Published in final edited form as:

Anal Biochem. 2011 January 1; 408(1): 71–85. doi:10.1016/j.ab.2010.08.010.

A lectin affinity workflow targeting glycosite-specific, cancer-related carbohydrate structures in trypsin-digested human plasma¹

Penelope M. Drake^a, Birgit Schilling^b, Richard K. Niles^a, Miles Braten^a, Eric Johansen^a, Haichuan Liu^a, Michael Lerch^a, Dylan J. Sorensen^b, Bensheng Li^b, Simon Allen^a, Steven C. Hall^a, H. Ewa Witkowska^a, Fred E. Regnier^c, Bradford W. Gibson^{b,d,*}, and Susan J. Fisher^{a,*}

^aDepartment of Obstetrics, Gynecology and Reproductive Sciences, 513 Parnassus Ave., Box 0556, University of California San Francisco, San Francisco, CA 94143

^bBuck Institute for Age Research, 8001 Redwood Blvd., Novato, CA 94945

^c201 S. University St. HANS B054, Purdue University, West Lafayette, IN 47907

^dDepartment of Pharmaceutical Chemistry, Box 0446, University of California, San Francisco, CA 94143

Abstract

Glycans are cell-type specific, post-translational protein modifications that are modulated during developmental and disease processes. As such, glycoproteins are attractive biomarker candidates. Here, we describe a mass spectrometry-based workflow that incorporates lectin affinity chromatography to enrich for proteins that carry specific glycan structures. As increases in sialylation and fucosylation are prominent among cancer-associated modifications, we focused on *Sambucus nigra* agglutinin (SNA) and *Aleuria aurantia* lectin (AAL), lectins which bind sialic acid- and fucose-containing structures, respectively. Fucosylated and sialylated glycopeptides from human lactoferrin served as positive controls, and high mannose structures from yeast invertase served as negative controls. The standards were spiked into Multiple Affinity Removal System (MARS) 14-depleted, trypsin-digested human plasma from healthy donors. Samples were loaded onto lectin columns, separated by HPLC into flow-through and bound fractions, and treated with peptide: *N*-glycosidase F to remove *N*-linked glycans. The deglycosylated peptide fractions were interrogated by ESI HPLC-MS/MS. We identified a total of 122 human plasma glycoproteins containing 247 unique glycosites. Importantly, several of the observed glycoproteins (*e.g.*, cadherin 5 and neutrophil gelatinase-associated lipocalin) typically circulate in plasma at low ng/mL levels. Together, these results provide mass spectrometry-based evidence of the utility of incorporating lectin-separation platforms into cancer biomarker discovery pipelines.

¹This work was supported by an NCR shared instrumentation grant S10 RR024615 (to B.W. Gibson) and by grants from the National Cancer Institute, U24 CA126477 (S.J. Fisher, UCSF) and a U24 Subcontract (B.W. Gibson, Buck Institute for Age Research), which are part of the NCI Clinical Proteomic Technologies for Cancer initiative (<http://proteomics.cancer.gov>).

© 2010 Elsevier Inc. All rights reserved.

* To whom correspondence should be address: Susan Fisher, phone: (415) 476-5297, fax: 415-476-5623, sfisher@cgl.ucsf.edu; Bradford Gibson, phone: (415) 209-2032, fax: 415 209-2231, bgibson@buckinstitute.org.

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.

Keywords

Lectin chromatography; glycopeptide; plasma; cancer; biomarker; mass spectrometry

Introduction

Advances in mass spectrometry (MS)-based proteomics, including nano liquid chromatography electrospray ionization (ESI) interfaces, faster and more sensitive mass analyzers, and robust bioinformatics approaches, have brought the unbiased discovery of disease biomarkers within reach. These technological improvements have led to a new era of research aimed at improving prognoses, diagnoses, and monitoring responses to therapy through detection of biomarkers in human body fluids (*i.e.*, plasma, saliva and urine). However, making this an effective strategy requires panels of verified disease-specific reporter molecules that, as yet, do not exist. Therefore, there is a tremendous interest in discovery efforts.

Overall, investigators who focus on the early stages of discovery pipelines use two approaches. The first is brute force protein identification to determine differences between samples obtained from patient subjects and control individuals. The inherent complexity of body fluids requires extensive sample separation, usually achieved by a series of orthogonal/complimentary chromatographic steps. Successful studies typically require large amounts of starting material, time, and expertise [1]. Second, other investigators have used targeted approaches to reduce complexity. One way to direct these experiments is by considering the biology of the disease of interest, in this case, cancer. In this regard, post-translational modifications (PTMs) are especially interesting because they are linked to the disease process and in some cases play a causal role [2]. Since these modifications can appear at multiple positions on a protein scaffold and on multiple protein backbones, the expression of these motifs is usually greatly amplified as compared to that of single proteins. Thus, through the use of affinity capture reagents, specific PTMs may be exploited as targets to enrich molecules that are a signature of a particular disease state.

Aberrant carbohydrate modifications have been recognized as a hallmark of cancer for over 30 years [3]. Intriguingly, many of the oldest and most widely used clinical diagnostic tests detect glycoproteins. These include carcinoembryonic antigen (CEA), commonly used as a marker of colorectal cancer; CA-125, frequently employed to diagnose ovarian cancer; and prostate-specific antigen (PSA) [4-6]. Interestingly, many of the most informative tests directly assess the expression of a particular class of carbohydrates termed Lewis (Le) blood group antigens, which exhibit unique biological functions [7]. Anti-sialyl Le^a (CA 19-9), -Le^x, -sialyl Le^x, and -Le^y antibodies are used in the evaluation of biopsy specimens from breast, bladder, colorectal, esophageal and non-small cell lung carcinoma [8-15]. In all instances, Le antigen expression is correlated with increased metastasis, advanced stage of disease and reduced survival time. The fact that cancer-related carbohydrate changes are correlated with clinically relevant outcomes such as metastasis and survival enhances their utility as biomarkers. Indeed, studies have already shown that selectively enriching cancer-related protein glycoforms affords the possibility of increasing diagnostic sensitivity and specificity. For example, separation of serum PSA by the *Maackia amurensis* agglutinin lectin, which specifically binds α 2,3-linked sialic acid, allows discrimination ($p < 0.001$) between blood samples from individuals with benign prostatic hypertrophy and prostate cancer patients [16], which standard PSA tests fail to do [17].

The relationship between saccharide expression and disease progression has a biological basis as glycans regulate many processes involved in tumorigenesis. For instance,

extravasation, a critical step in metastasis, is initiated by shear stress-induced interactions between selectin family endogenous lectins and their cell surface carbohydrate ligands. The selectin family consists of three members with differential expression patterns. Endothelial cells display E- and P-selectin, platelets express P-selectin, and leukocytes present L-selectin [18; 19]. Carbohydrate ligands for selectins are modified Le blood group antigens, which are abundant in malignant tissues. Evidence is accumulating that the clinical correlation between tumor expression of selectin ligands and metastasis is a reflection of a causal relationship [20]. Mice lacking L- or P-selectin, singly or in combination, exhibit greater resistance to the metastasis of colon carcinoma cells than do wild-type animals [21; 22], and E-selectin is also implicated in tumor metastasis [23]. Additionally, cancer-related protein scaffolds have been identified as selectin ligand carriers, and thus as participants in the extravasation process. For instance, CEA, the glycoprotein commonly used as a marker for colorectal cancer, and CD44, a highly glycosylated cell surface adhesion protein, can serve as E- and L-selectin ligands [20].

In light of the strong biological and clinical correlations between tumors and specific types of glycosylation, several laboratories have applied a variety of glycan-based enrichment strategies to cancer biomarker discovery. In general, these studies have tended to be small-scale proof-of-principle endeavors. It is noteworthy that, on the whole, they succeeded in identifying target biomarker candidates for further validation. Thus, the published literature supports the theory that focusing the selection process on cancer-relevant glycans increases the likelihood of identifying potential disease sentinels and/or targets. In terms of specific methods, a variety of protocols, including lectin-, antibody- and chemical-based approaches, have been developed for the capture of glycoproteins and glycopeptides. Lectin-based protocols include those using single species [24-26], as well as mixtures, *e.g.*, concanavalin A, wheat germ agglutinin, and *Artocarpus integrifolia* (jacalin). The latter method has been dubbed multi-lectin affinity chromatography [27-30]. Less common are protocols employing carbohydrate-reactive antibodies such as those recognizing Le blood group antigens [31]. In addition, two similar approaches, involving hydrazide or boronic acid chemistry, capitalize on the *cis*-diols in monosaccharides to covalently couple the glycans to a derivatized support [32; 33].

Here, we present the results of an extensive method development effort in which we used glycopeptide standards as positive and negative controls to optimize an automated workflow for the rapid enrichment of selectively glycosylated peptides from depleted, trypsin-digested human plasma. Specifically, we performed lectin affinity chromatography at the glycopeptide level. Using a number of requirements, including the presence of an *N*-linked consensus motif (NXS/T) and a +1 Da mass shift at the modified Asn residue after deglycosylation with PNGase F, we identified glycosites enriched by a specific lectin, an indication that they carried specialized carbohydrate modifications. Using commercially available preparations of *Sambucus nigra* agglutinin (SNA), which binds sialic acid α 2,6- and α 2,3-linked to terminal galactose, and *Aleuria aurantia* lectin (AAL), which binds α -linked fucose at both core and branch positions, we identified a total of 227 glycosites in 119 glycoproteins from human plasma. These included low-abundance proteins such as cadherin-5 and neutrophil gelatinase-associated lipocalin. Now, we are poised to compare families of SNA- and AAL-binding glycoproteins in cancer-relevant human samples.

Experimental Procedures

For details, also see a Standard Operating Procedure (SOP) provided as Supplementary Methods.

Human plasma

Pooled human plasma collected from equal numbers of male and female healthy donors was obtained from Bioreclamation (Westbury, NY).

Multiple Affinity Removal System (MARS)-14 Depletion of abundant plasma proteins

Batches of 200 μ l of human plasma were processed according to the manufacturer's instructions using a MARS-Hu-14 (10 $\times$ 100 mm) column from Agilent Technologies (Santa Clara, CA) on a Paradigm MG4 HPLC system (Michrom Bioresources, Auburn, CA). The protein flow-through fraction was collected and re-adjusted to the original volume using a 3 or 5 kDa molecular weight cut-off (MWCO) centrifugal concentrator (Sartorius AG, Goettingen, Germany). Hereafter, a plasma volume refers to the original plasma volume.

Trypsin digestion

MARS-14 depleted plasma was denatured with 6 M urea, reduced with 20 mM DTT (30 min at 37 °C), alkylated with 50 mM iodoacetamide (30 min at RT), and incubated overnight at 37°C with sequencing grade trypsin (Promega, Madison, WI) added at a 1:50 enzyme:substrate ratio (wt/wt) as previously described [34]. Following digestion, the samples were acidified with formic acid, and desalted using HLB Oasis SPE cartridges (Waters, Milford, MA) that were prepared by washing with 80% acetonitrile in 1% formic acid prior to equilibration with 0.1% formic acid. Samples were eluted with 80% acetonitrile in 0.1% formic acid, neutralized by the addition of 0.5 M ammonium bicarbonate and concentrated by vacuum centrifugation. Peptides were stored at -80 °C until use.

