Glycoproteomics assays for prostate cancer biomarker discovery

Wang, W.

Chapter 1

Introduction

Part of this introduction was adapted with permission from CE-MS approaches for Glyco(proteo)mic Analysis, Wang Wei, Wenjun Wang, and Guinevere SM Lageveen-Kammeijer, *Capillary Electrophoresis-Mass Spectrometry for Proteomics and Metabolomics: Principles and Applications*, 2022, DOI: 10.1002/9783527833092.ch13

Diagnosis of a disease is based on a wide variety of parameters, in which the role of molecular markers is increasingly important. None of the current medical tests is perfect in terms of sensitivity and specificity and commonly additional tools are required for confirmation of diagnosis, for example via imaging-based technology or via a more invasive approach such as a biopsy. The need for new and complementary biomarkers in less invasive tests is unmet and requires the development of new highly sensitive and specific analytical methods.

Altered protein glycosylation, one of the most complex post-translational modifications, has been reported in various human diseases[1, 2]. The investigation of glycosylation could facilitate the understanding of biological processes and human health conditions. The study of aberrant glycosylation has great potential in harboring clinical markers or providing therapy targets for human diseases[1, 3].

The research reported in this thesis is focused on two different proteins that have been applied as clinical markers for several decades. These clinically relevant proteins in fact consist of a plethora of isoforms, recently more precisely referred to as proteoforms[4]. For the identification of various glycoproteoforms of prostatic acid phosphatase (PAP) and prostate-specific antigen (PSA) the development of glycoproteomic assays is needed. The goal is to study the glycosylation of PAP derived from urine and PSA derived from seminal plasma and blood plasma and investigate their biomarker potential for male infertility and prostate cancer (PCa) by applying the developed assays to specific cohort samples.

This introduction provides background on male infertility and PCa as well as the basic principles of glycosylation and the analytical approaches to study this protein modification.

1.1 Male infertility

Infertility is a disease of the male or female reproductive system and is defined by World Health Organization (WHO) as the inability to conceive after at least 12 months of regular, unprotected sexual intercourse[5]. Infertility is a major health problem worldwide and is estimated to affect 8-12% of couples of reproductive age. In 50% of the cases, the male contributed to the infertility and in 20-30% of cases, the male were found to be solely responsible for the infertility[6]. Based on the findings from a study on the global burden of disease (GBD) in 2017, Sun *et al.,* reported the age-standardized prevalence rate of infertility increased by 0.29% per year for males from 1990 to 2017 globally[7]. The increased male infertility is explained by numerous different risk factors which can be classified to congenital factors (*e.g.*, congenital absence of *vas deferens*, chromosomal or genetic abnormalities, Y chromosome microdeletions), acquired factors (*e.g.,* varicocele, testicular trauma, testicular torsion) and environmental factors (*e.g.,* smoking, alcohol consumption, recreational drug use, obesity and psychological stress)[8]. Male infertility could be a herald to more serious health conditions, for example,

it has been associated with higher risk of mortality compared to fertile male in a prospective study[9] as well as a greater incidence of cancer[10, 11]. Here, infertile men were found to be 2.6 times more likely to be diagnosed with high-grade PCa compared to age-matched and geography-matched general population[12]. Infertile men with abnormal semen analyses have a 20-fold greater incidence of testicular cancer compared to age- and race-matched men from general population[13]. Bladder cancer and melanoma were also outlined to be related to male infertility[14]. Male infertility may be an early and identifiable risk factor for the development of clinically significant cancers[14]. Besides, malfunctioning of the male reproductive system (including poorer semen parameters and lower testosterone levels) accounts for a higher Charlson Comorbidity Index, which is a reliable proxy of a decreased general male health[15, 16]. Next to associating with poor physical health, infertility also negatively influences the patients both psychologically and economically. Appropriate management and an early diagnosis can moderate these effects[8].

Although a series of diagnostic tests are available, it remains challenging to diagnose whether a male is infertile, as multiple organs are involved as well as two individuals[8]. The first test to evaluate the male infertility is conventional semen analysis. It assesses a variety of semen parameters including volume, sperm concentration, total sperm count, sperm mobility, vitality and morphology[17]. Sperm function tests have been developed to provide the functional aspects of spermatozoa[18]. Further evaluation consists of physical examination, hormonal evaluation, genetic testing and imaging. Nonetheless, about 15% of infertile men, whose sperm meet WHO criteria, still face problems with conceiving, and are classified as unexplained male infertility (UMI). In the case of UMI, possible factors that might explain the difficulties to conceive include the presence of anti-sperm antibodies, sperm DNA damage, elevated levels of reactive oxygen species, and sperm dysfunction[19]. Still for a large group of UMI patients, the infertility factor remains unexplained.

Aberrant glycosylation (details will be described in section 1.3 of this Introduction) has been observed in glycoproteins involved in fertilization process, such as increased crotylation in seminal plasma of subfertile men[20]. In addition, α2,6-sialylation in total seminal plasma *N*-glycome has been found to be positively associated with sperm motility, while low α2,3-sialylation has been associated with a reduced sperm count[21]. Various observations indicate the potential value of glycosylation as a biomarker of male subfertility prediction and diagnosis[21-23]. Besides, sperm surface is highly glycosylated and some of its glycoproteins are considered as important factors for gamete binding[24, 25]. It is interesting to explore the glycosylation of sperm or sperm surface in different infertility and fertility conditions, to investigate if altered glycosylation occurs and how the altered glycosylation may disturb the interaction between sperm and the female immune system and contribute to UMI.

