Glycoconj J (2016) 33:937–951
DOI 10.1007/s10719-016-9677-z

ORIGINAL ARTICLE

Site-specific characterization of N-linked glycosylation in human urinary glycoproteins and endogenous glycopeptides

Rebeca Kawahara[1] · **Joyce Saad**[1] · **Claudia Blanes Angeli**[1] · **Giuseppe Palmisano**[1]

Received: 2 February 2016 / Revised: 9 May 2016 / Accepted: 10 May 2016 / Published online: 27 May 2016

Abstract Glycosylation is a very important post-translational modification involved in various cellular processes, such as cell adhesion, signal transduction and immune response. Urine is a rich source of glycoproteins and attractive biological fluid for biomarker discovery, owing to its availability, ease of collection, and correlation with pathophysiology of diseases. Although the urinary proteomics have been explored previously, the urinary glycoproteome characterization remains challenging requiring the development and optimization of analytical and bioinformatics methods for protein glycoprofiling. This study describes the high confident identification of 472 unique N-glycosylation sites covering 256 urinary glycoproteins. Besides, 202 unique N-glycosylation sites were identified in low molecular weight endogenous glycopeptides, which belong to 90 glycoproteins. Global site-specific characterization of the N-linked glycan heterogeneity was achieved by intact glycopeptide analysis, revealing 303 unique glycopeptides most of them displaying complex/hybrid glycans composed by sialic acid and fucose. These datasets consist in a valuable resource of glycoproteins and N-glycosylation sites found in healthy human urine that can be further explored in different disorders, in which the N-linked glycosylation may be aberrant.

Keywords N-glycosylation · Glycopeptides · Urine · Glycoforms

Introduction

Glycosylation is a very important post-translational modification, which occurs in approximately half of all mammalian proteins [1]. Glycoproteins is involved in a number of physiological processes, including protein folding and trafficking [2], immune response [3] and cell-cell and cell-matrix interaction [4, 5]. In cancer, there is increasing evidence pertaining to the role of glycosylation in tumour formation and metastasis [6–8]. It was demonstrated that glycoproteins have aberrant glycosylation patterns in malignancy, producing differential occupancy of glycosylation sites or variability in attached glycan structures [9–11].

Recent developments in MS-based technologies have enabled large-scale characterization of glycoproteins [12, 13]. Generally, glycoproteomic strategies rely on the analysis of released glycans or deglycosylated peptides [14–16]. Although the removal of glycans simplifies the identification of glycosylation sites by LC–MS/MS, limited information correlating glycan structural heterogeneity to the protein attachment site are obtained.

Recently, novel strategies were developed to enrich and analyze intact N and O-linked glycopeptides, for example, using metabolic labeling of the glycoproteome combined with chemical enrichment using an isotopic recoding affinity probe [17], or chemoenzymatic method for glycopeptide enrichment on solid phase [18]. Alternatively, several MS dissociation methods have been applied to study intact glycopeptides such as collision induced dissociation (CID), high-energy collision dissociation (HCD) and electron capture/transfer dissociation (ECD/ETD) [19–21]. Despite the recent advances, the

Electronic supplementary material The online version of this article (doi:10.1007/s10719-016-9677-z) contains supplementary material, which is available to authorized users.

✉ Giuseppe Palmisano
palmisano.gp@usp.br

[1] GlycoProteomics Laboratory, Departamento de Parasitologia, ICB, University of São Paulo, São Paulo, Brazil

detection of intact glycopeptides often requires increased sample quantities and an enrichment step, especially because of the glycan heterogeneity and the low ionization efficiency of glycoconjugates [22]. Besides, the computational identification via database searching are still challenging due to the heterogeneity and nontemplated nature of glycopeptides. Collectively, these challenges have hindered widespread use of intact glycopeptides analysis in the glycoproteome field.

Hydrophilic interaction chromatography (HILIC) has been used increasingly to separate and enrich glycopeptides and shows some advantages over other enrichment techniques such as broad specificity towards different glycan structures, good solubility of polar samples and greater compatibility with electrospray MS [12]. Hagglund *et al.* explored the use of HILIC followed by partial deglycosylation to identify N-glycosylation sites in complex biological samples [23] and Mysling improved the HILIC-glycopeptides enrichment efficiency by modifying the TFA-containing mobile phases [24]. This approach was also applied in the study of intact glycopeptides and glycoproteins in different biological matrices [21, 25–30].

Urine is a rich source for glycoproteins derived from renal- [31] and distal organs [32], and has been considered as an attractive biologic fluid for disease biomarkers discovery, since it can easily be collected continuously and noninvasively [33, 34]. Although urinary proteomics approaches have been applied in biomarker discovery studies of several disorders [32, 35–37], only a few studies aimed at identifying urinary glycoproteins [38, 39]. Moreover, information about site-specific glycan-peptide micro-heterogeneity is still not well explored.

Considering the urinary glycoproteomic studies that performed site-specific glycosylation analyses, Saraswat *et al.* [40] reported 51 N-glycosylation sites belonging to 37 glycoproteins in urinary exosomes and Halim *et al.* [41] showed 58 N- and 63 O-linked glycopeptides from 53 glycoproteins in the urine from one healthy person.

This study presents a simple and straightforward methodology without extensive sample preparation processes and pre-fractionation, to comprehensively characterize N-linked glycosylation sites of urinary glycoproteins as well as low molecular weight (<10 kDa) endogenous glycopeptides. Besides, by analyzing intact glycopeptides, high confident site-specific characterization of the N-linked glycans micro-heterogeneity was reported.

Material and methods

Urine sample collection and protein and peptide extraction

Twenty ml of urine (first morning void) was collected from a healthy male consenting individual in two different days and stored on ice prior to processing. The healthy volunteer was 35 years old, had no medical history, and did not take any prescription medicines. The samples were centrifuged at 3000 g for 10 min at 4 °C. The cleared supernatant was further filtered using 0.22 μm filter (Millipore, Billerica, MA). Twenty milliliters of filtered urine was concentrated using 10 kDa cutoff filters (Millipore, Billerica, MA). The concentrated urine, that is, the retentate, was stored immediately at −20 °C until further use. The fraction which passed through the 10 kDa cutoff filter was also collected and the peptides (<10 kDa) were extracted and concentrated using hydrophilic − lipophilic-balanced (HLB) solid phase (SPE; Waters).

Protein digestion and desalting

The proteins were reduced by addition of DTT to a final concentration of 10 mM and incubation for 30 min. The proteins were alkylated prior to digestion by the addition of IAA to a final concentration of 40 mM and incubation for 30 min in the dark at room temperature. To quench the reaction, DTT was added to a final concentration of 5 mM. Trypsin (1:50, *w*/w) was added, and the mixture incubated overnight at 37 °C. The reaction was stopped with 0.4 % TFA and desalted with hydrophilic − lipophilic-balanced (HLB) solid phase extraction (SPE; Waters).

Glycopeptide enrichment using HILIC

The samples were reconstituted in 100 μl of 80 % (*v*/v) acetonitrile, 10 % dimethyl sulfoxide (DMSO) and 1 % (*v*/v) trifluoroacetic acid. Peptides were loaded onto an in-house PolyHYDROXYETHYL A™ HILIC resin (PolyLC Inc) packed onto a C8 disk (Empore) in a p200 pipette tip. The flow-through and wash fraction (80 % (*v*/v) acetonitrile and 1 % (*v*/v) trifluoroacetic acid) were collected and analyzed by LC-MS/MS. The enriched glycopeptides were eluted with 50 μL 0.1 % (*v*/v) TFA followed by 50 μL of 25 mM NH_4HCO_3 and finally 50 μL of 50 % (*v*/v) acetonitrile. The three eluted fractions were combined and dried by vacuum centrifugation.

PNGase F deglycosylation

An aliquot of the HILIC-enriched glycopeptides (from protein trypsin digestion and endogenous glycopeptides) was resuspended in 50 mM ammonium bicarbonate, pH 8.0 and deglycosylated with 500 U of PNGase F (New England Biolabs, Ipswich, MA) for 12 h at 37 °C. After incubation, the peptides were dried by vacuum centrifugation and reconstituted in 50 μl of 0.1 % formic acid prior to LC-MS/MS analysis.

Mass spectrometry analysis

The resulting peptide mixture was analyzed on a LTQ Velos Orbitrap mass spectrometer (Thermo Fisher Scientific) coupled with LC-MS/MS by an EASY-nLC system (Thermo Fisher Scientific) through a nanoelectrospray ion source. Sample concentration and desalting were performed online using a pre-column (2 cm; 100 μm ID; 5 μm C18-A1; Thermo). Separation was accomplished on Acclaim™ PepMap™ 100 C18 column (10 cm; 75um ID; 3um C18-A2; Thermo) using a linear gradient of A and B buffers (buffer A: A = 0.1 % formic acid; Buffer B = 95 % ACN, 0.1 % formic acid) from 1 % to 50 % buffer B over 60 min at a flow rate of 0.3 μL/min to elute peptides into the mass spectrometer. Columns were washed and re-equilibrated between LC–MS/MS experiments. Mass spectra were acquired in the positive-ion mode over the range m/z 400–1200 at a resolution of 30,000 (full width at half-maximum at m/z 400) and AGC target $>1 \times e^6$. The 20 most intense peptide ions with charge states ≥2 were sequentially isolated to a target value of 15,000 and isolation width of 2 and fragmented in the linear ion trap using low-energy CID (normalized collision energy of 35 %) with activation time of 10 ms. For intact glycopeptides analysis, the 7 most intense peptide ions with charge states ≥2 were sequentially isolated to a target value of 15,000 and isolation width of 2 and fragmented in the linear ion trap using high-energy HCD (normalized collision energy of 35 %) with activation time of 0.1 ms. Dynamic exclusion was enabled with an exclusion size list of 500, exclusion duration of 15 s, and a repeat count of 1.

Analysis of intact N-glycopeptides

Intact glycopeptide spectra from each sample were searched against database using Byonic software v2.6.46 (Protein Metrics, http://www.proteinmetrics.com/) [42]. Searches were performed with the following fixed modifications: precursor mass tolerance of 10 ppm, product ion mass tolerance of 0.02 Da, carbamidomethylation of Cys, and fully trypsin specific cleavage with a maximum of two missed cleavages. Searches were also conducted with the following variable modifications: oxidation of methionine (15.994 Da), deamidation NQ (+0.984) and glycosylation of Asn with N-glycan database (309 mammalian no sodium) available within Byonic. The list of glycoproteins identified in the formerly N-linked peptides generated after PNGase F treatment fraction were used as protein database. All searches were filtered to <1 % false discovery rate (FDR) at protein level and 0 % at peptide level [43]. Moreover Byonic software calculates two-dimensional posterior error probability (PEP 2D), meaning that the protein identity (dimension one) and quality of peptide-spectrum match are taken into account for a correct sequence assignment. The probabilities are computed by the usual target/decoy method with the decoys being reversed protein sequences [44]. Extracted ion chromatograms using the HexNAc and NeuAc oxonium ions were performed using the Xcalibur software (Thermo). Manual inspection was performed in the MS/MS spectra of all intact glycopeptides identified.

Protein identification and glycosylation site analysis

For protein identification and glycosylation site analysis, raw files were imported into MaxQuant version 1.5.2.8 [45] and the database search engine Andromeda [46] was used to search MS/MS spectra against a database composed of the Uniprot Human Protein Database (release April 15, 2015; 45,185 entries) with a tolerance level of 4.5 ppm for MS and 20 ppm and 0.5 Da for MS/MS in HCD and CID mode, respectively. Enzyme specificity was set to trypsin with a maximum of two missed cleavages. For the database search of N-formerly endogenous glycopeptides no enzyme was selected and for the N-formerly endogenous glycopeptides treated with trypsin, semi-specific free N-terminus digestion with a maximum of two missed cleavages was considered. Carbamidomethylation of cysteine (57.021 Da) was set as a fixed modification, and oxidation of methionine (15.994 Da), deamidation NQ (+0.984 Da) and protein N-terminal acetylation (42.010 Da) were selected as variable modifications. All identifications were filtered in order to achieve a FDR of 1 %.

