

(19) AUSTRALIAN PATENT OFFICE

(54) Title
Methods for quantitative proteome analysis of glycoproteins

(51)⁶ International Patent Classification(s)
G01N 33/68 (2006.01) 1/34
C12Q 1/34 (2006.01) 20060101ALI2005100
C12Q 1/37 (2006.01) ^{8BMEP} **C12Q**
G01N 33/68 1/37
20060101AFI2005100 20060101ALI2005100
^{8BMEP} **C12Q** ^{8BMEP}
PCT/US03/17635

(21) Application No: 2003249692

(22) Application Date: 2003 .06 .03

(87) WIPO No: W003/102018

(30) Priority Data

(31) Number	(32) Date	(33) Country
60/469,361	2003 .05 .09	US
60/385,707	2002 .06 .03	US

(43) Publication Date : 2003 .12 .19

(43) Publication Journal Date : 2004 .02 .12

(71) Applicant(s)
The Institute for Systems Biology

(72) Inventor(s)
Aebersold, Rudolf H., Zhang, Hui

(74) Agent/Attorney
Griffith Hack, GPO Box 3125, Brisbane, QLD, 4001

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



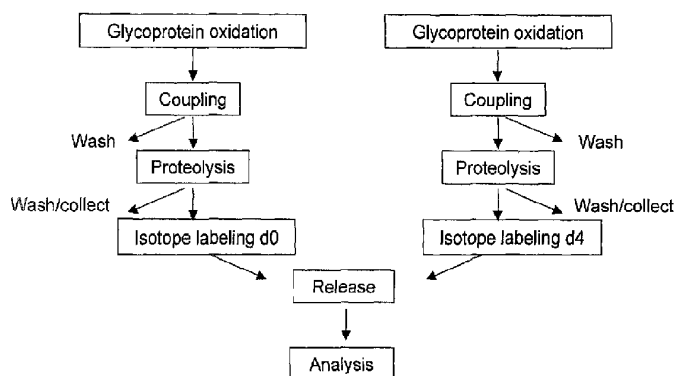
(43) International Publication Date
11 December 2003 (11.12.2003)

PCT

(10) International Publication Number
WO 03/102018 A2

- (51) International Patent Classification: **C07K**
- (21) International Application Number: PC1/US03/17635
- (22) International Filing Date: 3 June 2003 (03.06.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/385,707 3 June 2002 (03.06.2002) US
60/469,361 9 May 2003 (09.05.2003) US
- (71) Applicant (for all designated States except US): **THE INSTITUTE FOR SYSTEMS BIOLOGY** [US/US]; 1441 North 34th Street, Seattle, WA 98103-1299 (US).
- (72) Inventors; and
(75) Inventors/Applicants (for US only): **AEBERSOLD, Rudolf, H.** [CH/US]; 8609 SE 78th Street, Mercer Island, WA 98040 (US). **ZHANG, Hui** [CN/US]; 3221 N.E. 96th Street, Seattle, WA 98115 (US).
- (74) Agents: **CADENA, Deborah, L.** et al.; McDermott, Will & Emery, 4370 La Jolla Village Drive, 7th Floor, San Diego, CA 92122 (US).
- (81) Designated States (national): AL, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS FOR QUANTITATIVE PROTEOME ANALYSIS OF GLYCOPROTEINS



(57) Abstract: The invention provides a method for identifying and quantifying polyglycopeptides in a sample. The method can include the steps of immobilizing glycopolypeptides to a solid support; cleaving the immobilized glycopolypeptides, thereby releasing non-glycosylated peptides and retaining immobilized glycopeptides; releasing the glycopeptides from the solid support; and analyzing the released glycopeptides. The method can further include the step of identifying one or more glycopeptides, for example, using mass spectrometry.

METHODS FOR QUANTITATIVE PROTEOME ANALYSIS OF
GLYCOPROTEINS

BACKGROUND OF THE INVENTION

5

The present invention relates generally to the field of proteomics and more specifically to quantitative analysis of glycoproteins.

10

Complete genomic sequences and large partial (EST) sequence databases potentially identify every gene in a species. However, the sequences alone do not explain the mechanism of biological and clinical processes because they do not explain how the genes and their products cooperate to carry out a specific process or function. Furthermore, the gene sequence does not predict the amount or the activity of the protein products nor does it answer the questions of whether, how, and at what position(s) a protein may be modified.

20

Quantitative protein profiling has been recognized as an important approach for profiling the physiological state or pathological state of cells or organisms. Specific expectations of quantitative protein profiles include the possibility to detect diagnostic and prognostic disease markers, to discover proteins as therapeutic targets or to learn about basic biological mechanisms.

25

30

Not only do the amounts and type of proteins expressed vary in different pathological states, post-translational modifications of proteins also vary

depending on the physiological or pathological state of cells or organisms. Thus, it is important to be able to profile the amount and types of expressed proteins as well as protein modifications.

5

Glycosylation has long been recognized as the most common post-translational modification affecting the functions of proteins, such as protein stability, enzymatic activity and protein-protein interactions.

10

Differential glycosylation is a major source of protein microheterogeneity. Glycoproteins play key roles in cell communications, signaling and cell adhesion. Changes in carbohydrates in cell surface and body fluid are demonstrated in cancer and other disease states and highlights their importance. However, studies on protein glycosylation have been complicated by the diverse structure of protein glycans and the lack of effective tools to identify the glycosylation site(s) on proteins and of glycan structures. Oligosaccharides can be linked to serine or threonine residues (O-glycosylation) or to asparagine residues (N-glycosylation), and glycoproteins can have different oligosaccharides attached to any given possible site(s).

25

Among the many post-translation modifications of proteins, glycosylation is a modification that is common to proteins that are exposed to an extracellular environment. For example, proteins expressed on the surface of a cell are exposed to the external environment such as blood or surrounding tissue. Similarly, proteins that are secreted from a cell, for example, into the bloodstream, are commonly glycosylated.

30

Among the diverse types of proteins expressed by cells, proteins that are integral to or associated with lipid membranes perform a wide range of essential cellular functions. Pores, channels, pumps and
5 transporters facilitate the exchange of membrane impermeable molecules between cellular compartments and between the cell and its extracellular environment. Transmembrane receptors sense changes in the cellular environment and, typically via associated proteins,
10 initiate specific intracellular responses. Cell adhesion proteins mediate cell-specific interactions with other cells and the extracellular matrix. Lipid membranes also provide a hydrophobic environment for biochemical reactions that is dramatically different from that of the
15 cytoplasm and other hydrophilic cellular compartments.

Membrane proteins, in particular those spanning the plasma membrane, are also of considerable diagnostic and therapeutic importance, which is further reinforced due
20 to their easy accessibility. Antisera to proteins that are selectively expressed on the surface of a specific cell type have been used extensively for the classification of cells and for their preparative isolation by fluorescent activated cell sorting or
25 related methods. Membrane proteins, as exemplified by Her2/neu, the abundance of which is modulated in the course of certain diseases such as breast cancer, are commonly used as diagnostic indicators and, less frequently, as therapeutic targets. A humanized
30 monoclonal antibody (Herceptin, Genentech, Palo Alto, CA) that specifically recognizes Her2/neu receptors is the basis for a successful therapy of breast cancer, and

antibodies to other cell surface proteins are also undergoing clinical trials as anticancer agents. Moreover, the majority of current effective therapeutic agents for diseases such as hypertension and heart disease are receptor antagonists that target and selectively modify the activity of specific membrane proteins. It is therefore apparent that a general technique capable of systematically identifying membrane proteins and of accurately detecting quantitative changes in the membrane protein profiles of different cell populations or tissues would be of considerable importance for biology and for applied biomedical research.

In addition to membrane bound proteins, proteins secreted by cells or shed from the cell surface, including hormones, lymphokines, interferons, transferrin, antibodies, proteases, protease inhibitors, and other factors, perform critical functions with respect to the physiological activity of an organism. Examples of physiologically important secreted proteins include the interferons, lymphokines, protein and peptide hormones. Aberrant availability of such proteins can have grave clinical consequences. It is therefore apparent that the ability to precisely quantitatively profile secreted proteins would be of great importance for the discovery of the mechanisms regulating a wide variety of physiological processes in health and disease and for diagnostic or prognostic purposes. Such secreted proteins are present in body fluids such as blood serum and plasma, cerebrospinal fluid, urine, lung lavage, breast milk, pancreatic juice, and saliva. For example,

the presence of increased levels of prostate-specific antigen has been used as a diagnostic marker for prostate cancer. Furthermore, the use of agonists or antagonists or the replacement of soluble secreted proteins is an
5 important mode of therapy for a wide range of diseases.

Quantitative proteomics requires the analysis of complex protein samples. In the case of clinical diagnosis, the ability to obtain appropriate specimens
10 for clinical analysis is important for ease and accuracy of diagnosis. As discussed above, a number of biologically important molecules are secreted and are therefore present in body fluids such as blood and serum, cerebrospinal fluid, saliva, and the like. In addition
15 to the presence of important biological molecules, body fluids also provide an attractive specimen source because body fluids are generally readily accessible and available in reasonable quantities for clinical analysis. It is therefore apparent that a general method for the
20 quantitative analysis of the proteins contained in body fluids in health and disease would be of great diagnostic and clinical importance.

A key problem with the proteomic analysis of serum
25 and many other body fluids is the peculiar protein composition of these specimens. The protein composition is dominated by a few proteins that are extraordinarily abundant, with albumin alone representing 50% of the total plasma proteins. Due to the abundance of these
30 major proteins as well as the presence of multiple modified forms of these abundant proteins, the large number of protein species of lower abundance are obscured

or inaccessible by traditional proteomics analysis methods such as two-dimensional electrophoresis (2DE).

The classes of proteins described above, membrane
5 proteins, secreted proteins, and proteins in body fluids
have in common that they have a high propensity for being
glycosylated, that is, modified post translationally with
a carbohydrate structure of varying complexity at one or
10 several amino acid residues. Thus, the analysis of
glycoproteins allows characterization of important
biological molecules.

Thus, there exists a need for methods of high
15 throughput and quantitative analysis of glycoproteins and
glycoprotein profiling. The present invention satisfies
this need and provides related advantages as well.

20 SUMMARY OF THE INVENTION

According to one aspect of the present invention
there is provided a method for identifying and quantifying
a glycopolypeptide in a sample, comprising:

- 25 (a) derivatizing carbohydrate groups of
glycopolypeptides in a polypeptide sample;
(b) immobilizing said derivatized carbohydrate groups
of said glycopolypeptides to a solid support;
(c) cleaving said immobilized glycopolypeptides,
30 thereby releasing non-glycosylated peptide fragments and
retaining immobilized glycopeptides fragments;
(d) labeling said immobilized glycopeptides fragments
with an isotope tag;
(e) releasing said glycopeptides fragments from said
35 solid support, thereby generating released glycopeptides
fragments;
(f) analyzing said released glycopeptide fragments

- 6a -

using mass spectrometry;

(g) identifying a released glycopeptide fragment;

(h) quantifying the amount of said glycopeptide fragment identified in step (g); and

5 (i) identifying a parent glycopolypeptide from which the glycopeptide fragment identified in step (g) was derived, thereby identifying and quantifying a glycopolypeptide.

According to a further aspect of the present
10 invention there is provided a method for identifying and quantifying a glycopolypeptide in a sample, comprising:

(a) immobilizing glycopolypeptides in said sample to a solid support wherein the carbohydrate groups of said glycopolypeptides are derivatized;

15 (b) cleaving said immobilized glycopolypeptides, thereby releasing non-glycosylated peptides and retaining immobilized glycopeptides;

(c) labeling said immobilized glycopeptides with an isotope tag;

20 (d) releasing said glycopeptides from said solid support; and

(e) analyzing said released glycopeptides, wherein said analyzing identifies and quantifies a glycopolypeptide in said sample.

25 According to a still further aspect of the present invention there is provided a method for identifying and quantifying a glycopolypeptide in a sample, comprising:

(a) immobilizing glycopolypeptides in said sample to a solid support wherein the carbohydrate groups of said
30 glycopolypeptides are derivatized;

(b) cleaving said immobilized glycopolypeptides, thereby releasing non-glycosylated peptides and retaining immobilized glycopeptides;

35 (c) releasing said glycopeptides from the solid support; and

(d) analyzing said released glycopeptides, wherein said analyzing identifies and quantifies a

- 6b -

glycopolypeptide in said sample.

According to yet another aspect of the present invention there is provided a method of identifying a diagnostic marker for a disease, comprising:

- 5 (a) immobilizing glycopolypeptides from a test sample to a first solid support;
- (b) immobilizing glycopolypeptides from a control sample to a second solid support;
- (c) cleaving said immobilized glycopolypeptides,
10 thereby releasing non-glycosylated peptide and retaining immobilized glycopeptides;
- (d) labeling said immobilized glycopeptides on said first and second supports with differential isotope tags on the respective supports;
- 15 (e) releasing the glycopeptides from said solid supports; and
- (f) analyzing said released glycopeptides; and
- (g) identifying one or more glycosylated polypeptides having differential glycosylation between the test sample
20 and the control sample.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a schematic diagram of an exemplary method of identifying and quantifying glycopolypeptides/glycoproteins and for determining quantitative changes in the glycosylation state of proteins.

Figure 2 shows oxidation of a carbohydrate to an aldehyde followed by covalent coupling to hydrazide beads.

Figure 3 shows representative chemical reagents that have been tested and proved to be able to label amino groups of glycopeptides. The structures of labeled peptide are listed in the right column.

Figure 4 shows total protein staining or glycoprotein staining of crude serum before (-) and after immobilization (+) of glycoproteins to hydrazide resin. Proteins were separated by SDS-PAGE and stained with silver (left) or Gel Code Blue glycoprotein staining reagent (right).

Figure 5 shows an outline and comparison of the results of glycopeptide analysis of serum proteins observed with three methods: cysteine capture with extensive separation, glycopeptide capture and single liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS), and cysteine capture and single LC-MS/MS.

Figure 6 shows identification of glycosylated proteins secreted from macrophages. Glycoproteins were identified from secreted proteins of untreated or LPS-treated RAW macrophage cells.

5

Figure 7 shows comparison of protein/peptide identification from the microsomal fraction of the prostate cancer cell line LNCaP using an ICAT™ reagent or selective isolation of N-glycosylated peptides.

10

Figure 8 shows subcellular location of glycoproteins identified from a crude microsomal fraction of LNCaP prostate epithelial cells.

15

Figure 9 shows the chemistry and schematic diagram of isotopically labeling the N-termini of the immobilized glycopeptides by attaching differentially isotopically labeled forms of the amino acid phenylalanine (Phe) to their N-termini.

20

Figure 10 shows isotopic labeling with Phe and identification of glycopeptides (SEQ ID NOS:1-10) using MS/MS. The glycopeptides were isolated from 1 µl of mouse ascites fluid.

25

Figure 11 shows collision-induced dissociation (CID) spectrum of one of the peptides (SEQ ID NO:7) identified in Figure 10 (circled).

30

Figure 12 shows reconstructed ion chromatograms for the peptide measured in Figure 11. The ratio of the calculated peak area for the heavy and light form of the

isotope tagged peptides was used to determine the relative peptide abundance in the original mixtures.

Figure 13 shows the quantification for a single peptide pair. A single scan of the mass spectrometer at spot 28 from a MALDI plate in MS mode identified eight paired signals with a mass difference of four units (indicated with *).

Figure 14 shows analysis of a precursor ion by MS/MS. Sequence database searching of the resulting spectrum identified the peptide sequence as IYSGILN#LSDITK from human plasma kallikrein, a serum protease. N# indicates the modified asparagine in the peptide sequence.

Figure 15 shows the patterns of aligned sequences. For each position in the aligned sequence, the height of each letter is proportional to its frequency, and the most common one is on top. There was high preference of N at position 21 (removed to show the detail of other positions). The preference of N was followed by S or T at position 23 (removed to show residues in other positions).

Figure 16 shows proteins identified from extracellular matrix of normal and prostate cancer tissues.

Figure 17 shows the total peptides present in a single LC-MS/MS run and the identified peptides by CID

acquired during the LC-MS/MS run followed by a search using SEQUEST.

Figure 18 shows a schematic diagram of the strategy used to profile glycopeptides present in serum and identify biomarkers.

Figure 19 shows the signal intensity of peptides during the elution of an LC-MS/MS run. N1 and N2 were from normal mouse serum, and T1 and T2 were glycopeptides from mouse serum with skin cancer.

Figure 20 shows the intensity of deconvoluted peptides during different elution time from serum of normal mice and mice with skin cancer. The left panel shows peptides in normal mouse. The right panel shows peptides in cancer mouse.

Figure 21 shows normalized peptide abundance between cancer and normal mouse. The relative peptide intensity of cancer mouse to normal mouse.

Figure 22 shows clustering analysis of normal mice and mice with cancer. Automatic, whole feature clustering of mouse serum distinguishes cancer from healthy. All the cancer mice clustered together (indicated as 11A, 12A, 13A in experiment one, upper panel; and M11, M12, M13 in experiment two, lower panel).

Figure 23 shows clustering analysis of samples from individuals before and after overnight fasting. Automatic clustering of serum from three individuals

before and after overnight fasting consistently separates individuals (experiment one, upper panel; experiment two, lower panel). Serum samples from the same person cluster together.

5

Figure 24 shows a schematic diagram of a glycosylation occupancy study of serum from congenital disorders of glycosylation (CDG) patients.

10

Figure 25 shows a schematic diagram of a study on total level of glycosylation using serum from obese and normal mice.

15

Figure 26 shows sequences of heavy isotope labeled synthetic peptide standards (SEQ ID NOS:11-19) identified by mass spectrometry. V* is the heavy valine and F# is the heavy phenylalanine.

20

Figure 27 shows peptides (SEQ ID NOS:20-29) identified from a series of enzymatic cleavages to release O-linked glycopeptides from hydrazide resin after N-linked glycopeptides were released.

25

Figure 28 shows identified N-linked glycopeptides (SEQ ID NOS:30-48), with the consensus NXT/S motif highlighted.

30

Figure 29 shows peptides (SEQ ID NOS:49-63) identified with O-linked oligosaccharides. These were generated by the removal of the O-linked oligosaccharide chains in the electrospray source. The site of carbohydrate attachment is characterized by a loss of

water at Ser or Thr to which the O-linked oligosaccharides were linked. The serine or threonine residues with the 18 Dalton water loss are circled.

5

DETAILED DESCRIPTION OF THE INVENTION

The invention provides methods for quantitative profiling of glycoproteins and glycopeptides on a proteome-wide scale. The methods of the invention allow the identification and quantification of glycoproteins in a complex sample and determination of the sites of glycosylation. The methods of the invention can be used to determine changes in the abundance of glycoproteins and changes in the state of glycosylation at individual glycosylation sites on those glycoproteins that occur in response to perturbations of biological systems and organisms in health and disease.

The methods of the invention can be used to purify glycosylated proteins or peptides and identify and quantify the glycosylation sites. Because the methods of the invention are directed to isolating glypolypeptides, the methods also reduce the complexity of analysis since many proteins and fragments of glycoproteins do not contain carbohydrate. This can simplify the analysis of complex biological samples such as serum (see below). The methods of the invention are advantageous for the determination of protein glycosylation in glycome studies and can be used to isolate and identify glycoproteins from cell membrane or body fluids to determine specific glycoprotein changes related to certain disease states or

cancer. The methods of the invention can be used for detecting quantitative changes in protein samples containing glycoproteins and to detect their extent of glycosylation. The methods of the invention are
5 applicable for the identification and/or characterization of diagnostic biomarkers, immunotherapy, or other diagnostic or therapeutic applications. The methods of the invention can also be used to evaluate the effectiveness of drugs during drug development, optimal
10 dosing, toxicology, drug targeting, and related therapeutic applications.

In one embodiment, the cis-diol groups of carbohydrates in glycoproteins can be oxidized by
15 periodate oxidation to give a di-aldehyde, which is reactive to a hydrazide gel with an agarose support to form covalent hydrazone bonds. The immobilized glycoproteins are subjected to protease digestion followed by extensive washing to remove the non-
20 glycosylated peptides. The immobilized glycopeptides are released from beads by chemicals or glycosidases. The isolated peptides are analyzed by mass spectrometry (MS), and the glycopeptide sequence and corresponding proteins are identified by MS/MS combined with a database search.
25 The glycopeptides can also be isotopically labeled, for example, at the amino or carboxyl termini to allow the quantities of glycopeptides from different biological samples to be compared.

30 The methods of the invention are based on selectively isolating glycosylated peptides, or peptides that were glycosylated in the original protein sample,

from a complex sample. The sample consists of peptide fragments of proteins generated, for example, by enzymatic digestion or chemical cleavage. A stable isotope tag is introduced into the isolated peptide fragments to facilitate mass spectrometric analysis and accurate quantification of the peptide fragments.

The invention provides a method for identifying and quantifying glycopolypeptides in a sample. The method can include the steps of derivatizing glycopolypeptides in a polypeptide sample, for example, by oxidation; immobilizing the derivatized glycopolypeptides to a solid support; cleaving the immobilized glycopolypeptides, thereby releasing non-glycosylated peptide fragments and retaining immobilized glycopeptide fragments; optionally labeling the immobilized glycopeptide fragments with an isotope tag; releasing the glycopeptide fragments from the solid support, thereby generating released glycopeptide fragments; analyzing the released glycopeptide fragments or their de-glycosylated counterparts using mass spectrometry; and quantifying the amount of the identified glycopeptide fragment. The released glycopolypeptides can be released with the carbohydrate still attached (the glycosylated form) or with the carbohydrate removed (the de-glycosylated form).

An embodiment of the present invention is depicted in Figure 1. A sample containing glycopolypeptides is chemically modified so that carbohydrates of the glycopolypeptides in the sample can be selectively bound to a solid support. For example, the glycopolypeptides can be bound covalently to a solid support by chemically

modifying the carbohydrate so that the carbohydrate can covalently bind to a reactive group on a solid support. In the embodiment depicted in Figure 1, the carbohydrates of the sample glycopolypeptides are oxidized. The carbohydrate can be oxidized, for example, to aldehydes. The oxidized moiety, such as an aldehyde moiety, of the glycopolypeptides can react with a solid support containing hydrazide or amine moieties, allowing covalent attachment of glycosylated polypeptides to a solid support via hydrazine chemistry. The sample glycopolypeptides are immobilized through the chemically modified carbohydrate, for example, the aldehyde, allowing the removal of non-glycosylated sample proteins by washing of the solid support. If desired, the immobilized glycopolypeptides can be denatured and/or reduced. The immobilized glycopolypeptides are cleaved into fragments using either protease or chemical cleavage. Cleavage results in the release of peptide fragments that do not contain carbohydrate and are therefore not immobilized. These released non-glycosylated peptide fragments optionally can be further characterized, if desired.

Following cleavage, glycosylated peptide fragments (glycopeptide fragments) remain bound to the solid support. To facilitate quantitative mass spectrometry (MS) analysis, immobilized glycopeptide fragments can be isotopically labeled. If it is desired to characterize most or all of the immobilized glycopeptide fragments, the isotope tagging reagent contains an amino or carboxyl reactive group so that the N-terminus or C-terminus of the glycopeptide fragments can be labeled (see Figures 1,

3 and 9). The immobilized glycopeptide fragments can be
cleaved from the solid support chemically or
enzymatically, for example, using glycosidases such as N-
glycanase (N-glycosidase) or O-glycanase (O-glycosidase).
5 The released glycopeptide fragments or their
deglycosylated forms can be analyzed, for example, using
MS.

As used herein, the term "polypeptide" refers to a
10 peptide or polypeptide of two or more amino acids. A
polypeptide can also be modified by naturally occurring
modifications such as post-translational modifications,
including phosphorylation, fatty acylation, prenylation,
sulfation, hydroxylation, acetylation, addition of
15 carbohydrate, addition of prosthetic groups or cofactors,
formation of disulfide bonds, proteolysis, assembly into
macromolecular complexes, and the like. A "peptide
fragment" is a peptide of two or more amino acids,
generally derived from a larger polypeptide.

20 As used herein, a "glycopolypeptide" or
"glycoprotein" refers to a polypeptide that contains a
covalently bound carbohydrate group. The carbohydrate
can be a monosaccharide, oligosaccharide or
25 polysaccharide. Proteoglycans are included within the
meaning of "glycopolypeptide." A glycopolypeptide can
additionally contain other post-translational
modifications. A "glycopeptide" refers to a peptide that
contains covalently bound carbohydrate. A "glycopeptide
30 fragment" refers to a peptide fragment resulting from
enzymatic or chemical cleavage of a larger polypeptide in
which the peptide fragment retains covalently bound

carbohydrate. It is understood that a glycopeptide fragment or peptide fragment refers to the peptides that result from a particular cleavage reaction, regardless of whether the resulting peptide was present before or after the cleavage reaction. Thus, a peptide that does not contain a cleavage site will be present after the cleavage reaction and is considered to be a peptide fragment resulting from that particular cleavage reaction. For example, if bound glycopeptides are cleaved, the resulting cleavage products retaining bound carbohydrate are considered to be glycopeptide fragments. The glycosylated fragments can remain bound to the solid support, and such bound glycopeptide fragments are considered to include those fragments that were not cleaved due to the absence of a cleavage site.

As disclosed herein, a glycopolypeptide or glycopeptide can be processed such that the carbohydrate is removed from the parent glycopolypeptide. It is understood that such an originally glycosylated polypeptide is still referred to herein as a glycopolypeptide or glycopeptide even if the carbohydrate is removed enzymatically and/or chemically. Thus, a glycopolypeptide or glycopeptide can refer to a glycosylated or de-glycosylated form of a polypeptide. A glycopolypeptide or glycopeptide from which the carbohydrate is removed is referred to as the de-glycosylated form of a polypeptide whereas a glycopolypeptide or glycopeptide which retains its carbohydrate is referred to as the glycosylated form of a polypeptide.

As used herein, the term "sample" is intended to mean any biological fluid, cell, tissue, organ or portion thereof, that includes one or more different molecules such as nucleic acids, polypeptides, or small molecules.

5 A sample can be a tissue section obtained by biopsy, or cells that are placed in or adapted to tissue culture. A sample can also be a biological fluid specimen such as blood, serum or plasma, cerebrospinal fluid, urine, saliva, seminal plasma, pancreatic juice, breast milk,
10 lung lavage, and the like. A sample can additionally be a cell extract from any species, including prokaryotic and eukaryotic cells as well as viruses. A tissue or biological fluid specimen can be further fractionated, if desired, to a fraction containing particular cell types.

15

As used herein, a "polypeptide sample" refers to a sample containing two or more different polypeptides. A polypeptide sample can include tens, hundreds, or even
20 thousands or more different polypeptides. A polypeptide sample can also include non-protein molecules so long as the sample contains polypeptides. A polypeptide sample can be a whole cell or tissue extract or can be a biological fluid. Furthermore, a polypeptide sample can
25 be fractionated using well known methods, as disclosed herein, into partially or substantially purified protein fractions.

The use of biological fluids such as a body fluid as
30 a sample source is particularly useful in methods of the invention. Biological fluid specimens are generally readily accessible and available in relatively large

quantities for clinical analysis. Biological fluids can be used to analyze diagnostic and prognostic markers for various diseases. In addition to ready accessibility, body fluid specimens do not require any prior knowledge
5 of the specific organ or the specific site in an organ that might be affected by disease. Because body fluids, in particular blood, are in contact with numerous body organs, body fluids "pick up" molecular signatures indicating pathology due to secretion or cell lysis
10 associated with a pathological condition. Body fluids also pick up molecular signatures that are suitable for evaluating drug dosage, drug targets and/or toxic effects, as disclosed herein.

15 Quantitative proteomics, defined as the comparison of relative protein changes in different proteomes, has been recognized as an important component of the emerging science of functional genomics. The technology is expected to facilitate the detection and identification
20 of diagnostic or prognostic disease markers, the discovery of proteins as therapeutic targets and to provide new functional insights into biological processes. Two methods have been used preferentially to generate quantitative profiles of complex protein
25 mixtures. The first and most commonly used is a combination of two-dimensional gel electrophoresis (2DE) and mass spectrometry (MS). The second is a more recently developed technique based on stable isotope tagging of proteins and automated peptide tandem mass
30 spectrometry (Oda et al., Proc. Natl. Acad. Sci. USA 96:6591-6596 (1999); Veenstra et al., J. Am. Soc. Mass. Spectrom. 11:78-82 (2000); Gygi et al., Nat. Biotechnol.

17:994-999 (1999)). To date, neither method has
succeeded in determining the complete proteome of any
species. This is mainly due to the "top down" mode of
operation of either method in which the most abundant
5 proteins are preferentially or exclusively analyzed.

