Zhao J, Patwa TH, Qiu W, Shedden K, Hinderer R, Misek DE, Anderson MA, Simeone DM, Lubman DMGlycoprotein microarrays with multi-lectin detection: unique lectin binding patterns as...

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · MAY 2007

Impact Factor: 4.25 · DOI: 10.1021/pr070062p · Source: PubMed

9 AUTHORS, INCLUDING:

David E Misek
University of Michigan
108 PUBLICATIONS **10,467** CITATIONS
SEE PROFILE

David M Lubman
University of Michigan
289 PUBLICATIONS **7,809** CITATIONS
SEE PROFILE

research articles Journal of proteome research

Glycoprotein Microarrays with Multi-Lectin Detection: Unique Lectin Binding Patterns as a Tool for Classifying Normal, Chronic Pancreatitis and Pancreatic Cancer Sera

Jia Zhao,[†,‡] Tasneem H. Patwa,[†,‡] Weilian Qiu,[†] Kerby Shedden,[§] Robert Hinderer,[||] David E. Misek,[||] Michelle A. Anderson,[⊥] Diane M. Simeone,[||] and David M. Lubman*[,†,||,#]

Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109, Department of Surgery, University of Michigan Medical Center, Ann Arbor, Michigan 48109, Department of Internal Medicine, University of Michigan Medical Center, Ann Arbor, Michigan 48109, and Comprehensive Cancer Center, University of Michigan Medical Center Ann Arbor, Michigan 48109

Received February 1, 2007

Pancreatic cancer is the fourth leading cause of cancer-related death in the United States, with a 5-year survival rate of less than 4%. Effective early detection and screening are currently not available, and tumors are typically diagnosed at a late stage, frequently after metastasis. Existing clinical markers of pancreatic cancer lack specificity, as they are also found in inflammatory diseases of the pancreas and biliary tract. In the work described here, naturally occurring glycoproteins were enriched by using lectin affinity chromatography and then further resolved by nonporous reversed-phase chromatography. Glycoprotein microarrays were then printed and probed with a variety of lectins to screen glycosylation patterns in sera from normal, chronic pancreatitis, and pancreatic cancer patients. Ten normal, 8 chronic pancreatitis, and 6 pancreatic cancer sera were investigated. Data from the glycoprotein microarrays were analyzed using bioinformatics approaches including principal component analysis (PCA) and hierarchical clustering (HC). Both normal and chronic pancreatitis sera were found to cluster close together, although in two distinct groups, whereas pancreatic cancer sera were significantly different from the other two groups. Both sialylation and fucosylation increased as a function of cancer on several proteins including Hemopexin, Kininogen-1, Antithrombin-III, and Haptoglobin-related protein, whereas decreased sialylation was detected on plasma protease C1 inhibitor. Target alterations on glycosylations were verified by lectin blotting experiments and peptide mapping experiments using μLC−ESI−TOF. These altered glycan structures may have utility for the differential diagnosis of pancreatic cancer and chronic pancreatitis and identify critical differences between biological samples from patients with different clinical conditions.

Keywords: AUTHOR PLEASE PROVIDE KEYWORDS

1. Introduction

Pancreatic cancer, currently the fourth most frequent cause of cancer-related death in the USA, is generally incurable by available treatment modalities, with a 5-year survival rate of <4%.[1] The poor prognosis results from the biologically aggressive nature of this disease, with rapid metastasis, combined with the late clinical presentation of the malignancy. Current existing markers for pancreatic cancer are inadequate for early diagnosis, distinguishing between pancreatic cancer and chronic pancreatitis, and for the efficacious targeting of therapeutics.[2,3] CA19−9 has been tested for its utility as an early detection marker in pancreatic cancer.[2−4] However, the sensitivity and specificity of this marker is not high, and serum levels are also significantly increased in inflammatory diseases of the pancreas and biliary tract. Therefore, CA19−9 is not useful for early diagnosis, mass screening, or for distinguishing between pancreatic cancer and chronic pancreatitis.

There is currently great interest in developing protein-based serum markers for the early detection of cancer. Identification of useful serum markers is essential not only for the early detection of pancreatic cancer but also for the differential diagnosis of this malignancy from chronic pancreatitis. The glyco-proteome is one of the major sub-proteomes of human

* To whom correspondence should be addressed. Dr. David M. Lubman, The University of Michigan Medical Center, Department of Surgery, MSRB1, Rm A510B, 1150 West Medical Center Drive, Ann Arbor, Michigan 48109-0656, U.S.A. E-mail: dmlubman@umich.edu. Fax: +1-734-615-2088.
[†] Department of Chemistry.
[‡] These two authors equally contributed to this work.
[§] Department of Statistics.
[||] Department of Surgery.
[⊥] Department of Internal Medicine.
[#] Comprehensive Cancer Center.

10.1021/pr070062p CCC: $37.00
Published on Web 04/12/2007

serum. Serum glycoproteins may be secreted (or released) into the blood.[5] Both N-linked and O-linked glycan variants of glycoproteins on the cell surface and in plasma have been shown to correlate with the progression of cancer and other diseases.[6−10] Changes in glycosylation patterns have been associated with prostate cancer,[11,12] colorectal cancer,[13,14] and breast cancer.[15] For example, the glycosylation of prostate-specific antigen (PSA) secreted by the prostate tumor cell line LNCaP differs significantly from that found in seminal plasma (normal).[11] As glycan differences can distinguish PSA from normal and tumor origins, these differences may have utility for early diagnosis of prostate cancer. Glycosylation changes in a tumor-secreted protein may reflect fundamental changes in enzyme levels (or enzyme activities) involved in the glycosylation pathway. The ability to efficiently profile protein glycosylation variation may ultimately lead to the identification of disease-associated glycan alterations and new diagnostic markers in pancreatic cancer as well as in other types of cancer.

Protein microarrays are becoming increasingly popular in high throughput proteomic analysis due to their ability to screen large numbers of arrayed samples for a property of interest.[16−18] Current research in this area has focused on a variety of applications ranging from functional analysis to a more diagnostic-type approach. Functional approaches have focused on studying interactions of proteins with a variety of other molecules such as other proteins, lipids, drugs, and DNA.[19,20] Diagnostic applications involve immobilization of antibodies on high density arrays, which can be probed with biological fluids or cell lysates to monitor antigen−antibody interactions.[21] Further, protein microarrays arrayed with naturally produced proteins have been developed to assist in finding novel disease-associated proteins[22,23] using multidimensional liquid-based separation of a proteome, followed by the arraying of all proteins found in the individual fractions. The resulting microarrays can subsequently be probed with a variety of detection agents, including lectins for glycoprotein detection.

Lectins specifically and reversibly bind glycans with different structural moieties and, thus, have utility in screening glycosylation differences between various samples. Lectin glycoarrays can be used for the rapid profiling of glycan expression patterns of various glycoproteins.[24] We have utilized glycoarrays to discern differences in the glycosylation structural patterns of serum glycoproteins specific for pancreatic cancer and chronic pancreatitis. To this end, following immunodepletion to remove high abundance proteins from serum (and to facilitate our ability to detect low abundance glycoproteins), we enriched the remaining N-linked glycoproteins using a general multi-lectin column. These enriched glycoproteins were then separated using nonporous silica reverse phase high performance liquid chromatography (NPS-RP-HPLC). The resolved glycoproteins were then arrayed on nitrocellulose-coated slides and probed with a variety of lectins to screen the glycosylation structures of the serum glycoproteins. The glycoprotein-lectin interaction was assessed using a biotin-streptavidin system that had low femtomole limits of detection.

All data was subjected to bioinformatics analysis to handle and display efficiently the large datasets generated in this manuscript. To compare the overall pattern of protein expression in the samples, each sample was normalized and aligned with the corresponding UV peak area from the RP-HPLC chromatograms. A correlation matrix was obtained by calculating the Pearson correlations among the samples. These correlation matrices were then visualized, using either principal components analysis (PCA) or hierarchical clustering (HC) techniques, allowing multivariate relationships to be explored to highlight relationships present in the sample sets. Quantitative measurements were also facilitated, because normalization based on UV peak areas eliminated any concentration-dependent variability that existed in the fractionated glycoproteins. We calculated differential glycan expression by interrogating Z-value information. The individual glycoproteins with altered glycan structures were then identified by mass spectrometry. These glycan structural alterations may have utility for the early detection of pancreatic cancer and for the differential diagnosis of pancreatic cancer and chronic pancreatitis.

