Exploring the Sialiome Using Titanium Dioxide Chromatography and Mass Spectrometry*

Martin R. Larsen‡§, Søren S. Jensen‡, Lene A. Jakobsen‡, and Niels H. H. Heegaard¶

Strategies for biomarker discovery increasingly focus on biofluid protein and peptide expression patterns. Post-translational modifications contribute significantly to the pattern complexity and thereby increase the likelihood of obtaining specific biomarkers for diagnostics and disease monitoring. Glycosylation is a common post-translational modification that plays a role e.g. in cell adhesion and in cell-cell and receptor-ligand interactions. Abnormal protein glycosylation has important disease associations, and the glycoproteome is therefore a target for biomarker discovery. Here we present a simple and highly selective strategy for purification of sialic acid-containing glycopeptides (the sialiome) from complex peptide mixtures. The approach utilizes a high and selective affinity of sialic acids for titanium dioxide under specific buffer conditions. In combination with mass spectrometry we used this strategy to characterize the human plasma and saliva sialiomes where 192 and 97 glycosylation sites, respectively, were identified. Furthermore we illustrate the potential of this method in biomarker discovery. Molecular & Cellular Proteomics 6:1778–1787, 2007.

Protein glycosylation is among the most common post-translational modifications known in nature. Glycosylation is difficult to analyze by biochemical methods due to chemically very similar monosaccharide building blocks and pronounced heterogeneity and microheterogeneity of the carbohydrate chains with respect to branching patterns and monosaccharide composition. Glycoproteins containing N- and O-linked glycans are in general either secreted in soluble form from cells or tissues into body fluids or localized to the cell surface. O-Linked glycans are also observed intracellularly where they appear to participate in regulation of signaling pathways (1–3).

The glycan moieties of extracellular glycoproteins may stabilize the conformation of proteins and confer proteolytic resistance and influence protein turnover, receptor-ligand interactions, cell-cell signaling, and adhesion, but functions are in many instances unknown (4). However, it is well known that glycosylation patterns in chronic disease can be highly aber-
taining glycopeptides. In addition, we show for the first time a map of the sialome, defined as all the SA-containing glycosylated proteins, from plasma and saliva. Finally to illustrate the potential of the new approach in biomarker discovery, it is used to compare the differences of the plasma sialome of a control individual and a patient with advanced bladder cancer.

**EXPERIMENTAL PROCEDURES**

**Materials**

Modified trypsin was from Promega (Madison, WI). Fetuin and RNase B were from Sigma. Poros R2 and Poros Oligo R3 reversed phase material were from PerSeptive Biosystems (Framingham, MA). GE Loader tips were from Eppendorf (Hamburg, Germany). 2.5-Dihydroxybenzoic acid (DHB) was from Fluka (St. Louis, MO). 3M Empore™ C3 disk was from 3M Bioanalytical Technologies (St. Paul, MN). Syringes for HPLC loading were from Scientific Glass Engineering (Victoria, Australia). The water was from a Milli-Q system (Millipore, Bedford, MA). Titanium dioxide beads were obtained from a disassembled TiO₂ column (1350L250W046 Tiansphere, 5 μm, 250 × 4.6 mm) purchased from GL Sciences Inc. (Tokyo, Japan).

**Test Proteins**

A peptide mixture originating from tryptic digestions of 10 standard proteins was generated. The 10 standard proteins were: serum albumin (bovine), β-lactoglobulin (bovine), carbonic anhydrase (bovine), α-casein (bovine), ribonuclease B (bovine pancreas), alcohol dehydrogenase (bakers’ yeast), lysosome (chicken), α-amylase (Bacillus species), and fetuin (fetal calf serum). All were from Sigma. Each protein was dissolved in 50 mM ammonium bicarbonate, pH 7.8, including 10 mM DTT and incubated at 56 °C for 30 min. The proteins were subsequently alkylated using 40 mM iodoacetamide for 1 h at room temperature. The reaction was quenched with addition of 10 mM DTT, and the solution was digested using trypsin (2%, w/w) at 37 °C for 12 h. The peptide mixture was generated by mixing the peptides originating from the proteins in equal amounts (0.5 pmol/μl) and stored at −20 °C until further use.

