Utilizing Human Blood Plasma for Proteomic Biomarker Discovery †

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · AUGUST 2005

Impact Factor: 4.25 · DOI: 10.1021/pr0500657 · Source: PubMed

7 AUTHORS, INCLUDING:

Jon M Jacobs
Pacific Northwest National Laboratory
83 PUBLICATIONS **2,984** CITATIONS
SEE PROFILE

Wei-Jun Qian
Pacific Northwest National Laboratory
132 PUBLICATIONS **5,763** CITATIONS
SEE PROFILE

Tao Liu
Pacific Northwest National Laboratory
75 PUBLICATIONS **3,281** CITATIONS
SEE PROFILE

Richard D Smith
Pacific Northwest National Laboratory
1,135 PUBLICATIONS **45,939** CITATIONS
SEE PROFILE

reviews Journal of proteome research

Utilizing Human Blood Plasma for Proteomic Biomarker Discovery†

Jon M. Jacobs, Joshua N. Adkins, Wei-Jun Qian, Tao Liu, Yufeng Shen, David G. Camp II, and Richard D. Smith*

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352

Received March 15, 2005

Candidate proteomic biomarker discovery from human plasma holds both incredible clinical potential as well as significant challenges. The dynamic range of proteins within plasma is known to exceed 10^{10}, and many potential biomarkers are likely present at lower protein abundances. At present, proteomic based MS analyses provide a dynamic range typically not exceeding $\sim 10^3$ in a single spectrum, and $\sim 10^4-10^6$ when combined with on-line separations (e.g., reversed-phase gradient liquid chromatography), and thus are generally insufficient for low level biomarker detection directly from human plasma. This limitation is providing an impetus for the development of experimental methodologies and strategies to increase the possible number of detections within this biofluid. Discussed is the diversity of available approaches currently used by our laboratory and others to utilize human plasma as a viable medium for biomarker discovery. Various separation, depletion, enrichment, and quantitative efforts as well as recent improvements in MS capabilities have resulted in measurable improvements in the detection and identification of lower abundance proteins (by $\sim 10-10^2$). Despite these improvements, further advances are needed to provide a basis for discovery of candidate biomarkers at very low levels. Continued development of depletion and enrichment techniques, coupled with improved pre-MS separations (both at the protein and peptide level) holds promise in extending the dynamic range of proteomic analysis.

Keywords: biomarker • plasma • serum • human • proteomics • mass spectrometry • MS

Introduction

In just the past few years, there has been an explosion of new developments in the field of proteomics with new technologies being applied across multiple biological models and conditions. Mass spectrometry (MS), coupled with pre-MS high efficiency separations, has played an ever increasing role in this expansion as strategies are created to address complex and challenging biological samples. One application which has benefited from this expansion of proteomic analyses is the study of human plasma. Human plasma is collected on a large scale, with relatively large protein quantities (mg amounts) widely available for the study of numerous disease states and conditions.[1] Since blood plasma or serum may contain some residual and potentially detectable combination of all the differentiated sub-proteomes of the body, plasma analyses may provide information regarding these tissues, and in addition be potentially informative in regards to almost any disease state. In short, human plasma is potentially the single most informative sample that can be collected from an individual that describes their current state of health. Blood plasma would appear ideal for proteomic analysis with nearly limitless clinical applications, but the proteomic investigation of human plasma has proved to be more challenging when compared with other proteomic applications, i.e., microbial or cell based studies.[2−4]

For decades, the human plasma proteome has been studied mostly by using traditional chemical fractionation and electrophoresis separation techniques. In the mid-1970s, the application of high resolution 2-dimensional (2D) gel electrophoresis separation methods, including seminal work by Anderson et al., contributed largely to the understanding of plasma proteins at the time.[5] The eventual introduction of mass spectrometry, initially coupled to 2D gel analysis, greatly improved the identification of proteins from plasma. More recently, coupling such techniques with other chromatographic separations (anion exchange and size-exclusion chromatography) has been able to generate the detection of 3700 different species as "spots" and identification of 325 proteins.[6] These and other past plasma based 2D electrophoresis works have been reviewed by Anderson et al.[1] Later, by combining liquid-based chromatography efforts with MS, through the instrumental development of electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) interfaces, the throughput and sensitivity of peptide-level protein identifications using MS was greatly improved. The utilization of digested peptide fragmentation patterns generated by tandem

† Part of the Biomarkers special issue.

* To whom correspondence should be addressed. Dr. Richard D. Smith. Biological Sciences Division, Pacific Northwest National Laboratory, P.O. Box 999, MSIN: K8−98, Richland WA, 99352.

10.1021/pr0500657 CCC: $30.25

Published on Web 05/20/2005

MS (MS/MS) studies now allow the comparison of observed fragmentation spectra with a database of theoretical (i.e., presumed possible) peptides for their identification. This combination of multidimensional liquid chromatography (LC) separations coupled to MS/MS analysis and post-data processing, often termed "shotgun" proteomics,[3] was originally demonstrated in the yeast model but has now been applied to a number of model systems which have an appropriate protein sequence database. Large scale application of this method toward human samples became feasible largely due to the DNA sequencing efforts of the human genome which has provided a comprehensive database of human protein sequences.

The first global shotgun proteomic application toward human plasma/serum was performed by Adkins et al. in 2002 where 490 proteins were identified in human serum.[7] This initial study applied a standard shotgun proteomic analysis, offline strong cation exchange (SCX) LC prior to the reversed phase separation (RPLC) coupled to LCQ MS/MS analysis, in the analysis of an immunoglobulin depleted serum sample. This survey of human serum was only an initial step into using LC−MS/MS based techniques for the detection and identification of plasma proteins and demonstrated both the potential and limitations that are possible with the sample. What follows is a brief review of recent studies involving both proteomic plasma identifications, and the strategies being employed for protein detection. The challenges associated with detecting biomarkers in plasma will be discussed followed by specific approaches being initially explored in our laboratory and others for improving the detection, identification, and quantitation of potential biomarkers.

Recent Advances in Plasma Proteome Characterization. In the short time since the Adkins et al. work, numerous studies have now expanded the plasma proteomic field and provided insights into both novel protein identifications and separation/depletion/enrichment techniques for targeting sub-proteomes of interest. In 2004, a compiled nonredundant list of detected plasma proteins from 4 different sources, three previously published proteomic studies and one comprehensive literature search, resulted in the total identification of 1175 unique proteins.[8] Interestingly, from this list of 1175 proteins only 195 proteins (17%) were observed in at least two of the previous studies with only 46 proteins (4%) identified in all four. This represented a surprisingly low overlap which was attributed to three main sources, different separation and detection methods between the data sets, different serum/plasma samples utilized, and misidentifications within each of the studies. It is likely that all three of these conditions contributed to the reduced overlap of the samples, but with improved plasma proteome efforts it is increasingly clear that the identification criteria of the MS datasets used at the time can be significantly improved.

Additional work performed by Shen et al. applied nano-LC separations methods to increase the dynamic range of detection of plasma proteins[9]. This study demonstrated that both improved separations and measurement sensitivity provide improved dynamic range of detection and identification of plasma proteins, and thus overall proteome coverage. No depletion strategy was utilized and only 365 μg of plasma protein was needed for the entire study. Also reported was a possible range of detected plasma proteins, between 800 and 1682 proteins, simply by using slightly different data cutoff values, all of which were based upon previously published criteria.

A large scale approach for protein identification in plasma utilized a large volume of pooled samples (2.5 L) for separation and depletion approaches for identification of >700 proteins and polypeptides from human plasma.[10] The approach mainly targeted lower molecular weight proteins and, if sample size and time permitted, could represent an alternative approach for protein detection. Other researchers have also attempted to target the low molecular weight fraction of plasma as many lower abundance proteins of potential interest are believed to specifically populate this sub-proteome.[11−14] The physiological uptake of proteins from plasma is size based, with a kidney filtration cutoff value of approximately 45 kDa, so any protein or peptide smaller than this value should be rapidly cleared from the plasma volume. It is believed that the circulation half-life of this low molecular weight fraction is directly related to its binding affinity to large high abundant carrier proteins.[15] This has been supported by proteomic based studies that have demonstrated significant detection of low molecular weight proteins in specific carrier protein (serum albumin) fractions.[16,17] There has also been a variety of depletion strategies proposed specific for plasma analysis both in the published literature and available commercially.[18−21] These approaches will be discussed in more detail later.

Multiple plasma specific separation protocols have been proposed, with most based mainly from variations off of standard biochemical separations techniques.[22−24] Also, one specific enrichment technique, N-linked glycopeptide capture has been shown to target proteins containing the N-linked glycosylation modification, potentially enriching for all extracellular and secreted proteins which fall in this category, a potentially rich source of protein biomarkers in plasma.[25] This technique is not specific for plasma, as the chemistry can be applied with any protein sample, but its application appears well suited to enrich the extracellular sub-proteome of biofluids. In addition, there are other studies, though not directly related to proteomic analysis of plasma, that merit further mention. Initial work involving the extensive identification of the peptide component in plasma was performed previously by Richter and Schulz-Knappe et al.. Using MADLI time-of-flight (TOF) instrumentation, they report a comprehensive database of both identified peptides and protein fragments using human hemofiltrate samples, approximately 5000 different peptide masses detected.[26] A majority of identified peptide sequences corresponded to known high abundance plasma proteins in addition to peptide hormones, cytokines, and novel peptide identifications, revealing the diversity of the peptide subpopulation found in plasma. Seeberger and co-workers investigation of glycosylation moieties has greatly increased the understanding of the role of carbohydrate structure in biological systems.[27,28] New developments are underway for the synthentic development of glycosylated structures that can be utilized as both potential therapeutic agents and as microassay components to further elucidate carbohydrate binding interactions in biological systems.

