Human Urinary Glycoproteomics; Attachment Site Specific Analysis of N- and O-Linked Glycosylations by CID and ECD*§

Adnan Halim‡, Jonas Nilsson‡, Ulla Rüetschi‡, Camilla Hesse‡, and Göran Larson‡§

Urine is a complex mixture of proteins and waste products and a challenging biological fluid for biomarker discovery. Previous proteomic studies have identified more than 2800 urinary proteins but analyses aimed at unraveling glycan structures and glycosylation sites of urinary glycoproteins are lacking. Glycoproteomic characterization remains difficult because of the complexity of glycan structures found mainly on asparagine (N-linked) or serine/threonine (O-linked) residues. We have developed a glycoproteomic approach that combines efficient purification of urinary glycoproteins/glycopeptides with complementary MS-fragmentation techniques for glycopeptide analysis. Starting from clinical sample size, we eliminated interfering urinary compounds by dialysis and concentrated the purified urinary proteins by lyophilization. Sialylated urinary glycoproteins were conjugated to a solid support by hydrazide chemistry and trypsin digested. Desialylated glycopeptides, released through mild acid hydrolysis, were characterized by tandem MS experiments utilizing collision induced dissociation (CID) and electron capture dissociation fragmentation techniques. In CID-MS², Hex₅HexNAc₄-N-Asn and HexHexNAc-O-Ser/Thr were typically observed, in agreement with known N-linked biantennary complex-type and O-linked core 1-like structures, respectively. Additional glycoforms for specific N- and O-linked glycopeptides were also identified, e.g. tetra-antennary N-glycans and fucosylated core 2-like O-glycans. Subsequent CID-MS³, of selected fragment-ions from the CID-MS² analysis, generated peptide specific b- and γ-ions that were used for peptide identification. In total, 58 N- and 63 O-linked glycopeptides from 53 glycoproteins were characterized with respect to glycan- and peptide sequences. The combination of CID and electron capture dissociation techniques allowed for the exact identification of Ser/Thr attachment site(s) for 40 of 57 putative O-glycosylation sites. We defined 29 O-glycosylation sites which have, to our knowledge, not been previously reported. This is the first study of human urinary glycoproteins where “intact” glycopeptides were studied, i.e. the presence of glycans and their attachment sites were proven without doubt. Molecular & Cellular Proteomics 11: 10.1074/mcp.M111.013649, 1–17, 2012.

In search of disease biomarkers, urine qualifies as an important biologic fluid that can easily be collected by repeated and noninvasive sampling from single individuals. Proteins present in urine are derived not only from glomerular ultrafiltration of plasma but also from tubular secretion of soluble proteins, detachment of glycosylphosphatidylinositol anchored proteins and exosome shedding through the urothelium (1). For healthy individuals, 30% of the urinary proteome has been estimated to originate from the plasma filtrate whereas the remaining 70% is believed to be derived from the kidneys and the urothelium (2). Until 2005, ~800 urinary proteins had been identified by various proteomic approaches (3–7). In 2006, a comprehensive proteomic study identified more than 1500 proteins from healthy human urine samples, simultaneously reflecting the complexity and the potential information concealed in the urinary proteome (8). In 2009, Kentsis et al. reported the hitherto largest data set for the urinary proteome, unveiling more than 2300 protein identities (9). The “core urinary proteome” was recently defined as a common set of nearly 600 urinary proteins with a dynamic concentration range spanning five orders of magnitude (10). Interestingly, the authors also reported that the 20 most abundant proteins, which were estimated to constitute 2/3 of the core urinary proteome by mass, were glycoproteins with serum albumin being the only exception.

Glycoproteins are characterized by the presence of oligosaccharides linked to the peptide backbone primarily through N- or O-glycosidic bonds at asparagine or serine/threonine residues, respectively (11). N- and mucin-type O-glycosylations are widely accepted as the most common and structurally diverse post-translational modifications found on secreted proteins and on the extracellular parts of membrane bound proteins (12). Given that protein glycosylation is involved in various cellular processes (13–16), the site-specific characterization of N- and O-linked glycosylations and identification of the modified proteins is becoming increasingly important. Urine is potentially a rich source for N- and O-linked glycoproteins derived from renal- and distal organs and represents an interesting subproteome for structural charac-

From the ‡Department of Clinical Chemistry and Transfusion Medicine, Institute of Biomedicine, Sahlgrenska Academy at the University of Gothenburg, Sweden
Received August 24, 2011, and in revised form, December 9, 2011
Published, MCP Papers in Press, December 14, 2011, DOI 10.1074/mcp.M111.013649
CID and ECD of N- and O-glycosylated Urinary Glycoproteins

Characterization of human glycoproteins. However, glycoproteomic characterization of urine is lacking and only a few proteomic studies aimed at identifying urinary glycoproteins have been reported (17–20). In these studies, the glycan moieties were either cleaved off or not studied at all. It is, however, important to analyze qualitative glycan differences in glycoproteomes because changes associated with the carbohydrate moieties may reflect physiological status (21–23). Perhaps more importantly for the urinary proteome, the study of intact glycopeptides could reveal not only the glycoprotein origin but potentially also provide information regarding pathological changes of its original tissue (24, 25). By analyzing tryptic glycopeptides originating from urinary glycoproteins both the glycan structures and glycosylation sites of proteins may be addressed. However, a highly purified mixture of glycopeptides is the prerequisite for such studies because of the general phenomena of ion suppression and stoichiometric effects in the mass spectrometric analysis of complex mixtures (26–28).

Enrichment methods for the isolation of formerly N-linked glycopeptides from biological sources have been described using hydrazide chemistry, TiO₂ affinity purification, lectin chromatography and hydrophilic interaction liquid chromatography (HILIC) (29–33). The N-glycans are typically removed by PNGase F treatment during these protocols and the site-specific information of N-glycan structures is usually not addressed. Only a few glycoproteomic studies, aimed at analyzing intact N-glycopeptides from biological samples, have been published (34, 35). Also, by comparison to N-glycosylation, characterization of protein O-glycosylation is analytically more challenging for several reasons, e.g. due to the heterogeneity associated with O-glycan core structures (36). Although collision-induced dissociation (CID)¹-based MSⁿ strategies are well capable of revealing both O-glycan- and peptide sequences for intact glycopeptides (37) the site-specific information of the modified amino acid is however usually lost. This is because of predominant glycosidic fragmentation of the precursor during MSⁿ, and peptide fragmentation occurring mainly for the deglycosylated peptide ion in the MS³. Additionally, the exact glycosylation site of identified peptides containing several Ser/Thr residues cannot be predicted due to the lack of a consensus sequence for mucin-type O-glycosylation. The alternative fragmentation techniques electron capture dissociation (ECD) (38, 39) and electron transfer dissociation (ETD) (40) have been introduced for site-specific analysis of CID-labile PTMs but characterization of protein O-glycosylations using ECD/ETD have generally been limited to synthetic glycopeptides or single glycoproteins (41–45). Thus, investigation of protein O-glycosylation has lagged behind and relatively little is known about O-linked glycans with respect to their protein carriers and amino acid attachment sites. Recently, Darula and Medzihradszky used lectin enrichment with jacalin, recognizing core 1 O-glycans (Galβ3–1-3GalNAcα–O-Ser/Thr), and identified 21 O-glycosylation sites from bovine serum glycoproteins by combining ETD and exoglycosidase digestion (46). We have previously developed a sialic acid specific capture-and-release protocol for the enrichment of both N- and O-glycosylated peptides from sialylated glycoproteins in biological samples using hydrazide chemistry (37). Only CID based characterization was employed in our previous study and assignment of O-glycan attachment sites was therefore not possible for most O-glycosylated peptides. The low sensitivity and fragmentation yield for ECD/ETD compared with CID make it advantageous to use highly enriched samples of O-glycosylated peptides. We tested the sialic acid capture-and-release protocol on human serum samples but, as expected, N-glycosylated peptides completely dominated the LC-MS/MS chromatograms (Halim et al., unpublished). We then turned our attention to urine, with ambitions to characterize N- and O-glycosylated peptides, since urine also may serve as a sample source for biomedical diagnosis. However, because urine contains much salts and pigments, which could interfere with the periodate oxidation step in our protocol, we first developed a simple method to remove low-molecular waste products and attain pure protein samples suitable for redox chemistry and proteomics purposes. In this study, we have thus extended our protocol (Fig. 1 and supplemental Fig. S1) to include a unique dialysis procedure for isolation of human urinary proteins prior to the sialic acid capture-and-release method. In addition to the CID-based approach, we also included ECD for the characterization of O-glycan attachment sites and as a complementary peptide fragmentation mode for the identification of urinary glycopeptides.

EXPERIMENTAL PROCEDURES

Collection and Preparation of Human Urine—First morning, midstream urine was obtained from a healthy male individual during five consecutive days and prepared separately. Immediately after collection, 50 ml de-identified urine was separated from intact cells and debris by centrifugation at 3000 × g, 4 °C for 20 min. The uppermost 20 ml were frozen at −20 °C and used for further analysis. Routine clinical chemistry analyses of all five samples were all within the reference range (U-Albumin (<5.4 mg/L), U-Creatinine (mean 17.4 mmol/L: range 12–28 mmol/L), U-Bilirubin, U-Urobilinogen, U-Acetone, U-Glucose, U-Erythrocytes, U-Leukocytes, U-nitrite were all negative).

