

Multilectin Affinity Chromatography for Characterization of Multiple Glycoprotein Biomarker Candidates in Serum from Breast Cancer Patients

ZIPING YANG,¹ LYND SAY E. HARRIS,² DARRYL E. PALMER-TOY,³ and WILLIAM S. HANCOCK^{1*}

Background: Glycoproteins are often associated with cancer and are important in serum studies, for which glycosylation is a common posttranslational modification.

Methods: We used multilectin affinity chromatography (M-LAC) to isolate glycoproteins from the sera of breast cancer patients and controls. The proteins were identified by HPLC–tandem mass spectrometry (MS/MS) analysis of the corresponding tryptic digests. We used the FuncAssociate Gene Ontology program for association analysis of the identified proteins. Biomarker candidates in these groups were comparatively quantitated by use of peak area measurements, with inclusion of an internal standard. We analyzed data for concordance within the ontology association groups for vector of change with the development of breast cancer.

Results: Detection of the known low-concentration biomarker HER-2 (8–24 $\mu\text{g/L}$) enabled us to establish a dynamic range of 10^6 , relative to the amount of albumin, for the depletion step. We then used ELISA to confirm this range. Proteins associated with lipid transport and metabolism, cell growth and maintenance, ion homeostasis, and protease inhibition were found to be differentially regulated in serum from women with breast cancer compared with serum from women without breast cancer.

Conclusions: M-LAC for isolation of the serum glycoproteome, coupled with liquid chromatography–MS/MS and the use of gene ontology associations, can be used to characterize large panels of candidate markers, which can then be evaluated in a particular patient population.

© 2006 American Association for Clinical Chemistry

Although breast cancer is implicated in ~40 000 deaths each year in the US, this disease is highly curable if diagnosed at an early stage (1), but screening mammography is likely to miss early cancers (2). Proteomic profiling may identify tumor markers in blood that could aid in early diagnosis and subsequent treatment (3, 4), but characterizing such biomarkers in early-stage breast cancer is challenging. During early-stage disease, only small amounts of material may be released into the bloodstream, and analysis of these low-abundance proteins is complicated by the large dynamic range of serum proteins and the complexity of posttranslational modifications, especially glycosylation. The removal of the most abundant protein(s) by use of an immunoaffinity approach can enhance the identification of low-abundance proteins in plasma and serum (5), but depletion of proteins such as albumin that form complexes with low-abundance proteins can lead to loss of potential biomarkers (6).

We previously developed a multilectin affinity system for efficient and specific enrichment of human serum glycoproteins (7) and demonstrated that a multilectin affinity chromatography (M-LAC)⁴ column containing Jacalin, concanavalin A, and wheat germ agglutinin could overcome the low affinity and lack of complete glycopro-

¹ Barnett Institute and Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA.

² Dana-Farber Cancer Institute, Boston, MA.

³ Renal Unit, Department of Medicine, Massachusetts General Hospital, Harvard University, Cambridge, MA.

*Address correspondence to this author at: Barnett Institute and Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115. Fax 617-373-2855; e-mail wi.hancock@neu.edu.

Received December 23, 2005; accepted July 11, 2006.

Previously published online at DOI: 10.1373/clinchem.2005.065862

⁴ Nonstandard abbreviations: M-LAC, multilectin affinity chromatography; DCIS, ductal carcinoma in situ; IC, invasive breast cancer; ACN, acetonitrile; MS, mass spectrometry; MS/MS, tandem mass spectrometry; EIC, extracted ion chromatogram; GO, gene ontology; LC, liquid chromatography.

tein capture that typically occurs with a single-lectin affinity selector. This approach was specific to glycoproteins, yielded good recovery, and was reproducible, markedly improving the dynamic range of serum proteomic analysis and minimizing nonspecific losses associated with the depletion of abundant carrier proteins.

Because glycoproteins make up at least 50% of the blood proteome (7), and blood biomarkers of breast cancer that are glycoproteins have already been identified, such as HER-2 (8) and carcinoembryonic antigen (9, 10), we used M-LAC to investigate the glycoproteome in serum of breast cancer patients.

Materials and Methods

MATERIALS

Agarose-bound concanavalin A with a protein concentration of 6 g lectin/L of gel and a binding capacity of >4 g ovalbumin/L of gel, agarose-bound wheat germ agglutinin with a protein concentration of 7 g lectin/L of gel and a binding capacity of 8 g *N*-acetylglucosamine/L of gel, and agarose-bound Jacalin lectin with a protein concentration of 4 g lectin/L of gel and a binding capacity of >4 g monomeric IgA/L of gel were from Vector Laboratories. Trypsin (sequence grade) was from Promega.

SAMPLES

Study samples were collected from May 2000 to February 2002 from patients with ductal carcinoma in situ (DCIS) or invasive breast cancer (IC) [American Joint Committee on Cancer (11) stage IV]. We obtained control samples from healthy, cancer-free women who donated blood to the Brigham and Women's Hospital Blood Bank from January 2001 to June 2002. All participants gave informed consent in accordance with protocol 93-085 of the Partners Institutional Review Board. Samples were archived under protocols approved by the Human Subjects Committee of the Partners HealthCare System (Boston, MA) and the

Institutional Review Board for the Dana-Farber Cancer Institute, Harvard Medical School. Samples used for this study had been stored for 1–3 years (controls) or 2–3 years (patients). Patient disease data are listed in Table 1. Blood samples were collected in red-top tubes (Becton Dickinson) and processed in a CLIA-approved facility (Dana-Farber Cancer Clinical Laboratory). After separation from clot, serum was divided into aliquots, placed in cryovials, and stored in the Specialized Programs of Research Excellence Breast Cancer Research Bank. Samples were thawed once and divided into 100- μ L aliquots. Each aliquot was used only once per analysis, so only 1 freeze-thaw cycle per aliquot was required. Serum samples were obtained from 5 patients with IC, 5 patients with DCIS, and 5 controls. We pooled the same amount of protein from each sample and processed the 3 pooled samples and 15 individual samples with the following procedures.

