Human Serum Proteins Fractionated by Preparative Partition Chromatography Prior to LC−ESI−MS/MS

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · APRIL 2009

Impact Factor: 4.25 · DOI: 10.1021/pr8005217 · Source: PubMed

7 AUTHORS, INCLUDING:

Monika Tucholska
Mount Sinai Hospital, Toronto
23 PUBLICATIONS **168** CITATIONS
SEE PROFILE

Peter Bowden
Ryerson University
22 PUBLICATIONS **154** CITATIONS
SEE PROFILE

Shirley Furesz
MAPI Group
6 PUBLICATIONS **306** CITATIONS
SEE PROFILE

John Gregory Marshall
Ryerson University
197 PUBLICATIONS **1,142** CITATIONS
SEE PROFILE

research articles Journal of proteome research

Human Serum Proteins Fractionated by Preparative Partition Chromatography Prior to LC-ESI-MS/MS

Monika Tucholska, Peter Bowden, Kellie Jacks, Peihong Zhu, Shirley Furesz, Mila Dumbrovsky, and John Marshall*

Department of Chemistry and Biology, Ryerson University, 350 Victoria Street, Toronto, Canada M5B 2K3

Received July 10, 2008

Many proteomics studies are limited to the identification of only the most abundant proteins in a sample due to the high sample complexity in most proteomes. We have here addressed this problem by prefractionation of human blood samples using microchromatography. We show that our approach resulted in high-stringency tryptic peptides identified by LC-ESI-MS/MS. Serum proteins were fractionated by batch and stepwise preparative chromatography using various types of chromatography resins (propyl sulfate, quaternary amine, diethylaminoethanol, cibachron blue, phenol Sepharose, carboxy methyl sepharose, hydroxyl apatite, heparin, concanavalin A and protein G) that were compared. The efficacy of sample fractionation was determined by protein assays, electrophoresis, and mass spectrometry. Tryptic peptides were separated by C18 liquid chromatography with electrospray ionization via metal needle at 2 μL/min with ion trap tandem mass spectrometry. The MS/MS spectra were correlated to some 4396 distinct sequences of the human forward RefSeq by X!TANDEM. Of these, 61% have been detected by other algorithms, but 3219 (73%) were never previously reported from blood by X!TANDEM. The use of a simple apparatus for making gravity microchromatography columns that permits the rapid side-by-side fractionation of many serum samples is described. Disposable microcolumns rapidly prepared blood samples for LC ESI-MS/MS that detected both tissue and cell leakage proteins known to exist in the $\sim$1 ng/mL range and some circulating receptor sequences. Our results demonstrate that the depletion of albumin or IgG was not necessary prior to LC-MS/MS and that multiple forms of protein chromatography will be useful for complete identification of blood proteins.

Keywords: Human serum • Preparative Chromatography • tryptic digest • LC-MS/MS • X!TANDEM • liquid chromatography

Introduction

The ability to detect proteins at low concentrations in blood is critical to the discovery of new biomarkers of disease. While mass spectrometry is an ideal tool to identify and quantify proteins in blood, abundant peptides may be detected with greater intensity and therefore mask less abundant proteins. Fractionating a sample using liquid chromatography or electrophoresis may help detect proteins present at low concentrations.[1−3] Liquid chromatography has been shown to be more sensitive than electrophoresis.[4] Overall, electrophoresis identifies fewer proteins from blood compared to LC-ESI-MS/MS;[5,14] separation by high-resolution 2D IEF/SDS-PAGE gels only identified about 300 proteins by MASCOT.[15] Many polypeptides[6] have been purified to near homogeneity from serum or plasma by sequential partition chromatography[7] using resins such as heparin,[8] DEAE,[9,10] quaternary amine,[11] propyl sulfate,[12] concanavalin A[13] or other resins. The use of multiple partition chromatography resins is a powerful approach for identifying blood proteins.[5]

Many methods have been used to separate proteins from blood samples with results that vary both from the approach and sampling error between determinations. LC-MS/MS may approximate a random sampling of ions as they elute from the LC column. Variation between technical replicates and variation between individual samples both contribute to sampling error in the LC-MS/MS experiment. Hence, LC-MS/MS suffers from error at the level of the peptides sampled and the proteins identified. While digestion of blood with trypsin produced many tryptic peptides, digestion with chymotrypsin yielded fewer chymotryptic peptides, although this may be a function of ionization and not digestion efficiency.[5,16] Depletion of abundant proteins by protein A/G resin prior to 2D LC of the resulting tryptic peptides with nano C18 ESI-MS/MS yielded about 500 proteins including Interleukins, PI3K, and reportedly prostate-specific antigen from nontryptic peptides with moderate X-Corr and delta Corr scores.[17] About 300 proteins were identified from low molecular mass polypeptides isolated by ultrafiltration in organic solvents and tryptic digestion but with nontryptic correlation searches.[18] Ultra Pressure Liquid Chro-

* To whom correspondence should be addressed. John Marshall, Department of Chemistry and Biology, Faculty of Engineering and Applied Science, Ryerson University, 350 Victoria Street, Toronto Ontario, M5B 2K3. Phone: 416-979-5000, x4219. E-mail: 4marshal@ryerson.ca.

10.1021/pr8005217 CCC: $40.75

Published on Web 02/12/2009

matography, either without sample preparation (800−1682 proteins)[19] or with affinity depletion and 2D LC of tryptic peptides by PS and C18 (953−2258 proteins),[20] produced different sets of nontryptic peptides including reported proteins in the <30 pg/mL range.[19,20] Multidimensional chromatography prior to LC-ESI-MS/MS was used to detect 1444 blood proteins.[21] Twenty-five microliters of serum separated over 100 μL columns of DEAE, and other partition chromatography resins, produced a list of some 651 fully tryptic peptides with high stringency when calculated in single discrete LC-MS/MS runs including cytokine binding proteins[5] and when calculated jointly with a unified scoring algorithm[22,23] produced some 2571 fully tryptic peptides.[14,24] Multidimensional HPLC of intact proteins produced 1008 identifications.[25] A consortium of 35 laboratories produced a list of 9303 proteins from plasma by various methods.[26,27] Multiple depletion columns were compared for the analysis of serum.[28] Batch preparation over magnetic beads produced poor results when compared to HPLC.[29] Partition chromatography of intact proteins by Mono Q (QA) followed by POROS reversed phase HPLC resulted in 2138 proteins by X!Tandem.[30] The use of a peptide library attached to resin as a stationary phase prior to PAGE was observed to work well.[31] A filter apparatus, similar to the hemo filtrate approach,[32] also reportedly permitted the identification of blood proteins thought to be present in low concentrations.[33]

One practical issue common to most chromatographic separation techniques is the potential to foul the column with sample material. Column clogging at nanoflow rates is the major source of downtime for analysis of blood proteins[34] and does not allow for the robust stability[35] required for clinical comparisons. Because nonprotein components including lipids comprise a significant proportion of blood analytes,[36] the direct application of blood to an analytical scale LC system is likely to foul the column and lead to backpressure and ultimately poor separations. The use of a disposable preparative step would not only largely prevent cross-contamination, but act as a guard column for subsequent HPLC separation, preventing rapid fouling by the crude samples.

Blood contains a small number of superabundant proteins and many of these can be isolated into discrete fractions by partition microchromatography.[5] Analogous to the use of SEP Pak cartridges prior to MS analysis of small molecules,[37] or the use C18 zip tips for peptides,[38] it would be prudent to use a disposable micro to milli-scale chromatography to prepare blood fluids prior to analytical separation by HPLC. Disposable microchromatography resin can be discarded after each clinical sample, preventing cross-contamination of the analytical equipment. Digesting the blood before separation will tend to increase the number of polypeptides from ultra high-abundance proteins in subsequent fractions. In contrast, the separation of intact proteins by partition chromatography may remove many nonproteins and isolate high-abundance proteins prior to subsequent sensitive and reproducible HPLC. Fractionating intact blood proteins prior to digestion will likely decrease detection limits by preventing the suppressive effect of abundant peptides competing for ionization at the same time.[1,5,29,30,39]

A rational and integrated strategy for disposable sample preparation, documentation, analysis, computation and database comparison is required for large-scale studies of clinical blood samples. The success and reproducibility of microscale fractionation may be documented by protein assays, SDS-PAGE and MALDI spectra prior to LC-MS/MS. When intact proteins are fractionated prior to digestion, only the LC-MS/MS runs originating from a single source fraction need to be correlated together achieve the appropriate scale of computation efficiency and sensitivity. The MS/MS spectra themselves[40] and the correlated peptides and proteins may be compared using relational databases such as SQL[41,42] or the BLAST[43] algorithm.[14] At the end of the experiment, the same preparative fraction used for discovery by LC-MS/MS may be thawed and used for confirmation and quantification by Western blot or mass accurate analysis.[38,44] Here, we report the development of a rapid and simple procedure that permits detection of low-abundance proteins from blood in the concentration range of tissue leakage proteins and ligand receptors.

Materials and Methods

Materials. Quaternary amine (QA), propyl sulfate (PS), hydroxyl appatite (HA), diethylamino ethanol (DEAE), CBBR conjugated to DEAE (DEAEB), and chromatographic resins were obtained from Bio-Rad (Hercules, CA). Phenol sepharose (HIC), carboxy methyl cellulose (CMC), heparin (HEP), concanavalin A (ConA) and protein G (ProG) were obtained from Pharmacia (Upsalla, Sweden). CBBR, silver nitrate, CHCA, TFA and other buffers and reagents were obtained from Sigma-Aldrich (St. Louis, MO). Cibachron blue was obtained from Fluka (St. Louis, MO). Solvents were optical grade or better and obtained from Caledon (Georgetown, Ontario, Canada).

