Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Journal of Chromatography A, 1053 (2004) 79–88

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column

Ziping Yang, William S. Hancock*

Barnett Institute and Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115, USA

Abstract

Glycosylation is one the most common post-translational modifications (PTM) and glycoproteins play fundamental roles in a diversity of biological processes. The development of an analytical approach to the study of variation of glycosylation patterns in serum samples has been hindered by the structural heterogeneity of this post-translational modification and the complexity of serum proteome. We have used the ability of different lectins to recognize specific glycosylation motifs to develop a specific affinity system that can achieve a comprehensive capture of serum glycoproteins. In a preliminary investigation, we evaluated the ability of five commonly used immobilized lectins to capture glycoproteins from human serum. SDS-PAGE analysis showed each lectin was able to enrich a subset of the serum glycoproteome and overlaps in lectin specificity were indeed observed. Based on these results and with the goal of studying the extent of the human serum glycoproteome, we then developed a multi-lectin affinity column containing Concanavalin A (Con A), Wheat germ and Jacalin lectin. The selection of lectins was also based on the known N-linked and O-linked glycan structures that are considered representative of the serum proteome. We then demonstrated that the capture of glycoproteins was specific, efficient and reproducible with this multi-lectin column. The results obtained with this affinity step indicated that about 10% of human serum proteins are glycosylated (weight/weight) and, after removal of six high abundance proteins, including albumin, at least 50% of the remaining proteins were glycosylated. We then evaluated the use of this affinity column to monitor changes in the pattern of glycosylation in serum samples by a controlled, stepwise release of the captured proteins from the multi-lectin affinity column with specific displacers.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Glycosylation; Glycoprotein; Multi-lectin affinity column; Concanavalin A; Wheat germ; Jacalin lectins; Proteomics

1. Introduction

Glycosylation, one of the most common post-translational modifications (PTM), plays a fundamental role in a diverse set of biological processes [1], such as the immune response [2] and cellular regulation [3]. Glycosylation is also a good indicator of environmental effects on cellular processes [4]. It has been shown that glycosylation is involved in signaling pathways associated with the transformation of a normal cell to a cancer cell [5], and glycosylation has been intimately associated with cancer, for example, glycosylation can affect tumor antigen interactions with receptors, e.g. CA125 with galectin [6].

While serum is the most common diagnostic fluid, there is currently a lack of global methods for the characterization of glycoproteins, the 'glycoproteome'. Serum can reflect disease in different organs and tissues with a change in secreted proteins, and it has been recognized that glycoproteins compromise a major part of the serum proteome [7].

In the past individual serum glycoproteins have been enriched using lectins which can recognize and bind to specific carbohydrate residues [8,9]. In this study, five lectins, Concanavalin A (Con A), Wheat germ agglutinin (WGA), Jacalin, Lentil lectin (LCA), and peanut lectin (PNA), have been selected to individually capture different glycoproteins from human serum. It has been shown that Con A predominately recognizes alpha-mannose [10], which is very common in N-linked glycans. WGA recognizes *N*-acetyl-glucosamine (GlcNAc) and was also found to have affinity to sialic acid

* Corresponding author. Tel.: +1 617 373 4881; fax: +1 617 373 2855.
E-mail address: wi.hancock@neu.edu (W.S. Hancock).

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.08.150

[11]. The specificity of Jacalin lectin is to galactosyl (β-1,3) *N*-acetyl-galactosamine (GalNAc) and has been used to capture O-linked glycoproteins [12]. LCA has the specificity similar to Con A but with lower affinity [13], although it has an useful affinity for branched fucose. The specificity of PNA is the same as Jacalin lectin, but the affinity is affected by sialic acid associated with galactose [12]. The results from this preliminary study have shown that each individual lectin captured a different subset of glycoproteins from serum, and Con A, WGA and Jacalin enriched larger amounts of glycoproteins than LCA and PNA, although overlaps in specificity were observed.

In order to capture and study a significant part of the human serum glycoproteome, a multi-lectin affinity column containing Con A, WGA, and Jacalin lectin was designed based on a consideration of common N-linked and O-linked glycan structures present in serum proteins. The following carbohydrate moieties are common to serum proteins: firstly N-linked glycans have a core structure which includes three mannosyl residues and two GlcNAc residues. Further biosynthetic processing from this core has been shown to result in high mannose, hybrid, and complex glycan structures [14]. O-linked glycans are very heterogeneous and only their core structures have been classified. In these core structures, galactosyl (β-1,3) GalNAc is very common, mannose and GlcNAc are frequently present in the O-linked core structures and most of the glycans terminate with sialic acid [15].

The multi-lectin column was then used to enrich glycoproteins from human serum. After tryptic digest, the samples were analyzed by LC/MS/MS and the proteins were identified using the SEQUEST algorithm. The unbound and captured fractions were compared to indicate the specificity and efficiency of the multi-lectin affinity column. In order to investigate reproducibility of the enrichment procedure, the complete experiment was repeated, and the capture was found to be reproducible. It was found that the multi-lectin column was highly specific for glycoproteins and extended the dynamic range of the proteome measurement. In addition, a depleted serum sample (removal of six high abundance serum proteins) was tested on the multi-lectin affinity column, and better protein identifications were observed, although the number of proteins sequenced was not improved. Further, glycoproteins captured by the multi-lectin column were fractionated by the sequential use of different displacers, and the distribution of glycoproteins among these fractions was found to be related to the glycosylation pattern of the glycoproteins.