GLYCOPROTEIN ENRICHMENT BY M-LAC COLUMN

We prepared the multilectin column by mixing equal amounts of agarose-bound concanavalin A, agarose-bound wheat germ agglutinin, and agarose-bound Jacalin in an empty PD-10 disposable column (GE Healthcare). The 100- μ L serum sample was diluted with multilectin column equilibrium buffer (20 mmol/L Tris, 0.15 mol/L NaCl, 1 mmol/L Mn^{2+} , and 1 mmol/L Ca^{2+} , pH 7.4) to a volume of 1 mL and loaded on a newly packed multilectin affinity column. After a 15-min incubation, the unbound proteins were eluted with 10 mL of equilibrium buffer, and the captured proteins were released with 12 mL of a specific displacer solution (20 mmol/L Tris, 0.5 mol/L NaCl, 0.17 mol/L methyl- α -D-mannopyranoside, 0.17 mol/L *N*-acetylglucosamine and 0.27 mol/L galactose, pH 7.4). The captured fraction was collected from the multilectin affinity column and concentrated with 15-mL, 10 kD Amicon filters (Millipore).

Table 1. Description of the patients.

Sample	Stage ^a	Histology	Receptors ^b	Serum HER-2, μ g/L ^c	
				Mean	SD
BC1	0	Comedo and cribriform DCIS	NA ^d	8.5	0.2
BC2	0	Comedo DCIS	ER ⁺ , PR ⁻ , HER-2 ⁻	10.7	0.2
BC3	0	High-grade DCIS	NA	12.8	0.3
BC4	0	Comedo and cribriform DCIS	ER/PR ⁺ , HER-2 ⁻	13.1	0.7
BC5	0	Intermediate-grade, clinging DCIS	NA	9.0	0.2
BC6	IV	Ductal	NA	10.5	0.7
BC7	IV	Lobular	ER/PR ⁺ , HER-2 ⁻	14.6	1.2
BC8	IV	Ductal	ER/PR ⁺ , HER-2 ⁻	12.5	0.8
BC9	IV	Ductal	ER/PR ⁻ , HER-2 ⁻	10.4	0.2
BC10	IV	Ductal	ER/PR ⁺ HER-2 ³⁺	24.4	0.1

^a Stage according to the American Joint Committee on Cancer classification system.

^b ER, estrogen receptor; PR, progesterone receptor; HER-2, human epidermal growth factor receptor 2. All markers determined by immunohistochemistry.

^c Serum HER-2 extracellular domain.

^d NA, not available.

LIQUID CHROMATOGRAPHY-TANDEM MASS SPECTROMETRY

We used a previously described procedure (7) for trypsin digestion of the glycoprotein fractions. The trypsin-digested peptides were then separated and analyzed on an Ettan MDLC system (GE Healthcare) coupled to an LTQ linear ion trap (ThermoElectron). Approximately 2 μg of each sample were injected onto a Peptide Captrap column (Michrom Bioresources, Inc.) with the autosampler of the MDLC system. To desalt the sample, the trap column was washed with H_2O with 0.1% formic acid at a flow rate of 10 $\mu\text{L}/\text{min}$ for 4 min. Flow was directed to the solvent waste through a 10-port valve. After the sample was desalted, the valve was switched to direct the flow to the separation column. The desalted peptides were then released and separated on a C18 capillary column (packed in-house; Magic C18; 150×0.075 mm). The flow rate was maintained at 400 nL/min and monitored with a flow meter. The gradient was started at 0% acetonitrile (ACN) with 0.1% formic acid and linearly increased to 35% ACN in 120 min, then to 60% ACN in 40 min, and to 90% ACN in another 20 min. The gradient was then maintained at 90% ACN for 20 min. The Ettan MDLC was operated with UNICORNTM control software (GE Healthcare). We analyzed the resolved peptides on an LTQ linear ion trap mass spectrometer or the corresponding hybrid system coupled to a Fourier Transform mass spectrometry (MS) with a nanoelectrospray ionization ion source. The Fourier Transform was set for full mass scan, and the LTQ was set up for data-dependent tandem MS (MS/MS) fragmentation. The temperature of the ion transfer tube was controlled at 200 °C and the spray voltage was 2.0 kV. The normalized collision energy was set at 35% for MS/MS. Data-dependent ion selection was monitored to select the most abundant 7 ions from an MS scan for MS/MS analysis. Dynamic exclusion was continued for 2 min.

BIOINFORMATICS

Peptide sequences were identified with the Sequest algorithm (Version C1) incorporated in BioWorks software (Version 3.1 SR, ThermoElectron), and the spectra were searched against the human Rapido (November 2005) database. Only peptides resulting from tryptic cleavage were searched. To generate a large pool of candidate peptides we used the following search parameters: Xcorr $\geq 1.9, 2.2, 3.75$ for 1⁺, 2⁺, 3⁺ charged ions; $\Delta\text{Cn} \geq 0.1$; and Rsp ≥ 4 .

The MS/MS spectra of key peptide markers were further confirmed by manual interpretation of the cleavage pattern as well as inspection of the number of assigned b and y fragments and the presence of a high signal-to-noise ratio. An example of this process is depicted in Fig. 1, which shows the MS/MS spectra attributed to the peptide ¹⁴⁴SLTEILKGGVLIQR¹⁵⁷, presenting in the extracellular domain of the surface protein HER-2.

