Universitat de Girona

Facultat de Ciències

Memòria del Treball Final de Grau

Títol del treball:

GLYCOMIC AND GLYCOPROTEOMIC APPROACHES: POTENTIAL TOOLS FOR THE IMPROVEMENT OF THE GOLD STANDARD BIOMARKER FOR PROSTATE CANCER DIAGNOSIS

Estudiant: Heura Domènech García

Grau en Biotecnologia

Correu electrònic: u1929120@campus.udg.edu

Tutor: Dra. Esther Llop

Cotutor*:

Empresa / institució:

Vistiplau tutor (i cotutor*):

Nom del tutor: ESTHER LLOP

Nom del cotutor*:

Empresa / institució:

Correu(s) electrònic(s): esther.llop@udg.edu

*si hi ha un cotutor assignat

Data de dipòsit de la memòria a secretaria de coordinació: 20/07/2017

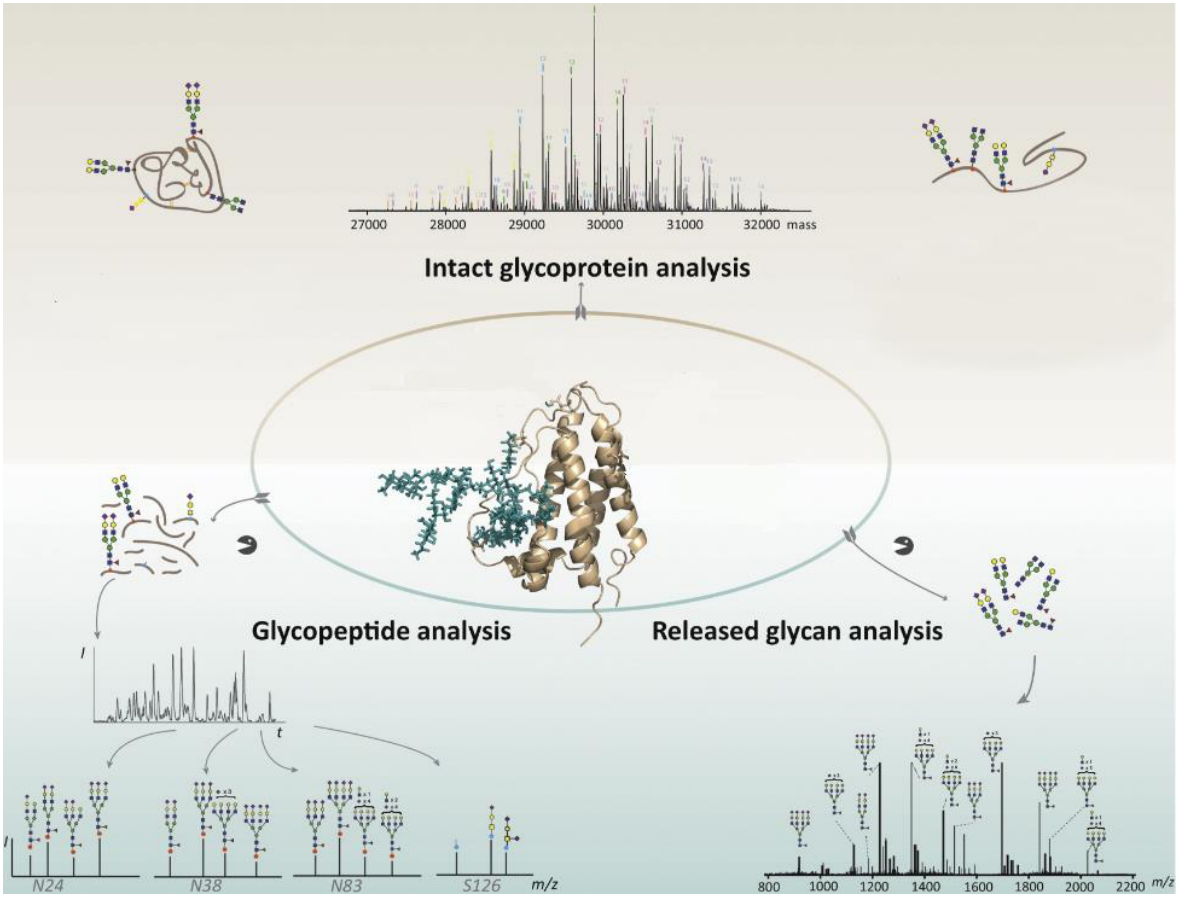
Bachelor Thesis

GLYCOMIC AND GLYCOPROTEOMIC APPROACHES: POTENTIAL TOOLS FOR THE IMPROVEMENT OF THE GOLD STANDARD BIOMARKER FOR PROSTATE CANCER DIAGNOSIS