1.2 Prostate cancer

PCa is second most frequently diagnosed cancer and the fifth leading cause of cancer mortality in males worldwide. It was estimated that more than 1.4 million new cases (33.5% in Europe, 26.2% in Asia, 16.9% in Northern America, 15.2% in Latin America and the Caribbean, 6.6% in Africa and 1.6% in Oceania) would be detected in 2020[26]. The leading risk factors of PCa are advanced age, ethnic lines and germline mutation background as well as family history of cancers[27]. Other factors such as exposure to smoking, obesity, diet and vitamin D deficiency might also have an influence on the occurrence and progression of PCa[28, 29]. Besides, notable geographical differences are observed in PCa incidence and mortality analyzing the GLOBOCAN 2020 estimates, an online database providing global cancer statistics and estimates of incidence and mortality in 185 countries for 36 types of cancer and for all cancer sites combined. The data is demonstrated online at the Global Cancer Observatory (GCO) (gco.iarc.fr)[30]. Incidence rates vary from 6.3 to 83.4 per 100,000 males across regions, and the highest incidence of PCa was observed in more developed regions. In regard to mortality, the highest rates were observed in the Caribbean and Middle Africa followed by Southern Africa, while the lowest was observed in South-Central Asia and Eastern Asia. Multiple factors could cause this difference, such as, genetic differences, varying policies regarding PSA screening, different age-specific PSA cut-off values, adequate health insurance coverage and treatment choices[31-33].

PCa occurs in the prostate, which is a walnut-sized accessory gland in the males reproductive system, and contributes to the production of seminal fluid that nourishes and transports sperm. The prostate is located beneath the bladder surrounding the urethra. It can be divided into five regions including central zone, transition zone, peripheral zone and fibromuscular region and periurethral region[34, 35]. Among these regions the peripheral zone is the most common site of origin for prostate tumors[36]. The initiation of PCa is primarily driven by genetic mutations that occur in basal or luminal prostate epithelial cells[37-39]. PCa in its early stages is asymptomatic[40, 41] and may not cause serious harm. Therefore no or minimum treatment is needed and if treatment is required, the cancer can be treated effectively. In case of cancer progression, the symptoms (*e.g.* trouble with urinating, weak stream of urine, blood in the urine, blood in the semen, bone pain, weight loss and erectile dysfunction) could largely decrease life quality of the patients. More importantly, advanced PCa is less curable and life expectancy drops dramatically from 99% over ten years for early stage PCa to 30% at five years for advanced PCa[27].

It is clear that early detection of PCa is of critical importance for treatment efficacy, curativeness and patients survival. In clinics, the diagnosis of PCa is a multistep procedure that starts with the serum-based PSA level test. PSA screening is the most common technique to detect early stage PCa in asymptomatic individuals. This method measures the serum concentration of PSA and was approved by

the United States Food and Drug Administration (FDA) to screen men for PCa already in 1986. With this test, elevated PSA levels (PSA concentration of > 3 ng/mL (the Netherlands) or > 4 ng/mL (United States)) could be an indicator of risk of PCa. When elevated levels of PSA are observed, the person is advised, to assess risks and inform treatment, to have a digital rectal examination (DRE) and a follow-up prostate biopsy to determine the presence of PCa[42-44]. However, PSA screening is considered controversial as it has not unambiguously proven its overall benefits[44]. Namely, rising levels of PSA also occur with advanced age and benign prostate diseases, such as, benign prostatic hyperplasia (BPH) and prostatitis[45]. Therefore the serum-based PSA test is unable to discriminate PCa from other prostate-related conditions. Besides, it exhibits low sensitivity and it has a poor predictive value due to the low specificity, leading to overdiagnosis and unnecessary (invasive) treatments[46], especially when the PSA concentration falls within the grey zone (4-10 ng/mL). The screening test performance characteristics of PSA are variable depending on the PSA cutoff values used, for example a 30% specificity and 90% sensitivity are observed using the PSA cutoff values 3 ng/ml)[47]. Therefore, the unmet clinical needs in the PCa clinical pathway remain to be the (1) early diagnosis and prognosis of PCa with high sensitivity and specificity and (2) discriminating aggressive from non-aggressive PCa.

In recent years, altered glycosylation has been closely correlated to human PCa and has shown its biomarker potential[48]. The attention of the PCa biomarker field has been drawn to glycosylation characteristics of a variety of specimens, such as, total urine *N*-glycome, total serum *N*-glycome and *N*-glycosylation of prostate specific glycoproteins[49-55]. Sialylation, fucosylation, branching and LacdiNAc (antenna structure where *N*-acetylgalactosamine is β1,4-linked to *N*-acetylglucosamine) motifs are important glycosylation features assessed in PCa studies (**Table 1**). However, none of them have made it (yet) into a clinical setting. It is worth to mention that an innovative glycan-based non-invasive test by Glycanostics Ltd. for early stage PCa diagnosis with up to 90% accuracy is in pre-clinical validation stage, which is much more accurate than the conventional serum PSA test and only requires 10 µL of blood. The test is based on free PSA (fPSA) glycoprofiling, using the combination of age and fPSA% and four specific glycosylation features including fucosylation, LacdiNAc, α 2,3-sialylation and α 2,6-sialylation[56].