Given the complexities of global proteome analysis,
several studies have adopted a "divide and conquer"
strategy to handle the "top down" problem by
10 comprehensively analyzing specific subsets of the
proteome that are selectively isolated. Such studies
include the analysis of functional multiprotein complexes
such as the ribosome (Link et al., Nat. Biotechnol.
17:676-682 (1999)), spliceosome (Rappsilber et al.,
15 Genome Res. 12:1231-1245 (2002); Zhou et al., Nature
419:182-185 (2002)), and nuclear pore complex (Rout et
al., J. Cell Biol. 148:635-651 (2000)), or organelles,
such as mitochondria (Fountoulakis et al.,
Electrophoresis 23:311-328 (2002)), peroxisomes (Yi et
20 al., Electrophoresis 23:3205-3216 (2002)), microsomes
(Han et al., Nat. Biotechnol. 19:946-951 (2001)) and
nuclei (Bergquist et al., J. Neurosci. Methods 109:3-11
(2001)). Alternatively, proteins that contain common
distinguishing structural features, such as phosphate
25 ester groups ((Ficarro et al., Nat. Biotechnol. 20:301-
305 (2002); Oda et al., Nat. Biotechnol. 19:379-382
(2001); Zhou et al., Nat. Biotechnol. 19:375-378 (2001)),
cysteine residues (Gygi et al. supra (1999); Spahr et
al., Electrophoresis 21:1635-1650 (2000)) or have the
30 ability to specifically bind to certain compounds
(Haystead et al., Eur. J. Biochem. 214:459-467 (1993);
Adam et al., Nat. Biotechnol. 20:805-809 (2002)) have

been selectively enriched prior to MS analysis. These strategies have in common that they focus on the in-depth analysis of sub-proteomes of rich biological context, thus minimizing the repeated analyses of abundantly
5 expressed proteins.

The methods of the invention utilize the selective isolation of glycopolypeptides coupled with chemical modification to facilitate MS analysis. Proteins are
10 glycosylated by complex enzymatic mechanisms, typically at the side chains of serine or threonine residues (O-linked) or the side chains of asparagine residues (N-linked). N-linked glycosylation sites generally fall into a sequence motif that can be described as N-X-S/T,
15 where X can be any amino acid except proline. Glycosylation plays an important function in many biological processes (reviewed in Helenius and Aebi, Science 291:2364-2369 (2001); Rudd et al., Science 291:2370-2375 (2001)).

20 Protein glycosylation has long been recognized as a very common post-translational modification. As discussed above, carbohydrates are linked to serine or threonine residues (O-linked glycosylation) or to
25 asparagine residues (N-linked glycosylation) (Varki et al. Essentials of Glycobiology Cold Spring Harbor Laboratory (1999)). Protein glycosylation, and in particular N-linked glycosylation, is prevalent in proteins destined for extracellular environments (Roth,
30 Chem. Rev. 102:285-303 (2002)). These include proteins on the extracellular side of the plasma membrane, secreted proteins, and proteins contained in body fluids,

for example, blood serum, cerebrospinal fluid, urine,
breast milk, saliva, lung lavage fluid, pancreatic juice,
and the like. These also happen to be the proteins in
the human body that are most easily accessible for
5 diagnostic and therapeutic purposes.

Due to the ready accessibility of body fluids
exposed to the extracellular surface of cells and the
presence of secreted proteins in these fluids, many
10 clinical biomarkers and therapeutic targets are
glycoproteins. These include Her2/neu in breast cancer,
human chorionic gonadotropin and α -fetoprotein in germ
cell tumors, prostate-specific antigen in prostate
cancer, and CA125 in ovarian cancer. The Her2/neu
15 receptor is also the target for a successful
immunotherapy of breast cancer using the humanized
monoclonal antibody Herceptin (Shepard et al., J. Clin.
Immunol. 11:117-127 (1991)). In addition, changes in the
extent of glycosylation and the carbohydrate structure of
20 proteins on the cell surface and in body fluids have been
shown to correlate with cancer and other disease states,
highlighting the clinical importance of this modification
as an indicator or effector of pathologic mechanisms
(Durand and Seta, Clin. Chem. 46:795-805 (2000); Freeze,
25 Glycobiology 11:129R-143R (2001); Spiro, Glycobiology
12:43R-56R (2002)). Therefore, a method for the
systematic and quantitative analysis of glycoproteins
would be of significance for the detection of new
potential diagnostic markers and therapeutic targets.

30

Disclosed herein is a method for quantitative
glycoprotein profiling. In one embodiment, the method is

based on the conjugation of glycoproteins to a solid support using hydrazide chemistry, stable isotope labeling of glycopeptides, and the specific release of formerly N-linked glycosylated peptides via Peptide-N-Glycosidase F (PNGase F). The recovered peptides are then identified and quantified by tandem mass spectrometry (MS/MS). The method was applied to the analysis of cell surface and serum proteins, as disclosed herein.

10

To selectively isolate glycopolypeptides, the methods utilize chemistry and/or binding interactions that are specific for carbohydrate moieties. Selective binding of glycopolypeptides refers to the preferential binding of glycopolypeptides over non-glycosylated peptides, as demonstrated in Example II. The methods of the invention can utilize covalent coupling of glycopolypeptides, which is particularly useful for increasing the selective isolation of glycopolypeptides by allowing stringent washing to remove non-specifically bound, non-glycosylated polypeptides.

The carbohydrate moieties of a glycopolypeptide are chemically or enzymatically modified to generate a reactive group that can be selectively bound to a solid support having a corresponding reactive group. In the embodiment depicted in Figure 2, the carbohydrates of glycopolypeptides are oxidized to aldehydes. The oxidation can be performed, for example, with sodium periodate. The hydroxyl groups of a carbohydrate can also be derivatized by epoxides or oxiranes, alkyl halogen, carbonyldiimidazoles, N,N'-disuccinimidyl

25
30

carbonates, N-hydroxycuccinimidyl chloroformates, and the like. The hydroxyl groups of a carbohydrate can also be oxidized by enzymes to create reactive groups such as aldehyde groups. For example, galactose oxidase oxidizes
5 terminal galactose or N-acetyl-D-galactose residues to form C-6 aldehyde groups. These derivatized groups can be conjugated to amine- or hydrazide- containing moieties.

The oxidation of hydroxyl groups to aldehyde using
10 sodium periodate is specific for the carbohydrate of a glycopeptide. Sodium periodate can oxidize hydroxyl groups on adjacent carbon atoms, forming an aldehyde for coupling with amine- or hydrazide-containing molecules. Sodium periodate also reacts with hydroxylamine
15 derivatives, compounds containing a primary amine and a secondary hydroxyl group on adjacent carbon atoms. This reaction is used to create reactive aldehydes on N-terminal serine residues of peptides. A serine residue is rare at the N-terminus of a protein. The oxidation to
20 an aldehyde using sodium periodate is therefore specific for the carbohydrate groups of a glycopolypeptide.

Once the carbohydrate of a glycopolypeptide is modified, for example, by oxidation to aldehydes, the
25 modified carbohydrates can bind to a solid support containing hydrazide or amine moieties, such as the hydrazide resin depicted in Figure 2. Although illustrated with oxidation chemistry and coupling to hydrazide, it is understood that any suitable chemical
30 modifications and/or binding interactions that allows specific binding of the carbohydrate moieties of a glycopolypeptide can be used in methods of the invention.

The binding interactions of the glycopolypeptides with the solid support are generally covalent, although non-covalent interactions can also be used so long as the glycopolypeptides or glycopeptide fragments remain bound
5 during the digestion, washing and other steps of the methods.

The methods of the invention can also be used to select and characterize subgroups of carbohydrates.
10 Chemical modifications or enzymatic modifications using, for example, glycosidases can be used to isolate subgroups of carbohydrates. For example, the concentration of sodium periodate can be modulated so that oxidation occurs on sialic acid groups of
15 glycoproteins. In particular, a concentration of about 1 mM of sodium periodate at 0°C can be used to essentially exclusively modify sialic acid groups.

Glycopolypeptides containing specific
20 monosaccharides can be targeted using a selective sugar oxidase to generate aldehyde functions, such as the galactose oxidase described above or other sugar oxidases. Furthermore, glycopolypeptides containing a subgroup of carbohydrates can be selected after the
25 glycopolypeptides are bound to a solid support. For example, glycopeptides bound to a solid support can be selectively released using different glycosidases having specificity for particular monosaccharide structures.

30 The glycopolypeptides are isolated by binding to a solid support. The solid support can be, for example, a bead, resin, membrane or disk, or any solid support

material suitable for methods of the invention. An advantage of using a solid support to bind the glycopolypeptides is that it allows extensive washing to remove non-glycosylated polypeptides. Thus, in the case of complex samples containing a multitude of polypeptides, the analysis can be simplified by isolating glycopolypeptides and removing the non-glycosylated polypeptides, thus reducing the number of polypeptides to be analyzed.

The glycopolypeptides can also be conjugated to an affinity tag through an amine group, such as biotin hydrazide. The affinity tagged glycopeptides can then be immobilized to the solid support, for example, an avidin or streptavidin solid support, and the non-glycosylated peptides are removed. The glycopeptides immobilized on the solid support can be cleaved by a protease, and the non-glycosylated peptide fragments can be removed by washing. The tagged glycopeptides can be released from the solid support by enzymatic or chemical cleavage. Alternatively, the tagged glycopeptides can be released from the solid support with the oligosaccharide and affinity tag attached (see Example XV and Figures 28 and 29).

Another advantage of binding the glycopolypeptides to the solid support is that it allows further manipulation of the sample molecules without the need for additional purification steps that can result in loss of sample molecules. For example, the methods of the invention can involve the steps of cleaving the bound glycopolypeptides as well as adding an isotope tag, or

other desired modifications of the bound glycopolypeptides. Because the glycopolypeptides are bound, these steps can be carried out on solid phase while allowing excess reagents to be removed as well as
5 extensive washing prior to subsequent manipulations.

The bound glycopolypeptides can be cleaved into peptide fragments to facilitate MS analysis. Thus, a polypeptide molecule can be enzymatically cleaved with
10 one or more proteases into peptide fragments. Exemplary proteases useful for cleaving polypeptides include trypsin, chymotrypsin, pepsin, papain, *Staphylococcus aureus* (V8) protease, Submaxillaris protease, bromelain, thermolysin, and the like. In certain applications,
15 proteases having cleavage specificities that cleave at fewer sites, such as sequence-specific proteases having specificity for a sequence rather than a single amino acid, can also be used, if desired. Polypeptides can also be cleaved chemically, for example, using CNBr, acid
20 or other chemical reagents. A particularly useful cleavage reagent is the protease trypsin. One skilled in the art can readily determine appropriate conditions for cleavage to achieve a desired efficiency of peptide cleavage.

25
Cleavage of the bound glycopolypeptides is particularly useful for MS analysis in that one or a few peptides are generally sufficient to identify a parent polypeptide. However, it is understood that cleavage of
30 the bound glycopolypeptides is not required, in particular where the bound glycopolypeptide is relatively small and contains a single glycosylation site.

Furthermore, the cleavage reaction can be carried out after binding of glycopolypeptides to the solid support, allowing characterization of non-glycosylated peptide fragments derived from the bound glycopolypeptide.

5 Alternatively, the cleavage reaction can be carried out prior to addition of the glycopeptides to the solid support. One skilled in the art can readily determine the desirability of cleaving the sample polypeptides and an appropriate point to perform the cleavage reaction, as
10 needed for a particular application of the methods of the invention.

If desired, the bound glycopolypeptides can be denatured and optionally reduced. Denaturing and/or
15 reducing the bound glycopolypeptides can be useful prior to cleavage of the glycopolypeptides, in particular protease cleavage, because this allows access to protease cleavage sites that can be masked in the native form of the glycopolypeptides. The bound glycopeptides can be
20 denatured with detergents and/or chaotropic agents. Reducing agents such as β -mercaptoethanol, dithiothreitol, tris-carboxyethylphosphine (TCEP), and the like, can also be used, if desired. As discussed above, the binding of the glycopolypeptides to a solid
25 support allows the denaturation step to be carried out followed by extensive washing to remove denaturants that could inhibit the enzymatic or chemical cleavage reactions. The use of denaturants and/or reducing agents can also be used to dissociate protein complexes in which
30 non-glycosylated proteins form complexes with bound glycopolypeptides. Thus, the use of these agents can be used to increase the specificity for glycopolypeptides by

washing away non-glycosylated polypeptides from the solid support.

5 Treatment of the bound glycopolypeptides with a cleavage reagent results in the generation of peptide fragments. Because the carbohydrate moiety is bound to the solid support, those peptide fragments that contain the glycosylated residue remain bound to the solid support. Following cleavage of the bound
10 glycopolypeptides, glycopeptide fragments remain bound to the solid support via binding of the carbohydrate moiety. Peptide fragments that are not glycosylated are released from the solid support. If desired, the released non-glycosylated peptides can be analyzed, as described in
15 more detail below.

 The methods of the invention can be used to identify and/or quantify the amount of a glycopolypeptide present in a sample. A particularly useful method for
20 identifying and quantifying a glycopolypeptide is mass spectrometry (MS). The methods of the invention can be used to identify a glycopolypeptide qualitatively, for example, using MS analysis. If desired, an isotope tag can be added to the bound glycopeptide fragments, in
25 particular to facilitate quantitative analysis by MS.

 As used herein an "isotope tag" refers to a chemical moiety having suitable chemical properties for
30 incorporation of an isotope, allowing the generation of chemically identical reagents of different mass which can be used to differentially tag a polypeptide in two samples. The isotope tag also has an appropriate

composition to allow incorporation of a stable isotope at one or more atoms. A particularly useful stable isotope pair is hydrogen and deuterium, which can be readily distinguished using mass spectrometry as light and heavy forms, respectively. Any of a number of isotopic atoms can be incorporated into the isotope tag so long as the heavy and light forms can be distinguished using mass spectrometry, for example, ^{13}C , ^{15}N , ^{17}O , ^{18}O or ^{34}S . Exemplary isotope tags include the 4,7,10-trioxa-1,13-tridecanediamine based linker and its related deuterated form, 2,2',3,3',11,11',12,12'-octadeutero-4,7,10-trioxa-1,13-tridecanediamine, described by Gygi et al. (Nature Biotechnol. 17:994-999 (1999)). Other exemplary isotope tags have also been described previously (see WO 00/11208, which is incorporated herein by reference).

In contrast to these previously described isotope tags related to an ICAT-type reagent, it is not required that an affinity tag be included in the reagent since the glycopolypeptides are already isolated. One skilled in the art can readily determine any of a number of appropriate isotope tags useful in methods of the invention. An isotope tag can be an alkyl, akenyl, alkynyl, alkoxy, aryl, and the like, and can be optionally substituted, for example, with O, S, N, and the like, and can contain an amine, carboxyl, sulfhydryl, and the like (see WO 00/11208). Exemplary isotope tags include succinic anhydride, isatoic-anhydride, N-methyl-isatoic-anhydride, glyceraldehyde, Boc-Phe-OH, benzaldehyde, salicylaldehyde, and the like (Figure 3). In addition to Phe, as shown in Figures 3 and 9, other amino acids similarly can be used as isotope tags.

Furthermore, small organic aldehydes, similar to those shown in Figure 3, can be used as isotope tags. These and other derivatives can be made in the same manner as that disclosed herein using methods well known to those skilled in the art. One skilled in the art will readily recognize that a number of suitable chemical groups can be used as an isotope tag so long as the isotope tag can be differentially isotopically labeled.

The bound glycopeptide fragments are tagged with an isotope tag to facilitate MS analysis. In order to tag the glycopeptide fragments, the isotope tag contains a reactive group that can react with a chemical group on the peptide portion of the glycopeptide fragments. A reactive group is reactive with and therefore can be covalently coupled to a molecule in a sample such as a polypeptide. Reactive groups are well known to those skilled in the art (see, for example, Hermanson, Bioconjugate Techniques, pp. 3-166, Academic Press, San Diego (1996); Glazer et al., Laboratory Techniques in Biochemistry and Molecular Biology: Chemical Modification of Proteins, Chapter 3, pp. 68-120, Elsevier Biomedical Press, New York (1975); Pierce Catalog (1994), Pierce, Rockford IL). Any of a variety of reactive groups can be incorporated into an isotope tag for use in methods of the invention so long as the reactive group can be covalently coupled to the immobilized polypeptide.

To analyze a large number or essentially all of the bound glycopolypeptides, it is desirable to use an isotope tag having a reactive group that will react with the majority of the glycopeptide fragments. For example,

a reactive group that reacts with an amino group can react with the free amino group at the N-terminus of the bound glycopeptide fragments. If a cleavage reagent is chosen that leaves a free amino group of the cleaved peptides, such an amino group reactive agent can label a large fraction of the peptide fragments. Only those with a blocked N-terminus would not be labeled. Similarly, a cleavage reagent that leaves a free carboxyl group on the cleaved peptides can be modified with a carboxyl reactive group, resulting in the labeling of many if not all of the peptides. Thus, the inclusion of amino or carboxyl reactive groups in an isotope tag is particularly useful for methods of the invention in which most if not all of the bound glycopeptide fragments are desired to be analyzed.

In addition, a polypeptide can be tagged with an isotope tag via a sulfhydryl reactive group, which can react with free sulfhydryls of cysteine or reduced cysteines in a polypeptide. An exemplary sulfhydryl reactive group includes an iodoacetamido group (see Gygi et al., supra, 1999). Other exemplary sulfhydryl reactive groups include maleimides, alkyl and aryl halides, haloacetyls, α -haloacyls, pyridyl disulfides, aziridines, acryloyls, arylating agents and thiomethylsulfones.

A reactive group can also react with amines such as the α -amino group of a peptide or the ϵ -amino group of the side chain of Lys, for example, imidoesters, N-hydroxysuccinimidyl esters (NHS), isothiocyanates, isocyanates, acyl azides, sulfonyl chlorides, aldehydes,

ketones, glyoxals, epoxides (oxiranes), carbonates, arylating agents, carbodiimides, anhydrides, and the like. A reactive group can also react with carboxyl groups found in Asp or Glu or the C-terminus of a peptide, for example, diazoalkanes, diazoacetyls, carbonyldiimidazole, carbodiimides, and the like. A reactive group that reacts with a hydroxyl group includes, for example, epoxides, oxiranes, carbonyldiimidazoles, N,N'-disuccinimidyl carbonates, N-hydroxysuccinimidyl chloroformates, and the like. A reactive group can also react with amino acids such as histidine, for example, α -haloacids and amides; tyrosine, for example, nitration and iodination; arginine, for example, butanedione, phenylglyoxal, and nitromalondialdehyde; methionine, for example, iodoacetic acid and iodoacetamide; and tryptophan, for example, 2-(2-nitrophenylsulfonyl)-3-methyl-3-bromoindolenine (BNPS-skatole), N-bromosuccinimide, formylation, and sulfonylation (Glazer et al., *supra*, 1975). In addition, a reactive group can also react with a phosphate group for selective labeling of phosphopeptides (Zhou et al., Nat. Biotechnol., 19:375-378 (2001)) or with other covalently modified peptides, including lipopeptides, or any of the known covalent polypeptide modifications. One skilled in the art can readily determine conditions for modifying sample molecules by using various reagents, incubation conditions and time of incubation to obtain conditions suitable for modification of a molecule with an isotope tag. The use of covalent-chemistry based isolation methods is particularly useful due to the highly specific nature of the binding of the glycopolypeptides.

The reactive groups described above can form a covalent bond with the target sample molecule. However, it is understood that an isotope tag can contain a reactive group that can non-covalently interact with a sample molecule so long as the interaction has high specificity and affinity.

Prior to further analysis, it is generally desirable to release the bound glycopeptide fragments. The glycopeptide fragments can be released by cleaving the fragments from the solid support, either enzymatically or chemically. For example, glycosidases such as N-glycosidases and O-glycosidases can be used to cleave an N-linked or O-linked carbohydrate moiety, respectively, and release the corresponding de-glycosylated peptide(s). If desired, N-glycosidases and O-glycosidases can be added together or sequentially, in either order. The sequential addition of an N-glycosidase and an O-glycosidase allows differential characterization of those released peptides that were N-linked versus those that were O-linked, providing additional information on the nature of the carbohydrate moiety and the modified amino acid residue. Thus, N-linked and O-linked glycosylation sites can be analyzed sequentially and separately on the same sample, increasing the information content of the experiment and simplifying the complexity of the samples being analyzed.

In addition to N-glycosidases and O-glycosidases, other glycosidases can be used to release a bound glycopolypeptide. For example, exoglycosidases can be

used. Exoglycosidases are anomeric, residue and linkage specific for terminal monnosaccharides and can be used to release peptides having the corresponding carbohydrate.

5

In addition to enzymatic cleavage, chemical cleavage can also be used to cleave a carbohydrate moiety to release a bound peptide. For example, O-linked oligosaccharides can be released specifically from a polypeptide via a β -elimination reaction catalyzed by alkali. The reaction can be carried out in about 50 mM NaOH containing about 1 M NaBH₄ at about 55°C for about 12 hours. The time, temperature and concentration of the reagents can be varied so long as a sufficient β -elimination reaction is carried out for the needs of the experiment.

In one embodiment, N-linked oligosaccharides can be released from glycopolypeptides, for example, by hydrazinolysis. Glycopolypeptides can be dried in a desiccator over P₂O₅ and NaOH. Anhydrous hydrazine is added and heated at about 100°C for 10 hours, for example, using a dry heat block.

In addition to using enzymatic or chemical cleavage to release a bound glycopeptide, the solid support can be designed so that bound molecules can be released, regardless of the nature of the bound carbohydrate. The reactive group on the solid support, to which the glycopolypeptide binds, can be linked to the solid support with a cleavable linker. For example, the solid support reactive group can be covalently bound to the

solid support via a cleavable linker such as a photocleavable linker. Exemplary photocleavable linkers include, for example, linkers containing o-nitrobenzyl, desyl, trans-o-cinnamoyl, m-nitrophenyl, benzylsulfonyl groups (see, for example, Dorman and Prestwich, Trends Biotech. 18:64-77 (2000); Greene and Wuts, Protective Groups in Organic Synthesis, 2nd ed., John Wiley & Sons, New York (1991); U.S. Patent Nos. 5,143,854; 5,986,076; 5,917,016; 5,489,678; 5,405,783). Similarly, the reactive group can be linked to the solid support via a chemically cleavable linker. Release of glycopeptide fragments with the intact carbohydrate is particularly useful if the carbohydrate moiety is to be characterized using well known methods, including mass spectrometry. The use of glycosidases to release de-glycosylated peptide fragments also provides information on the nature of the carbohydrate moiety.

Thus, the invention provides methods for identifying a glycopolypeptide and, furthermore, identifying its glycosylation site. The methods of the invention are applied, as disclosed herein, and the parent glycopolypeptide is identified. The glycosylation site itself can also be identified and consensus motifs determined (Example VII), as well as the carbohydrate moiety, as disclosed herein. The invention further provides glycopolypeptides, glycopeptides and glycosylation sites identified by the methods of the invention.

Glycopolypeptides from a sample are bound to a solid support via the carbohydrate moiety. The bound

glycopolypeptides are generally cleaved, for example, using a protease, to generate glycopeptide fragments. As discussed above, a variety of methods can be used to release the bound glycopeptide fragments, thereby
5 generating released glycopeptide fragments. As used herein, a "released glycopeptide fragment" refers to a peptide which was bound to a solid support via a covalently bound carbohydrate moiety and subsequently released from the solid support, regardless of whether
10 the released peptide retains the carbohydrate. In some cases, the method by which the bound glycopeptide fragments are released results in cleavage and removal of the carbohydrate moiety, for example, using glycosidases or chemical cleavage of the carbohydrate moiety. If the
15 solid support is designed so that the reactive group, for example, hydrazide, is attached to the solid support via a cleavable linker, the released glycopeptide fragment retains the carbohydrate moiety. It is understood that, regardless whether a carbohydrate moiety is retained or
20 removed from the released peptide, such peptides are referred to as released glycopeptide fragments.

After isolating glycopolypeptides from a sample and cleaving the glycopolypeptide into fragments, the
25 glycopeptide fragments released from the solid support and the released glycopeptide fragments are identified and/or quantified. A particularly useful method for analysis of the released glycopeptide fragments is mass spectrometry. A variety of mass spectrometry systems can
30 be employed in the methods of the invention for identifying and/or quantifying a sample molecule such as a released glycopolypeptide fragment. Mass analyzers

with high mass accuracy, high sensitivity and high resolution include, but are not limited to, ion trap, triple quadrupole, and time-of-flight, quadrupole time-of-flight mass spectrometers and Fourier transform ion cyclotron mass analyzers (FT-ICR-MS). Mass spectrometers
5 are typically equipped with matrix-assisted laser desorption (MALDI) and electrospray ionization (ESI) ion sources, although other methods of peptide ionization can also be used. In ion trap MS, analytes are ionized by
10 ESI or MALDI and then put into an ion trap. Trapped ions can then be separately analyzed by MS upon selective release from the ion trap. Fragments can also be generated in the ion trap and analyzed. Sample molecules such as released glycopeptide fragments can be analyzed,
15 for example, by single stage mass spectrometry with a MALDI-TOF or ESI-TOF system. Methods of mass spectrometry analysis are well known to those skilled in the art (see, for example, Yates, J. Mass Spect. 33:1-19 (1998); Kinter and Sherman, Protein Sequencing and Identification Using Tandem Mass Spectrometry, John Wiley & Sons, New York (2000); Aebersold and Goodlett, Chem. Rev. 101:269-295 (2001)).
20

For high resolution polypeptide fragment separation,
25 liquid chromatography ESI-MS/MS or automated LC-MS/MS, which utilizes capillary reverse phase chromatography as the separation method, can be used (Yates et al., Methods Mol. Biol. 112:553-569 (1999)). Data dependent collision-induced dissociation (CID) with dynamic
30 exclusion can also be used as the mass spectrometric method (Goodlett, et al., Anal. Chem. 72:1112-1118 (2000)).

Once a peptide is analyzed by MS/MS, the resulting CID spectrum can be compared to databases for the determination of the identity of the isolated glycopeptide. Methods for protein identification using single peptides has been described previously (Aebersold and Goodlett, Chem. Rev. 101:269-295 (2001); Yates, J. Mass Spec. 33:1-19 (1998)). In particular, it is possible that one or a few peptide fragments can be used to identify a parent polypeptide from which the fragments were derived if the peptides provide a unique signature for the parent polypeptide. Thus, identification of a single glycopeptide, alone or in combination with knowledge of the site of glycosylation, can be used to identify a parent glycopolypeptide from which the glycopeptide fragments were derived. Further information can be obtained by analyzing the nature of the attached tag and the presence of the consensus sequence motif for carbohydrate attachment. For example, if peptides are modified with an N-terminal tag, each released glycopeptide has the specific N-terminal tag, which can be recognized in the fragment ion series of the CID spectra. Furthermore, the presence of a known sequence motif that is found, for example, in N-linked carbohydrate-containing peptides, that is, the consensus sequence NX_S/T, can be used as a constraint in database searching of N-glycosylated peptides.

In addition, the identity of the parent glycopolypeptide can be determined by analysis of various characteristics associated with the peptide, for example, its resolution on various chromatographic media or using

various fractionation methods. These empirically determined characteristics can be compared to a database of characteristics that uniquely identify a parent polypeptide, which defines a peptide tag.

5

The use of a peptide tag and related database is used for identifying a polypeptide from a population of polypeptides by determining characteristics associated with a polypeptide, or a peptide fragment thereof, comparing the determined characteristics to a polypeptide identification index, and identifying one or more polypeptides in the polypeptide identification index having the same characteristics (see WO 02/052259). The methods are based on generating a polypeptide identification index, which is a database of characteristics associated with a polypeptide. The polypeptide identification index can be used for comparison of characteristics determined to be associated with a polypeptide from a sample for identification of the polypeptide. Furthermore, the methods can be applied not only to identify a polypeptide but also to quantitate the amount of specific proteins in the sample.

The methods for identifying a polypeptide are applicable to performing quantitative proteome analysis, or comparisons between polypeptide populations that involve both the identification and quantification of sample polypeptides. Such a quantitative analysis can be conveniently performed in two separate stages, if desired. As a first step, a reference polypeptide index is generated representative of the samples to be tested, for example, from a species, cell type or tissue type under investigation, such as a glycopolypeptide sample,

30

as disclosed herein. The second step is the comparison of characteristics associated with an unknown polypeptide with the reference polypeptide index or indices previously generated.

5

A reference polypeptide index is a database of polypeptide identification codes representing the polypeptides of a particular sample, such as a cell, subcellular fraction, tissue, organ or organism. A polypeptide identification index can be generated that is representative of any number of polypeptides in a sample, including essentially all of the polypeptides potentially expressed in a sample. In methods of the invention directed to identifying glycopolypeptides, the polypeptide identification index is determined for a desired sample such as a serum sample. Once a polypeptide identification index has been generated, the index can be used repeatedly to identify one or more polypeptides in a sample, for example, a sample from an individual potentially having a disease. Thus, a set of characteristics can be determined for glycopeptides that can be correlated with a parent glycopolypeptide, including the amino acid sequence of the glycopeptide, and stored as an index, which can be referenced in a subsequent experiment on a sample treated in substantially the same manner as when the index was generated.

The incorporation of an isotope tag can be used to facilitate quantification of the sample glycopolypeptides. As disclosed previously, the incorporation of an isotope tag provides a method for

30

quantifying the amount of a particular molecule in a sample (Gygi et al., supra, 1999; WO 00/11208). In using an isotope tag, differential isotopes can be incorporated, which can be used to compare a known amount
5 of a standard labeled molecule having a differentially labeled isotope tag from that of a sample molecule, as described in more detail below (see Example XIII). Thus, a standard peptide having a differential isotope can be added at a known concentration and analyzed in the same
10 MS analysis or similar conditions in a parallel MS analysis. A specific, calibrated standard can be added with known absolute amounts to determine an absolute quantity of the glycopolypeptide in the sample. In addition, the standards can be added so that relative
15 quantitation is performed, as described below.

