

Site-specific O-Glycosylation Analysis of Human Blood Plasma Proteins*[§]

Marcus Hoffmann‡, Kristina Marx§, Udo Reichl‡¶, Manfred Wuhrer||, and Erdmann Rapp‡**‡‡

Site-specific glycosylation analysis is key to investigate structure-function relationships of glycoproteins, e.g. in the context of antigenicity and disease progression. The analysis, though, is quite challenging and time consuming, in particular for O-glycosylated proteins. In consequence, despite their clinical and biopharmaceutical importance, many human blood plasma glycoproteins have not been characterized comprehensively with respect to their O-glycosylation. Here, we report on the site-specific O-glycosylation analysis of human blood plasma glycoproteins. To this end pooled human blood plasma of healthy donors was proteolytically digested using a broad-specific enzyme (Proteinase K), followed by a precipitation step, as well as a glycopeptide enrichment and fractionation step via hydrophilic interaction liquid chromatography, the latter being optimized for intact O-glycopeptides carrying short mucin-type core-1 and -2 O-glycans, which represent the vast majority of O-glycans on human blood plasma proteins. Enriched O-glycopeptide fractions were subjected to mass spectrometric analysis using reversed-phase liquid chromatography coupled online to an ion trap mass spectrometer operated in positive-ion mode. Peptide identity and glycan composition were derived from low-energy collision-induced dissociation fragment spectra acquired in multistage mode. To pinpoint the O-glycosylation sites glycopeptides were fragmented using electron transfer dissociation. Spectra were annotated by database searches as well as manually. Overall, 31 O-glycosylation sites and regions belonging to 22 proteins were identified, the majority being acute-phase proteins. Strikingly, also

11 novel O-glycosylation sites and regions were identified. In total 23 O-glycosylation sites could be pinpointed. Interestingly, the use of Proteinase K proved to be particularly beneficial in this context. The identified O-glycan compositions most probably correspond to mono- and disialylated core-1 mucin-type O-glycans (T-antigen). The developed workflow allows the identification and characterization of the major population of the human blood plasma O-glycoproteome and our results provide new insights, which can help to unravel structure-function relationships. The data were deposited to ProteomeXchange PXD003270. *Molecular & Cellular Proteomics* 15: 10.1074/mcp.M115.053546, 624–641, 2016.

Human blood plasma harbors arguably the most complex yet also the most informative proteome present in the human body (1). A significant impact on its clinical relevance and diagnostic potential is attributed to the features and functions of a plethora of proteins (60–80 mg protein per ml plasma), covering a dynamic concentration range of more than ten orders of magnitude (2). The majority, that is 99%, of these proteins are classical blood plasma proteins, like albumins, (immuno)globulins, clotting factors, and proteins of the complement system; however, also a lower abundant but—no less meaningful—fraction of nonclassical proteins is present that comprises a multitude of cytokines as well as tissue leakage proteins. Several clinical studies could show that qualitative and quantitative alterations of these proteins (and peptides)—analyzed individually or in their entirety as a proteome (or peptidome)—can directly reflect pathophysiological states, and can serve as biomarkers for the onset and progression of a number of diseases (3–5). In recent years the focus of in-depth analyses of the human blood plasma proteome has evolved from the identification and quantification of the entire proteome (or peptidome) (6–10) toward the analysis of sub-proteomes like the interactome (11), phosphoproteome (12, 13) or the glycoproteome (14). The latter has received particular interest in recent years, because the majority of blood plasma proteins is N- and/or O-glycosylated (2). Although the comprehensive analysis of the N-glycoproteome is already quite advanced (15), even in complex samples like human blood plasma (16, 17), similar analyses of the O-glycoproteome - though arguably equally important and relevant - are still lagging behind. The most ubiquitously found and func-

From the ‡Max Planck Institute for Dynamics of Complex Technical Systems, Bioprocess Engineering, 39106 Magdeburg, Germany; §Bruker Daltonik GmbH, 28359 Bremen, Germany; ¶Otto von Guericke University Magdeburg, Chair of Bioprocess Engineering, 39106 Magdeburg, Germany; ||Center for Proteomics and Metabolomics, Department of Rheumatology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands; **glyXera GmbH, Leipziger Strasse 44 (Zenit), 39120 Magdeburg, Germany

✂ Author's Choice—Final version free via Creative Commons CC-BY license.

Received July 14, 2015, and in revised form, November 16, 2015
Published, MCP Papers in Press, November 23, 2015, DOI 10.1074/mcp.M115.053546

Author contributions: M.H., M.W., and E.R. designed research; M.H. performed research; M.H. and E.R. contributed new reagents or analytic tools; M.H., K.M., and M.W. analyzed data; M.H., K.M., U.R., M.W., and E.R. wrote the paper.

tionally relevant form of O-glycosylation, as shown by a number of O-glycan-related (clinical) studies (18–23), is the mucin-type O-glycosylation (O-GalNAc), in particular the core-1 and core-2 types (24, 25). The predominantly clustered occurrence of mucin-type O-glycans on proteins is described to confer overall stability and proteolytic protection (26). Apart from this global impact, recent studies could link the presence of O-glycans in the proximity of regulatory domains to proteolysis events involved in protein maturation (proprotein-converterase-processing) (27). To better understand these protective and regulatory capabilities and to move the mucin-type O-glycoproteome from form to function comprehensive site-specific O-glycosylation analyses are required.

One of the main obstacles in site-specific mucin-type O-glycosylation analyses relates to the lack of a predictable O-glycan consensus-motif within the peptide backbone as it can be found for N-glycans (28). The initial attachment of the N-acetylgalactosamine monosaccharide to the hydroxyl group of either serine or threonine, but also to tyrosine or hydroxylysine, is governed by a family of 20 distinct polypeptide GalNAc-transferase isoenzymes (GalNAc-Ts) with different but partially overlapping peptide specificities and tissue expression patterns. This dynamic regulation, in turn, contributes to the complexity of the mucin-type O-glycoproteome. However, previous studies could show that mucin-type O-glycans are primarily attached to serine or threonine in regions with a high content of serine, threonine and proline (Ser/Thr-X-X-Pro, Ser/Thr-P and Pro-Ser/Thr) (29, 30). As O-glycosylation is a postfolding event, taking place in the Golgi apparatus, the attachment is depended on protein surface accessibility and is thus predominantly found in coil, turn, and linker regions (31). Additional confounding factors during mucin-type O-glycosylation analyses are the clustered occurrence of O-glycans and the lack of a universal endo-O-glycosidase that enables the release of intact O-glycans from the proteins; though, chemical O-glycan release methods do exist (28).

Mass spectrometry has proven to be the core technique in site-specific N- and O-glycosylation analyses. A generic O-glycoproteomic workflow usually starts with the isolation, enrichment or prefractionation of a single glycoprotein or a group of glycoproteins. In subsequent steps, (glyco)peptides are generated by proteolytic digestion primarily using specific proteases like trypsin. Apart from this, also broad- and non-specific proteases like Proteinase K or Pronase E were successfully employed in recent years (32–34). Essential to nearly every glycoproteomic approach is the removal of high-abundant and interfering nonglycosylated peptides by selective enrichment of the usually lower abundant glycopeptides. The repertoire of glycopeptide enrichment and separation techniques covers different solid phase extraction and chromatography based methods such as hydrophilic liquid interaction chromatography (HILIC) (35, 36). The most frequently used setup for the measurement of enriched (glyco)peptides

is liquid chromatography (LC)¹ coupled online to electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS). Recent advances in instrumentation, in particular the development of electron-transfer/electron-capture dissociation (ETD/ECD) (37, 38), and high resolution orbital mass analyzers, have paved the way for the mapping of thousands of occupied N- and O-glycosylation sites as recently shown (17, 27). Combined workflows using ETD/ECD fragmentation along with (multistage, MSⁿ) fragmentation with high- and/or low collisional induced dissociation energy (HCD/CID) can provide compositional (structural) information on the glycan moiety as well as information on the peptide sequence and the glycosylation site (39, 40). Recent advances in mass spectrometry driven O-glycoproteomics have been reviewed in detail elsewhere (41, 42). Owing to the amount and complexity of O-glycoproteomic data a number of bioinformatic software tools for the prediction of mucin-type O-glycosylation sites (27) as well as for the database assisted interpretation and annotation of glycan and glycopeptide fragment spectra have been developed (43, 44). Moreover, reporting guidelines for collecting, sharing, integrating, and interpreting mass spectrometry based glycomics data have been specified by the MIRAGE consortium (minimum information required for a glycomics experiment) (45, 46).

The aim of our study was to develop a glycoproteomic workflow that allows the explorative nontargeted analysis of O-glycosylated human blood plasma proteins, which are known to carry mainly short mono- and disialylated mucin-type core-1 and -2 O-glycans. To achieve this, we have combined O-glycopeptide selective offline-HILIC fractionation of Proteinase K digested peptides with nano-reversed-phase liquid chromatography coupled online to multistage ion-trap mass spectrometry (nanoRP-LC-ESI-IT-MS: CID-MS²/MS³, ETD-MS²). The workflow has been applied to investigate the mucin-type O-glycoproteome of a pooled blood plasma sample derived from 20 healthy donors. Based on the mass spectrometric analysis of intact O-glycopeptides, we were able to characterize the O-glycosylation (*i.e.* peptide, site, and attached O-glycans) of a number of major human blood glycoproteins, including many acute phase proteins such as fibrinogen and plasminogen. Overall, the site-specific glycosylation analysis of human blood plasma glycopeptides revealed exclusively mono- and disialylated core-1 mucin-type O-glycopeptides. Interestingly, also a few novel O-glycosylation sites could be identified.

EXPERIMENTAL PROCEDURES

All chemicals and solvents were of the highest purity available. Purified water used for sample preparation and HILIC

¹ The abbreviations used are: LC, liquid chromatography; ECD, Electron-capture dissociation; ETD, Electron-transfer dissociation; FT-ICR, Fourier transform ion cyclotron resonance; HILIC, Hydrophilic interaction liquid chromatography; IAA, Iodoacetamide; ICC, Ion charge control; LTQ, Linear trap quadrupole.

fractionation was freshly prepared using a Milli-Q water purification system (referred to as “Milli-Q water”, $18.2 \text{ m}\Omega \times \text{cm}^{-1}$ at 25°C , Total Organic Carbon 3 ppb; Merck Millipore, Darmstadt, Germany). For preparation of LC-MS solvents ultrapure water was used, which was freshly prepared using the same system but equipped with an additional filter (referred to as “Milli-Q water MS”; LC-Pak Polisher, Merck Millipore).

Sample Preparation—Human blood plasma (pooled sample, derived from 20 healthy donors) was purchased from Affinity Biologicals Inc. (VisuCon-F, Frozen Normal Control Blood, FRNCP0125; Ancaster, ON, Canada). To $25 \mu\text{l}$ of the sample (about 2 mg protein) $25 \mu\text{l}$ 100 mM ammonium bicarbonate_(aq) (NH_4HCO_3 , pH 8.0) (Sigma Aldrich, Steinheim, Germany) was added to obtain a final concentration of 50 mM NH_4HCO_3 _(aq) (pH 8.0). Disulfide bonds were reduced by addition of $6.25 \mu\text{l}$ 100 mM 1,4-dithiothreitol (DTT; Sigma Aldrich) dissolved in 50 mM NH_4HCO_3 _(aq) (pH 8.0), to a final concentration of 10 mM DTT. The sample was incubated for 45 min at 60°C , and subsequently allowed to cool down to room temperature. Cysteine alkylation was achieved by addition of $12.5 \mu\text{l}$ 100 mM iodoacetamide (IAA; Sigma Aldrich) dissolved in 50 mM NH_4HCO_3 _(aq) (pH 8.0), to a final concentration of 16.67 mM IAA. The sample was incubated at room temperature for 20 min under light exclusion. The alkylation reaction was quenched by addition of $2.5 \mu\text{l}$ 100 mM DTT dissolved in 50 mM NH_4HCO_3 _(aq) (pH 8.0), followed by addition of $3.75 \mu\text{l}$ 50 mM NH_4HCO_3 _(aq) (pH 8.0), before placing the sample under a fluorescent lamp (gas-discharge lamp) for 15 min to decompose the light-sensitive IAA. By adding $169 \mu\text{l}$ 50 mM NH_4HCO_3 _(aq) (pH 8.0) the sample was brought to a final volume of 250 μl .

Proteinase K Digestion—Proteolytic digestion was achieved by addition of Proteinase K (Sigma Aldrich), a serine protease with a broad specificity that cleaves primarily after aliphatic, aromatic and hydrophobic amino acids. The pooled blood plasma sample (about 2 mg protein in 250 μl buffer) was supplemented with 66 μg Proteinase K dissolved in 122 μl 50 mM NH_4HCO_3 _(aq) (pH 8.0) in order to obtain a final enzyme/protein ratio of 1:30 (w/w, 0.033 mg enzyme per mg protein). The sample was incubated for 16 h at 37°C with gentle agitation (200 rpm).

Acetonitrile Precipitation—For post-digestion cleanup the sample was precipitated using acetonitrile (ACN; Sigma Aldrich). To this end four volumes of ACN were added and the sample was centrifuged for 10 min at $2880 \times g$ (Centrifuge 5804 R; Eppendorf, Hamburg, Germany). The supernatant was transferred and dried by vacuum centrifugation (RVC 2–33 CDplus, ALPHA 2–4 LDplus; Martin Christ GmbH, Osterode am Harz, Germany).

Glycopeptide Enrichment and Fractionation via HILIC-HPLC—The dried Proteinase K digest was resuspended in 500 μl 80% ACN in 50 mM NH_4HCO_3 _(aq) (v/v, pH 8.0) and subsequently centrifuged for 10 min at $20,238 \times g$ to remove any particles (Centrifuge 5424; Eppendorf). The supernatant, containing about 2 mg peptides and glycopeptides, was subjected to HILIC-HPLC (UltiMate™ Nano HPLC-System: Thermo Scientific/Dionex, Dreieich, Germany; HILIC Column: ACQUITY UPLC BEH HILIC Column, 130\AA , $1.7 \mu\text{m}$, $2.1 \text{ mm} \times 100 \text{ mm}$; Waters, Manchester, UK) for fractionation and glycopeptide enrichment.

The HPLC system was operated using a binary gradient of 100% ACN (v/v; solvent A) and 50 mM ammonium formate_(aq) (NH_4FA , pH 4.4; solvent B, Sigma Aldrich). After sample injection (500 μl) 20% solvent B was applied isocratically for 5 min, followed by a linear gradient to 50% solvent B within 25 min, both using a constant flow rate of 250 $\mu\text{l}/\text{min}$. Subsequently, a linear gradient went to 90% solvent B within 1 min, while reducing the flow rate to 150 $\mu\text{l}/\text{min}$. To wash the column solvent B was kept at 90% for 9 min. Column re-equilibration was achieved by isocratic elution with 20% solvent B

for 20 min; (the flow rate was increased to 250 $\mu\text{l}/\text{min}$ after 10 min). During the separation the column temperature was kept constant at 40°C . The elution profile was monitored by UV absorption at 214 nm. Fractions were collected every 2 mins from 0 min to 34 min. The fractions were dried by vacuum centrifugation and reconstituted in 50 μl Milli-Q water.

nanoRP-LC-ESI-IT-MSⁿ (CID,ETD)—HILIC fractions were analyzed by reversed-phase nano-LC-MSⁿ using an Ultimate3000 nanoHPLC system (Thermo Scientific/Dionex) coupled online to an ion trap mass spectrometer (AmaZon ETD, Bruker Daltonics, Bremen, Germany). Within the first 2 mins after sample injection, (glyco)peptides were loaded isocratically on a C18 μ -precolumn (Acclaim PepMap100, C18, $5 \mu\text{m}$, 100\AA , $300 \mu\text{m}$ i.d. $\times 5 \text{ mm}$; Thermo Scientific/Dionex). During this pre-concentration and desalting step, “loading pump solvent 1” (98% Milli-Q water MS, 2% ACN, 0.05% trifluoroacetic acid (Sigma Aldrich)) was used at a flow rate of 7 $\mu\text{l}/\text{min}$. Subsequently, the C18 μ -precolumn was switched in line with the C18 nano-separation column (Acclaim PepMap RSLC, C18, $2 \mu\text{m}$, 100\AA , $75 \mu\text{m}$ i.d. $\times 15 \text{ cm}$; Thermo Scientific/Dionex) for gradient elution. Here, the following solvents were used at a constant flow rate of 300 nL/min: “A” (98% Milli-Q water MS, 2% ACN, 0.1% formic acid (Sigma Aldrich)); “B” (10% Milli-Q water MS, 10% 2,2,2-trifluoroethanol (Merck), 80% ACN, 0.1% formic acid (Sigma Aldrich)). A binary gradient was applied as follows: 4% B for 2 min; linear gradient to 30% B within 30 min; isocratic washing step at 90% B for 5 min, finally 20 min re-equilibration at 4% B. After 42 min the C18 μ -precolumn was switched back into loading-pump flow, in order to be washed for 3 min at 100% “loading pump solvent 2” (20% Milli-Q water MS, 80% ACN, 0.05% trifluoroacetic acid (Sigma Aldrich)), and eventually to be re-equilibrated for 15 min at 100% “loading pump solvent 1,” both at 7 $\mu\text{l}/\text{min}$ flow rate.