After excluding peptides identified as potential contaminants or in reverse database, we manually filtered the peptides containing the deamidation sites within the glycosylation motif NxS/T/C. Motif-X algorithm was used to identify sequence motifs from the endogenous N-glycopeptides, using the ipi.HUMAN.fasta database as background. Only motifs with $p < 10E6$ were considered.

In order to evaluate the chemical (spontaneous) deamidation rate, we also performed database search using MaxQuant in the HILIC-enriched glycopeptide fraction without treatment with PNGase F (intact glycopeptides fraction) [47].

To determine biological processes and tissue protein localization statistically over-represented in a given protein list InnateDB [48] and Enrichr [49] software were used.

MotifX (http://motif-x.med.harvard.edu/) and Sequence2Logo (http://www.cbs.dtu.dk/biotools/Seq2Logo/) were used to create the sequence motifs for endogenous glycopeptides. For MotifX analysis, peptide sequences were centered at the glycosylation site. For Sequence2Logo analysis, peptide sequences with six aminoacids were loaded.

Results and Discussion

Herein, we employed a simple and straightforward method to characterize the N-linked glycosylation sites of glycoproteins

and endogenous glycopeptides in human urine. For that, human urine was collected and concentrated using ultrafiltration. After protein digestion, glycopeptides were enriched using HILIC and N-glycans were removed by PNGase F treatment. The filtrate (peptides <10 kDa) was subjected to solid phase extraction, followed by HILIC enrichment and deglycosylation with PNGase F. Intact glycopeptides were also analyzed in order to site-specifically characterize the N-linked glycosylation heterogeneity. The glycoproteomic workflow is summarized in Fig. 1.

Overview of the urinary glycoproteome

Considering the formerly N-linked peptides generated after PNGase F treatment, 1092 redundant and 472 unique peptides were identified within the NxS/T/C motif (where X is not proline), covering 532 unique glycosylation sites from 256 glycoproteins (Fig. 2a and Supplemental Table S1 and S2). Considering the unmodified peptides or the peptides with deamidation modification in N or Q outside the NxS/T/C motif, we identified 422 redundant peptides. Therefore, using this strategy we observed a percentage of glycopeptides enrichment of about 72 % (1092/1514).Using the fraction of HILIC-enriched peptides without PNGase F treatment, we searched for peptides bearing a spontaneously deamidated asparagine within the N-linked glycosylation motif, however no peptides were found following this criteria.

The overlap of glycoproteins identified in two different days of urine collection was about 49 % (126 glycoproteins) (Fig. 2b). The non-modified peptides that did not bind to HILIC (considered here as the urine proteome) were also analyzed by LC-MS/MS, resulting in 391 proteins identified with 57 % overlap (223 proteins) between the two different days (Fig. 2c and Supplemental Table S3). In total, 562 proteins were identified (Fig. 2a); however, the glycoproteome fraction brought out 171 exclusive proteins (Fig. 2d). A higher proportion of proteins displaying transmembrane regions (36 %) and signal peptides (70 %) were observed in the glycoproteome-enriched fraction compared to the proteome fraction (24 % and 50 %, respectively) (Fig. 2e and Supplemental Table S2 and S3); thus secreted and cell surface proteins can be more efficiently identified using glycopeptide enrichment strategies. The top ten most significant biological processes overrepresented in the glycoproteome database were: cell adhesion, negative regulation of endopeptidase activity, extracellular matrix organization, angiogenesis, platelet degranulation, regulation of proteolysis, receptor-mediated endocytosis, platelet activation, blood coagulation and heterophilic cell-cell adhesion (Supplemental Table S4). To assess the localization and distribution of the glycoproteins in human tissues we performed enrichment analysis using Human Gene Atlas [50] and Human Proteome Map [51] data-set libraries available within Enrichr software. The glycoproteins identified in the urine of a male individual had a significant enrichment localization relative to the total number

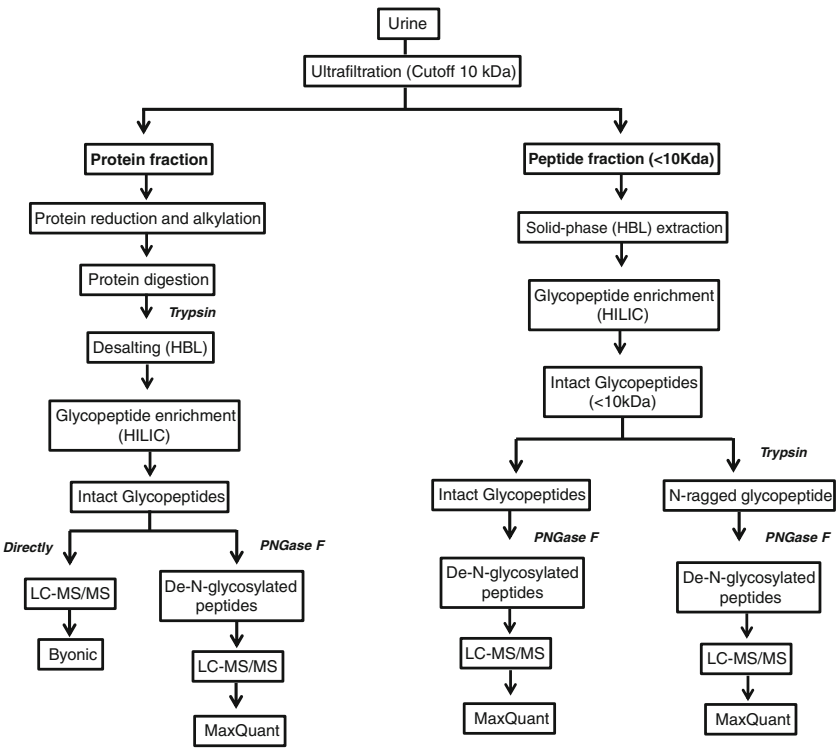

Fig. 1 Schematic workflow for preparation of urinary proteins and endogenous peptides for site-specific N-linked glycosylation analysis. The glycopeptides were enriched using polyLC-HILIC affinity chromatography. The glycans were released through treatment with PNGase F and the retained glycopeptide fractions as well as the non-modified peptides (flow through) were analyzed by LC-MS/MS employing CID fragmentation. The retained glycopeptides fraction was also directly analyzed by LC-MS/MS employing HCD fragmentation to site-specific characterization of the glycan microheterogeneity

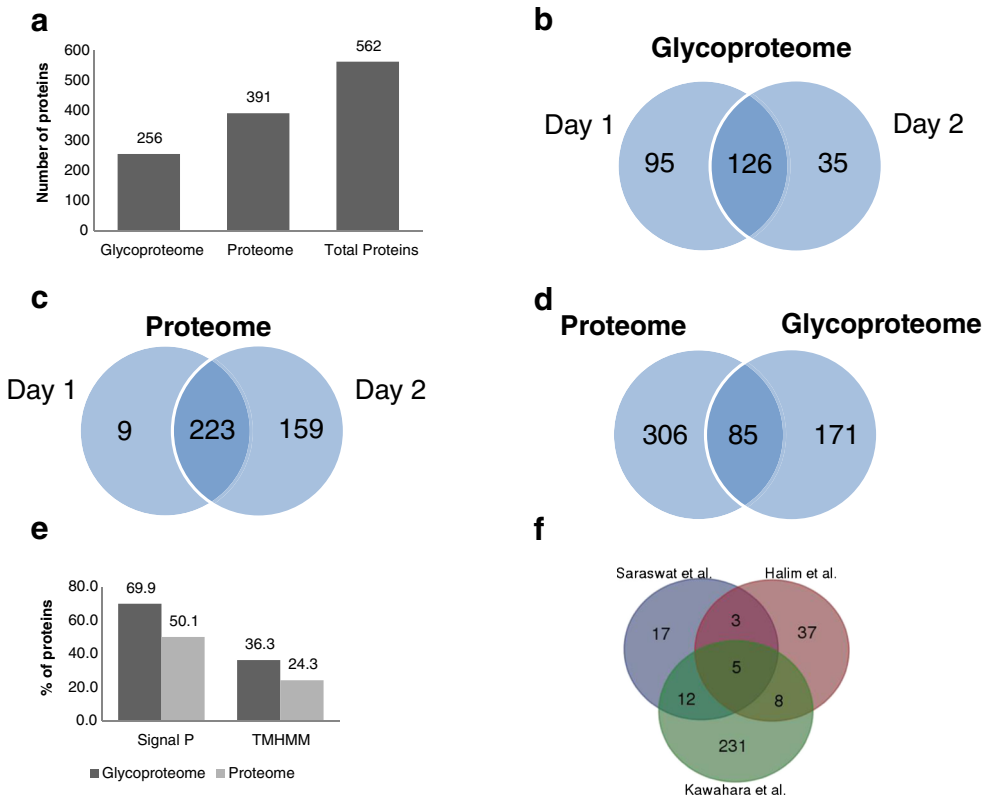

Fig. 2 Overview of the urinary proteome and glycoproteome. **a** Number of glycoproteins identified considering the formerly N-linked glycosylated peptides generated after PNGase F treatment (glycoproteome); number of proteins identified in the non-modified peptide fraction that did not bind to HILIC (proteome); number of total proteins identified the human urine considering the glycoproteome and proteome. The two biological replicates (urine collected at separate points over two different day) were considered to the total number of proteins identified in each fraction. **b** Distribution of N-glycoproteins (glycoproteome) identified in the urine collected at separate points over two different days. **c** Distribution of proteins (proteome) identified in the urine collected at separate points over two different days. **d** Overlap between the proteins identified in the glycoproteome and proteome fraction. **e** Percentage of proteins with signal peptides and transmembrane domains, considering the proteins identified in the glycoproteome or proteome fraction. **f** Venn diagram of the comparisons of the observed glycoproteome with those published by Saraswat *et al.* [40] and Halim *et al.* [41]

of genes (Human Gene Atlas) or proteins (Human Proteome map) identified in each tissue, for liver (p = 0.004, gene names: TF, BTD, AMBP, APOH, LCAT, HRG, EPHA1, KNG1, CPN2), adipocyte (p = 0.01, gene names: LRP1, FZD4, MXRA8, TIMP1, FSTL1), smooth muscle (p = 0.04, gene names: MRC2, HEG1, QSOX1, TIMP1, FSTL1, FBN1), kidney (p = 0.03, gene names: CUBN, EGF, DPEP1, FOLR1) (Humana Gene Atlas, Supplemental Fig. 1a) and adult prostate tissue (p = 0.0004, gene names: CDH2, KLK1, CADM2, AXL, QPCT, DPEP1, SERPINI1, CPZ, NEO1, PVRL1, KLK11) (Human Proteome Map, Supplemental Fig. 1b). These results suggest that the urinary glycoproteome may be a good biofluid to study disorders related not only to urological disease but also other tissues.

The list of glycoproteins identified in our study covered 17 out of 37 (46 %) glycoproteins identified in human urinary exosomes [40] and 16 out of 53 (25 %) glycoproteins identified in the previous healthy urinary glycoproteomic study [41]. However our study showed 231 exclusive glycoproteins that were not identified in the previous urinary glycoproteome studies (Fig. 2f). The modest overlap of glycoproteins identified in our study comparing with these two other studies were expected since biological as well methodological variation may produce different sets of enriched glycoproteins/glycopeptides. For example, the study by Halim *et al.*, used hydrazine capture to enrich sialylated glycoproteins in human urine and the intact glycopeptides were analysis by CID and ECD fragmentation method. In the study by Saraswat *et al.*, they focused on the urinary exosomes and the glycopeptides were enriched using SNA affinity and analyzed by CID-tandem MS. However, despite all these variation we could still observe common glycoproteins with these studies and demonstrate that our approach brings exclusive identifications.

Intact glycopeptides analysis

Site-specific glycan micro-heterogeneity of the urinary glycoproteins was investigated by analyzing an aliquot of the HILIC-enriched glycopeptides without prior PNGase F treatment, using high resolution LC-MS/MS and higher energy

collision dissociation (HCD) (Fig. 1). Using the intense diagnostic saccharide oxonium ions such as *m/z* 138.06/204.09 for *N*-acetylhexosamine (HexNAc) and/or *m/z* 274.09/292.10 for *N*-acetylneuraminic acid (NeuAc), we compared the glycosylation profile of intact glycopeptides expressed in the human urine collected in two different days (Fig. 3a).

