A novel four-dimensional strategy combining protein and peptide separation methods enables detection of low-abundance proteins in human plasma and serum proteomes

Hsin-Yao Tang, Nadeem Ali-Khan, Lynn A. Echan, Natasha Levenkova, John J. Rux and David W. Speicher

The Wistar Institute, Philadelphia, PA, USA

A novel strategy, termed protein array pixelation, is described for comprehensive profiling of human plasma and serum proteomes. This strategy consists of three sequential high-resolution protein prefractionation methods (major protein depletion, solution isoelectrofocusing, and 1-DE) followed by nanocapillary RP tryptic peptide separation prior to MS/MS analysis. The analysis generates a 2-D protein array where each pixel in the array contains a group of proteins with known $pI$ and molecular weight range. Analysis of the HUPO samples using this strategy resulted in 575 and 2890 protein identifications from plasma and serum, respectively, based on HUPO-approved criteria for high-confidence protein assignments. Most importantly, a substantial number of low-abundance proteins (low ng/mL – pg/mL range) were identified. Although larger volumes were used in initial prefractionation steps, the protein identifications were derived from fractions equivalent to approximately 0.6 $\mu$L (45 $\mu$g) of plasma and 2.4 $\mu$L (204 $\mu$g) of serum. The time required for analyzing the entire protein array for each sample is comparable to some published shotgun analyses of plasma and serum proteomes. Therefore, protein array pixelation is a highly sensitive method capable of detecting proteins differing in abundance by up to nine orders of magnitude. With further refinement, this method has the potential for even higher capacity and higher throughput.

Keywords:
Liquid chromatography-tandem mass spectrometry / Plasma proteome / Protein arrays / Proteomics methods / Serum proteome

1 Introduction

There is considerable interest in systematically analyzing the human plasma proteome to identify novel biomarkers that can be used for improved early diagnosis of a wide range of diseases. Plasma or serum is easily and widely collected and its proteome contains thousands of proteins including proteins secreted or shed by most cells and tissues as well as proteins that leak into the blood from damaged tissue [1].
The presence or change in concentration of blood proteins is likely to reflect the state of health of an individual. A number of proteins discovered through targeted studies are currently being used as diagnostic markers for diseases such as acute myocardial infarction (creatine kinase MB, myoglobin, and troponin T [2]), prostate cancer (prostate-specific antigen [3]), and ovarian cancer (CA125 [4]). However, it is likely that the blood contains many additional disease biomarkers that will have greater diagnostic value than the handful of biomarkers discovered.

While the human plasma proteome potentially contains many different important biomarkers for most human diseases, several factors make it difficult to characterize. Plasma proteins are present in a very wide dynamic range, varying by a factor of at least $10^{10}$ in abundance, and many of these proteins have a high degree of heterogeneous PTMs [1]. The ability to identify low-abundance plasma proteins is particularly severely limited by several major proteins that are present at >1 mg/mL. For example, albumin together with immunoglobulins contributes to more than 80% of the total plasma proteins at about 40 and 12 mg/mL, respectively [1, 5]. In contrast, many bioactive proteins and potential biomarkers of disease are low-abundance proteins that are typically found at ng/mL – pg/mL levels or less.

The strategies that have been most frequently used to overcome the dynamic range problem of plasma proteins are to fractionate the plasma proteome into smaller subsets, and/or to deplete one or more of the major proteins, particularly albumin and immunoglobulins [5–12]. Numerous dye-based and immunoaffinity methods for major protein depletion have been described and are available commercially. Immunoaffinity methods are preferred, as they provide the most efficient depletion of targeted major proteins with reduced nonspecific binding of other proteins [7–9]. Alternatively, albumin can be efficiently separated based on its pI by microscale solution IEF (MicroSol-IEF) into a single fraction [10, 11]. Both major protein depletion and MicroSol-IEF methods have resulted in increased detection of lower abundance proteins when analyzed by 2-DE [8, 9, 11]. While removal of major proteins is beneficial, multiple orthogonal fractionation steps have been used to further facilitate detection of low-abundance proteins [8, 12].

A popular alternative to 2-DE is the shotgun or multidimensional protein identification technology (MudPIT) approach which involves proteolytically digesting complex protein mixtures into peptides that are further subjected to multidimensional separations prior to analysis by ESI-MS/MS [13]. The most common form of multidimensional separations involves strong cation exchange (SCX) chromatography followed by RP-LC [12, 14, 15]. Alternate peptide separation strategies, such as ampholyte-free liquid-phase IEF [16] and CZE [17], have also been used in the analysis of human serum proteome.

Compared to 2-DE, the MudPIT approach has the potential of higher throughput and is capable of identifying more proteins from the plasma proteome. In a 2-DE study, 325 proteins were identified from human serum after 3-D fractionation using immunodepletion of nine abundant proteins, anion-exchange, and SEC [8]. In comparison, 490 proteins were identified with the MudPIT technique using immunoglobulin depletion and 2-D peptide separations by SCX and RP-LC [12]. While the 2-DE technology is relatively mature, the MudPIT method is constantly improving due to technological advances mainly to the LC and MS components of the system. In a recent study using ultra-high-performance SCX/RP-LC coupled to MS/MS, at least 800 proteins (depending on the criteria used) were identified from human plasma proteome [15]. These proteins were identified without prior depletion of major proteins, indicating that the improvement to the LC system and the longer gradient used were capable of overcoming the dynamic range problem of the plasma proteome to a certain degree. However, since immunoglobulins, which contain highly variable regions, were not depleted in the study, many of the proteins identified (up to 38%) belong to the immunoglobulin group [15].

The total number of proteins in the human plasma proteome is unknown but has been estimated to contain up to 10 000 proteins [18]. A recent analysis of the human plasma proteome by combining four separate sources of protein identification, including a 2-DE and two separate MudPIT experiments, has resulted in a conservative nonredundant list of 1175 proteins [19]. Interestingly, only 46 proteins are common to all four sources. This indicates that current methodologies cannot consistently provide comprehensive coverage of the human plasma proteome. Clearly, further reduction in the complexity of the human plasma proteome by additional more efficient fractionation steps is required to effectively mine the lower abundance proteins that have potential to be the next generation of disease biomarkers. Realizing the need for better methodology to analyze the human plasma proteome, HUPO has established the Plasma Proteome Project (PPP), and one of its aims is to determine the best technology platform for comprehensive profiling of the human plasma and serum proteomes [20].