2. Experimental

2.1. Materials

Human serum, dithiothreitol (DTT), iodoacetamide (IAA), sodium chloride, manganese chloride tetrahydrate, magnesium chloride, guanidinium hydrochloride, sodium azide, *N*-acetyl-glucosamine, galactose, methyl-α-mannopyranoside and calcium chloride were purchased from Sigma–Aldrich (St. Louis, MO). Ultra pure Tris and ammonium bicarbonate were purchased from ICN Biomedicals Inc. (Aurora, OH). Agarose bound Concanavalin A with protein concentration of 6 mg lectin/mL gel and binding capacity of more than 4 mg ovalbumin/mL gel, agarose bound Wheat germ agglutinin with protein concentration of 7 mg lectin/mL gel and binding capacity of 8 mg NGA/mL gel, agarose bound Jacalin with protein concentration of 4 mg lectin/mL gel and binding capacity of more than 4 mg monomeric IgA/mL gel, agarose bound peanut agglutinin (PNA) with protein concentration of 5 mg lectin/mL gel and binding capacity of more than 4 mg asialo-fetuin/mL gel and agarose bound lens culinaris agglutinin (LCA) with protein concentration of 3 mg lectin/ml gel and binding capacity of more than 3 mg mannosyl glycoprotein were obtained from Vector Laboratories (Burlingame, CA). Trypsin (sequence grade) was purchased from Promega (Madison, WI). Buffers A and B of the multiple affinity removal system were obtained from Agilent Technologies (Palo Alto, CA). NuPAGE MOPS SDS running buffer and molecular mass standards were purchased from Invitrogen (Carlsbad, CA).

2.2. Preparing lectin affinity columns

Single lectin, Con A, WGA, PNA, Jacalin, and LCA, affinity columns were prepared by adding 1 mL of corresponding agarose bound lectin to empty PD-10 disposable columns (Amersham Biosciences, Piscataway, NJ). The multi-lectin column was prepared by mixing 0.5 mL of agarose bound Con A, 0.5 ml agarose bound WGA, and 0.5 ml agarose bound Jacalin in an empty PD-10 disposable column. The agarose gel was then fixed between two frits. The columns were either immediately used or stored in a buffer (20 mM Tris, pH 7.4, 0.15 M NaCl, 0.08% sodium NaN_3) at 4 °C. The flow through the column was gravity driven. The columns were not regenerated with the view of avoiding carryover between different sample analyses.

2.3. Isolating glycoproteins using a single-lectin affinity column

100 μL human serum was diluted 10 times with the equilibration buffer for that lectin (see Table 1) and loaded on the affinity column. After 15 min reaction, the unretained proteins were eluted with 8 mL of equilibration buffer, and the flow-through was collected. The captured glycoproteins were then released with 8 mL of the elution buffer specific for that lectin (Table 1), and the eluted fraction was collected. The flow-through and eluted fractions were concentrated using a 10 kD Amicon filter (4 mL, Millipore, Billerica, MA) and stored at −70 °C until further use.

Table 1
The binding buffers and elution solutions used for single-lectin affinity columns

Lectin[a]	Binding buffer[b] (pH 7.4)	Sugars contained in elution buffer[c] (pH 7.4)
Con A	20 mM Tris, 0.15 NaCl, 1 mM Ca^{2+}, 1 mM Mn^{2+}	0.5 M methyl-α-D-mannopyranoside
WGA	20 mM Tris, 0.15 NaCl,	0.5 M *N*-acetyl-glucosamine
JAC	20 mM Tris, 0.15 NaCl,	0.8 M galactose
PNA	20 mM Tris, 0.15 NaCl, 1 mM Ca^{2+}, 1 mM Mg^{2+}	0.5 M lactose
LCA	20 mM Tris, 0.15 NaCl, 1 mM Ca^{2+}, 1 mM Mn^{2+}	0.5 M methyl-α-D-mannopyranoside

[a] The lectin immobilized in a single-lectin column.
[b] Binding buffer was used to prepare samples and remove non-specifically bound proteins.
[c] Elution buffer contained 20 mM Tris and 0.5 M NaCl.

2.4. Isolating glycoproteins using a multi-lectin affinity column

A depleted serum sample (see Section 2.5) or 200 μL of undepleted serum (containing about 6.7 mg of protein) was diluted with multi-lectin column equilibration buffer (20 mM Tris, 0.15 M NaCl, 1 mM Mn^{2+} and 1 mM Ca^{2+}, pH 7.4) to a volume of 2 mL, and was loaded on a multi-lectin affinity column. After 15 min reaction, the unbound proteins were eluted with 10 mL of equilibration buffer, and the captured proteins were released with 12 mL of elution solution (20 mM Tris, 0.5 M NaCl, 0.17 M methyl-α-D-mannopyranoside, 0.17 M *N*-acetyl-glucosamine and 0.27 M galactose, pH 7.4). The flow-through and eluted fractions were both collected and concentrated with 10 kD Amicon filters (15 mL, Millipore, Billerica, MA). The total amount of protein loaded on the column, the amount of protein collected in the flow through and the amount of protein collected in eluted fraction were measured using the Bradford assay. The recovery from the multi-lectin column was calculated using the equation:

$$\text{Recovery\%} = \left(\frac{\text{flow through} + \text{eluted protein}}{\text{total amount of protein loaded}}\right) \times 100$$

The collected samples were stored at −70 °C until further testing. The same procedure was repeated using undepleted human serum including packing a multi-lectin affinity column.

To fractionate the proteins captured by the multi-lectin column, proteins bound to Jacalin lectin were first released with 4 mL of 0.8 M galactose in 20 mM Tris buffer pH 7.4 containing 0.15 M NaCl. Then Con A selected proteins were released with 4 mL of 0.5 M methyl-α-D-mannopyranoside in a 20 mM Tris buffer, pH 7.4 containing 0.15 M NaCl. Finally, the WGA selected proteins were released with 4 mL of 0.5 M *N*-acetyl-glucosamine in 20 mM Tris buffer, pH 7.4 containing 0.15 M NaCl. The three fractions were concentrated with a 10 kD Amicon filter (4 mL, capacity). The collected samples were stored at −70 °C until further testing.

