

Investigation of the Mouse Serum Proteome

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

*Laboratory of Proteomics and Analytical Technologies, SAIC-Frederick, Inc., National Cancer Institute at
Frederick, PO Box B, Frederick, Maryland 21702*

Received April 18, 2005

With the rapid assimilation of genomic information and the equally impressive developments in the field of proteomics, there is an unprecedented interest in biomarker discovery. Although human biofluids represent increasingly 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. One of the most extensively used animal models for studying human disease is mouse because, unlike humans, they represent a highly controllable experimental model system. Unfortunately, very little is known about the proteomic composition of mouse serum. In this study, a multidimensional fractionation approach on both the protein and the peptide level that does not require depletion of highly abundant serum proteins was combined with tandem mass spectrometry to characterize proteins within mouse serum. Over 12 300 unique peptides that originate from 4567 unique proteins—approximately 16% of all known mouse proteins—were identified. The results presented here represent the broadest proteome coverage in mouse serum and provide a foundation from which quantitative comparisons can be made in this important animal model.

Keywords: mouse • serum proteomics • multidimensional fractionation • tandem mass spectrometry

Introduction

With continuing advances in proteomic techniques, the hope of discovering and developing novel biomarkers becomes increasingly tangible. The ability to detect these potentially low abundant species amid a plethora of highly abundant proteins becomes critical in diagnostics for the early detection of diseases, such as cancer.¹ Comprehensive proteomic analyses of various biofluids (i.e., serum, plasma, and cerebrospinal fluid) hold promise in the search for biomarkers as these samples are easily obtainable, contain high concentrations of proteins and perfuse key tissues, undoubtedly endowing these fluids with protein species indicative of disease states.^{2–5} Indeed, increased serum levels of CA125⁶ and prostate-specific antigen⁷ are currently used for the detection of ovarian and prostate cancer, respectively. Detection of changes in the abundance or alterations in the integrity of different protein species in biofluids may provide the means for disease detection enabling earlier intervention in disease progression, with the hope of dramatically increasing the success rate of currently available therapeutics.

Unfortunately, the sheer complexity and dynamic range of protein concentration in a biofluid such as serum makes its characterization, let alone discovery of potential biomarkers, very challenging. Within human serum for example, twenty-

two species comprise greater than 90% of the total protein mass, with human serum albumin (HSA) alone constituting approximately 50% of this mass.¹ To overcome the difficulties presented by this wide dynamic range of protein concentration, a variety of separation methods have been employed to reduce sample complexity and allow for the detection of the remaining low abundance proteins in these biofluids.

Ultimately, the identification of unique peptides from any complex mixture will be hindered by several factors. Obviously, as a sample is increasingly pre-fractionated more peptides will be lost due to effects of, for example, sample handling and dilution. Additionally, limitations associated with current analytical instrumentation prevent the complete identification of all species within complex peptide mixtures. Logically, depletion of the high abundance, high molecular weight proteins would assist the detection and identification of low abundance species. Numerous affinity purification and extraction experiments have been developed to enrich the less-abundant proteins and peptides residing in complex biofluids.^{2,8–12} Hence, initial results at characterizing human serum focused on the removal of the high abundance proteins such as albumin, transferrin, immunoglobins and lipoproteins via affinity reagents or ultracentrifugation.^{2,8,12} A recent investigation revealed that a number of proteins/peptides are co-depleted with many of the highly abundant proteins commonly targeted for removal (e.g., IgG, HSA, and transferrin) suggesting that serum is actually a complex system comprised of multiple protein–protein interactions.⁹ While these investigations have revealed critical dynamic information regarding the content and nature

* To whom correspondence should be addressed. Laboratory of Proteomics and Analytical Technologies, National Cancer Institute at Frederick, SAIC-Frederick, Inc. P.O. Box B, Frederick, MD 21702. Tel: (301) 846-7353. Fax: (301) 846-6037. E-mail: conrads@ncifcrf.gov.

of serum, it is evident that depletion of the high abundance proteins prior to an in-depth analysis would result in the loss of valuable information.

There is an increasing effort in using proteomic technologies in cancer research to discover novel biomarkers in serum/plasma that serve to detect diseases earlier with higher accuracy. Mouse models, either transgenic or xenograft, represent an invaluable experimental system for understanding human cancer pathogenesis, however, there are often fundamental differences in how the process of tumorigenesis occurs in mice and humans.^{13–16} Consequently, the direct relevance of findings forthcoming from such analyses is still debated, although efforts to validate these results are progressing.^{17–19} One obvious advantage in using mouse models compared to human clinical trials is that 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. The ability to reduce these artifacts enables a more direct comparison between serum/plasma from control and cancer mouse models than possible in the analysis of human samples. Despite the mouse genome being approximately 14% smaller than the human genome, the two systems are very similar in content and chromosomal organization.²⁰ This similarity is reflected in an approximate 40% alignment of the mouse and human genomes at the nucleotide level. Indeed approximately 80% of mouse coding regions have a direct orthologue in the human genome where both species share a large majority of the estimated 200 000 exons.^{20,21} Clearly, evaluation of these similarities to the wide variety of biological processes and functionality is fundamental to the comparative question, with the eventual ability to elucidate carcinogenesis in mouse model systems and translate that information for human relevance.