2. Method

2.1. Serum Samples. Serum was obtained at the time of diagnosis following informed consent using IRB-approved guidelines. Sera were obtained from 6 patients with a confirmed diagnosis of pancreatic adenocarcinoma in the Multidisciplinary Pancreatic Tumor Clinic at The University of Michigan Hospital. These sera were randomly selected from a clinic population that sees, on average, at the time of initial diagnosis, 15% of pancreatic adenocarcinoma patients presenting with early stage (i.e., stage 1/2) disease and 85% presenting with advanced stage (i.e., stage 3/4). Inclusion criteria for the study included patients with a confirmed diagnosis of pancreatic cancer, the ability to provide written, informed consent, and the ability to provide 40 mL of blood. Exclusion criteria included inability to provide informed consent, patient's actively undergoing chemotherapy or radiation therapy for pancreatic cancer, and patients with other malignancies diagnosed or treated within the last 5 years. Sera were also obtained from 8 patients with chronic pancreatitis who were seen in the Gastroenterology Clinic at University of Michigan Medical Center and from 10 control healthy individuals collected at University of Michigan under the auspices of the Early Detection Research Network (EDRN). The mean age of the tumor group was 65.4 years (range 54−74 years) and from the chronic pancreatitis group was 54 years (range 45−65). The sera from the normal subject group was age and sex-matched to the tumor group. All of the chronic pancreatitis sera were collected in an elective setting in the clinic in the absence of an acute flare. All sera were processed using identical procedures. The samples were permitted to sit at room temperature for a minimum of 30 min (and a maximum of 60 min) to allow the clot to form in the red-top tubes,and then centrifuged at 1300× g at 4 °C for 20 min. The serum was removed, transferred to a polypropylene, capped tube in 1 mL aliquots, and frozen. The frozen samples were stored at −70 °C until assayed. All serum samples were labeled with a unique identifier to protect the confidentiality of the patient. The handling of all serum samples was similar in that none of the samples were thawed more than twice before analysis to minimize protein degradation and precipitation.

2.2. Immunodepletion of High Abundance Proteins. Each serum sample (125 μL of each) was depleted using the ProteomeLab IgY-12 proteome partitioning kit (Beckman Coulter, Fullerton, CA), following centrifugation using a 0.45 μm spin filter for 1 min at 9200× g, according to manufacturer's protocols. This column facilitates removal of albumin, IgG, α1-antitrypsin, IgA, IgM, transferrin, haptoglobin, α1-acid glycoprotein, α2-macroglobin, apolipoprotein A-I, apolipoprotein A-II, and fibrinogen in a single step. The final volume of each serum sample following immunodepletion was concentrated

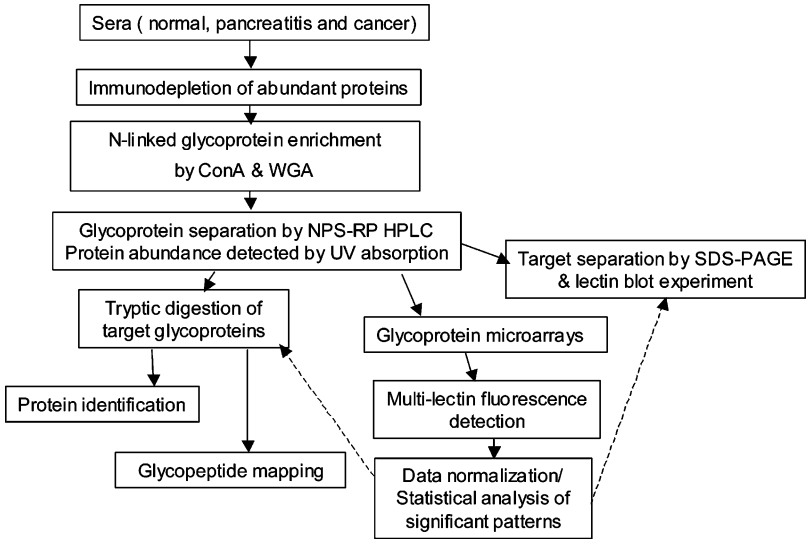


Figure 1. Strategy used to screen the glycosylation patterns and characterize the target glycoproteins using samples of normal, chronic pancreatitis, and pancreatic cancer sera.

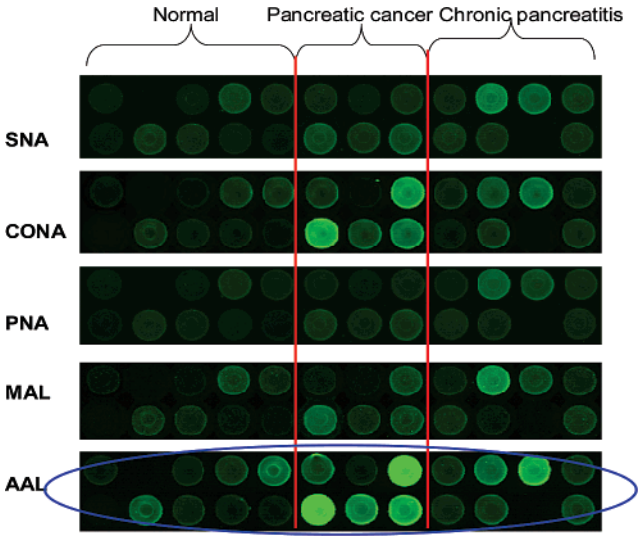


Figure 2. Sections of glycoprotein microarray showing comparison of one fraction from NPS-RP-HPLC across all 24 samples. Each panel is a section of identical arrays probed with lectin indicated on the left side of the panel. It was observed that this fraction contained proteins that were predominantly mannosylated and fucosylated. It was also observed that the level of glycosylation (based on raw microarray data) was higher in cancer samples compared to the controls.

to 500 μL using 15 mL 10 kDa Amicon filters (Millipore, Billerica, MA). Protein assays were carried out in a 250 μL transparent 96 well plate (Fisher, Barrington, IL) according to the Bradford assay.

2.3. Lectin Affinity Glycoprotein Extraction. Agarose-bound wheat germ agglutinin (WGA) and agarose-bound Concanavalin A (ConA) were purchased from Vector Laboratories (Burlingame, CA). Agarose-bound WGA (350 μL) and 250 μL of agarose-bound ConA were packed into disposable screw end-cap spin column with filters at both ends. The binding and elution process has been described elsewhere.[7] The binding buffer contained 20 mM Tris, 1 mM $MnCl_2$, 1 mM $CaCl_2$, and 0.15 M NaCl, pH 7.4. The immunodepleted serum proteins were resuspended in binding buffer and then passed through the lectin affinity column. The captured serum glycoproteins were released with 250 μL of elution buffer (0.3 M *N*-acetyl-glucosamine and 0.3 M methyl-α-D-mannopyroside in 20 mM Tris and 0.5 M NaCl, pH 7.0). This step was repeated twice and the eluted fractions were pooled.

2.4. RP-HPLC Separation of Lectin-Bound Glycoproteins. The lectin-enriched glycoprotein fraction was concentrated to ~100 μL with a 10k MW centrifugal filter (Millipore) and re-diluted with deionized water. Approximately 30 μg of protein sample was loaded in 800 μL of water onto a nonporous silica reverse phase high-performance liquid chromatography (NPS-RP-HPLC) column (ODSII (4.6 × 33 mm^2) column (Eprogen, Inc., Darien, IL) packed with 1.5 μm nonporous silica) for separation. The reverse-phase separation was performed at 0.5 mL/min and monitored at 214 nm using a Beckman 166 Model UV detector (Beckman-Coulter). Proteins eluting from the column were collected by an automated fraction collector (Model SC 100; Beckman-Coulter), controlled by an in-house designed DOS-based software program. The reversed phase column was heated to 60 °C by a column heater (Jones Chromatography, Model 7971). Both mobile phase A (water) and B (ACN) contained 0.1% v/v TFA. The gradient profile used was as follows: 5−15% B in 1 min, 15−25% B in 2 min, 25−30% B in 3 min, 30−41% B in 15 min, 41−47% B in 4 min, 47−67% B in 5 min, and 67−100% B in 2 min.

2.5. Glycoprotein Microarrays. Purified and separated glycoproteins were printed on nitrocellulose slides (Whatman, Keene, NH) using a noncontact printer, Nanoplotter 2.0 (GeSIM, Germany). Prior to printing, the proteins were dried down in a 96 well plate and resuspended in 15 μL of printing buffer with stirring overnight at 4 °C. The printing buffer contained 65 mM Tris-HCl, 1% SDS, 5% dithiothreitol (DTT), and 1% glycerol. Each spotting event resulted in approximately 500 pL of sample being deposited by a piezoelectric mechanism. The event was programmed to occur 5 times per spot to ensure that approximately 2.5 nL were being spotted per sample. The resulting spots were approximately 450 μm in diameter, with the spacing between spots being maintained at 600 μm. After printing, the slides were allowed to dry for 24 h. Blocking was

achieved by incubation with 1% Bovine serum albumin (BSA) and 0.1% Tween-20 in 1× phosphate buffered saline (PBS) overnight. Blocked slides were probed with biotinylated lectin in a solution of PBS−T (0.1% Tween 20 in 1X PBS). The lectins used in the study were biotinylated Peanut Agglutinin (PNA), *Sambucus Nigra* bark lectin (SNA), *Aleuria Aurentia* (AAL), Concanavalin A (ConA), and *Maackia Amurensis* lectin II (MAL), all purchased from Vector Laboratories (Burlingame, CA). The working concentration of all lectins was 5 μg/mL, with the exception of SNA (10 μg/mL, as per manufacturer's protocols). After primary incubation, all slides were washed with PBS−T 5 times for 5 min each. Detection was achieved using a streptavidin-AlexaFluor555 conjugate (Invitrogen, Carlsbad, CA) at 1 μg/mL in PBS containing 0.5% BSA and 0.1% Tween-20. The slides were washed 5 times for 5 min each in PBS−T and then completely dried by centrifugation. The dried slides were scanned using an Axon 4000A scanner in the green channel. GenePix Pro 6.0 software (Molecular Devices, Sunnyvale, CA) was used for data acquisition and analysis.