Alternatively, parallel glycosylated sample molecules can be labeled with a different isotopic label and compared side-by-side (see Gygi et al., supra, 1999).
20 This is particularly useful for qualitative analysis or quantitative analysis relative to a control sample. For example, a glycosylated sample derived from a disease state can be compared to a glycosylated sample from a non-disease state by differentially labeling the two
25 samples, as described previously (Gygi et al., supra, 1999). Such an approach allows detection of differential states of glycosylation, which is facilitated by the use of differential isotope tags for the two samples, and can thus be used to correlate differences in glycosylation as
30 a diagnostic marker for a disease (see Examples VIII, IX, XI and XII).

The methods of the invention provide numerous advantages for the analysis of complex biological and clinical samples. From every glycoprotein present in a complex sample, only a few peptides will be isolated since only a few peptides of a glycoprotein are glycosylated. Therefore, by isolating glycopeptide fragments, the composition of the resulting peptide mixture is significantly simplified for mass spectrometric analysis. For example, every protein on average will produce dozens of tryptic peptides but only one to a few tryptic glycosylated peptides. For example, the number of glycopeptides is significantly lower than the number of tryptic peptides or Cys-containing peptides in the major plasma proteins (see Table 1). Thus, analysis of glycopolypeptides or glycopeptides reduces the complexity of complex biological samples, for example, serum.

Table 1
Five major plasma proteins represent more than 80% total protein

Protein	Number of Peptides		
	Tryptic	Cys	Glyco
Albumin (40mg/ml)	82	27	0
α 1-antitrypsin (3mg/ml)	39	2	3
α 2-macroglobulin	125	24	8
Transferrin	79	37	2
γ -Globulin	15	4	2
Total	340	94	15

Another advantage of the methods of the invention is the use for analysis of body fluids as a clinical specimen, in particular serum. Five major plasma proteins represent more than 80% of the total protein in

plasma, albumin, α 1 antitrypsin, α 2 macroglobulin, transferrin, and γ -globulins. Of these, albumin is the most abundant protein in blood serum and other body fluids, constituting about 50% of the total protein in plasma. However, albumin is essentially transparent to the methods of the invention due to the lack of N-glycosylation. For example, no tryptic N-glycosylated peptides from albumin were observed when the methods of the invention were applied and a N-glycosidase was used to release the N-linked glycopeptides. This is all the more significant because more than 50 different albumin species have been detected by 2D gel electrophoresis that collectively obscure a significant part of the gel pattern and the analysis of less abundant serum proteins having clinical significance. Therefore, the methods of the invention that allow analysis of glycosylated proteins compensate for the dominance of albumin in serum and allow the analysis of less abundant, glycosylated proteins present in serum. As disclosed herein, the methods of the invention allowed the identification of many more serum proteins compared to conventional methods (see Example II). The methods of the invention also allow the analysis of less abundant serum proteins. These low abundance serum proteins are potential diagnostic markers. Such markers can be readily determined by comparing disease samples with healthy samples, as disclosed herein (see Examples VIII, IX, XI and XII).

Additionally, the known sequence motif for N-glycosylation (N-X-S/T) serves as a powerful sequence database search constraint for the identification of the

isolated peptides. This can be used to facilitate the identification of the polypeptide from which the glycopeptide fragment was derived since a smaller number of possible peptides will contain the glycosylation motif.

The methods of the invention are also advantageous because they allow fast throughput and simplicity. Accordingly, the methods can be readily adapted for high throughput analysis of samples, which can be particularly advantageous for the analysis of clinical samples. Furthermore, the methods of the invention can be automated to facilitate the processing of multiple samples (see Example XVI). As disclosed herein, a robotic workstation has been adapted for automated glycoprotein analysis (Example XVI).

In addition to the analysis of body fluids for the reasons described above, the methods of the invention are also advantageous for the analysis of proteins contained in the plasma membrane. The methods of the invention allow for the selective separation of cell surface proteins and secreted proteins based on the fact that the proteins most likely contaminating such specimens, intracellular proteins, are very unlikely to be glycosylated. Thus, the methods of the invention can be used to more accurately reflect proteins representative of the sample rather than contaminants from cell lysis.

Such an analysis can be optionally combined with subcellular fractionation for the analysis of glycopolypeptides (Example IV).

As described above, non-glycosylated peptide fragments are released from the solid support after proteolytic or chemical cleavage (see Figure 1). If desired, the released peptide fragments can be characterized to provide further information on the nature of the glycopolypeptides isolated from the sample. A particularly useful method is the use of the isotope-coded affinity tag (ICATTM) method (Gygi et al., Nature Biotechnol. 17:994-999 (1999) which is incorporated herein by reference). The ICATTM type reagent method uses an affinity tag that can be differentially labeled with an isotope that is readily distinguished using mass spectrometry. The ICATTM type affinity reagent consists of three elements, an affinity tag, a linker and a reactive group.

One element of the ICATTM type affinity reagent is an affinity tag that allows isolation of peptides coupled to the affinity reagent by binding to a cognate binding partner of the affinity tag. A particularly useful affinity tag is biotin, which binds with high affinity to its cognate binding partner avidin, or related molecules such as streptavidin, and is therefore stable to further biochemical manipulations. Any affinity tag can be used so long as it provides sufficient binding affinity to its cognate binding partner to allow isolation of peptides coupled to the ICATTM type affinity reagent. An affinity tag can also be used to isolate a tagged peptide with magnetic beads or other magnetic format suitable to isolate a magnetic affinity tag. In the ICATTM type reagent method, or any other method of affinity tagging a peptide, the use of covalent trapping, for example, using

a cross-linking reagent, can be used to bind the tagged peptides to a solid support, if desired.

A second element of the ICATTM type affinity reagent is a linker that can incorporate a stable isotope. The linker has a sufficient length to allow the reactive group to bind to a specimen polypeptide and the affinity tag to bind to its cognate binding partner. The linker also has an appropriate composition to allow incorporation of a stable isotope at one or more atoms. A particularly useful stable isotope pair is hydrogen and deuterium, which can be readily distinguished using mass spectrometry as light and heavy forms, respectively. Any of a number of isotopic atoms can be incorporated into the linker so long as the heavy and light forms can be distinguished using mass spectrometry. Exemplary linkers include the 4,7,10-trioxa-1,13-tridecanediamine based linker and its related deuterated form, 2,2',3,3',11,11',12,12'-octadeutero-4,7,10-trioxa-1,13-tridecanediamine, described by Gygi et al. (supra, 1999). One skilled in the art can readily determine any of a number of appropriate linkers useful in an ICATTM type affinity reagent that satisfy the above-described criteria, as described above for the isotope tag.

The third element of the ICATTM type affinity reagent is a reactive group, which can be covalently coupled to a polypeptide in a specimen. Various reactive groups have been described above with respect to the isotope tag and can similarly be incorporated into an ICAT-type reagent.

The ICAT™ method or other similar methods can be applied to the analysis of the non-glycosylated peptide fragments released from the solid support. Alternatively, the ICAT™ method or other similar methods
5 can be applied prior to cleavage of the bound glycopolypeptides, that is, while the intact glycopolypeptide is still bound to the solid support.

10 The method generally involves the steps of automated tandem mass spectrometry and sequence database searching for peptide/protein identification; stable isotope tagging for quantification by mass spectrometry based on stable isotope dilution theory; and the use of specific
15 chemical reactions for the selective isolation of specific peptides. For example, the previously described ICAT™ reagent contained a sulfhydryl reactive group, and therefore an ICAT™-type reagent can be used to label cysteine-containing peptide fragments released from the
20 solid support. Other reactive groups, as described above, can also be used.

The analysis of the non-glycosylated peptides, in conjunction with the methods of analyzing glycosylated
25 peptides, provides additional information on the state of polypeptide expression in the sample. By analyzing both the glycopeptide fragments as well as the non-glycosylated peptides, changes in glycoprotein abundance as well as changes in the state of glycosylation at a
30 particular glycosylation site can be readily determined.

If desired, the sample can be fractionated by a number of known fractionation techniques. Fractionation techniques can be applied at any of a number of suitable points in the methods of the invention. For example, a sample can be fractionated prior to oxidation and/or binding of glycopolypeptides to a solid support. Thus, if desired, a substantially purified fraction of glycopolypeptide(s) can be used for immobilization of sample glycopolypeptides. Furthermore, fractionation/purification steps can be applied to non-glycosylated peptides or glycopeptides after release from the solid support. One skilled in the art can readily determine appropriate steps for fractionating sample molecules based on the needs of the particular application of methods of the invention.

Methods for fractionating sample molecules are well known to those skilled in the art. Fractionation methods include but are not limited to subcellular fractionation or chromatographic techniques such as ion exchange, including strong and weak anion and cation exchange resins, hydrophobic and reverse phase, size exclusion, affinity, hydrophobic charge-induction chromatography, dye-binding, and the like (Ausubel et al., Current Protocols in Molecular Biology (Supplement 56), John Wiley & Sons, New York (2001); Scopes, Protein Purification: Principles and Practice, third edition, Springer-Verlag, New York (1993)). Other fractionation methods include, for example, centrifugation, electrophoresis, the use of salts, and the like (see Scopes, supra, 1993). In the case of analyzing membrane glycoproteins, well known solubilization conditions can

be applied to extract membrane bound proteins, for example, the use of denaturing and/or non-denaturing detergents (Scopes, supra, 1993).

5 Affinity chromatography can also be used including, for example, dye-binding resins such as Cibacron blue, substrate analogs, including analogs of cofactors such as ATP, NAD, and the like, ligands, specific antibodies
10 useful for immuno-affinity isolation, either polyclonal or monoclonal, and the like. A subset of glycopolypeptides can be isolated using lectin affinity chromatography, if desired. An exemplary affinity resin includes affinity resins that bind to specific moieties that can be incorporated into a polypeptide such as an
15 avidin resin that binds to a biotin tag on a sample molecule labeled with an ICATTM-type reagent. The resolution and capacity of particular chromatographic media are known in the art and can be determined by those skilled in the art. The usefulness of a particular
20 chromatographic separation for a particular application can similarly be assessed by those skilled in the art.

 Those of skill in the art will be able to determine
25 the appropriate chromatography conditions for a particular sample size or composition and will know how to obtain reproducible results for chromatographic separations under defined buffer, column dimension, and flow rate conditions. The fractionation methods can
30 optionally include the use of an internal standard for assessing the reproducibility of a particular chromatographic application or other fractionation

method. Appropriate internal standards will vary depending on the chromatographic medium or the fractionation method used. Those skilled in the art will be able to determine an internal standard applicable to a method of fractionation such as chromatography.
5 Furthermore, electrophoresis, including gel electrophoresis or capillary electrophoresis, can also be used to fractionate sample molecules.

10 The invention also provides a method for identifying and quantifying glycopeptides in a sample. The method includes the steps of immobilizing glycopolypeptides to a solid support; cleaving the immobilized glycopolypeptides, thereby releasing non-glycosylated
15 peptides and retaining immobilized glycopeptides; labeling the immobilized glycopeptides with an isotope tag; releasing the glycopeptides from the solid support; and analyzing the released glycopeptides.

20 The methods of the invention can be used in a wide range of applications in basic and clinical biology. The methods of the invention can be used for the detection of changes in the profile of proteins expressed in the plasma membrane, changes in the composition of proteins
25 secreted by cells and tissues, changes in the protein composition of body fluids including blood and seminal plasma, cerebrospinal fluid, pancreatic juice, urine, breast milk, lung lavage, and the like. Since many of the proteins in these samples are glycosylated, the
30 methods of the invention allow the convenient analysis of glycoproteins in these samples. Detected changes observed in a disease state can be used as diagnostic or

prognostic markers for a wide range of diseases, including congenital disorders of glycosylation (Example XI) or any disorder involving aberrant glycosylation; cancer, such as skin, prostate, breast, colon, lung, and others (Examples VIII and IX); metabolic diseases or processes such as diabetes (Example XII) or changes in physiological state (Example X); inflammatory diseases such as rheumatoid arthritis; mental disorders or neurological processes; infectious disease; immune response to pathogens; and the like. Furthermore, the methods of the invention can be used for the identification of potential targets for a variety of therapies including antibody-dependent cell cytotoxicity directed against cell surface proteins and for detection of proteins accessible to drugs.

Thus, the methods of the invention can be used to identify diagnostic markers for a disease by comparing a sample from a patient having a disease to a sample from a healthy individual or group of individuals. By comparing disease and healthy samples, a diagnostic pattern can be determined with increases or decreases in expression of particular glycopolypeptides correlated with the disease, which can be used for subsequent analysis of samples for diagnostic purposes (see Examples VIII, IX, XI and XII). The methods are based on analysis of glycopolypeptides, and such an analysis is sufficient for diagnostic purposes.

Thus, the invention provides a method for identifying diagnostic glycopolypeptide markers by using a method of the invention and comparing samples from

diseased individual(s) to healthy individual(s) and identifying glycopolypeptides having differential expression between the two samples, whereby differences in expression indicates a correlation with the disease
5 and thus can function as a diagnostic marker. The invention also provides the diagnostic markers identified using methods of the invention.

Furthermore, glycopolypeptides exhibiting
10 differential expression are potential therapeutic targets. Because they are differentially expressed, modulating the activity of these glycopolypeptides can potentially be used to ameliorate a sign or symptom associated with the disease. Thus, the invention
15 provides a method for identifying therapeutic glycopolypeptide targets of a disease. Once a glycopolypeptide is found to be differentially expressed, the potential target can be screened for potential therapeutic agents that modulate the activity of the
20 therapeutic glycopolypeptide target. Methods of generating libraries and screening the libraries for potential therapeutic activity are well known to those skilled in the art. Methods for producing pluralities of compounds, including chemical or biological molecules
25 such as simple or complex organic molecules, metal-containing compounds, carbohydrates, peptides, proteins, peptidomimetics, glycoproteins, lipoproteins, nucleic acids, antibodies, and the like, are well known in the art (see, for example, in Huse, U.S. Patent No.
30 5,264,563; Francis et al., Curr. Opin. Chem. Biol. 2:422-428 (1998); Tietze et al., Curr. Biol., 2:363-371 (1998); Sofia, Mol. Divers. 3:75-94 (1998); Eichler et al., Med.

Res. Rev. 15:481-496 (1995); Gordon et al., J. Med. Chem. 37: 1233-1251 (1994); Gordon et al., J. Med. Chem. 37: 1385-1401 (1994); Gordon et al., Acc. Chem. Res. 29:144-154 (1996); Wilson and Czarnik, eds., Combinatorial Chemistry: Synthesis and Application, John Wiley & Sons, New York (1997)). The invention additionally provides glycopolypeptide therapeutic targets identified by methods of the invention.

10 The methods can be used for a variety of clinical and diagnostic applications. Known therapeutic methods effected through glycopolypeptides can be characterized by methods of the invention. For example, therapies such as Enbrel™ and Herceptin function through glycoproteins.
15 The methods of the invention allow characterization of individual patients with respect to glycoprotein expression, which can be used to determine likely efficacy of therapy involving glycoproteins.

20 Thus, the methods of the invention can be used in a variety of applications including, but not limited to, the following applications. The methods of the invention can be used, for example, for blood serum profiling for the detection of prognostic and diagnostic protein
25 markers (see Examples VIII, IX, XI and XII). The methods of the invention can also be used for quantitative profiling of cell surface proteins for the detection of diagnostic/prognostic protein markers and the detection of potential targets of therapy (Example IV). For
30 example, the methods of the invention can be used for antibody-dependent cellular cytotoxicity (ADCC) or other types of therapy. The methods of the invention are

applicable in clinical and diagnostic medicine, veterinary medicine, agriculture, and the like. For example, the methods of the invention can be used to identify and/or validate drug targets and to evaluate drug efficacy, drug dosing, and/or drug toxicity. In such a case, the blood proteome, that is serum, can be analyzed using the methods disclosed herein to look for changes in serum glycopolypeptide profiles associated with drug administration and correlated with the effects of drug efficacy, dosing and/or toxicity, and/or validation of drug targets. Such a correlation can be readily determined by collecting serum samples from one or more individuals administered various drug doses, experiencing drug toxicity, experiencing a desired efficacy, and the like. In addition, a serum profile can be generated in combination with the analysis of drug targets as a way to rapidly and efficiently validate a particular target with the administration of a drug or various drug doses, toxicity, and the like. Thus, serum (blood samples) provide a surrogate marker for the status of an individual and his or her ability to respond to a pharmacological intervention.

The methods of the invention can additionally be used for quantitative protein profiling in various body fluids in addition to blood plasma, including CSF, pancreatic juice, lung lavage fluid, seminal plasma, urine, breast milk, and the like. The methods of the invention can also be used for quantitative protein profiling of proteins secreted by cells or tissues for the detection of new protein and peptide hormones and other factors. Thus, the invention provides a method to

generate quantitative profiles of glycoproteins. The invention also provides a method for quantifying a glycopolypeptide in a sample, as disclosed herein. The invention further provides a method for the detection of
5 prognostic or diagnostic patterns in blood serum and other body fluids. The invention additionally provides a method for the detection of secreted protein hormones and regulatory factors. Thus, the invention provides a method for profiling glycopolypeptides from body fluids,
10 secreted proteins and cell surface proteins.

The methods of the invention are also applicable to the detection of changes in the state of glycosylation of proteins based on the concurrent application of protein
15 abundance measurement and measurement of protein glycosylation on the same sample. Thus, the invention provides a method to detect quantitative changes in the glycosylation pattern of specific proteins.

20 The invention also provides a method for the systematic detection of glycosylation sites on proteins. Because the methods of the invention allow the identification of peptide fragments that are glycosylated, this also serves as the identification of
25 the site of glycosylation (Example VII).

Although the methods disclosed herein have generally been described for the analysis of glycopolypeptides, similar methods are also applicable to the analysis of
30 other carbohydrate-containing molecules. Because the methods are based on the specific binding of carbohydrate moieties, the methods of modification and/or isolation

can similarly be applied to other carbohydrate-containing molecules. For example, method steps analogous to those disclosed herein can be applied to the identification and quantification of glycosylated molecules such as
5 glycolipids, glycosphingolipids, and the like.

The invention also provides reagents and kits for isolating and quantifying glycopolypeptides. The kit can contain, for example, hydrazide resin or other suitably
10 reactive resin for solid phase capture of glycopolypeptides, a reagent for modification of carbohydrate moieties, for example, an oxidizing reagent such as periodate, and a set of two or more
15 differentially labeled isotope tags for coupling to two different samples, which are particularly useful for quantitative analysis using mass spectrometry. In one embodiment, the invention provides a kit comprising a hydrazide resin, periodate, and a pair of differentially
20 labeled isotope tags. The contents of the kit of the invention, for example, any resins or labeling reagents, are contained in suitable packaging material, and, if desired, a sterile, contaminant-free environment. In addition, the packaging material contains instructions
25 indicating how the materials within the kit can be employed to label sample molecules. The instructions for use typically include a tangible expression describing the reagent concentration or at least one assay method parameter, such as the relative amounts of reagent and
30 sample to be admixed, maintenance time periods for reagent/sample admixtures, temperature, buffer conditions, and the like.

The methods of the invention can be facilitated by the use of combinations of hardware and software suitable for analysis of methods of the invention. For example, a robotics workstation was developed to facilitate automated glycopeptide analysis (Example XVI). A computer program can be used to find patterns of proteins and/or peptides that are specifically present or present at specific abundances in a sample from a person with a specific disease (see Examples). For example, a number of serum samples can be analyzed and compared to serum samples from healthy individuals. An algorithm is used to find those peptides and/or proteins that are either individually or collectively diagnostic for the disease or the stage of the disease being examined.

In another embodiment, the invention provides a method for identifying and quantifying glycopeptides in a sample. The method can include the steps of immobilizing glycopolypeptides to a solid support; cleaving the immobilized glycopolypeptides, thereby releasing non-glycosylated peptides and retaining immobilized glycopeptides; releasing the glycopeptides from the solid support; and analyzing the released glycopeptides. The method can further include the step of identifying one or more glycopeptides, for example, using mass spectrometry.

In still another embodiment, the invention provides a method of identifying a diagnostic marker for a disease. The method can include the steps of immobilizing glycopolypeptides from a test sample to a first solid support; immobilizing glycopolypeptides from a control sample to a second solid support; cleaving the

immobilized glycopolypeptides, thereby releasing non-glycosylated peptides and retaining immobilized glycopeptides; labeling the immobilized glycopeptides on the first and second supports with differential isotope tags on the respective supports; releasing the glycopeptides from the solid supports; analyzing the released glycopeptides; and identifying one or more glycosylated polypeptides having differential glycosylation between the test sample and the control sample. Alternatively, the test and control samples can be run in parallel and analyzed separately. In such a case, the glycopeptides are identified and compared without using differential isotope tagging.

The test sample can be, for example, a specimen from an individual having a disease. The control sample can be, for example, a corresponding specimen obtained from a healthy individual. The sample can be, for example, serum or a tissue biopsy, as described herein.

Differential glycosylation can be a qualitative difference, for example, the presence or absence of a glycopolypeptide in the test sample compared to the control sample. Differential glycosylation can also be a quantitative difference. The determination of quantitative differences can be facilitated by the labeling with differential isotope tags such that the samples can be mixed and compared side-by-side, as disclosed herein and described in Gygi et al., *supra*, 1999. One or more glycopolypeptides exhibiting differential glycosylation are potential diagnostic markers for the respective disease. Such a method provides a glycopolypeptide disease profile, which can be

used subsequently for diagnostic purposes. Accordingly, rather than using one or a few diagnostic markers, the methods of the invention allow the identification of a profile of diagnostic markers, which can provide more
5 detailed information on the type of disease, the stage of disease, and/or the prognosis of a disease by determining profiles correlated with the type, stage and/or prognosis of a disease.

10 In yet another embodiment, the invention provides a method of diagnosing a disease. The method can include the steps of immobilizing glycopolypeptides from a test sample to a solid support; cleaving the immobilized glycopolypeptides, thereby releasing non-glycosylated
15 peptides and retaining immobilized glycopeptides; releasing the glycopeptides from the solid support; analyzing the released glycopeptides; and identifying one or more diagnostic markers associated with a disease, for example, as determined by methods of the invention, as
20 described above.

A test sample from an individual to be tested for a disease or suspected of having a disease can be processed as described for glycopeptide analysis by the methods
25 disclosed herein. The resulting glycopeptide profile from the test sample can be compared to a control sample to determine if changes in glycosylation of diagnostic markers has occurred, as discussed above. Alternatively, the glycopeptide profile can be compared to a known set
30 of diagnostic markers or a database containing information on diagnostic markers.

In another embodiment, the method of diagnosing a disease can include the step of generating a report on the results of the diagnostic test. For example, the report can indicate whether an individual is likely to have a disease or is likely to be disease free based on the presence of a sufficient number of diagnostic markers associated with a disease. The invention further provides a report of the outcome of a method of diagnosing a disease. Similar reports and preparation of such reports are provided for other methods of the invention.

It is understood that the methods of the invention can be performed in any order suitable for glycopolypeptide analysis. One skilled in the art can readily determine an appropriate order of carrying out steps of methods of the invention suitable for glycopeptide analysis.

It is understood that modifications which do not substantially affect the activity of the various embodiments of this invention are also provided within the definition of the invention provided herein. Accordingly, the following examples are intended to illustrate but not limit the present invention.

EXAMPLE I

Quantitative Analysis of Glycopeptides

This example describes purification of glycopeptides and differential labeling with isotope tags.

An embodiment of a method of the invention is schematically illustrated in Figure 1. The method can include the following steps: (1) Glycoprotein oxidation: Oxidation, for example, with periodate, converts the cis-
5 diol groups of carbohydrates to aldehydes (Figure 2); (2) Coupling: The aldehydes react with hydrazide groups immobilized on a solid support to form covalent hydrazone bonds (Figure 2). Non-glycosylated proteins are removed; (3) Proteolysis: The immobilized glycoproteins are
10 proteolyzed on the solid support. The non-glycosylated peptides are removed by washing and can be optionally collected for further analysis, whereas the glycosylated peptides remain on the solid support; 4) Isotope labeling: The α amino groups of the immobilized
15 glycopeptides are labeled with isotopically light (d0, contains no deuteriums) or heavy (d4, contains four deuteriums) forms of succinic anhydride after the ϵ -amino groups of lysine are converted to homoarginine (Figure 3); (5) Release: Formerly N-linked glycopeptides are
20 released from the solid-phase by PNGase F treatment; (6) Analysis: The isolated peptides are identified and quantified using microcapillary high performance liquid chromatography electrospray ionization tandem mass spectrometry (μ LC-ESI-MS/MS) or μ LC separation followed
25 by matrix-assisted laser desorption/ionization (MALDI) MS/MS. The data are analyzed by a suite of software tools.

Proteins from a sample, for example, a complex
30 biological sample, were changed to buffer containing 100 mM NaAc, 150 mM NaCl, pH 5.5 (coupling buffer). Sodium periodate solution at 15 mM was added to the samples.

The cap was secured and the tube was covered with foil. The sample was rotated end-over-end for 1 hour at room temperature. The sodium periodate was removed from the samples using a desalting column (Econo-Pac 10DG column).

5 Hydrazide resin (Bio-Rad; Hercules CA) equilibrated in coupling buffer was added to the sample (1 ml gel/5 mg protein). The sample and resin were capped securely and rotated end-over-end for 10-24 hours at room temperature.

10 After the coupling reaction was complete, the resin was spun down at 1000xg for 10 min, and non-glycoproteins were washed away extensively by washing the resin 3 times with an equal volume of 8M urea/0.4M NH_4HCO_3 . The proteins on the resin were denatured in 8M urea/0.4M

15 NH_4HCO_3 at 55°C for 30 min, followed by 3 washes with the urea solution. After the last wash and removal of the urea buffer, the resin was diluted 4 times with water. Trypsin was added at a concentration of 1 µg of trypsin/100 µg of protein and the bound proteins digested

20 at 37°C overnight. If desired, the peptides can be reduced by adding 8 mM TCEP (Pierce, Rockford IL) at room temperature for 30 min, and alkylated by adding 10 mM iodoacetamide at room temperature for 30 min. The trypsin released peptides were removed and collected for

25 labeling with ICAT™ reagent or other tagging reagent, if desired. The resin was washed with an equal volume of 1.5 M NaCl 3 times, 80% acetonitrile (MeCN)/0.1% trifluoroacetic acid (TFA) 3 times, 100% methanol 3

30 times, and 0.1 M NH_4HCO_3 6 times. N-linked glycopeptides were released from the resin by digestion with peptide-N-glycosidase F (PNGase F) overnight. The resin was spun and the supernatant was saved. O-linked glycopeptides

can be released from the resin by using combination of neuraminidase/O-glycosidase. The resin was washed twice with 80% MeCN/0.1% TFA and combined with the supernatant. The peptides were dried and resuspended in 0.4% acetic acid for LC-MS/MS analysis.

Alternatively, the glycopeptides can be released from the resin chemically. The N-linked glycopeptide can be released by hydrazinolysis. Glycopeptides are dried in a desiccator over P_2O_5 and NaOH. The reaction is carried out in an air-tight screw-cap tube using anhydrous hydrazine. The reaction is carried out at 100°C for about 10 hours using a dry heat block. The release of O-linked glycopeptide is carried out in 50 mM NaOH containing 1 M $NaBH_4$ at 55°C for about 18h.

For isotopic labeling of glycopeptides with succinic anhydride (Figure 3), the glycopeptides on the beads were washed twice with 15% NH_4OH in water (pH > 11). Methylisourea at 1 M in 15% NH_4OH ($NH_4OH/H_2O = 15/85$ v/v) was added in 100 fold molar excess over amine groups and incubated at 55°C for 10 minutes. Beads were then washed twice with water, twice with dimethylformamide (DMF)/pyridine/ $H_2O=50/10/40$ (v/v/v) and resuspended in DMF/pyridine/ $H_2O=50/10/40$ (v/v/v). Succinic anhydride solution was added to a final concentration of 2 mg/ml. The sample was incubated at room temperature for 1 hour, followed by washing three times with DMF, three times with water, and six times with 0.1M NH_4HCO_3 . The peptides were released from the beads using PNGase F as describe above.

Alternatively, the glycopeptides can be labeled with other reagents at amine groups of glycopeptides while the peptides are still conjugated to the hydrazide beads. A list of chemicals that have been tested and proved to be able to label the amino groups is listed in Figure 3. The structures of labeled peptide are listed at the right column. Once the glycopeptides were labeled isotopically, PNGase F was added to release the peptides from the solid support and analyzed by mass spectrometry.

