Evaluation of Multiprotein Immunoaffinity Subtraction for Plasma Proteomics and Candidate Biomarker Discovery Using Mass Spectrometry*

Tao Liu, Wei-Jun Qian, Heather M. Mottaz, Marina A. Gritsenko, Angela D. Norbeck, Ronald J. Moore, Samuel O. Purvine, David G. Camp II, and Richard D. Smith‡

Strategies for removal of high abundance proteins have been increasingly utilized in proteomic studies of serum/plasma and other body fluids to enhance the detection of low abundance proteins and achieve broader proteome coverage; however, both the reproducibility and specificity of the high abundance protein depletion process still represent common concerns. Here we report a detailed evaluation of immunoaffinity subtraction performed applying the ProteomeLab IgY-12 system that is commonly used in human serum/plasma proteome characterization in combination with high resolution LC-MS/MS. Plasma samples were repeatedly processed using this approach, and the resulting flow-through fractions and bound fractions were individually analyzed for comparison. The removal of target proteins by the immunoaffinity subtraction system and the overall process was highly reproducible. Non-target proteins, including one spiked protein standard (rabbit glyceraldehyde-3-phosphate dehydrogenase), were also observed to bind to the column at different levels but also in a reproducible manner. The results suggest that multiprotein immunoaffinity subtraction systems can be readily integrated into quantitative strategies to enhance detection of low abundance proteins in biomarker discovery studies. Molecular & Cellular Proteomics 5:2167–2174, 2006.

Regardless of the success of single protein depletion schemes (e.g. removing only albumin or immunoglobulin G), biomarker detection is usually facilitated by removal of as many high abundance proteins as possible. Several antibody-based approaches aimed at simultaneously removing multiple high abundance proteins have been developed to achieve a more comprehensive survey of the human plasma proteome. For example, Pieper et al. (2) reported an approach that is capable of removing 10 high abundance proteins in a single step, and two other recently commercialized systems offer the capability of simultaneously removing six (3) and 12 (4) of the most abundant proteins. These antibody-based depletion systems have been demonstrated to be highly efficient in removing the specifically targeted proteins. However, questions still arise with regard to their reproducibility and selectivity (3–8). In particular there are two primary questions: whether these systems are capable of removing their target proteins in a reproducible fashion and whether there is any loss of non-target proteins along with the high abundance proteins through non-selective binding or physiologically relevant association/interaction. Ideally potential losses of non-target proteins should be minimized during removal of multiple high abundance proteins, but in cases where losses do occur, such losses should be reproducible if the depletion strategy is to be applied for comparative quantitative biomarker discovery studies.

To address these concerns, we evaluated the performance of one widely used immunoaffinity subtraction system, the Beckman Coulter ProteomeLab™ IgY-12 system, for separation of the protein mass of this biofluid (1). Hence removal of these high abundance proteins in the serum/plasma sample preparation process is becoming increasingly widespread to provide higher sensitivity for achieving broader proteome coverage, particularly of low abundance proteins that are normally present in the concentration range of ng/ml and lower.

Human body fluids, especially blood plasma and serum, serve as the most important and readily available sources for discovering candidate disease biomarkers. However, the detection of novel protein biomarkers typically present at low concentrations is hampered by the “masking” effect caused by a number of highly abundant proteins. For example, the 22 most abundant proteins are responsible for ~99% of the bulk mass of the total proteins in human plasma; this leaves perhaps hundreds of thousands of other proteins in only ~1% of the biological mass of this biofluid (1). Hence removal of these high abundance proteins in the serum/plasma sample preparation process is becoming increasingly widespread to provide higher sensitivity for achieving broader proteome coverage, particularly of low abundance proteins that are normally present in the concentration range of ng/ml and lower.

From the Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352
Received, April 3, 2006, and in revised form, July 18, 2006
Published, MCP Papers in Press, July 19, 2006, DOI 10.1074/mcp.T600039-MCP200

Molecular & Cellular Proteomics 5.11 2167

This paper is available online at http://www.mcponline.org

†The abbreviations used are: IgY, immunoglobulin yolk; 2D, two-dimensional; G3PDH, glyceraldehyde-3-phosphate dehydrogenase; SPE, solid-phase extraction; ΔCn, delta correlation; Xcorr, cross-correlation score; MARS, multiple affinity removal system; D, distribution.
rating high abundance proteins. Our results indicate that high abundance protein separation using the IgY-12 system is highly reproducible, and the low level binding of non-target proteins in this system occurred in a reproducible fashion.

**EXPERIMENTAL PROCEDURES**

Approval for the conduct of this study was obtained from the Institutional Review Boards of the Stanford University School of Medicine and the Pacific Northwest National Laboratory in accordance with federal regulations.