2.6. Data Analysis and Clustering. All the microarray spot intensities were normalized with corresponding UV peak area. For data visualization, average linkage hierarchical clustering (HC) and principal component analysis (PCA) were used to provide graphical representations of the relationships among the samples. In these unsupervised approaches, 48 serum samples (10 normal, 8 chronic pancreatitis, and 6 pancreatic cancer, all processed in duplicate) and the replicate averages of the 24 distinct biological specimens were placed either in a hierarchical relationship (HC) or as points in a 2-dimensional scatterplot (PCA) based on similarities in normalized glycoform abundances. For differential abundance analysis, Z-statistics and Wilcoxon rank sum statistics for each protein detected by each lectin were calculated. Comparisons were made of cancer versus chronic pancreatitis and normal combined, and of chronic pancreatitis and cancer combined versus normal.

2.7. Protein Digestion by Trypsin. Fractions obtained from NPS-RP-HPLC were concentrated down to approximately 20 μL using a SpeedVac concentrator (Thermo, Milford, MA) operating at 45 °C. Twenty microliters of 100 mM ammonium bicarbonate (Sigma) was then mixed with each concentrated sample to obtain pH 7.8. TPCK modified sequencing grade porcine trypsin (0.5 μL) (Promega, Madison, WI) was added and briefly vortexed prior to a 12−16 h incubation at 37 °C on an agitator.

2.8. Mass Spectrometry. Protein Identification by LC−MS/MS. Digested peptide mixtures from NPS-RP-HPLC collection were separated using a reverse phase column attached to a Paradigm HPLC pump (Michrom Bio Resources Inc, Auburn, CA). For nanoLC−ESI−MS/MS experiments, a nanotrap platform (Michrom) was set up prior to the electrospray source. It included a peptide nanotrap (0.2 × 50 mm^2, Michrom) and a separation column (0.1 mm × 150 mm, C18, Michrom). The peptide sample was injected and first desalted on the trap column with 5% solvent B (0.3% formic acid in 98% ACN) at 50 μL/min for 5 min. The peptides were then eluted using a 45 min gradient from 5 to 95% B at a flow rate of 0.25 μL/min where solvent A was 0.3% formic acid in HPLC grade water.

A Finnigan LTQ mass spectrometer (Thermo) was used to acquire spectra. A 75 μm metal spray tip (Michrom) was used, and spray voltage was set at 2.5 kV. The instrument was operated in data-dependent mode with dynamic exclusion enabled. The MS/MS spectra on the five most abundant peptide ions in full MS scan were obtained. All MS/MS spectra were searched against the human protein database from SwissProt using SEQUEST algorithm incorporated in Bioworks software (Thermo). Oxidized methionine and *N*-acetylation were used as variable modifications during the database search. Trypsin was used as a specific protease with two missed cleavages allowed. Positive protein identification was accepted for a peptide with X_{corr} of greater than or equal to 3.0 for triply, 2.5 for doubly, and 1.9 for singly charged ions. ΔCn cutoff was set as 0.1. Positive protein identification was validated by TransProteomics pipeline. This software includes both the PeptideProphet and ProteinProphet programs that were developed by Keller et al. (http://peptideprophet.sourceforge.net/).[25] All the reported proteins have an identification probability higher than 95%.

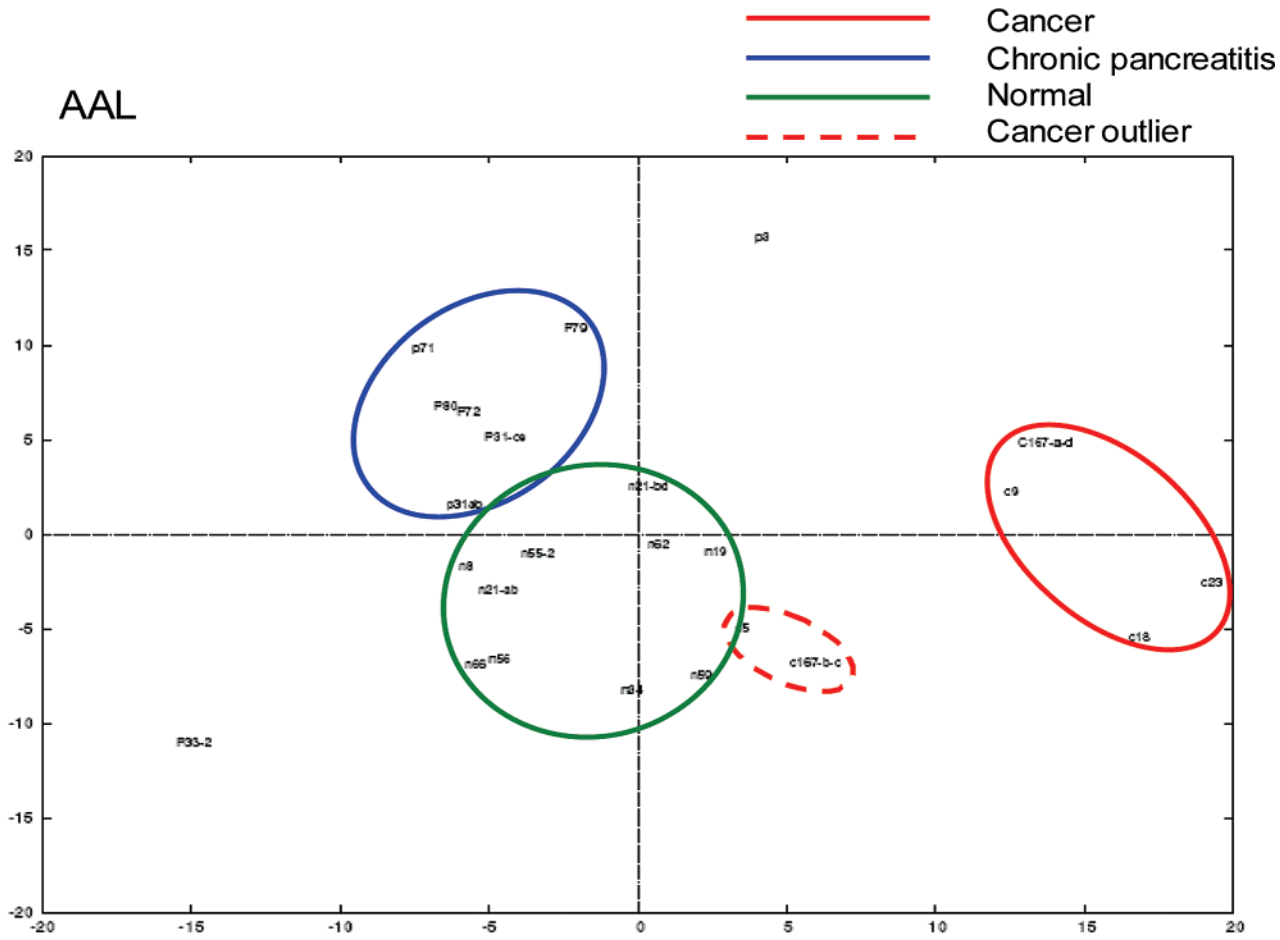
Glycopeptide Mapping. Digested peptide mixtures from target glycoproteins were separated by a capillary RP column (C18, 0.2 × 150 mm^2) (Michrom, Auburn, CA) on a capillary pump (Ultra-Plus II MD, Micro-Tech Scientific, Vista, CA). The capillary column was directly mounted to a micro-injector with a 500 nL internal sample loop (Valco Instruments, Houston, TX). The flow from the solvent delivery pump was split precolumn to generate a flow rate of approximately 4 μL/min. The gradient started at 5% ACN, was ramped to 60% ACN in 25 min, and finally ramped to 90% in another 5 min. Both solvent A (water) and B (ACN) contain 0.3% formic acid. The resolved peptides were detected by an ESI-TOF spectrometer (LCT premier, Micromass/Waters, Milford, MA). The capillary voltage for electrospray was set at 3200 V, and for the sample cone at 45 V. Desolvation was accelerated by maintaining the desolvation temperature at 200 °C and source temperature at 100 °C. The desolvation gas flow was 250 L/h. The data was acquired in "V" mode and the TOF was externally calibrated using a Sodium Iodide and Cesium Iodide mixture. The instrument was controlled by MassLynx 4.0 software. The experimental masses were matched with theoretical glycopeptide masses of target glycoproteins using GlyMod tool (http://www.expasy.ch/tools/glycomod/).

