Shotgun Glycopeptide Capture Approach Coupled with Mass Spectrometry for Comprehensive Glycoproteomics

Bingyun Sun, Jeffrey A. Ranish, Angelita G. Utleg, James T. White, Xiaowei Yan, Biaoyang Lin‡, and Leroy Hood§

We present a robust and general shotgun glycoproteomics approach to comprehensively profile glycoproteins in complex biological mixtures. In this approach, glycopeptides derived from glycoproteins are enriched by selective capture onto a solid support using hydrazide chemistry followed by enzymatic release of the peptides and subsequent analysis by tandem mass spectrometry. The approach was validated using standard protein mixtures that resulted in a close to 100% capture efficiency. Our capture approach was then applied to microsomal fractions of the cisplatin-resistant ovarian cancer cell line IGROV-1/CP. With a Protein Prophet probability value greater than 0.9, we identified a total of 302 proteins with an average protein identification rate of 136 ± 19 (n = 4) in a single linear quadrupole ion trap (LTQ) mass spectrometer nano-LC-MS experiment and a selectivity of 91 ± 1.6% (n = 4) for the N-linked glycoconsensus sequence. Our method has several advantages. 1) Digestion of proteins initially into peptides improves the solubility of large membrane proteins and exposes all of the glycosylation sites to ensure equal accessibility to capture reagents. 2) Capturing glycosylated peptides can effectively reduce sample complexity and at the same time increase the confidence of MS-based protein identifications (more potential peptide identifications per protein). 3) The utility of sodium sulfite as a quencher in our capture approach to replace the solid phase extraction step in an earlier glycoprotein chemical capture approach for removing excess sodium periodate allows the overall capture procedure to be completed in a single vessel. This improvement minimizes sample loss, increases sensitivity, and makes our protocol amenable for high throughput implementation, a feature that is essential for biomarker identification and validation of a large number of clinical samples. 4) The approach is demonstrated here on the analysis of N-linked glycopeptides; however, it can be applied equally well to O-glycoprotein analysis. Molecular & Cellular Proteomics 6:141–149, 2007.

Glycosylation is one of the most important and abundant post-translational modifications in nature (1). Glycoproteins play important roles during molecular and cellular recognition in development, growth, and cellular communication and in particular are involved in cancer progression and immune responses (2, 3). Glycoproteins have been used as therapeutic targets and biomarkers for cancer prognosis, diagnosis, and monitoring. Examples include the carcinoembryonic antigen in colon, breast, pancreatic, and lung cancers; Her2/neu in breast cancer; β human chorionic gonadotropin and α-fetoprotein in germ cell tumors; prostate-specific antigen in prostate cancer; and CA-125 in ovarian cancer (4, 5). As systems biology begins to revolutionize our understanding of biology and biomedical sciences (6, 7), the ability to efficiently and comprehensively profile glycoproteins in biological samples of interest (such as cells extracts and body fluids) is critical to many biological and clinical researchers.

Tandem mass spectrometry with its superior sensitivity, accuracy, and throughput in protein and peptide identification is currently the most sophisticated and powerful tool for global proteomics studies including glycoproteome analysis. Because the enormous dynamic range of protein concentrations in biological samples is far beyond the analysis range of most techniques (10^6 in mammalian cells and 10^10 in blood), low abundant proteins are masked by dominant proteins in global proteomics analysis (6, 8). Indeed just 22 proteins constitute about 99% of the blood protein mass (albumin alone is more than 50% of the mass). Front end enrichment and fractionation methods prior to MS analysis are necessary to enhance the detection sensitivity to low abundance proteins, a category that holds promising diagnostic and biological information (9). An effective enrichment of glycosylated proteins is important to decrease sample complexity and helps to unfold the glycoproteome comprehensively (10). Two strategies have emerged to enrich glycoproteins and/or glycopeptides: one is the “top down” strategy in which glycoproteins are enriched at the protein level and then digested into peptides (e.g. the lectin affinity capture (11) and glycoprotein chemical capture (5) approaches); the other is the “bottom up” strategy in which glycopeptides are digested first into peptides and then enriched directly (e.g. glycopeptide enrichment by chromatography (12–16)).

Despite the versatility of current glycoenrichment approaches, for complex biological samples such as sera and...
cell lysates, it is cumbersome to unravel glycoproteome completely. For instance, the top down strategy suffers from solubility problems and sterical hindrance when capturing proteins in their native forms. Moreover proteolysis of complex protein mixtures with trypsin (a commonly used proteolytic enzyme for tandem MS analysis) typically produces 20 or more peptides per protein; this results in increased sample complexity and is thereof not suitable for the analysis of low abundance proteins in complex samples. Further enrichment of glycosylated peptides after glycoprotein capture has been studied both by lectin affinity capture (17) and glycoprotein chemical capture (18) approaches. Although lectin affinity capture is the most widely used approach due to its ease of implementation, the binding selectivity of lectins to specific conformations of different carbohydrate moieties has limited the utility of lectin in global glycoprotein analysis (19, 20). The glycoprotein chemical capture approach developed by Zhang et al. (5) is generally applicable to all types of glycoproteins, but the complicated steps to implementation (12) and the relative low yields (depicted in Fig. 1, A and B) lead to this approach not being used as widely as the lectin capture approach. In the bottom up strategy, proteins are digested into peptides, and glycosylated peptides are separated from their unglycosylated counterparts by chromatography (12–16). Although this approach is direct, simple, and rapid, the separation based on different physical and chemical properties usually results in only a modest efficiency (12, 14).

