Glycoproteomics assays for prostate cancer biomarker discovery

Wang, W.

Chapter 4

Human Prostate-Specific Antigen Carries N-glycans with Ketodeoxynononic Acid

Wei Wang[1], Tao Zhang[1], Jan Nouta[1], Peter A. van Veelen[1], Noortje de Haan[1,2], Theo M. de Reijke[3], Manfred Wuhrer[1], Guinevere S.M. Lageveen-Kammeijer[1*]
1 Leiden University Medical Center, Center for Proteomics and Metabolomics, Leiden, the Netherlands
2 Current address: University of Copenhagen, Copenhagen Center for Glycomics, Copenhagen, Denmark
3 Amsterdam UMC, Department of Urology, University of Amsterdam, Amsterdam, the Netherlands

Reprinted and adapted with permission from Engineering, 2023, DOI: 10.1016/j.eng.2023.02.009

Abstract: Ketodeoxynononic acid (Kdn) is a rather uncommon sialic acid class in mammals. However, associations have been found between elevated concentrations of free or conjugated Kdn in relation to human cancer progression. Hitherto, there is a lack of conclusive evidence that Kdn occurs on (specific) human glycoproteins (conjugated Kdn). Here, we report for the first time that Kdn is expressed on prostate-specific antigen (PSA) *N*-linked glycans derived from human seminal plasma and urine. Interestingly, Kdn was found only in an α2,3-linkage configuration to an antennary galactose, indicating a highly specific biosynthesis. This unusual glycosylation feature was also identified in an urinary PSA cohort in relation to prostate cancer (PCa), although no differences were found between PCa and non-PCa patients. Further research is needed to investigate the occurrence, biosynthesis, biological role and biomarker potential of the free as well as conjugated Kdn in human.

Keywords: Ketodeoxynononic acid; Kdn; Glycosylation; Prostate cancer; Prostate-specific antigen.

4.1 Introduction

Ketodeoxynononic (Kdn) is a deaminated neuraminic acid and is widely distributed amongst bacteria and vertebrates[270]. Kdn is part of the structurally highly diverse sialic acid monosaccharide family[98]. Since its first discovery in rainbow trout eggs, it has been found on a broad range of glycoconjugates including glycolipids, *N*- and *O*-linked glycoproteins as well as capsular polysaccharides, in diverse linkages (α2,3, α2,4, α2,6, and α2,8)[271-274]. In the cortical alveolar polysialoglycoprotein of rainbow trout eggs it was suggested that the attachment of the Kdn, in an α2,8-linkage to the nonreducing terminal *N*-glycolylneuraminic acid (Neu5Gc) in *O*-glycans, to be a capping reaction that prevents bacterial sialidases to hydrolyze the oligo/polysialic acid chains, implying that they play an important role in the early embryonic development of salmonid fish[275]. Later research suggested that the various linkages of Kdn are "chain stoppers" that control the extent of the chain elongation of sialyl groups during its biosynthetic process[276]. In addition, stage-specific and site-specific expression of different Kdn-containing gangliosides was observed in the spermatogenesis of rainbow trout, indicating that Kdn-containing gangliosides may play an important role in this process[277, 278]. Kdn is also considered as an important sialic acid component of skin mucus and other organs of fish[270, 279]. Besides, the presence of Kdn in cell wall polymers of pathogenic *Streptomyces* strains suggests that Kdn is involved at an early stage of the attachment of pathogenic bacteria to host plant cells, causing scab disease of potatoes and root crops[280, 281].

In contrast to the biosynthesis in bacteria and most vertebrates, the biosynthesis of Kdn is inactivated in most mammals by substitution of a methionine to a threonine at position 42 of the cytosolic protein, *N*-acetylneuraminic acid 9-phosphate synthase (Neu5Ac-9-P synthase)[116]. Which is the key enzyme for the synthesis of Kdn. It is known that the activity of this cytosolic protein is restored independently for the synthesis of Kdn 9-phosphate in two mammalian lineages: Artiodactyla and primates (including humans)[116, 117]. As a result, Kdn can still be produced in humans, albeit at low concentrations as the activity of Neu5Ac-9-P synthase for the production of Kdn is much lower than its activity for *N*-acetylneuraminic acid (Neu5Ac)[282].

Various studies report on the presence of Kdn as a free monosaccharide in mammals[118-120] and only small amounts occur as cytidine monophosphate (CMP)-Kdn[118] or conjugated to glycolipids and glycoproteins[121]. Kawanishi *et al.* demonstrated that the production of Kdn is linked to the ingestion of mannoses leading to an increased excretion of free Kdn in urine[116]. It was also reported that serum levels of free Kdn were significantly elevated in patients with end stage renal disease compared to healthy volunteers[116]. Moreover, elevated levels of the free Kdn monosaccharide and CMP-Kdn were found in ovarian tumor tissues as well as in ascites cells when compared to normal controls[118]. Similarly, substantial amounts of free Kdn-containing *N*-glycans were found in human

prostate cancer (PCa) tissues[122].

Several studies have investigated the presence of protein or lipid-linked Kdn. Typical amounts of free and conjugated Kdn are found to be 100 up to 1000 fold less compared to Neu5Ac in various human cancer cells, cancer tissues and normal tissues[118-121]. Due to its low abundance, conjugated Kdn has only been investigated after hydrolysis[118-121]. This inherently results in the loss of information about the carrier of Kdn. Nonetheless, the presence of Kdn has been associated with a variety of human cancers *in vitro* and *in vivo*. For instance, Kdn was observed in human lung carcinoma cell lines (A125, A549, Calu, Hotz) after hydrolysis of glycolipids and glycoproteins which were extracted from the cells[121]. The presence of Kdn was also found in throat tissues for head and neck cancer[120] but it remains unclear whether Kdn was already free or whether it was conjugated, as the tissues were treated with a mild acid hydrolysis and no extraction step was performed prior to the hydrolysis. In the human ovarian teratocarcinoma cell line PA-1, free Kdn was observed as well as two different Kdn-containing glycoproteins; one soluble (49 kDa) and the other membrane-associated (30 kDa), which were detected by Western blot[119]. The presence of Kdn in the 30 kDa glycoprotein was indicated using a monoclonal antibody against the Kdn(α2,3)Gal(β1)-epitope as well as using a KDNase. The 49 kDa protein was not reactive to the monoclonal antibody, but did react on the KDNase, suggesting that this glycoprotein contains Kdn in another linkage than the α2,3-linkage. While this is the first study to report on mammalian Kdn-containing glycoproteins, the authors were unable to identify the glycoproteins.

In this study, we identified and characterized for the first time the expression of Kdn on the mammalian glycoprotein prostate-specific antigen (PSA). This glycoprotein is used as a diagnostic tool for PCa as elevated concentrations in the circulation (>3 ng/mL) indicate an increased risk of PCa. However, due to its low specificity, there is an urgent need for a better biomarker that can prevent overdiagnosis[283, 284]. Various studies have demonstrated that alterations in the *N*-glycomic profile of PSA might be a better indicator for the occurrence of PCa[58, 59, 285], particularly sialylation is considered to be a promising glycosylation feature in this regard[58, 285]. As mentioned above, an accumulation of free Kdn-containing *N*-glycans has been reported in human PCa tissues, indicating that Kdn might play a role in PCa[122] and could be exploited as a diagnostic marker but further research is required. Here, we report on the presence of conjugated Kdn-containing *N*-glycans on PSA derived from seminal plasma as characterized by tandem mass spectrometry (MS/MS) of both released *N*-glycans and glycopeptides. Moreover, the abundance of Kdn on PSA is investigated in a clinical cohort using urinary PSA from patients with elevated serum PSA concentrations.