2.9. SDS-PAGE and Lectin Blotting of Separated Fractions. The fractions collected from RP-HPLC were further separated by 1-D SDS-PAGE, run in a Mini-Protean cell (Bio-Rad, Hercules, CA) at 80 V. The resolved proteins were transferred onto a PVDF membrane (Bio-Rad). The PVDF membrane was rehydrated in methanol, rinsed, and then blocked in PBS, containing 1% BSA (Roche, Indianapolis, IN) and 0.1% Tween20. The membrane was then washed in PBS−T 3 times for 1 min and then incubated with biotinylated *Aleuria aurentia* lectin (5 μg/mL in PBS−T containing 1% BSA) for 1 h at room temp. Following incubation with the lectin, the membrane was washed 3 times for 2 min each in PBS−T. Detection was with a 200 ng/mL streptavidin-HRP in PBS−T containing 1% BSA. The membrane was washed in PBS−T 5 times for 5 min each followed by one wash with PBS for 5 min. Chemiluminescence was accomplished using an ECL analysis system (Amersham, Piscataway, NJ) and detected on XAR-5 X-ray film (Kodak). The film was digitized using a high-resolution digital camera.

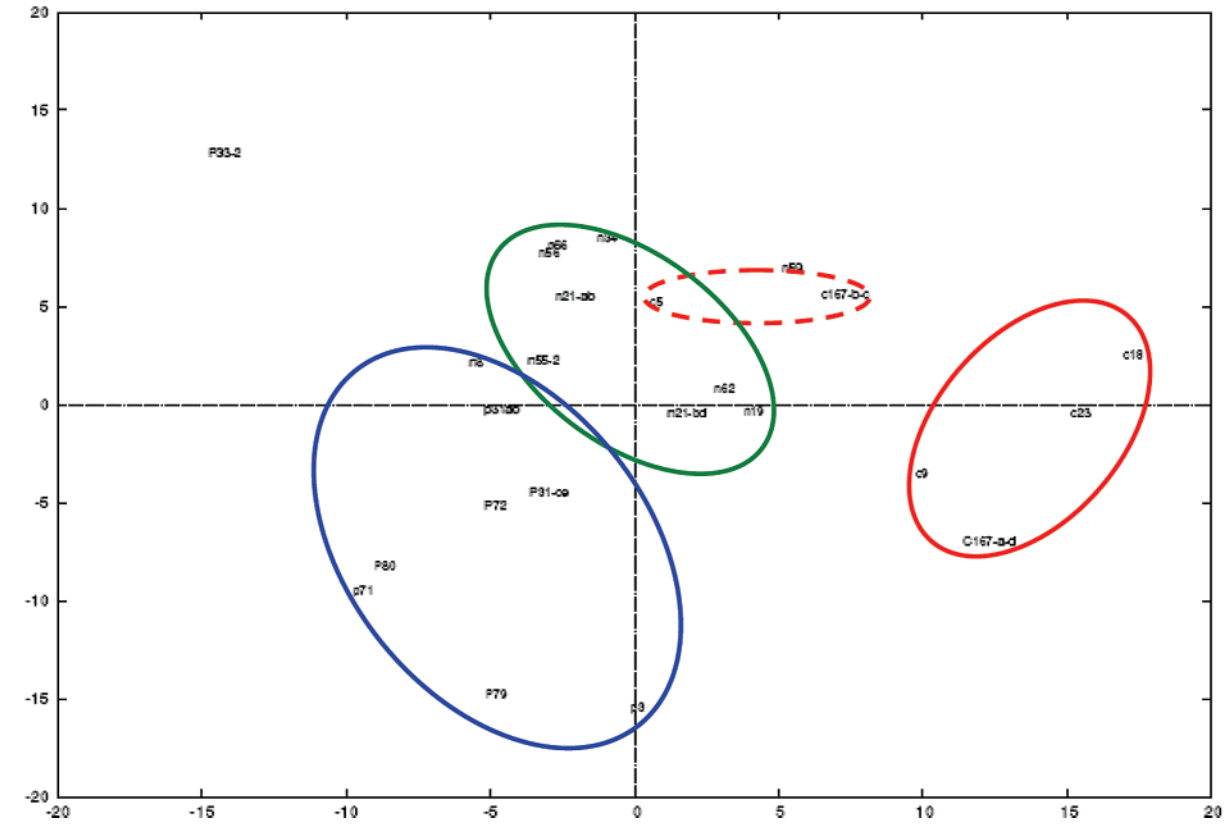
3. Results and Discussion

3.1. Glycoprotein Enrichment, Depletion, and Separation. The analytical work flow is depicted diagrammatically in Figure 1. Analysis of 10 normal, 8 chronic pancreatitis, and 6 pancreatic cancer serum samples was performed using glycoprotein extraction followed by liquid separation and microarray spotting of the separated glycoprotein fractions. Each serum sample

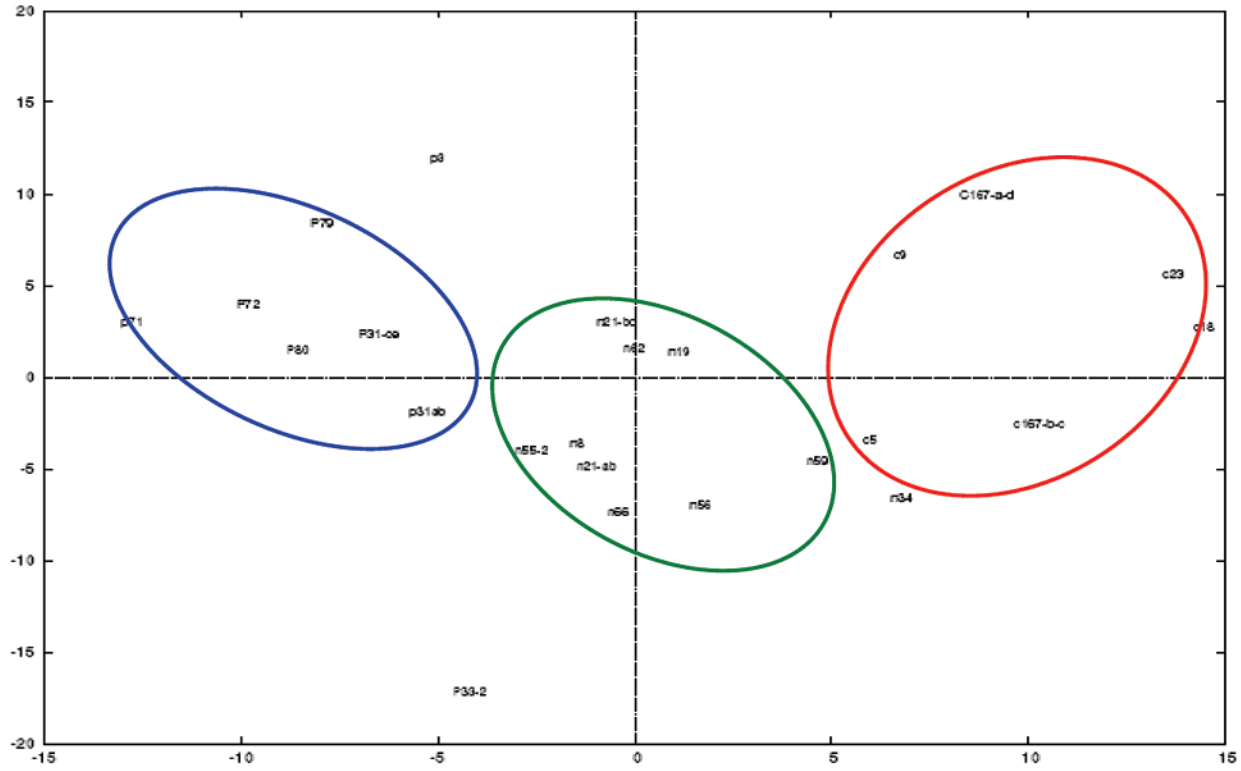
(a)


Cancer
Chronic pancreatitis
Normal
Cancer outlier
AAL
p3
P79
p71
P80
P72
P31-ce
p31ab
n21-bd
n62
n19
n55-2
n8
n21-ab
c5
c167-b-c
C167-a-d
c9
c23
c18
n66
n56
n34
n52
P93-2