Table 1 Examples of earlier observed protein glycosylation changes in PCa.

Glycosylation feature	Specimen	Glycosylation changes	Reference
Fucosylation	Serum *N*-Glycome, *N*-glycans on serum-derived PSA	↑	57, 58
α2,3-sialylation	Serum *N*-Glycome, *N*-glycans on serum-derived PSA	↑	57, 58
LacdiNAc	*N*-glycans on serum-derived PSA	↑	59, 60
Branching	Serum *N*-Glycome	↑	61

1.3 Glycosylation and glyco(proteo)mic approaches

Glycans are carbohydrate or oligosaccharide structures that are covalently attached to other biomolecules such as protein or lipids to form glycoconjugates. It is important to note that all cells are covered with glycans and most of the cell surface proteins are glycosylated. Moreover, many of these glycans play an important role in modulating or mediating many biological processes, such as cell-cell recognition, cell-protein and cell-matrix interactions[2, 62-65]. Glycoconjugates can be divided into three distinct groups; glycoproteins, glycolipids and glycosaminoglycans. Glycans can be attached to a protein via a nitrogen atom of an asparagine residue (referred to as *N*-linked glycosylation) or can be attached to a protein via an oxygen atom of an threonine or serine residue (referred to as *O*-linked glycosylation). This thesis will focus on the analysis of *N*-linked glycoproteins. *N*-linked glycans are often branched structures and are mainly located on the plasma membrane or secreted in extracellular matrices[66]. Multiple enzymes (transferases) are involved in covalently binding the glycan moiety to the protein, which occurs co- and post-translationally - a process known as glycosylation. Given the fact that the glycan moiety can contribute for a big proportion to the total size of the conjugate, the presence of a glycan will have a substantial influence on the physiological and biochemical properties of the conjugates, such as solubility, stability and function[67].

Due to the often large influence of glycans on e.g. protein function it is not surprising that glycosylation plays an important role in human health and disease. Especially, alterations in the glycosylation pattern are related to various human diseases including autoimmune diseases, congenital disorders of glycosylation (CDGs) and cancers[1, 68, 69]. Glycosylation studies have resulted in the discovery of targets with potential for therapies, where the glycans can either be an essential part of the therapeutic or be the main therapeutic agent as glycan-based pharmaceuticals[70-76]. Furthermore, abnormal glycosylation is predominantly being studied for biomarker discovery for disease diagnosis, prognosis and treatment monitoring[1, 77, 78]. Still, it often remains unclear whether the altered glycosylation is the cause or the consequence of a disease and in-depth studies are needed to understand causality.

1.3.1 *N*-linked glycosylation

In humans, the most common monosaccharides in *N*- glycans are α-L-fucose (Fuc, F), β-D-glucose (Glu, H), β-D-galactose (Gal, H), β-D-mannose (Man, H), β-*N*-acetyl-D-glucosamine (GlcNAc, N), β-*N*-acetyl-D-galactosamine (GalNAc, N), α-*N*-acetylneuraminic acid (NeuAc). *N*-linked glycan structures are categorized in three different types, namely oligomannosidic type, hybrid type and complex type (**Figure 1**). The occurrence of *N*-linked glycosylation happens exclusively on a consensus sequence, namely, asparagine-x-serine/threonine (Asn-X-Ser/Thr)[79], where X represents any of the amino acids except for proline (Pro).

The biosynthesis of *N*-linked glycans starts at the cytoplasmic side of the endoplasmic reticulum[2], and further processing takes place in the lumen of the endoplasmic reticulum where the precursor glycan is attached to a protein. After the addition of the glycan, the protein is properly folded and the glycoconjugate with glycan composition $GlcNAc_2Man_{8\text{-}9}$ will be translocated to the Golgi apparatus where the oligomannosidic *N*-glycan will be trimmed down. From this point onwards the formation of hybrid and complex type glycans is initiated.

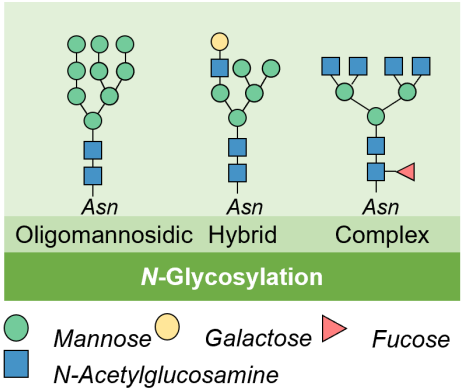

Figure 1. Overview of the *N*-linked glycosylation types in human.