2.5. Human serum depletion

A multiple affinity column (4.6 mm × 100 mm) (Agilent technologies, Palo Alto, CA) was used to remove albumin, IgG, antitrypsin, IgA, transferrin and haptoglobin from human serum. The depletion procedure was performed on a HP 1090 LC system (Hewlett-Packard, Palo Alto, CA). Briefly, human serum was diluted five times with Buffer A of the multiple affinity removal system and injected on the depletion column with an injection volume of 100 μL. Then the unbound serum proteins were eluted with Buffer A at a flow rate at 0.25 mL/min, and the flow through was collected. The column was then regenerated with Buffer B of the multiple affinity removal system before the next injection. The flow through was concentrated with a 10 kD amicon filter (15 mL capacity). A sample with approximately 1.3 mg of depleted serum proteins (compared with approximately 6.5 mg of undepleted human serum protein) was diluted to a volume of 2 mL with the multi-lectin binding buffer, and was further fractionated using the multi-lectin affinity column with the procedure described in Section 2.4.

2.6. SDS-PAGE

The glycoprotein fractions isolated from the fractionation of serum on the single-lectin affinity columns and the corresponding human serum sample were analysed on a NuPAGE 4-12% Bis–Tris gel (1.0 mm × 10 well) (Invitrogen, Carlsbad, CA) with loading amount of 15 μg of protein for each fraction (Fig. 1). The two glycoprotein fractions isolated from the single step elution of the multi-lectin affinity columns, the flow-through fraction from one multi-lectin column and a set of three glycoprotein fractions from the use of different displacers on the multi-lectin affinity column were also analyzed on the NuPAGE system (15 μg in each case, Fig. 2). The proteins were resolved with the NuPAGE MOPS SDS running buffer in a Novex Mini-Cell system (Invitrogen) at 200 volts (Powerpac power supply, Bio-Rad, Hercules, CA). The proteins were visualized by staining with a glycoprotein detection kit (Sigma, St. Louis, MO) containing Schiff's reagent, which is a specific stain for glycoproteins. The staining was performed using the protocol suggested by the manufacturer. To ensure the specificity of the staining, horseradish peroxidase, a glycoprotein, was analyzed on the same gel as positive control and a series of non-glycosylated molecular mass standards were used as negative control for the staining procedure.

2.7. Tryptic digestion

The glycoprotein fractions (100 μg) from single- and multi-lectin affinity columns and the flow-through fraction

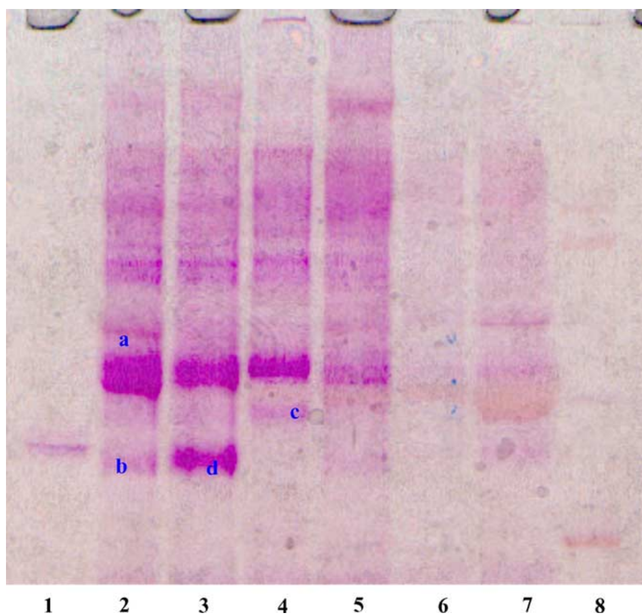

Fig. 1. SDS-PAGE of single-lectin binding proteins in human serum stained with Schiff's reagent. (1) Horseradish peroxidase used as positive control. (2) Con A bound proteins. (3) WGA bound proteins. (4) Jacalin lectin bound proteins. (5) LCA bound proteins. (6) PNA bound proteins. (7) Human serum. (8) Molecular mass standards used as negative control. The positive control developed color, while the negative control did not stain indicating the specificity of Schiff's reagent to glycoproteins.

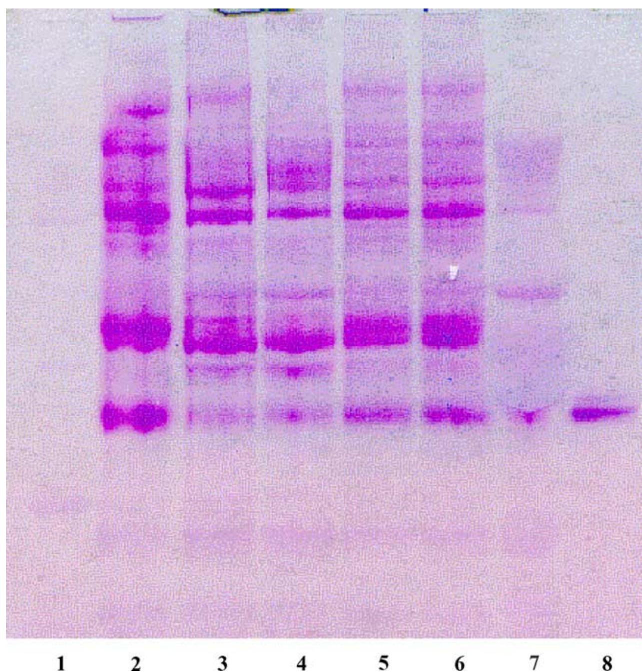

Fig. 2. SDS-PAGE of proteins captured from human serum by the multi-lectin column, stained with Schiff's reagent. (1) A series of molecular weight standards as a negative control. (2) WGA selected proteins captured by the multi-lectin column. (3) Con A selected proteins captured by the multi-lectin column. (4) Jacalin selected proteins captured by the multi-lectin column. (5) The elution sample from the replicate experiment, which had similar profiles to lane 6. (6) First elution sample. (7) Flow through. (8) Horseradish peroxidase as positive control.

from a multi-lectin column were digested with trypsin, using a procedure described previously [16]. Proteins were denatured with 6 M guanidine chloride in 0.1 M ammonium bicarbonate buffer, pH 8 and reduced by incubating with 5 mM DTT at 75 °C for 1 h, then alkylated for 2 h with 0.02 M iodoacetamide. The samples were solvent exchanged with a 10 kD Amicon filter (0.5 ml capacity, Millipore, Billerica, MA). The samples were then adjusted with 0.1 M ammonium bicarbonate buffer to a protein concentration of 0.5 mg/mL. Then 1 μg trypsin was added to each sample and incubated at ambient temperature overnight. For complete digestion, another aliquot of 1 μg trypsin was added, and the digestion was continued for a total of 24 h.