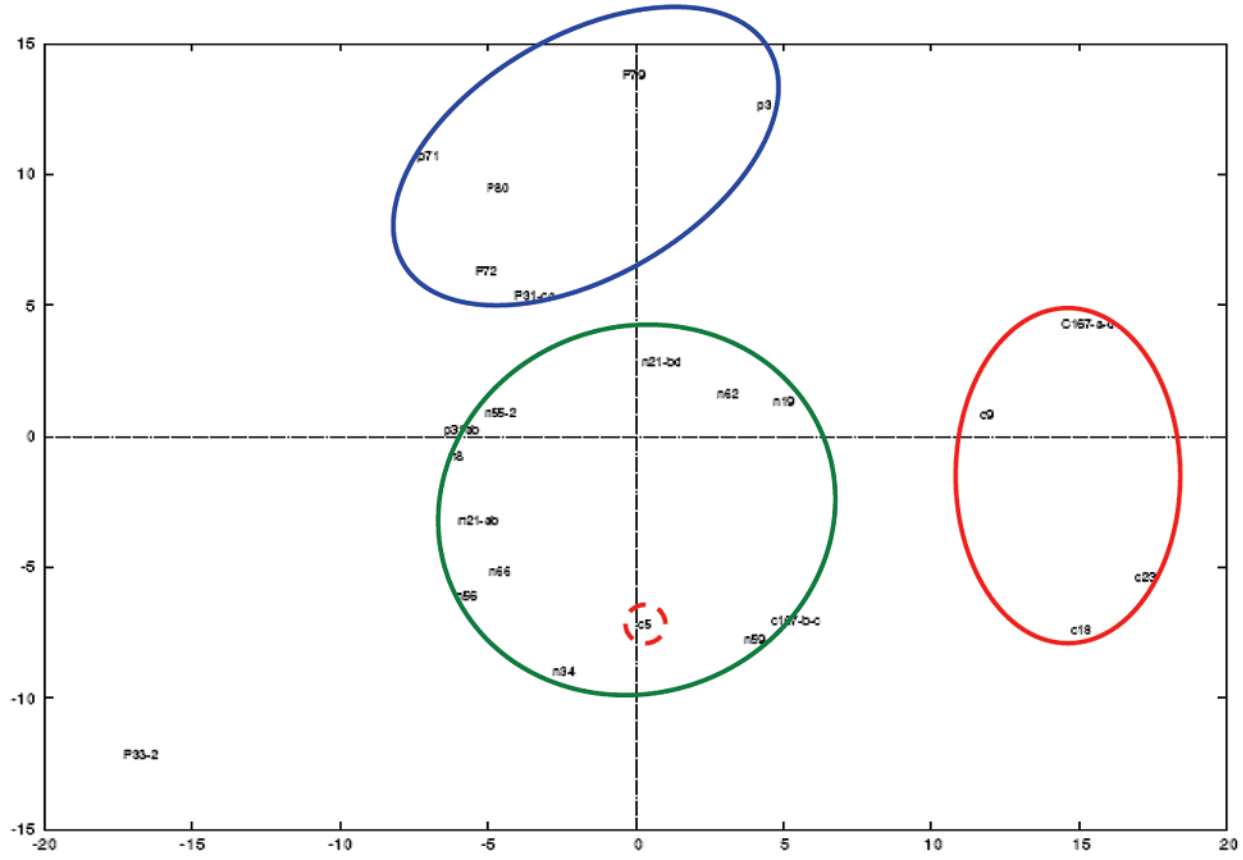
(b) ConA


P93-2
n56
n34
n52
n21-ab
c5
c167-b-c
n55-2
n8
n62
n21-bd
n19
p31ab
c18
c23
c9
C167-a-d
P31-ce
P72
P80
p71
P79
p3

(c) MAL


20
15
10
5
0
-5
-10
-15
-20
-15
-10
-5
0
5
10
15
p9
P79
P72
p71
P80
P31-cg
p31ab
n21-bd
n62
n19
n8
n55-2
n21-ab
n66
n56
n59
n34
C167-a-d
c9
c23
c18
c5
c167-b-c
P33-2

(d) SNA


15
10
5
0
-5
-10
-15
-20
-15
-10
-5
0
5
10
15
20
P79
p9
p71
P80
P72
P31-cg
n21-bd
n62
n19
n55-2
p31ab
n8
n21-ab
n66
n56
c5
n59
n34
C167-a-d
c9
c23
c18
P33-2

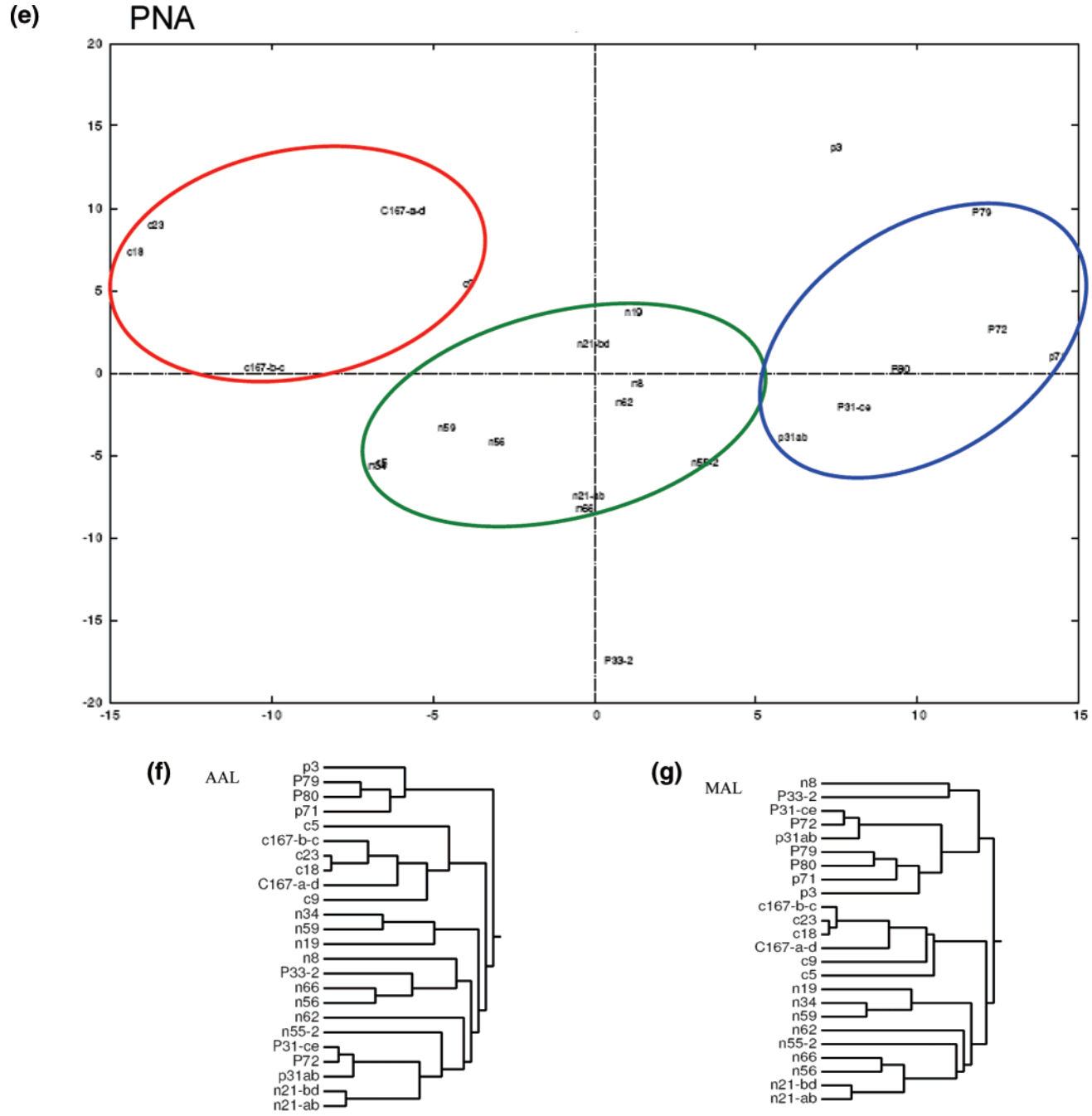


Figure 3. The normalized glycoprotein microarray responses to lectins (a) AAL, (b) ConA, (c) MAL, (d) SNA, and (e) PNA were visualized by principal component analysis (PCA). Twenty-four serum samples (10 normal, 8 chronic pancreatitis, and 6 pancreatic cancers) were studied. (f−g) Average linkage hierarchical clustering (HC) of the array responses to (f) AAL and (g) MAL were shown to provide graphical representations of the relationships among the samples. The figure shows the clustering of serum samples obtained from patients with pancreatic cancer, chronic pancreatitis, or from normal subjects.

(125 μL of each) was subjected to immunodepletion prior to the lectin extraction step to facilitate detection of lower abundance proteins. The immunodepletion was performed using the IgY-12 column (Beckman-Coulter), thereby removing the 12 most abundant serum proteins (albumin, IgG, α1-antitrypsin, IgA, IgM, transferrin, haptoglobin, α1-acid glycoprotein, α2-macroglobin, apolipoprotein A-I and A-II, and fibrinogen). Supplementary Figure 1a (see Supporting Information) shows the UV chromatogram of the depletion process where the immunodepleted fraction elutes at around 8 min. Approximately 8% of total serum proteins were retained in the immunodepleted serum fraction.

Glycoproteins retained in the immunodepleted serum were subsequently enriched using a multi-lectin affinity column composed of WGA and ConA. ConA recognizes α-linked mannose, including high mannose-type and mannose core structures which are common to N-linked glycosylated proteins. WGA can interact with some glycoproteins via sialic acid residues and it also binds oligosaccharides containing terminal *N*-acetylglucosamine.[26] Thus, a majority of the complex type

Table 1. *Z* Value of the Altered Glycosylations Detected by Five Lectins ($Z > 2$ or $Z < -2$ Corresponds to $P < 0.05$)

	AAL		MAL		SNA		ConA		PNA	
protein ID/acc #	normal	cancer	normal	cancer	normal	cancer	normal	cancer	normal	cancer
beta-2-glycoprotein 1 (P02749)		2.49		3.3		2.08		2.07	−2.47	2.13
hemopexin (P02790)		6.15		2.85		3.24		3.01		
haptoglobin-related protein (P00739)	−3.6	2.82	−2.49		−3.71	2.41	−3.08		−3.63	
serum amyloid P-component (P02743)	−4.96	4.02	−4.96	2.85	−5.28	4.11	−5.31	3.59	−5.96	3.12
clusterin (P10909)	−2.22	2.92	−2.52			2.22			−2.08	
antithrombin-III (P01008)	−3.5	3.18	−2.9	3.28	−2.93	2.58	−3.24	2.63	−3.44	2.56
kininogen-1 (P01042)	−2.69	4.31		2.39	−2.64	3.98	−3.06	3.95	−2.1	2.14
plasma protease C1 inhibitor (P05155)						−2.97				

glycans can interact with WGA. Combining these two lectins facilitated the extraction of most of the N-linked glycoproteins in serum. Approximately 70% protein recovery was achieved from each immunodepleted serum using this lectin affinity column, and ~90% of N-linked glycoproteins (>2 peptide hits) were reproducible from nano LC−MS/MS analysis of three dual-lectin extractions (data not shown).