As a participant of the HUPO PPP, in this report we describe a novel 4-D separation strategy to analyze the human plasma and serum proteomes that combines many of the benefits of 2-DE and MudPIT approaches. This strategy, termed protein array pixelation, consists of three sequential protein fractionation methods (major protein depletion, MicroSol-IEF fractionation, and SDS-PAGE). The result is a 2-D array of pixels or gel slices that is conceptually equivalent to a low-resolution 2-D gel. That is, each pixel in the array contains a group of proteins in a gel slice with a known pI and molecular weight (MW) range. Each pixel is then digested with trypsin followed by RP-LC peptide separation prior to ESI-MS/MS analysis. Using HUPO plasma and serum samples, we demonstrate that the protein array pixelation strategy is a highly sensitive method capable of detecting proteins that differ in abundance up to nine orders of magnitude.
2 Materials and methods

2.1 Materials

Human plasma (Caucasian American Sample Set; Lot # BDCA02-Heparin) and serum (Caucasian American Sample Set; Lot # BDCA02-Serum) were obtained from the HUPO Specimen Collection and Handling Committee [20]. The total protein concentration of plasma and serum was estimated using a BCA Protein Assay (Pierce Chemical, Rockford, IL, USA). Trypsin digest was performed with porcine sequencing grade modified trypsin (Promega, Madison, WI, USA). HPLC-grade ACN was obtained from J. T. Baker (Phillipsburg, NJ, USA). All reagents and buffers were prepared with Milli-Q water (Millipore, Bedford, MA, USA).

2.2 Top six protein depletion

Removal of the six most abundant proteins in human plasma was achieved with a single 4.6 × 50 mm multiple affinity removal system (MARS) HPLC column (Agilent Technologies, Wilmington, DE, USA). The MARS column contains polyclonal antibodies to human albumin, transferrin, haptoglobin, α-1 antitrypsin, IgG, and IgA. Typically, plasma was diluted five-fold with the manufacturer’s equilibration buffer and filtered through a 0.22 μm microcentrifuge filter tube, and aliquots containing ~1 mg total protein were injected onto the antibody column. A total of 193 mL (14.5 mg) of plasma was depleted. The flow-through fractions from sequential injections were collected, pooled, and concentrated to 200 μL using a 5 K MWCO spin concentrator (Millipore). Affinity-bound major proteins were eluted with the manufacturer’s elution buffer, neutralized with 1 M NaOH, concentrated as above, and stored at −70°C. The concentrated unbound fraction (depleted plasma) was reduced with 10 mM DTT for 1 h at 23°C in 1 mL of buffer (final volume) containing 8 M urea, 10 mM glycine, pH 8.5. The reaction volume was subsequently reduced to 200 μL using a 5 K MWCO spin concentrator, and alkylated with 25 mM iodoacetamide in 1 mL of buffer (final volume) containing 8 mM urea, 10 mM glycine, pH 8.5, for 2 h at 23°C. Reaction was quenched by adding DTT to 1% final concentration. Prior to MicroSol-IEF, salts and reagents were removed by buffer exchange using a 5 K MWCO spin concentrator.

For the analysis of human serum, the major proteins from 415 μL (35.3 mg) of serum were depleted using two MARS columns, where the 50 mm column was connected in tandem to another 4.6 × 100 mm column. All buffers used in the depletion were supplemented with protease inhibitors (1 mM DFP, 1 μg/mL leupeptin, 1 μg/mL pepstatin, 5 mM EDTA). Each injection contained ~200 μL of five-fold diluted serum. The unbound fractions were pooled and concentrated to 240 μL (4.3 mg). Proteins were reduced and alkylated in the presence of 100 mM Tris-Cl, 8 M urea, pH 8.3 with 20 mM DTT, and 60 mM iodoacetamide for 1 h at 37°C each. Reaction was terminated with 60 mM DTT for 15 min at 37°C. Salts and reagents were removed by precipitation with 9 vol of acetone.

2.3 MicroSol-IEF fractionation

MicroSol-IEF was performed using a ZOOM IEF Fractionator (Invitrogen, Carlsbad, CA, USA). Depleted plasma was fractionated on a seven-chamber device, separated by immobile gel membranes having pH values of 3.0, 4.4, 4.9, 5.4, 5.9, 6.4, 8.1, and 10.0. The pH 3.0 and 10.0 membranes were obtained commercially (Invitrogen), while the remaining membranes were prepared as described previously [21]. Reduced and alkylated plasma was diluted to 3.5 mL and adjusted to the same constituent concentrations as the MicroSol buffer, i.e., 8 M urea, 2 M thiourea, 4% CHAPS, 1% DTT, 0.2% carrier ampholytes, pH 3–10L. Aliquots (700 μL) of the sample were loaded into the inner five chambers, and the remaining two outer chambers were filled with MicroSol buffer without sample. Depleted serum was fractionated on a five-chamber device, with pH 3.0, 4.6, 5.4, 6.2, 7.0, and 10.0 membranes obtained commercially (Invitrogen). Acetone-precipitated serum was dissolved in 700 μL of MicroSol buffer and was loaded into the central chamber of the device only.

2.4 Protein array pixelation

Following MicroSol-IEF, the fractions were separated by 1-D PAGE in individual lanes on short minigels. In some cases, proteins were extracted from the membrane partitions by two sequential incubations with 400 μL of MicroSol buffer. Each fraction was run on separate gels to avoid possible cross-contamination from other fractions. The highest possible protein amounts that did not cause extensive band distortion were loaded into 10-well 10% NuPAGE (Invitrogen) SDS gels and electrophoresed using MOPS running buffer until the tracking dye had migrated 4 cm (plasma analysis) or 6 cm (serum analysis) into the gels. In fractions containing very low amounts of proteins (F1 and M1 of serum analysis), the proteins were concentrated by precipitation with 9 vol of acetone and electrophoresed for only 2 cm. Proteins were visualized by staining with Colloidal Blue (Invitrogen). Each lane was subsequently cut into uniform 1 mm gel lanes were pixelated in a variable manner depending on the band intensity. Intense bands were digested as 1 mm pixels, while regions of the lane without much staining were digested as 4 mm pixels.