10

For isotopic labeling of glycopeptides with Phe (see Figure 9), 0.22 M of Boc-d0-Phe-OH (Nova Biochem) or Boc-d5-Phe-OH (CDN Isotopes) were dissolved in anhydrous, N,N-dimethylformamide. 1,3-Diisopropylcarbodiimide was added to a final concentration of 0.2 M. The reaction was carried out at room temperature for 2 hours. The glycopeptides on the beads were washed with 0.5 M NaHCO₃ three times and resuspended to a 50% slurry. The same volume of Boc-Phe-anhydride was added to the glycopeptides on the beads, and the beads were incubated at room temperature for 30 min. The beads were washed with 80% MeCN/0.1% TFA three times and dried. The Boc protection group was removed by incubating with TFA for 30 min at room temperature. The beads were washed with glycosidase buffer, followed by release of the labeled glycopeptides with glycosidases, as described above.

25

This example describes purification of glycopeptides and differential labeling with an isotope tag.

30

EXAMPLE II**Quantitative Glycopeptide Profiling in Human Blood Serum**

This example describes profiling of glycoproteins in
5 human blood serum.

To assess the potential of the glycopeptide capture
method for serum protein profiling, the specificity and
efficiency of conjugation was first determined. Human
10 serum proteins were coupled to the hydrazide beads.
Identical aliquots (1 μ l) were removed from the sample
before ("- beads") or after capture of glycoproteins to
hydrazide resin ("+" beads"). The samples were separated
by 9% sodium dodecyl sulfate-polyacrylamide gel
15 electrophoresis (SDS-PAGE) and stained with silver (total
protein stain) or with a glycoprotein-staining reagent
(Figure 4).

Isolation of glycopolypeptides was performed
20 essentially as described in Example I. For analysis of
serum samples, 2.5 ml of human serum (200 mg total
protein) were changed to buffer containing 100 mM NaAc,
150 mM NaCl, pH 5.5 using a desalting column (Bio-Rad).
Sodium periodate solution at 15 mM was added to the
25 samples. The cap was secured and the tube was covered
with foil. The sample was rotated end-over-end for 1
hour at room temperature. The sodium periodate was
removed from the samples using a desalting column. A 50
 μ l aliquot of the sample was taken before coupling the
30 sample. To the sample was added 8 ml of coupling buffer
equilibrated hydrazide resin (Bio-Rad). The sample and
resin were capped securely and rotated end-over-end for

10-24 hours at room temperature. After the coupling
reaction was complete, the resin was spun down at 1000xg
for 10 min, and non-glycoproteins in the supernatant were
removed. A 50 µl aliquot of the post conjugation sample
5 was taken.

A portion of each of the aliquots taken before and
after coupling (1 µl) was analyzed on a 9% sodium dodecyl
sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel
10 and stained. For total proteins and glycoproteins,
silver staining or GelCode Glycoprotein staining reagent,
respectively, were used to determine the specificity and
efficiency of glycoprotein isolation.

15 As shown in Figure 4, the supernatant prior to
addition to the beads (-) contains a number of proteins,
many of which stain as glycoproteins (right panel, "-"
lane). After addition and incubation with the beads,
most of the glycopolypeptides are removed (left and right
20 panels, "+" lanes). These results show that the
hydrazide beads efficiently bind the glycopolypeptides
from the serum sample. Note that the major serum
protein, albumin, remained in the supernatant (left
panel, "+" lane) and was not stained with the
25 glycoprotein stain (right panel, "-" lane). Thus,
albumin does not appear to be glycosylated. Since
albumin is the major serum protein (>50%), the use of
carbohydrate-specific binding provides a method to
efficiently analyze low abundance, glycosylated
30 polypeptides present in serum.

The following is apparent from the experiment shown in Figure 4. First, as expected, the serum sample contains a considerable amount of glycosylated proteins (glycoprotein stain, "- beads" lane). Second, the majority of the protein bands were essentially depleted by the coupling reaction (silver stained bands "+/- beads" lanes). Third, as far as could be determined from the different staining intensities of the two staining methods used, glycosylated proteins were quantitatively depleted and bands containing glycosylated proteins were preferentially removed by the coupling reaction. Fourth, the major band representing serum albumin was not depleted by the coupling reaction and did not stain with the glycoprotein-staining reagent. Collectively, these results show that the hydrazide beads bind the oxidized glycoproteins from the serum sample efficiently and specifically. They also show that the major serum protein, albumin, predominantly remained in the supernatant (left panel, "+ beads" lane) and was not stained with the glycoprotein stain (right panel, "+/- beads" lane). The use of carbohydrate-specific isolation of serum glycoproteins therefore provides a more economical, simpler and more reproducible method for serum albumin removal than the affinity depletion methods commonly used. Since the present method is also compatible with the immobilization of denatured proteins, it reduces the possibility that the selective removal of albumin also removes albumin-associated proteins.

Non-specific proteins bound to the resin were washed away extensively by washing the resin 3 times with an equal volume of 8M urea/0.4M NH_4HCO_3 . The proteins on the

resin were denatured in 8M urea/0.4M NH_4HCO_3 at 55°C for 30 min, followed by 3 washes with the urea solution. After the last wash and removal of the urea buffer, the resin was diluted 4 times with water. Trypsin was added
5 at a concentration of 1 µg of trypsin/100 µg of protein and digested at 37°C overnight. The trypsin released peptides were removed by washing the resin with an equal volume of 1.5 M NaCl for 3 times, 80% MeCN/0.1% TFA for 3
10 times, 100% methanol for 3 times, and 0.1 M NH_4HCO_3 for 6 times. N-linked glycopeptides were released from the resin by digestion with PNGase F at 37°C overnight. The resin was spun and the supernatant saved. The resin was washed twice with 80% MeCN/0.1% TFA and combined with the supernatant. The resin was saved for O-linked
15 glycopeptide release later.

The peptides were dried in 17 tubes, and one tube was resuspended 50 µl of 0.4% acetic acid. A 3 µl aliquot of the sample (from 9 µl of serum) was loaded on
20 a capillary column for µLC-MS/MS analysis. CID spectra were searched against a human database using SEQUEST (Eng et al., J. Am. Soc. Mass. Spectrom. 5:976-989 (1994)) to identify the glycopeptides and glycoproteins (Figure 5, middle panel).

25 To determine whether the reduced peptide sample complexity achieved by glycopeptide capture and release allowed the identification of more serum proteins compared to conventionally prepared control samples if
30 comparable µLC-MS/MS protocols were applied, the number of glycopeptides and glycoproteins identified as described above was compared to the number of serum

proteins identified from other methods using the same
μLC-MS/MS protocols. Control samples were generated by
selectively isolating cysteine-containing peptides using
the ICAT reagent method (Gygi et al., Nat. Biotechnol.
5 17:994-999 (1999)). These were analyzed either using the
same μLC-ESI-MS/MS method as for the analysis of the
peptides isolated by the glycopeptide capture method
(Figure 5, right panel) or via extensive, three
dimensional (cation exchange/biotin affinity/reverse
10 phase liquid chromatography (RP-LC)) in which the peptide
mixture was fractionated into 17 cation exchange
fractions that were sequentially analyzed by μLC-ESI-
MS/MS (Figure 5, left panel; Han et al., Nat. Biotechnol.
19:946-951 (2001)).

15

Using a single μLC-ESI-MS/MS run requiring
approximately two hours of mass spectrometer time, 145
unique peptides mapping to 57 unique serum proteins were
identified with the glycopeptide capture method (2.5
20 peptides/protein). When comparable MS methods were
applied for the analysis of cysteine tagged peptides, 72
unique peptides mapping to 23 unique proteins were
identified, of which 15 were also identified via the
glycopeptide capture method (Figure 5, right panel).
25 Using the extensive peptide separation protocol for the
analysis of cysteine tagged peptides that required
approximately 34 hours of mass spectrometer time, 356
unique peptides mapping to 97 serum proteins were
identified. Of the 57 proteins isolated by the
30 glycopeptide capture method and identified by single
dimensional LC-MS/MS, 23 proteins were not seen by the
extensive μLC-ESI-MS/MS based protocol of cysteine tagged

peptides (Figure 5, left panel). These data demonstrate the increased efficiency of serum analysis provided by the glycopeptide capture method.

5 As the current "gold standard method " for serum protein analysis is based on high resolution two-dimensional electrophoresis (2DE) and MS, the number of proteins identified by a single LC-MS/MS analysis of peptides isolated by the glycopeptide capture method was
10 also related with the number of proteins that are annotated in the most up-to-date 2DE plasma protein map on SWISS-2DPAGE (us.expasy.org/cgi-bin/get-ch2d-table.pl). The 2DE map identifies 58 unique proteins from 626 detected spots. Of these, 270 spots represent 8
15 different forms of immunoglobulin chains. Glycopeptide capture and single dimensional LC-MS/MS analysis identified 57 proteins, of which 7 are different immunoglobulin chains and 16 proteins are not included in SWISS-2DPAGE. Four major conclusions can be drawn that
20 are relevant for assessing the potential of each method for serum protein profiling, even though, for reasons of sample and experimental variability, the data obtained from the three methods are not directly comparable. First, both the 2DE/MS based method and the cysteine
25 tagging method are substantially limited by the presence of a number of high abundance proteins (that is, the "top down" problem in its extreme), which include the five major plasma proteins representing more than 80% of the total plasma protein mass (albumin, α -1-antitrypsin, β -2-macroglobulin, transferrin, and γ -globulins). When the
30 cysteine tagged peptides were analyzed, the mass spectrometer spent over one third of the acquisition time

on CID spectra of albumin (39% of peptides identified by the cysteine tagging method were from albumin). In contrast, the glycopeptide capture method selected against albumin with only 1% of peptides identified from
5 albumin.

Second, proteins that were not identified by either of the traditional methods were readily identified following glycopeptide capture (Figure 5). This attests
10 to the potential of the glycopeptide capture method to achieve deeper serum protein coverage within a dramatically reduced data acquisition time. The limited diversity of the proteins analyzed by the traditional methods is further illustrated by the observation that of
15 the 63 proteins that were only identified using cysteine reactive tags, 18 were different immunoglobulins. The glycopeptide capture method identified only peptides from the constant region of immunoglobulin and thus limited the number of immunoglobulin-derived peptides (7
20 immunoglobulin chains identified by the glycopeptide capture method, which were also identified by the cysteine tagging method).

Third, the glycopeptide capture method reduced the
25 sample complexity; an average of 2.5 peptides per protein were detected. Fourth, the presence of the N-glycosylation sequence motif in the identified peptides provided further validation of specific isolation and increased the confidence in database searching results.
30 Therefore, the reduction in sample complexity achieved by the glycopeptide capture method provides a substantial

advance for the analysis of blood serum and other body fluids of similar protein composition.

Peptides isolated from 2.5 ml of serum using the glycopeptide capture method were further separated by cation exchange fractionation (Han et al. Nat. Biotechnol. 19:946-951 (2001)). Four of seventeen tubes containing peptides released from hydrazide resin as described above (equivalent of 600 µl serum) were separated by cation exchange chromatography to 38 fractions and resuspended in 20 µl of 0.4% acetic acid solution. A 5 µl aliquot of each fraction was loaded on a capillary column for µLC-MS/MS analysis. CID spectra were searched against a human database using SEQUEST (Eng et al., J. Am. Soc. Mass. Spectrom. 5:976-989 (1994)) to identify the glycopeptides and glycoproteins.

A large number of glycoproteins and their glycopeptides from a human serum sample were released with PNGase F. A total 1011 proteins identified had a protein probability score of at least 0.5. Based on the distribution of sensitivity and error rate at different protein and peptide probability score (Keller et al., Anal. Chem. 74:5383-5392 (2002)), there were 832 correctly identified proteins with a protein probability score of at least 0.5 (Table 2).

Table 2. Estimated sensitivity, error rates, number of correct and incorrect proteins with different protein probability score.

Minimum probability	sensitivity	error rate	Number of correct proteins	Number of incorrect proteins
1.00	0.264	0.000	288	0
0.99	0.291	0.001	318	0
0.98	0.315	0.002	344	1
0.97	0.331	0.004	362	1
0.96	0.344	0.005	375	2
0.95	0.356	0.007	388	3
0.90	0.417	0.019	455	9
0.80	0.522	0.049	570	29
0.70	0.607	0.084	662	61
0.60	0.684	0.126	746	108
0.50	0.762	0.177	832	179
0.40	0.835	0.235	911	279
0.30	0.907	0.304	990	432
0.20	1.000	0.406	1091	746

5

These results show that the glycopeptide capture method also removes albumin from the analysis of serum proteins, thereby allowing the analysis of less abundant serum proteins. The methods allowed the identification of a number of serum proteins that were not easily identified with other methods.

10

Blood serum is a complex body fluid that contains enormous information about body health. When blood

circulates through the body, proteins secreted from cells, shredded from cell surface proteins, and released from dead cells from all tissues are deposited to the blood serum. Blood serum is also the most easily
5 accessible specimen for diagnostic purpose. DNA array technology is not capable of analyzing serum samples since there is not a particular tissue sample from which to extract RNA. The analysis of plasma or serum proteins has also been a focus of proteomics. The two-dimensional
10 electrophoretic technique has been used in the analysis of human plasma proteins since 1977 (Anderson and Anderson, Proc. Natl. Acad. Sci. USA 74:5421-5425 (1977)). To date, 289 plasma proteins have been identified using the 2DE method (Anderson and Anderson,
15 Mol. Cell. Proteomics 1:845-687 (2002)). Recently, direct analysis of serum proteins with mass spectrometry was used to analyze proteins in human serum. In this analysis, abundant immunoglobulin proteins were first affinity depleted from serum sample. The resulting
20 peptides were separated by strong cation exchange chromatography into distinct fractions prior to analysis. 490 serum proteins were identified by on-line reversed-phase microcapillary liquid chromatography coupled with ion trap mass spectrometry (Adkins et al., Mol. Cell.
25 Proteomics 1947-1955 (2002)).

While the use of more extensive separation protocols for the formerly N-glycosylated peptides will increase the depth of serum protein coverage, tryptic peptides
30 that are too short or too long to fall within the detection range of the mass spectrometer used will not be identified. This can be overcome, at least in part, by

the use of proteases with cleavage specificities
different from that of trypsin.

5 The increased number of serum proteins identified
using the glycopeptide capture method compared to other
proteomics methods so far shows that the glycopeptide
method is an efficient method to analyze serum proteins
and has the capacity to identify low abundance proteins
as disease biomarkers in serum.

10

EXAMPLE III

Quantitative Profiling of Glycoproteins Secreted by Macrophages

15

This example describes the preparation of secreted
protein sample from stimulated RAW 264.7 mouse
monocyte/macrophage cell line.

20

Briefly, 10^9 RAW cells were used. On day 1, cells
were plated at a density of 2.5×10^5 cells/cm² with 10 nM
phorbol 12-myristate- 13-acetate (PMA). On day 2, the
media was removed, and new media was added without PMA.
On day 3, the cells were washed three times with serum-
free media.

25

Lipopolysaccharide (LPS) was added as stimulant to
the experimental cells with serum-free, PMA-free media.
The cells were incubated at 37°C for 4 hours. The
supernatant was removed, and the cells were centrifuged
at 3,000 xg for 5 minutes to remove cells and large

30

debris. The supernatant was centrifuged at 100,000 xg for 1 hour to remove debris.

5 The supernatant was concentrated with an 80 mL Centricon concentrator, with 300 mL concentrated to <1 mL for each condition. The final concentration of proteins was at least 2 mg/mL.

10 One mg of proteins secreted from unstimulated and stimulated macrophages was changed to buffer containing 100 mM NaAc, 150 mM NaCl, pH 5.5, using a desalting column (Bio-Rad). Sodium periodate solution at 15 mM was added to the samples. The cap was secured and the tube covered with foil. The sample was rotated end-over-end
15 for 1 hour at room temperature. Sodium periodate was removed from the samples using a desalting column. A 50 µl aliquot of the sample was taken before coupling the sample. To the sample was added 0.2 ml of coupling buffer equilibrated hydrazide resin (Bio-Rad). The resin
20 and sample were capped securely and rotated end-over-end for 10-24 hours at room temperature. After the coupling reaction was complete, the resin was spun down at 1000xg for 10 min, and non-glycoproteins in the supernatant were removed. An aliquot of 50 µl of the post conjugation
25 sample was taken. An aliquot of the samples before and after binding to the resin were analyzed on a 9% SDS-PAGE gel and stained for total proteins using silver staining reagent to determine the specificity and efficiency of glycoprotein isolation.

30

Non-specific proteins bound to the resin were washed away extensively by washing the resin 3 times with an

equal volume of 8M urea/0.4M NH_4HCO_3 . The proteins on the resin were denatured in 8M urea/0.4M NH_4HCO_3 at room temperature for 30 min, followed by 3 washes with the urea solution. After the last wash and removal of the urea buffer, the resin was diluted 4 times with water. Trypsin was added at a concentration of 1 μg of trypsin/100 μg of protein and digested at 37°C overnight. The trypsin released peptides were removed by washing the resin with an equal volume of 1.5 M NaCl for 3 times, 80% MeCN/0.1% TFA for 3 times, 100% methanol for 3 times, 0.1 M NH_4HCO_3 for 6 times. N-linked glycopeptides were released from the resin by digest with N-glycosidase at 37°C overnight. The resin was spin and the supernatant was saved. The resin was washed twice with 80% MeCN/0.1% TFA and combined with the supernatant. The resin was saved for O-linked glycopeptide release later.

The peptides were dried and resuspended in 50 μl of 0.4% acetic acid. 3 μl of sample was loaded on a capillary column for $\mu\text{LC-MS/MS}$ analysis. CID spectra were searched against a mouse database using SEQUEST to identify the glycopeptides and glycoproteins.

Figure 6 shows glycoproteins identified from secreted proteins of untreated or LPS-treated RAW macrophage cells. A total of 32 proteins were identified. Nineteen secreted glycosylated proteins were identified in both untreated and treated cells. Eight proteins were identified in untreated cells, and five proteins were identified in treated cells. One of the known macrophage secreted proteins, tumor necrosis factor (TNF), was positively identified in media from RAW cells after LPS

treatment. These results show that glycopolypeptides can be selectively isolated from a secreted proteins from cells in an efficient and specific manner.

5 For isotopic labeling of glycopeptides with succinic anhydride (Figure 3), the dried peptides released from hydrazide resin were resuspended in DMF/pyridine/H₂O=50/10/40 (v/v/v). Succinic anhydride solution was added to a final concentration of 2 mg/ml.
10 The sample was incubated at room temperature for 1 hour, followed by purification of peptides using C18 column. Labeled peptides are analyzed by mass spectrometry.

These results demonstrate that glycosylated secreted
15 proteins can be isolated, identified and quantified.

EXAMPLE IV

Quantitative Glycopeptide Profiling of Cell Surface 20 Proteins

This example describes profiling of cell surface glycoproteins.

25 To assess the potential of the glycopeptide capture method for the analysis of cell surface proteins, a crude membrane fraction from the LNCaP prostate cancer epithelial cell line was used to select and identify peptides containing N-linked glycosylation sites
30 (Horoszewicz et al., Prog. Clin. Biol. Res. 37:115-132 (1980)). The released peptides isolated from 60 µg of a

crude membrane fraction were analyzed by single dimension μ LC-MS/MS and the data were processed.

Briefly, glycopolypeptides were isolated essentially
5 as described in Example I. For the analysis of cell
surface proteins, 4 mg of crude membrane fraction from
the prostate cancer cell line, LNCaP (grown in RPMI
medium supplemented with 10% fetal bovine serum), were
dissolved in 1% NP40, 6 M urea, 100 mM Tris buffer, pH
10 8.3. The buffer was changed to coupling buffer
containing 100 mM NaAc, 150 mM NaCl, pH 5.5, using a
desalting column (Bio-Rad; Hercules CA). Sodium
periodate solution was added at 15 mM to the samples.
The cap was secured and the tube was covered with foil.
15 The sample was rotated end-over-end for 1 hour at room
temperature. The sodium periodate was removed from the
samples using a desalting column. A 50 μ l aliquot was
taken before coupling the sample. To the sample was
added 1 ml of coupling buffer equilibrated hydrazide
20 resin (Bio-Rad). The resin and sample were capped
securely and rotated end-over-end for 10-24 hours at room
temperature.

After the coupling reaction was complete, the resin
25 was spun down at 1000xg for 10 min, and non-glycoproteins
were washed away extensively by washing the resin 3 times
with an equal volume of 8M urea/0.4M NH_4HCO_3 . The
proteins on the resin were denatured in 8M urea/0.4M
 NH_4HCO_3 at 55°C for 30 min, followed by 3 washes with the
30 urea solution. After the last wash and removal of the
urea buffer, the resin was diluted 4 times with water.
Trypsin was added at a concentration of 1 μ g of

trypsin/100 µg of protein and digested at 37°C overnight. The trypsin released peptides were removed by washing the resin with an equal volume of 1.5 M NaCl for 3 times, 80% MeCN/0.1% TFA for 3 times, 100% methanol for 3 times, and
5 0.1 M NH₄HCO₃ for 6 times. N-linked glycopeptides were released from the resin by digestion with N-glycosidase overnight. The resin was spun and the supernatant saved. The resin was washed twice with 80% MeCN/0.1% TFA and combined with the supernatant. The resin was saved for
10 O-linked glycopeptide release later.

The peptides were dried in 4 tubes, and one tube was resuspended in 50 µl of 0.4% acetic acid. An aliquot of 3 µl of sample (from 60 µg original microsomal proteins)
15 was loaded on a capillary column for µLC-MS/MS analysis. CID spectra were searched against a human database using SEQUEST (Eng et al., supra, 1994) to identify the glycopeptides and glycoproteins (see Figures 7 and 8 and Table 3).

20

As shown in Figure 7, 1203 unique proteins were identified from the microsomal fraction of LNCaP cells using ICAT reagent followed by intensive 3D chromatography to fractionate the peptide mixture. Using
25 glycopeptide analysis, 64 unique proteins were identified. Of these, 35 glycopolypeptides were identified that were not identified from the total microsome fraction analysis. Table 3 shows glycoproteins and glycopeptides (SEQ ID NOS:64-174) as well as the
30 subcellular localization from a crude membrane fraction of the prostate cancer cell line LNCaP. The

glycopeptides contain the conserved N-linked glycosylation motif (NXS/T) (indicated in bold).

The subcellular localization of the identified proteins was further analyzed using information from SWISS-PROT database (www.expasy.org/sprot/) or prediction tool, PSORT II (psort.ims.u-tokyo.ac.jp/). As shown in Figure 8, of a total of 64 identified glycoproteins, 45 (70%) were bona fide or predicted transmembrane proteins. The non-transmembrane proteins were mostly designated as either extracellular (7 proteins, 11%) or lysosomal (9 proteins, 14%), two cellular compartments known to be enriched for glycoproteins. Only three proteins were assigned as cytoplasmic proteins (5%). Interestingly, two previously identified antigens, melanoma-associated antigen ME491 (CD63) and prostate-specific membrane antigen I (FOH1) were also identified in this experiment. These data indicate a marked improvement in selectivity for cell surface proteins over the analysis of crude microsomal fractions. Over 40% of the proteins identified were not membrane proteins in analysis of crude microsomal fraction (Han et al., *Nat. Biotechnol.* 19:946-951 (2001)). The data also indicate that proteins of high molecular weight and extreme pI, typically underrepresented in analyses performed using 2DE, are readily identified by this method. This is exemplified by the identification of basement membrane-specific heparan sulfate proteoglycan core protein (gene name SW: PGBM), a 470 kDa extracellular protein, and the acidic (pI = 4.39) transmembrane protein signal sequence receptor α subunit (gene name SW: SSRA). These results indicate that the glycopeptide capture method is also

effective for the selective analysis of proteins
contained in the plasma membrane. Furthermore, proteins
that were not detectable in analysis of a total microsome
fraction were readily identified (see Figure 7). These
5 results indicate that the methods can be used to analyze
glycopolypeptides not otherwise amenable to analysis of a
total microsome protein fraction.

Table 3. Subcellular location of glycoproteins identified from LNCap cells

Gene Name a	Protein Name	Subcellular Location b	Peptide Sequence c
GP:AB002313_1	mRNA for KIAA0315 gene	Transmembrane, ER/Golgi/Plasma membrane a	K.LHVTLNCSFGR.S
GP:AB033767_1	BSCv mRNA	Transmembrane, ER	R.SINVTGQGFSLIQR.F R.TEAGAFYYVDPDPTFENFTGGVKK.Q R.AGPNGTFLVADAYK.G
GP:AB045981_1	hFKBP65 mRNA for FK506 binding protein	Extracellular	K.LLLSSETPIEGKNMSFVNDLTVTQDGR.K R.YHYNGSLMDGTILFDSSYSR.N
GP:AF089745_1	FK506-binding protein (FKBP63) mRNA	Transmembrane, ER/Mitochondrial/Cytoplasmic	R.YHYNGTFLDGTILFDSSHNR.M
GP:AF302102_1	costimulatory molecule mRNA	Transmembrane, ER/Golgi/Plasma membrane	R.YHYNGTLLDGTILFDSSYSR.N R.TALFPDLLAQGNASLR.L
GP:AJ245820_1	mRNA for type I transmembrane receptor (psk-1 gene)	Transmembrane, ER/Golgi/Plasma membrane	R.LLANSSMLGEGQVLR.S
GP:AY032885_1	Similar to gp25L2 protein	Transmembrane, ER/Golgi	K.QVALQTFGNQITIPAGGAGYK.V R.FFTTSHTPGEHQICLHSNSTK.F
GP:BC001123_1	Similar to gp25L2 protein	Transmembrane, ER/Golgi/Plasma membrane	K.IFENQITGIEAK.K
GP:BC001615_1	Similar to hypothetical protein FLJ22625	Transmembrane, Cytoplasmic/Vesicles of secretory system	
GP:BC001740_1		Extracellular	K.AVLVNITTGER.L R.LQDDVLQFQKNQTNLER.K K.VVMDIPYELWNETSAEVDLK.K
GP:BC004423_1	clone MGC:3530 IMAGE:2819660	Transmembrane, Nuclear/mitochondrial	K.LNITNIWVLDYFGGPK.I K.YHYNASLIJLDDGTILDDSTWNLGK.T
GP:BC006786_1	Similar to FK506 binding protein 9	Extracellular	
GP:BC007443_1	serine carboxypeptidase 1	ER/Mitochondrial/Cytoplasmic	R.KTITWLQAASLLFVDNPVGTGFSYVNGSG AYAK.D
GP:BC010078_1	hypothetical protein GL012	Transmembrane, Golgi/ER/Mitochondrial	R.CFATTYYLSEGGGLIFRNVTGEPNCRPPT R.G
GP:BC015678_1		Mitochondrial/Cytoplasmic	R.YHYNGTLLDGTSEFDSYSK.G R.GILNATISVAENHPVTYK.T
GP:BC016467_1	mRNA for gp180-carboxypeptidase D-like	Extracellular	
GP:D85390_1		Transmembrane, ER	

	enzyme			R.GLVNMYPIITNLNLGQSTEYR.H R.LI.NTTDYYLLPSLNPDGFER.A R.ALGFENATQALGR.A R.VFPYISVMVNNNGSLSYDHSK.D
PIR2:A47161 PIR2:G01447	Mac-2-binding glycoprotein		Extracellular Transmembrane, Cytoplasmic/Vesicles of secretory system	R.YHYNCSLLDGTQLFTSHDYGAPEATLG ANK.V R.YSLNVTNYPVHYFDGR.K
PIR2:T42709	hypothetical protein DKFZp38610821		Cytoplasmic	R.DIENLKDASSFLAEWQNIK.G R.LLIAGTNSDDLQQLSLLESNK.D K.SLVTQYLNATGNR.W K.I LAPAYFILGNGQSGEGC*VITR.D R.TVLENSTSYEEAK.N R.FKLEWLGNCSGLNDETYGYK.E
PIR2:T47140 SW:4F2_HUMAN	hypothetical protein DKFZp761K1115.1 4f2 cell-surface antigen heavy chain (4f2hc)		Transmembrane, Vesicles of secretory system/nuclear	R.VLGFKPKPKNESLETYPVMK.Y K.YLOPLLAVOFTNLTMDEIR.I K.LHVGYLQPLVAVQVSPAPNNTGK.E
SW:ASAH_HUMAN SW:ATNB_HUMAN	acid ceramidase sodium/potassium-transporting apase beta- 1 chain		Lysosomal Type II membrane protein	K.LHVGYLQPLVAVQVSPAPNNTGKEVTVE CK.I K.IILLTCSLNDSTATEVTGHR.W
SW:ATND_HUMAN	sodium/potassium-transporting apase beta- 3 chain		Type II membrane protein	K.ITDSEDKALMNGSISR.I R.LLDAENKVVVANGTGTQGLK.V K.VVANGTGTQGLK.V K.IIISPENVTLTCTAENQLER.T
SW:BASI_HUMAN	basigin precursor (leukocyte activation antigen m6)		Type I membrane protein	K.LGDCISDSYPDGNITWYR.N R.LNLSENYTSLISNAR.I R.IVNSLNVSAISPEHDEADEISDENRE.E R.IVNSLNVSAISPEHDEADEISDENREK.V K.GSLSYLNVTR.K
SW:BGLR_HUMAN SW:C166_HUMAN	beta-glucuronidase cd166 antigen precursor (activated leukocyte-cell adhesion molecule)		Lysosomal Type I membrane protein	
SW:CATD_HUMAN	cathepsin d		Lysosomal	

