



Published in final edited form as:

J Chromatogr B Analyt Technol Biomed Life Sci. 2014 March 1; 0: 135–142. doi:10.1016/j.jchromb.2014.01.037.

Targeting Deeper the Human Serum Fucome by a Liquid-phase Multicolumn Platform in Combination with Combinatorial Peptide Ligand Libraries

Subhashini Selvaraju and Ziad El Rassi*

Department of Chemistry, Oklahoma State University, Stillwater, OK 74078-3071

Abstract

Combinatorial peptide ligand library (CPLL) was evaluated as an off line step to narrow the differences of protein concentration in human serum prior to the capturing of human fucose from disease-free and breast cancer sera by a multicolumn platform *via* lectin affinity chromatography (LAC) followed by the fractionation of the captured glycoproteins by reversed phase chromatography (RPC). Two monolithic lectin columns specific to fucose, namely *Aleuria aurantia* lectin (AAL) and *Lotus tetragonolobus* agglutinin (LTA) columns were utilized to capture the fucose, which was subsequently fractionated by RPC yielding desalted fractions in volatile acetonitrile-rich mobile phase, which after vacuum evaporation were subjected to tryptic digestion prior to LC-MS/MS analysis. AAL has a strong affinity towards core fucosylated N-glycans and has a weak binding towards fucose in the outer arm while LTA can bind to glycans having fucose present in the outer arm. The combined strategy consisting of the CPLL, multicolumn platform and LC-MS/MS analysis permitted the identification of the differentially expressed proteins (DEPs) in breast cancer serum yielding 58 DEPs in both the LTA and AAL fractions with 6 DEPs common to both lectins. 17 DEPs were of the low abundance type, 16 DEPs of the borderline abundance type, 4 DEPs of the medium abundance type and 15 DEPs of the high abundance type. The remaining 6 DEPs are of unknown concentration. Only proteins exhibiting 99.9% protein identification probability, 95% peptide identification probability, and a minimum of 5 unique peptides were considered in finding the DEPs *via* scatterplots.

Keywords

Breast Cancer Serum; Fucosylation; Glycoproteins; Lectin Affinity Chromatography; Monolithic Columns; ProteoMiner™

1. Introduction

Protein glycosylation represents the most abundant post-translational modification (PTM) causing over 70% of all human proteins to bear this important and complex PTM [1]. One of the most important aspects of this PTM is its alteration in diseases and especially in cancers. Thus, the urgent need for tools and strategies for isolating and assessing disease induced

© 2014 Elsevier B.V. All rights reserved.

Corresponding author: Prof. Ziad El Rassi, Department of Chemistry, Oklahoma State University, Stillwater, OK 74078-3071. Phone: (405) 744-5931; Fax: (405) 744-6007; elrassi@okstate.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

alteration in glycosylation. Fucosylation of glycoproteins (or the so-called fucome) has been recognized as one of the most common aberration in cancerous cells [2–6]. This research article is aimed at introducing the use of combinatorial peptide ligand libraries (CPLL) to narrow the protein concentration range in human serum prior to capturing and fractionating the human fucome followed by comparative analysis of fucome in cancer serum with respect to disease free serum *via* liquid chromatography – tandem mass spectrometry (LC-MS/MS). Thus, this investigation describes a novel approach incorporating the following: (i) narrowing serum proteins dynamic concentration range by CPLL beads, a solid-phase extraction technique based on peptide affinity, (ii) capturing the human fucome by lectin affinity chromatography (LAC) and (iii) fractionation of the captured fucome by reversed phase chromatography (RPC). Steps (ii) and (iii) are integrated in a liquid-phase multicolumn platform that is a slightly modified version of the previously reported one [7] in the sense that the depletion columns that were previously used online to remove albumin and immunoglobulins (Ig's) have been replaced by an off line protein equalization *via* the CPLL approach, which has been shown to be very effective in narrowing the protein concentration range in many biological fluids and extracts [8–11], thus allowing an in-depth proteomics profiling.

To capture a given sub-glycoproteomics, e.g., the fucome, LAC has been shown recently to offer the potential to achieve this goal [7, 12–17]. Two fucose specific lectins namely, *Aleuria aurantia* lectin (AAL) and *Lotus tetragonolobus* agglutinin (LTA) were immobilized onto the surface of glyceryl methacrylate (GMM)/pentaerythritol triacrylate (PETA) monolith, which was very recently introduced by Gunasena and El Rassi for performing immuno affinity chromatography at reduced nonspecific interactions [18]. AAL has a strong affinity towards N-glycans possessing a fucose residue attached to the innermost GlcNAc (represented as $\text{Fuc}\alpha 1 \rightarrow 6\text{GlcNAc} \rightarrow \text{R}$) of the N-linked-core structure (i.e., core fucosylated N-glycans) and has weak binding towards fucose in the outer arm such as $\text{Fuc}\alpha 1 \rightarrow 2\text{Gal}\beta 1 \rightarrow 4\text{GlcNAc}\beta 1 \rightarrow \text{R}$, $\text{Gal}\beta 1 \rightarrow 4(\text{Fuc}\alpha 1 \rightarrow 3)\text{GlcNAc} \rightarrow \text{R}$ and $\text{Gal}\beta 1 \rightarrow 3(\text{Fuc}\alpha 1 \rightarrow 4)\text{GlcNAc} \rightarrow \text{R}$, where R = H or sugar [19]. On the other hand, immobilized LTA can bind to glycans having fucose present in the outer arm including $\text{Fuc}\alpha 1 \rightarrow 3/1 \rightarrow 4\text{GlcNAc}$ and $\text{Fuc}\alpha 1 \rightarrow 2\text{Gal}$. LTA also has an affinity for glycans containing the Le^x determinant represented as $\text{Gal}\beta 1 \rightarrow 4(\text{Fuc}\alpha 1 \rightarrow 3)\text{GlcNAc}\beta 1 \rightarrow \text{R}$ [20]. The haptenic sugar for AAL and LTA is α -L-fucose.

The importance and significance of the combination of the CPLL technology with the multicolumn platform stem from facilitating the capturing and fractionating the human fucome at all level of protein abundance prior to LC-MS/MS analysis. The CPLL technology, also called ProteoMiner™, should allow in principle the processing of human serum prior to the multicolumn platform to yield representative serum protein samples that have much narrower concentration range than the originally existing range that usually spans over 10 to 12 orders of magnitude. This wide dynamic concentration range represents a real challenge for existing instrumentation and methods including LC-MS/MS. It is expected that upon narrowing the concentration range differences between the serum proteins, the lectin column sites in the multicolumn platform will be available to all fucosylated proteins on a more or less an equitable ground, a fact that should in principle facilitate the capturing of these proteins at all level of original abundance. In other words, the lectin column binding sites will not be primarily occupied by the high abundance fucosylated proteins and to lesser extent by the medium abundance fucosylated proteins but they will be available to all fucosylated proteins at equitable level from the serum treated by the CPLL beads. This should facilitate the capturing of more fucosylated proteins by the multicolumn platform. Thus, a wider range of DEPs should be readily detected by LC-MS/MS, a fact that may lead to approaching the next generation DEPs; meaning low abundance DEPs.