Glycopeptide standards

Human lactoferrin (L0520) and yeast invertase (I0408) were purchased from Sigma-Aldrich (St. Louis, MO). Glycoproteins were individually trypsin-digested as described above. Since invertase was not expected to bind to the lectin columns, no further purifications were carried out. The lactoferrin digest (50 μ g) was chromatographed on either an SNA or an AAL column to enrich the subset of glycopeptides that had glycan structures that interacted with the carbohydrate-binding proteins. The bound material was collected, and desalted as described below before use. Lactoferrin- and invertase-derived glycopeptide standards were characterized by matrix-assisted laser desorption/ionization–time-of-flight (MALDI-TOF) MS and MS/MS, also described below.

Preparation of lectin columns

SNA and AAL were purchased from Vector Laboratories (Burlingame, CA). Lectins (1.5–6 mg) were suspended at 10–20 mg/mL in PBS. 100 mg POROS-AL beads (Applied Biosystems, Foster City, CA) were washed 2 times in 1 mL PBS before addition of the lectin solution. Sodium cyanoborohydride was added to a final concentration of 50 mM and the sample was tumbled overnight at room temperature. The beads were washed once with 1 mL 1M Tris-HCl, pH 7.4. The remaining aldehydes were blocked by incubation for 30 min at room temperature in the same buffer containing 50 mM sodium cyanoborohydride. Unconjugated protein was removed by washing the beads (5 $\times$ 1 mL of 1 M sodium chloride). Then, the lectin-conjugated beads, approximately 300 μ l in volume, were packed into 4.6 $\times$ 50 mm PEEK HPLC columns. Routinely, columns were stored in PBS with 0.02% sodium azide at 4 °C for up to 6 months. Individual columns were used for up to 75 affinity separations without changes in the performance characteristics.

Lectin chromatography

A Michrom Paradigm MG4 HPLC system equipped with a CTC PAL robot configured as an autosampler and fraction collector (Michrom Bioresources, Auburn, CA) was employed.

Initially, we tested sample loads, buffer compositions, and column washing conditions using glycopeptide standards spiked into MARS 14-depleted trypsin-digested plasma. We began by varying the amount of sodium chloride (0-500 mM) in Buffer A (25 mM Tris, pH 7.4, 10 mM calcium chloride, 10 mM magnesium chloride). Buffer B was 0.5 M acetic acid. Next, we optimized the wash step, in terms of timing and flow rate: ~40 column volumes at 2 mL/min vs. 0-0.75 column volumes at 50 μ L/min. Conditions were assessed in terms of the binding characteristics of the glycopeptide standards, recovery of plasma glycopeptides in the bound fraction, and reproducibility (determined by monitoring peptide elution at 280 nm). Lastly, we explored sample loading (10-200 μ L) in terms of bound glycopeptides.

The optimized protocol was as follows. Routinely, ~1 mg (200 μ L depleted plasma) was diluted into Buffer A containing 50 mM sodium chloride, and spiked with 50 pmol of the invertase digest and 25 pmol of the lectin-enriched lactoferrin. Then, the mixture was applied to the lectin column and separated using a 3 step gradient: 1) Collection of the unbound fraction: Buffer A for 9.0 min at 50 μ L/min; 2) Elution: Buffer B for 4.8 min at 500 μ L/min; and 3) Re-equilibration: Buffer A for 6.0 min at 3000 μ L/min. The flow-through (FT) fraction, collected from 2.8 to 6.9 min, and the bound fraction, collected from 9.0 to 11.6 min, were desalted using Oasis HLB cartridges as described above. Eluted samples were neutralized by the addition of 0.5 M ammonium bicarbonate and concentrated to < 100 μ L by vacuum centrifugation.

PNGase F digestion

N-linked glycopeptides in the FT and bound fractions were deglycosylated by treatment with PNGase F (Glycerol-free, New England Biolabs; Ipswich, MA). A small volume was removed to estimate pH. If outside the desired range of 7.0-8.0, 50 mM ammonium bicarbonate was added. Initial experiments assessed the optimal enzyme to substrate ratio (1 U/ μ g-20 U/ μ g) and the effect of enzyme concentration (25 U/ μ L-200 U/ μ L) on deglycosylation efficiency. Final reaction conditions were as follows. Five μ L (2,500 U) of enzyme was added to each sample in a volume of 50-100 μ L prior to incubation overnight at 37 °C. Following deglycosylation, samples were desalted and concentrated using C18 ZipTips® (Millipore; Billerica, MA).

MALDI TOF-TOF mass spectrometric analyses

Mass spectra were acquired by matrix-assisted laser desorption ionization–time-of-flight (MALDI-TOF) MS using a 4800 MALDI TOF-TOF Proteomics Analyzer (Applied Biosystems, Foster City, CA; USA/MDS Sciex, Concord, ON, Canada) equipped with TOF/TOF™ ion optics, a 200 Hz Nd:YAG laser, controlled by the 4000 Series Explorer Software V3.5.28193. The positive ion MS spectra in the m/z range of 800-6000 Da were manually acquired by accumulating 1500-2000 shots at a fixed laser intensity (4000-4200). The accelerating voltage was 20 kV. Other instrumental settings were optimized to afford a resolution of >15,000 at m/z = 2093.89 Da. MS/MS spectra were generated by employing collision-induced dissociation (CID) using the following settings: collision cell floated at 1 kV; resolution of precursor ion selector = 300 or 400 full-width half-maximum (FWHM); metastable suppressor: on; total shots per spectrum = 4000; and fixed laser intensity = 5000-5500. No collision gas was used. Manufacturer-supplied Plate Model and Default MS Calibration algorithms were employed to generate external calibration of the MS mode using monoisotopic molecular ions from a combination of six peptide standards (des-Arg¹-bradykinin, angiotensin I, [Glu¹]-fibrinopeptide B, and three ACTH peptides: 1-17, 18-39 and 7-38). Acceptance criteria thresholds for a generation of MS calibration files required at least four standard molecular ions with S/N of $\geq$ 300, mass tolerance of 0.5 Da, and maximum outlier error of 25 ppm. Default calibration of the MS/MS spectra used a minimum of five matched fragment ions of [Glu¹]-fibrinopeptide B that were detected with

a minimum S/N of 120, within a mass tolerance window of ± 1 Da and a maximum outlier error of 20 ppm. Data were processed using Data Explorer Software Version 4.9 (Applied Biosystems). Samples were desalted and concentrated using C18 ZipTips® (Millipore). Then aliquots of 0.5 μ l (~ 15 nM – 3 μ M) in 0.1% TFA were mixed on a MALDI target with the matrix (α -cyano-4-hydroxycinnamic acid, 5 mg/ml in 80% ACN/0.1% TFA/10 mM dibasic ammonium phosphate) at a 1:1 ratio (v/v). All data were manually processed and interpreted.

ESI-QqTOF mass spectrometric analyses

The proteolytic peptide mixtures obtained after tryptic digestion, lectin enrichment, PNGase F treatment, and desalting were analyzed by reversed-phase nano-HPLC-ESI-MS/MS using an Eksigent nano-LC 2D HPLC system (Eksigent, Dublin, CA), which was directly connected to a quadrupole time-of-flight (QqTOF) QSTAR Elite mass spectrometer (MDS SCIEX, Concord, CAN). To compensate for differences in the amount of glyco/peptides in the various lectin fractions, we injected 10-20% of the bound and 5% of the flow-through material. Briefly, peptides were applied to a guard column (C18 Acclaim PepMap100, 300 μ m I.D. $\times$ 5 mm, 5 μ m particle size, 100 Å pore size, Dionex, Sunnyvale, CA) and washed with the aqueous loading solvent (2% solvent B in A, flow rate: 20 μ L/min) for 10 min. Subsequently, samples were transferred onto the analytical C18-nanocapillary HPLC column (C18 Acclaim PepMap100, 300 μ m I.D. $\times$ 15 cm, 3 μ m particle size, 100 Å pore size, Dionex, Sunnyvale, CA) and eluted at a flow rate of 300 nL/min using one of two gradients: Gradient 1 (~ 3 h): at 2% solvent B in A for 5 min, 2-40% solvent B in A (from 5-125 min), 40-90% solvent B in A (from 125-140 min), and at 90% solvent B in A (from 140-149 min), with a total runtime of 194 min (including column equilibration), or Gradient 2 (~ 2 h): 2-40% solvent B in A (from 0-60 min), 40-90% solvent B in A (from 60-75 min), and at 90% solvent B in A (from 75-85 min), with a total runtime of 120 min (including column equilibration). Solvent A consisted of 0.1% formic acid in 98% H₂O/2% acetonitrile and solvent B was 0.1% formic acid in 98% acetonitrile/2% H₂O. Mass spectra (ESI-MS) and tandem mass spectra (ESI-MS/MS) were recorded in positive-ion mode with a resolution of 12,000-15,000 FWHM. The nanospray needle voltage was typically 2,300 V in HPLC-MS mode. Spectra were calibrated in static nanospray mode using MS/MS fragmentations of a Glu-Fibrinogen B peptide standard (y_1 fragment ion with m/z at 175.1195, and y_{11} fragment ion with m/z at 1285.5444). For collision induced dissociation tandem mass spectrometry (CID-MS/MS), the mass window for precursor ion selection of the quadrupole mass analyzer was set to ± 1 m/z . The precursor ions were fragmented in a collision cell using nitrogen. Advanced information dependent acquisition was employed for MS/MS data collection, using QSTAR Elite (Analyst QS 2.0) specific features, including “Smart Collision” and “Smart Exit” (fragment intensity multiplier set to 2.0 and maximum accumulation time at 2 sec) to obtain MS/MS spectra for the six most abundant parent ions following each survey scan. Dynamic exclusion features, based on value M, were set to a width of 50 mDa and exclusion duration of 120 sec. For further details see Supplementary Methods. To increase overall sampling efficiencies, two or three replicate nano-HPLC-MS/MS experiments were performed per sample.

Database searches

Mass spectrometric data were analyzed with a bioinformatics database search engine system Protein Pilot 3.0 (revision 114732) using the Paragon Algorithm 3.0.0.0 11342 developed by ABI [35]. Peaklists for the QSTAR Elite LC-MS/MS data sets were generated using Protein Pilot 3.0 (revision 114732). The following search parameters were used: trypsin enzyme specificity, carbamidomethyl (Cys) as a fixed modification, “thorough search effort setting” allowing for “biological modifications”. In contrast to other search engines, the “thorough” mode of Paragon Algorithm determines modifications using feature probabilities based on

the specified settings [35]. The Protein Pilot “Modifications Catalog and Translations.xls” file, and the “ParameterTranslation.xml” file, that were used in this study, were uploaded to Tranche (see below). The mass tolerance for the QSTAR Elite data were generally considered <50 ppm on the precursor and the fragment ion levels. In addition, during the search, Protein Pilot 3.0 performs a mass recalibration of the datasets based on highly confident peptide spectra. Specifically, a first search iteration is done to select high confidence peptide identifications that are used to recalibrate both the MS and MS/MS data, which is automatically re-searched.

