High Throughput Quantitative Analysis of Serum Proteins Using Glycopeptide Capture and Liquid Chromatography Mass Spectrometry*[S]

Hui Zhang‡§, Eugene C. Yi‡, Xiao-jun Li‡, Parag Mallick‡, Karen S. Kelly-Spratt¶, Christophe D. Masselon‖, David G. Camp II‖, Richard D. Smith‖, Christopher J. Kemp¶, and Ruedi Aebersold‡

It is expected that the composition of the serum proteome can provide valuable information about the state of the human body in health and disease and that this information can be extracted via quantitative proteomic measurements. Suitable proteomic techniques need to be sensitive, reproducible, and robust to detect potential biomarkers below the level of highly expressed proteins, generate data sets that are comparable between experiments and laboratories, and have high throughput to support statistical studies. Here we report a method for high throughput quantitative analysis of serum proteins. It consists of the selective isolation of peptides that are *N*-linked glycosylated in the intact protein, the analysis of these now deglycosylated peptides by liquid chromatography electrospray ionization mass spectrometry, and the comparative analysis of the resulting patterns. By focusing selectively on a few formerly *N*-linked glycopeptides per serum protein, the complexity of the analyte sample is significantly reduced and the sensitivity and throughput of serum proteome analysis are increased compared with the analysis of total tryptic peptides from unfractionated samples. We provide data that document the performance of the method and show that sera from untreated normal mice and genetically identical mice with carcinogen-induced skin cancer can be unambiguously discriminated using unsupervised clustering of the resulting peptide patterns. We further identify, by tandem mass spectrometry, some of the peptides that were consistently elevated in cancer mice compared with their control littermates. ***Molecular & Cellular Proteomics 4: 144–155, 2005.***

There is growing interest in testing the hypothesis that the serum[1] proteome contains protein biomarkers that are useful for classifying the physiological or pathological status of an individual. Such markers are expected to be useful for the prediction, detection, and diagnosis of disease as well as to follow the efficacy, toxicology, and side effects of drug treatment (1). The idea of reading diagnostic or prognostic signatures from human body fluids is neither new nor original. Early attempts using high resolution two-dimensional gel electrophoresis were described more than 2 decades ago (2–4). Renewed interest in this idea has emerged due to recent advances in proteomic technologies (5), intriguing initial results from analyzing serum protein patterns using mass spectrometry (1), and the clinical validation and use of a number of diagnostic disease markers including CA125 for ovarian cancer, prostate-specific antigen for prostate cancer, and carcinoembryonic antigen for colon, breast, pancreatic, and lung cancer (6).

A number of new approaches that differ from the traditional two-dimensional gel electrophoresis method for the discovery of protein biomarkers in serum have recently been described (1). These include surface-enhanced laser desorption ionization mass spectrometry (SELDI-MS)[2] (7), liquid chromatography tandem mass spectrometry (LC-MS/MS) of serum proteome digests (8–10), two- or three-dimensional (chromatography/gel electrophoresis) protein separation analyzed by differential fluorescent staining (11, 12), fractionation of the serum proteome on surface-modified magnetic beads followed by matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) (13), and combinations and variations of these approaches.

From the ‡Institute for Systems Biology, Seattle, Washington 98103, ¶Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, ‖Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352, and **Swiss Federal Institute of Technology (ETH) Zurich, and Faculty of Natural Sciences, University of Zurich, CH-8093 Zurich, Switzerland

Received, July 12, 2004, and in revised form, November 30, 2004
Published, MCP Papers in Press, December 17, 2004, DOI 10.1074/mcp.M400090-MCP200

[1] In this paper, the term serum is used to indicate serum or plasma.

[2] The abbreviations used are: SELDI, surface-enhanced laser desorption ionization; MS, mass spectrometry; LC, liquid chromatography; ESI, electrospray ionization; MS/MS, tandem mass spectrometry; MALDI, matrix-assisted laser desorption ionization; CID, collision-induced dissociation; TOF, time-of-flight; QTOF, quadrupole time-of-flight; CV, coefficient of variance; DMBA, 7,12-dimethylbenz[*a*]anthracene; HPLC, high performance liquid chromatography.

Any study of the serum proteome is confronted with the peculiar properties of serum samples. First, human blood serum is assumed to consist of minimally tens of thousands of different protein species that span a concentration range of an estimated 10 orders of magnitude (14). Second, the serum proteome is dominated by a few highly abundant proteins, *i.e.* the 22 most abundant human serum proteins combined constitute 99% of total protein mass (9). Indeed almost one-half of total serum protein mass is represented by just one protein, albumin. Third, many of the serum proteins show complex two-dimensional electrophoretic patterns, suggesting that they are extensively post-translationally modified with glycosylation apparently being the major source of protein heterogeneity (14). In fact when protein spots from two-dimensional electropherograms of serum were systematically identified by mass spectrometry, five to seven protein spots on average were identified as products of the same gene (15). Fourth, the serum proteome varies over time in an individual and among individuals in a population.

Useful platforms for serum proteome analysis should thus have minimally the following properties: first, sufficient analytical depth to reliably detect relatively low abundance proteins; second, quantitative accuracy to determine changes in the proteome pattern; third, reproducibility and robustness to detect disease-specific changes in a background of pattern changes unrelated to disease; fourth, the ability to identify distinct peptides for their cross-validation on different analytical platforms and comparison of results obtained from different research groups, studies, and diseases; and fifth, high sample throughput to support studies with sufficient statistical power.

Here we describe a new method for quantitative serum proteome analysis. It is based on the selective isolation of those peptides from serum proteins that are *N*-linked glycosylated in the native protein and the analysis of the complex peptide mixture representing the now deglycosylated forms of these peptides by LC-MS and MS/MS. By selectively isolating this subset of peptides, the procedure achieves a significant reduction in analyte complexity at two levels: first, a reduction of the total number of peptides due to the fact that every serum protein on average only contains a few *N*-linked glycosylation sites, and second, a reduction of pattern complexity by removing the oligosaccharides that contribute significantly to the peptide pattern heterogeneity. We go on to show that this method is reproducible and achieves increased analytical depth and higher throughput compared with the analysis of samples without selective analyte enrichment. Furthermore we demonstrate that, in a controlled experiment, peptide patterns distinguishing the serum proteome of cancer-bearing mice from genetically identical untreated normal mice could be detected, and discriminatory peptides could be subsequently identified. At present, this method affords one of the most comprehensive routine and high throughput analyses of the serum proteome. We therefore believe that it will find broad application in serum marker discovery research.

EXPERIMENTAL PROCEDURES

Materials and Reagents

For all chromatographic steps, HPLC grade reagents were purchased from Fisher Scientific (Pittsburgh, PA, USA). Peptide-*N*-glycosidase F was from New England Biolabs (Beverly, MA). Hydrazide resin was from Bio-Rad. All other chemicals were purchased from Sigma.

Chemical Induction of Mouse Skin Tumors

Male mice of strain NIHO1a were subjected to the two-stage skin carcinogenesis protocol (16). Five littermates were used: two were untreated, and three were treated with carcinogen. The shaved backs of three 8-week-old mice were treated with a single dose of the carcinogen 7,12-dimethylbenz[*a*]anthracene (DMBA) (Sigma; 25 mg in 200 ml of acetone). Initiated cells were promoted with 12-*O*-tetradecanoylphorbol-13-acetate twice a week for 15 weeks, giving rise to papillomas that were hyperplastic, well differentiated, benign lesions consisting of keratinocytes together with stromal tissue. Papillomas appeared as early as 8 weeks after DMBA initiation and continued to grow for the next several months. A small percentage of these benign papillomas progressed to squamous cell carcinomas. At week 22 after DMBA initiation, all mice were sacrificed, and whole blood was collected by heart puncture with a 21-gauge needle and 1-ml syringe. Blood was allowed to clot for 1 h at room temperature. Sera were collected by centrifugation at 3000 rpm. The untreated mice contained no tumors, while the DMBA/12-*O*-tetradecanoylphorbol-13-acetate-treated mice each had at least one carcinoma as confirmed by histological analysis.