4.2 Materials and methods

4.2.1. Chemicals and Standards

PSA standard derived from seminal plasma was purchased from Lee BioSolutions (Maryland Heights, MO). TPCK treated trypsin (bovine pancreas) was purchased from Worthington Biochemical Corporation (Lakewood Township, NJ). Peptide N-glycosidase F (PNGase F) was purchased from Roche Diagnostics (Mannheim, Germany). Ammonium bicarbonate (ABC), ammonium acetate, Dowex® cation-exchange resin (50W-X8), NaCl, hydroxylamine solution 50% in water, hydrochloride (HCl), sodium tetraborohydride ($NaBH_4$), dithiothreitol (DTT), iodoacetamide (IAA), dimethyl sulfoxide (DMSO), ammonium hydroxide solution and 2-methylpyridine borane complex (PB) were acquired from Sigma-Aldrich (Steinheim, Germany). Isopropanol, formic acid (FA), glacial acetic acid, potassium hydroxide (KOH) and water of LC-MS grade H_2O were obtained from Honeywell Fluka (Seelze, Germany). Beta-mercaptoethanol (βME) and nonidet P-40 substitute (NP40) were purchased from VWR (Cleveland, OH). From Fluorochem (Hadfield, United Kingdom) 1-ethyl-3-(3-(dimethylamino)propyl)-carbodiimide hydrochloride (EDC) was obtained and acetonitrile (MeCN) of LC-MS grade from Biosolve (Valkenswaard, the Netherlands). Anthranilic acid (2-AA), 40% dimethylamine in water, 20% sodium dodecyl sulfate solution in water (SDS) and 1-hydroxybenzotriazole hydrate (HOBt) were obtained from Sigma-Aldrich (Buchs, Switzerland). Sodium bicarbonate (NaBC), sodium hydroxide (NaOH), methanol, ethanol, 2-propanol and trifluoroacetic acid (TFA), sodium phosphate dibasic dihydrate (Na_2HPO_4) and monopotassium phosphate (KH_2PO_4) were acquired from Merck (Darmstadt, Germany). Milli-Q water (MQ) was generated from a Q-Gard 2 system (Millipore, the Netherlands) maintained at ≥ 18 MΩ. Five times concentrated PBS (5xPBS) was made out of 0.16 M Na_2HPO_4, 0.02 M KH_2PO_4 and 0.73 M NaCl (pH 7.3). A dilution from the 5xPBS was made with MQ to obtain 1xPBS (pH 7.6). Conical 96-well Nunc plates were bought from Thermo Scientific (Roskilde, Denmark). Solid phase extraction (SPE) bulk sorbent Carbograph were obtained from Grace Discovery sciences (Maryland, USA). Multiscreen HTS 96 multiwell plates (pore size 0.45 µm) with high protein-binding membrane (hydrophobic Immobilon-P PVDF membrane) and 96-well PP Microplate were acquired from Millipore (Amsterdam, the Netherlands). From Orochem Technologies (Naperville, IL) a 96-well PP filter plate was obtained. TMTzero™ label reagent was purchased from Themo Fisher (Germany).

4.2.2. Clinical Samples

Urine samples were collected from individuals suspected of PCa (based on currently used PCa biomarker named PSA serum concentration, if its level >3 ng/mL, the individual is recommended for further examination) at the Amsterdam University Medical Centers (Amsterdam UMC, location AMC, Amsterdam, the Netherlands). Patients with cystitis, undergoing chemotherapy, using 5-alpha reductase inhibitors or having a history or presence of cancer or non-prostate urological disorders were excluded from the study. In total, 52 urine samples were collected prior to prostate biopsies. The collection of the urine samples

and clinical information was approved by the medical ethical committee of the Amsterdam UMC, location AMC (W16_010#16.020). For clinical information of the patients (e.g., age, serum PSA concentration, prostate volume, ethnicity, clinical tumor stage and Gleason score) see **Supplementary Information, Table S-1**. After urine donation (5-96 mL), the urine was cooled down to room temperature (RT) and stored at -80˚C. For more information in regard to the anti-PSA antibodies and the PSA capture from urine samples see **Supplementary Information, Section S-1.1 and S-1.2.** The urinary samples were processed in two batches, the samples of batch 1 were prepared and analyzed in our previous study[286] and the remaining solution (fully prepared samples; captured and digested) were stored for approximately six months and underwent three freeze and thaw cycles prior to the current measurement. Batch two was only stored for a maximum of two weeks and underwent one freeze and thaw cycle. This cohort study was a collaboration between Leiden University Medical Center (LUMC) and Amsterdam UMC. Clinical information of the patients as well as the urine collection was performed by Amsterdam UMC. The urinary PSA assay was developed using a female urine pool (FUP) consisting of 10 healthy female urine samples and was collected at LUMC that also used as negative and as positive controls (spiked PSA standard) for the clinical cohort analysis.

4.2.3. 2-AA Labeling of Released PSA N-glycans for Positive Mode Analysis by RP-nanoLC-MS/MS (Q-ToF)

A total of 10 µL of PSA standard (pre-concentrated to 7 µg/µL) was added to 20 µL 2% SDS and was shaken for 5 min on a plate shaker at max speed. Followed by incubation for 10 min at 60˚C. Samples were cooled down to RT and 20 µL of release mixture (10 µL 4% NP40, 10 µL acidic 5xPBS and 1 µL PNGase F) were added to the samples. Samples were shaken for another 5 min on a plate shaker at max speed. Subsequently, the samples were incubated overnight (ON) at 37˚C. Prior to 2-AA labeling, PSA released *N*-glycans were purified by cotton HILIC SPE as described before[168]. Briefly, 1 µL of PSA released *N*-glycans was purified and eluted with 10 µL MQ. Labeling was performed as described previously[287, 288]. Shortly, 50 µL of released PSA *N*-glycans was added to 50 µL labeling solution consisting of 25 µL 2-AA solution (48 µg/µL in 85% DMSO:15% acetic acid) and 25 µL PB solution (107 µg/µL in DMSO). After mixing, the samples were incubated for 2 h at 60˚C. Prior to injection, PSA glycans were diluted to 3.5 ng/µL with H_2O.

4.2.4. Released PSA *N*-glycans for Negative Mode Analysis by PGC-nanoLC-MS/MS (Ion trap)

The PSA *N*-glycan release was performed using a 96-well plate sample preparation method as previously described[289]. Briefly, 20 µL of PSA standard (1.5 µg/µL) were loaded on a hydrophobic Immobilon-P PVDF membrane in a 96-well plate format. Subsequently, denaturation was performed by applying 75 µL denaturation mix (72.5 µL of 8 M GuHCl and 2.5 µL of 200 mM DTT) and the plate

was shaken for 15 min at RT. Afterwards, the samples were incubated for 30 min at 60˚C in a humid box. Any unbound material was removed by centrifugation (500 g for 2 min). *N*-glycan release was achieved by adding PNGase F (2 U of enzyme diluted with MQ to 15 µL) to each well and ON incubation was performed at 37˚C. Released *N*-glycans were eluted from the PVDF plate by centrifugation (500 g for 2 min). The glycosylamines of the released *N*-glycans were hydrolyzed by adding 20 µL of 100 mM ammonium acetate (pH 5) to the sample and an additional incubation of 1 h at RT, followed by evaporation at 35˚C in a SpeedVac concentrator 5301 (Eppendorf, Hamburg, Germany). To reduce and desalt the released *N*-glycans, PGC-SPE cleanup was performed. For this purpose 20 µL of 1 M $NaBH_4$ in 50 mM KOH was added to each well using a 96-well polymerase chain reaction plate based protocol[289.] Additionally, samples were dried in a SpeedVac concentrator and reconstituted in 10 µL MQ and stored at -20˚C prior to the analysis by PGC-nanoLC-MS/MS.

4.2.5. Tryptic PSA Glycopeptides

For more information about the digestion procedure see **Supplementary Information, Section S-1.3**.

4.2.6. Derivatization on Tryptic PSA Glycopeptides: DA for Positive Mode Analysis by CE-MS/MS (Q-ToF)

After digestion, tryptic PSA glycopeptides from the PSA standard (seminal plasma) were derivatized to stabilize the sialic acids and enable the distinction between different linkages as described before[168]. One µL of the tryptic PSA digest was added to 20 µL DA reagent (250 mM 1-ethyl-3-(3-(dimethylamino)propyl)-carbodiimide hydrochloride (EDC), 500 mM HOBt and 250 mM dimethylamine in DMSO) and incubated together at 60˚C for 1 h in a 96-well plate. After adding 8 µL of 28% ammonium hydroxide solution, samples were incubated for an additional 2 h at 60˚C. The PSA glycopeptides were enriched from the reaction mixture using cotton HILIC SPE. Glycopeptides were eluted in 10 µL H_2O and stored at -20 ˚C prior to their analysis by CE-MS.

4.2.7. Derivatization on Tryptic PSA Glycopeptides: TMT Labeling for Positive Ionization Mode Analysis by RP-nanoLC-MS/MS (Orbitrap)

The amine-reactive tandem mass tag TMTzero was applied to increase hydrophobicity of tryptic PSA glycopeptides. TMT labels were designed for MS-level protein quantification which modifies lysine residues and the peptide *N*-termini. In brief, a mixture of five µL tryptic PSA (200 ng/µL) and 2 µL of TMTzero™ label reagent in MeCN (4 µg/µL) was made and incubated for 2 h at 37˚C. The reaction was quenched by adding 1.2 µL of 5% hydroxylamine, followed by an additional incubation of 15 min at RT.

4.2.8. CE-MS/MS

All CE-MS measurements were conducted on a CESI 8000 system (SCIEX, Framingham, MA), mounted with a neutral OptiMS cartridge (non-captured samples) or an in-house ultratrol coated BFS OptiMS capillary (captured samples) both capillaries were 91 cm long with an i.d. of 30 µm and o.d. of 150 µm (SCIEX). The CESI 8000 system was coupled to an UHR-QqToF maXis Impact HD mass spectrometer (Bruker Daltonics) using an OptiMS Bruker MS adapter (SCIEX) employed with an internal polymer cone onto the porous tip housing to enable the usage of a DEN gas[290]. Prior to analysis the neutral OptiMS cartridge was conditioned as stated by the manufacturers' protocol and the BFS OptiMS capillary was conditioned and coated in-house with ultratrol dynamic pre-coat LN (Target Discovery, Palo Alto, CA) as described by Lageveen-Kammeijer *et al.,*[291].