**Alkaline Phosphatase Treatment**

To avoid co-purification of phosphorylated peptides, the tryptic peptides originating from fetuin and serum/plasma/saliva proteins were treated with 0.5 and 2 units of alkaline phosphatase, respectively, in 50 mM NH₄HCO₃, pH 7.8, at 37 °C for 2 h.

**Plasma**

Plasma samples from control individuals and bladder cancer (stage IV) patients were depleted of the six highest abundance proteins, albumin, IgG, IgA, haptoglobin, transferrin, and α₁-antitrypsin, using immobilized affinity-purified antibodies according to the manufacturer’s protocol (Agilent). The depleted plasma was concentrated using Microcon 10-kDa spin columns to 50 μl. The concentrated proteins were diluted with 50 μl of 50 mM NH₄HCO₃, reduced and alkylated as described above, and digested overnight with 20 μg of trypsin/200 μg of protein at 37 °C. Prior to purification of SA-containing glycopeptides with TiO₂ the samples were treated with alkaline phosphatase as described above.

The serum from the bladder cancer patient was obtained in connection with ongoing microarray studies of gene expression in bladder cancer patients (T. Ørntoft). Patients were staged according to international criteria, and plasma was secured and frozen at −20 °C until used.

**Saliva**

Saliva was collected more than 2 h after food intake in a test tube containing Complete™ protease inhibitors (Roche Diagnostics GmbH). The proteins were precipitated by adding 5 volumes of acetone followed by incubation for 1 h at −80 °C. The precipitated proteins were redissolved in 8 M urea and diluted to 1 mM urea using 100 mM NH₄HCO₃ followed by reduction and alkylation as described above. The proteins were digested overnight with 20 μg of trypsin/200 μg of protein at 37 °C. Prior to purification of SA-containing glycopeptides with TiO₂ the sample was treated with alkaline phosphatase as described above.

**18O Labeling of Tryptic Peptides**

The tryptic peptides were labeled as described previously (33). The incorporation efficiency and the 1:1 ratio were tested using MALDI MS.

**Purification of SA-containing Glycopeptides Using TiO₂ Microcolumns**

TiO₂ microcolumns were packed in GE Loader tips essentially as described previously (34) or in p10 micropipette tips (depending on the amount of material to be purified). Briefly a small plug of C₂₅ material was stapled out of a 3M Empore C₁₈ extraction disk (e.g., using an HPLC syringe needle or similar device) and placed at the end of the tip. TiO₂ beads were suspended in acetonitrile and packed on top of the C₁₈ disk using a 1-ml disposable syringe. Peptide mixtures were diluted five times in loading buffer (1 mM glycolic acid in 80% acetonitrile, 2–5% TFA) and loaded onto the TiO₂ microcolumn (2% TFA was used for the fetuin sample, whereas 5% TFA was applied for the standard peptide mixture and plasma and saliva peptide mixtures). The column was washed with 10 μl of loading buffer followed by 40 μl of washing buffer (80% acetonitrile, 2% TFA). The SA-containing glycopeptides were eluted using 20–40 μl of ammonium-water (10 μl of 25% ammonia solution in 490 μl of water), pH 11. A small aliquot of each of the eluates was acidified with 100% formic acid, purified using a Poros R3 reversed phase microcolumn, and analyzed by MALDI MS. The remaining eluate was lyophilized for N-glycosidase F digestion.

**Glycosidase Treatments**

**Neuraminidase**—D Dephosphorylated tryptic peptides originating from fetuin were lyophilized and redissolved in 50 mM NH₄ acetate, pH 6. Neuraminidase (Roche Diagnostics GmbH) was added to the peptide mixture (0.005 unit), and the sample was incubated overnight at 37 °C.

**N-Glyc.osidase F**—Glycosylated tryptic peptides purified by TiO₂ were lyophilized and redissolved in 40 μl of 50 mM NH₄HCO₃, pH 7.8, containing 0.5 unit of N-glycosidase F (Roche Diagnostics GmbH). To increase the efficiency of N-glycosidase F, sialidase A (Glyko, GK80040) (5 milliunits) and 1,6-fucosidase (Glyko, X5006) (10 milliunits) were added to the incubation buffer. The deglycosylation was performed for 2–12 h at 37 °C.