Preparation of Peptide Samples for Mass Spectrometry Analysis

Formerly *N*-linked glycosylated peptides were isolated and labeled using the *N*-linked glycopeptide capture procedure as described previously (17). Proteins from 100 μl of serum were used in isolation and isotope labeling of formerly *N*-linked glycopeptides, and peptides from 5 μl of original serum were used in each mass spectrometry analysis.

To prepare tryptic peptides from serum proteins, proteins from 1 μl (80 μg) of mouse serum were denatured in 20 μl of 8 M urea, 0.4 M NH_4HCO_3 for 30 min at room temperature. The proteins were diluted four times with water after which 1 μg of trypsin was added, and the proteins were digested at 37 °C overnight. The peptides were then reduced by adding 8 mM Tris(2-carboxyethyl)phosphine (Pierce) at room temperature for 30 min and alkylated by adding 10 mM iodoacetamide at room temperature for 30 min. The peptides were dried and resuspended in 0.4% acetic acid. Peptides from 0.05 μl of original serum (4 μg of original serum proteins) were used for each LC-MS analysis.

Analysis of Peptides by Mass Spectrometry

The peptides and proteins were identified using MS/MS analysis using an LCQ ion trap mass spectrometer (Thermo Finnigan, San Jose, CA) as described previously (18). For quantitative analysis of peptides using LC-MS, an ESI-QTOF mass spectrometer (Waters, Beverly, MA) was used. In both systems, peptides isolated from 5 μl of serum sample using the glycopeptide capture method were injected into a home-made peptide trap packed with Magic C_{18} resin (Michrome Bioresources, Auburn, CA) using a FAMOS autosampler (DIONEX, Sunnyvale, CA) and then passed through a 10-cm $\times$ 75-μm-inner diameter microcapillary HPLC column packed with Magic C_{18} resin (Michrome Bioresources). The effluent from the microcapillary HPLC column entered a home-built electrospray ionization

source in which peptides were ionized and passed directly into the respective mass spectrometer. The C_{18} peptide trap cartridge, μ-ESI emitter/microcapillary HPLC pulled tip column combination, a high voltage line for ESI, and the waste line were each connected to separate ports of a four-port union (Upchurch Scientific, Oak Harbor, WA) constructed entirely out of polyetheretherketone (19). A linear gradient of acetonitrile from 5–32% over 100 min at a flow rate of ~300 nl/min was applied. During the LC-MS mode, data were acquired with a profile mode in the mass range scan between *m/z* 400 and 2000 with 3.0-s scan duration and 0.1-s interscan. After completion of the LC/MS runs, inclusion peptide mass lists were created from data analysis software. The inclusion lists were then used for targeted LC-MS/MS analysis for peptide/protein identifications with the remaining of samples.

Data Analysis

For ESI-QTOF data analysis, a suite of software tools was developed or optimized in-house to analyze LC-MS data for this project and will be published separately.[3] The software tools use LC-MS data generated by ESI-QTOF analysis of formerly *N*-linked glycopeptides from serum samples and sequentially perform the following tasks to determine peptides that are of different abundance in cancer and normal mice, respectively.

Peptide List—A list of peptide peaks was generated from each LC-ESI-MS run. The tool performing this operation was a straightforward extension of a previous tool for the analysis of LC-MALDI-MS data (20). That tool was modified to take into account the fact that in ESI-MS peptides are observed in different charge states. Peaks were selected if the signal to noise ratio exceeded 2.

Peptide Alignment—Peptides detected in individual LC-MS patterns were aligned mainly based on peptide mass. The retention time was then used to align peptides with the same *m/z* value. The software tool accounted for shifts in the retention time in different LC-MS analyses during peptide alignment. Peptide alignment was facilitated by the following factors: i) the glycopeptide capture procedure significantly simplifies the sample complexity, ii) the high mass accuracy achieved in the ESI-QTOF instrument, and iii) the optimized HPLC system that produced highly consistent and reproducible peptide patterns. In the mouse studies, peptides that appeared at least in two of three analyses in either group were selected for further quantitative analysis.

Peptide Abundance Ratio—An abundance ratio of matched peptides in different samples was determined for each peptide peak using the same method as described in the ASAPRatio software tool developed for LC-ESI-MS/MS data (21). Briefly the software uses spectra from multiple LC-MS analyses of a peptide peak (with same mass-to-charge ratio (*m/z*), same charge state, and close retention time) and calculates one ratio for each peptide peak. In the present study, ratios calculated for different charge states of the same peptide were not combined. The algorithm also estimates a noise background level in each spectrum and subtracts that value from the signal intensities when calculating the peak area.

Clustering Analysis—The lists of matched peptides with their relative signal intensities were subjected to unsupervised hierarchical clustering (22) to identify peptides distinguishing cancer samples from normal samples. Prior to clustering, the data were transformed to log value, and the mean intensity of each peptide across all samples was normalized. Peptides present at least in 50% of the total samples were used for clustering analysis.

[3] X. J. Li, E. C. Yi, H. Zhang, and R. Aebersold, manuscript in preparation.

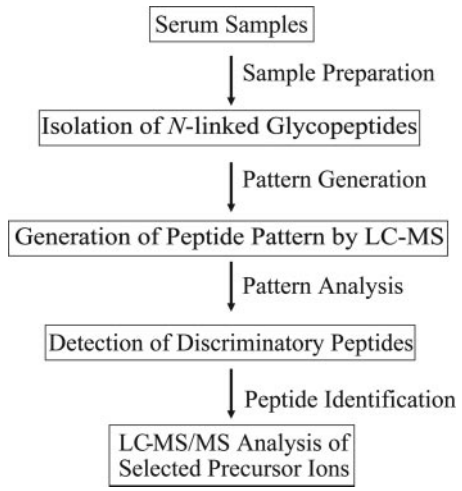

FIG. 1. **Schematic presentation for high throughput analysis of serum proteins using glycopeptide capture and LC-MS.**

RESULTS

Principle of the Method

The objective of the method is the generation of reproducible peptide patterns representing the serum proteome, leading to the detection of peptides that discriminate between related groups of proteomes and the subsequent identification of these discriminatory peptides. The method is schematically illustrated in Fig. 1 and consists of four steps.

Sample Preparation—Peptides that contain *N*-linked carbohydrates in the native protein were isolated in their deglycosylated form using a recently described solid-phase capture-and-release method (17).

Pattern Generation—Isolated peptides were analyzed by LC-MS to generate three-dimensional (retention time, *m/z*, and intensity) patterns.

Pattern Analysis—Peptide patterns obtained from different samples were compared, and the discriminatory peptides were determined.

Peptide Identification—Discriminatory peptides and the proteins from which they originated were identified by tandem mass spectrometry and sequence data base searching.