2. Materials and methods

2.1 Materials

The unconjugated lectins namely, AAL and LTA were purchased from Vector Laboratories (Burlingame, CA, USA). Pooled breast cancer serum from six donors (stages 2, 3 and 4) and pooled disease-free human serum from six donors (same age group and race as the cancer serum) were purchased from Bioreclamation (Jericho, NY, USA). Stainless steel tubing of 4.6 mm ID was obtained from Alltech Associates (Deerfield, IL, USA). The ProteoMiner™ bulk beads were purchased from Bio-Rad (Hercules, CA, USA). Glycerylmethacrylate (GMM) was purchased from Monomer-polymers & Dajac Labs (Feaster-Ville, PS, USA). The AcroSep™ SDR columns were purchased from Pall Life Sciences (Port Washington, NY, USA). Pentaerythritol triacrylate (PETA), 2,2'-azobis(isobutyronitrile) (AIBN) and 1-dodecanol were purchased from Aldrich Chemical Co. (Milwaukee, WI, USA). Sodium periodate, sodium cyanoborohydride, trifluoroacetic acid (TFA) and L-(-)-fucose were purchased from Sigma-Aldrich (St. Louis, MO, USA). Cyclohexanol and HPLC grade acetonitrile were obtained from Fisher Scientific (Fair Lawn, NJ, USA). The reversed phase chromatography column (ProSwift™ RP-1S) was purchased from Dionex Corporation (Sunnyvale, CA, USA).

2.2 Methods

2.2.1 Monolithic affinity columns with surface immobilized lectins—A

polymerization mixture weighing 5g and containing 7.6% w/w GMM, 7% w/w PETA, 59.1% w/w cyclohexanol 22.9% w/w dodecanol and 3.4% w/w water containing 1.0% w/w AIBN with respect to the monomers [21] was sonicated for 15 min, purged with nitrogen for 5 min and introduced into a stainless steel columns of 25 cm × 4.6 mm ID each that function as molds for the monoliths and were heated at 60 °C for 15 h in a gas chromatography oven. The resulting monolithic columns were washed extensively with acetonitrile followed by water. The monolithic support was transferred from the 25 cm column to a shorter column (3 cm × 4.6 mm ID) by connecting the two columns with a ¼"-union and running water through the columns at flow rate of 3.0 mL/min until the modified monolithic support is transferred. Two 3 cm × 4.6 mm ID columns were prepared by this procedure.

The 3 cm monolithic columns were perfused with a freshly prepared 0.1 M NaIO₄ solution for 2 h at room temperature to convert the diol into aldehyde groups followed by a 5 min water wash at 1 mL/min using a reciprocating HPLC pump. Then, the lectin immobilization was done on column by recycling 2 times a 0.5 mL immobilization solution of the lectin of interest through the column at 0.5 mL/h with a micro-syringe pump at 4 °C. This was followed by an additional treatment for overnight of the column with the same lectin immobilization solution at 0.05 mL/h with the same micro-syringe at 4°C. The lectin immobilization solution was made of 1 mg of AAL or LTA in 0.5 mL of 0.1 M sodium acetate pH 6.4, containing 0.1 M fucose and 50 mM sodium cyanoborohydride. After this lectin immobilization step, a 10 mL solution containing 0.4 M Tris/HCl, pH 7.2, and 50 mM sodium cyanoborohydride was pumped with a micro-syringe through the columns at 3 mL/h for 3 h at room temperature. The immobilized lectin columns thus obtained were stored in 20 mM Tris-HCl, pH 7.4, and 0.08% NaN₃ at 4 °C until use. More information on lectin immobilization on various supports can be found in a recent review article [22].

2.2.2 Sample treatment with the combinatorial peptide ligand libraries—The

bulk ProteoMiner™ beads were swelled in 20% v/v aqueous ethanol (the storage solution) at 4° C for 12 h. 500 µL of the final slurry, which have 100 µL of beads, were introduced into a single spin column. Six spin columns were prepared in this way. Out of these six spin columns, 3 columns were used to treat the disease-free serum and 3 columns were used for

the breast cancer serum treatment. For each spin column, the storage solution was removed by centrifuging at $1000 \times g$ for 60 sec. Then, 600 μL of 10 mM NaH_2PO_4 , 150 mM NaCl, pH 7.4 (wash buffer) were added, and the spin column was rotated end-to-end several times for 5 min, followed by centrifugation at $1000 \times g$ for 60 sec. This washing step was repeated once again. The serum samples were centrifuged at $3650 \times g$ for 10 min. Thereafter, 1 mL of the centrifuged serum was treated with 100 μL of the settled beads (i.e., a spin column) by rotating each of the spin columns end-to-end at room temperature for 2 h. The unbound proteins were centrifuged at $1000 \times g$ for 60 sec. To each of the spin columns, 600 μL of wash buffer were added and rotated end-to-end for 5 min followed by centrifugation at $1000 \times g$ for 60 sec. This wash was repeated for another three times. The same procedure was followed with 600 μL of deionized water. This was followed by elution with $3 \times 100 \mu\text{L}$ of 9 M urea, 2% (w/v) CHAPS in 5% (v/v) acetic acid (elution reagent). The elution reagent was added to the column and incubated at room temperature for 15 min with light vortexing in between. The three elution volumes from the three spin columns for disease free serum and the three spin columns for the breast cancer serum were pooled and stored at -20°C until further use. These equalized protein solutions were treated with detergent removal AcroSep™ SDR columns to remove the urea and CHAPS according to the manufacturer instructions. The protein solution which was deprived of the detergent and urea was brought up to $\sim \text{pH } 7$ by adding Tris to it. The resulting solutions of 5.6 mL for each type of sera were injected into the lectin columns (see next section).

2.2.3 Chromatographic conditions: Enrichment of the fucosylated glycoproteins—

The HPLC setup consisted of a quaternary solvent delivery system Q-grad pump from Lab Alliance (State College, PA, USA), a solvent delivery system Model CM4000, a Model 3100 UV-Vis variable wavelength detector from Milton Roy, LDC division (Riviera Beach, FL, USA) and a Rheodyne injector Model 7010 (Cotati, CA, USA) equipped with a 1-mL loop. All the switching valves were from Rheodyne, while one 3-way valve was from SSI (State College, PA, USA). The Spectra/Chrom CF-1 fraction collector was from spectrum chromatography (Houston, TX, USA). All mass spectra were obtained using a hybrid LTQ-Orbitrap mass spectrometer from Thermo Fisher Scientific (Waltham, MA, USA).