All data were searched using the publicly available SwissProt database release version 57.6 (28-Jul-2009). For plasma samples, a merged database of 20337 sequences was generated that contained sequences from 3 sources: 1) human; 2) *Saccharomyces cerevisiae* invertase standards: [P10594 (INV1_YEAST), P00724 (INV2_YEAST), P10595 (INV3_YEAST), P00724 (INV2_YEAST), and P10597 (INV5_YEAST)]; and 3) the *Elizabethkingia miricola* (Chryseobacterium miricola) PNGase F enzyme precursor P21163|PNGF_ELIMR.

To assess and restrict rates of false positive peptide/protein identifications, we used the Proteomics System Performance Evaluation Pipeline (PSPEP) tool available in ProteinPilot 3.0. This tool automatically creates a concatenated forward and reverse decoy database, and provides an Excel output of the experimentally determined false discovery rate at the spectral, peptide and protein levels with standard statistical errors [36]. The discriminating variable for the Paragon Algorithm is the peptide confidence value, which is a 0-99.0 scaled real number [35]. For database searches, a cut-off peptide confidence value of 95 was chosen with the following justification. Typically, for the lectin-enriched plasma datasets, the Protein Pilot false discovery rate (FDR) analysis tool (PSPEP) algorithm [36] provides a local FDR at 5% for a peptide confidence score of 94, and a global FDR at 1% for a peptide confidence score of 90. For all further data processing, the cut-off peptide confidence value of 95 was chosen. A key requirement for the positive identification of a PNGaseF-deglycosylated peptide was the motif NXS/T (in which X represents any amino acid except P), where N was converted to D; the N→D conversion was reported by a search engine as N deamidation. Based on these criteria, these peptides, which were classified as “deglycosylated”, are listed in Supplementary Tables 2 and 3 as well as in Supplementary Tables 4 and 5 (the latter showing only the deglycosylated peptides). To ensure inclusion of glycosites containing lysine and/or arginine in the X position (e.g., NKT), which were likely to have been cleaved by trypsin, the amino-acid residue following the carboxy-terminal cleavage site was also considered. Peptides containing the motif NGS or NGT were excluded due to the fact that asparagine residues in that sequence are prone to chemical deamidation during overnight trypsin digestion [37]. The latter peptides are listed in Supplementary Tables 4 and 5 (see worksheet 2). For all deglycosylated peptides we examined the corresponding MS/MS spectra to ensure correct assignment. The MS/MS spectra were manually inspected based on an adaptation of previously published criteria [38] and the peptide was confirmed or deleted from the table. Specifically, deglycosylated peptides were validated according to the following criteria: (a) a good signal-to-noise ratio for the majority of fragment ions ($S/N > 3$); (b) a minimum of 3 consecutive y- or b-ions; (c) fragment ions that clearly determine the site of N-deamidation/deglycosylation; (d) intense y-ions from N-terminal proline; (e) an immonium ion for Arg (m/z 175.12) or Lys (m/z 147.11) consistent with the expected C-terminal tryptic site; and (f) a clean precursor ion isotope pattern demonstrating quantitative N-deamidation/deglycosylation.

Data accession

The data associated with this manuscript may be downloaded from ProteomeCommons.org Tranche using the following hash: udWtxXPMwG0TDQ7KrbC8o

+14nI1aEGopRFn44unPCO119rWL8a2GWUH2edn3N76L2PAgiUsPuXT7wdKLH3rLC2c
oEAMAAAAAAACVBA==

The hash may be used to prove exactly what files were published as part of this manuscript's data set, and the hash may also be used to check that the data has not changed since publication. Accessible information includes all raw data files and processed data report files, as well as detailed data file annotation legends.

Data analysis/bioinformatics

Detailed information on identified peptides was extracted from ProteinPilot 3.0 result files. The peptide data generated in all experiments were assembled into a single table using software developed in-house. Retention times (RT) in different runs were aligned across all runs using a linear trend calculated by fitting a line to the RT pairs for all common observations among runs. All distinct peptides that were observed in at least one run and were identified with a confidence score of at least 95 were combined to generate a Master List. The results of each individual run were then compared to the Master List. If a peptide was identified in a run with a score of 95 or above, it was retained. Peptides with scores of < 50 were discarded, while those with scores between 50 and 95 were included if their aligned RTs differed from those of a high-confidence (>95) peptide identification by ≤ 25 seconds. The program totaled the number of all peptides and deglycosylated peptides observed in each sample. These totals, along with the percentage of deglycosylated peptides relative to total peptides observed for each sample were displayed. Due to a combination of missed tryptic cleavage sites (partial digestion) and non-specific cleavages, we observed multiple peptides covering the same protein glycosylation site. To ensure that only distinct glycosylation sites were counted, each deglycosylated peptide was annotated with the amino acid number of the glycosylated residue (the glycosite). Then, the number of distinct glycosites observed was reported. Finally, the program monitored proteins with at least one deglycosylated peptide and produced a table of glycoproteins (and the corresponding glycosites) observed in the samples.

Results

Glycopeptide standards demonstrate the specificity of lectin binding

In these experiments, we took the approach of developing, as a first step, standards that allowed us to optimize the conditions that favor glycopeptide enrichment, which was a key element in our experimental design. As a test of lectin specificity, we interrogated the binding characteristics of the columns by chromatographing positive controls that met the structural requirements for interaction and negative controls that did not. We selected standards with well-characterized glycan structures [39-41] that reflected the specificity of the lectins we employed. Specifically, SNA binds sialic acid α 2,6- and α 2,3-linked to terminal galactose, and AAL binds fucose α 1,6-linked to *N*-acetylglucosamine and α 1,3-linked to *N*-acetylglucosamine. Accordingly, we expected that sialylated and fucosylated human lactoferrin glycopeptides [39; 40] would bind to both SNA and AAL. By contrast, yeast invertase glycopeptides, which carry only high mannose structures, should not interact with either lectin [41]. Indeed, this was the case for both lectins, as exemplified by the AAL data shown in Fig. 1 (and substantiated by the MS data shown in Supplementary Figures 1A-1K). In the case of lactoferrin, glycopeptides were concentrated in the bound fraction and not observed in the flow-through. The principle species detected were two overlapping peptide sequences that were modified with various fucose-containing glycoforms (as confirmed by MALDI-MS/MS analyses, Supplementary Figures 1A-1G, and Supplementary Table 1). Conversely, invertase-derived glycopeptides were observed predominantly in the flow-through fraction. To demonstrate that the lectin chromatographic separation was

glycan-dependent and reproducible, we rechromatographed the glycopeptide-containing fractions from each standard and demonstrated consistent behavior. Moreover, the lactoferrin glycopeptides recovered in the bound fraction were further enriched relative to non-glycosylated species. Accordingly, we used unseparated invertase glycopeptides and lectin-enriched lactoferrin glycopeptides as internal standards in all subsequent experiments.

Internal standards enabled the optimization of chromatographic parameters for separating human plasma glycopeptides

Next, we optimized the conditions for HPLC separation of human plasma glycopeptides (Fig. 2). Specifically, we digested MARS 14-depleted human plasma with trypsin, spiked in the glycopeptide standards and fractionated the mixture by using lectin chromatography. To prove binding specificity and facilitate peptide identification, we used PNGase F to remove *N*-linked glycans. This enzyme, an amidase, catalyzes the hydrolysis of the amide bond on the asparagine side chain, releasing the glycan and converting asparagine to aspartic acid. This modification produces a +1 Da increment in the molecular mass of the deglycosylated peptide. When observed in the context of an *N*-linked consensus sequence (NXS/T, where X is any residue except proline), this mass shift strongly suggested that the peptide was *N*-glycosylated. Due to technical constraints, lectin enrichment of *O*-linked glycopeptides was not addressed in these studies, and all further references to glycopeptides pertain to *N*-linked species unless otherwise stated.

Initial optimization focused on the PNGaseF digestion conditions. Specifically, a wide range of enzyme:substrate ratios and reaction volumes was assessed. The extent of deglycosylation of glycopeptide standards and trypsin-digested MARS 14-depleted plasma was monitored by MS. For the standards, MALDI MS was used to detect the loss of ions corresponding to glycopeptides, and the corresponding increases in signals from the deglycosylated peptides (data not shown). For plasma, ESI-HPLC MS/MS was used to assess the extent of deglycosylation. Over the range of enzyme concentrations tested, 1 –20 U/μg of peptide, we observed only a modest increase (max. $\Delta = 2\%$) in product formation. Accordingly, we used the highest enzyme concentration that was practical given cost considerations (2.5 U/μg).

Next, we optimized the affinity chromatography methods. First, we focused on buffer A (25 mM Tris HCl, 10 mM CaCl₂, 10 mM MgCl₂), which was supplemented with increasing amounts of NaCl (0-500 mM) in an effort to identify the maximum salt concentration at which lectin-specific interactions were maintained as determined by the partitioning of lactoferrin and invertase glycopeptides to the bound and flow-through fractions, respectively. Perturbation of lactoferrin binding was first observed at 100 mM NaCl and was exacerbated at higher salt concentrations. Therefore 50 mM was chosen as the working concentration. We also optimized the column wash step, *i.e.*, the elution time between collection of the flow-through and bound fractions with the goal of maximizing recovery of bound glycopeptides, as determined by LC MS/MS, and to increase the reproducibility of the elution profiles. Initially, we used a high-flow, high-pressure wash step (2 mL/min at 1600 psi) of approximately 40 column volumes. However, under these conditions, glycopeptides eluted from the column before application of the elution buffer, indicating that some glycan-mediated interactions did not withstand extensive washing. Furthermore, the UV traces of the wash were highly variable. Therefore, we significantly modified this step by reducing both the flow rate and the volume. Ultimately, we used 0.5 column volumes under low-flow and low-pressure conditions (50 μL/min at 30 psi), which preserved glycopeptide binding and stabilized the UV-monitored elution profile.

Lectin-specific differences in the specificity of glycopeptide binding correlated with the amount of starting material

As a last step, we optimized the sample load with the goal of maximizing the number of glycopeptides recovered in the bound fraction. We tested a range (10-200 μ l) of plasma volumes (as defined in Experimental Procedures) corresponding to ~50-1000 μ g of peptides, and determined the number of glycopeptides and the percent glycopeptide enrichment (total glycopeptides/total peptides) in the flow-through and bound fractions (Table 1). Interestingly, the results differed between AAL and SNA, perhaps reflecting the relative availability of their target carbohydrate modifications among human plasma glycoproteins, as well as the significant structural differences in the lectin molecules themselves. For example, at low peptide loads, SNA selected a large number of glycopeptides (>150) at a high percent enrichment (>50%), while AAL produced a similar level of enrichment (>40% enrichment), but bound very few glyco- and non-glycopeptides. In both cases, applying increasing amounts of starting material to the column enabled the identification of greater numbers of bound glycopeptides. However, SNA-specific glycopeptide binding plateaued at 500 μ g of sample, whereas the number of glycopeptides that bound to AAL continued increasing through the highest amount tested (*i.e.*, 1 mg).

For both lectins, the percent enrichment decreased as more sample was applied to the column. The flow-through fractions also reflected differences in lectin binding. At low peptide concentrations, fewer glycopeptides were detected in the unbound SNA fraction, as determined by the absolute glycopeptide numbers and percent enrichment. As the peptide load increased, the glycopeptide numbers and enrichment levels became more similar for the two lectins. As the goal of this workflow was to enable robust and comprehensive biomarker discovery efforts, we elected to sacrifice specificity of glycopeptide enrichment in favor of increasing the total number of bound glycopeptides. Accordingly, 200 μ l of plasma (~1 mg of peptides) was chosen as the final volume of starting material. The protocol was provided online as a video demonstration intended to enable viewers to develop their own lectin chromatography systems [42], however the lectin affinity datasets, specific results and conclusions are exclusively described in this paper.