After thawing, 10 ml of each sample was dialyzed against 14 × 2 L of tap water at 4 °C using Spectra/Pol MWCO 12–14 kDa (Spectrum Laboratories) for 7 days (Fig. 1). The urine samples were lyophilized, dissolved in 6 ml 5% sodium-dodecyl sulfate (SDS) and dialyzed against 2 × 2 L of 1.5% SDS at 60 °C for 24 h. The SDS was subsequently removed by dialysis against 2 × 2 L Milli Q deionized H₂O (dH₂O) at room temperature for 24 h. Finally, the samples were lyophilized and dissolved in 0.5 ml dH₂O. Protein content was determined using the BCA-1 protein assay (Sigma-Aldrich) on a NanoDrop

¹ The abbreviations used are: Con A, concanavalin A; dHex, deoxyhexose; ECD, electron capture dissociation; Gal, galactose; GalNAc, N-acetylgalactosamine; GlcNAc, N-acetylglucosamine; Hex, hexose; HexNAc, N-acetylhexosamine; PNGase F, peptide N-glycosidase F; WGA, wheat germ agglutinin.
1000 spectrophotometer (Thermo Scientific) according to the manufacturer’s protocol.

**Protein Separation**—For protein separation prior to in-gel trypsin digestion 80 μg of urinary proteins were dissolved in NuPage LDS-sample buffer (Invitrogen, Carlsbad, CA) supplemented with 50 mM dithiothreitol, reduced and denatured at 70 °C for 10 min. Protein samples were then separated on 4–12% Bis-Tris precasted polyacrylamide gels (Invitrogen). SeeBlue Plus2 pre-stained standard (Invitrogen) was used as molecular weight marker and proteins were visualized by Coomassie colloidal blue staining. For in-gel trypsin digestion one gel lane was divided into 15 equally sized gel slices and subjected to automated trypsin digestion (supplemental Fig. S1A) on a BioMek 2000 work station equipped with a vacuum manifold. 96-well plates supplemented with a 7 μl volume of C18 reversed phase chromatographic resin were used for vacuum filtration and sample clean-up. The work-flow essentially followed the protocol previously described (47) except that the peptide extraction was performed twice with 0.2% trifluoroacetic acid to allow for peptide binding to the C18 resin of the filter plates. Finally, peptides were eluted twice in 40 μl of 60% acetonitrile in 0.1% trifluoroacetic acid and the eluted fractions were evaporated to dryness in a vacuum centrifuge. Prior to liquid chromatography/tandem MS (LC-MS/MS) analysis samples were redissolved in 0.1% formic acid.

For electrophoretic analysis of repeatedly dialyzed urine samples, 30 μg of urinary proteins were denatured by heating (100 °C, 5 min) in 1% SDS and 100 mM dithiothreitol and separated on a 4–12% Bis-Tris precasted polyacrylamide gel (Invitrogen). SeeBlue Plus2 prestained standard (Invitrogen) was used as molecular weight marker and proteins were visualized by Coomassie colloidal blue staining (supplemental Fig. S2C).

**Glycopeptide Enrichment Procedure**

**Hydrazide Capture**—Capture of sialylated glycoproteins to hydrazide beads (supplemental Fig. S1B) was done as previously described (37) with minor modifications. One hundred μg protein in 1 ml dH2O was oxidized with 2 mM periodic acid for 15 min at 0 °C. The reaction was quenched by the addition of 5 μl 99% glycerol and buffer exchanged to 2.5 ml coupling buffer (100 mM acetate, 150 mM NaCl, pH 5.2) (36) using Sephadex PD-10 columns (GE Healthcare). One hundred μl hydrazide beads (Bio-Rad) in coupling buffer was added and agitated for 16 h at room temperature in the dark. The beads were then washed with 3 × 3 ml 0.1% Tween 20 in PBS, pH 7.4 and finally with 2 × 3 ml of 50 mM NH4HCO3, pH 8.0.

**Reduction, Alkylation and Trypsin Digestion**—The glycoproteins captured on the beads were then incubated with 0.3 ml 10 mM dithiothreitol for 1 h at 37 °C in the dark. Following a washing step (50 mM NH4HCO3, pH 8.0), 0.3 ml 55 mM iodoacetamide (Sigma Aldrich) was added and incubated for 30 min at room temperature and in the dark. The beads were then washed with 2 × 3 ml of 8 M urea, 50 mM NH4HCO3, pH 8.0 and with 2 × 3 ml of 1% SDS in dH2O with gentle agitation. Finally, five washing steps with 3 ml of 50 mM NH4HCO3, pH 8.0 were performed. Captured glycoproteins were digested with 1 μg sequencing grade porcine trypsin (Promega, Madison, WI) in 70 μl 50 mM NH4HCO3, pH 8.0, at 37 °C for 18 h. The trypsin-released peptides were transferred to pre lubricated eppendorf tubes (Costar). Any remaining peptides were extracted once with 100 μl 50% acetonitrile, pooled and lyophilized together with the trypsin released peptides and subjected to mass spectrometric analysis (supplemental Fig. S1B).

**Release of Glycopeptides**—The beads were initially washed once with 3 ml of 50% acetonitrile in dH2O, once with 3 ml dH2O and once with 3 ml 1.5 M NaCl in dH2O. The beads were then washed 3 × 3 ml dH2O, 2 × 3 ml 50% acetonitrile in dH2O, 2 × 3 ml with 25% acetonitrile in dH2O and finally with 2 × 3 ml dH2O. One hundred μl 0.1 M formic acid was added to the beads and incubated for 1 h at 80 °C (supplemental Fig. S1C). The released glycopeptides were transferred to pre lubricated eppendorf tubes (Costar, Cambridge, MA). Any remaining glycopeptides were extracted once with 50 μl 50% acetonitrile in dH2O, pooled and lyophilized together with the formic acid released glycopeptides and subjected to mass spectrometric analysis.

**LC-MS/MS Analysis**—Tryptic peptides, obtained either from in-gel digestion of electrophoretically separated urinary proteins (supplemental Fig. S1A), from unglycosylated peptides released by trypsin digestion of hydrazide captured glycoproteins (supplemental Fig. S1B) or glycopeptides released through formic acid hydrolysis (supplemental Fig. S1C) were separated by reversed phase chromatography on a 15 cm capillary column (Zorbax SB300 C18, 0.075 mm ID). Peptides/glycopeptides were reconstituted in 40 μl 0.1% formic acid, 20 μl was loaded onto the column in eluent A (0.1% formic acid) and separated with a linear gradient from 5% to 60% eluent B (84% acetonitrile in 0.1% formic acid) at a flow rate of 250–300 nL/min. Gradient lengths were either 50 min, for the analysis of the peptide fraction, or 150 min, for the glycopeptide fraction and the in-gel digested fractions. The LC system (Ettan MDLC, GE Healthcare) was coupled in-line with a LTQ-FTICR instrument (Thermo Fisher Scientific) via a nanoelectrospray source (Thermo Fisher Scientific). The source was operated at 1.4 kV, with no sheath gas flow and with the ion transfer tube at 200 °C. The mass spectrometer was programmed for acquisition in a data dependent mode. The survey scans were acquired in the FTICR mass analyzer and covered the m/z range 300–2000. For the analysis of peptides the seven most intense peaks in each full mass scan, with charge state ≥2 and intensity above a threshold of 100, were selected for fragmentation in the linear ion trap (LTQ) by CID. Glycopeptides were analyzed with two independent methods, one based on CID fragmentation and the other on ECD fragmentation. For the CID method the most intense peak in each FTICR full scan was selected for fragmentation in the linear ion trap (LTQ) followed by subsequent selection and fragmentation of the five most intense MS2 fragment ions. For the ECD method the two most intense peaks in each FTICR full scan was selected for fragmentation in the ICR cell. CID fragmentation was performed with normalized collision energy of 35% activation, q = 0.25, activation time of 30 ms and three microscans. ECD fragmentation was performed with a relative energy of 4 and 5 in subsequent scans and a duration of 70 ms and three microscans. For all fragmentation events dynamic exclusion was enabled with a repeat count of 2. Peaks selected for fragmentation more than twice within a 30 s interval were excluded from selection (20 ppm window) for 180 s and the maximum number of excluded peaks was 200. AGC settings were 1000000 (FTMS full scan), 30000 (ion trap), 10000 (ion trap MS2), and 500000 (FTMS ECD).

**Data Analysis**

**Protein Identification**—Raw data containing centroid MS/MS spectra, from the analysis of tryptic peptides, were converted into .dta format by the Bioworks software (version 3.3.1) utility extract.msn (Thermo Fisher Scientific) and analyzed with an in-house version of the Mascot software (Mascot ver. 2.3.01, http://www.matrixscience.com). Search parameters were set as follows: peptide tolerance, 10 ppm; MS/MS tolerance, 0.5 Da; enzyme, trypsin, one missed cleavage allowed; fixed carbamidomethyl modification of cysteine; variable oxidation of methionine; database, IPI human version 3.72 (86,392 sequences). Fragment ions from the b- and y-series, including losses of ammonia or water, were used for scoring. Minimal requirement for each protein identification was two unique peptide hits with scores above the significance threshold (p < 0.05).

**Protein Clustering**—Mascot results, including information on identified proteins and peptides, were imported into the ProteinCenter database.
**CID and ECD of N- and O-glycosylated Urinary Glycoproteins**

software (Proxeon Bioinformatics). Data was filtered so that each identified protein contained at least two unique peptides and identified proteins were clustered, based on peptide sharing, into groups of indistinguishable proteins. Lists of protein identifiers from two independent studies (8, 9) were also imported into the ProteinCenter software and comparisons of the three data sets were performed.

**Glycopeptide Characterization Using CID**—Glycopeptide identification and relative quantification of N- and O-glycan microheterogeneity was done as previously described (37). N- and O-linked glycan sequences were manually verified in CID-MS² spectra for each glycopeptide by tracing peaks corresponding to the loss of individual monosaccharides. Manually selected MS² spectra, corresponding to the fragmentation of unmodified peptides for O-glycopeptides, were individually converted to .mzXML format via the Readaw application (http://www.proteomecenter.org). Each .mzXML file was individually visualized with the mm toss (version 2.4) application (48) and searched with the Mascot algorithm. The peptide monoisotopic mass was manually defined for each search by subtracting the monoisotopic mass of the glycan from the FTICR-MS1 measured precursor. Search parameters were set as follows: peptide tolerance, 10 ppm; MS/MS tolerance, 0.6 Da; enzyme, trypsin, one missed cleavage allowed; fixed carbamidomethyl modification of cysteine; variable oxidation of methionine and variable loss of NH₃ (-17.0266 Da) at N-terminal cysteine and glutamine; taxonomy, human, 20,259 sequences (protein entries); database, SwissProt 101005. Peptides were considered as positive identifications if the ion score was above the significance threshold (p < 0.05). For MS² spectra that did not yield positive identifications, in the above described procedure, the peak list of individual glycopeptides were manually exported from the mMass application as .txt files and analyzed with an in-house version of the Mascot software (Mascot version 2.3.01, www.matrixscience.com). The precursor mass was manually defined in each .txt file so that it would match the monoisotopic mass of the peptide as described above. Enzyme specificity was set to semi-trypsin or no enzyme to account for peptides with a single or no tryptic sites, respectively.