2.5 LC-ESI-MS/MS methods

Tryptic peptides from pixelation of the fractionated plasma sample were analyzed on an LCQ Deca XP+ IT mass spectrometer (Thermo Electron, San Jose, CA, USA) interfaced with a MicroPro pump (Eldex, Napa, CA, USA) and an auto-
sample. Serum tryptic digests were analyzed on an LTQ linear IT mass spectrometer (Thermo Electron) coupled with a NanoLC pump (Eksigent Technologies, Livermore, CA, USA) and autosampler. For each pixel, 5 μL (plasma samples) or 7 μL (serum samples) of the tryptic digest (total ~30 μL) was analyzed. Tryptic peptides were separated by RP-HPLC on a nanocapillary column, 75 μm i.d. × 20 cm PicoFrit (New Objective, Woburn, MA, USA), packed with MAGIC C18 resin, 5 μm particle size (Michrom Bioresources, Auburn, CA, USA). In some of the initial optimization experiments, POROS R2 C18 resin, 10 μm particle size (Applied Biosystems, Foster City, CA, USA) was used. Solvent A was 0.58% acetic acid in Milli-Q water, and solvent B was 0.58% acetic acid in ACN. Peptides were eluted into the mass spectrometer at 200 nL/min using an ACN gradient. Each RP-LC run consisted of a 10 min sample load at 1% B; a 75 min total gradient consisting of 1–28% B over 50 min, 28–50% B over 14 min, 50–80% B over 5 min, 80% B for 5 min before returning to 1% B in 1 min. To minimize carryover, a 36 min blank cycle was run between each sample. Hence, the total sample-to-sample cycle time was 121 min. In some optimization experiments, a 49 min gradient (1–28% B over 27 min, 28–50% B over 11 min, 50–80% B over 5 min, 80% B for 5 min before returning to 1% B in 1 min) was used instead of the 75 min gradient.

The mass spectrometers were set to repetitively scan m/z from 375 to 1600 followed by data-dependent MS/MS scans on the three most intense (LCQ Deca XP+) or the ten most abundant (LTQ) ions with dynamic exclusion enabled. In some experiments, gas-phase fractionation using different m/z ranges was performed as described in Fig. 2B.

2.6 Data analysis

Proteins from each pixel were identified from the MS/MS spectra using the SEQUEST Browser program (Thermo Electron). DTA files were generated from MS/MS spectra using an intensity threshold of 500,000 (Deca XP+) data or 5000 (LTQ data), and minimum ion count of 30. The DTA files generated were processed by the ZSA, CorrectIon, and IonQuest algorithms of the SEQUEST Browser program, and searched against the International Protein Index (IPI) human protein database [23] version 2.21 (July, 2003) containing 56,530 entries as requested by HUPO. In some of the optimization experiments (data sets for Fig. 2A), the National Center for Biotechnology Information non-redundant database (01/15/2004) was also used. To reduce database search time, the databases were indexed with the following parameters: average mass range of 500–3500, length of 6–100, tryptic cleavages with 1 (for LTQ analysis) or 2 (for LCQ Deca XP+ analysis) internal missed cleavage sites, static modification of Cys by carboxymethylation, and dynamic modification of Met to methionine sulfoxide (+16 Da). The DTA files were searched with a 2.5 Da peptide mass tolerance and 0 Da fragment ion mass tolerance. Other search parameters were identical to those used for database indexing. For each pixel, the peptides identified were assembled into the minimum number of unique proteins using SEQUEST SUMMARY with a depth of 3. Perl programs were developed for parsing, storing, analyzing, and retrieving SEQUEST results. Data from SEQUEST SUMMARY were stored in a relational database (Oracle 9i) with Perl Object layer.

The peptides from each protein were initially filtered using the following criteria: $X_{corr} \geq 1.9$ (z = 1), $2.3$ (z = 2), $3.75$ (z = 3) and $\Delta C_n \geq 0.1$; or $S_p \geq 0.7$. Further data analysis used the HUPO defined criteria where peptides were filtered using $X_{corr} \geq 1.9$ (z = 1), $2.2$ (z = 2), $3.75$ (z = 3) and $\Delta C_n \geq 0.1$; and $R_{Sp} \leq 4$. For both criteria, redundant peptides with the same accession number were removed and different forms (charge states and modification) of the same peptide were counted as a single-peptide hit. Keratins were also excluded from all data sets.

3 Results and discussion

3.1 Protein array pixelation strategy

We previously showed that MicroSol-IEF is capable of providing high-resolution fractionation of serum samples, resulting in albumin being confined into a single fraction [10, 11]. This fractionation approach has substantially expanded the number of proteins that can be detected by 2-DE since higher protein loads can be analyzed, in most fractions, without interference from the highly abundant albumin. We have also examined a number of commercially available methodologies for depleting abundant proteins from human plasma/serum and found that the Agilent MARS column is highly efficient in depleting the six most abundant proteins (albumin, transferrin, haptoglobin, α1-antitrypsin, IgG, and IgA) with minimal nonspecific binding of other proteins [9]. Removal of the major proteins allowed higher amounts of serum or plasma to be loaded onto 2-D gels. However, when the minor protein spots were analyzed by LC-ESI-MS/MS, most of these proteins turned out to be proteolytic products of major proteins [9]. To further enhance detection of lower abundance proteins, the depleted plasma/serum was subjected to MicroSol-IEF fractionation followed by 2-DE analysis. This very time-consuming series of 2-D gels only moderately increased the number of protein spots detected (data not shown). Hence, 2-DE is not an efficient method for detecting lower abundance proteins (<μg/mL) of the human plasma/serum proteome.

To overcome these limitations of 2-DE, we developed the protein array pixelation strategy for comprehensive profiling of the human plasma proteome (Fig. 1). The first step is major protein depletion using the Agilent MARS column. Following reduction and alkylation of the unbound (depleted) proteins, MicroSol-IEF is used as the second fractionation step to further reduce the complexity of the plasma proteome. Each fraction is subsequently electrophoresed on 1-D
gels, sliced into pixels, digested individually with trypsin, and analyzed by LC-ESI-MS/MS. In initial analysis of the data, proteins were identified from peptides that passed the stringent criteria of $X_{corr} \geq 1.9 (z = 1), 2.3 (z = 2), 3.75 (z = 3)$ and $ΔC_n \geq 0.1$, which is based on a commonly used relatively stringent published criteria [13]. Due to the concern that the strict $X_{corr}/ΔC_n$ used may eliminate some correctly identified low-level proteins, we also incorporated an additional scoring scheme, $S_f$ (final score), which was developed by William Lane at Harvard University and is available in the commercial version of SEQUEST Browser. The $S_f$ score examines the $X_{corr}$, $ΔC_n$, $S_p$, $R_s$, and Ions scores of SEQUEST using a neural network and combines them into a single score that reflects the strength of peptide assignment on a scale of 0–1. Peptides with $S_f$ score $\geq 0.7$ were considered to have a high probability of being correct (William Lane, personal communication). Therefore, peptide assignments by SEQUEST were also considered positive if they had an $S_f$ value of $\geq 0.7$, regardless of the $X_{corr}/ΔC_n$ scores.

In this study, emphasis is given to proteins identified by multiple peptides ($\geq 2$ peptides) because the chance multi-peptide proteins are false positives decreases exponentially with each additional peptide identified [24]. Since multiple peptides with lower $X_{corr}$ values can provide the same confidence as a single peptide with a high $X_{corr}$ value [24], the inclusion of the $S_f$ score in our analysis should not generate a significant increase in false identifications of multi-peptide proteins.