SW:CATL_HUMAN	cathepsin 1	Lysosomal	K.YYKGSLSYLVNTR.K
SW:CD63_HUMAN	cd63 antigen (melanoma-associated antigen me491)	Integral membrane protein; Lysosomal	K.YSVANDTGFVDPK.Q
SW:CLUS_HUMAN	clusterin	Extracellular	R.QQMENYPKNHTASILDR.M
SW:DRN2_HUMAN	deoxyribonuclease ii (lysosomal dnase ii)	Lysosomal	K.MLNTSSLLLEQLNEQFNWVSR.L
SW:DSG2_HUMAN	desmoglein 2	Type I membrane protein	K.GHHVSVQEPWNSSITLTSQAGAVFQSPAK.F
SW:ENPL_HUMAN	endoplasmic	ER	K.DTIGELNVTSILDREETPFLLTGYALDAR.G
SW:FOH1_HUMAN	folate hydrolase (prostate-specific membrane antigen 1)	Type II membrane protein	K.HNNDTQHIWESDSNEFSVIADPR.G
			K.YLNFVKGVVDSDDLPLNVSRE
			K.FLYNFTQIPHLAGTEQNFQLAK.Q
			R.GVAYINADSSIEGNYTLR.V
SW:GL6S_HUMAN	n-acetylglucosamine-6-sulfatase	transmembrane, Lysosomal	K.JYVSFSDSLFSVKNFTEIASK.F
SW:GLCM_HUMAN	glucosylceramidase	transmembrane, lysosomal	R.VDCTPLMAYSLVHNLTK.E
			K.VPYNVGGFTGNFSTQK.V
			K.TPMTNNSIQFLDNAFR.K
			K.YYNYTLSSNGK.A
			R.DLGPTLANSTHNVRL
			R.MELSMGPQANHTGTGLLLTLQPEQK.F
			R.RMELSMGPQANHTGTGLLLTLQPEQK.F
			R.IYTYADTPDDFQLHNFSLPEEDTK.L
			R.DIVGNLLESEDIQIEALLMR.A
SW:GLG1_HUMAN	golgi sialoglycoprotein ng-160 (cysteine-rich fibroblast growth factor receptor)	Type I membrane protein, Golgi	K.LDSFGPNPTLNTTYSFLTTFEK.E
SW:HEXB_HUMAN	beta-hexosaminidase beta chain	Lysosomal	R.TAADTTGLQPLNOFTPANISR.Q
SW:ITAV_HUMAN	integrin alpha-v	type I membrane protein	R.LTGSDVYNLLAENLLSPEDMVLHNLTPR
SW:LDLR_HUMAN	low-density lipoprotein receptor	type I membrane protein	R.G
SW:LMG1_HUMAN	laminin gamma-1 chain	transmembrane, extracellular	K.LLNNLTSIK.I
SW:LMP1_HUMAN	lysosome-associated membrane glycoprotein 1	Type I membrane protein	R.GHTLTILNFTNR.N
SW:LMP2_HUMAN	lysosome-associated membrane glycoprotein 2	Type I membrane protein	K.SGPKNMTFDLPSDATVVLNR.S
			K.IAVQFGFGFSWIANFTK.A
			K.WQMNFIVR.Y

SW:LU_HUMAN SW:LYAG_HUMAN SW:LYII_HUMAN	lutheran blood group glycoprotein lysosomal alpha-glucosidase lysosome membrane protein ii	Type I membrane protein Lysosomal Type II membrane protein; lysosomal	R:IQNFTLLVQGSPELK.T R:GVFIINETGOPLIGK.V K:CNMINGTDGDSFHLITK.D R:TMVFPVMYLNESVHDK.E R:TMVFPVMYLNESVHDKETASR.L R:LEHQFAVGEDSGRNLSAPVILNR.D R:ATLITFLCDRDAGVGPEYQEDNSTYNF R:W R:HGNLYDLKPLGLNDTIVSAGEYTYFFR.V K:IKTNITLVCKPGDLESAPVLR.T R:SLLEFNTVSCDQOQGINHR.V K:TNITLVCKPGDLESAPVLR.T K:LVLPAKNTITNLK.T R:SHGVQTMVVLNNLEPNTTYEIR.V R:SCINESAIDSR.G K:VMELEKEIANATAKPEDR.N R:TSLELWMHTDPVSKNESVR.N R:AEPLNASASDQGEK.V K:LGNTISSLFGGGTTDPDAKENGTDITVQEEB ESPAEGSK.D R:LSALLDNLLNHSSMFLK.G R:QTVHFQISSQLQFSPEEVLGMVNLNYSR.S R:VFGSQNLTTVK.L K:VINETWAWK.N K:VINETWAWKNA.TLAEQAK.L R:LPQVSPADSGEYVCRVENGSGPK.E
SW:MA2B_HUMAN SW:MPRI_HUMAN	lysosomal alpha-mannosidase cation-independent mannose-6-phosphate receptor	<i>transmembrane, Lysosomal</i> Type I membrane protein; lysosomal	
SW:NCM2_HUMAN SW:NEP_HUMAN SW:NICA_HUMAN SW:OXRP_HUMAN	neural cell adhesion molecule 2 nephrilysin nicastrin 150 kda oxygen-regulated protein	Type I membrane protein Type II membrane protein Type I membrane protein <i>transmembrane, ER</i>	
SW:FGBM_HUMAN SW:PPT_HUMAN SW:PTK7_HUMAN SW:SAP_HUMAN SW:SEIL_HUMAN	basement membrane-specific heparan sulfate proteoglycan core protein palmitoyl-protein thioesterase tyrosine-protein kinase-like 7 precursor (Colon carcinoma kinase-4) P07602 h proactivator polypeptide sel-1 homolog precursor (suppressor of lin- 12-like protein)	<i>transmembrane, extracellular surface</i> <i>Extracellular/vacuolar/mitochondria</i> Type I membrane protein <i>transmembrane, lysosomal</i> Integral membrane protein	K:FLNDSIVDPVDSEWFGFYR.S R:MHIFQNGSLVIHDVAPEDSGR.Y K:DVVYTAAGDMLKDNTATEEELVYLEK.T R:NLEKNSTKQEILAALEK.G K:GQTALGFLYASGLGVNSSQAK.A

SW:SPHM_HUMAN	n-sulphoglucosamine sulphohydrolase	Lysosomal	R.DAGVLNDTLVIFTSDNGIPPPGR.T R.NALLLLADDGGFSGAYNNSAIAIPILD ALAR.R
SW:SSRA_HUMAN	signal sequence receptor alpha subunit	Type I membrane protein; ER	R.YPDYQFYIQNFTALPLNTVVPPQR.Q
SW:SSRB_HUMAN	signal sequence receptor beta subunit	Type I membrane protein; ER	K.AGYFNFTSAITITLAQEDGPVVGSTAP GQGILAQR.E
SW:TPPI_HUMAN	tripeptidyl-peptidase i	Lysosomal	R.IAPASNVSHITVVLRLK.A
SW:STMI_HUMAN	stromal interaction molecule 1	Type I membrane protein cell surface	K.FLSSPHLPSSSYFNASGR.A RLAVTNTMTMTGTVLK.M

a: Gene name is from human NCBI protein database (www.ncbi.nlm.nih.gov).
b: Subcellular locations in *italic* letter are predicted by PSORT, and b) in regular letter are from SWISS-PROT
c: The consensus motif for N-linked glycosylation is highlighted.

The total number of proteins identified in this experiment is relatively small but consistent with the number of unique proteins identified from complex samples using LC-MS/MS without extensive separation. Because of
5 the "top down" mode of precursor ion selection in the mass spectrometer, the most abundant proteins are preferentially identified. To identify a higher number of proteins, the sample would have to be more extensively fractionated prior to mass spectrometric analysis.

10

The method provides for quantitative profiling of glycoproteins or glycopeptides. The method allows the identification and quantification of glycoproteins containing N-linked carbohydrate in a complex sample and
15 the determination of the site(s) of glycosylation. The selectivity of the method makes it ideally suited for the analysis of samples that are enriched in glycosylated proteins. These include cell membranes, body fluids and secreted proteins. Such samples are of great biological
20 and clinical importance, in particular for the identification of diagnostic biomarkers and targets for immunotherapy or pharmacological intervention.

By combining this method with the cysteine tagging
25 method using ICAT reagents (Gygi et al., *supra*, 1999), the occupancy of individual N-linked glycosylation sites and changes thereof can also be determined. This is of particular interest in studies in which changes of glycosylation occupancy are suspected, as exemplified by
30 patients with Type I Congenital Disorders of glycosylation, in which the pathway of N-linked

glycosylation is deficient (Aebi and Hennet, Trends Cell Biol. 11:136-141 (2001)).

5 The selectivity of the method also substantially
reduces the complexity of the peptide mixture if complex
protein samples are being analyzed because glycoproteins
generally only contain a few glycosylation sites. The
method is focused on the analysis of N-linked
glycosylation sites. Analogous strategies can be devised
10 to also analyze O-glycosylated peptides and in fact, a
protein sample, once immobilized on a solid support, can
be subjected to sequential N-linked and O-linked
glycosylation peptide release, thus further increasing
the resolution of the method and the information contents
15 of the data obtained by it. Therefore, the method has
wide applications in proteomics research and diagnostic
applications.

20 These results show that membrane glycopolypeptides
can be readily analyzed. Furthermore, glycopolypeptides
that were not detectable in analysis of a total microsome
fraction were readily identified (see Figure 7). These
results indicate that the methods can be used to analyze
glycopolypeptides not otherwise amenable to analysis of a
25 total microsome protein fraction. Also note that the
method simplifies the analysis and focus on proteins
located in plasma membrane and extracellular surface,
which have therapeutic value for easy drug accessibility
and antibody directed therapy.

30

EXAMPLE V
Quantitative Glycopeptide Profiling of Mouse Ascites
Fluid

5 This example describes profiling of glycoproteins
from mouse ascites fluid.

 Glycopolypeptides were purified essentially as
described in Example I. For the analysis of ascites
10 fluid, 20 µl of mouse ascites fluid (600 µg total
protein) were changed to buffer containing 100 mM NaAc,
150 mM NaCl, pH 5.5, using a desalting column (Bio-Rad).
Sodium periodate solution was added at 15 mM to the
samples. The cap was secured and the tube was covered
15 with foil. The sample was rotated end-over-end for 1
hour at room temperature. The sodium periodate was
removed from the samples using a desalting column. An
aliquot of 20 µl of coupling buffer equilibrated
hydrazide resin (Bio-Rad) was added to the sample. The
20 sample and resin were capped securely and rotated end-
over-end for 10-24 hours at room temperature.

 After the coupling reaction was complete, the resin
was spun down at 1000xg for 10 min, and non-glycoproteins
25 were washed away extensively by washing the resin 3 times
with an equal volume of 8M urea/0.4M NH₄HCO₃. The
proteins on the resin were denatured in 8M urea/0.4M
NH₄HCO₃ at 55°C for 30 min, followed by 3 washes with the
urea solution. After the last wash and removal of the
30 urea buffer, the resin was diluted 4 times with water.
Trypsin was added at a concentration of 1 µg of
trypsin/100 µg of protein and digested at 37°C overnight.

The trypsin released peptides were removed by washing the resin with an equal volume of 1.5 M NaCl for 3 times, 80% MeCN/0.1% TFA for 3 times, 100% methanol for 3 times, and 0.5 M NaHCO₃ three times, and the resin was resuspended in 20 µl of 0.5 M NaHCO₃, pH 8.0.

For modification of peptides, 0.22 M of Boc-d0-Phe-OH (Nova Biochem) or Boc-d5-Phe-OH (CDN Isotopes) were dissolved in anhydrous N,N-Dimethylformamide. 1,3-Diisopropylcarbodiimide was added to a final concentration of 0.2 M, and the reaction was carried out at room temperature for 2 hours. A 10 µl aliquot of Boc-Phe-anhydride heavy or light forms was added to 10 µl of glycopeptides on the beads and incubated at room temperature for 30 min. The beads were washed with 80% MeCN/0.1% TFA three times, combined and dried. The Boc was removed by incubating with TFA for 30 min at room temperature. The beads were washed with glycosidase buffer, followed by release of the labeled glycopeptides with N-glycosidases at 37°C overnight. N-glycopeptides were dried and resuspended in 20 µl of 0.4% of acetic acid. A 2 µl aliquot was analyzed by LC-MS/MS to determine the quantification of N-terminal labeling of glycopeptides by Phe (see Figures 9-12).

Mass spectrometry analysis of the peptide by LCQ and searching protein database by Sequest resulted in the identification of N-glycosylated peptides with the conserved N-glycosylation motif NXS/T. More than 50 glycoproteins were identified from 20 µl of mouse ascetic fluid, indicating the method is sensitive and useful for

the identification of the glycoproteins from biological samples.

As shown in Figure 9, isotopic labeling with Phe was
5 performed with two equal amounts of mouse ascites fluid
(1 μ l), and the formerly N-linked glycopeptides were
identified using MS/MS. Figure 10 shows the list
identified peptides after isotopically labeling with Phe.
The corresponding collision-induced dissociation (CID)
10 spectrum of one of the identified peptides, indicated by
a circle, was shown in Figure 11.

Figure 12 shows reconstructed ion chromatograms for
the peptide measured in Figure 11. The ratio of the
15 calculated peak area for the heavy and light form of the
isotope tagged peptides was used to determine the
relative peptide abundance in the original mixtures
(light scan: mass 1837.0; heavy scan: mass 1842.0). The
ratio (0.81:1) agreed reasonably well with the expected
20 ratio of 1 to 1.

These results show that glycopolypeptides from
complex body fluids can be analyzed, identified and
quantified. Using isotope tags, two samples were
25 compared and the relative amount of peptide in the
original mixtures was determined, showing that the
methods can be used quantitatively.

EXAMPLE VI

**Quantitative Glycopeptide Analysis of Control
Glycoproteins with a Known Ratio**

5 This example describes quantitative analysis of glycoproteins from a pure glycoprotein mix with a known ratio and from two equal amounts of a human serum protein mix.

10 Two mixtures containing the same three glycoproteins at different amounts were prepared. The proteins were purchased from Calbiochem (San Diego, CA). The amount of each protein (μ g) in mixture A and B were: α -1-antitrypsin (50, 10), α -2-hs-glycoprotein (10, 30), and
15 α -1-antichymotrypsin (2, 2). Formerly N-linked glycosylated peptides from the two protein mixtures were purified and labeled as described in Example I.

 Formerly N-glycosylated peptides were analyzed by
20 μ LC-ESI-MS/MS and identified. Table 4 shows the identified sequences (SEQ ID NOS:175-179) and the observed d0/d4 peptide ratio for each identified peptide from two experiments. Of the four identified N-glycosylation sites, three have been described previously
25 (Yoshioka et al., J. Biol. Chem. 261:1665-1676 (1986); Mills et al., Proteomics 1:778-786 (2001); Baumann et al., J. Mol. Biol. 218:595-606 (1991)), while N# in the sequence FN#LTETSEAEIHQSFQH (SEQ ID NO:180) represents a glycosylation site in α -1-antichymotrypsin that has not
30 been described previously. The abundance ratios calculated from the isotopic ratios agreed reasonably with the expected values. These results indicate that the

method selectively isolates and quantifies N-linked glycopeptides from mixtures of glycoproteins.

Table 4. Quantitative analysis of glycoproteins in glycoprotein mixture

Protein Name	Sequences of Identified Peptides	Glycosylation Sites	Observed Peptide Ratio (A/B)	Protein Ratio (A/B)	Expected Protein Ratio (A/B)
	K.FN#LTETSEAEIHQSFQHL	Novel	0.69; 0.91		
α -1-antichymotrypsin	F.LSLGAHN#TTLTEILK.G	Known	1.63; 1.35	1.09 $\pm$ 0.39	1.00
	L.SISTALAEFLSLGAHN#TTLTEILK.G	Known	0.88		
α -1-antitrypsin	R.QLAHQSN#STNIFF.S	Known	6.47; 4.06	5.27 $\pm$ 1.70	5.00
α -2-hs-glycoprotein	K.AALAAFNAAQNN#GSNFQLEISR.A	Known	0.34; 0.51	0.42 $\pm$ 0.12	0.33

The specific capture of glycoproteins is based on the oxidation of hydroxyl groups on adjacent carbon atoms of carbohydrates to aldehydes by sodium periodate as previously described (Bobbitt, Adv. Carbohydr. Chem. 11, 1-41 (1956)). The aldehydes in turn covalently couple to amine- or hydrazide-containing molecules (Bayer et al., Anal. Biochem. 170:271-281 (1988)). Under the conditions used, the only expected side reaction of sodium periodate oxidation resulting in aldehydes is the oxidation of polypeptides containing a primary amine and a secondary hydroxyl group on adjacent carbon atoms, as exemplified by N-terminal serine residues (Geoghegan and Stroh, Bioconjug. Chem. 3:138-146 (1992)). This constellation is rare in proteins. The attachment of periodate oxidized proteins to hydrazide resin is therefore quite specific for glycoproteins containing N-linked and/or O-linked carbohydrates. Different types of oligosaccharides oxidize at different periodate concentrations and reaction conditions. The conditions used here (15 mM sodium periodate, room temperature for one hour) were chosen to assure oxidation of all types of oligosaccharides with hydroxy groups on adjacent carbon atoms. The enzyme catalyzed release of formerly N-glycosylated peptides by PNGase F provides specificity for N-linked glycopeptides and -N-linked glycosylation sites (Maley et al., Anal. Biochem. 180:195-204 (1989)). PNGase F will not, however, release N-linked oligosaccharides containing core fucosylation.

It was also determined whether the glycopeptide selection method could be used for detecting quantitative changes in the profiles of N-linked glycopeptides

isolated from different samples of human serum. In a proof-of-principle experiment, glycopeptides from two equal amounts of human serum (1 mg total protein) were isotopically labeled with either light (d0) or heavy (d4) forms of succinic anhydride at N-termini after C-terminal lysine residues were converted to homoarginines as described in Example I. The lysine-to-homoarginine conversion facilitated detection by MALDI quadrupole time-of-flight (MALDI-QqTOF) mass spectrometry and the stable isotope tag was incorporated for quantification. After labeling, the beads containing the two samples were combined, and the formerly N-linked glycopeptides were released. A fraction of the sample, equivalent to 1.25 µl of serum, was fractionated to 29 spots on a MALDI plate by RP-LC and analyzed by MALDI-QqTOF MS and MS/MS. The experiment was repeated and analyzed by ESI-QqTOF MS, and the results were comparable to those identified by MALDI-QqTOF MS. Table 5 lists the identified peptides (SEQ ID NOS:181-197), the proteins from which they originated and their observed quantitative ratio from two experiments. Generally, the observed ratios were close to the expected ratio of 1. The differences between the observed and expected ratio ranged between 0%-29% with a mean of 8%. This indicates that the glycopeptide capture method allows reasonable quantification if combined with stable isotope tagging.

The quantification is further illustrated for a single peptide pair in Figure 13. A single scan of the mass spectrometer at spot 28 in MS mode identified eight paired signals with a mass difference of four units (indicated with *, Figure 13). An expansion of the mass

range between $m/z=1577$ and $m/z=1590$ resolved the natural isotopic distribution of a peptide pair with monoisotopic peaks at 1579.74 and 1583.78, in which the signals had a quantitative ratio of 1.11.

Table 5. Quantitative analysis of glycoproteins from two identical serum samples

Gene Name a	Protein Names	Sequences of identified peptides b	Observed Ratio (Mean±SD)	Expected Ratio	% Error
GP:AF384856_1	peptidoglycan recognition protein L	R.GFGVAIVGN#YTAALPTEAALR.	0.95±0.02	1	5
GP:M36501_1	α-2-macroglobulin	Y.VLDYLN#ETQQLTPEIK.S	0.93±0.03	1	7
SW:AIAG_HUMAN	α-1-acid glycoprotein 1	N.LVPVPTN#ATLDQITGK.W	1.05±0.11	1	5
SW:AIAT_HUMAN	α-1-antitrypsin	K.YLGN#ATAIFLPDEGK.L	1.10±0.11	1	10
SW:AIAT_HUMAN	α-1-antitrypsin	R.QLAHQSN#STNIFF.S	1.00±0.01	1	0
SW:AACT_HUMAN	α-1-antichymotrypsin	K.YTGN#ASALFILPDQDK.M	1.05±0.03	1	5
SW:CO3_HUMAN	complement c3	N.HMGN#VFTTIPANR.E	0.91±0.02	1	9
SW:CO4_HUMAN	complement c4	R.FSDGLESN#SSTQFEVK.K	0.93±0.07	1	7
SW:HPT1_HUMAN	haptoglobin-1	K.VVLHPN#YSQVDIGLIK.L	1.04±0.03	1	4
SW:HPT1_HUMAN	haptoglobin-1	K.NLFLN#HSEN#ATAK.D	1.29±0.33	1	29
SW:IC1_HUMAN	plasma protease c1 inhibitor	R.VLSN#NSDANLELINTWVAK.N	0.90±0.03	1	10
SW:ITH1_HUMAN	inter-α-trypsin inhibitor heavy chain h1	H.FFAPQN#LTNMNK.N	0.96±0.01	1	4
SW:ITH2_HUMAN	inter-α-trypsin inhibitor heavy chain h2	K.GAFISN#FSMTVDGK.T	1.08±0.12	1	8

SW:ITH4_HUMAN	inter- α -trypsin inhibitor heavy chain I α 4	NQLVDALTTWQN#K.T	1.01 $\pm$ 0.13	1	1
SW:KAIN_HUMAN	Kallistatin	K.FLN#DTMAVYEAK.L	1.24 $\pm$ 0.30	1	24
SW:KAL_HUMAN	plasma kallikrein	R.IYSGILN#LSDITK.D	1.06 $\pm$ 0.08	1	6
SW:KNG_HUMAN	Kininogen	K.LNAENN#ATFYFK.I	0.94 $\pm$ 0.10	1	6

a: Gene name is from human NCBI protein database (www.ncbi.nlm.nih.gov).

b: The consensus motif for N-linked glycosylation is highlighted and the asparagine residues to which carbohydrate linked are N#.

These results indicate that the method selectively isolates and quantifies N-linked glycopeptides from mixtures of glycoproteins.

5

EXAMPLE VII**Identification of N-linked Glycosylation Sites and
Consensus Motif for N-linked Glycosylation**

10 This example describes the identification of asparagine residues that are occupied by N-linked carbohydrates in the native protein and determination of consensus motif from the alignment of identified N-linked glycosylation sites.

15

Glycoproteins were conjugated to hydrazide resin and released from the solid support by PNGase F as described in Example I. PNGase F catalyzed cleavage of oligosaccharides from glycoproteins deaminates the linker
20 asparagine to aspartic acid causing a mass shift of one mass unit. The single mass unit differences between asparagine and aspartic acid were detected by mass spectrometers and identify the asparagine residues to which the oligosaccharides were attached.

25

The one mass unit difference caused by conversion of asparagine to aspartic acid after cleavage of oligosaccharides from glycoproteins was specified in Sequest search parameter during database search of the
30 MS/MS spectra. The acquired MS/MS spectra were searched against the human protein database from NCBI. For MS/MS spectra acquired by MALDI QqTOF (MDS SCIEX; Concord,

Ontario CA), the mass window for the singly-charged ion of each peptide being searched was given a tolerance of 0.08 Da between the measured monoisotopic mass and the calculated monoisotopic mass, and the b, y, and z ion series of the database peptides were included in the Sequest analysis. For MS/MS spectra acquired by a Finnigan LCQ ion trap mass spectrometer, the mass window for each peptide being searched was given a tolerance of 3 Da between the measured average mass and the calculated average mass, and the b and y ion series were included in the Sequest analysis. The sequence database was set to expect the following possible modifications to certain residues: carboxymethylated cysteines, oxidized methionines and an enzyme catalyzed conversion of Asn to Asp at the site of carbohydrate attachment. There were no other constraints included in the Sequest search.

The precursor ion with $m/z = 1579.74$ identified in Example VI was further analyzed by MS/MS and sequence database searching of the resulting spectrum, and it was identified with peptide sequence IYSGILN#LSDITK from human plasma kallikrein, a serum protease (Figure 14). N# indicates the modified asparagine in the peptide sequence. The series of y ions from this peptide confirmed the match and indicates that the single mass unit difference between asparagine and aspartic acid can be easily detected by MALDI QqTOF mass spectrometry, thus confirming the precise glycosylation site within the peptide as N7.

The peptides identified with N to D conversion were aligned using Sequence Logos (Schneider and Stephens,

Nucleic Acids Res. 18:6097-6100 (1990)). Figure 15 shows the patterns of aligned sequences. For each position in the aligned sequence, the height of each letter is proportional to its frequency, and the most common one is on top. As expected, there was high preference of N at position 21 in Figure 15 (removed to show the detail of other positions). The preference of N was followed by S or T at position 23 (removed to show residues in other positions). This is a known consensus N-linked glycosylation motif. In addition, the preference of L, V, A, S, G at positions 9, 15, 20, 22, 24, 28, 29 was identified.

The identified glycopeptides were used to build a glycopeptide database. When searching a human database for potential N-linked glycosylation motifs with the previously defined NXS/T sequence, sixty percent of human proteins contain the consensus N-linked glycosylation motif. The alignment of identified N-linked glycopeptides by the glycopeptide capture method described here refined and extended the consensus N-linked glycosylation motif. The refined motif is used to generate an algorithm to search the entire database for possible N-linked glycosylation sites. This increases the database searching constraints and reduces the propensity of false identifications. Protein topology of known proteins or predicted protein topology from prediction programs such as PSORT II can be used to further increase the confidence of the predicted N-linked glycosylation motif since it is known that N-linked glycosylation occurs on extracellular domains and on the protein surface.

The increased prediction power for N-linked glycosylation sites can be used to search the candidate genes specific to ovarian cancer from microarray data of normal and ovarian cancer samples. The predicted N-linked glycosylation peptides are synthesized with the incorporation of stable isotope amino acids. 500 fmole of synthetic peptides are mixed with peptides purified from normal and ovarian cancer serum using the glycopeptide capture method described in Example I. The relative abundance of candidate peptides in normal and cancer patients are quantified with high accuracy and sensitivity. Since the peptide mass and MS/MS spectra of each synthetic peptide are known, the mass spectrometer can be set to run in single reaction monitor mode (SRM) with increased sensitivity and accuracy of quantification.

This example describes an exemplary method to identify glycosylation sites.

EXAMPLE VIII

Quantitative Profiling of Glycoproteins in Extracellular Matrix of Prostate Cancer

Prostate cancer is the most common cancer in men in the Western world, and the second leading cause of cancer mortality. The prostate is remarkably prone to developing cancer, and because little is known about the cause, no preventive measures can be formulated. With the use of prostate-specific antigen (PSA)-based

screening, 80% of prostate cancer can be detected at a stage where it can be treated by local therapies. However, the rate of treatment failure as indicated by rising PSA levels can range from 10% to 40%. Apparently
5 the escape of cancer cells from the prostate is an early event, and many patients test positive for these cells in their blood and bone marrow. A challenge in the diagnosis and treatment of prostate cancer is to develop better markers for cancer diagnosis to detect the disease
10 at an early, more curable stage; to molecularly define prostate cancer progression for more accurate prognosis; and to identify cancer cell surface specific antigens as therapeutic targets.

15 Tumor and benign tissue samples from the peripheral zone of the prostate of the same patients were handled under sterile conditions. Tissue specimens were minced and digested with collagenase in RPMI-1640 medium supplemented with 10^{-8} M dihydrotestosterone (Liu et al.,
20 Prostate 40:192-199 (1999)). The digestion medium was saved, and glycoproteins were isolated as described in Example I.

The extracellular matrix protein species from
25 patient-matched normal and cancer samples were processed by the glycopeptide capture method as described in Example I. The peptides released from the hydrazide resin were resuspended in 20 μ l of 0.4% acetic acid. A 5
30 μ l aliquot of sample was analyzed by μ LC-MS/MS analysis, and the CID spectra were searched against the Human NCI database using Sequest.

Figure 16 shows the proteins identified from normal and cancer tissues. Two cancer specific proteins, prostate-specific antigen (PSA) and prostatic acid phosphatase (PAP), were readily detected in cancer tissues.

The formerly N-linked glycosylated peptides are labeled with light and heavy succinic anhydride as described in Example I, and peptides from normal (labeled with light succinic anhydride) and cancer (labeled with heavy succinic anhydride) samples are combined and analyzed by LC-MS/MS. The CID spectra are searched against a human database, and the identified proteins are quantified using stable isotope quantification software tools such as ASASratio, and Express (Han et al., Nat. Biotechnol. 19:946-951 (2001)).

Since the concentration of specific proteins at the cancer tissue is much higher than that in blood serum, the cancer specific surface proteins are easily detected. The identified proteins can serve as cancer cell surface specific therapeutic targets. To determine the existence of cancer specific proteins in serum of prostate cancer patients, synthetic peptides are mixed with glycopeptides isolated from serum and analyzed by mass spectrometry. SRM mode analysis is used in the analysis and it increases the specificity and sensitivity of detecting the peptides in patient serum for early detection markers.

This example describes the identification of markers from cancer samples as potential diagnostic markers and/or therapeutic targets.