We report here a novel chemical capture approach that focuses on a very efficient glycopeptide enrichment. Our approach provides optimized and robust selectivity for glycosylated peptides, improved identification of glycosylated membrane proteins, and enhanced MS detection sensitivity and accuracy to low abundance but multiglycosylated proteins. The strategy is illustrated in Fig. 1C. We demonstrated the feasibility and characterized the capture efficiency of this approach using chicken avidin, a monoglycosylated protein, and on a protein mixture consisting of five different glycoproteins containing up to 13 glycosylation sites. Close to 100% capture efficiency was obtained both for chicken avidin (using MALDI-TOF/TOF) and for the five-glycoprotein mixture (using LTQ LC-MS/MS). We also applied our capture approach to a complex and challenging biological mixture, the microsomal fraction from an ovarian cancer cell line, IGROV-1/CP (a cisplatin-resistant ovarian cancer cell line derived from the cisplatin-sensitive ovarian-cancer cell line IGROV-1). During a typical nano-LC-MS analysis, we identified a total of 156 unique proteins and 311 unique peptides, which includes 68 proteins with multiple peptide hits. The glycopeptide specificity of our approach is 91%. From four LTQ experiments of two biological replicates (two repeated LTQ runs per each biological replicate), we identified a total of 302 proteins with an average protein identification rate of 136 ± 19 per LTQ MS run (n = 4) and a selectivity of 91 ± 1.6% for the N-linked glycoconsensus sequence (n = 4). We used 0.9 as a cutoff of the Protein Prophet probability value in all of the analyses, and on average the error rate of our analysis was as small as 0.006 in all four MS runs, and the averaged number of incorrectly identified peptides was 1 of 136 by statistical analysis (21).

**MATERIALS AND METHODS**

All glycoproteins were obtained from Sigma. MALDI matrix was from Agilent Technologies (Palo Alto, CA). Bradford assay reagent, sodium periodate, and hydrazide resin were obtained from Bio-Rad. PNGase F1 was from New England Biolabs. The cell culture reagents and media were from Invitrogen. TCEP was from Pierce, and trypsin was from Promega. Rapigest and C18 columns were from Waters. All other chemicals were purchased from Fisher Scientific.

**Glycopeptide Capture—**In our glycopeptide capture approach, biological samples were subjected to denaturation and trypsin digestion first. In a typical procedure, a biological sample was reconstituted in a denaturing buffer of 5 mM EDTA, 40 mM tris, 10 mM TCEP, 0.5% Rapigest at pH 8.3 and heated at 100 °C for 10 min. After allowing the sample to cool to room temperature, urea was added to 8 M, and the solution was incubated at 37 °C for 30 min. To prevent disulfide bond formation, cysteine residues were modified by alklylation with iodoacetamide. Iodoacetamide was added to the sample solution in at least a 6-fold molar excess over the free sulfhydryls in the sample. For an unknown protein mixture, an estimation of 6 cysteines per protein was used for calculating the molar concentration of sulfhydryls. A 30-min incubation in the dark, at room temperature, with end-over-end rotation was carried out for cysteine derivatization. The reaction was quenched by the addition of DTT at half of the molar concentration of the iodoacetamide for 10 min. After iodoacetamide deactivation, the sample solution was diluted 10-fold with 40 mM Tris buffer, pH 8.3, and 1 mg of trypsin/20–50 mg of protein was added into the sample solution, and the sample mixture was digested at 37 °C overnight. To avoid a large volume for trypsin digestion, the denatured sample was kept at 4–6 mg/ml. Rapigest was degraded by acidifying the trypsinized sample mixture to pH ~1 with HCl and incubation at 37 °C for 1 h. The hydrophobic residues of Rapigest were precipitated out and removed from the sample by centrifugation, and the supernatant was passed over a C18 column to remove extra urea, DTT, and Tris. Tryptic peptides were eluted from the column with 80% acetonitrile in 0.1% TFA and dried in a SpeedVac® (Thermo Savant, Holbrook, NY) concentrator.

**Tryptic Digestion of Samples—**In our glycopeptide capture approach, biological samples were subjected to denaturation and trypsin digestion first. In a typical procedure, a biological sample was reconstituted in a denaturing buffer of 5 mM EDTA, 40 mM tris, 10 mM TCEP, 0.5% Rapigest at pH 8.3 and heated at 100 °C for 10 min. After allowing the sample to cool to room temperature, urea was added to 8 M, and the solution was incubated at 37 °C for 30 min. To prevent disulfide bond formation, cysteine residues were modified by alklylation with iodoacetamide. Iodoacetamide was added to the sample solution in at least a 6-fold molar excess over the free sulfhydryls in the sample. For an unknown protein mixture, an estimation of 6 cysteines per protein was used for calculating the molar concentration of sulfhydryls. A 30-min incubation in the dark, at room temperature, with end-over-end rotation was carried out for cysteine derivatization. The reaction was quenched by the addition of DTT at half of the molar concentration of the iodoacetamide for 10 min. After iodoacetamide deactivation, the sample solution was diluted 10-fold with 40 mM Tris buffer, pH 8.3, and 1 mg of trypsin/20–50 mg of protein was added into the sample solution, and the sample mixture was digested at 37 °C overnight. To avoid a large volume for trypsin digestion, the denatured sample was kept at 4–6 mg/ml. Rapigest was degraded by acidifying the trypsinized sample mixture to pH ~1 with HCl and incubation at 37 °C for 1 h. The hydrophobic residues of Rapigest were precipitated out and removed from the sample by centrifugation, and the supernatant was passed over a C18 column to remove extra urea, DTT, and Tris. Tryptic peptides were eluted from the column with 80% acetonitrile in 0.1% TFA and dried in a SpeedVac® (Thermo Savant, Holbrook, NY) concentrator.