The ion trap mass spectrometer was interfaced with a nanoFlow ESI Sprayer (Bruker Daltonics) and was operated in positive ion mode. For electrospray ionization the following parameters were used: capillary voltage ($-4,500 \text{ V}$), end plate offset (-500 V), N_2 dry gas (5 L/min), nebulizer (8 psi), dry gas temperature (220°C). The (glyco)peptides were fragmented via CID using multistage fragmentation (CID-MS², CID-MS³ experiments) and ETD-MS². For negative-mode chemical ionization during ETD measurements methane was supplied at 4 bar.

CID experiments were carried out using the following precursor scan settings: precursor scan mass range (m/z 100–2500); ion charge control (ICC) target (300, 000); maximum accumulation time (200 ms); averages (5); rolling averaging (off); target mass for smart parameter settings (m/z 850). CID-MS² experiments were conducted using a data dependent fragmentation routine. The top four most intense precursor ions, in the range of m/z 500–1500, were subjected to CID fragmentation in the ion trap mass analyzer (MS/MS fragmentation amplitude 1.20 V). The relative intensity threshold for fragmentation was set to 5%. Singly charged ions were excluded and selected precursors were actively excluded for 0.15 min after acquiring two fragment spectra. Charge state preference was set to “none.” Recorded scan range, ICC target and maximum accumulation time were the same as for the precursor scan. In CID-MS³ experiments precursor selection and fragmentation was applied manually. The fragmentation amplitude was set to 1.20 V. The recorded scan range was set individually with respect to the m/z of the precursor. ICC target and maximum accumulation time were the same as for the precursor scan. In both CID-MS² and CID-MS³ experiments the following CID parameters were used: cut-off selection (default); smart fragmentation (on); start amplitude (30%); end amplitude (200%); reaction time (40 ms). All CID experiments were carried out using the enhanced resolution mode. For CID-MS² measurements 1 μl of each HILIC

fraction was injected. For CID-MS³ measurements 5 μ l were used, respectively.

ETD experiments were carried out using the following precursor scan settings: precursor scan mass range (m/z 400–2500); ICC target (200, 000); max. accu. Time (50 ms); averages (5); rolling averaging (on, number: 1); target mass for smart parameter settings (m/z 850), enhanced resolution mode. Fragment spectra were acquired using a data dependent fragmentation routine in the ultrascan mode. The top three most intense precursor ions were subjected to ETD fragmentation in the ion trap mass analyzer. The relative intensity threshold for fragmentation was set to 1%. Singly charged ions were excluded and selected precursors were actively excluded for 0.15 min after acquiring two fragment spectra. Charge state preference was set to “none”. Fragment ions between m/z 100–3000 were detected. ICC target was set to 400,000 and max. accu. time was set to 100 ms. The following parameters were used for the EDT reagent: ICC target (500,000); max. accu. time (10 ms); Remove $\leq m/z$ 210 (On); Max. ETD Precursor (m/z 1200), cut-off (m/z 160); reaction time (160 ms); smart decomposition (auto). For ETD measurements 1 μ l of each HILIC fraction was injected.

All MS parameters were tested and optimized using *N*-glycopeptides derived from human IgG (sample preparation according to Selman *et al.* (47)) as well as *O*-glycopeptides from erythropoietin (Protea Biosciences, Morgantown, WV) (data not shown).

Data Analysis—Two separate strategies were followed for the spectra analysis. The first approach focuses exclusively on the identification of nonglycosylated peptides, whereas the second approach aims for the characterization and identification of the glycopeptides.

Analysis of Nonglycosylated Peptides—Fragment spectra (MS²) acquired with CID and ETD were searched for nonglycosylated peptides. To this end spectra were processed in DataAnalysis software 4.0 (Bruker Daltonics) using a built-in function for MSⁿ spectra processing (“processautomsn”; compound detection: standard settings). Processed spectra were imported into ProteinScape 3.1 (Bruker Daltonics) and were searched against a UniProtKB/Swiss-Prot database (SwissProt 51.6; 257964 sequences; 93947433 residues; downloaded February, 2013) using MASCOT version 2.2.04 (Matrix Science, London, UK). The following search parameters were applied: taxonomy (human); enzyme (none); fixed modifications (carbamidomethylation of cysteine residues); variable modifications (deamidation of asparagine and/or glutamine; methionine oxidation); precursor ion mass tolerance (± 0.3 Da, with #¹³C = 1; monoisotopic mass); fragment ion mass tolerance (CID: ± 0.5 Da; ETD: ± 1.3 Da); preferred charge state (2⁺/3⁺); peptide decoy search (1% FDR). Proteins and peptides with a MASCOT ion score higher than 50 and 25 were accepted, respectively.

Analysis of Glycopeptides—CID and ETD fragment spectra (MS²/MS³) were manually analyzed assisted by the DataAnalysis software 4.0 (Bruker Daltonics) without any pre-processing. Fragmentation of glycopeptides using low-energy CID almost exclusively yields fragment ions derived from the glycan moiety. This allows filtering of CID-MS² spectra for the presence of low-molecular weight fragment ions derived from the nonreducing end of the glycan (48) (B-ions, oxonium ions; [M+H]⁺; e.g. Hex { m/z 163.06}; NeuAc-H₂O { m/z 274.09}; NeuAc { m/z 292.10}; Hex₁NeuAc₁ { m/z 454.16}; HexNAc₁Hex₁-NeuAc₁ { m/z 657.24}; tolerance: m/z ± 0.3) using dedicated extracted ion chromatograms (EICs). In addition to this CID-MS² glycopeptide spectra feature multiply charged fragment ions (Y-ions) that show characteristic mono(oligo)-saccharide mass differences caused by the consecutive fragmentation of the glycan moiety down to the deglycosylated peptide. Both features were used to deduce the glycan composition along with the putative peptide mass in CID-MS² glycopeptide spectra. To identify the peptide moiety the putative peptide mass was used to trigger manual CID-MS³ fragmentation in

a separate run. In rare cases the peptide mass with an additional HexNAc had to be used for CID-MS³ fragmentation. CID-MS³ fragment spectra were exported to BioTools software 3.2 (Bruker Daltonics). Subsequent peptide identification was conducted using MASCOT. The spectra were searched against a UniProtKB/Swiss-Prot database using the following parameters: taxonomy (human); enzyme (none); fixed modifications (carbamidomethylation of cysteine residues); variable modifications (deamidation of asparagine and/or glutamine; methionine oxidation); precursor ion mass tolerance (± 0.3 Da, with #¹³C = 1; monoisotopic mass); fragment ion mass tolerance (CID: ± 0.35 Da); preferred charge state (2⁺/3⁺); MASCOT significance threshold (0.05); maximum number of reported hits: 10.

Peptides with a MASCOT ion score greater than 20 were considered; in very rare cases also lower scored peptides were accepted. Peptide identification was supported by the presence of a glycosylation consensus motif within the putative peptide sequence (*N*-glycosylation: Asn-X-Ser/Thr; *O*-glycosylation: Ser/Thr). Furthermore, knowledge, derived from public databases (UniProtKB and UniCarbKB) on already described *N*-/*O*-glycosylation sites within the putative peptide sequence or within the entire protein, was used to validate a peptide/protein hit. ETD-MS² fragment spectra of identified and characterized glycopeptides were annotated manually with respect to the presence of glycan fragment ions (Y-ions). Subsequently, the spectra were exported to BioTools software 3.2 (Bruker Daltonics) to identify the glycosylation site(s). The peptide sequences, proposed by CID-MS³ measurements, were modified *in silico* with the corresponding glycan compositions inferred from CID-MS², taking into account all the potential glycosylation sites. Fragment ions (c- and z-type ions) derived from these *in silico* glycopeptide sequences were then matched to their counterparts in the measured ETD-MS² spectra. The accuracy of this annotation was validated using the BioTools score along with manual inspection of the respective spectra. The entire glycopeptide data analysis workflow is briefly summarized in Fig. 1.

All mass spectrometry raw data as well as (glyco)peptide identifications and spectra annotations have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the MassIVE repository with the dataset identifier PXD002315 or MSV000079141, respectively.

RESULTS

Central to this study is the explorative, nontargeted analysis of *O*-glycosylated blood plasma glycoproteins. To this end a glycoproteomics approach was applied, that includes the identification of the peptide moiety as well as a characterization and localization of the *O*-glycosylation sites with the characterization of the corresponding *O*-glycans. HILIC enriched glycopeptides derived from a broad-specific proteolytic digest of human blood plasma proteins were analyzed by reversed-phase liquid chromatography combined with multistage mass spectrometry (CID-MS²/MS³, ETD-MS² fragmentation).

Reproducibility of the Proteinase K Digest—Previous studies on single glycoproteins could show the successful application of Proteinase K in the context of *N*- and *O*-glycoproteomics (32, 49–52). However, its application on complex samples, like human blood plasma, has not been described so far. Here, we have employed Proteinase K to generate (glyco)peptides from the entire (glyco)proteome of a pooled human blood plasma sample that was derived from 20 healthy

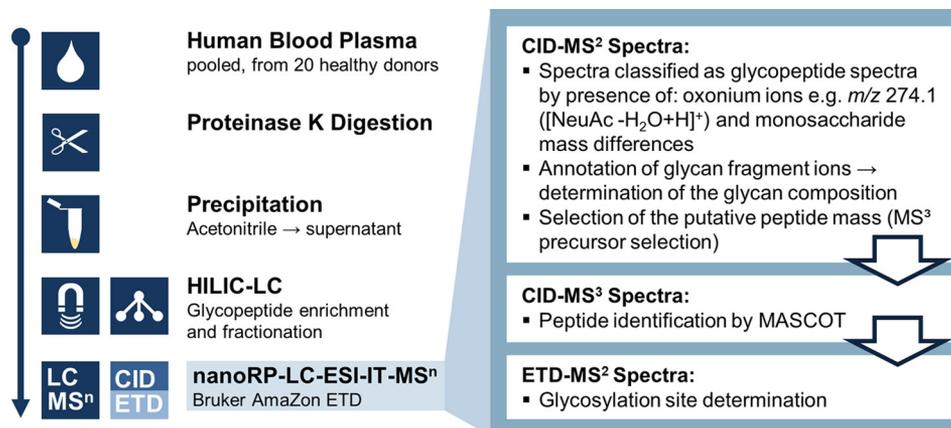


FIG. 1. Left: O-Glycoproteomic workflow for the analysis of human blood plasma glycoproteins. Right: LC-MSⁿ measurement and data analysis workflow. RP = reversed-phase; CID = Collision induced dissociation; ETD = Electron transfer dissociation.

donors. To assess the reproducibility of such a digest, five independent Proteinase K treated blood plasma samples (technical replicates) were measured with nanoRP-LC-ESI-IT-MS/MS in preliminary experiments. A comparison of the resulting base peak chromatograms revealed a high reproducibility of these digests, as shown in supplemental Fig. S1.

Glycopeptide Enrichment and Fractionation via HILIC-HPLC—The HILIC-HPLC fractionation carried out in the present study was optimized for the enrichment of O-glycosylated peptides (data not shown). In total 17 HILIC fractions were collected and were analyzed by nanoRP-LC-ESI-IT-MS² (CID). The acquired fragment spectra were manually screened for the presence of N- and O-glycopeptides—relying on the detection of diagnostic oxonium ions (B-ions, e.g. HexNAc₁-Hex₁NeuAc₁; m/z 657.24) and characteristic mono(oligo)-saccharide neutral loss fragment ions (Y-ions). Glycopeptides were detected in five HILIC fractions (#13–#17) (Fig. 2). The glycopeptides eluted in the range of 9–32 min in RP-LC-MS and clusters of glycopeptides were registered between 12–18 min, 20–22 min, and 25–29 min (exemplarily shown for fraction #15, Fig. 3).

Determination of the Glycan Composition—CID-MS² spectra were carefully inspected and manually annotated with respect to the glycan composition. Major signals in these spectra resulted from consecutive neutral losses (singly and doubly charged species) of the monosaccharides hexose, N-acetylhexosamine and N-acetylneuraminic acid from the intact glycopeptide and most of the time the applied collision energy induced the complete fragmentation of the glycan moiety while leaving the de-glycosylated peptide intact. These fragment ions along with corresponding oxonium ions, allowed inferring the glycan composition and the putative peptide mass (Fig. 4A). Detailed analysis revealed, that exclusively mucin-type core 1 mono- and disialylated O-linked glycopeptides ((di)sialyl-T-antigen) were present. For the glycan annotation a mass error of ± 0.3 Da was accepted. This parameter was justified as the observed mass errors were

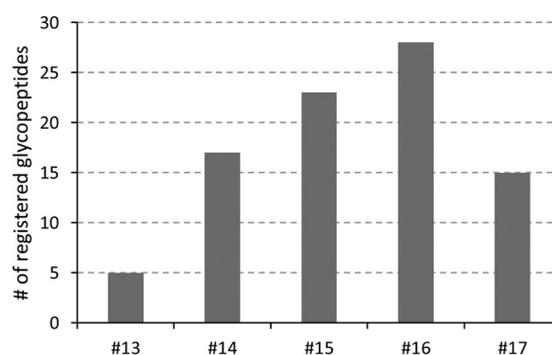


FIG. 2. Number of detected glycopeptides in HILIC fractions #13–#17.

about 0.07 Da (median value). In total 88 O-glycopeptides were detected and characterized with respect to their glycan composition. The registered glycopeptides covered an m/z range of 507–945 (average m/z 728) and were either doubly (55 peptides) or triply charged (33 peptides).

Identification of the Peptide Moiety—To complement the deduced glycan composition with peptide sequence information, CID-MS³ experiments were conducted on putative peptide masses, which were derived from CID-MS² spectra (Fig. 4A). In separate LC-MS runs the selected peptide precursor masses (predominantly singly charged) were used to trigger manual CID-MS³ fragmentation. In rare cases peptide+HexNAc was selected for fragmentation, because of low signal intensity of the peptide species in MS². CID-MS³ spectra were searched against the human subset of the highly curated and nonredundant protein database UniProtKB/Swiss-Prot. Notably, also in some CID-MS² spectra b- and y-ions derived from peptide backbone cleavages were detected, which enabled peptide identification (e.g. supplemental Fig. S5: α -2-HS-glycoprotein m/z 623.23³⁺).