Finally, variable phosphorylation at serine or threonine residues was used in selected cases. All CID-MS² spectra that resulted in positive identifications were also converted to .mgf files according to the same procedures as above and Mascot searched against a decoy database (taxonomy, human, 20,245 sequences (protein entries); database, SwissProt 110817) using the same search parameters as described for N-linked glycopeptides above.

**Glycopeptide Characterization Using ECD**—The precursor ion masses of ECD spectra were matched to precursor ion masses of glycopeptides that had been identified by the CID-MS² approach. Peak lists of c, (c - 1), z and (z+1)-ions were prepared for candidate glycopeptides using the MS-product tool (http://prospector.ucsf.edu). Glycopeptide identifications were verified and O-glycan attachment sites were pinpointed manually to unique Ser/Thr residues by tracing c- and z-ion peaks that contained or lacked the anticipated glycan(s). Also, the Mascot distiller program (version 2.3.2.0, Matrix Science) was used for peak picking and to prepare Mascot files from the ECD spectra. Subsequent MS² spectra at relative energy 4 and 5 were aggregated and the ions presented as singly protonated in the output Mascot files. Search parameters were set as follows: peptide tolerance, 10 ppm; MS/MS tolerance, 0.03 Da; enzyme, trypsin, one missed cleavage allowed; fixed carbamidomethyl modification of cysteine; variable modification of HexHexNAc (365.1322 Da), Hex.HexNAc (730.2644 Da) and dHex.HexNAc2 (876.3223 Da) of serine, threonine and tyrosine; variable Hex.HexNAc (1622.5816 Da) modification of asparagine; variable oxidation of methionine and variable loss of NH₃ (-17.0266 Da) at N-terminal cysteine and glutamine; taxonomy, human, 20,259 sequences; database, SwissProt 101005. Instrument was set to match 1+ ions of the c, z and z+1 series (c, z+1 and z+2 using Mascot terminology). We did not observe any y-ions and these were thus not considered in the scoring. Acceptance criteria for a positive identification was based on scoring above the significance threshold value (p < 0.05). The Mascot files were analyzed with the in-house version of the Mascot software (Mascot version 2.3.01).