The proteomics field in general is maturing rapidly. Post-MS data processing has greatly improved in both throughput and accuracy of results as there is now a deeper understanding of the confidence levels of "filtering" criteria for peptide/protein identification from MS/MS results, especially for the large mammalian systems.[29−32] For example, the need was evident for more stringent and adaptive criteria for MS/MS based identifications when searching larger databases, as there is a proportionally greater propensity for false positive identifica-

tions. However, even when using the same database, noticeably different false positive rates were observed between different sample types (i.e., plasma and human cell lines), illustrating that the proper criteria is often sample dependent,[29] as well as method and instrument dependent, etc. The field is now moving beyond the standard reporting of mere "parts lists" of specific biological systems and is playing an ever increasing role in specific biological applications for the elucidation of proteins or protein modifications in comparative studies. This shift has centered attention on the need for accurate quantitative proteomic information which in the past has been challenging to obtain in an MS based system of this complexity. There are now an array of techniques which allows the generation of reproducible MS quantitative information that can be applied to virtually any biological model system. Additionally, developments continue to increase sensitivity, dynamic range, and throughput, resulting in the detection and identification of more components from smaller samples, increasing the potential to obtain meaningful comparative results from multiple samples not previously accessible to proteomic methodologies i.e., analysis of small tissue biopsies, flow-cytometry collections, and small populations of cells.

One of the largest plasma proteomics efforts has been undertaken by the Human Proteome Organization (HUPO). The Plasma Proteome Project (PPP), guided by Gil Omenn, is one of the most mature of the major initiatives of HUPO, which also include initiatives in proteomics standards and proteomics investigations of the liver and brain. The PPP initiative has three stated long-term goals: (1) determine the extent of variation across individual plasmas within and across populations; (2) determine the extent and source of variation in an individual's plasma through time; and (3) to make a comprehensive analysis of the protein constituents of plasma.[33] The pilot effort has involved eighteen laboratories from seven nations who have contributed mass spectrometry-based peptide/protein identifications. As shown previously, plasma identifications between laboratories analyzing different plasma samples with different techniques, and different protein search databases resulted in a poor concordance.[34] As part of the HUPO PPP, a limited number of samples were made available to the participating laboratories. This allowed the pilot effort to make some assessments of the overall concordance when using the same sample, a single protein database for searches, and in many cases similar approaches.[35] The final results of this effort are still forthcoming, but initial assessments suggested a limited concordance of characterized plasma proteins when using well accepted approaches and criteria. Taken together with other studies, this suggests that plasma sample analysis differs in some fashion from typical proteomics experiments. Our laboratory contribution to this analysis was a pilot analysis of the accurate mass and time (AMT) tag approach which resulted in a limited number of multiple peptide per protein identifications[36] from a search database made up of two previous analyses.[7,9]

Challenges of Plasma Proteomics. Typical mammalian cellular based proteome identification approaches are now routinely obtaining thousands of protein detections per study, which appears to provide a sufficient depth and dynamic range to observe the possible significant changes in most comparative studies.[4,37−39] When using similar techniques, the number of confidently identified plasma proteins has been consistently much less,[7,9] an interesting observation considering that plasma is believed to be the most complex human proteome.[1] The unique nature of plasma, coupled with the very low levels of likely candidate biomarkers of interest present significant analytical challenges for proteomics, but the enormous clinical and diagnostic potential of such measurements dictates that efforts continue. The overriding challenge in proteomic plasma analysis lies in the steep dynamic range present in the sample, which begins with serum albumin at ~45 mg/mL and ends with a series of detected cytokines (and potentially many other proteins) at 1−10 pg/mL and possibly beyond. The limit of LC−MS analysis does not lie in the absolute detection amounts, i.e., FTICR−MS can routinely detect individual species present in the attomole[40] or even zeptomole range with advanced instrumentation,[41,42] but is limited by the overall sample quantity used in an analysis. These parameters define the dynamic range of the analysis for a specific sample type. A single mass spectrum provides a dynamic range of no more than $\sim 10^3$ and with combined on-line separations (e.g., reversed phase gradient liquid chromatography RPLC) is limited to a dynamic range of $<10^4$ to 10^5 (~5.0−0.5 μg/mL protein in plasma limit in the presence of albumin). While an LC−FTICR analysis can stretch this range to $>10^6$ (<50 ng/mL protein in plasma limit), this still falls well short of the $>10^{10}$ dynamic range desired to detect many known or suspected candidate plasma proteins. There are techniques which can assist in bridging this remaining dynamic range gap (these will be discussed in depth later) but generally an increase in the separation quality prior to MS analysis can allow for more complex samples to be addressed, while the lack of dynamic range capability in MS can be compensated, to a certain extent, by a more intensive pre-MS separation (SCX, RPLC).[43]

Figure 1 shows the comparison of a number of antibody detected proteins in plasma and shows the striking range of proteins of relatively high abundance with a swift drop in concentration levels for other proteins in plasma.[44] Contained above this observed sigmoidal curve within the $\sim 10^6$ dynamic range (~50 ng/mL) is only a limited fraction of the total protein species, most of which are of very high abundance, and thus unlikely to include many new potential proteins of interest. These higher abundance proteins can be classified as "classic" plasma proteins, those which are secreted, predominately by the liver and intestines, to actively perform their function in the blood stream. Virtually all proteins found in plasma at a concentration >1 μg/mL are classic plasma proteins involved in proteolysis, inhibition, binding, transport, coagulation, and immune response, among other functions.

So what other proteins are present in human plasma? Generally, the plasma proteome can be broken down into two main categories, (Anderson et al. further divided these into eight separate functional groups).[1] These include (1) the "classic" proteins whose function depends on their presence in plasma, and (2) transient, cellular leakage, shed, or other secreted proteins whose function is not tied directly to their presence in plasma (at least not yet to our knowledge), but which utilize plasma for either transportation purposes for localization or for dilution and filtered removal effects. Each protein category is found at differing concentration levels in plasma. For instance, as previously discussed, virtually all high abundance proteins fall under the "classic" category. These proteins are very well characterized and many have been utilized as diagnostic indicators of various conditions and/or diseases, i.e., serum amyloid A and C-reactive proteins for acute phase response,[45−47] Apo B_{100} as marker of LDL particle levels for predictor of atherosclerosis,[48] ferritin and tranferrin as

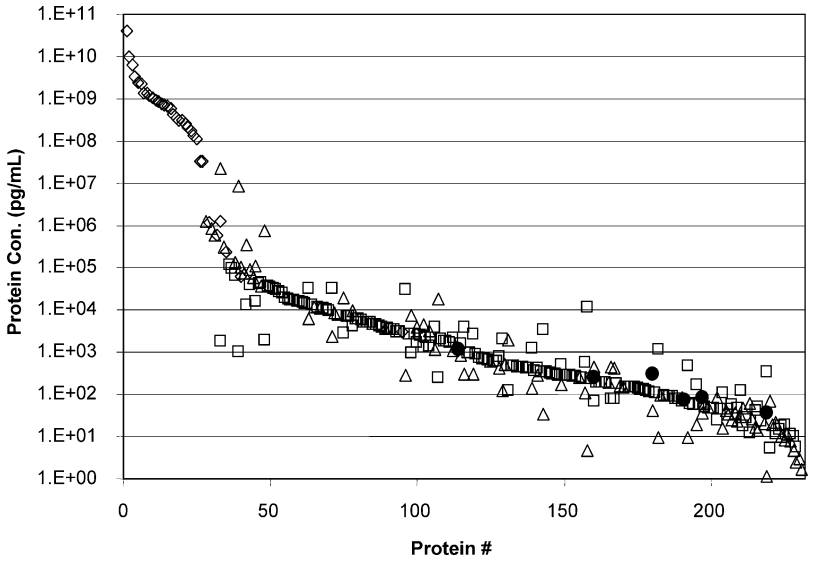

Figure 1. Compilation of plasma protein concentrations measured by multiple immunoassay and antibody microarray assays from multiple laboratories. The geometric mean concentration over all the samples is plotted for each of 295 quantitative assays for 231 unique proteins. For analytes measured by more than one laboratory, the geometric mean concentration derived by each laboratory is displayed. Observed is a dramatic sigmoidal curve representing the concentration differential between the higher abundant "classic" plasma components and other lower abundant proteins (This figure was reproduced with the permission of Brian Haab).[44]

indicators of iron deficiency and anemia[49], and elevated myoglobin levels for diagnosis of myocardial infarction.[50]

Also contained within this "classical" category are the immunoglobulin (Ig) classes of proteins which represent roughly 25% of the total protein content of plasma.[1] The intrinsic nature of Ig proteins themselves represents challenges based upon the inherent diversity and complexity of their complementarity-determining regions (CDRs). For peptide based proteomic studies, immunoglobulin proteins, if retained in the analysis, have the potential to generate a library of random peptide sequences which can greatly increase the number of "background" species present to challenge peptide detection and identification, thus presenting not only an overall abundance challenge (~20 mg/mL concentration) but a complexity issue as well.

The second class of proteins is the largest, and from which most new proteomic identifications originate. Included are the subgroup of known secreted proteins (cytokines, receptor ligands, hormones) which are generally considered passenger proteins (some more transient than others) that utilize plasma for transportation, localization, and mediation of cellular responses. Other detected intracellular proteins are believed to originate from leakage of cells due to injury or damage. Due to the flow of species from the volume of interstitial fluid into the circulating plasma, there is a higher propensity of extracellular, outer-membrane, or locally secreted proteins to be detected. This also applies to both cerebral spinal fluid and synovial fluid. Also, cellular proteins exclusively localized to either the vascular tissue or the circulating cellular components of the blood (erythrocytes, lymphocytes, monocytes, platelets, etc.) are also likely to have higher representation, especially for extra-cellular or outer-membrane attached proteins. Overall, this second group is the least characterized, but possibly most interesting for the area of biomarker discovery with protein concentrations believed to range from low μg/mL to pg/mL levels, and possibly extending to levels below the detection limits of traditional ELISA assays (~1 pg/mL).[51]

Biomarker Discovery Strategies. Most proteomic studies that are utilizing human plasma are not intentionally targeting the high to moderately abundant proteins commonly found at the currently detectable dynamic range, but are attempting to focus on specific predictors or pre-disposition identifiers (i.e., candidate biomarkers) which can potentially distinguish different disease states or predict their early onset. It is the general understanding, whether correct or not, that these biomarkers of interest will not originate from classic plasma secretions but most likely from leakage, secretion, or shedding of proteins from the specific affected tissue, cell type, or cellular pathway. But once such proteins are deposited in the ~ 6 Ls of plasma, they have been severely diluted from their original localized concentration and will be only found at low concentration. It is this biomarker hypothesis that drives the need to extend the dynamic range of protein detection. This does not rule out the possibility of an aberrant splice variant or post-translationally modified higher abundance classic protein from taking on the role of a biomarker; however, the discovery of such a potentially useful biomarker at a high abundance level is probably unrealistic for most diseases. In such cases, it is unlikely that an overt biological role can be properly identified, which greatly increases the likelihood that the observed correlation (e.g., disease vs normal) arose due to chance or bias of some sort in the sample analysis.[52] Such potential candidate biomarkers are likely to be ineffective in accurately distinguishing between normal and disease states unless prior orthogonal validation has occurred to specifically link the perturbation to a biological role for the disease state.