2.8. LC/MS/MS

The trypsin digested peptides were separated on a C18 capillary column (Biobasic C-18, Thermo, 180 μm × 10 cm) using a ProteomeX system (ThermoFinnigan, San Jose, CA). The flow rate was maintained at 2 μL/min. The gradient was started at 5% acetonitrile (ACN) with 0.1% formic acid and a linear gradient to 40% ACN was achieved in 120 min, then ramped to 80% ACN in 5 min and kept at 80% ACN for 20 min to wash the column. Then 15 μL of each sample containing 3.5 μg of protein was injected on the column from a Surveyor autosampler (ThermoFinnigan, San Jose, CA) using the no-waste injection mode. The resolved peptides were analyzed on an LCQ DECA XP ion trap mass spectrometer (ThermoFinnigan, San Jose, CA) with an ESI ion source. The temperature of the ion transfer tube was controlled at 185 °C and the spray voltage was 3.3 kV. The normalized collision energy was set at 35% for MS/MS. Data dependent ion selection was monitored to select the most abundant four ions from a MS scan for MS/MS analysis. Dynamic exclusion was continued for duration of 5 min.

2.9. Bioinformatics

Peptide sequences were identified using SEQUEST algorithm (Version C1) incorporated in BioWorks software (Version 3.1) (ThermoFinnigan). Only peptides resulting from tryptic cleavages were searched. The SEQUEST results were filtered by Xcorr versus charge state. Xcorr was used for a match with 1.5 for singly charged ions, 2.0 for doubly charged ions, and 2.5 for triply charged ions. The protein identification was made based on the corresponding peptide identification. In this research, proteins with two or more peptide identifications were considered as positive identifications.

3. Results

3.1. Proteins captured by a single-lectin affinity column

To overcome the inability of a single lectin to completely capture a glycoprotein sample, five commonly used lectins,

Con A, WGA, Jacalin, LCA and PNA, were examined separately for the capture of glycoproteins from human serum. Each lectin immobilized on agarose was separately packed in columns in which flow was driven by gravity. An equivalent sample of human serum was loaded on each of these single-lectin affinity columns, and the captured proteins were eluted with the elution buffer (Table 1) specific for the corresponding lectin. After being concentrated, the captured protein samples were analyzed on SDS-PAGE and visualized using a Schiff's reagent which specifically stains glycoproteins (lanes 2–6 in Fig. 1). On the same gel, a positive control (lane 1 in Fig. 1), horseradish peroxidase (glycoprotein), and a negative control (lane 8 in Fig. 1) containing non-glycosylated protein molecular weight standards, were stained with the Schiff's reagent. In this case the glycoprotein standard developed color while the negative control did not stain, which confirmed the specificity of the Schiff's reagent for glycoproteins.

The proteins captured by the different lectins showed different SDS-PAGE profiles, although they were captured from the same human serum sample (Fig. 1). Some gel bands were only shown in one sample, such as band "a" in the lane of Con A bound proteins and band "c" in the lane of WGA bound proteins. Meanwhile, some bands from different samples had different staining intensity although they migrated with the same approximate molecular weight, such as band "b" in the lane of Con A bound proteins and band "d" in the lane of Jacalin bound proteins. Based on the staining intensity, peanut agglutinin did not capture as much glycoprotein from serum as the other four lectins. In addition, the corresponding gels that were visualized with a Coomaissie stain (data not shown) suggested that there was more non-specific binding with both the PNA and LCA columns than other three lectins (Con A, WGA, and Jacalin). An unfractionated serum sample was also analyzed on the gel (lane 7 in Fig. 1) and showed that the serum proteins were not as strongly stained by the Schiff's reagent relative to the lectin fractions. This is due to the presence of large amounts of non-glycosylated proteins such as albumin in the serum sample, therefore the glycosylated proteins were only a small fraction of the loaded sample. These SDS-PAGE results indicated that each lectin can enrich specific sets of glycoproteins with some of overlap glycoprotein specificities. Also, as expected, none of the individual lectins could achieve a complete capture of glycoproteins from human serum, and this lead to the concept of a multi-lectin affinity enrichment.

3.2. *Isolation of glycoproteins using the multi-lectin affinity column*

A multi-lectin affinity column was prepared from a physical mixture of immobilized Con A, WGA and Jacalin lectins and the column was used to enrich glycoproteins from human serum. These lectins were selected because their affinities cover most of the common sugar residues present in O- and N-linked glycans contained in serum proteins. We did not include PNA and LCA in the multi-lectin column because their specificities are similar to Jacalin and Con A, respectively.

Human serum was loaded on the multi-lectin column. After washing out any unbound proteins (termed the 'flow-through' fraction), the captured glycoproteins (termed 'bound') were eluted with a specific displacer (see Section 2.4). The collected fractions were then digested with trypsin and analyzed with capillary reversed phase LC with MS/MS detection. The proteins were identified from the LC/MS/MS data by using the SEQUEST algorithm. In addition, the SEQUEST rank of each identified protein was used as an indication of the relative quantities of a given set of proteins. The specificity of this multi-lectin column was examined by an analysis of the glycosylation patterns of proteins identified in the bound (captured proteins) fraction (Table 2). Of the 51 proteins identified (with two or more peptide identifications) from the captured protein fraction, 50 were glycoproteins (including subunits of glycoproteins, such as light chains of immunoglobulins), according to the Swissprot database (http://us.expasy.org/sprot/). Albumin, the exception, is not glycosylated but was found at low levels in the bound fraction (307 versus 22 hits in the flow through versus bound fraction) indicating that albumin is largely removed by the multi-lectin column. The presence of small amounts of albumin could either be due to a low level of non-specific binding or due to the formation of complexes with glycoproteins, such as IgA and IgG [17].