**RESULTS**

**Protein Yields and Identifications**—Starting from 10 ml urine we used dialysis against water to remove salts and pigments but this was found to yield inadequate sample purity. However, after a second dialysis against 1.5% SDS at 60 °C the procedure was satisfactorily efficient in removing pigments (Fig. 1 and supplemental Fig. S2). We recovered 31 ± 10 μg/ml protein (mean ± SD) from the dialyzed urine samples. One dialyzed urine sample was analyzed by GeLC-MS/MS (supplemental Fig. S1A). Applying the criteria of at least two uniquely identified peptides per identified protein, we identified 989 urinary proteins that were grouped into 413 protein groups of indistinguishable proteins by clustering based on peptide sharing (Supplementary excel Table, Gel-based proteomics). Following hydrazide capture (supplemental Fig. S1B), 63 proteins were either identified only from peptides found in the tryptic digests of captured proteins (n = 10), only from the covalently linked glycopeptides released through acid hydrolysis (n = 36) (supplemental Fig. S1C) or from both of these procedures (n = 17). Thus, 53 glycopeptid proteins could be identified solely based on the identification of unique glycopeptides and for 17 of those glycopeptides the identities were also supported by peptide identifications (supplemental Table S1 and supplemental Fig. S3). Altogether, 26 urinary glycoproteins were identified from 122 un-
Identification of O-Linked Glycopeptides by CID—We identified 63 glycopeptides, corresponding to 49 differently O-glycosylated peptides originating from 40 urinary glycoproteins (0.0% false positive identifications). These are presented, together with their identified O-linked glycans, their attachment sites and Mascot scores of individual glycopeptides in Table I. Annotated CID-MS² and ECD spectra for each O-glycopeptide is presented in supplemental Fig. S5. The relative abundance of specific glycoforms at each O-glycan attachment site are listed in supplemental Table S2. Typical CID-MS² experiments for three O-linked glycopeptides constituting the same tryptic peptide are presented in Figs. 2A–2C to illustrate our strategies for glycan fragmentation analysis and manual identification of O-linked glycopeptides. Doubly (dashed line) and triply (solid line) protonated precursor ions of co-eluting glycoforms in the ion chromatograms (Fig. 2E and supplemental Fig. S4) were accurately mass measured (±10 ppm) in the ICR cell. CID-MS² of the HexHexNAc₂ glycoform resulted in fragmentation of the glycan part into Y-type ions (nomenclature according to Domon and Costello (50) and a B-type ion corresponding to the [HexHexNAc⁺H]⁺ oxonium ion at m/z 366 (Fig. 2A). The most abundant Y-type ions were frequently observed as charge reduced species, e.g. at m/z 1010.4 and m/z 929.8 (Fig. 2A and supplemental Fig. S5). The neutral loss of one and two Hex residues (m/z 741.7 and m/z 687.8, respectively) followed by the loss of one HexNAc residue (m/z 620.3 and m/z 929.8) and finally the loss of the final HexNAc residue (m/z 828.4) demonstrated the HexHexNAc₂ composition. For HexHexNAc₂ containing glycopeptides we could not distinguish two separate HexHexNAc-O-Ser/Thr core 1-like glycans from one Hex(2HexHexNAc)HexNAc-O-Ser/Thr core 2-like structure solely based on the Y-type ions. In CID-MS³ and MS⁵, these glycoforms could be differentiated by the presence of glycan fragments (B and internal B/Y-type ions) exceeding the HexHexNAc (m/z 366) composition, e.g. by the presence of diagnostic ions at m/z 407 corresponding to an internal HexNAcHexNAc fragment (51). The CID-MS³ spectrum of the HexHexNAc₂ glycoform in Fig. 2A did not contain a fragment ion at m/z 407, suggesting that two separate core 1-like glycans occupied two individual Ser/Thr residues within the glycopeptide. Conversely, in other cases core 2-like glycans were indeed identified (Fig. 3, see below). For the HexHexNAc₂ glycoform (Fig. 2A) the intact peptide ion (Y₀⁻ ion) was observed as the fifth most intense ion (for z ≥ 2 ions) at m/z 828.4 and peptide fragmentation was obtained in the final CID-MS⁵ spectrum. The HexHexNAc₂ glycoform was the next glycopeptide that eluted (m/z 741.7, Fig. 2E) and the CID-MS³ spectrum (Fig. 2B) showed an intense charge reduced fragment ion at m/z 929.1 corresponding to the loss of HexHexNAc and a proton from the precursor ion. Additional charge reduced fragment ions at m/z 1010.6 and 827.7 showed the loss of HexNAc and HexHexNAc₂, respectively. CID-MS⁴ of Y₀ at m/z 827.7 resulted in peptide fragmentation.
| UniProtKB accession | Glycoprotein                        | Peptide sequence | Attachment site | Glycan       | Mascot score | Mascot threshold | ECD   |
|---------------------|------------------------------------|------------------|----------------|-------------|--------------|-----------------|-------|
| **O-linked Glycopeptides** |                                    |                  |                |             |              |                 |       |
| P02765              | Alpha-2-HS-glycoprotein            | R.TVQPSVGAAGPVVPPCPGR.I | 346           | HexHexNAc   | 25           | >18             | Yes   |
| P02666              | Apolipoprotein C-III               | D.PEVRPTSJVAA.-- | 94             | HexHexNAc   | 19           | >16             | Yes   |
| Q13790              | Apolipoprotein F                   | K.DANISOPETKEGL.R | 256           | HexHexNAc   | 57           | >21             | Yes   |
| P98160              | Basement membrane-specific heparan sulfate proteoglycan core protein | R.AYDGLSLPEDIETFASQMR.W | 42            | HexHexNAc   | 93           | >20             | Yes   |
| P26842              | CD27 antigen                       | D.PLLPSLTAR.S    | 127           | HexHexNAc   | 16           | >14             | Yes   |
| P16070              | CD44 antigen                       | S.EGEGANTTSQGIR.T | 637–638       | HexHexNAc   | 43           | >33             | Yes   |
| P00742              | Coagulation factor X               | R.SVAQTASSGEAPDSITWKPYYAADLD.P | 183–203 | HexHexNAc   | 41           | >38             | No    |
| P00742              | Coagulation factor X               | K.SHAPEVTSSPLK   | 476–485       | HexHexNAc   | 37           | >33             | No    |
| P39069              | Collagen-alpha-1(XV) chain         | E.ILEAVTDASQP.K.E | 265           | HexHexNAc   | 68           | >32             | Yes   |
| P10643              | Complement component C7            | N.PLLQAVPK.C     | 696           | HexHexNAc   | 39           | >29             | No    |
| Q68CJ9              | Cyclic AMP-responsive element-binding protein 3-like protein 3 | R.VAADAVPGSEAPGPRPEADITR.E | 379           | HexHexNAc   | 47           | >21             | Yes   |
| Q13508              | Ecto-ADP-ribosyltransferase 3      | K.SQGNINNTPGVPVPGPK.S | 346           | HexHexNAc   | 59           | >20             | Yes   |
| P98095              | Fibulin-2                          | R.AEAGARPENILDAQATS.R.S | 347–348       | HexHexNAc   | 54           | >24             | No    |
| Q8NFU4              | Follcular dendritic cell secreted peptide | R.RNFPPIPSAPTTPITPLSE.K | 75–83         | HexHexNAc   | 64           | >31             | No    |
| P78423              | Fractalkine                        | K.AQDGPVGTLEF.R | 183           | HexHexNAc   | 79           | >22             | Yes   |
| P78423              | Fractalkine                        | R.WGGQGQSRPFNSL.R.E | 253           | HexHexNAc   | 35           | >21             | Yes   |
| P78423              | Fractalkine                        | R.LGLUVTPVDQAAI.R.R | 329–338       | HexHexNAc   | 43           | >13             | No    |
| P04921              | Glycoporphin-C                     | D.PGMSWPDGR.M    | 42            | HexHexNAc   | 59           | >30             | No    |
| Q8TDQ0              | Hepatitis A virus cellular receptor 2 | R.DFTAAFP.R.M | 145           | HexHexNAc   | 32           | >21             | No    |
| P04233              | HLA class II histocompatibility antigen gamma chain | S.LEQKPTDAPPK.V | 203           | HexHexNAc   | 37           | >32             | No    |
| P04233              | HLA class II histocompatibility antigen gamma chain | D.PSSGLGTVKQDLGPVP.M.- | 281–287       | HexHexNAc   | 42           | >40             | No    |
| P01344              | Insulin-like growth factor II      | R.DVSTPPTYLPDNPFR.Y | 96           | HexHexNAc   | 67           | >21             | Yes   |
| P01344              | Insulin-like growth factor II      | R.DVSTPPTYLPDNPFR.Y | 96           | HexHexNAc   | 63           | >21             | Yes   |
| P01344              | Insulin-like growth factor II      | R.DVSTPPTYLPDNPFR.Y | 96           | HexHexNAc   | 39           | >21             | Yes   |
| P01344              | Insulin-like growth factor II      | P.UALTDQ.T.P | 163           | HexHexNAc   | (38)         | >42             | No    |
| P19823              | Inter-alpha-trypsin inhibitor heavy chain H2 | K.WPDSTPSWANPSPPVITSM.L.A | 665–679       | HexHexNAc   | (31)         | >33             | No    |
| Q14624              | Inter-alpha-trypsin inhibitor heavy chain H2 | K.IEETTMITQTAPIPAIQPASAILPQGSVER.L | 720           | HexHexNAc   | 49           | >18             | Yes   |
| Q14624              | Inter-alpha-trypsin inhibitor heavy chain H4 | K.IEETTMITQTAPIPAIQPASAILPQGSVER.L | 722–723       | HexHexNAc   | 30           | >17             | Yes   |
| Q14624              | Inter-alpha-trypsin inhibitor heavy chain H4 | K.IEETTMITQTAPIPAIQPASAILPQGSVER.L | 719–725       | HexHexNAc   | 30           | >17             | Yes   |
| O95988              | Interleukin-18-binding protein     | D.PCPSGPPVFPAAQ.K | 53            | HexHexNAc   | 17           | >17             | No    |
| O95988              | Interleukin-18-binding protein     | D.PCPSGPPVFPAAQ.K | 53            | HexHexNAc   | 60           | >33             | Yes   |
| P09603              | Macrophage colony-stimulating factor 1 | K.GQQADVATOTALPR.V | 363           | HexHexNAc   | 57           | >20             | No    |
| Uniprot/KB accession | Glycoprotein                                      | Peptide sequence | Attachment site | Glycan       | Mascot score | Mascot threshold | ECD |
|----------------------|--------------------------------------------------|------------------|----------------|--------------|--------------|------------------|-----|
| P09603               | Macrophage colony-stimulating factor 1           | R.ISSLRPQGGSNPSTLSAQPQLSR.S | 406–426         | HexHexNAc   | 55           | >22              | No  |
| Q13361               | Microfibrillar-associated protein 5              | D.PATDETVALA.V   | 54             | HexNAc      | (37)         | >40              | Yes |
| Q13361               | Microfibrillar-associated protein 5              | D.PATDETVALA.V   | 54             | HexHexNAc   | 42           | >42              | Yes |
| Q13361               | Microfibrillar-associated protein 5              | D.PATDETVALA.V   | 54             | HexHexNAc   | 42           | >38              | No  |
| Q6UXB8               | Peptidase inhibitor 16                          | E.LQATLDHTGHTSSK.S | 386–395        | HexHexNAc   | 34           | >33              | No  |
| Q96FE7               | Phosphoinositide-3-kinase-interacting protein 1  | R.EQOTSPAPGLR.C  | 39             | HexHexNAc   | 51           | >20              | Yes |
| P05155               | Plasma protease C1 inhibitor                     | K.VATTVISK.M     | 47–48          | HexHexNAc   | 21           | >8               | Yes |
| P05155               | Plasma protease C1 inhibitor                     | K.VATTVISK.M     | 47–48          | HexHexNAc   | 16           | >8               | Yes |
| P05154               | Plasma serine protease inhibitor                 | R.VEDLHVGATAPRSSR | 39           | HexHexNAc   | 66           | >20              | Yes |
| P01133               | Pro-epidermal growth factor                      | K.NQQVPTDLSK.T   | 801–807        | HexHexNAc   | 47           | >20              | No  |
| P01133               | Pro-epidermal growth factor                      | R.LSEPGLICPDSTPPPLR.E | 954–955       | HexHexNAc   | 57           | >20              | Yes |
| P01133               | Pro-epidermal growth factor                      | R.LSEPGLICPDSTPPPLR.E | 954–955       | HexHexNAc   | 72           | >20              | Yes |
| Q99075               | Proheparin-binding EGF-like growth factor        | D.PPTVSTDQPLLPGGR,D | 44           | HexHexNAc   | 86           | >31              | Yes |
| Q99075               | Proheparin-binding EGF-like growth factor        | D.PPTVSTDQPLLPGGR,D | 44, 47       | HexHexNAc   | 61           | >31              | Yes |
| Q9UHG2               | ProSAAS                                           | R.GLSAASPPLEGTAPPR.R | 53           | HexHexNAc   | 60           | >21              | Yes |
| Q9UHG2               | ProSAAS                                           | R.AADHDVGELPPGVLGALL.R.V | 228         | HexHexNAc   | 84           | >20              | Yes |
| Q9UHG2               | ProSAAS                                           | K.RLETPAPQPR.R   | 247           | HexHexNAc   | 42           | >17              | Yes |
| P80370               | Protein delta homolog 1                          | R.ALSQPQFV.T.L   | 256           | HexHexNAc   | 30           | >27              | Yes |
| Q9ULJ9               | Protein HEG homolog 1                            | R.EPPFPVR.PR.R   | 67            | HexHexNAc   | 25           | >23              | No  |
| Q9GZM5               | Protein YIPF3                                     | K.AAVATLQSH.-    | 346           | HexHexNAc   | 43           | >19              | Yes |
| Q9GZM5               | Protein YIPF3                                     | K.AAVATLQSH.-    | 346           | HexHexNAc   | 32           | >19              | Yes |
| Q9GZM5               | Protein YIPF3                                     | K.AAVATLQSH.-    | 346           | HexHexNAc   | 43           | >21              | Yes |
| Q9GZM5               | Protein YIPF3                                     | K.AAVATLQSH.-    | 346           | HexHexNAc   | 33           | >20              | Yes |
| Q9GZM5               | Protein YIPF3                                     | K.AAVATLQSH.-    | 346           | dHexHexNac   | 8           | >18              | No  |
| Q16849               | Receptor-type tyrosine-protein phosphatase-like N | K.AARPPVTPV.LLE.K | 441           | HexHexNAc   | 21           | >20              | Yes |
| Q4LDE5               | Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1 | Y.DDFLDTVQETATSIGNAK.S | 887–894 | HexHexNAc | 119 | >33 | No |
| P34741               | Syndecan-2                                        | K.IPAQTKSPEETDK.E | 101           | HexHexNAc   | 30           | >25              | Yes |
| Q6UWD8               | Transmembrane protein C16orf54                   | M.PLTPPEPSGR.V   | 4       | HexHexNAc   | 30           | >22              | Yes |
| P25445               | Tumor necrosis factor receptor superfamily member 6 | A.QVDINSK,G     | 28           | HexHexNAc   | 30           | >29              | Yes |
| Q9UFP1               | Protein FAM198A                                   | D.PGPMEPGQGAGAPTHIR.Q | 53–58       | HexHexNAc   | 77           | >41              | No  |
| P04070               | Vitamin K-dependent protein C                     | G.TPAPLDSVFS.SER.A | 19           | HexHexNAc   | 67           | >34              | Yes |
| P04070               | Vitamin K-dependent protein C                     | G.TPAPLDSVFS.SER.A + Phosphorylation | 19           | HexHexNAc   | 49           | >39              | Yes |
| Uniprot/KB accession | Glycoprotein                      | Peptide sequence         | Attachment site | Glycan                      | Mascot score | Mascot threshold* | ECD |
|----------------------|----------------------------------|--------------------------|-----------------|-----------------------------|--------------|-------------------|-----|
| P02763               | Alpha-1-acid glycoprotein 1      | N.LVPV/PI1NATLDQITGK.W   | 33              | Hex$_3$HexNAc$_4$           | 33           | >17               | No  |
| P19652               | Alpha-1-acid glycoprotein 2      | N.LVPV/PI1NATLDR.I       | 33              | Hex$_3$HexNAc$_4$           | 22           | >17               | No  |
| P01009               | Alpha-1-antitrypsin              | R.QLA/HS2NSTIFFSPVSIATAFAMLSLGTK.A | 70       | Hex$_3$HexNAc$_4$           | 19           | >14               | No  |
| P01009               | Alpha-1-antitrypsin              | K.YLG/NA1AIFGFLPGK.L     | 271             | Hex$_3$HexNAc$_4$           | 31           | >15               | No  |
| P02765               | Alpha-2-HS-glycoprotein          | K.VCQDCPLAP1NDTR.V       | 156             | Hex$_3$HexNAc$_4$           | 26           | >15               | No  |
| P02765               | Alpha-2-HS-glycoprotein          | F.NAQ/NGSNFQLEIS.R       | 176             | Hex$_3$HexNAc$_4$           | 35           | >27               | No  |
| P05090               | Apolipoprotein D                 | R.CIOANYS.VL.MENG.I      | 65              | Hex$_3$HexNAc$_4$           | 36           | >8                | No  |
| P05090               | Apolipoprotein D                 | R.ADGTV.MQIEATPVLLEPAK.L | 98              | dHex$_1$Hex$_3$HexNAc$_4$  | 18           | >10               | No  |
| Q96F04               | Carboxypeptidase B2              | C.SVLLADV.DS.IQQQSH.DTVSPR.A | 108        | Hex$_3$HexNAc$_4$           | 20           | >19               | No  |
| P01876               | Ig alpha-1 chain C region        | R.PALE/LLGSEAN.TCTTGLR.D | 144             | Hex$_3$HexNAc$_4$           | 31           | 23                | No  |
| O95998               | Interleukin-18-binding protein   | R.FPNFSILY.WL.GNFSFIEHLPGR.L | 103        | Hex$_3$HexNAc$_4$           | 10           | >4                | No  |
| O95998               | Interleukin-18-binding protein   | K.ALVLE/QT.PAL1SNT/NSCVL.DPEQVDQQR.H | 147     | dHex$_1$Hex$_3$HexNAc$_4$  | 19           | >11               | No  |
| Q96F87               | Phosphoinositide-3-kinase-        | R.CUNVLUDAQ.SG.LA.SPGNS.GHNSYCR.N | 66         | dHex$_1$Hex$_3$HexNAc$_4$  | 36           | >10               | No  |
| P05155               | Plasma protease C1 inhibitor     | S.NP/NTSSSSDOPESLQD.R.G  | 25              | Hex$_3$HexNAc$_4$           | 69           | >22               | No  |
| P15151               | Poliovirus receptor              | R.VEDE/NGYTCI.FVTPQPS.G.S | 120            | Hex$_3$HexNAc$_4$           | 28           | >7                | No  |
| P41222               | Prostaglandin-H2 H$_2$-isomerase  | K.SWAPAT.GGLN.LSTFLR.K | 78              | dHex$_1$Hex$_3$HexNAc$_4$  | 13           | >6                | No  |
| P02760               | Protein AMBP                      | K.WN/ITESV.HIN/TY.DYE.HFL.TK.K | 115        | Hex$_3$HexNAc$_4$           | 15           | >13               | No  |
| P02760               | Protein AMBP                      | R.YFY/NTSMACF.EQ.Q       | 250             | Hex$_3$HexNAc$_4$           | 21           | >13               | No  |
| P00734               | Prothrombin                      | R.GHV/INTR.S             | 121             | Hex$_3$HexNAc$_4$           | 19           | >4                | Yes |
| P00734               | Prothrombin                      | R.YPHKPEI/STFPHGQ.LQENFCR.N | 143        | Hex$_3$HexNAc$_4$           | 30$^d$       | >18$^d$           | Yes |
| P07911               | Uromodulin                       | R.CNTAAP.WM.GH.PSHDEGIVSR.K | 132       | Hex$_3$HexNAc$_4$           | 17           | >15               | No  |
| P07911               | Uromodulin                       | K.QDFN.TD.LS.LHRE.R.L   | 322             | dHex$_1$Hex$_3$HexNAc$_4$  | 13           | >10               | No  |
| P07911               | Uromodulin                       | R.NETHYAT.SLY.L          | 396             | Hex$_3$HexNAc$_4$           | 19           | >19               | No  |
| O6EMK4               | Vasinor                          | R.LHEYE/NETF.EQ.G       | 117             | Hex$_3$HexNAc$_4$           | (11)$^d$     | >15$^d$           | Yes |
| P25311               | Zinc-alpha-2-glycoprotein        | R.FGCEIENR.S            | 128             | Hex$_3$HexNAc$_4$           | 14           | >11               | No  |