Selectivity of the N-Linked Glycopeptide Capture Method

To determine the selectivity of the glycopeptide capture method for serum protein analysis, serum samples from four genetically identical mouse littermates were individually processed using the *N*-linked glycopeptide capture-and-release method, and the isolated peptides were analyzed by LC-MS/MS. The resulting collision-induced dissociation (CID) spectra were searched against the mouse International Protein Index sequence data base (Version 1.24), and the data base search results were further statistically analyzed using the PeptideProphet software tool (23). From four LC-MS/MS analyses of the mouse sera, 1722 CID spectra resulted in peptide identifications from the data base search with peptide prob-

TABLE I
Total number of peptide identifications, unique peptides, and unique proteins and the proportion of each that contain NX(T/S) motif

	Total peptides	Peptides containing NX(T/S) motif	Percentage of motif-containing peptides
			%
Number of identifications	1722	1611	93.6
Number of unique peptides identified	319	261	81.8
Number of unique proteins identified	93	87	93.5

ability scores of at least 0.99 (corresponding to a false positive error rate of 0.0007 (23)). The identified sequences were then examined for the presence of the known consensus *N*-linked glycosylation motif (N*X*(T/S) where *X* = any amino acid except proline). The number of proteins represented by the selected peptides were determined using INTERACT (24). The full list of peptides and proteins identified are given in Supplemental Table 1 on line, and the CID spectra of these peptides are given in Supplemental Fig. 1 on line. The number of identified proteins and peptides are summarized in Table I. A total of 319 unique peptides were identified, representing 93 unique proteins. 93.6% of the identifications, 81.8% of unique peptides, and 93.5% of identified proteins contained the consensus *N*-linked glycosylation motif (Table I).

The peptides identified as not containing the consensus *N*-linked glycosylation motif can be grouped into two pools. The first contains peptides that are correctly identified, and the second contains peptides that are incorrectly identified by SEQEST search (false positives). In the present analysis, the false positive error rate was estimated by the PeptideProphet statistical model. To further estimate the selectivity of the isolation method, we calculated the fraction of peptides identified without the consensus *N*-linked glycosylation motif as a function of the PeptideProphet probability values. The data are shown in Fig. 2. It is apparent that the fraction of peptides without the N*X*(S/T) motif decreases as the stringency of the identification criteria increases. Concurrently, as expected, the number of false positive peptide identifications also decreases. Significantly and consistent with the data in Table I, the percentage of peptides without the N*X*(S/T) motif plateaus out at ~6.4% as the false positive error rate approaches 0. We therefore conclude that the peptide isolation method used has a selectivity that is not lower than 93.6%.

Reduction in the Complexity of Serum-derived Peptide Mixtures Obtained via the Glycopeptide Capture-and-release Method

We used the data described above to estimate the reduction in sample complexity achieved via the glycopeptide capture-and-release method. A total of 93 proteins was identified collectively from the four serum samples analyzed (Supplemental Table 1 on line). Disregarding the complexity caused by protein post-translational modifications, the 93 identified proteins were expected to generate an average of 28.8 tryptic peptides per protein (Supplemental Table 1, column A). Of

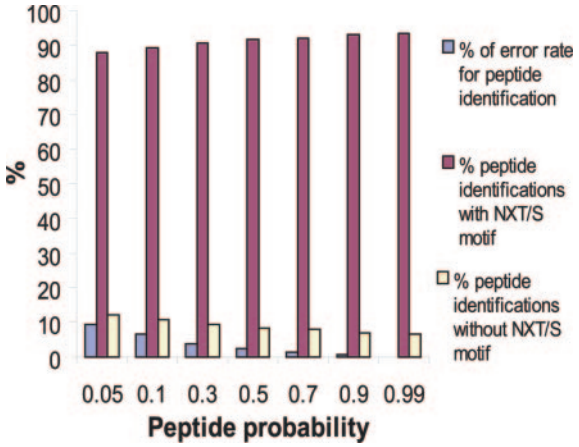

FIG. 2. **Ratio of peptides identified without the N*X*(S/T) glycosylation motif as a function of peptide identification stringency.** The fraction of peptides identified with (*red*) or without (*yellow*) the glycosylation consensus motif is shown for different PeptideProphet (23) probabilities. The false positive error rates estimated by PeptideProphet are indicated as *blue bars*.

these, 3.6 peptides on average contained the N*X*(S/T) motif and were therefore designated potentially *N*-linked glycosylated peptides (Supplemental Table 1, column B). Among the 93 identified proteins in this study, an average of 3.6 peptides representing 1.8 unique *N*-linked glycosylation sites per protein were actually identified (Supplemental Table 1, columns C and E). By comparing the number of unique *N*-linked glycosylation sites identified (Supplemental Table 1, column E) with the number of predicted peptides containing the consensus *N*-linked glycosylation motif (Supplemental Table 1, column B), we therefore found that 50% of the predicted glycosylated peptides had been detected. Interestingly an analysis of the actual occupancy rate of potential *N*-linked glycosylation sites in glycoproteins in the crystallographic data base showed ~65% site occupancy (25). Collectively these data indicate that the glycopeptide capture and release from serum proteins significantly reduces sample complexity and that the method captured a significant fraction of the potentially available *N*-linked glycosylated peptides.

To determine whether the increased sensitivity achieved by reducing sample complexity was sufficient to detect serum protein biomarkers of clinically relevant concentration, we related data obtained in this study to the concentrations of human serum marker proteins (26, 27). We are not aware of a direct comparison of the protein compositions between the

TABLE II
Peak intensities of formerly N-linked glycopeptides identified from mouse sera and the reported concentration of their corresponding proteins in human serum

Protein name	IPI number	Peptide sequences[a]	μg/ml	Intensity
Kallikrein B, plasma I	IPI00113057	R.IVGGTN#ASLGEWPWQVSLQVK.L	50	1.50×10^7
		K.LQTPLN#YTEFQKPICLPSK.A		3.30×10^7
Coagulation factor II	IPI00114206	R.CAMDLGVNYLGTVN#VTHTGIQCQLWR.S	20	1.30×10^7
		R.WVLTAAHCILYPPWDKN#FTENDLLVR.I		2.90×10^7
Coagulation factor V	IPI00117084	K.SN#ETALSPDLN#QTSPSM*STDR.S	20	1.50×10^6
Similar to carboxypeptidase N	IPI00119522	E.ITGSPVSN#LSAHIFSN#LSSLEK.L	35	1.10×10^8
		R.DGSDSAAM*VYN#SSQEWGLR.S		3.20×10^7
Epidermal growth factor receptor	IPI00121190	R.DCVSCQN#VSR.G		8.30×10^6
		R.DIVQNVFM*SN#M*SM*DLQSHPSSCPK.C		1.80×10^7
		K.DTLSIN#ATNIK.H		1.10×10^7
Coagulation factor XIII, β subunit	IPI00122117	K.EQETCLAPELEHGN#YSTTQR.T	10	5.30×10^6
		R.TYEN#GSSVEYR.C		8.40×10^6
Coagulation factor XII (Hageman factor)	IPI00125393	R.HN#QSCEWCQTLAVR.S	30	3.30×10^7
Interferon (α and β) receptor 2	IPI00132817	K.SGPPAN#YTLWYTVM*SK.D		1.70×10^7
Serum amyloid P-component	IPI00267939	K.LIPHLEKPLQN#FTLCFR.T	20	7.00×10^7
Average				2.70×10^7
Background				3.00×10^4
SNR[b]				8.99×10^2

[a] N#, *N*-linked glycosylation site; M*, oxidized methionine; peptide sequences between the two periods are identified by MS/MS.
[b] SNR, signal to noise ratio.

human and mouse serum proteomes. However, the serum two-dimensional maps of human and mouse are sufficiently similar to allow an approximate comparison of the concentrations of the proteins identified in this study between human and mouse (28). From the 93 proteins identified above (Supplemental Table 1), several proteins are known to be present in human serum at low μg/ml concentration (Table II). These include carboxypeptidase N and coagulation factors II, V, XII, and XIII. Except for epidermal growth factor receptor and serum amyloid P-component, none of the other proteins listed in Table II have been identified in the previous mouse two-dimensional map, suggesting that they are present at low abundance in mouse serum (28). To estimate the detection sensitivity, the peak intensities of the identified peptides from these proteins were calculated using the intensities of the chromatographic peaks at the charge states used for peptide identification. Examination of the peak intensities indicated an average peptide peak intensity of 2.7×10^7, which is ~900 times greater than the observed background signal for these experiments (Table II). This indicates that even without multidimensional separation, serum proteins at concentrations on the order of ng/ml may be detected by LC-MS of formerly *N*-linked glycopeptides.