**Reversed Phase Microcolumn Purification**

Reversed phase microcolumn purification was performed as described previously (34).

**MALDI MS**

MALDI MS was performed using a Voyager STR (PerSeptive Biosystems, Framingham, MA) equipped with delayed extraction. Spectra were obtained in positive reflector or linear ion mode using an accelerating voltage of 20 kV. MALDI MS data analysis was performed using the MoverZ software (Genomic Solutions).
Identification of Sialic Acid-containing Peptides Using TiO₂

Nanoscale LC-MSMS

Nanoscale liquid chromatography-tandem mass spectrometry was performed using a Q-TOF Ultima mass spectrometer (Waters/Micromass UK Ltd., Manchester, UK) or an LTQ-FT instrument (hybrid two-dimensional linear quadrupole ion trap-FTICR mass spectrometer (Thermo Electron)). A nano-flow HPLC system (Ultimate, Switchos2, Famos, Dionex/LC Packings, Amsterdam, The Netherlands) was used for chromatographic separation of the peptide mixture prior to MS detection on the Ultima Q-TOF system, whereas an Agilent 1100 nanoflow LC system was used for peptide separation prior to detection on the LTQ-FT instrument. The peptides were concentrated and desalted on a 1.5-cm precolumn (75-μm inner diameter, 360-μm outer diameter, ReproSil-Pur C₁₈ AQ 3 μm (Dr. Maisch)) and eluted at 200 nL/min by an increasing concentration of acetonitrile (1%/min gradient) onto a 8-cm analytical column (50-μm inner diameter, 360-μm outer diameter, ReproSil-Pur C₁₈ AQ 3 μm). The mass spectrometers were operated in the data-dependent mode selecting three parent ions for collision-induced dissociation per full scan.

Database Searching

Raw data files from the Q-TOF Ultima instrument were processed into pkl files using the ProteinLynx program. On each MSMS spectrum the background was subtracted (40%), and smoothing was performed (Savitzky-Golay; iteration, 2; window, 3 channels). In addition deisotoping was performed using the following parameters: minimum peak width, four channels; centroid top, 80%; TOF resolution, 10,000; NP multiplier (number of pushes correction factor), 0.7. Raw data from the LTQ-FT instrument were processed using the DTA SuperCharge (version 1.07) program (SourceForge) and converted into Mascot generic format (mgf) files according to the protocol. Deisotoping was performed using the software default settings. Database searching was performed using the pkl and mgf files using an in-house Mascot database search program (version 2.1). The searched databases were National Center for Biotechnology Information non-redundant (NCBI-nr, March 26, 2007 (4,761,919 sequences; 1,643,098,755 residues), human) and International Protein Index (IPI-human human_20060517 (64,191 sequences; 27,188,565 residues)) and International Protein Index (IPI-human human_20060517 (64,191 sequences; 27,188,565 residues). The database searches were performed with fixed modification as carbamidomethyl (Cys) and variable modifications as oxidation (Met) and deamidation (Asn) or deamidation in [18O]water. Enzyme specificity was selected to semitrypsin. For the data obtained using the FT-MS instrument a mass accuracy of 10 ppm was used on the parent ion, and 0.6 Da was used on the fragment ions. For data obtained on the Q-TOF Ultima instrument a mass accuracy of 70 ppm was used on the parent ion, and 0.2 Da was used on the fragment ions.

All identified glycosylated peptides were manually evaluated according to the following criteria: peptides had to include peptide sequence tags (4–5 amino acids) assigned by abundant Y- or B-ions from the higher mass area, and if prolines were present in the sequence an abundant fragment ion had to be assigned to the predominant fragmentation N-terminal to the proline residue. In addition, the charge states were evaluated according to the number of basic charges in the sequence.

Annotated spectra from each identified peptide have been included in the supplemental data. Annotated spectra were generated either in the MSQuant program (SourceForge) or from Mascot data including Y- and B-ion masses that have been assigned to the peptide.