N-linked glycans appear in all eukaryotes and are known to play a critical role in the regulation of many intracellular and extracellular functions[2, 80, 81]. With that, *N*-linked glycosylation profiles have extensively been investigated with regard to various human diseases, e.g. different types of cancer[59, 82], CDGs[83], diabetes[84], kidney disease[85, 86], inflammatory[87] and autoimmune diseases[88]. Significant features that seem to be altered between healthy and disease conditions include fucosylation [89, 90], sialylation[60, 84, 91] and bisection[92]. To only name a few examples, in colorectal cancer, a significant increase of sialylated Lewis-type structures and decrease of bisecting GlcNAc structures were observed in cancer tissues compared to control tissues. Sulfated structures and paucimannosidic structures were also increased[82]. In a large case-control study, fucosylation, α2,3-linked sialylation of triantennary glycans and bisection of diantennary glycans were decreased in the plasma *N*-glycome of type 2 diabetes patients, while total and α2,6-linked sialylation were increased[84]. Studying the serum *N*-glycome, antennary fucosylation of complex *N*-glycans, branching of complex *N*-glycans and α2,6-linked sialylation were found increased in pancreatic ductal adenocarcinoma compared to control samples[93]. For the treatment of MPI (mannose-6-phosphate isomerase)-CDG, a disorder of the mannose metabolism during glycan production, mannose dietary supplement is approved[69]. Overall, the above mentioned glycosylation signatures can be exploited as a biomarker for disease diagnosis, prognosis or as a target with therapeutic purposes.

1.3.2 Sialylation

Sialylation is an interesting glycosylation feature in relation to human health. Sialylation is defined as the covalent attachment of sialic acid (SA) to a glycoconjugate (glycoprotein; glycolipid) by a sialyltransferase, at the non-reducing end of the glycan chain via a repertoire of different linkages. For example, SA commonly links to Gal or GalNAc via α2,3- or α2,6-linkage[94, 95],

whereas sialylation at another SA occurs through an α2,8- or α2,9-linkage[96, 97]. SA consists out of a large family of neuraminic acid (Neu) varients of which the three major naturally occurring forms are Neu5Ac, *N*-glycolylneuraminic acid (Neu5Gc) and ketodeoxynononic acid (Kdn; deaminated neuraminic acid) (**Figure 2**).

Sialylation is known to play an essential role in many biology processes including cell signaling, cellular recognition and communication, cellular aggregation and development[98]. Sialylation is involved in mediating bacterial and viral infections as sialylated glycoconjugates are used as an entry receptor by many different viruses[99]. Via lectin and receptor binding, sialylated glycans are impacting key processes in health and disease, for instance, the immune response (SA-Siglecs, SA-Selectins), tumor growth and metastasis (SA-Selectins, hypersialylation of β1-integrin and Fas receptors)[100-103]. Besides, the expression of poly-sialic acid epitopes have been strongly correlated with various types of late stage, highly metastatic cancers[104, 105]. Above mentioned findings make SA a promising tumor marker and target for therapies. It is of particular importance to distinguish the diverse SA linkages as they are involved in different biological process and exhibit different functions[2, 106].

The most common SA in humans is Neu5Ac, whereas Neu5Gc is observed in minor amounts in certain human cell types, particularly epithelia and endothelia[107]. Remarkably, the presence of Neu5Gc is unexpected since in humans a 92 base pair deletion in the cytidine monophospho-*N*-acetylneuraminic acid hydroxylase (CMAH) gene has occurred that excludes biosynthesis of Neu5Gc[108]. Several

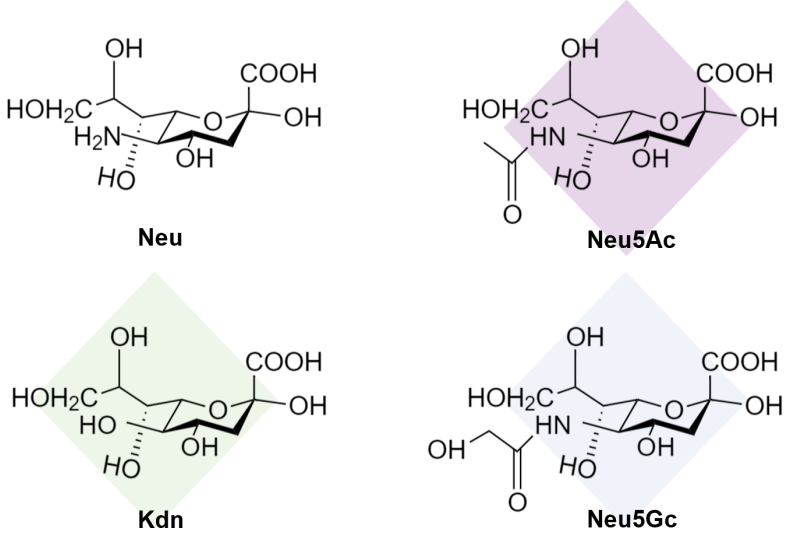

Figure 2. Neu and the three major forms of its varients (Neu5Ac, Neu5Gc, Kdn).