SNA and AAL affinity chromatography identified low-abundance plasma glycoproteins

We evaluated performance of the optimized lectin affinity step by separating MARS 14-depleted, trypsin-digested plasma from a pool of healthy donors. Numerous replicates were performed to assess the reproducibility of the protocol. Specifically, sample separation on AAL was repeated 14 times, and on SNA, 10 times; each bound fraction was subjected to either duplicate or triplicate LC-MS/MS analyses. The entire dataset for flow-through and bound fractions is presented in Supplementary Tables 2 and 3. This includes all peptide sequences ($\pm$ modifications) with confidence scores and retention times. In addition, for both fractions of each experimental replicate, the table summarizes the number of peptides, glycopeptides, and the % glycopeptide enrichment. Please note that glycopeptides containing NGS/T in the *N*-linked consensus sequence were excluded from the list, as asparagines neighboring glycines are easily deamidated during trypsin digestion [37]. Additionally, we noted 4 putative glycopeptides containing the sequence NXC, which has recently been proposed as an alternative *N*-linked consensus sequence [43]. MS/MS spectra for these unusual glycosites is included in the supplement, however they were excluded from the list of glycosites observed. For easy reference, the glycoproteins and their glycosites are summarized at the bottom of each table. Spectra documenting each of the observed glycopeptides are included as Supplementary Figure. 2. Finally, separate spreadsheets listing exclusively the observed glycopeptides for AAL and SNA chromatography, and the excluded NGS/NGT-containing peptides are supplied as Supplementary Tables 4 and 5.

Key elements of the data are summarized as follows. Overall, we used two binary gradient elution profiles (2 h or 3 h) for the reversed-phase separation prior to MS/MS. The total number of peptides identified in each run varied accordingly—563-842 and 1035-1726 for the shorter and longer times, respectively. With respect to distinct glycosites, 48-69 were observed in the flow-through fraction, 53-128 bound to AAL, and 117-145 interacted with SNA. The percent glycopeptide/total peptide values were 4-5% in the flow-through fraction for each lectin, increasing to 8-15% in the AAL bound portion, and 11-18% in the subset that interacted with SNA. In total, we identified 119 human plasma glycoproteins, containing 227 distinct glycosites, of which 62 (52%) bound both lectins; 37 (31%) bound to AAL and 20 (17%) interacted with SNA (Fig. 3A). The glycosites demonstrated a slightly different distribution; 117 (51%) bound to both lectins, while AAL and SNA each bound 55 (24%), possible evidence of heterogeneity of glycan placement along the protein backbone (Fig. 3B). In support of this, we observed that 22 glycoproteins contained multiple glycosites that were differentially identified by the two lectins. For example, complement C2 contained 3 glycosites observed in this study—N29, N621 and N651. While the first glycosite was captured by both lectins, N621 was detected only with SNA, and N651 only with AAL (Table 2).

As to the reproducibility of this workflow, similar glycosites were identified in replicate lectin chromatography experiments as determined by two measures—the percentage of total glycosites observed in a given number of experiments (Fig. 4A), and pairwise repeatability, or the percentage of glycosites present in two replicate datasets (Fig. 4B). With respect to the former, fully 35% of all glycosites detected in the AAL workflow were found in the bound fractions of all 14 experiments, and >65% were observed in at least half. SNA chromatography was even more robust, with 60% of all glycosites observed in all 10 experiments, and close to 90% observed in 5 or more. With respect to the latter, the median pairwise repeatability was 73% for AAL and 80% for SNA.

All plasma glycoproteins and glycosites bound by AAL and SNA are listed in Table 2, along with information (where available) regarding the extent of protein glycosylation and published concentration ranges in human plasma or serum. We were particularly interested in the latter values, as the major goal of this work was to employ lectin chromatography followed by MS to capture and identify low-abundance glycopeptides from biological samples. Two measures suggest we met this objective. First, of the 119 identified glycoproteins, 55 (46%) were not counted among the top 150 most-abundant plasma proteins, which are present at concentrations greater than 600-900 ng/mL [44]. Second, at lower abundances, we detected, with high confidence, proteins that circulate at levels that are comparable to those that are currently used as biomarkers (Fig. 5). These included cadherin-5, which is found at 0.3-4.9 ng/mL, neutrophil gelatinase-associated lipocalin with levels estimated at ~2.5 ng/mL, and hepatocyte growth factor activator, which is present at 3-6 ng/mL [45-47].

Discussion

Although lectin chromatography had been previously used as a method for glycoprotein/ glycopeptide enrichment [48; 49], five criteria, chosen to broaden and strengthen the utility of this protocol, shaped our approach. First, our goal was to develop a robust, easily adaptable method that would be within the financial and technical reach of most laboratories engaged in biomarker discovery. Second, we focused on biologically relevant species, *i.e.*, carbohydrate modifications that were associated with cancer. Third, to establish lectin specificity, we included positive and negative glycopeptide standards. Fourth, we used relatively small plasma volumes that are more easily accessible from banked collections of high-quality patient and control samples. Fifth, and most importantly, we sought to develop

a method that could access low-abundance glycoproteins in a complex mixture (*i.e.*, human plasma) for the purpose of biomarker discovery.

We accomplished these goals by optimizing specific components of the workflow. The final protocol required a minimum number of procedures and no specialized equipment beyond that found in a typical protein research laboratory that uses standard chromatographic separation strategies and has access to routine biological MS technology. The lectins chosen for initial studies, AAL and SNA, recognize fucose- and sialic acid-containing carbohydrates, respectively, since expression of the glycosyltransferases that add these terminal residues is frequently upregulated in cancer cells [50]. We used glycopeptide standards as quality control measures to monitor the performance of the lectin chromatography step. For this purpose, we used fucosylated and sialylated *N*-linked lactoferrin peptides, which met the structural requirements for lectin binding, and high-mannose species from invertase, which did not. In parallel, numerous parameters were assessed to develop a robust and reproducible protocol that uses relatively small plasma volumes, an important issue given the paucity of well-annotated patient and control specimen banks. Finally, using established MS platforms, our method enabled routine detection of numerous low-abundance (ng/mL) plasma glycoproteins from a pooled sample of plasmas obtained from healthy donors. Thus, by optimizing each step of the workflow, we developed a method that can be used to exploit relatively underappreciated features of cancer-related post-translational modifications, such as disease-related structural differences in glycosylation and their attachment sites in terms of the tryptic peptide sequences.

The sensitivity and specificity inherent in these data are a largely unexplored biomarker discovery space that commonly used methods (*e.g.*, lectin chromatography at the glycoprotein level, and hydrazide- or boronic acid-mediated chemical capture of glycoproteins/glycopeptides) do not typically access (Table 3). Specifically, our ability to study oligosaccharide changes at the glycosite level allows a powerful means to assess subtle disease-related modifications in carbohydrate structure and placement along the protein backbone. Accordingly, this study identified 23 glycoproteins from healthy pooled human plasma in which distinct glycosites were detected using different lectins (Table 2). We anticipate that this precision will make our workflow highly responsive to cancer-related variations in oligosaccharide structure.

As we work at the glycopeptide level, the % glycopeptide enrichment in our workflows cannot be directly compared with that of lectin chromatography performed at the glycoprotein level. The latter methods commonly use both peptides and glycopeptides to identify enriched proteins. As a majority of human plasma proteins are glycosylated [50], both specific and non-specifically bound species are likely to be derived from glycoproteins, with the latter subset artificially contributing to the reported enrichment. By contrast, our workflow set a very high bar for the identification of a lectin-bound glycoprotein—we required that the mass spectrometric data specifically identified the glycosite itself (as a deamidated asparagine after PNGase F treatment). Importantly, the MS/MS spectra for each glycosite-containing peptide were manually inspected as described in the Experimental Procedures (excluding NGS/T sequences as a conservative measure).

During method optimization, we found that AAL and SNA behaved differently with respect to glycopeptide capture from a given sample volume. Specifically, SNA bound more glycopeptides from a smaller sample volume than did AAL. This effect likely reflected the greater abundance of sialylated as compared to fucosylated species in human plasma, although other factors such as differences in carbohydrate binding kinetics between these two lectins could also play a role. As the amount of sample loaded onto the column was increased, the number of glycopeptides identified by both lectins rose, however the binding

specificity decreased. Because reaching the biologically-relevant glycoproteome was our primary goal, we pursued the method that allowed deeper glycopeptide coverage. This was a feasible strategy as our glycopeptide standards confirmed that the columns were indeed selecting fucosylated and sialylated species, and the presence of an *N*-linked consensus sequence deglycosylated at the Asn residue allowed facile detection of deglycosylated peptides. Preliminary experiments in which we assessed the chromatographic efficiency of various column sizes suggested that the observed enrichment effect was related to the peptide content of the sample in relationship to the column bed volume (data not shown). Although we selected final conditions that optimized detection of a maximum number of lectin-bound glycopeptides from a minimal amount of plasma using relatively inexpensive amounts of lectins, it is likely that these parameters could be modified to suit a variety of endpoints. For instance, investigators that are not limited by sample volume or cost might scale-up to larger columns to achieve higher yields of enriched glycopeptides. Thus, in accord with the fact that individual lectin species have unique structural features [51] and distinct saccharide binding motifs, our work shows that the performance of each carbohydrate-binding protein can and should be optimized using the appropriate glycopeptide standards.

We elected to begin these studies with two lectins of complimentary glycan specificity, which detected distinct cancer-related carbohydrate modifications. This choice was motivated by the expectation that these lectins would enable identification of distinct, but overlapping, plasma glycopeptide subsets. Accordingly, we anticipated that this approach would allow us to identify greater numbers of low-abundance glycoproteins, thus reaching deeper into the plasma glycoproteome. Indeed, these hypotheses proved to be correct. Of the 119 plasma glycoproteins observed in the bound species, approximately 52% were selected by both lectins, 31% by AAL, and 17% by SNA. Importantly, 90% of the glycoproteins identified using AAL alone and 30% of those observed with SNA alone were low-abundance relative to the previously described set of proteins occupying the upper 4 log concentration range in human plasma [44]. Furthermore, the glycoproteins comprised a total of 227 unique glycosites, 6% of which had not been previously annotated as such, that were selected by lectin chromatography. Of these, 40% were bound by both AAL and SNA, while the remainder were selected by a single lectin. This observation illustrates that the described method can be used to detect glycan heterogeneity at distinct sites along a peptide backbone, in contrast to alternative techniques involving chemical capture or lectin chromatography at the glycoprotein level. Our findings also support the notion that judicious selection of lectins with regard to the biology of the system results in enriched capture of disease- or process-specific glycoproteins. Finally, the overall strategy we developed including the chromatographic enrichment protocol is easily adapted for other protein-based affinity selectors (our unpublished observations). Currently, we are focusing on carbohydrate-specific antibodies for the enrichment of glycosylated species bearing biologically-interesting saccharide motifs as demonstrated by the use of relevant and irrelevant glycopeptide standards [50].