RESULTS AND DISCUSSION

Principle of the Method—Previously we published a highly selective strategy for purification of phosphorylated peptides using TiO₂ microcolumns (34). We found that peptide loading onto a TiO₂ matrix under highly acidic conditions in a solution containing DHB or phthalic acid resulted in a selective purification of phosphorylated peptides with little co-purification of non-phosphorylated peptides. We proposed that the adsorption of phosphate anions to the surface of TiO₂ is mediated by a bridging bidentate binding (Fig. 1A). This type of binding is much stronger than the chelating bidentate binding that DHB can make with TiO₂ (Fig. 1B), which would mimic the binding that acidic amino acids would be able to make to TiO₂.

The highly negatively charged SA, which contains both carboxylic acid and hydroxyl functionalities (Fig. 1C), presumably interacts with TiO₂ molecules via a multipoint binding, similar to a multidentate binding (35) (Fig. 1D), in which the carboxylic acid and the hydroxyl groups contribute to the binding. Therefore, SA either free or present at the non-reducing end of glycans on glycopeptides should very efficiently be retained on TiO₂ (or ZrO₂) under conditions where non-modified peptides or neutral glycopeptides will not bind, similar to the interaction of phosphorylated peptides with TiO₂ (or ZrO₂). Therefore, in combination with phosphatase treatment to remove the phosphorylated peptides that otherwise would saturate the TiO₂ column, this should provide a purified fraction of sialylated peptides.

To examine this scheme, we tested the purification of SA-containing N-linked glycopeptides from bovine fetuin. Fetuin contains 359 amino acid residues and three N-linked and four O-linked glycans. All the N-linked sites contain SA. Here we only show results from the analysis of the N-linked sites where
Identification of Sialic Acid-containing Peptides Using TiO₂

the most efficient deglycosylation enzymes are available. Fetuin is also phosphorylated at several serine residues, and these groups were removed prior to the TiO₂ chromatography to ensure optimal binding of the SA-containing glycopeptides. A MALDI MS peptide mass map of the raw mixture of tryptic peptides from fetuin is shown in Fig. 2A. Several peptide signals can be detected including some in the higher mass range (>3000 Da), which is known to belong to SA-containing glycopeptides. An aliquot (5 pmol) of the dephosphorylated tryptic peptide mixture from fetuin was diluted in 5% TFA, 80% ACN and applied onto a TiO₂ microcolumn. After washing and elution the purified peptides were analyzed by MALDI MS (Fig. 2B). Here an enrichment of SA-containing glycopeptides was observed due to the signals in the higher mass range; however, a significant amount of signals originating from nonspecific binding was observed. An aliquot (10 pmol) of the dephosphorylated tryptic peptide mixture from fetuin was diluted in the loading buffer containing 1 M glycolic acid and applied onto the TiO₂ microcolumn. After washing and elution, half of the eluted peptides were analyzed by MALDI MS in linear positive ion mode (Fig. 2C). This peptide mass map is dominated by peptide signals in the high mass range >3000 Da indicating efficient purification of the glycosylated peptides. The remaining eluate was lyophilized and subsequently treated with N-glycosidase F for deglycosylation. The deglycosylated peptides were analyzed by MALDI MS in linear positive ion mode (Fig. 2D). In this peptide mass map only four signals were detected, corresponding to peptides expected to carry N-linked glycans (Table I). No non-glycosylated peptides were detected, indicating a highly efficient and selective purification of the glycopeptides. The selectivity can also be achieved using other acids with functionalities similar to glycolic acid (e.g. DHB, phthalic acid, gallic acid, etc.).