### 3.2 Optimization of protein array pixelation

A number of parameters that could affect the performance of the protein array pixelation were examined to optimize the method (Fig. 2). The first consideration is the size of the pixel used for trypsinic digestion. The smaller the pixel size, the more total samples will need to be analyzed by LC-ESI-MS/MS, thereby substantially increasing the total time needed to completely analyze a proteome. To determine the effect of pixel size, a test sample of nondepleted serum was loaded on multiple lanes of a 1-D gel and electrophoresed for the full distance. When the same 4 mm region of a gel lane was examined with pixel size of 1 mm (four pixels total), 2 mm (two pixels total), and 4 mm (one pixel total), the largest number of nonredundant proteins was identified from the four 1 mm pixels analyzed separately. When 2 mm pixels were used the number of identified proteins decreased moderately but when a 4 mm pixel was used the decrease was dramatic (Fig. 2A, columns 1–3). Even though the total analysis time decreased four-fold with the single 4 mm pixel analysis versus four 1 mm pixels, the 58% decrease in high-confidence protein identifications ($\geq 2$ peptides) is clearly unacceptable. The 2 mm pixel size is a good compromise between the total analysis time and the number of proteins detected, since compared with 1 mm pixels, the analysis time was reduced by 50% and the high-confidence identifications were reduced by only 15% (Fig. 2A).

In these analyses, the number of protein identifications could be improved by increasing the sample injection volume from 2 to 4 μL (21% increase in high-confidence proteins; Fig. 2A, columns 3 and 4). Although this is a modest increase, it does not increase the analysis time and is therefore a positive factor. Increasing the RP-LC gradient time increased the high-confidence protein identification by 36% and the analysis time by 27% (Fig. 2A, columns 3 and 5). Hence, this change had a marginal advantage. A greater increase was observed when 10 μm C18 particle size POROS R2 resin was replaced with 5 μm MAGIC C18 particle size resin where a 57% increase in high-confidence proteins was obtained for a constant analysis time (Fig. 2A, columns 6 and 7). Extending the column length from 10 to 20 cm did not appreciably increase the number of proteins identified (6% increase in high-confidence proteins), but substantially improved protein coverage, since a 30% increase in the number of proteins with $\geq 3$ peptides was observed with the 20 cm column (Fig. 2A, columns 7 and 8).
Figure 2. Parameters affecting the efficiency of the protein array pixelation strategy. (A) Bar chart displaying the effect of pixel size (columns 1–3), sample injection volume (columns 3 and 4), RP-LC time (columns 4 and 5), type of C18 resin (columns 6 and 7), and column length (columns 7 and 8) on the number of non-redundant proteins identified. P, POROS R2 C18 10 µm; M, MAGIC C18 5 µm. (B) Bar charts showing the effect of gel separation distance and gas-phase fractionation on the number of non-redundant proteins identified. Number of proteins identified from the human plasma F3 MicroSol-IEF fraction electrophoresed for 1 cm (10 × 1 mm sizepixel) and 4 cm (20 × 2 mm size pixel) on 1-D gels are shown. Gas-phase fractionation of the F3–7 pixel from the human plasma sample was analyzed using the full m/z range of 375–1600, or with three separate m/z ranges as indicated. Last column shows the combined number of non-redundant proteins identified from the three separate m/z ranges. Number of proteins identified by 1, 2, and ≥3 unique peptides are indicated by the white, black, and gray bars, respectively.

We also examined the effect of 1-D gel separation distance on the number of proteins identified. For this analysis, a MicroSol-IEF fraction of the major protein-depleted plasma sample (F3, see below) was electrophoresed for a total distance of 4 or 1 cm (see Fig. 3B). The 4 cm lane was divided into 2 mm pixels for a total of 20 pixels, whereas the 1 cm lane was analyzed as 1 mm pixels for a total of 10 pixels. The total number of nonredundant proteins identified from the 4 cm lane was 56% greater than the 1 cm lane and the high-confidence identifications increased by 14% (Fig. 2B). Because longer gel separation distances are likely to increase the total number of analyses per proteome, the benefits of increased SDS gel separation distance are ambiguous. If a substantial number of the identifications based on one peptide are correct, the increased analysis time may be worthwhile.

A well-known major factor that limits peptide identification capability of complex peptide mixtures using LC-ESI-MS/MS is coelution of more peptides from the RP column than the mass spectrometer can analyze. One method of addressing this problem is gas-phase fractionation, where a single sample is repeatedly analyzed using different segments of the full m/z range in each run [25]. To test the utility of gas-phase fractionation in the current method, a 2 mm pixel (F3 pixel 7, depleted plasma sample; see below) was analyzed using the unsegmented m/z of 375–1600 approach and compared with gas-phase fractionation using three separate m/z segments of 375–780, 780–1200, and 1200–1600 (Fig. 2B). In the gas-phase fractionation experiment, most proteins were identified using the m/z range of 780–1200. In contrast, least proteins were identified using m/z of 1200–1600, and all proteins identified in this segment were also found in the other two segments (data not shown). However, peptides identified in the m/z 1200–1600 segment are important because they increased sequence coverage of many proteins. By combining the three m/z segments, the total number of nonredundant proteins identified increased by 47% compared to the single unsegmented analysis. However, the number of high-confidence proteins increased by a marginal 6%. Taking into consideration the three-fold increase in the analysis time, the segmented approach does not appear to be an efficient strategy for comprehensive proteome analysis using the protein array pixelation strategy.

3.3 Total analysis time for protein array pixelation of human plasma proteome

As emphasized in the above discussion of separation parameters, a major consideration for any comprehensive proteome analysis strategy is the total time required to analyze an entire proteome. Some improvements such as increased injection volume, smaller resin size, and longer column length can be implemented without affecting the total run time and therefore even modest improvements in protein coverage or number of proteins identified are considered positive improvements. However, increasing the number of pixels per SDS gel lane, increasing the RP-LC gradient time for better peptide separation, and increasing gel separation distances will increase total proteome analysis time as well as
increase the number of proteins identified. Therefore, a practical compromise between improved number of proteins detected and increased analysis time has to be achieved. We felt that a generally acceptable time frame for complete proteome analysis should be similar to the time required to perform a MudPIT analysis of the human plasma/serum proteome [12, 15]. Therefore, based upon the optimization results discussed in Section 3.2, we decided to fractionate the depleted plasma using MicroSol-IEF into seven fractions, followed by 1-DE of each fraction for a total distance of 4 cm (Fig. 1). Each gel lane is sliced into 2 mm size pixels to produce a total of 140 pixels. Following tryptic digestion, each pixel is analyzed by LC-ESI-MS/MS with an RP-LC gradient time of 75 min. However, to minimize carryover from the previous run especially with increased sample injection volume, a short blank gradient is run after each analytical run. Hence, the total RP-LC run time from sample to sample is 121 min (see Section 2.5). In total, 11.8 days will be required to complete the analysis of the 140 pixels from the plasma proteome.