In summary, this study design can serve as a blue print for the development of lectin-based workflows for capturing disease-related changes in the glycoproteome, with subsequent MS-identification of the modified protein substrates. Several features make glycoproteins especially attractive as biomarkers. First, biology can be used to guide the experimental design, and specific carbohydrate motifs can be targeted in a rational manner that is derived from the expression of the relevant glycosyltransferases in affected cells/tissues. Because glycosylation, like phosphorylation, is dynamic in nature and responsive to numerous biological stimuli [52], saccharide-motifs offer a window for viewing many normal and pathological processes. Second, as multiple saccharide-modifications can decorate a single protein scaffold, and as a single glycoform can be appended to numerous protein backbones,

carbohydrate-based signals can be greatly amplified relative to a single protein sequence. Third, carbohydrate modifications offer some protection from proteolytic degradation, which raises the possibility that glycosylated molecules may be more biologically stable than their unmodified counterparts. Thus, we are very interested in applying this workflow and modifications, which utilize other lectins, for biomarker discovery projects that employ cell lines, tissues, and biological fluids in discovery-phase analyses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol.* 2006; 24:971–83. [PubMed: 16900146]
2. Arnold JN, Saldivo R, Hamid UM, Rudd PM. Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics.* 2008; 8:3284–93. [PubMed: 18646009]
3. Meezan E, Wu HC, Black PH, Robbins PW. Comparative studies on the carbohydrate-containing membrane components of normal and virus-transformed mouse fibroblasts. II. Separation of glycoproteins and glycopeptides by sephadex chromatography. *Biochemistry.* 1969; 8:2518–24. [PubMed: 4307997]
4. Hammarstrom S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin Cancer Biol.* 1999; 9:67–81. [PubMed: 10202129]
5. Meany DL, Zhang Z, Sokoll LJ, Zhang H, Chan DW. Glycoproteomics for prostate cancer detection: changes in serum PSA glycosylation patterns. *J Proteome Res.* 2009; 8:613–9. [PubMed: 19035787]
6. Moss EL, Hollingworth J, Reynolds TM. The role of CA125 in clinical practice. *J Clin Pathol.* 2005; 58:308–12. [PubMed: 15735166]
7. Sanders DS, Kerr MA. Lewis blood group and CEA related antigens; coexpressed cell-cell adhesion molecules with roles in the biological progression and dissemination of tumours. *Mol Pathol.* 1999; 52:174–8. [PubMed: 10694936]
8. Friedrich MG, Hellstern A, Hautmann SH, Graefen M, Conrad S, Huland E, Huland H. Clinical use of urinary markers for the detection and prognosis of bladder carcinoma: a comparison of immunocytology with monoclonal antibodies against Lewis X and 486p3/12 with the BTA STAT and NMP22 tests. *J Urol.* 2002; 168:470–4. [PubMed: 12131290]
9. Nakagoe T, Fukushima K, Nanashima A, Sawai T, Tsuji T, Jibiki M, Yamaguchi H, Yasutake T, Ayabe H, Matuo T, Tagawa Y, Arisawa K. Expression of Lewis(a), sialyl Lewis(a), Lewis(x) and sialyl Lewis(x) antigens as prognostic factors in patients with colorectal cancer. *Can J Gastroenterol.* 2000; 14:753–60. [PubMed: 11064310]
10. Narita T, Funahashi H, Satoh Y, Watanabe T, Sakamoto J, Takagi H. Association of expression of blood group-related carbohydrate antigens with prognosis in breast cancer. *Cancer.* 1993; 71:3044–53. [PubMed: 8490832]
11. Ogawa J, Sano A, Inoue H, Koide S. Expression of Lewis-related antigen and prognosis in stage I non-small cell lung cancer. *Ann Thorac Surg.* 1995; 59:412–5. [PubMed: 7847958]
12. Ohyama C. Glycosylation in bladder cancer. *Int J Clin Oncol.* 2008; 13:308–13. [PubMed: 18704630]
13. Pode D, Golijanin D, Sherman Y, Lebensart P, Shapiro A. Immunostaining of Lewis X in cells from voided urine, cytopathology and ultrasound for noninvasive detection of bladder tumors. *J Urol.* 1998; 159:389–92. discussion 393. [PubMed: 9649245]
14. Steplewska-Mazur K, Gabriel A, Zajecki W, Wylezol M, Gluck M. Breast cancer progression and expression of blood group-related tumor-associated antigens. *Hybridoma.* 2000; 19:129–33. [PubMed: 10868792]

15. Tauchi K, Kakudo K, Machimura T, Makuuchi H, Mitomi T. Immunohistochemical studies of blood group-related antigens in human superficial esophageal carcinomas. *Cancer*. 1991; 67:3042–50. [PubMed: 2044050]
16. Ohyama C, Hosono M, Nitta K, Oh-eda M, Yoshikawa K, Habuchi T, Arai Y, Fukuda M. Carbohydrate structure and differential binding of prostate specific antigen to Maackia amurensis lectin between prostate cancer and benign prostate hypertrophy. *Glycobiology*. 2004; 14:671–9. [PubMed: 15044396]
17. Vashi AR, Wojno KJ, Henricks W, England BA, Vessella RL, Lange PH, Wright GL Jr, Schellhammer PF, Weigand RA, Olson RM, Dowell BL, Borden KK, Oesterling JE. Determination of the “reflex range” and appropriate cutpoints for percent free prostate-specific antigen in 413 men referred for prostatic evaluation using the AxSYM system. *Urology*. 1997; 49:19–27. [PubMed: 9000179]
18. Rosen SD. Ligands for L-selectin: homing, inflammation, and beyond. *Annu Rev Immunol*. 2004; 22:129–56. [PubMed: 15032576]
19. Witz IP. The selectin-selectin ligand axis in tumor progression. *Cancer Metastasis Rev*. 2008; 27:19–30. [PubMed: 18180878]
20. Thomas SN, Zhu F, Schnaar RL, Alves CS, Konstantopoulos K. Carcinoembryonic antigen and CD44 variant isoforms cooperate to mediate colon carcinoma cell adhesion to E- and L-selectin in shear flow. *J Biol Chem*. 2008; 283:15647–55. [PubMed: 18375392]
21. Borsig L, Wong R, Feramisco J, Nadeau DR, Varki NM, Varki A. Heparin and cancer revisited: mechanistic connections involving platelets, P-selectin, carcinoma mucins, and tumor metastasis. *Proc Natl Acad Sci U S A*. 2001; 98:3352–7. [PubMed: 11248082]
22. Borsig L, Wong R, Hynes RO, Varki NM, Varki A. Synergistic effects of L- and P-selectin in facilitating tumor metastasis can involve non-mucin ligands and implicate leukocytes as enhancers of metastasis. *Proc Natl Acad Sci U S A*. 2002; 99:2193–8. [PubMed: 11854515]
23. Biancone L, Araki M, Araki K, Vassalli P, Stamenkovic I. Redirection of tumor metastasis by expression of E-selectin in vivo. *J Exp Med*. 1996; 183:581–7. [PubMed: 8627169]
24. Comunale MA, Wang M, Hafner J, Krakover J, Rodemich L, Kopenhagen B, Long RE, Junaidi O, Bisceglie AM, Block TM, Mehta AS. Identification and development of fucosylated glycoproteins as biomarkers of primary hepatocellular carcinoma. *J Proteome Res*. 2009; 8:595–602. [PubMed: 19099421]
25. Jung K, Cho W, Regnier FE. Glycoproteomics of plasma based on narrow selectivity lectin affinity chromatography. *J Proteome Res*. 2009; 8:643–50. [PubMed: 19099503]
26. Xiong L, Andrews D, Regnier F. Comparative proteomics of glycoproteins based on lectin selection and isotope coding. *J Proteome Res*. 2003; 2:618–25. [PubMed: 14692455]
27. Calvano CD, Zamboni CG, Jensen ON. Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry. *J Proteomics*. 2008; 71:304–17. [PubMed: 18638581]
28. Heo SH, Lee SJ, Ryoo HM, Park JY, Cho JY. Identification of putative serum glycoprotein biomarkers for human lung adenocarcinoma by multilectin affinity chromatography and LC-MS/MS. *Proteomics*. 2007; 7:4292–302. [PubMed: 17963278]
29. Plavina T, Wakshull E, Hancock WS, Hincapie M. Combination of abundant protein depletion and multi-lectin affinity chromatography (M-LAC) for plasma protein biomarker discovery. *J Proteome Res*. 2007; 6:662–71. [PubMed: 17269723]
30. Yang Z, Harris LE, Palmer-Toy DE, Hancock WS. Multilectin affinity chromatography for characterization of multiple glycoprotein biomarker candidates in serum from breast cancer patients. *Clin Chem*. 2006; 52:1897–905. [PubMed: 16916992]
31. Cho W, Jung K, Regnier FE. Use of glycan targeting antibodies to identify cancer-associated glycoproteins in plasma of breast cancer patients. *Anal Chem*. 2008; 80:5286–92. [PubMed: 18558770]
32. Zhang H, Li XJ, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol*. 2003; 21:660–6. [PubMed: 12754519]

33. Sparbier K, Koch S, Kessler I, Wenzel T, Kostrzewa M. Selective isolation of glycoproteins and glycopeptides for MALDI-TOF MS detection supported by magnetic particles. *J Biomol Tech.* 2005; 16:407–13. [PubMed: 16522863]
34. Keshishian H, Addona T, Burgess M, Kuhn E, Carr SA. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics.* 2007; 6:2212–29. [PubMed: 17939991]
35. Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP, Hunter CL, Nuwaysir LM, Schaeffer DA. The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol Cell Proteomics.* 2007; 6:1638–55. [PubMed: 17533153]
36. Tang WH, Shilov IV, Seymour SL. Nonlinear fitting method for determining local false discovery rates from decoy database searches. *J Proteome Res.* 2008; 7:3661–7. [PubMed: 18700793]
37. Krokhn OV, Antonovici M, Ens W, Wilkins JA, Standing KG. Deamidation of -Asn-Gly-sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. *Anal Chem.* 2006; 78:6645–50. [PubMed: 16970346]
38. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR 3rd. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol.* 1999; 17:676–82. [PubMed: 10404161]
39. Matsumoto A, Yoshima H, Takasaki S, Kobata A. Structural study of the sugar chains of human lactoferrin: finding of four novel complex-type asparagine-linked sugar chains. *J Biochem.* 1982; 91:143–55. [PubMed: 7068558]
40. Spik G, Strecker G, Fournet B, Bouquelet S, Montreuil J, Dorland L, van Halbeek H, Vliegthart JF. Primary structure of the glycans from human lactotransferrin. *Eur J Biochem.* 1982; 121:413–9. [PubMed: 7060557]
41. Trimble RB, Atkinson PH. Structure of yeast external invertase Man8-14GlcNAc processing intermediates by 500-megahertz ¹H NMR spectroscopy. *J Biol Chem.* 1986; 261:9815–24. [PubMed: 3525534]
42. Johansen, E.; Schilling, B.; Lerch, M.; Niles, RK.; Liu, H.; Li, B.; Allen, S.; Hall, SC.; Witkowska, HE.; Regnier, FE.; Gibson, BW.; Fisher, SJ.; Drake, PM. A lectin HPLC method to enrich selectively-glycosylated peptides from complex biological samples. *J Vis Exp.* 2009. <http://www.jove.com/index/Details.stp?ID=1398>Video Article
43. Zielinska DF, Gnadt F, Wisniewski JR, Mann M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell.* 2010; 141:897–907. [PubMed: 20510933]
44. Hortin GL, Sviridov D, Anderson NL. High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance. *Clin Chem.* 2008; 54:1608–16. [PubMed: 18687737]
45. Mishra J, Dent C, Tarabishi R, Mitsnefes MM, Ma Q, Kelly C, Ruff SM, Zahedi K, Shao M, Bean J, Mori K, Barasch J, Devarajan P. Neutrophil gelatinase-associated lipocalin (NGAL) as a biomarker for acute renal injury after cardiac surgery. *Lancet.* 2005; 365:1231–8. [PubMed: 15811456]
46. Soeki T, Tamura Y, Shinohara H, Sakabe K, Onose Y, Fukuda N. Elevated concentration of soluble vascular endothelial cadherin is associated with coronary atherosclerosis. *Circ J.* 2004; 68:1–5. [PubMed: 14695457]
47. Watanabe T, Sakata Y, Matsubara S, Yamagishi T, Nagaike K, Kuwata T, Suzuki M. Changes in plasma levels of hepatocyte growth factor and its associated factors during pregnancy. *J Obstet Gynaecol Res.* 2006; 32:10–4. [PubMed: 16445520]
48. Ito S, Hayama K, Hirabayashi J. Enrichment strategies for glycopeptides. *Methods Mol Biol.* 2009; 534:195–203. [PubMed: 19277551]
49. Madera M, Mechref Y, Novotny MV. Combining lectin microcolumns with high-resolution separation techniques for enrichment of glycoproteins and glycopeptides. *Anal Chem.* 2005; 77:4081–90. [PubMed: 15987113]
50. Drake PM, Cho W, Li B, Prakobphol A, Johansen E, Anderson NL, Regnier FE, Gibson BW, Fisher SJ. Sweetening the Pot: Adding Glycosylation to the Biomarker Discovery Equation. *Clin Chem.* 2010; 56:223–236. [PubMed: 19959616]

