nature biotechnology

© 2003 Nature Publishing Group http://www.nature.com/naturebiotechnology

# Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry

Hui Zhang, Xiao-jun Li, Daniel B Martin & Ruedi Aebersold

**Quantitative proteome profiling using stable isotope protein tagging and automated tandem mass spectrometry (MS/MS) is an emerging technology with great potential for the functional analysis of biological systems and for the detection of clinical diagnostic or prognostic marker proteins. Owing to the enormous complexity of proteomes, their comprehensive analysis is an as-yet-unresolved technical challenge. However, biologically or clinically important information can be obtained if specific, information-rich protein classes, or sub-proteomes, are isolated and analyzed. Glycosylation is the most common post-translational modification. Here we describe a method for the selective isolation, identification and quantification of peptides that contain N-linked carbohydrates. It 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 MS/MS. We applied the approach to the analysis of plasma membrane proteins and proteins contained in human blood serum.**

Quantitative proteomics, defined as the comparison of relative protein differences 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 of diagnostic or prognostic disease markers and 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 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 MS/MS[1–3]. To date, neither method has succeeded in determining the complete proteome of any species owing to the 'top down' mode of operation of each method in which the most abundant proteins are preferentially or exclusively analyzed.

Accepting that global proteome analyses will remain difficult, several studies have adopted a 'divide and conquer' strategy to handle the 'top down' problem by comprehensively analyzing specific subsets of the proteome that are selectively isolated. Such studies include the analysis of functional multiprotein complexes such as the ribosome[4], spliceosome[5,6], and nuclear pore complex[7], or organelles, such as mitochondria[8], peroxisomes[9], microsomes[10] and nuclei[11]. Alternatively, proteins that contain common distinguishing structural features, such as phosphate ester groups[12–14] or cysteine residues[3,15], or that have the ability to specifically bind to certain compounds[16,17] have been selectively enriched before MS analysis. These strategies have in common that they focus on the in-depth (ideally complete) analysis of sub-proteomes of rich biological context, thus minimizing the repeated analyses of abundantly expressed proteins.

Protein glycosylation has long been recognized as a very common post-translational modification. Typically, carbohydrates are linked to serine or threonine residues (O-linked glycosylation) or to asparagine residues (N-linked glycosylation)[18]. N-linked glycosylation sites generally fall into the N-X-S/T sequence motif in which X denotes any amino acid except proline[19]. Protein glycosylation, and in particular N-linked glycosylation, is prevalent in proteins destined for extracellular environments[20]. These include proteins on the extracellular side of the plasma membrane, secreted proteins and proteins contained in body fluids (such as blood serum, cerebrospinal fluid, urine, breast milk, saliva, lung lavage fluid or pancreatic juice). These also happen to be the proteins in the human body that are most easily accessible for diagnostic and therapeutic purposes. It is therefore no surprise that many clinical biomarkers and therapeutic targets are glycoproteins. These include Her2/neu in breast cancer, β human chorionic gonadotropin and α-fetoprotein in germ cell tumors, prostate-specific antigen in prostate cancer and CA125 in ovarian cancer. The Her2/neu receptor is also the target for a successful immunotherapy of breast cancer using the humanized monoclonal antibody Herceptin[21]. 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[22–24].

Institute for Systems Biology, 1441 N 34th Street, Seattle, Washington 98103-8904, USA. Correspondence should be addressed to R.A. (raebersold@systemsbiology.org).

© 2003 Nature Publishing Group http://www.nature.com/naturebiotechnology

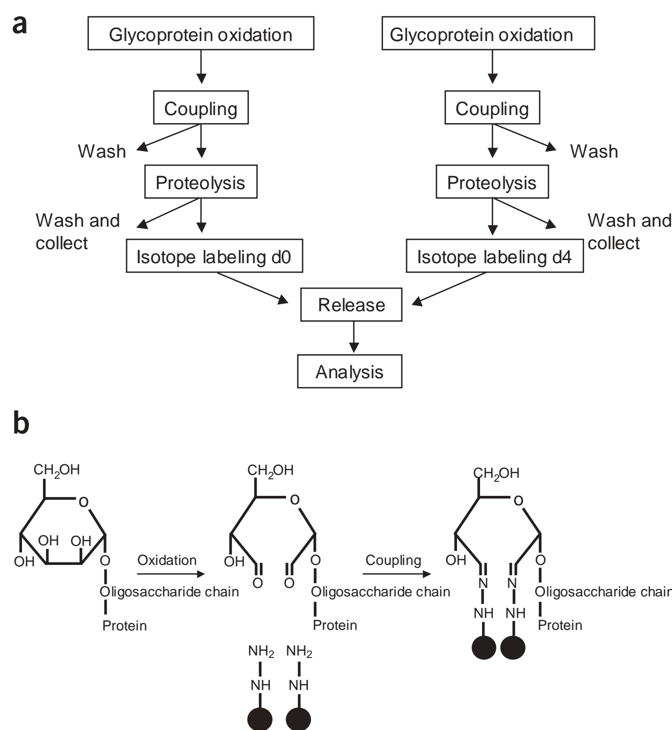


**Figure 1** Schematic diagram of quantitative analysis of N-linked glycopeptides. (**a**) Strategy for quantitative analysis of glycopeptides. Proteins from two biological samples are oxidized and coupled to hydrazide resin. Nonglycosylated peptides are removed by proteolysis and extensive washes. The nonglycosylated peptides are optionally collected and analyzed. The N-terminus of glycopeptides are isotope labeled by succinic anhydride carrying either d0 or d4. The beads are then combined and the isotopically tagged peptides are released by PNGase F and analyzed by MS. (**b**) Oxidation of a carbohydrate to an aldehyde followed by covalent coupling to hydrazide resin.