5

EXAMPLE IX**Quantitative Profiling of Glycopeptides and Identification of Biomarkers from Mice with Skin Cancer**

10 This example describes identification of biomarkers associated with skin cancer.

Mass spectrometry has recently been used as a platform for protein-based biomarker profiling (Petricoin et al. Lancet 359:572-577 (2002)). It has been shown
15 that pathological changes of tissues and organs are reflected in serum protein changes while blood circulates in the body. The reduced sample complexity and enriched biological information from the glycopeptide capture method provides advantages for the systematic
20 investigation of serum protein expression patterns of thousands of proteins in serum.

Several advantages of using the glycopeptide capture method and peptide mass to identify serum biomarkers are
25 as follows. (1) It is fast and obviates the need for extensive separation methods. Because of the top down mode of operation of tandem mass spectrometry in the time available during a LC-MS/MS experiment, only a fraction of the peptides present is selected for CID to identify
30 the peptide sequence. Consequently, if peptides are detected by their mass only, in the time period of a LC-MS/MS experiment a significantly higher number of

peptides can be detected than sequenced. This is illustrated in Figure 17. The total number of peptides present in a single LC-MS/MS run is shown, and identified peptides during the run. It was consistently found that
5 less than 10% of total peptides were identified for complex biological samples. (2) The glycopeptide capture method simplified the total peptides present in serum after protease digestion and removed the heterogeneity of peptides caused by different oligosaccharides
10 modifications and break down during MS analysis. (3) The majority of proteins and peptides in biological samples were unchanged in different states of the samples. Analyzing the relative abundance of all the peptides present in LC-MS, the peptides that change in abundance
15 can be identified and the CID analysis focused on the differentially expressed proteins for identification.

The strategy used to identify the biomarkers in serum is shown schematically in Figure 18. Glycopeptides
20 from 100 µl of serum from 10 normal and 3 diseased mice were purified as described in Example I. The peptides were resuspended in 30 µl 0.4% acetic acid, and 5 µl of samples were analyzed by LC-MS/MS. Figure 19 shows the signal intensity of peptides during the elution of the
25 LC-MS/MS run. N1 and N2 were from normal mice, and T1 and T2 were glycopeptides from mice serum with skin cancer. Reproducible patterns of peptides from individual mice were observed during the LC-MS/MS runs.

30 Peptide peaks from different charge states in the entire run were deconvoluted to signal charged peptides. Figure 20 shows the deconvoluted peptides intensity

during different elution time from normal and skin cancer mice. About 3000 peptides were consistently observed in different samples. The peptides were then aligned by elution time using in-house developed software, and
5 normalized to background to reduce the variation of different runs. The relative peptide intensity of cancer mouse to normal mouse was calculated and shown in Figure 21.

10 To facilitate the quantification, an equal amount of peptides from all 13 mice was mixed and analyzed by mass spectrometry as control. The relative intensity of each peptide to the control after alignment was obtained. The relative peptide intensities from all 13 mice from two
15 different experiments was analyzed by unsupervised hierarchical clustering ((Eisen et al., Proc. Natl. Acad. Sci. USA 95:14863-14868 (1998))). No predefined reference vectors and prior knowledge of normal or cancer were used. In this clustering analysis, relationships among
20 peptides were represented by a tree whose branch lengths reflect the degree of similarity between the objects. As shown in Figure 22, all the cancer mice were found clustered together (indicated as 11A, 12A, 13A in experiment one, and M11, M12, M13 in experiment two).
25 The peptide intensity shown in dark gray indicates that the peptide abundance is lower than the corresponding peptide intensity in the common control, and the peptide intensity shown in light gray indicates a higher abundance of the peptide compared to the common control
30 of the mixture.

This example shows that peptides isolated by the glycocapture method contain markers for cancer. The analysis of formerly N-linked glycopeptides using peptides mass and retention time increases the
5 information of peptides during the mass spectrometry analysis. This approach is capable of distinguishing the difference between normal mice and mice with cancer and identifying cancer markers from serum.

10

EXAMPLE X**Quantitative Profiling of Glycopeptides from Human Serum
Samples Obtained Before and After Overnight Fasting**

15

This example describes the quantitative profiling and clustering analysis of glycopeptides from serum samples of three individuals before and after overnight fasting.

20

25

Glycopeptides from 100 μ l of serum from three persons before and after overnight fasting were purified as described in Example I. The peptides were resuspended in 30 μ l 0.4% acetic acid, and a control sample was made by mixing an equal amount (1 μ l) of every glycopeptide from all 6 samples. A 5 μ l aliquot of samples was
analyzed by LC-MS/MS. The peptide peaks were
deconvoluted to single charged peptides. After alignment and normalization of different runs, the relative intensity of each peptide to the common control sample was determined.

30

The relative peptide intensities from three individuals before and after overnight fasting were

determined from each experiment and were analyzed by unsupervised hierarchical clustering without prior knowledge of any specificity and conditions of the individual samples ((Eisen et al., Proc. Natl. Acad. Sci. USA 95:14863-14868 (1998))). In this clustering analysis, relationships among peptide were represented by a tree whose branch lengths reflect the degree of similarity between the objects. As shown in Figure 23, it was found that, in both experiments, serum samples from each individual before and after breakfast clustered together (indicated by person 1-3) in both experiments. The color coding is similar to that shown in Figure 22.

These results show that peptides isolated by the glycocapture method from serum samples of each individual before and after overnight fasting are most closely related. The analysis of formerly N-linked glycopeptides using peptide mass and retention time increases the information on the peptides during the mass spectrometry analysis. This approach is capable of automatically distinguishing the most significant differences between the samples. This shows that glycopeptides from individual serum samples contain the characteristic features that can be used to assess the physiological state of an individual.

EXAMPLE XI

Determination of Glycosylation Occupancy from Serum
Samples Obtained from Healthy Individuals or Patients
with Type I Congenital Disorders of Glycosylation (CDG)

5

This example describes glycopeptide profiling of
individuals with disorders of glycosylation.

Quantitative analysis of N-linked glycosylation is
capable of determining the relative N-linked
glycosylation in different proteomes. The cysteine
tagging method can be used to determine the relative
protein changes in different proteomes (Gygi et al., Nat.
Biotechnol. 17:994-999 (1999)). By combining
quantitative analysis of N-linked glycosylation with
cysteine tagging, the occupancy of individual N-linked
glycosylation sites and changes thereof can also be
determined. This is of particular interest in studies in
which changes of glycosylation occupancy are suspected,
as exemplified by patients with Type I Congenital
Disorders of glycosylation (CDG), in which the pathway of
N-linked glycosylation is deficient (Aebi and Hennet
Trends Cell Biol. 11:136-141 (2001)). In addition,
changes in the extent of glycosylation and the
carbohydrate structure of proteins on the cell surface
and in body fluids have been shown to correlate with
cancer and other disease states, highlighting the
clinical importance of this modification as an indicator
or effector of pathologic mechanisms (Spiro, Glycobiology
12:43R-56R (2002); Freeze Glycobiology 11:129R-143R
(2001); Durand and Seta, Clin. Chem. 46:795-805 (2000)).

The glycosylation occupancy study of serum from CDG patients is described in Figure 24. The ratio of total serum protein level of an individual was quantified using the ICAT reagent, and the ratio of N-linked glycopeptides of the individual is determined by glycopeptide capture followed by N-terminal isotopic labeling. The glycosylation occupancy is determined by the ratio of each N-linked glycopeptides divided by total protein ratio of the proteins.

To determine the relative ratio of total protein, the ICAT reagent was used to label the protein. Seven samples containing 0.5 mg of serum proteins from normal person #1 was labeled with the ICAT light reagent, and 0.5 mg of serum proteins from normal person #1, normal person #2, CDG 1a patient #1, CDG 1g patient #2, CDG 1b patient #1, CDG 1b patient #2, and CDG 1b patient #3 were labeled with ICAT heavy reagent. The ICAT reagent was purchased from Applied Biosystems, and labeling was performed following the manufactory's instruction.

Briefly, serum proteins (0.5 mg, 6.25 μ l) were added to 0.5 ml of ICAT labeling buffer (6M urea, 0.05% SDS, 200 mM Tris, 5 mM EDTA, pH 8.3). The samples were reduced by adding 8 mM tris-carboxyethylphosphine (TCEP) and incubating at 37°C for 45 minutes. Five fold excess of light and heavy ICAT reagents was added, and labeling was performed in the dark at 37°C for 2 hours. The seven samples labeled with heavy ICAT reagent were mixed with one of seven normal samples labeled with light ICAT reagent. The seven mixed samples were diluted ten fold, and 5 μ g of trypsin was added and incubated at 37°C

overnight. The ICAT labeled tryptic peptides were purified by avidin affinity chromatography using a Vision chromatography workstation from Applied Biosystems (Foster City, CA). The peptides were resuspended in 20 μ l of 0.4% acetic acid, and 5 μ l of peptides were analyzed by Finnigan LCQ ion trap mass spectrometer (Finnigan, San Jose, CA). The CID spectra were searched against the human NCI database using Sequest. A suite of software tools developed at the Institute for Systems Biology were used to analyze protein identification and relative expression ratio, including probability analysis of peptide and protein identification, and expression ratio of each protein (Eng et al., J. Am. Soc. Mass Spectrom. 5:976-989 (1994); Han et al., Nat. Biotechnol. 19:946-951 (2001); Keller et al., Anal. Chem. 74:5383-5392 (2002)). The protein ratio from normal person #1 to normal person #1 is shown in Table 6. The ratio agreed well with the expected 1:1 ratio.

20

Table 6. Protein expression ratio determined by ICAT labeling

Protein name	ratio of protein expression
GP:A00279_1	0.803 +- 0.268
GP:AB064062_1	0.826 +- 0.140
GP:AB064121_1	0.975 +- 0.091
GP:AJ390244_1	1.464 +- 0.269
PIR2:A37927	0.795 +- 0.268
SW:A2HS_HUMAN	0.847 +- 0.224
SW:A2MG_HUMAN	0.967 +- 0.138

SW:ALBU_HUMAN	1.166 +- 0.036
SW:ALC1_HUMAN	1.352 +- 0.115
SW:ALC2_HUMAN	1.327 +- 0.539
SW:AMBP_HUMAN	1.208 +- 0.292
SW:APA2_HUMAN	1.047 +- 0.178
SW:APOH_HUMAN	0.991 +- 0.176
SW:CFAB_HUMAN	1.388 +- 0.757
SW:CFAH_HUMAN	1.043 +- 0.111
SW:CO3_HUMAN	0.896 +- 0.063
SW:FIBB_HUMAN	1.093 +- 0.218
SW:FIBG_HUMAN	0.949 +- 0.184
SW:FINC_HUMAN	1.391 +- 0.126
SW:GC1_HUMAN	1.195 +- 0.058
SW:GC2_HUMAN	1.047 +- 0.249
SW:GC4_HUMAN	1.195 +- 0.425
SW:HEMO_HUMAN	1.057 +- 0.086
SW:HPT1_HUMAN	1.152 +- 0.091
SW:HPTR_HUMAN	1.377 +- 0.206
SW:ITH2_HUMAN	1.330 +- 0.314
SW:KAC_HUMAN	1.072 +- 0.146
SW:LAC_HUMAN	1.063 +- 0.592
SW:MUCB_HUMAN	1.191 +- 0.536
SW:MUC_HUMAN	0.986 +- 0.221
SW:TRFE_HUMAN	1.183 +- 0.061
SW:VTDB_HUMAN	1.394 +- 0.204

To determine the relative glycosylation ratio of each N-linked glycosylation site, seven aliquots of 1 mg (12.5 μ l) from normal person #1, and 1 mg of serum from
5 normal person #1, normal person #2, CDG 1a patient #1, CDG 1g patient #2, CDG 1b patient #1, CDG 1b patient #2, and CDG 1b patient #3 were subjected to the glycopeptide

capture method as described in Example I. Glycopeptides from the seven samples from normal person #1 were labeled with light succinic anhydride, and the other samples were labeled with heavy succinic anhydride while the
5 glycopeptides were still attached to solid beads. The paired normal and seven individuals are mixed, and formerly N-linked glycosylated peptides are released. The peptides were resuspended in 20 µl of 0.4% acetic acid and 5 µl of peptides were analyzed by Finnigan LCQ
10 ion trap mass spectrometer (Finnigan, San Jose, CA). The CID spectra were searched against the human NCI database using Sequest, a suite of software tools developed in Institute for Systems Biology were used to analyze the peptide and protein probability and protein expression
15 ratio using ASAPratio. The ratio of glycosylated peptides is divided by the total protein ratio, and the glycosylation occupancy is determined for each N-linked glycosylation sites.

20

EXAMPLE XII**Determination of the Level of Glycosylation from Diabetic Obese Mouse Serum**

This example describes the determination of
25 glycosylation in a model of diabetes.

Nonenzymatic glycation in diabetes results from the reaction between glucose and primary amino groups on proteins to form glycated residues. The glycated
30 proteins and the later-developing advanced glycation end-products have been mechanistically linked to the pathogenesis of diabetic nephropathy. Glycated albumin

has been causally linked to the pathobiology of diabetic renal disease (Cohen and Ziyadeh, J. Am. Soc. Nephrol. 7:183-190 (1996)).

5 Other proteins in serum may be also responsible for the development of diabetic complications. Samples are analyzed for changes in carbohydrate modified serum proteins. Serum from wild type litter mates and diabetic obese mice from BTBR mouse strain are labeled with light and heavy ICAT reagent as shown in Figure 25. The
10 labeled serum samples are divided into two equal fractions, and paired light and heavy serum from normal and diabetic obese mouse samples are mixed. One mixture is used to determine the total serum protein ratio using
15 the ICAT measurement. The second mixture is conjugated to a solid support using hydrazide chemistry. The cysteine containing peptides from glycoproteins are released by trypsin and isolated by avidin chromatography column using the Vision chromatography workstation (ABI).
20 The relative abundance of glycoproteins between normal and diabetic mice is determined. After normalization to the total protein in serum, the changes of glycosylation are determined.

25 This experiment shows that the glycopeptide capture method can be used to analyze enzymatically glycosylated proteins as well as non-enzymatic lysine glycation of proteins. The level of non-enzymatic glycation increases in certain diseases caused by diabetes due to the high
30 glucose levels in the patient's blood serum.

EXAMPLE XIII**Quantification of N-linked Glycopeptides Using Heavy Isotope Labeled Synthetic Peptide Standards**

5 This example describes quantification using labeled synthetic peptide standards.

10 Table 7 shows several synthetic peptides (SEQ ID NOS: 198-209) identified from human serum, as described in Example II. The peptides were synthesized using standard solid phase synthesis chemistry with the carbon 13 amino acid incorporated in the valine residues at the underlined position. The glycosylated asparagines were also changed to aspartic acid. 500 fmol of each peptide

15 was mixed and run separately on LC-MS/MS analysis to determine the retention time and CID spectra. The same amount of peptides was mixed with human serum samples from three individuals to determine the relative amount of these glycopeptides in serum.

20

Figure 26 shows the synthetic peptides identified by mass spectrometry.

25 Table 7. Synthetic heavy isotope labeled peptide standards.

Protein Name	Peptide Sequence
Plasma protease C1 inhibitor precursor	GVTSVSQIFHSPDLAIRDTFVDASR
Angiotensinogen precursor [Contains:	
Angiotensin I	VYIHPFHLV <u>I</u> HDESTCEQLAK
Pigment epithelium-derived factor precursor	V <u>T</u> QDLTLIEESLTSEFIHDIDR
Serum amyloid A-4 protein precursor	SRVYLQGLIDYLLFGDSSTV <u>L</u> EDSK

Complement component C9 precursor	AVDITSENLIDDV <u>V</u> SLIR
Biotinidase precursor	YQFNTNVVFSNDGTL <u>V</u> DR
Coagulation factor XIII B chain precursor	HGVIISSTVDTYEDGSS <u>V</u> EYR
Alpha-1-acid glycoprotein 2 precursor	QNQCFYDSSYL <u>N</u> VQR
Plasma serine protease inhibitor precursor	VVG <u>V</u> FPYQGDATALFILPSEGK
Aminopeptidase N	GPSTPLPEDPNWD <u>V</u> TEFHTTPK
Antithrombin-III precursor	SLTFDETYQDISEL <u>V</u> YGAK
ICOS ligand precursor	TDNSLLDQALQDDT <u>V</u> FLNMR

The samples are analyzed for glycopeptides as described in Example I. These results show that a known amount of synthetic peptides can be used to determine the relative or absolute amount of the same glycopeptides in individual serum samples.

EXAMPLE XIV

10 Identification of O-linked Glycopeptides Using Enzymatic Cleavage

This example describes identification of O-linked glycopeptides.

15

Analogous strategies to those described herein for analysis of N-linked glycosylation sites can be used to also analyze O-glycosylated peptides. In fact, a protein sample, once immobilized on a solid support, can be subjected to sequential N-linked and O-linked glycosylation peptide release, thus further increasing the resolution of the method and the information contents

of the data obtained by it. There is no enzyme comparable to PNGase F for removing intact O-linked sugars. To release O-linked oligosaccharides, monosaccharides are sequentially removed by using a panel
5 of exoglycosidases until only the Gal β 1,3GalNAc core remains attached to the serine or threonine residue. The core can then be released by O-glycosidase. Since not all O-linked oligosaccharides contain this core structure, a chemical method, such as β -elimination can
10 be more general and effective for the release of the formerly O-linked glycosylated peptides.

After releasing N-linked glycopeptides, 100 μ l of hydrazide resin was washed with 1 ml of 1.5 M NaCl twice,
15 1 ml of 100% methanol twice, and 1 ml of water twice. O-linked glycopeptides were cleaved by a set of enzymes (Calbiochem), including Endo- α -N-acetylgalactosaminidase, Neuraminidase, β 1,4-Galactosidase, and β -N-Acetylglucosaminidase. The released peptides were dried
20 and resuspended in 0.4% acetic acid for LC-MS/MS analysis.

Figure 27 shows the identified peptides from the series of enzymatic cleavages from hydrazide resin after
25 N-linked glycopeptides were released. Unlike the N-linked glycosylation, in which PNGase F converts the glycosylated N to D after release of oligosaccharides, O-linked glycosylated serine or threonine remained unchanged. There are no known consensus motifs available
30 for O-linked glycosylation. To date, the serine or threonine residues to which the O-linked oligosaccharides were attached have not been identified.

EXAMPLE XV

Identification of Glycopeptides Isolated by Biotin Tagged
Hydrazide

The same procedure described in Example I was also performed in solution phase using biotin tagged hydrazide (PIERCE) with some modifications. After proteins were oxidized and conjugated to biotin hydrazide, the proteins were denatured in 0.5% SDS and 8M urea in 0.4 M NH_4CO_3 for 30 minutes at room temperature. The samples were diluted 4 times with water, and trypsin was added at a final concentration of 1:100. The trypsin digest was performed overnight at room temperature. The glycopeptides conjugated to biotin hydrazide were purified by an avidin column using the Vision chromatography workstation. The glycopeptides were isolated with oligosaccharides still attached to the peptides. The peptides were dried and resuspended in 0.4% acetic acid and analyzed by mass spectrometry.

When high spray voltage was used in ESI-LC-MS/MS analysis (2.0 kv), the oligosaccharides were separated from peptide at the source. It resulted in the analysis of N-linked peptides and O-linked peptides by mass spectrometry. The identified N-linked glycopeptides are

shown in Figure 28, with the consensus NXT/S motif highlighted. The O-linked oligosaccharides were removed in the source with a loss of water. This left the formerly O-linked glycosylation Ser or Thr with 18 Dalton less than the unmodified Ser or Thr. This is represented in Figure 29 at S or T without modification (shown in circles).

These results show that the glycopeptide capture method can also be performed via affinity reactive tags attached to the protein by solution chemistry. The glycopeptides isolated by this method can have oligosaccharide chains attached to the glycopeptides. Both N-linked and O-linked glycopeptides can be isolated and analyzed simultaneously.

EXAMPLE XVI

Automation of the Glycopeptide Capture Method Using a TECAN Workstation

20

This example describes adaptation of the glycopeptide analysis method to automation.

To improve the throughput and reproducibility of the method of glycopolypeptide analysis, an automated robotic workstation was designed to perform the sequence of reactions for glycopeptide isolation. The workstation is particularly useful for all applications requiring high sample throughput. The procedure described in Example I is tested in solid phase extraction format for automation. in serum biomarker identification

25
30

A TECAN workstation was designed for the glycopeptide capture procedure. The workstation is used to automate sampling and analysis of glycopeptides. The workstation can be readily adapted to diagnostic
5 applications, for example, the analysis of a large number of serum samples or other biological samples of diagnostic interest.

Throughout this application various publications
10 have been referenced. The disclosures of these publications in their entireties are hereby incorporated by reference in this application in order to more fully describe the state of the art to which this invention
15 pertains. Although the invention has been described with reference to the examples provided above, it should be understood that various modifications can be made without departing from the spirit of the invention.

What is claimed is:

1. A method for identifying and quantifying a glycopolypeptide in a sample, comprising:
 - 5 (a) derivatizing carbohydrate groups of glycopolypeptides in a polypeptide sample;
 - (b) immobilizing said derivatized carbohydrate groups of said glycopolypeptides to a solid support;
 - (c) cleaving said immobilized glycopolypeptides,
 - 10 thereby releasing non-glycosylated peptide fragments and retaining immobilized glycopeptides fragments;
 - (d) labeling said immobilized glycopeptides fragments with an isotope tag;
 - (e) releasing said glycopeptides fragments from said
 - 15 solid support, thereby generating released glycopeptides fragments;
 - (f) analyzing said released glycopeptide fragments using mass spectrometry;
 - (g) identifying a released glycopeptide fragment;
 - 20 (h) quantifying the amount of said glycopeptide fragment identified in step (g); and
 - (i) identifying a parent glycopolypeptide from which the glycopeptide fragment identified in step (g) was derived, thereby identifying and quantifying a
 - 25 glycopolypeptide.

2. A method for identifying and quantifying a glycopolypeptide in a sample, comprising:
 - (a) immobilizing glycopolypeptides in said sample to
 - 30 a solid support wherein the carbohydrate groups of said glycopolypeptides are derivatized;
 - (b) cleaving said immobilized glycopolypeptides, thereby releasing non-glycosylated peptides and retaining immobilized glycopeptides;
 - 35 (c) labeling said immobilized glycopeptides with an isotope tag;
 - (d) releasing said glycopeptides from said solid

support; and

(e) analyzing said released glycopeptides, wherein said analyzing identifies and quantifies a glycopolypeptide in said sample.

5

3. A method for identifying and quantifying a glycopolypeptide in a sample, comprising:

(a) immobilizing glycopolypeptides in said sample to a solid support wherein the carbohydrate groups of said glycopolypeptides are derivatized;

(b) cleaving said immobilized glycopolypeptides, thereby releasing non-glycosylated peptides and retaining immobilized glycopeptides;

(c) releasing said glycopeptides from the solid support; and

(d) analyzing said released glycopeptides, wherein said analyzing identifies and quantifies a glycopolypeptide in said sample.

20

4. A method of identifying a diagnostic marker for a disease, comprising:

(a) immobilizing glycopolypeptides from a test sample to a first solid support;

(b) immobilizing glycopolypeptides from a control sample to a second solid support;

(c) cleaving said immobilized glycopolypeptides, thereby releasing non-glycosylated peptide and retaining immobilized glycopeptides;

(d) labeling said immobilized glycopeptides on said first and second supports with differential isotope tags on the respective supports;

(e) releasing the glycopeptides from said solid supports; and

(f) analyzing said released glycopeptides; and

(g) identifying one or more glycosylated polypeptides having differential glycosylation between the test sample and the control sample.

5. The method of any one of claims 1 to 4, wherein said glycopolypeptides are oxidized.

5 6. The method of any one of claims 1 to 4, wherein said solid support comprises a hydrazide moiety.

7. The method of any one of claims 1 to 4, wherein said glycopolypeptides are oxidized with periodate.

10 8. The method of any one of claims 1 to 4, wherein said glycopeptides are released from said solid support using a glycosidase.

15 9. The method of any one of claims 1 to 4, wherein said glycosidase is an N-glycosidase or an O-glycosidase.

10. The method of any one of claims 1 to 4, wherein said glycopeptides are released from said solid support using sequential addition of N-glycosidase or an O-glycosidase.

11. The method of any one of claims 1 to 4, wherein said glycopeptides are released from said solid support using chemical cleavage.

12. The method of any one of claims 1 to 4, wherein said glycopolypeptides are cleaved with trypsin.

30 13. The method of any one of claims 2, 3 or 4, wherein said released non-glycosylated peptides are analyzed by mass spectrometry.

35 14. The method of claim 1, wherein said released non-glycosylated peptides are isotopically labeled and analyzed by mass spectrometry.

15. The method of claim 1, wherein said sample is selected from a body fluid, secreted proteins, and cell surface proteins.

5 16. The method of claim 4, wherein the disease is cancer.

10 17. A kit when used in a method as claimed in claim 1 comprising a solid support, a reagent for derivatizing carbohydrate groups of glycopolypeptides and an isotope tag for labeling immobilized glycopeptide fragments.

15 18. A kit as claimed in claim 17 comprising a pair of differentially labelled isotope tags.

19. A kit as claimed in claim 17 or 18 wherein the solid support comprises a hydrazide resin and the reagent comprises periodate.