51. De Hoff PL, Brill LM, Hirsch AM. Plant lectins: the ties that bind in root symbiosis and plant defense. *Mol Genet Genomics*. 2009; 282:1–15. [PubMed: 19488786]
52. Alavi A, Axford JS. Sweet and sour: the impact of sugars on disease. *Rheumatology (Oxford)*. 2008; 47:760–70. [PubMed: 18375404]

Fig. 1. Glycopeptide standards bound to AAL, which interacts with fucosylated oligosaccharides

Glycopeptides from human lactoferrin, which carries fucosylated and sialylated structures, served as positive controls; glycopeptides from invertase, which carries high-mannose structures, served as negative controls. Lactoferrin (A-C) and invertase (D-F) glycopeptides were affinity selected on AAL-conjugated POROS. Flow-through (FT) and bound (BD) fractions were collected and analyzed by MALDI-MS or MALDI-MS/MS. (A) MS spectrum of the lactoferrin FT1 fraction. (B) MS spectrum of the BD1 fraction. Peaks corresponding to various fucosylated glycoforms of three related peptides were observed (1) TAGWNVPIGTLRPFLNWTGPPEIEA AAVAR, (2) PFLNWTGPPEIEA AAVAR, and (3) LRPFLNWTGPPEIEA AAVAR; where N corresponded to glycosylated Asn with NWT as the consensus glycosylation sequence. (C) The BD1 fraction was rechromatographed and subjected to MS analysis; consistent with the initial results, glycopeptides were detected in the BD2 but not in the FT2 fraction (data not shown). The peptide backbone for these structures was confirmed by MS/MS. Monosaccharide composition and, in some cases, partial sequence information was also obtained for the carbohydrate structures (Supplementary Table 1 and Supplementary Figs. 1A-1G). These data (and the depicted glycans) were consistent with previously identified lactoferrin carbohydrate structures, and the fact that this molecule is highly fucosylated [40; 41]. (D) MS spectrum of the invertase FT1 fraction. Peaks corresponded to high-mannose structures appended to peptide “4” (AEPILNISNAGPSR). (E) MS spectrum of the BD1 fraction. The glycopeptides were also observed but at approximately 10-fold less intensity. (F) The FT1 fraction was rechromatographed and subjected to MS analysis. The four molecular ions, which were also observed in the FT1 fraction (compare panels D and F), were interrogated by MS/MS and the peptide backbone and monosaccharide composition were confirmed (Supplementary Table 1 and Supplementary Figs. 1H-1K). These data (and the depicted glycans) are consistent with the fact that invertase presents high-mannose structures [42]. No invertase glycopeptides were detected in the rechromatographed BD2 fraction (data not shown). (■, N-acetylglucosamine; ●, galactose; ○, mannose; △, fucose; ◆, sialic acid.)

Fig. 2. Lectin chromatography enhances the recovery of glycopeptides carrying specific carbohydrate modifications

Positive (lactoferrin) and negative (invertase) glycoprotein standards were trypsin digested (1) and spiked into MARS 14-depleted, trypsin-digested human plasma (2). Then, the sample was subjected to lectin chromatography (3) on immobilized SNA (which binds sialic acid) or AAL (which binds fucose). The flow-through fraction (3a) contained peptides and glycopeptides not recognized by the lectin column. After a wash step (3b), elution buffer was applied and the bound fraction (3c) containing sialylated or fucosylated peptides and non-specifically bound species was collected. *N*-glycosylated peptides were deglycosylated with PNGase F (4), a peptidase that removes the intact glycan via cleavage of the amide bond on the asparagine side chain. This results in a +1 Da mass shift (indicated by an *) that is detectable by MS. Importantly, *N*-glycans are restricted to Asn residues within the sequence NXS/T, where X is any residue except proline. Thus, the presence of both features—the consensus sequence and mass shift—indicates that a peptide was formerly glycosylated.

Fig. 3. AAL and SNA captured distinct glycoprotein and glycosite subsets

Venn diagrams depict the total number of glycoproteins (A) or unique glycosites (B) captured from MARS 14-depleted, trypsin-digested human plasma by AAL, SNA, or both lectins using the finalized method.

Fig. 4. Glycosite identifications were highly reproducible among individual AAL and SNA datasets

(A) Survival curve depicting the number of AAL lectin chromatography experiments (x-axis) in which a given percentage of all glycosites (y-axis) were observed. (B) Box-whisker plot of the percentage overlap between all the possible pairs of datasets. The AAL and SNA datasets had $14 \times 13/2 = 91$ and $10 \times 9/2 = 45$ pairs to compare, respectively. For a pair of data sets the overlap was computed as $|A \cap B| / |A \cup B|$. For example, if one of the pairs consists of 154 and 151 glycosites respectively, with 141 in common, then the percentage overlap would be $141/164 = 86\%$.

Fig. 5. High-confidence identification of a glycoprotein that typically circulates in human blood at ng/mL levels

ESI-MS/MS spectrum resulting from CID on the molecular ion MH_2^{2+} , selected at m/z 806.91 $^{2+}$ ($M=1611.80$). The spectrum was searched using Protein Pilot and peptide EVYPWYN*LTVEAK was identified (N^* = deamidated N) with a peptide confidence of 99. Fragment ions y_6 and y_7 clearly show an increment mass difference of $\Delta M = 115$ Da, localizing the deamidation to N442, the glycosite. Similar MS/MS spectra for all other glycosylated peptides are provided in Supplementary Fig. 2.

Table 1
Chromatography on AAL or SNA enabled glycopeptide/glycoprotein capture

Plasma loaded (~(μ g) ^a			50	100	200	300	400	500	1000
	Fraction^b								
% Glycopeptides	AAL FT		3 $\pm$ 0 ^c	2.5 $\pm$ 1	3 $\pm$ 0	3.5 $\pm$ 1	4 $\pm$ 0	4 $\pm$ 0	3 $\pm$ 0
	SNA FT		0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	2 $\pm$ 0	3 $\pm$ 0	3.5 $\pm$ 1	4 $\pm$ 0
	AAL BD		43 $\pm$ 20	16.5 $\pm$ 9	9.5 $\pm$ 8	12 $\pm$ 6	17.5 $\pm$ 4	16 $\pm$ 0	13.5 $\pm$ 1
	SNA BD		59.5 $\pm$ 9	53 $\pm$ 1	35 $\pm$ 1	27.5 $\pm$ 4	27.5 $\pm$ 1	25 $\pm$ 0	18 $\pm$ 0
# Glycopeptides	AAL FT		26 $\pm$ 4	27.5 $\pm$ 4	30.5 $\pm$ 4	38 $\pm$ 3	42 $\pm$ 1	48.5 $\pm$ 2	38.5 $\pm$ 1
	SNA FT		8.5 $\pm$ 1	5 $\pm$ 3	5.5 $\pm$ 1	39.5 $\pm$ 2	54.5 $\pm$ 5	61.5 $\pm$ 2	72.5 $\pm$ 1
	AAL BD		4.6 $\pm$ 1	10 $\pm$ 8	15.5 $\pm$ 16	21.5 $\pm$ 15	21 $\pm$ 6	48 $\pm$ 4	66.54 $\pm$ 8
	SNA BD		163 $\pm$ 3	182.5 $\pm$ 2	253.5 $\pm$ 21	228.5 $\pm$ 13	242 $\pm$ 0	248 $\pm$ 3	211 $\pm$ 8
# Glycoproteins	AAL FT		19.5 $\pm$ 4	20 $\pm$ 1	22.5 $\pm$ 2	25 $\pm$ 1	27 $\pm$ 1	28 $\pm$ 1	24 $\pm$ 1
	SNA FT		7 $\pm$ 1	5.5 $\pm$ 4	4.5 $\pm$ 1	25.5 $\pm$ 2	33.5 $\pm$ 2	36.5 $\pm$ 2	40.5 $\pm$ 2
	AAL BD		3.5 $\pm$ 1	6.5 $\pm$ 4	10.5 $\pm$ 9	14 $\pm$ 10	14 $\pm$ 3	27.5 $\pm$ 1	36.5 $\pm$ 2
	SNA BD		67.5 $\pm$ 1	76.5 $\pm$ 1	81 $\pm$ 1	77 $\pm$ 7	82.5 $\pm$ 1	83 $\pm$ 3	75 $\pm$ 1

^aMARS 14-depleted, digested human plasma was separated by lectin chromatography.

^bFT, flow-through (in italics); BD, bound.

^cNumbers represent average values $\pm$ standard deviation from duplicate experiments.