For 88 detected glycopeptides, 60 corresponding peptides could be identified unambiguously (Table I, Table II). These 60 peptides could be linked to 22 different proteins, most of them being acute phase proteins. As the protein identification is

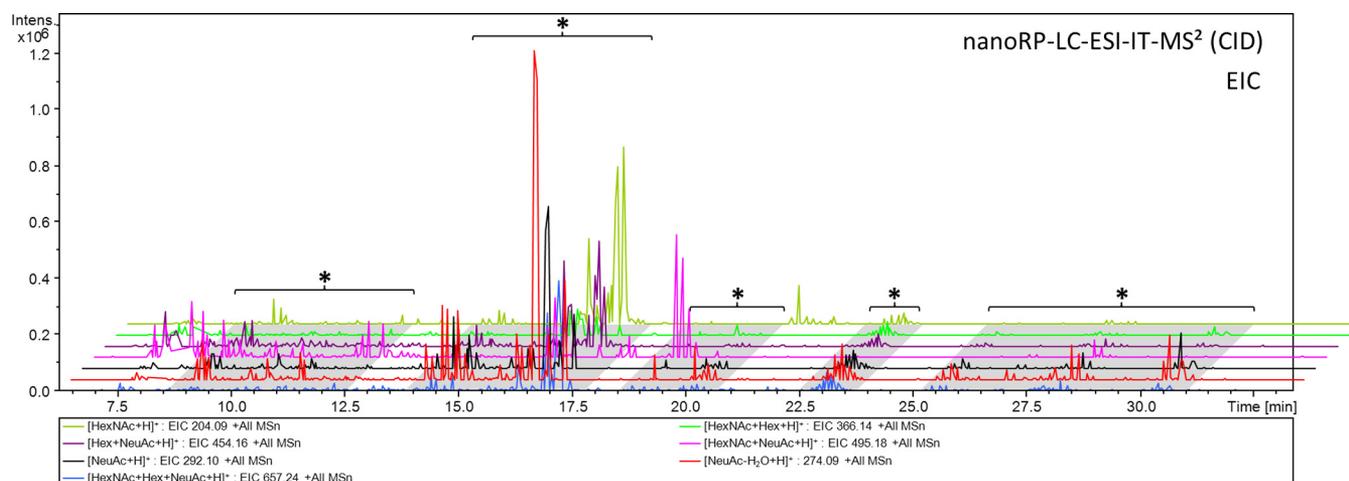


FIG. 3. Extracted ion chromatograms (EICs) of diagnostic glycan oxonium ions (e. g. $[\text{HexNAc}+\text{Hex}+\text{NeuAc}+\text{H}]^+$; EIC 657.24) reveal the clustered elution of O-glycopeptides (*) on a C18 reversed-phase column. EICs of HILIC fraction #15 are shown as an example.

based on a single peptide, validation of the potential peptide hits is of utmost importance. Here, in particular, the protein inference problem (53), which is intrinsic to bottom-up proteomic approaches, had to be considered. To cope with this, peptide spectra were manually revised and only peptide hits with a MASCOT ion score of greater than 20 were considered; only in rare cases, and supported by other evidences, also lower scored peptides were accepted. Furthermore, peptide hits needed to exhibit at least one potential O-glycosylation site (Ser/Thr). If available, knowledge derived from public databases (UniProtKB and UniCarbKB) on already described O-glycosylation sites within the putative peptides or within the entire protein was used to support a potential hit. The peptide identification was further corroborated by redundant identifications, that is the multiple occurrence of: (1) the same glycopeptide in different HILIC fractions, (2) or the same peptide but with a different glycan moiety, (3) or the identification of a peptide harboring the same glycosylation site, but differing in peptide length; the latter being attributed to the broad-specific proteolysis (e.g. alpha-2-HS-glycoprotein, ${}_{341}\text{TVVQPS}[\text{HexNAc}_1\text{Hex}_1\text{NeuAc}_1]\text{VG}_{348}$ derived from HILIC fraction #13 and ${}_{342}\text{VVQPS}[\text{HexNAc}_1\text{Hex}_1\text{NeuAc}_1]\text{VG}_{348}$ from fraction #14). In some cases, though, peptide identification was hampered or inconclusive. One of the main obstacles here was the frequent occurrence of prolines within the (glyco)peptide sequence, which was also described in literature. The cyclic structure of proline, gives rise to a high signal of the preceding y-ion but precludes in most cases the generation of a subsequent b-ion—thus introducing a sequence gap (54). This in turn leads to incomplete peptide fragment ion series and the occurrence of dipeptide fragment ions (e.g. PS and SP), which may result in ambiguity in peptide identification. This effect is particularly critical for short peptide sequences, as usually obtained by a broad- or nonspecific digest. The average peptide length of glycopeptides identified in this study is 10 amino acids (aa). This is significantly shorter than

the average length of tryptic peptides (14 aa, based on an *in-silico* digestion of the human UniProtKB database (55), supplemental Fig. S2). All this—in conjunction with a nonspecific peptide search—makes a reliable peptide identification challenging.

To complement the identified O-glycopeptides with nonglycosylated peptides that are also present in blood plasma, CID and ETD fragment spectra of the corresponding HILIC fractions (#1–17) were searched against the human subset of the UniProtKB/Swiss-Prot protein database. In total 111 proteins were identified. CID and ETD spectra provided complementary results; 54 and 45 proteins were identified, respectively, and only 12 proteins were identified with both modes. Compared with ETD, significantly more peptides were identified with CID (321 versus 150), though. The majority of peptides were derived from immunoglobulins, serotransferrin, haptoglobin and serumalbumin (supplemental Table S1). Notably, also nonglycosylated peptides corresponding to previously identified O-glycopeptides, e.g. of plasminogen and hemopexin, were identified (Table I).

Localization of the O-Glycosylation Sites—To further characterize the identified O-glycopeptides, the corresponding O-glycosylation sites needed to be localized. In a few cases the use of Proteinase K, already generated glycopeptides that exhibit only one possible O-glycosylation site, e.g. ${}_{132}\text{EGPVT}[\text{HexNAc}_1\text{Hex}_1\text{NeuAc}_1]\text{A}_{138}$ and ${}_{567}\text{DLIAT}[\text{HexNAc}_1\text{Hex}_1\text{NeuAc}_2]\text{M}_{572}$ from kininogen-1 or ${}_{234}\text{APT}[\text{HexNAc}_1\text{Hex}_1\text{NeuAc}_1]\text{HPAPPGLH}_{244}$ from selenoprotein P. Noteworthy, in the first example a tryptic digest would have generated a peptide with a length of 43 aa (${}_{119}\text{FSVATQTCQITPAEGPVVTAQYDC-LGCVHPISQSPDLEPILR}_{161}$), harboring 8 potential O-glycosylation sites. This clearly illustrates a benefit of the Proteinase K digest for the O-glycan site identification.

When the O-glycosylation sites could not be inferred directly, glycopeptides were subjected to ETD fragmentation in a separate LC-MS run (Fig. 4B). The most prominent

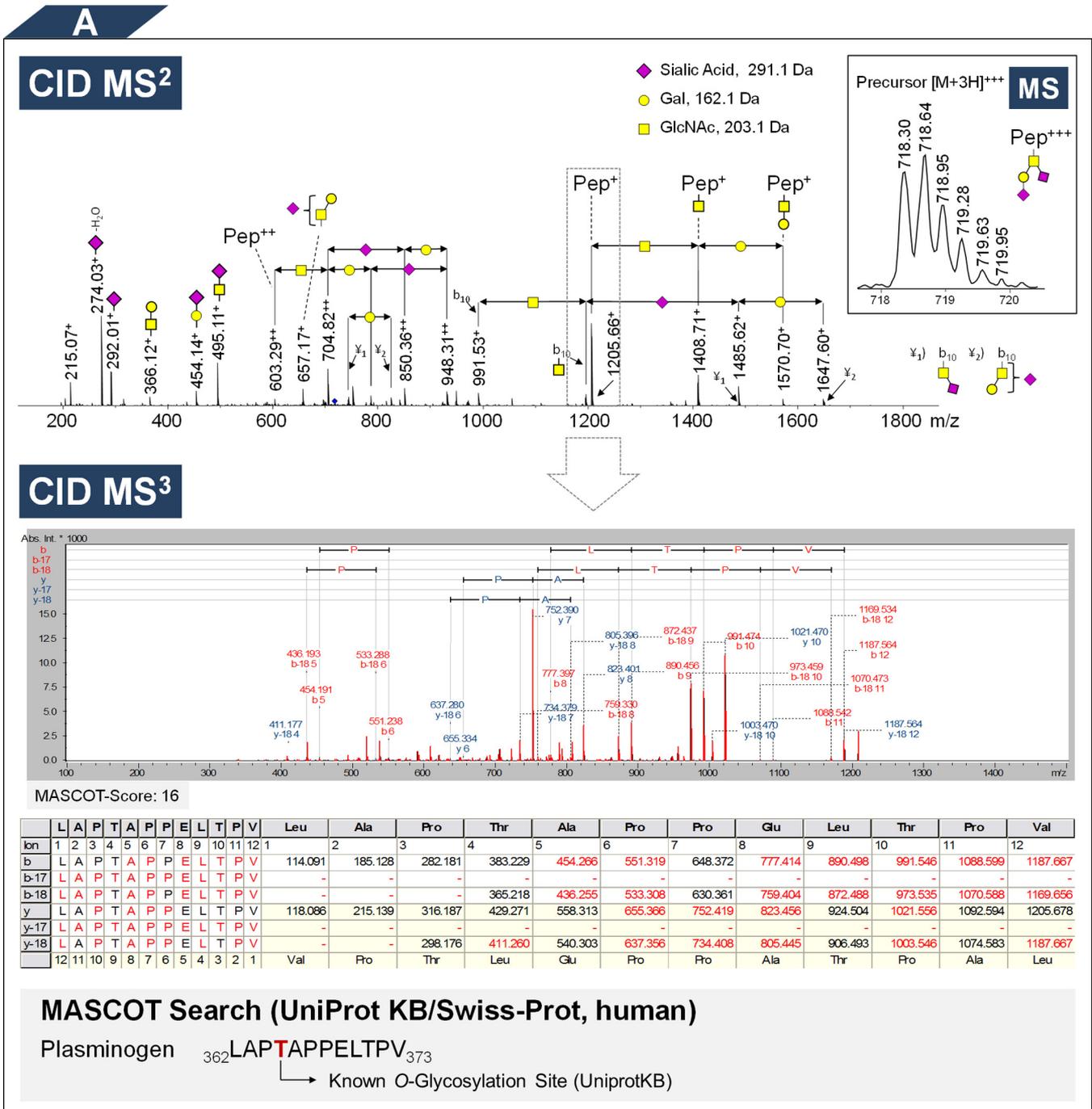
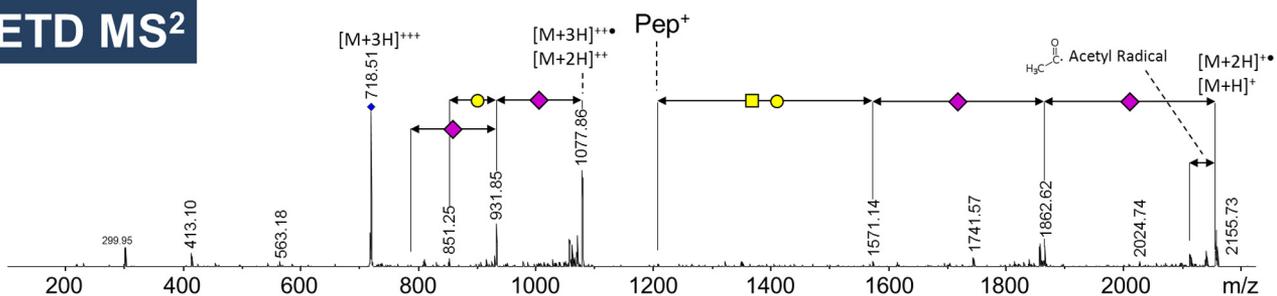
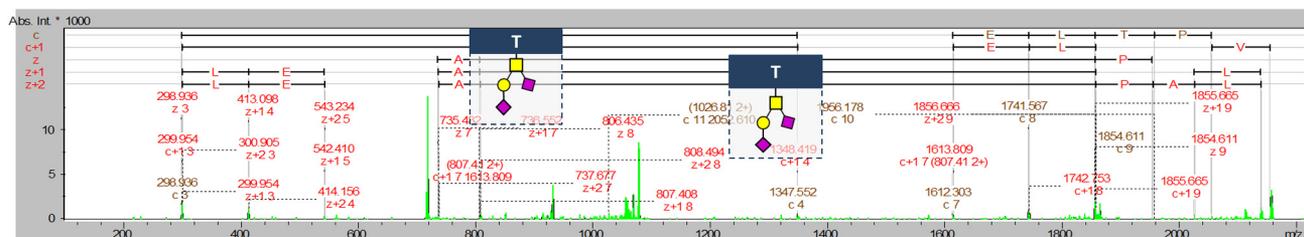
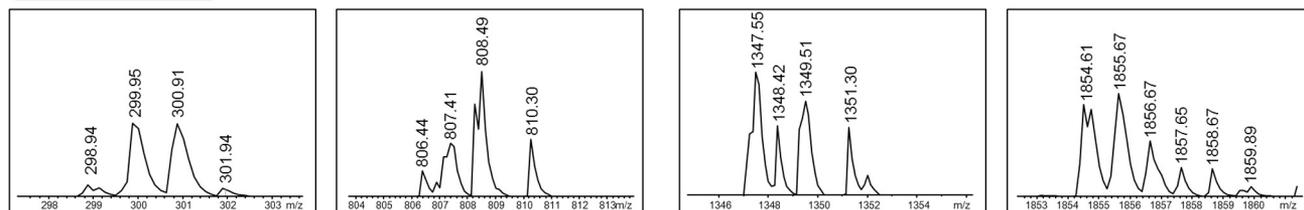


FIG. 4. Fragment ion spectra of the Proteinase K generated plasminogen O-glycopeptide ₃₆₂LAP**T**APPELTPV₃₇₃ measured with nanoRP-LC-ESI MSⁿ (positive ion mode, CID and ETD). A (top), For the given O-glycopeptide the CID-MS² spectrum is shown together with its corresponding precursor ion *m/z* 718.30 [M+3H]³⁺ (inset). The spectrum allows the elucidation of the O-glycan composition (here disialylated T-antigen). In addition, also some internal glycopeptide fragments have been detected (e.g. b₁₀+HexNAc). A (bottom): The putative peptide mass (*m/z* 1205.66 [M+H]⁺) of the given O-glycopeptide was subjected to CID-MS³ fragmentation. The peptide was identified by MASCOT search (Score: 16, UniProt KB/Swiss-Prot, human). B, The O-glycosylation site (here Thr₃₆₅) was pinpointed by means of ETD (Biotoools-Score: 150). Magnified regions show the isotope pattern of selected peptide fragment ions, confirming the annotation. In addition to peptide fragment ions also fragment ions derived from the glycan moiety were detected, allowing a verification of the glycan composition. Furthermore, a neutral loss of an acetyl radical from the intact O-glycopeptide was observed, which is typically seen in ETD spectra of glycopeptides.

B

 ETD MS²

 Biotoools: Peptide Spectra Annotation ${}_{362}\text{LAPT}(\text{HexNAc}_1\text{Hex}_1\text{NeuAc}_2)\text{APPELTPV}_{373}$


Biotoools-Score: 150



	L	A	P	T	A	P	P	E	L	T	P	V	Leu	Ala	Pro	Thr	Ala	Pro	Pro	Glu	Leu	Thr	Pro	Val
Ion	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
c	L	A	P	T	A	P	P	E	L	T	P	V	131.118	202.155	299.208	1347.578	1418.616	1515.668	1612.721	1741.764	1854.848	1955.895	2052.948	2152.017
c+1	L	A	P	T	A	P	P	E	L	T	P	V	132.126	203.163	300.216	1348.586	1419.623	1516.676	1613.729	1742.772	1855.856	1956.903	2053.956	2153.024
z	L	A	P	T	A	P	P	E	L	T	P	V	101.060	198.112	299.160	412.244	541.287	638.340	735.392	806.429	1854.800	1951.853	2022.890	2135.974
z+1	L	A	P	T	A	P	P	E	L	T	P	V	102.068	199.120	300.168	413.252	542.295	639.347	736.400	807.437	1855.808	1952.861	2023.898	2136.982
z+2	L	A	P	T	A	P	P	E	L	T	P	V	103.075	200.128	301.176	414.260	543.302	640.355	737.408	808.445	1856.816	1953.869	2024.906	2137.990
	12	11	10	9	8	7	6	5	4	3	2	1	Val	Pro	Thr	Leu				Ala	Thr	Pro	Ala	Leu

Fig. 4. Continued

peaks in the acquired ETD glycopeptide spectra were the unfragmented precursor ion along with charge-reduced species; minor peaks were derived from c- and z-type peptide backbone cleavages. Furthermore, fragment ions indicating either the loss of 43.018 Da ($\text{C}_2\text{H}_3\text{O}$) from the radical cationic species or 42.016 Da ($\text{C}_2\text{H}_2\text{O}$) from the even electron species $[\text{M}+\text{H}]^+$ were consistently detected. In the literature this spectral feature was attributed to the loss of an acetyl-radical from the N-acetyl group of a HexNAc (56, 57). This in turn can support the discrimination of ETD spectra derived from glycosylated and nonglycosylated species. Strikingly, and in contrast to the general mode of action of ETD, also fragmentations of the glycan moiety along the intact peptide backbone were observed, leading to a complete loss of the O-glycosylated Ser/Thr side-chain. Nevertheless, the resulting fragment ions enabled a verifi-

cation of the glycan composition as well as the peptide mass.