The efficient detection and identification of specific biomarkers in human plasma depends on many factors related to both the plasma sample itself, as well as sample collection and processing methods. Any pre-MS sample handling protocol has direct influence upon the accuracy of the generated downstream data for the comparative analysis between samples. As for the plasma sample itself, as previously discussed, the overriding challenge is its large protein dynamic range, but other

intrinsic characteristics also hamper analysis. Proteolysis present in plasma generates multiple protein fragments that add to the complexity of the sample, essentially increasing the background noise of detection and reducing the number of identifications possible. Also, many plasma proteins are highly modified (i.e. posttranslational modifications, alternate splicing events, various isoforms) also increasing the complexity of the sample and often making it difficult to correctly identify the peptide/protein of interest. Furthermore, it is understood that many candidate biomarkers will likely exist at levels which lie beyond the capabilities of current proteomic techniques, so what can be done to overcome such challenges?

There are to date very few proteomic studies which have effectively attempted to detect lower level biomarkers directly from serum or plasma. One strategy in a number of recent examples involves using a secondary tissue or fluid of interest to first identify potential candidates followed by a screening of the complementary plasma sample for their detection. Ding et al. used 2D gel electrophoresis and MALDI-TOF instrumentation with cell cultures to help identify the overexpression of CK-19 in carcinoma metastasis, with subsequent demonstration of detectability in mouse serum samples.[53] Borozdenkova et al. utilized cardiac biopsies, 2D gel electrophoresis and ELISA verification in serum for identification of potential biomarkers of cardiac allograph rejection.[54] Both Liao et al. (using LC−MS/MS analysis) and Drynda et al. (using 2D gel electrophoresis) examined synovial fluid for identification of protein biomarkers of rheumatoid arthritis followed by verification in sera and plasma samples respectively.[55,56] Other studies of interest include Gravett et al.[57] which studied amniotic fluid using an array of separation and MS techniques for identification of polypeptide biomarkers for intra-amniotic infections and Celis et al.[58] which performed proteomic analysis on tumor interstitial fluid.

The use of surface-enhanced laser desorption and ionization (SELDI)-TOF approaches for the pattern recognition and diagnosis of various cancers has also been described.[59] Such an approach is based upon pattern recognition of generated MS spectra and potentially holds promise as a new area of discovery. Though not geared toward potential biomarker identification, when attempts have been made to identify peaks corresponding to patterns of interest, they have been often determined to be fragments of higher abundance plasma proteins.[60] This should not be surprising, as the dynamic range of detection with such approaches is very limited, and hence limits the proteins that fall within the detection window. Of vital importance is not only to effectively detect such variations in comparative analysis, but to correctly identify the species and link the perturbations to the disease state of interest.

Interestingly, when direct comparative plasma/serum results have been reported for a disease or condition, the vast majority of observed changes include only higher abundance classic plasma proteins.[61−64] This is a striking observation, but consistent with the known limited dynamic range of the current methods. Furthermore, a recent comprehensive study performed in our laboratory from a lipopolysaccharide treatment time course study using multidimensional LC separations prior to MS/MS for human plasma samples revealed few changes in most of the known higher abundance proteins (see Figure 2),[65,66] with the majority of changes observed involving midrange abundance proteins with known biological relevance to the experiment, i.e., acute phase and inflammatory response proteins. With the use of pre-MS separations, a dynamic range of greater than 10^6 was indicated for quantitative measurements without the use of depletion techniques for abundant proteins.

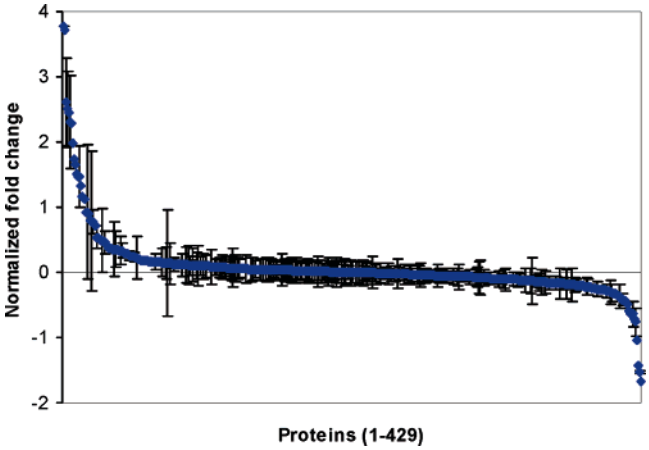

Figure 2. Normalized fold changes for 429 quantified proteins following LPS administration using ^{18}O stable isotope labeling. Results are the comparison between the baseline 0 h time point and 9 h after LPS administration. Abundance ratio for each protein shown was normalized to zero ($R - 1$). For ratios smaller than 1, normalized inverted ratios were calculated as $[1 - (1/R)]$. Error bar for each protein indicates the standard deviation for the abundance ratios from multiple peptides. Proteins without error bars were identified with single peptides. Observed is the overwhelming majority of protein concentrations that do not change, most of which are in the high abundant range.[66]

To address this challenge, direct proteomic plasma analysis can be approached in two different fashions. The first is to improve the technology itself (separations, mass spectrometry) to allow detection and identification over a greater dynamic range of relative protein abundances. The second is to effectively improve the plasma sample by enriching the proteins of potential interest and thus decreasing the dynamic range needed to detect more candidate biomarker proteins.

The first option would generally imply that technology advances would increase the range of relative peptide MS signal intensities obtainable for a given LC separation. Clearly, any sensitivity improvements would be helpful since the overall achievable dynamic range is defined by the sample loading (e.g., amount injected onto the LC column) and the minimum detection level for MS detection. Indeed there have been a number of very impressive advancements in both separations and MS technologies, i.e., the superseding of the ion trap instrumentation with the greatly improved linear ion trap mass spectrometers,[67] and a steady decrease in LC-capillary column diameters with an accompanying increase in sensitivity and capability for quantitation,[43,68] but further large improvements that will result in orders of magnitude increases in detectable dynamic range with similar sample sizes are unlikely. Leading edge LC−FTICR−MS technology provides at this time the most sensitive and powerful measurement platform, and some significant potential exists for increasing dynamic range when limited by sample complexity, rather than detection limits. We have demonstrated approaches for selectively excluding high abundance ions during accumulation prior to characterization.[69] Termed Dynamic Range Enhancement Applied to MS (DREAMS), a 35% increase in protein identifications was observed, but further improvements in this area can be made.

There are also efforts in our laboratory, as well as in others, to apply various novel intact protein separation and detection strategies for the intensive separation of higher abundant proteins, e.g., isoelectric focusing.[70] If proteins can be effectively separated prior to peptide digestion, then the potential dynamic range of detection can be extended.

Plasma Protein Detection Strategies. A. Top-Down vs Bottom-Up. As previously reviewed, there are several different proteomic platforms with multiple variations being utilized by researchers for samples analysis. The current proteomic methodology utilized in our laboratory as well as in most other laboratories is a "bottom up" approach where the protein mixtures are digested into their respective peptide forms in the first step, followed by the detection and identification of the unique peptides, and by inference the parent protein, by high resolution LC−MS/MS. Once a mixture of peptides has been identified, we create an extensive database composed of these peptide identifications (based upon both their exact mass and the observed elution time during the LC separation). This accurate mass and time (AMT) tag database is essentially used as a "look-up" table to speed peptide identification when using LC−FTICR for high throughput and highly sensitive quantitative measurements. This approach has been shown beneficial for increasing throughput (peptides need only to be MS/MS identified once) and in providing better quantitative information along with peptide identifications, based upon either direct LC−MS peak intensity values or in combination with various stable isotopic labeling methods.[66,71,72] The downside to a bottom-up approach is that there is an overall increase in the complexity of the sample as each protein species is reduced into multiple peptide components, and upon this reduction, some information concerning specific proteins is lost. To retain the protein specific information, often a "top-down" approach is applied (most commonly used with 2D gel electrophoresis) in efforts to characterize identifications at the protein level, then infer, based upon mass and elution characteristics, any potential variations in comparison to their known sequence. An LC−MS directed "top-down" approach is currently being pursued in our laboratory but still remains limited in the extent of protein coverage obtainable. The increased molecular weight of the species and the variations of the potential detected masses both present some challenges for correctly identifying intact proteins. The benefits of the "top-down" approach include the overall reduced complexity of the sample (requiring less intensive upstream separations and/or higher relative dynamic range of detection), and potentially improved quantitative results as the relative intensity values will directly correlate to the protein and not indirectly dependent upon multiple digested peptides. The retention of specific protein information will also potentially distinguish different protein isoforms or post-translational variations from the sample protein, i.e., multiple phosphorylation events on one protein. Such approaches have the potential to greatly enhance the application of proteomics in the biomarker discovery field, especially when searching can include specifically altered proteins or modifications.[73,74]

A further development being pursued in our laboratory as an addition to the AMT tag approach is a feature comparison and identification methodology which is directed toward those detected species in LC−FTICR that are not identified during the AMT database searching. For each detected isotopic distribution in an LC−MS analysis, we generate an accurate mass and specific elution time for that unique identification which is then used to correctly identify the sequence of the peptide. For each analysis there are a finite number of these detected isotopic features that are not identified by the AMT tag database, but whose presence and intensity could still be utilized for comparative analysis of samples. We are currently developing the informatics capabilities to align with very high reproducibly the measurements from multiple analyses to select "interesting" features (i.e., potential biomarkers) which can then be targeted for later identification. As mentioned previous, the eventual identification of the detected features of interest is an integral component of our overall strategy and we are currently investigating multiple methods for a broad based approach.[75] Figure 3 shows a 2D visualization of the targeted MS/MS approach for the identification of previously detected peptide pairs of interest. This approach is similar to the previously discussed SELDI method, but instead we are using the high efficiency LC separations and high resolution and mass accuracy LC−FTICR platform for the more accurate detection (and identification) of features of interest over the much more complex 2-D separation space. So far, both the bioinformatics and MS platforms developed for the AMT tag approach in our laboratory has shown to be well suited for the implementation of this type of analysis.