Specificity toward Sialic Acid-containing Glycopeptides—To assess the specificity of TiO₂ for SA-containing glycopeptides versus neutral glycopeptides under the specified loading conditions, dephosphorylated tryptic peptides from fetuin were labeled with ¹⁸O by using ¹⁸O-buffer during trypsinization. An aliquot of the normal ¹⁶O-containing peptides was treated with neuraminidase to cleave off SA and mixed with the ¹⁸O-labeled peptide mixture containing intact SA-carrying glycopeptides in a 1:1 ratio. This mixture was subjected to TiO₂ chromatography. The eluted glycopeptides were deglycosylated using N-glycosidase F and subsequently analyzed by MALDI MS in reflector positive ion mode (Fig. 2E). Again signals from all four N-linked glycosylated peptides were detected (Table I). More than 90% of the signals detected in this MALDI mass spectrum originated from the ¹⁸O-labeled peptides (Table I). Because a 50:50% distribution of the heavy and light peptides would be expected with no SA selectivity of TiO₂, this result indicates a highly selective purification of the SA-containing glycopeptides over neutral glycopeptides. The ¹⁶O-glycopeptides appearing in this experiment.
Identification of Sialic Acid-containing Peptides Using TiO₂

Identification of SA-containing Glycopeptides from Proteins in Biofluids—After establishing the method using a single model glycoprotein, the method was tested on one of the most complex biofluids used in proteomics, i.e. plasma. Our ultimate aim is to completely characterize the sialiome in plasma, i.e. all the SA-containing glycoproteins in plasma.

Plasma was depleted for the six most abundant proteins. An aliquot (1%) of the dephosphorylated tryptic peptides originating from 60 μl of depleted plasma was analyzed by MALDI MS (Fig. 3A). The remaining peptides were submitted to TiO₂ chromatography. The SA-containing glycopeptides were profiled by MALDI MS as shown (Fig. 3B) using 5% of the eluate from the TiO₂ chromatography for the analysis. In the higher mass range (>3000 Da) broad signals are observed indicating the heterogeneous SA-containing glycopeptides, whereas only a very few signals are observed in the lower mass area mainly originating from phosphopeptides that have escaped the phosphatase treatment. The large amount of glycopeptides present together with the often pronounced microheterogeneity observed on each glycosylation site will result in a multitude of signals from SA-containing glycopeptides explaining the badly resolved material in the MALDI mass spectrum of Fig. 3B. After treating the remaining eluate from the TiO₂ chromatography with N-glycosidase F, an aliquot (5%) was analyzed by MALDI MS (Fig. 3C). All the ion signals in the high mass area disappeared after this treatment, and new peaks were confined to the lower mass range (<4000 Da) indicating an efficient purification of SA-containing glycopeptides from depleted human plasma. The remaining deglycosylated peptide eluate was analyzed by LC-MSMS.

Supplemental Table S1 shows a list of the SA-containing glycopeptides and their corresponding glycosylation sites identified using 60 μl of depleted plasma as starting material. The list includes only non-redundant peptides. If peptide sequences are shared by more than one protein entry only one of the protein entries is included in the list. A total of 192 SA-containing glycosylation sites in 100 proteins were identified in this way. Most of the proteins are secreted proteins. However, some are membrane proteins presumably shed from plasma membranes from cells or tissues. A total of 127

<table>
<thead>
<tr>
<th>Peptide sequence*</th>
<th>Monoisotopic mass (M + H)⁺</th>
<th>Average mass (M + H)⁺</th>
<th>¹⁶O:¹⁸O ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLDPTPLAN*CSVR</td>
<td>1442.8</td>
<td>1443.7</td>
<td>3:97</td>
</tr>
<tr>
<td>LCPDCPLLAPLN*DSR</td>
<td>1741.8</td>
<td>1743.0</td>
<td>7:93</td>
</tr>
<tr>
<td>VVHAVEALATFNAESN*GSYLQVEISR</td>
<td>3017.5</td>
<td>3019.4</td>
<td>8:92</td>
</tr>
<tr>
<td>RPTGEVYDIEIDTLETTCVLDPTPLAN*CVR</td>
<td>3672.7</td>
<td>3675.1</td>
<td>4:96</td>
</tr>
</tbody>
</table>

* N* denotes the site of glycan attachment.