At first, ETD generated glycopeptide spectra were searched against the human subset of the UniProtKB/Swiss-Prot database using MASCOT, under consideration of the O-glycan modification (theoretical glycan mass used as variable modification of Ser/Thr). However, this strategy failed because of the presence of intense signals in the ETD spectrum, which correspond to: (I) the precursor ion, (II) the charge reduced precursor ion, (III) acetyl radicals ions, (IV) or glycan fragment ions. These ions might be erroneously interpreted as peptide derived fragment ions by the search engine, because ETD is supposed to solely produce peptide fragment ions while keeping fragile side-chain modifications, like the glycosylation, intact. To overcome this, glycopeptide spectra were exported to Bruker BioTools for manual spectra annotation.

TABLE I

Site-specific O-glycan composition of identified human blood plasma glycoproteins. Glycoproteins are listed with their UniprotKB accession number as well as the number of identified glycopeptides. O-glycosylated sites or regions are indexed with respect to the attached O-glycans (mono- and/or disialylated mucin-type core 1 O-glycans). O-glycosylation sites in bold have been pinpointed within this study. Previously unknown sites and regions are indicated by underlining. Curled brackets mark regions with several possible O-glycosylation sites. Superscript numbers indicate literature references. For every protein the number of registered, previously known, as well as new O-glycosylation sites and regions are given. For underlined proteins, glycosylated as well as non-glycosylated peptides were identified (supplemental Table S1). In addition previously reported plasma concentrations are given. HexNAc (N-acetylhexosamine), Hex (hexose), NeuAc (N-acetylneuraminic acid, sialic acid)

Identified Proteins	Accession	# Identified Glycopeptides	Site-Specific O-Glycan Composition				Registered / Previously Known / New O-Glycosylation Sites (Regions)	Reported Plasma Concentrations [ng/mL]
			I: HexNAc ₁ Hex ₁ NeuAc ₁	II: HexNAc ₁ Hex ₁ NeuAc ₂	I & II	2x HexNAc ₁ Hex ₁ NeuAc ₁		
1 Alpha-2-HS-glycoprotein	P02765	19	T(256) ⁽⁵⁹⁾ S(346) ^(30,58)	-	T(270) ⁽⁵⁹⁾	-	3 / 3 / 0	6.1*10 ⁽¹⁵⁾ / 8.2*10 ⁽¹⁷⁾
2 Apolipoprotein C-III	P02656	2	T(94) ^(30,58,61,87)	-	-	-	1 / 1 / 0	1.2*10 ⁽¹⁵⁾ / 1.7*10 ⁽¹⁷⁾
3 Apolipoprotein-L domain-containing protein 1	Q96LR9	1	-	T(133), T(135), S(136)	-	-	1 / 0 / 1	-
4 Complement C4-B	P0COL5	1	(S(1242), T(1244), S(1251)) ⁽⁶⁰⁾	-	-	-	1 / 1 / 0	9.0*10 ⁽¹⁷⁾
5 Fibrinogen alpha chain	P02671	1	(S(524), T(525), T(528)) ⁽⁵⁸⁾	-	-	-	1 / 6 / 0	2.7*10 ⁽¹⁵⁾ / 1.3*10 ⁽¹⁷⁾
6 Fibrinogen beta chain	P02675	1	S(58) ⁽⁵⁸⁾	-	-	-	1 / 1 / 0	9.7*10 ⁽¹⁵⁾ / 1.3*10 ⁽¹⁷⁾
7 GTP-binding protein 1	O00178	1	(T(119), S(122))	-	-	-	1 / 0 / 1	-
8 Hemopexin	P02790	1	(T(24) ^(30,60,61) , T(29) ⁽³⁰⁾ , S(30))	-	-	-	1 / 3 / 0	7.5*10 ⁽¹⁵⁾ / 1.8*10 ⁽¹⁷⁾
9 Immunoglobulin J Chain	P01591	2	T(97)	-	-	-	1 / 0 / 1	3.5*10 ⁽¹⁵⁾ / 5.6*10 ⁽¹⁷⁾
10 Insulin-like growth factor-binding protein 6	P24592	1	T(126) ^(30,61)	-	-	-	1 / 5 / 0	1.1*10 ⁽¹⁷⁾
11 Inter-alpha-trypsin inhibitor heavy chain H4	Q14624	4	(T(722) ⁽⁵⁸⁾ , T(723) ⁽⁵⁸⁾ , T(725) ⁽⁵⁸⁾ , S(733))	-	-	-	2 / 2 / 0	1.8*10 ⁽¹⁵⁾ / 4.2*10 ⁽¹⁷⁾
12 Inter-alpha-trypsin inhibitor heavy chain H2	P19823	3	T(691) ^(60,66,67)	-	-	(S(673), T(675)) ^(58,66,67)	3 / 4 / 0	2.1*10 ⁽¹⁵⁾ / 2.1*10 ⁽¹⁷⁾
13 Kininogen-1	P01042	6	(S(150), T(151)) ^(30,61) S(577) ⁽⁶²⁾	S(604) T(671) ⁽⁶²⁾	T(137) ^(30,61)	-	5 / 9 / 1	7.0*10 ⁽¹⁵⁾ / 2.8*10 ⁽¹⁷⁾
14 Legumain	Q99538	1	(T(305), S(307))	-	-	-	1 / 0 / 1	-
15 Plasminogen	P00747	4	-	-	T(365) ⁽⁵⁹⁾	-	1 / 2 / 0	1.4*10 ⁽¹⁵⁾ / 2.5*10 ⁽¹⁷⁾
16 Protein AMBP	P02760	6	T(24) ⁽³⁰⁾	-	-	-	1 / 2 / 0	6.0*10 ⁽¹⁵⁾ / 4.8*10 ⁽¹⁷⁾
17 Protocadherin beta-11	Q9Y5F2	1	S(24)	-	-	-	1 / 0 / 1	-
18 Protocadherin-16	Q96JQ0	1	T(692)	-	-	-	1 / 0 / 1	-
19 Selenoprotein P	P49908	1	T(236)	-	-	-	1 / 0 / 1	5.1*10 ⁽¹⁷⁾
20 Tau-tubulin kinase 2	Q6IQ55	1	T(820)	-	-	-	1 / 0 / 1	-
21 Tenascin-R	Q92752	1	-	-	T(1288)	-	1 / 2 / 1	-
22 Tenascin-X	P22105	1	-	-	T(3586)	-	1 / 0 / 1	7.0*10 ⁽¹⁷⁾
23 not unambiguously identified	-	28	18x	10x	-	-	-	-

Here, the identified glycopeptides were built *in silico*, taking into account the corresponding O-glycan moieties as well as all possible O-glycosylation sites. Subsequently, the resulting *in silico* fragment ions (c- and z-type ions) were matched to their counterparts in the measured ETD-MS² spectra. To evaluate the spectra annotation and to discern the correct O-glycosylation site, the BioTools spectra matching score along with manual inspection of the respective spectra were considered. Furthermore, public repositories, namely UniProtKB and UniCarbKB, were queried with respect to known O-glycosylation sites within the peptide in question. To further assess the validity of the O-glycosylation site annotation, the site occupancy was predicted using NetOGlyc—an online tool, based on machine-learning algorithms, which allows the prediction of mucin-type O-glycosylation sites (27). For 36 of 60 identified glycopeptides the quality of the corresponding ETD spectra was acceptable - in terms of signal intensity and the number of fragment ions. Overall, 31 O-glycosylation sites and regions were detected, of which 23 sites could be pinpointed (Tables I and II). Strikingly, 11 previously unknown O-glycosylation sites and regions were registered, of which 8 sites could be pinpointed. Generally, O-glycosylation on threonine residues was observed more frequently than on serine (16× Thr, 7× Ser). In accordance with literature, prolines were frequently found in close vicinity to the O-glycosylation site (positions *n* - 1, *n* + 1, *n* + 3), e.g. 267AVPT[HexNAc₁Hex₁NeuAc₁]PV₂₇₂, 343VQPS[HexNAc₁Hex₁NeuAc₁]VGA₃₄₉ from alpha-2-HS-glycoprotein (30). In addition also prolines in

position *n* + 2 were found occasionally, e.g. 20GPVPT [HexNAc₁Hex₁NeuAc₁]PPDNI₂₉ from alpha-1 microglycoprotein (protein AMBP).

Identified Glycoproteins: Selected Examples—In the following selected examples of identified O-glycopeptides are detailed, that feature novel O-glycosylation sites or exhibit remarkable fragmentation characteristics.

α-2-HS-glycoprotein—In this study, the majority of identified O-glycopeptides were derived from α-2-HS-glycoprotein, also known as fetuin-A. Fetuin-A is a negative acute phase glycoprotein that is highly abundant in fetal blood plasma. It is involved in transport and storage of substances and features three O-glycosylation sites (Thr₂₅₆, Thr₂₇₀, Ser₃₄₆), which are decorated with sialylated mucin-type core 1 O-glycan structures (30, 58, 59). In contrast to previous reports (30, 58), intact O-glycopeptides identified and characterized in the present study describe all three known fetuin-A O-glycosylation sites including the attached O-glycans (mono- and disialylated mucin-type core 1 O-glycans), respectively. By pinpointing O-glycosylation sites using ETD, the reported ETD Biotools scores can be misleading. This for instance holds true for the fetuin-A O-glycopeptide 252TQPVTSQPQPE₂₆₂ (*m/z* 623.23³⁺) and its three potential O-glycosylation sites: Thr₂₅₂ (669), Thr₂₅₆ (412), Thr₂₅₇ (362) (supplemental Fig. S7). According to the score values T₂₅₂ would be the occupied site; the presence of characteristic ETD fragment ions at *m/z* 344.01¹⁺ (c₃), 1200.45¹⁺ (c₅), 1287.47¹⁺ (c₆), 1525.57¹⁺ (z+1₈), 1751.51¹⁺ (z+2₁₀), though, clearly indicates the oc-

Site-specific O-Glycosylation Analysis of Human Blood Plasma Proteins

TABLE II

Detailed overview of all identified human blood plasma O-glycopeptides. For each O-glycopeptide, the corresponding glycoprotein including the UniProtKB accession number, the identified O-glycosylation site(s)/regions as well as the O-glycan composition are given, respectively. Likewise, the LC-MS retention time, the mass of the intact glycopeptide precursor, the measured peptide mass as well as the corresponding error is listed. The peptide identification using CID-MS³ was validated by the MASCOT peptide score and the Biotools score (CID). ETD based determination of the O-glycosylation site(s) was validated by the Biotools score (CID) as well as the NetOGlyc 4.0 score