Therefore, it can be expected that a method for the systematic and quantitative analysis of glycoproteins would be very useful for the detection of new potential diagnostic markers and therapeutic targets.

Here we describe a method for quantitative glycoprotein profiling. It 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 PNGase F. The recovered peptides are then identified and quantified by MS/MS. The method was applied to the analysis of cell surface and serum proteins.

## RESULTS

### Principle of the method

The method consists of six steps (**Fig. 1a**). (i) Glycoprotein oxidation: periodate oxidation converts the *cis*-diol groups of carbohydrates to aldehydes (**Fig. 1b**). (ii) Coupling: the aldehydes react with hydrazide groups immobilized on a solid support to form covalent hydrazone bonds (**Fig. 1b**). Nonglycosylated proteins are removed. (iii) Proteolysis: the immobilized glycoproteins are proteolyzed on the solid support. The nonglycosylated peptides are removed by washing and can be collected, if desired, for further analysis, whereas the glycosylated peptides remain on the solid support. (iv) Isotope labeling: the α-amino groups of the immobilized 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[25]. (v) Release: formerly N-linked glycopeptides are released from the solid-phase by PNGase F treatment. (vi) Analysis: the isolated peptides are identified and quantified using microcapillary high-performance liquid chromatography electrospray ionization (μLC-ESI) MS/MS or μLC separation followed by matrix-assisted laser desorption/ionization (MALDI) MS/MS. The data are analyzed by a suite of software tools[10,26,27].

To illustrate the new method, two mixtures (A and B) that contained the three glycoproteins, α-1-antichymotrypsin, α-1-antitrypsin and α-2-hs-glycoprotein, at different concentrations were treated as described above. Formerly N-glycosylated peptides were analyzed by μLC-ESI-MS/MS and identified. **Table 1** shows the identified sequences 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[28–30], whereas N# in the sequence FN#LTETSEAEIHQSFQH represents a glycosylation site in α-1-antichymotrypsin that has not 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.

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[31]. The aldehydes, in turn, covalently couple to amine- or hydrazide-containing molecules[32]. 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[33]. 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

**Table 1 Quantitative analysis of glycoproteins in glycoprotein mixture**

| Protein name | Sequences of identified peptides[a] | Glycosylation sites | Observed peptide ratio (A/B)[b] | Protein ratio (A/B)[c] | Expected protein ratio (A/B) |
|---|---|---|---|---|---|
| α-1-antichymotrypsin | K.F**N#LT**ETSEAEIHQSFQH.L | Novel | 0.69; 0.91 | 1.09 ± 0.39 | 1.00 |
| | F.LSLGAH**N#TT**LTEILK.G | Known[30] | 1.63; 1.35 | | |
| | L.SISTALAFLSLGAH**N#TT**LTEILK.G | Known[30] | 0.88 | | |
| α-1-antitrypsin | R.QLAHQS**N#ST**NIFF.S | Known[29] | 6.47; 4.06 | 5.27 ± 1.70 | 5.00 |
| α-2-hs-glycoprotein | K.AALAAFNAQN**N#GS**NFQLEEISR.A | Known[28] | 0.34; 0.51 | 0.42 ± 0.12 | 0.33 |

[a] Sequences of identified peptides from the control glycoprotein mixture are shown in single letter code. N# indicates the asparagine with consensus N-linked glycosylation motif. [b]The observed peptide ratio shows the measured d0/d4 ratio of each identified peptide. [c]The protein ratio show the average ratio of the peptides from each protein.

periodate concentrations and reaction conditions. The conditions we used here (15 mM sodium periodate, room temperature for 1 h) 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[34]. PNGase F will not, however, release N-linked oligosaccharides containing core fucosylation. There is no enzyme comparable to PNGase F for removing intact O-linked sugars. To successfully release O-linked oligosaccharides, it is necessary to sequentially remove monosaccharides by using a panel of exoglycosidases until only the Galβ1,3GalNAc core remains attached to the serine or threonine residue. The core can then be released by O-glycosidase. Because not all O-linked oligosaccharides contain this core structure, a chemical method, such as β-elimination, may be more generally useful and effective for the release of the formerly O-linked glycosylated peptides.

