Glycoproteomics assays for prostate cancer biomarker discovery

Wang, W.

Chapter 6

Discussion and Future Perspectives

The work described in this thesis specifically focused on assay development for the analysis of the glycosylation features present on urinary prostatic acid phosphatase (PAP), seminal plasma prostate-specific antigen (PSA) and blood plasma PSA. The research was performed on the glycopeptide level with the aim to discover biomarkers for the diagnosis of male infertility and prostate cancer (PCa). In the following sections, the challenges, technological limitations, and opportunities that were encountered during the research will be addressed. A special focus will be on the future clinical applicability of the developed methods, by critically discussing the use of different biofluids as a valuable source for biomarkers in specific contexts, including male infertility and PCa.

6.1 MS workflows in prostate-focused glycoproteomics

6.1.1 The choice of analytical workflow

In this thesis a variety of different analytical techniques were employed, with each of them having their own advantages and disadvantages in terms of complexity of sample preparation, sample throughput as well as assay sensitivity (**Table 1**).

6.1.1.1 CE-MS[1]

In **Chapter 2**, to study the glycoprofile of PAP at low concentration (nanograms per mL) in urine, capillary electrophoresis–mass spectrometry (CE-MS) was selected as analytical platform due to its ability to perform in-depth glycomic analysis on a glycopeptide level as demonstrated by several other studies[168, 333-338]. Here, CE provides excellent separation of different glycan species on the same peptide backbone[156], but -in combination with MS- also high sensitivity (detection of glycans/glycopeptides from picograms to nanograms of protein), especially when a dopant enriched nitrogen (DEN) gas is employed at the interface between the CE and the MS[290, 339, 340]. The advantage of the here presented workflow is demonstrated as follows: analysis by CE-MS does not require an additional clean-up step compared to other platforms (e.g. liquid chromatography (LC)–MS) and the sample can directly be analyzed after immunoaffinity capture and proteolytic cleavage, enabling the analysis of peptides and glycopeptides in a single analysis[156]. Therefore, this workflow is simplest with respect to sample preparation compared to other workflows presented in this thesis. The separation of glycopeptides in CE is influenced by both the peptide backbone and the glycan moiety. In the presence of neutral glycans, no additional charge is added by the glycan and the separation will be mainly driven by charge of the peptide backbone and the hydrodynamic volume. In normal polarity (cathode at the MS side of the capillary), the net charge should be positive to enable the

1 Part of this section was adapted with permission from: CE-MS approaches for Glyco(proteo)mic Analysis, W. Wang, W. Wang, G.S.M. Lageveen-Kammeijer, *Capillary Electrophoresis-Mass Spectrometry for Proteomics and Metabolomics,* 2022, p.p. 35-375 https://doi.org/10.1002/9783527833092.ch13.

Table 1. Overview of established MS workflows in prostate-focused glycoproteomics.

Chapter	Workflow	Platform	Sample processing	Throughput	Sensitivity	Isomer differentiation	Future investigation
# 2	Capture, digestion	CE-MS[a]	Facile	Low	Medium	Achieved via CE separation	Inline sample trap capillary, multi-segment injection, sample multiplexing with e.g. TMT[b] labeling
# 3	Capture, digestion, DA[c]	MALDI-MS[d]	Medium	High	Low	Achieved by DA reaction	Automation
# 5	Capture, digestion, TMT labeling, DA	RP-LC-MS[e]	Complex	Medium	High	Achieved by DA reaction	Automation; DEN[f] gas incorporation; CE coupling

[a] *Capillary electrophoresis–mass spectrometry*

[b] *Tandem mass tags*

[c] *Dimethylamidation with amidation modification on sialic acids*

[d] *Matrix-assisted laser desorption/ionization–mass spectrometry*

[e] *Reverse phase-liquid chromatography–mass spectrometry*

[f] *Dopant enriched nitrogen*

observation of the analytes. The presence of basic amino acids will provide a faster migration while the presence of acidic amino acids will slow down the migration. The same accounts for acidic monosaccharides or the addition of other post-translational modifications (PTMs), such as sulfation or phosphorylation, as these PTMs add a negative charge to the glycopeptide thus slowing down the migration. Distinct clusters were observed between the glycopeptides with a neutral glycan moiety, mono-, di- and tri-sialylated glycan moiety as well as for the sulfated non- and mono-sialylated glycan moieties (**Figure 1**)[341]. It is important to note that in **Chapter 2** the (in some cases baseline) separation of α2,3- and α2,6-sialylated isomers was achieved without any additional sample treatment after the proteolytic digestion. One study by Lageveen-Kammeijer *et al.* indicated that the separation might be driven by a slight difference in pK_a value (Δ $3.4 \cdot 10^{-2}$), but further research is needed to investigate whether the isomers differ in hydrodynamic volume and whether this also contributes to its separation[156].