Author: Heura Domènech García

Bachelor degree in Biotechnology

Academic year 2016-2017



20-07-2017

UNIVERSITY OF GIRONA
Faculty of Science
Under the supervision of: Dra. Esther Llop

ABSTRACT

Biomarkers of cancer have become the central stage for many scientists since last years. Because of the rapid emerging technologies, scientists can now understand these biological markers and apply them for the early detection of cancer. Thus, providing potential noninvasive, sensitive and reliable assays that, in fact, could reduce cost in health care and increase quality of life. A broadly used serum marker is the prostate-specific antigen (PSA), which is referred as the *gold standard* tumor marker for the diagnosis of prostate cancer (PCa). The PSA test consists in the monitoring of PSA levels in sera and an increase of its levels may indicate that an individual can have PCa. However, serum PSA levels are also high in other diseases, such as benign prostatic hyperplasia (BPH). Moreover, PSA assay cannot distinguish between non-aggressive and aggressive tumors. Many scientists, however, have started to focus on studying some biochemical characteristics of PSA, especially the glycosylation pattern of PSA which has shown to be altered in cancer. The recent fields of Glycomics and Glycoproteomics, which encompass glycan or glycoprotein enrichment and proteomics technologies, are developing more sensitive and high-throughput techniques for an in-depth and quantitative identification of specific glycosylation abnormalities in a complex biological sample. This work focuses on a broad literature review about several glycoproteomic and glycomic methodologies applied for improving the specificity of PSA as well as on searching through databases and using some bioinformatic tools, which are a key element for the analysis and characterization of PSA. Moreover, three approaches - ELLA, CE-ESI-MS and MALDI-TOF-MS - have been explained in-detail and discussed. Particularly, a comparison has been stablished between several methods and strategies by mentioning the different advantages and limitations that these approaches present. This paper shows that through the application of the distinct approaches, a tremendous amount of data about the heterogenous glycan composition of PSA was revealed and significant alterations in the glycosylation pattern were reported, when comparing samples of PSA driven from healthy or BPH individuals with PCa patients. Hence, suggesting that the glycosylation pattern of PSA might be more reliable and sensitive for PCa diagnosis than PSA test. Nonetheless, more efforts are required to consider PSA glycosylation as the new golden standard.

RESUM

Els biomarcadors tumorals han guanyat interès durant els darrers anys. Gràcies al desenvolupament de noves tecnologies, els marcadors tumorals es poden estudiar en detall i utilitzar per a la detecció precoç del càncer. Així doncs, proporcionant una alternativa als mètodes de detecció convencionals invasius. De fet, la utilització d'aquests en la pràctica clínica podria reduir els costos en l'atenció sanitària i augmentar l'esperança de vida. L'antigen específic de la pròstata (PSA) és un dels marcadors més utilitzats en per a la detecció del càncer de pròstata (CP). El test de PSA consisteix en el seguiment del nivells de PSA en sèrum, els quals solen ser molt baixos en condicions fisiològiques. Si la concentració és veu augmentada, això podria indicar que l'individu té CP. Tanmateix, els nivells de PSA també s'han vist incrementats en malalties com en la hiperplàsia prostàtica benigna (HPB). A més, l'assaig de PSA no pot diferenciar entre tumors no agressius i agressius. Per tal de resoldre aquesta manca d'especificitat, s'han començat a estudiar algunes característiques bioquímiques del PSA, com el patró de glicosilació del PSA, entre d'altres, que s'ha vist alterat en càncer. Les àrees de "Glycomics i Glycoproteomics", estan desenvolupant tècniques cada vegada més sensibles i automatitzades per a la identificació i quantificació d'anomalies específiques de glicosilació en una mostra biològica. Aquest treball es centra en una àmplia revisió bibliogràfica sobre diverses metodologies glicoproteòmiques i glicòmiques aplicades per millorar l'especificitat del PSA, així com en la recerca de bases de dades i l'ús d'eines bioinformàtiques, que són un element clau per a l'anàlisi i la caracterització del PSA. Altrament, tres estratègies diferents - ELLA, CE-ESI-MS i MALDI-TOF-MS – han estat analitzades i comparades, tot fent esment dels avantatges i limitacions d'aquestes. Mitjançant l'aplicació de diferents estratègies, s'ha pogut obtenir una gran quantitat de dades sobre la composició heterogènia dels glicans del PSA, com també s'han trobat alteracions significatives en el patró de glicosilació, en comparar mostres de PSA d'individus sans o amb HPB amb pacients amb CP. Aquests resultats suggereixen que el patró de glicosilació de PSA podria incrementar l'especificitat i sensibilitat del diagnòstic de CP. No obstant, faran falta més esforços per al desenvolupament de noves estratègies més sensibles, per tal de poder considerar la glicosilació del PSA com el nou biomarcador estàndard.