## Qualitative analysis of glycoproteins in human serum

To assess the potential of the glycopeptide capture method for serum protein profiling, we first determined the specificity and efficiency of conjugation. Human serum proteins were coupled to the hydrazide beads. Identical aliquots (1 μl) were removed from the sample before ('– beads') or after ('+ beads') capture of glycoproteins to hydrazide resin. The samples were separated by 9% (wt/vol) sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and stained with silver (total protein stain) or with a glycoprotein-staining reagent (**Fig. 2a**). The following is apparent from this experiment. 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, '+ beads' lane) and was not stained with the glycoprotein stain (right, '+/– beads' lane). Because the present method is compatible with the immobilization of denatured proteins (data not shown), it reduces the possibility that the affinity depletion of albumin also removes albumin-associated proteins.

We then determined whether the reduced peptide complexity achieved by the glycopeptide capture method allowed us to identify a larger number of serum proteins compared to conventionally prepared control samples if comparable μLC-MS/MS protocols were applied. A serum sample was processed and the formerly N-linked glycosylated peptides were analyzed by μLC-ESI-MS/MS (**Fig. 1**). The resulting collision-induced dissociation (CID) spectra were searched using Sequest against the human National Center for Biotechnology Information (NCBI) sequence database (**Fig. 2b**, middle). Control samples were generated by selectively isolating cysteine-containing peptides using the isotope-coded affinity tagged (ICAT) reagent method[3]. These were analyzed either using the same μLC-ESI-MS/MS method used for the analysis of the peptides isolated by the glycopeptide capture method (**Fig. 2b**, right) or through extensive, three-dimensional (cation exchange–biotin affinity–reverse phase liquid chromatography (RP-LC)) separation in which the peptide mixture was fractionated into 17 cation exchange fractions that were sequentially analyzed by μLC-ESI-MS/MS[10] (**Fig. 2b**, left). Using a single μLC-ESI-MS/MS run requiring approximately 2 h of mass spectrometer time, we identified 145 unique peptides mapping to 57 unique serum proteins with the glycopeptide capture method (2.5 peptides per protein); when comparable MS methods were applied for the analysis of cysteine-tagged peptides, we identified 72 unique peptides mapping to 23 unique proteins, of which 15 were also identified via the glycopeptide capture method (**Fig. 2b**, right). Using the extensive peptide separation protocol for the analysis of cysteine-tagged peptides that required approximately 34 h of mass spectrometer time, we identified 356 unique peptides mapping to 97 serum proteins. Of the 57 proteins isolated by the glycopeptide capture method and identified by single-dimensional LC-MS/MS, 23 proteins were not seen by extensive μLC-ESI-MS/MS-based protocol of cysteine-tagged peptides (**Fig. 2b**, left). These data demonstrate the increased efficiency of serum analysis provided by the glycopeptide capture method.

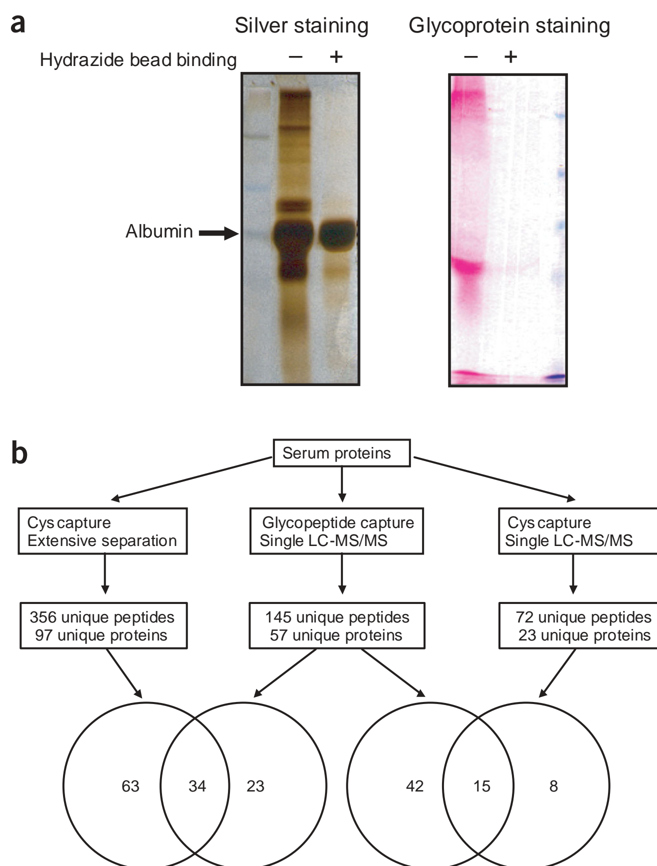


**Figure 2** Isolation of glycoproteins from serum. (**a**) Specific capture of glycoproteins by hydrazide resin. Total protein staining or glycoprotein staining of serum before (–) and after (+) capture of glycoproteins by hydrazide resin. Proteins were separated by SDS-PAGE and stained with silver (left) or Gel Code Blue glycoprotein staining reagent (right). (**b**) Comparison of total number of proteins or peptides identified from serum samples isolated by either cysteine-reactive tag or glycopeptide capture method.