The efficiency of this multi-lectin affinity column was further shown by the absence of a majority of the captured glycoproteins (41/51) in the flow-through fraction. In the case of three abundant glycoproteins (serotransferrin, alpha-1-antitrypsin, Ig alpha-1-chain c region) there was a significant enrichment in the bound fraction relative to the flow-through fraction based on the number of peptide identifications for each protein (Table 2). The incomplete removal of these proteins could be due to the presence of non-glycosylated isoforms in these proteins and will be the subject of future investigations. In addition, the 1D-gel of the bound (lane 6 in Fig. 2) and the flow-through (lane 7 in Fig. 2) fractions showed that the bound fraction had a more intense and discrete banding pattern of glycoproteins relative to the flow-through fraction.

To demonstrate reproducibility of the multi-lectin affinity chromatography procedure for glycoprotein enrichment, the procedure was repeated (including packing of another multi-lectin column). The bound sample (glycoproteins captured by the second column) was analyzed by SDS-PAGE (lane 5 in Fig. 2). The replicate showed a similar separation profile to that obtained with the first multi-lectin column (lane 6 in Fig. 2). The LC/MS/MS analysis identified 50 proteins in the captured fraction and included 49 glycoproteins. Of these 50 proteins, 47 were in common with the glycoprotein fraction in the first experiment (46 out of these 47 proteins were glycoproteins, see Table 3) with albumin again as the only outlier.

The rank of identified proteins as determined by the SEQUEST algorithm can represent an approximate relative

Table 2
Proteins captured from human serum by the multi-lectin affinity column

Rank[a]	ID	Reference	Hits (EL)[b]	Hits (FL)[c]	Glycoprotein[d]
1	A2MG	Alpha-2-macroglobulin	86	0	Yes
2	TRFE	Serotransferrin	61	14	Yes
3	HPT2	Haptoglobin-2	60	0	Yes
4	HPTR	Haptoglobin-related protein	60	0	Yes
5	A1AT	Alpha-1-antitrypsin	53	7	Yes
6	HEMO	Hemopexin (beta-1B-glycoprotein)	51	0	Yes
7	ALC1	IG alpha-1 chain C region	42	1	Yes
8	CO3	complement C3	34	0	Yes
9	KAC	IG kappa chain C region	34	14	Yes
10	APA1	Apolipoprotein A-I	24	17	Yes
11	ALBU	Serum albumin	22	307	No
12	AACT	Alpha-1-antichymotrypsin	19	0	Yes
13	A2HS	Alpha-2-HS-glycoprotein	18	0	Yes
14	MUC	IG MU chain C region	16	0	Yes
15	ITH1	Inter-alpha-trypsin inhibitor heavy chain	14	0	Yes
16	PZP	Pregnancy zone protein	14	0	Yes
17	CERU	Ceruloplasmin	13	0	Yes
18	CFAH	Complement factor H	13	0	Yes
19	CO4	Complement C4	12	0	Yes
20	HRG	Histidine-rich glycoprotein	12	0	Yes
21	GC1	IG gamma-1 chain C region	12	19	Yes
22	GC4	IG gamma-4 chain C region	11	12	Yes
23	A1AG	Alpha-1-acid glycoprotein 1	11	0	Yes
24	ITH2	Inter-alpha-trypsin inhibitor heavy chain H2	11	0	Yes
25	VTNC	Vitronectin	9	0	Yes
26	A1AH	Alpha-1-acid glycoprotein 2	8	0	Yes
27	APA2	Apolipoprotein A-II	7	6	Yes
28	LAC	IG lambda chain C regions	7	11	Yes
29	CLUS	Clusterin	7	0	Yes
30	A1BG	Alpha-1B-glycoprotein	6	0	Yes
31	CFAB	Complement factor B	6	0	Yes
32	APOH	Beta-2-glycoprotein-I	6	0	Yes
33	PLMN	Plasminogen	6	0	Yes
34	APD	Apolipoprotein D	6	0	Yes
35	HBB	Hemoglobin beta chain	5	0	Yes
36	KNG	Kininogen	5	0	Yes
37	ANT3	Antithrombin-III	5	0	Yes
38	ITH4	Inter-alpha-trypsin inhibitor heavy chain H4	4	0	Yes
39	IC1	Plasma protease C1 inhibitor	4	0	Yes
40	AMBP	AMBP protein	3	0	Yes
41	C4BP	C4B-binding protein alpha chain	3	0	Yes
42	SAMP	Serum amyloid P-component	3	0	Yes
43	HBA	Hemoglobin alpha chain	3	0	Yes
44	C1QA	Complement C1Q subcomponent	3	0	Yes
45	KV3G	IG kappa chain V-III region	2	0	Yes
46	HV3P	IG heavy chain V-III region	2	0	Yes
47	HV1F	IG heavy chain V-I region	2	0	Yes
48	THRB	Prothrombin	2	0	Yes
49	LV3B	IG lambda chain V-III region	2	0	Yes
50	APC3	Apolipoprotein C-III	1	0	Yes
51	APB	Apolipoprotein B-100	1	0	Yes

[a] The rank is related to the probability of the MS assignment.
[b] The number of peptides identified for a given protein captured by multi-lectin column.
[c] The number of peptides identified in the protein in the flow-through fraction from multi-lectin column; 0 stands for not detected.
[d] Whether the protein is or is not a glycoprotein or a subunit of a glycoprotein is derived from the Swissprot database.

abundance of the protein in the sample (see Table 4 for a comparison of relative abundances of the proteins). Among the proteins identified in both elution samples, 38 out of 50 proteins (75%) were ranked in a consistent manner (difference of <5) and 46 proteins (90%) were found with a rank difference less or equal to 10. Finally all of the 50 proteins were identified with rank difference less than 20. Here, we used conservative protein identifications, namely two or more tryptic peptide hits with a conservative MS filter (see Section 2). Our experience has suggested that the reproducibility of

Table 3
The number of proteins identified in two elution samples captured from replicate multi-lectin affinity columns[a]

Elution sample[b]	Protein ID (>2 hits)[c]	Glycoproteins[d]
1	51	50
2	50	49
Proteins in common[e]	47	46

[a] The proteins captured by the multi-lectin column were trypsin digested and analyzed by LC/MS/MS. The proteins were identified by using SEQUEST database search.