20

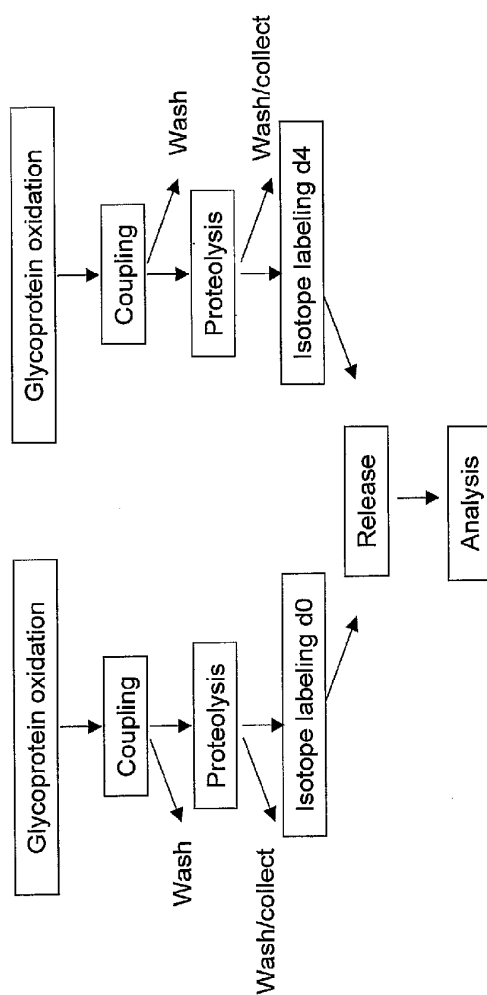


Figure 1



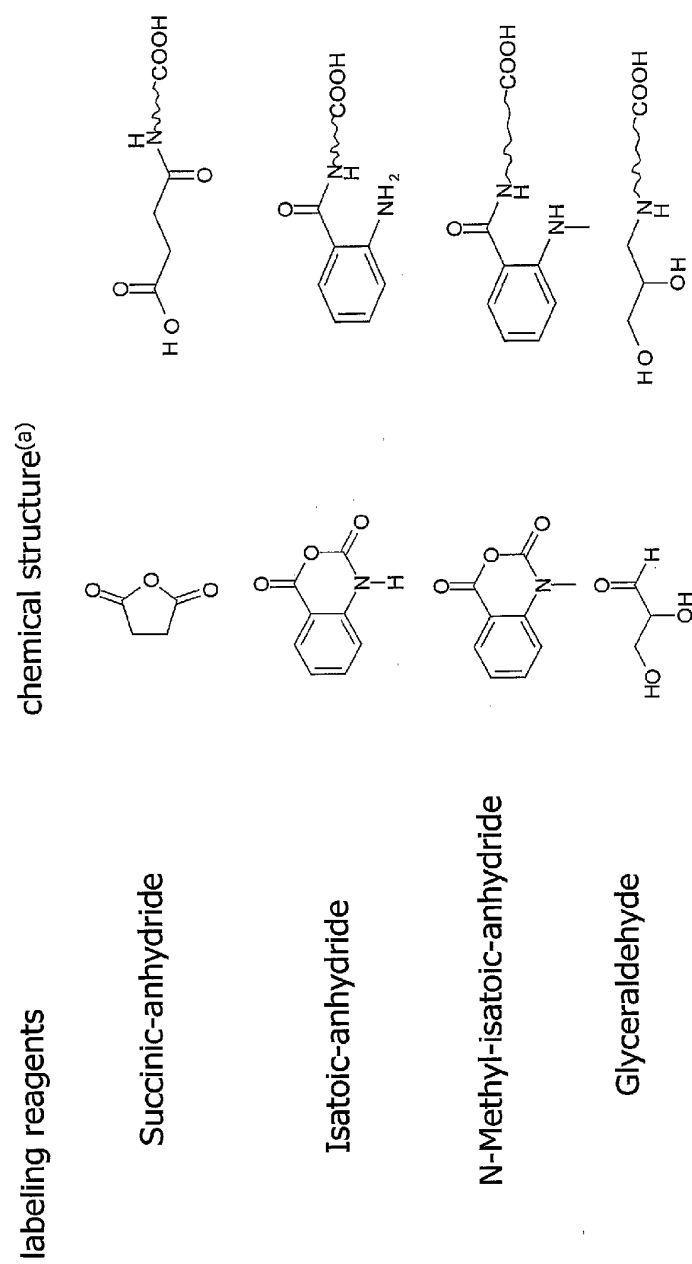


Figure 3

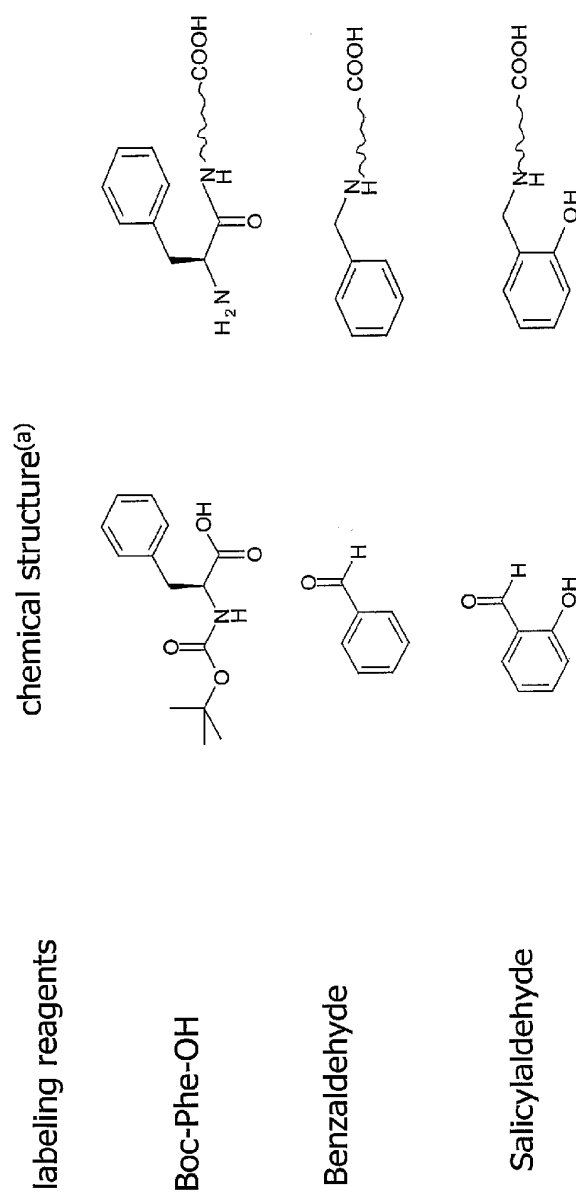


Figure 3 continued

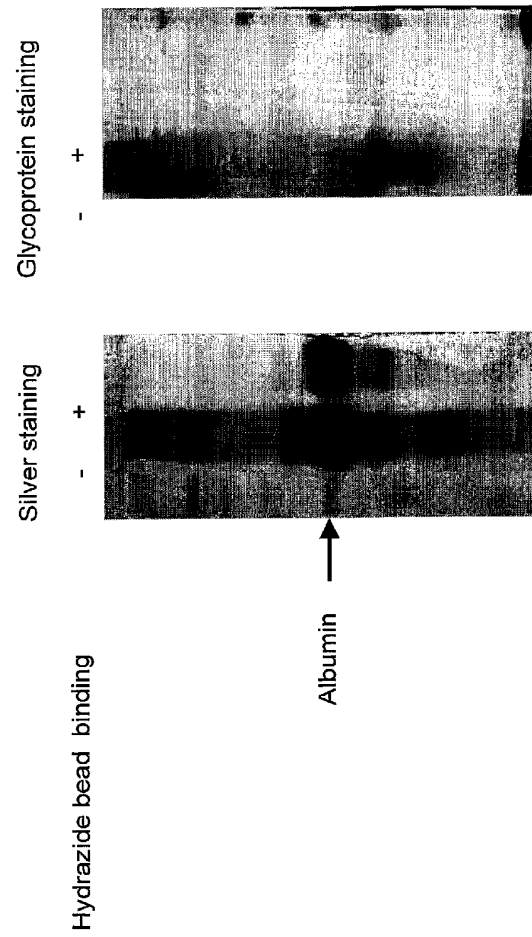


Figure 4

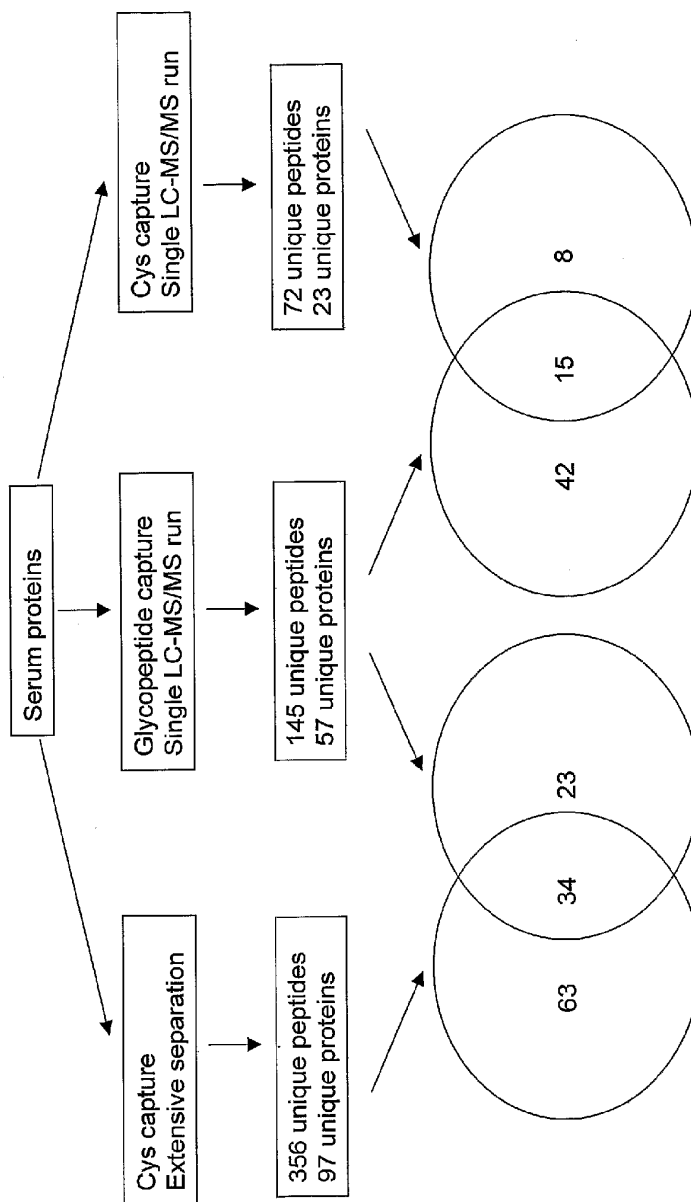


Figure 5

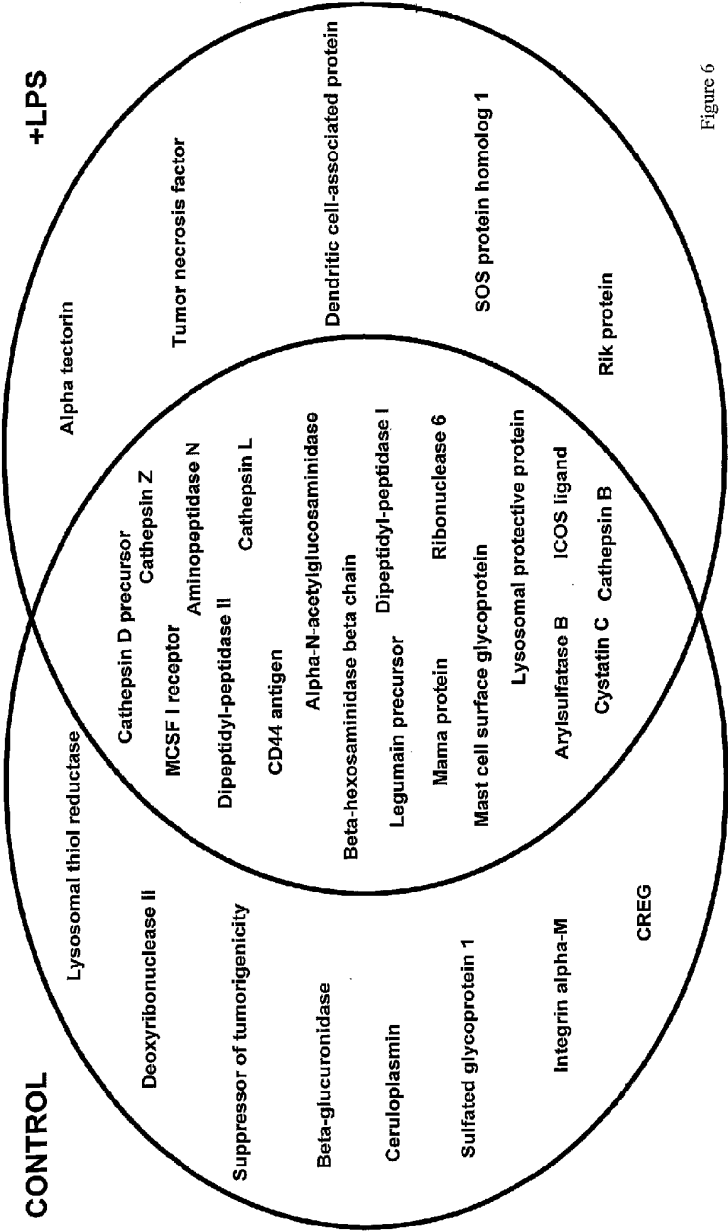


Figure 6

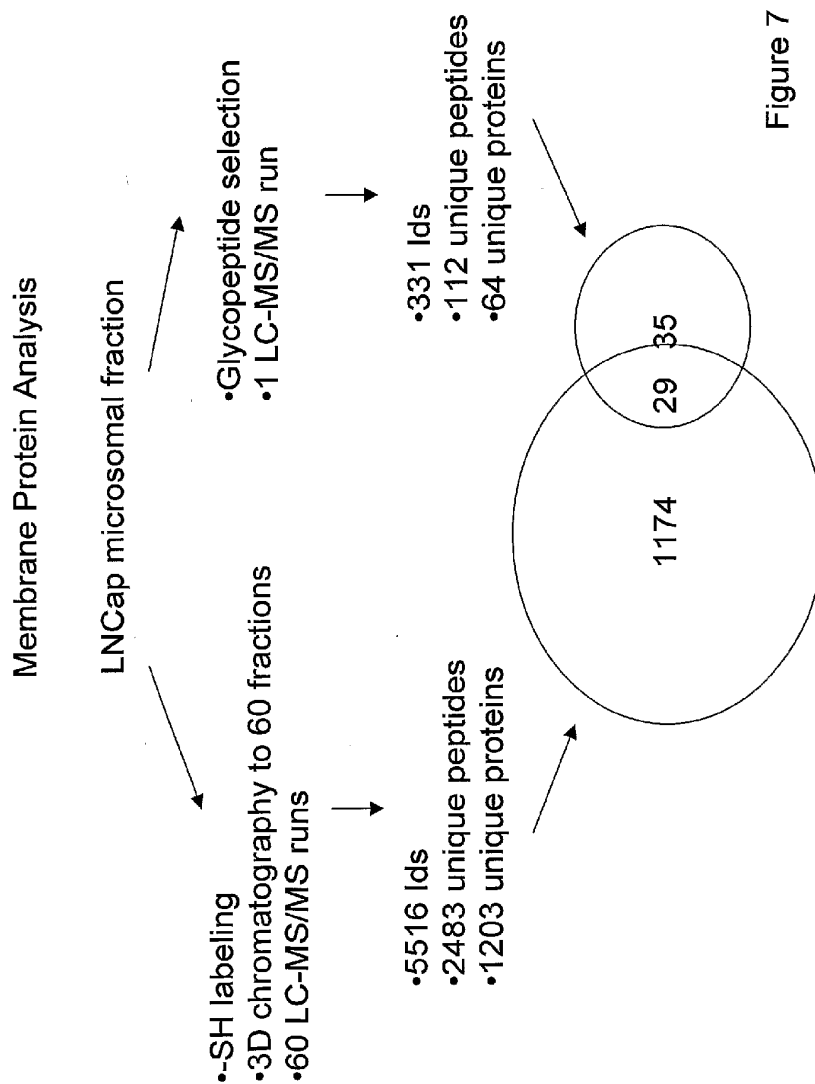


Figure 7

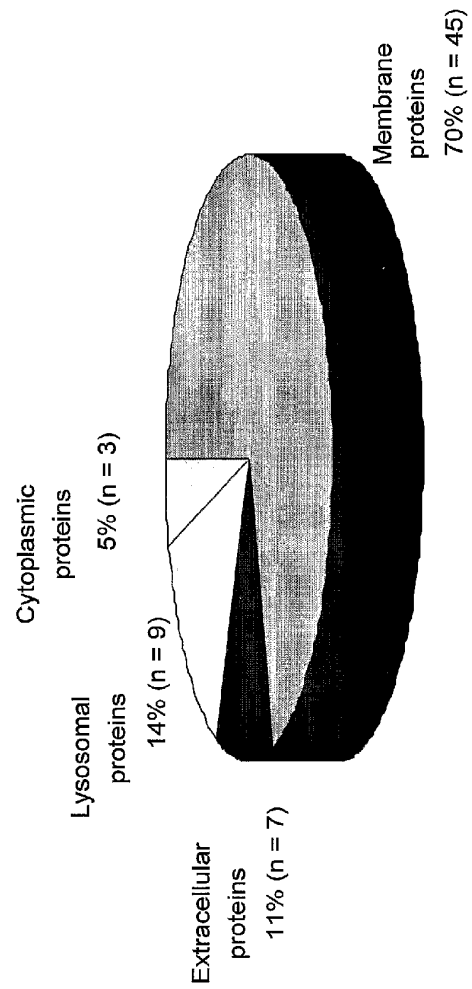


Figure 8

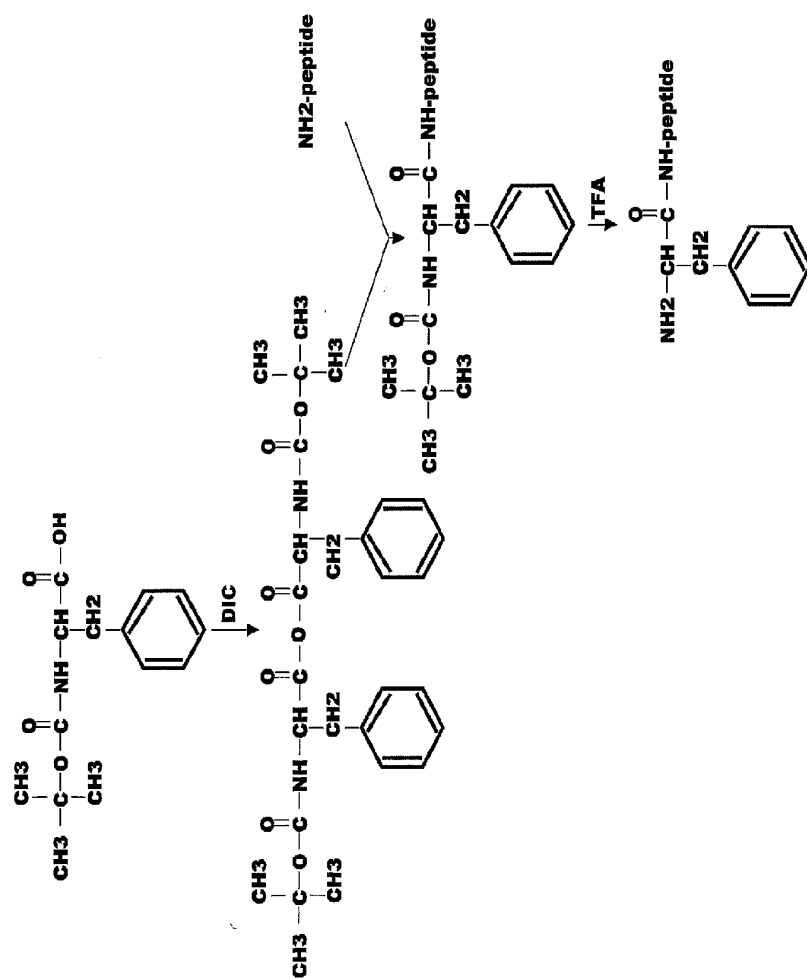
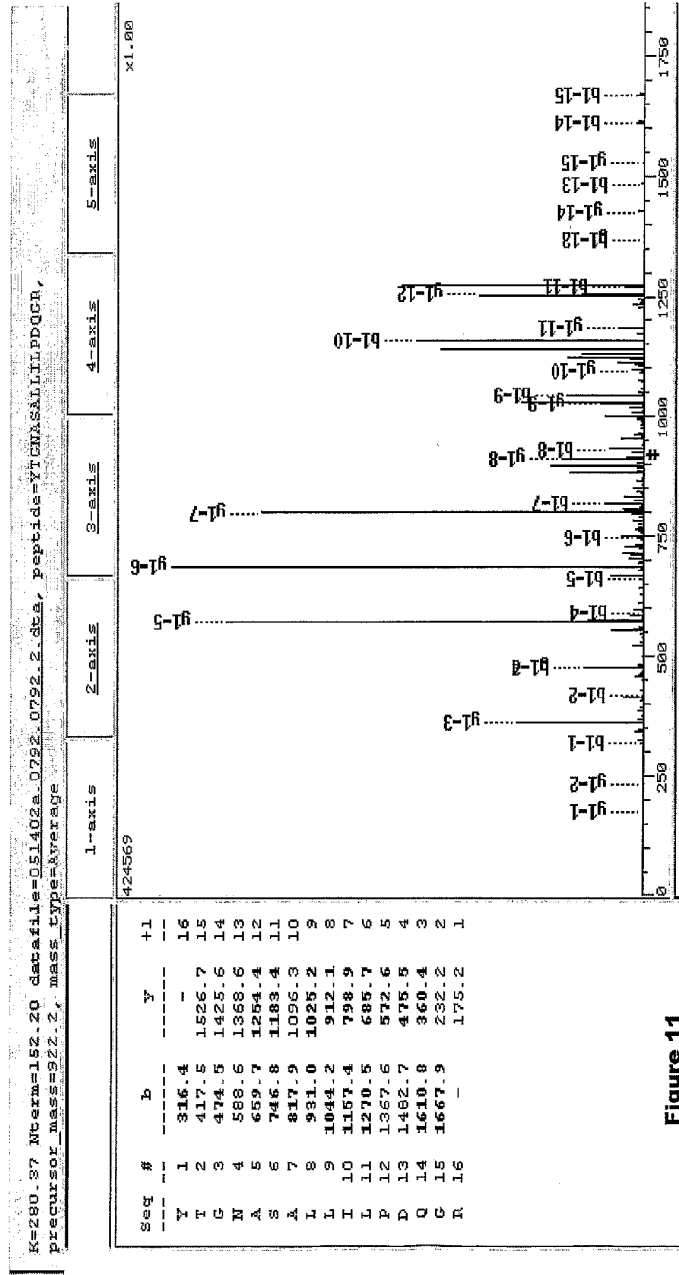


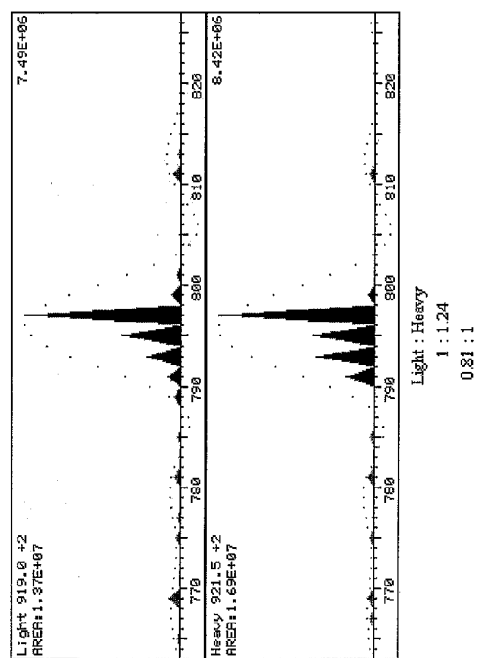
Figure 9

8 unique proteins; 6 single hits

1. 0.9914	283	/051402a.0830.0830.2	1796.1 (+0.6)	2.7967	0.163	1183.3	1	17/ 24	GPM-BC022986_1	+2	K.NSTLC*DLCTICP
2. 0.9938	324	/051402a.0924.0924.2	2392.8 (+0.1)	3.8459	0.347*	902.7	1	17/ 36	FR2-I56481		L.MMDYIGENATATF
3. 0.9998	279	/051402a.0822.0822.2	1876.1 (+0.9)	3.7204	0.378	1121.3	1	20/ 30	FR2-JH0424	+3	K.YTCASALFLLP
4. 0.9957	254	/051402a.0782.0782.2	2479.8 (+1.8)	3.1847	0.235	604.5	1	19/ 36	SU-ADPH_MOUSE	+2	K.DYDFZAGNNELTQ
5. 0.9950	194	/051402a.0700.0700.2	1908.1 (+1.9)	2.7744	0.259	1287.4	1	17/ 24	SU-CEAH_MOUSE		K.LTEFTHNSIMDY
6. 0.9992	246	/051402a.0722.0722.2	2159.5 (+0.1)	4.6248	0.402	1128.2	1	22/ 30	SU-CO1_MOUSE	+7	K.MLNDYVSNOTOF
7. 0.9987	260	/051402a.0722.0722.2	1842.1 (+1.3)	3.6408	0.254	1056.8	1	18/ 30	SU-CO1_MOUSE	+6	K.YTCASALFLLP
8. 0.9922	287	/051402a.0828.0828.2	2045.5 (+0.1)	4.7472	0.301	1998.5	1	23/ 30	SU-HPT_MUSSA	+1	K.VVLPNHSKVDIC
9. 0.9992	231	/051402a.0752.0752.2	1787.0 (+1.3)	4.3536	0.433	3162.3	1	23/ 24	SU-HPT_MUSSA	+1	K.NIFLHHSKIASAK
10. 0.9962	188	/051402a.0690.0690.2	2102.2 (+1.9)	4.6955	0.420	1748.6	1	22/ 30	SU-RMG_MOUSE	+1	K.HSIEHFNNITDI

Figure 10



**Figure 12**

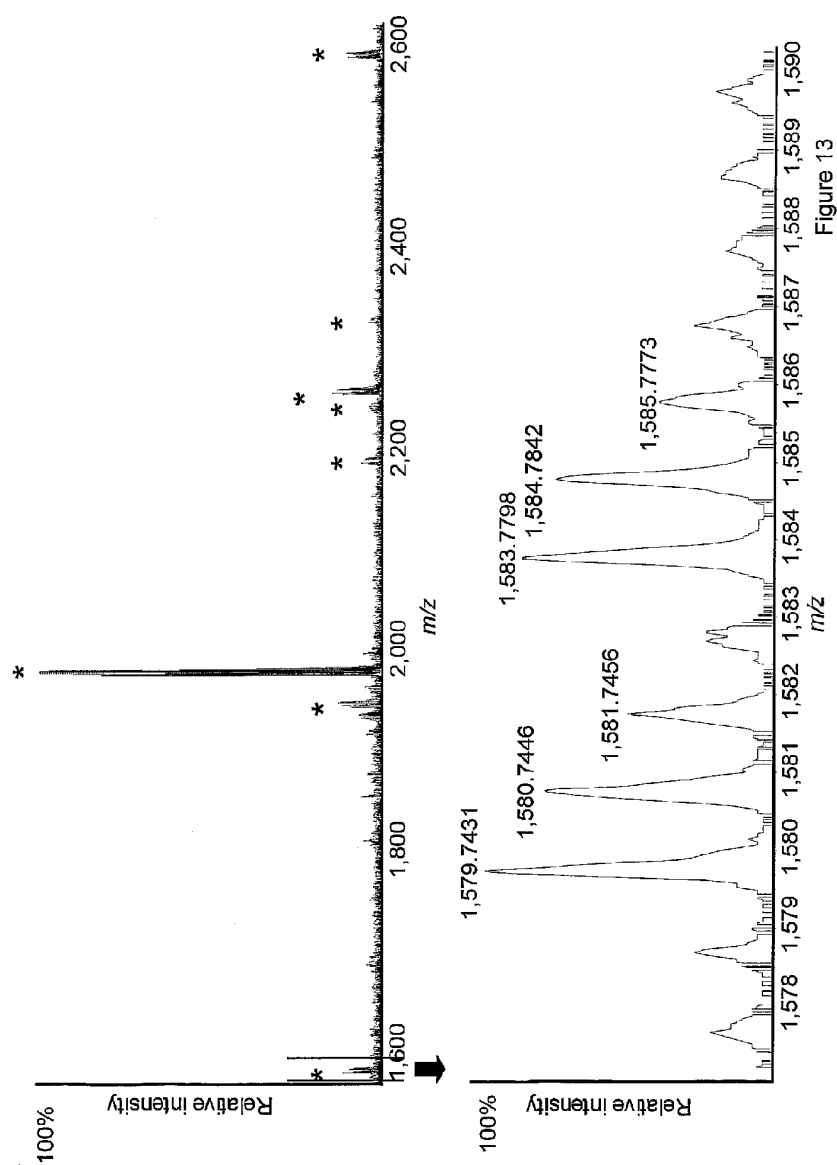


Figure 13

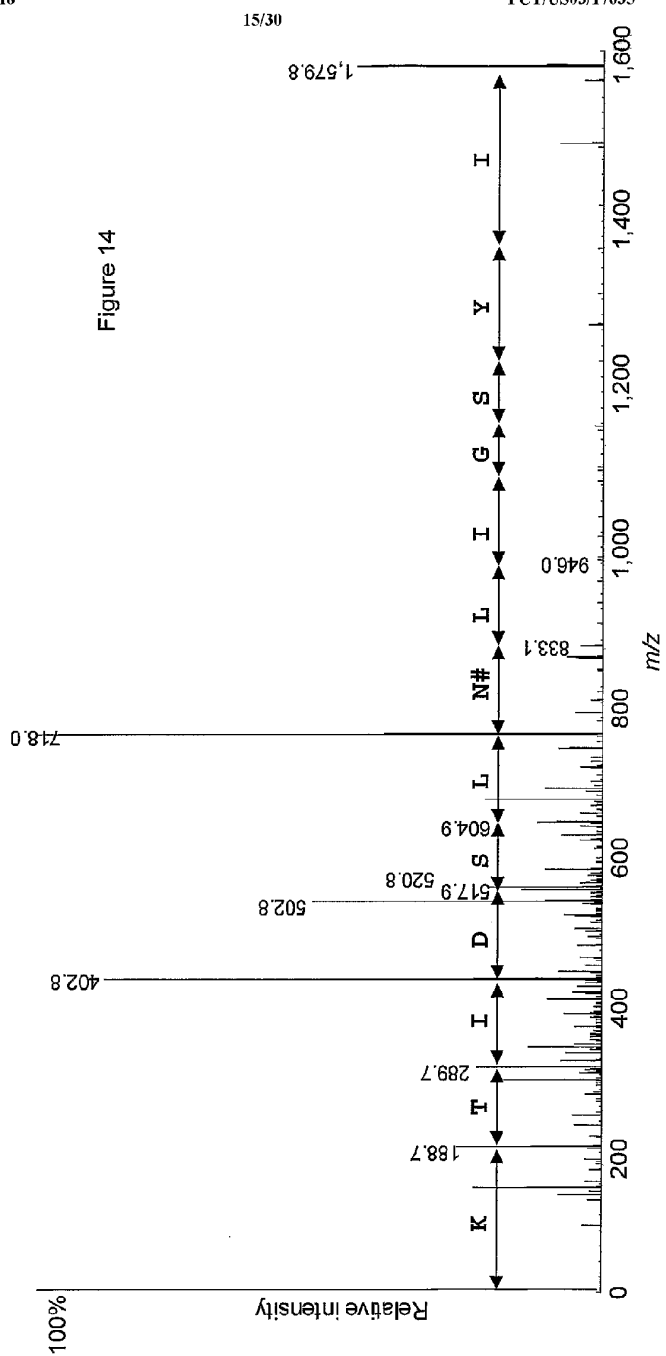
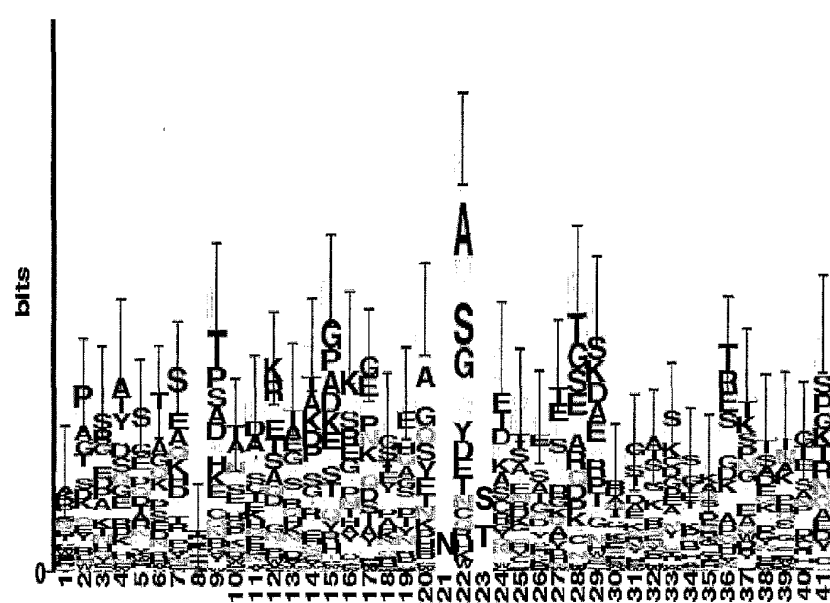
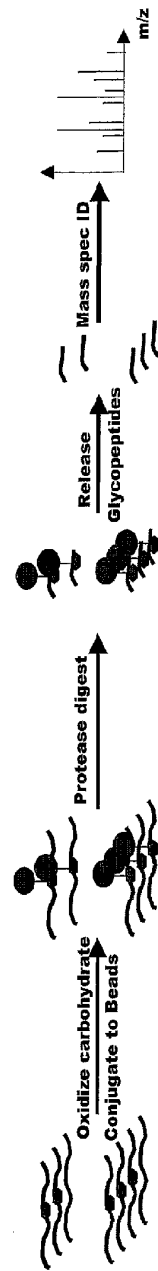