Four major conclusions can be drawn that are relevant for assessing the potential of each method for serum protein profiling. First, the cysteine-tagging method is substantially limited by the presence of a number of high-abundance proteins that include the five major plasma proteins, representing more than 80% of the total plasma

© 2003 Nature Publishing Group http://www.nature.com/naturebiotechnology

**Table 2 Quantitative analysis of glycoproteins from two identical serum samples**

| Gene name [a] | Protein names | Sequences of identified peptides[b] | Observed ratio (Mean ± s.d.) | Expected ratio | % Error |
|---|---|---|---|---|---|
| *PGLYRP* (GP:AF384856_1) | Peptidoglycan recognition protein L | R.GFGVAIVG**N#YT**AALPTEAALR.T | 0.95 ± 0.02 | 1 | 5 |
| *A2M* (GP:M36501_1) | α-2-macroglobulin | Y.VLDYL**N#ET**QQLTPEIK.S | 0.93 ± 0.03 | 1 | 7 |
| *ORM1* (SW:*A1AG_HUMAN*) | α-1-acid glycoprotein 1 | N.LVPVPIT**N#AT**LDQITGK.W | 1.05 ± 0.11 | 1 | 5 |
| *SERPINA1* (SW:*A1AT_HUMAN*) | α-1-antitrypsin | K.YLG**N#AT**AIFFLPDEGK.L | 1.10 ± 0.11 | 1 | 10 |
| *SERPINA1* (SW:*A1AT_HUMAN*) | α-1-antitrypsin | R.QLAHQS**N#ST**NIFF.S | 1.00 ± 0.01 | 1 | 0 |
| *SERPINA3* (SW:*AACT_HUMAN*) | α-1-antichymotrypsin | K.YTG**N#AS**ALFILPDQDK.M | 1.05 ± 0.03 | 1 | 5 |
| *C3* (SW:*CO3_HUMAN*) | Complement c3 | N.HMG**N#VT**FTIPANR.E | 0.91 ± 0.02 | 1 | 9 |
| *C4B* (SW:*CO4 HUMAN*) | Complement c4 | R.FSDGLES**N#SS**TQFEVK.K | 0.93 ± 0.07 | 1 | 7 |
| *HP* (SW:*HPT1_HUMAN*) | Haptoglobin-1 | K.VVLHP**N#YS**QVDIGLIK.L | 1.04 ± 0.03 | 1 | 4 |
| *HP* (SW:*HPT1_HUMAN*) | Haptoglobin-1 | K.NLFL**N#HS**E**N#AT**AK.D | 1.29 ± 0.33 | 1 | 29 |
| *SERPING1* (SW:*IC1_HUMAN*) | Plasma protease c1 inhibitor | R.VLS**N#NS**DANLELINTWVAK.N | 0.90 ± 0.03 | 1 | 10 |
| *IGHEP1* (SW:*ITH1_HUMAN*) | Inter-α-trypsin inhibitor heavy chain h1 | H.FFAPQ**N#LT**NMNK.N | 0.96 ± 0.01 | 1 | 4 |
| *IGHEP2* (SW:*ITH2_HUMAN*) | Inter-α-trypsin inhibitor heavy chain h2 | K.GAFIS**N#FS**MTVDGK.T | 1.08 ± 0.12 | 1 | 8 |
| *ITIH4* (SW:*ITH4_HUMAN*) | Inter-α-trypsin inhibitor heavy chain h4 | N.QLVDALTTWQ**N#K.T** | 1.01 ± 0.13 | 1 | 1 |
| *SERPINA4* (SW:*KAIN_HUMAN*) | Kallistatin | K.FL**N#DT**MAVYEAK.L | 1.24 ± 0.30 | 1 | 24 |
| *KLKB1* (SW:*KAL_HUMAN*) | Plasma kallikrein | R.IYSGIL**N#LS**DITK.D | 1.06 ± 0.08 | 1 | 6 |
| *KNG* (SW:*KNG_HUMAN*) | Kininogen | K.LNAEN**N#AT**FYFK.I | 0.94 ± 0.10 | 1 | 6 |

[a]Gene name is from the HUGO Gene Nomenclature Committee database (http://www.gene.ucl.ac.uk/nomenclature/) (and the human NCBI protein database (http://www.ncbi.nlm.nih.gov/)). [b]The consensus motif for N-linked glycosylation is highlighted and the asparagine residues to which carbohydrate is linked are N#.

© 2003 Nature Publishing Group http://www.nature.com/naturebiotechnology

protein mass (albumin, α-1-antitrypsin, β-2-macroglobulin, transferrin and γ-globulins). When the 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 albumin. Second, proteins that were not identified by the cysteine-tagging method were readily identified after glycopeptide capture (**Fig. 2b**). This attests to the potential of the glycopeptide capture method to achieve deeper serum protein coverage within a substantially reduced data acquisition time. Third, the glycopeptide capture method reduced the 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. 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.

However, neither this nor any other method is currently capable of profiling all the proteins present in serum. Although the use of more extensive separation protocols for the formerly N-glycosylated peptides will increase the depth of serum protein coverage, tryptic peptides that are too short or too long to fall within the detection range of the mass spectrometer used will not be identified. This limitation may be overcome, at least in part, by the use of proteases with cleavage specificities different from that of trypsin.