In this manuscript, a global characterization of the mouse serum proteome is described with the identification of 4567 unique proteins from 12 389 unique peptides. Proteins from all functional classes and localizations were detected, as well as the identification of numerous low abundance proteins without the need for prior removal of albumin. Classification of the identified proteins by gene ontology demonstrated that over 40% originate from the membrane. This analysis establishes the basis for the development of additional quantitative proteomic techniques to allow for the detection of potential biomarkers in mouse models to determine the presence or absence of various molecular species between control and interrogated animals. In addition, future cross comparison of these results with those obtained from the recent Human Proteome Organization (HUPO) Plasma Proteome Project (PPP)²² (when released) will enable determination of the utility of mouse models for detection of biomarkers for human diseases.

Experimental Section

Materials. Pooled standard mouse serum, 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).

Protein Fractionation. Ion exchange HPLC (HP 1090 HPLC, Agilent Technologies, Palo Alto, CA) of mouse serum proteins was performed using weak anion exchange (WAX) and weak cation exchange columns (WCX) (4.6×200 mm, $5 \mu\text{m}$, 1000 Å pore size, PolyLC, Inc., Columbia, MD). Mobile phase A was 5% acetonitrile and mobile phase B was 5% acetonitrile containing 0.6 M ammonium acetate (pH 6). Two hundred microliters of serum was diluted to 1 mL with water. With the WAX and WCX columns connected in series, diluted serum was loaded and the columns were washed for 30 min with 2% B. The two columns were disconnected, and captured proteins were separately eluted from the two columns using multistep salt gradients (2% B for 10 min, up to 100% B in 70 min and held at 100% B for 6 min). Fractions were collected every min (flow rate = 1 mL/min) for each of the WAX and WCX elutions. The above fractionation procedure was repeated four times. On the basis of an evaluation of the complexity of the separations, the five serial WAX LC runs were combined and pooled into seven sub-fractions, while the five serial WCX LC runs were combined and pooled into five sub-fractions.

Tryptic Digestion. Both the WAX and WCX serum protein sub-fractions were lyophilized and re-suspended in 0.1% formic acid, 20% acetonitrile and boiled for 10 min. Total protein in each sample was quantified by the BCA protein assay (Pierce, Rockford, IL). The samples were lyophilized and re-suspended in 50 mM NH_4HCO_3 , pH 8.4, and the final pH in each sample was adjusted to 8.4 by adding concentrated ammonium hydroxide (NH_4OH) if necessary. Trypsin was added to each sample at an enzyme-to-protein ratio of 1:50 and digested for 16 h at 37 °C. Samples were lyophilized and re-suspended in 200 μL of 0.1% formic acid, 20% acetonitrile for strong cation exchange (SCX) fractionation.

Strong Cation Exchange Fractionation. Strong cation exchange LC (HP 1090 LC, Agilent) of the WAX or WCX fractions tryptic peptide digestates was performed using a polysulfoethyl A column (4.6×200 mm, $5 \mu\text{m}$, 300 Å pore size, PolyLC, Inc.). The separation was monitored by fluorescence (280/250 nm). The following $\text{NH}_4\text{HCO}_2/\text{CH}_3\text{CN}$ multistep gradient at a flow rate of 1 mL/min was used to elute the peptides from the column: 2% mobile phase B (25% CH_3CN , 0.5 M NH_4HCO_2 , pH 3.0) for 2 min, followed by 32% B in 60 min, then 100% B in 13 min and maintained at 100% B for 11 min. Mobile phase A was 25% CH_3CN . Fractions were collected every minute, lyophilized and reconstituted in 30 μL of 0.1% TFA prior to microcapillary (μ) reversed-phase (RP) liquid chromatography (LC) – tandem mass spectrometry (MS/MS) analysis.

Microcapillary Reversed-Phase Liquid Chromatography Online with Tandem Mass Spectrometry. Microcapillary RPLC was performed using an Agilent 1100 capillary LC system (Agilent Technologies) coupled online to an ion trap (IT) mass spectrometer (LCQ DecaXP, Thermo Electron, San Jose, CA) with the nanoelectrospray interface supplied by the manufacturer. Separations were performed using 75 μm i.d. \times 360 o.d. \times 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) and peptides were eluted using a linear gradient of 2% mobile phase B (0.1% formic acid in acetonitrile)

Table 1. SEQUEST Filter Criteria for Peptide Identification.

charge state ^a	X_{corr}^b	proteolytic constraint
+1	≥ 1.9	fully tryptic
+1	≥ 2.1	fully chymotryptic and/or elastic
+1	≥ 2.2	partially tryptic, chymotryptic, and/or elastic
+1	≥ 2.2	no protease constraint
+2	≥ 2.2	fully tryptic
+2	≥ 2.2	fully chymotryptic and/or elastic
+2	≥ 2.4	partially tryptic, chymotryptic, and/or elastic
+2	≥ 3.0	no protease constraint
+3	≥ 3.5	fully tryptic
+3	≥ 3.5	fully chymotryptic and/or elastic
+3	≥ 3.75	partially tryptic, chymotryptic, and/or elastic
+3	≥ 3.75	no protease constraint

^a Charge state of the peptide molecular ion chosen for tandem MS.
^b SEQUEST cross correlation score.

to 40% mobile phase B in 110 minutes, then to 98% B in an additional 30 min, all at a constant flow rate of 0.5 μL/min.

The IT-MS was operated in a data dependent tandem MS (MS/MS) mode in which each full MS scan was followed by three MS/MS scans where the three most abundant peptide molecular ions are dynamically selected for collision-induced dissociation (CID) using a normalized collision energy of 38%. Dynamic exclusion was utilized to prevent redundant acquisition of peptides previously selected for MS/MS. The heated capillary temperature and electrospray voltage were set at 160 °C and 1.7 kV, respectively.