Thirty micrograms of protein from each sample of lectin-enriched glycoproteins were further separated on a nonporous reverse-phase HPLC (NPS-RP-HPLC) C18 column. High-resolution separation of intact proteins was achieved, with the eluting proteins being detected by UV absorption at 214 nm. Supplementary Figure 1b (see Supporting Information) shows the UV map consisting of the chromatograms of three selected samples. Generally, we observed a high level of reproducibility in the UV traces among the different samples in the same group, although slight retention time shifts were observed. The UV peak area variance was within 10% for serum samples from different individuals. Protein fractions were collected by peak. The immunodepletion step simplified the serum sample, thereby making the collected UV peaks relatively pure compared to the non-immunodepleted sample. Although we occasionally observed more than one protein per UV peak, it was generally found that the dominant protein was responsible for the UV absorption or glycan expression change.[7]

3.2. Lectin Glycoarrays for Differential Detection of Changes in Glycan Structure. The intact glycoproteins were separated and collected, and the peaks were spotted on nitrocellulose slides using a noncontact microarray spotter. The microarrays were then hybridized against various lectins for differential glycan expression analysis. Five lectins (AAL, MAL, SNA, PNA, and ConA) were used to detect different glycan moieties. AAL recognizes fucose linked (α-1,6) to *N*-acetylglucosamine or to fucose linked (α-1,3) to *N*-acetyllactosamine. MAL can detect glycans containing NeuAc-Gal-GlcNAc with sialic acid at the 3 position of galactose whereas SNA binds preferentially to sialic acid attached to terminal galactose in an (α-2,6) and, to a lesser degree, an (α-2,3) linkage.[2728] In contrast, PNA binds desialylated exposed galactosyl (*β*-1,3) *N*-acetylgalactosamine. In fact, sialic acid in close proximity to the PNA binding site will inhibit PNA binding. ConA was also used to detect high mannose structures. Greater than 95% of N-glycan types can be covered using these five lectins. The glycoproteins were hybridized with lectins to probe differences in glycan content between normal, chronic pancreatitis and pancreatic cancer sera and the binding was visualized using a biotin-streptavidin-AlexaFluor555 interaction.[24]

Figure 2 shows sections of 5 microarrays probed with 5 different lectins. The left 5 lanes contain normal samples, the middle 3 lanes contain cancer samples and the right 4 lanes contain the chronic pancreatitis samples. The array data suggests that this particular fraction contains glycan structures consisting primarily of mannose and fucose residues as reaction with ConA and AAL were significant. Further, it appears that the overall levels of mannosylation and fucosylation are higher in cancer samples compared to normal. However, the raw microarray data in this figure was not normalized and should be analyzed with caution.

As only changes induced by variations in glycan expression are of interest, all array spot intensities were normalized with respect to their corresponding UV peak areas from the chromatograms to mitigate protein abundance differences. The normalized data was used for cluster analysis.

3.3. Bioinformatic Analysis of the Glycoprotein Patterns. Bioinformatic analysis of the glycoprotein arrays was performed to highlight lectin response patterns that grouped different disease states together. For data visualization, average linkage HC and PCA were used to provide graphical representations of the relationships among the samples. In these unsupervised approaches, the samples were placed either in a hierarchical relationship or as points in a 2-dimensional scatterplot (PCA) based on similarities in normalized glycoform abundances. For visualization, the data for each of the 5 lectins used in our experiments were analyzed separately. All 24 sera (8 chronic pancreatitis, 6 pancreatic cancer, and 10 normal), each assayed in duplicate, were analyzed using unsupervised visualization approaches and supervised differential abundance analyses. Separate PCA and HC results were generated for all 48 samples and for the replicate averages of the 24 distinct biological specimens. The normalized abundances were log transformed, then their pairwise correlations were used to carry out HC, and their pairwise co-variances were used for PCA. The results on the set of 48 samples clearly showed excellent concordance between replicates from the same biological source (see supplementary data Figure 2a−e, Supporting Information). Therefore, for biological inferences, we focused on the results for the 24 specimen-wise averages. The scatter plots where the duplicate averaged samples tend to cluster separately are illustrated in Figure 3 a−e. The pancreatic cancer samples clustered further away from the normal sample than the chronic pancreatitis sample pool, especially in response to AAL, ConA, and SNA. There were some outliers in the cancer pool that fell into the chronic pancreatitis pool. However, it was seen that this behavior always occurred with sera from the same 3 patients, indicating that the outliers were likely due to individual patient heterogeneity. It was also observed that the glycan expression of chronic pancreatitis serum glycoproteins were more similar to glycoproteins from the normal sera than to glycoproteins from the cancer sera, as shown in the fucosylated (Figure 3a) and high mannose glycan expression (Figure 3b). The hierarchical clustering results of average samples detected by ConA and MAL are shown in Figure 3f

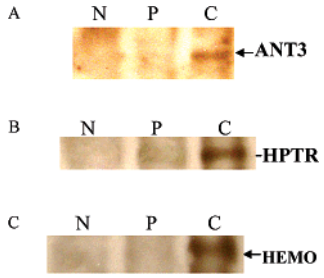


Figure 4. AAL lectin blot analysis of (a) Antithrombin-III, (b) Haptoglobin-related protein, (c) Hemopexin in N (normal), P (chronic pancreatitis), and C (pancreatic cancer) serum.

and g. The clustering results for fucosylated and sialylated glycan expression patterns generally distinguished the three clinical groups. Results with some lectins more clearly distinguished cancer from chronic pancreatitis/normal, whereas other lectins more clearly distinguished normal from cancer/chronic pancreatitis (shown in the supplementary data Figure 3a−c, Supporting Information).

3.4. Proteins with Altered Glycan Structures in Pancreatic Cancer Serum. For differential abundance analysis, we calculated Z-statistics and Wilcoxon rank-sum statistics for the normalized array spot intensities from each LC fraction, as detected by each lectin. Comparisons were made of pancreatic cancer versus chronic pancreatitis and normal combined and normal versus chronic pancreatitis and cancer combined. Only Z values higher than 2 or lower than −2, meaning only <5% would be expected by chance, were considered significant. A positive Z value indicates overexpression and a negative Z value indicates underexpression. The data suggests that all of the lectins have substantial power for identifying cancer samples relative to control or normal samples.

Proteins with significant changes ($P < 0.05$) in chronic pancreatitis or pancreatic cancer serum were digested and identified by peptide sequencing using nanoLC−MS/MS. Positive protein identification was validated by the Trans-Proteomics pipeline which includes both PeptideProphet and ProteinProphet software.[25] PeptideProphet automatically validates peptide assignment to MS/MS spectra made by a database search program such as SEQUEST. For each dataset, it calculates the distribution of search scores and peptide properties among correct and incorrect peptides, and uses those distributions to compute for each obtained peptide sequence a probability that it is correct. ProteinProphet takes the peptides and search results and statistically validates the identifications at the protein level. The altered protein IDs together with their Z-statistics ($Z > 2$ or $Z < -2$) are summarized in Table 1. The positive Z score in "cancer" is indicative that this glycosylation is specifically overexpressed in cancer compared to normal and pancreatitis combined. The negative Z score in "normal" is indicative that this glycosylation is underexpressed in the normal sample compared to pancreatitis and cancer combined. Thus, the differences shown are cancer specific.

The identification probabilities of all these reported protein IDs are higher than 95%. In certain fractions, more than one protein was identified due to coelution during LC separation. These fractions were further separated by SDS-PAGE and analyzed by lectin blots to determine which of the coeluting proteins was responsible for the differential lectin responsiveness.