The focal point of the remaining discussion will be based upon the commonly used "bottom-up" approach and the application of various detection strategies for LC−MS based detection and quantitative analyses. Upstream processing of plasma samples can generally be divided into two main categories, depletion techniques for the reduction of specific high abundant proteins and enrichment techniques for the selective sequestering of proteins of interest. Each method contains both advantages and disadvantageous and what follows is a discussion of a few specific strategies representative of both methods which are now being commonly used by researchers and/or that we have found useful in our applications.

B. Major Protein Depletion Strategies. In a perfect world the best protein depletion option for human plasma would be to remove the top 50−100 most abundant proteins (the classic plasma proteins of little interest to the biomarker researcher) to greatly extend the dynamic range of proteome detection deep into the pg/mL range of plasma proteins for greater coverage and diversity of protein identifications. This of course is not currently practical, but the intent has been well taken by others as there are a number of basic techniques that have been used to eliminate the high abundance proteins in plasma. The two largest contributors to the dynamic range challenge of plasma are human serum albumin (HSA) at ∼50% of the total protein mass, and the immunoglobulin proteins which represent 20−25% of the total protein mass. Each has relatively high molecular weights, 64 kDa and ∼150 kDa respectively, which have spawned methods to help selectively target them. Organic solvent precipitation (methanol, acetonitrile), affinity dye based depletion (Cibacron Blue), and simple ultra-filtration are commonly used general methods to either remove whole proteins entirely or remove the higher molecular weight component. Often, a series of increasingly concentrated acetonitrile fractions or "cuts" are taken to selectively precipitate out the higher molecular protein fraction. Such techniques are effective but act in a nonspecific manner and lack in reproducibility, especially concerning the quantitation of proteins remaining in solution. The introduction of affinity chromatography for depletion purposes now allows for the selective

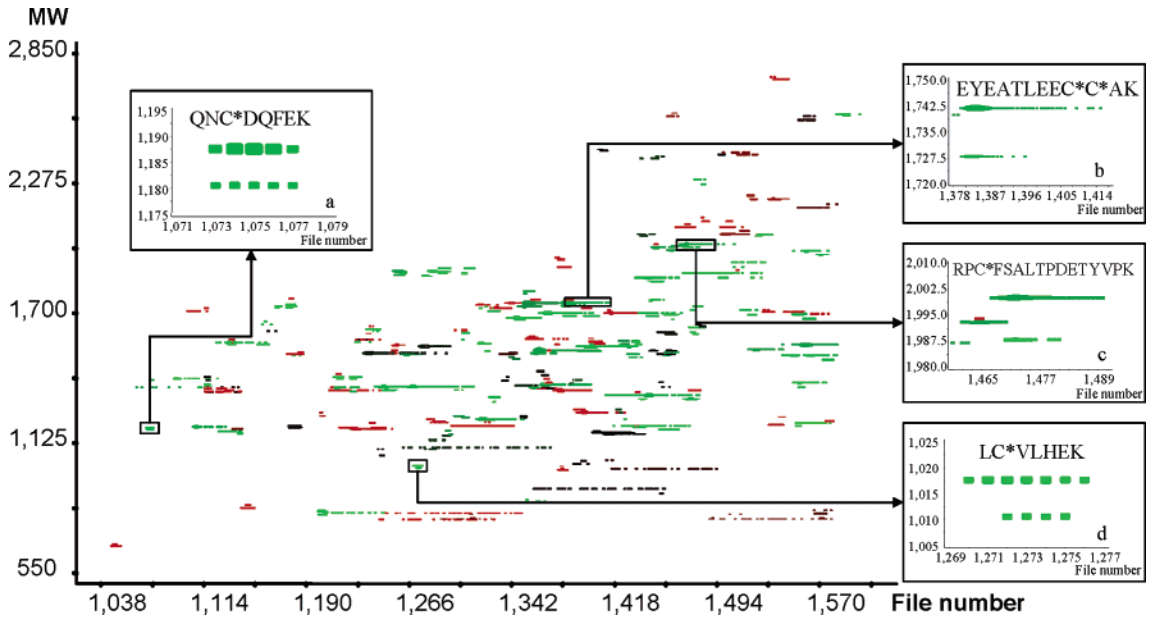

Figure 3. 2D-display of LC−FTICR analysis of a standard protein mixture labeled with solid phase-ICAT reagent to produce heavy and light labeled pairs. Among the 209 detected pairs of these dataset, 46 pairs were targeted for MS/MS analysis. Insets (a−d) show examples of targeted pairs and the identifications obtained during the subsequent MS/MS analysis (asterisks represent ICAT labels). Axes correspond to the elution time (spectrum number) and molecular weight of the detected feature.[75]

removal of virtually any protein of interest (although with imperfect efficiency) and when coupled to a high pressure liquid chromatography (HPLC) system, can provide a reasonably reproducible platform for repeated quantitative analysis.

One concern for depletion methodologies is that for every sample manipulation, there is a small but accumulative rate of error which is propagated into the analysis. This principle hold true for each step in the sample processing pipeline, especially for any depletion process. The hope is that any introduced error will be minimal in affecting experimental accuracy in comparison to the benefits gained from increasing the number of overall identifications. A second concern of the process is the loss of potential proteins of interest, whether due to selective or nonselective binding to the depleted protein or column surface, during the depletion method. For example, HSA is a known nonspecific binder as part of its biological role is to act as a molecular "sponge" of sorts to assist in the circulation and distribution of other proteins.

Depletion of the immunoglobulin class of proteins (IgG, IgA, IgM, IgE, etc..) is actually quite straightforward as researchers for some time have been extensively targeting this group of proteins to develop multiple antibodies (i.e., Goat α-human IgG) and binding proteins (Concanavalin A, Protein G, Protein A, Protein L) for the selective binding, and eluting, of the various Ig classes. Though not as abundant as HSA, the removal of Ig proteins accomplishes not only a reduction of total protein content, but also greatly reduces the overall complexity and potential random background sequence generation that occurs once the CDR of the immunoglobulin has been reduced to peptide form.

We have utilized both HSA depletion and IgG depletion methods in attempts to increase the potential dynamic range of detection for human plasma proteomics.[17] Reducing both species effectively removes ∼75% of the total protein content of the sample, but still leaves some higher abundant species in the 2−3 mg/mL range. In our studies, even after depletion, we were still able to detect a significant number of peptides from both HSA and Ig proteins from the "depleted" sample. The columns tested in our laboratory (Poros@ anti-HSA and protein G affinity depletion cartridges, Applied Biosystems) both claimed to have > 99% depletion of the specific component. This initially appears adequate for a depletion study, but when compared in depth, 99% depletion for HSA at ∼45 mg/mL would correspond to on average 450 ug/mL of the protein remaining in the sample, still maintaining HSA near the top of the most abundant proteins present in plasma. This is not even taking under consideration likely fragmentation and partial protein components, which are proportionally present for every protein, but for albumin would constitute a significant percentage of the total protein content. Partial protein retention is also very likely for depletion of immunoglobulin proteins as well, especially when using Protein A/G depletion, which binds to a specific site on the Fc portion of the immunoglobulin, and would not recognize or bind to any separate light chain, fragment or portion that does not contain the binding site.

Also notably observed when analyzing the retentate from each depletion was a significant number of non-HSA and non-Ig proteins retained on the column, a major concern of depletion protocols. Most of these extra identifications corresponded to higher abundance proteins, which suggests that the nonspecific interactions occurring during depletion (either with the column itself or with the retained HSA and IgG) are concentration dependent. However, this supposition is difficult to prove because potentially retained lower abundant species would be difficult to detect in the enriched presence of HSA and immunoglobulins. Regardless, by using such depletion techniques, we were able to increase our overall protein coverage of human plasma compared to the results using only nondepleted samples.

The obvious next step or progression is not only to remove the top two most abundant proteins but to attempt a simultaneous and reproducible removal of all of the highest abundant proteins. By combining various affinity materials into one depletion column, the depletion of multiple components can

be accomplished in one step, simplifying the sample preparation and potentially decreasing errors in reproducibility. Such columns are now commercially available from multiple suppliers; MIXED12 column from GenWay and the MARS column from Agilent are two which have been used in our laboratory (Liu et al., manuscript in preparation). When using such columns, we have been able to identify a number of new proteins not previously identified using only HSA or Ig depletion methods and have observed similar results of reproducibility and depletion levels as previously published results.[76] Depletion results dealing with specific human plasma studies have revealed numerous secreted and known lower abundance proteins which have been very beneficial in ascertaining the approximate increase in dynamic range of our detections (Liu et al., manuscript in preparation). Utilizing plasma samples has been very beneficial to use for such studies due to the large number of proteins with known plasma concentrations, making it very straightforward to infer, based upon the identification of these known proteins, the dynamic range of the detection methods.

Beyond what has already been discussed, there still remain some technical challenges associated with such multiple protein depletion strategies. First, most have very limited loading capacities, 15−25 μL volume of total plasma for the small columns and 30−40 μL for the larger columns, and since only 5−10% protein amount is expected to be recovered, multiple depletions steps for one sample are often required to generate enough sample for the following analysis. For a depletion of 1 mL of plasma, 20−40 separate HPLC runs need to be performed for each sample. Considering the recommended lifetime of each column is ∼100−200 runs, this can be very limiting. Add into this the cost of each column, both small and large, can range between sixteen hundred to several thousand dollars each, this method, especially if you are conducting in-depth studies with several plasma samples, can greatly increase the overall baseline cost of analysis. Despite this, it is becoming clear that such experiments are necessary in certain proteomic applications for the enhancement of detection of low abundance proteins. In our laboratory, we have shown that such depletion techniques can be beneficial in reducing the dynamic range issue with some samples, and that our preliminary results suggest that the increase in peptide identifications and coverage is a reasonable tradeoff in most model systems to the errors in reproducibility introduced through the depletion step.