The equalized serum fucosylated glycoproteins were enriched and fractionated by the following order of tandem columns, LTA (3 cm \times 4.6 mm ID) \rightarrow AAL (3 cm \times 4.6 mm ID) \rightarrow RPC column (5 cm \times 4.6 mm ID). The chromatographic set-up is shown in Fig. 1. The RPC column, which was initially stored in ACN, was washed first with 20 mL of the aqueous mobile phase (mobile phase A) that consisted of $\text{H}_2\text{O}/\text{ACN}$ (95:5 v/v) containing 0.1% v/v TFA. The eluting mobile phase (mobile phase B) for the RPC column consisted of $\text{ACN}/\text{H}_2\text{O}$ (95:5 v/v) containing 0.1% v/v TFA. The binding mobile phase for the lectin columns consisted of 20 mM Tris containing 0.3 M NaCl, pH 7.4, while the eluting mobile phase consisted of 5 mM fucose in 20 mM Tris, pH 7.4.

The equalized serum was injected onto the lectin columns, followed by washing with the binding mobile phase for 20 min at a flow rate of 1 mL/min. Then, the proteins from LTA and AAL column were transferred to the RPC column by passing 20 mL of the eluting mobile phase. The RPC column was washed with mobile phase A to remove the salts for 20 min at 1 mL/min. This was followed by a linear gradient elution of the RPC column at increasing ACN concentration in the mobile phase. The linear gradient consisted of increasing the % of mobile phase B from 0% to 75% v/v in 12 min. This was followed by 2 min at 75% mobile phase B, and then returning to initial condition (0% mobile phase B or 100% A) in 1 min. Before the next gradient run, the RPC column was washed with 100% mobile phase A for 20 min. Now, the accumulated proteins from the AAL lectin column were eluted onto the RPC column. The elution of the RPC column was performed in the

same way as described above. The proteins from the RPC column were collected every 24 s and evaporated to dryness in a SpeedVac from Savant Instruments, Inc. (Holbrook, NY) and stored at -20°C until further use.

2.3 Other methods

The descriptions of (i) the digestion of protein fractions by trypsin, (ii) LC-MS/MS methodology, and (iii) LC-MS/MS data analysis that are pertinent to this investigation are provided on line as “Supporting Information”.

3. Results and Discussion

3.1 Protein identification by LC-MS/MS analysis

3.1.1 Proteins identified in the RPC fractions captured by the LTA column—

The proteins captured by the lectin columns and their subsequent RPC fractionation were obtained *via* the platform shown in Fig. 1. The operational aspects of the platform are detailed in the experimental section as well as in the legend of Fig. 1. The RPC chromatograms obtained from the LTA captured proteins from disease-free and cancer sera are shown in Fig. 2. This figure shows clearly that more proteins were captured from the cancer serum when compared to disease-free serum. In fact, 188 and 212 non-redundant proteins were identified by LC-MS/MS in the disease-free and cancer sera, respectively. When compared, the 188 and 212 non-redundant proteins revealed the presence of 19 and 43 non-redundant unique proteins in disease-free and cancer sera, respectively, and 169 common proteins. Only proteins that exhibited protein and peptide identification probability of at least 99 % and 95 %, respectively, and containing at least two unique peptides were considered and are listed in Table S-1 (see Supporting information). The concentrations of some of the identified proteins such as ADAMTS-like protein 2, ADAMTS-like protein 4, calumenin, drebrin-like protein, hyaluronidase-1, neuropilin-2 and out at first protein homolog have not been reported yet in the literature. Apart from some Ig's identified in the RPC fractions, 75 proteins were reported to be in the range of a few ng to $< 2\ \mu\text{g/mL}$ [23], and these protein have been marked as low abundance (la) proteins in Table S-1. Also, some of the identified proteins in the LTA fractions were not listed in the compilation of human plasma proteins reported in Ref. [23]. These proteins have been marked in the Table S-1 as not listed (nl). The SWISSPROT database that provides the *N*- and *O*-glycosylation of proteins and NetNGlyc and NetOGlyc software, which could predict the potential *N*-glycosylation and *O*-glycosylation sites in proteins, respectively, were used to identify the percentage of glycoproteins captured by the lectin columns. 85.6% of the identified proteins in the disease-free serum were glycoproteins and the remaining 14.4% were non-glycoproteins. In the case of cancer serum, 88.2% of the proteins were glycoproteins and 11.8% of the proteins were non-glycoproteins. As reported earlier by us [7], the capturing of non glycoproteins by the two lectin columns might be due in part to protein-protein interactions that are often observed in complex serum samples and protein-based affinity assays [24, 25].

3.1.2 Proteins identified in the RPC fractions captured by the AAL column—

For comparison, the RPC chromatograms obtained from the disease-free and cancer sera for the AAL captured proteins are shown in Fig. 3. As in the case of the LTA captured proteins, it can be seen from this figure that higher amounts of proteins were captured from the cancer serum when compared to the disease free serum. In fact, 188 and 211 non-redundant proteins were identified by LC-MS/MS in the RPC fractions from the disease-free and cancer sera, respectively, with 164 common proteins for both sera (see Table S-2 in Supporting information). Thus, 24 and 47 unique proteins were found in the disease-free and cancer sera, respectively. The concentrations of some of the proteins such as calumenin,

arginase-1, nucleobindin-2 and ran-specific GTPase-activating proteins were known. Some of the proteins such as dermatopontin, extracellular glycoprotein lacritin, golgi apparatus protein 1, haptoglobin-related protein, papilin, pregnancy-specific beta-1-glycoprotein 1 and 6, reticulocalbin-1, serum paraoxonase/arylesterase 1, ubiquitin-60S ribosomal protein L40, UV excision repair protein RAD23 homolog A and some keratins were not listed in the Ref. [23]. Also, 87 proteins were reported to be in the range of a few ng to $< 2 \mu\text{g/mL}$ [23], and these proteins have been marked as low abundance (la) proteins in Table S-2 (see Supporting information). Some of the proteins were not listed in the Ref. [23], and these have been marked nl in Table S-2. From the SWISSPROT database and using NetNGlyc and NetOGlyc software, it was found that out of the 188 proteins that were identified in the disease-free serum, 84% were glycoproteins and the remaining 16% were non-glycoproteins. In the case of cancer serum, 83.4% were found to be glycoproteins and 16.6% were non-glycoproteins. As stated above, the binding of non-glycoproteins to the lectin columns may be attributed to protein-protein interactions [24, 25].