hypotheses have been proposed to explain the presence of Neu5Gc in human tissues that need further investigation[109]. So far, it has been shown that the incorporation of Neu5Gc into human glycoconjugates can happen through our diet, such as red meat consumption, and via cell feeding experiments using human epithelial carcinoma cells, normal skin fibroblasts and neuroblastomas [109,110]. A previous study demonstrated that the incorporation of Neu5Gc promotes inflammation and tumor progression[111] and higher quantities have been found in human tissue and serum of cancer patients[110, 112-115]. The Kdn structure is often seen in high amounts in bacteria and lower vertebrates. For most mammals the biosynthesis is inactivated by the substitution of a methionine to a threonine at position 42 of *N*-acetylneuraminic acid 9-phosphate synthase (Neu5Ac-9-P synthase) which is the key enzyme for the synthesis of Kdn[116]. In mammalian cells and tissues, Kdn can still be detected as the activity of Neu5Ac-9-P synthase for synthesizing Kdn is restored exclusively in two mammalian lineages: Artiodactyla and primates (including humans)[116, 117]. The majority of Kdn occurs as the free monosaccharide and minority occurs conjugated to glycoproteins and/or glycolipids[118-121]. Elevated level of free Kdn, cytidine monophosphate (CMP)-Kdn and Kdn-containing *N*-glycans has been observed in several human diseases, such as end stage renal disease[116], ovarian cancer[118] and PCa[122]. One study proposed a hypothesis on elevated free Kdn expression in human cancers cells: *the common feature of advanced localized tumor-hypoxia induces the activity of crucial enzymes including Neu5Ac-9-P synthase and phosphomannoisomerase (PMI) and enhance the expression of mRNA for corresponding enzymes in Kdn biosynthetic pathway*[123]. However, further investigations are needed to truly understand how Kdn is incorporated in the human glycome and its role in human diseases.

1.3.3 Glycosylation workflows

To investigate the glycosylation of a glycoprotein several different approaches can be employed, including released glycan analysis, a bottom-up approach and intact analysis. Depending on the sample type and research questions, different strategies can be applied. The most commonly performed procedure is the enzymatic release of *N*-linked glycans[124]. In this case the entire glycan moiety is separated from the protein. Released glycan analysis is the preferred choice for analyzing complex samples to provide an overall glycomic profile of the complex mixture, for instance, a total plasma glycome. Compared to a bottom-up glycopeptide-based approach, the acquired data is less complex as the heterogeneity of the protein carriers is removed and solely the glycan moieties are being studied. However this also results in loss of information such as site occupancy and origin of the glycoproteins. The prominent advantage of performing intact glycoprotein analysis is that it covers the whole analyte while it requires no proteolytic digestion, allowing the (glyco)protein to be analyzed in its native or denatured form[125, 126]. Though intact analysis is able to identify the

co-existence of PTMs including glycosylation, phosphorylation and acetylation, the information about site specificity is often limited. Overall, intact analysis is a promising field to identify the coexistence of PTMs[127].

Using a bottom-up glycoproteomics strategy both site- and protein-specificity are obtained, provided that the peptide backbone is unique for the protein. The work described in this thesis is based on these glycopeptide analysis strategies in which the peptide portion remains attached to the glycan moiety. To generate peptides and glycopeptides, a variety of proteases can be employed to digest the glycoprotein into smaller parts. The bottom-up approach comes with its own challenges. For example, the proteolytic enzyme can be hampered by the large size of the glycan moiety resulting in missed cleavages and generating large and often heterogeneous peptide backbones. In addition, the diversity of glycosylation on a single site (microheterogeneity) distributes the signal of the peptide backbone over several species, which requires high sensitivity to enable the detection of these analytes. Also glycan occupancy often happens at multiple glycosylation sites. Moreover, the presence of other PTMs present on the protein, further complicate data-analysis. The presence of non-glycosylated peptides can suppress the ionization and detection of glycosylated peptides by MS. The improvements in glycopeptides enrichment and separation strategies have contributed greatly to improve sensitivity and coverage in this respect[128]. In addition, until recently, it was difficult to gain simultaneous information on the peptide and glycan moiety using tandem mass spectrometry (MS/MS). However, the emergence of fragmentation methods such as stepping energy in collision induced dissociation (CID)[129] and higher-energy C-trap dissociation (HCD) [130] as well as the development of targeted bioinformatics tools [131] makes the interpretation of glycopeptide structures more feasible and easier for the end user.

As mentioned above, the glycan moiety is highly heterogeneous as different glycan structures can be attached. Due to their diversity in linkages and composition of monosaccharides, elucidation of the structural composition of a glycan is often required and further sample preparation are needed such as derivatization or enrichment strategies. Especially in the case for MS analysis, enrichment strategies might be needed to prevent ion suppression by other analytes[128]. To distinguish diverse linkages of SA and to stabilize SAs for subsequent MS measurement, a series of chemical strategies for SA linkage-specific derivatization have been developed and adapted in many diseases studies by achieving different derivatization of the carboxyl groups of the isomers, such as, permethylation, methylamidation and dimethylamidation[132-140]. The dimethylamidation approach was established to explore IgG glycosylation on glycopeptides level and under the reaction conditions, α2,6-linked SAs form a dimethylamide, while α2,3-linked SAs react with the neighboring galactose and form a lactone. The product can be further stabilized with an additional

amidation step which would open the lactone ring and form a primary amine. The peptide moiety will also be modified at the C-terminal and when carboxyl group containing amino acids are present, such as glutamic acid and aspartic acid. It should be kept in mind that side products might be observed when a free primary amine, hydroxyl or carboxyl group is available caused by both inter- and intramolecular reactions.

The absolute quantification of glycans/glycopeptides remains challenging due to their heterogeneity and the lack of proper internal standards for all glycan species, thus relative quantification is commonly performed[141]. For glycopeptide quantification, labeling strategies are often employed including metabolic incorporation, isotopic labeling and isobaric labeling[142]. For example, tandem mass tag (TMT) is a widely used, amine reactive isobaric label of which the mass reporter will be cleaved off during MS fragmentation and form a reporter ion for analyte quantification, while the mass normalizer enables multiplexed sample analysis. TMT has been applied in various biological samples for glycoproteomic profiling[143-146].