Fig. 3. Characterization of the sialiome from depleted human plasma. Human plasma was depleted for the six most abundant proteins (albumin, IgG, IgA, haptoglobin, transferrin, and α₁-antitrypsin) using an antibody affinity depletion column. An aliquot (1%) of the dephosphorylated tryptic peptides originating from 60 μl of depleted plasma was analyzed by MALDI MS (A). B illustrates the MALDI MS peptide mass map of 5% of the eluate obtained from the TiO₂ purification of the remaining peptide mixture. SA-containing glycopeptides are shown in the high mass region of the spectrum. The eluate from the TiO₂ purification was treated with N-glycosidase F (PNGase F), and 5% was analyzed by MALDI MS (C). Signals in the low mass range of the spectrum are shown here. All spectra were obtained in linear positive ion mode.
non-glycosylated peptides (that did not have an NX(S/T/C) consensus site) were identified due to co-purification with the SA-containing peptides, indicating a purification efficiency of 60%. Of the 192 glycosylation sites, 28 were listed in the Swiss-Prot and TrEMBL databases as potential glycosylation sites. These sites are now verified in human plasma. Fig. 4A shows a pie chart describing the location of the sialylated proteins identified in plasma according to the Swiss-Prot database. As expected, the majority of the identified proteins are secreted proteins; however, a significant number of the proteins localize to membranes and could originate from tissue leakage.

In a recent much more time-consuming and complex approach using immunofinity subtraction, hydrazide chemistry, strong cation exchange fractionation, reversed phase chromatography, and mass spectrometry a total of 639 glycosylation sites were identified using 16 times more starting material than used in this study (36). Hydrazine chemistry targets all glycosylated peptides, and we therefore expect more sites using this method than the present SA-specific method.

Identification of SA-containing Glycopeptides in Proteins in Human Saliva—Human saliva is secreted from multiple salivary glands including parotid, submandibular, sublingual, and other minor glands lying beneath the oral mucosa. Saliva is a potential rewarding source of disease biomarkers. Whereas many proteins have been identified in saliva, very little is known about the modification status of those proteins. Recently 128 N-glycosylated peptides were identified from 45 unique N-glycoproteins in saliva using hydradzine chemistry in combination with in-solution isoelectric focusing and mass spectrometry (37).

Here we collected saliva from a healthy individual directly into a protease inhibitor solution. The proteins were precipitated using ice-cold acetone, and the redissolved proteins were digested using trypsin and dephosphorylated using alkaline phosphatase. The SA-containing glycopeptides were isolated using the described protocol, and the deglycosylated peptides were identified by LC-MSMS. Supplemental Table S2 shows the identified SA-containing glycopeptides and their glycosylation sites. A total of 97 glycosylation sites were found in 45 proteins. The list includes only non-redundant peptides. If peptide sequences are shared by more than one protein entry only one of the protein entries is included in the list. In addition to the glycosylated peptides, we identified 17 non-glycosylated peptides that did not contain an N-linked glycosylation consensus sequence, indicating a purification efficiency of 85% for saliva. Of the 97 sites, 29 sites were new sites according to the Swiss-Prot and TrEMBL databases. In addition, a number of hypothetical proteins with no known function were found to be glycosylated. The pie chart showing the location distribution of the sialylated proteins identified in saliva is shown in Fig. 4B. As expected the majority of the proteins are also secreted; however, almost 20% originate from cellular surfaces, i.e. transmembrane proteins and glycosylphosphatidylinositol-anchored proteins.

Application of SA-specific TiO₂ Chromatography to the Discovery of SA-containing Glycopeptide Profiles in Plasma—The sialiome of plasma from a healthy individual was compared with plasma from a patient with advanced stage bladder cancer to validate the potential for applying this method in biomarker discovery. The two samples were immunoaffinity-depleted for the six most abundant plasma proteins and then reduced, S-alkylated, trypsinized, and dephosphorylated exactly as described above for normal plasma. The tryptic peptides were lyophilized and redigested using trypsin in normal ¹⁶O-buffer (control) and ¹⁸O-buffer (cancer) (Fig. 5A). Using this strategy most of the internal tryptic peptides originating from the cancer sample were labeled with two heavy oxygen molecules, resulting in a mass increase of 4 Da for those peptides. Inadequate labeling can be observed if an acidic amino acid is located close to the tryptic cleavage site. The control and cancer peptides were mixed in a 1:1 ratio, and the SA-containing glycopeptides were purified by TiO₂ chromatography. Half of the eluate from the TiO₂ purification was analyzed directly by LC-MSMS to obtain mass and fragmentation information on the intact glycosylated pep-
Identification of Sialic Acid-containing Peptides Using TiO₂