Gravity Chromatography Columns. A simple and reproducible means of making and running preparative chromatography is to strike 100 μL scale columns in 1−3 mL plastic transfer pipettes stoppered with glass wool.[5,38] The tops of the plastic disposable transfer pipettes were cut off with scissors to form an upper reservoir. The stoppered pipettes were placed within 15 mL disposable tubes to strike the columns (see Supplemental Figure 1 in Supporting Information). The 15 mL tube, and therefore the pipet, was filled to near the top with water. A 200 μL volume of a 50% slurry of each resin was added to the top of the water column in the transfer pipet and allowed to settle to the glass wool at the bottom. The columns were drained and equilibrated with the appropriate loading buffer. The columns were loaded with 25 μL of serum diluted 20-fold in loading buffer to the practical limit before washing in 5 column volumes and elution. A column volume of 100−200 μL can be run by gravity alone without the aid of a frit or centrifugation. Given a binding capacity of 50 mg/mL, such a column could bind up to 5 mg of protein, which is more than sufficient for SDS-PAGE, MALDI and 1D or 2D LC-MS/MS (Figure 1). A microcolumn can be rapidly used to reproducibly fractionate serum[5,38,45] and the results of protein fractions eluted in high concentrations of salts can still be directly compared by tricine electrophoresis to ensure comparable fractionation (see Supplemental Figure 2 in Supporting Information).

Preparative Fractionation. A batch or stepwise partition chromatography was performed with organic buffers and salts. The loading and eluting buffers were usually pH 2 formic, pH 3−4 acetic, pH 4−6 citrate, PBS pH 7, pH 8 tricine, pH 9 tris, and pH 10−12 ethanolamine at 20−50 mM. The columns were loaded, washed extensively and subsequently eluted with the appropriate buffer. The failure of chromatography to provide selectivity may often be the result of insufficient washing.[29,46] The paper dot blot protein assay invented by Dumbroff[47] can be rapidly performed to determine if further washing is required and to quantify eluates in the SDS-PAGE sample

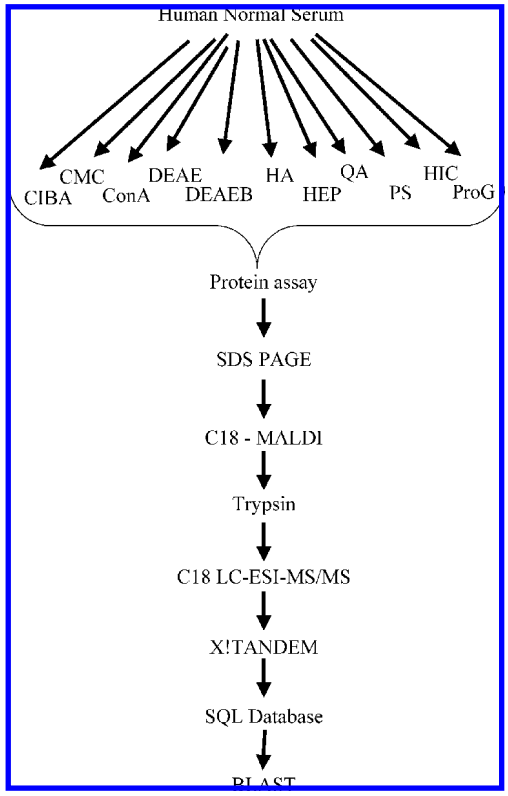

Figure 1. Order of operations for preparative chromatography experiment. The loading and elution conditions in the range described in Table 1 can be tested for selectivity by protein assays and using tricine SDS-PAGE (see Supporting Information) followed by MALDI prior to and after digestion to document the process. The results may be searched with a correlation algorithm and the results refined using SQL and BLAST analysis.

buffer. The useful conditions were determined by first loading the column at one pH extreme or hypotonic salt condition or with progressive elution by salt or pH. A narrow range of salt or pH conditions may be used to reveal the alternative binding conditions that provide differential selectivity. Once the binding conditions were established for the resin, the proteins were loaded and eluted in narrow fractions that provide the desired selectivity. The columns were loaded and washed with the same buffer, eluted with the buffers indicated and were also scrubbed of bound proteins using 1% SDS or Triton X-100 to determine how much protein was irreversibly bound. A similar procedure was followed to determine selective binding and elution conditions for all columns. The various resins were loaded and eluted over a wide combination of pH and salt conditions to yield a reproducible and selective fraction that can be documented by MALDI and SDS-PAGE prior to analysis of tryptic peptide by LC-ESI-MS/MS.[5,38,45]

Protein Assays. The concentration of proteins in each fraction was determined by dot blotting on filter paper alongside BSA standards using the method of Dumbroff.[47] Briefly, the proteins were mixed 1:1 with 2× SDS sample loading buffer and dotted on Whatman number 1 blotting paper alongside BSA standards in the same detergent buffer before drying, rinsing in methanol, and staining with CBBR in water, methanol and acetic acid. The blots were then scanned into an electronic form and concentrations read using Java Image downloaded from the National Center for Biotechnology Information (Bethesda, MD). The Dumbroff method has an advantage prior to electrophoresis since the sample can be measured in detergent and then directly loaded on the gel.

Protein Electrophoresis. Discontinuous polyacrylamide gels with 4% monomer stacking gels with Tris pH 6.8 and 7% monomer separating gels with Tris pH 8.8 were run in tricine anode and cathode buffers.[48] Fractions were resolved and compared by SDS-PAGE on tricine gels stained with Coomassie brilliant blue[49] or diamine silver staining.[50] Tricine electrophoresis was used to resolve the protein fractions for the comparison of selectivity since they often contained salts. Tricine polyacrylamide gels can tolerate large amounts of blood proteins and salts with effective containment of albumin compared to glycine electrophoresis[51] (see Supplemental Figure 2 in Supporting Information). The column chromatography system followed by tricine SDS-PAGE was the basis for the identification of complement fragmentation patterns[38] and many other proteins[5] in blood as subsequently illustrated.[45]

MALDI-TOF. Chromatographic samples for MALDI were diluted 10-fold in 0.2% formic acid. C18 resin was wet in 65% ACN with 0.2% formic acid and then equilibrated in 5% ACN with 0.2% formic acid in water prior to collection of the polypeptides. The peptides were collected with C18 zip tips, washed with over 100 column volumes,[52] and eluted onto gold plated target spots in 65% ACN in 0.2% formic acid. The spots were matrixed with CHCA in 50% ACN with 0.1% TFA (Agilent, Santa Clara, CA). The spots were read with a PBSII MALDI TOF (Ciphergen Biosystems, Freemont, CA).

Trypsin Digestion. One hundred micrograms of each sample was digested in Tris pH 8.8 in 600 mM Urea and 5% ACN with 1 μg of trypsin for 12 h followed by reduction in 2 mM DTT for 30 min at 37 °C prior to another digestion with trypsin. Proteins fractionated over each resin were digested with trypsin and the peptides collected and desalted over C18 using zip tips.[38]

LC-MS/MS. The separation of trytpic peptides by LC-MS/MS was performed exactly as described.[5] All digests of preparative fractions were collected on C18 resin prior to injections.[52] Briefly, the samples were analyzed on an XP-100 Paul ion trap via a microelectrospray ionization head (Thermo Electron Corporation, Waltham, MA) through a metal needle exactly as described by the manufacturer.[35] A 300 μm i.d. column was run at 2 μL/min via an Agilent 1100 pump loaded manually with a 20 μL Rheodyne injection loop. Digested peptides were collected over C18 zip tips and the column was loaded at 5% ACN close to its practical limit with about 5 μg of peptides per LC-MS/MS run. The sample was separated over a linear gradient from 12% to 35% over 90 min, increased to 45% over 30 min, regenerated with a 5 min gradient to 65% ACN, washed at 65% ACN for 5 min, before returning to 5% ACN over 5 min, and equilibration for 5 min prior to the next run.

MS/MS Correlation. The MS/MS spectra were correlated to the tryptic products [R/K]|[X] of the human transcripts in RefSeq downloaded September 2007 using X!TANDEM 2006.06.01.2 where mass errors and putative isotopic forms of the same peptides from −1 to +3 Da from the monoisotopic parent mass and 0.4 Da on the fragment mass were considered together with up to 3 missed cleavages. No modifications or losses other than water, protonation, acetylation or methionine oxidation were considered. The correlations for each column were made using X!TANDEM on generic IBM-compatible personal computers.

SQL and BLAST Comparison. The resulting XML files from individual column fractions were parsed into an SQL database.[41] The redundant sets of related peptides identified by the various methods were collapsed into the set of distinct peptides.

Table 1. Column Binding, Washing and Elution Buffers for Comparisons of Selectivity by SDS-PAGE, MALDI-TOF and LC-MS/MS[a]

resin	binding buffers	elution buffers
Ciba	20 mM Sodium Acetate pH 4.5	20 mM PBS pH 7.0
CMC	1× PBS pH 6.0	1× PBS pH 6.0, 0.4−1.0 M NaCl
CON A	1× PBS	1× PBS, 0.4−1.0 M NaCl
DEAE or DEAEB	20 mM pH 8.5 Tricine, 50−100 mM NaCl	20 mM pH 8.5 Tricine, 0.15−1.0 M NaCl
HA	20 mM PBS pH 7.2	0.5 M PBS pH 6.9
Hep	1× PBS	1× PBS, 600 mM NaCl
QA	20 mM Ethanolamine pH 9.5, PBS	20 mM Ethanolamine, pH 9.5, PBS, 0.4−1.0 M NaCl
QA	20 mM Tricine pH 8.5	20 mM Tricine pH 8.5, 0.4−1.0 M NaCl
PS	20 mM Sodium Acetate pH 4.5, PBS	20 mM Sodium Acetate pH 4.5, PBS, 0.4−1.0 M NaCl
PS	50 mM Citrate pH 4.2	50 mM Citrate pH 4.2, 0.4 − 1.0 M NaCl
HIC	1 M Ammonium Sulfate, 50 mM PBS pH 7.0	0.35 M Ammonium Sulfate 50 mM PBS pH 7.0
ProG	1× PBS*	1× PBS, 0.4 − 1.0 M NaCl

[a] One hundred microlite columns were struck as described and equilibrated, loaded and washed with the range of binding buffer indicated. The samples were eluted with 500 μL (5 column volumes) of the range of elution buffers indicated. The tabulated columns and buffers reflect the range of conditions that have been tested to date and do not represent protocols that have been optimized.