1.3.4 Analytical approaches

MS is a powerful analytical tool for the identification and characterization of complex glycan structures and larger glycoproteins and glycoconjugates. By applying MS/MS specific structural details are obtained that result in confident structure assignments. In general, a mass spectrometer consists of an ion source, a mass analyzer and a detector. As the field is rapidly evolving, the systems are continuously improved providing higher sensitivities and resolving powers, which is needed in the glycoproteomics field. Some monosaccharides residues are labile, such as fucoses can be re-arranged and SAs can be detached during ionization[147, 148]. Hence milder ionization techniques is used to get the analytes into the gas phase. The most widely applied ionization techniques are matrix-assisted laser desorption/ionization (MALDI) and (electrospray ionization) ESI. MALDI-MS has been extensively used for glycomic profiling and disease biomarker discovery from biological specimen[149, 150] due to its ability for high-throughput analysis. Prior to measurement, samples are generally co-crystalized with a matrix that ensures that the laser energy for desorption and protonation/ionization is passed onto analytes. In MALDI, the presence of (multiple) SAs on glycans or glycoconjugates may considerably influence ionization properties in both positive- and negative-ion mode. However, as a glycomic mixture frequently contains a combination of sialylated and non-sialylated glycans, analysis is often operated in positive ionization mode after a derivatization procedure to neutralize the sialylated species. Moreover, derivatization can prevent in-source fragmentation of the sialylated species[147]. MS lacks the ability to distinguish isomeric species as the same *m/z* ratio will be observed, hampering a full identification and presence of glycan species in a sample. This can be partly solved using the aforementioned derivatization procedures. Prior to the MS analysis, a separation technique is

often employed. This extra dimension usually results in higher sensitivity and more exhaustive glycan information in combination with high accuracy, high resolution mass detectors. For this purpose, ESI is commonly applied rather than MALDI as ESI has the ability to be hyphenated online to a separation platform.

Liquid chromatography (LC)-ESI-MS is the mostly applied strategy in glycomics and glycoproteomics due to its versatility, robustness and separation power. Chromatography is achieved based on the interaction of the analytes and the stationary phase. The separation efficiency and sensitivity can be even further improved by using ultra high performance LC (UHPLC) and nanoLC systems. Various chromatographic stationary phases have been applied to glycosylation analysis, including reverse phase LC (RPLC), hydrophilic interaction liquid chromatography (HILIC), and porous graphitized carbon (PGC) chromatography. RPLC uses non-polar, hydrophobic materials such as C18 or C8-bonded silica as stationary phase and organics as mobile phase. Here, hydrophobic compounds will retain on the non-polar material non-covalently. However, due to their hydrophilic properties, retention of the glycan part on a RPLC column is very limited. Therefore, in the case of glycopeptides the retention depends on their hydrophobic peptide moiety or, in case of released glycans, an additional modification is required such as a hydrophobic tag or permethylation[151]. Once retention is achieved the glycan part affects the retention of the analyte, by the presence of different monosaccharides, their positions as well as their linkages. The separation abilities of PGC and HILIC have been found to be outperform RPLC in separating glycan compositions and their structural isomers[152-154]. Especially the PGC platform does not require an additional derivatization or labeling step, leading to minimal sample preparation time as well as less sample loss. Overall, HILIC has high peak capacities and with that it is a suitable method for the analysis of complex biological mixtures, while, PGC has excellent separation power in regard to isomeric species[155].

Capillary electrophoresis (CE)-MS is another powerful tool for glyco(proteo)mic applications due to its ability to characterize and identify released glycans up to intact glycoproteins and more importantly its baseline separation capability. While CE-MS was already reported in the early 90s, it took sometime before the interest in CE-MS for glycomic studies grew. Just as in regular CE applications, the sample can be loaded in two different ways, namely, through a hydrodynamic injection (applying an external pressure) or electrokinetic injection (applying an electric field). Separation is often achieved by applying a capillary voltage from 10 up to 30 kV and the result is based on charge-to-size ratio of the analyte, the viscosity of the buffer and the hydrodynamic volume of the analyte. In combination with MS, a platform is created which combines the advantages of both techniques, whereas CE provides excellent separation power, enabling baseline separation of isomeric species and high sensitivity, MS provides the ability to elucidate co-migrating, non-isomeric, complex compounds. An excellent

example of the strength of CE-MS is its ability to separate α2,3- and α2,6-sialylated isomers, without the need for any additional sample pre-treatment next to a conventional proteolytic digestion[156]. The different migration behavior between the two isomeric variants has been correlated to a minor difference in Pk_a values, moreover it is likely that the separation is also driven by a slight difference in hydrodynamic volume but this needs further investigation.