Increased fucosylation and sialylation in pancreatic cancer sera were detected on most of the differentially glycosylated

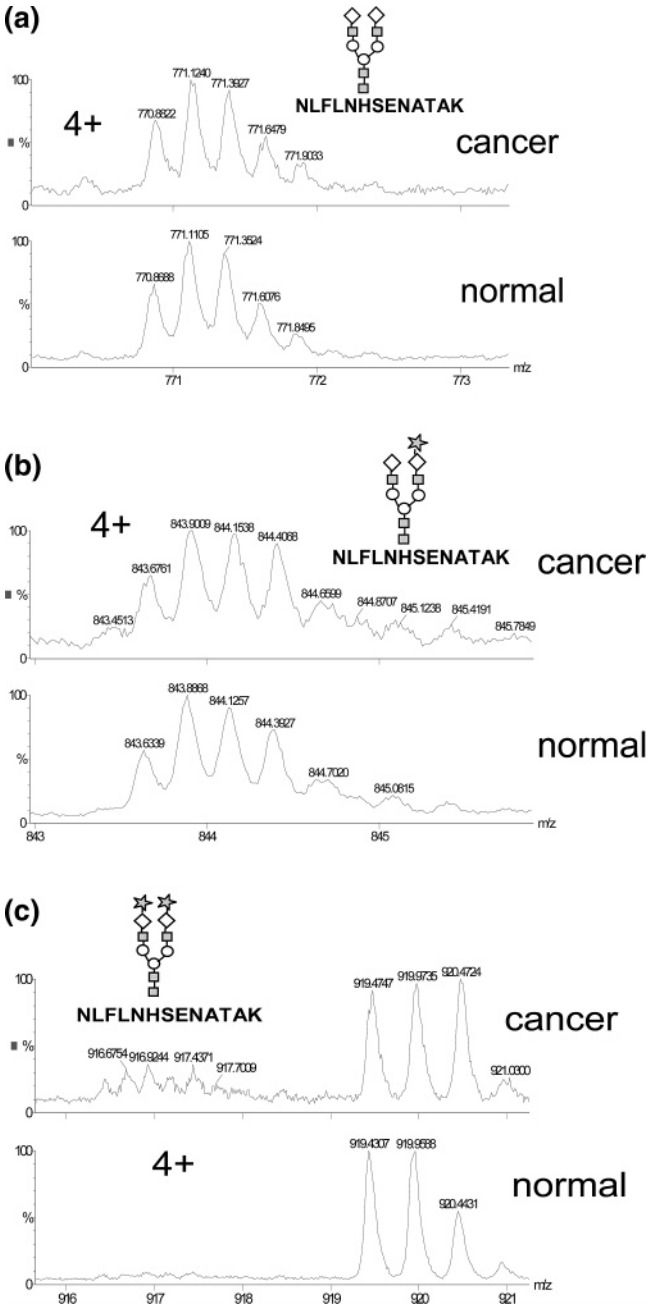


Figure 5. Peptide mapping of Haptoglobin-related protein. (a) Glycopeptide NLFL*N*HSE*N*ATAK(145−157) + $(Hex)_2(HexNAc)_2$ + $(Man)_3(GlcNAc)_2$, (b) glycopeptides NLFL*N*HSE*N*ATAK(145−157) + $(Hex)_2(HexNAc)_2(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$, and (c) glycopeptides NLFL*N*HSE*N*ATAK(145−157) + $(Hex)_2$ $(HexNAc)_2(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$ were detected as multiple charged peaks in normal and pancreatic cancer serum.

proteins, including Hemopexin, Beta-2-glycoprotein 1, serum amyloid P-component, Antithrombin-III, and Haptoglobin-related protein. Decreased sialylation was detected on plasma protease C1 inhibitor, as detected previously.[7] The immunodepletion of the abundant serum proteins in combination with further separation and lectin detection enabled the observation of glycosylation alteration in less-abundant proteins that had previously been difficult to detect. Some proteins have been suggested to be potential marker proteins in cancer. Beta-2-glycoprotein has been observed to be overexpressed in

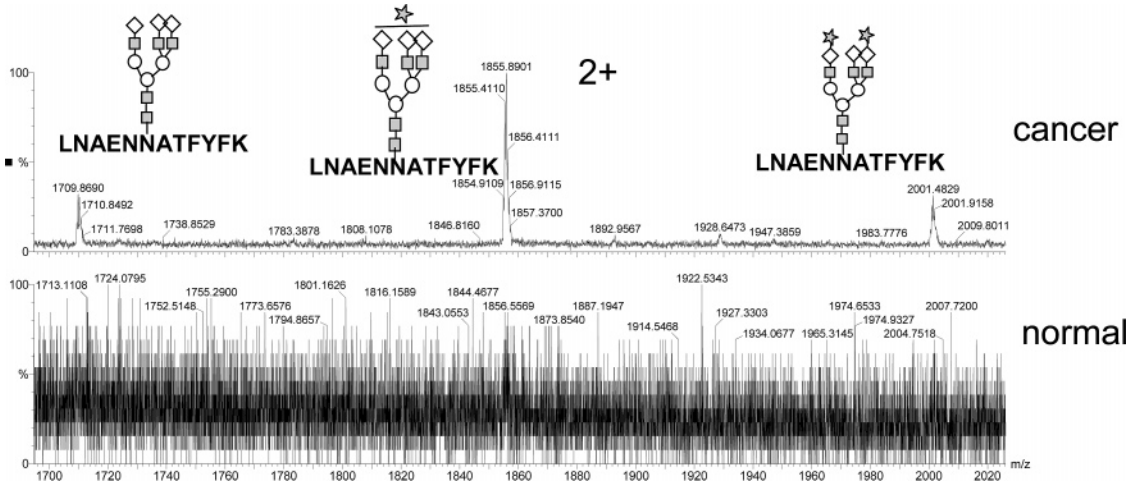


Figure 6. Peptide mapping of Kininogen-1 (P01042). Glycopeptide LNAEN*N*ATFYFK(289−300) + $(Hex)_3(HexNAc)_3$ + $(Man)_3(GlcNAc)_2$, LNAEN*N*ATFYFK(289−300) + $(Hex)_3(HexNAc)_3(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$, and LNAEN*N*ATFYFK(289−300) + $(Hex)_3(HexNAc)_3(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$ were detected as doubly charged peaks.

breast cancer serum[29] and serum amyloid P-component has been found down-regulated in stomach cancer tissue.[30] However, the glycosylation pattern alteration of these proteins has not been widely studied in sera from other cancer types.

Increased sialylation and fucosylation of these proteins in cancer serum lends support to the theory that glycosylation changes may have clinical utility for the identification of markers for early cancer detection. To verify the glycosylation changes that we observed, lectin immunoblotting and glycopeptide mapping experiments were performed on selected LC fractions. AAL lectin was used to examine the fucosylation expression level of target proteins. Figure 4a shows the lectin blot results of fucosylated Antithrombin-III. It was observed to be up-regulated in cancer serum compared to normal and chronic pancreatitis serum. Peptide mapping experiments were performed on the tryptic peptides from the LC fractions using μLC−ESI-TOF. As shown in Supplementary Figure 4a (see Supporting Information), very similar patterns were observed for the unmodified peptides for cancer versus normal samples from Antithrombin-III. However, an overexpressed fucosylated mono-sialylated glycopeptide was detected in cancer serum (Supplementary Figure 4b, see Supporting Information). This is consistent with the up-regulation of fucosylation and sialylation on Antithrombin-III observed in the microarray experiment (Table 1). These results highlight the potential utility of using altered glycosylation patterns, rather than absolute protein levels, as markers for early cancer detection.

Haptoglobin-related protein epitope expression is a clinically important predictor of the recurrence of cancer in patients with early breast cancer, especially in combination with progesterone-receptor status.[31] In our study, the overexpression of fucosylated haptoglobin-related protein in pancreatic cancer serum was also verified by both glycoprotein microarrays and lectin blot experiments (see Figure 4b and Table 1). The up-regulation of fucosylated haptoglobin in pancreatic cancer serum has been reported previously,[32] although this protein was removed during the immunodepletion step; thus, its glycosylation changes were not analyzed in this study. In a peptide mapping experiment, similar levels of a desialylated glycan structure and a mono-sialylated glycan structure were observed on the peptide NLFLNHSENATAK from haptoglobin-related protein in cancer and normal samples (Figure 5a and b), whereas the fully sialylated glycan structure on this peptide was found to be up-regulated in cancer sample as shown in Figure 5c. These results confirmed the increased response of SNA lectin and unaltered response of PNA lectin on this protein (Table 1).

The overexpression of fucosylation in hemopexin in hepatocellular carcinoma has been previously reported.[33] In pancreatic cancer serum, significant up-regulation with a *Z* score of 6.15 was observed on fucosylated Hemopexin. The lectin blotting experiment verified this alteration where an increased response to AAL was observed as compared to chronic pancreatitis and normal serum (Figure 4c). Overexpressed desialylated and partially sialylated glycopeptides were also observed on Kininogen-1 in pancreatic cancer sera (shown in Figure 6). This is consistent with the result from glycoprotein microarrays where an increased response of this protein to SNA, MAL, and PNA was detected in cancer samples.

In some cases, the glycan moieties that were detected by the five lectins including sialylation, fucosylation, galactosylation, and mannosylation were all up-regulated in pancreatic cancer. For instance, the glycosylation of both Serum amyloid P-component and Beta-2-glycoprotein 1 were found to be up-regulated as detected by all five lectins. This may be due to the increased branching of glycans that has been associated with metastasis and has been correlated with tumor progression in human cancers of pancreas, breast, colon, and melanomas.[34,35,9,12] Fucosyltransferase 3 has been shown to be overexpressed, and several isoforms of mannosidase have been shown to have decreased expression in pancreatic cancer (as compared to chronic pancreatitis and normal pancreata).[36] Fucosyltransferases increase fucosylation in selected proteins. The overexpression of highly branched glycosylation either implicates increased activity of certain glycosyltransferases (which may lead to increased expression of certain terminal glycans such as sialic acid and fucosyl residues) or to decreased mannosidase activity (leading to decreased trimming of high mannose structures, with corresponding increased branching of the high mannose core) in cancer. However, all that being said, no pancreatic proteins were found to show large glycosylation changes in the present study. In fact, the majority of interesting proteins are secreted by the liver. It is well-established that pancreatic cancer is a highly inflammatory

neoplasm and, as such, may elicit inflammatory cytokine production with an associated acute phase response from the liver. This acute phase response may, in fact, contribute to the synthesis of altered glycan moieties on the secreted liver glycoproteins.