This compares favorably with ~9.8 days to analyze the 135 SCX fractions of human serum proteome [12], and ~13.9 days for 77 LC-ESI-MS/MS runs from two cycles of SCX-LC of the human plasma proteome [15].

### 3.4 Systematic protein array pixelation of the human plasma proteome

The profiling of human plasma proteome began with major protein depletion from a total of 193 μL (14.5 mg) plasma (BDCA02-Heparin) using the MARS antibody column (Fig. 3A). Following depletion, 2.4 mg of unbound proteins were recovered, indicating that the six targeted proteins constituted approximately 83% of plasma proteins in this sample. Analysis of the bound fraction by 2-DE showed that the bound proteins were the six targeted proteins with no apparent evidence of other proteins [9].

The depleted plasma was then fractionated by MicroSol-IEF into seven fractions (Fig. 3B). Based on the 1-D gel analysis of the MicroSol-IEF fractions, the plasma proteins were well distributed throughout the seven fractions although the terminal fractions (F1 and F7) have the least amount of proteins as judged by the staining intensity (Fig. 3B). Many protein bands (including nondepleted abundant proteins) were present only in a specific fraction, indicating that MicroSol-IEF effectively separated proteins based on their pI. For example, a major protein with apparent MW of approximately 25 kDa was located almost exclusively in F4 (pH 5.4–5.9) with minor amounts found in more acidic fractions as observed by 1-D gel (Fig. 3B). Subsequent analysis by MS/MS identified this protein as apolipoprotein A-I precursor with calculated MW of 30.8 kDa and pI of 5.6. The good agreement with the observed values, confirmed the effectiveness of the pI and MW separation in this strategy. Hence, MicroSol-IEF not only further reduced the complexity of the plasma proteome, but also confined most remaining abundant proteins into specific fractions. This allows higher amounts of samples to be analyzed in downstream processes and permits us to dig deeper into the plasma proteome for lower abundance proteins.

Following MicroSol-IEF fractionation, the seven fractions were further separated by 1-D SDS-PAGE for a total distance of 4 cm (Fig. 3B). The gel lanes were then sliced and analyzed as uniform 2 mm pixels for a total of 140 pixels. Each pixel was digested in-gel with trypsin and analyzed by LC-ESI-MS/MS on an LCQ Deca XP+ mass spectrometer. In order to obtain a better correlation between the observed and the calculated MW of the proteins identified, the amount of sample loaded on the gel was limited to avoid overloading and to provide the optimal resolution of protein bands. In addition, the edge of the gel lanes where some degree of vertical smearing is frequently

![Figure 3. Major protein depletion and MicroSol-IEF separation of human plasma proteins. (A) 1-D gel showing the plasma proteins [P] before depletion, and unbound [UB] and bound [B] proteins from the MARS antibody column. Tr, transferrin; Alb, albumin; αT, antitrypsin; HC, Ig heavy chain; Hp, haptoglobin; LC, Ig light chain. (B) Seven MicroSol-IEF fractions of the depleted plasma proteins were subjected to 1-D gel separations for a total distance of 4 cm from the bottom of the wells. Separation of F3 fraction for 1 cm is shown in the right panel. Proteins were separated on 10% bis-Tris NuPage gels using MOPS buffer, and stained with Colloidal blue.](image-url)
observed was excluded when the lane was cut. Depending
upon protein concentration, between 1.3 and 2.8% (average
1.9%) of each MicroSol-IEF fraction was loaded onto 1-D
gels used for pixelation. This average amount is equivalent
to approximately 3.7 μL (278 μg) of the original plasma
sample. Following tryptic digestion of the pixels, only
16.7% of each digestion mixture was analyzed by LC-ESI-
MS/MS analysis. Hence, an amount equivalent to 0.6 μL
(45 μg) of the original plasma sample was actually con-
sumed in the final analysis.

From the LC-ESI-MS/MS analysis of the 140 pixels, a
2-D array of the human plasma proteome was generated
(Fig. 4A). Each pixel in the array has a distinct range of
MW and pI as shown and contains a group of identified
proteins (from 3 to 36 proteins) defined by one or more
peptides that passed the $X_{corr}/\Delta C_n/S_f$ criteria. Each pixel
was assigned a name in the format $F_x-y$, where $x$ is the
MicroSol-IEF fraction (1–7), and $y$ is the MW fraction
from 1 (largest) to 20 (smallest). In general, the number of
proteins identified in the pixels corresponds roughly to the
staining density of the gel (Figs. 3B, 4A). A total of
744 nonredundant proteins defined by 3235 nonredundant
peptides were identified from all the 140 pixels. Of these,
185 proteins (24.9%) were identified by at least two differ-
ent peptides (high-confidence) whereas the majority
(75.1%) was single-peptide proteins (Table 1).

A unique feature of this method is that the 2-D array can
also be used to display the distributions of specific proteins
that provide insight into their MW, pI, and the presence of
alternate forms of each protein such as alternate splices and
proteolytic fragments (Fig. 4B–D). Of course due to the fact
that many plasma proteins are heterogeneously modified
such as by glycosylation and proteolytic processing, the
observed MW and pI are not expected to closely match the
values derived from amino acid sequences. Due to the high
sensitivity of the mass spectrometer, the high- and moderate-
abundance proteins (mg/mL – μg/mL) were commonly
found in more than one pixel. Since the relative abundance of
a specific protein can be roughly determined from the num-
er of unique peptides identified [26], the primary position of
an abundant protein is determined by the pixel containing
the maximum number of peptides. The distribution of three
proteins with varying abundance (apolipoprotein B-100,
720 μg/mL; ceruloplasmin, 210 μg/mL; metalloproteinase
inhibitor I, 14 ng/mL [27]) is shown in Fig. 4. The distribu-
tion of apolipoprotein B-100 in the array indicated that the

© 2005 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
www.proteomics-journal.de
Table 1. Number of nonredundant proteins identified from human plasma/serum using different filters