The peptide sequences served as search strings to locate exact matches in a full-length FASTA library of RefSeq proteins.[53−55] The redundant set of protein FASTA sequences were collapsed into the set of distinct FASTA sequences containing the peptide search strings. The distinct FASTA sequences were then distinguished and compared among preparative chromatography columns using SQL, and the results compared to those of BLAST[43] at 75% full length and 20 contiguous amino acids.[14]

Results

Sample Preparation. In preparation for comparative LC-MS/MS, the success and quality of the sample preparations were documented by protein assays, SDS-PAGE and MALDI. Almost any chromatography resin can be used to reduce albumin while simultaneously providing fractions selectively enriched in lower abundance proteins. The different steps required to determine the selective binding and elution conditions of a chromatography resin for one or many fractions are illustrated using PS resin as the example (Figure 2). Some albumin eluted at pH 2 or 2.5 (Figure 2, panel A). Blood proteins remained bound to PS resins at pH 3.0−4.5 and some protein first eluted at pH 5 (Figure 2, panel A) to pH 10 (Figure 2, panel B). Subsequently, proteins were observed to bind well at pH 4−5 and near pH 7.5−8.0 on PS resin (Figure 2, panel C). Binding at pH 4 acetic with progressive elution to pH 8.8 provided a useful range of fractionation (Figure 2, panel D). Holding the pH at 4.5 and increasing salt showed the best selectivity is between the 600 and 1000 mM for the strong cation exchange resin (Figure 2, panel E). Loading and eluting at one-half pH units in the presence of salts provides a clear demonstration of the excellent selectivity that may be achieved by disposable preparative chromatography resins (Figure 2, panels F and G). The final conditions produce a flow through, elution and tightly bound fraction that can be analyzed and compared (Figure 2, panel H). Since every protein is unique, the best binding and elution conditions can be determined for a given protein, but there is no single optimal binding and elution condition for any column. The columns were run as batch and stepwise elutions with the appropriate buffers over a range of loading and unloading conditions of salt and pH as illustrated for PS (Figure 2). The sample binding and elution conditions that resulted in a selective fractionation of the serum were chosen by loading and eluting the samples over a range of salt and pH values prior to comparison of selectivity by SDS-PAGE. Table 1 displays a range of suggested sample binding and elution conditions.

SDS-PAGE. SDS-PAGE on 7% tricine gels stained with CBBR or diamine silver staining showed different protein profiles consistent with selective chromatography compared to crude serum (Figure 3). We selected conditions that reduced albumin while permitting the detection of lower abundance proteins as detected by CBBR or silver staining. Butyl, octyl and phenol sepharose all yielded similar results and those of Phenol sepharose are shown. Several different sources of HA were compared with similar results. On the basis of SDS-PAGE, it appears that each preparative gravity chromatography was able to selectively fractionate serum in a different manner using different combinations of resins, loading and elution buffers.

MALDI-TOF. Fractions from the various columns were compared by MALDI-TOF. The MALDI spectra showed the differential presence of ions that were observed in the specific elution fractions. Inspection of the MALDI spectra revealed many differences in relative ion intensities among columns. The various gravity separations were apparently selective based on MALDI spectra (Figure 4). All resins could retain and elute some peptides. We observed that PS resin was capable of capturing most peptides from serum but tended to be more retentive compared to other columns. C18 resins alone, without prefractionation, sometimes clogged or failed to provide sensitive detection of peptides with high signal-to-noise ratios. After preparative fractionating, strong signals and a much wider variety of endogenous peptides could be consistently observed (Figure 4).

LC-MS/MS. The intact proteins collected over disposable preparative microchromatography were digested and separated by LC-MS/MS. The selectivities of the chromatographic supports were reflected in the TIC traces generated by LC-ESI-MS/MS with an ion trap (Figure 5). In general, we observed that a 90 min gradient from 12% to 35% ACN permitted the identification of peptides from digests of human serum. Above 45% ACN, complex ion series emerged, perhaps indicating that analytes too large to effectively fragment by CID were eluting from the C18 column. The TIC traces from the peptide digests of the different columns all showed different profiles of ion current over time. Under these conditions, the analytical column was not fouled, showed little change in backpressure and could be used for hundreds of runs at the practical load limit. Our results suggest that the preparative chromatography methods used here produced a selective fractionation of serum proteins by SDS-PAGE, MALDI-TOF and HPLC-ESI-MS.

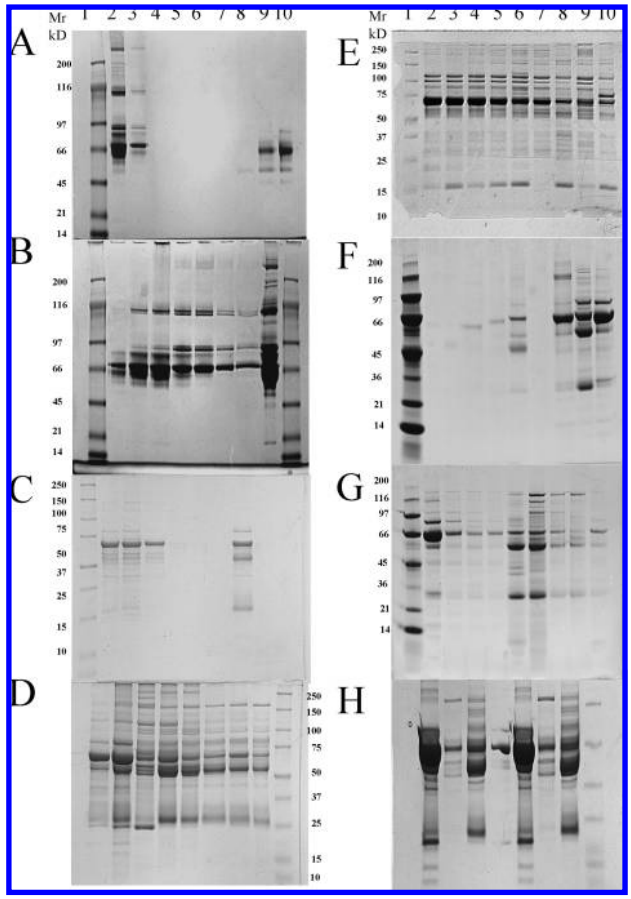

Figure 2. Illustration of the process for determining the binding and elution conditions for PS preparative chromatography. (A) A PS column loaded in 0.2% TFA and eluted one-half pH units. Lanes: (1) marker; (2) pH 2.0 formic; (3) pH 2.5 formic; (4) pH 3 acetic; (5) pH 3.5 acetic; (6) pH 4 acetic; (7) pH 4.5 citrate; (8) pH 5 citrate; (9) pH 5.5 citrate; (10) pH 6 citrate; (B) A PS column loaded in 0.2% TFA and eluted one-half pH units. Lanes: (1) marker; (2) pH 6.5 PBS; (3) pH 7.0 PBS; (4) pH 7.5 PBS; (5) pH 8.0 tricine; (6) pH 8.5 tricine; (7) pH 9.0 tris; (8) pH 10 ethanolamine; (9) crude NHS; (10) marker. (C) Binding conditions for PS with elution at 1 M NaCl. Lanes: (1) marker; (2) 50 mM citrate pH 4.0; (3) 50 mM citrate pH 5.0; (4) 50 mM citrate pH 6.0; (5) 20 mM phosphate pH 6.5; (6) 20 mM phosphate pH 7.0; (8) 20 mM tricine pH 7.5; (9) 20 mM tricine pH 8.0; (10) 20 mM tricine pH 8.5. (D) Binding at pH 4 acetic with elution using different pH buffers. Lanes: (1) 50 mM citrate pH 5.0; (2) 50 mM citrate pH 6.0; (3) 20 mM phosphate pH 6.5; (4) 20 mM phosphate pH 7.0; (5) 20 mM tricine pH 7.5; (6) 20 mM tricine pH 8.0; (7) 20 mM tricine pH 8.5; (8) 20 mM tris pH 8.8. (E) Binding at 20 mM acetate pH 4.5 with different levels of salt with elution at 1000 mM NaCl. Lanes: (1) marker; (2) 20 mM acetate pH 4.5 with 400 mM NaCl; (3) 500 mM NaCl; (4) 600 mM NaCl; (5) 650 mM NaCl; (6) 700 mM NaCl; (8) 750 mM NaCl; (9) 800 mM NaCl; (10) 1000 mM NaCl. (F) Binding and elution separated by one-half pH values. Lanes: (1) marker; (2) loaded in 0.2% TFA and eluted in pH 2.0 formic; (3) pH 2.5 formic; (4) pH 3 acetic; (5) pH 3.5 acetic; (6) pH 4 acetic; (7) pH 4.5 citrate; (8) pH 5 citrate; (9) pH 5.5 citrate; (10) pH 6 citrate. (G) Binding and elution separated by one-half pH values. Lanes: (1) marker; (2) bound in pH 6 and eluted at pH 6.5 PBS; (3) eluted at pH 7.0 PBS; (4) 7.5 PBS; (5) pH 8.0 tricine; (6) pH 8.5 tricine; (7) pH 9.0 tris; (8) pH 9.5 tris; (10) pH 10 ethanolamine. (H) Two independent replicates of NHS fractionated over a propyl sulfate column showing the flow through, elution and scrub fractions as described in Table 1. Lanes: (1) empty; (2) flow through; (3) elution; (4) scrub; (5) empty; (6) flow through; (7) elution; (8) scrub; (9) marker; (10) empty. Note irreversible binding to PS resin lanes 4 and 8.