Bioinformatic Analysis. The μRPLC-MS/MS analyses conducted in this investigation resulted in approximately 2.3 million tandem mass spectra that were searched against the UniProt mouse proteomic database (12/18/2003 release) from the European Bioinformatics Institute (<http://www.ebi.ac.uk/Databases/>) with SEQUEST operating on an 18 node Beowulf cluster. For a peptide to be considered legitimately identified, it had to achieve stringent charge state and proteolytic cleavage-dependent cross correlation (X_{corr}) scores similar to those previously reported (Table 1) and a minimum delta correlation (ΔC_n) of 0.08.

Unique proteins were categorized by gene ontology (www.geneontology.org). Membrane composition of unique proteins was determined using TMHMM, a transmembrane hidden Markov model prediction algorithm (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>),^{23–25} and PHOBIUS (<http://phobius.binf.ku.dk/>), a transmembrane topology, and signal peptide prediction algorithm.²⁶ Pathway analysis was performed using Ingenuity Pathways analysis software (<http://analysis.ingenuity.com>).

Results

Mouse Serum Proteome Analysis. A pooled mouse serum standard was initially fractionated at the protein level by weak anion exchange (WAX) and weak cation exchange (WCX) chromatography (Figure 1). To avoid protein loss, the columns were connected in series such that the flow through during loading of the sample onto the WAX column was applied directly to the WCX column. Following loading of the serially connected WAX/WCX columns, the columns were separated

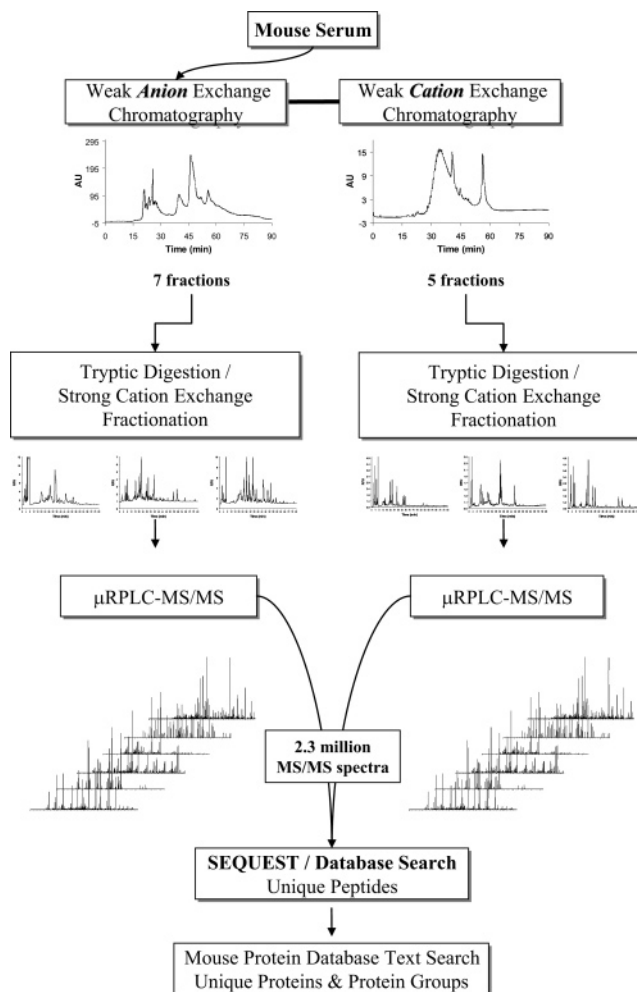


Figure 1. Schematic representation of the experimental design utilized to investigate the mouse serum proteome. Serum was fractionated at the protein level, followed by tryptic digestion and strong cation exchange (SCX) fractionation of the resulting peptides. SCX fractions were analyzed by μRPLC-MS/MS and data was searched against the mouse protein database using SEQUEST to identify peptides. Unique peptides were searched against the UniProt mouse protein database to identify unique proteins and protein groups.

and bound proteins were separately eluted from each of the columns using linear step gradients optimized for each column. Fractions from each of these separations were pooled into seven subfractions from the WAX column and five subfractions from the WCX column. These subfractions were tryptically digested and each of the digestates was further resolved by strong cation exchange (SCX) chromatography (Figure 1). On the basis of the fluorescence intensity, fractions were either analyzed individually or combined prior to μRPLC-MS/MS analysis. This analysis resulted in the accumulation of ~2.3 million tandem MS spectra that were searched against the UniProt mouse in silico translated protein database using the SEQUEST algorithm. Peptides were considered legitimately identified according to strict X_{corr} and ΔC_n scores as previously determined (Table 1).²⁷

Using the fractionation protocol described, 12 389 unique peptides (Supporting Information Tables 1 and 2) were identified that correspond to 4567 unique proteins (Supporting Information Tables 1 and 3) and 1057 protein groups (as