Assessment of Reproducibility of LC-MS Patterns following Glycopeptide Capture and Release of Serum Proteins

Of the 319 peptides and 93 proteins identified by four LC-MS/MS analyses (Supplemental Table 1), 109 unique peptides and 52 unique proteins were identified from all four analyses. The number of peptides identified in all four LC-MS/MS runs is low compared with the total number of unique peptides identified (34.2%). We used the Pep3D software tool (29) to determine whether these observations were due to peptide undersampling in the LC-MS/MS experiment or whether they indicated poor pattern reproducibility. From the data shown in Fig. 3, the following is apparent. First, as expected, the LC-MS patterns of the peptides from individual mouse serum were consistent (Fig. 3, peptides presented by *black spots*). Second, due to the complexity of the sample, not all peptides in a given analysis were selected for MS/MS analyses and subsequently identified. Peptides selected for MS/MS analyses are marked with *blue*, and peptides identified from these analyses are marked with *red* or *green*, depending on the confidence of peptide identifications. Third, as far as could be determined from the difference between the number of identified peptides from MS/MS analysis and total peptides present in a sample from MS analysis, only a small portion of peptides, predominantly the high abundance peptides from each sample, were selectively identified by MS/MS analyses. Fourth, the differences between peptide/protein identifications by MS/MS analyses between different samples were caused mainly by the fact that only a fraction of total peptides was identified by MS/MS analysis in the data-dependent mode of operation (Fig. 3, *bottom panel*, where peptides present in both mouse samples but only identified in one of the samples are indicated by *arrows*). Collectively these results suggest that LC-MS analyses of glycopeptides isolated from genetically identical mice are reproducible. However, peptide/protein identifications using MS/MS analyses, due to peptide undersampling, results in a relatively small number of peptide identifications and a seemingly poor reproducibility of the method.

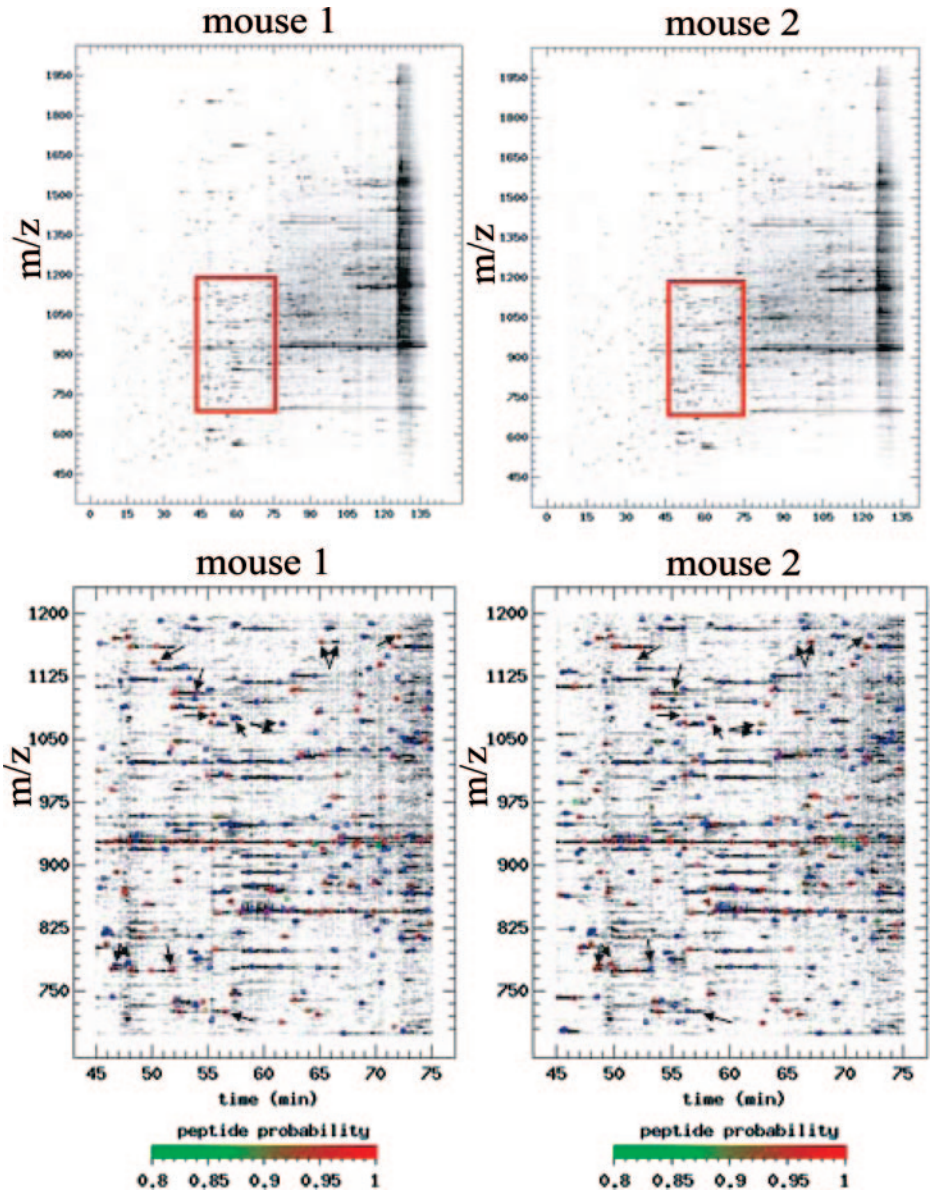

FIG. 3. **Peptide patterns and peptide identification reproducibility.** The Peptide3D (29) display of peptides detected by LC-MS/MS analysis of formerly *N*-linked glycopeptides isolated from mouse serum is shown. The detected peptides are indicated as *black dots*. The *bottom panel* represents an enlarged view of the *boxed area* in the *upper panel*. In the enlarged area, the peptides that were selected as precursors for MS/MS analysis are indicated as *blue dots* and peptides that were identified by data base search are indicated as *red* or *green dots* depending on the probability scores of identified peptides.

We then examined the reproducibility of the peptide patterns obtained by LC-MS. Four 50-μl aliquots from a single serum sample were processed in parallel to generate four isolates and then analyzed by LC-MS. First, to assess LC-MS reproducibility, we combined equal amounts of each isolate and analyzed the combined sample nine times by LC-MS using a 100-min reverse phase gradient. In-house-developed software tools were used to detect peaks in the resulting patterns, to measure peak intensity, and to align corresponding peptide peaks between multiple patterns.[3] From these data, we calculated the average intensity, standard deviation of intensity, and coefficient of variance (CV) for each peptide. A histogram of CV from the nine repeat analyses of identical samples by LC-MS is shown in *blue* in Fig. 4. The mean and median CV values observed in the nine repeat LC-MS analyses of the same sample were 28.3 and 21.8%, respectively. We next analyzed glycopeptides from the four individual isolates as described above to determine reproducibility with respect to peptide isolation. This data is shown in *red* in Fig. 4. The mean and median CV values for the four replicate sample preparations were 25.7 and 21.6%, respectively, and therefore comparable to the analogous values from repeat LC-MS analysis of identical samples. These values indicate that sample preparation does not significantly contribute to the variability of observed peptide patterns.