Table 2
Lectin chromatography enabled capture and MS identification of low-abundance human plasma glycoproteins

Selecting Lectin	# of glycosites ^a	Known glycosite ^b	Polypeptide	Gene Name	Gene	Conc.(mg/mL) ^c	Among top 150 plasma proteins ^d	Predicted MW (kD) ^e	CHO % ^f	Reference
SNA	3	Y, N70 Y, N107 Y, N271	Alpha-1-antitrypsin	A1AT_HUMAN	SERPINA1	0.9-2.0	Y	46.7	7	40
AAL, SNA	4, 4	Y, N44 Y, N179 Y, N363 Y, N371	Alpha-1B-glycoprotein	A1BG_HUMAN	A1BG	0.15-0.3	Y	54.2	13	40
AAL, SNA	1, 2	N, N79 S Y, N186	Leucine-rich alpha-2-glycoprotein	A2GL_HUMAN	LRG1	0.02-0.03	Y	38.2	20	40
AAL, SNA	1, 1	Y, N126	Alpha-2-antiplasmin	A2AP_HUMAN	SERPINF2	0.02-0.07	Y	54.5	19	40, 47
AAL, SNA	1, 1	Y, N1424	Alpha-2-macroglobulin	A2MG_HUMAN	A2M	1.3-3.0	Y	163	n/a	40
AAL, SNA	5, 5	Y, N33 Y, N93 Y, N106 Y, N127 Y, N271	Alpha-1-antichymotrypsin	AACT_HUMAN	SERPINA3	0.3-0.6	Y	47.6	27	40
AAL, SNA	2, 2	Y, N33 Y, N402	Afamin	AFAM_HUMAN	AFM	0.05-0.12	Y	69	8	40
SNA	1	Y, N115	Protein AMBP	AMBP_HUMAN	AMBP	0.1-0.2	Y	39	0	40
AAL, SNA	1, 1	Y, N47	Angiotensinogen	ANGT_HUMAN	AGT	0.06	Y	53.1	18.5	40
AAL, SNA	2, 3	Y, N128 Y, N187 Y, N224 S	Antithrombin-III	ANT3_HUMAN	SERPINC1	0.17-0.30	Y	52.6	19	40
AAL, SNA	3, 5	Y, N983 S Y, N1523 S Y, N3411 Y, N3465 Y, N3895	Apolipoprotein B-100	APOB_HUMAN	APOB	0.33-1.29	Y	515	6.5	40, 48
AAL, SNA	4, 4	Y, N162 Y, N183 Y, N193 Y, N253	Beta-2-glycoprotein 1	APOH_HUMAN	APOH	0.15-0.3	Y	38.3	24	40, 49
AAL, SNA	1, 3	Y, N416 S Y, N731 Y, N1259 S	Attractin	ATRN_HUMAN	ATRN	n/a	Y	158	10	40, 50
AAL	1	N, N707	A disintegrin and metalloproteinase with thrombospondin motifs	ATS13_HUMAN	ADAMTS13	0.5-1.0 (µg/mL)	N	153	19	51, 52

Selecting Lectin	# of glycosites ^a	Known glycosite ^b	Polypeptide	Gene Name	Gene	Conc.(mg/mL) ^c	Among top 150 plasma proteins ^d	Predicted MW (kD) ^e	CHO % ^f	Reference
AAL, SNA	4, 4	Y, N119 Y, N150 Y, N349 Y, N402	Biotinidase	BTD_HUMAN	BTD	n/a	Y	61.1	10	40, 53
AAL	2	Y, N105 N, N123	Scavenger receptor cysteine-rich type 1 protein	C163A_HUMAN	CD163	0.73-4.69 (µg/mL)	N	125	19	54, 55
AAL, SNA	1, 1	Y, N146	Complement C1q subcomponent subunit A	C1QA_HUMAN	C1QA	0.045-0.075	Y	26	4	40, 56
SNA	1	Y, N242	Complement C1r subcomponent-like protein	C1RL_HUMAN	C1RL	0.004-0.005 (serum)	N	21.5	70	57
AAL, SNA	1, 1	Y, N174	Complement C1s subcomponent	C1S_HUMAN	C1S	0.08	Y	76.6	1	40, 58
AAL, SNA	1, 3	Y, N221 Y, N506 S Y, N528 S	C4b-binding protein alpha chain	C4BPA_HUMAN	C4BPA	0.2-0.53	Y	67	10	40, 59
SNA	1	Y, N98	C4b-binding protein beta chain	C4BPB_HUMAN	C4BPB	0.03-0.07	Y	28.3	37	40
AAL	1	Y, N442	Cadherin-5	CADH5_HUMAN	CDH5	0.3-4.9(ng/mL)	N	87.5	35	42, 60
AAL, SNA	2, 2	Y, N31 Y, N96	Corticosteroid-binding globulin	CBG_HUMAN	SERPINA6	0.023-0.045	Y	45.1	13	40, 61
AAL, SNA	1, 2	Y, N44S Y, N73	Carboxypeptidase B2	CBPB2_HUMAN	CPB2	0.004-.015	Y	48.4	16	40, 62
AAL	1	Y, N419	CD109 antigen	CD109_HUMAN	CD109	n/a	N	161	5	63
SNA	1	Y, N151	Monocyte differentiation antigen CD14	CD14_HUMAN	CD14	2-6 (µg/mL)	N	40	27	64, 65
AAL	1	Y, N57	CD44 antigen	CD44_HUMAN	CD44	177-396 (ng/mL)	N	81.5	10	66, 67
AAL, SNA	4, 4	Y, N138 Y, N358 Y, N397 Y, N762	Ceruloplasmin	CERU_HUMAN	CP	0.2-0.6	Y	122	10	40, 68
AAL, SNA	3, 3	Y, N122 Y, N285 Y, N378	Complement factor B	CFAB_HUMAN	CFB	0.18-0.46	Y	85.5	15	40
AAL, SNA	6, 6	Y, N529 N, N822 Y, N882 Y, N911 Y, N1029 N, N1095	Complement factor H	CFAH_HUMAN	CFH	0.24-0.76	Y	139	10	40, 69
AAL, SNA	2, 2	Y, N177 Y, N464	Complement factor I	CFAL_HUMAN	CFI	0.03	Y	65.7	24	40, 70
AAL	2	Y, N87 Y, N767	Neural cell adhesion molecule L1-like protein	CHL1_HUMAN	CHL1	n/a	N	135	0	n/a

Selecting Lectin	# of glycosites ^a	Known glycosite ^b	Polypeptide	Gene Name	Gene	Conc.(mg/mL) ^c	Among top 150 plasma proteins ^d	Predicted MW (kD) ^e	CHO % ^f	Reference
AAL, SNA	1, 1	Y, N85	Cholinesterase	CHLE_HUMAN	BCHE	0.003-0.007	Y	68.4	20	40, 71
AAL, SNA	3, 4	Y, N103 Y, N145 S Y, N354 Y, N374	Clusterin	CLUS_HUMAN	CLU	0.025-0.09	Y	52.4	0	40, 72
AAL, SNA	1, 1	Y, N382	Beta-Ala-His dipeptidase	CNDP1_HUMAN	CNDP1	0.004	Y	56.6	27	40, 73
AAL	1	Y, N457	Contactin-1	CNTN1_HUMAN	CNTN1	n/a	N	113	0	n/a
AAL, SNA	1, 2	N, N29 Y, N651 A	Complement C2	CO2_HUMAN	C2	0.011-0.03	Y	83.2	16	40
SNA	3	Y, N226 Y, N1328 Y, N1391	Complement C4-A	CO4A_HUMAN	C4A	0.04-0.16	Y	193	0	40
AAL, SNA	2, 3	Y, N226 Y, N1328 Y, N1391 S	Complement C4-B	CO4B_HUMAN	C4B	0.025-0.09	Y	193	5	40, 74
SNA	1	Y, N911	Complement C5	CO5_HUMAN	C5	0.03-0.07	Y	188	1	40
AAL, SNA	1, 1	Y, N437	Complement component C8 alpha chain	CO8A_HUMAN	C8A	0.028-0.066	Y	65.1	0	40
AAL, SNA	1, 2	Y, N277 S Y, N415	Complement component C9	CO9_HUMAN	C9	0.031-0.075	Y	63.1	20	40
AAL, SNA	3, 4	N, N74 N, N119 S Y, N348 Y, N359	Carboxypeptidase N subunit 2	CPN2_HUMAN	CPN2	n/a	Y	60.6	27	40, 77
AAL	1	Y, N623	Complement receptor type 2	CR2_HUMAN	CR2	30-90 (ng/mL) (serum)	N	113	22	78, 79
AAL	1	Y, N239	Cysteine-rich secretory protein	CRIS3_HUMAN	CRISP3	4.92-7.72 (µg/mL)	N	27.6	5	80
AAL	1	Y, N153	Macrophage colony stimulating factor 1 receptor	CSF1R_HUMAN	CSF1R	n/a	N	108	14-20	81
AAL	2	Y, N182 Y, N462S	Desmoglein-2	DSG2_HUMAN	DSG2	n/a	N	122	26	83
AAL, SNA	1, 1	Y, N444	Extracellular matrix protein 1	ECM1_HUMAN	ECM1	n/a	Y	60.6	29	40, 84
SNA	2	Y, N126 Y, N450	Coagulation factor XI	FA11_HUMAN	F11	0.004-0.006	N	70.1	19.5	40
AAL, SNA	1, 1	Y, N433	Coagulation factor XII	FA12_HUMAN	F12	0.02-0.04	N	67.8	15	40
SNA	2	Y, N1559 Y, N1703	Coagulation factor V	FA5_HUMAN	F5	0.004-0.014	Y	251	24	40
AAL	1	Y, N98	Fibulin-1	FBLN1_HUMAN	FBLN1	0.03-0.05	Y	77.2	4	40
AAL	3	Y, N1317	IgG Fc-binding protein	FCGBP_HUMAN	FCGBP	n/a	N	572	0	n/a

Selecting Lectin	# of glycosites ^a	Known glycosite ^{a,b}	Polypeptide	Gene Name	Gene	Conc.(mg/mL) ^c	Among top 150 plasma proteins ^d	Predicted MW (kD) ^e	CHO % ^f	Reference
		Y, N1830S Y, N5186								
AAL, SNA	1, 1	Y, N189	Ficolin-3	FCN3_HUMAN	FCN3	0.007-0.023	Y	32.9	3	40
AAL, SNA	1, 1	Y, N156	Alpha-2-HS-glycoprotein	FETUA_HUMAN	AHSG	0.36-1.2	Y	39.3	0	40
SNA	1, 1	Y, N37 Y, N136	Fetuin-B	FETUB_HUMAN	FETUB	0.004-0.006	Y	42	31	40, 85, 86
AAL, SNA	2, 2	Y, N126 Y, N194	Complement factor H-related protein	FHR1_HUMAN	CFHR1	20-50 (µg/mL)	N	37.6	0-16	87
AAL, SNA	1, 1	Y, N126	Complement factor H-related protein 2	FHR2_HUMAN	CFHR2	n/a	N	30.6	17	87
SNA	1	Y, N394	Fibrinogen beta chain	FIBB_HUMAN	FGB	0.52-1.42	Y	55.9	0	40
AAL, SNA	1, 1	Y, N78	Fibrinogen gamma chain	FIBG_HUMAN	FGG	0.49-1.42	Y	51.5	0	40
AAL, SNA	2, 4	Y, N430 S Y, N528 Y, N542 Y, N1007 S	Fibronectin	FINC_HUMAN	FN1	0.3	Y	262	10	40
AAL	1	Y, N37	Platelet glycoprotein Ib alpha chain	GP1BA_HUMAN	GP1BA	n/a	N	68.9	49	88
AAL	1	Y, N181	Platelet glycoprotein V	GPV_HUMAN	GP5	78 ng/mL	N	60.9	34	89, 90
AAL, SNA	1, 1	N, N279	Protein HEG homolog 1	HEG1_HUMAN	HEG1	n/a	N	147	0	n/a
AAL, SNA	2, 2	Y, N187 Y, N453	Hemopexin	HEMO_HUMAN	HPX	0.50-1.15	Y	51.6	18	40
AAL, SNA	1, 2	Y, N49 Y, N188 S	Heparin cofactor 2	HEP2_HUMAN	SERPIND1	0.09	Y	57	17	40,
SNA	1	Y, N468	Hepatocyte growth factor activator	HGFA_HUMAN	HGFAC	3-6 (ng/mL)	N	70.6	26	43, 92
SNA	1	Y, N296	Hepatocyte growth factor-like protein	HGFL_HUMAN	MST1	0.02 (µg/mL)	N	80.3	17	93, 94
SNA	3	Y, N207 Y, N211 Y, N241	Haptoglobin	HPT_HUMAN	HP	0.2-1.5	Y	45.2	16	40, 95
AAL, SNA	1, 2	Y, N63 S Y, N125	Histidine-rich glycoprotein	HRG_HUMAN	HRG	0.05-0.15	Y	59.5	10	40
AAL, SNA	1, 1	Y, N862	Hypoxia up-regulated protein 1	HYOU1_HUMAN	HYOU1	n/a	N	111	26	n/a
AAL, SNA	3, 1	N, N116 Y, N136 A Y, N199 A	Insulin-like growth factor-binding protein 3	IBP3_HUMAN	GFBP3	0.003-0.007	Y	31.6	25	40
AAL, SNA	3, 3	Y, N25 Y, N253 Y, N352	Plasma protease C1 inhibitor	IC1_HUMAN	SERPING1	0.19-0.37	Y	55.1	41-50	40