The extracted ion chromatogram for the HexNAc and NeuAc oxonium ions showed similar profile between the two different days (Fig. 3a). Interestingly, the diagnostic ion at *m/z* 274.091/292.10 pointed out to large amount of glycopeptides containing NeuAc in human urine.

To assess site-specific glycan composition, intact glycopeptides were identified using Byonic software. 1437 and 1038 spectra were assigned to a glycopeptide, in the first and second day, respectively (data not shown). Considering the non-glycosylated peptides, 172 and 91 were identified in the first and second day, respectively, corresponding to a percentage of glycopeptide enrichment in almost 90 %. However, for high confident identification, we filtered only the glycopeptides identified with 0 % FDR, resulting in 156 and 402 identifications in the first and second day, respectively (Supplemental Table S5). From the 156 redundant intact glycopeptides identified with 0 % FDR in the first day, all of them had PEP 2D <0.001 (less than 0.1 % chance that the match is wrong). In the second day, from 402 redundant intact glycopeptides identified with 0 % FDR, 257 had PEP 2D <0.001 and 142 had PEP 2D <0.01 (less than 1 % chance of a wrong match) and only 3 had PEP 2D <0.1 (less than 10 % chance of a wrong match). The glycopeptides identified with the lowest PEP score had 14 fragments assigned (not shown). These parameters were chosen for stringent and confident assignments [17, 52]. However, more efforts are needed to establish a method for intact glycopeptide FDR calculation.

In total, 303 unique intact glycopeptides were identified, with an overlap of 20 % between the first and second day (Fig. 3b). These glycopeptides covered a total of 60 glycoproteins, of which 30 (50 %) were identified in both days (Fig. 3c). Unfortunately, the FDR and scores only assesses the correctness of the peptide sequence [44], which means that the glycopeptide identification can be partially correct, *e.g.* correct peptide sequence and glycosylation site but still remain incorrect at the identified monosaccharide composition. For example, we observed that from the 303 unique intact glycopeptides, 61 showed the glycan NeuGc, which is uncommon in human. We manually inspect all MS/MS spectra to search for its oxonium ions at m/z 308/290 however we did

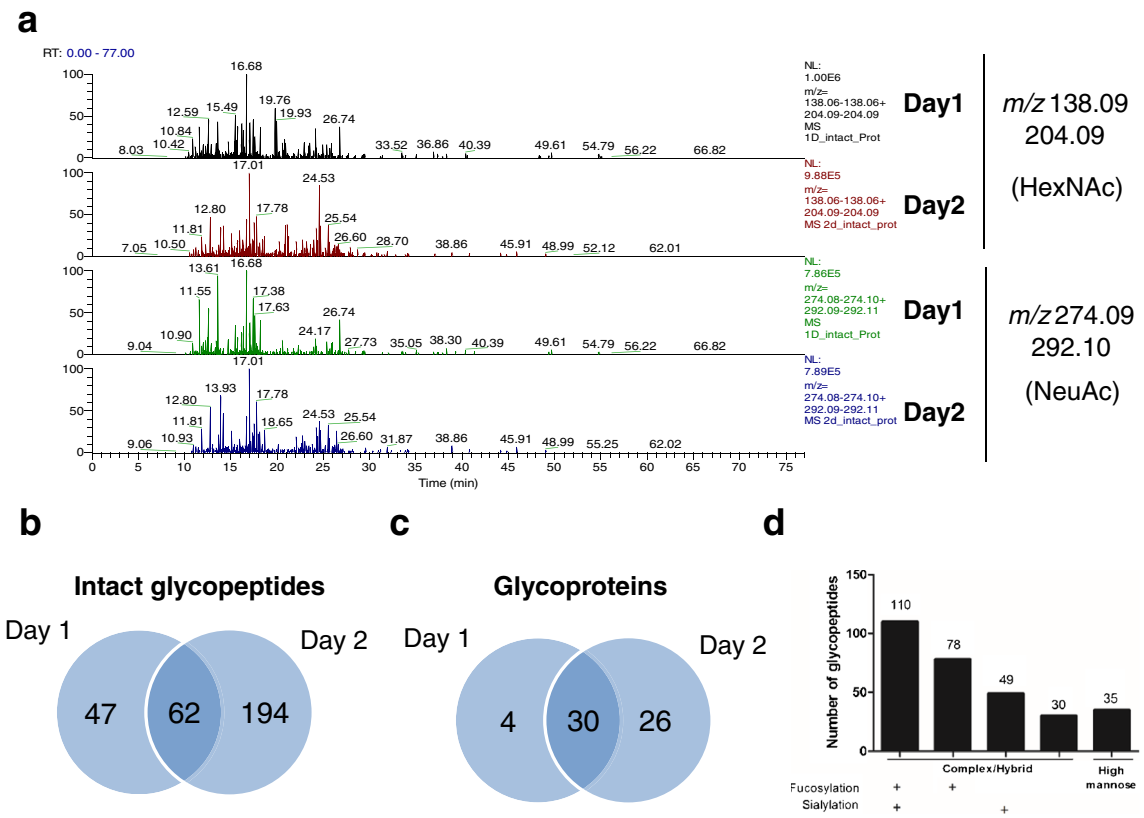

Fig. 3 Investigation of the site-specific N-linked glycans of the human urinary glycoproteins. **a** Extracted ion chromatogram of HexNAc (*m/z* 138.06/204.09) and NeuAc (*m/z* 274.09 and 292.10) oxonium ions with a mass tolerance of 20 ppm. **b** Distribution of unique N-linked intact glycopeptides in the urine collected in two different days. **c** Distribution of glycoproteins covered by the identified intact glycopeptides in the urine collected in two different days. **d** Distribution of the number of intact glycopeptides identified according to the class (high mannose or complex/hybrid) and features

Table 1 N-linked intact glycopeptides identified in human urinary glycoproteins with the possible glycan structure selected from UnicarbKB database

Modified Sequence	Sequence	Glycans composition	Deamidation (NQ) Probabilities	Protein name
R.YEQLQN[+1216.42286]ETR.Q	YEQLQNETR	HexNAc(2)Hex(5)	YEQLQ(0.181)N(0.819)ETR; YEQ(0.248)LQ(0.798)N(0.954)ETR	B7Z7D2 Lysosomal acid phosphatase
R.AQHELN[+1216.42286]C[+57.02146]SLQDPR.S	AQHELNCSLQDPR		AQ(0.411)HELN(0.589)CSLQ(0.001)DPR	E9PI73 Transmembrane protein 25 (Fragment)
R.VFLN[+1216.42286]GSRQER.V	VFLNGSRQER		VFLN(0.5)GSRQ(0.5)ER	EGF Pro-epidermal growth factor
R.GPGIKPN[+1216.42286]QTSK.M	GPGIKPNQTSK		GPGIKPN(0.999)Q(0.001)TSK	H7C3P4 N-acetylglucosamine-6-sulfatase
K.SELVN[+1216.42286]ETR.W	SELVNETR		SELVN(1)ETR	SUSD2 Sushi domain-containing protein 2
R.VFLN[+1378.47569]GSRQER.V	VFLNGSRQER	HexNAc(2)Hex(6)	VFLN(0.5)GSRQ(0.5)ER	EGF Pro-epidermal growth factor
R.GPGIKPN[+1378.47569]QTSK.M	GPGIKPNQTSK		GPGIKPN(0.999)Q(0.001)TSK	H7C3P4 N-acetylglucosamine-6-sulfatase
R.EDQYHYLLDRN[+1378.47569]LSC[+57.02146]R.V	EDQYHYLLDRNLSCR		EDQYHYLLDRN(1)LSCR;EDQ(1)YHYLLDRN(1)LSCR	IPSP Plasma serine protease inhibitor
R.RGVFITN[+1378.47569]ETGQPLIGK.V	RGVFITNETGQPLIGK		RGVFITN(0.977)ETGQ(0.023)PLIGK;RGVFITN(1)ETGQ(1)PLIGK	LYAG Lysosomal alpha-glucosidase
R.VFLN[+1540.52851]GSRQER.V	VFLNGSRQER	HexNAc(2)Hex(7)	VFLN(0.5)GSRQ(0.5)ER	EGF Pro-epidermal growth factor
R.GPGIKPN[+1540.52851]QTSK.M	GPGIKPNQTSK		GPGIKPN(0.999)Q(0.001)TSK	H7C3P4 N-acetylglucosamine-6-sulfatase
K.GSLSYLN[+1540.52851]VTRK.A	GSLSYLNVTRK		GSLSYLN(1)VTRK	H7C469 Uncharacterized protein (Fragment)
R.RGVFITN[+1540.52851]ETGQPLIGK.V	RGVFITNETGQPLIGK		RGVFITN(0.977)ETGQ(0.023)PLIGK;RGVFITN(1)ETGQ(1)PLIGK	LYAG Lysosomal alpha-glucosidase
R.KN[+1702.58133]LTLMATTSQLPK.L	KNLTLMATTSQLPK	HexNAc(2)Hex(8)	KN(1)LTLMATTSQLPK	B7Z7D2 Lysosomal acid phosphatase
R.VFLN[+1702.58133]GSRQER.V	VFLNGSRQER		VFLN(0.5)GSRQ(0.5)ER	EGF Pro-epidermal growth factor
R.KN[+1864.63416]LTLMATTSQLPK.L	KNLTLMATTSQLPK	HexNAc(2)Hex(9)	KN(1)LTLMATTSQLPK	B7Z7D2 Lysosomal acid phosphatase
R.AGFN[+1403.50732]ASFHK.S	AGFNASFHK	HexNAc(3)Hex(4)Fuc(1)	AGFN(1)ASFHK	CUBN Cubilin
K.TKPREEQYN[+1403.50732]STYR.V	TKPREEQYNSTYR		TKPREEQ(0.012)YN(0.988)STYR;TKPREEQ(1)YN(1)STYR	IGHG1 Ig gamma-1 chain C region
K.YKN[+1694.60274]NSDISSTR.G	YKNNSDISSTR	HexNAc(3)Hex(4)Fuc(1)NeuAc(1)	YKN(0.987)N(0.013)SDISSTR	A0A075B6N9 Ig mu chain C region (Fragment)
K.YKN[+1710.59765]NSDISSTR.G	YKNNSDISSTR	HexNAc(3)Hex(4)Fuc(1)NeuAc*(1)	YKN(0.987)N(0.013)SDISSTR	A0A075B6N9 Ig mu chain C region (Fragment)
R.LTYRN[+1710.59765]LSGPDKR.L	LTYRNLSGPDKR		LTYRN(1)LSGPDKR	VASN Vasorin
K.YKN[+1564.53974]NSDISSTR.G	YKNNSDISSTR	HexNAc(3)Hex(4)NeuAc*(1)	YKN(0.987)N(0.013)SDISSTR	A0A075B6N9 Ig mu chain C region (Fragment)
K.YKN[+1565.56014]NSDISSTR.G	YKNNSDISSTR	HexNAc(3)Hex(5)Fuc(1)	YKN(0.987)N(0.013)SDISSTR	A0A075B6N9 Ig mu chain C region (Fragment)
R.LGAELRN[+1710.59765]ASLR.M	LGAELRNASLR	HexNAc(3)Hex(5)NeuAc(1)	LGAELRN(1)ASLR	A0A0A0MSA9 Poliovirus receptor
R.LTYRN[+1581.55506]LSGPDKR.L	LTYRNLSGPDKR	HexNAc(3)Hex(6)	LTYRN(1)LSGPDKR	VASN Vasorin
K.YKN[+1872.65048]NSDISSTR.G	YKNNSDISSTR	HexNAc(3)Hex(6)NeuAc(1)	YKN(0.987)N(0.013)SDISSTR	A0A075B6N9 Ig mu chain C region (Fragment)
R.LTYRN[+1872.65048]LSGPDKR.L	LTYRNLSGPDKR		LTYRN(1)LSGPDKR	VASN Vasorin
R.DYNTKN[+1444.53387]GTIK.W	DYNTKNGTIK	HexNAc(4)Hex(3)Fuc(1)	DYN(1)TKN(1)GTIK	DSG1 Desmoglein-1
K.TKPREEQYN[+1444.53387]STYR.V	TKPREEQYNSTYR		TKPREEQ(0.012)YN(0.988)STYR;TKPREEQ(1)YN(1)STYR	IGHG1 Ig gamma-1 chain C region
K.TKPREEQFN[+1444.53387]STFR.V	TKPREEQFNSTFR		TKPREEQ(0.21)FN(0.79)STFR	IGHG2 Ig gamma-2 chain C region
K.TKPREEQFN[+1444.53387]STYR.V	TKPREEQFNSTYR		TKPREEQ(0.017)FN(0.983)STYR	IGHG4 Ig gamma-4 chain C region
R.EDQYHYLLDRN[+1444.53387]LSC[+57.02146]R.V	EDQYHYLLDRNLSCR		EDQYHYLLDRN(1)LSCR;EDQ(1)YHYLLDRN(1)LSCR	IPSP Plasma serine protease inhibitor
R.SSC[+57.02146]GKEN[+1444.53387]TSDPSLVIAFGR.G	SSCGKENTSDPSLVIAFGR		SSCGKEN(1)TSDPSLVIAFGR	LAMP1 Lysosome-associated membrane glycoprotein 1
K.TKPREEQYN[+1460.52878]STYR.V	TKPREEQYNSTYR	HexNAc(4)Hex(4)	TKPREEQ(0.012)YN(0.988)STYR;TKPREEQ(1)YN(1)STYR	IGHG1 Ig gamma-1 chain C region
R.DYNTKN[+1606.58669]GTIK.W	DYNTKNGTIK	HexNAc(4)Hex(4)Fuc(1)	DYN(1)TKN(1)GTIK	DSG1 Desmoglein-1
K.TKPREEQYN[+1606.58669]STYR.V	TKPREEQYNSTYR		TKPREEQ(0.012)YN(0.988)STYR;TKPREEQ(1)YN(1)STYR	IGHG1 Ig gamma-1 chain C region
K.TKPREEQFN[+1606.58669]STFR.V	TKPREEQFNSTFR		TKPREEQ(0.21)FN(0.79)STFR	IGHG2 Ig gamma-2 chain C region
K.TKPREEQFN[+1606.58669]STYR.V	TKPREEQFNSTYR		TKPREEQ(0.017)FN(0.983)STYR	IGHG4 Ig gamma-4 chain C region