4. Conclusion

We have demonstrated the utility of glycoprotein microarrays as a tool to differentiate serum samples from patients with pancreatic cancer, chronic pancreatitis, or normal subjects. Analysis of multiple normal, chronic pancreatitis, and pancreatic cancer sera showed distinct segregation of each state following PCA and HC analysis. Normal and chronic pancreatitis sera were closer in similarity to each other, whereas pancreatic cancer sera were distinct from the other two groups. Sialylation and fucosylation were the dominant glycosylation differences seen to change with the progression of pancreatic cancer. Both an increase and decrease in glycosylation levels of different proteins were observed as a function of disease. Many proteins whose glycosylation patterns changed as a function of disease have been previously implicated in cancer. The results from this study confirm previously implicated changes[9,12] in that not only do protein abundances change as a function of cancer, but more importantly, modifications such as changes in glycosylation patterns of a serum glycoproteome may indicate presence or absence of a disease. This change in disease specific glycosylation was further confirmed for selected proteins by glycopeptide mapping experiments using a μLC−ESI−MS platform that was able to show distinct glycopeptide differences between different sample types. The ability to screen serum glycosylation patterns for sample classification and detect the location of the altered glycosylations by further mass spectrometric validation may have utility for the early detection of cancer.

Acknowledgment. This work was supported in part by the National Cancer Institute under grant R01CA106402 (D.M.L.), the National Institutes of Health under grant R01GM49500 (D.M.L.), and a Michigan Economic Development Grant MEDC03-622 (D.M.S.). Support was also generously provided by Eprogen, Inc.

Supporting Information Available: Supplementary Figures 1−4. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Jemal, A.; Murray, T.; Ward, E.; Samuels, A.; Tiwari, R. C.; Ghafoor, A.; Feuer, E. J.; Thun, M. J. *CA Cancer J. Clin.* **2005**, *55*, 10−30.
(2) Dalgleish, A. G. *Bmj* **2000**, *321*, 380.
(3) Duffy, M. J. *Ann. Clin. Biochem.* **1998**, *35*(Pt 3), 364−370.
(4) Nazli, O.; Bozdag, A. D.; Tansug, T.; Kir, R.; Kaymak, E. *Hepatogastroenterology* **2000**, *47*, 1750−1752.
(5) Anderson, N. L.; Anderson, N. G. *Electrophoresis* **1998**, *19*, 1853−1861.
(6) Durand, G.; Seta, N. *Clin. Chem.* **2000**, *46*, 795−805.
(7) Zhao, J.; Simeone, D. M.; Heidt, D.; Anderson, M. A.; Lubman, D. M. *J. Proteome Res.* **2006**, *5*, 1792−1802.
(8) Peracaula, R.; Royle, L.; Tabares, G.; Mallorqui-Fernandez, G.; Barrabes, S.; Harvey, D. J.; Dwek, R. A.; Rudd, P. M.; de Llorens, R. *Glycobiology* **2003**, *13*, 227−244.
(9) Block, T. M.; Comunale, M. A.; Lowman, M.; Steel, L. F.; Romano, P. R.; Fimmel, C.; Tennant, B. C.; London, W. T.; Evans, A. A.; Blumberg, B. S.; Dwek, R. A.; Mattu, T. S.; Mehta, A. S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 779−784.
(10) Zhao, J.; Qiu, W.; Simeone, D. M.; Lubman, D. M. *J. Proteome Res.* **2007**, *6*, 1126−1138.
(11) Peracaula, R.; Tabares, G.; Royle, L.; Harvey, D. J.; Dwek, R. A.; Rudd, P. M.; de Llorens, R. *Glycobiology* **2003**, *13*, 457−470.
(12) Drake, R. R.; Schwegler, E. E.; Malik, G.; Diaz, J.; Block, T.; Mehta, A.; Semmes, O. J. *Mol. Cell. Proteomics* **2006**, *5*, 1957−1967.
(13) Kasbaoui, L.; Harb, J.; Bernard, S.; Meflah, K. *Cancer Res.* **1989**, *49*, 5317−5322.
(14) Kellokumpu, S.; Sormunen, R.; Kellokumpu, I. *FEBS Lett.* **2002**, *516*, 217−224.
(15) Ng, R. C.; Roberts, A. N.; Wilson, R. G.; Latner, A. L.; Turner, G. A. *Br. J. Cancer* **1987**, *55*, 249−254.
(16) Uttamchandani, M.; Wang, J.; Yao, S. Q. *Mol. Biosyst.* **2006**, *2*, 58−68.
(17) Sobek, J.; Bartscherer, K.; Jacob, A.; Hoheisel, J. D.; Angenendt, P. *Comb. Chem. High Throughput Screen* **2006**, *9*, 365−380.
(18) Hultschig, C.; Kreutzberger, J.; Seitz, H.; Konthur, Z.; Bussow, K.; Lehrach, H. *Curr. Opin. Chem. Biol.* **2006**, *10*, 4−10.
(19) Pal, M.; Moffa, A.; Sreekumar, A.; Ethier, S. P.; Barder, T. J.; Chinnaiyan, A.; Lubman, D. M. *Anal. Chem.* **2006**, *78*, 702−710.
(20) Wang, D.; Liu, S.; Trummer, B. J.; Deng, C.; Wang, A. *Nat. Biotechnol.* **2002**, *20*, 275−281.
(21) Haab, B. B.; Dunham, M. J.; Brown, P. O. *Genome Biol.* **2001**, *2*, RESEARCH0004.
(22) Yan, F.; Sreekumar, A.; Laxman, B.; Chinnaiyan, A. M.; Lubman, D. M.; Barder, T. J. *Proteomics* **2003**, *3*, 1228−1235.
(23) Davies, D. H.; Liang, X.; Hernandez, J. E.; Randall, A.; Hirst, S.; Mu, Y.; Romero, K. M.; Nguyen, T. T.; Kalantari-Dehaghi, M.; Crotty, S.; Baldi, P.; Villarreal, L. P.; Felgner, P. L. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 547−552.
(24) Patwa, T. H.; Zhao, J.; Anderson, M. A.; Simeone, D. M.; Lubman, D. M. *Anal. Chem.* **2006**, *78*, 6411−6421.
(25) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. *Anal. Chem.* **2002**, *74*, 5383−5392.
(26) Bakry, N.; Kamata, Y.; Simpson, L. L. *J. Pharmacol. Exp. Ther.* **1991**, *258*, 830−836.
(27) Shibuya, N.; Goldstein, I. J.; Broekaert, W. F.; Nsimba-Lubaki, M.; Peeters, B.; Peumans, W. J. *Arch. Biochem. Biophys.* **1987**, *254*, 1−8.
(28) Wang, W. C.; Cummings, R. D. *J. Biol. Chem.* **1988**, *263*, 4576−4585.
(29) Alsabti, E. A.; Muneir, K. *Jpn. J. Exp. Med.* **1979**, *49*, 235−240.
(30) Jang, J. S.; Cho, H. Y.; Lee, Y. J.; Ha, W. S.; Kim, H. W. *Oncol. Res.* **2004**, *14*, 491−499.
(31) Kuhajda, F. P.; Piantadosi, S.; Pasternack, G. R. *N. Engl. J. Med.* **1989**, *321*, 636−641.
(32) Okuyama, N.; Ide, Y.; Nakano, M.; Nakagawa, T.; Yamanaka, K.; Moriwaki, K.; Murata, K.; Ohigashi, H.; Yokoyama, S.; Eguchi, H.; Ishikawa, O.; Ito, T.; Kato, M.; Kasahara, A.; Kawano, S.; Gu, J.; Taniguchi, N.; Miyoshi, E. *Int. J. Cancer* **2006**, *118*, 2803−2808.
(33) Comunale, M. A.; Lowman, M.; Long, R. E.; Krakover, J.; Philip, R.; Seeholzer, S.; Evans, A. A.; Hann, H. W.; Block, T. M.; Mehta, A. S. *J. Proteome Res.* **2006**, *5*, 308−315.
(34) Kim, Y. J.; Varki, A. *Glycoconj. J.* **1997**, *14*, 569−576.
(35) Dube, D. H.; Bertozzi, C. R. *Nat. Rev. Drug Discovery* **2005**, *4*, 477−488.
(36) Logsdon, C. D.; Simeone, D. M.; Binkley, C.; Arumugam, T.; Greenson, J. K.; Giordano, T. J.; Misek, D. E.; Kuick, R.; Hanash, S. *Cancer Res.* **2003**, *63*, 2649−2657.

PR070062P