While, CE-MS shows excellent separation and sensitivity, the sample loadability is a limiting factor. In general, 1 - 2% of the capillary volume is recommended for sample loading to avoid loss of resolution caused by peak broadening. In **Chapter 2**, an online pre-concentration technique is implemented, also known as *transient* isotachophoresis (*t*-ITP), and can increase the loadability to up

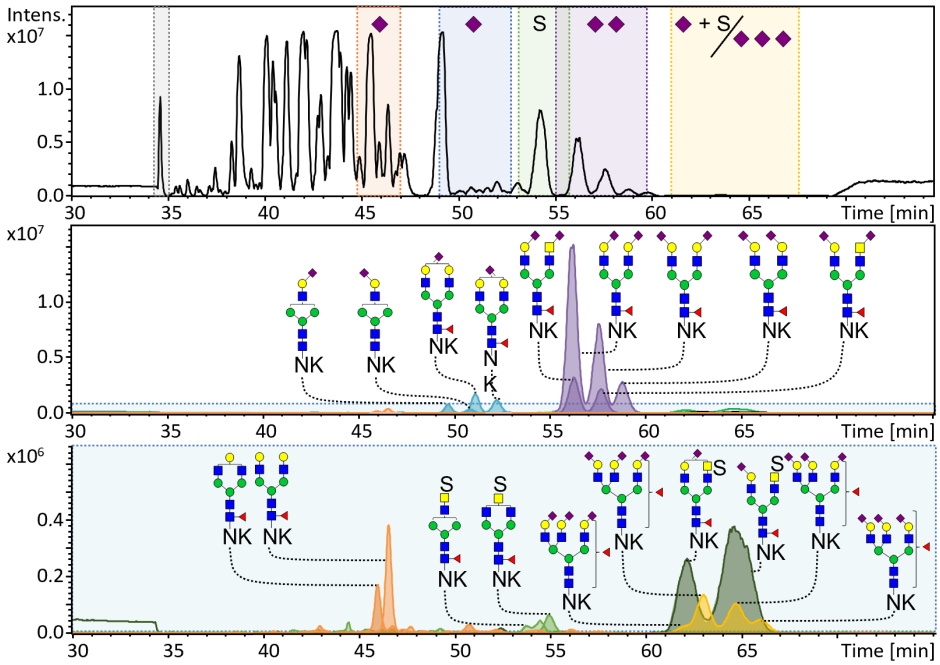

Figure 1. The migration pattern and separation of tryptic PSA glycopeptides with CE-MS using a neutral coated capillary. (A) Representative basepeak electropherogram of digested PSA standard using porcine trypsin. Distinct clusters can be observed for non-, mono-, di- and tri-sialylated glycopeptides highlighted in orange, blue, purple and yellow, respectively. In addition the green and yellow color highlight the sulfated non- and mono-sialylated glycopeptides, respectively. The grey box indicates the migration time of sodium acetate clusters which can be used for external calibration after analysis. (B) Extracted ion electropherograms (EIEs) of the two most abundant mono- and di-sialylated glycopeptides present in the PSA standard. (C) EIEs of the two most abundant non-sialylated as well as the most abundant tri-sialylated and sulfated non- and mono-sialylated glycopeptides. As illustrated by this example, sialylation and sulfation are the main contributors on the migration and separation behavior of the glycopeptides, where sulfation is the biggest influence. In addition, differently linked sialic acids are separate where the α2,6-linked sialylated species migrate prior to the α2,3-linked sialylated species. NK: peptide moiety of PSA glycopeptides, S: sulfation.

to 30 - 50% of the capillary volume while maintaining proper resolution[342, 343]. However, the loadability of CE (nLs) is still insignificant compared with nanoLC (µLs). To further increase the CE capillary loadability, the use of an inline sample trap capillary can be explored which has not been tested for glycopeptide analysis[344-346]. Moreover, different CE interfaces and different types of capillary was used in the case of inline sample trap capillary other than the sheathless interface and CE cartridge used in this thesis. To implement the inline sample trap column, adaptations in both capillary and cartridge configurations would be necessary, along with the need to address potential technical challenges, such as determining the optimal placement and integration of the inline sample trap

column on the CE cartridge and ensuring compatibility with the CESI 8000 Plus system.

To obtain optimal isomer separation of glycopeptides, most CE-MS measurements require relatively long analysis time (*e.g.,* up to 80 min), which could be seen as another constraining factor, especially when large sample sets need to be analyzed. For the workflow presented in **Chapter 2**, approximately 18 samples could be processed in one batch (approximately one and a half days). To increase sample throughput of the assay, the usage of multisegment injection could be investigated, as the added value of this procedure has previously been demonstrated for metabolomics[347], bottom-up proteomics[348] as well as for *N*-glycan analysis[349]. This techniques allows to significantly diminish the analysis time as samples are injected in series, as illustrated by the study of Kovács and co-workers, where 96 IgG *N*-glycan samples were analyzed in a single run within a duration of 4 h using CE-laser-induced fluorescence detection (LIF)[349]. While such an approach may significantly increase throughput, the method might compromises on isomer separation due to short separation time window (1.14 min). Further studies are needed to investigate its potential for glycopeptide research. Another possibility would be the usage of isobaric tags for relative and absolute quantification (iTRAQ) and tandem mass tags (TMT) that have already proven their values in glycoproteomic research[350, 351]. TMT was investigated in this thesis and will be further discussed in section *6.1.1.3 LC-MS*.