The identification of key peptides could be missed by MS/MS scan because of the time constraints of a flowing

system in which peptides are selected for fragmentation on the basis of abundance. To avoid possible false negatives, the m/z signal of the peptide with a mass window of 0.5 atomic mass units (amu) was selected for an extracted ion chromatogram (EIC). The presence of the peptide in the sample was considered to be confirmed by the observation of an EIC peak (signal-to-noise ratio, >5) within 0.5 min of the same retention time at which the peptide was detected, an approach similar to the use of accurate mass tags (12).

COMPARATIVE QUANTIFICATION BY PEAK AREA MEASUREMENT

Among the peptides identified in a given protein, we selected the one identified with highest confidence as a diagnostic peptide. For proteins identified with multiple peptides, we used additional peptides for peak area measurement and the mean for comparative quantification. The m/z signals of peptides were extracted from the liquid chromatography (LC)-MS EIC, and the peaks were integrated at the appropriate retention times. Each LC-MS/MS assay time was 2 h, and the reproducibility of between-assay retention times was achieved with an SD of 0.23 min. To monitor the variation between sample preparation and LC-MS analysis, we added standard bovine fetuin (5% of the total protein content of the sample) as an internal standard. The internal standard was digested together with the sample, and the peptides were separated and detected by LC-MS/MS. For normalization, we found the mean of the peak areas of 5 peptides with retention times covering the range at which the majority of the peptides were eluted. After normalization to eliminate the differential in ionization efficiency, the relative SD of these 5 peak areas was $<30\%$.

USE OF GENE ONTOLOGY TERMS TO SEARCH FOR BIOMARKER CANDIDATES

The list of proteins identified in the pooled control, DCIS, and IC samples (659 proteins identified with 3 or more peptides and 154 proteins with 2 peptides) were submitted to the FuncAssociate on-line program (<http://llama.med.harvard.edu/cgi/func/FuncAssociate>) for Gene Ontology (GO; March 2005) attribute characterization. This program groups related proteins on the basis of GO attributes. The GO attributes associated with the queue of the gene group were listed and ranked according to the significance of the association. We compared the rankings of GO attributes ($P < 0.001$ based on Fisher exact test) of control, DCIS, and invasive pools, and we selected the GO attributes with significant rank differences (arbitrarily set at >10) as target associations. We used peak area measurements to comparatively quantify the proteins associated with these attributes, selecting the proteins showing the greatest changes between pooled control and disease samples, with changes having the same trend as the attribute rankings. Then, we used peak area measurements to quantify these proteins in individual patient samples, found the mean of the normalized peak areas,