**Fig. 5.** Characterization of SA-containing glycopeptides from plasma from healthy and bladder cancer individual. The strategy used to differentially display the peptides originating from tryptic digestion control and cancer depleted plasma using ¹⁸O labeling is shown in A. After labeling of the cancer peptide mixture with ¹⁸O, the two peptide mixtures were mixed in a ratio 1:1, and the SA-containing glycopeptides were purified using the TiO₂ method. Half of the eluate from the TiO₂ purification was analyzed directly by LC-MSMS. The LC-MS ion trace is shown in B. The mass spectrum of the SA-containing glycopeptides eluting at time point 1 (1; 40 min) is shown in C. The masses of the glycopeptides carrying various glycan structures are labeled. The enlargement of the doubly charged signal at m/z 1173/1175 showing an equal degree of sialylation is illustrated in D. The mass spectrum of the SA-containing glycopeptides eluting at time point 2 (2; 60 min) is shown in E. The masses of the glycopeptides carrying various glycan structures are labeled. The enlargement of the triply charged signal at m/z 1423.92/1425.26 showing a marked increase in peptide amount in the cancer sample is illustrated in F.

tides. The LC-MSMS ion chromatogram is shown in Fig. 5B. Material eluting at two different times (marked 1 and 2 in the figure) was chosen to illustrate the complexity of the sample and to directly compare some SA-containing glycopeptides from the healthy sample and the cancer sample.

The mass spectrum of the SA-containing glycopeptide eluting after 40 min (time point 1) and 60 min (time point 2), respectively, are shown in Fig. 5, C and E. Here all the mass-labeled peaks represent doubly charged species and illustrate a unique glycopeptide with various glycan structures attached at each time point resulting in a significant heterogeneity. An enlargement of the mass area around the signal at 1173 Da at time point 1 (Fig. 5D) shows a doubly charged peak with 2-Da spacing where the light peak originates from the healthy sample and the heavy peak originates from the cancer sample. The intensity of the two peaks is very similar indicating that no quantitative change can be observed between the normal sample and the cancer sample for this SA-containing glycopeptide. In contrast, the SA-containing glycopeptide eluting at time point 2 was found to be more abundant in the glycopeptide fraction from cancer plasma as illustrated in Fig. 5F by the intensity of the triply charged heavy signal compared with the light fraction signal in the enlargement of the material at m/z 1424. Obviously the increase in the signal could be due to more sialylation of the peptides as well as more of a given sialylated glycoprotein in the cancer sample.

To identify the peptide sequence of the more abundant sialylated glycopeptide, the fragment ion spectrum of the triply charged glycopeptide signal at 1425.26 (Fig. 5F, heavy labeled) was interrogated. The fragment ion spectrum is shown in Fig. 6A.

The fragment ions can be assigned to the loss of single
Identification of Sialic Acid-containing Peptides Using TiO₂

monosaccharides from the glycopeptide, indicating that the charge is predominantly retained on the peptide (Y-type fragments (23)). Three oxonium ions are present that are diagnostic for SA-containing glycans (m/z 274.07, 292.09, and 657.23 (38)). The mass of the peptide can be deduced when taking into account the common core structure of N-linked glycans, i.e. (Man)₂(GlcNAc)₃ (illustrated in the figure as losses of 3 × 162 and 2 × 203 Da, respectively) (for further information see Ref. 39). The mass of the naked peptide was found to be m/z 1412.84.