For the analysis on the neutral OptiMS capillary, the CE separation was conducted at 20 kV and 20°C using a background electrolyte (BGE) consisting of 20% acetic acid (pH 2.3). Prior to injection, 8 µL of the sample was mixed with 4 µL leading electrolyte (LE, 1200 mM ammonium acetate, pH 3.2). All standards (PSA derived from seminal plasma) were injected by applying 10 psi pressure for 60 s, corresponding to 13.5% of the total capillary volume (87 nL). After sample injection, a BGE post plug was injected by applying 2.5 psi for 10 s (0.56% capillary volume). Analysis on the in-house coated BFS OptiMS capillary was performed at 30 kV at 24°C using a 20% acetic acid with 10% MeOH as BGE (pH 2.0). A total of 1.5 µL of the clinical sample was used and 1.0 µL of 250 mM ammonium acetate, pH 4.0 (LE) was added. Injection was performed hydro-dynamically with 1 psi pressure for 60 s, corresponding to 1.4% of the total capillary volume (9 nL). Followed by a BGE post-plug using 0.5 psi for 25 seconds (0.3% capillary volume).

All experiments were performed in positive ionization mode. Electrospray ionization (ESI) was achieved using a glass capillary voltage between 1100 V and 1300 V. MeCN was used as dopant for the DEN gas at 0.2 bar. The temperature and drying gas (nitrogen) flow rate were set at 150°C and 1.2 L/min, respectively. MS spectra were acquired between *m/z* 200-2000, using 1 Hz as spectral acquisition rate.

4.2.9. RP-nanoLC-MS/MS (Q-ToF)

The analyses of 2-AA labeled released *N*-glycans were performed on an ultimate 3000 high-performance liquid chromatography (HPLC) system (Dionex Corporation, Sunnyvale, CA), coupled to a Maxis Impact HD quadrupole time of flight mass spectrometry (Q-ToF-MS) device (Bruker Daltonics, Bremen, Germany), equipped with a Dionex Acclaim PepMap100 C18 trap column (5 mm × 300 µm, Thermo Fisher Scientific, Breda, the Netherlands) and an Ascentis Express C18 nanoLC analytical column (50 mm × 75 µm with 2.7 µm fused core particles, Supelco, Bellefonte, PA). Mobile phase A consisted of an aqueous solution (0.1% TFA; *v/v*) while mobile phase B consisted of 95 % MeCN in MQ

(*v/v*). Gradient speed was set to increase mobile phase B by 1.73% per min. To enhance the sensitivity, DEN gas, using MeCN as a dopant, was employed at 0.2 bar. The dry gas flow was set at 3.0 L/min and the dry temperature at 180˚C. MS as well as MS/MS were obtained in positive ionization mode between *m/z* 150–2800 using a spectral acquisition frequency of 1 Hz.

4.2.10. RP-nanoLC-MS/MS (Orbitrap)

TMT labeled tryptic PSA glycopeptides were measured on a nanoLC-MS system composed of an Easy nLC1200 gradient HPLC system (Thermo, Bremen, Germany) coupled to an Orbitrap Fusion LUMOS MS (Thermo). The TMT labeled tryptic PSA glycopeptides (10 ng/µL in 20 µL water) were lyophilized and dissolved in 10 µL solvent A (water/MeCN/FA; 95/3/0.1; *v/v/v*) and loaded onto an in-house packed C18 precolumn (100 µm × 15 mm; Reprosil-Pur C18-AQ 3 µm, Dr. Maisch, Ammerbuch, Germany) and separated on a homemade analytical nanoLC column (50 cm × 75 µm; Reprosil-Pur C18-AQ 3 µm). For the elution of the analytes a linear gradient was set that went from 10 % to 40 % in 20 min for solvent B (water/MeCN/FA; 20/80/0.1; *v/v/v*). The nanoLC column was drawn to a tip of ~5 µm and acted as the electrospray needle of the MS source. The MS was operated in data-dependent MS/MS (top-20 mode). The MS1 were acquired with mass range of *m/z* 750–1800 at the AGC standard target value combined with a maximum accumulation time of 50 ms. The resolution setting for MS1 scans was set at 120,000. A dynamic exclusion was set for 10 s with a single repeat count, and charge states in the range 2-3 were included for MS/MS. The resolution of MS/MS scans was 30,000 at an AGC target of 2×10^4 with maximum fill time of 60 ms. MS/MS spectra were generated from precursors isolated with the quadrupole with an isolation width of 1.2 Da at a scan range (*m/z*) of 110-3500. An exclusion list of the 12 most abundant tryptic PSA glycopeptides was used (**Supplementary Information, Table S-2**). Upon detection of the *N*-acetylhexosamine (HexNAc) oxonium ion at *m/z* 204.087 (with a tolerance of 15 pm), an HCD scan was triggered with stepped normalized energies of 25%, 32% and 39%, at a normalized AGC target of 200% in combination with a maximum fill time of 200 ms[292].

4.2.11. PGC-nanoLC-ESI-MS/MS

The measurements of released PSA *N*-glycans were conducted on an ultimate 3000 UHPLC system (Dionex/Thermo), coupled to an amazon ETD speed ion trap (Bruker, Bremen, Germany), fitted with a home-packed PGC trap column (5 µm Hypercarb, 320 µm x 30 mm) and a home-packed PGC nano-column (3 µm Hypercarb 100 µm x 150 mm). Solvent A consisted out of 10 mM ABC and solvent B of 60% MeCN and 10 mM ABC (*v/v*). To analyze the released *N*-glycans, 3 µL of the 10 µL sample was injected and loaded on to the trap column using a 6 µL/min loading flow in 2% buffer B for 5 min. To achieve separation, a multi-step gradient was employed, in which solvent B increased from 2% up to 9% in

1 min, followed by an increase up to 49% in 80 min. After each analysis a 10 min washing step was performed using 95% of solvent B at a flow rate of 0.6 µL/min. Ionization was facilitated with the nanoBooster source using isopropanol as dopant. A column temperature of 45°C, capillary voltage of 1000 V, dry gas temperature of 280°C at 5 L/min and nebulizer at 3 psi were applied. Smart parameter setting was set to *m/z* 1200. MS spectra were acquired in enhanced mode using negative ion mode within a *m/z* range of 500-1850. MS/MS spectra were acquired for the three most abundant ions in intensity within a *m/z* range of 100-2500.

4.2.12 Data Analysis.

The analysis of the raw data obtained from Bruker MS instruments (RP-nanoLC-MS(/MS), PGC-nanoLC-MS(/MS) and CE-MS(/MS)) was performed manually by DataAnalysis 5.0 (Build 203, Bruker Daltonics) for glycopeptide and glycan discovery and identification. CE-MS(/MS) data was internally recalibrated with sodium adducts prior to peak integration. PSA glycopeptides and *N*-glycan compositions were extracted by the first three isotopes of the singly, doubly and triply charged analytes using a *m/z* window of ± 0.02 to acquire extracted ion chromatograms/electropherograms (EIC/EIE; smoothed with a Gaussian fit). The EICs and EIEs were used to evaluate the elution time (LC) and migration time (CE) of the analytes, the accurate mass (± 20 ppm), and the isotopic peak pattern (comparison to the theoretical isotopic pattern), followed by the integration of the peak area. Relative quantitation was performed by integrating the total area of all analytes within a single analysis. Observed *N*-glycan structures in the PGC-nanoLC-MS/MS analysis were assigned based on known MS/MS fragmentation patterns in negative-ion mode as well as by using Glycoworkbench[293] and Glycomod[294] software.

For the clinical samples, analytes were included for further data analysis when their migration time belonged to the correct migration clusters, ppm errors observed between -10 and 10 and at least present in more than 2/3 of the spectra. Spectral curation was performed based on the number of analytes identified in the spectrum. Spectra were excluded if the number of analytes identified was lower than 60 (2/3 of the highest number of analytes detected in the individual spectra). Finally, the absolute areas of the included analytes were normalized to the summed absolute area of all analytes (total area normalization). Further data analysis for the clinical samples was carried out using R (version 3.4.3, R Foundation for Statistical Computing) together with RStudio (version 1.2.1335). The data was visualized by principle component analysis (PCA; R, pcaMethods) of the two batches of samples.

The high mass accuracy analysis of the raw RP-nanoLC-MS(/MS) data of using Thermo MS instruments was performed manually by Xcalibur (Thermo Xcalibur 2.2 SP1.48) for more confident glycopeptide confirmation.