Identified Proteins	Accession	O-Glycosylation Site (Region)	# Identified Peptides	O-Glycan Composition	Retention Time [min]	Precursor Ion	Measured peptide mass [M+H] ⁺ [Da]	Error [Da]	Peptide Score (CID)	Biotools Score (CID)	Biotools Score (ETD)	NetOGlyc 4.0 Score
1 Alpha-2-HS-glycoprotein	P02765	T(256)	19 252TQPVTSGPOPE ₂₅₂	HexNAc(1)Hex(1)NeuAc(1)	14.1-14.3	[934.36] ²⁺	1211.55	-0.04	28	402	T(252): 12 / T(256): 19 / S(257): 18	0.81 / 0.90 / 0.87
				HexNAc(1)Hex(1)NeuAc(2)	13.9-14.0	[623.23] ²⁺	1211.49	-0.11	42	46	T(252): 669 / T(256): 412 / S(257): 362	0.87 / 0.90 / 0.87
				HexNAc(1)Hex(1)NeuAc(2)	20.25-20.37	[620.26] ²⁺	583.33	-0.09	20	1	44	0.87
		T(270)	267AVFTPV ₂₇₂	HexNAc(1)Hex(1)NeuAc(1)	19.8-20.2	[765.77] ²⁺	583.27	-0.07	18	1	6	0.87
				HexNAc(1)Hex(1)NeuAc(1)	22.1-22.2	[833.3] ²⁺	1009.43	-0.09	19	8	141	0.87
				HexNAc(1)Hex(1)NeuAc(1)	27.5-28.5	[775.67] ²⁺	1668.84	-0.04	54	20	T(270): 153 / S(280): 75	0.87 / 0.96
		267AVFTPVDDPAPPSPPL ₂₈₃	HexNAc(1)Hex(1)NeuAc(2)	31.9-32.0	[872.72] ²⁺	1668.75	-0.13	59	10	T(270): 39 / S(280): 34	0.87 / 0.96	
			HexNAc(1)Hex(1)NeuAc(2)	31.3-31.7	[872.73] ²⁺	1668.77	-0.12	65	15	T(270): 39 / S(280): 21	0.87 / 0.96	
			HexNAc(1)Hex(1)NeuAc(1)	20.4-20.7	[684.76] ²⁺	712.35	-0.04	18	2	7	0.87	
		246EAVPTPV ₂₇₂	HexNAc(1)Hex(1)NeuAc(2)	21.9-22.2	[830.34] ²⁺	712.32	-0.07	16	38	-	0.87	
			HexNAc(1)Hex(1)NeuAc(1)	31.5-31.7	[818.68] ²⁺	1797.75	-0.18	11	12	-	0.87 / 0.96	
			HexNAc(1)Hex(1)NeuAc(2)	32.0	[915.71] ²⁺	1797.73	-0.10	46	23	-	0.87 / 0.96	
		S(346)	255NEAVPTPV ₂₇₂	HexNAc(1)Hex(1)NeuAc(1)	20.1-20.4	[741.8] ²⁺	826.38	-0.05	43	158	44	0.87
				HexNAc(1)Hex(1)NeuAc(1)	17.8-17.9	[721.78] ²⁺	786.39	-0.07	24	76	T(341): 533 / S(346): 187	0.93 / 0.94
				HexNAc(1)Hex(1)NeuAc(1)	16.7-16.8	[721.76] ²⁺	786.43	-0.01	25	149	-	0.93 / 0.94
				HexNAc(1)Hex(1)NeuAc(1)	17.3-17.4	[757.3] ²⁺	857.43	-0.05	20	36	-	0.93 / 0.94
				HexNAc(1)Hex(1)NeuAc(1)	14.0-14.1	[657.2] ²⁺	657.29	-0.11	20	15	-	0.94
				HexNAc(1)Hex(1)NeuAc(1)	15.3-15.7	[671.26] ²⁺	685.35	-0.03	44	13	-	0.94
				HexNAc(1)Hex(1)NeuAc(1)	16.3-16.4	[706.77] ²⁺	756.37	-0.04	19	18	-	0.94
HexNAc(1)Hex(1)NeuAc(1)	25.2-25.4			[681.25] ²⁺	1385.57	-0.17	35	21	T(94): 46 / S(95): 40	0.51 / 0.51		
HexNAc(1)Hex(1)NeuAc(1)	24.6-25.1			[681.35] ²⁺	1385.71	0.04	46	15	T(94): 124 / S(95): 148	0.51 / 0.51		
3 Apolipoprotein-I domain-containing protein 1	Q96LR9			T(133), T(135), S(136)	1 131AVTITSDL ₁₃₈	HexNAc(1)Hex(1)NeuAc(2)	24.3-24.5	[883.82] ²⁺	819.35	-0.09	31	10
4 Complement C4-B	P0C0L5	S(1242), T(1244), S(1251)	1 1242VSPTPAPRNPSDDPMPQ ₁₂₅₈	HexNAc(1)Hex(1)NeuAc(1)	17.8-18.1	[782.97] ²⁺	1690.59	-0.23	34	30	S(1242): 36 / T(1244): 49 / S(1251): 145	0.96 / 0.94 / 0.51
5 Fibrinogen alpha chain	P02671	S(524), T(525), T(528)	1 524STGKTFPG ₅₃₁	HexNAc(1)Hex(1)NeuAc(1)	12.3-12.5	[725.78] ²⁺	794.34	-0.06	28	61	-	0.76 / 0.75 / 0.58
6 Fibrinogen beta chain	P02675	S(58)	1 58EAPSLRPAPPPI ₆₇	HexNAc(1)Hex(1)NeuAc(1)	23.3-23.5	[706.27] ²⁺	1460.67	-0.10	22	14	S(58): 34 / S(67): 13	0.95 / 0.86
7 GTP-binding protein 1	O00178	T(119), S(122)	1 119YATVKSM ₁₂₃	HexNAc(1)Hex(1)NeuAc(1)	15.4	[728.75] ²⁺	800.36	0.96	27	9	-	0.24 / 0.17
8 Hemopexin	P02790	T(24), T(29), S(30)	2 24TLPPT ₃₀	HexNAc(1)Hex(1)NeuAc(1)	14.4-14.6	[684.73] ²⁺	712.35	-0.09	19	8	-	0.51 / 0.83 / 0.96
				HexNAc(1)Hex(1)NeuAc(1)	17.7-17.8	[608.71] ²⁺	560.28	0.02	27	15	-	0.11
9 Immunoglobulin J chain	P01591	T(97)	2 95DPT ₉₉	HexNAc(1)Hex(1)NeuAc(1)	18.6-18.8	[608.72] ²⁺	560.29	0.03	-	1	-	0.11
				HexNAc(1)Hex(1)NeuAc(1)	17.7-17.8	[608.71] ²⁺	560.28	0.02	27	15	-	0.11
10 Insulin-like growth factor-binding protein 6	P24592	T(126)	1 122KPKQAGTA ₁₂₇	HexNAc(1)Hex(1)NeuAc(1)	16.3-16.6	[664.69] ²⁺	672.30	-0.07	10	18	22	0.88
11 Inter-alpha-trypsin inhibitor heavy chain H4	Q14624	S(640), T(722), T(723), T(725)	4 640ASFSPR ₆₄₄	HexNAc(1)Hex(1)NeuAc(1)	11.0-11.3	[660.72] ²⁺	664.30	-0.07	16	51	S(640): 11 / S(642): 7	0.66 / 0.80
				HexNAc(1)Hex(1)NeuAc(1)	18.5-18.9	[742.76] ²⁺	828.42	-0.03	27	33	-	0.61 / 0.64 / 0.90
				HexNAc(1)Hex(1)NeuAc(1)	15.3-15.8	[806.77] ²⁺	956.51	0.00	26	4	-	0.61 / 0.64 / 0.90
				HexNAc(1)Hex(1)NeuAc(1)	16.0-16.2	[623.27] ²⁺	1211.58	0.07	31	23	T(722): 152 / T(723): 34 / T(725): 19 / T(733): 5	0.61 / 0.64 / 0.90 / 0.58
12 Inter-alpha-trypsin inhibitor heavy chain H2	P19823	S(673), T(675), T(691)	3 689WANPSP ₆₇₇	2x HexNAc(1)Hex(1)NeuAc(1)	21.0-21.1	[760.92] ²⁺	968.40	-0.08	30	1	-	0.79 / 0.88
				HexNAc(1)Hex(1)NeuAc(1)	12.5-12.7	[760.24] ²⁺	863.37	-0.06	32	65	S(690): 14 / T(691): 22	0.88 / 0.93
				HexNAc(1)Hex(1)NeuAc(1)	12.4-12.9	[507.15] ²⁺	863.34	-0.11	-	8	S(690): 65 / T(691): 84	0.88 / 0.93
13 Kininogen-1	P01042	T(137)	6 132EGPVVTA ₁₃₈	HexNAc(1)Hex(1)NeuAc(1)	16.6-16.7	[664.7] ²⁺	672.33	-0.03	-	43	-	0.51
				HexNAc(1)Hex(1)NeuAc(2)	17.4	[874.28] ²⁺	800.25	-0.02	19	19	-	0.51
		S(150), T(151)	6 146VHPISTQ ₁₅₂	HexNAc(1)Hex(1)NeuAc(1)	11.0-11.2	[719.23] ²⁺	781.37	-0.02	24	29	-	0.30 / 0.34
				HexNAc(1)Hex(1)NeuAc(2)	26.1-26.5	[805.73] ²⁺	663.34	-0.09	28	16	-	0.69
		T(571)	6 573MPPISAPIQ ₆₈₃	HexNAc(1)Hex(1)NeuAc(1)	25.5-25.9	[853.81] ²⁺	1050.54	-0.10	40	11	22	0.87
				HexNAc(1)Hex(1)NeuAc(2)	31.8-32.2	[734.26] ²⁺	1253.49	-0.08	24	21	S(604): 44 / T(609): 22 / T(610): 14	0.64 / 0.69 / 0.51
		14 Legumain	Q99538	T(305), S(307)	1 305HLDLTPSD ₃₀₉	HexNAc(1)Hex(1)NeuAc(1)	14.0-14.7	[550.86] ²⁺	994.38	-0.10	17	9
15 Plasminogen	P00747	T(365)	4 365LAPTAPPELTPV ₃₇₃	HexNAc(1)Hex(1)NeuAc(1)	27.9-28.8	[931.4] ²⁺	1205.64	-0.07	16	51	T(365): 115 / T(371): 173	0.98 / 0.45
				HexNAc(1)Hex(1)NeuAc(2)	28.2-28.5	[718.3] ²⁺	1205.63	-0.01	16	47	T(365): 150 / T(371): 36	0.98 / 0.45
				HexNAc(1)Hex(1)NeuAc(1)	25.6-26.1	[583.58] ²⁺	1092.56	-0.03	19	24	T(365): 65 / T(371): 199	0.98 / 0.45
HexNAc(1)Hex(1)NeuAc(1)	25.8-26.2	[874.88] ²⁺	1092.60	0.01	24	326	T(365): 177 / T(371): 34	0.98 / 0.45				
16 Protein AMBP	P02760	T(24)	6 24GPVPTPPDNI ₂₉	HexNAc(1)Hex(1)NeuAc(1)	25.4-25.7	[831.84] ²⁺	1006.51	-0.01	22	36	33	0.55
				HexNAc(1)Hex(1)NeuAc(1)	22.8-23.7	[895.87] ²⁺	1134.58	0.00	8	196	34	0.55
				HexNAc(1)Hex(1)NeuAc(1)	24.1	[895.85] ²⁺	1134.48	-0.10	10	7	80	0.55
				HexNAc(1)Hex(1)NeuAc(1)	23.0-23.5	[597.58] ²⁺	1134.55	-0.03	11	45	117	0.55
				HexNAc(1)Hex(1)NeuAc(1)	27.6-28.1	[945.37] ²⁺	1233.63	-0.02	12	32	-	0.55
				HexNAc(1)Hex(1)NeuAc(1)	27.5-28.5	[630.57] ²⁺	1233.58	-0.07	18	9	-	0.55
17 Protocadherin beta-11	Q9Y5F2	S(24)	1 24LLGMSQ ₂₅	HexNAc(1)Hex(1)NeuAc(1)	10.0-10.4	[652.74] ²⁺	648.35	-0.06	28	14	-	0.04
18 Protocadherin-16	Q96IQ0	T(592)	1 592EGTQPG ₅₉₄	HexNAc(1)Hex(1)NeuAc(1)	11.6-11.8	[622.72] ²⁺	588.26	0.00	29	44	-	0.05
19 Selenoprotein P	P49908	T(236)	1 236APTHPAPPLH ₂₄₄	HexNAc(1)Hex(1)NeuAc(1)	13.8-14.2	[584.24] ²⁺	1094.51	-0.07	43	517	57	0.99
20 Tau-tubulin kinase 2	Q6IQ55	S(817), T(819), T(820)	1 817KDHSAETPEL ₈₂₃	HexNAc(1)Hex(1)NeuAc(1)	19.9-20.0	[877.82] ²⁺	1098.59	0.04	22	313	S(817): 72 / T(819): 285 / T(820): 399	0.77 / 0.94 / 0.90
21 Nascin-R	Q92752	T(1288)	1 1288DNDVAVTNC ₁₃₀₀	HexNAc(1)Hex(1)NeuAc(2)	23.0-23.2	[652.93] ²⁺	1009.47	0.10	18	27	15	0.42
22 Nascin-X	P22105	T(3586)	1 3586TAPAEAP ₃₅₉₆	HexNAc(1)Hex(1)NeuAc(2)	26.8-27.4	[680.61] ²⁺	1092.53	-0.03	15	49	26	0.62
23 not unambiguously identified	-	-	23 23QVPTPPDNI ₂₉	HexNAc(1)Hex(1)NeuAc(1) (18x)	-	-	-	-	-	-	-	-
Coagulation factor V	P12259	S(1455)	1 1455QISPPDL ₁₄₆₀	HexNAc(1)Hex(1)NeuAc(1)	9.0-9.3	[761.78] ²⁺	866.41	-0.06	15	48	-	0.92
Myosin-3	Q5VT55	T(699), S(701)	1 699ALATPSA ₇₀₂	HexNAc(1)Hex(1)NeuAc(1)	8.2-8.3	[643.65] ²⁺	630.30	-0.05	24	28	-	0.51 / 0.51
Ig kappa chain V4 region Ka	P01603	S(7), S(9), S(12), S(14)	1 7QDeamid, SPSTLSV ₁₄	HexNAc(1)Hex(1)NeuAc(1)	13.3	[781.76] ²⁺	906.36	-0.09	32	18	-	0.21 / 0.29 / 0.18 / 0.25

cupancy of Thr₂₅₆, which is in agreement with literature findings. For the two other described fetuin-A O-glycosylation sites Thr₂₇₀ and Ser₃₄₆ ETD fragmentation was actually not mandatory, because corresponding O-glycopeptides were identified that solely harbor one O-glycosylation site (e.g. Thr₂₇₀:₂₆₇AVPTPV₂₇₂, Ser₃₄₆:₃₄₂VVQPSVG₃₄₈), respectively. Also of note, with respect to the peptide identification, b- and y-ions were detected in the CID-MS² fragment spectra of the fetuin-A O-glycopeptides ₂₅₂TQPVT SQPQPE₂₆₂ (*m/z* 623.23³⁺), ₂₆₇AVPTPVDPDAPPSPPL₂₈₃ (*m/z* 872.72³⁺) and ₂₆₆EAVPTPWDPDAPPSPPL₂₈₃ (*m/z* 915.71³⁺), which already permit the unambiguous peptide identification without consideration of CID-MS³ spectra (supplemental Figs. S5–S7). Furthermore, internal glycopeptide fragment ions resulting from concerted fragmentations along the peptide backbone and along the glycan moiety were detected in the same CID-MS² spectra - a low-energy CID glycopeptide fragmentation event that is rarely described in literature (e.g. ₂₅₂TQPVT(HexNAc₁Hex₁NeuAc₁)SQPQPE₂₆₂ → ₂₅₂TQPVT(HexNAc)SQ₂₅₈ *m/z* 945.35¹⁺) (supplemental Fig. S7).

Kininogen-1—The human KNG1 gene codes for two splicing variants of kininogen, namely low-molecular and high-molecular weight kininogen. The latter is involved in blood coagulation and the assembly of the kallikrein-kinin system and was identified in the present study by six O-glycopeptides. Currently nine O-glycosylation sites/regions are described in literature for kininogen-1—presumably all being decorated with mucin-type core 1 or possibly core 8 O-glycans (30, 60–62). Experimental glycoproteomic evidence on the macro and microheterogeneity of kininogen-1 is still missing, though. In the present study, four kininogen-1 O-glycosylation sites, including one novel site (Ser₆₀₄), could be pinpointed and described with respect to the composition of the attached O-glycans (Table I). The identified O-glycopeptide ₆₀₀FNPIISDFPDTT₆₁₀ (*m/z* 734.26³⁺) carries a disialylated T-antigen and harbors three potential O-glycosylation sites. ETD analysis implies the occupancy of Ser₆₀₄, because of the presence of a signal at *m/z* 1638.30¹⁺, corresponding to a c₆ ion (supplemental Fig. S6). Also of note, in previous studies the use of trypsin did not allow to pinpoint occupied O-glycosylation sites in the region aa119–161 (30, 61). Proteinase K, however, generated two distinct O-glycopeptides (₁₃₂EGPWVTA₁₃₈ *m/z* 664.70²⁺, ₁₄₆VHPISTQ₁₅₂ *m/z* 719.23²⁺) that allowed pinpointing the site Thr₁₃₇ and the region Ser₁₅₀/Thr₁₅₁. For the latter, unfortunately, the ETD spectrum quality did not allow localizing the exact site. The peptide ₁₃₂EGPVVTA₁₃₈ (*m/z* 664.70²⁺) could not be identified correctly by MASCOT database search, because of missing fragment ions. However, the peptide could be identified via manual *de novo* annotation supported by mass tag ([283.0 Da]VTA) assisted *de novo* sequencing using the tool MS-Homology (<http://prospector.ucsf.edu/prospector>) (supplemental Fig. S5). The peptide identity was further verified by the identification of the glycosylated peptide ₁₃₂EGPVVTAQ₁₃₈ (*m/z*

874.28²⁺) in a subsequent HILIC fraction (supplemental Fig. S7).

Immunoglobulin J Chain—The immunoglobulin J chain (joining chain) participates in the effective di-/polymerization of either IgA or IgM and is essential for the secretion of these immunoglobulins into the mucosa. In literature the J chain was reported to be N-glycosylated at Asn₄₉ (60, 63, 64); however, O-glycosylation has hitherto not been described for the molecule. Interestingly, two O-glycopeptides detected in HILIC fractions #13 and #14 might correspond to the J chain and suggest O-glycosylation at Thr₉₇ (₉₅DPTEV₉₉ *m/z* 608.72²⁺, *m/z* 608.71²⁺) (supplemental Figs. S3 and S5). This potentially new O-glycosylation site is in close vicinity to a cysteine (Cys₉₁) that can form a disulfide-bridge to IgM molecules. Hence, one might speculate that an occupied O-glycosylation site in this region might function in the establishment/preservation of this inter-molecular bond. However, the number of present fragment ions in the corresponding CID-MS³ spectra did not allow an unambiguous identification of the peptide, as evidenced by several potential peptide hits being equally scored by the search engine. Manual fragment spectra annotation, though, suggest the identification of immunoglobulin J chain—nevertheless, this identification deserves further validation. Both identified O-glycopeptides were found to be decorated with monosialylated T-antigens.

Inter- α -trypsin Inhibitor Heavy Chain H4—For the protease inhibitor inter-alpha-trypsin inhibitor heavy chain H4 two O-glycosylation sites/regions, Ser₆₄₀ and Thr_{722/723} have been described in literature (58, 65). In agreement with recent findings by Chandler *et al.*, Ser₆₄₀ was found to be O-glycosylated. The O-glycopeptide ₆₃₉ASFSPR₆₄₄ (*m/z* 660.72²⁺) harbors two potential O-glycosylation sites and the occupied site could be clearly inferred from the ETD spectra by the presence of a signal at *m/z* 490.22¹⁺, corresponding to a z+1₄ ion (supplemental Fig. S5). In contrast to Chandler *et al.*, but in agreement with Halim *et al.*, ETD data of the O-glycopeptide ₇₂₂TTQTPAPIQAPS₇₃₃ (*m/z* 623.27³⁺) suggested the occupancy of the sites Thr_{722/723} (58, 65) (supplemental Fig. S5). Unfortunately, none of the two potential O-glycosylation sites could be clearly ruled out by the detected fragment ions. Both sites/regions Ser₆₄₀ and Thr_{722/723} were decorated with a monosialylated T-antigen. This contrasts previous findings by Chandler *et al.* who also observed a disialylated T-antigen on S₆₄₀.

