T.; Russell, S.; Wattawa, J. L.; Goehle, G. R.; Knight, R. D.; Ahn, N. G. Improving Reproducibility and Sensitivity in Identifying Human Proteins by Shotgun Proteomics. *Anal. Chem.* **2004,** *76,* 3556–3568.
7. Yu, L. R.; Conrads, T. P.; Uo, T.; Kinoshita, Y.; Morrison, R. S.; Lucas, D. A.; Chan, K. C.; Blonder, J.; Issaq, H. J.; Veenstra, T. D. Global Analysis of the Cortical Neuron Proteome. *Mol. Cell. Proteom.* **2004,** *3,* 896–907.
8. Anderson, N. L.; Anderson, N. G. The Human Plasma Proteome—History, Character, and Diagnostic Prospects. *Mol. Cell. Proteom.* **2002,** *1,* 845–867.
9. Rothemund, D. L.; Locke, V. L.; Liew, A.; Thomas, T. M.; Wasinger, V.; Rylatt, D. B. Depletion of the Highly Abundant Protein Albumin from Human Plasma Using the Gradiflow. *Proteomics* **2003,** *3,* 279–287.
10. Pieper, R.; Su, Q.; Gatlin, C. L.; Huang, S. T.; Anderson, N. L.; Steiner, S. Multi-Component Immunoaffinity Subtraction Chromatography: An Innovative Step Towards a Comprehensive Survey of the Human Plasma Proteome. *Proteomics* **2003,** *3,* 422–432.
11. Liotta, L. A.; Ferrari, M.; Petricoin, E. Clinical Proteomics: Written in Blood. *Nature* **2003,** *425,* 905.
12. 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. Use of Proteomic Patterns in Serum to Identify Ovarian Cancer. *Lancet* **2002,** *359,* 572–577.
13. Basso, D.; Valerio, A.; Brigato, L.; Panozzo, M. P.; Miola, M.; Lucca, T.; Ujka, F.; Zaninotto, M.; Avogaro, A.; Plebani, M. An Unidentified Pancreatic Cancer Cell Product Alters Some Intracellular Pathways of Glucose Metabolism in Isolated Rat Hepatocytes. *Pancreas* **1997,** *15,* 132–138.
14. Rubin, R. B.; Merchant, M. A Rapid Protein Profiling System that Speeds Study of Cancer and other Diseases. *Am. Clin. Lab.* **2000,** *19,* 28–29.
15. Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. Proteolytic $^{18}O$ Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus. *Anal. Chem.* **2001,** *73,* 2836–2842.
16. Schnolzer, M.; Jedrzejewski, P.; Lehmann, W. D. Protease-Catalyzed Incorporation of $^{18}O$ into Peptide Fragments and its Application for Protein Sequencing by Electrospray and Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Electrophoresis* **1996,** *17,* 945–953
17. Rose, K.; Simona, M. G.; Offord, R. E.; Prior, C. P.; Otto, B.; Thatcher, D. R. A New Mass-Spectrometric C-Terminal Sequencing Technique Finds a Similarity Between γ-Interferon and α2-Interferon and Identifies a Proteolytically Clipped γ-Interferon that Retains Full Antiviral Activity. *Biochem. J.* **1983,** *215,* 273–277.
18. Chan, K. C.; Lucas, D. A.; Hise, D.; Schaefer, C. F.; Xiao, Z.; Janini, G. M.; Buetow, K. H.; Issaq, H. J.; Veenstra, T. D.; Conrads, T. P. Analysis of the Human Serum Proteome. *Clin. Proteom.* **2004,** *1,* 101–226.
19. Shen, Y.; Jacobs, J. M.; Camp, D. G., II; Fang, R.; Moore, R. J.; Smith, R. D.; Xiao, W.; Davis, R. W.; Tompkins, R. G. Ultra-High-Efficiency Strong Cation Exchange LC/RPLC/MS/MS for High Dynamic Range Characterization of the Human Plasma Proteome. *Anal. Chem.* **2004,** *76,* 1134–1144.
20. Anderson, N. L.; Polanski, M.; Pieper, R.; Gatlin, T.; Tirumalai, R. S.; Conrads, T. P.; Veenstra, T. D.; Adkins, J. N.; Pounds, J. G.; Fagan, R.; Lobley, A. The Human Plasma Proteome: A Nonredundant List Developed by Combination of Four Separate Sources. *Mol. Cell. Proteom.* **2004,** *3,* 311–326.
21. Zhang, Z.; Bast, R. C., Jr.; Yu, Y.; Li, J.; Sokoll, L. J.; Rai, A. J.; Rosenzweig, J. M.; Cameron, B.; Wang, Y. Y.; Meng, X. Y.; Berchuck, A.; Van Haaften-Day, C.; Hacker, N. F.; de Bruijn, H. W.; van der Zee, A. G.; Jacobs, I. J.; Fung, E. T.; Chan, D. W. Three Biomarkers Identified from Serum Proteomic Analysis for the Detection of Early Stage Ovarian Cancer. *Cancer Res.* **2004,** *64,* 5882–5890.
22. Zelvyte, I.; Wallmark, A.; Piitulainen, E.; Westin, U.; Janciauskiene, S. Increased Plasma Levels of Serine Proteinase Inhibitors in Lung Cancer Patients. *Anticancer Res.* **2004,** *24,* 241–247.
23. Blonder, J.; Conrads, T. P.; Yu, L. R.; Terunuma, A.; Janini, G. M.; Issaq, H. J.; Vogel, J. C.; Veenstra, T. D. A Detergent- and Cyanogen Bromide-Free Method for Integral Membrane Proteomics: Application to Halobacterium Purple Membranes and the Human Epidermal Membrane Proteome. *Proteomics* **2004,** *4,* 31–45.
24. Senger, D. R.; Galli, S. J.; Dvorak, A. M.; Perruzzi, C. A.; Harvey, V. S.; Dvorak, H. F. Tumor Cells Secrete A Vascular Permeability Factor That Promotes Accumulation Of Ascites Fluid. *Science* **1983,** *219,* 983–985.
25. Hiratsuka, S.; Nakamura, K.; Iwai, S.; Murakami, M.; Itoh, T.; Kijima, H.; Shipley, J. M.; Senior, R. M.; Shibuya, M. MMP9 Induction by Vascular Endothelial Growth Factor Receptor-1 is Involved in Lung-Specific Metastasis. *Cancer Cell* **2002,** *2,* 289–300.