4.3 Results

Here, we present a mass spectrometry-based study demonstrating the presence of Kdn sialylation on human PSA from different biofluids (seminal plasma and urine; **Table 1**). Four Kdn-containing *N*-glycans on PSA were identified, being H5N4F1S1K1, H5N4F1K1 (H: hexose, N: *N*-acetylhexosamine, F: fucose, S: *N*-acetylneuraminic acid/Neu5Ac, K: Kdn) and their non-fucosylated variants (**Figure 1a**). The characterization and relative quantification was performed using four complementary analytical platforms being reversed phase nano-liquid chromatography (RP-nanoLC) either coupled to an Orbitrap or a quadrupole-time-of-flight (Q-ToF)-MS in positive ion mode, capillary electrophoresis (CE)-MS (positive ion mode; Q-ToF) and porous

graphitized carbon (PGC)-nanoLC-MS (negative ion mode; ion trap) with acquisition of collision-induced MS/MS spectra throughout (**Table 1, Supplementary Table S-3**). Released *N*-glycans as well as glycopeptides were assessed both with and

Table 1. The identification and key information obtained of Kdn-containing PSA glycans/glycopeptides with different analytical approaches using MS2.

Analytical platform	Mass analyzer	Ionization mode	Analyte class	Analyte modification	Key information
CE-MS	Q-ToF	Positive	Glycopeptides	NA	Kdn expression indicated by the presence of the oxonium ion of H1N1K1 antennae in MS/MS spectra of four tryptic glycopeptides of PSA (glycosylation site N_{69}). Site-specific glycosylation of PSA site N_{69} indicated via tryptic NK glycopeptide.
				Dimethyl-amidation followed by amidation	Sialic acid derivatization of Kdn on PSA demonstrates to be solely α2,3-linked.
RP-nanoLC-MS	Q-ToF	Positive	Released *N*-glycans	2AA labeling	Kdn expression is further supported by the presence of oxonium ions of H1N1K1 and H2N1K1 in MS/MS spectra.
PGC-nanoLC-MS	Ion trap	Negative	Released *N*-glycans	Reduction	MS/MS analysis reveals the Kdn monosaccharide as a deprotonated ion. Arm occupancy by Kdn is elucidated.
RP-nanoLC-MS	Orbitrap	Positive	Glycopeptides	TMT labeling	MS/MS analysis reveals the oxonium ion of the Kdn monosaccharide at high mass accuracy.

without linkage-specific sialic acid derivatization. For structural elucidation, reducing-end labeled and reduced N-glycans were analyzed by RP-nanoLC-MS/MS and PGC-nanoLC-MS/MS, respectively. Furthermore, the presence of Kdn on PSA as well as its association with PCa was explored by investigating urinary PSA derived from men with elevated serum PSA concentrations (>3 ng/mL). While the Kdn containing *N*-glycans were identified on a PSA standard, which is rather pure, it cannot be assumed that is also the case of captured PSA from a biofluid. Namely, aspecific binding of other glycoproteins could lead to a co-capture and, in the case of an *N*-glycan release, all the *N*-glycans present in the sample would be released. This would result in a biased glycomic profile due to contamination of the other glycoproteins. In the case of contamination glycomic profiles of PSA would be biased. By performing a proteolytic cleavage, PSA glycopeptides could be specifically targeted as the peptide moiety could be used as an identifier of PSA.

4.3.1. Expression of Kdn on mammalian PSA

The four Kdn-containing *N*-glycans on PSA were initially observed with CE-MS. CE-MS is very well-known for its high sensitivity which enables the detected of low abundance analytes like conjugated Kdn. Notably, inspection of the MS/MS of non-derivatized PSA glycopeptides by CE-MS revealed distinct B- and Y-ions which were uncommon for *N*-glycans retrieved from a human source (*e.g.* B-ion at *m/z* 616.208^{1+} and Y-ions at *m/z* 1606.621^{1+}, *m/z* 1752.679^{1+}, *m/z* 1768.674^{1+} and *m/z* 1914.732^{1+}; **Figure 1** and **Supplementary Figure S-1**). The observation of oxonium ion at *m/z* 616.208^{1+} suggests the presence of a Kdn modification. In this case, the oxonium ion at *m/z* 616.208^{1+} represents the composition H1N1K1 and the distinct Y-ions at *m/z* 1606.621^{1+}, *m/z* 1752.679^{1+}, *m/z* 1768.674^{1+} and *m/z* 1914.732^{1+} could be assigned to H3N3K1[NK], H3N3F1K1[NK], H4N3K1[NK] and H4N3F1K1[NK], respectively ([NK] indicates the asparagine-lysine peptide portion). Interestingly, several isomers were found by generating the extracted ion electropherogram (EIE) of *m/z* 1140.436^{2+}, corresponding to the mass of Kdn-glycopeptide H5N4F1K1[NK] (**Supplementary Information, Figure S-1a**). MS/MS showed, next to the Kdn-glycopeptide (**Supplementary Information, Figures S-1e-g**; with oxonium ion at *m/z* 616.208^{1+} for H1N1K1), the presence of the isomeric hybrid-type *N*-glycan, H6N3F1S1[NK] (assignment supported by H1N1S1 at *m/z* 657.235^{1+}, H5N3F1[NK] at *m/z* 1826.716^{1+} (loss of a H1S1) and H6N3F1[NK] at *m/z* 1988.769^{1+} (loss of S1; **Supplementary Information, Figures S-1f-h**) which was in line with our previous findings[286]. An according range of variants was detected lacking fucose. Namely, the fragmentation of *m/z* 1067.407^{2+}, corresponding to peak 1.4.1 (**Supplementary Information, Figure S-1i**), revealed the presence of both the Kdn-containing glycopeptide (H5N4K1[NK]; H1N1K1 at *m/z* 616.208^{1+}) as well as a hybrid-type *N*-glycan (H6N3S1[NK]; H1N1S1 at *m/z* 657.235^{1+}, H5N3[NK] at *m/z* 1680.658^{1+} and H6N3[NK] at *m/z* 1842.711^{1+}) while the fragmentation of *m/z* 1067.407^{2+}, corresponding to peak 1.4.2, indicated the

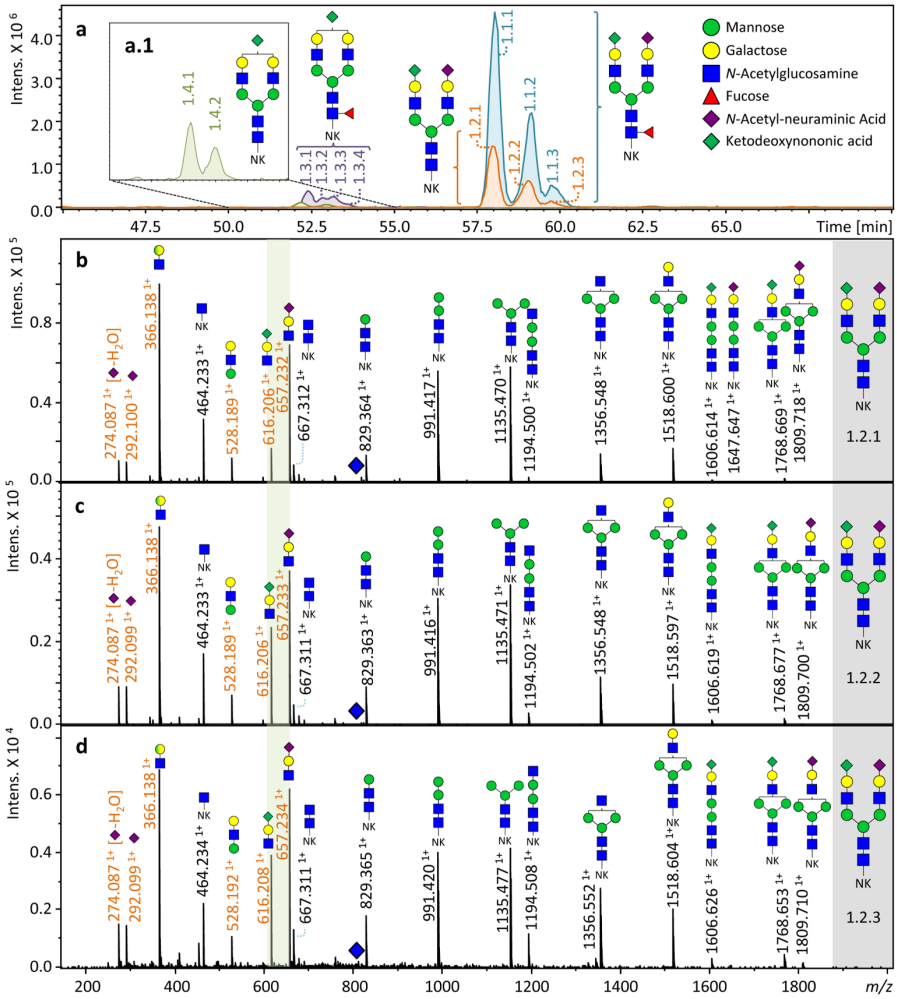

Figure 1. Identification of the Kdn motif (H1N1K1) on native tryptic PSA glycopeptides using CE-MS/MS in positive ionization mode. (a) Extracted ion electropherograms of four glycopeptides containing the H1N1K1 motif; H5N4F1S1K1[NK] (three isomers; 1.1.1, 1.1.2, 1.1.3), H5N4S1K1[NK] (three isomers; 1.2.1, 1.2.2, 1.2.3), H5N4F1K1[NK] (four isomers; 1.3.1, 1.3.2, 1.3.3, 1.3.4), H5N4K1[NK] (two isomers; 1.4.1, 1.4.2 (a.1)). Tandem mass spectra corresponding to the three isomers of H5N4S1K1[NK] are given in panels b, c and d. Oxonium (B-)ions and Y-ions are highlighted in orange and black, respectively. Blue diamond marks the precursor ion. Green panel highlights the Kdn (H1N1K1) and Neu5Ac (H1N1S1) antennae at *m/z* 616.208 and 657.242, respectively. [NK]: peptide portion of the tryptic PSA glycopeptides. H: hexose. N: *N*-acetylhexosamine. F: fucose. S: *N*-acetylneuraminic acid (Neu5Ac). K: ketodeoxynononic acid. The Kdn structure is not specifically assigned to the α1,6-antenna or the α1,3-antenna. For the fragmentation data of the other isomers see Supplementary Information, Figure S-1.

presence of solely H6N3S1[NK] (**Supplementary Information, Figures S-1j**).