1.4 PSA glycosylation

PSA, also known as human kallikrein 3 or gamma-seminoprotein, is a member of kallikrein related peptidase family which is encoded by KLK3 gene in humans. It is synthesized in the prostate gland and secreted by its epithelial cells. In healthy males, PSA is normally confined to the prostate gland. However, it may diffuse from the lumen into the bloodstream in nanogram per milliliter concentrations either bound to another protein or as fPSA. PSA is considered to be prostate-specific, although small amounts of PSA can be found in other bio-fluids like breast milk, amniotic fluid and female ejaculate[157-160]. Due to its specificity, it has been extensively studied as a potential biomarker to indicate the occurrence of prostate related diseases, especially regarding PCa, but also others such as, prostatitis[161-163], BPH[164, 165] and male infertility[166-168].

PSA is one of the most abundant proteins in the secretion of normal human prostate epithelium and seminal plasma. It plays an essential role in reproduction as it exhibits chymotrypsin-like enzymatic activity that can digest seminal vesicle secreted proteins, such as semenogelins[169, 170], resulting in the liquefication of semen and enhancing sperm motility. Apart from facilitating coagulum liquefaction, PSA is also important for sperm fertility as its physiological substrate, semenogelins, were found to play significant roles in sperm fertility by affecting hyaluronidase activity, capacitation and motility[170-176]. The exact role of PSA glycosylation in male fertility is still not clear and more efforts should be taken to study role of PSA glycosylation in relation to the fertilization process.

As described previously, PSA is used in the clinic as a diagnostic biomarker for PCa *via* a blood-based PSA level test. Unfortunately, the PSA test shows low specificity and is unable to discriminate PCa from other prostate related diseases[46]. Therefore, more accurate and advanced biomarkers are needed for the early detection of PCa. Extensive efforts have explored whether altered glycosylation of PSA could serve as a more specific and promising PCa biomarker, especially in the last 10 to 15 years[43, 58, 59, 177]. Literature suggests that the glycosylation of PSA possess great potential for PCa early detection and prognosis[178-185] and the glycome of various specimens have been investigated, such as urine[177, 186-188], serum[58, 60, 189] and tissues[190, 191]. The most abundant *N*-glycan type in PSA derived from urine and seminal plasma was found to be the complex type *N*-glycans, following by minor amounts of hybrid and oligomannosidic *N*-glycans[168, 177]. While, some similarities were found between the urinary and seminal *N*-glycome to the

serum derived PSA[60, 168, 177], serum PSA also revealed high amounts of sialylated tri- and tetra-antennary complex *N*-glycans[60]. All biofluids have a high abundance of core fucosylation, while PCa cell lines, such as LNCaP and PC-3, also revealed antennae fucosylation[190, 191]. This shows that PSA derived from these cell lines might not be a representative PSA *N*-glycome derived from biofluids. Indicating that PSA glycosylation is highly dependent on the source of its production and secretion, which is interesting to take along in PSA glycomics studies.

One of the most intriguing changes in the glycosylation of PSA is the variation in linkages of the SAs and their abundance in relation with PCa. The percentage of α2,3-linked SAs of PSA in serum was evaluated by Llop *et al.* as a potential glycan biomarker. In this lectin study, high-risk PCa patients could be distinguished from intermediate-risk PCa, low-risk PCa and BPH[58]. In another study, the overexpression of α1,6 fucosyltransferase was found in metastatic and aggressive primary PCa implying increased level of fucosylation[192]. Later on, in a lectin study researchers showed that the level of serum fucosylated PSA is correlated with the tumor Gleason score (GS)[193]. Namely, aggressive PCa (GS 7-9) could be differentiated from non-aggressive PCa (GS 6) with area-under-the-curve (AUC) of 0.7056 and 0.7662 for fucosylated PSA and percentage of fucosylated PSA, respectively. Another interesting glycosylation feature is the LacdiNAc motif which revealed to be upregulated in clinically significant PCa as compared to those in biopsy negative men and active surveillance eligible PCa by using a lectin assay (WFA) in a large study (n = 442)[194]. It should be noted that, using a lectin-based assay, most likely the sialylated LacdiNAc motifs are not bound and are not taken along into the equation. However, it was demonstrated that in most cases the LacdiNAc antennae is occupied with an α2,6-linked SA[195]. Especially the LacdiNAc motif seems to be an important feature to distinguish PCa from BPH[59, 196]. For example, Haga *et al*. found that multisialylated LacdiNAc structures are significantly upregulated in the PCa group compared to the BPH group[60]. High branching structures have also been observed in PCa[197].

While the previous findings provide promising results, it should be noted that most of these studies have been making use of lectins or immunoassays and, as stated earlier, the main limitation of these approaches are the lack of in-depth characterization and the potential of non-specific binding. Therefore, more in-depth characterization is needed to gain better insights as well as better defined glycan biomarkers that will hopefully show potential in predicting the progression of a disease.

1.5 PAP glycosylation

PAP is a glycoprotein that is synthesized in prostate epithelial cells. It exists in two different forms, namely intracellular PAP and secreted PAP (sPAP), of which the latter normally forms a 100 kDa dimer and acts as a phosphatase. The serum concentration of PAP was clinically used as a biomarker of PCa, but was later

replaced by PSA with improved test sensitivity and specificity[198, 199]. While later on Kong et al. demonstrated that PAP has a higher correlation with PCa progression compared to PSA[200] and due to the current challenges PSA still faces as a biomarker[201], it would be worthwhile to investigate whether the glycosylation of PAP could contribute to the early discovery of PCa. Especially because previous studies have already shown a correlation between an altered glycosylation profile of PAP and PCa[202-205]. By using SA and fucose binding lectins, a significant decrease in fucosylation and sialylation was observed in PAP that was derived from aggressive expressed prostatic secretions (EPS) urine pool upon comparison of indolent PCa *versus* non-PCa pooled EPS urine samples (each pool contained ten individual samples)[202]. In addition, a decrease in fucosylated di- and tetra-antennary classes in PCa patients as well as a reduction in the oligomannosidic species (Man6) and an increase of less complex *N*-glycan structures (e.g. H4N3 which consists out of one galactose, three *N*-acetylglucosamines and three mannoses) was reported in pooled seminal PAP compared to normal control and BPH pool[203].