Application of the Method to Distinguish Sera from Normal and Skin Cancer-bearing Mice

To test the hypothesis that the serum proteome profiles from individuals in different physiological states can be differentiated, we applied the glycopeptide capture-and-release

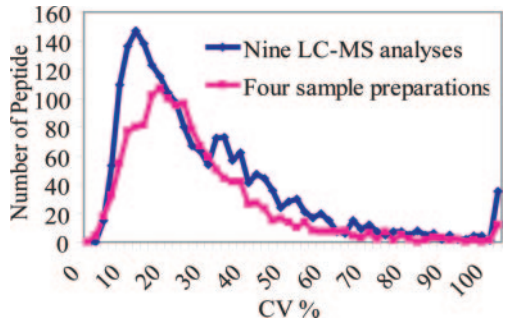

FIG. 4. **Reproducibility of the high throughput serum analysis method.** The distribution of CV from nine repeated LC-MS analyses of the same glycopeptide mixture (*blue line*) and the distribution of CV from four repeated sample preparations using the glycopeptide capture-and-release method and LC-MS analysis (*red line*) are shown.

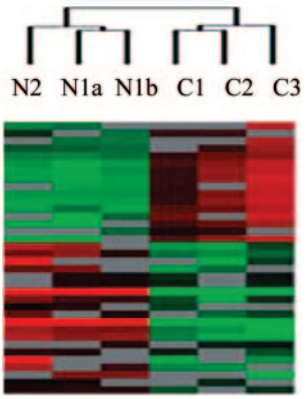

FIG. 5. **Unsupervised hierarchical clustering of *N*-linked glycopeptides distinguishes treated cancer-bearing mice (C1, C2, and C3) from untreated normal mice (N1a, N1b, and N2).** A section of the clustered peptides illustrating a cluster of discriminatory peptides is shown in which each *row* represents a peptide and each *column* represents a different serum sample. *Red* indicates a peptide with higher than average abundance in a specific sample, *green* indicates a peptide with lower than average abundance in a specific sample, *black* indicates no changes in abundance, and *gray* indicates the data in a specific sample is missing.

method to serum samples from mice in which skin tumors had been induced and from normal untreated littermates. Skin tumors were induced in a well established skin carcinoma model via topical treatment of the skin with a single dose of DMBA followed by repeated treatments with the tumor promoter 12-*O*-tetradecanoylphorbol-13-acetate (16). This treatment gives rise to papillomas that are hyperplastic and well differentiated benign lesions of the skin, each one originating from a single initiated cell (30, 31). After a latency period of several months, a small percentage of these lesions progress to squamous cell carcinomas.

From the sera of three cancer-bearing male mice (C1, C2, and C3) and two untreated normal male mice (N1 and N2) from the same litter, glycopeptides were isolated and analyzed by LC-MS as described above. The sample from N1 was analyzed by LC-MS twice (N1a and N1b); thus a total of six LC-MS patterns were generated. After aligning peptide peaks from all six patterns, over 3000 peptide peaks were found to occur in at least two of the three analyses from either normal or cancer-bearing mice. The six LC-MS patterns consisting of the peptide peaks matched between the samples and their associated intensities were next subjected to unsupervised hierarchical clustering (22). Neither predefined reference vectors nor prior knowledge about the nature of each pattern (untreated normal *versus* cancer-bearing) was used. The results of this unsupervised hierarchical clustering analysis are represented by the tree structure (Fig. 5). A section of the clustered peptides illustrating a cluster of discriminatory peptides is also shown in Fig. 5 in which each *row* represents a peptide and each *column* represents a different serum sample. The lengths of the branches among different samples are proportional to the similarity of the obtained peptide patterns. From this clustering, it is apparent that the cancer-bearing mice (C1, C2, and C3) were clustered together and clearly differentiated from the patterns obtained from their sex- and litter-matched normal mice (N1a, N1b, and N2), which were also clustered together.

To test whether the same serum samples could be equally differentiated without applying the glycopeptide capture-and-release enrichment method, we subjected tryptic peptides from 50 nl of each unprocessed serum sample to the same LC-MS and pattern analysis procedure. Peptide peaks were aligned from the resulting patterns, and a similar number of peptide peaks were detected as for the glycopeptide-enriched samples. In contrast to the glycopeptide-enriched samples, unsupervised clustering of the total serum peptide patterns did not differentiate the cancer group from the normal group (data not shown). These results indicate that the larger number of proteins and/or the deeper penetration into the serum proteome achieved by the glycopeptide selection chemistry is critical to the successful differentiation between serum samples according to the clinical state of the individuals.

The glycopeptide-enriched samples were then further analyzed by MS/MS to identify peptides that increase in abundance in cancer-bearing mice compared with untreated normal animals. The *m/z* and retention time coordinates of these peptides were added to the inclusion list on a tandem mass spectrometer and identified by LC-MS/MS and sequence data base searching. Fig. 6*A* shows a peptide at *m/z* of 709.7 (eluted at ~65 min) that, while showing variation between individuals, also clearly showed consistently increased abundance in cancer-bearing mice (C1, C2, and C3) compared with normal animals (N1a, N1b, and N2). The signal at *m/z* of 709.7 was subsequently identified as a peptide with the amino acid sequence LIPHLEKPLQN#FTLCFR (in which N# indicates the formerly *N*-linked glycosylation site) derived from serum amyloid P-component in mouse. This is an acute phase protein whose expression is known to be elevated during inflammation (32).

We further verified the differential abundance of the identi-

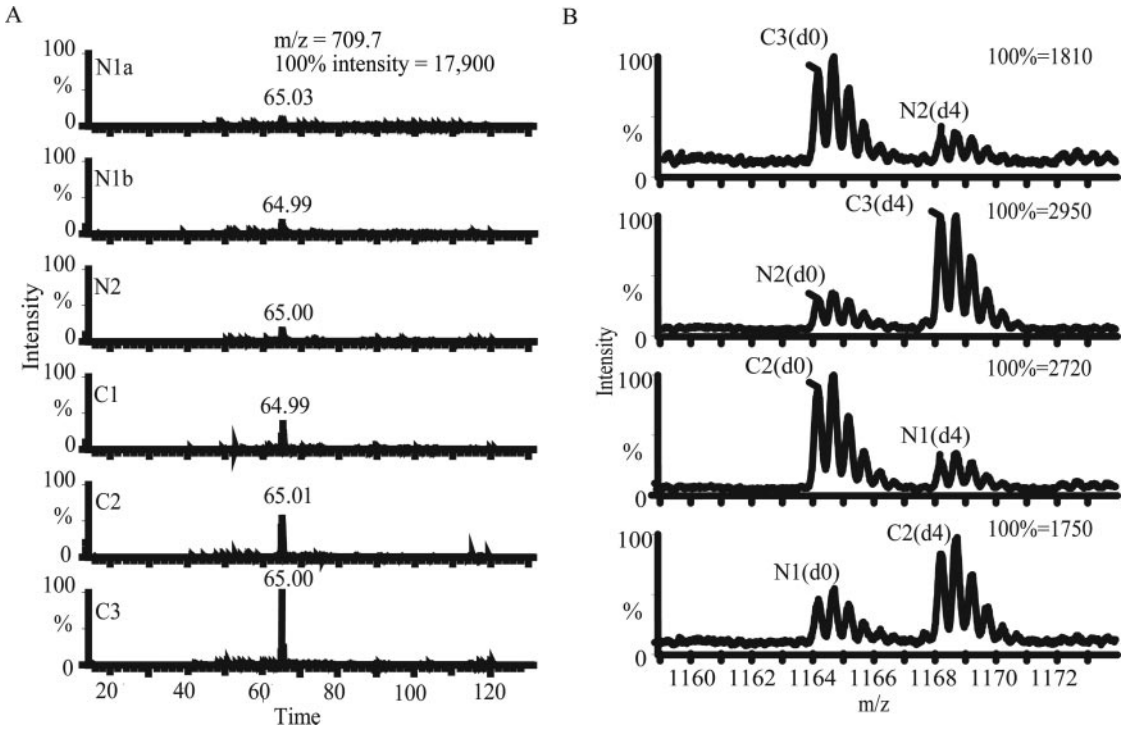

FIG. 6. **Identification of peptides exhibiting increased abundance in treated cancer-bearing mice.** *A*, normalized abundances of the peptide at *m/z* value of 709.7 observed in sera of normal (N1a, N1b, and N2) and cancer-bearing mice (C1, C2, and C3) determined by LC-MS analysis. *B*, validation of differential abundance of the same peptide shown in *A* using isotopic labeling of N termini.