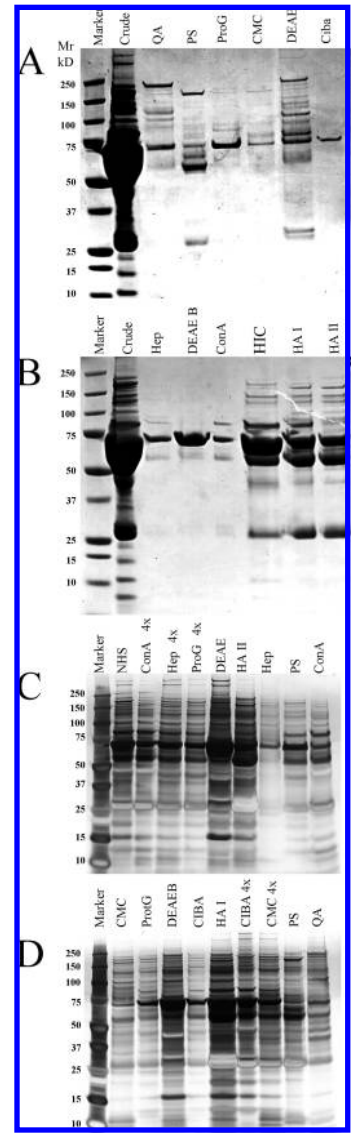

Figure 3. SDS-PAGE illustrating the selectivity that may be achieved using micropreparative batch chromatography of normal human serum. Proteins were separated over resins as indicated using the loading and elution conditions in Table 1. Except where indicated, all fractions are elutions. (A and B) Protein fractions loaded by volume of eluate per lane alongside NHS and stained with CBBR. HA I and II were variants of HA resin. (C and D) Protein fractions loaded by equal volume of eluate stained with diamine silver stain. The 4× indicates that four times the volume was loaded.

MS/MS Correlation. Calculating the apparently random MS/MS spectra that resulted from the blank runs produced few or no correlations by X!TANDEM, similar to previous blank runs analyzed by SEQUEST.[5] All columns provided roughly similar numbers of identified proteins when an equal number of runs were compared, with diminishing marginal returns from additional runs. However, PS showed a large fraction of proteins that remained irreversibly bound to the resin (Figure 4, panel H, lanes 4 and 8). Under our conditions, Apolipoprotein AI (and not albumin, tranferin, IgG or other commonly depleted serum proteins) was by far the most intensely analyzed protein with or without depletion of albumin with DEAEB/CIBA or IgG with ProG (see Supplemental Tables 1 and 2 in Supporting Information). The specific depletion of albumin and other major serum

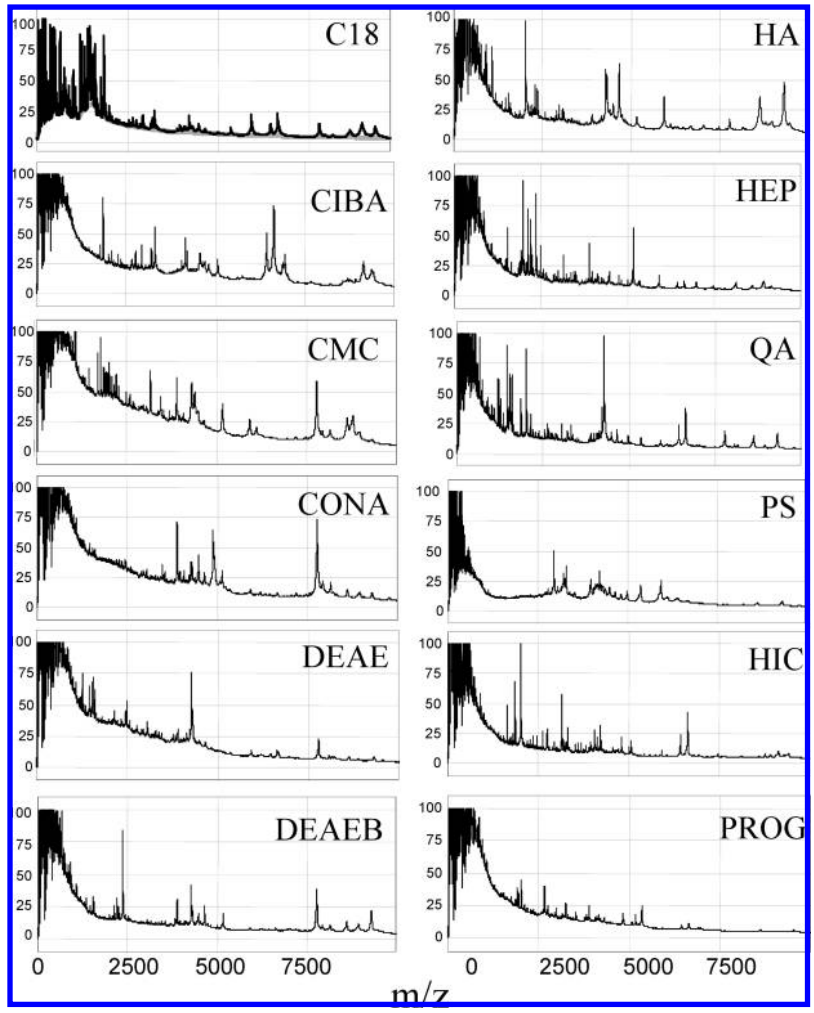

Figure 4. MALDI-TOF spectra illustrating the selectivity achieved using micropreparative batch chromatography of normal human serum. The microcolumns were loaded and eluted as described in Table 1. The eluted product was *not* digested with trypsin prior to collection over C18, followed by elution and drying on gold plated MALDI targets, prior to the application of CHCA matrix and MALDI-TOF with a Ciphergen PBSII.

proteins was not required to analyze blood proteins by preparative chromatography prior to LC-MS/MS which is in agreement with our previous results.[5]

Preparative partition chromatography prior to digestion and collection over C18 identified many proteins from the serum samples by C18 LC-ESI-MS/MS and X!TANDEM correlation. Heparin, QA, PS, ConA and DEAE were complementary in terms of their MALDI and SDS-PAGE profiles and the ability to detect some apparent lower abundance peptides. These were selected to break serum into as many as 15 fractions by either salt and/or pH under various binding and elution conditions. The most productive samples were analyzed numerous times in proportion to the number of proteins observed in the subfraction and/or the tryptic peptide subfractionated over PS.[56] The resulting spectra from all columns were then searched by X!TANDEM. The protein sequences were collapsed into the set of distinct FASTA sequences that contained the correlated peptides returned from X!TANDEM. The many-to-one collapse of the peptide and protein sequences usually resulted in only one representative protein product carrying forward to the final database. In many cases, however, there was more than one sequence that contained some or all of the peptide search strings in related isoforms or splice variants of the protein (e.g., see fibronectin 1 in Supplemental Table 1 of Supporting Information). After collapse, as few as 82 total proteins per column remained in the case of CIBA, and as many as 1433 distinct proteins were identified by preparation with QA. Prefractionation of proteins by ConA, DEAE, QA, PS, and Hep resins each produced correlations to many hundreds or even thousands of proteins by X!TANDEM (Table 2). In total, the MS/MS spectra were correlated to 4396 different FASTA sequence entries of the human forward RefSeq by X!TANDEM. We estimate that about 61% of these proteins have been previously detected with MASCOT, SEQUEST or other algorithms. However, 3219 of the 4396 distinct proteins have not been previously correlated from blood fluids by rigorous X!TANDEM correlation after examining the blood data made publicly available from the General Proteomic Machine.[40]

Comparison of Proteins Identified by Chromatography Resins. Many of the proteins were apparently best observed with particular chromatographic preseparations and not the others. The SQL database[41] was used to estimate the total number of proteins that may be implicated by the MS/MS spectra.[57] The proteins can also be further collapsed using BLAST at 75% full length and 20 contiguous amino acid residues and selecting the top hit.[14] The the overlap between columns was calculated to be about 20−35% by either SQL or BLAST. It appeared that the selectivity of the column presepa-