**Plasma Sample**—The human blood plasma sample supplied by the Stanford University School of Medicine (Palo Alto, CA) was obtained from a single, healthy volunteer. An initial plasma protein concentration of 65 mg/ml was determined by the Coomassie protein assay (Pierce). To generate a spiked plasma sample, bovine carbonic anhydrase II, rabbit glyceraldehyde-3-phosphate dehydrogenase (G3PDH), and bovine β-lactoglobulin (Sigma) were added to unprocessed human plasma samples to final concentrations of 40 μg/ml, and equine skeletal muscle myoglobin and chicken ovalbumin (Sigma) were added to final concentrations of 4 μg/ml. Unless otherwise noted, all protein sample processing was performed at 4 °C.

**Immunoadfinity Subtraction**—Twelve high abundance plasma proteins (albumin, IgG, α1-antitrypsin, IgA, IgM, transferrin, haptoglobin, α1-acid glycoprotein, α2-macroglobulin, apolipoprotein A-I, apolipoprotein A-II, and fibrinogen) that constitute ~95% of the total protein mass of human plasma were simultaneously separated from other proteins using a packed 7 × 52-mm (loading capacity, 25 μl of plasma) IgY-12 affinity LC column (Beckman Coulter, Fullerton, CA; previously known as Seppro™ MIXED12 column from GenWay BioTech, San Diego, CA) using an Agilent 1100 series HPLC system (Agilent, Palo Alto, CA). All IgY-12 separations were performed within the manufacturer-instructed column usage and loading capacity. Three buffers (dilution/washing buffer: 10 mM Tris-HCl, 150 mM NaCl, pH 7.4 (TBS); stripping buffer: 100 mM glycine, pH 2.5; neutralization buffer: 100 mM Tris-HCl, pH 8.0) were used in a separation scheme that consisted of sample loading-washing-eluting-neutralization followed by a re-equilibration scheme for a total cycle time of 48 min. The spiked human plasma sample was subjected to five individual IgY-12 separations, and the resulting flow-through fraction and the bound/eluted fraction were collected separately. To generate sufficient samples for subsequent two-dimensional (2D) LC-MS/MS analysis, a non-spiked plasma sample was repeatedly processed eight times, and corresponding flow-through and bound fractions were pooled. All fractions were then individually concentrated in iCon tech, San Diego, CA using an Agilent 1100 series HPLC system at a flow rate of 200 μl/min. Exponential gradient elution was performed by increasing the mobile phase composition from 0 to 70% B over 150 min. To identify the eluting peptides, the linear ion trap mass spectrometer was operated in a data-dependent MS/MS mode (LTQ; ThermoElectron) via an electrospray ionization interface manufactured in-house. The reversed-phase capillary column was prepared by slurry-packing 3-μm Jupiter C18 bonded particles (Phenomenex, Torrance, CA) into a 65-cm-long, 150-μm-inner diameter × 360-μm-outlet diameter fused silica capillary (Polymeric Technologies, Phoenix, AZ). The mobile phase consisted of 0.2% acetic acid and 0.05% TFA in water (A) and 0.1% TFA in 90% ACN, 10% water (B). After loading 10 μl of peptides onto the column, the mobile phase was held at 100% A for 20 min. Exponential gradient elution was performed by increasing the mobile phase composition from 0 to 70% B over 150 min. To identify the eluting peptides, the linear ion trap mass spectrometer was operated in a data-dependent MS/MS mode (m/z 400–2000) in which a full MS scan was followed by 10 MS/MS scans. The 10 most intensive precursor ions were dynamically selected in the order of highest intensity to lowest intensity and subjected to collision-induced dissociation using a normalized collision energy setting of 35%. A dynamic exclusion duration of 1 min was used. The temperature of the heated capillary and the ESI voltage were 200 °C and 2.2 kV, respectively.

**Data Analysis**—The SEQUEST algorithm (part of the Bioworks software package, Version 3.1 SR1; ThermoFinnigan) was used to search all MS/MS spectra independently against the human International Protein Index (IPI) database (Version 3.05 that consists of 49,161 protein entries; available on line at www.ebi.ac.uk/IPI) supplied with the sequences of the five protein standards and the corresponding reversed human IPI database with no enzyme constraint. Dynamic carboxamidomethylation of cysteine and oxidation of methionine were used during the database search. The reversed human protein database was created as reported previously (10) by reversing the order of the amino acid sequences for each protein. The false discovery rates of peptide identifications were estimated as described previously by dividing the number of unique peptides from the reversed database search by the number of unique peptides from the normal database search (10).