3.2 Differentially expressed proteins in breast cancer serum

Proteins exhibiting at least 99.9% protein identification probability, 95% peptide identification probability, and a minimum of 5 unique peptides were considered for the analysis of DEPs. The DEPs were identified *via* the Q-Q scatterplot whereby the normalized MS spectral counts of the proteins found in the cancer serum are plotted against the normalized MS spectral counts of these same proteins found in disease-free serum. The proteins that were more than two standard deviations away from being the same in both categories were considered as DEPs. In addition, all these proteins had a p-value < 0.05 using the t-test. Four typical scatterplots are shown in Fig. 4. In two recent reports from our laboratories, the Q-Q plots provided reliable basis for the identification of DEPs [7, 16]. The Q-Q plots shown in Fig. 4 belong to proteins found in the RPC fractions 3 and 6 of the LTA captured proteins and to proteins found in the RPC fractions 3 and 7 of the AAL captured proteins. It is not surprising to notice that only 7 DEPs are found in fraction 3 versus twice as much DEPs in fraction 6 (i.e., 14 DEPs) captured by the LTA column as this fact is well reflected in the intensity increment of peaks corresponding to fractions 3 and 6 shown in Fig. 2 for cancer serum versus disease-free serum. For the fractions 3 and 7 originating from the AAL column an equivalent number of DEP (7 DEPs in each fraction) are found in those fractions and this is well substantiated by the intensity of the peaks corresponding to the fractions 3 and 7 for cancer serum when compared to those of the disease free serum. Thus, the quick assessment of the difference between breast cancer and disease-free sera by a simple visual inspection of the corresponding RPC chromatograms of targeted captured proteins may prove useful in expediting the screening of a large number of biological samples. As already reported in our very recent contribution [7] to the identification of DEPs, this quick visual inspection is very instrumental in reducing the exuberant cost and length of the LC-MS/MS analysis in the sense that only those fractions corresponding to the RPC peaks with apparent intensity difference would be analyzed by LC-MS/MS rather than considering the analysis of all collected fractions.

The LC-MS/MS analysis of the fractions obtained from the RPC fractionation of the LTA captured proteins from the cancer serum show 35 proteins that were either up or down-regulated in the cancer serum with respect to disease-free serum, see Table 1. Out of these 35 DEPs, 10 proteins were predicted to be at the concentration level ranging from a few ng/mL to $< 2 \mu\text{g/mL}$ (i.e., low abundance proteins). In these 10 DEPs, the proteins C4b-binding protein alpha and beta chain, EGF-containing fibulin-like extracellular matrix protein 1, extracellular matrix protein 1, extracellular superoxide dismutase [Cu-Zn], fibrillin-1, fibulin-1 and thrombospondin-1 are glycoproteins and only 2 proteins (desmoplakin and nucleobindin) are non-glycoproteins. The low abundance protein thrombospondin-1, and the

two borderline low abundance proteins coagulation factor XIII B chain and pregnancy zone protein have been reported to be potential cancer biomarkers [23].

Upon a visual examination of the RPC chromatograms in Fig. 2 generated for the LTA captured proteins, it can be seen that in some of the fractions such as 3, 6 and 7, the peak intensity of the fractions from the cancer serum is higher as compared to those from the normal serum. In these fractions, a total of 12 proteins were found to be up-regulated. Out of these 12 proteins, 6 proteins are of the low abundance proteins, 3 proteins are borderline abundance between medium and low abundance proteins, 2 proteins are medium abundance proteins, and the concentration of actin cytoplasmic 1 is unknown.

There were 29 DEPs in the RPC fractions captured by the AAL column from cancer serum. Nine of these DEPs are in the concentration range of few ng/mL to < 2 µg/mL (i.e., low abundance proteins), and out of these 9 DEPs, the proteins C4b-binding protein alpha chain, latent-transforming growth factor beta-binding protein 1, prothrombin, SPARC-like protein 1, thrombospondin-1, vitronectin and von Willebrand factor are glycoproteins while the remaining 2 proteins namely, hornerin and protein disulfide-isomerase are non-glycoproteins. Also, the proteins thrombospondin-1 and von Willebrand factor have been reported to be potential cancer biomarkers [23].

Six proteins were found to be commonly differentially expressed in both LTA and AAL columns. These common DEPs are C4b-binding protein alpha chain, complement C3, complement C5, fibronectin, inter-alpha-trypsin inhibitor heavy chain H4 and thrombospondin-1, see Table 1. Upon quick visual inspection of the RPC chromatograms in Fig. 3 corresponding to the AAL captured proteins, one can see that some of the peaks in the RPC chromatogram of the cancer serum are of higher intensity as compared to those of the RPC chromatogram of the disease-free serum. Some of the proteins that were found to be up-regulated in fractions 3, 5, 6 and 7 included some of the low abundance proteins such as C4b-binding protein alpha chain, thrombospondin-1, protein disulfide-isomerase, vitronectin and von Willebrand factor, some of the borderline abundance proteins between low and medium abundance proteins such as fibronectin, apolipoprotein C-II, apolipoprotein E and vimentin and the high abundance protein apolipoprotein A-I.

Overall, a total of 58 proteins were found to be differentially expressed in both the LTA and AAL fractions with 6 DEPs common to both lectins. 17 DEPs were of the low abundance type, 16 DEPs of the borderline abundance type, 4 DEPs of the medium abundance type and 15 DEPs of the high abundance type. The remaining 6 DEPs are of unknown concentration or are not reported yet in the literature.

In a recent work from our laboratories [7] that involved the online (i) depletion of albumin and Ig's from both disease-free and cancer sera, (ii) capturing/enrichment of fucosylated glycoproteins using LTA and AAL lectin columns and (iii) subsequent RPC fractionation of the lectin captured proteins, 35 proteins were found to be differentially expressed in cancer serum when compared to normal serum. Upon comparing the DEPs from this earlier study to those of the current study, it was found that 23 DEPs (see Table 2) were common to both studies. Also, there were 12 and 35 DEPs that were unique to the earlier and current study, respectively. Out of the 35 DEPs that were unique to the current study, 14 proteins were in low abundance range, 12 proteins were borderline abundance between low and medium abundance proteins, 2 proteins were medium abundance proteins and 4 proteins were high abundance proteins. The proteins actin cytoplasmic 1, papilin and serum paraoxonase/arylesterase 1 were not listed in ref [23] and their concentration is unknown. Whereas, out of the 12 proteins that were unique to the earlier study [7], 8 proteins were on the borderline between low and medium abundance proteins, one protein was medium abundance protein,

two proteins were high abundance proteins and only one protein was low abundance protein. Clearly, this indicates that the treatment of the serum with the CPLL beads allowed for the enrichment of the low abundance proteins as well as high and medium abundance proteins. This enrichment of the low abundance proteins allowed the identification of a larger number of DEPs in the low abundance range. In the earlier study just mentioned from our lab [7], only three DEPs were in the low abundance range. This is because even though that albumin and Ig's were removed from the serum, still the medium abundance proteins have probably masked the interactions of the low abundance proteins with the lectins making the detection of the low abundance proteins a difficult task. But in the current study where the proteins were equalized in terms of their concentrations, low, medium and high abundance DEPs were readily identified.

4. Conclusions

It is clear from the above results and discussion that the combination of CPLL with a liquid phase multi column platform composed of LAC for the selective enrichment of human fucose followed by RPC fractionation represents an effective approach to facilitate the detection of DEPs by LC-MS/MS that cover a wide range of abundance spanning from very low- to borderline low-, medium- and high-abundance proteins. The combination of CPLL with the multicolumn platform is superior to the multicolumn platform that involves an additional 3 depletion columns for albumin and Ig's in terms of the total number of DEPs. However, both approaches may be viewed as complementary to each other bringing the total numbers of DEPs to 70 proteins, thus widening the range of DEPs to a multiplicity of protein abundance levels. It should be emphasized that this investigation has provided an optimized workflow for large-scale studies that could validate the reported findings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The financial support of this research by a Grant No. 1R15GM096286-01 from the National Institutes of Health is greatly appreciated.