Table 1 (continued)

Modified Sequence	Sequence	Glycans composition	Deamidation (NQ) Probabilities	Protein name
R.EDQYHYLLDRN[+1606.58669]LSC[+57.02146]R.V	EDQYHYLLDRNLSCR		EDQYHYLLDRN(1)LSCR;EDQ(1)YHYLLDRN(1)LSCR	IPSP Plasma serine protease inhibitor
K.TKPREEQYN[+1897.68211]STYR.V	TKPREEQYNSTYR	HexNAc(4)Hex(4)Fuc(1)NeuAc(1)	TKPREEQ(0.012)YN(0.988)STYR;TKPREEQ(1)YN(1)STYR	IGHG1 Ig gamma-1 chain C region
R.EDQYHYLLDRN[+1897.68211]LSC[+57.02146]R.V	EDQYHYLLDRNLSCR		EDQYHYLLDRN(1)LSCR;EDQ(1)YHYLLDRN(1)LSCR	IPSP Plasma serine protease inhibitor
R.EDQYHYLLDRN[+1751.62420]LSC[+57.02146]R.V	EDQYHYLLDRNLSCR	HexNAc(4)Hex(4)NeuAc(1)	EDQYHYLLDRN(1)LSCR;EDQ(1)YHYLLDRN(1)LSCR	IPSP Plasma serine protease inhibitor
R.SRYPHKPEIN[+1622.58161]STTHPGADLQENFC[+57.02146]R.N	SRYPHKPEINSTTHPGADLQENFCR	HexNAc(4)Hex(5)	SRYPHKPEIN(0.984)STTHPGADLQ(0.016)EN(0.001)FCR;SRYPHKPEIN(1)STTHPGADLQ(0.022)EN(0.978)FCR;SRYPHKPEIN(0.999)STTHPGADLQ(0.623)EN(0.378)FCR;SRYPHKPEIN(0.188)STTHPGADLQ(0.899)EN(0.914)FCR	C9JV37 Prothrombin (Fragment)
R.AGFN[+1622.58161]ASFHK.S	AGFNASFHK		AGFN(1)ASFHK	CUBN Cubilin
R.LRN[+1622.58161]VSWATGR.S	LRNVSWATGR		LRN(1)VSWATGR	D6RFL4 Monocyte differentiation antigen CD14 (Fragment)
R.IIVPLNNREN[+1622.58161]ISDPTSPLR.T	IIVPLNNRENISDPTSPLR		IIVPLN(0.216)N(0.858)REN(0.926)ISDPTSPLR	D6RHJ6 Immunoglobulin J chain (Fragment)
R.DYNTKN[+1622.58161]GTIK.W	DYNTKNGTIK		DYN(1)TKN(1)GTIK	DSG1 Desmoglein-1
R.TQPSSEN[+1622.58161]KTAHLHK.G	TQPSSENKTAHLHK		TQPSSEN(1)KTAHLHK	H0YFL7 Complement C1r subcomponent-like protein (Fragment)
K.TKPREEQYN[+1622.58161]STYR.V	TKPREEQYNSTYR		TKPREEQ(0.012)YN(0.988)STYR;TKPREEQ(1)YN(1)STYR	IGHG1 Ig gamma-1 chain C region
R.EDQYHYLLDRN[+1622.58161]LSC[+57.02146]R.V	EDQYHYLLDRNLSCR		EDQYHYLLDRN(1)LSCR;EDQ(1)YHYLLDRN(1)LSCR	IPSP Plasma serine protease inhibitor
K.FLN[+1622.58161]ESYKHEQVYIR.S	FLNESYKHEQVYIR		FLN(0.999)ESYKHEQ(0.001)VYIR;FLN(1)ESYKHEQ(1)VYIR	PPAP Prostatic acid phosphatase
R.LHEITN[+1622.58161]ETFR.G	LHEITNETFR		LHEITN(1)ETFR	VASN Vasorin
K.YKN[+1768.63952]NSDISSTR.G	YKNNSDISSTR	HexNAc(4)Hex(5)Fuc(1)	YKN(0.987)N(0.013)SDISSTR	A0A075B6N9 Ig mu chain C region (Fragment)
R.AGFN[+1768.63952]ASFHK.S	AGFNASFHK		AGFN(1)ASFHK	CUBN Cubilin
K.TAIFAFN[+1768.63952]ISHVSNK.V	TAIFAFNISHVSNK		TAIFAFN(1)ISHVSNK;TAIFAFN(1)ISHVSN(1)K	FBN1 Fibrillin-1
K.TKPREEQYN[+1768.63952]STYR.V	TKPREEQYNSTYR		TKPREEQ(0.012)YN(0.988)STYR;TKPREEQ(1)YN(1)STYR	IGHG1 Ig gamma-1 chain C region
K.TKPREEQFN[+1768.63952]STFR.V	TKPREEQFNSTFR		TKPREEQ(0.21)FN(0.79)STFR	IGHG2 Ig gamma-2 chain C region
K.TKPREEQFN[+1768.63952]STYR.V	TKPREEQFNSTYR		TKPREEQ(0.017)FN(0.983)STYR	IGHG4 Ig gamma-4 chain C region
K.FLN[+1768.63952]ESYKHEQVYIR.S	FLNESYKHEQVYIR		FLN(0.999)ESYKHEQ(0.001)VYIR;FLN(1)ESYKHEQ(1)VYIR	PPAP Prostatic acid phosphatase
K.NYHQPAILN[+1768.63952]SSALR.Q	NYHQPAILNSSALR		NYHQPAILN(1)SSALR;N(0.896)YHQ(0.104)PAILN(1)SSALR	QPCT Glutaminyl-peptide cyclotransferase
R.KTN[+1768.63952]ESVSEPR.K	KTNESVSEPR		KTN(1)ESVSEPR	S39 A6 Zinc transporter ZIP6
R.N[+1768.63952]GSQAFVHWQEPR.A	NGSQAFVHWQEPR		N(0.821)GSQ(0.179)AFVHWQEPR	UFO Tyrosine-protein kinase receptor UFO
R.SLHVPGLN[+1768.63952]KTSSFSC[+57.02146]EAHNAK.G	SLHVPGLNKTSSFSCEAHNAK		SLHVPGLN(0.996)KTSSFSCEAHN(0.004)AK;SLHVPGLN(1)KTSSFSCEAHN(1)AK	UFO Tyrosine-protein kinase receptor UFO
R.LHEITN[+1768.63952]ETFR.G	LHEITNETFR		LHEITN(1)ETFR	VASN Vasorin
K.YKN[+2059.73493]N[+0.98402]SDISSTR.G	YKNNSDISSTR	HexNAc(4)Hex(5)Fuc(1)NeuAc(1)	YKN(0.987)N(0.013)SDISSTR	A0A075B6N9 Ig mu chain C region (Fragment)
K.AREDIFMETLKDIVEYYN[+2059.73493]DSNGSHVLQGR.F	AREDIFMETLKDIVEYYNDSNGSHVLQGR		AREDIFMETLKDIVEYYN(0.033)DSN(0.952)GSHVLQ(0.015)GR;AREDIFMETLKDIVEYYN(0.857)DSN(0.973)GSHVLQ(0.17)GR	C9JEV0 Zinc-alpha-2-glycoprotein
K.ALVLEQLTPALHSTN[+2059.73493]FSC[+57.02146]VLVDPEQVVQR.H	ALVLEQLTPALHSTNFSCVLVDPEQVVQR		ALVLEQ(0.124)LTPALHSTN(0.875)FSCVLVDPEQVVQR;ALVLEQ(0.021)LTPALHSTN(0.979)FSCVLVDPEQ(0.8)VVQ(0.2)R;ALVLEQ(1)LTPALHSTN(0.998)FSCVLVDPEQ(0.002)VVQR	I18BP Interleukin-18-binding protein
R.IQLEN[+2366.82526]VTLLNPDPAEGPKPR.P	IQLENVTLLNPDPAEGPKPR	HexNAc(4)Hex(5)Fuc(1)NeuAc*(2)	IQLEN(0.991)VTLLN(0.009)PDPAEGPKPR;IQ(0.997)LEN(0.996)VTLLN(0.007)PDPAEGPKPR	B4DYV8 Roundabout homolog 4
R.TQPSSEN[+2350.83035]KTAHLHK.G	TQPSSENKTAHLHK		TQPSSEN(1)KTAHLHK	

Table 1 (continued)