Criteria that would yield an overall confidence of over 95% for peptide identification at the unique peptide level were established for filtering raw peptide identifications. For example, with delta correlation (ΔCn) value of ≥0.1, the following cross-correlation score (Xcorr) cutoffs were used: for the +1 charge state, Xcorr ≥ 1.5 for fully tryptic peptides and Xcorr ≥ 3.0 for partially tryptic peptides; for the +2 charge state, Xcorr ≥ 2.7 for fully tryptic peptides and Xcorr ≥ 3.7 for partially tryptic peptides; and for the +3 charge state, Xcorr ≥ 3.3 for fully tryptic peptides and Xcorr ≥ 4.5 for partially tryptic peptides. Non-tryptic peptides were not considered. Two additional ΔCn cutoff values of 0.05 and 0.16 were applied to reduce false negative iden-
iffications while maintaining a 95% level of confidence for peptide assignments. For $\Delta Cn \geq 0.05$, the minimum acceptable $Xcorr$ value was raised to achieve a comparable percentage of false positive rate identifications, and similarly for $\Delta Cn \geq 0.16$, the minimum acceptable $Xcorr$ value was reduced. In an attempt to remove redundant protein entries in the reported results, ProteinProphet software was used as a clustering tool to group similar or related protein entries into a "protein group" (11). All unique peptides that passed the filtering criteria were assigned an identical peptide probability score of 1 and entered into the software program (solely for clustering analysis) to generate a final list of nonredundant proteins or protein groups. One protein identification was randomly selected to represent each corresponding protein group that contains member database entries.

RESULTS

Reproducibility of the Immunoaffinity Separations—Although the reproducibility of the immunoaffinity subtraction systems for separating target high abundance proteins from other proteins has been previously tested using conventional gel-based techniques (e.g. SDS-PAGE and two-dimensional gel electrophoresis), the reproducibility for detecting low abundance proteins and the potential nonspecific binding of proteins has remained unclear. To clarify this issue, we evaluated the reproducibility of the IgY-12 column using a plasma sample spiked with five protein standards and high resolution capillary LC coupled to linear ion trap tandem MS (LC-MS/MS), a technique capable of identifying thousands of peptides in a single analysis. Separations with the IgY-12 column were repeated five times, and the corresponding flow-through fraction and bound fraction were individually analyzed under the same conditions.

In the LC-MS/MS analyses of complex peptide mixtures, the selection of ions for MS/MS is not totally randomized but rather dependent on the width of the chromatographic peaks or the concentration of peptides delivered to the mass spectrometer. Thus, data acquisition can be biased against the low abundance species, and peptide ions from more abundant proteins are selected more frequently for MS/MS (12). The amount of proteins in the original plasma sample. An average of 5.5 and 75.4% of proteins were recovered in the flow-through and bound fractions, respectively.

In the LC-MS/MS analyses of complex peptide mixtures, the selection of ions for MS/MS is not totally randomized but rather dependent on the width of the chromatographic peaks or the concentration of peptides delivered to the mass spectrometer. Thus, data acquisition can be biased against the low abundance species, and peptide ions from more abundant proteins are selected more frequently for MS/MS (12). The number of spectra acquired for a given protein (spectral count) was shown to accurately reflect its relative abundance with a linear correlation over 2 orders of magnitude of linear dynamic range (12). It has also been demonstrated that spectral counting and MS-derived peak areas strongly correlate for determining quantitative changes in protein expression (13, 14). In this study, raw peptide identifications were filtered using a set of stringent criteria providing >95% confidence at the unique peptide level, and the spectral counts for each protein were then calculated by summing the numbers of observations of all filtered peptides derived from the protein for comparison of relative protein abundances in the five replicate analyses.

An average of 54 and 122 proteins were identified from the IgY-12-bound fraction and flow-through fraction, respectively (only proteins with two or more unique peptide identifications from all the replicate analyses are considered; the ambiguous immunoglobulin identifications were removed). The numbers of proteins identified in the five replicate analyses are fairly consistent with a coefficient of variance of less than 4% for both fractions. Fig. 1 displays linear correlations of the spectral counts between pairs of analyses among the five replicates of the flow-through fractions. Note that a correlation coefficient ($R^2$) of greater than 0.98 was obtained from each comparison, and an averaged coefficient of variance of spectral count of 17% was obtained for proteins identified from flow-through fractions in the five replicate analyses. Similar results were obtained from the replicate analyses of the bound fractions (data not shown), demonstrating the excellent reproducibility of high abundance protein separation using the IgY-12 system.

The detection of the five non-human protein standards (i.e. 40 $\mu$g/ml each of bovine carbonic anhydrase II, rabbit G3PDH, and bovine $\beta$-lactoglobulin and 4 $\mu$g/ml each of equine myoglobin and chicken ovalbumin) from the replicate LC-MS/MS analyses is summarized in Table I. Almost all of the peptides from the five non-human proteins were observed exclusively in the flow-through fraction except for rabbit G3PDH. Again high reproducibility was demonstrated based on the spectral counts of protein standards in each LC-MS/MS analysis. Although nonspecific binding of rabbit G3PDH was observed (i.e. the protein was identified in both bound and flow-through fractions), the extent of nonspecific binding for this protein was reproducible, which is key for quantitative studies.

Recoveries of sample processing were determined for both the flow-through and bound fractions (Table II). In the first stage, the amount of proteins recovered after IgY-12 separation, concentrating, and buffer exchange was compared with the amount of proteins in the original plasma sample. An average of 5.5 and 75.4% of proteins were recovered in the flow-through fraction and the bound fraction, respectively. In the second stage, 60–70% of peptides were recovered after denaturation, alkylation, tryptic digestion, and SPE cleanup. Recoveries of proteins in the flow-through and bound fractions in different stages remained consistent, which contributes to the overall high reproducibility of the analysis.