The sequence of this peptide was determined after deglycosylation of the remaining SA-containing glycopeptides from the TiO₂ purification and subsequent LC-MSMS analysis of the deglycosylated peptides. The fragment ion spectrum of the peptide at m/z 1413.84 is shown in Fig. 6B (note that deglycosylation of N-glycosylation with N-glycosidase F results in conversion of asparagine to aspartic acid and a mass increase of 1 Da). From these data the sequence VPVPITN*ATLDR could be read where N* indicates the glycan attachment site. This sequence belongs to α₁-acid glycoprotein 2, also known as orosomucoid 2 (Swiss-Prot entry Q5T538), which is a well known acute-phase reactant with increased synthesis in response to inflammation and tissue damage. As many cancers arise from sites of infection, chronic irritation, and inflammation (e.g. Ref. 40) an increase in acute-phase proteins may correlate with the tumor progression. Because α₁-acid glycoprotein is one of the most prominent acute-phase reactants and is known to exhibit pronounced microheterogeneous glycan structures shifting in the normal and the acute-phase form of the protein, it is very likely that the findings are due to an increased concentration of the protein in this sample compared with a normal control. Also altered glycan antennary structures (leading to changes in SA content) associated with different microheterogeneity profiles might very well contribute to the results even though the largest effect would be due to entirely new sites being glycosylated in diseased orosomucoid. This, however, to our knowledge has not been reported.

Conclusion—We describe a highly selective method that uses TiO₂ chromatography for purification and characterization of SA-containing glycopeptides. The outcome of the present study clearly illustrates the potential of using TiO₂ chromatography in combination with sensitive and high resolution mass spectrometry for detection and quantification of SA-containing glycopeptides from dephosphorylated complex mixtures. We believe that we would achieve a similar enrichment using other oxidized metal surfaces such as ZrO₂.

The highly acidic buffer conditions in combination with the DHB, phthalic acid, or glycolic acid was efficiently used to selectively purify SA-containing glycopeptides from both simple and highly complex samples such as human plasma depleted for the six most abundant proteins. Only a very little co-purification of non-modified peptides was detected from complex mixtures. We speculate that this selectivity is a result of the capability of SA to form a multipoint binding to the TiO₂ surface, whereas acidic non-glycosylated peptides and neutral glycopeptides are eliminated by the competitive binding of the substituted aromatic acid (DHB, phthalic acid, or glycolic acid) to the TiO₂ surface. Other acids that resemble DHB, phthalic acid, or glycolic with respect to the functional groups could also be included in the loading procedure to exclude non-sialylated peptides. This approach should also be applicable to other SA-containing compounds, e.g. free
Identification of Sialic Acid-containing Peptides Using TiO₂

glycans. Contamination from sulfated glycostructures cannot be ruled out using the current method, and more work is needed to examine the binding of such glycan structures to TiO₂.

The method described here has been applied to characterization of the sialome of depleted human plasma and saliva. In total, we have identified 192 and 97 SA-containing glycosylation sites in proteins from depleted human plasma and saliva, respectively. Experiments with differential display of SA-containing glycopeptides from a healthy and a cancer depleted plasma sample have shown that some SA-containing glycoproteins are differentially expressed/sialylated in cancer compared with the healthy individual. Significant changes were observed in some of the acute-phase proteins that are expected to be changed because the plasma originated from a patient with advanced bladder cancer.

Besides being simple, fast, and efficient in enrichment of SA-containing glycopeptides from complex biofluids, the presented method is extremely tolerant toward salts and other low molecular weight contaminants such as detergents. Combining the method with prepurification methods either on the protein or peptide level, such as isoelectric focusing of proteins or peptides, would most likely increase the number of identified SA-containing glycosylation sites. The presented results also clearly prove the concept of using this method for discovery of potential SA-containing disease biomarkers. In further work the choice of appropriate controls, large sample sets, and stringent validation will need to be applied for this aspect to be developed in a useful way, but the simplicity of the procedure ensures that methodological obstacles will not be prohibitive for such development. We therefore believe that the method could find wide application in glycoproteomics research and biomarker discovery.

* This work was supported by the European Foundation for the Study of Diabetes (to M. R. L.), Danish Natural Science Research Council Grant 21-03-0167, and The Danish Strategic Research Council Grant 21-03-0167, and The Danish Strategic Research Council Grant 21-03-0167, and The Danish Strategic Research Council Grant 21-03-0167 and The Danish Strategic Research Council Grant 21-03-0167 and The Danish Strategic Research Council Grant 21-03-0167.

REFERENCES

Identification of Sialic Acid-containing Peptides Using TiO₂