## Quantitative analysis of glycoproteins

We also determined whether the glycopeptide selection method could be used to detect 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 N-terminally labeled with either light (d0) or heavy (d4) forms of succinic anhydride after C-terminal lysine residues were converted to homoarginines. The lysine-to-homoarginine conversion facilitated detection by MALDI quadrupole time-of-flight (MALDI-QqTOF) MS[25] 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 portion 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[35,36]. The experiment was repeated and samples were analyzed by ESI-QqTOF MS with comparable results to those obtained by MALDI-QqTOF MS. **Table 2** lists the identified peptides, 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% and 29% with a mean of 8%. This indicates that the glycopeptide capture method allows reasonable quantification if combined with stable isotope tagging.

The quantification and identification process is further illustrated for a single peptide pair (**Fig. 3**). 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 *, **Fig. 3a**). An expansion of the mass range between $m/z$ = 1,577 and $m/z$ = 1,590 resolved the natural isotopic distribution of a peptide pair with monoisotopic peaks at 1,579.74 and 1,583.78, in which the signals had a quantitative ratio of 1.11. The precursor ion with $m/z$ = 1,579.74 was further analyzed by MS/MS and sequence database searching of the resulting spectrum, and it was identified as peptide sequence IYSGILN#LSDITK from human plasma kallikrein, a serum protease (**Fig. 3b**). Cleavage of oligosaccharides from glycoproteins catalyzed by PNGase F deaminates the linker asparagine to aspartic acid, causing a mass shift of 1 mass unit[37]. The series of y ions from this peptide confirmed the match, and indicated that the 1–mass unit difference between asparagine and aspartic acid can be easily detected by MALDI QqTOF MS, thus confirming the precise glycosylation site within the peptide as N7.

© 2003 Nature Publishing Group http://www.nature.com/naturebiotechnology

We carried out guanidination of lysine residues and isotope-tag addition while the glycopeptides were still bound to the solid support. Solid-phase attachment facilitates the reactions and the removal of excess reagents and is therefore compatible with a range of different tagging chemistries. The results show that glycopeptide capture, stable isotope tagging and MS/MS constitute a method for quantitative profiling of the glycoproteins in serum.

**Glycopeptide analysis of cell surface proteins**

To assess the potential of the glycopeptide capture method for the analysis of cell surface proteins, we used a crude membrane fraction from the LNCaP prostate cancer epithelial cell line[38] to select and identify peptides containing N-linked glycosylation sites. The released peptides isolated from 60 µg of a crude membrane fraction were analyzed by single-dimensional µLC-MS/MS and the data were processed as described above.

The experiment identified 104 unique peptides mapping to 64 unique proteins (1.6 peptides per protein) (see **Supplementary Table 1** online). All the peptides identified contained the conserved N-linked glycosylation motif (N-X-S/T), indicating that formerly N-glycosylated peptides were isolated with high selectivity. We further analyzed the subcellular localization of the identified proteins using information from the SWISS-PROT database (http://www.expasy.org/sprot/)[39]or the prediction tool PSORT II[40–42] (http://psort.ims.u-tokyo.ac.jp/). Of a total of 64 identified glycoproteins, 45 (70%) were bona fide or predicted transmembrane proteins (**Fig. 4**). The nontransmembrane proteins were mostly designated as either extracellular (7 proteins, 11%) or lysosomal (9 proteins, 14%), thus belonging to 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 (*FOLH1*), 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, in which over 40% of the proteins identified were not membrane proteins[10]. 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 (encoded by the gene *HSPG2*, also known as *PGBM*), a 470-kDa extracellular protein[39], and the acidic (pI = 4.39) transmembrane protein signal sequence receptor α subunit (encoded by *SSR1*, also known as *SSRA*)[39]. These results indicate that the glycopeptide capture method is also effective for the selective analysis of proteins contained in the plasma membrane.

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 the 'top down' mode of precursor ion selection in the mass spectrometer, the most abundant proteins are preferentially identified. To identify a larger number of proteins, the sample would have to be more extensively fractionated before mass spectrometric analysis.