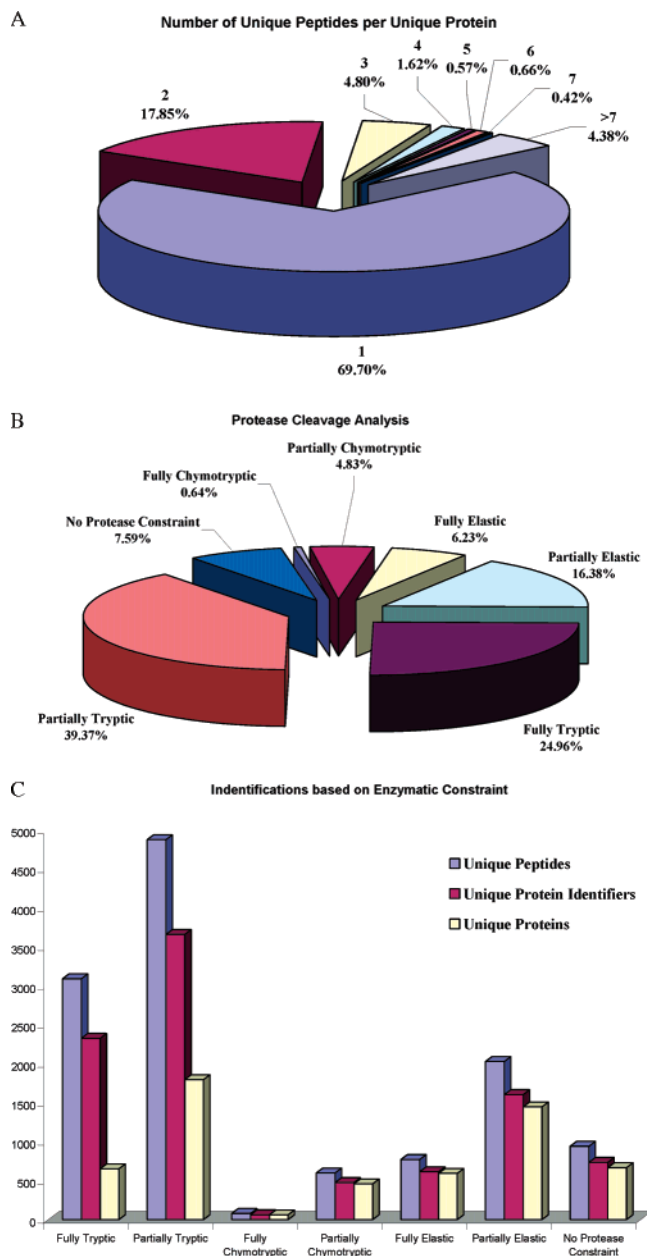


Figure 2. Analysis of the mouse serum proteome. (A) Number of unique peptides used to determine a single unique protein. (B) Percentages of enzymatic constraints observed for unique proteins. (C) Number of unique peptides, unique protein identifiers (peptides with a match to a single protein), and unique proteins observed for all enzymatic constraints.

defined below) (Supporting Information Tables 1 and 4). Of these uniquely identified proteins (Supporting Information Table 3), 70% were identified by a single peptide (Figure 2A). Unique protein identifications for each enzyme constraint are shown in Figure 2B, where approximately 65% of the observed unique peptides were either fully or partially tryptic, with elastase cleavage sites comprising almost 23% of the remainder. The total number of unique peptides, unique protein identifiers (peptides that identify only one protein in the mouse protein database) and unique proteins observed for each enzyme constraint are shown in Figure 2C. The use of a modified search (see below) has removed redundancy in the protein identifica-

tions as determined by SEQUEST (>5700 proteins, Supporting Information Table 1).

Reproducibility of μ RPLC-MS/MS. With the fractionation procedure utilized in the present investigation, a high-resolution analysis of undepleted serum has been obtained, allowing a substantial number of peptide identifications. Although a large number of peptides were identified, we sought to ascertain the reproducibility of the overall analysis and the extent to which an increase in the number of unique peptides may be identified per μ RPLC-MS/MS analysis. A second μ RPLC-MS/MS run of the WCX/SCX serum peptide fractions resulted in the identification of an average of 20% more unique peptides, the 3rd analysis yielded ~12% more, and the 4th analysis yielded ~8% more. Comparison of unique peptides identified in consecutive fractions from a single SCX separation show minor overlap between fractions with 20–25% of observed unique peptides identified in successive fractions.

False Positive Evaluation. The total peptide output from the SEQUEST searches was filtered as discussed above using software developed in-house to filter out false positive peptide identifications and to remove redundant peptide and protein identifications. Further validation of peptides and determination of false positive peptide identifications was performed by searching a subset of the data (>10% or 230 000 MS/MS spectra) against a ‘reversed’ mouse database.²⁸ This database was generated from the UniProt mouse protein database by reversing the peptide sequences for each protein (e.g., N term–C term to C term–N term), against which MS/MS data were searched using the same SEQUEST algorithm and X_{corr} and ΔC_n filtering criteria. The resulting peptide identifications were compared to those from the forward database search, selecting those having the same scan number and charge state. Cases where two peptides had been identified from a single tandem mass spectrum were evaluated based on their respective X_{corr} scores and fragment ion series. Positive X_{corr} score differences (of the forward database identification *minus* the X_{corr} of the reverse database identification) represent correct peptide identifications, while negative differences (or a value of zero) represent false identifications. Utilizing the SEQUEST X_{corr} filter criteria in Table 1, our analysis resulted in approximately 5% false positives ((140/2563)*100 = 5.46%). Increasing the cross correlation score filtering of peptides to 2.0 for $[M+H]^+$ peptide molecular ions (fully tryptic), and 2.5 for $[M+2H]^{2+}$ peptide molecular ions (all enzyme constraints) resulted in the loss of 1652 unique peptides (~13%), representing a decrease in the total unique proteins to 3863. It is clear that the selection of filtering criteria and peptide identification is a continual concern in global proteomic data analyses where the impact of increasingly strict criteria results in loss of potential legitimately identified peptides and less strenuous conditions result in increased false positive peptide identifications. The present analysis encompasses both concerns, providing abundant identifications while striving to reduce the amount of false positives.