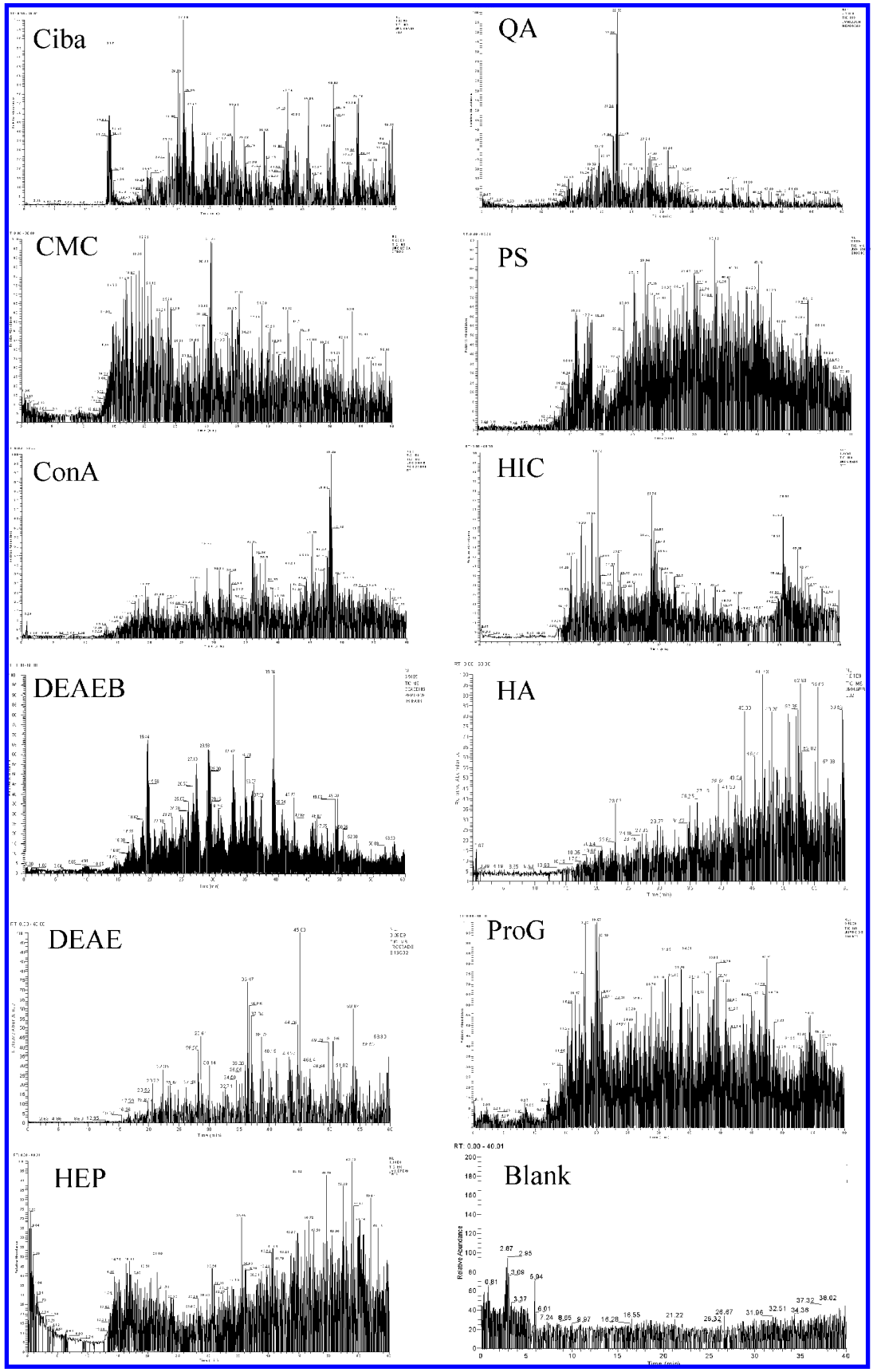

Figure 5. LC-TIC chromatograms illustrating the selectivity achieved using micropreparative batch chromatography of normal human serum. The microcolumns were loaded and eluted as described in Table 1. The eluted fraction was digested with trypsin, reduced in DTT, redigested with trypsin prior to collection over C18, followed by elution, dilution in 20 μL of 0.1% acetic acid, with injection and separation at 2 μL on a 300 μm i.d. $\times$ 15 cm analytical C18 column with ionization by microspray ESI into a Thermo Electron Corp LCQ XP-100 ion trap.

Table 2. Redundant and Distinct Counts of Both Peptides and Proteins[a]

chromatography	runs	peptides total	peptides distinct	proteins total	proteins distinct	proteins/run	reported
Ciba	5	900	411	1399	82	16.4	83.3
CMC	8	953	432	1481	138	17.2	86.99
ConaA	172	29950	2881	46229	1259	17.5	76.27
DEAE	187	35950	2910	58956	1185	6.3	78.27
DEAEB	17	2174	719	3102	158	9.2	78.95
HA	18	1552	616	2288	174	9.6	78.63
Hep	124	33227	2573	51336	1048	8.5	77.47
QA	242	36949	3039	55654	1433	5.9	78.06
PS	174	4705	1338	7015	503	2.9	79.09
HIC	29	1780	702	3415	266	9.1	75.38
ProG	19	3228	633	5074	217	11.4	77.53
totals		151368	16254	235949	6463		61.11
unique totals			**8910**		**4369**		

[a] X!Tandem results were imported into an SQL database system and the redundant and distinct counts of both peptides and proteins where determined using SQL methods. Peptide search strings were used to locate all full-length FASTA files that contained the peptide in RefSeq. The distinct counts across all resins were also calculated. Distinct means nonredundant in that the peptide or protein differs by at least one amino acid from any other tabulated sequence. A run means a batch or stepwise fraction from the indicated column followed by a 1D LC-MS/MS experiment. Reported shows the percentage of proteins found to be previously discovered from blood by 75% full-length homology and 20 contiguous amino acids.

Table 3. Resin Cross-Comparison Using Distinct Proteins by SQL[a]

resin	distinct	Ciba	CMC	ConA	DEAE	DEAEB	HA	Hep	QA	PS	HIC	ProG
Ciba	82	82	48	60	61	46	49	65	57	56	51	46
CMC	138	48	138	87	94	72	58	85	83	79	69	68
ConA	1259	60	87	1259	282	102	90	227	295	154	105	106
DEAE	1185	61	94	282	1185	105	84	253	277	157	113	102
DEAEB	158	46	72	102	105	158	61	97	107	84	82	68
HA	174	49	58	90	84	61	174	80	90	79	61	68
Hep	1048	65	85	227	253	97	80	1048	241	139	97	85
QA	1433	57	83	295	277	107	90	241	1433	167	104	101
PS	503	56	79	154	157	84	79	139	167	503	122	100
HIC	266	51	69	105	113	82	61	97	104	122	266	71
ProG	217	46	68	106	102	68	68	85	101	100	71	217

[a] The comparison and calculation of overlap between preparative chromatography resins using SQL. A distinct (i.e., nonredundant) database table of protein sequences was generated for each resin. The resultant tables were then cross-compared using SQL to determine the degree of overlap between the resins. The set of distinct peptides was located in the FASTA sequences of RefSeq. The set of FASTA sequences was rendered distinct by SQL. The set of distinct full-length proteins was compared between columns by SQL and exact matches between resins are shown.

ration observed by SDS-PAGE, MALDI and LC TIC traces was also reflected in the composition of the protein lists obtained from each resin. Aligning the peptides obtained from different proteins shows that even abundant proteins, and most low-abundance proteins, were better detected using specific column(s) that yielded a greater number of peptides than others. We observed that the ion exchangers DEAE, QA, and CMC performed well, although PS (sometimes called SCX) may require a stronger elution buffer. ConA was simple and effective for examining glycoproteins in the absence of complex hydrazide chemistry.[58] HEP revealed proteins that interacted with sulfated carbohydrates, such as ALPHA 1 TYPE XIII COLLAGEN and many others (see Supplemental Table 1 in Supporting Information). For example, almost 500 peptides from fibronectins were identified from Hep resin while only 3 fibronectin peptides were observed from QA or ConA. The peptide coverage for many of these proteins is so low that it is often not possible to determine with certainty which different related proteins or splice variants were actually the ones detected in the sample. The differences in fibronectin and other abundant proteins observed between columns cannot be attributed to sampling error (Tables 3 and 4). Therefore, different columns showed distinct selectivity and resulted in the identification of different sets of proteins. In the case of human blood proteins searched against NR RefSeq by X!TANDEM and made distinct by SQL based on full-length protein sequences, the results of the mathematically independent SQL and BLAST analyses are nearly identical (cf. Tables 3 and 4). It should be possible to examine and compare the coverage of important or differing isoforms directly from the peptide coverage to determine which isoforms cannot be ruled out. The overlap between proteins identified from the resins was typically between 20% and 35% by BLAST and SQL. All the resins were complimentary and were required for the comprehensive investigation of serum by LC-MS/MS.

Protein Distributions. The distribution of scores to the same set of MS/MS spectra from all DEAE files were arbitrarily compared between searches of human forward versus human reverse protein libraries (see Supplemental Figure 3 in Supporting Information). The forward search was clearly distinguished from the reverse search below a score of log −1.6 to −2.0. There were few or no correlations to the reverse library below these cutoff values since the proteins rarely showed two and almost never three peptides correlated to the same protein.[59] In contrast, about 1094 of the proteins in the human forward search, and few in the control search, showed more than one peptide by X!TANDEM and therefore were nearly certain.[59] Applying the human reverse decoy cutoff score to the whole database of 4396 proteins results in some ∼2257 of the proteins that were clearly distinguished from the arbitrary

Table 4. Resin Cross-Comparison Using BLAST 75% Identity, 20 AA with Top Hits Rendered Distinct by SQL[a]

resin	distinct	Ciba	CMC	ConA	DEAE	DEAEB	HA	Hep	QA	PS	HIC	ProG
Ciba	82	82	48	60	62	48	51	65	57	56	52	47
CMC	138	48	138	89	95	76	59	85	87	81	71	70
ConA	1259	61	88	1257	291	106	97	241	314	162	107	109
DEAE	1185	62	94	291	1185	110	90	265	293	165	117	105
DEAEB	158	46	72	102	106	158	61	99	107	86	82	68
HA	174	49	58	92	86	61	174	80	91	83	62	70
Hep	1048	65	85	240	255	100	84	1048	254	143	101	87
QA	1433	61	87	319	298	113	99	259	1433	181	109	108
PS	503	56	79	159	164	87	85	141	173	503	126	102
HIC	266	51	69	107	114	83	65	104	110	132	266	72
ProG	217	46	68	109	104	68	70	90	103	103	72	217