6.1.1.2 MALDI-MS

As demonstrated in **Chapter 3**, matrix-assisted laser desorption/ionization (MALDI)-MS is an attractive option for glycopeptide profiling offering an easily accessible and high-throughput platform. The established MALDI-Fourier transform ion cyclotron resonance (FTICR)-MS workflow has shown its capability to characterize the *N*-glycan heterogeneity of purified PSA at the glycopeptide level. Here, the biggest advantage of the established MALDI-MS workflow is its high sample throughput enabling the preparation and analysis of 96 samples in two and a half days. Measurement of full clinical cohorts can be done in a short time period which minimizes possible batch effects. In MALDI-MS analysis, glycans and glycopeptides are singly charged which can immensely reduce the spectra complexity compared with CE/LC-MS analysis (via electrospray ionization; ESI) in which glycans and glycopeptides are detected in multiple charge stages. In general, MALDI-FTICR-MS workflows exhibited high repeatability, allowing to pick up biological variations between individuals and, thus suitable to apply to biomarker discovery studies. Besides, the developed workflow is also compatible with MALDI-time-of-flight(TOF)-MS (despite with lower mass resolution and potential more ambiguous glycoform assignments) making the method accessible to a broader range of laboratories.

As presented in this thesis, PSA glycoprofiles show high level of sialylation

which is an important glycosylation feature in cancer studies. To overcome the hampered analysis of sialylated glycoconjugates (by instability of sialic acid (SA) residues, ionization bias and salt adduct formation[352, 353]) as well as to distinguish SA linkage isomers, extra sample processing is required (e.g. linkage-specific SA derivatization) prior to MALDI-MS analysis[132-140]. Moreover, glycopeptide enrichment and purification is important as MALDI-MS requires clean samples to form proper sample-matrix crystals, and to avoid the suppression of glycopeptide ionization by strongly ionizing coexisting peptides, the presence of nonvolatile salts as well as detergents. Successful implementation of a high-throughput glycomics protocol on an automated liquid-handling robot system has been reported[354] that provides an opportunity to further improve the MALDI-MS workflow of **Chapter 3** regarding both throughput and reproducibility.

6.1.1.3 LC-MS

While first steps were made in establishing a high-performance PSA Glycomics Assay (PGA) using urine analysis by CE-MS[177], no biomarkers for PCa assessment were found in urine. Nonetheless, studies have clearly demonstrated that PCa patients show an altered glycosylation profile of PSA, which is derived from the circulation[56, 58, 60]. Therefore, we intended to further expand the established PGA to plasma and investigate its biomarker potential for the early diagnosis of PCa as well as for the differentiation between aggressive and non-aggressive PCa, especially for the individuals with PSA concentrations that fall in the so-called grey zone (3-10 ng/mL). With such low PSA concentrations, an ultrahigh sensitivity platform is needed that will allow the analysis of PSA glycosylation of a few mLs of plasma. For this purpose, a RP-nanoLC-MS approach was explored in **Chapter 5**, in which a workflow was developed to analyze the glycosylation of blood plasma PSA and in future, the method can be exploited to explore potential PCa biomarkers for early diagnosis and prognosis. Although many challenges were encountered, the developed assay demonstrates a high sensitivity of analyzing and quantifying the top 20 PSA glycopeptides using 4 mL plasma with PSA concentration of 6 ng/mL.

To allow multiplexing and provide retention of tryptic PSA glycopeptides (Asn-Lys) on the C18 column, TMT label is implemented in **Chapter 5.** More importantly, TMT labels allow absolute quantification based on MS/MS reporter ions. To use as a standard, a reference sample (in this case PSA standard[2]-spiked female plasma) can be introduced and processed along with the real samples during multiplexing. Differentially labeled (glyco)peptides have identical masses that are indistinguishable in the MS spectrum. Via tandem mass analysis, the labile mass reporters with different masses will be fragmented to generate reporter ions. Intensities of reporter ions are used for quantification of the glycoforms. Here,

2 The PSA standard employed in this research originates from seminal plasma. Notably, the glycoprofile of seminal PSA differs from that of plasma PSA. Highly branched complex glycans that reported in plasma PSA are largely absent in seminal plasma PSA.

absolute quantification is defined as the accurate determination of the quantity of PSA captured from real plasma samples. While RP-nanoLC-MS is highly sensitive and able to separate glycan isomers to a certain extent, the linkage-specific SA derivatization developed in **Chapter 3** was incorporated for obtaining facile isomer discrimination in MS, leading to an additional sample cleanup step. Due to the various sample preparation steps, the developed workflow is the most tedious and time-consuming among all methods developed in this thesis. Especially the multiple sample processing steps are challenging in terms of sample recovery, assay sensitivity and robustness, as it is well-known that more steps will result in more sample loss as well as more human intervention introduces more variations.