C. Enrichment Strategies at the Peptide Level. An alternate method for improving the detection of lower abundance proteins of interest is to target a subset of proteins using some method for their enrichment. There are two main approaches which have been taken by us that merit further discussion.

One method is to specifically enrich at the peptide level based upon the use of a specific amino acid residue found in only a fraction of the digested peptides. Commonly a cysteine residue is targeted, but other amino acids have been found to be effectively enriched.[77−79] By selectively targeting a peptide-subset, the overall complexity of the sample is reduced to allow the potential detection of novel and lower abundance species. Often such enrichment techniques contain an element of stable isotopic labeling for quantitative measurements, i.e., ICAT and the more recent QCET approach.[72,80] The basis of the approach relies on the specificity of the capture step to highly enrich for the amino acid containing peptide/or protein. These approaches for application with human plasma present certain challenges considering that HSA contains 35 cysteine residues itself; however, by coupling an effective depletion method prior to the peptide enrichment step, one can gain both a reduction of the dynamic range from the depletion step and a reduction of the complexity of the sample with the enrichment. In the past, we have observed a large benefit to reducing the overall complexity of the sample using cysteine enrichment techniques for better MS peak detection, identification, and quantitation efforts. Recently, we have performed such a study in our laboratory by applying a multicomponent protein depletion strategy followed by a cysteinyl-peptide enrichment of the digested peptides coupled with an off-line standard SCX−RPLC−MS/MS shotgun peptide identification approach. Preliminary results indicate that an increase of 1−2 orders of dynamic range was observed with reproducible and consistent identification of proteins found in the low ng/mL to high pg/mL range.

Other peptide based enrichment techniques are now commonly used in the context of proteomic studies for enrichment purposes. Most are beyond the scope of this work to extensively discuss, but briefly. A more commonly pursued enrichment methodology is to target phosphorylated peptides in both targeted and global approaches to elucidate the phosphorylation sites of proteins to, for example, uncover novel signaling pathways and mechanisms. Selective separation columns, immunoaffinity approaches, and bioinformatic applications are all areas actively used in attempts to identify these peptides of interest.[81−83] Additionally, approaches have been developed for the specific N-terminal labeling at the protein level, to theoretically isolate only a single peptide from each protein, reducing the complexity and hence increasing protein detection coverage.[84] Such an approach could be beneficial in plasma analysis as many proteins have altered N-terminal cleavage sites (signal peptide cleavage, alternate splicing events) and the identification of these sites can be a great value-added approach combined with the reduction in complexity of the sample.

Another novel approach, first demonstrated by Zhang et al. is to specifically target the glycoprotein sub-proteome present in plasma.[25] This approach follows the rational that often the lower abundance proteins of interest in plasma originate from secretion or outer membrane shedding events which are often heavily glycosylated. By targeting these proteins, a potentially informative fraction of human plasma can be enriched, helping to overcome the dynamic range of detection and increasing the chances to detect interesting proteins. The methodology has been demonstrated, and involves the capture of the N-linked glycosylation moiety on the protein using hydrazide chemistry. This is followed by peptide digestion and release of the remaining glycopeptides using a deglycosylase. The obtained peptides are then analyzed via MS methods and quantitated using peak intensity values with stable isotope labeling methods. The initial methodology utilized an N-terminal sequence labeling approach for quantitation but other stable isotopic approaches such as $^{16}O/^{18}O$ labeling are also amenable to this process. A further application of this approach has been demonstrated in a comparative study using LC−MS analysis and a mouse skin carcinoma model system.[85] The selective isolation of only N-linked glycopeptides was shown to reduce the complexity of the sample, and increase the sensitivity toward these select peptides allowing unambiguous discrimination between the sera of the control and skin cancer mouse model.

One caveat to using the N-linked glycosylation capture approach is that by specifically targeting only the glycosylated extracellular sub-proteome, there is the possibility of limiting the detection of other biologically pertinent proteins. This obviously depends on the model system or paradigm of interest and if there is a need to only focus on secreted/shed cellular-proteins or a requirement for a more diverse characterization. When observing the general global identification of plasma proteins (see Figure 6), the majority of new identifications originate specifically from the intracellular region, alluding that there is a significant presence of nonglycosylated proteins in circulation, probably due to multiple mechanisms, which should not be excluded when attempting to detect and identify potential biomarkers. This would be especially critical when searching for potential outcome prognosticators due to injury, trauma, microbial infections, or any condition which would induce a disruption of cellular components into the blood stream.

D. Quantitative Strategies. To effectively identify candidate biomarkers, it is useful to precisely or accurately quantitate their abundance values, and to ideally accomplish this across a large series of measurements. Recent developments, coupled with more established techniques have resulted in the development of many different methods to assist in the quantitation of MS data. We have generally utilized two approaches depending upon sample type, the data generated, and the information desired. A "semi"-quantitative technique being used with proteomic MS/MS peptide identification data is the comparative analysis of the total number of peptide identifications between samples to obtain some relative quantitative information. This method is becoming more commonly used[86,87] and we have found it to be reasonably helpful when applied in the proper setting and with the adequate thresholds.[39,65] A recent example is the initial comparative proteome analysis of human plasma from lipopolysaccharide (LPS) administration. In this study, standard SCX−LC−MS/MS analysis was preformed for both a 0 timepoint and 9 h timepoint after LPS administration. The number of MS/MS analysis and sample amounts was carefully controlled so the results could be compared relatively. Both peptide peak areas from the MS/MS ion-trap analysis and the total number of peptide identifications per protein were used to obtain a relative quantitative measure between the control and 9 h time points. Most known inflammatory response and acute phase proteins found at the mid-abundance range were observed to be up-regulated upon LPS administration including other proteins of interest. A general correlation was also seen when comparing the total peptide identifications per protein with the absolute abundance values of the identified proteins in plasma (see Figure 4).[65]

An alternative approach for LC−MS quantitation is to use the actual peak intensity information to determine the relative abundance of species in the biological sample. In doing so, there are a number of issues which can affect the intensity value of any detected peptide, of which include electrospray ionization efficiency, separations, and peptide digestion efficiencies to name the most dominant factors.[43] These factors are peptide dependent, which often make it challenging to surmise an abundance for protein when there is disparity in the peptide values. Strategies to avoid or evade these issues include performing the pre-MS separations in a lower flow (nano-flow) regime which alleviates potential electrospray ion suppression effects and by using smaller i.d. capillary columns that will also improve sensitivity and separation efficiencies.[43,68] Also, the

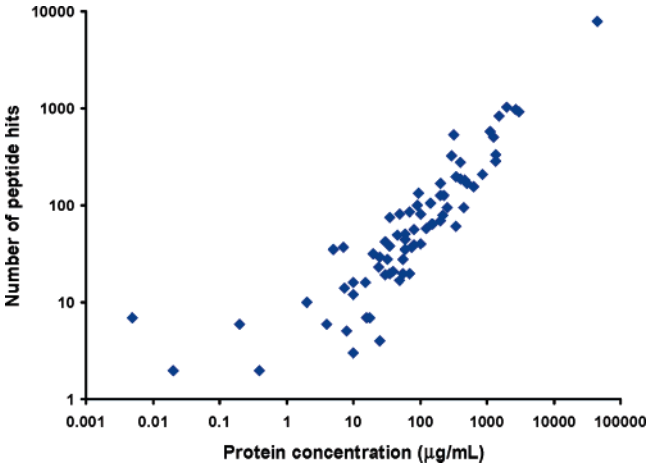

Figure 4. Correlation between peptide hits for each protein from Qian et al. with a known plasma concentration and the known concentrations obtained from the literature.[65]

utilization of stable isotopic labeling methods allows the direct peak intensity comparison of both control and test samples concurrently at the peptide level, allowing for a comparison that can more effectively be "rolled-up" to a protein abundance ratio (with the appropriate caveats). Recent reports involving this methodology (e.g., using ^{18}O, ^{15}N labeling) have demonstrated the incorporation of the label virtually anywhere through the stages of sample preparation. For application with human plasma though, ^{15}N metabolic labeling is generally precluded, but interestingly there are reports of ^{15}N labeling of whole organisms (*Caenorhabditis elegans, Drosophila melanogaster,*[88] and *Rattus norvegicus*[89]). Such studies have demonstrated that these methods are feasible, but in regards to mammalian systems it is not likely that this will become widely used due to the sheer expense. Regardless, there are other stable isotopic labeling methods which are effective in determining peptide abundance information in proteomic studies.

An approach commonly utilized in our laboratory is to enzymatically incorporate ^{18}O atoms into the C-terminus of tryptically cleaved peptides for comparative quantitative analysis. By using a previously reported post-digestion strategy which drives the incorporation of 2 ^{18}O atoms into the carboxy terminus,[90,91] a 4 Da mass shift is observed, allowing an appropriate separation between the isotopic distributions of the paired peaks and the accurate quantitation of their respective intensities (see Figure 5). A recent application of ^{18}O labeling methodology is the previously described LPS administration study.[66] By using the quantitative ^{18}O labeling in conjunction with LC−FTICR−MS analysis, the obtained results correlated well with the previous "semi"quantitative study,[65] validating our previous methods as well as providing a more accurate quantitation of the observed up-regulated proteins.