Modified Sequence	Sequence	Glycans composition	Deamidation (NQ) Probabilities	Protein name
		HexNAc(4)Hex(5)Fuc (1)NeuAc(2)		H0YFL7 Complement C1r subcomponent-like protein (Fragment)
K.ALVLEQLTPALHSTN[+2350.83035] FSC[+57.02146]VLVDPEQVVQR.H	ALVLEQLTPALHSTNFSCVLVDPEQVVQR		ALVLEQ(0.124)LTPALHSTN(0.875)FSCVLVDPEQVVQR; ALVLEQ(0.021)LTPALHSTN(0.979)FSCVLVDPEQ(0.8) VVQ(0.2)R;ALVLEQ(1)LTPALHSTN(0.998)FSCVLVDPEQ (0.002)VVQR	I18BP Interleukin-18-binding protein
R.EEQFN[+2350.83035]STFR.V	EEQFNSTFR		EEQ(0.001)FN(0.999)STFR	IGHG2 Ig gamma-2 chain C region
R.N[+2350.83035]KSVILLGR.H	NKSVILLGR		N(1)KSVILLGR	M0QZF9 Prostate-specific antigen (Fragment)
R.KFLN[+2350.83035]ESYKHEQVYIR.S	KFLNESYKHEQVYIR		KFLN(1)ESYKHEQVYIR	PPAP Prostatic acid phosphatase
K.YLGN[+1914.69743] ATAIFFLPDEGKLQHLENELTHDIITK.F	YLGNATAIFFLPDEGKLQHLENELTHDIITK	HexNAc(4)Hex(5)Fuc(2)	YLGN(0.998)ATAIFFLPDEGKLQ(0.013)HLEN(0.99)ELTHDIITK; YLGN(0.999)ATAIFFLPDEGKLQHLEN(0.001)ELTHDIITK	A0A0G2JRN3 Alpha-1-antitrypsin
R.SRYPHKPEIN[+1914.69743] STTHPGADLQENFC[+57.02146]R.N	SRYPHKPEINSTTHPGADLQENFCR		SRYPHKPEIN(0.984)STTHPGADLQ(0.016)EN(0.001)FCR; SRYPHKPEIN(1)STTHPGADLQ(0.022)EN(0.978)FCR; SRYPHKPEIN(0.999)STTHPGADLQ(0.623)EN(0.378)FCR; SRYPHKPEIN(0.188)STTHPGADLQ(0.899)EN(0.914)FCR	C9JV37 Prothrombin (Fragment)
K.VC[+57.02146]QDC[+57.02146] PLLAPLN[+1914.69743]DTRVVHAAK.A	VCQDCPLLAPLNDTRVVHAAK		VCQDCPLLAPLN(1)DTRVVHAAK	FETUA Alpha-2-HS-glycoprotein
K.YLGN[+2205.79284] ATAIFFLPDEGKLQHLENELTHDIITK.F	YLGNATAIFFLPDEGKLQHLENELTHDIITK	HexNAc(4)Hex(5)Fuc (2)NeuAc(1)	YLGN(0.998)ATAIFFLPDEGKLQ(0.013)HLEN(0.99)ELTHDIITK; YLGN(0.999)ATAIFFLPDEGKLQHLEN(0.001)ELTHDIITK	A0A0G2JRN3 Alpha-1-antitrypsin
R.IQLEN[+ 2205.79284]VTLLNPDPAEGPKPR.P	IQLENVTLLNPDPAEGPKPR		IQLEN(0.991)VTLLN(0.009)PDPAEGPKPR;IQ(0.997)LEN(0.996) VTLLN(0.007)PDPAEGPKPR	B4DYV8 Roundabout homolog 4
K.AREDIFMETLKDIVEYYNDSN [+2205.79284]GSHVLQGR.F	AREDIFMETLKDIVEYYNDSNGSHVLQGR		AREDIFMETLKDIVEYYN(0.033)DSN(0.952)GSHVLQ(0.015)GR; AREDIFMETLKDIVEYYN(0.857)DSN(0.973)GSHVLQ(0.17)GR	C9JEV0 Zinc-alpha-2-glycoprotein
R.SRYPHKPEIN[+2205.79284] STTHPGADLQENFC[+57.02146]R.N	SRYPHKPEINSTTHPGADLQENFCR		SRYPHKPEIN(0.984)STTHPGADLQ(0.016)EN(0.001)FCR; SRYPHKPEIN(1)STTHPGADLQ(0.022)EN(0.978)FCR; SRYPHKPEIN(0.999)STTHPGADLQ(0.623)EN(0.378)FCR; SRYPHKPEIN(0.188)STTHPGADLQ(0.899)EN(0.914)FCR	C9JV37 Prothrombin (Fragment)
R.YPHKPEIN[+2205.79284] STTHPGADLQENFC[+57.02146]R.N	YPHKPEINSTTHPGADLQENFCR		YPHKPEIN(0.92)STTHPGADLQ(0.058)EN(0.022)FCR	C9JV37 Prothrombin (Fragment)
R.IIVPLNNREN[+2205.79284] ISDPTSPLR.T	IIVPLNNRENISDPTSPLR		IIVPLN(0.216)N(0.858)REN(0.926)ISDPTSPLR	D6RHJ6 Immunoglobulin J chain (Fragment)
K.VC[+57.02146]QDC[+57.02146] PLLAPLN[+2205.79284]DTRVVHAAK.A	VCQDCPLLAPLNDTRVVHAAK		VCQDCPLLAPLN(1)DTRVVHAAK	FETUA Alpha-2-HS-glycoprotein
K.APDKNVIFSPLSISTALAFLSLGAHN [+2205.79284]TTLTEILK.G	APDKNVIFSPLSISTALAFLSLGAHNTTLTEILK		APDKNVIFSPLSISTALAFLSLGAHN(1)TTLTEILK	G3 V595 Alpha-1-antichymotrypsin (Fragment)
R.KN[+ 2205.79284]QSVNVFLGHTAIDEMLK.L	KNQSVNVFLGHTAIDEMLK		KN(0.621)Q(0.149)SVN(0.23)VFLGHTAIDEMLK	H0YFL7 Complement C1r subcomponent-like protein (Fragment)
K.ALVLEQLTPALHSTN[+2205.79284] FSC[+57.02146]VLVDPEQVVQR.H	ALVLEQLTPALHSTNFSCVLVDPEQVVQR		ALVLEQ(0.124)LTPALHSTN(0.875)FSCVLVDPEQVVQR; ALVLEQ(0.021)LTPALHSTN(0.979)FSCVLVDPEQ(0.8) VVQ(0.2)R;ALVLEQ(1)LTPALHSTN(0.998) FSCVLVDPEQ(0.002)VVQR	I18BP Interleukin-18-binding protein
R.KFLN[+2205.79284]ESYKHEQVYIR.S	KFLNESYKHEQVYIR		KFLN(1)ESYKHEQVYIR	PPAP Prostatic acid phosphatase
R.IQLEN[+ 2351.85075]VTLLNPDPAEGPKPR.P	IQLENVTLLNPDPAEGPKPR	HexNAc(4)Hex(5)Fuc (3)NeuAc(1)	IQLEN(0.991)VTLLN(0.009)PDPAEGPKPR;IQ(0.997)LEN(0.996) VTLLN(0.007)PDPAEGPKPR	B4DYV8 Roundabout homolog 4
R.IIVPLNNREN[+2351.85075]ISDPTSPLR.T	IIVPLNNRENISDPTSPLR		IIVPLN(0.216)N(0.858)REN(0.926)ISDPTSPLR	D6RHJ6 Immunoglobulin J chain (Fragment)
K.ALVLEQLTPALHSTN[+2351.85075] FSC[+57.02146]VLVDPEQVVQR.H	ALVLEQLTPALHSTNFSCVLVDPEQVVQR		ALVLEQ(0.124)LTPALHSTN(0.875)FSCVLVDPEQVVQR;ALVLEQ (0.021)LTPALHSTN(0.979)FSCVLVDPEQ(0.8)VVQ(0.2)R; ALVLEQ(1)LTPALHSTN(0.998)FSCVLVDPEQ(0.002)VVQR	I18BP Interleukin-18-binding protein
K.YLGN[+1913.67702] ATAIFFLPDEGKLQHLENELTHDIITK.F	YLGNATAIFFLPDEGKLQHLENELTHDIITK	HexNAc(4)Hex(5)NeuAc(1)	YLGN(0.998)ATAIFFLPDEGKLQ(0.013)HLEN(0.99)ELTHDIITK; YLGN(0.999)ATAIFFLPDEGKLQHLEN(0.001)ELTHDIITK	A0A0G2JRN3 Alpha-1-antitrypsin

Table 1 (continued)

Modified Sequence	Sequence	Glycans composition	Deamidation (NQ) Probabilities	Protein name
R.SRYPHKPEIN[+1913.67702]STTHPGADLQENFC[+57.02146]R.N	SRYPHKPEINSTTHPGADLQENFCR		SRYPHKPEIN(0.984)STTHPGADLQ(0.016)EN(0.001)FCR; SRYPHKPEIN(1)STTHPGADLQ(0.022)EN(0.978)FCR; SRYPHKPEIN(0.999)STTHPGADLQ(0.623)EN(0.378)FCR; SRYPHKPEIN(0.188)STTHPGADLQ(0.899)EN(0.914)FCR	C9JV37 Prothrombin (Fragment)
R.YPHKPEIN[+1913.67702]STTHPGADLQENFC[+57.02146]R.N	YPHKPEINSTTHPGADLQENFCR		YPHKPEIN(0.92)STTHPGADLQ(0.058)EN(0.022)FCR	C9JV37 Prothrombin (Fragment)
R.DGPC[+57.02146]GTVLTRN[+1913.67702]ETHATYSN[+0.98402]TLYLADEIIIRDLNIK.I	DGPCGTVLTRNETHATYSNTLYLADEIIIRDLNIK		DGPCGTVLTRN(0.967)ETHATYSN(0.033)TLYLADEIIIRDLNIK	UROM Uromodulin
K.YLGN[+2204.77244]ATAIFFLPDEGKLQHLENELTHDIITK.F	YLGNATAIFFLPDEGKLQHLENELTHDIITK	HexNAc(4)Hex(5)NeuAc(2)	YLGN(0.998)ATAIFFLPDEGKLQ(0.013)HLEN(0.99)ELTHDIITK; YLGN(0.999)ATAIFFLPDEGKLQHLEN(0.001)ELTHDIITK	A0A0G2JRN3 Alpha-1-antitrypsin
K.YLGN[+2204.77244]ATAIFFLPDEGKLQHLEN[+0.98402]ELTHDIITK.F	YLGNATAIFFLPDEGKLQHLENELTHDIITK		YLGN(0.998)ATAIFFLPDEGKLQ(0.013)HLEN(0.99)ELTHDIITK; YLGN(0.999)ATAIFFLPDEGKLQHLEN(0.001)ELTHDIITK	A0A0G2JRN3 Alpha-1-antitrypsin
R.VYIHPFHLVIHN[+2204.77244]ESTC[+57.02146]EQLAK.A	VYIHPFHLVIHNESTCEQLAK		VYIHPFHLVIHN(0.966)ESTCEQ(0.034)LAK	ANGT Angiotensinogen
R.SRYPHKPEIN[+2204.77244]STTHPGADLQENFC[+57.02146]R.N	SRYPHKPEINSTTHPGADLQENFCR		SRYPHKPEIN(0.984)STTHPGADLQ(0.016)EN(0.001)FCR; SRYPHKPEIN(1)STTHPGADLQ(0.022)EN(0.978)FCR; SRYPHKPEIN(0.999)STTHPGADLQ(0.623)EN(0.378)FCR; SRYPHKPEIN(0.188)STTHPGADLQ(0.899)EN(0.914)FCR	C9JV37 Prothrombin (Fragment)
R.LRN[+2204.77244]VSWATGR.S	LRNVSWATGR		LRN(1)VSWATGR	D6RFL4 Monocyte differentiation antigen CD14 (Fragment)
R.KVC[+57.02146]QDC[+57.02146]PLLAPLN[+2204.77244]DTRVVHAAK.A	KVCQDCPLLAPLNDTRVVHAAK		KVCQDCPLLAPLN(1)DTRVVHAAK	FETUA Alpha-2-HS-glycoprotein
K.VC[+57.02146]QDC[+57.02146]PLLAPLN[+2204.77244]DTRVVHAAK.A	VCQDCPLLAPLNDTRVVHAAK		VCQDCPLLAPLN(1)DTRVVHAAK	FETUA Alpha-2-HS-glycoprotein
R.TQPSSEN[+2204.77244]KTAHLHK.G	TQPSSENKTAHLHK		TQPSSEN(1)KTAHLHK	H0YFL7 Complement C1r subcomponent-like protein (Fragment)
R.LHEITN[+2204.77244]ETFR.G	LHEITNETFR		LHEITN(1)ETFR	VASN Vasorin
R.LTYRN[+2204.77244]LSGPDKR.L	LTYRNLSGPDKR		LTYRN(1)LSGPDKR	VASN Vasorin
K.ALVLEQLTPALHSTN[+2092.74516]FSC[+57.02146]VLVDPEQVVQR.H	ALVLEQLTPALHSTNFSCVLVDPEQVVQR	HexNAc(4)Hex(7)Fuc(1)	ALVLEQ(0.124)LTPALHSTN(0.875)FSCVLVDPEQVVQR; ALVLEQ(0.021)LTPALHSTN(0.979)FSCVLVDPEQ(0.8)VVQ(0.2)R;ALVLEQ(1)LTPALHSTN(0.998)FSCVLVDPEQ(0.002)VVQR	I18BP Interleukin-18-binding protein
R.DYNTKN[+1647.61324]GTIK.W	DYNTKNGTIK	HexNAc(5)Hex(3)Fuc(1)	DYN(1)TKN(1)GTIK	DSG1 Desmoglein-1
K.TKPREEQYN[+1647.61324]STYR.V	TKPREEQYNSTYR		TKPREEQ(0.012)YN(0.988)STYR;TKPREEQ(1)YN(1)STYR	IGHG1 Ig gamma-1 chain C region
K.TKPREEQFN[+1647.61324]STFR.V	TKPREEQFNSTFR		TKPREEQ(0.21)FN(0.79)STFR	IGHG2 Ig gamma-2 chain C region
K.HTAEYAAN[+1663.60816]ITK.S	HTAEYAANITK	HexNAc(5)Hex(4)	HTAEYAAN(1)ITK	AMPE Glutamyl aminopeptidase
K.KLN[+1663.60816]YTLSQGHR.V	KLNYTLSQGHR		KLN(1)YTLSQGHR	AMPN Aminopeptidase N
K.KLN[+1809.66607]YTLSQGHR.V	KLNYTLSQGHR	HexNAc(5)Hex(4)Fuc(1)	KLN(1)YTLSQGHR	AMPN Aminopeptidase N
R.LQDERAN[+1809.66607]LTTLAGTHTNIPK.L	LQDERANLTTLAGTHTNIPK		LQ(0.022)DERAN(0.978)LTTLAGTHTNIPK	DPEP1 Dipeptidase 1
R.AN[+1809.66607]LTTLAGTHTNIPK.L	ANLTTLAGTHTNIPK		AN(1)LTTLAGTHTNIPK	DPEP1 Dipeptidase 1
R.DYNTKN[+1809.66607]GTIK.W	DYNTKNGTIK		DYN(1)TKN(1)GTIK	DSG1 Desmoglein-1
K.TKPREEQYN[+1809.66607]STYR.V	TKPREEQYNSTYR		TKPREEQ(0.012)YN(0.988)STYR;TKPREEQ(1)YN(1)STYR	IGHG1 Ig gamma-1 chain C region
R.LHEITN[+1809.66607]ETFR.G	LHEITNETFR		LHEITN(1)ETFR	VASN Vasorin
R.DYNTKN[+2100.76148]GTIK.W	DYNTKNGTIK	HexNAc(5)Hex(4)Fuc(1)NeuAc(1)	DYN(1)TKN(1)GTIK	DSG1 Desmoglein-1
R.KVN[+2100.76148]ASVPR.G	KVNASVPR		KVN(1)ASVPR	K7EMN4 Amyloid-like protein 1 (Fragment)
R.LHEITN[+2100.76148]ETFR.G	LHEITNETFR		LHEITN(1)ETFR	VASN Vasorin
R.LQDERAN[+1955.72397]LTTLAGTHTNIPK.L	LQDERANLTTLAGTHTNIPK	HexNAc(5)Hex(4)Fuc(2)	LQ(0.022)DERAN(0.978)LTTLAGTHTNIPK	DPEP1 Dipeptidase 1
R.KVN[+1955.72397]ASVPR.G	KVNASVPR		KVN(1)ASVPR	K7EMN4 Amyloid-like protein 1 (Fragment)