Nonstandard abbreviations

AAL	<i>Aleuria aurantia</i> lectin
AIBN	2,2'-azobis(isobutyronitrile)
CPLL	combinatorial peptide ligand library
DEPs	differentially expressed proteins
Fuc	fucose
Gal	galactose
GlcNAc	<i>N</i> -acetylglucosamine
GMM	glycerylmethacrylate
Ig's	immunoglobulins
LAC	lectin affinity chromatography
LTA	<i>Lotus tetragonolobus</i> agglutinin

PETA	pentaerythritol triacrylate
RPC	reversed phase chromatography
TFA	trifluoroacetic acid

References

1. Apweiler R, Hermjakob H, Sharon N. *Biochim Biophys Acta*. 1999; 1473:4–8. [PubMed: 10580125]
2. Block TM, Comunale MA, Lowman M, Steel LF, Romano PR, Fimmel C, Tennant BC, London WT, Evans AA, Blumberg BS, Dwek RA, Mattu TS, Mehta AS. *Proc Natl Acad Sci, USA*. 2005; 102:779–784. [PubMed: 15642945]
3. Listinsky JL, Siegal GP, Listinsky CM. *Am J Transl Res*. 2011; 3:292–322. [PubMed: 21904652]
4. Mann B, Madera M, Klouckova I, Mechref Y, Dobrolecki LE, Hickey RJ, Hammoud ZT, Novotny MV. *Electrophoresis*. 2010; 31:1833–1841. [PubMed: 20446296]
5. Miyoshi E, Moriwaki K, Terap N, Tan CC, Terao M, Nakagawa T, Matsumoto H, Shinzaki S, Kamada Y. *Biomolecules*. 2012; 2:34–45.
6. Muinelo-Romay L, Villar-Portela S, Gil-Martin E, Fernandez-Briera A. *BMC Cancer*. 2011; 11:508. [PubMed: 22152070]
7. Selvaraju S, El Rassi Z. *Proteomics*. 2013; 13:1701–1713. [PubMed: 23533108]
8. Righetti PG, Boschetti E, Lomas L, Citterio A. *Proteomics*. 2006; 6:3980–3992. [PubMed: 16800034]
9. Righetti PG, Fasoli E, Boschetti E. *Electrophoresis*. 2011; 32:960–966. [PubMed: 21449065]
10. Selvaraju S, El Rassi Z. *Electrophoresis*. 2010; 32:674–685. [PubMed: 21365658]
11. Selvaraju S, El Rassi Z. *Electrophoresis*. 2012; 33:74–88. [PubMed: 22125262]
12. Abbott KL, Aoki K, Lim JM, Porterfield M, Johnson R, O'Regan RM, Wells L, Tiemeyer M, Pierce M. *J Proteome Res*. 2008; 7:1470–1480. [PubMed: 18271524]
13. Comunale MA, Wang M, Hafner J, Krakover J, Rodemich L, Kopenhaver B, Long RE, Junaidi O, Bisceglie AM, Block TM, Mehta AS. *J Proteome Res*. 2009; 8:595–602. [PubMed: 19099421]
14. Qiu R, Regnier FE. *Anal Chem*. 2005; 77:2802–2809. [PubMed: 15859596]
15. Schwientek T, Mandel U, Roth U, Müller S, Hanisch FG. *Proteomics*. 2007; 7:3264–3277. [PubMed: 17708590]
16. Selvaraju S, El Rassi Z. *J Sep Sc*. 2012; 35:1785–1795. [PubMed: 22807361]
17. Yang Z, Hancock WS. *J Chromatogr A*. 2004; 1053:79–88. [PubMed: 15543974]
18. Gunasena D, El Rassi Z. *Electrophoresis*. 2012; 33:251–261. [PubMed: 22147366]
19. Kobata, A.; Yamashita, K. Fractionation of oligosacchrides by serial affinity chromatography with use of immobilized lectin columns. In: Fukuda, M.; Kobata, A., editors. *Glycobiology A Practical Approach*. IRL Press; Oxford: 1993. p. 103-125.
20. Yan L, Wilkins PP, Alvarez-Manilla G, Do SI, Smith DF. *Glycoconjugate J*. 1997; 14:45–55.
21. Gunasena D, El Rassi Z. *J Sep Sci*. 2011; 34:2097–2105.
22. Monzo A, Bonn GK, Guttman A. *Trends Anal Chem*. 2007; 26:423–432.
23. Farrah T, Deutsch EW, Omenn GS, Campbell DS, Sun Z, Bletz JA, Mallick P, Katz JE, Malmstrom J, Ossola R, Watts JD, Lin B, Zhang H, Moritz RL, Aebersold RH. *Molecular & Cellular Proteomics*. 2011; 10:1–14.
24. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE. *Cell*. 2005; 122:957–968. [PubMed: 16169070]
25. Stumpf MPH, Throne T, de Silva E, Stewart R, Jun An H, Lappe M, Wiuf C. *Proc Natl Acad Sci, USA*. 2008; 105:6959–6964. [PubMed: 18474861]
26. Cho W, Jung K, Regnier FE. *J Proteome Res*. 2010; 9:5960–5968. [PubMed: 20858014]

27. Jia W, Lu Z, Fu Y, Wang HP, Wang LH, Chi H, Yuan ZF, Zheng ZB, Song LN, Han HH, Liang YM, Wang JL, Cai Y, Zhang YK, Deng YL, Ying WT, He SM, Qian XH. *Mol Cell Proteomics*. 2009; 8:913–923. [PubMed: 19139490]

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at

HIGHLIGHTS

- Novel approach incorporating CPLL and a multicolumn platform was introduced
- Identification of differentially expressed fucosylated glycoproteins (the fucome)
- Fucome was captured by lectin columns and fractionated by RPC for LC-MS/MS analysis
- 58 differentially expressed proteins (DEPs) were identified
- The 58 DEPs can be viewed as 58 candidate breast cancer biomarkers

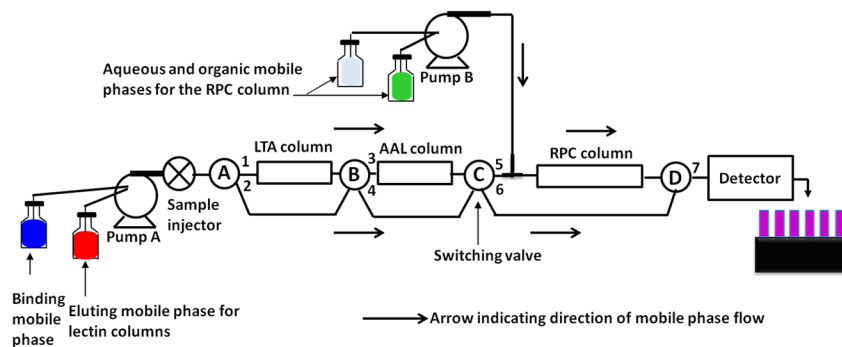


Figure 1.