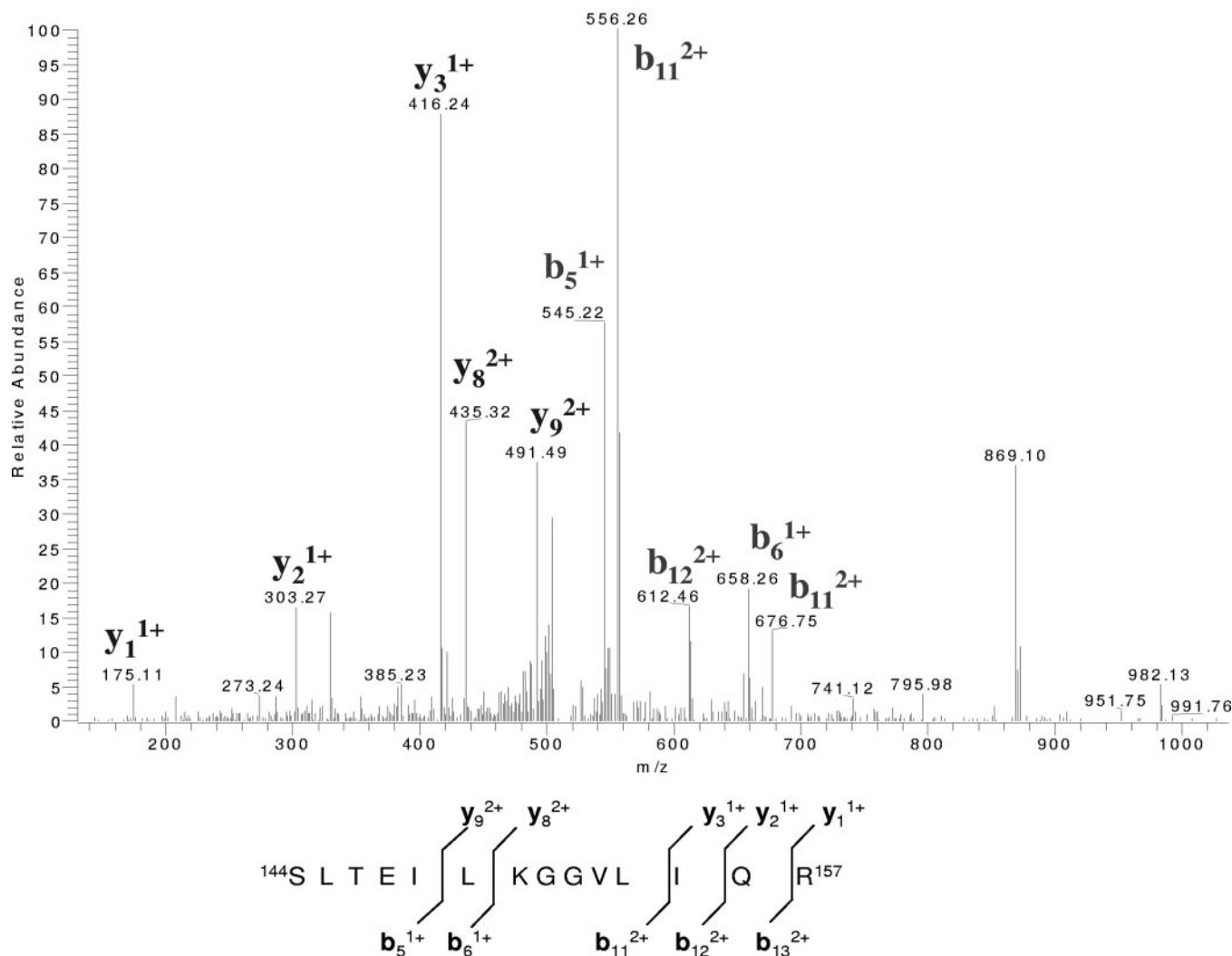


Fig. 1. The fragmentation (MS/MS) spectrum of 3+ charge state of the peptide $^{144}\text{SLTEILKGGVLIQR}^{157}$, derived from the protein HER-2.

and calculated the SDs. To measure the protein concentration difference between disease and control samples, we normalized all peak area values relative to the mean control values.

HER-2 ELISA IN BREAST CANCER SERA

We measured HER-2/ECD (extracellular domain) with a sandwich immunoassay (Oncogene Science, Inc). In brief, 200 μL of serum were incubated in a well coated with the capture antibody, reacting with the detector antiserum. The amount of detector antibody bound to antigen was measured by streptavidin/horseradish peroxidase conjugate. All samples were analyzed in duplicate. The inter-assay CV was 3.1%.

Results

GLYCOPROTEINS IDENTIFIED IN SERUM FROM BREAST CANCER PATIENTS

We used M-LAC columns to selectively capture the glycoprotein fraction of serum samples. Overall protein

recovery (captured plus unbound fraction) was 88% (CV <5%), and the bound glycoprotein fraction made up 12% of the total sample (approximately half of the remaining amount of proteins, excluding albumin). A total of 813 high-probability nonredundant proteins with 2 or more peptide identifications were found in the breast cancer serum samples (aggregated from all the individual and pooled sample analyses). The most abundant proteins present in the serum glycoprotein fraction included α -2-macroglobulin, serotransferrin, haptoglobin, hemopexin, α -1-antitrypsin, complement C3, apolipoprotein A-I, α -1-acid glycoprotein, ceruloplasmin, and complement factor H. Because of the depletion of abundant nonglycosylated proteins, especially serum albumin, any given glycoprotein marker was identified with a higher sequence coverage than in the nondepleted serum samples (see Table 2). For example, lower-abundance serum glycoproteins such as neuropilin-1 and pregnancy zone protein could not be detected in the nondepleted serum, but they were found with multiple peptides after the M-LAC enrichment step.

Table 2. A comparison of protein identifications between M-LAC-enriched fractions and nondepleted serum samples.^a

Protein	Sequence coverage, % ^b		Number of peptide identifications ^c	
	M-LAC-enriched	Nondepleted	M-LAC-enriched	Nondepleted
Apolipoprotein C III	27	16	2	1
Ceruloplasmin	45	29	30	19
Prothrombin	21	20	8	7
Fibrinogen β -chain	23	13	9	5
Neuropilin-1	4	0	2	0
α -1-Antichymotrypsin	33	17	9	5
Pregnancy zone protein	9	0	7	0

^a The sample was a pool of sera from 5 control patients.

^b Number of peptides measured vs total. Percentage of recovery is rounded to whole numbers.

^c Number of nonredundant peptides detected by MS/MS analysis.

For medium- or high-abundance proteins, such as α -1-antichymotrypsin and ceruloplasmin, the quality of identification (higher sequence coverage) was similarly improved by the M-LAC enrichment step (data not shown). These observations were consistent for the analysis of both pooled and individual samples.

IDENTIFICATION OF A KNOWN LOW-ABUNDANCE BIOMARKER, HER-2

To investigate the dynamic range of our proteomic study and to assess the performance of the M-LAC depletion procedure, we evaluated the preliminary identification of a low-abundance breast cancer biomarker in individual serum samples collected from patients with breast cancer. The HER-2 peptide (residues 144–157) was identified multiple times by MS/MS sequencing in the 10 breast cancer samples (nonpooled). The results of an MS/MS analysis performed on the most abundant 3⁺ charge state of the peptide [caused by improved ionization and fragmentation of a peptide with charged internal residues(13, 14)] are shown in Fig. 1. Similar findings were obtained when analysis was performed on the 2⁺ ion.

COMPARATIVE QUANTIFICATION BY NORMALIZED PEAK AREA (WITH AN INTERNAL STANDARD)

Comparative quantification approaches, such as stable isotope labeling or the incorporation of mass-tagged derivatives, (15, 16), have been used to compare protein concentrations between samples. These approaches are less suitable for studies on serum samples because isotope labels cannot be incorporated into human samples and because increased side reactions occur in the chemical derivatization of low-abundance peptides (13). For these reasons, we used the measurement of peak area of a diagnostic peptide in an EIC to determine relative protein abundance in different samples. This approach involved normalization for variations in sample preparation and in peptide ionization in HPLC-MS measurement by adding a known amount of a calibration protein to each sample at the beginning of the isolation protocol. This approach was used for the analysis of all individual and pooled samples. To minimize interference between the sample and the

internal standard in peptides peak area measurement, we selected bovine fetuin as the internal standard because it is glycosylated and contains many unique tryptic peptides relative to human proteins. The tryptic peptides generated from this internal standard elute over a range of retention times that cover the organic modifier concentrations used in the separation. Use of a range of peptide standards enables monitoring for any system fluctuations during the 2-h HPLC separation process. Furthermore, with a complex sample such as serum, not all low-abundance peptides can be consistently identified in a given LC-MS analysis, owing to the time constraints of a flowing system in which peptides are selected for fragmentation on the basis of abundance. In such situations, however, selected peptides can still be detected and the peak area measured in EICs at a predetermined mass and retention time window.