<table>
<thead>
<tr>
<th>Sample</th>
<th>Filter</th>
<th>Number of nonredundant proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>Plasma</td>
<td>$S^a_\pi$</td>
<td>744</td>
</tr>
<tr>
<td>Serum</td>
<td>$S^a_\pi$</td>
<td>4377</td>
</tr>
<tr>
<td>Plasma</td>
<td>HUPO$^b_\Pi$</td>
<td>575</td>
</tr>
<tr>
<td>Serum</td>
<td>HUPO$^b_\Pi$</td>
<td>2890</td>
</tr>
<tr>
<td>Plasma-serum</td>
<td>HUPO$^b_\Pi$</td>
<td>319</td>
</tr>
<tr>
<td>Serum-serum</td>
<td>HUPO$^b_\Pi$</td>
<td>319</td>
</tr>
<tr>
<td>Combined</td>
<td>HUPO$^b_\Pi$</td>
<td>3146</td>
</tr>
<tr>
<td>Without Ig</td>
<td>HUPO$^b_\Pi$</td>
<td>3104</td>
</tr>
</tbody>
</table>

a) Filter used: $X_{\text{corr}} \geq 1.9$ (z = 1), 2.3 (z = 2), 3.75 (z = 3) and $\Delta C \geq 0.1$; or $S \geq 0.7$
b) Filter used: $X_{\text{corr}} \geq 1.9$ (z = 1), 2.2 (z = 2), 3.75 (z = 3) and $\Delta C \geq 0.1$; and $e_{\text{m}} \leq 4$
c) $\geq 3$, 2, and 1 indicate the number of unique peptides per protein
d) Proteins in plasma that are also identified in the serum data set
e) Proteins in serum that are also identified in the plasma data set
f) Both plasma and serum data sets were combined for analysis
g) Combined data set with immunoglobulin entries removed

Protein was present in at least two major forms (Fig. 4B). Both major forms were larger than 200 kDa; the smaller form (in F4–2 and F5–2) had a pI in the range of pH 5.4–6.4, whereas the pI of the larger form (in F3–1) is between pH 4.9 and 5.4. The observed MW of the protein and the multiple forms observed are consistent with the calculated MW of 515.6 kDa, and the reported forms of the protein such as B-74, B-48, and B-26 with apparent MW of 400, 259, and 140 kDa, respectively [28]. The observed pI of the protein is slightly lower than the theoretical pI of 6.6, which could be caused by heterogeneous modifications such as glycosylation [28]. The moderate-abundance protein, ceruloplasmin, was found mainly in F4–5 which is consistent with the calculated MW of 122.2 kDa and the theoretical pI of 5.4 (Fig. 4C). Unlike high- and moderate-abundance proteins, low-abundance proteins were usually identified by a single peptide that was found in only one or two pixels. For example, metalloproteinase inhibitor 1 precursor identified by the single-peptide GFQALGDAADIR is found only in pixel F6–15 at $\sim 30$ kDa and pI between 6.4 and 8.1 (Fig. 4D). These values are close to the expected MW of 23.2 kDa and pI of 8.5 for the protein. The MS/MS spectrum of this peptide was verified by manual inspection (see also Fig. 7). Hence, the MW and pI values derived from the 2-D array can be used to reinforce the protein identifications made by SEQUEST, especially for proteins identified by a single peptide, which is the group of proteins that predominates in most shotgun proteomics approaches.

3.5 Systematic protein array pixelation of the human serum proteome

Protein array pixelation of the HUPO serum sample, BDC02-Serum, was performed using a method similar to that used for the plasma sample, except this method incorporated several refinements to further improve coverage of the proteome (Fig. 5A). The major protein depletion was performed on 415 µL (35.3 mg) of serum using a dual MARS column. A total of 4.3 mg of unbound proteins were recovered, indicating that approximately 88% of the total serum protein content was removed in this sample compared with the removal of 83% of total plasma proteins. This difference was at least partially due to more effective removal of targeted major proteins using the dual column compared with the single MARS column depletion of the plasma. This is consistent with the number of albumin peptides observed in both samples after depletion, where 9.2% sequence coverage of albumin was obtained from the depleted serum compared to 59.1% sequence coverage from the depleted plasma sample. Similarly, serotransferrin was identified with 40.7% sequence coverage in the depleted plasma sample but was not detected in the depleted serum sample. Since minor amounts of major proteins such as albumin (40 mg/mL concentration) are still major components of the sample, they will still interfere with the overall analysis. Hence, it is better to use a longer antibody column and under-load the column to ensure the most effective depletion of targeted proteins as possible.

Following reduction and alkylation of the depleted serum proteins, the sample was fractionated into five pH fractions by MicroSol-IEF (Fig. 5B). The fractionation was performed using the commercially available pH membrane partitions, which greatly simplify the MicroSol-IEF procedure. To compensate for the reduced MicroSol-IEF fractions, the majority of the fractions (F2–F5) were separated on 1-D gels for a total distance of 6 cm. Compared to the 4 cm separation of the plasma sample, the longer separation distance should allow for increased sample loading (up to 50%) without overloading the gels. Due to the lower amount of proteins in F1, this fraction was concentrated by acetone precipitation and electrophoresed for only 2 cm to minimize empty regions in the gel lane (Fig. 5B). In addition, we also extracted proteins from the membrane partition (M1) between the anode buffer and F1 to detect proteins that might be trapped in the membrane. The M1 fraction was also concentrated by acetone precipitation prior to 1-D gel separation for 2 cm (Fig. 5B).

Following gel electrophoresis, the gel lanes containing M1 and F1 were analyzed as 2 mm size pixels. In the initial optimization studies presented above, more unique proteins were identified from four 1 mm size pixels than a single 4 mm pixel (Fig. 2A). To potentially increase the number of...
proteins identified, gel lanes containing the F2-F5 fractions were pixelated in a variable manner (1–4 mm size pixel) depending on the band intensity. Regions of the gel with intense staining were analyzed as 1 mm pixels, and regions without much staining were analyzed as 4 mm pixels (Figs. 5B, 6A). For a direct comparison of pixel size using current methods, fraction F3 was also reanalyzed as uniform 2 mm pixels. In total, 159 pixels were generated for tryptic digestion and analyzed by LC-ESI-MS/MS with a total analysis time of 13.4 days (121 min RP-LC total run time per sample). Depending upon protein concentration, between 1.1 and 5.2% (2.5% average) of each soluble MicroSol-IEF fraction was used for gel pixelation. The concentration of each fraction was equivalent to 10.4 μL (885 μg) of the original serum sample. After tryptic digestion, 23.3% of the digested material was injected and analyzed by LC-ESI-MS/MS. Therefore, protein identification was performed using an amount equivalent to approximately 2.4 μL (204 μg) of the original serum sample.