* Site occupancy reported in the UniProtKB/Swiss-Prot database, glycan unknown.

* Site occupancy and glycan not reported in the UniProtKB/Swiss-Prot database.

* For N-glycan microheterogeneity, see supplementary Table III.

* Mascot score and threshold values obtained for ECD data.
CID and ECD of N- and O-glycosylated Urinary Glycoproteins

Fig. 2. LTQ-FTICR mass spectrometry of urinary O-linked glycopeptides derived from Insulin-like growth factor II. A, CID-MS² spectrum of the Hex₃HexNAc₂ glycoform (m/z 795.7093²). B, CID-MS² spectrum of the HexHexNAc₂ glycoform (m/z 741.6934⁴). C, CID-MS² spectrum of the HexHexNAc₂ glycoform (m/z 673.9967²). D, CID-MS² spectrum of the unmodified peptide at m/z 828.4⁺ from panel A. E, Extracted base peak chromatograms showing the elution profile and intensity of triply charged (solid line) and doubly charged (dashed line) parent ions. F, CID-MS² spectrum of the unmodified peptide at m/z 827.8⁺ from panel C. G, ECD-MS² spectrum of the triply charged Hex₃HexNAc₂ glycoform (m/z 795.7085²) showing the effective dissociation of the precursor, which revealed the novel glycosylation site at Thr²⁶. H, ECD-MS² spectrum of the triply charged HexHexNAc₂ glycoform (m/z 741.6914⁴). I, ECD-MS² spectrum of the triply charged HexHexNAc₂ glycoform (m/z 673.9979²), which confirms the previously reported glycosylation site at Thr²⁶. The isolated ions subjected to CID-MS²/ECD-MS² fragmentation are boxed and schematically illustrated in each panel. Circle, Hex; square, HexNAc; bold line, D³⁶-STPPTVLPDNFPR peptide. Potential hexasaccharide rearrangements products are depicted with asterisk.

(see below). The Y-type fragment ion at m/z 687.5 corresponding to [peptide+HexNAc₂+3H]⁺ showed that two HexNAc residues were attached to the peptide but did not reveal if they were located on individual Ser/Thr or linked in a core 2-like manner. Again, a diagnostic [HexNAcHexNAc+H]⁺ ion at m/z 407 was not observed, indicating that the HexNAc residues were located on separate Ser/Thr residues. Approximately 1 min later the HexHexNAc glycoform eluted (m/z 671.0 in Fig. 2E) and the CID-MS² spectrum (Fig. 2C) showed intense Y-ions corresponding to the loss of Hex (m/z 620.1 and m/z 929.3) and HexHexNAc (m/z 552.6 and m/z 827.8) from the precursor. CID-MS³ fragmentation of the peptide ion (Yc⁻ion) at m/z 828.4 (Fig. 2A) and m/z 827.8 (Fig. 2C) resulted in b- and y-ions, shown in Figs. 2D and 2F, which were used for peptide identification through the Mascot algorithm. The CID-MS³ spectra of the Yc⁻ions in Figs. 2A–2C (m/z 828) were all matched to the tryptic D³⁶-STPPTVLPDNFPR⁰⁷ peptide of Insulin-like growth factor II (IGF-II, UniProt/KB accession P01344) with ion scores of 67 (p < 0.05 threshold; >26), 39 (p < 0.05 threshold; >21) and 63 (p < 0.05 threshold; >21) for the HexHexNAc₂ (Fig. 2D), HexHexNAc₂ (not shown), and HexHexNAc (Fig. 2F) glycoform, respectively (Table I).

Assignment of Glycan Attachment Sites by ECD—We also acquired ECD-MS² spectra of the triply charged D³⁶-STPPTVLPDNFPR⁰⁷ glycopeptides from IGF-II, with HexHexNAc₂ (Fig. 2G), HexHexNAc₂ (Fig. 2H) and HexHexNAc (Fig. 2I) glycans. Fragmentation of triply charged precursors generated sufficient c- and z-ions to be used for glycosylation site identification purposes. For the ECD-MS² of the triply charged HexHexNAc₂ glycoform (Fig. 2G), the c₅⁻ion was observed without the additional mass of any glycan (m/z 319.16) indicating that Ser²⁶ was not modified. The c₁⁻ion, however, was detected with the additional mass of HexHexNAc₂ (m/z 1445.62) showing that the glycan(s) had to reside within the Thr²⁶-Pro-Pro-Thr²⁶ sequence. The cyclic structure of proline precludes ECD induced N-terminal cleavage and ca, ca, z₁₀, and z₁₁ ions can thus not be observed. The only fragment ions that can resolve the glycan attachment site(s) are therefore za and ca. A glycosylated ca fragment was indeed observed at m/z 979.44 (Fig. 2G), which showed that
CID and ECD of N- and O-glycosylated Urinary Glycoproteins

One glycosylation site on the Thr346 residue was mapped for Thr346 harbored a single HexHexNAc. The c7 fragment was observed at m/z 1445.62, which mapped the second HexHexNAc to Thr96. The glycan sequence, determined as two separate HexHexNAc-O-Ser/Thr structures by CID-MS2 (Fig. 2A), was thus mapped by ECD-MS2 (Fig. 2G) to two individual amino acids, i.e., Thr96 and Thr99 of IGF-II. The ECD-MS2 spectrum of the triply charged HexHexNAc2 glycoform (Fig. 2H) allowed us to verify the peptide sequence and the presence of a HexHexNAc moiety within the Asp93-Val94-Ser95-Thr96-Pro97-Pro98-Thr99 region. However, we did not detect any fragment ions that could differentiate whether Thr96 or Thr99 was modified with the single HexNAc. For the HexHexNAc glycoform (Fig. 2f) the c7 ion was once again observed without the additional mass of the carbohydrate, showing that Ser95 was not modified. Furthermore, the c7 was detected at m/z 614.31 and was thus not glycosylated and showed that Thr96 was not the glycosylation site. In contrast, the c7 ion was detected with the additional mass of HexNAc at m/z 1080.49, thereby pinpointing the glycosylation site to Thr99 as previously described (52). Taken together, these experiments also revealed the site occupancy (macroheterogeneity) within the D35VSTPP7VLDPNPFPR107 trypsin glycopeptide, i.e., the initial HexHexNAc glycosylation occurs at Thr99 whereas the second HexHexNAc is attached to Thr96.