ANALYSIS OF POOLED SAMPLE SETS AND OF GO ATTRIBUTES

The proteomics analysis of several serum samples rapidly produces a large data set, which then becomes an informatics challenge, particularly if the analysis depends on the study of lists of unrelated proteins. In this study, therefore, we used analysis of GO-based associations to group protein identifications and decreased the complexity of the data sets for comparison. Results of triplicate analysis of the pools of control and disease samples for proteins selected to represent the GO categories are shown in Table 3. Variability of the peak area measurement was <30% (Table 3). Although it is a concern that pools of 5 alternative individuals may show different values because of biological variability (17), we used the pools in the following manner. We used pooled samples to perform several replicates, estimate the variability of the analytical measurement, and establish a baseline between control and disease samples. Assessment of true biological variability would require a much larger sample set and will be the focus of future investigations. Some markers showed 2–4-fold changes (increased or decreased) in the IC relative to the DCIS and control samples. When multiple peptides were suitable for peak

Table 3. Peak areas of selected proteins in the pooled samples.^a

GO attributes	No. of proteins ^b	Selected protein	Selected peptide	Control		DCIS		Invasive	
				PA ^c	RSD% ^d	PA	RSD%	PA	RSD%
Lipid transport	15	Apolipoprotein C III	DALSSVQESQVAQQAR	0.95–1.05	7	1.32–2.30	15	1.85–2.01	6
Cell ion homeostasis	7	Ceruloplasmin	DLYSGLIGPLIVCR	1.05–0.95	7	1.14–1.53	15	1.91–2.32	10
Cell growth and maintenance	125	Prothrombin	LAVTTHGLPCLAWASAQAK	0.86–1.14	19	1.64–2.05	16	1.46–1.73	12
Fibrinogen complex	3	Fibrinogen β -chain	NYCGLPGEYWLGNNDK	0.9–1.10	15	0.12–0.14	8	ND ^e	0
Lipid metabolism	31	α -1-Antichymotrypsin	LYGSEAFATDFQDSAAAK	0.80–1.20	29	1.12–1.34	10	1.47–2.61	28
Wide-spectrum protease inhibitor	41	Pregnancy zone protein	AGAFCLSEDAAGLIGSSTASLR	0.93–1.08	11	1.55–2.13	16	3.83–4.50	8
Cell fraction ^f	2	Neuropilin-1	FKIGYSNNGSDWK	0.83–1.10	19	0.13–0.16	4	ND	NA

^a In each pool, 5 samples were combined. All analyses were performed in triplicate.

^b Total number of proteins identified in the GO association.

^c Peak area (PA) was normalized using bovine fetuin as an internal standard. Average PA of the control triplicates was nominally set at 1 unit, and the range of values observed for triplicate analyses is listed.

^d Relative standard deviation (RSD%) of the triplicate analyses.

^e ND, not detected; NA, not applicable.

^f Includes plasma membrane fraction.

area quantification, the area measurement was shown to be consistent in replicates (Table 3). An additional criterion for protein representative selection in a given ontological association was availability of literature suggesting an association with breast cancer.

Discussion

M-LAC DETECTION OF LOW-ABUNDANCE SERUM PROTEINS

Our M-LAC system displayed ease of implementation and high throughput, enabling parallel sample processing and rapid capture and elution, e.g., 8 h for a 15-sample set. Samples of HER-2, a glycosylated protein shown to be correlated with tumor load and clinical stage (18), were selected to provide a suitable test for the MS methodology. We were unable to detect this protein without the M-LAC depletion step. We used an ELISA to confirm proteomic identification and reported values for HER-2 of 8–24 $\mu\text{g}/\text{L}$ in the individual patient samples. Compared with a high-abundance protein such as serum albumin (present at $\sim 50 \text{ g}/\text{L}$), the HER-2 marker is present at a concentration $< 1/10^6$ (by weight). Using a similar LC-MS analysis, we previously measured a dynamic range of 1 in 10^4 for a standard protein (growth hormone) added to plasma (13); thus, the M-LAC step represents a marked improvement in depth of analysis.

GO ANNOTATION

Our approach, which is based on the understanding that the serum glycoproteome originates from a wide range of tissues and proteins can be annotated to various biological pathways (19), clusters glycoproteins on the basis of their GO attributes to allow direct comparison of the ranking of such groupings in control and disease samples. A concern in the use of these associations is that some GO assignments are based on computerized sequence similarity

searches on data that are not well curated. We minimize such errors in the database by performing our own curation of proteins of interest that met our threshold for differential production between the sample groups. An additional benefit of such a focused analysis is a marked decrease in the number of proteins to be examined relative to the initial list of identified proteins.

The study revealed that the following ontology-based associations showed consistent differences in the ranks of attributes between the control and disease sets:

- Increased in IC: lipid transport and metabolism, cell ion homeostasis, and wide-spectrum protease inhibitor.
- Increased in DCIS: actin binding and cell growth and maintenance.
- Decreased in DCIS and IC: cell fraction, circulation, and fibrinogen complex.