[a] The comparison and calculation of overlap between preparative chromatography resins using SQL and BLAST. Distinct protein sequence FASTA files were generated for each resin using SQL and then blasted against one another to produce an 11 × 11 matrix of BLAST results in the m8 BLAST report format. After importing the matrix report into a database, hits with less than 75% full length and less than 20 contiguous amino acids were discarded. Additional filtering removed all but the top hit for any protein−protein comparison between any two resins. The remaining records were then cross-compared using SQL to determine the overlap between the resins.

pseudo control of reverse correlation with an empirically derived cutoff (see Supplemental Figure 3 in Supporting Information). However, all the 4396 proteins have an expectation value of E-1 or lower (90% confidence) as calculated by X!TANDEM that appears to be reliable.[57,60,61]

After removing redundant identifications to yield the set of 4396 distinct protein sequences, the length distribution of proteins from human blood was compared to that of the starting RefSeq database (Figure 6). The proteins were binned in lengths of 100 amino acids. A preponderance of correlations to extremely long proteins might be indicative of false identifications.[27] The human RefSeq database contains thousands of distinct proteins with lengths below 1000 amino acids (Figure 6, Human). RefSeq contains relatively few proteins of greater than 3000 amino acids in length. The redundant blood proteins showed tens of thousands of proteins identified below 1000 amino acids in length (Figure 6, Redundant). After collapse into the set of distinct proteins, there were several hundred proteins in the bins below 1000 amino acids and relatively few proteins more than 3000 amino acids in length. However, there was a relatively small minority of proteins of greater than 3000 amino acids in length. In general, blood proteins appeared on average to be slightly longer than typical human proteins.

Sensitivity and Specificity. The analysis detected the classic abundant serum proteins and moderately low-abundance blood proteins that are supposed to result from tissue leakage.[62] Moderately low-abundance proteins were observed including some kallikreins, plasminogen, retinal binding protein, myoglobin, thyroglobin, neuron specific enolase, retinoic acid receptor (see Supplemental Figure 4 in Supporting Information) and others that may exist in concentrations of $\sim 10^3$ to 10^4 pg/mL (see Supplemental Tables 1 and 2 in Supporting Information). These proteins have also been observed in other LC-MS/MS experiments from depleted serum or plasma.[14] TNF associated factors, ligand binding proteins and a variety of other receptors were observed with weak correlations in agreement with results from other groups (see Supplemental Table 3 in Supporting Information). However, the preparative prefractionation of 25 μL of serum analyzed at 2 μL/min with an LCQ ion trap should not be sufficient to expose extreme low-abundance proteins such as troponins, prostate-specific antigen, tissue factor, interleukins and others.

In general, highly abundant and moderately abundant proteins were identified by relatively short tryptic peptides[5] with few missed cleavage sites[14] and with many of the fragment ions accounted for (cf. high-abundance inter alpha globulin Inhibitor verses other lower abundance peptides in Supplemental Figure 4 in Supporting Information). Moreover, the reproducibility of the spectra was good with similar fragments observed from replicate kallikrein and retinoic acid receptor spectra (see Supplemental Figure 4 in Supporting Information). On the basis of the known concentration of the lowest abundance proteins confidently observed, we estimate that the sensitivity of the preparative chromatography alone was on the order of thousands of pg/mL. A sensitivity estimate of 1000 pg/mL is consistent with the paucity of extreme low-abundance proteins confidently identified.

Discussion

Blood comes in contact with all cells and organs; it is a convenient test fluid that would be the most universal source of markers for broad screening.[62] While urine and other fluids also hold great promise,[63−65] many larger proteins may not enter the urine,[66] thus, partially limiting its utility. Great efforts have been made to enumerate the proteins in human blood that might contain biomarkers.[5] Peptides from ~18 000 distinct proteins may have been observed in blood to date[14] and that figure is likely to increase substantially. Changes in the host response reflected by the presence, or modification, of low-abundance polypeptides discovered by LC-MS/MS and subsequently confirmed and measured by targeted analysis or with good parent mass accuracy[67−69] may eventually permit direct examination of disease states in blood samples. Hence, the identities of blood proteins, the columns that best capture them, and a list of the most detectable peptide ions and their fragments are all highly desirable information. Preparative microchromatography using a variety of resins[5] permits the small-scale primary discovery experiments on a limited number of control and disease blood samples *and* the capacity to repetitively generate only a single subfraction hundreds of times, for subsequent MS or immunological confirmation and measurement.

LC-MS/MS is more convenient, cost-effective and sensitive than SDS-PAGE.[4,5,30] Few of the proteins identified by early multiple depletion HPLC and 2D LC studies were from fully tryptic peptides.[17,19,20] In contrast, proteins fractionated by partition chromatography prior to LC-ESI-MS/MS yielded much higher quality results with 100% tryptic peptides.[5] Simplifying blood fluids by trypsin digestion prior to separation

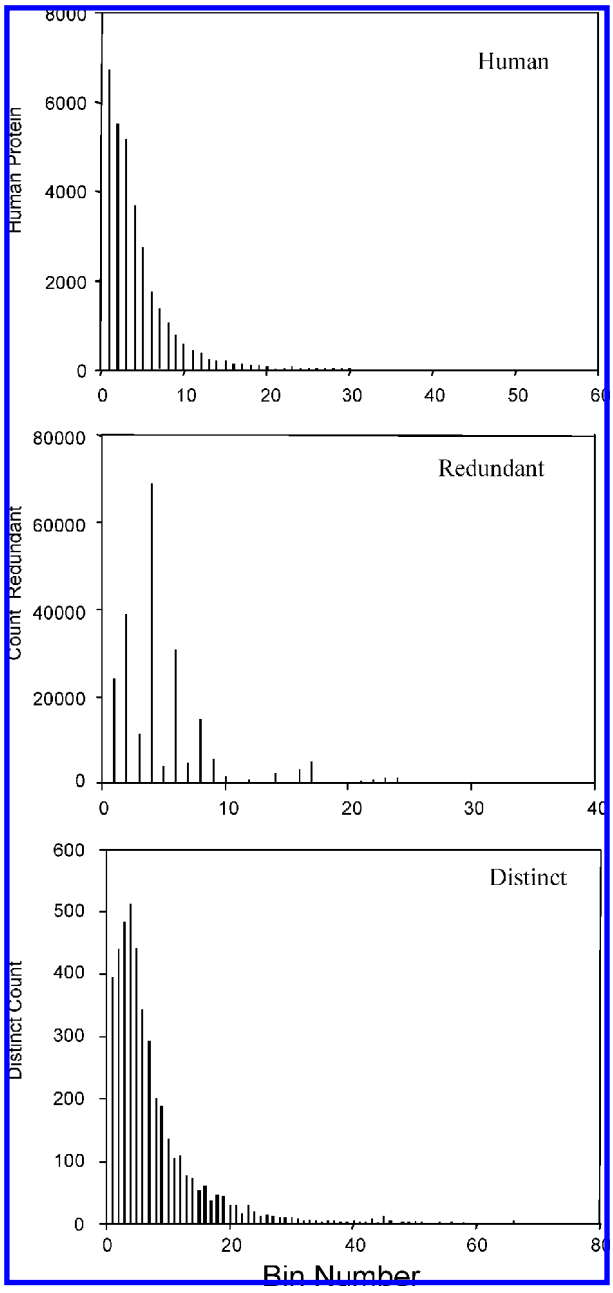

Figure 6. Protein length distribution histogram of the redundant and distinct data compared the distribution of RefSeq. The full-length FASTA sequences were binned into groups by 100 amino acid increments in the length of the encoded amino acid chain. The distribution of the lengths are shown alongside the distribution of RefSeq (Human). The distribution is shown along the steps from the redundant LC-MS/MS correlation produced by X!TANDEM (Redundant) to the set of nonredundant FASTA sequences (Distinct).

is not possible, as digestion will multiply the high molar abundance species and ensure that peptides from high-abundance proteins will interfere with the ionization of low-abundance peptides during subsequent separation and analysis. Instead, a more rational approach may be to separate the full-length proteins prior to digestion and LC-MS/MS.[1,5,29−31,39] The proteins may be selectively captured, bound and eluted from a preparative chromatography resin prior to analysis to make reproducible and repetitive fractionation of selected blood proteins from nonproteinaceous materials.[5,38,45]

Protein prefractionation ensures that only peptides originating from one starting fraction appear in any one, or set of, LC-MS/MS run(s). Since every protein has unique biochemical characteristics, there may not be a single universal sample preparation system[56] for preparing proteins from higher Eukaryotes for LC-MS/MS that is analogous to the single set of general methods for working with nucleic acids.[70−76] Many low-abundance proteins had been purified to homogeneity from blood using chromatography[7−13] before the advent of proteomics.[28] Thus, partition microchromatography may be a simple and rapid method to reproducibly fractionate serum by a variety of orthogonal methods in a comparative manner for protein discovery and validation.[5,38] Preparative chromatography may also serve as an orthogonal prefractionation for both capillary electrophoresis[77] and free flow electrophoresis,[78] as well as LC-MS/MS. Note that the selectivity between adjacent column fractions provided by the judicious use of preparative disposable resin rivaled the separations achieved by HPLC in the study by Faca et al.[30] and similar uses of binary HPLC buffer systems. The results of this study are a conservative estimate of the number of proteins in the concentration range of known biomarkers that may be identified from blood by these preparative chromatography methods. There is no doubt that a continued focus on perfecting preparative microchromatography fractionation prior to analytical HPLC will permit the detection of many more blood proteins using LC-MS/MS.