Schematic representation of the platform for the simultaneous enrichment of fucosylated glycoproteins and the subsequent RPC fractionation of the captured proteins. When the switching valve (SV) A, B, and C were in 1, 3 and 6 positions, respectively, the equalized serum was injected onto the lectin columns by-passing the RPC column, followed by washing with the binding mobile phase (BMP) using pump A. Then, the eluting mobile phase for the LTA column was passed by keeping the SV-A position in 1 and changing the SV-B position to 4 and that of SV-C to 5, thus by-passing the AAL column and passing through the RPC column. This was followed by washing, eluting and re-equilibrating the RPC column using the mobile phase from pump B. Then, the AAL column was eluted by changing the SV-A position to 2, keeping the SV-B position in 3 and SV-C position in 5. This was again followed by washing, eluting and re-equilibrating the RPC column and re-equilibrating the lectin columns with the BMP.

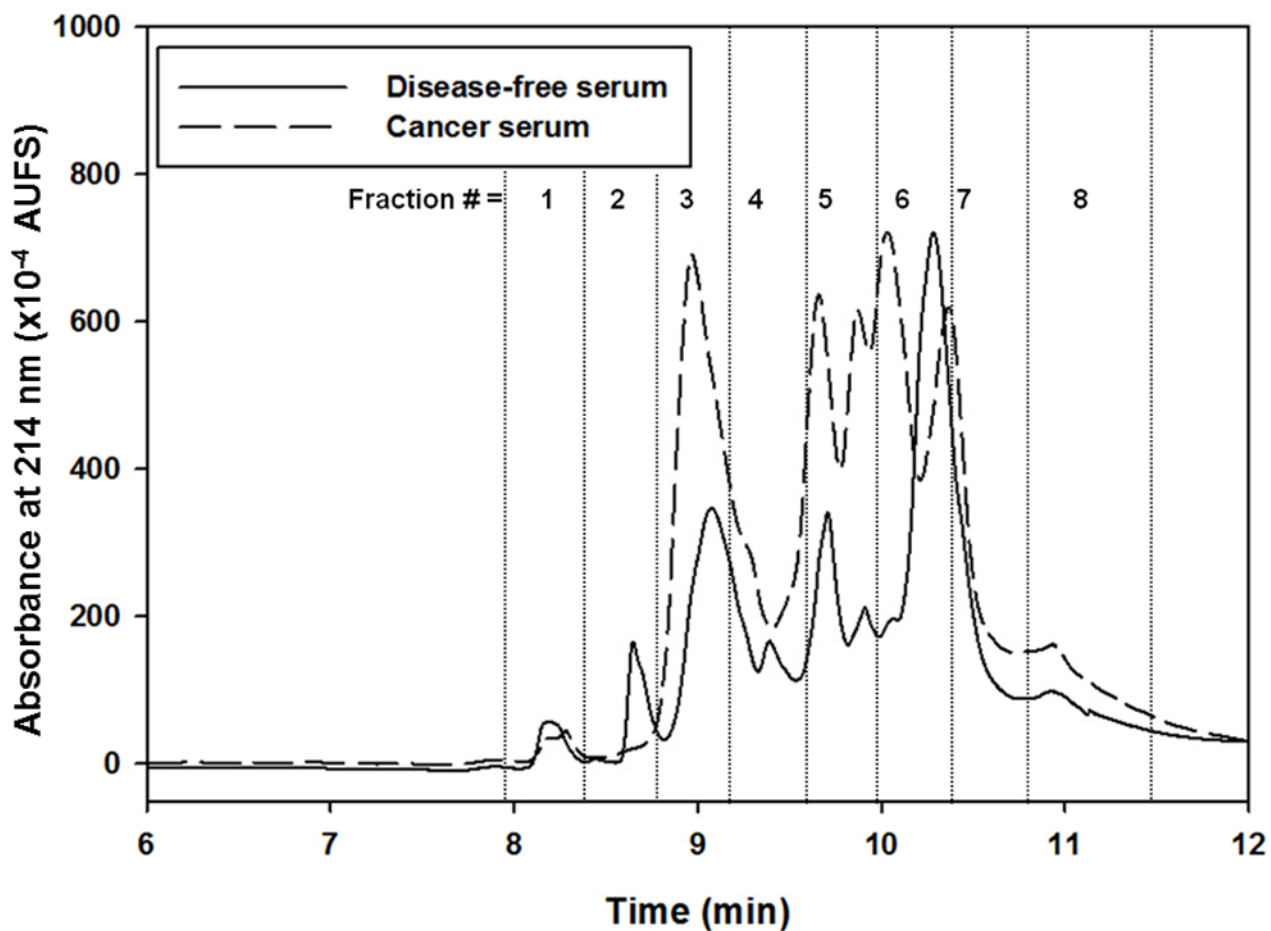


Figure 2.

Chromatograms of RPC gradient for LTA captured band from equalized disease-free and cancer serum. The RPC gradient consisted of 0 % to 75 % of mobile phase B in 12 min. Mobile phase A consisted of H₂O/ACN (95:5 v/v) containing 0.1% TFA and mobile phase B consisted of ACN/H₂O (95:5 v/v) containing 0.1% TFA. Flow rate 1 mL/min and wavelength 214 nm.

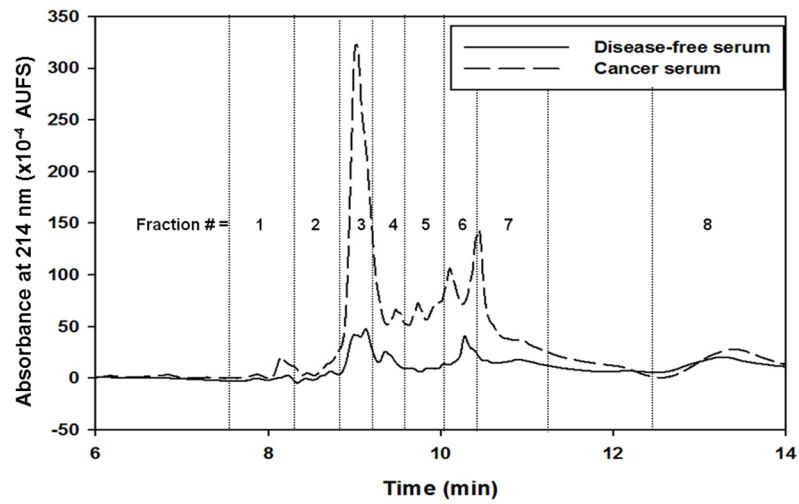


Figure 3. Chromatograms of RPC gradient for AAL captured band from equalized disease-free and cancer serum. Other conditions same as Fig. 1.