[b] The glycoprotein isolation experiment was repeated including packing a replicate multi-lectin column. Therefore two elution samples were collected.

[c] Number of proteins identified with two or more than two peptide identification.

[d] Number of glycoproteins in the identified proteins with two or more peptide identification.

[e] Number of proteins found in both elution samples.

the ranking is improved by the use of these conservative filters.

Meanwhile, the amount of proteins loaded on the multi-lectin column and the amount of proteins recovered (flow through and bound) were measured using the Bradford assay. A study of the recovery achieved with the multi-lectin affinity column showed that about 0.6 mg of glycoproteins were enriched in the bound fraction from a total of approximately 6.7 mg of serum protein and about 5.0 mg of protein was collected in the flow-through fraction. Therefore, the recovery of this multi-lectin enrichment procedure was about 84%. The sample loss was primarily due to the 10 kD MWCO filtration step used to concentrate the fractions (see Section 2). From these data we concluded that at least 10% of human serum proteins were glycosylated, considering potential sample losses during the process, while most of the large amount of non-glycosylated material was due to albumin.

3.3. Isolating glycoproteins from depleted serum

In order to improve dynamic range of glycoprotein identifications on a multi-lectin column, human serum was first depleted with a multiple affinity removal column that specifically removes albumin, IgG, IgA, antitrypsin, transferrin, and haptoglobin. This depletion step indeed removed approximately 80% of serum proteins and yielded a sample with about 1.3 mg of depleted serum proteins (from 6.5 mg of serum protein). Thus sample was then loaded onto the multi-lectin column and a 0.6 mg glycoprotein fraction was captured (50% of the total protein present in the depleted serum sample). In the LC/MS/MS results, 42 proteins were identified in this captured fraction and the six high abundance proteins that were expected to be removed were indeed not observed. The 42 proteins observed in the depleted sample were also observed in the analysis of the original serum sample. Despite the lack of increase in identifications on depletion of the six most abundant proteins, the glycoproteins were observed with greater sequence coverage in the latter experiment. The lack of increase in dynamic range may be due to the fact that the lectin column itself represents a significant depletion step. Another factor is that there are many more abundant proteins than the few depleted in this study, for example alpha-2-macroglobulin was present in large amounts in the purified glycoprotein sample.

Table 4
The rank differences for same proteins identified in two elution samples captured by replicate multi-lectin affinity columns[a]

Rank difference[b]	≤5[c]	≤10[d]	≤0[e]
Number of proteins[f]	38	46	50
Total proteins (%)[g]	75	90	100

[a] The rank of the protein can represent the approximate relative abundance of the protein in the sample.

[b] The difference in rank of the same protein in the two elution samples.

[c] The rank difference was equal or less than 5.

[d] The rank difference was equal or less than 10.

[e] The rank difference was equal or less than 20.

[f] The number of proteins identified with two or more peptide identifications.

[g] The percentage of the number of proteins in each rank relative to the total number of proteins identified with two or more peptide identifications.

3.4. Fractionation of proteins captured by the multi-lectin column with a sequential use of the three displacers specific for each lectin

The proteins captured by the multi-lectin column from non-depleted human serum were fractionated by sequential use of displacers specific for each lectin. The sequence of specific displacers was as follows: Jacalin, Con A, and finally WGA lectin. The three fractions containing the eluted glycoproteins were analyzed by SDS-PAGE (lanes 2–4 in Fig. 2) which showed that each of the three fractions had a significantly different 1D-gel profile (lane 2–4 in Fig. 1). These three fractions were also analyzed by LC/MS/MS and Table 5 gives the distribution of the proteins in the three displacement fractions with approximate abundance (as measured by LC/MS) of each protein (+ to ++++) observed in a given fraction. These results show that many glycoproteins were concentrated in a specific displacer fraction and this distribution could be correlated with known glycosylation structures (listed in the Swissprot database). Some proteins were eluted in all displacer fractions and by inference were selected by all the three lectins, such as IgA and haptoglobin. IgA has five O-linked glycosylation sites and two N-linked glycosylation sites. With the extensive O-glycosylation and high abundance in human serum, it is not surprising that IgA was one of the most abundant proteins in the Jacalin displacement fraction (ranked at position 1). Haptoglobin was highly enriched in WGA displacement fraction, which may be related to a high abundance of GlcNAc or sialic acid residues in the oligosaccharides of this protein. Some proteins were found to be highly abundant in only one or two displacement fractions. Alpha-1-antitrypsin, which is normally N-glycosylated at three asparagine residues (46, 83, 247), was, however, enriched in the Jacalin fraction as well as the Con A fraction, which may suggest the existence of O-linked glycosylation or

Table 5
Significant proteins eluted by each displacer specific for a given lectin in the multi-lectin column

ID	Reference	Abundance[a] (Jacalin)[b]	Abundance (ConA)[c]	Abundance (WGA)[d]
A1AT	Alpha-1-antitrypsin	+	+	
A1BG	Alpha-1B-glycoprotein		+++	
A2HS	Alpha-2-HS-glycoprotein	++		+++
A2MG	Alpha-2-macroglobulin		+	+
AACT	Alpha-1-antichymotrypsin		++++	
ALBU	Serum albumin	+	+	++++
ALC1	IG alpha-1 chain C region	+	++	+
APA1	Apolipoprotein A-I		+++	
APC3	Apolipoprotein C-III	++		
APE	Apolipoprotein E	+++		
CBP8	Carboxypeptidase N 83 KDA chain			++
CERU	Ceruloplasmin		++++	
CFAB	Complement factor B		++++	
CFAH	Complement factor H		++	
CO3	Complement C3		+	
GC1	IG gamma-1 chain C region		+++	
GC3	IG gamma-3 chain C region	++++		
HEMO	Hemopexin		+++	+
HPT2	Haptoglobin-2	+++	++	+
HPTR	Haptoglobin-related protein	++	++	+
IC1	Plasma protease C1 inhibitor			++++
ITH1	Inter-alpha-trypsin inhibitor heavy chain H1	++		+++
ITH2	Inter-alpha-trypsin inhibitor heavy chain H2	++++		
MUC	IG MU chain C region		++++	++
PZP	Pregnancy zone protein		++++	++
TRFE	Serotransferrin	++++	+	
VTNC	Vitronectin			++

[a] The relative abundance was defined by grouping the SEQUEST ranking (five rankings per group). Proteins ranked 1–5 have an abundance of +, ranked 6–10 have an abundance of ++, ranked 11–15 have an abundance +++, and ranked 16–20 have an abundance of ++++. The empty boxes denote protein rankings that were not in the top 20.