Isomeric separation was observed for all four *N*-glycan compositions in CE-MS/

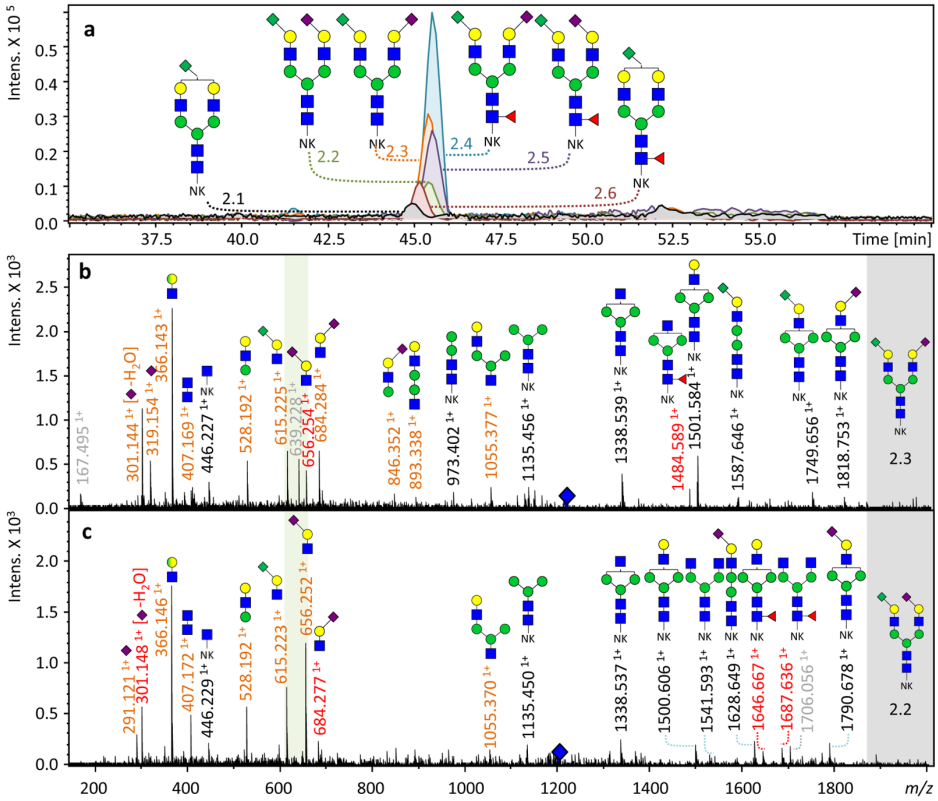

Figure 2. Identification of the Kdn motif (H1N1K1) on derivatized PSA glycopeptides with CE-MS/MS in positive ionization mode. (a) Extracted ion electropherograms of the four glycopeptides containing the H1N1K1 motif after sialic acid derivatization; H5N4K1[NK] (2.1), H5N4F1K1[NK] (2.6), H5N4F1S1K1[NK] (two isomers; 2.4 and 2.5) and H5N4S1K1[NK] (two isomers; 2.2 and 2.3). Tandem mass spectra of the three isomers of (b) H5N4A1D1[NK] (2.3) and (c) H5N4A2[NK] (2.2). Oxonium (B-)ions and Y-ions are highlighted in orange and black, respectively. Ions highlighted in grey are ions from co-isolated analytes which have a similar *m/z* value as the fragmented precursor ion. Blue diamond marks the precursor ion. The presence of the oxonium ions at *m/z* 615.224^{1+} (H1N1Ka1) and absence of *m/z* 643.256^{1+} (H1N1Kd1) indicate the occurrence of the H1N1K1 motif and that the Kdn-linkage is exclusively present in a α2,3-linkage (green panel). [NK]: peptide portion of tryptic PSA glycopeptides. H: hexose, N: *N*-acetylhexosamine. F: fucose. S: *N*-acetylneuraminic acid. K: ketodeoxynononic acid. Kd: dimethylamidated α2,6-ketodeoxynononic acid. Ka: amidated α2,3-ketodeoxynononic acid. The Kdn structure is not specifically assigned to the α1,6-antenna or the α1,3-antenna. For the fragmentation data of analytes 2.1, 2.4-2.6, see Supplementary Information, Figure S-2.

MS (**Figure 1** and **Supplementary Information, Figures S-1**). Based upon the separation characteristics of CE and its ability to separate differently linked sialic acids[156], it was assumed that most of these structures would show variance in their sialic acid linkages. Hence, to further characterize these isomers, linkage-

specific sialic acid derivatization was performed on tryptic PSA glycopeptides[168] followed by CE-MS. Similar to Neu5Ac, differently linked Kdn have identical mass values. After derivatization α2,6-linked Kdn got dimethylamine amidation (+27.0473 Da) and α2,3-linked Kdn got ammonia amidation (-0.9840 Da). The resulting mass difference enables isomer distinction. The Kdn-containing *N*-glycan compositions without the presence of Neu5Ac (H5N4Ka1 and H5N4F1Ka1; Ka: amidated α2,3-linked Kdn) resulted in a single peak per species, whilst the *N*-glycan compositions with both Kdn and Neu5Ac each showed up as two separate isomers (H5N4Ka1A1, H5N4Ka1D1, H5N4F1Ka1A1 and H5N4F1Ka1D1; A: amidated α2,3-linked Neu5Ac; D: dimethylamidated α2,6-linked Neu5Ac; **Figure 2; Supplementary Information, Figures S-2, S-3** (reaction scheme) and **Table S-3**). Interestingly, the Kdn sialylated antenna was only detected in an α2,3-linkage (Ka; amidated) configuration which was further supported by MS/MS (**Figures 2b** and **c**). Specifically, the oxonium ion at *m/z* 615.224^{1+} confirms the H1N1Ka1 antenna, while the oxonium ion that corresponds to H1N1Kd1 (Kd: dimethylated, α2,6-linked Kdn, expected mass at *m/z* 643.256^{1+}) was not detected (**Figure 2**). The fragmentation of the precursor ion at *m/z* 1144.955^{2+} further confirmed the co-elution and presence of both a complex type *N*-glycan (H5N4F1Ka1) and a hybrid type *N*-glycan (H6N3F1A1) attached to the dipeptide [NK]. Several diagnostic ions were found in the MS/MS spectrum supporting the co-existence of these glycans (e.g. *m/z* 1605.626^{1+}, 615.229^{1+}, 656.256^{1+}, 1379.483^{1+} corresponding to the Y-ion H5N2F1[NK] and B-ions H1N1Ka1, H1N1A1 and H6N2, respectively; **Supplementary Information, Figure S-2e**), which is in accordance with results of the non-derivatized samples.

To further explore the presence of isomers, and to gain more insights which isomer is corresponding to which composition and linkage, *N*-glycans were released from PSA, labeled with 2-anthranilic acid (2-AA) and analyzed by RP-nanoLC-MS (Q-ToF) in positive ionization mode with MS/MS (**Supplementary information, Figure S-4** and **Table S-3**). Released *N*-glycans were not retained on the C18 column and could not be directly analyzed by RP-LC-MS. To enhance the hydrophobicity, a 2-AA label (a well-developed and commonly used glycan label with relatively high label efficiency[295, 296]) was added to the *N*-glycans. With this approach, all four Kdn-containing *N*-glycans were identified and confirmed by the presence of the oxonium ion at *m/z* 616.208^{1+} (H1N1K1) in the MS/MS spectra. The presence of the Kdn sialylated antenna was further confirmed by the oxonium ion at *m/z* 778.261^{1+} (H2N1K1). Other distinct Y-ions, like H3N3F1K1[2AA] at *m/z* 1631.593^{1+}, H4N3F1K1[2AA] at *m/z* 1793.646^{1+} and their non-fucosylated variants were also detected. For the *N*-glycan with composition H5N4F1S1K1[2AA], the Y-ions for H5N4F1S1[2AA] at *m/z* 1100.406^{2+} and H5N4F1K1[2AA] at *m/z* 1079.893^{2+} were also observed. Furthermore, the fragmentation of the peaks *k*, *l* and *j* with precursor ion at *m/z* 1079.893^{2+} contained ions that indicated the presence of both a hybrid type *N*-glycan H6N3F1S1[2AA] as well as H5N4F1K1[2AA] (peak *k)* or solely one of them (peak *l*

and *j*, respectively) due to the observation of H1N1K1 at *m/z* 616.208^{1+}, H1N1S1 at *m/z* 657.235^{1+} and H5N2F1[2AA] at *m/z* 1502.550^{1+} (**Supplementary Figures S-4j-l**) which is consistent with the CE-MS results.