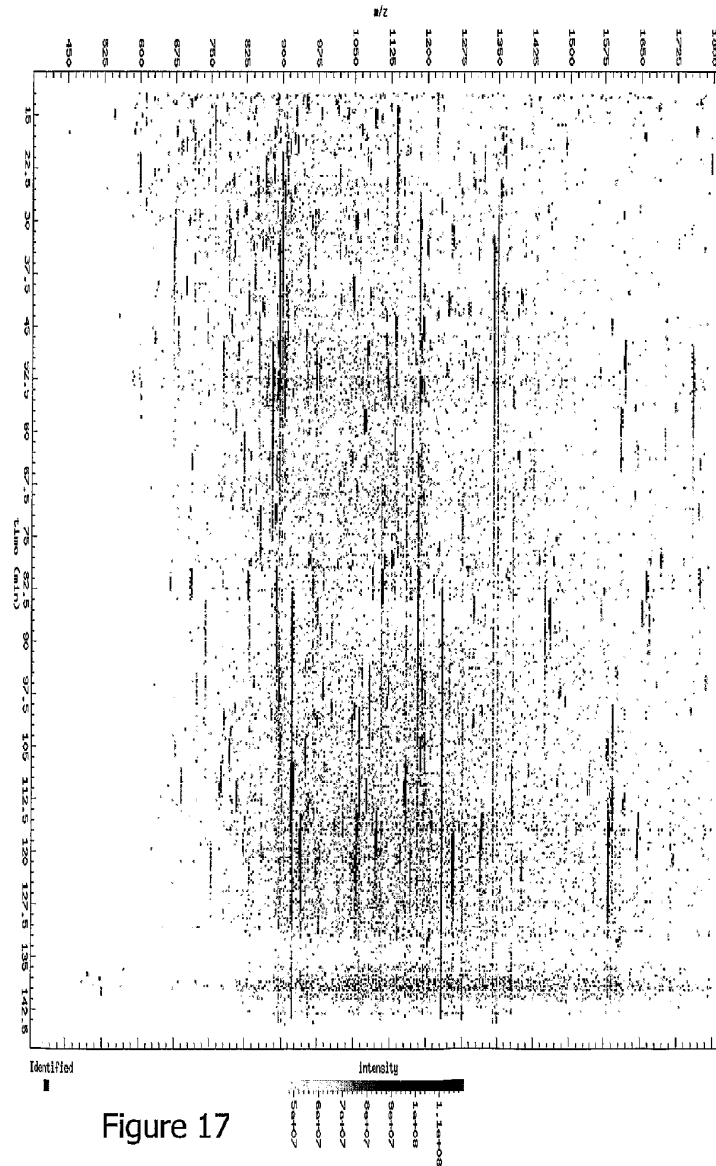
Figure 14





Glycoproteins	Normal	Cancer
Ig gamma-2 chain C region	+	+
Lumican precursor	+	+
Serum amyloid A-4 protein precursor	+	+
Alpha-1-antitrypsin precursor	+	+
Plasma protease C1 inhibitor precursor	+	+
Complement C3 precursor	+	+
Alpha-2-macroglobulin precursor	+	+
Haptoglobin-1 precursor; Haptoglobin-2 precursor	+	+
AMBP protein precursor	+	+
Alpha-1-antichymotrypsin precursor	+	+
Carboxypeptidase N 83 kDa chain (Fragment)	+	+
Alpha-1-acid glycoprotein 1 precursor; Alpha-1-acid glycoprotein 2 precursor	+	+
Complement C4 precursor	+	+
Apolipoprotein B-100 precursor	+	+
Kininogen precursor; Kininogen	+	+
Inter-alpha-trypsin inhibitor heavy chain H4 precursor	+	+
Complement C1q subcomponent; A chain precursor	+	+
Peptidoglycan recognition protein 1 precursor	+	+
Membrane copper amine oxidase	+	+
Prostatic acid phosphatase precursor	+	+
Microfibril-associated glycoprotein 4 precursor	+	+
Prostate specific antigen precursor	+	+
Zinc-alpha-2-glycoprotein precursor	+	+
Golgi membrane protein GP73	+	+
Collagen alpha 1(V) chain precursor	+	+
Lamin gamma-1 chain precursor	+	+
Acid ceramidase precursor	+	+

Figure 16



Methodology

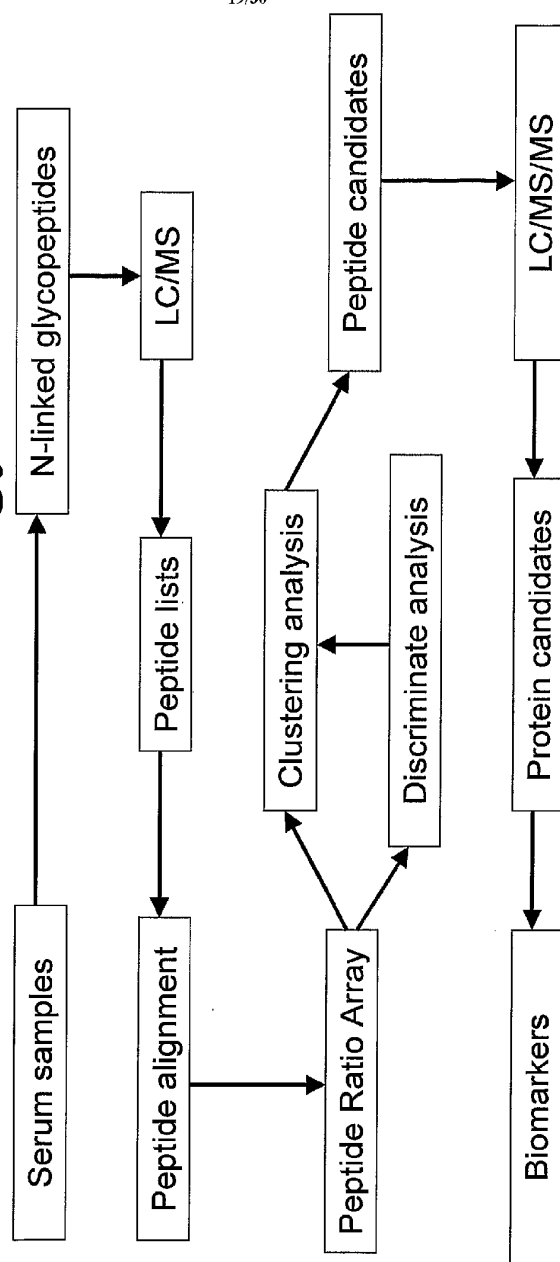


Figure 18

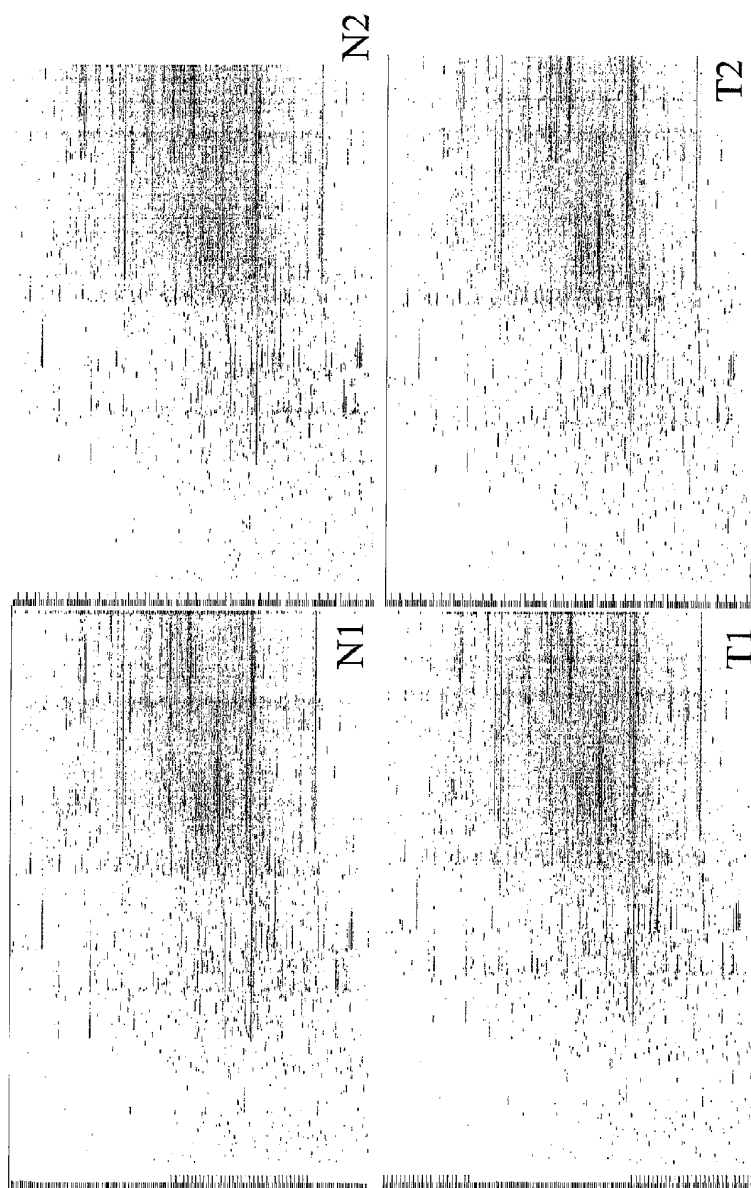


Figure 19

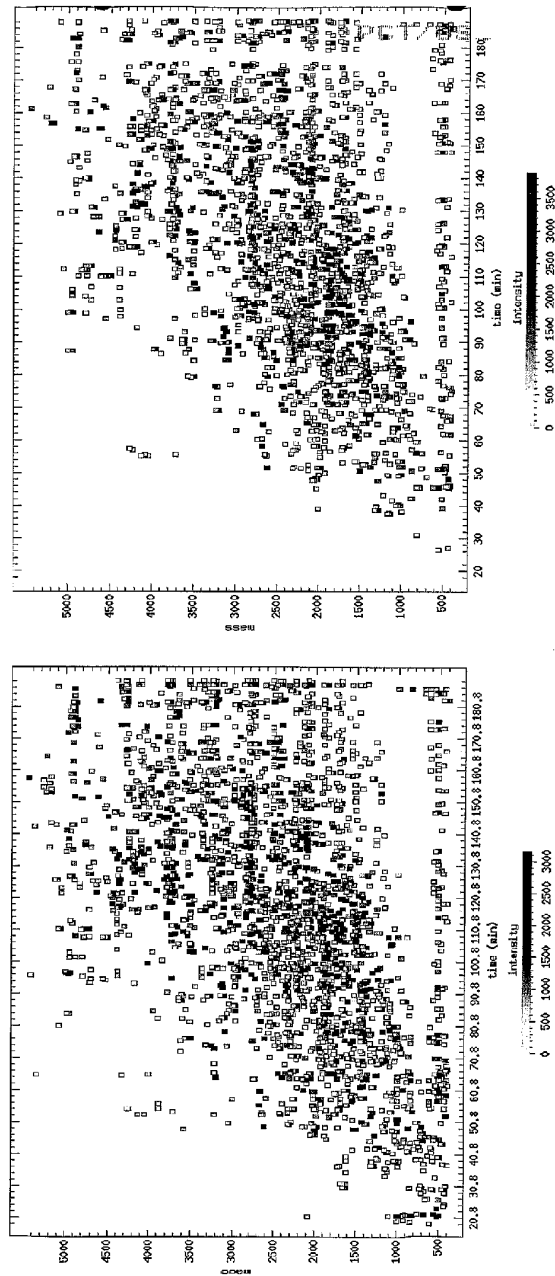


Figure 20

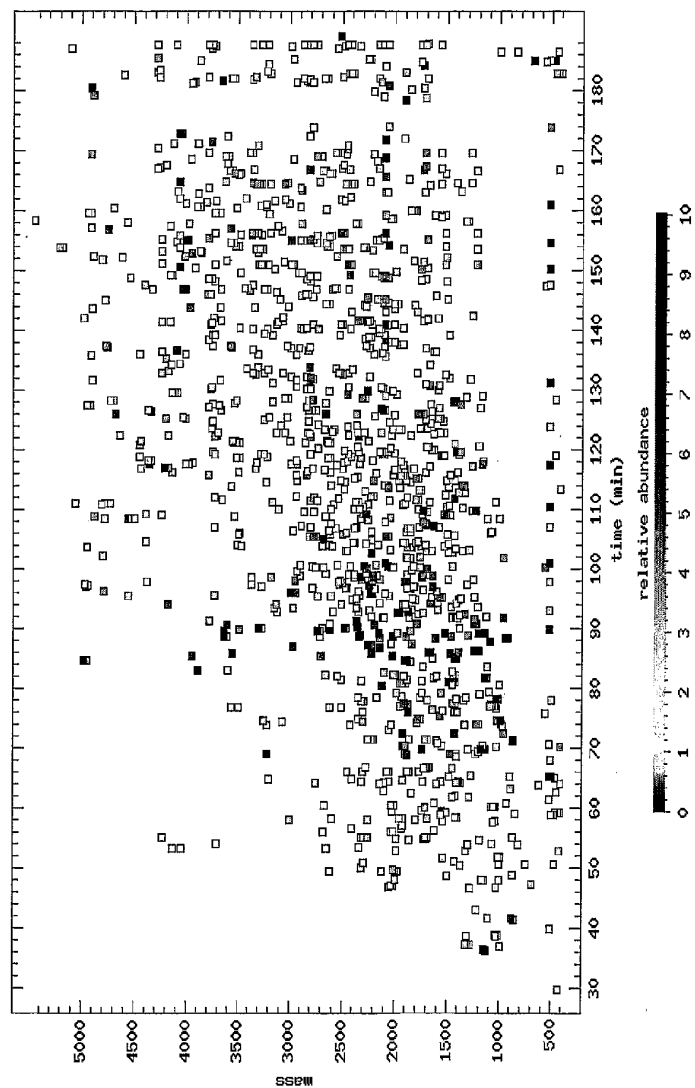


Figure 21

Figure 22

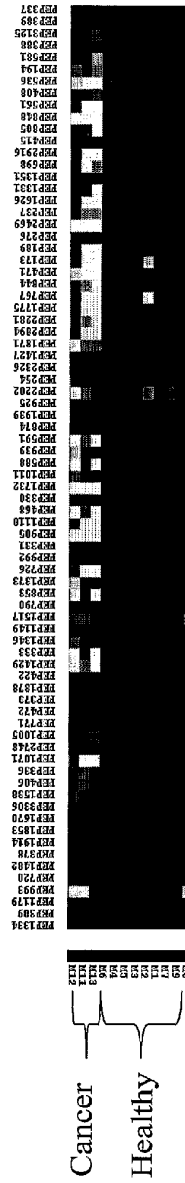


Figure 23



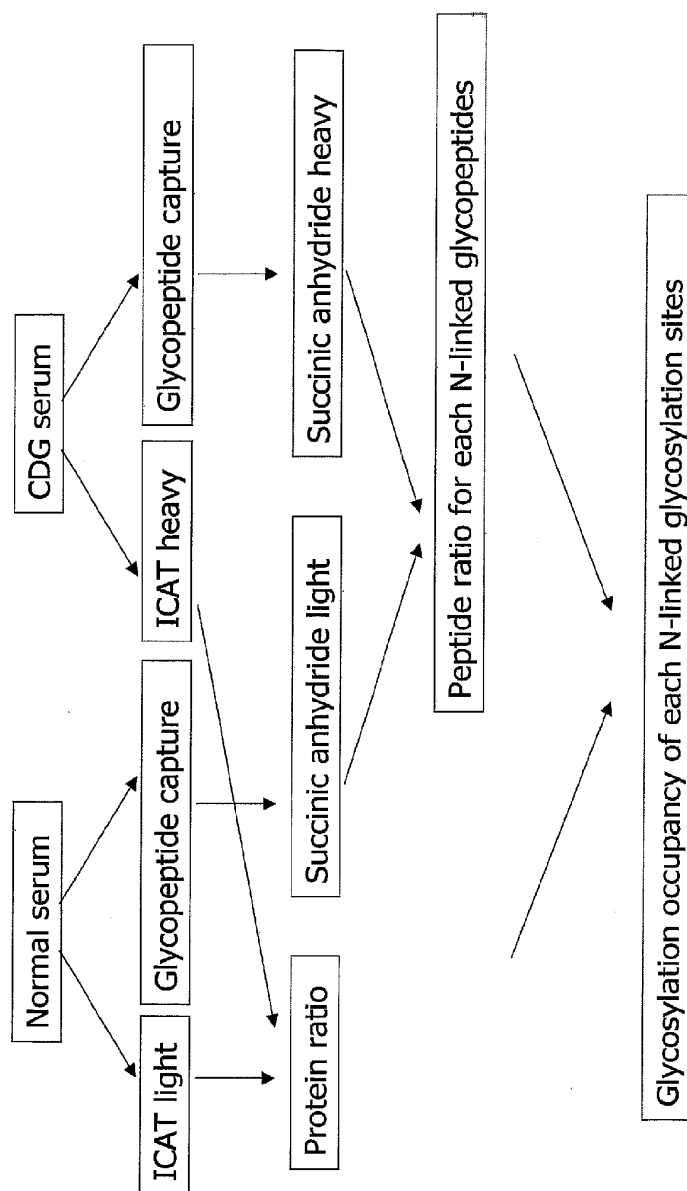


Figure 24

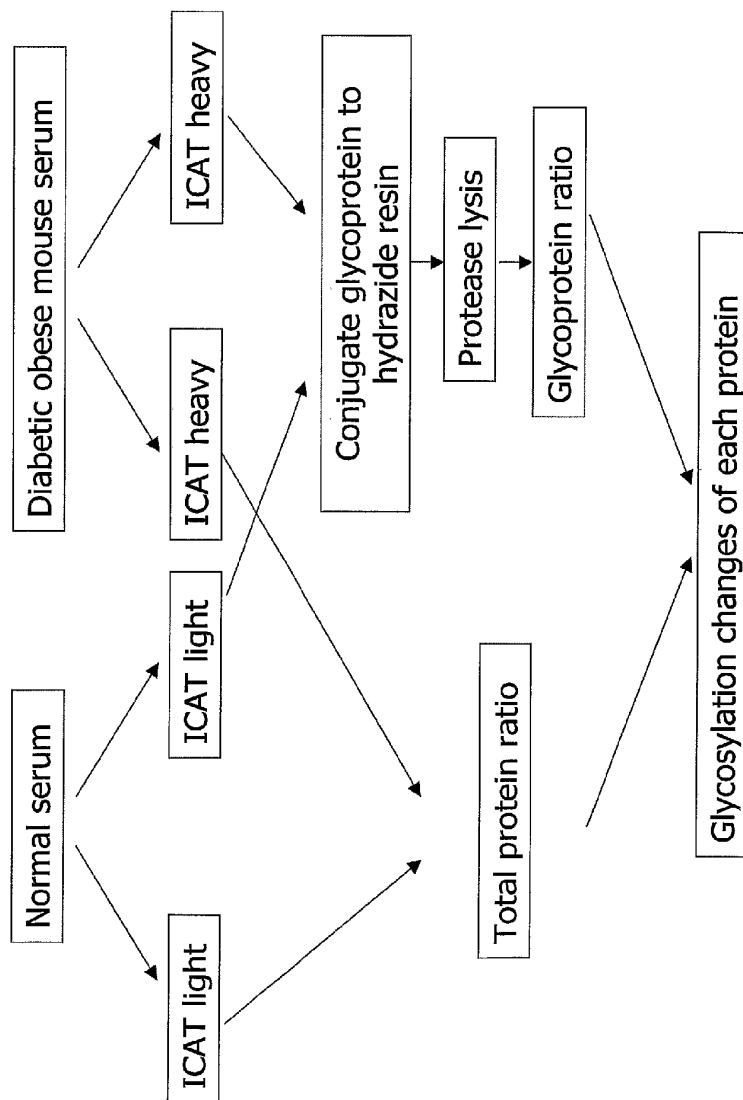


Figure 25

3.9298	0.272*	1497.6	1	<u>35/ 80</u>	<u>IP:IP100007221.1</u>	R.VVGVPTQGNRTAAT#IIPSEGR.H
4.4259	0.257*	1176.7	1	<u>21/ 40</u>	<u>IP:IP100007221.1</u>	R.VVGVPTQGNRTAAT#IIPSEGR.H
2.9818	0.133	914.0	1	<u>27/ 95</u>	<u>IP:IP100007222.1</u>	K.QVTSVSGIIPHSDDLAINDTFV*NDER.T
4.0143	0.382*	1060.5	1	<u>34/ 80</u>	<u>IP:IP100007240.1</u>	K.HGVISSTVDYENSGSV*EYR.C
4.9100	0.463*	1952.7	1	<u>28/ 40</u>	<u>IP:IP100007240.1</u>	K.HGVISSTVDYENSGSV*EYR.C
3.3810	0.115*	711.5	1	<u>31/ 84</u>	<u>IP:IP100010234.1</u>	K.CPSIDLPDPDUNGV*TEPHTRK.H
3.4283	0.303*	1361.0	1	<u>22/ 28</u>	<u>IP:IP100020091.1</u>	R.QNOCETNSYLNV*OR.N
3.9313	0.326*	2485.2	1	<u>33/ 68</u>	<u>IP:IP100022325.1</u>	R.AVNGIITSNLIDDDV*EIR.G
2.8455	0.224*	920.8	6	<u>17/ 34</u>	<u>IP:IP100022325.1</u>	R.AVNGIITSNGLIDDDV*VSLTR.G
4.1656	0.405*	1927.5	1	<u>24/ 38</u>	<u>IP:IP100024416.1</u>	K.TDMSLDDALONEMTV*FLMHR.G
4.9909	0.131*	2059.3	1	<u>22/ 72</u>	<u>IP:IP100022179.1</u>	K.SLTFNNVTYQDISRIY*YGAR.L
4.2321	0.433*	1682.5	1	<u>23/ 36</u>	<u>IP:IP100032179.1</u>	K.SLTFNNVTYQDISRIY*YGAR.L
5.3725	0.306*	1781.2	1	<u>38/ 80</u>	<u>IP:IP100032220.1</u>	R.VYTHDPHLV*THNGSTCROLAK.A
4.1820	0.337*	1796.7	1	<u>24/ 40</u>	<u>IP:IP100032220.1</u>	R.VYTHDPHLV*THNGSTCROLAK.A

Figure 26

2.7417	0.161*	520.4	1	16/ 24	IP1:IP100004608.1 f1	K.TKDRHQFMSER.V
3.3107	0.211*	577.3	1	17/ 20	IP1:IP100004609.1 f1	K.PAREQFMSER.V
3.2942	0.270*	728.1	1	16/ 20	IP1:IP100004638.1 f2	K.YKM#NSDLSER.C
2.0628	0.166	668.5	1	13/ 16	IP1:IP100004641.1	K.TPATAN#TK.S
3.3788	0.341*	545.3	1	17/ 42	IP1:IP100006662.1	R.ADCFWKQIRCHAFPPVNTPEAK.L
3.7782	0.403*	765.0	1	20/ 42	IP1:IP100006662.1	R.ADCFWKQIRCHAFPPVNTPEAK.L
4.4292	0.312*	978.6	1	18/ 36	IP1:IP100007222.1	R.VLSNNSDANLELINTWYAK.N
3.8572	1.000*	1054.7	2	19/ 36	IP1:IP100007222.1	R.VLSNNSDANLELINTWYAK.N
4.1419	0.335*	1051.3	1	20/ 36	IP1:IP100007222.1	R.VLSNNSDANLELINTWYAK.N
2.7875	0.139	195.2	90	26/124	IP1:IP100013441.1	R.AGEN#DFSTWYTPRESAOLWLGPAKIN#HOCK.P
1.1511	0.157*	132.4	156	7/ 10	IP1:IP100013712.1	R.ESTVNM#.E
2.7042	0.160*	1206.6	1	15/ 20	IP1:IP100018570.1 f2	L.KIN#KKN#AK.D
2.2302	0.170*	885.6	1	15/ 20	IP1:IP100018570.1 f2	L.FAHHS#HAK.D
2.6114	0.253*	267.9	3	15/ 42	IP1:IP100018570.1 f2	K.H-YSHN#PTGATLEH#QWHLFTYAK.N

Figure 27

2.1575	1.000*	539.9	21	21/ 68	IFI-IP100006142.1	M. NENHSEMYENOCRLQVTH+N. K
3.5797	0.345*	963.4	1	20/ 40	IFI-IP100007222.1	K. VCOLQISENHSLSKILYDMLK. H
3.2974	0.211*	659.1	4	15/ 36	IFI-IP100007222.1	R. VLSNHSDBMLLNTWYAK. N
6.0560	0.354*	1460.7	2	31/ 92	IFI-IP100019570.1	K. MYSEHNLTTGATLNNQVLLTAK. N
5.3020	0.401*	2024.6	1	25/ 45	IFI-IP100019570.1	K. MYSEHNLTTGATLNNQVLLTAK. N
5.1176	0.337*	1874.1	4	35/ 92	IFI-IP100019570.1	K. MYSEHNLTTGATLNNQVLLTAK. N
4.7364	0.473*	824.7	1	24/ 46	IFI-IP100019570.1	K. MYSEHNLTTGATLNNQVLLTAK. N
4.4349	0.310*	1071.3	1	32/ 26	IFI-IP100019570.1	K. MYSEHNLTTGATLNNQVLLTAK. N
2.7879	0.302*	1194.7	3	16/ 20	IFI-IP100019570.1	K. MYSEHNLTTGATLNNQVLLTAK. N
3.3246	0.344*	1119.0	1	12/ 26	IFI-IP100021002.1	K. CAPISNFSH+YVDEK. T
5.5646	1.000*	3308.3	3	38/ 84	IFI-IP100022431.1	K. ALAALAFACNNH+SNFQERISD. A
5.5250	0.346*	2636.1	4	40/ 92	IFI-IP100022453.1	K. ILRQOCHLIGSNH+PTDCSNCICFR. S
4.4199	0.320*	1879.6	2	29/ 80	IFI-IP100022453.1	R. QOHLFGSNH+PTDCSNCICFR. S
3.2847	0.352*	937.1	2	14/ 26	IFI-IP100022453.1	G. SNHPTDCSNCICFR. S
2.6890	0.217*	512.2	3	17/ 40	IFI-IP100022453.1	R. QOHLFGSNH+PTDCSNCICFR. S
4.2571	0.308*	1747.0	1	22/ 38	IFI-IP100023520.1	K. M+LNTSSLLLEONHNFNVSR. L
3.6581	0.347*	1037.1	2	19/ 26	IFI-IP100023520.1	R. LANHFTQCDQYTLR. V
2.0746	0.140*	191.0	192	11/ 52	IFI-IP100024501.1	K. AKENHDEHRCCTTIVTVYENHSEKASDML. I
3.6310	0.164*	700.6	1	28/ 64	IFI-IP100027265.1	R. FSDCLSEHSHSTQPFVERK. Y
5.9257	0.237*	1155.6	2	42/120	IFI-IP100032180.1	K. ADTHDILLEGANFNH+PTFPAQHCHQRELL. R
5.2182	0.413*	2539.8	1	46/124	IFI-IP100032180.1	R. ADTHDILLEGANFNH+PTFPAQHCHQRELL. R
4.7681	0.422*	496.6	2	20/ 62	IFI-IP100032180.1	R. ADTHDILLEGANFNH+PTFPAQHCHQRELL. R
5.0103	0.453*	1250.1	3	31/116	IFI-IP100032180.1	R. QLAHNSNTNFFSIVSIATAFAN+ISICTE. A

Figure 28

[illegible]

Figure 29