Although the exact number of identified proteins in a given attribute is a function of the dynamic range of the triplicate measurements, this variable can serve as another valuable discriminator between the control and disease samples. For example, in a total of 18 identified proteins associated with actin-binding function, 11 were found only in the disease samples; similarly, 49 of 125 identified proteins involved in cell growth and maintenance were also found only in the disease samples (Table 3). As described in "Results", an increased ranking of ontology associations of interest was also supported by quantitative measurement of key protein members (Table 3). Examples of appreciable up-regulation in IC samples included ceruloplasmin, pregnancy-zone protein, and α -1-antichymotrypsin; examples of down-regulation included fibrinogen β -chain and neuropilin-1. Peak-area changes were consistent with changes in the corresponding ranking of the GO attributes. Proteins listed in Table 3 range from higher-abundance blood proteins, such as

apolipoprotein C-III, to lower-abundance proteins, such as pregnancy-zone protein; in all cases, protein identification was based on 2 or more peptides.

When we repeated the analysis with 15 individual patient samples to assess the variability in abundance of the proteins listed in Table 3 in different individuals and to examine the overlap between the control and disease (DCIS and IC) samples (Fig. 2), the results for disease samples revealed an increasing trend for protein groups a–e (apolipoprotein C-III, ceruloplasmin, pregnancy zone protein, α -1-antichymotrypsin, and prothrombin) and a decreasing trend for protein groups f and g (neuropilin-1 and fibrinogen β -chain). As expected, values for various control and patient samples overlapped somewhat, and, in general, the DCIS samples showed smaller differences from the controls than did the IC samples. The results were consistent with the results for the pooled samples.

EVALUATION OF POTENTIAL BIOMARKERS WITH DISEASE AND PATIENT INFORMATION

Cancer is a complex set of diseases that can lead to systemic changes (20, 21). Serum proteomic detection of early cancer would be enhanced by the identification of specific metabolic or immunological alterations that do not depend on the presence of notable tumor mass (22).

This study provides evidence that serum samples from patients with cancer show increases in specific GO associations, including lipid metabolism and transport, cell ion homeostasis, and wide-spectrum protease inhibitors. Other changes in metabolism are indicated by decreases in proteins in the coagulation and fibrinogen complex category.

Tumor growth and metastasis have been associated with high-fat diet and specific metabolic control by di-

etary substrates (23, 24). Clinical findings have supported these observations; for example, studies have demonstrated increased concentrations of apolipoprotein B (25) and apolipoprotein C III (26) in patients with breast cancer. For the purposes of this study, we selected apolipoprotein C-III as an example of increased serum lipid transport proteins in disease samples; however, other apolipoproteins, such as Lp(a), gave similar results. Ceruloplasmin is another example of an upregulated transport protein in the current study, and this marker was also found to be increased in another study of metastatic breast cancer (27, 28).

The significantly higher serum concentrations of protease inhibitors (Table 3 and Fig. 2) are in agreement with other studies of breast cancer (29). One study showed that α -1-antichymotrypsin was synthesized by MCR-7 breast cancer cells (30, 31) and that synthesis was stimulated by epidermal growth factor (32) and insulin-like growth factors (33), which may protect the tumor from invading leukocytes (34, 35).

Patients with cancer exhibit changes in coagulation and fibrinolysis, (36) and increased serum concentrations of prothrombin (37). Thrombin induces production of vascular endothelial growth factor mRNA and is a survival factor for malignant cells (38). In addition to increased prothrombin in breast cancer sera reported here (Table 3), other proteins within the same ontological association group showed similar changes, including transthyretin and plasminogen. Neuropilin was decreased in the disease samples (Table 3 and Fig. 2), a finding consistent with an observation of deficient concentrations in breast carcinoma cells (39). Neuropilin-1 exerts an apoptotic effects in endothelial cells (40) and activation of a p53-dependent apoptotic pathway, (41, 42).

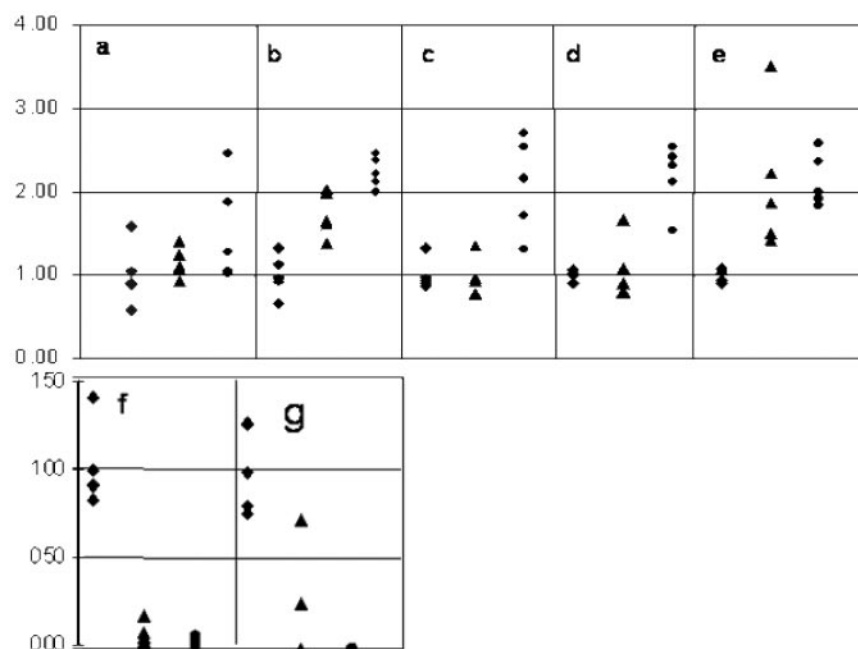


Fig. 2. Amount changes of the proteins listed in Table 3 in individual patient samples.

◆, control; ▲, DCIS; ●, IC. The protein identifications are as follows: (a), apolipoprotein-C III; (b), ceruloplasmin; (c), pregnancy zone protein; (d), α -1-antichymotrypsin; (e), prothrombin; (f), neuropilin-1; (g), fibrinogen β -chain. Peak areas were normalized by use of the mean peak area of the 5 control samples to indicate changes in protein amount between disease and control samples.