Proteins of moderate low abundance, such as those lost from organs, tissues or cells, were identified as expected and with convincing CID spectra by disposable micropreparative chromatography followed by LC-ESI-MS/MS at 2 μL on a Paul ion trap. Tryptic peptides from superlow abundance proteins such as PSA, chemokines, cytokines, interferon and interleukins, previously implicated from nontryptic peptide correlations,[17,19,20] were not detected from 25 μL of serum in these experiments. Proteins at the lower end of reliable detection in this study such as myosins, enolase, myoglobins, retinol binding protein, kallikreins, thyroglobulin and retinoic acid receptor are in concentrations of 1000−10 000 pg/mL. Therefore, on the order of tens to hundreds of femtomoles of peptide species should have been loaded onto the LC column consistent with measured sensitivity of the microflow LC-ESI-Paul-ion trap system. The agreement on the presence of myosins, enolase, myoglobins, retinol binding protein, and thyroglobulin between published LC-MS/MS experiments is also consistent with their reliable identification.[14] We conclude that the moderate low-abundance proteins reported were most often correct and the calculated probability of correlation by X!TANDEM appears reliable.[57,60] Presently, mass spectrometry only detects on the order of 0.1 to 10% of ions[79] and this should improve over time, making the direct analysis of blood even more attractive.

Correlation algorithms can obtain greater apparent sensitivity by the cumulative summation of correlation scores in a joint calculation of many different LC-MS/MS runs.[14] Algorithms such as X!TANDEM[57] and Paragon[80] are presently embodied in applications that are amenable to calculating each run separately using modest computers and subsequently combining the results without altering apparent sensitivity.[81] In this regard, the approach of fractionating intact proteins is also more amenable to computation since all the peptides from a given protein remain together and in a limited number of LC-MS runs, and can thus be conveniently calculated on simple low cost computer systems and later assembled in SQL. SQL and BLAST can be used to reduce the often highly redundant

results of correlation analysis to smaller sets of nonredundant proteins implicated by the data. The peptides and proteins discovered can be compared and contrasted between populations of LC-MS/MS runs by SQL and BLAST to reveal the instances where one treatment data set qualitatively differs from another.[14] Subtraction of populations of LC-MS/MS runs at the level of identified proteins would permit researchers to focus on a small number of proteins that differ among large populations of LC-MS/MS runs for further examination. It has been suggested that the protein length distribution obtained by MS/MS correlation should match that of the search database.[82] The rules introduced to implement SQL or BLAST procedures (i.e., collapsing to the longest homologue) might impose a bias on protein length that must be accounted for in subsequent calculations. The data may indicate that blood proteins show a bias toward larger sizes consistent with the requirement to avoid excretion through the kidneys.[66]

Albumin and other major blood proteins must be removed prior to 2D PAGE[15,35,83] but not tricine SDS-PAGE.[5] In terms of the number of proteins identified, dealbuminization resins such as DEAEB and CIBA were not much different from other resins when compared on an equal number of runs. The results here clearly show that the removal of abundant blood proteins by depletion chromatography is not required prior to LC-MS/MS. All the chromatography resins tested could be run with loading and elution conditions that markedly diminish albumin and some other major proteins while selectively enhancing a subset of lower abundance proteins. Microscale partition chromatography of intact proteins can be used directly to discover proteins from blood without specific dealbuminization, and different proteins are best enriched with different chromatographic resins. This is in agreement with numerous previous experiments prior to the invention of LC-MS/MS.[7−13] Reduction and alkylation of blood proteins may be avoided prior to LC-MS/MS with good results[5] since albumin and IgG, two of the major contaminants in blood, are both held together in their own three-dimensional structure by multiple intramolecular disulfide links and are thus naturally resistant to trypsin. In fact, the reduction and alkylation protocol was specifically optimized to linearize and digest albumin.

Intact blood proteins may also need to be fractionated by HPLC techniques.[1,29,30] A reproducible preparative chromatography step is a practical consideration prior to HPLC. Preparative batch fractionation may act as a disposable guard column for each clinical sample. Lipids, carbohydrates, salts and other sample components that might foul HPLC resins or LC-ESI were apparently removed by disposable preparative resin, permitting hundreds of separations on the same column under stable conditions at 2 μL/min. Optimized loading, washing and elution conditions are essential for chromatography to achieve selectivity.[29] The disposable preparative microchromatography system described here can produce up to 5 mg of intact proteins from each resin by this rapid and reproducible method. Subsequent analytical HPLC chromatography may provide much greater purification, and therefore sensitivity, by separation at the milligram scale and should permit effective ionization of ultralow abundance proteins from blood via nanospray 1D or 2D LC-ESI-MS/MS.[28,34]

Conclusion

A reliable preparative chromatography system, as documented by protein assays, SDS-PAGE and MALDI, will be a crucial step toward the repetitive fractionation of clinical blood samples for direct discovery and validation of proteins in the concentration range of many presently utilized biomarkers. The development of a wide array of disposable micropartition chromatography protocols will be a key step in the practical analysis of clinical blood samples by LC-ESI-MS/MS. Stepwise and batch elution using gravity over microchromatography columns identified peptides from some 4396 distinct proteins of the human forward RefSeq by X!TANDEM. Peptides from proteins in the concentration ranges of tissue or cell leakage and receptors were detected.

Abbreviations: ACN, acetonitrile; CIBA, Cibachron blue resin; CBBR, Coomassie brilliant blue reagent; CHCA, cyano-4-hydroxy cinnamic acid; CMC, carboxy methyl cellulose; ConA, Concanavalin A; HA, Hydroxy Apatite; ESI; electrospray ionization; MALDI-TOF, matrix-assisted laser desorption/ionization-time-of-flight; PS, Propyl Sulfate; QA, Quaternary Amine; DEAE, Diethylamino Ethanol; DEAEB, Diethylamino Ethanol Blue; HIC, Phenol Sepharose; HEP, Heparin; i.d., inner diameter; ProG, Protein G; PAGE, polyacrylamide gel electrophoresis; PBS, phosphate buffered saline; LC-MS/MS, liquid chromatography tandem mass spectrometry; RefSeq, the Reference Sequence Database; TFA, trifluoroacetic acid.

Acknowledgment. This work was supported with a grant from the Heart and Stroke Foundation of Ontario and Canada.

Supporting Information Available: Supplemental information in the form of additional figures, tables and SQL databases is provided. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Gao, J.; et al. Biomarker discovery in biological fluids. *Methods* **2005**, *35*, 291–302.
(2) Issaq, H. J.; Conrads, T. P.; Janini, G. M.; Veenstra, T. D. Methods for fractionation, separation and profiling of proteins and peptides. *Electrophoresis* **2002**, *23*, 3048–61.
(3) Luque-Garcia, J. L.; Neubert, T. A. Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry. *J. Chromatogr., A* **2007**, *1153*, 259–76.
(4) Koller, A.; et al. Proteomic survey of metabolic pathways in rice. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11969–74.
(5) Marshall, J.; et al. Human serum proteins preseparated by electrophoresis or chromatography followed by tandem mass spectrometry. *J. Proteome Res.* **2004**, *3*, 364–82.
(6) Putnam, F. *The Plasma Proteins: Structure Function, and Genetic Control*, 2nd ed.; Academic Press: New York, 1975.
(7) Duckert, F.; Koller, F.; Matter, M. Purification and physiological properties of factor VII from plasma and serum; separation from prothrombin. *Proc. Soc. Exp. Biol. Med.* **1953**, *82*, 259–61.
(8) Tollefsen, D. M.; Majerus, D. W.; Blank, M. K. Heparin cofactor II. Purification and properties of a heparin-dependent inhibitor of thrombin in human plasma. *J. Biol. Chem.* **1982**, *257*, 2162–9.
(9) Kanai, M.; Raz, A.; Goodman, D. S. Retinol-binding protein: the transport protein for vitamin A in human plasma. *J. Clin. Invest.* **1968**, *47*, 2025–44.
(10) McConnell, D. J.; Mason, B. The isolation of human plasma prekallikrein. *Br. J. Pharmacol.* **1970**, *38*, 490–502.
(11) Vasileva, R.; Jakab, M.; Hasko, F. Application of ion-exchange chromatography for the production of human albumin. *J. Chromatogr.* **1981**, *216*, 279–84.
(12) Poffenbarger, P. L. The purification and partial characterization of an insulin-like protein from human serum. *J. Clin. Invest.* **1975**, *56*, 1455–63.
(13) Murthy, R. J.; Hercz, A. Purification of human alpha 1-antitrypsin by affinity chromatography on sepharose bound concanavalin A. *FEBS Lett.* **1973**, *32*, 243–6.
(14) Zhu, P.; et al. Comparison of protein expression lists from mass spectrometry of human blood fluids using exact peptide sequences versus BLAST. *Clin. Proteomics* **2007**, *2*, 185.
(15) Pieper, R.; et al. The human serum proteome: display of nearly 3700 chromatographically separated protein spots on two-