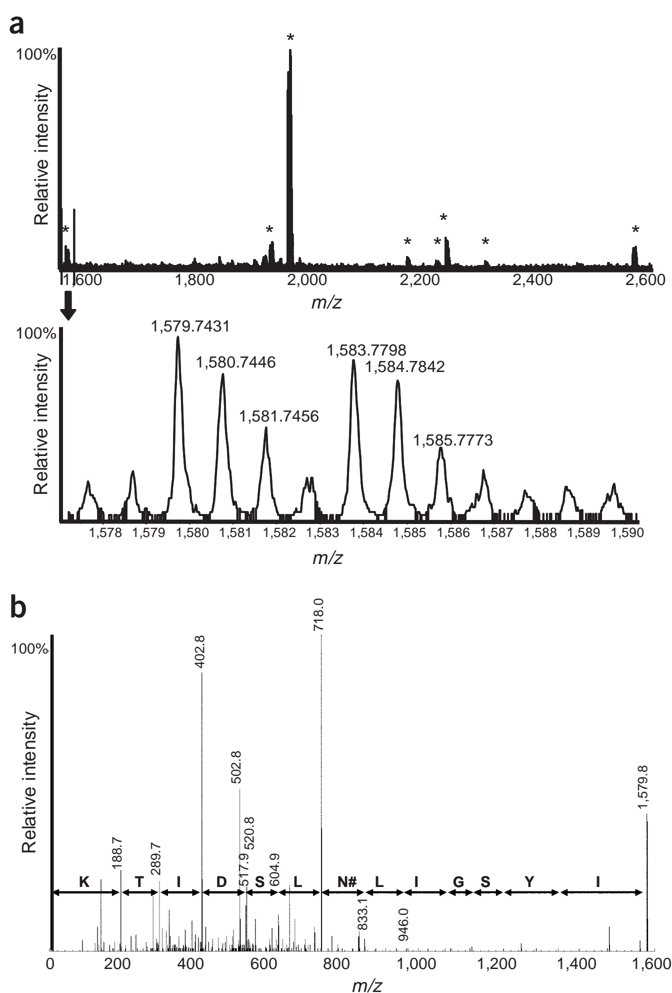


**Figure 3** Quantitative analysis of formerly N-linked glycosylated peptides by MALDI QqTOF MS. (**a**) The peptides were analyzed by MALDI-QqTOF MS with an initial MS scan to quantify the protein. The ratio of each peptide was analyzed by comparison of the total area of the first three monoisotopic peaks for the d0 and d4 forms of the peptide. (**b**) An example of a peptide with specific *m/z* value (1579.7431) was selected for CID spectrum to identify the peptide and N-linked glycosylation site.

**DISCUSSION**

We describe a method 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 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 and clinical importance, especially for the identification of diagnostic biomarkers and targets for immunotherapy or pharmacological intervention.

By combining this method with the cysteine-tagging method using ICAT reagents[3], the occupancy of individual N-linked glycosylation sites and changes thereof can also be determined. This would be of particular interest in studies in which changes of glycosylation occupancy are suspected, as exemplified by patients with type I congenital disorders of glycosylation, in which the N-linked glycosylation pathway is deficient[43].

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 as described in this paper is focused on the analysis of N-linked glycosylation sites. Analogous strategies can be devised to analyze O-glycosylated peptides as well; in fact, a protein sample, once immobilized on a solid support, can be subjected to

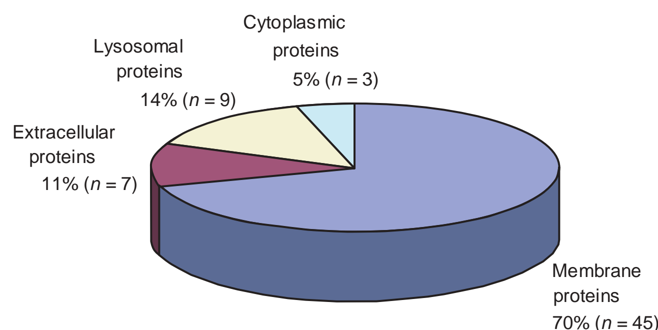


**Figure 4** Subcellular location of glycoproteins identified from a crude microsomal fraction of LNCaP prostate epithelial cells. For known and characterized proteins, protein subcellular localization was from the SWISS-PROT database (indicated by normal letter in **Supplementary Table 1** online). For protein sequences from hypothetical proteins and functionally uncharacterized proteins, PSORT II and algorithms included therein were used to predict cellular localization (indicated by italic letters in **Supplementary Table 1** online).

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. We therefore believe that this method will find wide application in proteomics research.

## METHODS

**Materials and reagents.** For all chromatographic steps, high-performance LC-grade reagents were purchased from Fisher Scientific. MALDI matrix (α-cyano-4-hydroxycinnamic acid) was from Agilent Technologies. Succinic-d4-anhydride was from C/D/N Isotopes. Gel-CODE glycoprotein staining kit was from Pierce. PNGase F was from New England Biolabs. Hydrazide resin was from Bio-Rad. All cell culture reagents and media were from Gibco-BRL. All other chemicals and the human serum sample used in this study were purchased from Sigma.

**Purification of formerly N-linked glycosylated peptides from serum.** Serum glycoproteins in coupling buffer (100 mM sodium acetate and 150 mM NaCl, pH 5.5) were oxidized by adding 15 mM of sodium periodate at room temperature for 1 h. After removal of sodium periodate, the sample was conjugated to the hydrazide resin at room temperature (20 °C) for 10–24 h. Nonglycoproteins were then removed by washing the resin six times with an equal volume of urea solution (8 M urea/0.4 M $NH_4HCO_3$, pH 8.3). After the last wash and removal of the urea solution, the resin was diluted with 3 bed volumes of water. Trypsin was added at a concentration of 1 µg of trypsin/200 µg of serum protein and digested at 37 °C overnight. The peptides were reduced by adding 8 mM TCEP (Pierce) at room temperature (20 °C) for 30 min, and alkylated by adding 10 mM iodoacetamide at room temperature (20 °C) for 30 min. The trypsin-released peptides were removed by washing the resin three times with 1.5 M NaCl, 80% acetonitrile/0.1% trifluoroacetic acid (vol/vol), 100% (vol/vol) methanol and six times with 0.1 M $NH_4HCO_3$. N-linked glycopeptides were released from the resin by addition of PNGase F (at a concentration of 1 µl of PNGase F per 40 mg of serum protein) overnight. The released peptides were dried and resuspended in 0.4% acetic acid for LC-MS/MS analysis.