Table 1 (continued)

Modified Sequence	Sequence	Glycans composition	Deamidation (NQ) Probabilities	Protein name
K.KLN[+1825.66098]YTLSQGHR.V	KLNYTLSQGHR	HexNAc(5)Hex(5)	KLN(1)YTLSQGHR	AMPN Aminopeptidase N
R.AGFN[+1825.66098]ASFHK.S	AGFNASFHK		AGFN(1)ASFHK	CUBN Cubilin
K.YKN[+1971.71889]NSDISSTR.G	YKNNSDISSTR	HexNAc(5)Hex(5)Fuc(1)	YKN(0.987)N(0.013)SDISSTR	A0A075B6N9 Ig mu chain C region (Fragment)
K.KLN[+1971.71889]YTLSQGHR.V	KLNYTLSQGHR		KLN(1)YTLSQGHR	AMPN Aminopeptidase N
R.EGN[+1971.71889]ATGHLVGR.Y	EGNATGHLVGR		EGN(1)ATGHLVGR	CUBN Cubilin
R.AGFN[+1971.71889]ASFHK.S	AGFNASFHK		AGFN(1)ASFHK	CUBN Cubilin
R.LQDERAN[+1971.71889]LTTLAGTHTNIPK.L	LQDERANLTTLAGTHTNIPK		LQ(0.022)DERAN(0.978)LTTLAGTHTNIPK	DPEP1 Dipeptidase 1
R.LHEITN[+1971.71889]ETFR.G	LHEITNETFR		LHEITN(1)ETFR	VASN Vasorin
K.YKN[+2262.81431]NSDISSTR.G	YKNNSDISSTR	HexNAc(5)Hex(5)Fuc(1)NeuAc(1)	YKN(0.987)N(0.013)SDISSTR	A0A075B6N9 Ig mu chain C region (Fragment)
K.YLGN[+2262.81431]ATAIFFLPDEGKLQHLENELTHDIITK.F	YLGNATAIFFLPDEGKLQHLENELTHDIITK		YLGN(0.998)ATAIFFLPDEGKLQ(0.013)HLEN(0.99)ELTHDIITK;YLGN(0.999)ATAIFFLPDEGKLQHLEN(0.001)ELTHDIITK	A0A0G2JRN3 Alpha-1-antitrypsin
R.LHEITN[+2262.81431]ETFR.G	LHEITNETFR		LHEITN(1)ETFR	VASN Vasorin
R.EGN[+2117.77680]ATGHLVGR.Y	EGNATGHLVGR	HexNAc(5)Hex(5)Fuc(2)	EGN(1)ATGHLVGR	CUBN Cubilin
R.LQDERAN[+2117.77680]LTTLAGTHTNIPK.L	LQDERANLTTLAGTHTNIPK		LQ(0.022)DERAN(0.978)LTTLAGTHTNIPK	DPEP1 Dipeptidase 1
K.ALVLEQLTPALHSTN[+2117.77680]FSC[+57.02146]VLVDPEQVVQR.H	ALVLEQLTPALHSTNFSCVLVDPEQVVQR		ALVLEQ(0.124)LTPALHSTN(0.875)FSCVLVDPEQVVQR;ALVLEQ(0.021)LTPALHSTN(0.979)FSCVLVDPEQ(0.8)VVQ(0.2)R;ALVLEQ(1)LTPALHSTN(0.998)FSCVLVDPEQ(0.002)VVQR	I18BP Interleukin-18-binding protein
R.LQDERAN[+2116.75640]LTTLAGTHTNIPK.L	LQDERANLTTLAGTHTNIPK	HexNAc(5)Hex(5)NeuAc(1)	LQ(0.022)DERAN(0.978)LTTLAGTHTNIPK	DPEP1 Dipeptidase 1
K.AREDIFMETLKDIVEYYN[+2407.85181]DSNGSHVLQGR.F	AREDIFMETLKDIVEYYNDSNGSHVLQGR	HexNAc(5)Hex(5)NeuAc(2)	AREDIFMETLKDIVEYYN(0.033)DSN(0.952)GSHVLQ(0.015)GR;AREDIFMETLKDIVEYYN(0.857)DSN(0.973)GSHVLQ(0.17)GR	C9JEV0 Zinc-alpha-2-glycoprotein
R.LHEITN[+2133.77171]ETFR.G	LHEITNETFR	HexNAc(5)Hex(6)Fuc(1)	LHEITN(1)ETFR	VASN Vasorin
R.TQPSSEN[+3007.05796]KTAHLHK.G	TQPSSENKTAHLHK	HexNAc(5)Hex(6)Fuc(1)NeuAc(3)	TQPSSEN(1)KTAHLHK	H0YFL7 Complement C1r subcomponent-like protein (Fragment)
R.KFLN[+2279.82962]ESYKHEQVYIR.S	KFLNESYKHEQVYIR	HexNAc(5)Hex(6)Fuc(2)	KFLN(1)ESYKHEQVYIR	PPAP Prostatic acid phosphatase
K.FLN[+2279.82962]ESYKHEQVYIR.S	FLNESYKHEQVYIR		FLN(0.999)ESYKHEQ(0.001)VYIR;FLN(1)ESYKHEQ(1)VYIR	PPAP Prostatic acid phosphatase
R.TQPSSEN[+2570.92504]KTAHLHK.G	TQPSSENKTAHLHK	HexNAc(5)Hex(6)Fuc(2)NeuAc(1)	TQPSSEN(1)KTAHLHK	H0YFL7 Complement C1r subcomponent-like protein (Fragment)
K.FLN[+2570.92504]ESYKHEQVYIR.S	FLNESYKHEQVYIR		FLN(0.999)ESYKHEQ(0.001)VYIR;FLN(1)ESYKHEQ(1)VYIR	PPAP Prostatic acid phosphatase
R.TQPSSEN[+2861.00005]KTAHLHK.G	TQPSSENKTAHLHK	HexNAc(5)Hex(6)NeuAc(3)	TQPSSEN(1)KTAHLHK	H0YFL7 Complement C1r subcomponent-like protein (Fragment)
K.IITKEN[+1850.69261]ISQPR.G	IITKENISQPR	HexNAc(6)Hex(3)Fuc(1)	IITKEN(1)ISQPR;IITKEN(1)ISQ(1)PR	EGF Pro-epidermal growth factor
R.EDQYHYLLDRN[+1850.69261]LSC[+57.02146]R.V	EDQYHYLLDRNLSCR		EDQYHYLLDRN(1)LSCR;EDQ(1)YHYLLDRN(1)LSCR	IPSP Plasma serine protease inhibitor
R.LAN[+2141.78803]LTQGEDQYYLR.V	LANLTQGEDQYYLR	HexNAc(6)Hex(3)Fuc(1)NeuAc(1)	LAN(0.999)LTQ(0.001)GEDQYYLR;LAN(1)LTQ(0.221)GEDQ(0.779)YYLR	CLUS Clusterin
R.SKIITKEN[+2141.78803]ISQPR.G	SKIITKENISQPR		SKIITKEN(0.951)ISQ(0.049)PR	EGF Pro-epidermal growth factor
R.VVGVPYQGN[+2142.80843]ATALFILPSEGK.M	VVGVPYQGNATALFILPSEGK	HexNAc(6)Hex(3)Fuc(3)	VVGVPYQ(0.304)GN(0.696)ATALFILPSEGK;VVGVPYQ(1)GN(1)ATALFILPSEGK	IPSP Plasma serine protease inhibitor

*NeuAc +16 Da. Manual inspection of the Y1 ion in the MS/MS spectra revealed the +16 Da located on the peptide part

not find these ions but the NeuAc oxonium ion instead (some examples are provided in Supplemental Fig. 3). Since the mass difference between NeuAc and NeuGc is 16 Da, we checked if the Y1 ion of these peptides also contained the addition of +16 Da. Interestingly, from the 43 peptide sequences assigned with NeuGc, we were able to find in 22 of them the Y1 + 16 in the MS/MS spectra. Therefore, the peptides identified with NeuGc have a modification of +16 Da in the peptide sequence (Supplemental Fig. 2 A-U).

Most of the glycan structures were classified as hybrid/complex; amongst them 110 glycopeptides were identified with fucose and sialic acid (36 %), 78 with only fucose (25 %), 49 with only sialic acid (16 %) and 30 (10 %) composed by only HexNAc and Hex. High mannose was observed in 35 glycopeptides (11 %) (Fig. 3d and Supplemental Table S5). Interestingly, Gao *et al.* showed very similar distribution of N-glycans composition in human serum, where complex/hybrid glycans containing fucose and sialic acid were the main composition, followed by complex/hybrid with only sialic acid or only fucose, undecorated complex/hybrid and in less proportion the high mannose [53]. We also performed charge state distribution analysis of the intact glycopeptides identified in each class. We observed that glycopeptides containing NeuAc presented the highest percentage of >4 charges, while high mannose glycopeptides were identified mainly with +2 and +3 charges (Supplemental Fig. 3).