To gain further insights into arm occupancy, enzymatically released and reduced PSA *N*-glycans were measured by PGC-nanoLC-MS in negative ion mode (**Figure 3**). As the ion trap used in the PGC-nanoLC-MS platform is not as sensitive to the Q-TOF used for CE-MS, high concentration injections was performed. The same as CE-MS, three isomers were observed for the composition H5N4F1K1S1 by PGC-nanoLC-MS (**Figure 3, Supplementary Information, Figure S-5**), as well as the non-fucosylated glycan (H5N4K1S1). The occurrence of the Kdn moiety on PSA was supported by distinct ions such as the B1-ion at *m/z* 249.06^{1-} (K1), B3-ion at *m/z* 614.19^{1-} (H1N1K1) and Y6-ion at *m/z* 2037.73^{1-} (H5N4F1K1), but also the loss of Kdn (*m/z* 2078.75^{1-} (H5N4F1S1)) and the observation of Neu5Ac supports this (*m/z* 290.09^{1-} (S1) and 655.22^{1-} (H1N1S1)). Three distinct isomers were observed by PGC-nanoLC-MS/MS and the D-ions in the fragmentation spectra revealed on which arm Kdn was present for each isomer. Whereas the D-ion at *m/z* 938.30^{1-} reveals the Kdn on the α1,6-arm (isomer 2; **Figure 3c**), the other D-ion at *m/z* 979.33^{1-} shows that the α1,6-arm is occupied with Neu5Ac and the Kdn is present on the α1,3-arm (isomer 3; **Figure 3d**). As can be seen from **Figure 3b**, the first peak contains both ions with *m/z* 938.30^{1-} and *m/z* 979.33^{1-}, suggesting that it is a co-elution of two analytes that with a H5N4F1S1K1 structure that has Kdn either on the α1,3-arm or on the α1,6-arm. The ratio between the ions imply that the Kdn was mainly on the α1,3-arm and to a lesser extent on the α1,6-arm. Unfortunately, no conclusive information regarding arm occupancy could be obtained from the fragmentation spectra of H5N4S1K1, although three isomers could be observed (**Supplementary Information, Figure S-5**).

To further support our findings, high mass accuracy MS/MS was performed by RP-nanoLC-MS (Orbitrap) (**Figure 4** and **Supplementary Information, Figure S-6**). To enable the analyze of PSA glycodipeptides on RP C18 column by providing sufficient retention, a tandem mass tag named TMTzero™ label was introduced which also serves as a 'light' tag designed for MS-level peptide quantitation. Moreover, the TMT label was intended to improve ionization efficiency of the analytes. Here, distinct oxonium ions were found for the Kdn feature in the low mass range (*m/z* 251.0760^{1+} [K1], *m/z* 215.0549^{1+} [K1 – 2·H_2O] and *m/z* 616.2082^{1+} [H1N1K1]). The identified Kdn-containing glycopeptides and Kdn-containing fragment ions were all found to be within the low ppm error range (± 5 ppm; **Supplementary Information, Figure S-7**).

Based on the CE-MS result (without derivatization), the overall relative abundances of the four Kdn-containing glycans (H5N4K1, H5N4F1K1, H5N4K1S1 and H5N4F1K1S1) with isomer distinction were all below 0.80% of all glycopeptides identified in the PSA standard (derived from seminal plasma; **Supplementary Table S-4**). The glycan structure of the majority of PSA glycopeptides (non-

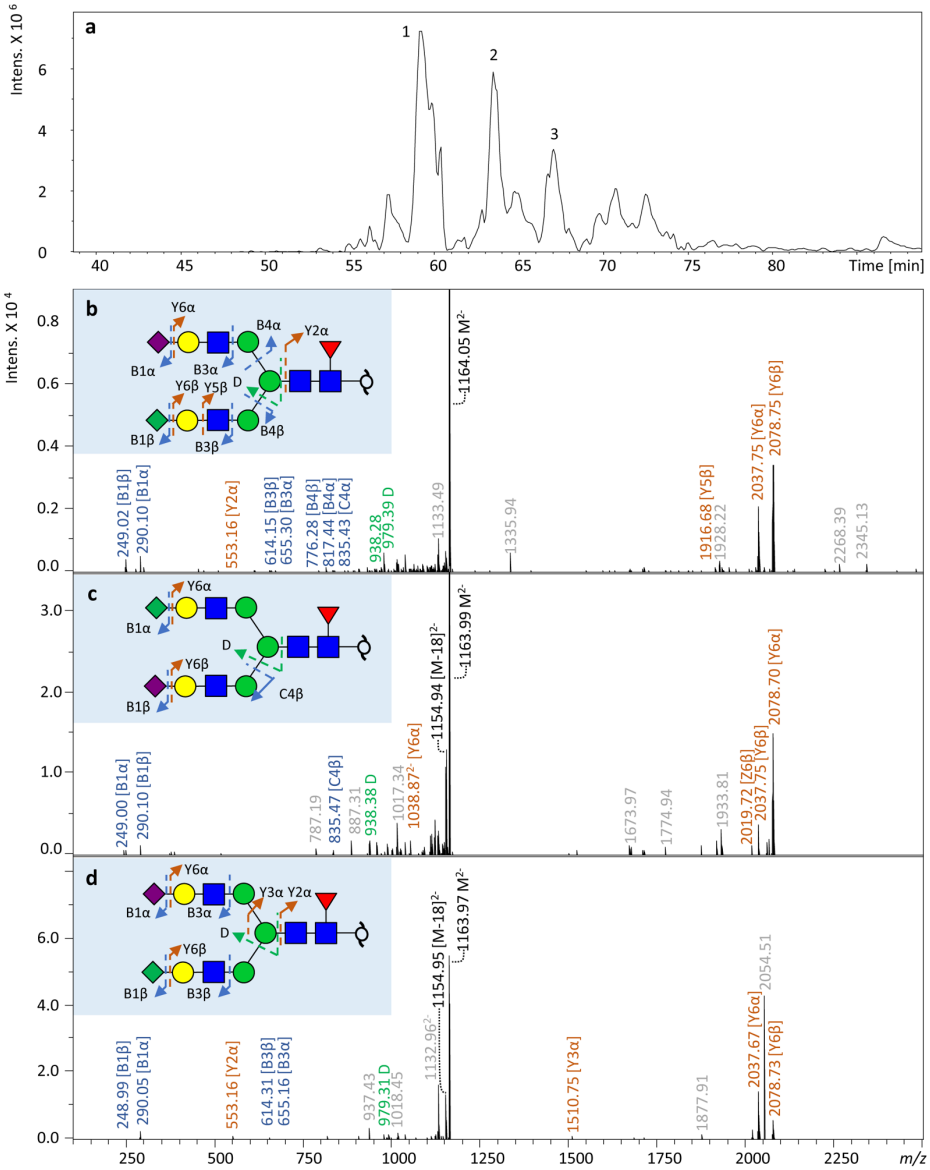

Figure 3. Identification of the Kdn motif on released *N*-glycans from PSA by PGC-nanoLC-MS/MS in negative ion mode. (a) Extracted ion chromatogram of released *N*-glycans from PSA containing the Kdn motif. Fragmentation spectra of the released *N*-glycan at mass 1163.91^{2-} indicate that this belongs to the released *N*-glycan with composition (b, c and d) H5N4F1K1S1. B and C-ions are highlighted in blue. Y-ions are indicated in red. D ions in green indicating arm occupancy. M indicating the precursor ion. H: hexose. N: *N*-acetylhexosamine. F: fucose. S: *N*-acetylneuraminic acid (Neu5Ac). K: ketodeoxynononic acid. Note: In the spectrum of (b) isomer one, the ions with *m/z* 979.33 and 938.30 were both observed indicating that this isomer is a mixture (co-elution) of H5N4F1K1S1 with Kdn either on the α1,3 or on α1,6 arm. (c) indicated the glycan structure with Kdn on the α1,6 arm and (d) on the α1,3 arm.