Selecting Lectin	# of glycosites ^a	Known glycosite ^b	Polypeptide	Gene Name	Gene	Conc.(mg/mL) ^c	Among top 150 plasma proteins ^d	Predicted MW (kD) ^e	CHO % ^f	Reference
AAL	1	Y, N267	Intercellular adhesion molecule 1	ICAM1_HUMAN	ICAM1	100-254 (ng/mL)	N	57.8	28	96, 97
AAL	2	Y, N82 Y, N176	Intercellular adhesion molecule 2	ICAM2_HUMAN	ICAM2	n/a	N	30.6	49	98
AAL	1	Y, N227	Ig gamma-3 chain C region	IGHG3_HUMAN	IGHG3	n/a	N	41.2	21	99
AAL	1	N, N209	Ig mu chain C region	IGHM_HUMAN	GHM	0.3-1.7	Y	49.3	30	40
SNA	1	Y, N262	Plasma serine protease inhibitor	IPSP_HUMAN	SERPINA5	0.005	Y	45.7	14	40
AAL, SNA	2, 2	Y, N285 Y, N588	Inter-alpha-trypsin inhibitor heavy chain H1	ITIH1_HUMAN	ITIH1	0.12-0.24	Y	101	0	40
SNA	1	Y, N118	Inter-alpha-trypsin inhibitor heavy chain H2	ITIH2_HUMAN	ITIH2	0.14-0.28	Y	106	0	40
AAL, SNA	4, 4	Y, N81 Y, N207 Y, N517 Y, N577	Inter-alpha-trypsin inhibitor heavy chain H4	ITIH4_HUMAN	ITIH4	0.12-0.24	Y	103	40	40
AAL, SNA	4, 4	Y, N33 Y, N108 Y, N157 Y, N238	Kallistatin	KAIN_HUMAN	SERPINA4	0.014	Y	48.5	25	40
AAL, SNA	2, 4	Y, N308 S Y, N396 S Y, N453 Y, N494	Plasma kallikrein	KLKB1_HUMAN	KLKB1	0.035-0.045	Y	71.3	20	40
AAL, SNA	4, 4	Y, N48 Y, N169 Y, N205 Y, N294	Kininogen-I	KNG1_HUMAN	KNG1	0.18	Y	33	25	40
AAL, SNA	1, 1	Y, N257	Lysosome-associated membrane glycoprotein 2	LAMP2_HUMAN	LAMP2	0.77-1.82 (µg/mL)	N	44.9	30	100, 101
AAL, SNA	2, 3	Y, N398 S Y, N551 Y, N580	Galectin-3-binding protein	LG3BP_HUMAN	LGALS3BP	5.1-5.7 (µg/mL) serum	N	65.3	29	102, 103
AAL, SNA	3, 1	Y, N88 Y, N127 A Y, N160 A	Lumican	LUM_HUMAN	LUM	n/a	Y	38.4	43-61	40, 104
AAL, SNA	2, 1	Y, N60 A Y, N104	L-selectin	LYAM1_HUMAN	SELL	0.67 (µg/mL)	N	42.1	60	105, 106
AAL	1	Y, N178	Mannan-binding lectin serine protease	MASP1_HUMAN	MASP1	1.5-12.8 (µg/mL) (serum)	N	79.2	16	107
AAL	3	Y, N114 Y, N120 Y, N136	Multimerin-1	MMRN1_HUMAN	MMRN1	n/a	N	138	11	n/a

Selecting Lectin	# of glycosites ^a	Known glycosite ^b	Polypeptide	Gene Name	Gene	Conc.(mg/mL) ^c	Among top 150 plasma proteins ^d	Predicted MW (kD) ^e	CHO % ^f	Reference
AAL	1	Y, N845	Multimerin-2	MMRN2_HUMAN	MMRN2	n/a	N	104	0	n/a
AAL	1	Y, N38	Neuroblastoma suppressor of tumorigenicity 1	NBL1_HUMAN	NBL1	n/a	N	19.2	0	n/a
AAL	1	Y, N85	Neutrophil gelatinase-associated lipocalin	NGAL_HUMAN	LCN2	2.5 (ng/mL)	N	22.5	10	41, 108
AAL	1	Y, N150	Neuropilin-1	NRP1_HUMAN	NRP1	n/a	N	103	21	109
AAL, SNA	1, 1	Y, N285	Pigment epithelium derived factor	PEDF_HUMAN	SERPINF1	0.005	Y	46.3	4	40
AAL, SNA	1, 3	Y, N77 Y, N367 S Y, N485 S	N-acetylmuramoyl-L-alanine amidase	PGRP2_HUMAN	PGLYRP2	n/a	N	62.2	19	110
AAL, SNA	1, 2	Y, N307 S Y, N568	Phosphatidylinositol 1-glycan-specific phospholipase D	PHLD_HUMAN	GPLD1	0.02-0.09	Y	92.3	28	40, 111
AAL	3	N, N114 N, N403 N, N409	Peptidase inhibitor 16	PII6_HUMAN	PII6	384-2072 (ng/mL)serum	N	49.4	28	112
AAL	3	Y, N64 Y, N143 Y, N398	Phospholipid transfer protein	PLTP_HUMAN	PLTP	0.005-0.020	N	54.7	27	113, 114
AAL, SNA	1, 1	Y, N253	Serum paraoxonase/aryle sterase 1	PON1_HUMAN	PON1	0.017-0.52	Y	39.7	16	40
AAL	1	Y, N428	Properdin	PROP_HUMAN	CFP	0.025	Y	51.2	7	40
AAL	2	Y, N51 Y, N78	Prostaglandin-H2 D-isomerase	PTGDS_HUMAN	PTGDS	0.1-0.9 µg/mL (serum)	N	21	16	115, 116
AAL	1	Y, N160	Plexin domain-containing protein 2	PXDC2_HUMAN	PLXDC2	n/a	N	59.5	0	n/a
SNA	1	Y, N94	Serum amyloid A-4 protein	SAA4_HUMAN	SAA4	0.05	Y	14.8	0	40
AAL	1	Y, N215	Proactivator polypeptide	SAP_HUMAN	PSAP	n/a	N	58.1	32-26	117
AAL, SNA	3, 3	Y, N83 Y, N119 Y, N174	Selenoprotein P	SEPP1_HUMAN	SEPP1	0.003	Y	43.1	32	40, 118, 119
SNA	1	Y, N380	Sex hormone-binding globulin	SHBG_HUMAN	SHBG	0.001-0.012	Y	43.7	20	40
AAL	1	Y, N184	Tenascin	TENA_HUMAN	TNC	0.47-1.47 (µg/mL)	N	240	10-20	120, 121
AAL	1	N, N3900	Tenascin-X	TENX_HUMAN	TNXB	269-657 (ng/mL) serum	N	464	n/a	122
AAL, SNA	1, 2	Y, N36 Y, N165 S	Thyroxine-binding globulin	THBG_HUMAN	SERPINA7	0.011-0.027	Y	46.3	15	40
AAL, SNA	2, 2	Y, N143 Y, N416	Prothrombin	THRB_HUMAN	F2	0.1	Y	70	15	40
AAL	1	Y, N53	Metalloproteinase inhibitor 1	TIMP1_HUMAN	TIMP1	48-119 (ng/mL)	N	23.1	7-23	123, 124

Selecting Lectin	# of glycosites ^a	Known glycosite ^b	Polypeptide	Gene Name	Gene	Conc.(mg/mL) ^c	Among top 150 plasma proteins ^d	Predicted MW (kD) ^e	CHO % ^f	Reference
SNA	2	Y, N432 Y, N630	Serotransferrin	TRFE_HUMAN	TF	2.0-3.6	Y	77	3	40
AAL	1	Y, N117	Vasorin	VASN_HUMAN	VASN	n/a	N	71.7	n/a	n/a
AAL	1	Y, N527	Vascular endothelial growth factor receptor 3	VGFR3_HUMAN	FLT4	0.2-33.2 (ng/mL) serum	N	145	25	125, 126
SNA	2	Y, N283 Y, N353	Pantetheinase	VNN1_HUMAN	VNN1	n/a	N	57	n/a	n/a
AAL, SNA	2, 2	Y, N86 Y, N242	Vitronectin	VTNC_HUMAN	VTN	0.06-0.17	Y	54.3	30	40
AAL	2	Y, N666 Y, N1515	von Willebrand factor	VWF_HUMAN	VWF	0.01-0.02	Y	309	0	40
AAL, SNA	1, 1	Y, N106	Zinc-alpha-2-glycoprotein	ZA2G_HUMAN	AZGP1	0.03-0.06	Y	33.8	22	40

“n/a”, information not available.

^aAs identified by AAL or SNA (first and second number) in our datasets.

^cLiterature values obtained from the listed references.

^bAs annotated in the Unipep or Swissprot databases. Previously unknown glycosites are in boldface. Where a glycoprotein was identified by both lectins, but a glycosite by only one lectin, the selecting lectin is noted by A or S.

^dAs denoted in Hortin *et al.*, 2008 (40).

^eBased on the amino acid sequence reported in the Swissprot database. When multiple isoforms exist, MW reflects sequence of the largest isoform.

^fEstimated or known contribution of glycosylation to observed MW.

Table 3
Lectin chromatography at the peptide level allows for the discrimination of carbohydrate motifs on individual glycosites

Mode of capture ^a	Glycosite determination?	Sensitive to carbohydrate structure?
Chemical: protein level	Possible	No
Chemical: peptide level	Yes	No
Lectin: protein level	No	Yes
Lectin: peptide level	Yes	Yes

^aChemical capture - Hydrazide or boronic acid chemistry.

Lectin capture - Single lectin or multiple lectins combined in one column (MLAC).