All samples from this serum analysis were analyzed using a Thermo Electron linear IT LTQ mass spectrometer, which is more sensitive and has a faster scan rate than the LCQ Deca XP+ [29]. The number of proteins identified for all 159 pixels is shown as a heat map in the 2-D array (Fig. 6A). Each pixel contained between 13 and 199 proteins that pass the Xcorr/ΔCn/Sv criteria defined above. Comparison of the uniform (F3f) and variable (F3v) pixelation methods of the F3 fraction indicated that the variable pixelation method did not offer any improvement over the uniform pixelation method (Fig. 6B). In fact, uniform pixelation identified 6.9% more high-confidence proteins compared to the variable pixelation method. In addition, the uniform pixelation method is easier and quicker to perform, as there is no need to correlate the pixel size with band intensity. The total number of nonredundant proteins identified from the 159 pixels was 4377 from a total of 9393 nonredundant peptides. Of these, 752 proteins (17.2%) were identified as high-confidence and the majority (82.8%) was single-peptide proteins (Table 1). Therefore, the overall improvements due to further refinement of the method and, most importantly, use of the highly sensitive LTQ mass spectrometer, resulted in about fourfold increase in the number of high-confidence proteins identified in serum compared to the plasma analysis.

The establishment of the pH gradient during MicroSol-IEF is dependent on the membrane partitions between each fraction [21]. During MicroSol-IEF, some proteins can be partially or completely trapped in the membrane partitions and are therefore excluded from the soluble fractions. To investigate this possibility, proteins were extracted from the five membrane partitions and analyzed on 1-D gels (data not shown). Except for M1 (pH 3.0 membrane partition between anode buffer and F1), all protein bands from the other membrane partitions appeared to have corresponding protein bands in adjacent soluble fractions. Furthermore, the protein bands from the membrane partitions are much less intensely stained than their soluble fractions counterparts.

---

[Diagram of protein analysis process]

**Figure 5.** Protein array pixelation of the HUPO serum sample (BDCA02-Serum). (A) Diagram showing the improved methodologies for analysis of the human serum sample. Steps identical to those used for analysis of the plasma sample are shown in gray. (B) 1-D gel showing the five MicroSol-IEF fractions (F1–F5) of major protein-depleted human serum. M1 shows protein extracted from the pH 3.0 membrane. Separation distances for each fraction are as indicated. Proteins were separated on 10% bis-Tris NuPage gels using MOPS buffer, and stained with Colloidal blue.
The only exception is apolipoprotein B-100 which is found predominately in the membrane partition between F3 and F4, presumably due to its large size of ~540 kDa [28]. The tentative conclusion that most proteins trapped in membrane partitions were also partially recovered in adjacent fractions was further supported by parallel analysis of membrane and soluble fractions from similar serum separations on 2-D gels (data not shown). In M1, however, two apparently unique protein bands were observed in the 1-D gel analysis (Fig. 5B shows the concentrated M1 proteins). Pixelation of the M1 fraction and comparison with F1 fraction indicated that 42 high-confidence proteins were identified in M1, but only 7 were not identified in F1 (Fig. 6C). Out of these seven high-confidence M1 proteins, six were not identified elsewhere in the entire serum proteome analysis. Hence, a few acidic proteins were found exclusively in the pH 3 membrane partition, and it is likely that a small number of unique proteins are in other membrane partitions and could be detected if higher sensitivity methods like LC-ESI-MS/MS are used instead of 1- or 2-D gels. Therefore, it may be advantageous to include membrane partition extracts in the analyses to provide more comprehensive coverage of proteins.

3.6 Analyses of human plasma and serum proteomes using HUPO filter criteria

All the data described above were analyzed using the $X_{corr}/\Delta C_n/S_r$ criteria. From experience, we know that a substantial number of the proteins identified by a single peptide using these criteria are incorrect and the probability of a protein being correctly identified increases with the number of unique peptides identified. However, for biomarker discovery it is better to have a less stringent filter so that potentially interesting low-abundance proteins will not be excluded from the analysis. This, however, will inevitably increase the number of false positives and will require more efforts to verify the data generated. As the aim of the HUPO PPP is to provide the most accurate description of the human serum/plasma proteome possible, a more stringent filter ($X_{corr} \geq 1.9$ (z = 1), 2.2 (z = 2), 3.75 (z = 3) and $\Delta C_n \geq 0.1$; and $R_{sp} \leq 4$) was selected by HUPO to minimize false identifications. The analysis of our plasma and serum data sets using the HUPO criteria is summarized in Table 1. With the more stringent HUPO criteria, the number of nonredundant proteins identified by $\geq 2$ peptides from the smaller plasma data set was reduced by only 5.9% while the single-peptide proteins were reduced by 28.3%. A larger reduction was observed with the serum data set, where proteins identified by $\geq 2$ peptides and a single peptide were reduced by 30.9 and 34.6%, respectively. In both data sets, the reduction in proteins identified by $\geq 2$ peptides was mainly contributed by the two-peptide proteins (Table 1). This is consistent with the expectation that identification errors are more likely to happen for single-peptide proteins followed by two-peptide proteins, and least likely for proteins identified by more than two peptides.

In this study, nonredundant proteins are defined as proteins with different accession number. The in-house software used for the analysis of these datasets did not eliminate potential redundancy caused by SEQUEST assignment of the same peptide in different pixel to different but homologous...
database entries. To address this issue, the plasma and serum datasets were reanalyzed using the DTASelect program [30] that is capable of grouping redundant identifications. The program was used to filter peptides using the HUPO high stringency criteria. Proteins that were subsets of others of contained the description ‘keratin’ were removed. A total of 576 and 2725 nonredundant proteins were reported for the plasma and serum datasets, respectively, using the DTASelect program compared with 575 and 2890 proteins using our in-house software. Therefore, the redundancy in our analysis is very minimal.

When both data sets were combined, a total of 3146 nonredundant proteins were identified using the HUPO criteria (Table 1). Of these, 567 (18.0%) were identified by ≥2 unique peptides and 82.0% were single-peptide proteins. Since immunoglobulins were depleted in both samples, they constituted only 1.3% of the total nonredundant proteins (or 3.2% of proteins with ≥2 peptides) identified in the combined data set (Table 1). The number of proteins that are common to both data sets is only 319, and is limited by the lower sensitivity method used in the plasma analysis (Table 1). In addition, 92.5% of the proteins with ≥2 peptides in plasma were identified in the serum analysis, but only 40.6% of serum proteins with ≥2 peptides were identified in the plasma analysis. However, 49.5% of the common proteins identified in plasma are single-peptide proteins. Since these single-peptide proteins were identified using a different instrument and sample, it is likely that a large percentage of the single-peptide proteins identified in plasma are probably correct. In support of this, many of the single-peptide proteins identified in plasma, as well as in serum, have rich MS/MS fragmentation patterns that agree well with peptide sequences assigned by SEQUEST. Examples of the MS/MS spectra for single-peptide proteins identified in both data sets are shown in Fig. 7. All of the major peaks in both MS/MS spectra can be accounted for by fragment ions from the predicted peptide sequences, indicating that the peptide assignment is correct. Of particular interest is the protein creatine kinase M which, in the MB isoform, is an important serum marker for myocardial infarction [31]. Therefore, even though the single-peptide protein category contains the most false positives, it also contains many important correct entries that cannot be ignored.