Unique Protein Identification. An appreciable number of peptides identified by MS/MS in many proteome experiments do not possess sequences sufficiently divergent in the context of the total proteome (particularly in large mammalian proteomes) to uniquely identify only one protein. To address this issue, we have developed a search method based on a simple text search to match only those peptides that possess unique sequences in a given proteome. By application of this method, the common over-representation of the number of uniquely

identified proteins can be avoided where the result is a more accurate determination of the truly unique protein identifications. To find the unique proteins identified, uniquely identified peptides (Supporting Information Table 2) were queried against the UniProt mouse protein database using a simple text search. Individual peptides were searched for complete (e.g., 100%) identity to sequences in proteins within the database. Peptides whose sequence is present in only one protein represent unique protein identifiers, while those peptides that identify multiple proteins are considered nonunique and represent a protein group (orthologues, classes, families, etc.).

The benefit of using this search for the identification of proteins from the unique peptide list is 2-fold. The most significant result is that it provides a list of proteins that can be specifically and definitively identified by any given single peptide sequence (e.g., that result from searching MS/MS data), what we term a unique protein identifier. In addition to the uniquely identified proteins, the search also indicates groups of proteins that are identifiable by a given peptide, where that specific peptide sequence is present within several proteins (e.g., especially in the case in paralogs). These peptides, however, cannot be used to definitively identify a specific member in that protein group. Hence, these protein groups must be reported as such—they represent proteins that possess in common a given peptide sequence that was identified by MS/MS, but cannot be listed as uniquely identified (Supporting Information Table 4).

In some instances, however, a member of a protein group may have also been identified by a unique protein identifier (i.e., peptide). This situation has been observed in the case of TCF7, a protein that has a unique protein identifier as well as a peptide whose sequence is present in another eight different proteins (Figure 3). The likelihood that the common peptide is in fact from TCF7 and not one of the other proteins is greater, but just as another protein cannot be definitively identified from that group, the presence of one of those proteins in the group cannot be ruled out. The end result is that of the 12 389 uniquely identified peptides, 2911 of these peptides identify 1057 protein groups.

Gene ontology analysis of the unique proteins identified (Supporting Information Table 3) in this study revealed an extensive coverage of proteins from all aspects of biological processes, molecular functions and cellular localization (data not shown).²⁹ Interestingly, a large percentage of membrane proteins, approximately 41%, were observed in this study without the aid of currently established methods for enrichment of this important class of proteins.^{30–32} This result suggests that a large portion of the mouse serum proteome constituents originate from membrane proteins. Furthermore, utilizing the predicted membrane protein topology derived from a TMHMM analysis of these membrane proteins reveals that 66% of the peptides identified originate from extracellular domains. This finding is exemplified in the case where four peptides were identified from the glutamate receptor KA-2, a multipass integral membrane protein that belongs to the glutamate-gated ion channel family. Of these four peptides, three are predicted to arise from the extracellular domain (Figure 4). This finding shows that clipped or shed proteins from the surface of cells proximal to the circulatory system contribute significantly to the makeup of this biofluid proteome.

Discussion

Current hypotheses predict that serum perfusion through tissues should endow an abundance of information regarding cellular homeostasis or pathophysiology.³³ Until recently, however, our previous understanding of the human serum proteome was limited to only a few hundred proteins, most of which were predicted to be present in serum.¹ A number of published and ongoing proteomic investigations have been or are being conducted of human serum that will usher in a new paradigm of understanding of this complex and vital clinical sample. 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 due to the intrinsic variability in each serum sample attributed to genetics, lifestyle, and environmental differences among people.

For these reasons, mouse models of human disease generated either by transgenic or xenograft techniques have become increasingly critical for experimentally isolating the disease mechanism from artifacts introduced due to sample heterogeneity prevalent in human-derived samples. Our understanding of mouse serum and hence the extent to which proteomic investigations of serum derived from mouse models of human diseases are translatable to humans remains unclear. This situation is largely due to the lack of a basic understanding of the mouse serum proteome. In an effort to redress this issue, we have undertaken a global proteomic investigation of mouse serum that resulted in the identification of over 4500 unique proteins (Supporting Information Table 3).

Classification of the proteins identified in this investigation by a number of bioinformatic methods demonstrate that, in addition to the expected proteins in serum, a large number are present with origins from nearly every compartment of the cell and every functionality. Indeed, an astonishingly high number of membrane proteins (~41%) were identified. Arising from their overall insolubility in aqueous environments, it is likely that a significant portion of the membrane proteins identified in this investigation were due to the presence of clipped or shed fragments from the parent protein. If this situation is the case, then we hypothesize that the preponderance of the peptides identified should arise from the extracellular rather than the intracellular or transmembrane domains of these proteins. We find that, by mapping the peptides identified from the putative membrane proteins, 66% of the peptides identified do indeed arise from the extracellular domains. Hence, it is evident from the identified proteins that mouse serum represents a unique looking glass into the organism whereby no specific cellular genome is expressed *per se*, but rather the protein content is likely contributed by the summation of all cellular genomic expression in the organism. The collection of information in serum is comprised not only of the expected circulatory proteins, such as complement factors, immunoglobulins, and binding proteins, but also of peptides and proteins that are likely secreted and shed into the blood from cells in every tissue. It is within this information that future discovery of biomarkers that possess greater power for disease diagnosis may be accomplished.