The 303 unique intact glycopeptides covered to 104 unique peptide sequences. These sequences had an overlap of 58 % (60 peptides) with the peptides identified in the deglycosylated fraction, in which the peptide harbored the deamidation

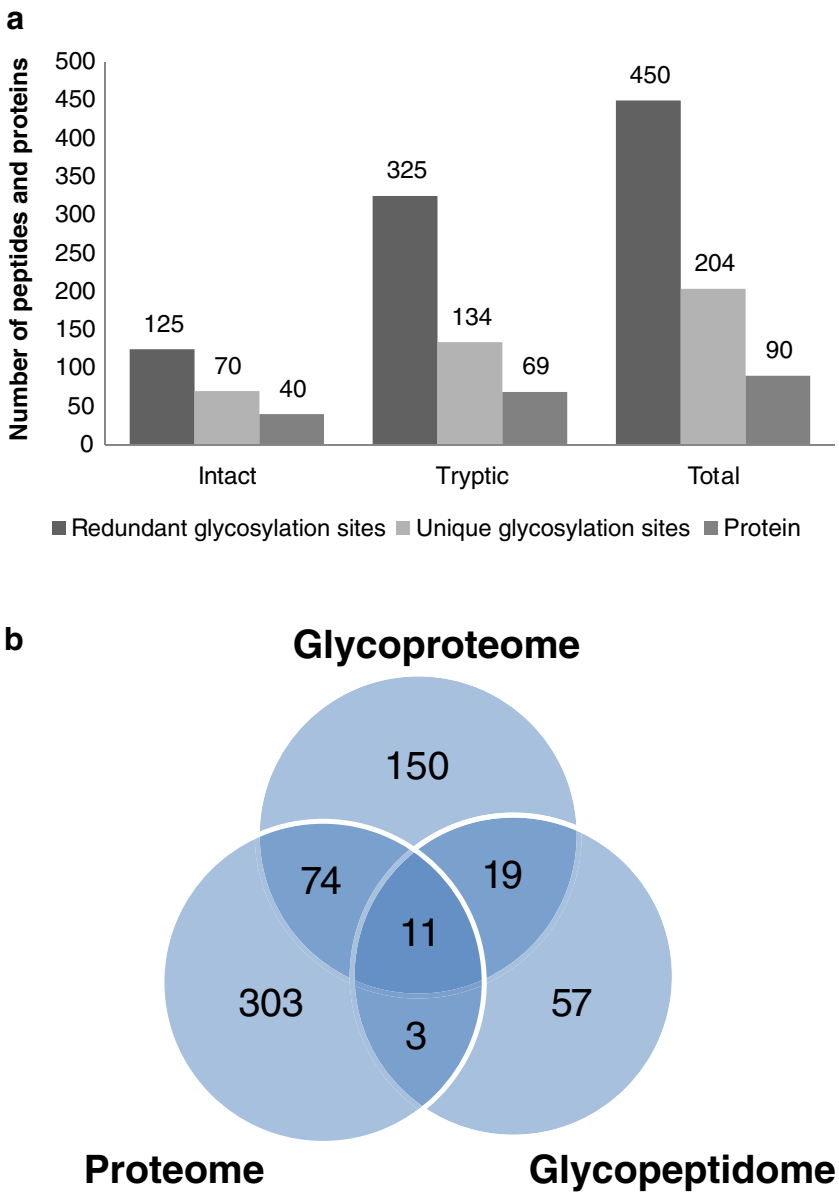

Fig. 4 N-glycosylation site analysis of endogenous glycopeptides (<10 kDa). **a** Distribution of identified formerly N-linked glycosylated peptides generated after PNGase F treatment in intact or tryptic endogenous glycopeptides. Only deamidation sites within NxS/T/C motif where x is not P were considered. **b** Distribution of proteins identified in the glycoproteome, proteome and glycopeptidome fraction

site in the consensus NXS/T/C motif (where X is not proline) (Supplemental Fig. 4). Using UnicarbKB database, we manually checked the possible structures for each glycan assigned to the 186 intact glycopeptides, whose peptide sequences were also identified in the formerly N-linked peptide fraction (Supplemental Table S6). From 57 glycan unique structures, 48 were found in the UnicarbKB database [54] (http://www.unicarbkb.org/) (Table 1). The glycopeptides showed in Table 1 were considered here as high confident identified glycopeptides in human urine, since they were evidenced by two different MS-based approaches and had the glycan structure annotated in a public database. Examples of annotated spectra of N-linked glycopeptides, directly from the Xcalibur program, with typical oxonium ions, glycosidic fragments and b/y-ions, are provided in Supplemental Fig. 5a-d.

N-linked glycosylation sites analysis of urinary endogenous peptides

Low molecular weight glycopeptides that were not retained by the 10 kDa-cutoff membrane was extracted and enriched using HILIC. The formerly N-linked peptides generated after PNGase F treatment were analyzed by LC-MS/MS. We identified 125 glycosylated peptides covering 70 unique glycosylation sites within NxS/T/C motifs, from 40 glycoproteins (Fig. 4a and Supplemental Table S7).

We further performed amino acid frequency analysis in order to search for specific protease preference cleavage sites that may be generating the endogenous glycopeptides. We found that leucine is the main amino acid occurring in the the NxT motif from the glycosylation site to the N-terminal of the peptide. On the other hand, in the NxS motif we did not observe the same preference and leucine was observed in the vicinity of the S to the C-terminal direction (Supplemental Fig. 6a and b). It was demonstrated that some metalloproteases (MMPs 1, 2, 3, 7, 9, 12, 13, 14) and ADAMs (ADAM17, ADAM10) have preferences for hydrophobic residues, such as leucine in the cleavage site [55–58]. Moreover, we found that leucine and isoleucine were the main aminoacids present at the N-terminus of the endogeneous glycopeptides (Supplemental Fig. 6c).

Since the peptide dabatabase search is compromised by large peptide sequences and no digestion mode, we performed tryptic digestion of the HILIC-enriched N-glycopeptides in order to increase the coverage of N-linked glycosylation sites and reduce false positives. Using this strategy we were able to increase the number of unique glycosylation sites to 134, which covered to 69 glycoproteins (Fig 4a and Supplemental Table S8).

In total, 204 unique N-linked glycosylation sites were identified in the human endogenous glycopeptide, covering90 glycoproteins. Interestingly, when we compared the overlap of proteins identified in the glycoproteome, proteome and glycopeptidome fraction, we observed that the glycopeptidome were able to retrieve exclusive proteins (57 out of 90) (Fig. 4b).

Chemical artifacts were searched using the intact glycopeptides fraction and the peptides bearing a spontaneously deamidated asparagine within the N-linked glycosylation motif were reported in Supplemental Table S1 and S2.

Conclusion

To our knowledge, the present dataset provides a highly confident characterization of N-linked glycosylation sites in human urinary proteins, including site-specific glycan microheterogeneity and for the first time the N-glycosylation sites from endogenous glycopeptides. This dataset, as well the methodology showed in this study, consist in a valuable resource for future MS-based glycoproteomic studies concerned with disorders in which N-linked protein glycosylation may be aberrant and an alternative approach for searching for novel candidate biomarkers in human urine.

Acknowledgments CNPq (GP 441878/2014-8) and FAPESP (GP: 2014/06863-3), Rebeca Kawahara is supported the Capes, PNPD, and FAPESP (2015/02866-0). Joyce Saad is supported by the "Programa Unificado de Bolsas de Estudo". The facility Biomass at CEFAP-USP is acknowledged for the analyses.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Apweiler R., Hermjakob H., Sharon N.: On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim. Biophys. Acta. **1473**(1), 4–8 (1999)
2. Ferris S.P., Kodali V.K., Kaufman R.J.: Glycoprotein folding and quality-control mechanisms in protein-folding diseases. Dis. Model. Mech. **7**(3), 331–341 (2014). doi:10.1242/dmm.014589
3. Marth J.D., Grewal P.K.: Mammalian glycosylation in immunity. Nat. Rev. Immunol. **8**(11), 874–887 (2008). doi:10.1038/nri2417
4. Cummings R.D., Pierce J.M.: The challenge and promise of glycomics. Chem. Biol. **21**(1), 1–15 (2014). doi:10.1016/j.chembiol.2013.12.010
5. Crocker P.R., Feizi T.: Carbohydrate recognition systems: functional triads in cell-cell interactions. Curr. Opin. Struct. Biol. **6**(5), 679–691 (1996)
6. Pinho S.S., Reis C.A.: Glycosylation in cancer: mechanisms and clinical implications. Nat. Rev. Cancer. **15**(9), 540–555 (2015). doi:10.1038/nrc3982
7. Fuster M.M., Esko J.D.: The sweet and sour of cancer: glycans as novel therapeutic targets. Nat. Rev. Cancer. **5**(7), 526–542 (2005). doi:10.1038/nrc1649