fied peptides by applying accurate quantitative analysis using stable isotope labeling. In these experiments, the amino groups of the glycopeptides were isotopically labeled with d_0- and d_4-succinic anhydride, respectively, while the peptides were still attached to the solid support during their isolation (17). Equal aliquots of samples from two cancer-bearing mice (C2 and C3) and two normal mice (N1 and N2) were reverse labeled with either the d_0- and d_4-succinic anhydride, and the released peptides were combined in the following way: sample N1 (d_0) was paired with sample C2 (d_4), sample C2 (d_0) was paired with sample N1 (d_4), sample N2 (d_0) was paired with sample C3 (d_4), and sample C3 (d_0) was paired with sample N2 (d_0). The combined samples were analyzed by LC-MS/MS. The *m/z* values of peptides identified with higher abundance in cancer-bearing mice using LC-MS analysis and pattern matching were selected, and the corresponding masses for light and heavy succinic anhydride-labeled peptides were included in the mass inclusion list (with a 100-Dalton addition for the light form of succinic anhydride and a 104-Dalton addition for the heavy succinic anhydride labeling), then sequenced by MS/MS analysis using ESI-QTOF, and identified by data base searching. Table III lists the identified peptides and proteins with elevated protein level in the cancer-bearing mouse group detected by LC-MS analysis and verified by reverse stable isotope labeling. The CID spectra are given in Supplemental Fig. 1 on line. The LC-MS spectrum obtained for the same peptide from serum amyloid P-component is shown in Fig. 6*B*. The increased level of this peptide in cancer-bearing mice quantified by isotopic labeling was consistent with that determined by LC-MS analysis (Fig. 6*A*). Collectively these data indicate that the LC-MS-based analysis of isolated formerly *N*-linked glycosylated peptides reproducibly detected peptides of different abundance in serum samples of cancer and normal mice and that the discriminatory peptides could be identified by MS/MS analysis.

DISCUSSION

We describe a method for high throughput quantitative analysis of serum proteins using glycopeptide capture and LC-MS. It consists of the selective and reproducible isolation of those peptides from the serum proteome that are modified by *N*-linked glycosylation in the intact protein. The complex mixture of the deglycosylated forms of these peptides was then analyzed by LC-MS. The mass of discriminatory peptides was determined using pattern matching software, and these peptides were subsequently identified by MS/MS. Our results indicate that the glycopeptide capture-and-release method is specific for the isolation of *N*-linked glycopeptides. On average, 3.6 peptides were isolated per protein representing an average of 1.8 glycosylation sites per protein. This is contrasted with a predicted 28.8 unique tryptic peptides per protein calculated from the pool of identified proteins. The data also indicate that this reduced sample complexity resulted in an increase in sensitivity compared with the analysis

TABLE III
Identification of peptides and proteins with elevated abundance in treated cancer-bearing mice

Protein name	IPI number	Peptide sequences[a]	CID spectrum number given in Supplemental Fig. 1 on line
Ig γ-1 chain C region secreted form	IPI00109911	R.EEQFN#STFR.S	332
Serum amyloid P-component	IPI00267939	K.LIPHLEKPLQN#FTLCFR.T	333
Haptoglobin	IPI00274017	K.NLFLN#HSETASAK.D	334
		K.N#LTSPVGVQPILNEHTFCAGLTK.Y	335
Leucine-rich α-2-glycoprotein	IPI00129250	R.SLPPGLFSTSAN#LSTLVLR.E	336
Complement component factor h	IPI00130010	K.DNSCVDPPHVPN#ATIVTR.T	337
Fetuin β	IPI00134837	R.VLYLPAYN#CTLRPVSK.R	338
		R.RVLYLPAYN#CTLRPVSK.R	339

[a] N#, *N*-linked glycosylation site; peptide sequences between the two periods are identified by MS/MS.

of non-selected serum digests using an identical analytical platform. To test its suitability for analysis of disease, the method was applied to the differentiation of sera from genetically identical mice that were either untreated normal or cancer-bearing. The resulting peptide patterns could clearly and correctly be differentiated into two groups via unsupervised clustering. Some of the discriminatory peptides were further identified by MS/MS, and their differential abundance in cancer *versus* control mice was verified by accurate quantification using stable isotope labeling.

Ideally, for the detection and validation of protein biomarkers in serum, the complete serum proteomes of multiple individuals representing different clinical states would be completely and quantitatively analyzed. Due to the enormous complexity of the serum proteome and technical limitations, all the current proteomic technologies for such analyses can only sample a small part of the proteome, predominantly the most abundant proteins (14, 33). For example, two-dimensional gel electrophoresis-based studies have identified about 300 serum proteins collectively (14, 15, 34). It has also been estimated that SELDI-TOF approaches have limited detection of low abundance proteins due to the high dynamic range of serum proteins and the limited binding capacity of the SELDI chip (6). In the method presented here, the selective isolation of the *N*-linked glycosylated peptides resulted in a substantial improvement in the concentration limit of protein detected due to the reduction in sample complexity.

A number of factors contribute to this effect. First, the number of peptides per protein isolated after applying the glycopeptide capture-and-release method is significantly reduced. The 93 proteins identified in this study are predicted to generate an average of 28.8 tryptic peptides per protein. Of these, only 3.6 on average contain the *N*-linked glycosylation consensus motif and can be potentially glycosylated, and an average of 3.6 peptides representing 1.8 unique *N*-linked glycosylation sites per protein were actually identified. By comparison, a similar number of *N*-linked glycosylation sites identified per protein was reported by Kaji and colleagues (35) (1.8 sites per protein) in a study in which *N*-linked glycopeptides were isolated from *Caenorhabditis elegans* proteins using lectin enrichment. Second, the most abundant serum protein, albumin, does not contain *N*-linked glycosylation motifs and therefore is effectively transparent to the analysis. Since albumin itself comprises almost 50% of total serum protein content, exclusion of albumin eliminates numerous peptides that otherwise dominate serum peptide samples. Indeed quantitative removal of albumin, a goal that is normally attempted by use of costly affinity depletion methods (36), is an automatic by-product of the glycopeptide capture method. Third, the method only selects peptides from the constant region of immunoglobulins and thus dramatically reduces the number of immunoglobulin-derived peptides. This is important since immunoglobulins constitute ~20% of total protein mass in serum (26) and comprise a population of an estimated 10 million different molecules (14). The difficulty of penetrating the population of immunoglobulins in unbiased serum proteome analyses was recently illustrated in a study in which a tryptic digest of serum was analyzed by ultrahigh efficiency strong cation exchange LC/reversed-phase LC/MS/MS. Of the 1061 plasma protein identifications reported, 38% were immunoglobulins (10). It is also likely that an even more significant fraction of peptides observed in LC-MS patterns of unbiased serum protein digests are derived from immunoglobulins since nucleic acid and protein sequence data bases dramatically underreport the contribution of somatic combinatorial gene rearrangement to immunoglobulin diversity. Fourth, many serum proteins are post-translationally altered by phosphorylation, glycosylation, acetylation, methionine oxidation, protease processing, and other mechanisms, resulting in multiple forms for each protein. It has been estimated that one protein may generate on the order of 100 species (14). In the case of glycosylation, the oligosaccharide structures attached at each site are typically diverse, compounding the complexity of the peptide mixture. The peptides isolated by the glycopeptide capture method remove the heterogeneous oligosaccharides and thus, by isolating a few peptides per protein only, also eliminate other significant sources of pattern heterogeneity.