PAP has three (partial) occupied *N*-linked glycosylation sites[203, 206, 207]. In 2009, White *et al.*, studied the glycomic profile of PAP derived from seminal plasma and reported 21 *N*-glycan structures (*N*-glycan release) by MALDI coupled to time-of-flight MS (TOF-MS)[203]. In addition, to identify the glycan feature per site, glycopeptide analysis was performed and revealed that site N_{333} contained only oligomannosidic structures, while sites N_{94} and N_{220} were occupied with complex type *N*-glycans. Within a later study it was also found that *N*-glycosylation site N_{333} was occupied with oligomannosidic type *N*-glycans, but also contained complex type *N*-glycans[202]. This discrepancy could be related to the difference in biofluid, namely this study investigated PAP from direct EPS urine rather than seminal fluid. While these studies provide insights in the *N*-glycome of PAP, limited information was provided (e.g. a low number of identified *N*-glycans and *N*-glycosylation site N_{220} was not resolved in some studies) as well as a complex dataset was obtained due to missed cleavages observed on all three glycosylation sites[202, 203]. More importantly, no distinction could be made between sialylated isomeric species, which is of crucial importance in cancer studies as α2,6-linked isomers were presumed to be involved in blocking galectin binding which enhances the tumor cell survival while α2,3-linked isomers are considered as hallmark of malignant types of cancers[208-211]. As far as we know, no study has investigated the PAP glycosylation profiles at an individual level. Therefore, to gain a further understanding of PAP and its biomarker potential in relation to PCa, an in-depth characterization of the PAP *N*-glycome is of critical importance. As is the case for PSA, it is important to take into account the source of PAP production and secretion in its glycoproteomics studies.

Scope

Protein glycosylation is vastly important in all organisms. Its identification and characterization facilitates the understanding of biological processes in general and in healthcare allows for the development of clinical markers for various diseases. The aim of the work described in this thesis is to explore the glycosylation of PAP and PSA obtained from different biofluids including seminal plasma, urine and blood plasma, and to investigate whether or not glycosylation profiles differ between diseased and healthy individuals. For this purpose, three glycoproteomic assays were developed to discover potential glycan biomarkers for improved detection, stratification and prognosis of disease with a specific focus on PCa and male infertility.

In **Chapter 1** an overview is provided of the general concept of glycosylation, its features and common analytical approaches. In addition, a birds-eye-view is given on the current clinical practice with regard to male infertility and PCa and the corresponding clinical needs. Emerging glycomic biomarkers are introduced including two specific prostate-secreted glycoproteins, namely PAP and PSA.

Chapter 2 aims to investigate the *N*-glycome of PAP derived from DRE urine by developing an in-depth glycoproteomic assay for urinary PAP. A high sensitivity and a powerful separation is achieved by using CE-MS as analytical platform, providing identification of all three glycosylation sites in PAP, including differentiation of SA linkages and separation of isomeric species.

In **Chapter 3** the development of a high-throughput glycoprofiling method for seminal PSA is described. The assay includes an immunocapture step, tryptic digestion, and an adapted version of a previously developed SA linkage-specific derivatization strategy. After analytical validation the assay is applied on a cohort consisting of fertile and infertile males with different origin of infertility to investigate the biomarker potential of altered PSA glycosylation derived from seminal fluid.

Sialylation is a recognized glycosylation feature in various human diseases. For example, elevated levels of free Kdn have been observed in several human cancers even though this SA species is rather uncommon in humans. **Chapter 4** investigates the presence of Kdn sialylated *N*-glycans on seminal PSA for the first time in mammals. Four different analytical approaches are applied to identify the observed *N*-glycans. Next to seminal fluid, urinary PSA is examined on the presence of Kdn and its biomarker potential is investigated by analyzing a cohort that consists of PCa and non-PCa samples.

The aim of **Chapter 5** is to further expand and improve on previously developed in-depth glycosylation assay for urinary PSA to allow the characterization of PSA glycosylation derived from plasma with PSA concentration > 3 ng/mL. TMT labeling is performed to enhance the hydrophobicity of derivatized tryptic PSA

glycopeptides and simultaneously facilitate sample multiplexing in MS/MS-based quantification and identification. Eventually the biomarker potential of paired urinary and plasma PSA can be evaluated by analyzing a cohort that consisted of samples from PCa and non-PCa patients.

Finally, in **Chapter 6** the work described in the previous chapters is discussed with a specific focus on the advantages and disadvantages of different MS workflows in prostate focused glycoproteomics and provide guidance on make suitable choices on analytical workflow. A discussion on the potential of altered PAP and PSA gylcosylation in a clinical setting for various diseases is also provided.