8. Dube D.H., Bertozzi C.R.: Glycans in cancer and inflammation–potential for therapeutics and diagnostics. Nat. Rev. Drug Discov. **4**(6), 477–488 (2005). doi:10.1038/nrd1751
9. Stowell S.R., Ju T., Cummings R.D.: Protein glycosylation in cancer. Annu. Rev. Pathol. **10**, 473–510 (2015). doi:10.1146/annurev-pathol-012414-040438
10. Gilgunn S., Conroy P.J., Saldova R., Rudd P.M., O'Kennedy R.J.: Aberrant PSA glycosylation–a sweet predictor of prostate cancer. Nat. Rev. Urol. **10**(2), 99–107 (2013). doi:10.1038/nrurol.2012.258
11. Hauselmann I., Borsig L.: Altered tumor-cell glycosylation promotes metastasis. Frontiers in oncology. **4**, 28 (2014). doi:10.3389/fonc.2014.00028
12. Zhang Y., Jiao J., Yang P., Lu H.: Mass spectrometry-based N-glycoproteomics for cancer biomarker discovery. Clin. Proteomics. **11**(1), 18 (2014). doi:10.1186/1559-0275-11-18
13. Thaysen-Andersen M., Packer N.H.: 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**(9), 1437–1452 (2014). doi:10.1016/j.bbapap.2014.05.002
14. Kaji H., Saito H., Yamauchi Y., Shinkawa T., Taoka M., Hirabayashi J., Kasai K., Takahashi N., Isobe T.: Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. Nat. Biotechnol. **21**(6), 667–672 (2003). doi:10.1038/nbt829
15. Morelle W., Faid V., Chirat F., Michalski J.C.: Analysis of N- and O-linked glycans from glycoproteins using MALDI-TOF mass spectrometry. Methods Mol. Biol. **534**, 5–21 (2009). doi:10.1007/978-1-59745-022-5_1
16. Jensen P.H., Karlsson N.G., Kolarich D., Packer N.H.: Structural analysis of N- and O-glycans released from glycoproteins. Nat. Protoc. **7**(7), 1299–1310 (2012). doi:10.1038/nprot.2012.063
17. Woo C.M., Iavarone A.T., Spiciarich D.R., Palaniappan K.K., Bertozzi C.R.: Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. Nat. Methods. **12**(6), 561–567 (2015). doi:10.1038/nmeth.3366
18. Sun S., Shah P., Eshghi S.T., Yang W., Trikannad N., Yang S., Chen L., Aiyetan P., Hoti N., Zhang Z., Chan D.W., Zhang H.: Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. Nat. Biotechnol. **34**(1), 84–88 (2016). doi:10.1038/nbt.3403
19. Medzihradszky K.F., Kaasik K., Chalkley R.J.: Tissue-specific glycosylation at the glycopeptide level. Mol. Cell. Proteomics: MCP. **14**(8), 2103–2110 (2015). doi:10.1074/mcp.M115.050393
20. Hoffmann M., Marx K., Reichl U., Wuhrer M., Rapp E.: Site-specific O-glycosylation analysis of human blood plasma proteins. Mol. Cell. Proteomics: MCP. **15**(2), 624–641 (2016). doi:10.1074/mcp.M115.053546
21. Parker B.L., Thaysen-Andersen M., Solis N., Scott N.E., Larsen M.R., Graham M.E., Packer N.H., Cordwell S.J.: Site-specific glycan-peptide analysis for determination of N-glycoproteome heterogeneity. J. Proteome Res. **12**(12), 5791–5800 (2013). doi:10.1021/pr400783j
22. Stavenhagen K., Hinneburg H., Thaysen-Andersen M., Hartmann L., Varon Silva D., Fuchser J., Kaspar S., Rapp E., Seeberger P.H., Kolarich D.: Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. J. Mass Spectrom.: JMS. **48**(6), 627–639 (2013). doi:10.1002/jms.3210
23. Hagglund P., Bunkenborg J., Elortza F., Jensen O.N., Roepstorff P.: A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. J. Proteome Res. **3**(3), 556–566 (2004)
24. Mysling S., Palmisano G., Hojrup P., Thaysen-Andersen M.: Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. Anal. Chem. **82**(13), 5598–5609 (2010). doi:10.1021/ac100530w
25. Li, X., Jiang, J., Zhao, X., Wang, J., Han, H., Zhao, Y., Peng, B., Zhong, R., Ying, W., Qian, X.: N-glycoproteome analysis of the secretome of human metastatic hepatocellular carcinoma cell lines combining hydrazide chemistry, HILIC enrichment and mass spectrometry. PloS one 8(12), e81921 (2013). doi:10.1371/journal.pone.0081921
26. Melo-Braga M.N., Schulz M., Liu Q., Swistowski A., Palmisano G., Engholm-Keller K., Jakobsen L., Zeng X., Larsen M.R.: Comprehensive quantitative comparison of the membrane proteome, phosphoproteome, and sialiome of human embryonic and neural stem cells. Mol. Cell. Proteomics: MCP. **13**(1), 311–328 (2014). doi:10.1074/mcp.M112.026898
27. Pompach P., Chandler K.B., Lan R., Edwards N., Goldman R.: Semi-automated identification of N-glycopeptides by hydrophilic interaction chromatography, nano-reverse-phase LC-MS/MS, and glycan database search. J. Proteome Res. **11**(3), 1728–1740 (2012). doi:10.1021/pr201183w
28. Cheng K., Chen R., Seebun D., Ye M., Figeys D., Zou H.: Large-scale characterization of intact N-glycopeptides using an automated glycoproteomic method. J. Proteome. **110**, 145–154 (2014). doi:10.1016/j.jprot.2014.08.006
29. Wuhrer M., de Boer A.R., Deelder A.M.: Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry. Mass Spectrom. Rev. **28**(2), 192–206 (2009). doi:10.1002/mas.20195
30. Yu Y.Q., Gilar M., Kaska J., Gebler J.C.: A rapid sample preparation method for mass spectrometric characterization of N-linked glycans. Rapid Commun. Mass Spectrom.: RCM. **19**(16), 2331–2336 (2005). doi:10.1002/rcm.2067
31. Shimwell N.J., Bryan R.T., Wei W., James N.D., Cheng K.K., Zeegers M.P., Johnson P.J., Martin A., Ward D.G.: Combined proteome and transcriptome analyses for the discovery of urinary biomarkers for urothelial carcinoma. Br. J. Cancer. **108**(9), 1854–1861 (2013). doi:10.1038/bjc.2013.157
32. Zhang H., Cao J., Li L., Liu Y., Zhao H., Li N., Li B., Zhang A., Huang H., Chen S., Dong M., Yu L., Zhang J., Chen L.: Identification of urine protein biomarkers with the potential for early detection of lung cancer. Sci. Rep. **5**, 11805 (2015). doi:10.1038/srep11805
33. Wu J., Chen Y.D., Gu W.: Urinary proteomics as a novel tool for biomarker discovery in kidney diseases. J. Zhejiang Univ. Sci. B. **11**(4), 227–237 (2010). doi:10.1631/jzus.B0900327
34. Thomas, C.E., Sexton, W., Benson, K., Sutphen, R., Koomen, J.: Urine collection and processing for protein biomarker discovery and quantification. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 19(4), 953–959 (2010). doi:10.1158/1055-9965.EPI-10-0069
35. Overbye A., Skotland T., Koehler C.J., Thiede B., Seierstad T., Berge V., Sandvig K., Llorente A.: Identification of prostate cancer biomarkers in urinary exosomes. Oncotarget. **6**(30), 30357–30376 (2015). doi:10.18632/oncotarget.4851
36. Haj-Ahmad T.A., Abdalla M.A., Haj-Ahmad Y.: Potential urinary protein biomarker candidates for the accurate detection of prostate cancer among benign prostatic hyperplasia patients. J. Cancer. **5**(2), 103–114 (2014). doi:10.7150/jca.6890
37. Jedinak A., Curatolo A., Zurakowski D., Dillon S., Bhasin M.K., Libermann T.A., Roy R., Sachdev M., Loughlin K.R., Moses M.A.: Novel non-invasive biomarkers that distinguish between benign prostate hyperplasia and prostate cancer. BMC Cancer. **15**, 259 (2015). doi:10.1186/s12885-015-1284-z

38. Wang L., Li F., Sun W., Wu S., Wang X., Zhang L., Zheng D., Wang J., Gao Y.: Concanavalin A-captured glycoproteins in healthy human urine. Mol. Cell. Proteomics: MCP. **5**(3), 560–562 (2006). doi:10.1074/mcp.D500013-MCP200
39. Yang N., Feng S., Shedden K., Xie X., Liu Y., Rosser C.J., Lubman D.M., Goodison S.: Urinary glycoprotein biomarker discovery for bladder cancer detection using LC/MS-MS and label-free quantification. Clinical cancer research: an official journal of the American Association for Cancer Research. **17**(10), 3349–3359 (2011). doi:10.1158/1078-0432.CCR-10-3121
40. Saraswat M., Joenvaara S., Musante L., Peltoniemi H., Holthofer H., Renkonen R.: N-linked (N-) glycoproteomics of urinary exosomes. [Corrected]. Mol. Cell. Proteomics: MCP. **14**(2), 263–276 (2015). doi:10.1074/mcp.M114.040345
41. Halim A., Nilsson J., Ruetschi U., Hesse C., Larson G.: : Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. Mol. Cell. Proteomics: MCP. **11**(4), M111 013649 (2012). doi:10.1074/mcp.M111.013649
42. Bern, M., Kil, Y.J., Becker, C.: Byonic: advanced peptide and protein identification software. Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis. .. [et al.] **Chapter 13**, Unit13 20 (2012). doi:10.1002/0471250953.bi1320s40
43. Bern M., Cai Y., Goldberg D.: Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. Anal. Chem. **79**(4), 1393–1400 (2007). doi:10.1021/ac0617013
44. Bern M.W., Kil Y.J.: Two-dimensional target decoy strategy for shotgun proteomics. J. Proteome Res. **10**(12), 5296–5301 (2011). doi:10.1021/pr200780j
45. Cox J., Mann M.: MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. **26**(12), 1367–1372 (2008). doi:10.1038/nbt.1511
46. Cox J., Neuhauser N., Michalski A., Scheltema R.A., Olsen J.V., Mann M.: Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. **10**(4), 1794–1805 (2011). doi:10.1021/pr101065j
47. Palmisano G., Melo-Braga M.N., Engholm-Keller K., Parker B.L., Larsen M.R.: Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. J. Proteome Res. **11**(3), 1949–1957 (2012). doi:10.1021/pr2011268
48. Breuer K., Foroushani A.K., Laird M.R., Chen C., Sribnaia A., Lo R., Winsor G.L., Hancock R.E., Brinkman F.S., Lynn D.J.: InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. Nucleic Acids Res. **41**(Database issue), D1228–D1233 (2013). doi:10.1093/nar/gks1147
49. Chen E.Y., Tan C.M., Kou Y., Duan Q., Wang Z., Meirelles G.V., Clark N.R., Ma'ayan A.: Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinf. **14**, 128 (2013). doi:10.1186/1471-2105-14-128
50. Su A.I., Wiltshire T., Batalov S., Lapp H., Ching K.A., Block D., Zhang J., Soden R., Hayakawa M., Kreiman G., Cooke M.P., Walker J.R., Hogenesch J.B.: A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. U. S. A. **101**(16), 6062–6067 (2004). doi:10.1073/pnas.0400782101
51. Kim M.S., Pinto S.M., Getnet D., Nirujogi R.S., Manda S.S., Chaerkady R., Madugundu A.K., Kelkar D.S., Isserlin R., Jain S., Thomas J.K., Muthusamy B., Leal-Rojas P., Kumar P., Sahasrabuddhe N.A., Balakrishnan L., Advani J., George B., Renuse S., Selvan L.D., Patil A.H., Nanjappa V., Radhakrishnan A., Prasad S., Subbannayya T., Raju R., Kumar M., Sreenivasamurthy S.K., Marimuthu A., Sathe G.J., Chavan S., Datta K.K., Subbannayya Y., Sahu A., Yelamanchi S.D., Jayaram S., Rajagopalan P., Sharma J., Murthy K.R., Syed N., Goel R., Khan A.A., Ahmad S., Dey G., Mudgal K., Chatterjee A., Huang T.C., Zhong J., Wu X., Shaw P.G., Freed D., Zahari M.S., Mukherjee K.K., Shankar S., Mahadevan A., Lam H., Mitchell C.J., Shankar S.K., Satishchandra P., Schroeder J.T., Sirdeshmukh R., Maitra A., Leach S.D., Drake C.G., Halushka M.K., Prasad T.S., Hruban R.H., Kerr C.L., Bader G.D., Iacobuzio-Donahue C.A., Gowda H., Pandey A.: A draft map of the human proteome. Nature. **509**(7502), 575–581 (2014). doi:10.1038/nature13302
52. Shah P., Wang X., Yang W., Toghi Eshghi S., Sun S., Hoti N., Chen L., Yang S., Pasay J., Rubin A., Zhang H.: Integrated proteomic and glycoproteomic analyses of prostate cancer cells reveal glycoprotein alteration in protein abundance and glycosylation. Mol. Cell. Proteomics: MCP. **14**(10), 2753–2763 (2015). doi:10.1074/mcp.M115.047928
53. Gao W.N., Yau L.F., Liu L., Zeng X., Chen D.C., Jiang M., Liu J., Wang J.R., Jiang Z.H.: Microfluidic Chip-LC/MS-based Glycomic analysis revealed distinct N-glycan profile of rat serum. Sci. Rep. **5**, 12844 (2015). doi:10.1038/srep12844
54. Campbell M.P., Peterson R., Mariethoz J., Gasteiger E., Akune Y., Aoki-Kinoshita K.F., Lisacek F., Packer N.H.: UniCarbKB: building a knowledge platform for glycoproteomics. Nucleic Acids Res. **42**(Database issue), D215–D221 (2014). doi:10.1093/nar/gkt1128
55. Eckhard U., Huesgen P.F., Schilling O., Bellac C.L., Butler G.S., Cox J.H., Dufour A., Goebeler V., Kappelhoff R., Keller U.A., Klein T., Lange P.F., Marino G., Morrison C.J., Prudova A., Rodriguez D., Starr A.E., Wang Y., Overall C.M.: Active site specificity profiling of the matrix metalloproteinase family: Proteomic identification of 4300 cleavage sites by nine MMPs explored with structural and synthetic peptide cleavage analyses. Matrix Biol.: journal of the International Society for Matrix Biology. **49**, 37–60 (2016). doi:10.1016/j.matbio.2015.09.003
56. Tucher J., Linke D., Koudelka T., Cassidy L., Tredup C., Wichert R., Pietrzik C., Becker-Pauly C., Tholey A.: LC-MS based cleavage site profiling of the proteases ADAM10 and ADAM17 using proteome-derived peptide libraries. J. Proteome Res. **13**(4), 2205–2214 (2014). doi:10.1021/pr401135u
57. Prudova A., auf dem Keller U., Butler G.S., Overall C.M.: Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. Mol. Cell. Proteomics: MCP. **9**(5), 894–911 (2010). doi:10.1074/mcp.M000050-MCP201
58. Schilling O., Overall C.M.: Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. Nat. Biotechnol. **26**(6), 685–694 (2008). doi:10.1038/nbt1408