**Cell Culture and Microsomal Fraction Extraction—**IGROV-1/CP cisplatin-resistant ovarian cancer cells were grown in RPMI 1640 medium (Invitrogen) containing 10% fetal bovine serum, 100 units/ml penicillin, and 100 units/ml streptomycin at 37 °C. A crude microsomal fraction of IGROV-1/CP was prepared as described previously (22, 23). The microsomal pellet contained plasma membranes, Golgi apparatus, endoplasmic reticulum, mitochondria, lysosomes, and all other membrane-bound vesicles separated from soluble cytosol. The Bradford protein assay was used to quantify the concentration of the extracted proteins. About 0.5–0.8 mg of crude microsomal membrane proteins was used to proceed with the glycopeptide capture.

**Materials and Methods**

1 The abbreviations used are: PNGase F, peptide-N-glycosidase F; n, number of experiments; TCEP, tris(2-carboxyethyl)phosphine hydrochloride; SPE, solid phase extraction; ADAM 10, a disintegrin and metalloproteinase domain 10; LTQ, linear quadrupole ion trap mass spectrometer; CD, cluster of differentiation.
following reactions. First, to oxidize the cis-diol groups of carbohydrates to aldehydes, sodium periodate at 10 mM final concentration was introduced into the peptide solution, and the sample was incubated in the dark at room temperature for 30 min with end-over-end rotation. Second, sodium sulfite was added to 20 mM final concentration and incubated for 10 min to deactivate the excess oxidant in the peptide solution. The coupling reaction was initiated by introducing hydrazide resin into the quenched peptide solution at 20 mg/ml resin, and extra coupling buffer was added to make a solid to liquid ratio of 1:5. The coupling reaction was performed at 37 °C overnight with end-over-end rotation.

After the coupling reaction, the resin was washed twice thoroughly and sequentially with deionized water, 1.5 n NaCl, methanol, and acetonitrile and was followed by a buffer exchange step to 100 mM NH4HCO3 (made fresh), pH = 8.0. Enzymatic cleavage of the N-linked peptides from the sugar moiety was carried out at 37 °C overnight by PNGase F at a concentration of 1 µl of PNGase F/2–6 mg of crude proteins. The supernatant, containing the released deglycosylated peptides, was collected by centrifugation and combined with the supernatant of an 80% acetonitrile wash. The peptide solution was dried, reconstituted with 1% acetonitrile in 0.1% formic acid, and subjected to MS analysis.

Analysis of Peptides by Mass Spectrometry—Peptide samples were analyzed by either a MALDI-TOF/TOF tandem mass spectrometer (ABI 4700 Proteomics Analyzer, Applied Biosystems, Foster City, CA) or by nano-LC-ESI-MS/MS using an LTQ linear ion trap mass spectrometer (Thermo Finnigan, San Jose, CA). For MALDI-TOF/TOF analysis, the peptide sample was purified with a ZipTip (Millipore) and reconstituted with 0.4% acetic acid prior to analysis. A 1:1 dilution of peptide solution with MALDI matrix solution (Agilent Technologies) was used for MALDI spotting.

For LTQ mass spectrometry analysis, an in-house fabricated nanoelectrospray source and an HP1100 solvent delivery system (Agilent Technologies) were coupled to LTQ. Samples were automatically injected by a FAMOS autosampler (LC Packings, San Francisco, CA) or by nano-LC-ESI-MS/MS using an LTO linear ion trap mass spectrometer (Thermo Finnigan, San Jose, CA). For MALDI-TOF/TOF analysis, the peptide sample was purified with a ZipTip (Millipore) and reconstituted with 0.4% acetic acid prior to analysis. A 1:1 dilution of peptide solution with MALDI matrix solution (Agilent Technologies) was used for MALDI spotting.

Application of the Glycopeptide Capture Approach to a Monoglycosylated Protein—To validate our glycopeptide capture approach, we first analyzed a mono-N-glycosylated protein, chicken avidin. Fig. 2A shows the MS spectrum of the enriched avidin deglycosylated glycopeptide collected by a MALDI-TOF/TOF mass spectrometer (ABI 4700 Proteomics Analyzer, Applied Biosystems). The two major peaks were from the same peptide (K.WTNDLSNMT/GAVNSR.G where the consensus sequence of N-glycosylation sites is bold and underlined and the periods indicate the peptide cleavage sites) as determined by MS/MS fragmentation (please see Supplemental Fig. 1 for the CID spectra). The mass of 1852 is contributed from the peptide with methionine oxidized to methionine sulfoxide (with a mass increase of 16 Da) (5) and the asparagine from the consensus sequence modified to aspartic acid (with a mass increase of 1 Da) after PNGase F deglycosylation (28). The mass of 1791 is attributed from the same peptide as shown by the CID spectrum with a cleavage between the γ carbon and the sulfur of the methionine side chain, which most likely happened during the mass spectrometry direction (33–35). The successful identification of the formally glycosylated peptide from chicken avidin indicated that the capture strategy we developed here was effective and that the sodium sulfite we introduced to reduce periodate did not interfere with the capture procedure. To evaluate the capture efficiency, we analyzed the non-captured avidin peptide mixture by MALDI-TOF/TOF as well. Because glycosylated peptides give low and complex signals in MS spectra,...
due to the heterogeneity of the glycan structure, we used PNGase F to remove all linked glycans prior to MS analysis and focused on the presence or absence of the deglycosylated glycopeptides from the non-captured fraction of chicken avidin. Fig. 2B shows the MS spectrum of the non-captured avidin peptides after pursuing the glycopeptide capture and PNGase F deglycosylation. The absence of glycopeptide signal indicates that the efficiency of our glycopeptide capture process is high based on the MALDI-TOF/TOF analysis.