Binding of Non-target Proteins—The potential for co-depleting non-target proteins has been a major concern for all systems designed to remove high abundance proteins. The Cibacron Blue dye-based system has been reported to bind many more non-target proteins than the Agilent multiple affinity removal system (MARS), indicating the superior specificity of antibody-based systems (3, 6). It has also been reported that only four non-target proteins were detected and identified in the bound fraction of the IgY-12 system using SDS-PAGE, in-gel digestion, and LC-MS/MS (4). However, an increased level of non-selective loss of non-target proteins was observed when 2D LC-MS/MS was used to compare the unprocessed and MARS-depleted human serum samples (7). To investigate the extent of potential binding of non-target proteins...
**TABLE I**

High abundance protein separation for a plasma sample spiked with five protein standards with the IgY-12 column. Spectral counts for each protein standard from the bound and flow-through fractions are listed.

| Replicate | Bound fraction, G3PDH | Flow-through fraction |
|-----------|-----------------------|-----------------------|
|           |                       | Carbonic anhydrase II | G3PDH | β-Lactoglobulin | Myoglobin | Ovalbumin |
| 1         | 3                     | 21                    | 5     | 23             | 9         | 8         |
| 2         | 2                     | 21                    | 4     | 20             | 7         | 6         |
| 3         | 2                     | 17                    | 4     | 19             | 6         | 6         |
| 4         | 2                     | 19                    | 4     | 18             | 5         | 6         |
| 5         | 3                     | 18                    | 5     | 23             | 6         | 7         |

**FIG. 1.** Linear correlations of the spectral counts between pairs of analyses among the five replicates of the flow-through fractions. Correlation coefficient ($R^2$) is shown for each comparison.
proteins to the IgY-12 system in more detail, we performed 1) five replicate LC-MS/MS analyses of individually prepared bound and flow-through fractions and 2) 2D LC-MS/MS analysis of bound and flow-through fractions from the IgY-12 system.

A total of 38 bound non-target proteins were identified in the replicate LC-MS/MS analyses of the five individually prepared bound and flow-through fractions of the IgY-12 column. To compare the extent of nonspecific binding in an individual high abundance protein separation, a distribution (D)-value was calculated by dividing the spectral counts for each protein from the bound fraction by the sum of spectral counts for the corresponding protein from the bound fraction and from the flow-through fraction. The D-value serves as an indicator for evaluating the reproducibility of protein partition between the bound and flow-through fractions. A larger D-value suggests relatively significant nonspecific binding. The D-values of each protein in the replicate analyses are listed in Table III; the numbers of unique peptides for specific proteins identified from the bound fraction and from the flow-through fraction are also provided. Details of these peptide and protein identifications in the replicate LC-MS/MS analyses can be found in Supplemental Table 2. A number of proteins bound significantly to the IgY-12 column, such as zinc-α2-glycoprotein, apolipoproteins (C-I, C-III, D, E, and L1), CD5 antigen-like protein, serum amyloid proteins A2 and A4, transthyretin, hemoglobin, serum paraoxonase/arylesterase 1, and leucine-rich α2-glycoprotein (Table III). One important finding from this set of analyses is that, although many non-target proteins were observed to bind to the IgY-12 systems at different levels, the fraction of binding for any specific non-target protein to the column appeared to be very reproducible (Table III).

Binding of one of the five spiked protein standards (rabbit G3PDH) was also observed (Table II), again in a reproducible manner.

The 2D LC-MS/MS analysis of the bound and flow-through IgY-12 fractions revealed a larger number of non-target protein identifications, presumably due to the increased dynamic range in detection. A total of 108 and 311 proteins (without counting immunoglobulin identifications) were identified from the bound fraction and the flow-through fraction, respectively, with at least two unique peptides identified from either fraction. These proteins along with their unique peptide counts and spectral counts in both the bound and flow-through fractions are listed in Supplemental Table 2. Details of the peptides and proteins identified from the bound and flow-through fractions are provided in Supplemental Table 3 and Supplemental Table 4, respectively. A total of 91 proteins were detected in common to both fractions, including various proteins from several major plasma protein groups, e.g. albumins/prealbumins, apolipoproteins, complement components and factors, coagulation factors, serum amyloid proteins, and protease inhibitors. Some of the proteins are similar in sequence to the target proteins and bound at high levels (e.g. apolipoproteins). Interestingly among the coagulation factors and complement components, a few bound at high levels (e.g. coagulation factor XIII B chain, C1q subcomponent C chain, and complement factor H-related protein), whereas the majority bound at fairly low levels or not at all, suggesting differential binding properties of these molecules and/or differences in physiological associations with the directly captured molecules. A total of 17 non-target proteins were observed only in the bound fraction (i.e. 100% removal), such as transthyretin (19 unique peptides), zinc-α2-glycoprotein (12 unique peptides), and CD5 antigen-like protein (11 unique peptides). A significant portion of proteins were observed to be common in both the bound and flow-through fractions based on both one-dimensional and 2D LC-MS/MS, suggesting that many non-target proteins bind to the columns to a certain extent either directly or indirectly. Meanwhile although the 12 target proteins were removed with high efficiency as reported previously (4) (Supplemental Table 5), none of them appeared to be removed at 100% efficiency when 2D LC-MS/MS was used for the analysis (Supplemental Table 2), presumably due to their high original concentration and the increased sampling depth of the flow-through fraction.