It is likely that development of disease diagnostics that rely on detection or quantitation of a single protein biomarker will lack the required sensitivity and specificity when applied to large heterogeneous populations. A number of studies promote the view that these limitations may be overcome using panels of biomarkers that together possess greater fitness for accurate

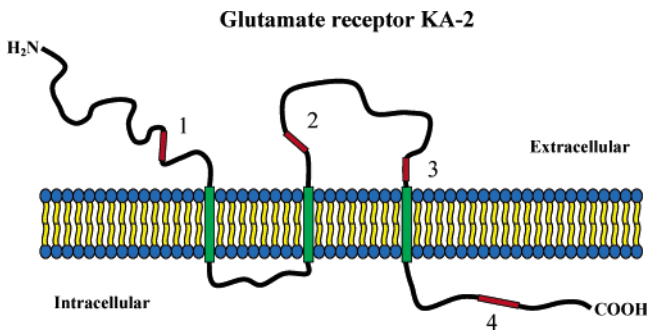


Figure 4. Mapping of observed peptides from the glutamate receptor KA-2. Four peptides were identified from this membrane protein that is predicted to possess three transmembrane domains (shown as green boxes). On the basis of the TMHMM topology prediction algorithm, three of the four observed peptides are predicted to arise from extracellular domains. The predicted peptide locations are highlighted in red and possess the following sequences: 1: ${}_{464}\text{LRLVEDGLYGAPEP}_{477}$; 2: ${}_{652}\text{VESADDLADQTNIEYGT}_{666}$; 3: ${}_{790}\text{KEEDHRAKGLGME}_{802}$; 4: ${}_{884}\text{LSNGKLYSAGAGGDAGAHG}_{903}$.

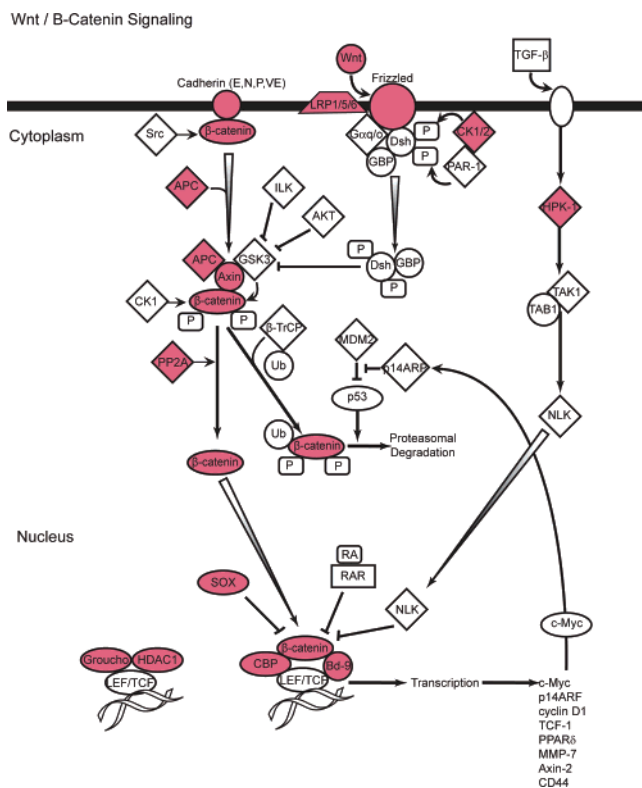


Figure 5. Pathway analysis of identified proteins. Representative protein pathway analysis of proteins identified in this sample (highlighted in red) from the Wnt/ β -catenin signaling pathway.

of the system as a whole, but also allow for detailed investigation of, for example, signal transduction pathway constituents that may hold the key for identifying dysfunctional or dysregulated cell growth, as seen in cancer. Clearly, if multiple proteins from deranged signal transduction pathways could be directly identified and utilized clinically as disease biomarkers, their overall diagnostic sensitivity and specificity would be expected to surpass that obtainable with those simply resulting from epiphenomena.

In 2002, the HUPO launched the PPP, which is a global effort involving 47 laboratories dedicated to conducting the most comprehensive proteomic characterization of human serum and plasma. The goal of the PPP is to gain an in depth understanding of the protein constituents in human plasma/serum, to identify sources of physiological, pathological, and pharmacological variability along with differences arising from genetics, nutrition, and lifestyle. The pilot phase of this project is anticipated to draw to conclusion with a series of publications describing the findings of the PPP consortium.²² Certainly, a greater understanding of human serum is needed and many interesting findings will result from this consortium and the results that ensue. Historically, the greatest challenge to our understanding of human serum has undoubtedly been complicated by artifactual heterogeneity arising from the inability to carefully control sources of random biological variability due to genetics, nutrition and lifestyle, as mentioned. 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. Hence, this study is timely as it represents the most extensive proteomic characterization of mouse serum that will serve as a benchmark database for conducting the largest comparison between the mouse and human serum proteomes. This study, in combination with that soon-to-be-released from the HUPO PPP, provides the critical mass of proteomic data required to evaluate the effectiveness of translating biomarker discoveries in mouse models to human diseases.

Abbreviations: CID, collision-induced dissociation; IT, ion trap; LC, liquid chromatography; μ , microcapillary; MS, mass spectrometry; MS/MS, tandem MS; SCX, strong cation exchange; WCX, weak cation exchange; WAX, weak anion exchange.