Application of the Glycopeptide Capture Strategy to a Five-glycoprotein Mixture—To further validate the glycopeptide capture approach and to characterize the capture efficiency, we applied our glycopeptide capture strategy to a protein mixture with five N-glycosylated proteins: invertase (yeast), α1-antitrypsin (human), conalbumin (chicken), ribonuclease B (bovine), and ovalbumin (chicken) (all purchased from Sigma). Table I lists the representative N-linked glycopeptides captured and identified by our approach using a nano-LC-MS/MS analysis on an LTQ linear ion trap mass spectrometer. All of the proteins were identified with a Protein Prophet value of 1.0. Strikingly all the previously identified glycosylation sites within the five glycoproteins were captured and identified except for some sites from invertase (Table I). For invertase

Fig. 2. Results from the MALDI-TOF analysis of the glycopeptide capture strategy applied to tryptic peptides of chicken avidin. A, deglycosylated N-glycopeptides captured from chicken avidin. B, non-captured tryptic peptides from avidin after sequential glycopeptide capture and deglycosylation by PNGase F. The inset shows the expanded m/z region from 1770 to 1955 of B.
that has 13 N-glycosylation sites, we identified a total of eight N-linked sites (shown in Table I). The remaining five N-glycosylation sites reside in large tryptic peptides with molecular masses above 3000 Da that were absent from our LTQ results. The absence of some of the large tryptic peptides with N-linked glycosylation sites is likely caused by insufficient ionization of these peptides.

To evaluate the capture efficiency, the total non-captured peptides from the five glycoproteins were analyzed by LTQ after applying PNGase F deglycosylation. Only one N-linked consensus sequence, NLS, from ovalbumin was identified, and all the other N-linked sequences were absent. This sequence from ovalbumin has been reported to be a transient glycosylation site that only occurs in hen oviduct (36). Because of the absence of all other glycosylation sequences from the five glycoproteins in the non-captured peptide mixture, the presence of this N-linked sequence in both the captured and non-captured fraction was likely due to the fact that it is partially glycosylated rather than a result of incomplete capture. The above results suggest that our glycopeptide capture approach can be used to comprehensively capture most (perhaps all) glycopeptides in mixtures.

Application of Glycopeptide Capture Approach to Ovarian Cell Microsomal Fractions—Analysis of membrane proteins by MS is challenging because the proteins easily aggregate and are difficult to dissolve in aqueous solutions (22, 37). To assess the applicability of the glycopeptide capture approach to membrane proteins, we analyzed the microsomal fraction from a cisplatin-resistant ovarian cancer cell line (IGROV-1/CP) that is rich in membrane proteins. The capture strategy was carried out on two microsomal fractions with 500 and 800 µg of crude protein, respectively, and one-fifth of the final captured peptides were analyzed by a single nano-LC-MS/MS and data-dependent acquisition, selecting three precursor ions for MS/MS with a dynamic exclusion setting of 1. The peptides that were identified by the SEQUEST database searching with a peptide probability score above 0.9 are displayed. The color indicates different probability values.

![Fig. 3. Peptide three-dimensional image of N-glycopeptides detected by LTQ nano-LC-MS/MS from microsomal fraction of cisplatin-resistant ovarian cancer cell line IGROV-1/CP. A gradient of 10–35% solvent B (100% acetonitrile) over 30 min was applied to a 75-µm × 10-cm fused silica capillary column packed with 100-Å pore size Magic C18AQ material (Michrom Bioresources). Eluting peptides were analyzed by nano-LC-MS and data-dependent acquisition, selecting three precursor ions for MS/MS with a dynamic exclusion setting of 1. The peptides that were identified by the SEQUEST database searching with a peptide probability score above 0.9 are displayed. The color indicates different probability values.](image)

| Protein name | Sequences of identified peptides | Reported glycosylation site | Detected glycosylation site |
|--------------|---------------------------------|-----------------------------|-----------------------------|
| Invertase    | S.MTNETSDFRPVHFTPKK.G           | 13                          | 8                           |
|              | K.NFVLANSTQFR.D                 |                |                             |
|              | Y.FVGYFNGITTEAFDQCR.V           |                |                             |
|              | K.AEPIHNAGPWSR.F                |                |                             |
|              | R.FAYMTQIL.K                    |                |                             |
|              | L.VYAVNPTQITK,S.R               |                |                             |
|              | Y.MTCAALGYSVNTT.C              |                |                             |
| α1 Antitrypsin| R.LHAHSSTNIIFFPSVIATAFAMLSGTK.A | 3               | 3                           |
|              | K.ATDHERLEGLNVLNFLLT          |                |                             |
|              | K.YCNCATAIFPLDEGK.L            |                |                             |
| Conalbumin  | G.WVMPNLKHPR.T                 | 1               | 1                           |
| Ribonuclease B| S.RNTRKDR.C                    | 1               | 1                           |
| Ovalbumin   | K.NLTCVILMAGITDVFSSANLSGIIASLKR.I | 2               | 2                           |

*The consensus sequence of N-glycosylation sites is highlighted, and the period indicates the peptide cleavage site.
ber of identified peptides. Combining the results of all four MS runs (two biological replicates), we identified a total of 302 proteins from the microsomal fractions of IGROV-1/CP (see the supplemental table) with an average identification rate of \(136 \pm 19\) (\(n = 4\)) proteins and glycopeptide selectivity (the number of peptides identified with the \(N\)-glycoconsensus sequence divided by the total number of identified peptides) of \(91.0 \pm 1.6\%\) (\(n = 4\)) per MS analysis.

Based on the results of previous studies, glycosylated tryptic peptides constitute 2–5% of the glycoproteins (12, 17). If the enrichment is characterized by the ratio between the percentage of glycopeptides in the sample after (91%) and before (2–5%) (assuming all the microsomal proteins are glycoproteins) the capture, then the enrichment factor of our glycopeptide capture approach is 19–45-fold. As cell microsomal fractions also include organelle and plasma proteins that are not glycosylated (22), the enrichment factor we estimated here is conservative and provides a good demonstration of the effectiveness of our capture approach in enrichment of glycopeptides from a complex biological sample.