**Isotopic labeling of glycopeptides with succinic anhydride.** The glycopeptides on the beads were washed twice with 15% (wt/vol) $NH_4OH$ 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 min. Beads were then washed twice with water, twice with DMF/pyridine/$H_2O$ (50%/10%/40% (vol/vol/vol)) and resuspended in DMF/pyridine/$H_2O$ (50%/10%/40% (vol/vol/vol)). Succinic anhydride solution was added to a final concentration of 2 mg/ml. The sample was incubated at room temperature for 1 h, then washed 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 described above.

**Analysis of three-control protein mix using the glycopeptide capture method.** Two mixtures containing differing amounts of the same three glycoproteins were prepared. The proteins were purchased from Calbiochem. The amounts of each protein (µg) in mixtures A and B were: α-1-antitrypsin (50, 10), α-2-hs-glycoprotein (10, 30) and α-1-antichymotrypsin (2, 2). Formerly N-linked glycosylated peptides from the two protein mixtures were purified and labeled as described above.

**Profiling serum proteins using cysteine-specific tagging.** Five milligrams of total serum protein was labeled using ICAT reagent as described previously[10]. The proteins were digested by trypsin and the peptides were fractionated by strong cation-exchange chromatography to 17 fractions. Each fraction was further separated by avidin chromatography and µLC-MS/MS[10].

**Purification of formerly N-linked glycosylated peptides from LNCaP crude microsomal membrane fractions.** LNCaP cells were grown in RPMI medium supplemented with 10% FBS. A crude microsomal fraction was prepared as described[10]. The protein amount was measured using a BCA protein analysis kit (Pierce). Sixty micrograms of crude microsomal membrane proteins was conjugated to hydrazide resin and processed as described above.

**Analysis of peptides by MS.** The peptides were analyzed by Finnigan LCQ ion trap mass spectrometer (Finnigan) or MALDI QqTOF mass spectrometer (MDS SCIEX) as described previously[3,35,36].

**Searching database using 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, the mass window for the single-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 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 tool was set to expect the following modifications: carboxymethylated cysteines, oxidized methionines and an enzyme-catalyzed conversion of asparagine to aspartic acid at the site of carbohydrate attachment. No other constraints were included in the Sequest search. All the MS/MS spectra were manually checked to verify the validity of the database search results.

*Note: Supplementary information is available on the Nature Biotechnology website.*

ACKNOWLEDGMENTS
This work was supported in part by grants from the National Cancer Institute (R33, CA93302) and by a sponsored research agreement from MacroGenics. The Institute for Systems Biology is supported by a generous gift from Merck & Co. D.B.M is supported by a National Institutes of Health career development award (K08CA97282-01).

COMPETING INTERESTS STATEMENT
The authors declare that they have no competing financial interests.

Received 7 November 2002; accepted 17 March 2003
Published online 18 May 2003; doi:10.1038/nbt827

1. Oda, Y., Huang, K., Cross, F.R., Cowburn, D. & Chait, B.T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. USA* **96**, 6591–6596 (1999).
2. Veenstra, T.D., Martinovic, S., Anderson, G.A., Pasa-Tolic, L. & Smith, R.D. Proteome analysis using selective incorporation of isotopically labeled amino acids. *J. Am. Soc. Mass Spectrom.* **11**, 78–82 (2000).
3. Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
4. Link, A.J. *et al.* Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17**, 676–682 (1999).