**DISCUSSION**

The development and application of advanced methods for specifically and efficiently removing high abundance proteins from various biofluids, in particular serum/plasma, is being broadly pursued to enhance discovery of candidate disease biomarkers through the application of proteomics-based technologies. Antibody-based immunoaffinity subtraction systems are at the forefront of removal methods due to their high efficacy, e.g. a new immunoaffinity subtraction system that removes the top 20 abundant proteins from serum/plasma (~99% of its protein mass) in a single step recently.
Evaluation of Immunoaffinity Subtraction

Identification of non-target human plasma proteins that bind to the IgY-12 column using replicate LC-MS/MS analysis

To evaluate the reproducibility and estimate the extent of nonspecific binding, a D-value was calculated for each protein by dividing the spectral counts for each protein from the bound fraction (B) by the sum of spectral counts for corresponding protein from the bound fraction and from the flow-through fraction (FT), i.e. B/(B+FT). Numbers of unique peptide identifications for specific proteins are listed in parentheses: numbers to the left of the comma are from the bound fraction, and numbers to the right of the comma are from the flow-through fraction. D-values in individual LC-MS/MS analysis (D1–5), averaged D-values (Avg), and S.D. of specific proteins are listed. HS, Heremans Schmid. AMBP, α₁-microglobulin/bikunin precursor.

| IPI        | Protein name                        | D1     | D2     | D3     | D4     | Avg   | S.D. |
|------------|-------------------------------------|--------|--------|--------|--------|-------|------|
| IPI00019943| Afamin                              | 0.21   | 0.4    | 0.11   | 0.2    | 0.22  | 0.1  |
| IPI00550991| α₁-Antichymotrypsin                 | 0.01   | 0.05   | 0.03   | 0.05   | 0.03  | 0.01 |
| IPI00022895| α₁-B-glycoprotein                   | 0.02   | 0.02   | 0.05   | 0.01   | 0.01  | 0.03 |
| IPI00029863| α₂-Antiplasmin                      | 0.07   | 0.16   | 0.11   | 0.11   | 0.14  | 0.11 |
| IPI00166729| α₂-Glycoprotein 1, zinc             | 1      | 1      | 1      | 1      | 1     | 1    |
| IPI00022431| α₂-HS-glycoprotein                  | 0.44   | 0.34   | 0.36   | 0.29   | 0.3   | 0.34 |
| IPI00022426| AMBP protein                        | 0.07   | 0.06   | 0.07   | 0.09   | 0.11  | 0.07 |
| IPI00032220| Angiotensinogen                     | 0      | 0.03   | 0.04   | 0.04   | 0.04  | 0.03 |
| IPI00032179| Antithrombin III variant            | 0.06   | 0.14   | 0.15   | 0.15   | 0.13  | 0.11 |
| IPI00022229| Apolipoprotein B-100                | 0.19   | 0.12   | 0.15   | 0.15   | 0.16  | 0.15 |
| IPI00021855| Apolipoprotein C-I                  | 1      | 1      | 1      | 1      | 1     | 1    |
| IPI00021857| Apolipoprotein C-III                | 1      | 1      | 1      | 1      | 1     | 1    |
| IPI00006662| Apolipoprotein D                    | 1      | 1      | 1      | 1      | 1     | 1    |
| IPI00021842| Apolipoprotein E                    | 0.87   | 0.75   | 0.88   | 0.81   | 0.87  | 0.83 |
| IPI00177869| Apolipoprotein L1 isofrom A         | 1      | 1      | 1      | 1      | 1     | 1    |
| IPI00025204| CDS antigen-like                    | 1      | 1      | 1      | 1      | 1     | 1    |
| IPI00017601| Ceruloplasmin                       | 0.01   | 0.01   | 0.01   | 0.01   | 0.01  | 0.01 |
| IPI00291262| Clusterin                           | 0.42   | 0.28   | 0.33   | 0.27   | 0.28  | 0.31 |
| IPI00164623| Complement C3                       | 0.1    | 0.09   | 0.11   | 0.11   | 0.11  | 0.11 |
| IPI00032258| Complement C4                       | 0.06   | 0.07   | 0.14   | 0.09   | 0.07  | 0.07 |
| IPI00218746| Complement component 1, Q subcomponent | 0.14  | 0.11  | 0.11  | 0.11  | 0.25  | 0.16 |
| IPI00029739| Complement factor H                 | 0.03   | 0.03   | 0.01   | 0.02   | 0.04  | 0.02 |
| IPI00027482| Corticosteroid-binding globulin     | 0.5    | 0.33   | 0.33   | 0.33   | 0.75  | 0.52 |
| IPI00022418| Fibronectin                         | 0      | 0      | 0.02   | 0.05   | 0.02  | 0.03 |
| IPI00218816| Hemoglobin β chain                  | 1      | 1      | 1      | 1      | 1     | 1    |
| IPI00022488| Hemopexin                           | 0.11   | 0.06   | 0.08   | 0.06   | 0.05  | 0.07 |
| IPI00022371| Histidine-rich glycoprotein         | 0.1    | 0.18   | 0.25   | 0.24   | 0.13  | 0.12 |
| IPI00218192| Inter-α -trypsin inhibitor heavy chain | 0    | 0      | 0.44   | 0.44   | 0.14  | 0.01 |
| H4         | IPI00022417| Leucine-rich α 2-glycoprotein       | 0.9    | 0.8    | 0.87   | 0.8    | 0.77  | 0.82 |
| IPI00019580| Plasminogen                         | 0.02   | 0.04   | 0.04   | 0.04   | 0.04  | 0.03 |
| IPI00025426| Pregnancy zone protein              | 0.68   | 0.75   | 0.61   | 0.7    | 0.64  | 0.67 |
| IPI00165421| SERPIN1 protein                     | 0.1    | 0.05   | 0.06   | 0.33   | 0.22  | 0.21 |
| IPI00006146| Serum amyloid A2                    | 1      | 1      | 1      | 1      | 1     | 1    |
| IPI00019399| Serum amyloid A4                    | 1      | 1      | 1      | 1      | 1     | 1    |
| IPI00021832| Serum paraoxonase/arylesterase 1    | 1      | 0.9    | 1      | 1      | 1     | 0.98 |
| IPI00022432| Transthyretin                        | 1      | 1      | 1      | 1      | 1     | 1    |
| IPI00029885| Vitamin D-binding protein           | 0      | 0      | 0.28   | 0.28   | 0.07  | 0.01 |
| IPI00298971| Vitronecin                          | 0.13   | 0.11   | 0.2    | 0.2    | 0.12  | 0.14 |