To classify the identified proteins by cellular function and to explore the biological significance of the glycoproteins we identified, we annotated our data using GoMiner (discover.nci.nih.gov/gominer) (40). For the analysis, Entrez Gene names were retrieved from the European Molecular Biology Laboratory International Protein Index (IPI) number, and redundant identifications were removed before GoMiner analysis. Among 302 proteins with Entrez Gene names, 251 proteins have been annotated as cellular components, and 244 proteins have been annotated for molecular function. As expected, the majority of identified proteins are membrane proteins (170 of 251). The major molecular functions among the identified proteins include ligand binding, catalytic activity, signal transduction activity, transporter activity, etc. (shown in Fig. 4). The results of this analysis are concordant with our knowledge of the main cellular location and functions of glycoproteins in the microsomal fractions.

We compared the proteins identified here with the proteins identified previously by the ICAT/MS/MS approach on a similar microsomal fraction of IGROV-1/CP cells (23). Interestingly only 46 proteins overlapped in the two datasets (302 proteins identified in glycopeptide capture dataset and 307 proteins identified in ICAT dataset), suggesting that these two approaches allow detection of different subsets of the microsomal proteome (shown in Fig. 5A) and complement each other for global proteomics. We also compared our identified proteins with glycoprotein list obtained by Zhang et al. (5) from microsomal fractions of the prostate cancer epithelial cell line LNCaP. Notwithstanding the difference in cell lines, the two microsomal glycoprotein datasets share 50 proteins in common (302 proteins in peptide capture dataset whereas only 64 proteins in protein capture dataset). That is 77% of the total identified proteins from the list obtained by Zhang et al. (5) but is only 31% of the total identified proteins from our list (Fig. 5B). Most shared proteins are common and abundant glycoproteins in cell microsomal fractions, such as integrin, sodium/potassium-transporting ATPase, cation-independent mannose 6-phosphate receptor, lysosome-associated membrane glycoprotein, glucuronidase, glycosidase, mannosidase, hexosaminidase, glucosylceramidase, etc. The much larger protein dataset including more biologically and clinically interesting glycoproteins (addressed in detail under “Discussion”) obtained from our capture approach indicates that we achieved a more comprehensive identification of the glycoproteome. At the same time, more glycosylation sites from the same protein were identified by our glycopeptide capture approach than from the protein capture approach. For exam-
ple, using glycoprotein capture, only one glycosylation site has been identified from γ-1 chain of laminin protein (5), whereas using the glycopeptide capture approach, another five additional glycosylation sites were identified (see the supplemental table).

**DISCUSSION**

The biological and clinical importance of glycoproteins raises the need for a comprehensive and robust molecular profiling of glycoproteins and accordingly ways to enrich glycoproteins. Moreover system approaches to biology and medicine benefit enormously from global approaches to the analysis of genes and proteins. We have developed an approach that digests proteins into peptides and enriches glycopeptides via hydrazide chemistry. Adapting the pros and avoiding the cons raised from both the top down and bottom up strategies, our enrichment approach is designed to have improved robustness and completeness for glycoproteomics. This approach appears to move us toward more global analyses of glycoproteins. The advantages of our approach over the existing methods are elaborated below.

First, digestion of proteins into peptides improves solubility of large membrane proteins and exposes all of the glycosylation sites to ensure equal accessibility to external capture reagents. Analyses of known structures of glycoproteins indicate that about one-third of glycosylation sites are buried inside of proteins (41). Therefore, the steric hindrance raised from protein topology can diminish the capture efficiency of many sites of glycosylation. Cleaving globular proteins to smaller peptides circumvents this shortcoming.

Second, capturing glycosylated peptides can effectively reduce the complexity of the sample and increase the confidence of using MS-based protein identifications. Although the protein capture strategy can effectively enrich glycoproteins from complicated samples, the peptides (20 or more tryptic peptides per protein) generated from proteolysis prior to MS analysis increase the sample complexity again and counteract the enrichment effect at the protein level. Given the fact that proteins can be identified by individual signature proteolytic peptides with MS and that identification from multiple peptides improves the confidence of protein assignment (21), it is ideal to use multiple peptides to identify a protein. Because glycoproteins are glycosylated at multiple sites in general (17, 5) and because the glycopeptides constitute only 2–5% of the full glycoprotein, enriching glycopeptides not only decreases sample complexity effectively but also provides multiple peptides for unambiguous protein identification. Using 0.9 as the protein probability cutoff score, on average the error rate was as small as 0.006 in all four MS runs, and the number of incorrectly identified peptides was 1 of 136 by statistical analysis (21).

Third, our capture approach using hydrazide chemistry provides good selectivity of glycopeptides over the non-glycosylated peptides. To date, different chromatographic separation techniques have been reported to enrich glycopeptides by the diverse physical and chemical properties of the glycopeptides (12–16). The selectivity, however, is very limited (12, 14). To overcome this problem, we take advantage of hydrazide chemistry, which allows us to selectively capture glycopeptides via a covalent bond formed between hydrazide and the aldehyde groups from oligosaccharides. Such chemistry is ubiquitous to all glycan structures, and the covalent bond formed tolerates extreme wash conditions. Virtually all non-covalently attached peptides can be removed from the solid support. Using our chemical capture approach, we have achieved a glycopeptide selectivity of 91% on the microsomal fraction of the ovarian cancer cells.