© 2003 Nature Publishing Group http://www.nature.com/naturebiotechnology

5. Rappsilber, J., Ryder, U., Lamond, A.I. & Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**, 1231–1245 (2002).
6. Zhou, Z., Licklider, L.J., Gygi, S.P. & Reed, R. Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**, 182–185 (2002).
7. Rout, M.P. *et al.* The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635–651 (2000).
8. Fountoulakis, M., Berndt, P., Langen, H. & Suter, L. The rat liver mitochondrial proteins. *Electrophoresis* **23**, 311–328 (2002).
9. Yi, E.C. *et al.* Approaching complete peroxisome characterization by gas-phase fractionation. *Electrophoresis* **23**, 3205–3216 (2002).
10. Han, D.K., Eng, J., Zhou, H. & Aebersold, R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **19**, 946–951 (2001).
11. Bergquist, J., Gobom, J., Blomberg, A., Roepstorff, P. & Ekman, R. Identification of nuclei associated proteins by 2D-gel electrophoresis and mass spectrometry. *J. Neurosci. Methods* **109**, 3–11 (2001).
12. Ficarro, S.B. *et al.* Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **20**, 301–305 (2002).
13. Oda, Y., Nagasu, T. & Chait, B.T. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* **19**, 379–382 (2001).
14. Zhou, H., Watts, J.D. & Aebersold, R. A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* **19**, 375–378 (2001).
15. Spahr, C.S. *et al.* Simplification of complex peptide mixtures for proteomic analysis: reversible biotinylation of cysteinyl peptides. *Electrophoresis* **21**, 1635–1650 (2000).
16. Haystead, C.M., Gregory, P., Sturgill, T.W. & Haystead, T.A. Gamma-phosphate-linked ATP-sepharose for the affinity purification of protein kinases. Rapid purification to homogeneity of skeletal muscle mitogen-activated protein kinase kinase. *Eur. J. Biochem.* **214**, 459–467 (1993).
17. Adam, G.C., Sorensen, E.J. & Cravatt, B.F. Proteomic profiling of mechanistically distinct enzyme classes using a common chemotype. *Nat. Biotechnol.* **20**, 805–809 (2002).
18. Varki, A. *et al. Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999).
19. Bause, E. Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* **209**, 331–336 (1983).
20. Roth, J. Protein N-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality control, and cell interactions. *Chem. Rev.* **102**, 285–303 (2002).
21. Shepard, H.M. *et al.* Monoclonal antibody therapy of human cancer: taking the HER2 protooncogene to the clinic. *J. Clin. Immunol.* **11**, 117–127 (1991).
22. Durand, G. & Seta, N. Protein glycosylation and diseases: blood and urinary oligosaccharides as markers for diagnosis and therapeutic monitoring. *Clin. Chem.* **46**, 795–805 (2000).
23. Freeze, H.H. Update and perspectives on congenital disorders of glycosylation. *Glycobiology* **11**, 129R–143R (2001).
24. Spiro, R.G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R–56R (2002).
25. Beardsley, R.L., Karty, J.A. & Reilly, J.P. Enhancing the intensities of lysine-terminated tryptic peptide ions in matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **14**, 2147–2153 (2000).
26. Eng, J., McCormack, A.L. & Yates, J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
27. Keller, A., Nesvizhskii, A.I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392 (2002).
28. Yoshioka, Y. *et al.* The complete amino acid sequence of the A-chain of human plasma $\alpha$2HS-glycoprotein. *J. Biol. Chem.* **261**, 1665–1676 (1986).
29. Mills, P.B., Mills, K., Johnson, A.W., Clayton, P.T. & Winchester, B.G. Analysis by matrix assisted laser desorption/ionisation-time of flight mass spectrometry of the post-translational modifications of $\alpha$1-antitrypsin isoforms separated by two-dimensional polyacrylamide gel electrophoresis. *Proteomics* **1**, 778–786 (2001).
30. Baumann, U. *et al.* Crystal structure of cleaved human $\alpha$1-antichymotrypsin at 2.7 A resolution and its comparison with other serpins. *J. Mol. Biol.* **218**, 595–606 (1991).
31. Bobbitt, J.M. Periodate oxidation of carbohydrates. *Adv. Carbohydr. Chem.* **11**, 1–41 (1956).
32. Bayer, E.A., Ben-Hur, H. & Wilchek, M. Biocytin hydrazide—a selective label for sialic acids, galactose, and other sugars in glycoconjugates using avidin-biotin technology. *Anal. Biochem.* **170**, 271–281 (1988).
33. Geoghegan, K.F. & Stroh, J.G. Site-directed conjugation of nonpeptide groups to peptides and proteins via periodate oxidation of a 2-amino alcohol. Application to modification at N-terminal serine. *Bioconjug. Chem.* **3**, 138–146 (1992).
34. Maley, F., Trimble, R.B., Tarentino, A.L. & Plummer, T.H., Jr. Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Anal. Biochem.* **180**, 195–204 (1989).
35. Griffin, T.J. *et al.* Quantitative proteomic analysis using a MALDI quadrupole time-of-flight mass spectrometer. *Anal. Chem.* **73**, 978–986 (2001).
36. Griffin, T.J. *et al.* Abundance ratio-dependent proteomic analysis by mass spectrometry. *Anal. Chem.* **75**, 867–874 (2003).
37. Gonzalez, J. *et al.* A method for determination of N-glycosylation sites in glycoproteins by collision-induced dissociation analysis in fast atom bombardment mass spectrometry: identification of the positions of carbohydrate-linked asparagine in recombinant $\alpha$-amylase by treatment with peptide-N-glycosidase F in 18O-labeled water. *Anal. Biochem.* **205**, 151–158 (1992).
38. Horoszewicz, J.S. *et al.* The LNCaP cell line—a new model for studies on human prostatic carcinoma. *Prog. Clin. Biol. Res.* **37**, 115–132 (1980).
39. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).
40. Nakai, K. & Horton, P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* **24**, 34–36 (1999).
41. Klein, P., Kanehisa, M. & DeLisi, C. The detection and classification of membrane-spanning proteins. *Biochim. Biophys. Acta* **815**, 468–876 (1985).
42. Horton, P. & Nakai, K. Better prediction of protein cellular localization sites with the k nearest neighbors classifier. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **5**, 147–152 (1997).
43. Aebi, M. & Hennet, T. Congenital disorders of glycosylation: genetic model systems lead the way. *Trends Cell Biol.* **11**, 136–141 (2001).

npg © 2003 Nature Publishing Group http://www.nature.com/naturebiotechnology