[b] The abundance score of proteins identified in the Jacalin displacement fraction.

[c] The abundance score of proteins identified in the Con A displacement fraction.

[d] The abundance score of proteins identified in the WGA displacement fraction.

galactosyl (β-1,3) GalNAc residues in its carbohydrate moieties. In contrast, alpha-2-macroglobulin, which has eight N-linked glycosylation sites, was found to be highly abundant in both the Con A and WGA displacement fractions. The absence of this protein in the Jacalin fraction might suggest the lack of O-linked glycosylation on this protein, particularly as it is a relatively high abundance protein in human plasma (2000 mg/L). Pregnancy zone protein was highly enriched in the WGA displacement fraction although it is a relatively low abundance plasma protein (8 mg/L [18]) and the result demonstrated the ability of the multi-lectin affinity column to enrich specific glycoproteins. Apolipoprotein E has only an O-linked glycosylation site (Thr 212) and was only found in the Jacalin fraction which further indicated that the expected lectin specificity was indeed observed in the fractionation process. These results indicated that the distribution of a glycoprotein into each of the three fractions is determined to a large extent by the glycosylation pattern of the glycoprotein. In addition, albumin was found in all three displacer fractions, but with the abundance sequentially decreased, which may suggest that the capture of albumin by the multi-lectin column was indeed due to non-specific binding.

The reproducibility of the fractionation procedure was further established by observation of a similar distribution of the glycoproteins in the three displacement fractions with other plasma and serum samples, and these results will be the subject of a future report.

4. Discussion

4.1. Enrichment of glycoproteins in human serum

With the goal of investigating the serum glycoproteome, we used a multi-lectin affinity column containing Con A, WGA and Jacalin lectin to capture glycoproteins from human serum, and the results showed that the enrichment procedure was specific, efficient, and reproducible. By using this multi-lectin affinity column, 10% of human serum proteins were found to be glycosylated (w/w basis). We also depleted serum of the six most abundant proteins, which together count for 80% of serum proteins (albumin, IgG, IgA, antitrypsin, transferrin, and haptoglobin), with a commercially available affinity column. The multi-lectin column was then used to capture the glycoprotein fraction of this depleted sample and

analysis of this fraction found that 50% of the remaining serum proteins were glycosylated (w/w). Since it is likely that the combined lectin column does not capture 100% of the glycoprotein fraction, the actual percentage is probably greater than 50%. We have also shown that this column gave a high degree of capture of the glycoprotein fraction, as determined by the lack of known glycoproteins in the unbound fraction and the absence of reaction of this fraction with specific carbohydrate stains.

The presence of high abundance proteins, such as albumin, can affect the identification of low abundance proteins, and it has been shown that depletion of high abundance proteins can improve the dynamic range of protein identification [19]. In this study, the multi-lectin column was shown to successfully remove a large portion of the serum albumin fraction (estimated at greater than 80% of the total) and resulted in more reproducible protein identifications (more peptides characterized per protein and increased consistency between serum proteomic analyses, data not shown). Therefore, the use of lectin based affinity columns can be used to extend the dynamic range of plasma/serum proteomic studies as well as be used for the specific capture of glycoproteins.

4.2. The reproducibility of the affinity enrichment procedure

The reproducibility of the multi-lectin enrichment procedure was shown by the results of replicates in which a similar ranked list of glycoproteins was achieved (see Tables 3 and 4). Such reproducibility of a method is necessary for a comparative proteome study, such as the search for disease specific biomarkers, where one is looking for differential regulation of a biomarker between normal and disease samples [20–22]. Isotope labeling has been used to explore possible changes in protein regulation, but is not useful in most cases to characterize differences in PTMs such as glycosylation. With the development of the multi-lectin column, however, one can obtain a preliminary estimate of differential expression at the glycoprotein level by comparing the LC/MS data (see above) and then use different displacers on the multi-lectin column to give information about glycosylation changes (see below).

4.3. Fractionation of the proteins captured by a multi-lectin column by sequential use of three displacers specific for each lectin

In this study we optimized the multi-lectin column for the analysis of serum glycoproteins by the selection of three different lectins containing complementary specificity for carbohydrate motifs known to be present in serum. These three lectins need different displacers to elute the affinity captured glycoproteins, which allows the sequential use of these displacers to elute different fractions. An advantage of this method is that it is not limited to a single elution step (a common shortcoming of affinity columns). The difference in 1D-gel results of the three lectin displacement fractions (lanes 2–4 in Fig. 2) from that of the proteins captured by the corresponding single-lectin column (lanes 2–4 in Fig. 1) can be related to the affinity of glycoproteins to more than one of the lectins immobilized in the multi-lectin column. For example, IgA, alpha-1-antitrypsin, and alpha-2-macroglobulin are observed in more than one displacer fraction (Table 5) and can be related to the presence of more than one type of sugar residue in a typical glycan structure. Such an overlap in specificity among the three lectins can result in a competition between the immobilized lectins to capture a given glycoprotein, provided the column capacity is significantly greater than the sample loading. One could expect that the distribution of a given glycoprotein in the three lectin displacement fractions would be determined by the glycosylation pattern (composition and structure of glycan) of that glycoprotein. If the glycosylation pattern of a given protein is changed due to biological changes, as some glycoprotein biomarkers in cancer or other diseases [23,24], the affinity of this protein for each lectin in the multi-lectin column may change, and the ratio captured by each lectin will be different. Thus one could expect that the relative abundance of the protein in each displacement fraction will change, and can be measured by LC/MS/MS analysis of the tryptic digest. Therefore, with the use of different displacer elution steps, the use of multi-lectin affinity column, with overlaps in lectin specificity for specific glycosylation motifs, will be useful to monitor changes in the pattern of glycosylation in a biological sample.