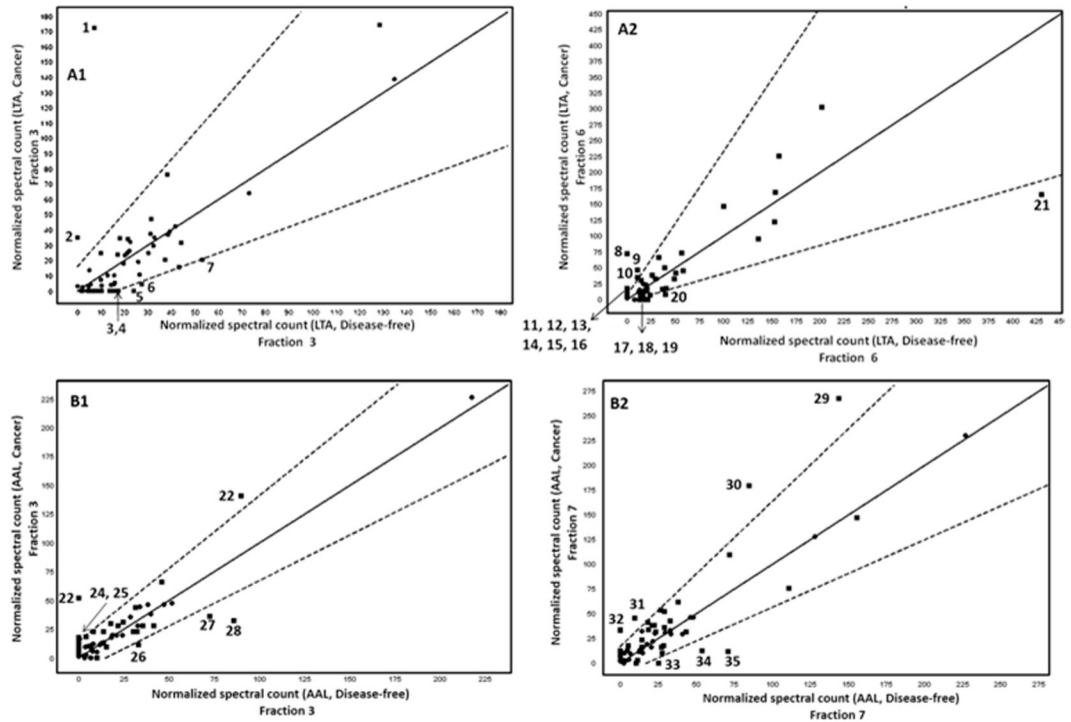


Figure 4.

Representative Q-Q scatterplots for the identification of DEPs. The squares that plot outside the dotted lines are more than two standard deviations away from being the same in both categories. The dotted lines delimit the upper and lower error bars. A1 and A2: Q-Q scatterplots for fractions 3 and 4 from RPC fractionation from LTA column; B1 and B2: Q-Q scatterplots for fractions 3 and 7 from RPC fractionation from AAL column. The DEPs in each of the scatterplots have been marked using numbers (1, 2, 3 etc.). The DEPs in A1 are as follows: 1, C4b-binding protein alpha chain; 2, C4b-binding protein beta chain; 3, apolipoprotein C-III; 4, nucleobindin-1; 5, haptoglobin-related protein; 6, alpha-1-antitrypsin; 7, complement C3. The DEPs in A2 are: 8, C4b-binding protein alpha chain; 9, inter-alpha-trypsin inhibitor heavy chain H4; 10, fibulin-1; 11, thrombospondin-1; 12, extracellular matrix protein 1; 13, pregnancy zone protein; 14, CD5 antigen-like; 15, coagulation factor V; 16, EGF-containing fibulin-like extracellular matrix protein 1; 17, angiotensinogen; 18, afamin; 19, complement C5; 20, alpha-1-antitrypsin; 21, alpha-1-antichymotrypsin. The DEPs in B1 are: 22, Vitronectin; 23, C4b-binding protein alpha chain; 24, von Willebrand factor; 25, apolipoprotein A-II; 26, complement component C6; 27, haptoglobin; 28, complement C3. The DEPs in B2 are: 29, fibronectin; 30, apolipoprotein A-I; 31, von Willebrand factor; 32, thrombospondin-1; 33, hornerin; 34, apolipoprotein B-100; 35, complement C5.

Table 1

DEPs in the LTA and AAL fractions.

DEPs unique to LTA fractions	DEPs unique to AAL fractions	DEPs common to both LTA and AAL fractions
Actin, cytoplasmic ¹ <i>nl</i>	Alpha-2-macroglobulin ¹ <i>ha</i> (CF, F)	C4b-binding protein alpha chain ¹ <i>a</i> (F)
Afamin <i>bla</i> (CF, F)	Apolipoprotein A-1 ¹ <i>ha</i> (F)	Complement C3 ¹ <i>ha</i> (F)
Alpha-1-antichymotrypsin ¹ <i>ha</i> (F)	Apolipoprotein A-II ¹ <i>ha</i> (F)	Complement C5 ¹ <i>nl</i> (F)
Alpha-1-antitrypsin ¹ <i>ha</i> (CF, F)	Apolipoprotein A-IV [*] , <i>ha</i>	Fibronectin <i>bla</i> (F)
Angiotensinogen <i>bla</i> (F)	Apolipoprotein B-100 <i>ha</i> (F)	Inter-alpha-trypsin inhibitor heavy chain H4 ¹ <i>ma</i> (F)
Apolipoprotein C-III ¹ <i>ma</i> (F)	Apolipoprotein C-II ¹ <i>bla</i> (F)	Thrombospondin-1 ¹ <i>a</i> (F)
Apolipoprotein D ¹ <i>ma</i> (CF, F)	Apolipoprotein E <i>bla</i> (F)	
C4b-binding protein beta chain ¹ <i>a</i> (F)	Ceruloplasmin ¹ <i>ha</i> (CF, F)	
CD5 antigen-like [*] , <i>bla</i>	Complement component C6 <i>bla</i> (F)	
Coagulation factor V ¹ <i>bla</i>	Haptoglobin ¹ <i>ha</i> (CF)	
Coagulation factor XIII B chain ¹ <i>bla</i>	Homerin ¹ <i>a</i> ,*	
Complement component C7 ¹ <i>bla</i> (CF, F)	Hyaluronan-binding protein 2 ¹ <i>bla</i>	
Complement factor H ¹ <i>bla</i> (CF, F)	Ig lambda-2 chain C regions ^{unb} ,*	
Desmoplakin ¹ <i>a</i> ,*	Inter-alpha-trypsin inhibitor heavy chain H2 ¹ <i>ha</i> (F)	
EGF-containing fibulin-like extracellular matrix protein 1 ¹ <i>a</i>	Latent-transforming growth factor beta-binding protein 1 ¹ <i>a</i>	
Extracellular matrix protein 1 ¹ <i>a</i> (CF, F)	Protein disulfide-isomerase ¹ <i>a</i> ,*	
Extracellular superoxide dismutase [Cu-Zn] ¹ <i>a</i>	Prothrombin ¹ <i>a</i> (CF, F)	
Fibrillin-1 ¹ <i>a</i>	Serotransferrin ¹ <i>ha</i> (CF, F)	
Fibrinogen alpha chain ¹ <i>ha</i> (F)	Serum albumin [*] , <i>ha</i>	
Fibulin-1 ¹ <i>a</i>	SPARC-like protein 1 ¹ <i>a</i>	
Haptoglobin-related protein ¹ <i>nl</i> (F)	Vimentin ¹ <i>bla</i>	
Ig mu chain C region ¹ <i>bla</i> (CF, F)	Vitronectin ¹ <i>a</i> (CF, F)	
Inter-alpha-trypsin inhibitor heavy chain H1 ¹ <i>ma</i> (CF, F)	von Willebrand factor ¹ <i>a</i> (CF)	