For any given disease, an individual protein marker may be suitable for only a portion of the patient population (43), and a panel of multiple biomarkers may be required to control bias in disease diagnosis. Such a panel increases the complexity of subsequent data analysis, a problem further magnified in genomic and proteomic studies, that discover large numbers of potential protein markers (44). Such complexity may be decreased by grouping potential biomarkers by GO terms (45, 46).

Our results indicate that isolation of the serum glycoproteome by M-LAC coupled with LC-MS/MS and the use of GO associations facilitates characterization of a large panel of candidate markers, which can then be evaluated in a particular patient population. We hope to stimulate additional interest in glycosylation and further development of technology for characterizing this important posttranslation modification. Although it was not the focus of this study, we developed the M-LAC approach to monitor shifts in glycosylation by elution of the bound glycoproteins with 3 separate displacer steps (47). Larger studies are required to assess the functional importance of changes seen in different proteins and pathways. We and others (48) believe that MS has great potential in the search for multiple novel disease-associated biomarkers, enabling identification of large panels of proteins not readily accessible by current ELISA technology.

We thank ThermoElectron and GE Healthcare for instrumentation support. We also thank Drs. Shiaw-Lin Wu and Barry Karger for helpful comments. Contribution number 874 from Barnett Institute.

References

- Mollick JA, Carlson RW. Rational surveillance programs for early stage breast cancer patients after primary treatment. *Breast Dis* 2004;21:47–54.
- Houssami N, Irwig L, Simpson JM, McKessar M, Blome S, Noakes J. The influence of knowledge of mammography findings on the accuracy of breast ultrasound in symptomatic women. *Breast J* 2005;11:167–72.
- Shin BK, Wang H, Hanash S. Proteomics approaches to uncover the repertoire of circulating biomarkers for breast cancer. *J Mammary Gland Biol Neoplasia* 2002;7:407–13.
- Li J, Zhang Z, Rosenzweig J, Wang YY, Chan DW. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem* 2002;48:1296–304.
- Pieper R, Su Q, Gatlin CL, Huang ST, Anderson NL, Steiner S. Multi-component immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome. *Proteomics* 2003;3:422–32.
- Rapkiewicz AV, Espina V, Petricoin EF 3rd, Liotta LA. Biomarkers of ovarian tumours. *Eur J Cancer* 2004;40:2604–12.
- Yang Z, Hancock WS. Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *J Chromatogr A* 2004;1053:79–88.
- Esteva FJ, Cheli CD, Fritsche H, Fornier M, Slamon D, Thiel RP, et al. Clinical utility of serum HER2/neu in monitoring and prediction of progression-free survival in metastatic breast cancer patients treated with trastuzumab-based therapies. *Breast Cancer Res* 2005;7:436–43.
- van Gisbergen KP, Aarnoudse CA, Meijer GA, Geijtenbeek TB, van Kooyk Y. Dendritic cells recognize tumor-specific glycosylation of carcinoembryonic antigen on colorectal cancer cells through dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin. *Cancer Res* 2005;65:5935–44.
- Martinez-Trufero J, de Lobera AR, Lao J, Puertolas T, Artal-Cortes A, Zorrilla M, et al. Serum markers and prognosis in locally advanced breast cancer. *Tumori* 2005;91:522–30.
- Carey LA, Metzger R, Dees EC, Collichio F, Sartor CI, Ollila DW, et al. American joint Committee on Cancer tumor-node-metastasis stage after neoadjuvant chemotherapy and breast cancer outcome. *J Natl Cancer Inst* 2005;97:1137–42.
- Strittmatter EF, Ferguson PL, Tang K, Smith RD. Proteome analyses using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry. *J Am Soc Mass Spectrom* 2003;14:980–91.
- Wu SL, Amato H, Biringer R, Choudhary G, Shieh P, Hancock WS. Targeted proteomics of low-level proteins in human plasma by LC/MSn: using human growth hormone as a model system. *J Proteome Res* 2002;1:459–65.
- Wu SL, Kim J, Hancock WS, Karger B. Extended range proteomic analysis (ERPA): a new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of β -casein and epidermal growth factor receptor (EGFR). *J Proteome Res* 2005;4:1155–70.
- Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999;17:994–9.
- Cagney G, Emili A. De novo peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging. *Nat Biotechnol* 2002;20:163–70.
- Kendzioriski CM, Zhang Y, Lan H, Attie AD. The efficiency of pooling mRNA in microarray experiments. *Biostatistics* 2003;4:465–77.
- Yuan P, Xu BH, Chu DT. Correlation between serum HER-2 oncoprotein and patients with breast cancer. *Chin Med Sci J* 2004;19:212–5.
- Ping P, Vondriska TM, Creighton CJ, Gandhi T, Yang Z, Menon R, et al. A functional annotation of subproteomes in human plasma. *Proteomics* 2005;5:3506–19.
- Caine GJ, Stonelake PS, Rea D, Lip GY. Coagulopathic complications in breast cancer. *Cancer* 2003;98:1578–86.
- Schor SL, Schor AM. Phenotypic and genetic alterations in mammary stroma: implications for tumour progression. *Breast Cancer Res* 2001;3:373–9.
- Tchabo NE, Liel MS, Kohn EC. Applying proteomics in clinical trials: assessing the potential and practical limitations in ovarian cancer. *Am J Pharmacogenomics* 2005;5:141–8.
- Rose DP, Connolly JM. Regulation of tumor angiogenesis by dietary fatty acids and eicosanoids. *Nutr Cancer* 2000;37:119–27.
- Cunningham BA, Moncur JT, Huntington JT, Kinlaw WB. "Spot 14" protein: a metabolic integrator in normal and neoplastic cells. *Thyroid* 1998;8:815–25.
- Launonen V, Laake K, Huusko P, Niederacher D, Beckmann MW, Barkardottir RB, et al. European multicenter study on LOH of APOC3 at 11q23 in 766 breast cancer patients: relation to clinical variables. *Breast Cancer Somatic Genetics Consortium. Br J Cancer* 1999;80:879–82.
- Lane DM, Boatman KK, McConathy WJ. Serum lipids and apolipoproteins in women with breast masses. *Breast Cancer Res Treat* 1995;34:161–9.
- Schapiro DV, Schapiro M. Use of ceruloplasmin levels to monitor