A further development involving the ^{18}O labeling technology was to couple it to a cysteine enrichment approach, as previously discussed in the enrichment section. Quantitative Cysteinyl-peptide Enrichment Technology (QCET)[72] is comparable to the popular ICAT technology except the quantitative labeling is performed by ^{18}O incorporation and no proprietary reagents are needed for the process. We are now routinely using the QCET approach in application for numerous samples. Since the specificity of cysteine enrichment is very high (>95%), and the efficiency of the enzymatic ^{18}O labeling is also very high

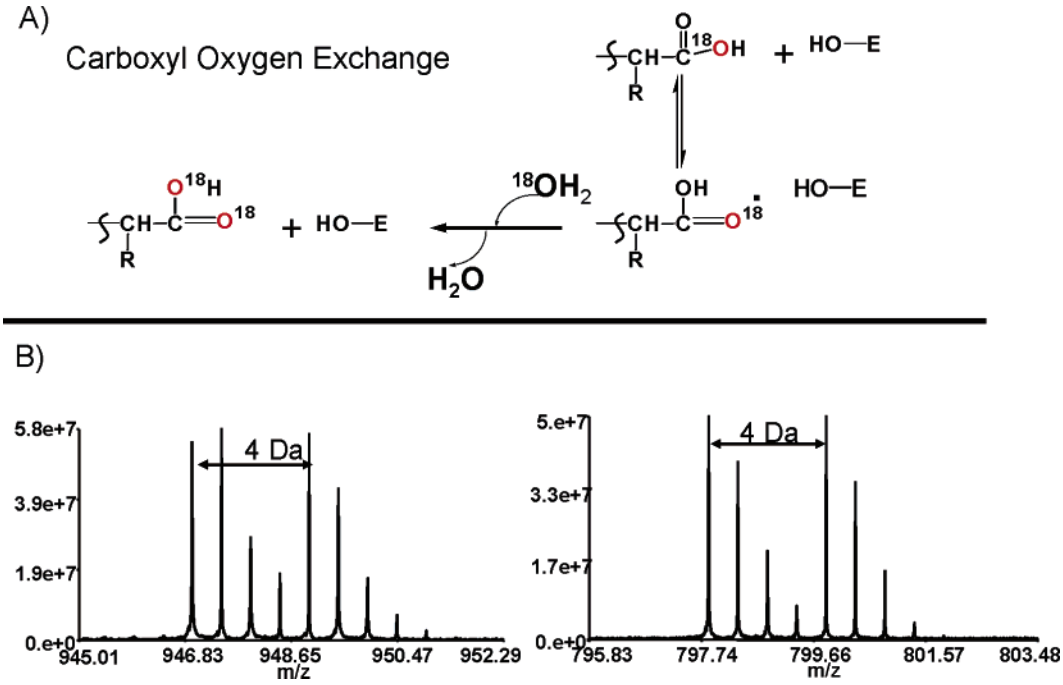

Figure 5. Methodology behind the post-digestion ^{18}O labeling for the incorporation of two ^{18}O atoms into each peptide for a consistent and total mass difference of 4 Da. (A) Description of the enzymatically driven exchange reaction after previous tryptic cleavage which provides the second incorporation of the ^{18}O molecule. (B) Isotopic chromatographic representation of a 1:1 distribution for $^{16}O/^{18}O$ quantitation providing a 4 Da mass shift.

compared to biochemical methods, we have found it to be beneficial for decreasing the overall complexity of the sample and increasing the confidence for peptide identification when utilizing LC−FTICR−MS analysis. Also, the sample preparation methodology allows the collection of the remaining noncysteine containing peptides which provides an appropriate "global" sample for analysis that directly complements the cysteine enriched fraction results.

Global Human Plasma Characterization. We have been generating a growing AMT tag database specifically for human plasma samples by combining all the identification data generated from among the extensive separations and enrichment methods previously discussed. This database is a significant resource as it allows the identification of peptides using the AMT tag approach from high throughput accurate mass and elution time LC−MS measurements. In the creation of this database, we have utilized different MS/MS proteomic data search criteria, but have observed a wide range of results when using the various reported search criteria from the literature, especially in application with mammalian samples and very large mammalian protein sequence databases. Considering that very little baseline work had been performed in describing the potential false positive identifications using such large databases, it was very likely such results contain a higher number of spurious results than anticipated. To address this problem for our future studies and to increase the confidence of our data, we began to utilize a full human reverse-database approach for the characterization of a measurable false positive rate for these identifications to alleviate possible misidentifications in the final results.[29] Incorporating these criteria provides a dramatic increase in the data quality, which is reflected in improvements to the reproducibility and quantitative results from the MS-based accurate mass and time measurements.

In looking at the database as a whole, and by applying these criteria, a total of >10 000 peptide mass and time tags have been identified with high confidence in plasma, resulting in the identification of 1427 proteins. This is directly comparable to Qian et al.[66] description of the plasma database (938 proteins) with the addition of the multiple analysis of the combined depletion and N-linked/cysteine enrichment approaches described here (Liu et al. manuscript in preparation). Figure 6 shows the breakdown distribution of the identified proteins revealing that a majority of the proteins still originate from the intracellular compartments as previously discussed. Due to the detection of proteins with known plasma concentrations, we believe that the limit to the dynamic range of detection is $\sim 10^8$ (500−200 pg/mL), and includes most known proteins at $\leq 10^7$ dynamic range (~1−5 ng/mL) that are consistently identified by our analyses. As we have discussed, this level is really only the tip of a rather large iceberg, with future gains in protein identifications likely coming only from lower abundant species after they have been enriched by one or more of the methods discussed earlier.

Conclusions and Perspective. Significant effort is now being placed into the proteomic detection of biologically relevant biomarkers, with human plasma now being viewed as a potential treasure-trove of candidate biomarkers. However, there are significant challenges at present which preclude the easy application of this biofluid for candidate biomarker discovery. The sheer dynamic range of proteins present coupled with the need to detect the lower end of this dynamic range for proteins of interest calls for better separation/enrichment approaches and analysis tools to effectively deal with these challenges.

Also evident is the statistical analysis needed to validate any particular result and the need for multiple replicates i.e., biological, sample preparation, instrumental, to address errors

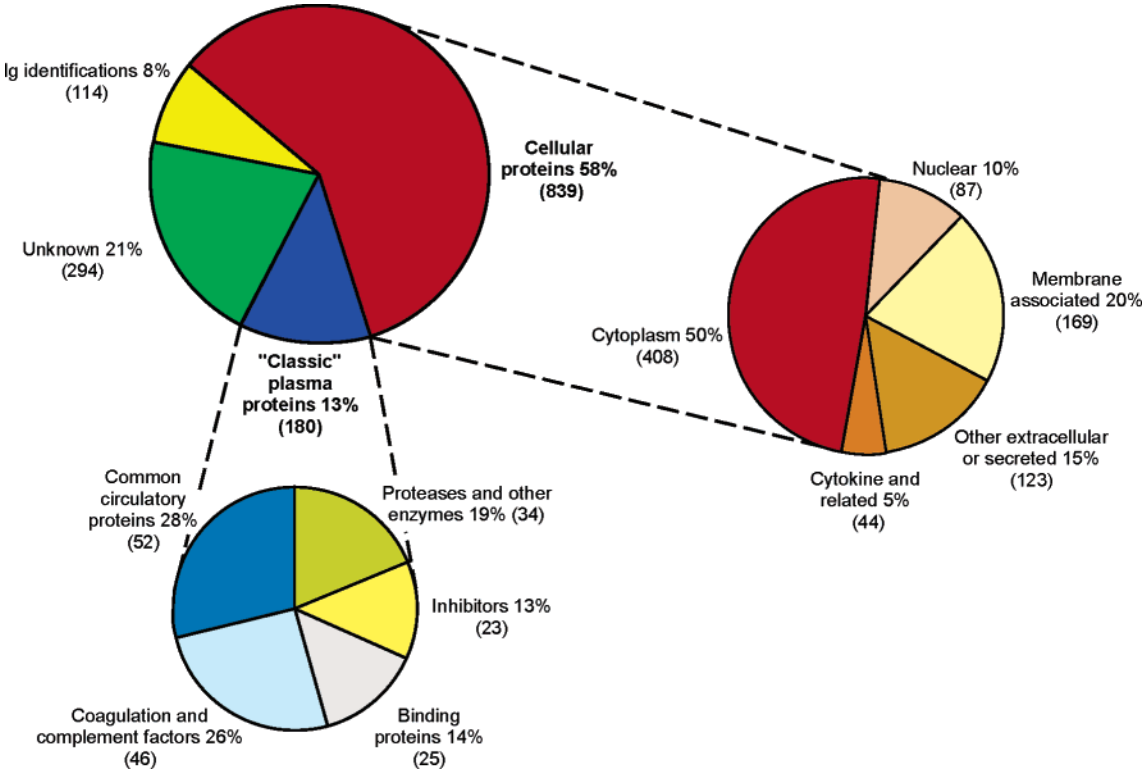

Figure 6. Categorization of 1427 high confidence plasma proteins that comprise the mass and time tag database and have passed 95% confidence limit filtering criteria. This total is based upon multiple analyses involving both depleted and nondepleted plasma samples.

or biases which could be introduced throughout the analysis pipeline. In addition, it is important that QA/QC measures be developed and applied throughout the sample collection, handling and analysis pipeline as well to ensure the quality and reproducibility of the efforts. It is common practice to regularly monitor the performance of the downstream instrumentation for high efficiency LC−MS operations, but often the upstream components of the analysis (i.e., biological sample selection, collection, handling, sample preparation), which are likely more important to the overall quality of the results, are not as effectively or reproducibly controlled, and often inadequately documented.

If extensive separations and enrichment strategies are needed to reproducibly identify those proteins of interest, a large amount of effort is often needed, both manually and instrumentally, for each biological replicate. This often precludes the utilization of the multiple (on the order of hundreds or thousands) biological replicates possibly needed to support biomarker development and validation. There is a tangible tradeoff between the number of biological samples that can be included in a proteomic study and the amount of effort each sample requires for both instrument analysis and sample preparation. This tradeoff can be addressed by either simplifying or automating the efforts needed to analyze each sample, or by increasing the number of samples that can be addressed within a certain time frame (increasing the throughput of the pipeline).

At present levels of throughput, it is not practical to study plasma samples from several hundred patients if relying on extensive pre-MS fractionation methods, and thus greatly increasing the number of samples, if needed to cover a sufficient fraction of the proteome. We experienced this in our initial clinical proteomic studies, encountering the need for increasing both the throughput of sample preparation by using automated instrumentation as well as speeding up the total analysis time per sample by using faster LC separations. In this regard, there appears to be significant potential in the use of new fast separation approaches coupled to MS analysis, e.g., using ion mobility separations.[92,93] By taking an approach to allow increasing numbers of analytical and/or biological replicates, and providing significant increases in throughput, the potential exists for avoiding the pitfalls associated with complexity and chance, and the development of high quality sets of candidate biomarkers which have eluded proteomic researchers to date.