became commercially available (Sigma ProteoPrep® 20 plasma immunodepletion kit), and LC columns that remove 50–100 proteins are anticipated in the near future. The reproducibility and selectivity for such systems, i.e. the risk of losing potential proteins of interest along with the target proteins, is critical for these systems to be broadly applied for quantitative candidate biomarker discovery or verification studies.

In this study, the widely used IgY-12 immunodepletion system was evaluated based on replicate depletion experiments and high resolution LC-MS/MS and 2D LC-MS/MS analyses. The spectral counting approach used here has been demonstrated previously to be applicable to comparative proteomic analysis (12–14). The high abundance protein separations were determined to be very reproducible based on the analyses of replicate flow-through fractions using LC-MS/MS and spectral counts (Fig. 1); similar results were obtained for the bound fractions. High reproducibility was also observed based on the direct measurements of protein recoveries at key sample processing steps (Table II) where an average of 5.5% of proteins were recovered from the flow-through fraction, consistent with the 95% protein mass removal estimated.
from known concentrations of the 12 target proteins in human plasma and the 95–99.5% depletion efficiency of the IgY-12 system determined in previous studies using SDS-PAGE and ELISA (4). The averaged recovery of flow-through proteins is similar to what was determined in a recent large scale plasma proteome profiling study (5.6%) utilizing the same IgY-12 system (15). Moreover high abundance protein separation on the IgY-12 column did not significantly deteriorate within the manufacturer-defined usage and sample load (data not shown). These results demonstrated that the performance of the IgY-12 column is highly consistent, and the sample processing procedures used in this study are robust and reproducible, thus suitable for comparative proteomic studies for candidate biomarker discovery. Recently a study reported that reproducible and high efficiency depletion can be obtained from the MARS column as determined by using ELISA (8). This complementary study demonstrates yet another example that the selective depletion of abundant plasma proteins by immunoaffinity chromatography is a valid approach for the enrichment of potential biomarkers sought by proteomics-based methodologies.

Estimating the percentage of protein partition between the bound and flow-through fractions is challenging due to the drastic component differences between the two fractions and nearly 20-fold difference in protein quantities as a result of the immunoaffinity separation itself. The calculated D-value does not reflect the true percentage of protein partition in the bound and flow-through fractions but still serves as a good indicator of the reproducibility of the nonspecific binding; the alternation of spectral counts for both the bound and flow-through fractions in the replicate analyses should be consistent if the high abundance protein separation process is reproducible. However, when the spectral counts from both fractions are small, the calculated D-values will be less reproducible due to the undersampling nature of tandem mass spectrometry measurements.