Acknowledgment. This work was supported by Federal funds from the National Cancer Institute, National Institutes of Health, under Contract NO1-CO-12400.

Supporting Information Available: Supporting Information Tables 1–4. 12 389 unique peptides identified that correspond to 4567 unique proteins and 1057 protein groups. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Anderson, N. L.; Anderson, N. G. *Mol. Cell Proteomics* **2002**, *1*, 845–867.
- 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.
- Tirumalai, R. S.; Chan, K. C.; Prieto, D. A.; Issaq, H. J.; Conrads, T. P.; Veenstra, T. D. *Mol. Cell. Proteomics* **2003**, *2*, 1096–1103.
- Carrette, O.; Demalte, I.; Scherl, A.; Yalkinoglu, O.; Corthals, G.; Burkhard, P.; Hochstrasser, D. F.; Sanchez, J. C. *Proteomics* **2003**, *3*, 1486–1494.
- Cepok, S.; Jacobsen, M.; Schock, S.; Omer, B.; Jaekel, S.; Boddeker, I.; Oertel, W. H.; Sommer, N.; Hemmer, B. *Brain* **2001**, *124*, (Pt 11), 2169–2176.
- Whitehouse, C.; Solomon, E. *Gynecol. Oncol.* **2003**, *88*, (1 Pt 2), S152–157.
- Grossklaus, D. J.; Smith, J. A.; Shappell, S. B.; Coffey, C. S.; Chang, S. S.; Cookson, M. S. *Urol. Oncol.* **2002**, *7*, 195–198.
- Rothmund, D. L.; Locke, V. L.; Liew, A.; Thomas, T. M.; Wasinger, V.; Rylatt, D. B. *Proteomics* **2003**, *3*, 279–287.
- Zhou, M.; Lucas, D. A.; Chan, K. C.; Issaq, H. J.; Petricoin, E. F., 3rd; Liotta, L. A.; Veenstra, T. D.; Conrads, T. P. *Electrophoresis* **2004**, *25*, 1289–1298.