Fig. 3. Microheterogeneity of the A342VAVTLQSH350 O-linked glycopeptide from urinary protein YIPFS. (A) CID-MS2 (m/z 564.79872) and (B) ECD-MS2 spectra (m/z 564.79872) of the HexNAc glycoform which pinpoint the novel glycosylation site to Thr99. The isolated ions subjected to CID-MS2/ECD-MS2 of the same fucosylated glycoform (m/z 828.39252) spectrum showed that the entire glycan moiety is attached to Thr99 whereas the second HexHexNAc is attached to Thr96.

O-linked Glycopeptide Microheterogeneity and Modifications—In a few instances, also other glycoforms apart from excess HexNAc-O-Ser/Thr or Ser95-Thr96-Thr99 glycopeptide, i.e., the initial HexHexNAc glycosylation occurs at Thr99 whereas the second HexHexNAc is attached to Thr96.

...
The presence of the HexNAcHexNAc B/Y-type ion (m/z 407, Fig. 3E and 3G; the HexHexNAcB/Y-type ion (m/z 569, Fig. 3G); Hex2HexNAc (m/z 731, Fig. 3G) and dHexHexHexNAc (m/z 877, Fig 3f) verified that these glycans exceeded the HexHexNAc structure in complexity and thus confirmed the presence of one as opposed to two glycosylation sites for this peptide. B/Y-type oxonium ions exceeding m/z 407, e.g. at m/z 569 equally well matched ions corresponding to [Hex-(HexNAc)-HexNAc + H]^+ and [Hex-NAc-Hex-HexNAc + H]^+, i.e. a branched or a linear glycan sequence, respectively. Thus, B/Y-type ions at m/z 569 were unable to differentiate core 2-like glycans from elongated (linear) core 1-like structures. The same limitation is true for B-type ions at m/z 731 (Fig. 3G), corresponding to the entire Hex2HexNAc moiety of O-linked glycopeptides. Y-type oxonium ions at m/z 528, corresponding to [Hex-HexNAc-Hex + H]^+, could potentially reveal a linear O-glycan sequence but such ions were not observed in any CID-MS^n experiments for Hex2HexNAc glycoforms in this study.

Additionally, we identified secondary modifications of some O-linked glycopeptides. The CID-MS^2 fragmentation spectrum of the HexHexNAc glycosylated P^32ATDELVLA^60 peptide (Microfibrillar-associated protein 5, UniProt/KB accession Q13361) (Fig. 4A) showed an initial loss of ~80 Da (m/z 641.0), which we tentatively assigned as a sulfate group (79.9568 Da), but which could in theory also be a phosphate group (79.9663 Da). The precursor ion (m/z 681.28042+, not shown) was found to deviate by 1.69 ppm (~5.28 ppm for a phosphorylated precursor ion) from the theoretical monoisotopic mass of a sulfated precursor ion. In addition to the oxonium ions at m/z 204 (HexNAc) and m/z 366 (HexHexNAc), a fragment ion at m/z 446 was also observed which indicated that the sulfate group resides on the glycan and not on the peptide (Fig. 4A and Fig. 4B). Co-eluting with the sulfated precursor, we also observed the nonsulfated glycoform, i.e. the HexHexNAc modified P^32ATDELVLA^60 peptide, which was also characterized by CID-MS^n and ECD-MS^2 fragmentation (supplemental Fig. S5). The FTICR-MS^1 measured mass difference between the sulfated (m/z 681.30192+) and nonsulfated (m/z 641.30192+) variants of the HexHexNAc glycosylated P^32ATDELVLA^60 peptide was found to be 79.9570 Da, which deviates from the theoretical value of a sulfate group (79.9568 Da) only by 0.0020 Da. Although the m/z 446 ion, corresponding to HexHexNAc + Sulf, was detected and mass measured in the ion trap, the accurate mass of the sulfated group was thus indirectly confirmed by the mass measurements of the precursor ions in the ICR cell. Unfortunately, whether the Hex or HexNAc was carrying the secondary modification could not be defined.

The CID-MS^2 spectrum of the HexHexNAc glycosylated T^19PAPLDSVFSSSER^32 peptide (Vitamin K-dependent protein C, UniProt/KB accession P04070) is shown in Fig. 4C. This glycopeptide was also detected with a mass increment of ~80 Da. However, the CID-MS^2 fragmentation of this glycopeptide resulted in an initial loss of Hex (to m/z 888.3) followed by a loss of HexNAc (to m/z 786.8), showing that the modification, tentatively assigned as a phosphorylation, was attached to the peptide and not to the glycan. The precursor ion (m/z 969.42002+, supplemental Fig. S5) was found to deviate by 3.30 ppm (8.20 ppm for a sulfated precursor ion) from the theoretical monoisotopic mass of a phosphorylated precursor ion. The results in Fig. 4C indicate that O-linked glycans are more susceptible to CID-induced fragmentation by comparison to phosphate groups. ECD-MS^2 fragmentation (supplemental Fig. S5) allowed us to pinpoint the HexHexNAc-O-
sequence to Thr\textsuperscript{19} but the phosphorylated serine residue, among the four possible, was not identified (Table I and supplemental Fig. S5).

CID- and ECD-fragmentation of N-linked Glycopeptides

Fifty-eight glycopeptides, corresponding to 25 differently N-linked peptides from 17 urinary glycoproteins were identified (0.0% false positive identifications) in the formic acid released glycopeptide fractions (supplemental Fig. S1C and supplemental Fig. S6). They are all listed together with their N-linked glycans, their attachment sites and Mascot scores of the dominating glycopeptides in Table I and in supplemental Table S3. As a general feature, we observed the presence of several glycoforms for each N-linked glycopeptide. The relative abundance of specific glycoforms was determined by integrating chromatographic peaks for individual peptide glycoforms and the values were used to estimate the relative distribution of N-glycan microheterogeneity at each site. Oligosaccharide composition corresponding to the biantennary complex type structure was typically dominating, although triantennary and fucosylated bi- and triantennary glycoforms were also identified (supplemental Table S3). Sialic acid micro-heterogeneity was not observable since sialic acids were hydrolyzed in the preparative procedure. To illustrate the used methodology, MS\textsuperscript{n} of N-linked glycopeptides originating from three well-known N-glycoproteins (apolipoprotein D, uromodulin and prothrombin) are described in more detail (Fig. 5).

CID and ECD of N- and O-glycosylated Urinary Glycoproteins

10.1074/mcp.M111.013649-12

Molecular & Cellular Proteomics 11.4
N-acetylhexosaminyl (NAc) and hexosaminyl (Hex) residues, respectively. The third most intense ion (m/z 1229.5) resulted from a glycosidic cleavage at the GlcNAcGlcNAc (N-linked) and GlcNAc (O-linked) core and corresponded to the [peptide+HexNAc+2H]^{2+} (Y_1) ion. CID-MS^n of the [peptide+HexNAc+2H]^{2+} ion (Fig. 5C) induced peptide backbone fragmentation into b- and y-ions and were used for identification of the glycan attachment site and peptide sequence by the Mascot algorithm.

Second, the CID-MS^n fragmentation of a precursor at m/z 1394.9, corresponded to a fucosylated tetra-antennary complex type N-glycopeptide from uromodulin (UniProtKB accession P07911) and resulted in a prominent charge reduced fragment ion at m/z 1909.8 because of the loss of a terminal HexHexNAc moiety and a proton (Fig. 5D). The second most intense fragment (m/z 1017.4) corresponded to [peptide+HexHexNAc+2H]^{2+}, indicating that the fucose residue on the asparagine linked GlcNAc. Additional fragment ions were visible at m/z 1836.8, m/z 1727.6, and m/z 1646.7 corresponding to the loss of dHexHexHexNAc, Hex_2HexNAc_2 and Hex_2HexNAc_2, respectively, and revealed partial structural information on the N-linked glycan. The CID-MS^3 spectrum at m/z 1909.8 (Fig. 5E) showed further sequential glycosidic fragmentation and the entire N-glycan sequence was verified. Ideally, the fragment ion corresponding to [peptide+HexHexNAc+2H]^{2+} at m/z 944.6 (Fig. 5D) would have been used for the peptide identification but because of its low abundance it was not selected for CID-MS^3 fragmentation. Low abundance of Y_1-ion peaks was found to be a common feature for core fucosylated N-glycopeptides in CID-MS^2 spectra (supplemental Fig. S6). Instead, the fragment ion corresponding to [peptide+HexHexNAc+2H]^{2+} (m/z 1017.4, Fig. 5D) was selected for CID-MS^3 fragmentation (Fig. 5F). We observed an intense peak at m/z 943.9 corresponding to the loss of dHex together with minor peaks corresponding to peptide fragmentation and the MS^3 spectrum was matched to the tryptic QDFN^{122}TDISLLEHR peptide of uromodulin, with a Mascot score of 13 (p < 0.05 threshold; >10).