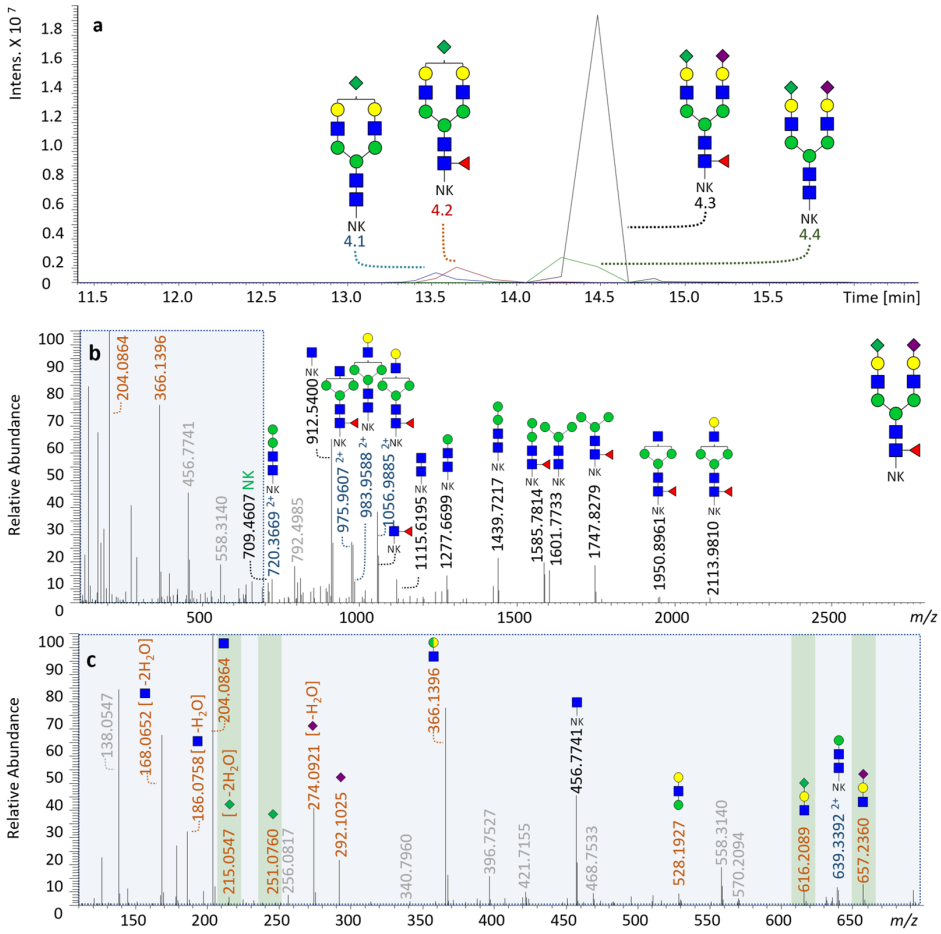

Figure 4. Identification of the Kdn motif (H1N1K1) on TMT-labeled PSA glycopeptides by RP-LC-MS/MS (Orbitrap) in positive ionization mode. (a) Extracted ion chromatograms of the four glycopeptides containing the H1N1K1 motif; H5N4K1[NK] (4.1), H5N4F1K1[NK] (4.2), H5N4F1S1K1[NK] (4.3) and H5N4S1K1[NK] (4.4). Full range (b) and magnified (c) tandem mass spectra of the glycopeptide H5N4F1S1K1 with precursor mass at *m/z* 1007.0930^{3+}. The fragments of the full fragmentation spectrum and expanded fragmentation spectrum in the low mass range indicate that this precursor *m/z* belongs to the glycopeptide with composition H5N4F1K1S1[NK-TMT]. Oxonium (B-) ions are highlighted in orange. Y-ions are highlighted in black and blue being either singly or doubly charged, respectively. Unassigned ions in grey are unknown. [NK]: peptide portion of tryptic PSA glycopeptides. H: hexose, N: *N*-acetylhexosamine. F: fucose. S: *N*-acetylneuraminic acid. K: ketodeoxynononic acid. The Kdn motif is not specifically assigned to the α1,6-antenna or the α1,3-antenna. For the fragmentation data of the mono-sialylated analytes see Supplementary Information, Figure S-6.

Kdn contained) were identified by CE-MS/MS in a previous study[177]. Among all identified Kdn-containing glycans, H5N4F1K$_{2,3}$1S$_{2,6}$1[NK] (isomer 1) revealed the highest relative abundance of 0.78%.

4.3.2. Expression of Kdn in a Urinary PSA Cohort

To investigate whether the Kdn motif could also be detected on PSA derived from patients' urine, our previous clinical cohort[286] was reanalyzed with additional patients being included (total of 52 individuals with elevated serum PSA concentrations >3 ng/mL). The repeatability of the applied urinary PSA assay was validated by determining intraday and interday precision over three different days with three replicates per day and a system suitability sample was measured prior and after each measurement batch to monitor the performance of the CE-MS. Male urine was considered unsuitable to use during the development of PSA capturing procedure as it contains PSA. In contrast, FUP contains no PSA and the matrix is the closest to that of male urine samples, hence it was chosen to use for protocol development. By spiking a PSA standard to the FUP we were able to validate the performance of the overall assay. The data of nine individuals were excluded as more than one third of all PSA glycopeptides were not observed due to low signal intensity. The remaining data of 43 patients' urine were included in the dataset (21 PCa and 22 non-PCa, for clinical information see **Supplementary Information, Table S-1**). This resulted in the overall identification of 71 PSA *N*-glycopeptides including the distinction of sialic acid isomers. The data was visualized by principle component analysis (PCA) and the clustering of the positive controls (female urine pool spiked with PSA standard from seminal plasma) from both batches indicate that no substantial differences were found between the batches, even though two different time points were used for the processing as well as the difference in storage and freeze and thaw cycles **(Supplementary Information, Figure S-8**). The clustering of the pooled patient sample (including 0.5 µL of all captured and digested samples of batch 1 and 2) within the individual patient samples illustrates the low technical variability of the CE-MS measurements. In agreement with the PSA derived from seminal plasma, Kdn sialylation also appeared to be overall low abundant in the clinical cohort (**Supplementary Table S-1, Supplementary Figure S-8**). In PCa samples, a relative abundance of 4.15% ± 1.58% was found for H5N4K1[NK], 0.15% ± 0.04% for H5N4F1K1[NK], 0.27% ± 0.15% for H5N4K1S1[NK] and 0.95% ± 0.39% for H5N4F1K1S1[NK]. In non-PCa samples, a relative abundance of 4.91% ± 1.48% was found for H5N4K1[NK], 0.15% ± 0.04% for H5N4F1K1[NK], 0.29% ± 0.14% for H5N4K1S1[NK] and 0.96% ± 0.24% for H5N4F1K1S1[NK]. A large variation in Kdn sialylation was observed with an average relative standard deviation (RSD) of 41.20% (PCa patients), 32.44% (non-PCa patients) and 36.70% (all patients) implying a large biological variation. Interestingly, tryptic PSA glycopeptides with the theoretical mass of 2133.806 Da (corresponding to the *N*-glycopeptide with composition H5N4K1[NK] as well as that of H6N3S1[NK]) was found to be 100 times higher in relative abundance in the cohort samples derived from urine than PSA standard derived from seminal plasma. The fragmented precursor mass, belonging to both H5N4K1[NK] and H6N3S1[NK], revealed the presence of both analytes in the sample (**Supplementary Figure S-9**). However, the fragments

derived from H6N3S1[NK] appear to be more abundant compared to those from H5N4K1[NK]. At last, no significant differences were found for the Kdn-features in urine between different disease conditions (PCa *vs* non-PCa).

4.4 Discussion

4.4.1. Expression of Kdn on mammalian PSA

Previously, we described the presence of analyte H6N3(F1)S2 in PSA derived from seminal plasma[168] as well as derived from urine of individuals with elevated PSA serum concentrations (>3 ng/mL)[286]. However, we were unable to elucidate the structure of these *N*-glycans, and it remained unclear whether these *N*-glycans were complex or hybrid types and where the postulated second sialic acid would be positioned. To figure this out, we studied the PSA standard derived from seminal plasma with tandem mass spectrometry and the unidentified *N*-glycans were detected in the PSA standard. By combining the evidence obtained with four different analytical approaches (**Table 1**), we were able to assign these *N*-glycans to the occurrence of Kdn sialylation. It should be noted that N1K1 and H1S1 compositions are isomeric and we could re-assign the previously assigned H6N3(F1)S1 and H6N3(F1)S2 *N*-glycans to four Kdn-containing *N*-glycans (H5N4K1, H5N4F1K1, H5N4K1S1 and H5N4F1K1S1). Interestingly, the distinct oxonium ion of H1N1K1 was observed with all four complementary platforms, while the distinct oxonium ion of K1 was observed only with PGC-nanoLC-MS and RP-nanoLC-MS (Orbitrap). This might be related to a difference in linkage stability between the Kdn and Neu5Ac monosaccharides which are attached to a galactose and warrants further exploration.