To reach the desired sensitivity (3 ng/mL), a thorough examination was performed of the complete multiple-step protocol. Details were considered, such as, the type of sample vials being used, sample storage conditions, the most suitable protease, reaction buffers and chemicals. While a sensitivity of analyzing PSA captured from 4 mL plasma with PSA concentration of 6 ng/mL was reached with the current set-up - the highest of all methods described – improvements are envisioned by coupling the CE (essential for high sensitivity and isomer separation) with Orbitrap (essential for TMT and sensitivity) using DEN gas, as the implementation of DEN gas showed a 2-fold enhancement in absolute peak area for PSA glycopeptides[290]. Moreover, in this case, SA derivatization step can be eliminated. To further improve the assay repeatability, automation of the sample processing procedure would be desirable to avoid the variation associated with human intervention.

6.1.2 Absolute quantification

Most glycomic biomarker studies focus on the difference in the relative abundance of various glycans and glycosylation features. However, it is the absolute, and not the relative, quantification that can provide a complete picture of the disease related changes in glycosylation which is important for understanding of the disease occurrence, development and potentially future treatment[141]. Instead of providing only the percentage of the individual glycans normalized to total glycan content by relative quantification, absolute quantification determines the exact concentrations for individual glycans in the sample. Another advantage of absolute quantification is that the absolute amounts of glycans provide information on the number of glycan epitopes available for interaction with targeting antibodies in future glycotherapy[141]. The ideal absolute quantification -knowing the exact amount of each glycan in the original sample in mg/mL- is challenging but can still be achieved by using heavy isotope-labeled internal standards. The biggest challenge rest in the demand of a large variety of standards due to the micro and marco heterogeneity of glycosylation in one glycoprotein. Ideally, heavy isotope-labeled internal standards can cover different glycoforms detected in the glycoprotein of interest.

In MS, relative quantification is compromised by the variation in ionization efficiencies of different glycan structures. The established assay in **Chapter 5** is a standardized approach, where the changes of glycan expression can be assessed per independent glycan from the complete glycoprofile by referring to the reference TMT channel (PSA standard-spiked female plasma), that can be used to perform absolute quantification of the amount of captured PSA. It should be noted that plasma PSA contains highly branched *N*-glycans that may not present in the PSA standard (derived from seminal plasma) used in the reference TMT channel which cannot be absolutely quantified. To control for capturing efficiency and achieve absolute quantification of the PSA amount in patient plasma samples, it is imperative to develop a stable isotope-labeled (SiL) PSA glycan standard which can be designed in similar way as SILuMAB (Stable-Isotope Labeled Universal Monoclonal Antibody Standard). It is widely utilized for quantitation of monoclonal antibodies and is commercially available. The SiL-PSA standard should contain the glycoforms of interest and the relative proportions of each glycoform within it must be precisely determined. The SiL-PSA standard can be spiked into female plasma, serving as a reference or control, and be processed along with patient plasma samples.

6.1.3 Data analysis

In this thesis, manual data processing was often performed with Compass DataAnalysis (Bruker). The analysis can be tedious and time-consuming, especially in the case of a large sample cohort. Besides, the cohort samples need to be processed in different batches due to the limited sample throughput (**Chapter 2**). Therefore, batch effects should be taken into account for further data analysis. Skyline is a free access MS data analysis software created by the MacCoss Lab[355] which is another option for processing CE- or LC-MS data and can partially reduce the laborious work. Namely, Skyline can automatically check through the complete spectra and find the best-fitting peak for the target analyte based on mass accuracy and isotopic patterns when an analyte list is provided (including analyte name, peptide composition, atom formula and mass values). Moreover, it is suitable for tandem MS analysis and, due to this, allows for TMT reporter ions-based quantification (**Chapter 5**). However, it should be noted that manual peak checking and curation of the data is still important and necessary for reliable assignments of the analytes. Other advanced software and data analysis tools as well as open databases are available to facilitate the data interpretation and some of these have been well summarized and evaluated by Lippold *et al.* [19]. However, they lack proper evaluation for their usage on CE-MS data. For example, LC shows higher repeatability than CE in regards to analyte elution or migration time. It should be investigated how these tools handle large migration shifts from CE while isomer separation is maintained. Besides, the large migration shifts can be corrected by applying specific tools and proper standards[356, 357].