There remains much to be accomplished to fully utilize plasma proteomics for biomarker discovery, but the increasingly clear understanding of the challenges that lay ahead will likely stimulate the development of novel methodologies and better refinement of existing technologies. These in turn will help to better utilize plasma and demonstrate that proteomic approaches can eventually help elucidate clinical biomarkers of interest.

Acknowledgment. Portions of this research were supported by the National Institute of General Medical Sciences (NIGMS, Large Scale Collaborative Research Grants U54 GM-62119-02), the NIH National Center for Research Resources (RR18522), and the Environmental Molecular Science Laboratory (a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory). Pacific Northwest National Laboratory is operated by Battelle

Memorial Institute for the U.S. Department of Energy under contract DE-AC05-76RLO-1830.

References

(1) Anderson, N. L.; Anderson, N. G. *Mol. Cell Proteomics* **2002**, *1*, 845−867.
(2) Lipton, M. S.; Pasa-Tolic, L.; Anderson, G. A.; Anderson, D. J.; Auberry, D. L.; Battista, J. R.; Daly, M. J.; Fredrickson, J.; Hixson, K. K.; Kostandarithes, H.; Masselon, C.; Markillie, L. M.; Moore, R.; Romine, M. F.; Shen, Y.; Strittmatter, E.; Tolic, N.; Udseth, H. R.; Venkateswaran, A.; Wong, K. K.; Zhao, R.; Smith, R. D. *Proc. Natl. Acad. Sci., U. S. A.* **2002**, *99*, 11, 049−011, 054.
(3) Washburn, M. P.; Wolters, D.; Yates, J. R. *Nat. Biotechnol.* **2001**, *19*, 242−247.
(4) Liu, T.; Qian, W.-J.; Chen, W.-N. U.; Jacobs, J. M.; Moore, R. J.; Anderson, D. J.; Gritsenko, M. A.; Monroe, M. E.; Thrall, B. D.; David G. Camp, I.; Smith, R. D. *Proteomics* **2005**, *5*, 1263−1273.
(5) Anderson, L.; Anderson, N. G. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 5421−5425.
(6) 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*, 1345−1364.
(7) 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.
(8) 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. *Mol. Cell. Proteomics* **2004**, *3*, 311−316.
(9) Shen, Y.; Jacobs, J. M.; Camp, D. G.; Fang, R.; Moore, R. J.; Smith, R. D.; Xiao, W.; Davis, R. W.; Tompkins, R. G. *Anal. Chem.* **2004**, *76*, 1134−1144.
(10) Rose, K.; Bougueleret, L.; Baussant, T.; Bohm, G.; Botti, P.; Colinge, J.; Cusin, I.; Gaertner, H.; Gleizes, A.; Heller, M.; Jimenez, S.; Johnson, A.; Kussmann, M.; Menin, L.; Menzel, C.; Ranno, F.; Rodriguez-Tome, P.; Rogers, J.; Saudrais, C.; Villian, M.; Wetmore, D.; Bairoch, A.; Hochstrasser, D. *Proteomics* **2004**, *4*, 2125−2150.
(11) Johnson, K. L.; Mason, C. J.; Muddiman, D. C.; Eckel, J. E. *Anal. Chem.* **2004**, *76*, 5097−5103.
(12) 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.
(13) Harper, R. G.; Workman, S. R.; Schuetzner, S.; Timperman, A. T.; Sutton, J. N. *Electrophoresis* **2004**, *25*, 1299−1306.
(14) Zhou, M.; Lucas, D. A.; Chan, K. C.; Issaq, H. J.; Petricoin, E.; Liotta, L. A.; Veenstra, T. D.; Conrads, T. P. *Electrophoresis* **2004**, *25*, 1289−1298.
(15) Liotta, L. A.; Ferrari, M.; Petricoin, E. *Nature* **2003**, *425*, 905.
(16) Mehta, A. I.; Ross, S.; Lowenthal, M. S.; Fusaro, V.; Fishman, D. A.; Petricoin, E. F.; Liotta, L. A. *Dis. Markers* **2003**, *19*, 1−10.
(17) Shen, Y.; Kim, J.; Strittmatter, E. F.; Jacobs, J. M.; Camp, D. G.; Fang, R.; Tolic, N.; Moore, R. J.; Smith, R. D. *Proteomics* **2005**, in press.
(18) Ahmed, N.; Barker, G.; Oliva, K.; Garfin, D.; Talmadge, K.; Georgiou, H.; Quinn, M.; Rice, G. *Proteomics* **2003**, *3*, 1980−1987.
(19) Pieper, R.; Su, Q.; Gatlin, C. L.; Huang, S. T.; Anderson, N. L.; Steiner, S. *Proteomics* **2003**, *3*, 422−432.
(20) Steel, L. F.; Trotter, M. G.; Najkajima, P. B.; Mattu, T. S.; Gonye, G.; Block, T. *Mol. Cell. Proteomics* **2003**, *2*, 262−270.
(21) Greenough, C.; Jenkins, R. E.; Kitteringham, N. R.; Pirmohamed, M.; Park, B. K.; Pennington, S. R. *Proteomics* **2004**, *4*, 3107−3111.
(22) Marshall, J.; Jankowski, A.; Furesz, S.; Kireeva, I.; Barker, L.; Dombrovsky, M.; Zhu, W.; Jacks, K.; Ingratta, L.; Bruin, J.; Kristensen, E.; Zhang, R.; Stanton, E.; Takahashi, M.; Jackowski, G. *J. Proteome Res.* **2004**, *3*, 364−382.
(23) Fujii, K.; Nakano, T.; Kawamura, T.; Usui, F.; Bando, Y.; Wang, R.; Nishimura, T. *J. Proteome Res.* **2004**, *3*, 712−718.
(24) Morris, D. L.; Sutton, J. N.; Harper, R. G.; Timperman, A. T. *J. Proteome Res.* **2004**, *3*, 1149−1154.
(25) Zhang, H.; Li, X.-j.; Martin, D. B.; Aerbersold, R. *Nat. Biotechnol.* **2003**, *21*, 660−666.
(26) Richter, R.; Schulz-Knappe, P.; Schrader, M.; Standker, L.; Jurgens, M.; Tammen, H.; Forssmann, W. G. *J. Chromatogr. B Biomed. Sci. Appl.* **1999**, *726*, 25−35.
(27) Ratner, D. M.; Adams, E. W.; Disney, M. D.; Seeberger, P. H. *ChemBioChem* **2004**, *5*, 1375−1383.
(28) Holemann, A.; Seeberger, P. H. *Curr. Opin. Biotechnol.* **2004**, *15*, 615−622.
(29) Qian, W.-J.; Liu, T.; Monroe, M. E.; Strittmatter, E. F.; Jacobs, J. M.; Kangas, L. J.; Petritis, K.; Camp II, D. G.; Smith, R. D. *J. Proteome Res.* **2005**, *4*, 53−62.
(30) Peng, J.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P. *J. Proteome Res.* **2003**, *2*, 43−50.
(31) Cargile, B. J.; Bundy, J. L.; Stephenson, J. L. *J. Proteome Res.* **2004**, *3*, 1082−1085.
(32) Lopez-Ferrer, D.; Martinez-Bartolome, S.; Villar, M.; Campillos, M.; Martin-Maroto, F.; Vazquez, J. *Anal. Chem.* **2004**, *76*, 6853−6860.
(33) Omenn, G. S. *Proteomics* **2004**, *4*, 1235−1240.
(34) 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. *Mol. Cell. Proteomics* **2004**, *3*, 311−326.
(35) Omenn, G. S. *Dis. Markers* **2004**, *20*, 131−134.
(36) Adkins, J. N.; Monroe, M. E.; Auberry, K. J.; Shen, Y.; Jacobs, J. M.; Camp, D. J. I.; Vitzthum, F.; Rodland, K. D.; Smith, R. D.; Pounds, J. G. *Proteomics* **2005**, in review.
(37) Jacobs, J. M.; Mottaz, H. M.; Yu, L. R.; Anderson, D. J.; Moore, R. J.; Chen, W. U.; Auberry, K. J.; Strittmatter, E. F.; Monroe, M. E.; Thrall, B. D.; Camp II, D. G.; Smith, R. D. *J. Proteome Res.* **2004**, *3*, 68−75.
(38) 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.
(39) Jacobs, J. M.; Diamond, D. L.; Chan, E. Y.; Gritsenko, M. A.; Qian, W. J.; Stastna, M.; Baas, T.; Camp, D. G.; Carithers, R. J.; Smith, R. D.; Katze, M. G. *J. Virol.* **2005**, in press.
(40) Martin, S. E.; Shabanowitz, J.; Hunt, D. F.; Marto, J. A. *Anal. Chem.* **2000**, *72*, 4266−4274.
(41) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1−35.
(42) Belov, M. E.; Gorshkov, M. V.; Udseth, H. R.; Anderson, G. A.; Smith, R. D. *Anal. Chem.* **2000**, *72*, 2271−2279.
(43) Smith, R. D.; Shen, Y.; Tang, K. *Acc. Chem. Res.* **2004**, *37*, 269−278.
(44) Haab, B. A.; Geierstanger, B. H.; Michailidis, G.; Vitzthum, F.; Forrester, S.; Okon, R.; Saviranta, P.; Brinker, A.; Sorette, M.; Perlee, L.; Suresh, S.; Drwal, G.; Adkins, J. N.; Omenn, G. S. *Proteomics* **2005**, in press.
(45) Ridker, P. M.; Hennekens, C. H.; Buring, J. E.; Rifai, N. *N. Engl. J. Med.* **2000**, *342*, 836−843.
(46) Van Amersfoort, E. S.; Van Berkel, T. J.; Kuiper, J. *Clin. Microbiol. Rev.* **2003**, *16*, 379−414.
(47) Dinarello, C. A. In *Cecil Textbook of Medicine*, 20th ed.; Bennett, J. C., Plum, F., Eds.; Saunders Co: Philadelphia, 1996; pp 1535−1537.
(48) Ginsberg, H. N.; Goldberg, I. J. In *Harrison's Principles of Internal Medicine*, 14th ed.; Fauci, A. S., Braunwald, E., Isselbacher, K., Eds.; McGraw-Hill Co.: New York, 1998; pp 2138−2148.
(49) Hillman, R. S. In *Harrison's Principles of Internal Medicine*, 14th ed.; Fauci, A. S., Braunwald, E., Isselbacher, K., Eds.; McGraw-Hill Co: New York, 1998; pp 638−645.
(50) Cloonan, M. J.; Bishop, G. A.; Wilton-Smith, P. D.; Carter, I. W.; Allan, R. M.; Wilcken, D. E. *Pathology* **1979**, *11*, 689−699.
(51) Schweitzer, B.; Roberts, S.; Grimwade, B.; Shao, W.; Wang, M.; Fu, Q.; Shu, Q.; Laroche, I.; Zhou, Z.; Tchernev, V. T.; Christiansen, J.; Velleca, M.; Kingsmore, S. F. *Nat. Biotechnol.* **2002**, *20*, 359−365.
(52) Ransohoff, D. *Science* **2003**, *299*, 1679−1680.
(53) Ding, S. J.; Li, Y.; Tan, Y. X.; Jiang, M. R.; Liu, Y. K.; Shao, X. X.; Ye, S. L.; Wu, J. R.; Zeng, R.; Wang, H. Y.; Tang, Z. Y.; Xia, Q. C. *Mol. Cell. Proteomics* **2004**, *3*, 73−81.
(54) Borozdenkova, S.; Westbrook, J. A.; Patel, V.; Wait, R.; Bolad, I.; Burke, M. M.; Bell, A. D.; Banner, N. R.; Dunn, M. J.; Rose, M. L. *J. Proteome Res.* **2004**, *3*, 282−288.
(55) Liao, H.; Wu, J.; Kuhn, E.; Chin, W.; Chang, B.; Jones, M. D.; O'Neil, S.; Clauser, K. R.; Karl, J.; Hasler, F.; Roubenoff, R.; Zolg, W.; Guild, B. C. *Arthritis Rheum.* **2004**, *50*, 3792−3803.
(56) Drynda, S.; Ringel, B.; Kekow, M.; Kuhne, C.; Drynda, A.; Glocker, M. O.; Thiesen, H. J.; Kekow, J. *Pathol. Res. Pract.* **2004**, *200*, 165−171.
(57) Gravett, M. G.; Novy, M. J.; Rosenfeld, M. J.; Reddy, A. P.; Jacob, T.; Turner, M.; McCormack, A.; Lapidus, J. A.; Hitti, J.; Eschenbach, D. A.; Roberts, C. T.; Nagalla, S. R. *JAMA* **2004**, *292*, 462−469.
(58) Celis, J. E.; Gromov, P.; Cabezon, T.; Moreira, J. M.; Ambartsumian, N.; Sandelin, K.; Rank, F.; Gromova, I. *Mol. Cell. Proteomics* **2004**, *3*, 327−344.