Inter- α -trypsin Inhibitor Heavy Chain H2—For the H2 heavy chain of the Inter-alpha-trypsin inhibitor a c-terminal cluster of mono- and disialylated mucin-type core 1 O-glycans (Thr₆₆₆, Ser₆₇₃, Thr₆₇₅ and Thr₆₉₁) has been described in literature (58, 60, 66, 67). These previously reported O-glycosylated sites, except for the site T₆₆₆, could be confirmed by the present study, albeit solely with monosialylated T-antigens. ETD spectra of the O-glycopeptide ₆₈₉ESTPPPHV₆₉₆ (*m/z* 507.15³⁺/760.24²⁺) enabled a clear identification of the occupied O-glycosylation site Thr₆₆₆. This finding is supported, in

DISCUSSION

particular, by a signal detected in the doubly charged species at m/z 1287.41¹⁺ which corresponds to a $z+1_e$ ion (supplemental Fig. S6). Remarkably, the CID-MS² spectrum of the O-glycopeptide ₆₆₉WANPSPTPV₆₇₇ (m/z 760.92³⁺) revealed that both O-glycosylation sites, Ser₆₇₃ and Thr₆₇₅, are occupied by a monosialylated T-antigen (supplemental Fig. S7). Moreover, the spectrum features signals indicating the presence of hexose rearrangement products, that is the transfer of an additional hexose either to the glycan or the peptide moiety, as described earlier (68, 69). The occurrence of these artifacts necessitates the careful interpretation of CID glycopeptide fragment spectra.

τ -Tubulin Kinase 2—The τ -tubulin kinase 2 (TTBK2) phosphorylates τ and tubulin, preferably in the nervous system. Aberrant TTBK2 activity was linked to the progression of the Alzheimer's disease (70, 71). The protein resides primarily in the cytosol; however, Böhm *et al.* could also detect TTBK2 in a secreted form in human tears (72, 73). Hitherto, no glycosylation of this protein has been described. CID-MS³ as well as ETD spectra of the O-glycopeptide ₈₁₄KDHSATTEPL₈₂₃+HexNAc₁Hex₁NeuAc₁ (m/z 877.82²⁺), though, suggest the O-glycosylation of Thr₈₂₀. ETD fragment ions at m/z 485.22¹⁺, 557.43¹⁺, and 1098.63¹⁺, corresponding to c_4 , $c+1_s$, and z_4 ions, allowed discerning the exact glycosylation site. As TTBK2 is involved in ciliogenesis (74, 75), a process which requires the vesicle transport from the Golgi to the basal bodies and cilia, we speculate that TTBK2 might become O-glycosylated during this process.

Fibrinogen α and β Chain—The blood clotting protein fibrinogen is known to be N-glycosylated at the β - and γ -chain. Interestingly, a recent study by Zauner *et al.* could also show O-glycosylated sites and regions, seven in total, within the molecule (51). In the present study O-glycosylation of the fibrinogen alpha region aa524–528 could be confirmed; pinpointing the exact O-glycosylation site was not possible, though (supplemental Fig. S5, ₅₂₄STGKTFPG₅₃₁, m/z 725.78²⁺). Nevertheless, O-glycosylation within the fibrinogen beta region aa58–67 could be confirmed and pinpointed. Here, the presence of the ETD fragment ions m/z 931.54¹⁺, 1300.54¹⁺, and 1915.50¹⁺, corresponding to $z+1_g$, c_6 , and c_{12} ions (supplemental Fig. S6), ₅₄EEAPSLRPAPPIS₆₇, m/z 706.27³⁺) indicates O-glycosylation at the site Ser₅₈. This contrasts recent findings by Bai *et al.* who reported the site Ser₆₇ to be O-glycosylated, but not the site Ser₅₈ (76). In agreement with previous findings, both fibrinogen O-glycopeptides (₅₂₄STGKTFPG₅₃₁, m/z 725.78²⁺, ₅₄EEAPSLRPAPPIS₆₇, m/z 706.27³⁺), detected in the present study, were found to be decorated with monosialylated T-antigens. Interestingly, the peptide ₅₄EEAPSLRPAPPIS₆₇ was also found in its nonglycosylated form (HILIC fractions #12–#15, CID, see supplemental Table S2), which suggests only a partial occupation of the O-glycosylation site Ser₅₈.

Over the last few years mass spectrometry based glycoproteomics has experienced significant advances in terms of instrumentation, methodology and bioinformatics; resulting in a variety of excellent glycoproteomic publications that highlight the merits of high resolution mass spectra, complementary fragmentation techniques, improved multidimensional glycopeptide enrichment and separation techniques as well as sophisticated software tools (41). However, despite these advances—and despite its enormous clinical and pharmaceutical relevance as well as diagnostic potential—our knowledge about the human blood plasma glycoproteome is still very limited. This holds particularly true for the human blood plasma O-glycoproteome. Here several important questions can be raised: Which proteins are O-glycosylated?, Which O-glycans are attached to which sites?, Which dynamics in terms of the O-glycan micro- and macroheterogeneity can be observed in a certain biological context?, What are the biological and biotechnological implications of O-glycosylation?

In the present study we have developed and employed an analytical workflow that allows the explorative, nontargeted analysis of the human blood plasma O-glycoproteome in a site-specific manner. To this end intact human blood plasma O-glycopeptides, generated by a broad-specific proteolytic digest via Proteinase K, were selectively enriched using HILIC fractionation in order to be analyzed by multistage nanoRP-LC-ESI-IT-MS using low-energy CID as well as ETD (CID-MS²/MS³, ETD-MS²). This combined workflow was applied on a pooled blood plasma sample derived from 20 healthy donors and allowed for the identification of 31 O-glycosylation sites in 22 proteins, including the detection of 11 previously unknown O-glycosylation sites. We were able to pinpoint 23 O-glycosylation sites, of which eight sites have been described for the first time. The identified O-glycan compositions most probably correspond to mono- and disialylated core-1 mucin-type O-glycans (T-antigen).

Other O-glycoproteomic Studies on Complex Biofluids—In the recent past efforts have been made to investigate the O-glycoproteome of different complex biological samples. Halim *et al.*, for instance, analyzed the O-glycoproteome of cerebrospinal fluid (CSF) using a sialic-acid capture-and-release protocol (30). This protocol is based on the sialic acid specific hydrazide capturing of periodate oxidized glycoproteins. Upon tryptic digestion the protocol allows the acid hydrolysis of sialic acid glycosidic bonds in order to release and analyze (formerly) sialylated glycopeptides. To focus on O-glycosylations the authors included a peptide N-glycosidase F (PNGase F) sample pretreatment step to remove N-glycans. The authors have used an automated CID-MS²/MS³ spectra search protocol for glycopeptide identification (Peptide-GalNAc-Gal) and have employed ECD and ETD to pinpoint the glycosylation sites. In total they have identified 106 O-glycosylation sites and could pinpoint 67 of these. The

identified CSF O-glycopeptides belong to 49 different proteins and were predominately decorated with structures corresponding to core-1 mucin-type O-glycans. In a previous study the same group has also investigated the human urinary N- and O-glycoproteome using the sialic-acid capture-and-release protocol (58). Unfortunately, the applied protocol does not allow the enrichment of nonsialylated glycoproteins nor does it give any information on the degree of sialylation of the attached O-glycan moieties. This limits the applicability of this procedure, as the degree of O-glycan sialylation is a crucial determinant in the pathogenesis of a number of diseases (22).

In another large-scale glycoproteomics study conducted by Hägglund *et al.* in 2007 human plasma proteins, derived from Cohn fraction IV of a plasma fractionation, were analyzed with respect to occupied N- and O-glycosylation sites (60). The analyzed Cohn fraction is supposed to contain mainly α -globulins, like plasminogen and haptoglobin, and is depleted from γ -globulins and serum albumin. The authors have employed two different enzymatic deglycosylation strategies to pinpoint occupied N-glycosylation sites: (1) PNGase F + $H_2^{18}O$; (2) endo- β -N-acetylglucosaminidases (Endo D and Endo H) + different exoglycosidases. These two strategies were applied on HILIC enriched tryptic (glyco)peptides, that were fractionated by strong cation exchange chromatography and eventually measured by LC-ESI-MS/MS using high-energy CID. The authors were able to identify 103 N-glycosylation sites as well as 23 O-glycosylation sites/regions derived from 61 and 11 human blood plasma proteins, respectively. Unfortunately, the occupied O-glycosylation sites could not be pinpointed and no information on the glycan moiety could be deduced.

In 2012 Darula *et al.* reported on the O-glycoproteomic analysis of bovine serum (77). In this study the authors have combined different protein- and peptide-level prefractionation and enrichment strategies, including jacalin lectin affinity chromatography, mixed-mode chromatography, and electrostatic repulsion hydrophilic interaction chromatography (ERLIC) to enrich tryptic mucin-type O-glycopeptides. After additional use of exoglycosidases to improve glycopeptide characterization, truncated glycopeptides were subjected to LC-ESI-MS/MS with HCD and ETD for automated peptide identification and glycosylation site determination. Overall, the authors could identify and pinpoint 124 glycosylation sites in 51 proteins, including many O-glycosylation sites that have not been described before—unfortunately, though, at the expense of the intact glycan structure.

In a recent publication from Bai *et al.* an analytical workflow is presented, which allows the mapping of mucin-type O-glycosylation sites on glycoproteins present in human blood plasma (76). The authors have used jacalin lectin affinity chromatography to enriched tryptic O-glycopeptides (peptide + GalNAc), which were treated with PNGase F and different exoglycosidases. In this study 49 O-glycopeptides, belonging to 36 human blood plasma glycoproteins, were identified by LC-ESI-MS/MS (CID). Overall, the authors could assign 13

O-glycosylation sites unambiguously, of which nine sites have not been described before.

Proteinase K Digest—The majority of large-scale glycoproteomic studies features trypsin for the generation of (glyco)peptides. Trypsin is the proteolytic gold standard in LC-MS/MS based peptide identification and quantification, as it reproducibly generates predictable peptides that can be readily retained on reversed-phase column and that give enough fragment ions for an unambiguous peptide identification, in most cases. In terms of glycoproteomics, though, the cleavage specificity of trypsin can be a limiting factor for the identification and the localization of certain glycosylation sites, in particular for densely clustered O-glycosylation sites. Hence, the use of broad- and nonspecific proteases, like Pronase E or Proteinase K was proposed, to reduce the number of nonglycosylated peptides and to make certain glycosylation sites analytically amenable (34). Proteinase K, for instance, has been successfully used in a number of publications that are centered on the O-glycoproteomic analysis of single proteins; though, the use of Proteinase K in large-scale glycoproteomic studies on complex samples has not been described so far. In the present study we could show that Proteinase K generates (glyco)peptides from a complex sample, like human blood plasma, in a reproducible and non-random manner, which is in agreement with a report from Hua *et al.* (34). We could show that, most of the time, Proteinase K generates shorter peptides compared with trypsin (supplemental Fig. S7), and that Proteinase K cleaves effectively in-between densely O-glycosylated regions—thus, rendering the determination of the occupied O-glycosylation site(s) less difficult. In fact we could show that Proteinase K can generate O-glycopeptides that exhibit only one potential O-glycosylation site, thus allowing for an unambiguous localization of the occupied site. We could clearly show that some O-glycosylation sites could only be identified and pinpointed by the use of Proteinase K, because tryptic peptides would have been too long and would have harbored too many potential O-glycosylation sites.

Glycopeptide Enrichment Via HILIC—Glycopeptides are usually under-represented in a peptide mixture, because of the glycan microheterogeneity. In a tryptic digest of a typical glycoprotein only about 2% to 5% of the peptides are glycopeptides (78). In addition, the ionization efficiency of glycopeptides is significantly lower compared with their nonglycosylated counterparts, thus making the efficient and selective enrichment of glycopeptides key to most glycoproteomics workflows. The use of HILIC based glycopeptide enrichment methods has proven to be a vital tool in glycoproteomics because of their broad glycan specificity, reproducibility and compatibility with mass spectrometry. In a previous report by Zauner *et al.* it could be shown, that Proteinase K-generated glycopeptides can be separated into earlier eluting O-glycopeptides and later eluting N-glycopeptides using HILIC (32). Based on this publication we have employed HILIC for the

selective enrichment and fractionation of human blood plasma O-glycopeptides. Here of particular importance is the removal of highly abundant nonglycosylated peptides derived from albumin and other major (glyco-)proteins. Careful manual inspection of CID-MS² fragment spectra of the acquired HILIC fractions revealed the efficient enrichment of glycopeptides - and indeed the presence of solely mucin-type core-1 O-glycosylated glycopeptides. N-glycopeptides were not detected, as they were expected to be present in the late eluting HILIC wash fractions because of their generally higher hydrophilicity compared with the most commonly found forms of mucin-type O-glycopeptides (non-, mono- and disialylated core-1 and -2 O-glycopeptides).

Identification of the O-glycan Composition—For an automated glycopeptide spectra filtering and glycan fragment annotation the use of commercial software tools was considered, but turned out to be too error-prone in our case (data not shown). Hence, in the present work we relied on manual annotation and interpretation of low-energy CID-MS² fragment spectra in order to elucidate the O-glycan composition—however, at the expense of throughput and the possibility to report false discovery rates. In total we were able to characterize 88 O-glycopeptides with respect to their O-glycan composition. The detected O-glycan compositions most likely correspond to mucin-type core-1 mono- and disialylated O-glycans ((di)sialyl-T-antigen). In agreement with literature, glycopeptides carrying disialylated O-glycans, were found in later eluting HILIC fractions (#15–#17), as the additional sialic acid renders the molecule more hydrophilic. Mono- and disialylated glycoforms could be usually discriminated by the presence of distinct oxonium ions: whereas fragmentation of monosialylated O-glycans generated a characteristic oxonium ion at m/z 454.16 (Hex₁NeuAc₁), disialylated O-glycans gave rise to an additional intense peak at m/z 495.18 (HexNAC₁NeuAc₁) (supplemental Fig. S6, ²⁶⁶EAVPTPVDPDAPSPPL₂₈₃, m/z 818.68³⁺, ²⁶⁷AVPTPVDPDAPSPPL₂₈₃, m/z 872.73³⁺). Furthermore, in disialylated species characteristic fragment ions of the peptide+HexNAC+NeuAc were observed. In a few cases the glycan annotation was compromised by the presence of fragment ions corresponding to hexose rearrangement products (68, 69). Generally, it is important to note, that low-energy CID-MS² fragmentation of glycopeptides does usually not produce fragment ions that relate to the linkage of the attached monosaccharides. Therefore, validation of the inferred O-glycan structures using dedicated O-glycomics approaches, including for instance (reductive) beta-elimination or hydrazinolysis, is recommended. However, our findings are in good agreement with literature, as mono- and disialylated mucin-type core-1 O-glycans are known to be present on the majority of secreted blood plasma glycoproteins, produced by hepatic cells of healthy individuals (79). Notably, a study on plasma-derived von Willebrand factor could show, that apart from mucin-type core 1 O-glycans (T-antigen), more complex O-glycan structures in-

cluding ABH blood group antigen containing mucin-type core-2 ((GalNAcβ1–6-(Galβ1–3)-GalNAcα-O-Ser/Thr)), can be present on human blood plasma glycoproteins, too (80). In the present work, analyzing the total human blood plasma O-glycoproteome, we could not detect any (glyco)peptide derived from von Willebrand factor, nor could we find any indication for the presence of fucosylated (ABH blood group antigens) and/or LacNAc extended mucin-type core-2 O-glycans.