Fourth, the utility of sodium sulfite as a quencher in our capture approach in place of the SPE step in the glycoprotein chemical capture approach, which removes the excess sodium periodate, allows the overall capture procedure to be completed in a single vessel. This modification prevents sample loss as well as saving labor and time. Sample loss is a nontrivial problem when the proteomics research is focused on low abundance proteins such as biomarkers. The more than 3-fold increase in the number of identified proteins by our approach compared with the glycoprotein capture approach of Zhang et al. (5) may be due in part to the avoidance of sample losses. Another reason for the limited yield in the glycoprotein capture approach of Zhang et al. (5) is due to the incomplete capture of glycopeptides inherent in the protein capture approach itself as illustrated in Fig. 1A. For a multiglycosylated protein, it is highly unlikely that all of the glycosylated sites can form chemical bonds with the solid support due to the globular structure of proteins (Fig. 1A). After on-support proteolysis and a series of washes to remove non-bonded peptides, only a fraction of the glycopeptides remains on the support (Fig. 1B). For example, using glycoprotein capture only one glycosylation site was identified from γ-1 chain of laminin protein (5), whereas using the glycopeptide capture approach, five additional glycosylation sites were identified. As the PeptideProphet and Protein Prophet analyses penalize single hit identifications and reward multihit identifications (21), the glycoprotein capture approach is likely to result in a lower protein identification rate (64 proteins in total) compared with our glycopeptide capture approach (302 proteins in total).

Our glycopeptide capture approach is adaptable to high throughput and automation because of the completion of capture in a single vessel. The first step proteolysis in our peptide chemical capture approach is compatible with quantitative proteomics analyses. Moreover the glycopeptide capture approach is complementary to the widely used ICAT approach that labels and enriches cysteine-containing peptides. With only a small fraction of the peptides overlapping, the number of proteins identified by our glycopeptide capture approach is similar to that of the ICAT approach. A total of 569 proteins were identified from the microsomal fractions of IGROV-1/CP by combining the ICAT and glycopeptide capture results, indicating that the use of both strategies in con-
Novel Glycopeptide Capture Approach

cert provides a powerful approach to global proteomics profiling of a complex biological medium.

Although the biological significance of the proteins identified in this study is not the focus of this report, many of the glycoproteins we identified have been implicated in ovarian carcinoma and cisplatin resistance. For instance, the folate receptor (42, 43), the insulin-like growth factor receptor (44), and the epidermal growth factor receptor (45) are overexpressed in cancer cells and are used as drug delivery targets. The tumor-associated calcium signal transducer 1 (46), tumor necrosis factor receptor (46–48), metastasis suppressor protein 1 (46), heat shock protein HSP 90 (49), laminin (46, 50, 51), and reticulocalbin-1 (52) have been reported to be associated with ovarian carcinogenesis. Increased expression of disulfide isomerase (46, 53) and ADAM 10 (54) is strongly correlated with cisplatin resistance. These observations validate the contention that the peptide glycocapture method we developed is a powerful approach to the discovery of potential biomarkers. Meanwhile CD proteins that play important immune functions in cells are a class of membrane proteins that are often glycosylated and also make good drug targets and biomarkers. We compared our protein dataset with the PROW (Protein Reviews on the Web) database for CD proteins (mpr.nci.nih.gov/prow/) (361 CD proteins in total) and biomarkers. We compared our protein dataset with the peptide glycocapture method we developed is a powerful approach to the discovery of potential biomarkers. Meanwhile CD proteins that play important immune functions in cells are a class of membrane proteins that are often glycosylated and also make good drug targets and biomarkers. We compared our protein dataset with the PROW (Protein Reviews on the Web) database for CD proteins (mpr.nci.nih.gov/prow/) (361 CD proteins in total) and biomarkers.

Although our approach can also capture O-linked glycosylated peptides, for ease of analysis we detected the N-linked glycopeptides only. With a proper combination of O-glycosidase or chemical cleavage such as β-elimination, the O-glycopeptides can also be released from the solid support and analyzed by MS. Due to technical limitations of MS analysis such as ionization efficiency of peptides, sample complexity and dynamic range, and mass accuracy and resolution of mass spectrometry itself (4, 8, 9), not all the tryptic glycopeptides can be detected. To use this approach to study global glycosylation site(s) of individual proteins for the purpose of investigating post-translational modifications, detection approaches other than MS would be necessary. Nonetheless to serve the purpose of global glycoproteomics, our strategy provides a comprehensive and robust methodology with improved accuracy and sensitivity.

Acknowledgments—We thank James Eddes for constructing a five-protein database with reversed yeast protein sequences. We also appreciate the constructive discussions with Dr. Hui Zhang, Dr. Lan Chun Tu, Dr. Zhiyuan Hu, Alex Scherl, Martin Sadilek, and David Stheyenberg.

* This work was supported by National Institutes of Health Grants 1U54DA021519 and 1U54CA119347 and NIGMS, National Institutes of Health Grant 1P50GM076547. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence may be addressed: The Inst. for Systems Biology, 1441 N. 34th St., Seattle, WA 98103. Tel.: 206-732-1297; Fax: 206-732-1299; E-mail: blin@systemsbiology.org.

§ To whom correspondence may be addressed: The Inst. for Systems Biology, 1441 N. 34th St., Seattle, WA 98103. Tel.: 206-732-1201; Fax: 206-732-1299; E-mail: lhoo@systemsbiology.org.