Isomeric separation was observed for all four Kdn-containing *N*-glycan compositions in CE-MS (**Figure 1** and **Supplementary Information, Figures S-1, Figure S-10**). However, some differences were observed, for example, three isomers were observed for the composition H5N4F1K1S1 by CE-MS (**Figure 1, Supplementary Information, Figure S-1**), whereas four isomers were observed with RP-nanoLC-MS (Q-ToF, **Supplementary Information, Figure S-4a**) and the same observation was found for the mass corresponding to the non-fucosylated *N*-glycan (H5N4K1S1). To obtain information whether the isomers were related to different sialic acid linkages on Kdn, sialic acid linkage-specific derivatized tryptic PSA glycopeptides were analyzed. These experiments suggest that the Neu5Ac sialylated antenna varied in linkage, while the Kdn sialylated antenna was only observed in α2,3-configuration (**Figure 2**). This is in contrast to a previous study which reported the accumulation of complex free *N*-glycans in human PCa tissues, with Kdn linked both α2,3- and α2,6 to galactose[122]. Moreover, the authors stated that the α2,6-Kdn *N*-glycans were the dominant form in these samples and that the accumulation was most probably related to insufficient degradation. It should be noted that α2,6-Kdn *N*-glycans could still be present on PSA but was below the limit of detection of our analytical methods. In addition,

Kdn antenna occupancy was resolved by PGC-nanoLC-MS analysis, revealing it as the main determinant of isomer separation of Kdn-containing glycans.

4.4.2. Expression of Kdn in a Urinary PSA Cohort

Glycosylation reflects human health and disease. Altered glycosylation is a feature of several diseases, although it is not clear whether the altered glycosylation is caused by or resulting in diseases[1]. The role of glycosylation in medicine is expanding with increasing knowledge on glycan structures and functions as many different technologies have been developed for the analysis of glycans and glycoconjugates. For example, MS techniques enable identification and characterization of glycans, whilst glycoengineering and chemoenzymatic techniques for the synthesis of glycans enable optimized therapeutics and development of new glycomedicines (preventive (vaccines), predictive and personalized glycomedicine)[3, 297, 298]. Aberrant glycosylation has shown its potential as biomarker for better targeting of disease diagnostics and prognosis and the possibility to serve as target for therapies[299]. Many glycan-based therapeutics are currently available or under development using glycans as therapeutic drugs or the carrier of therapeutic drugs[1, 69, 300, 301].

In this study, the Kdn containing *N*-glycans were first discovered on PSA derived from seminal plasma. As demonstrated previously, elevated levels of Kdn (free monosaccharide or conjugated) in human cancer cells as well as tissues implicate that the sugar is associated to cancer development [1]. Thus, Kdn modifications have biomarker potential and may eventually serve as a target for cancer therapy. Therefore, it is interesting to explore Kdn modifications on PSA derived from different biofluids, including urine and serum. In this study, we explored the urinary glycome as analyzing the glycomic profile of serum PSA remains rather challenging: 1) its concentrations are at least 100 times lower compared to urine[302, 303] requiring a highly sensitive analytical platform, 2) plasma is a highly complex and protein-rich matrix making purification or enrichment of PSA challenging, 3) PSA is relatively low abundant in serum/plasma compared to other proteins which increases the risk of aspecific binding and complicating the analysis. In addition, several studies have already shown the biomarker potential of altered urinary glycomic profiles[186, 304, 305] and, the collection of urine is less invasive for patients. Although no significant differences were found in urine between different disease conditions (PCa *vs* non-PCa), considering only a rather small cohort was explored and the diversity of the urine PSA samples (e.g. high variety in PSA concentrations, a relative low number of patients per Gleason score, different clinical stage of tumor and so on), a larger cohort would be recommended before strong conclusions can be drawn about the diagnostic value of this uncommon monosaccharide. In addition, as the PSA derived from seminal plasma (positive control) is obtained from healthy individuals, it would be interesting to investigate whether the urinary PSA *N*-glycome of healthy men also show a higher abundancy of the *N*-glycan H6N3S1[NK]/H5N4K1[NK] and

whether this is a feature related to individuals with elevated PSA concentrations. Moreover, based upon our findings, the *N*-glycan assignments of previous glycomic studies on PSA, which identified the presence of H6N3(F1)S1[168, 186], should be reevaluated taking into account that PSA could also contain *N*-glycans occupied with a Kdn motif and that the previously identified *N*-glycan might contain H5N4(F1)K1 or that both species are present. It should be noted that this may also apply to other, non-PSA related, *N*-glycomic studies and that these datasets require reevaluation as well as to investigate the presence of Kdn-containing *N*-glycans. In more general terms, Kdn may be way more common than currently thought, and considering Kdn in compositional assignment of mammalian N-glycans is warranted.

MS based methods possess great potential for human diseases diagnosis[306] allowing identification and quantification of active compounds in many clinical areas, such as newborn screening, therapeutic drug monitoring, drugs of abuse and clinical toxicology[307-309]. The development of advanced MS technologies has pushed clinical MS toward the analysis of peptides and proteins for diagnostic appliations, although the quantitative analysis of especially glycopeptides by MS is still challenging. A targeted approach is preferred for MS based diagnostic examination. Recently a number of MS-based in vitro diagnostics have been adopted in routine clinical practices, and more can be expected to undergo transition from lab to clinic in the near future[310, 311]. Its remarkably high sensitivity, specificity and capability to test for multiple diseases in a single measurement make this technology highly powerful in disease detection and therapy monitoring. In case of promising biomarkers discovered in free Kdn or free-/conjugated Kdn containing glycans for PCa, cohort studies and clinical trials of the MS-based diagnostic methods should be undertaken for its transition from discovery to standard clinical practices. Elevated levels of free and conjugated Kdn in human has been linked to the occurrence of antibody-mediated inflammation as antibodies against Kdn-glycans were detected in pooled human immunoglobulins[116]. Thus, pathological conditions that elevate Kdn levels could result in antibody-mediated inflammatory pathologies[116]. Kawanishi *et al.* has shown that the ingestion of mannoses leading to an increased excretion of free Kdn in human urine, they proposed the hypothesis that the biosynthesis of free Kdn is a critical metabolic pathway that buffers the potentially toxic concentration of cytosolic mannose 6-phosphate and fructose 6-phosphate under conditions of excess mannose in mammals [116]. A mechanism of elevated free Kdn expression in human cancers cells has been proposed by Go *et al*.[123]: Briefly, the common feature of advanced localized tumor-hypoxia induce the activities of crucial enzymes (Neu5Ac-9-P synthase and phosphomannoisomerase (PMI)) and the expression of mRNA for corresponding enzymes in Kdn biosynthetic pathway[123]. In addition, hypoxia enhances the mannose induced increase of free Kdn expression[123]. Literature has shown that the enzymes Neu5Ac-9-P synthase and CMP-sialic acid synthase, both involved in the biosynthesis of Neu5Ac- and

Neu5Gc-containing glycans in mammals, also catalyze the synthesis of Kdn and Kdn-containing glycans, but with much lower efficiency[119, 312]. Besides, a study on human cell lines showed that gene mutation (PMI-KO) and zebrafish *Cmas* gene transfection can force Kdn incorporation on cell membrane glycans thus enhancing Kdn utilization which implies the limited utility of Kdn in human glycosylation under natural conditions[116]. However, further investigations are needed to fully understand how Kdn is incorporated in the human glycome. Moreover, future endeavors should also focus on the presence of Kdn sialylation in other biofluids, and in other human glycoproteins. Finally, studies have shown elevated levels of free Kdn and Kdn sialylation in several human cancers and end stage renal disease. It is still unclear what role Kdn plays, therefore it should be addressed how Kdn is associated with those human diseases. And in relation to human diseases the potential of Kdn, free Kdn or free-/conjugated Kdn-containing glycans, as a valuable glycan biomarker or a target for cancer therapy should be exploited.

Conclusion

In this study, we provided evidence that Kdn sialylation occurs on *N*-linked glycans of PSA derived from different biofluids (urine as well as seminal plasma). Moreover, it is very likely that this monosaccharide is not solely present on PSA, but also on other mammalian glycoproteins and might have been misinterpreted in previous mass spectrometric studies as the N1K1 and H1S1 compositions share the same mass. Further studies are recommended to explore the distribution of Kdn in human glycoproteins and tissues, as well as the biosynthesis and biological relevance of Kdn glycoconjugates.

Supporting Information

The Supporting Information is available via https://www.engineering.org.cn/en/10.1016/j.eng.2023.02.009.

Data availability

The raw mass spectrometric data files that support the findings of this study and the clinical cohort data files are available in MassIVE in .mzXML format, with the identifier MSV000089140 [https:// doi:10.25345/C5BG2HF07]. All other data supporting the findings of this study are available from the corresponding author on request.

Acknowledgments

The authors would like to thank Arnoud de Ru, George M.C. Janssen, Carolien A.M. Koeleman and Irina Dragan for their contributions to this study. This research was supported by the Chinese Scholarship Council, Cure for Cancer Foundation and Astellas Pharma B.V.