Glycopeptide Identification—Low-energy CID-MS² fragmentation of glycopeptides, as employed in the present work, almost exclusively generates fragment ions corresponding to the fragmentation of the glycan moiety, while leaving the peptide backbone mainly intact. Thus, this type of fragmentation does usually not provide any information on the sequence of the peptide backbone nor on the occupied glycosylation site. To identify the peptide we have employed manual CID-MS³ fragmentation on the putative peptide mass, which has been inferred from the annotation of the corresponding CID-MS² spectra before. In a few of cases the signal of the putative peptide mass was too low to yield sufficient fragment ions. Consequently, the putative peptide+HexNAC ion was subjected to CID-MS³ fragmentation instead. We did not employ an automated CID-MS³ fragmentation procedure, e.g. fragmentation of the three most intense precursor ions in the CID-MS² spectrum, because we wanted to generate and sum up as many fragment spectra as possible from the selected putative peptide mass, in order to increase spectra quality and therefore the chance of successful peptide identification. By searching the acquired CID-MS³ fragment spectra against the human subset of the UniProtKB/Swiss-Prot protein database, a total of 60 peptides (of 88 detected O-glycopeptides) could be identified unambiguously. Notably, in a few cases also peptide fragment ions present in CID-MS² spectra allowed for an unambiguous peptide identification (supplemental Fig. S4, ²⁶⁷AVPTPVDPDAPSPPL₂₈₃, m/z 872.73³⁺). Overall, the identified peptides belong to 22 different proteins—primarily acute phase proteins. This constantly growing group of blood plasma proteins fulfills essential functions during inflammation (e.g. coagulation, anti-inflammatory and anti-pathogenic activity), and, accordingly, their expression is known to be either significantly up- or downregulated (positive and negative acute phase proteins) in this context. As a result, this group of proteins attracted a lot of attention as potential cancer biomarkers in recent years (5). Noteworthy, the identified proteins span a concentration range of 5 orders of magnitude. Therefore, the applied approach seems to be suitable to also detect lower abundant proteins or peptides.

A group of O-glycosylated proteins that have frequently been identified in other large-scale glycoproteomic studies are Coagulation factors (30, 58, 60, 77). In our study there is an indication for the presence of an O-glycosylated peptide derived from Coagulation factor V (HILIC fraction #15, m/z 761.78²⁺, ¹⁴⁵³QISPPDL₁₄₆₀+HexNAC₁Hex₁NeuAc₁, Table

II, supplemental Fig. S5). Interestingly, the detected Coagulation factor V O-glycosylation site (Ser₁₄₅₅) has not been described so far. Unfortunately, our data do not allow an unambiguous identification of this protein.

General Remarks on Immunoglobulin O-glycoproteomics—Another O-glycosylated protein that could not be identified in our study is Ig α -1 (IgA1). IgA1 is a high abundant human blood plasma glycoprotein that features a cluster of three to five mucin-type O-glycans in the hinge region of the heavy chain (81). This cluster harbors many prolines, hence corresponding Proteinase K generated peptides might have been not unambiguously identified (the tryptic IgA1 hinge region O-glycopeptide looks as follows: (K)₈₉HYTNPSQDVTVP-**PVPSTPPTSPSTPPTSPSCCHPR**₁₂₆). Furthermore, because of the densely clustered O-glycans a potential IgA1 O-glycopeptide carrying mucin-type O-glycans at each potential site, such as **PSTPPTSPSTPPTSPSCC**, might be too hydrophilic and consequently might have been among the (glyco)peptides present in the late eluting HILIC wash fraction. Worth mentioning, in our study we could detect the IgA1 peptide ₉₅QDVTVPCPVPS₁₀₅ in its nonglycosylated form (HILIC Fraction #11, CID, supplemental Table S2). Therefore, the O-glycosylation site S₁₀₅ seems to be only partially occupied. Surprisingly, human IgA1 O-glycopeptides have not been identified in any other large-scale glycoproteomic studies (30, 58, 60, 76, 77, 82). However, there is a targeted glycoproteomic study from Takahashi *et al.* focusing on IgA1 O-glycosylation (81). In this study the authors analyzed human plasma derived IgA1 O-glycopeptides (tryptic and nontryptic) with ESI-FT-ICR-MS/MS as well as ESI-LTQ-FT-MS/MS, both in online- and offline-Mode. To pinpoint the O-glycosylation sites the authors have employed activated ion-electron capture dissociation (AI-ECD) and ETD. Another immunoglobulin that is reported to carry mucin-type O-glycans in the hinge region is Ig delta (IgD) (83). The plasma concentration of IgD is much lower than the concentration of IgA, IgG, and IgM but higher than that of IgE (IgD represents 0.25% of total plasma immunoglobulins). Apart from the study conducted by Takayasu *et al.* from 1982 (83) on truncated O-glycopeptides (peptide+GalNAc), at present no O-glycoproteomic data do exist for intact human IgD O-glycopeptides. Also of particular interest is a recent finding by Plomp *et al.*: using a targeted glycoproteomics approach these authors could demonstrate, for the first time, that IgG3 is partially O-glycosylated in its hinge region (mucin-type core-1 O-glycans) (84).

Pinpointing of O-glycosylation Sites—Pinpointing the correct O-glycosylation sites is a crucial but very challenging task. Proteinase K, in this regard, proved to be beneficial as it can generate short glycopeptides, which exhibit only one potential O-glycosylation site. In case the occupied O-glycosylation site could not be inferred directly, we have employed ETD-MS² fragmentation. In first attempts database-assisted peptide identification via MASCOT was tested on the acquired ETD glycopeptide spectra, but turned out to be not

successful. One reason for this is the presence of intense signals in the ETD-MS² spectrum, which do not correspond to peptide fragment ions (e.g. unfragmented precursor ions, glycan fragment ions), and which thus compromise automated peptide identification (85). A possible solution for this is the (manual) removal of these additional *m/z*-values from the ETD-spectra before running the search algorithm. In the present study, however, this procedure did not improve the database-assisted peptide identification. For these reasons we relied on manual spectra annotation and interpretation using *DataAnalysis*, *Biotoools* as well as public repositories (UniProtKB and UniCarbKB). Furthermore, NetOGlyc 4.0 was employed to predict O-glycosylation sites and to support experimental findings. Predicted and experimentally determined O-glycosylation sites were mostly in good agreement for already known O-glycosylation sites—however, support for potentially novel sites could only be found in a few cases. A general shortcoming of glycopeptide enrichment methods is that they are biased toward glycosylated peptides, while underrepresenting potential corresponding aglycosylated counterparts. Hence, in the present study no conclusions with respect to the macro-heterogeneity of the glycoproteins (site-occupancy) can be drawn.

Caveats of the Approach—In contrast to tryptic (glyco)peptides, Proteinase K generated peptides and glycopeptides cannot be predicted because of the broad cleavage specificity of the enzyme. More importantly, though, is the reduced peptide length compared with a tryptic digest, as this can lead to an insufficient number of detected fragment ions to allow for unambiguous peptide identifications. This problem can be even more intensified by the frequent occurrence of prolines within mucin-type O-glycopeptide sequences, as prolines can introduce additional sequence gaps during mass spectrometry-based peptide sequencing. Also important to note is the increased search space of the search engine because of the use of a nonspecific enzyme, which results in an increased ambiguity with respect to the peptide identification (lower identification scores) and longer search times. A confounding factor that relates to the ETD analysis is the predominance of charge state 2⁺ among the measured O-glycopeptide precursor ions, because ETD fragmentation is more efficient for precursor charge states greater than 2⁺ (86). The predominance of charge state 2⁺ can be explained by a lack of ionizable/basic amino acids (lack of Arg, Lys, His) within the glycopeptides—a characteristic that can be linked to the broad-specific proteolytic digest by Proteinase K (87). Another caveat is related to the HILIC glycopeptide enrichment: this step was optimized to enrich O-glycopeptides carrying short mucin-type core-1 and -2 O-glycans, as they represent the vast majority of O-glycans on human blood plasma proteins (25). Hence, O-glycopeptides carrying bigger and thus more hydrophilic O-glycans, such as *N*-acetyl-lactosamine (LacNAc) extended mucin-type core-2 O-glycans, or O-glycopeptides carrying multiple mucin-type O-glycans, might elute in the sub-

sequent washing phase of the HILIC fractionation and as a consequence cannot be found during the analysis.

Summary and Outlook—In the present study we have investigated the human blood plasma mucin-type O-glycoproteome of healthy individuals in an explorative and nontargeted manner. To this end, we have conducted a site-specific large-scale O-glycoproteomic analysis, which combines a broad-specific proteolytic digest, with HILIC enrichment/fractionation and subsequent multistage mass spectrometry measurement (nano-RPLC-ESI-IT-MSⁿ) with CID and ETD. Centered on the characterization and identification of intact glycopeptides, we could demonstrate the in-depth O-glycoproteomic analysis of a number of important human blood plasma glycoproteins (mainly acute phase proteins), including alpha-2-HS-glycoprotein, fibrinogen, plasminogen and kininogen-1. Our results are in good agreement with previous findings by other research groups, but also add new aspects to the field, e.g. the identification of a couple of novel O-glycosylation site as well as the benefits and drawbacks of using Proteinase K in large-scale mass spectrometric glycoproteomic studies.

Explorative site-specific N- and O-glycoproteomic studies of biofluids, like human blood plasma, human milk, urine or cerebrospinal fluid hold an enormous potential to better understand the implications of protein glycosylation under normal physiological conditions, but also under pathophysiological conditions. By serving as a diagnostic tool, the detection/discovery of relevant glycopeptides (biomarker candidates) can be the basis for targeted quantitative glycoproteomic analyses, which allow for a site-specific monitoring of glycosylation alterations, e.g. during disease progression. Site-specific glycosylation analyses are, moreover, important to produce biopharmaceuticals according to quality by design requirements, in particular if these biopharmaceuticals are produced in heterologous expression systems. In this regard site-specific glycosylation analyses might also enable understanding/controlling important glycan-related features of the final product including its efficacy, half-life, or antigenicity.

Acknowledgments—We thank Dr. Gerhild Zauner and Dr. Maurice H.J. Selman for fruitful discussions.

* M.H., M.W., and E.R. acknowledge support by the European Union (Seventh Framework Programme HighGlycan project, grant number 278535).

§ This article contains supplemental Figs. S1 to S7 and Tables S1 to S3.

¶ To whom correspondence should be addressed: Dr. Erdmann Rapp, Max Planck Institute for Dynamics of Complex Technical Systems Bioprocess Engineering, Head of Bio/Process Analytics Sandtorstrasse 1, Magdeburg, 39106, Germany. Tel.: +49 391 6110 314; E-mail: rapp@mpi-magdeburg.mpg.de.

REFERENCES

- Anderson, N. L. (2002) The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol. Cell. Proteomics* **1**, 845–867
- Schaller, J., Gerber, S., Kämpfer, U., Lejon, S., and Trachsel, C. (2008) Blood Plasma Proteins. *Human Blood Plasma Proteins*, pp. 17–20, John Wiley & Sons, Ltd
- Anderson, L. (2014) Six decades searching for meaning in the proteome. *J. Proteomics* **107**, 24–30
- Ceciliani, F., and Pocacqua, V. (2007) The acute phase protein alpha1-acid glycoprotein: a model for altered glycosylation during diseases. *Curr. Protein Pept. Sci.* **8**, 91–108
- Polanski, M., and Anderson, N. L. (2007) A list of candidate cancer biomarkers for targeted proteomics. *Biomark Insights* **1**, 1–48
- Anderson, L., and Hunter, C. L. (2006) Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* **5**, 573–588
- Farah, T., Deutsch, E. W., Omenn, G. S., Campbell, D. S., Sun, Z., Bletz, J. A., Mallick, P., Katz, J. E., Malmstrom, J., Ossola, R., Watts, J. D., Lin, B., Zhang, H., Moritz, R. L., and Aebersold, R. (2011) A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol. Cell. Proteomics* **10**, M110 006353
- Haab, B. B., Geierstanger, B. H., Michailidis, G., Vitzthum, F., Forrester, S., Okon, R., Saviranta, P., Brinker, A., Sorette, M., Perlee, L., Suresh, S., Drwal, C., Adkins, J. N., and Omenn, G. S. (2005) Immunoassay and antibody microarray analysis of the HUPO Plasma Proteome Project reference specimens: Systematic variation between sample types and calibration of mass spectrometry data. *Proteomics* **5**, 3278–3291
- Hortin, G. L., Sviridov, D., and Anderson, N. L. (2008) High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance. *Clin. Chem.* **54**, 1608–1616
- Schenk, S., Schoenhals, G. J., de Souza, G., and Mann, M. (2008) A high confidence, manually validated human blood plasma protein reference set. *BMC Med Genomics* **1**, 41
- Zhou, M., Lucas, D. A., Chan, K. C., Issaq, H. J., Petricoin lii, E. F., Liotta, L. A., Veenstra, T. D., and Conrads, T. P. (2004) An investigation into the human serum “interactome”. *Electrophoresis* **25**, 1289–1298
- Carrascal, M., Gay, M., Ovelleiro, D., Casas, V., Gelpi, E., and Abian, J. (2010) Characterization of the human plasma phosphoproteome using linear ion trap mass spectrometry and multiple search engines. *J. Proteome Res.* **9**, 876–884
- Zawadzka, A. M., Schilling, B., Cusack, M. P., Sahu, A. K., Drake, P., Fisher, S. J., Benz, C. C., and Gibson, B. W. (2014) Phosphoprotein secretome of tumor cells as a source of candidates for breast cancer biomarkers in plasma. *Mol. Cell. Proteomics* **13**, 1034–1049
- Kim, Y. J., Zaidi-Ainouch, Z., Gallien, S., and Domon, B. (2012) Mass spectrometry-based detection and quantification of plasma glycoproteins using selective reaction monitoring. *Nat. Protoc.* **7**, 859–871
- Pasing, Y., Sickmann, A., and Lewandrowski, U. (2012) N-glycoproteomics: mass spectrometry-based glycosylation site annotation. *Biol. Chem.* **393**, 249–258
- Lee, H. J., Cha, H. J., Lim, J. S., Lee, S. H., Song, S. Y., Kim, H., Hancock, W. S., Yoo, J. S., and Paik, Y. K. (2014) Abundance-ratio-based semi-quantitative analysis of site-specific N-linked glycopeptides present in the plasma of hepatocellular carcinoma patients. *J. Proteome Res.* **13**, 2328–2338
- Zielinska, D. F., Gnad, F., Wisniewski, J. R., and Mann, M. (2010) Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897–907
- Pacchiarotta, T., Hensbergen, P. J., Wührer, M., van Nieuwkoop, C., Nevedomskaya, E., Derks, R. J., Schoenmaker, B., Koeleman, C. A., van Dissel, J., Deelder, A. M., and Mayboroda, O. A. (2012) Fibrinogen alpha chain O-glycopeptides as possible markers of urinary tract infection. *J. Proteomics* **75**, 1067–1073
- Gomes, C., Almeida, A., Ferreira, J. A., Silva, L., Santos-Sousa, H., Pinto-de-Sousa, J., Santos, L. L., Amado, F., Schwientek, T., Levery, S. B., Mandel, U., Clausen, H., David, L., Reis, C. A., and Osorio, H. (2013) Glycoproteomic analysis of serum from patients with gastric precancerous lesions. *J. Proteome Res.* **12**, 1454–1466
- Cazet, A., Julien, S., Bobowski, M., Burchell, J., and Delannoy, P. (2010) Tumour-associated carbohydrate antigens in breast cancer. *Breast Cancer Res.* **12**, 204
- Dube, D. H., and Bertozzi, C. R. (2005) Glycans in cancer and inflammation—potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* **4**, 477–488
- Rhodes, J., Campbell, B. J., and Yu, L.-G. (2010) Glycosylation and Disease. eLS, John Wiley & Sons, Ltd
- Ju, T., Wang, Y., Aryal, R. P., Lehoux, S. D., Ding, X., Kudelka, M. R., Cutler,