REFERENCES

1. Parodi, A. J. (2000) Protein glycosylation and its role in protein folding. Annu. Rev. Biochem. 69, 89–93
2. Helenius, A., and Aebi, M. (2001) Intracellular functions of N-linked glycans. Science 291, 2364–2369
3. Lowe, J. B. (2001) Glycosylation, immunity, and autoimmunity. Cell 104, 809–812
4. Diamandis, E. P. (2004) Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. Mol. Cell. Proteomics 3, 367–378
5. Zhang, H., Li, X. J., Martin, D. B., and Aebersold, R. (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat. Biotechnol. 21, 660–666
6. Hood, L. (2003) Systems biology: integrating technology, biology, and computation. Mech. Ageing Dev. 124, 9–16
7. Hood, L., Heath, J. R., Phelps, M. E., and Lin, B. (2004) Systems biology and new technologies enable predictive and preventative medicine. Science 306, 640–643
8. Aebersold, R., and Cravatt, B. F. (2002) Proteomics—advances, applications and the challenges that remain. Trends Biotechnol. 20, S1–S2
9. Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: history, character, and diagnostic prospects. Mol. Cell. Proteomics 1, 845–867
10. Novotny, M. V., and Mechref, Y. (2005) New hyphenated methodologies in high-sensitivity glycoprotein analysis. J. Sep. Sci. 28, 1956–1968
11. O’Shannessy, D. J., and Quarles, R. H. (1987) Labeling of the oligosacchare moieties of immunoglobulins. J. Immunol. Methods 99, 153–161
12. Alvarez-Manilla, G., Atwood, J. A., Guo, Y., Warren, N. L., Orlando, R., and Pierce, M. (2006) Tools for glycoproteomic analysis: size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites. J. Proteome Res. 5, 701–708
13. An, H. J., Peavy, T. R., Hedrick, J. L., and Lebrilla, C. B. (2003) Determination of N-glycosylation sites and site heterogeneity in glycoproteins. Anal. Chem. 75, 5628–5637
14. Hagglund, P., Bunkenberg, J., Elortza, F., Jensen, O. N., and Roepstorff, P. (2004) A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. J. Proteome Res. 3, 556–566
15. Larsen, M. R., Hofrup, P., and Roepstorff, P. (2005) Characterization of gel-separated glycoproteins using two-step proteolytic digestion combined with sequential microcolumns and mass spectrometry. Mol. Cell. Proteomics 4, 107–119
16. Wada, Y., Tajiri, M., and Yoshida, S. (2004) Hydrophilic affinity isolation and MALDI-multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics. Anal. Chem. 76, 6560–6565
17. Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi, N., and Itose, T. (2003) Lectin affinity capture, isoporte-coded tagging and mass spectrometry to identify N-linked glycoproteins. Nat. Biotechnol. 21, 667–672
18. Bobbitt, J. M. (1956) Periodate oxidation of carbohydrates. Adv. Carbohydr. Chem. 4, 1–41
19. Lis, H., and Sharon, N. (1998) Lectins: carbohydrate-specific proteins that mediate cellular recognition. Chem. Rev. 98, 637–674
20. Nilsson, C. L. (2003) Lectins: proteins that interpret the sugar code. Anal. Chem. 75, 548A–535A
21. Niewczikisz, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 75, 4646–4658
22. Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nat. Biotechnol. 19, 948–951
23. Stewart, J. J., White, J. T., Yan, X., Collins, S., Drescher, C. W., Urban,
Novel Glycopeptide Capture Approach

N. D., Hood, L., and Lin, B. (2006) Proteins associated with cisplatin resistance in ovarian cancer cells identified by quantitative proteomic technology and integrated with mRNA expression levels. Mol. Cell. Proteomics 5, 433–443

24. Yi, E. C., Lee, H., Aebbersold, R., and Goodlett, D. R. (2003) A microcapillary trap cartridge-microcapillary high-performance liquid chromatography electrospray ionization emitter device capable of peptide tandem mass spectrometry at the attomole level on an ion trap mass spectrometer with automated routine operation. Rapid Commun. Mass Spectrom. 17, 2093–2098

25. Gygi, S. P., Rochon, Y.,Franza, B. R., and Aebbersold, R. (1999) Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol. 19, 1720–1730

26. Keller, A., Eng, J., Zhang, N., Li, X.-j., and Aebbersold, R. (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol. Syst. Biol. 1, m04100024–E1–m04100024–E8

27. Pedrioli, P. G., Eng, J. K., Hubley, R., Vogelzang, M., Deutsch, E. W., Raught, B., Pratt, B., Nilsson, E., Angeletti, R. H., Apweiler, R., Cheung, K., Costello, C. E., Hermjakob, H., Huang, S., Julian, R. K., Kapp, E., McComb, M. E., Oliver, S. G., Ommenn, G., Patton, N. W., Simpson, R., Smith, R., Taylor, C. F., Zhu, W., and Aebbersold, R. (2004) A common open representation of mass spectrometry data and its application to proteomics research. Nat. Biotechnol. 22, 1459–1466

28. Carr, S. A., and Roberts, G. D. (1986) Carbohydrate mapping by mass spectrometry: a novel method for identifying attachment sites of Asn-linked sugars in glycoproteins. Anal. Biochem. 157, 396–406

29. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebbersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. 74, 5383–5392

30. Keil, B. (1992) Specifity of Proteolysis, p. 335, Springer-Verlag, New York

31. Eng, J. K., McCormack, A. L., and Yates, J. R., III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 767–869

32. Wu, T. P., Ruan, K. C., and Liu, W. Y. (1998) A fluorescence-labeling method for sequencing small RNA on polycrylamide gel. Nucleic Acids Res. 24, 3472–3473

33. Jiang, X., Smith, J. B., and Abraham, E. C. (1996) Identification of a MS-MS fragment diagnostic for methionine sulfoxide. J. Mass Spectrom. 31, 1309–1310

34. Lapko, V. N., Smith, D. L., and Smith, J. B. (2000) Identification of an artifact in the mass spectrometry of proteins derivatized with iodoacetamide. J. Mass Spectrom. 35, 572–575

35. Reid, G. E., Roberts, K. D., Simpson, R. J., and O’Hair, R. A. (2005) Selective identification and quantitative analysis of methionine containing peptides by charge derivatization and tandem mass spectrometry. J. Am. Soc. Mass Spectrom. 16, 1131–1150