The cumulative effect of these factors is the generation of a peptide sample from the serum proteome with a moderate redundancy of an average of 3.6 unique peptides per protein. Theoretically an average of 3.6 potential *N*-linked glycopep-

tides (containing an N*X*(T/S) motif) is predicted for the 93 identified serum proteins (Supplemental Table 1). However, not all of these potential *N*-linked glycosylation sites were observed. Some of these potential *N*-linked glycosylation sites may not actually be occupied (25), the peptides from certain sites may not be detectable by mass spectrometry, or protein digestion may be hindered by the protein post-translational modifications such as oligosaccharide attachment and/or disulfide bond formation. On the other hand, the number of peptides from each glycosylation site was increased due to other types of protein modifications (*e.g.* methionine oxidation and protease processing) in the glycosylation region. It is expected that the same factors would also lead to an inflation of the number of peptides observed if digests of non-selected serum samples were analyzed. In our analyses of peptides generated from 5 μl of mouse serum using the glycopeptide capture-and-release method, we were able to detect and quantify over 3000 peptide peaks that were present at least at two of three samples in either group with intensity at least at 2-fold above background noise level. In MS/MS analysis, only a small fraction of peptides (319 unique peptides, Supplemental Table 1) were identified. This was due to the complexity of the sample and the fact that the mass spectrometer only had time to sequence a small portion of the peptides, predominantly the highly abundant peptides in each sample (undersampling). The same undersampling factor was also the major cause of the inconsistency of protein identifications using LC-MS/MS. In this study, we used reproducible LC-MS for quantitative analyses, and this allowed us to analyze all the peptide ions in each sample, including those from proteins of low abundance.

While the reduction of peptide redundancy is beneficial for achieving higher coverage of the proteome per analysis, it is also apparent that it leads to the loss of some, potentially important information. First, non-glycosylated proteins are transparent in this system. While it is believed that the majority of serum-specific proteins are in fact glycosylated (37), intracellular proteins (typically non-glycosylated) that may represent a rich source of biomarkers if leaked into serum might go undetected. Second, the availability of fewer peptides per protein increases the challenge of identifying the corresponding protein. Third, this approach will reveal differences in protein level or glycosylation level (glycosylation site occupancy). Disease markers that alter other protein post-translational modifications including proteolytic processing will not be detected on a glycopeptide level. Finally collapsing peptides modified by different oligosaccharide structures into a single signal will obscure potential disease markers that are due to oligosaccharide structure alteration (37).

In this study, we successfully used the glycopeptide capture and LC-MS analysis platforms to differentiate serum from mice with chemically induced skin cancer from that of non-treated littermates. In this experiment, the mice with skin cancer and their untreated littermates had the same genetic background and lived in the same environment. The study therefore represents a controlled experiment with chemically induced skin cancer being the sole variable. The sera were clearly distinguished by numerous distinct peptides, the abundance of which was consistently increased or decreased between the cancer and control sera. While in this controlled experiment, the low number of samples was sufficient to detect disease-associated signatures, the application of the method to identification of potential biomarkers in much more variable human samples will require the analysis of much larger sample numbers to facilitate statistical validation of the data. The current method, at present, has sufficient throughput to perform studies involving a few hundred samples, a number that appears sufficient to generate statistically significant results within a reasonable time frame (38, 39). By developing a robotic procedure to allow automated sample preparation and by further optimizing LC-MS analysis procedures and the development of a robust, automated data analysis platform, we are further increasing the performance of the system.

In contrast to the widely used SELDI-TOF and similar polypeptide profiling methods, the signals detected in the present method are defined molecular species, mostly peptides ranging in length between 7 and 30 amino acids. These peptides, if selected for CID in a tandem mass spectrometer, are readily sequenced. By adding the coordinates of selected discriminatory peptides to an inclusion list, we have identified several serum proteins for which the abundance is increased in correlation with the chemical induction of skin cancer in mice (Table III). While these proteins are indicators of interesting biology and have been reported to change the abundance in different types of cancer (40), they are likely not markers for the specific diagnosis of skin cancer. Proteins useful for cancer detection, diagnosis, or stratification might be proteins released in small amounts from the primary lesion, indicators of a specific response of the system to the lesion, or other subtle changes in the serum proteome. For the reliable detection of such proteins or patterns of proteins, it is imperative that all candidate molecules are identified so that potential markers or signatures observed in different diseases, studies, and laboratories can be validated, correlated, and compared. This will allow the proteomic biomarker discovery community to establish defined molecular signatures as the currency of communication and to distinguish between true biomarkers and coincidental changes. The identification of discriminatory peptides in this study furthermore indicates that at least some of the proteins changing in abundance in the skin cancer model are moderately to highly expressed. In contrast, serum cancer markers currently in clinical use have concentrations in the ng/ml range. Diamandis (6) has argued that the SELDI-TOF method and by implication similar methods are about 3 orders of magnitude too insensitive from the sensitivity required to detect such proteins. The method presented here has the potential to reach ng/ml sensitivity levels

and even lower concentration limits if high performance Fourier transform ion cyclotron resonance-MS instruments are used (data not shown). For example, at a concentration of 4 ng/ml, 5 μl of serum sample contains ~20 pg (~700 amol) of prostate-specific antigen, an amount that is readily detected in a modern mass spectrometer. In comparison, if non-biased serum digests are analyzed on the same capillary LC-MS system, the total amount of serum that can be applied to the system would be 50 nl, and therefore the concentration limit of detection would be 100-fold reduced compared with the glycopeptide-selected sample. Thus prostate-specific antigen would be well outside the detection limit of such an analysis. If further increases in the concentration limit of detection were required, the glycopeptide capture-and-release method could easily be combined with other peptide fractionation methods, including electrophoresis (gel-based or free flow electrophoresis) chromatography or affinity depletion.

In summary, selectively isolating peptides from *N*-linked glycosylated serum proteins has been found to be a powerful method for the analysis of the serum proteome. Together with the high reproducibility of this method, the high level of serum proteome coverage achieved at a moderate throughput suggests that this method will be most useful for the detection of proteins or protein patterns that distinguish individuals in different physiological states.

Acknowledgments—We thank Julian Watts and Paul Loriaux for critical comments. We thank David J. Anderson, Yufeng Shen, Andrey N. Vilkov, and Nikola Tolic at Pacific Northwest National Laboratory for assistance with obtaining and analyzing the LC-Fourier transform ion cyclotron resonance data from the mouse serum samples.

* This work was supported in part by NCI, National Institutes of Health Grants R33 and CA93302, by Grant N01-HV-28179 from the NHLBI, National Institutes of Health, Proteomics Initiative, and by a sponsored research agreement from Industrial Technology Research Institute in Taiwan. The Institute for Systems Biology was supported by a generous gift from Merck & Co. The LC-FTICR analyses at PNNL were supported by NIH National Center for Research Resources (RR-18522). Pacific Northwest National Laboratory is operated by the Battelle Memorial Institute for the U. S. Department of Energy through Contract DE-AC06-76RLO1830. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

§ To whom correspondence should be addressed. E-mail: hzhang@systemsbiology.org.