Third, the CID-MS^2 fragmentation of a pentuply charged biantennary N-linked glycopeptide at m/z 867.6 (Fig. 5G) rendered in a different fragmentation pattern compared with a triply charged biantennary N-glycopeptide (compare Figs. 5B and 5G) because of the different charge states, 3+ versus 5+. For the pentuply charged precursor we observed abundant glycosidic fragmentation of the terminal HexHexNAc residues and no apparent ion intensity corresponding to the peptide+HexNAc fragment. Subsequent CID-MS^3 at m/z 993.3 in (Fig. 5H) allowed for verification of the biantennary glycan structure but the amino acid sequence remained unidentified because of the lack of CID-MS^3 data on the peptide+HexNAc fragment ion. However, considering the high charge state, and thus the relatively low m/z ratio, this glycopeptide was efficiently fragmented into c- and z-type ions by ECD-MS^n (Fig. 5I) and the peptide sequence was identified to originate from the tryptic YPHK-PEIN^{143}STTHPGADLQENFCR peptide from prothrombin (UniProtKB accession P00734). The combination of CID-MS^n with ECD-MS^n was found to be useful in the identification of an additional N-linked glycopeptide (supplemental Fig. S6), namely the tryptic LHEITN^{177}ETFR peptide of vasorin (UniProtKB accession Q6EMK4).

DISCUSSION

The production of urine takes place in the nephron and involves a complex process of ultrafiltration, reabsorption and secretion, eventually leading to the formation of a complex solution containing metabolic waste products, proteins and peptides (54). The high content of salt and metabolic waste products in human urine requires sample purification for the removal of interfering compounds and isolation of urinary proteins prior to proteomic analysis. As yet, there is no universal method that offers complete recovery of the urinary proteome. Various approaches have been investigated for this purpose with each method offering advantages and disadvantages when compared with each other (55). In our study, the choice of sample preparation method was important not only for qualitative recovery of urinary proteins but was also essential for our downstream application, i.e. mild periodic acid oxidation of sialic acids. Efficient and selective oxidation of sialic acids was critical for the enrichment procedure of urinary glycoproteins, a reaction conducted under mild conditions employing only 2 mM periodic acid. Thus, the sample preparation method had to offer qualitative recovery of the urinary proteome and deplete metabolic waste products that might interfere or quench the subsequent oxidation of sialic acids. Several sample preparation methods were examined for this purpose, including organic solvent precipitation (acetone and trichloroacetic acid), spin column purification, size exclusion and reversed phase (C18) chromatography (not shown). Unfortunately, all were found to yield inadequate sample purity and failed in removing residual urinary pigments, which interfered with the sialic acid oxidation.

Eventually, we explored dialysis followed by lyophilization as a way to isolate and concentrate urinary proteins in a two-step procedure. Dialysis of urine against water alone was inefficient (supplemental Fig. S2) but the addition of 1.5% SDS and dialysis at 60 °C was found to yield sufficient sample purity for subsequent sialic acid oxidation. The dilute dialysates were subsequently concentrated through lyophilization to minimize the risk of unnecessary sample losses. Albeit time consuming, the preparative procedure employed in this study was thus justified by the strict requirement of sample purity and qualitative protein recovery.

Given that the dialyzed samples would serve as the basis for enrichment of sialoglycoproteins, it was also important to validate the preparative procedure to ensure that a representative urinary proteome was isolated following dialysis and lyophilization. By comparing our data set with the comprehensive proteomic studies of Adachi et al. and Kentis et al.
CID and ECD of N- and O-glycosylated Urinary Glycoproteins

(8, 9), we concluded that 90% of our protein identifications showed a nearly uniform overlap with the data sets of these studies (supplemental Fig. 3A). This observation confirmed that the glycoproteomic data would not mirror an atypical urinary subproteome as a result of the preparative procedure. It should be stressed that our proteomic analysis was not intended to expand the urinary proteome coverage. Thus, in contrast to previous studies, we did not deplete or prefractionate the urine sample prior to the one-dimensional electrophoretic separation, which may explain the relatively low number of protein identifications in this study.

Subsequent enrichment of sialoglycoproteins from the dialysates was achieved through conjugation of oxidized sialic acids to hydrazide beads (supplemental Fig. S1B and S1C). Although side reactions with terminal Hex or HexNAc residues of nonsialylated glycoproteins cannot be completely avoided, the mild oxidation constitutes the first step of introducing specificity to the enrichment procedure. Under these mild conditions, oxidation takes place primarily at the glycerol side chain (C7–C9) of sialic acids. In other words, hydrazide reactive aldehyde groups are specifically introduced on sialic acid by periodic acid oxidation at 0 °C. Consequently, targeted enrichment of sialoglycoproteins is enabled by reducing sample complexity through sequential washes of the solid phase to remove nonglycosylated and nonsialylated urinary proteins.

Following trypsin digestion and peptide extraction, the solid phase was extensively washed to remove any remaining nonglycosylated peptides in order to avoid interference by e.g., ion suppression effects in downstream analyses. The covalently linked glycopeptides were subsequently released by mild formic acid hydrolysis for MS-analysis. The formic acid treatment results in specific hydrolysis of sialic acid glycosidic bonds without affecting linkages between dHex, Hex or HexNAc residues, and thereby represents the second step of specificity in the glycopeptide enrichment procedure. Only species sensitive to formic acid cleavage are released from the hydrazide beads, which includes glycopeptides conjugated through sialic acids and exclude nonsialylated glycopeptides. Thus, other biomolecules harboring hydrazide reactive groups but lacking formic acid sensitive linkages are also excluded in this step. The combination of both specificity steps, i.e. mild periodic acid oxidation and mild formic acid hydrolysis, thus allows for selective isolation of desialylated glycopeptides. Consistent with this statement, base peak chromatograms of formic acid released fractions revealed various N- and O-linked glycopeptides as the dominating components (supplemental Fig. S4) with >80% of the subsequent CID-MS3 spectra presenting typical glycopeptide fragmentation patterns accompanied by diagnostic carbohydrate oxonium ions (56).

Identification of glycan- and peptide sequences was enabled by subjecting enriched glycopeptides to multiple rounds of CID fragmentation. CID-MS3 spectra of HexHexNAc glycoforms displayed prominent Y1 and Y0 fragments that were used to identify HexHexNAc-O-Ser/Thr sequences. Weak fragment ions corresponding to the mass of peptide+Hex, indicated with an asterisk in Figs. 2 to 4, were also observed during CID-MS2. These observations may contradict the HexHexNAc-O-Ser/Thr sequence outlined above, suggesting a Hex residue as the internal peptide linked monosaccharide. However, migration of hexose residues upon CID of protonated N-glycans and N-glycopeptides has been previously observed (57, 58), resulting in fragment ions which may lead to incorrect structural predictions. We speculate that the weak peptide+Hex fragment ions generated upon CID of protonated O-linked glycopeptides are most likely caused by hexose migrations similar to those observed for protonated N-linked glycopeptides but further studies are needed to verify these findings.

O-linked glycopeptides containing the Hex3HexNAc2 glycoform generally required five CID-MS3 experiments to delineate glycan- and peptide sequences. For O-linked glycopeptides with more than four monosaccharide units, isolation of intact peptide ions for CID-MS3 fragmentation proved difficult because of the increasing dominance of glycosidic fragments in MS2 spectra. Thus, the characterization of glycan- and peptide sequences for O-linked glycopeptides glycosylated beyond the simple core 1-like structure was rapidly complicated by the increasing number of monosaccharides. This is in contrast to N-linked glycopeptides which are readily identified even though they contain 9–13 monosaccharide units. Difficulties in characterizing O-linked glycopeptides with the Hex3HexNAc2 glycoform arise not only from isolation of Y0 ions for CID-MS3, but also from assigning the correct glycan sequence for the carbohydrate moiety. Y-type fragments are usually unable to resolve complex O-glycan sequences since they are equally well matched to the fragmentation pattern of different glycoforms. Thus, the identification procedure for O-glycopeptides is not easily automated and careful manual annotation is still necessary for correct assignment of glycan sequences.

By combining the CID and ECD data for each precursor ion, complementary information of core glycosylation could be gathered. ECD induced peptide fragmentation of HexHexNAc2 glycoforms revealed if the oligosaccharide components were located on two separate amino acids, suggesting a macroheterogeneity with two core 1-like glycans (Fig. 2), or different glycans on one single amino acid, indicating site-specific microheterogeneity (Fig. 3). However, ECD fragmentation does not provide structural information on the glycan sequence per se and determination of glycan sequence was therefore mainly based on CID-MS3 data. Thus, the main purpose of the ECD experiments was to determine the amino acid attachment sites of O-linked glycans. Traditionally, O-linked glycans are attached to serine or threonine residues but recently we reported a tyrosine residue to be modified by a sialylated O-linked glycan on amyloid beta peptides in human cerebrospinal fluid (59). However, our ECD
experiments did not reveal any tyrosine glycosylated peptides in the urine samples, suggesting that complex tyrosine glycosylation is rare, and possibly more tissue specific, than mucin-type O-glycosylation on the serine and threonine residues.