DEPs unique to LTA fractions	DEPs unique to AAL fractions	DEPs common to both LTA and AAL fractions
Nucleobindin-1 <i>la</i> ,*		
Papilin ^{nl}		
Phosphatidylinositol-glycan-specific phospholipase D <i>bla</i>		
Pregnancy zone protein <i>bla</i>		
Serum paraoxonase/arylesterase 1 ^{nl} (F)		
Transferrin ^{ha} (F)		

la Low abundance (few ng/mL to < 2 µg/mL level);

bla borderline between low and medium abundance proteins (2 µg/mL to < 0.1 mg/mL);

ma medium abundance (0.1 mg/mL to < 0.2 mg/mL);

ha high abundance (0.2 mg/mL to > 1 mg/mL).

nl not listed. The scale for protein abundance used here is arbitrary and is simply introduced to facilitate the comparison and discussion.

CF core fucosylated;

F fucosylated;

* nonglycoprotein. The protein concentration used are from published or predicted values from Ref. [23]. Data on fucosylation are from Refs. [4, 12, 13, 26, 27] based on glycan analysis and/or the reactivity of the proteins with fucose specific lectins. All other glycoproteins listed in the table may be fucosylated but their fucosylation have not been reported yet in the literature.

Table 2

DEPs unique and common to the current study and previous studies.

#	DEPs unique to the current study (treated with CPLL beads)	DEPs common to both studies	DEPs unique to the previous study [7]
1	Actin, cytoplasmic β	Afamin ^{bla} (CF,F)	Alpha-2-antiplasmin ^{bla} (F)
2	Angiotensinogen ^{bla} (F)	Alpha-1-antichymotrypsin ^{bla} (F)	Alpha-2-HS-glycoprotein ^{bla} (CF, F)
3	Apolipoprotein A-II ^{bla} , (F)	Alpha-1-antitrypsin ^{bla} (CF,F)	Complement C1r subcomponent ^{bla} (F)
4	Apolipoprotein C-II ^{bla} (F)	Alpha-2-macroglobulin ^{bla} (CF,F)	Complement C1s subcomponent ^{bla} (F)
5	Apolipoprotein C-III ^{ma} (F)	Apolipoprotein A-I ^{bla} (F)	Complement C4-B ^{bla} (F)
6	Apolipoprotein D ^{ma} (CF, F)	Apolipoprotein A-IV ^{bla} , *	Hemopexin ^{bla} (CF)
7	Apolipoprotein E ^{bla} (F)	Apolipoprotein B-100 ^{bla} (F)	Lactotransferrin ^{la}
8	C4b-binding protein beta chain ^{la} (F)	C4b-binding protein alpha chain ^{la} (F)	Plasma kallikrein ^{bla} (F)
9	Coagulation factor V ^{bla}	CD5 antigen-like ^{bla} ,*	Plasminogen ^{ma} (F)
10	Coagulation factor XIII B chain ^{bla}	Ceruloplasmin ^{la} (CF,F)	Protein AMBP ^{bla}
11	Complement component C6 ^{bla} (F)	Complement C3 ^{la} (F)	Serum amyloid A-4 protein ^{bla} (F)
12	Complement component C7 ^{bla} (CF, F)	Complement C5 ^{la} (F)	Vitamin K-dependent protein S ^{bla} (F)
13	Complement factor H ^{bla} (CF)	Desmoplakin ^{la} , *	
14	EGF-containing fibulin-like extracellular matrix protein 1 ^{la}	Haptoglobin-related protein ^{ml} (F)	
15	Extracellular matrix protein 1 ^{la} (CF, F)	Ig lambda-2 chain C regions ^{smk} , *	
16	Extracellular superoxide dismutase [Cu-Zn] ^{la}	Inter-alpha-trypsin inhibitor heavy chain H1 ^{ma} (CF,F)	
17	Fibrillin-1 ^{la}	Inter-alpha-trypsin inhibitor heavy chain H2 ^{la} (F)	
18	Fibrinogen alpha chain ^{la} (F)	Inter-alpha-trypsin inhibitor heavy chain H4 ^{ma} (F)	
19	Fibronectin ^{bla} (F)	Phosphatidylinositol-glycan-specific phospholipase D ^{bla}	
20	Fibulin-1 ^{la} (CF)	Pregnancy zone protein ^{bla}	
21	Haptoglobin ^{la} (CF)	Prothrombin ^{la} (CF,F)	
22	Hormerin ^{la} , *	Serotransferrin ^{la} (CF,F)	
23	Hyaluronan-binding protein 2 ^{bla}	Serum albumin ^{la} , *	

#	DEPs unique to the current study (treated with CPL beads)	DEPs common to both studies	DEPs unique to the previous study [7]
24	Ig mu chain C region ^{<i>bla</i>} (CF, F)		
25	Latent-transforming growth factor beta-binding protein 1 ^{<i>la</i>}		
26	Nucleobindin-1 ^{<i>la</i>,*}		
27	Papilin ^{<i>nl</i>}		
28	Protein disulfide-isomerase ^{<i>la</i>,*}		
29	Serum paraoxonase/arylesterase 1 ^{<i>nl</i>} (F)		
30	SPARC-like protein 1 ^{<i>la</i>}		
31	Thrombospondin-1 ^{<i>la</i>}		
32	Transhyretin ^{<i>ha</i>} (F)		
33	Vimentin ^{<i>bla</i>}		
34	Vitronectin ^{<i>la</i>} (CF, F)		
35	von Willebrand factor ^{<i>la</i>} (CF)		

la Low abundance (few ng/mL to < 2 µg/mL level);

bla borderline between low and medium abundance proteins (2 µg/mL to < 0.1 mg/mL);

ma medium abundance (0.1 mg/mL to < 0.2 mg/mL);

ha high abundance (0.2 mg/mL to > 1 mg/mL).

nl not listed. The scale for protein abundance used here is arbitrary and is simply introduced to facilitate the comparison and discussion.

CF core fucosylated;

F fucosylated;

* nonglycoprotein. The protein concentration used are from published or predicted values from Ref.[23]. Data on fucosylation are from Refs. [4, 12, 13, 26, 27] based on glycan analysis and/or the reactivity of the proteins with fucose specific lectins. All other glycoproteins listed in the table may be fucosylated but their fucosylation have not been reported yet in the literature.