A relatively large number of non-target proteins that bind to the IgY-12 system have regions of amino acid sequence homologous to target proteins, e.g. apolipoproteins (homologous to apolipoproteins A-I and A-II), and these epitopes may be bound directly to the polyclonal antibodies. Furthermore many abundant proteins may bind to the column through nonspecific interactions with the bead surfaces; other proteins are most likely binding to one or more of the affinity column-bound target proteins (that act as “bait”) through mechanisms of specific and physiologically relevant protein-protein interactions under certain conditions, i.e. the effect of “interactome” (16, 17). The proteins identified from both the bound and flow-through fractions may provide interesting information related to specific protein-protein interactions (Table III and Supplemental Table 2). For example, hemoglobin and complement C3 are reported to bind to haptoglobin and IgG, respectively (18). Zinc-α2-glycoprotein, a protein that binds 100% to IgY-12 column (Table III and Supplemental Table 2), may interact with the targeted apolipoproteins and/or interact directly with the Fc region of IgG because it is a lipid-mobilizing factor similar to the Fc receptor and belongs to the Class I major histocompatibility complex immunoglobulin superfamily (19, 20). Transthyretin (prealbumin, 20–40 μg/ml), another relatively abundant protein, also binds to the IgY-12 column (Table III and Supplemental Table 2) possibly due to protein-protein interactions with the immunoaffinity-bound target proteins because both transthyretin and albumin are major transport proteins for the hydrophobic thyroid hormones. Alternatively a contaminating antigen used for antibody generation in the IgY-12 system may be present (4) because transthyretin binds to completion to the IgY-12 column but only at very low levels to the MARS column (data not shown). One of the five spiked protein standards, rabbit G3PDH (spiked at 40 μg/ml level), also reproducibly binds to the IgY-12 system (Table II) in the five replicate LC-MS/MS analyses. However, a survey of the non-spiked plasma sample using 2D LC-MS/MS revealed that human G3PDH does not bind to the IgY-12 system at all at its normal plasma concentration. Six G3PDH unique peptides were identified only from the flow-through fraction, and among them three peptides are unique to human G3PDH (Supplemental Table 4), although the protein sequences of rabbit and human G3PDH have 94% in common. These results suggest that the binding of G3PDH to the IgY-12 system may be concentration-dependent; however, the exact mechanism remains unclear.

Mild denaturing conditions (e.g. addition of 5–20% acetonitrile) may help to disrupt the binding of some proteins, particularly low molecular weight proteins, to the carrier proteins, thus decreasing the possibility of co-depleting non-target proteins of interest (21, 22). Although more stringent washing conditions may assist with the dissociation of certain tightly bound proteins, it is compromised by the need to maximize binding of target proteins to the depletion systems to achieve high efficiency immunodepletion. It is likely unavoidable that some of the non-target proteins will bind to the immunoaffinity subtraction systems without compromising the optimized depletion conditions. A practical solution to this scenario would be to characterize both the bound and flow-through fractions, i.e. use these systems as a method for “partitioning” instead of “depletion” so that virtually no protein of interest would be lost along the high abundance protein separation process. This approach would also create an opportunity for investigating the subinteractomes in the blood using individual antibody resin, e.g. albuminome, haptoglobinome, apolipoproteinome, etc. (4). An important observation made in this study is that the partitioning of any specific protein appeared to be very reproducible across replicate analyses regardless of whether some non-target proteins bound at varying levels, indicating the overall high reproducibility of the IgY-12 system as well as the sample processing procedures used in this study.

In summary, the present study demonstrated that immuno-
affinity subtraction systems, exemplified by the widely used IgY-12 system, can reproducibly and efficiently remove targeted high abundance proteins. Binding of non-target proteins to the system does occur but in a reproducible fashion under controlled conditions. These findings suggest that these immunoaffinity systems can be readily integrated into an analytical strategy for broader and deeper quantitative proteomic analysis without complicating the overall accuracy of the measurements. As a result, these systems should find broad application in candidate disease biomarker discovery or verification using clinically obtainable biofluids as samples.

* Portions of this research were supported by NIGMS, National Institutes of Health, Large Scale Collaborative Research Grant U54 GM-62119-02 and National Institutes of Health Health National Center for Research Resources Grant RR18522. Work was performed in the Environmental Molecular Science Laboratory, a United States Department of Energy (DOE) national scientific user facility located on the campus of Pacific Northwest National Laboratory (PNNL) in Richland, WA. PNNL is a multiprogram national laboratory operated by Battelle Memorial Institute for the DOE under Contract DE-AC05-76RLO-1830. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, P. O. Box 999, MSIN: K8-98, Richland, WA 99352. E-mail: rds@pnnl.gov.