REFERENCES

1. Wulfkuhle, J. D., Liotta, L. A., and Petricoin, E. F. (2003) Proteomic applications for the early detection of cancer. *Nat. Rev. Cancer* **3,** 267–275
2. Anderson, L., and Anderson, N. G. (1977) High resolution two-dimensional electrophoresis of human plasma proteins. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5421–5425
3. Merril, C. R., Goldman, D., Sedman, S. A., and Ebert, M. H. (1981) Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science* **211,** 1437–1438
4. Merril, C. R., Switzer, R. C., and Van Keuren, M. L. (1979) Trace polypeptides in cellular extracts and human body fluids detected by two-dimensional electrophoresis and a highly sensitive silver stain. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 4335–4339
5. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422,** 198–207
6. Diamandis, E. P. (2004) Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Mol. Cell. Proteomics* **3,** 367–378
7. Petricoin, E. F., Ardekani, A. M., Hitt, B. A., Levine, P. J., Fusaro, V. A., Steinberg, S. M., Mills, G. B., Simone, C., Fishman, D. A., Kohn, E. C., and Liotta, L. A. (2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359,** 572–577
8. Adkins, J. N., Varnum, S. M., Auberry, K. J., Moore, R. J., Angell, N. H., Smith, R. D., Springer, D. L., and Pounds, J. G. (2002) Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. *Mol. Cell. Proteomics* **1,** 947–955
9. Tirumalai, R. S., Chan, K. C., Prieto, D. A., Issaq, H. J., Conrads, T. P., and Veenstra, T. D. (2003) Characterization of the low molecular weight human serum proteome. *Mol. Cell. Proteomics* **2,** 1096–1103
10. Shen, Y., Jacobs, J. M., Camp, D. G., II, Fang, R., Moore, R. J., Smith, R. D., Xiao, W., Davis, R. W., and Tompkins, R. G. (2004) Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome. *Anal. Chem.* **76,** 1134–1144
11. Wang, H., and Hanash, S. (2003) Multi-dimensional liquid phase based separations in proteomics. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **787,** 11–18
12. Shin, B. K., Wang, H., and Hanash, S. (2002) Proteomics approaches to uncover the 31 repertoire of circulating biomarkers for breast cancer. *J. Mammary Gland Biol. Neoplasia* **7,** 407–413
13. Villanueva, J., Philip, J., Entenberg, D., Chaparro, C. A., Tanwar, M. K., Holland, E. C., and Tempst, P. (2004) Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. *Anal. Chem.* **76,** 1560–1570
14. Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **1,** 845–867
15. Pieper, R., Gatlin, C. L., Makusky, A. J., Russo, P. S., Schatz, C. R., Miller, S. S., Su, Q., McGrath, A. M., Estock, M. A., Parmar, P. P., Zhao, M., Huang, S. T., Zhou, J., Wang, F., Esquer-Blasco, R., Anderson, N. L., Taylor, J., and Steiner, S. (2003) The human serum proteome: display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins. *Proteomics* **3,** 1345–1364
16. Kemp, C. J., Donehower, L. A., Bradley, A., and Balmain, A. (1993) Reduction of p53 gene dosage does not increase initiation or promotion but enhances malignant progression of chemically induced skin tumors. *Cell* **74,** 813–822
17. Zhang, H., Li, X. J., Martin, D. B., and Aebersold, R. (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **21,** 660–666
18. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17,** 994–999
19. Yi, E. C., Lee, H., Aebersold, R., and Goodlett, D. R. (2003) A microcapillary trap cartridge-microcapillary high-performance liquid chromatography electrospray ionization emitter device capable of peptide tandem mass spectrometry at the attomole level on an ion trap mass spectrometer with automated routine operation. *Rapid Commun. Mass Spectrom.* **17,** 2093–2098
20. Griffin, T. J., Lock, C. M., Li, X. J., Patel, A., Chervetsova, I., Lee, H., Wright, M. E., Ranish, J. A., Chen, S. S., and Aebersold, R. (2003) Abundance ratio-dependent proteomic analysis by mass spectrometry. *Anal. Chem.* **75,** 867–874
21. Li, X. J., Zhang, H., Ranish, J. A., and Aebersold, R. (2003) Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. *Anal. Chem.* **75,** 6648–6657
22. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 14863–14868

23. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74,** 5383–5392
24. Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling of 33 differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **19,** 946–951
25. Petrescu, A. J., Milac, A. L., Petrescu, S. M., Dwek, R. A., and Wormald, M. R. (2004) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **14,** 103–114
26. Putnam, F. (1975) *The Plasma Proteins: Structure, Function, and Genetic Control*, 2nd Ed., Academic Press, New York
27. Lum, G., and Gambino, S. R. (1974) A comparison of serum versus heparinized plasma for routine chemistry tests. *Am. J. Clin. Pathol.* **61,** 108–113
28. Duan, X., Yarmush, D. M., Berthiaume, F., Jayaraman, A., and Yarmush, M. L. (2004) A mouse serum two-dimensional gel map: application to profiling burn injury and infection. *Electrophoresis* **25,** 3055–3065
29. Li, X. J., Pedrioli, P. G., Eng, J., Martin, D., Yi, E. C., Lee, H., and Aebersold, R. (2004) A tool to visualize and evaluate data obtained by liquid chromatography-electrospray ionization-mass spectrometry. *Anal. Chem.* **76,** 3856–3860
30. Brown, K., Quintanilla, M., Ramsden, M., Kerr, I. B., Young, S., and Balmain, A. (1986) v-ras genes from Harvey and BALB murine sarcoma viruses can act as initiators of two-stage mouse skin carcinogenesis. *Cell* **46,** 447–456
31. Quintanilla, M., Brown, K., Ramsden, M., and Balmain, A. (1986) Carcinogen-specific mutation and amplification of Ha-ras during mouse skin carcinogenesis. *Nature* **322,** 78–80
32. Mole, J. E., Beaulieu, B. L., Geheran, C. A., Carnazza, J. A., and Anderson, J. K. (1988) Isolation and analysis of murine serum amyloid P component cDNA clones. *J. Immunol.* **141,** 3642–3646
33. Zhang, H., Yan, W., and Aebersold, R. (2004) Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes. *Curr. Opin. Chem. Biol.* **8,** 66–75
34. Anderson, N. L., Polanski, M., Pieper, R., Gatlin, T., Tirumalai, R. S., Conrads, T. P., Veenstra, T. D., Adkins, J. N., Pounds, J. G., Fagan, R., and Lobley, A. (2004) The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* **3,** 311–326
35. Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi, N., and Isobe, T. (2003) Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotechnol.* **21,** 667–672
36. Pieper, R., Su, Q., Gatlin, C. L., Huang, S. T., Anderson, N. L., and Steiner, S. (2003) Multi-component immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome. *Proteomics* **3,** 422–432
37. Durand, G., and Seta, N. (2000) Protein glycosylation and diseases: blood and urinary oligosaccharides as markers for diagnosis and therapeutic monitoring. *Clin. Chem.* **46,** 795–805
38. Sullivan Pepe, M., Etzioni, R., Feng, Z., Potter, J. D., Thompson, M. L., Thornquist, M., Winget, M., and Yasui, Y. (2001) Phases of biomarker development for early detection of cancer. *J. Natl. Cancer Inst.* **93,** 1054–1061
39. Adam, B. L., Qu, Y., Davis, J. W., Ward, M. D., Clements, M. A., Cazares, L. H., Semmes, O. J., Schellhammer, P. F., Yasui, Y., Feng, Z., and Wright, G. L., Jr. (2002) Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res.* **62,** 3609–3614
40. Vejda, S., Posovszky, C., Zelzer, S., Peter, B., Bayer, E., Gelbmann, D., Schulte-Hermann, R., and Gerner, C. (2002) Plasma from cancer patients featuring a characteristic protein composition mediates protection against apoptosis. *Mol. Cell. Proteomics* **1,** 387–393