dimensional electrophoresis gels and identification of 325 distinct proteins. *Proteomics* **2003**, *3*, 1345–64.
(16) Olsen, J. V.; Ong, S. E.; Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **2004**, *3*, 608–14.
(17) Adkins, J. N.; et al. Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. *Mol. Cell. Proteomics* **2002**, *1*, 947–55.
(18) Tirumalai, R. S.; et al. Characterization of the low molecular weight human serum proteome. *Mol. Cell. Proteomics* **2003**, *2*, 1096–103.
(19) Shen, Y.; et al. Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome. *Anal. Chem.* **2004**, *76*, 1134–44.
(20) Shen, Y.; et al. Characterization of the human blood plasma proteome. *Proteomics* **2005**, .
(21) Chan, K.; 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. Analysis of the Human Serum Proteome. *Clin. Proteomics* **2004**, *1*, 101–225.
(22) Moore, R. E.; Young, M. K.; Lee, T. D. Qscore: an algorithm for evaluating SEQUEST database search results. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 378–86.
(23) Chelius, D.; et al. Analysis of the adenovirus type 5 proteome by liquid chromatography and tandem mass spectrometry methods. *J. Proteome Res.* **2002**, *1*, 501–13.
(24) Zhu P-H, K. K.; Jankowski, A.; Marshall, J. Comparison of published human serum/plasma data. *Mol. Cell. Proteomics* **2004**, *3*, S235.
(25) Jin, W. H.; et al. Human plasma proteome analysis by multidimensional chromatography prefractionation and linear ion trap mass spectrometry identification. *J. Proteome Res.* **2005**, *4*, 613–9.
(26) Omenn, G. S.; et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **2005**, *5*, 3226–45.
(27) States, D. J.; et al. Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat. Biotechnol.* **2006**, *24*, 333–8.
(28) Liu, T.; et al. Evaluation of multiprotein immunoaffinity subtraction for plasma proteomics and candidate biomarker discovery using mass spectrometry. *Mol. Cell. Proteomics* **2006**, *5*, 2167–74.
(29) Whiteaker, J. R.; et al. Head-to-head comparison of serum fractionation techniques. *J. Proteome Res.* **2007**, *6*, 828–36.
(30) Faca, V.; et al. Contribution of protein fractionation to depth of analysis of the serum and plasma proteomes. *J. Proteome Res.* **2007**, *6*, 3558–65.
(31) Sennels, L.; et al. Proteomic analysis of human blood serum using peptide library beads. *J. Proteome Res.* **2007**, *6*, 4055–62.
(32) Richter, R.; et al. Composition of the peptide fraction in human blood plasma: database of circulating human peptides. *J. Chromatogr., B: Biomed. Sci. Appl.* **1999**, *726*, 25–35.
(33) Tanaka, Y.; et al. A novel approach and protocol for discovering extremely low-abundance proteins in serum. *Proteomics* **2006**, *6*, 4845–55.
(34) Qian, W. J.; Jacobs, J. M.; Liu, T.; Camp, D. G., II.; Smith, R. D. Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications. *Mol. Cell. Proteomics* **2006**, *5*, 1727–44.
(35) Guzzetta, A. W.; Thakur, R. A.; Mylchreest, I. C. A robust microelectrospray ionization technique for high-throughput liquid chromatography/mass spectrometry proteomics using a sanded metal needle as an emitter. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 2067–72.
(36) Tietz. *Tietz Fundamentals of Clinical Chemistry*; W. B. Saunders: Philadelphia, PA, 2001; p 1091.
(37) Marshall, J.; et al. Involvement of cytosolic phospholipase A2 and secretory phospholipase A2 in arachidonic acid release from human neutrophils. *J. Immunol.* **2000**, *164*, 2084–91.
(38) Marshall, J.; et al. Processing of serum proteins underlies the mass spectral fingerprinting of myocardial infarction. *J. Proteome Res.* **2003**, *2*, 361–72.
(39) Sviridov, D.; Meilinger, B.; Drake, S. K.; Hoehn, G. T.; Hortin, G. L. Coelution of other proteins with albumin during size-exclusion HPLC: Implications for analysis of urinary albumin. *Clin. Chem.* **2006**, *52*, 389–97.
(40) Craig, R.; Cortens, J. P.; Beavis, R. C. Open source system for analyzing, validating, and storing protein identification data. *J. Proteome Res.* **2004**, *3*, 1234–42.
(41) Codd, E. F. A relational model of data for large shared data banks. 1970. *MD Comput.* **1998**, *15*, 162–6.
(42) Field, H. I.; Fenyo, D.; Beavis, R. C. RADARS, a bioinformatics solution that automates proteome mass spectral analysis, optimises protein identification, and archives data in a relational database. *Proteomics* **2002**, *2*, 36–47.
(43) Altschul, S. F.; et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–402.
(44) Anderson, L.; Hunter, C. L. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* **2006**, *5*, 573–88.
(45) Zhang, R.; et al. Mining biomarkers in human sera using proteomic tools. *Proteomics* **2004**, *4*, 244–56.
(46) Whiteaker, J. R.; et al. Antibody-based enrichment of peptides on magnetic beads for mass-spectrometry-based quantification of serum biomarkers. *Anal. Biochem.* **2007**, *362*, 44–54.
(47) Ghosh, S.; Gepstein, S.; Heikkila, J. J.; Dumbroff, E. B. Use of a scanning densitometer or an ELISA plate reader for measurement of nanogram amounts of protein in crude extracts from biological tissues. *Anal. Biochem.* **1988**, *169*, 227–33.
(48) Schagger, H.; von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **1987**, *166*, 368–79.
(49) Schagger, H.; Aquila, H.; Von Jagow, G. Coomassie blue-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for direct visualization of polypeptides during electrophoresis. *Anal. Biochem.* **1988**, *173*, 201–5.
(50) Dionne, R.; Forest, J. C.; Moutquin, J. M.; De Grandpre, P.; Masse, J. Electrophoretic method for separating small peptides in serum without extraction of macromolecules: application to the detection of preeclampsia. *Clin. Biochem.* **1994**, *27*, 99–103.
(51) Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970**, *227*, 680–5.
(52) Jin, X.; Chen, Y.; Lubman, D. M.; Misek, D.; Hanash, S. M. Capillary electrophoresis/tandem mass spectrometry for analysis of proteins from two-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 2327–34.
(53) Wheeler, D. L.; et al. Database resources of the National Center for Biotechnology. *Nucleic Acids Res.* **2003**, *31*, 28–33.
(54) Pruitt, K. D.; Tatusova, T.; Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **2007**, *35*, D61−5.
(55) Maglott, D. R.; Katz, K. S.; Sicotte, H.; Pruitt, K. D. NCBI's LocusLink and RefSeq. *Nucleic Acids Res.* **2000**, *28*, 126–8.
(56) Link, A. J.; et al. Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999**, *17*, 676–82.
(57) Craig, R.; Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20*, 1466–7.
(58) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **2003**, *21*, 660–6.
(59) Cargile, B. J.; Bundy, J. L.; Stephenson, J. L., Jr. Potential for false positive identifications from large databases through tandem mass spectrometry. *J. Proteome Res.* **2004**, *3*, 1082–5.
(60) Kapp, E. A.; et al. An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis. *Proteomics* **2005**, *5*, 3475–90.
(61) Brosch, M.; Swamy, S.; Hubbard, T.; Choudhary, J. Comparison of mascot and X!tandem performance for low and high accuracy mass spectrometry and the development of an adjusted mascot threshold. *Mol. Cell. Proteomics* **2008**, .
(62) Anderson, N. L.; Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **2003**, *2*, 50.
(63) Wait, R. Proteins of rat serum, urine, and cerebrospinal fluid: VI. Further protein identifications and interstrain comparison. *Electrophoresis* **2001**, *22*, 3043–52.
(64) Bergquist, J.; Palmblad, M.; Wetterhall, M.; Hakansson, P.; Markides, K. E. Peptide mapping of proteins in human body fluids using electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Mass Spectrom. Rev.* **2002**, *21*, 2–15.
(65) Adachi, J.; Kumar, C.; Zhang, Y.; Olsen, J. V.; Mann, M. The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *GenomeBiology* **2006**, *7*, R80.
(66) Hortin, G. L. The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. *Clin. Chem.* **2006**, *52*, 1223–37.

(67) Kulasingam, V.; et al. "Product ion monitoring" assay for prostate-specific antigen in serum using a linear ion-trap. *J. Proteome Res.* **2008**, *7*, 640–647.
(68) Keshishian, H.; Addona, T.; Burgess, M.; Kuhn, E.; Carr, S. A. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* **2007**, *6*, 2212–29.
(69) Hager, J. W.; Yves Le Blanc, J. C. Product ion scanning using a Q-q-Qlinear ion trap (Q TRAPTM) mass spectrometer. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1056–64.
(70) Brenner, S.; et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **2000**, *18*, 630–4.
(71) Cohen, S. N.; Chang, A. C. A method for selective cloning of eukaryotic DNA fragments in Escherichia coli by repeated transformation. *Mol. Gen. Genet.* **1974**, *134*, 133–41.
(72) Istrail, S.; et al. Whole-genome shotgun assembly and comparison of human genome assemblies. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 1916–21.
(73) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5463–7.
(74) Schena, M. Genome analysis with gene expression microarrays. *BioEssays* **1996**, *18*, 427–31.
(75) Smith, L. M.; et al. Fluorescence detection in automated DNA sequence analysis. *Nature* **1986**, *321*, 674–9.
(76) Mullis, K. B. Target amplification for DNA analysis by the polymerase chain reaction. *Ann. Biol. Clin.* **1990**, *48*, 579–82.
(77) Zhang, S.; Zhang, J.; Horvath, C. Rapid separation of peptides and proteins by isocratic capillary electrochromatography at elevated temperature. *J. Chromatogr., A* **2001**, *914*, 189–200.
(78) Nissum, M.; et al. Two-dimensional separation of human plasma proteins using iterative free-flow electrophoresis. *Proteomics* **2007**, *7*, 4218–27.
(79) Corr, J. J.; et al. Design considerations for high speed quantitative mass spectrometry with MALDI ionization. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1129–41.
(80) Shilov, I. V.; et al. The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics* **2007**, *6*, 1638–55.
(81) Tucholska, M.; et al. Endogenous peptides from biophysical and biochemical fractionation of serum analyzed by matrix-assisted laser desorption/ionization and electrospray ionization hybrid quadrupole time-of-flight. *Anal. Biochem.* **2007**, *370*, 228–45.
(82) Ping, P.; et al. A functional annotation of subproteomes in human plasma. *Proteomics* **2005**, *5*, 3506–19.
(83) Bjorhall, K.; Miliotis, T.; Davidsson, P. Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics* **2005**, *5*, 307–17.

PR8005217