36. Suzuki, T., Kitajima, K., Emori, Y., Inoue, Y., and Inoue, S. (1997) Site-specific de-N-glycosylation of diglycosylated ovalbumin in hen oviduct by endoglycosidase: N-glycanase as a quality control system for newly synthesized proteins. Proc. Natl. Acad. Sci. U. S. A. 94, 6244–6249

37. Wu, C. C., MacCoss, M. J., Howell, K. E., and Yates, J. R., III (2003) A method for the comprehensive proteomic analysis of membrane proteins. Nat. Biotechnol. 21, 532–538

38. Li, X. J., Pedrioli, P. G., Eng, J., Martin, D., Yi, E. C., Lee, H., and Aebbersold, R. (2004) A tool to visualize and evaluate data obtained by liquid chromatography-electrospray ionization-mass spectrometry. Anal. Chem. 76, 3856–3860

39. Li, X. J., Yi, E. C., Kemp, C. J., Zhang, H., and Aebbersold, R. (2005) A software suite for the generation and comparison of peptide arrays from sets of data collected by liquid chromatography-mass spectrometry, Mol. Cell. Proteomics 4, 1328–1340

40. Zeeberg, B. R., Feng, W., Wang, G., Wang, M. D., Fojo, A. T., Sunshine, M., Narasimhan, S., Kane, D. W., Reinhold, W. C., Lababidi, S., Bussey, K. J., Riss, J., Barrett, J. C., and Weinstein, J. N. (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol. 4, R28

41. Petrescu, A. J., Milic, A. L., Petrescu, S. M., Dwek, R. A., and Wormald, M. R. (2004) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. Glycobiology 14, 103–114

42. Hilgenbrink, A. R., and Low, P. S. (2005) Folate receptor-mediated drug targeting: from therapeutics to diagnostics. J. Pharm. Sci. 94, 2135–2146

43. Holm, J., Hansen, S. I., Hoier-Madsen, M., and Helkjaer, P. E. (1998) A folate binding protein in ascitic fluid, serum and ovarian tissue of patients with ovarian adenocarcinoma immuno-reacts with antibodies against human milk folate binding protein. Biosci. Rep. 18, 49–57

44. DaL MaSo, L., Augustin, L. S., Franceschi, S., Talamini, R., Polels, J. K., Kendall, C. W., Jenkins, D. J., Vidgen, E., and La Vecchia, C. (2004) Association between components of the insulin-like growth factor system and epithelial ovarian cancer risk. Oncology 67, 225–230

45. Oliveira, S., van Bergen en Henegouwen, P. M., Storm, G., and Schifflers, E. R. (2006) Molecular biology of epidermal growth factor receptor inhibition for cancer therapy. Expert Opin. Biol. Ther. 6, 605–617

46. Gagne, J. P., Gagne, P., Hunter, J. M., Bonicalzi, M. E., Lemay, J. F., Kelly, I., Le Page, C., Provencer, D., Mes-Masson, A. M., Droti, A., Bourgais, D., and Porier, G. G. (2003) Proteome profiling of human epithelial ovarian cancer cell line TOV-112D. Mol. Cell. Biochem. 275, 25–55

47. Debernardis, D., Stanzione, S., Ottoboni, C., Clerico, L., Mancuso, T., Parodi, S., and Russo, P. (1996) Endogenous tumour necrosis factor enhances topoisomerase II inhibitors activity in human ovarian cancer cell lines. J. Pharmacol. Exp. Ther. 279, 84–90

48. Kubbe, H., Hagemann, T., Szlosarek, P. W., Balkwill, F. R., and Wilson, J. L. (2005) The inflammatory cytokine tumour necrosis factor-alpha regulates chemokine receptor expression on ovarian cancer cells. Cancer Res. 65, 10355–10362

49. Zhao, R., and Houy, W. A. (2005) Hsp90: a chaperone for protein folding and gene regulation. Biochem. Cell Biol. 83, 703–710

50. Shin, B. K., Wang, H., Yim, A. M., Le Naour, F., Brichory, F., Jang, J. H., Zhao, R., Purwans, E., Tra, J., Michael, C. W., Misiek, D. E., and Hanash, S. M. (2003) Global profiling of the cell surface proteome of cancer cells uncovers an abundance of proteins with chaperone function. J. Biol. Chem. 278, 7607–7616

51. Yow, H. K., Wong, J. M., Chen, H. S., Lee, C. G., Davis, S., Steele, G. D., Jr., and Chen, L. B. (1988) Increased mRNA expression of a laminin-binding protein in human colon carcinoma: complete sequence of a full-length cDNA encoding the protein. Proc. Natl. Acad. Sci. U. S. A. 85, 6394–6398

52. Hirano, T., Kato, H., Maeda, M., Goy, Y., Shou, Y., Nakamura, M., Maeda, J., Yashima, K., Kato, Y., Akimoto, S., Ohira, T., Tsuobi, M., and Ikeda, N. (2005) Identification of postoperative adjuvant chemotherapy responders in non-small cell lung cancer by novel biomarker. Int. J. Cancer 117, 460–468

53. Stierum, R., Gaspari, M., Dommels, Y., Otata, T., Piuk, H., Jespersen, S., Vogels, J., Verhoeckx, O., Groten, J., and van Ommen, G. B. (2003) Proteome analysis reveals novel proteins associated with proliferation and differentiation of the colorectal cancer cell line Caco-2. Biochem. Biophys. Acta 1650, 73–91

54. Gough, P. J., Barton, J. K., Willc, P. R., Rychlewski, M., Dempsey, P. J., and Raines, E. W. (2004) A disintegrin and metalloproteinase 10-mediated cleavage and shedding regulates the cell surface expression of CXCR chemokine ligand 16. J. Immunol. 172, 3679–3685