6.1.4 Structural assignment of glycans

To assign the structure of observed glycans and glycopeptides, multiple analytical techniques can be performed, such as exoglycosidase treatment in combination with UV/fluorescence HPLC/CE detection[358-361], binding studies with fluorescently labeled lectins/antibodies[362, 363] and the use of MS[364-366]. The latter is the preferred option in the recent years, as it allows for in-depth characterization of the target glycome/glycoproteome[252, 253]. In **Chapter 2**, **Chapter 3** and **Chapter 5**, three different glycopeptide approaches were developed with complementary properties. These kind of complementary methods are needed, as exemplified by a separate set of techniques applied to elucidate the newly identified structure Ketodeoxynononic acid (Kdn) on PSA as show in **Chapter 4**. Here, four complementary MS-based analytical approaches were performed for the novel identification of Kdn on PSA *N*-glycans. The difficulty of identifying Kdn rests in the identical mass of compositions H2S1 and H1N1K1[3]. As demonstrated in **Chapter 4**, using positive or negative ion mode can result in different fragmentation of glycans; fragmentation in positive mode using HCD (higher energy collisional dissociation) generates high ion intensities, especially the signature oxonium ions at *m/z* 366.14^{1+} (H1N1), at *m/z* 657.23^{1+} (H1N1S1), etc., allowing sensitive glycan detection, while negative fragmentation mode yields more cross-ring cleavages of glycans that provides unique linkage and branching information[367].

MS is a core technology in the field of glycopeptide-based glycosylation profiling. However, identification and complete glycan structure elucidation of isomeric glycoforms remains a challenging task. Arm occupancy is often not resolved, especially when the analyte is low in abundance. An emerging and powerful alternative for distinction of structural isomer in glycomic, proteomic, and glycoproteomic studies is ion mobility spectrometry (IMS) integrated with MS [368, 369]. Here analytes are separated based on their charge, size and shape in the gas phase (ions). Moreover, the technique can be hyphenated with a separation technique such as LC or CE prior to performing IMS-MS[370-373]. Collision cross section (CCS) values can be determined by IMS-MS which are particularly valuable for structural elucidation as they reflect the size and shape of the ions. However, IMS-MS data can be complicated due to the existence of conformers that overlap with structural isomers. Apart from MS, non-MS technologies such as nuclear magnetic resonance (NMR) spectroscopy are also often applied for the qualitative and quantitative analysis of oligosaccharide[141, 374-376]. But for complex biological samples, such as plasma derived PSA (one big mixture of diverse polysaccharides/glycopeptides), NMR is not an option. Well-characterized glycan standards can also be used to validated the assignment of glycan structure by aligning the elution/migration time, the tandem MS patterns and/or conformer distribution[377-379]. Glycan structures are commercially available in a variety of different forms, unlabeled glycan

3 H: hexose. N: *N*-acetylglucosamine. F: fucose. S: *N*-acetylneuraminic acid (Neu5Ac). K: ketodeoxynononic acid (Kdn).

standards, stable isotope labeled glycan standards or glycan standards labeled with fluorescent labels, e.g. 2-aminobenzamide (2-AB), 2-aminobenzoic acid (2-AA), procainamide, 8-aminopyrene-1,3,6-trisulfonic acid (APTS), although only limited glycoforms are available. In general, for accurate identification of glycan structures, especially novel structures like Kdn, it is suggested to apply several different techniques which are complementary to each other to acquire as much information as possible.

6.2 PAP and PSA as targets for biomarker discovery

6.2.1 Biofluids

A clinical biomarker is described by the National Cancer Institute as "*a biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease. A biomarker may be used to see how well the body responds to a treatment for a disease or condition. Also called molecular marker and signature molecule*"[380]. Literature has underlined the fundamental role of several biological materials as sources of useful information for diagnostic, prognostic and predictive purposes, such as breath[381], sweat[382], feces[383], urine, blood, tissue[384] and extracellular vesicles[385]. In this thesis, PAP and PSA derived from different biofluids including urine, seminal plasma and blood plasma were explored in the search of new biomarkers for different medical conditions, namely male infertility and PCa. Each biofluid has its own advantages and challenges as a biomarker source and will be discussed below.

6.2.1.1 Urine

Urine is one of the most commonly chosen biofluids for biomarker discovery in general as it contains a variety of different proteins and peptides which can serve as a source of potential biomarkers of human diseases[386]. To study PAP glycosylation, urine was chosen as a source for its several advantages in **Chapter 2**. First of all, PAP concentration in urine is normally very high (on average 513 ng/mL from nine healthy male adults) which is roughly 100 times higher than what is expected in plasma[387, 388]. Second, urine is less rich in proteins (10^5 ng/mL) compared to blood plasma (6.0–8.0 x 10^7 ng/mL)[389-391]. Consequently, when compared to plasma, urine comes with less interferences for the purification of target proteins and the purification tends to be easier and more successful. This is exemplified, by the rather clean SDS-PAGE after performing a PSA immunoaffinity capture[177]. Last but not least, urine can be obtained non-invasively in large quantities causing no burden to the participants with the exception when urine is collected after a digital rectal examination (DRE).

Urinary PSA and its glycosylation has been investigated extensively in different PCa studies. However, contradictory results have been reported[177, 186, 191, 392, 393]. Comparing the findings from different studies, no general patterns of altered glycosylation in urinary PSA have been observed. This might be due to various

reasons: There is no uniform analytical approach for glyco(proteo)mics studies. Different research groups have their own approaches. E.g., different analytical techniques were used (lectin-based assay versus MS approaches, and different lectins in lectin assays); different types of urine samples were analyzed (non-DRE urine versus post-DRE urine; urine samples of different stages of PCa). The limited amount of cancer cell-derived PSA which finds its way into the urine could explain why the glycosylation analyses of non-DRE urinary PSA found no biomarkers for PCa.