- C., Zeng, J., Wang, J., Sun, X., Heimburg-Molinaro, J., Smith, D. F., and Cummings, R. D. (2013) Tn and sialyl-Tn antigens, aberrant O-glycomics as human disease markers. *Proteomics Clin. Appl.* **7**, 618–631
24. Hanisch, F. G. (2001) O-glycosylation of the mucin type. *Biol. Chem.* **382**, 143–149
25. Yabu, M., Korekane, H., and Miyamoto, Y. (2014) Precise structural analysis of O-linked oligosaccharides in human serum. *Glycobiology* **24**, 542–553
26. Jentoft, N. (1990) Why are proteins O-glycosylated? *Trends Biochem. Sci.* **15**, 291–294
27. Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Bennett, E. P., Mandel, U., Brunak, S., Wandall, H. H., Lavery, S. B., and Clausen, H. (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **32**, 1478–1488
28. Jensen, P. H., Kolarich, D., and Packer, N. H. (2010) Mucin-type O-glycosylation—putting the pieces together. *FEBS J.* **277**, 81–94
29. Strous, G. J., and Dekker, J. (1992) Mucin-type glycoproteins. *Crit. Rev. Biochem. Mol. Biol.* **27**, 57–92
30. Halim, A., Ruetschi, U., Larson, G., and Nilsson, J. (2013) LC-MS/MS characterization of O-glycosylation sites and glycan structures of human cerebrospinal fluid glycoproteins. *J. Proteome Res.* **12**, 573–584
31. Julenius, K., Molgaard, A., Gupta, R., and Brunak, S. (2005) Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology* **15**, 153–164
32. Zauner, G., Koelman, C. A., Deelder, A. M., and Wührer, M. (2010) Protein glycosylation analysis by HILIC-LC-MS of Proteinase K-generated N- and O-glycopeptides. *J. Sep. Sci.* **33**, 903–910
33. Nwosu, C. C., Seipert, R. R., Strum, J. S., Hua, S. S., An, H. J., Zivkovic, A. M., German, B. J., and Lebrilla, C. B. (2011) Simultaneous and extensive site-specific N- and O-glycosylation analysis in protein mixtures. *J. Proteome Res.* **10**, 2612–2624
34. Hua, S., Hu, C. Y., Kim, B. J., Totten, S. M., Oh, M. J., Yun, N., Nwosu, C. C., Yoo, J. S., Lebrilla, C. B., and An, H. J. (2013) Glyco-analytical multispecific proteolysis (Glyco-AMP): a simple method for detailed and quantitative Glycoproteomic characterization. *J. Proteome Res.* **12**, 4414–4423
35. Alpert, A. J., Shukla, M., Shukla, A. K., Zieske, L. R., Yuen, S. W., Ferguson, M. A. J., Mehler, A., Pauly, M., and Orlando, R. (1994) Hydrophilic-interaction chromatography of complex carbohydrates. *J. Chromatogr. A* **676**, 191–202
36. Zauner, G., Deelder, A. M., and Wührer, M. (2011) Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics. *Electrophoresis* **32**, 3456–3466
37. Hanisch, F. G. (2012) O-glycoproteomics: Site-specific O-glycoprotein analysis by CID/ETD electrospray ionization tandem mass spectrometry and top-down glycoprotein Sequencing by In-Source Decay MALDI Mass Spectrometry. *Methods Mol. Biol.*, pp. 179–189
38. Alley, W. R., Jr., Mechref, Y., and Novotny, M. V. (2009) Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun. Mass Spectrom.* **23**, 161–170
39. Saba, J., Dutta, S., Hemenway, E., and Viner, R. (2012) Increasing the Productivity of Glycopeptides Analysis by Using Higher-Energy Collision Dissociation-Accurate Mass-Product-Dependent Electron Transfer Dissociation. *Int. J. Proteomics* **2012**, 7
40. Singh, C., Zampronio, C. G., Creese, A. J., and Cooper, H. J. (2012) Higher energy collision dissociation (HCD) product ion-triggered electron transfer dissociation (ETD) mass spectrometry for the analysis of N-linked glycoproteins. *J. Proteome Res.* **11**, 4517–4525
41. Thaysen-Andersen, M., and Packer, N. H. (2014) Advances in LC-MS/MS-based glycoproteomics: Getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. *Biochim. Biophys. Acta* **1844**, 1437–1452
42. Lavery, S. B., Steentoft, C., Halim, A., Narimatsu, Y., Clausen, H., and Vakhrushev, S. Y. (2014) Advances in mass spectrometry driven O-glycoproteomics. *Biochim. Biophys. Acta* **1850**, 33–42
43. Dallas, D. C., Martin, W. F., Hua, S., and German, J. B. (2013) Automated glycopeptide analysis—review of current state and future directions. *Brief. Bioinform.* **14**, 361–374
44. Wu, S. W., Pu, T. H., Viner, R., and Khoo, K. H. (2014) Novel LC-MS2 product dependent parallel data acquisition function and data analysis workflow for sequencing and identification of intact glycopeptides. *Anal. Chem.* **86**, 5478–5486
45. Kolarich, D., Rapp, E., Struwe, W. B., Haslam, S. M., Zaia, J., McBride, R., Agravat, S., Campbell, M. P., Kato, M., Ranzinger, R., Kettner, C., and York, W. S. (2013) The minimum information required for a glycomics experiment (MIRAGE) project: improving the standards for reporting mass-spectrometry-based glycoanalytic data. *Mol. Cell. Proteomics* **12**, 991–995
46. York, W. S., Agravat, S., Aoki-Kinoshita, K. F., McBride, R., Campbell, M. P., Costello, C. E., Dell, A., Feizi, T., Haslam, S. M., Karlsson, N., Khoo, K. H., Kolarich, D., Liu, Y., Novotny, M., Packer, N. H., Paulson, J. C., Rapp, E., Ranzinger, R., Rudd, P. M., Smith, D. F., Struwe, W. B., Tiemeyer, M., Wells, L., Zaia, J., and Kettner, C. (2014) MIRAGE: the minimum information required for a glycomics experiment. *Glycobiology* **24**, 402–406
47. Selman, M. H., Hemayatkar, M., Deelder, A. M., and Wührer, M. (2011) Cotton HILIC SPE microtips for microscale purification and enrichment of glycans and glycopeptides. *Anal. Chem.* **83**, 2492–2499
48. Domon, B., and Costello, C. E. (1988) A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **5**, 397–409
49. Neue, K., Mormann, M., Peter-Katalinic, J., and Pohlentz, G. (2011) Elucidation of glycoprotein structures by unspecific proteolysis and direct nanoESI mass spectrometric analysis of ZIC-HILIC-enriched glycopeptides. *J. Proteome Res.* **10**, 2248–2260
50. Nwosu, C. C., Huang, J., Aldredge, D. L., Strum, J. S., Hua, S., Seipert, R. R., and Lebrilla, C. B. (2013) In-gel nonspecific proteolysis for elucidating glycoproteins: A method for targeted protein-specific glycosylation analysis in complex protein mixtures. *Anal. Chem.* **85**, 956–963
51. Zauner, G., Hoffmann, M., Rapp, E., Koelman, C. A., Dragan, I., Deelder, A. M., Wührer, M., and Hensbergen, P. J. (2012) Glycoproteomic analysis of human fibrinogen reveals novel regions of O-glycosylation. *J. Proteome Res.* **11**, 5804–5814
52. Plomp, R., Hensbergen, P. J., Rombouts, Y., Zauner, G., Dragan, I., Koelman, C. A. M., Deelder, A. M., and Wührer, M. (2014) Site-specific N-glycosylation analysis of human immunoglobulin e. *J. Proteome Res.* **13**, 536–546
53. Nesvizhskii, A. I., and Aebersold, R. (2005) Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* **4**, 1419–1440
54. Schwartz, B. L., and Bursey, M. M. (1992) Some proline substituent effects in the tandem mass spectrum of protonated pentaalanine. *Biol. Mass Spectrom.* **21**, 92–96
55. Burkhardt, J. M., Schumbrutzki, C., Wortelkamp, S., Sickmann, A., and Zahedi, R. P. (2012) Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *J. Proteomics* **75**, 1454–1462
56. Mormann, M., Paulsen, H., and Peter-Katalinic, J. (2005) Electron capture dissociation of O-glycosylated peptides: radical site-induced fragmentation of glycosidic bonds. *Eur. J. Mass Spectrom.* **11**, 497–511
57. Chamot-Rooke, J., van der Rest, G., Dalleu, A., Bay, S., and Lemoine, J. (2007) The combination of electron capture dissociation and fixed charge derivatization increases sequence coverage for O-glycosylated and O-phosphorylated peptides. *J. Am. Soc. Mass Spectrom.* **18**, 1405–1413
58. Halim, A., Nilsson, J., Ruetschi, U., Hesse, C., and Larson, G. (2012) Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. *Mol. Cell. Proteomics* **11**, M111013649
59. Watzlawick, H., Walsh, M. T., Yoshioka, Y., Schmid, K., and Brossmer, R. (1992) Structure of the N- and O-glycans of the A-chain of human plasma alpha 2HS-glycoprotein as deduced from the chemical compositions of the derivatives prepared by stepwise degradation with exoglycosidases. *Biochemistry* **31**, 12198–12203
60. Hagglund, P., Matthiesen, R., Elortza, F., Hojrup, P., Roepstorff, P., Jensen, O. N., and Bunkenborg, J. (2007) An enzymatic deglycosylation scheme enabling identification of core fucosylated N-glycans and O-glycosylation site mapping of human plasma proteins. *J. Proteome Res.* **6**, 3021–3031
61. Nilsson, J., Ruetschi, U., Halim, A., Hesse, C., Carlsohn, E., Brinkmalm, G., and Larson, G. (2009) Enrichment of glycopeptides for glycan structure

- and attachment site identification. *Nat. Meth.* **6**, 809–811
62. Lottspeich, F., Kellermann, J., Henschen, A., and Muller-Esterl, W. (1986) Amino acid sequence of the light chain of human high molecular mass kininogen. *Adv. Exp. Med. Biol.* **198**, 85–89
 63. Kristiansen, T. Z., Bunkenborg, J., Gronborg, M., Molina, H., Thuluvath, P. J., Argani, P., Goggins, M. G., Maitra, A., and Pandey, A. (2004) A proteomic analysis of human bile. *Mol. Cell. Proteomics* **3**, 715–728
 64. Deshpande, N., Jensen, P. H., Packer, N. H., and Kolarich, D. (2010) GlycoSpectrumScan: fishing glycopeptides from MS spectra of protease digests of human colostrum sIgA. *J. Proteome Res.* **9**, 1063–1075
 65. Chandler, K. B., Brnakova, Z., Sanda, M., Wang, S., Stalnaker, S. H., Bridger, R., Zhao, P., Wells, L., Edwards, N. J., and Goldman, R. (2014) Site-specific glycan microheterogeneity of inter-alpha-trypsin inhibitor heavy chain h4. *J. Proteome Res.* **13**, 3314–3329
 66. Olsen, E. H., Rahbek-Nielsen, H., Thogersen, I. B., Roepstorff, P., and Engild, J. J. (1998) Posttranslational modifications of human inter-alpha-inhibitor: identification of glycans and disulfide bridges in heavy chains 1 and 2. *Biochemistry* **37**, 408–416
 67. Flahaut, C., Capon, C., Balduyck, M., Ricart, G., Sautiere, P., and Mizon, J. (1998) Glycosylation pattern of human inter-alpha-inhibitor heavy chains. *Biochem. J.* **333**, 749–756
 68. Wuhler, M., Deelder, A. M., and van der Burgt, Y. E. (2011) Mass spectrometric glycan rearrangements. *Mass Spectrom. Rev.* **30**, 664–680
 69. Wuhler, M., Koeleman, C. A., and Deelder, A. M. (2009) Hexose rearrangements upon fragmentation of N-glycopeptides and reductively aminated N-glycans. *Anal. Chem.* **81**, 4422–4432
 70. Ikezu, S., and Ikezu, T. (2014) Tau-tubulin kinase. *Front Mol Neurosci* **7**, 33
 71. Gotz, J., Gladbach, A., Pennanen, L., van Eersel, J., Schild, A., David, D., and Ittner, L. M. (2010) Animal models reveal role for tau phosphorylation in human disease. *Biochim. Biophys. Acta* **1802**, 860–871
 72. Böhm, D., Keller, K., Pieter, J., Boehm, N., Wolters, D., Siggelkow, W., Lebrecht, A., Schmidt, M., Kölbl, H., Pfeiffer, N., and Grus, F.-H. (2012) Comparison of tear protein levels in breast cancer patients and healthy controls using a de novo proteomic approach. *Oncol. Rep.* **28**, 429–438
 73. Tagliabracci, V. S., Pinna, L. A., and Dixon, J. E. (2013) Secreted protein kinases. *Trends Biochem. Sci.* **38**, 121–130
 74. Goetz, Sarah C., Liem Jr., Karel F., and Anderson, Kathryn V. (2012) The Spinocerebellar Ataxia-Associated Gene Tau Tubulin Kinase 2 Controls the Initiation of Ciliogenesis. *Cell* **151**, 847–858
 75. Stenmark, H. (2009) Rab GTPases as coordinators of vesicle traffic. *Nat. Rev. Mol. Cell Biol.* **10**, 513–525
 76. Bai, X., Li, D., Zhu, J., Guan, Y., Zhang, Q., and Chi, L. (2015) From individual proteins to proteomic samples: characterization of O-glycosylation sites in human chorionic gonadotropin and human-plasma proteins. *Anal. Bioanal. Chem.* **407**, 1857–1869
 77. Darula, Z., Sherman, J., and Medzihradzky, K. F. (2012) How to dig deeper? Improved enrichment methods for mucin core-1 type glycopeptides. *Mol. Cell. Proteomics* **11**, O111 016774
 78. Alvarez-Manilla, G., Atwood, Guo, Y., Warren, N. L., Orlando, R., and Pierce, M. (2006) Tools for glycoproteomic analysis: Size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites. *J. Proteome Res.* **5**, 701–708
 79. Wopereis, S., Lefeber, D. J., Morava, É., and Wevers, R. A. (2006) Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: A review. *Clin. Chem.* **52**, 574–600
 80. Canis, K., McKinnon, T. A. J., Nowak, A., Panico, M., Morris, H. R., Laffan, M., and Dell, A. (2010) The plasma von Willebrand factor O-glycome comprises a surprising variety of structures including ABH antigens and disialosyl motifs. *J. Thromb. Haemost.* **8**, 137–145
 81. Takahashi, K., Wall, S. B., Suzuki, H., Smith, A. D. t., Hall, S., Poulsen, K., Kilian, M., Mobley, J. A., Julian, B. A., Mestecky, J., Novak, J., and Renfrow, M. B. (2010) Clustered O-glycans of IgA1: defining macro- and microheterogeneity by use of electron capture/transfer dissociation. *Mol. Cell. Proteomics* **9**, 2545–2557
 82. Darula, Z., and Medzihradzky, K. F. (2009) Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum. *Mol. Cell. Proteomics* **8**, 2515–2526
 83. Takayasu, T., Suzuki, S., Kametani, F., Takahashi, N., Shinoda, T., Okuyama, T., and Munekata, E. (1982) Amino acid sequence of galactosamine-containing glycopeptides in the hinge region of a human immunoglobulin D. *Biochem. Biophys. Res. Commun.* **105**, 1066–1071
 84. Plomp, R., Dekkers, G., Rombouts, Y., Visser, R., Koeleman, C. A. M., Kammeijer, G. S. M., Jansen, B. C., Rispen, T., Hensbergen, P. J., Vidarsson, G., and Wuhler, M. (2015) Hinge-Region O-Glycosylation of Human Immunoglobulin G3 (IgG3). *Mol. Cell. Proteomics* **14**, 1373–1384
 85. Zhu, Z., Su, X., Clark, D. F., Go, E. P., and Desaire, H. (2013) Characterizing O-Linked glycopeptides by electron transfer dissociation: Fragmentation rules and applications in data analysis. *Anal. Chem.* **85**, 8403–8411
 86. Good, D. M., Wirtala, M., McAlister, G. C., and Coon, J. J. (2007) Performance characteristics of electron transfer dissociation mass spectrometry. *Mol. Cell. Proteomics* **6**, 1942–1951
 87. Liu, H., Zhang, J., Sun, H., Xu, C., Zhu, Y., and Xie, H. (2011) The Prediction of Peptide Charge States for Electrospray Ionization in Mass Spectrometry. *Procedia Environ. Sci.* **8**, 483–491
 88. Nicolardi, S., van der Burgt, Y. E., Wuhler, M., and Deelder, A. M. (2013) Mapping O-glycosylation of apolipoprotein C-III in MALDI-FT-ICR protein profiles. *Proteomics* **13**, 992–1001
 89. Marti, T., Schaller, J., Rickli, E. E., Schmid, K., Kamerling, J. P., Gerwig, G. J., van Halbeek, H., and Vliegthart, J. F. (1988) The N- and O-linked carbohydrate chains of human, bovine and porcine plasminogen. Species specificity in relation to sialylation and fucosylation patterns. *Eur. J. Biochem.* **173**, 57–63
 90. Escribano, J., Lopex-Otin, C., Hjerpe, A., Grubb, A., and Mendez, E. (1990) Location and characterization of the three carbohydrate prosthetic groups of human protein HC. *FEBS Lett.* **266**, 167–170