(59) Petricoin, E. F.; Ardekani, A. M.; Hitt, B. A.; Levine, P. J.; Fusaro, V. A.; Steinberg, S. M.; Simone, C.; Fishman, D. A.; Kohn, E. C.; Liotta, L. A. *Lancet* **2002**, *359*, 572−577.
(60) Diamandis, E. P. *J. Natl. Cancer Inst.* **2004**, *96*, 353−356.
(61) Zhang, R.; Barker, L.; Pinchev, D.; Marshall, J.; Rasamoelisolo, M.; Smith, C.; Kupchak, P.; Kireeva, I.; Ingratta, L.; Jackowski, G. *Proteomics* **2004**, *4*, 244−256.
(62) Zhang, L. Y.; Ying, W. T.; Mao, Y. S.; He, H. Z.; Liu, Y.; Wang, H. X.; Liu, F.; Wang, K.; Zhang, D. C.; Wang, Y.; Wu, M.; Qian, X. H.; Zhao, X. H. *World J. Gastroenterol.* **2003**, *9*, 650−654.
(63) Steel, L. F.; Shumpert, D.; Trotter, M.; Seeholzer, S. H.; Evans, A. A.; London, W. T.; Dwek, R.; Block, T. M. *Proteomics* **2003**, *3*, 601−609.
(64) Kwak, J. Y.; Ma, T. Z.; Yoo, M. J.; Choi, B. H.; H. G., K.; Kim, S. R.; Yim, C. Y.; Kwak, Y. G. *Exp. Hematol.* **2004**, *32*, 836−842.
(65) Qian, W. J.; Jacobs, J. M.; Camp II, D. G.; 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*, 572−584.
(66) Qian, W. J.; Monroe, M. E.; Liu, T.; Jacobs, J. M.; Anderson, G. A.; Shen, Y.; Camp, D. G.; Moore, R. J.; Anderson, D. J.; Zhang, R.; Davis, R. W.; Tompkins, R. G.; Smith, R. D. *Mol. Cell. Proteomics* **2005**, in press.
(67) Mayya, V.; Rezaul, K.; Cong, Y. S.; Han, D. *Mol. Cell. Proteomics* **2005**, *4*, 214−223.
(68) Jacobs, J. M.; Monroe, M. E.; Qian, W. J.; Shen, Y.; Anderson, G. A.; Smith, R. D. *Int. J. Mass Spectrom.* **2005**, *240*, 195−212.
(69) Pasa-Tolic, L.; Harkewicz, R.; Anderson, G. A.; Tolic, N.; Shen, Y.; Zhao, R.; Thrall, B.; Masselon, C.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 954−963.
(70) Heller, M.; Michel, P. E.; Morier, P.; Crettaz, D.; Wenz, C.; Tissot, J. D.; Reymond, F.; Rossier, J. S. *Electrophoresis* **2005**, epub ahead of print.
(71) Wang, W.; Zhou, H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H. *Anal. Chem.* **2003**, *75*, 4818−4826.
(72) Liu, T.; Qian, W. J.; Strittmatter, E. F.; Camp, D. G.; Anderson, G. A.; Thrall, B. D.; Smith, R. D. *Anal. Chem.* **2004**, *76*, 5345−5353.
(73) Lee, S.-W.; Berger, S.; Martinovic, S.; Pasa-Tolic, L.; Anderson, G.; Shen, Y.; Zhao, R.; Smith, R. *Proc. Natl. Acad. Sci., U. S. A.* **2002**, *99*, 5942−5947.
(74) Bogdanov, B.; Smith, R. D. *Mass Spectrom. Rev.* **2005**, *24*, 168−200.
(75) Masselon, C. D.; Paša-Tolic', L.; Tolic', N.; Anderson, G. A.; Bogdanov, B.; Vilkov, A. N.; Shen, Y.; Zhao, R.; Qian, W.; Lipton, M. S.; Camp II, D. G.; Smith, R. D. *Anal. Chem.* **2005**, *77*, 400−406.
(76) Szafranski, C.; Bailey, J.; Christoph, W. T.; G., N.; Martosella, J.; Zolotarjova, N. *PharmaGenomics* **2004**, *September*, 40−46.
(77) Shen, M.; Guo, L.; Wallace, A.; Fitzner, J.; Eisenman, J.; Jacobson, E.; Johnson, R. S. *Mol. Cell. Proteomics* **2003**, *2*, 315−324.
(78) Weinberger, S. R.; Viner, R. I.; Ho, P. *Electrophoresis* **2002**, *23*, 3182−3192.
(79) Gevaert, K.; Van Damme, J.; Goethals, M.; Thomas, G. R.; Hoorelbeke, B.; Demol, H.; Martens, L.; Puype, M.; Staes, A.; Vandekerckhove, J. *Mol. Cell. Proteomics* **2002**, *1*, 896−903.
(80) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994−999.
(81) Zhou, H.; Watts, J. D.; Aebersold, R. *Nat. Biotechnol.* **2001**, *19*, 375−378.
(82) Oda, Y.; Nagasu, T.; Chait, B. *Nat. Biotechnol.* **2001**, *19*, 379−382.
(83) Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. *Nat. Biotechnol.* **2002**, *20*, 301−305.
(84) Gevaert, K.; Goethals, M.; Martens, L.; Van Damme, J.; Staes, A.; Thomas, G. R.; Vandekerckhove, J. *Nat Biotechnol* **2003**, *21*, 566−569.
(85) Zhang, H.; Yi, E. C.; Li, X. J.; Mallick, P.; Kelly-Spratt, K. S.; Masselon, C. D.; Camp, D. G.; Smith, R. D.; Kemp, C. J.; Aebersold, R. *Mol. Cell. Proteomics* **2005**, *4*, 144−155.
(86) Gao, J.; Opiteck, G. J.; Friedrichs, M. S.; Dongre, A. R.; Hefta, S. A. *J. Proteome Res.* **2003**, *2*, 643−649.
(87) Liu, H.; Sadygov, R. G.; Yates, J. R. *Anal. Chem.* **2004**, *76*, 4193−4201.
(88) Krijgsveld, J.; Ketting, R. F.; Mahmoudi, T.; Johansen, J.; Artal-Sanz, M.; Verrijzer, C. P.; Plasterk, R. H.; Heck, A. J. *Nat. Biotechnol.* **2003**, *21*, 927−931.
(89) Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Matthews, D. E.; Yates, J. R. *Anal. Chem.* **2004**, *76*, 4951−4959.
(90) Heller, M.; Mattou, H.; Menzel, C.; Yao, X. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 704−718.
(91) Yao, X.; Afonso, C.; Fenselau, C. *J. Proteome Res.* **2003**, *2*, 147−152.
(92) Tang, K.; Shvartsburg, A. A.; Lee, H.; Prior, D. C.; Buschbach, M. A.; Li, F.; Tolmachev, A.; Anderson, G. A.; Smith, R. D. *Anal. Chem.* **2005**, in press.
(93) Hoaglund-Hyzer, C. S.; Lee, Y. J.; Counterman, A. E.; Clemmer, D. E. *Anal. Chem.* **2002**, *74*, 992−1006.

PR0500657