The fractionation of glycoproteins in a given sample can be further optimized by consideration of known carbohydrate structures and the appropriate selection of lectins for immobilization. Another step in the optimization is the sequential use of displacers specific for each lectin to generate fractions suitable for the analysis of changes in protein glycosylation patterns in a comparative proteomic study. Such an analysis is particularly important in the case where the regulation of the protein amount is unchanged, a so-called 'silent change'. Furthermore, such a differential analysis is not readily achieved by use of a single ligand affinity column with a single elution step, unless the change in glycosylation pattern results in a complete loss or gain in affinity for a particular ligand. In this study we used a semiquantitive method, namely SEQUEST ranking, but in the future studies we will explore other methods of differential quantification, such as isotope labeling the proteins with O^{16}/O^{18} exchange before or after the sequential displacement steps or direct quantitation by FTMS [25].

5. Conclusions

In this study, a multi-lectin affinity column, which combines the affinity of Con A, WGA, and Jacalin, has been shown to capture the majority of glycoproteins present in human serum. Our results demonstrated that the multi-lectin column gives a more complete capture of the glycoproteins than a single-lectin column. By using this specific, efficient, and reproducible enrichment procedure, 10% of the total hu-

man serum proteins were found to be glycosylated (w/w) and after depletion of six highly abundant proteins, 50% of the remaining proteins were found to be glycosylated (w/w). The use of a multi-lectin column may allow for the detection and perhaps quantitation of "silent" changes in which a disease is associated not with a change in the amount of a given biomarker but rather in the structure of a given PTM, such as glycosylation.

Acknowledgements

The authors thank ThermoFinnigan for instrumental support and Agilent for providing the multiple affinity depletion column. The authors would like to thank Drs. Barry Karger, Billy Wu, Jo Tsai, and Darryl Palmer-Toy, and Mr. Haven Baker for helpful comments either during the experiments or during manuscript preparation.

References

[1] C.R. Bertozzi, L.L. Kiessling, Science 291 (2001) 2357.
[2] P.M. Rudd, T. Elliott, P. Cresswell, I.A. Wilson, R.A. Dwek, Science 291 (2001) 2370.
[3] N.E. Zachara, G.W. Hart, Chem. Rev. 102 (2002) 431.
[4] C.F. Goochee, T. Monica, Biotechnology (NY) 8 (1990) 421.
[5] J. Alper, Science 301 (2003) 159.
[6] C. Seelenmeyer, S. Wegehingel, J. Lechner, W. Nickel, J. Cell. Sci. 116 (2003) 1305.
[7] N.L. Anderson, N.G. Anderson, Electrophoresis 19 (1998) 1853.
[8] C.L. Nilsson, Anal. Chem. A-Page (2003) 348A.
[9] L. Xiong, D. Andrews, F. Regnier, J. Proteome Res. 2 (2003) 618.
[10] J.W. Becker, G.N.J. Reeke, J.L. Wang, B.A. Cunningham, G.M. Edelman, J. Biol. Chem. 250 (1975) 1513.
[11] N. Bakry, Y. Kamata, L.L. Simpson, J. Pharmacol. Exp. Ther. 258 (1991) 830.
[12] F.T. Saulsbury, J. Rheumatol. 24 (1997) 2246.
[13] F.P. Schwarz, K.D. Puri, R.G. Bhat, A. Surolia, J. Biol. Chem. 268 (1993) 7668.
[14] S.A. Brooks, M.V. Dwek, U. Schumacher, Functional & Molecular Glycobiology, BIOS Scientific Publishers, Oxford, 2002.
[15] P. Pahlsson, D.P. Blackall, M. Ugorski, M. Czerwinski, S.L. Spitalnik, Glycoconj J. 11 (1994) 43.
[16] S.L. Wu, H. Amato, R. Biringer, G. Choudhary, P. Shieh, W.S. Hancock, J. Proteome Res. 1 (2002) 459.
[17] J.S. Baumstark, Prep. Biochem. 13 (1983) 15.
[18] C.M. Petersen, P.H. Jensen, A. Bukh, L. Sunde, L.U. Lamm, J. Ingerslev, Scand. J. Clin. Lab. Invest. 50 (1990) 479.
[19] R. Pieper, Q. Su, C.L. Gatlin, S.T. Huang, N.L. Anderson, S. Steiner, Proteomics 3 (2003) 422.
[20] L.F. Steel, D. Shumpert, M. Trotter, S.H. Seeholzer, A.A. Evans, W.T. London, R. Dwek, T.M. Block, Proteomics 3 (2003) 601.
[21] L.V. Leak, E.F. Petricoin 3rd, M. Jones, C.P. Paweletz, A.M. Ardekani, V.A. Fusaro, S. Ross, L.A. Liotta, Ann. N.Y. Acad. Sci. 979 (2002) 211.
[22] P.R. Srinivas, S. Srivastava, S. Hanash, G.L. Wright Jr., Clin. Chem. 47 (2001) 1901.
[23] M. Butler, D. Quelhas, A.J. Critchley, H. Carchon, H.F. Hebestreit, R.G. Hibbert, L. Vilarinho, E. Teles, G. Matthijs, E. Schollen, P. Argibay, D.J. Harvey, R.A. Dwek, J. Jaeken, P.M. Rudd, Glycobiology 13 (2003) 601.
[24] K. Goswami, D.N. Nandakumar, B.C. Koner, Z. Bobby, S.K. Sen, Clin. Chim. Acta 337 (2003) 163.
[25] H.R. Padley, S. Bashir, T.D. Wood, Anal. Chem. 69 (1997) 2914.