- (10) Pieper, R.; Su, Q.; Gatlin, C. L.; Huang, S. T.; Anderson, N. L.; Steiner, S. *Proteomics* **2003**, *3*, 422–432.
- (11) Chertov, O.; Biragyn, A.; Kwak, L. W.; Simpson, J. T.; Boronina, T.; Hoang, V. M.; Prieto, D. A.; Conrads, T. P.; Veenstra, T. D.; Fisher, R. J. *Proteomics* **2004**, *4*, 1195–1203.
- (12) Sato, A. K.; Sexton, D. J.; Morganelli, L. A.; Cohen, E. H.; Wu, Q. L.; Conley, G. P.; Streltsova, Z.; Lee, S. W.; Devlin, M.; DeOliveira, D. B.; Enright, J.; Kent, R. B.; Wescott, C. R.; Ransohoff, T. C.; Ley, A. C.; Ladner, R. C. *Biotechnol. Prog.* **2002**, *18*, 182–192.
- (13) Hahn, W. C.; Counter, C. M.; Lundberg, A. S.; Beijersbergen, R. L.; Brooks, M. W.; Weinberg, R. A. *Nature* **1999**, *400*, 464–468.
- (14) Cohen, B. *Nat. Genet.* **1999**, *21*, 249–251.
- (15) Artandi, S. E.; DePinho, R. A. *Nat. Med.* **2000**, *6*, 852–855.
- (16) Cardiff, R. D. *Microsc. Res. Technol.* **2001**, *52*, 224–230.
- (17) Green, J. E.; Desai, K. V.; Ye, Y.; Kavanaugh, C.; Calvo, A.; Huh, J. I. *Toxicol. Pathol.* **2004**, *32*, (Suppl 1), 84–89.
- (18) Maronpot, R. R.; Flake, G.; Huff, J. *Toxicol. Pathol.* **2004**, *32*, (Suppl 1), 40–48.
- (19) Schuh, J. C. *Toxicol. Pathol.* **2004**, *32*, (Suppl 1), 53–66.
- (20) Waterston, R. H.; Lindblad-Toh, K.; Birney, E.; Rogers, J.; Abril, J. F.; Agarwal, P.; Agarwala, R.; Ainscough, R.; Alexandersson, M.; An, P.; Antonarakis, S. E.; Attwood, J.; Baertsch, R.; Bailey, J.; Barlow, K.; Beck, S.; Berry, E.; Birren, B.; Bloom, T.; Bork, P.; Botcherby, M.; Bray, N.; Brent, M. R.; Brown, D. G.; Brown, S. D.; Bult, C.; Burton, J.; Butler, J.; Campbell, R. D.; Carninci, P.; Cawley, S.; Chiaromonte, F.; Chinwalla, A. T.; Church, D. M.; Clamp, M.; Clee, C.; Collins, F. S.; Cook, L. L.; Copley, R. R.; Coulson, A.; Couronne, O.; Cuff, J.; Curwen, V.; Cutts, T.; Daly, M.; David, R.; Davies, J.; Delehaunty, K. D.; Deri, J.; Dermitzakis, E. T.; Dewey, C.; Dickens, N. J.; Diekhans, M.; Dodge, S.; Dubchak, I.; Dunn, D. M.; Eddy, S. R.; Elnitski, L.; Emes, R. D.; Eswara, P.; Eyas, E.; Felsenfeld, A.; Fewell, G. A.; Flicek, P.; Foley, K.; Frankel, W. N.; Fulton, L. A.; Fulton, R. S.; Furey, T. S.; Gage, D.; Gibbs, R. A.; Glusman, G.; Gnerre, S.; Goldman, N.; Goodstadt, L.; Grafham, D.; Graves, T. A.; Green, E. D.; Gregory, S.; Guigo, R.; Guyer, M.; Hardison, R. C.; Haussler, D.; Hayashizaki, Y.; Hillier, L. W.; Hinrichs, A.; Hlavina, W.; Holzer, T.; Hsu, F.; Hua, A.; Hubbard, T.; Hunt, A.; Jackson, I.; Jaffe, D. B.; Johnson, L. S.; Jones, M.; Jones, T. A.; Joy, A.; Kamal, M.; Karlsson, E. K.; Karolchik, D.; Kasprzyk, A.; Kawai, J.; Keibler, E.; Kells, C.; Kent, W. J.; Kirby, A.; Kolbe, D. L.; Korf, I.; Kucherlapati, R. S.; Kulbokas, E. J.; Kulp, D.; Landers, T.; Leger, J. P.; Leonard, S.; Letunic, I.; Levine, R.; Li, J.; Li, M.; Lloyd, C.; Lucas, S.; Ma, B.; Maglott, D. R.; Mardis, E. R.; Matthews, L.; Mauceli, E.; Mayer, J. H.; McCarthy, M.; McCombie, W. R.; McLaren, S.; McLay, K.; McPherson, J. D.; Meldrim, J.; Meredith, B.; Mesirov, J. P.; Miller, W.; Miner, T. L.; Mongin, E.; Montgomery, K. T.; Morgan, M.; Mott, R.; Mullikin, J. C.; Muzny, D. M.; Nash, W. E.; Nelson, J. O.; Nhan, M. N.; Nicol, R.; Ning, Z.; Nusbaum, C.; O'Connor, M. J.; Okazaki, Y.; Oliver, K.; Overton-Larty, E.; Pachter, L.; Parra, G.; Pepin, K. H.; Peterson, J.; Pevzner, P.; Plumb, R.; Pohl, C. S.; Poliakov, A.; Ponce, T. C.; Ponting, C. P.; Potter, S.; Quail, M.; Reymond, A.; Roe, B. A.; Roskin, K. M.; Rubin, E. M.; Rust, A. G.; Santos, R.; Sapojnikov, V.; Schultz, B.; Schultz, J.; Schwartz, M. S.; Schwartz, S.; Scott, C.; Seaman, S.; Searle, S.; Sharpe, T.; Sheridan, A.; Shownkeen, R.; Sims, S.; Singer, J. B.; Slater, G.; Smit, A.; Smith, D. R.; Spencer, B.; Stabenau, A.; Stange-Thomann, N.; Sugnet, C.; Suyama, M.; Tesler, G.; Thompson, J.; Torrents, D.; Trevaskis, E.; Tromp, J.; Ucla, C.; Ureta-Vidal, A.; Vinson, J. P.; Von Niederhausern, A. C.; Wade, C. M.; Wall, M.; Weber, R. J.; Weiss, R. B.; Wendl, M. C.; West, A. P.; Wetterstrand, K.; Wheeler, R.; Whelan, S.; Wierzbowski, J.; Willey, D.; Williams, S.; Wilson, R. K.; Winter, E.; Worley, K. C.; Wyman, D.; Yang, S.; Yang, S. P.; Zdobnov, E. M.; Zody, M. C.; Lander, E. S. *Nature* **2002**, *420*, 520–562.
- (21) Pennacchio, L. A. *Mamm. Genome* **2003**, *14*, 429–436.
- (22) Omenn, G. S. *Proteomics* **2004**, *4*, 1235–1240.
- (23) Sonnhammer, E. L.; von Heijne, G.; Krogh, A. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1998**, *6*, 175–182.
- (24) Moller, S.; Croning, M. D.; Apweiler, R. *Bioinformatics* **2001**, *17*, 646–653.
- (25) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. *J. Mol. Biol.* **2001**, *305*, 567–580.
- (26) Kall, L.; Krogh, A.; Sonnhammer, E. L. *J. Mol. Biol.* **2004**, *338*, 1027–1036.
- (27) 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. *Mol. Cell. Proteomics* **2004**, *3*, 896–907. Epub 2004 Jun 30.
- (28) Peng, J.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P. *J. Proteome Res.* **2003**, *2*, 43–50.
- (29) Chan, K. C.; Lucas, D. A.; Hise, D.; Schaefer, C. F.; Xiao, Z.; Janini, G. M.; Beutow, K. H.; Issaq, H. J.; Veenstra, T. D.; Conrads, T. P. *Clin. Proteomics* **2004**, *1*, 101–226.
- (30) Blonder, J.; Conrads, T. P.; Yu, L. R.; Terunuma, A.; Janini, G. M.; Issaq, H. J.; Vogel, J. C.; Veenstra, T. D. *Proteomics* **2004**, *4*, 31–45.
- (31) Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. *Nat. Biotechnol.* **2001**, *19*, 946–951.
- (32) Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Yates, J. R., 3rd. *Nat. Biotechnol.* **2003**, *21*, 532–538.
- (33) Liotta, L. A.; Ferrari, M.; Petricoin, E. *Nature* **2003**, *425*, 905.
- (34) Gumbiner, B. M. *Cell* **1996**, *84*, 345–357.
- (35) Gottardi, C. J.; Gumbiner, B. M. *Am. J. Physiol. Cell Physiol.* **2004**, *286*, C747–756. Epub 2003 Nov 12.

PR050107R