6.2.1.1 Seminal plasma

Seminal plasma has been explored for its biomarker potential in male infertility, reproductive disorders and prostate related diseases, such as, PCa and chronic prostatitis[394-396]. In **Chapter 3**, PSA derived from seminal plasma was studied in male infertility. PSA concentration in seminal plasma is approximately 1 x 10^3 times higher than its concentration in urine and estimated as high as 8 x 10^5 ng/mL. Even though a co-capture of semenogelin is observed in the MS spectra, it has no influence on the analysis of PSA glycopeptides. Although seminal plasma has its advantages for the analysis of PSA glycosylation, it is not a suitable material to explore PCa. Namely, PSA produced by different cells carries different signatures, PSA produced by 22Rv1 cells[4] has higher enzymatic activity than PSA produced by LNCaP cells[5], and this might also apply to glycan signatures[399]. As seminal fluids contains PSA produced by the whole organ, the majority of seminal PSA will be produced by healthy prostate cells and only a minor fraction of the PSA produced by cancer cells (potentially with altered glycosylation features). Besides, the expression level of PSA is lower in PCa than in normal prostate epithelium[400]. The cancer-type glycosylation features of PSA might be hidden by healthy-type glycan features. Therefore, seminal plasma was not explored for PCa in this study. The same also applies to urinary PSA as in **Chapter 3**, similar glycoprofiles of PSA were observed from urine and seminal plasma.

6.2.1.1 Blood plasma

From proteomics and glyco(proteo)mics point of view, blood is an attractive and a commonly used medium for the discovery of putative diagnostic and prognostic biomarkers of many human diseases as it is such a rich source of proteins and reasonably easy to obtain and process. Therefore, it is not surprising that blood and blood-derived fluids (plasma/serum) are already routinely utilized in clinical

4 22Rv1 cell line was derived from a human prostate carcinoma xenograft (CWR22R) that was serially propagated in nude mice after castration-induced regression and relapse of the parental , androgen-dependent CWR22 xenograft.397. Sramkoski, R. M.; Pretlow, T. G.; Giaconia, J. M.; Pretlow, T. P.; Schwartz, S.; Sy, M.-S.; Marengo, S. R.; Rhim, J. S.; Zhang, D.; Jacobberger, J. W., *In Vitro Cellular & Developmental Biology-Animal* 1999, 35, 403-409.

5 LNCaP cells are androgen-sensitive human prostate adenocarcinoma cells derived from the left supraclavicular lymph node metastasis from a 50-year-old caucasian male in 1977. 398. Horoszewicz, J. S.; Leong, S. S.; Kawinski, E.; Karr, J. P.; Rosenthal, H.; Chu, T. M.; Mirand, E. A.; Murphy, G. P., *Cancer research* 1983, 43 (4), 1809-1818.

settings as blood cellular and molecular composition reflect changes in an individual's health status[401]. In **Chapter 5**, a plasma PSA glycoproteomic assay was developed for the characterization of the glycosylation of plasma PSA and the assessment of its biomarker potential for PCa. According to literature, PSA derived from circulation clearly shows altered glycosylation in PCa[392]. Generally, it is believed that the presence of PSA in blood is due to the disrupted architecture of prostate tissue in the cancerous areas of the glands[43, 402]. This may explain why the blood PSA glycosylation change is more promising for PCa biomarker discovery than urinary and seminal PSA glycosylation.

6.2.2 PAP glycosylation as clinical biomarker

In order to explore the glycosylation of urinary PAP, a PAP glycoproteomics assay was developed (**Chapter 2**). The assay has proven its power in mapping the glycoprofile of urine PAP by being successfully applied to a male urine pool consists of nine post-DRE urine samples with unknown health conditions. Next, the validated assay will be employed to analyze a sample cohort to study glycosylation changes between PCa and non-PCa conditions. Before that, no conclusions can be drawn regarding its potential in PCa diagnosis or prognosis. The glycosylation of circulation-derived PAP also have potential value as PCa biomarker, however, glycoproteomic studies of circulation-derived PAP and its glycosylation in relation to PCa are lacking. Therefore, the further extension of the developed urinary assay to serum or plasma would be necessary.

As PAP is abundantly present in seminal plasma (1 mg/mL), it has been investigated for its potential role in male fertility since the 1980s[222, 403-407]. Singh *et al.,* demonstrated that sperm counts and motility are positively correlated with the level of PAP in men with very severe oligospermia (low sperm count)[404]. In another study, the level of PAP in azoospermic (semen that contains no sperm) group was found to be exclusively higher compared with groups normozoospermic (normal sperm parameters but infertility), asthenozoospermic (reduced sperm motility) and oligozoospermic (low numbers of spermatozoa)[406]. Recently, Kumar *et al*. studied on semen-derived amyloids and suggested that $PAP_{248-286}$ fragment can directly influence sperm motility and viability in a concentration-dependent manner[407]. The investigation of PAP functions in male fertility have been mainly focusing on its level and enzymatic activity in seminal plasma. In recent ten years, the glycosylation of PAP has been correlated to male subfertility[22, 408, 409]. However, there is a lack of in-depth examination of PAP glycosylation in relation to male infertility. Therefore, it is recommended to investigate whether PAP glycosylation plays a role in fertilization process and the developed PAP glycomics assay in this thesis (**Chapter 2**) could also be applied for these studies.