The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

Evaluation of Immunoaffinity Subtraction

REFERENCES
1. Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: history, character, and diagnostic prospects. Mol. Cell. Proteomics 1, 845–867
2. Pieper, R., Su, Q., Gattin, C. L., Huang, S. T., Anderson, N. L., and Steiner, S. (2003) Multi-component immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome. Proteomics 3, 422–434
3. Zolotoraja, N., Martosella, J., Nicol, G., Bailey, J., Boyes, B. E., and Barrett, W. C. (2005) Differences among techniques for high-abundant protein depletion. Proteomics 5, 3304–3313
4. Huang, L., Harvie, G., Fettelson, J. S., Gramatikoff, K., Herold, D. A., Allen, D. L., Amungama, R., Hagler, R. A., Pisano, M. R., Zhang, W. W., and Fang, X. (2005) Immunoaffinity separation of plasma proteins by IgY microbeads: meeting the needs of proteomic sample preparation and analysis. Proteomics 5, 3314–3328
5. Chromy, B. A., Gonzales, A. D., Perkins, J., Choi, M. W., Corzett, M. H., Chang, B. C., Corzett, C. H., and McCutchen-Malone, S. L. (2004) Proteomic analysis of human serum by two-dimensional differential gel electrophoresis after depletion of high-abundant proteins. J. Proteome Res. 3, 1120–1127
6. Echan, L. A., Tang, H. Y., Al-Khan, N., Lee, K., and Speicher, D. W. (2005) Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. Proteomics 5, 3292–3303
7. Yocum, A. K., Yu, K., Oe, T., and Blair, I. A. (2005) Effect of immunoaffinity depletion of human serum during proteomic investigations. J. Proteome Res. 4, 1722–1731
8. Brand, J., Haslberger, T., Zoigl, W., Pestiin, G., and Palme, S. (2006) Depletion efficiency and recovery of trace markers from a multiparameter immuno-depletion column. Proteomics 6, 3236–3242
9. Shen, Y., Zhao, R., Belov, M. E., Conrads, T. P., Anderson, G. A., Tang, K., Pasa-Tolic, L., Veenstra, T. D., Lipton, M. S., and Smith, R. D. (2001) Packed capillary reverse-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. Anal. Chem. 73, 1766–1775
10. Qian, W. J., Liu, T., Monroe, M. E., Strittmatter, E. F., Jacobs, J. M., Kangas, L. J., Petritis, K., Camp, D. G., II, and Smith, R. D. (2005) Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. J. Proteome Res. 4, 53–62
11. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 75, 4646–4658
12. Liu, H., Sadovyg, P. G. T., and Yates, J. R., III (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal. Chem. 76, 4193–4201
13. Zybyakov, B., Coleman, M. K., Flores, L., and Washburn, M. P. (2005) Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. Anal. Chem. 77, 6218–6224
14. Qian, J. W., Jacobs, J. M., Camp, D. G., II, Monroe, M. E., Moore, R. J., Gritsenko, M. A., Calvano, S. E., Lowry, S. F., Xiao, W., Moldawer, L. L., Davis, B. W., Tompkins, R. G., and Smith, R. D. (2005) Comparative proteome analyses of human plasma following in vivo lipopolysaccharide administration using multidimensional separations coupled with tandem mass spectrometry. Proteomics 5, 572–584
15. Liu, T., Qian, W. J., Gritsenko, M. A., Xiao, W. Z., Moldawer, L. L., Kaushal, A., Monroe, M. E., Varnum, S. M., Moore, R. J., Purvine, S. O., Maier, R. V., Davis, R. W., Tompkins, R. G., Camp, D. G., II, and Smith, R. D. High dynamic range characterization of the trauma patient plasma proteome. Mol. Cell. Proteomics 5, 1899–1913
16. Vidal, M. (2001) A biological atlas of functional maps. Cell 104, 333–339
17. Zhou, M., Lucas, D. A., Chan, K. C., Issaq, H. J., Petricoin, E. F., III, Liotta, L. A., Veenstra, T. D., and Conrads, T. P. (2004) An investigation into the human serum “interactome”. Electrophoresis 25, 1289–1298
18. Scopes, R. K. (1994) Protein Purification: Principles and Practice, 3rd Ed., Springer-Verlag, New York
19. Kennedy, M. W., Heikema, A., Cooper, A., Bjorkman, P. J., and Sanchez, L. M. (2001) Hydrophobic ligand binding by Zn-2-glycoprotein, a soluble fat-depleting factor related to major histocompatibility complex proteins. J. Biol. Chem. 276, 35008–35013
20. Bing, C., Yao, B., Jenkins, J., Sanders, P., Manieri, M., Cinti, S., Tisdale, M. J., and Trayhurn, P. (2004) Zinc-2-glycoprotein, a lipid mobilizing factor, is expressed in adipocytes and is up-regulated in mice with cancer cachexia. Proc. Natl. Acad. Sci. U. S. A. 101, 2500–2505
21. Huang, H. L., Stasiky, T., Morandell, S., Mogi, M., Schreiber, M., Feuerstein, I., Huck, C. W., Stecher, G., Bonn, G. K., and Huber, L. A. (2005) Enrichment of low-abundant serum proteins by albumin/immunoglobulin G immunoaffinity depletion under partly denaturing conditions. Electrophoresis 26, 2843–2849
22. Tirumalai, R. S., Chan, K. C., Prieto, D. A., Issaq, H. J., Conrads, T. P., and Veenstra, T. D. (2003) Characterization of the low molecular weight human serum proteome. Mol. Cell. Proteomics 2, 1096–1103