The majority of O-linked glycopeptides in Table I were thus identified with a single core 1-like glycan, which raises the issue of whether or not proteins O-glycosylated with core-1 like glycans are positively selected for by our approach. We argue that terminal sialic acids should be equally well oxidized by the periodic acid treatment, regardless of their core glycan structure, and that O-glycopeptides are equally well enriched on the hydrazide beads, given that they are sialylated to the same extent. The release mechanism should also not be dependent on the core glycan structure but only related to the hydrolysis of acid sensitive NeuAc-Gal or NeuAc-GalNAC glycosidic linkages. The subsequent detection of glycopeptides in LC-FTICR-MS is largely dependent on two factors: 1) the chromatographic properties of the peptide backbone, i.e. only glycopeptides of suitable length and hydrophobic character will be resolved by the C18 column; and 2) the physiochemical properties of the peptide backbone, which will dictate the extent of ionization and the stability of the parent ions. O-glycosylation microheterogeneity was found to have a minor impact on chromatographic retention times (Fig. 2E) with various peptide glycoforms eluting within a narrow time frame. The chromatography is thus not expected to favor any particular peptide glycoform since the retaining properties of the C18 column are generally dependent on the peptide composition rather than on the glycan structure. Thus, enrichment and characterization of O-glycan microheterogeneity, i.e. core 1-like versus core 2-like glycosylations, is probably not limited by the chromatographic resolution since different core glycans attached to the same peptide backbone are expected to be resolved equally well. Positive mode ionization of glycopeptides results in detection of \([M+nH]^{+}\) molecular ions, an outcome that is dependent on the proton affinity of the peptide backbone. This property justifies the comparison of signal intensities not only for detection of microheterogeneity but also for relative quantification of individual peptide glycoforms (60). We were also able to observe extensive microheterogeneity for specific O-glycopeptides, as demonstrated for the A342VAVTLQSH350 peptide of protein YIPF3 in Fig. 3. This O-glycopeptide was identified in five different core glycoforms ranging from a single HexNAc residue to a fucos containing pentasaccharide, clearly showing that our approach is not selective for O-glycopeptides occupied only by core 1-like glycans. Taken together, this indicates that the observed HexHexNAc core 1-like glycans are indeed the predominant O-glycans of the sialylated human urinary glycoproteome. In an earlier study the sialylated core 1 glycan was really shown to be the dominating O-glycan for uromodulin in nonpregnant female and male urine samples whereas Lewis structures on O-glycans were typical for uromodulin in pregnant female urine (61). We were unable to identify any O-linked glycopeptides from uromodulin in our study, which suggests that the O-glycans of uromodulin are located within trypsin-inaccessible regions of the protein. Alternatively, the trypsin digestion might also result in short, hydrophilic O-glycopeptides which were not retained by the C18 column and thus not detected during analysis. This limitation, which extends to all urinary glycoproteins and is valid for both N- and O-linked glycosylations, may be circumvented by the use of alternative proteases.

Several urinary glycoproteins, e.g. CD44, macrophage colony-stimulating factor 1, vasorin, complement component 7 and protein HEG homolog, identified as enriched glycopeptides in Table I, are each estimated to constitute less than 0.1–0.02% (by mass) of the core urinary proteome (10). This clearly shows that sialylated glycoproteins present in minute amounts in the urine are selectively made accessible for glycoproteomic characterization by the enrichment procedure. Notably, several other glycoproteins of Table I have been identified as potential biomarkers, e.g. elevated levels of urinary IGF-2 in urothelial carcinoma of the bladder (62) and it is not unlikely that these changes are accompanied by aberrant O-glycan profiles. The sialyl-Tn antigen (Neu5Acα2–6GalNAc–O–Ser/Thr) is a rare glycoepitope in normal tissue but high expression levels are known to occur in ovarian (63), gastric (64), colorectal (65) and pancreatic (66) carcinomas. Existing evidence also indicates that O-glycan occupancy is increased in cancer cells (67, 68). The ability to probe both these features simultaneously, i.e. site occupancy and O-glycan microheterogeneity, thus offers a unique opportunity to link aberrant glycans with distinct proteins. Although nonsialylated structures, e.g. Tn-antigen (GalNAc–O–Ser/Thr) or high-mannose type N-glycans are not enriched by the procedure, this analytical strategy could provide further insight into the process of pathogenesis for a wide range of diseases by identifying key proteins that are aberrantly glycosylated. Thus, the methodology and the results presented in this study should be of value for further exploration of the urinary glycoproteome in search of novel disease biomarkers.

* This study was supported by grants from the Swedish Research Council (project 8266), the Inga-Britt and Arne Lundberg Research Foundation, the Wilhelm and Martina Lundgren Foundation, the Torsten and Ragnar Söderberg Foundation and by governmental grants to the Sahlgrenska University Hospital.

§ To whom correspondence should be addressed: Department of Clinical Chemistry and Transfusion Medicine, Institute of Biomedicine, Sahlgrenska Academy at the University of Gothenburg, Sweden. Tel.: +46 31 342 1330; Fax: +46 31 82 84 58; E-mail: goran.larson@clinchem.gu.se.

Address Bruna Stråket 16, Sahlgrenska University Hospital, SE 413 45 Gothenburg, Sweden.

During the revision of this manuscript, the O-glycosites of Protein delta homolog 1(Th234), Uniprot/KB accession P80370) and Protein YIPF3 (Th346, Uniprot/KB accession Q9GZM5) were independently identified by Steentoft et al Nat Methods, 2011, Oct 9. doi: 10.1038/nmeth.1731.
REFERENCES

1. Hoorn, E. J., Piisipik, T., Zietse, R., Gross, P., Frokiaer, J., Wang, N. S., Gonzales, P. A., Star, R. A., and Knepper, M. A. (2005) Prospects for urinary proteome exomes as a source of urinary biomarkers. Ne- phrology 10, 283–290

2. Thongboonkerd, V., and Malsapin, P. (2005) Renal and urinary proteomics: current applications and challenges. Proteomics 5, 1033–1042

3. Heine, G., Raida, M., and Forsmann, W. G. (1997) Mapping of peptides and protein fragments in human urine using liquid chromatography-mass spectrometry. J. Chromatogr. A 776, 117–124

4. Spahr, C. S., Davis, M. T., McGriny, M. D., Robinson, J. H., Bures, E. J., Beierle, J., Mort, J., Courchesne, P. L., Chen, K., Wahl, C. R., Yu, W., Luethy, R., and Patterson, S. D. (2001) Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfraccionated tryptic digest. Proteomics 1, 93–107

5. Pieper, R., Gatlin, C. L., McGrath, A. M., Makusky, J. A., Mondal, M., Seonarain, M., Field, E., Schatz, C. R., Estock, M. A., Ahmed, N., Ander- son, N. G., and Steiner, S. (2004) Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots. Proteomics 4, 1159–1174

6. Piisipik, T., Shen, R. F., and Knepper, M. A. (2004) Identification and proteomic profiling of exomes in human urine. Proc. Natl. Acad. Sci. U. S. A 101, 13368–13373

7. Castagna, A., Cecconi, D., Sennels, L., Rappsilber, J., Guerrier, L., Fortis, F., Boschetti, E., Lomas, L., and Righetti, P. G. (2005) Exploring the hidden human urinary proteome via ligand library beads. J. Proteome Res. 4, 1917–1930

8. Adachi, J., Kumar, C., Zhang, Y., Olsen, J. V., and Mann, M. (2006) The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. Genome Biol. 7, R80

9. Kentsis, A., Monigatti, F., Dorff, K., Campagne, F., Bachur, R., and Steen, H. (2009) Hidden human urinary proteome via ligand library beads. J. Proteome Res. 4, 1917–1930

10. Tabak, L. A. (2010) The role of mucin-type O-glycans in eukaryotic development. Semin. Cell Dev. Biol. 21, 616–621

11. Scholdager, K. T., Vester-Christensen, M. B., Goth, C. K., Petersen, T. N., Brunak, S., Bennett, E. P., Lavery, S. B., and Clausen, H. (2011) A systematic study of site-specific GalNAc-Type O-glycosylation modulating proprotein convertase processing. J. Biol. Chem. 286, 325–334

12. Janik, M. E., Litynska, A., and Vereeken, P. (2010) Cell migration: the role of integrin glycosylation. Biochim. Biophys. Acta 1800, 545–555

13. Jensen, P. H., Kolarich, D., and Packer, N. H. (2010) Mucin-type O-glyco- sylation—putting the pieces together. FEBS J. 277, 81–94

14. Nilsson, J., Ruetschi, U., Halim, A., Carlssohn, E., Brinkmalm, G., and Larson, G. (2009) Enrichment of glycopeptides for glycan structure and attachment site identification. Nat. Methods 6, 809–811

15. Kelleher, N. L., Zubarev, R. A., Bush, K., Furie, B., Furie, B. C., McLafferty, F. W., and Walsh, C. T. (1999) Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxyglutamic acid. Proc. Natl. Acad. Sci. U. S. A. 96, 9528–9533

16. Deguchi, K., Ito, H., Baba, T., Hirabayashi, A., Nakagawa, H., Fumoto, M., Hinou, H., and Nishimura, S. (2007) Structural analysis of O-glycopeptides containing negative- and positive-ion multi-stage mass spectra obtained by collision-induced and electron-capture dissociations in lin-
ear ion trap time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. 21, 691–698
42. Perdivara, I., Petrovich, R., Allinquant, B., Deterding, L. J., Tomer, K. B., and Przybylski, M. (2009) Elucidation of O-glycosylation structures of the beta-amyloid precursor protein by liquid chromatography-mass spectrometry using electron transfer dissociation and collision induced dissociation. J. Proteome Res. 8, 631–642
43. Sibborn, C., van Dijk, J., Lidell, M. E., Noll, T., Hansson, G. C., and Backstrom, M. (2009) Localization of O-glycans in MUC1 glycoproteins using electron-capture dissociation fragmentation mass spectrometry. Glycobiology 19, 375–381
44. Christiansen, M. N., Kolarich, D., Hejna, H., Packer, N. H., and Jensen, P. H. (2010) Challenges of determining O-glycopeptide heterogeneity: a fungal glucanase model system. Anal. Chem. 82, 3500–3509
45. Takahashi, K., Wall, S. B., Suzuki, H., Smith, A. D. T., Kondo, A., Lattova, E., Mechref, Y., Miyoshi, E., Nakamura, K., Narimatsu, H., Novotny, M. V., Packer, N. H., Perreault, H., Peter-Katalinic, J., Pohlentz, G., Reinhold, V. N., Rudd, P. M., Suzuki, A., and Taniguchi, N. (2007) Comparison of the methods for profiling glycoprotein glycans—HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. Glycobiology 17, 411–422
46. Darula, Z., and Medzihradsky, K. F. (2009) Affinity enrichment and characterization of mucin core-1 type glycopolypeptides from bovine serum. Mol. Cell. Proteomics 8, 2515–2526
47. Forsman, A., Ruetschi, U., Ekholm, J., and Rymo, L. (2008) Identification of intracellular proteins associated with the EBV-encoded nuclear antigen 5 using an efficient TAP procedure and FT-ICR mass spectrometry. J. Proteome Res. 7, 2309–2319
48. Strohalm, M., Hassman, M., Kosata, B., and Kodicek, M. (2008) mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMass mMatt