6.2.3 PSA glycosylation as clinical biomarker

The glycosylation of PSA has proven its value as clinical marker for early detection

of PCa[56, 410]. A PSA glycan-based test, namely the Giasay prostate blood test, is currently available in selected urology clinics developed by Glycanostics[56, 410]. This is an ELISA-based test which specifically detects cancer-specific glycans in free PSA with a specificity of 87.1% and sensitivity of 64.3% while the conventional serum PSA test only shows sensitivity of 24.3% with specificity of 87.1%[56]. The test uses antibody-coated magnetic particles to selectively extract free PSA. The extraction is added to a microtiter plate that is modified with glycan recognizing biomolecules (specifically for the LacdiNAc motif in this case). The intended purpose of this PSA test is to be used as a second opinion test to increase the accuracy of correctly prescribing prostate biopsy for cancer detection. It means that the test can reduce unnecessary biopsies but the definitive diagnosis still relies on biopsy. How altered PSA glycosylation in cancer serves as a marker independently and how the glycosylation changes correspond to cancers in different stages should be further explored.

As mentioned above (non-DRE) urinary PSA glycosylation does not have any biomarker potential in early assessment of PCa[177]. The hypothesis is that urine might not contain a high amount of cancer cell-derived PSA, while PSA enters the circulation is more derived from altered cells and more PSA can leak into the circulation due to the distorted cell structures of PCa area. Thus urinary PSA may present normal type glycosylation and plasma PSA carries cancer type glycosylation. It is proposed that urinary PSA may provide an unaffected (by PCa) glycoprofile and act as a baseline or normal control to test whether the PSA glycosylation appears changed in the serum[42]. It is of great interest to investigate the glycosylation of PSA derived from paired urine and plasma samples which is also our ongoing project. By applying the glycoproteomic assay established in **Chapter 5**, we aim to further investigate this hypothesis by studying a cohort consisting of paired plasma and urine samples. For this purpose, it will be examined whether glycosylation differences are observed between the two biofluids and, more importantly, to exploit for potential markers to, on one hand, distinguish PCa from benign prostate diseases, and on the other hand to discriminate indolent PCa from aggressive PCa. Furthermore, such a process would be beneficial for personalized medicine considering the large biological variation between individuals on glycosylation. More ideally, the research can be performed in a longitudinal way meaning follows up on the same cohort and monitor the glycosylation pattern of urinary and plasma PSA of a single individual over years (for example, starting time point, 5 years, 10 years and so on). More importantly, this information should be combined with the health condition of the prostate at the same timepoint.

However, exploring the glycosylation of plasma derived PSA for its biomarker potential is more challenging than the glycosylation analysis of urinary PSA owing to the high complexity of protein content of plasma. The total protein concentration in plasma/serum is approximately 60–80 mg/mL, among which

the most abundant proteins are albumins (about 50-60%) and globulins (40%, 10-20% immunoglobulin G, IgG)[390, 391]. Moreover, the target protein, PSA, is often present in very low abundance (~1-100 ng/mL) and, compared to urine, the concentration can be tens to hundreds times lower. In our assay (**Chapter 5**), dominating proteins in plasma such as IgG, IgA and IgM were co-captured. The co-captured Ig proteins are also glycosylated, which complicates the analysis of PSA on released *N*-glycan level. Also different from urinary PSA, plasma PSA occurs in both free and complex forms complexed with anti-trypsin and/or anti-chymotrypsin. The complex on one hand prevents degradation of PSA, but on the other hand results in the co-capture of α1-anti-trypsin and/or α1-anti-chymotrypsin, which are also glycosylated. All of the above, will have a huge influence on the purity of captured PSA and to ensure only the glycosylation of PSA is being studied, it is recommended to analyze the sample on a glycopeptide level rather than released *N*-glycan level as the peptide backbone can serve as an identifier.

In addition, for PCa biomarker discovery it will be important to obtain a case-control study with a well-designed cohort. Namely, a series of factors can cause biases and should be considered to ensure the patient groups are properly matched, such as, populations[411], age[412], PSA concentrations, and clinical stages of the tumor. A large cohort with age-matched patients from similar population grouping by different Gleason scores (Gleason score ≤ 6 vs Gleason score 7 vs Gleason score ≥ 8) would be reasonable. The median and average of the PSA concentration should be similar between the different patient groups as concentration difference may play a role and bias the glycosylation analysis. Proper controls are also very important in case-control studies which are normally from healthy individuals. However, in the case of PCa, the risk of PCa increases with age, especially after age 50 and age-matched individuals (which were considered healthy) might have other (unknown) prostate related diseases that could bias the outcome of the study. Therefore, it is recommended, in this case, that patients with a known other prostate disease (e.g., BPH) should be used as control group and the glycosylation features are potentially disease specific.