Examples of low-abundance proteins identified in the plasma and serum samples using HUPO criteria are shown in Table 2. The list provides an estimate of the detection limit

### Table 2. Examples of low-abundance proteins (<100 ng/mL) and the corresponding peptides identified in the human plasma and serum samples

<table>
<thead>
<tr>
<th>Sample</th>
<th>Name</th>
<th>ng/mL</th>
<th>Sequence</th>
<th>z</th>
<th>Xcorr</th>
<th>ΔCn</th>
<th>RSp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasma</td>
<td>Vascular endothelial-cadherin</td>
<td>30</td>
<td>VHDVNDNWPFVTHR</td>
<td>3</td>
<td>4.08</td>
<td>0.52</td>
<td>1</td>
</tr>
<tr>
<td>Serum</td>
<td>Vascular endothelial-cadherin</td>
<td>30</td>
<td>DTGENLETPSSFTIK</td>
<td>2</td>
<td>4.32</td>
<td>0.43</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>EYFAIDNSGR</td>
<td>2</td>
<td>2.59</td>
<td>0.55</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KPLGTVLAM*DPDAAR</td>
<td>3</td>
<td>3.76</td>
<td>0.32</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>VDAETGDVFAIER</td>
<td>2</td>
<td>3.75</td>
<td>0.61</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>VHDVNDNWPFVTHR</td>
<td>2</td>
<td>3.45</td>
<td>0.51</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YEIVVEAR</td>
<td>2</td>
<td>2.39</td>
<td>0.27</td>
<td>1</td>
</tr>
<tr>
<td>Plasma</td>
<td>L-selectin</td>
<td>17</td>
<td>NKEDCVEIYIK</td>
<td>2</td>
<td>3.56</td>
<td>0.38</td>
<td>1</td>
</tr>
<tr>
<td>Serum</td>
<td>L-selectin</td>
<td>17</td>
<td>NKEDCVEIYIK</td>
<td>2</td>
<td>4.29</td>
<td>0.34</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SLTEEAENWGDGEPNNK</td>
<td>2</td>
<td>4.38</td>
<td>0.50</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SLTEEAENWGDGEPNNKK</td>
<td>2</td>
<td>5.22</td>
<td>0.61</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SYYWGIR</td>
<td>2</td>
<td>2.65</td>
<td>0.35</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TICESSGIWSNSPICQK</td>
<td>2</td>
<td>5.33</td>
<td>0.50</td>
<td>1</td>
</tr>
<tr>
<td>Plasma</td>
<td>Metalloproteinase inhibitor 1</td>
<td>14</td>
<td>GFQALGDAADIR</td>
<td>2</td>
<td>3.72</td>
<td>0.45</td>
<td>1</td>
</tr>
<tr>
<td>Serum</td>
<td>Metalloproteinase inhibitor 1</td>
<td>14</td>
<td>GFQALGDAADIR</td>
<td>2</td>
<td>3.05</td>
<td>0.26</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLACLPR</td>
<td>2</td>
<td>2.38</td>
<td>0.11</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LGSHTCLWTDQLLOGSEK</td>
<td>3</td>
<td>5.32</td>
<td>0.52</td>
<td>1</td>
</tr>
<tr>
<td>Serum</td>
<td>Vascular endothelial growth factor D</td>
<td>0.500</td>
<td>SEQQIRAASSEELLR</td>
<td>2</td>
<td>2.47</td>
<td>0.13</td>
<td>2</td>
</tr>
<tr>
<td>Serum</td>
<td>Calcitonin</td>
<td>0.190</td>
<td>SALESSPADPATLSEDEAR</td>
<td>2</td>
<td>2.24</td>
<td>0.15</td>
<td>3</td>
</tr>
<tr>
<td>Serum</td>
<td>Tumor necrosis factor (TNF-a)</td>
<td>0.041</td>
<td>PWYEPYLYGGVFQLEK</td>
<td>2</td>
<td>2.87</td>
<td>0.30</td>
<td>1</td>
</tr>
</tbody>
</table>

a) Concentration values were obtained from [27]

* Indicates methionine oxidation

© 2005 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
of the protein array pixelation strategy. Some proteins in the low ng/mL can be detected from 45 μg of plasma using the LCQ Deca XP+, whereas some proteins in the pg/mL can be detected from the 204 μg of serum analyzed using the LTQ mass spectrometer. Not surprisingly, the ability to detect low abundance proteins decreases with protein abundance. For example, among the low abundance proteins described by Haab et al. [27] in their Table 2, 14 out of the 20 proteins in the 1 to 100 ng/mL concentration range were detected in our serum analysis, whereas only 2 out of 19 proteins at concentrations below 1 ng/mL were detected. In addition, most of the lower abundance proteins identified in plasma are single-peptide proteins whereas the same proteins were identified with multiple peptides in the serum analysis using the more sensitive linear IT mass spectrometer. This indicates that the use of the highly sensitive LTQ mass spectrometer coupled with our optimized method allows detection of proteins up to a concentration range of 10⁹.

4 Concluding remarks

This study demonstrates the utility of a novel 4-D protein profiling strategy, protein array pixelation, for comprehensive profiling of human plasma and serum proteomes. The four separations used in this strategy greatly reduce plasma/serum complexity, allowing access to proteins differing in abundance by up to nine orders of magnitude. Using HUPO criteria for high-confidence protein identifications, this strategy has detected a total of 3104 nonredundant proteins, after excluding keratins and immunoglobulins. Although larger amounts of sample are used for early steps, the final LC-ESI-MS/MS analyses are based on very low amounts of sample (45 μg of plasma and 204 μg of serum). Of these identified proteins, 549 were identified with two or more unique peptides. The total time required for analyzing each sample was similar to MudPIT approaches described by others [12, 15]. Analysis of the HUPO serum sample (BDCA02) using the
highly sensitive LTQ mass spectrometer and an optimized method produced a very rich data set that contained >90% of the proteins with two or more peptides identified in the plasma sample. Most importantly, many low-abundance proteins (<100 ng/mL – pg/mL levels) were identified in this data set. In conclusion, the protein array pixelation strategy is a powerful method for comprehensive protein profiling and for protein biomarker discovery.

We would like to thank Brian Haab for sharing data prior to publication. We are also grateful to John Yates, III from The Scripps Research Institute for providing the DTASelect software. This work was supported in part by the National Institutes of Health Grants CA94360 and CA77048 to D.W.S., and institutional grants to the Wistar Institute including an NCI Cancer Grants CA94360 and CA77048 to D.W.S., and institutional grants to the Wistar Institute including an NCI Cancer and the Commonwealth Universal Research Enhancement Program, Pennsylvania Department of Health.

5 References