- response to therapy and predict recurrence of breast cancer. *Breast Cancer Res Treat* 1983;3:221–4.
28. Linder MC, Moor JR, Wright K. Ceruloplasmin assays in diagnosis and treatment of human lung, breast, and gastrointestinal cancers. *J Natl Cancer Inst* 1981;67:263–75.
29. Larsen MB, Stephens RW, Brunner N, Nielsen HJ, Engelholm LH, Christensen IJ, et al. Quantification of tissue inhibitor of metalloproteinases 2 in plasma from healthy donors and cancer patients. *Scand J Immunol* 2005;61:449–60.
30. Tamir S, Kadner SS, Katz J, Finlay TH. Regulation of antitrypsin and antichymotrypsin synthesis by MCF-7 breast cancer cell sublines. *Endocrinology* 1990;127:1319–28.
31. Bergman D, Kadner SS, Cruz MR, Esterman AL, Tahery MM, Young BK, et al. Synthesis of α 1-antichymotrypsin and α 1-antitrypsin by human trophoblast. *Pediatr Res* 1993;34:312–7.
32. Finlay TH, Tamir S, Kadner SS, Cruz MR, Yavelow J, Levitz M. α 1-Antitrypsin- and anchorage-independent growth of MCF-7 breast cancer cells. *Endocrinology* 1993;133:996–1002.
33. Confort C, Rochefort H, Vignon F. Insulin-like growth factors (IGFs) stimulate the release of α 1-antichymotrypsin and soluble IGF-II/mannose 6-phosphate receptor from MCF7 breast cancer cells. *Endocrinology* 1995;136:3759–66.
34. Gendler SJ, Dermer GB, Silverman LM, Tokes ZA. Synthesis of α 1-antichymotrypsin and α 1-acid glycoprotein by human breast epithelial cells. *Cancer Res* 1982;42:4567–73.
35. Gendler SJ, Tokes ZA. Expression of an active proteinase inhibitor, α 1-antichymotrypsin, by human breast epithelial cells. *Biochim Biophys Acta* 1986;882:242–53.
36. Korte W. Changes of the coagulation and fibrinolysis system in malignancy: their possible impact on future diagnostic and therapeutic procedures. *Clin Chem Lab Med* 2000;38:679–92.
37. Schiller H, Bartscht T, Arlt A, Zahn MO, Seifert A, Bruhn T, et al. Thrombin as a survival factor for cancer cells: thrombin activation in malignant effusions in vivo and inhibition of idarubicin-induced cell death in vitro. *Int J Clin Pharmacol Ther* 2002;40:329–35.
38. Yamahata H, Takeshima H, Kuratsu J, Sarker KP, Tanioka K, Wakimaru N, et al. The role of thrombin in the neo-vascularization of malignant gliomas: an intrinsic modulator for the up-regulation of vascular endothelial growth factor. *Int J Oncol* 2002;20:921–8.
39. Bachelder RE, Crago A, Chung J, Wendt MA, Shaw LM, Robinson G, et al. Vascular endothelial growth factor is an autocrine survival factor for neuropilin-expressing breast carcinoma cells. *Cancer Res* 2001;61:5736–40.
40. Barr MP, Byrne AM, Duffy AM, Condrón CM, Devocelle M, Harriott P, et al. A peptide corresponding to the neuropilin-1-binding site on VEGF(165) induces apoptosis of neuropilin-1-expressing breast tumour cells. *Br J Cancer* 2005;92:328–33.
41. Wang WJ, Kuo JC, Yao CC, Chen RH. DAP-kinase induces apoptosis by suppressing integrin activity and disrupting matrix survival signals. *J Cell Biol* 2002;159:169–79.
42. Levy D, Plu-Bureau G, Decroix Y, Hugol D, Rostene W, Kimchi A, et al. Death-associated protein kinase loss of expression is a new marker for breast cancer prognosis. *Clin Cancer Res* 2004;10:3124–30.
43. Espina V, Geho D, Mehta AI, Petricoin EF 3rd, Liotta LA, Rosenblatt KP. Pathology of the future: molecular profiling for targeted therapy. *Cancer Invest* 2005;23:36–46.
44. Carr KM, Rosenblatt K, Petricoin EF, Liotta LA. Genomic and proteomic approaches for studying human cancer: prospects for true patient-tailored therapy. *Hum Genomics* 2004;1:134–40.
45. Taghavi M, Siermala M, Lehtinen TO, Vihinen M. Dynamic covariation between gene expression and proteome characteristics. *BMC Bioinformatics* 2005;6:215.
46. Omenn GS. Exploring the Human Plasma Proteome. *Proteomics* 2005;5:3223–5.
47. Yang Z, WS. H. Monitoring glycosylation pattern changes of glycoproteins using multi-lectin affinity chromatography. *J Chromatogr A* 2005;1070:57–64.
48. Clarke W, Zhang Z, Chan DW. The application of clinical proteomics to cancer and other diseases. *Clin Chem Lab Med* 2003;41:1562–70.