This thesis is focused on the development of new assays to investigate new potential markers. However, finding marker candidates is just the beginning of a long journey which eventually leads to the final goal of clinical diagnosis and is a multiphases process[413]. For example, if any interesting differences are observed, a new set of samples (the so called validation cohort) should be analyzed to verify the findings. Additionally, there will be two validation phases, the analytical validation which assess the assay characteristics (sensitivity, specificity, accuracy, precision, *et al.*) and the clinical validation to define the diagnostic performance and clinical utility[413, 414]. At the moment a few glycan-based tests have made it to clinics, for instance, the above mentioned Giasay prostate blood test and the Glyco Liver Profile test[315-318] from Helena Bioscience, which can be taken as

examples on successful clinical translation.

6.2.4 Kdn

Interestingly, an uncommon SA in mammals - Kdn - was observed in human PSA derived from seminal plasma and urine in the form of Kdn-containing *N*-glycans (H5N4F1S1K1, H5N4F1K1 and their non-fucosylated variants)[6] in **Chapter 4**. Later on, while characterizing the glycosylation features of PAP, one Kdn-containing glycan (H6N5F1S2K1) was identified in both seminal and urinary PAP. Similar to their presence in PSA, the Kdn-containing glycans were present at low relative abundance in PAP (<1 %). Many questions were raised based on these findings, as there is limited knowledge on Kdn in human: *Are those Kdn-containing glycans organ-specific or more generally distributed in human? Is the attachment of Kdn to human glycoproteins done by human glycosyltransferase or by a bacterial glycosyltransferase due to an infection? How is Kdn being transferred onto human N-glycans? What are the biological roles of the Kdn/ Kdn-containing glycans present in human? Are there any correlations between Kdn expression and incorporation in the human glycome and human disease?* As described above, the difficulty of identifying Kdn is for the reason of shared mass of compositions H2S1 and H1N1K1 which may lead to misinterpretation of glycan structures. Therefore, it is worthwhile to re-analyze some datasets that have already been analyzed. Currently, we are examining publicly available glycoproteomic datasets on different human samples for the presence of Kdn, including prostate tissue[115], stem cell[415] and human plasma. So far, Kdn-containing *N*-glycans have been observed in prostate tissue on tissue-derived α1-acid glycoprotein and PAP and also in human plasma on hemopexin and paraoxonase[115]. Suggesting that Kdn-containing glycans are not organ-specific but rather more broadly distributed in the human body than we thought and more human glycoproteins may carry Kdn modifications. The observation of Kdn-containing glycans in human plasma samples (*unpublished dataset*) from healthy donors implies that the attachment of Kdn to human glycoproteins was not likely due to bacterial infection but is performed by human glycosyltransferase. With this knowledge, the hypothesis is that human sialyltransferases (e.g. ST3GAL) transferring Neu5Ac can also transfer Kdn as it has been reported that certain human enzymes active on Neu5Ac show low level of activity on Kdn[116, 117]. For example, human *N*-acetylneuraminic acid-9-phosphate synthase (Neu5Ac-9-P synthase) presents low level of activity on synthesizing Kdn-9-phosphate[116, 117]. To verify this, enzymatic synthesis of Kdn-containing glycans should be performed *in vitro*. In addition, Kdn belongs to the family of SA and it occurs in the same position and same linkages as Neu5Ac in mammals, it is interesting to investigate whether Kdn has similar functions and whether it has similar biological roles as Neu5Ac. Further investigations are needed to gain more knowledge on the

6 H: hexose, N: *N*-acetylhexosamine, F: fucose, S: *N*-acetylneuraminic acid (Neu5Ac), K: ketodeoxynononic acid (Kdn).

biosynthesis and biological relevance of Kdn/Kdn-containing glycans in human.

6.3 Conclusion

In this thesis, a panel of glycoproteomics assays was developed to study two prostate secreted glycoproteins (PAP and PSA). This provides the basics for glycomic biomarker discovery in PCa. Different mass spectrometric workflows were introduced in **Chapters 2, 3** and **5**. Those studies highlight the importance of study design that is guided by a specific research question, e.g. focusing on depth, throughput or sensitivity. The information provided in this thesis can help with choosing suitable workflows in glycoproteomics studies in general and aid to a better design of methods. One of the interesting findings in this thesis was that human derived urinary and seminal plasma PSA and PAP carried Kdn-containing *N*-glycans (**Chapter 2 and 4**), revealing a new aspect of the human glycome as Kdn-containing glycoproteins were never before unambiguously identified in humans, emphasizing the importance of keeping an open mind when interpreting experimental results. Both the fundamental and the clinical findings described in this thesis provided novel insights for the research field of glycoproteomics and provide a stepping stone to develop and apply new glycoanalytical methods as well as to investigate glycoproteomic changes in human diseases such as PCa and male infertility.