RESUMEN

En los últimos años, los biomarcadores del cáncer han sido de gran interés para muchos científicos. Gracias al desarrollo de nuevas tecnologías, los marcadores tumorales pueden ser estudiados con exactitud e, incluso, utilizados para la detección precoz del cáncer. Así pues, los marcadores tumorales podrían ser una alternativa a los métodos de detección convencionales invasivos. De hecho, la utilización de estos en la práctica clínica podría reducir los gastos en la atención sanitaria y aumentar la esperanza de vida. El antígeno específico de la próstata (PSA) es uno de los marcadores más utilizados en oncología para la detección del cáncer de próstata (CP). El test de PSA consiste en el seguimiento de los niveles de PSA en suero, los cuales suelen ser muy bajos en condiciones fisiológicas. Si la concentración de PSA aumentara, esto podría indicar que el individuo padece CP. Sin embargo, los niveles de PSA también se han visto incrementados en otras enfermedades como en hiperplasia prostática benigna (HPB). Además, el ensayo de PSA no puede distinguir los tumores no agresivos de los agresivos. Para poder resolver esta falta de especificidad, se ha empezado a analizar el patrón de glicosilación del PSA, entre otras características bioquímicas, el cual se muestra alterado en cáncer. Las áreas de "Glycomics y Glycoproteomics", están desarrollando técnicas cada vez más sensibles y automatizadas para la identificación y cuantificación de anomalías específicas de glicosilación en una muestra biológica. Este trabajo se centra en una amplia revisión bibliográfica de diversas metodologías utilizadas para mejorar la especificidad del PSA, así como en la búsqueda de bases de datos y el uso de herramientas bioinformáticas, que son un elemento clave para el análisis y la caracterización del PSA. Asimismo, tres estrategias distintas - ELLA, CE-ESI-MS y MALDI-TOF-MS - han sido analizadas y comparadas, resaltando sus ventajas y limitaciones. Mediante la aplicación de diferentes estrategias, se ha podido obtener una gran cantidad de datos sobre la composición heterogénea de los glicanos del PSA, como también se han encontrado alteraciones significativas en el patrón de glicosilación, al comparar muestras de PSA de individuos sanos o con HPB con pacientes con CP. Estos resultados sugieren que el patrón de glicosilación del PSA podría aumentar la especificidad y sensibilidad del diagnóstico de CP. Sin embargo, harán falta más esfuerzos en cuanto al desarrollo de nuevas tecnologías más sensibles, para poder considerar la glicosilación del PSA como el nuevo marcador estándar.

TABLE OF CONTENTS

1. INTRODUCTION....2
1.1 Cancer biomarkers....2
1.2 Glycomics and cancer....2
1.2.1 Glycobiology in cancer....3
1.2.2 Glycoprotein N- and O-glycans....4
1) -N-glycosylation....4
2) -O-glycosylation....5
1.2.3 Common glycosylation alterations in cancer....6
2. GLYCOSYLATION OF PROSTATE SPECIFIC ANTIGEN (PSA) AS A PROSTATE CANCER BIOMARKER....8
2.1 Biology of PSA....8
2.2 PSA and prostate cancer....9
2.3 PSA serum assay: a controversial topic....9
2.4 Targeting the PSA glycosylation pattern....10
2.4.1 PSA glycosylation in normal and malignant cells....11
3. OBJECTIVES....11
4. METHODOLOGY....11
5. RESULTS AND DISCUSSION....12
5.1 Bioinformatics search....12
5.2 Approaches to study PSA glycosylation....17
5.2.1 Methods used to determine PSA glycosylation....17
1) Lectin based-detection methods....17
2) Mass spectrometry-based methods....19
3) 2-Dimensional electrophoresis....20
4) Capillary electrophoresis....20
5) High-Performance Liquid Chromatography (HPLC)....21
5.2.2 Analysis of the glycoproteins....22
1) Analysis of the intact PSA....22
2) Analysis of the intact glycopeptides....23
6. ETHICS AND SUSTAINABILITY....28
7. CONCLUSIONS....28
8. BIBLIOGRAHY....29

1. INTRODUCTION

1.1 Cancer biomarkers

For the last three decades, studies in the field of biomarkers in cancer have become of special interest to many researchers, aiming to reduce the amount of deaths (Sivastrava and Gopal-Sivastrava, 2002). In fact, a tremendous amount of data about tumor markers is being published daily, because of the fast-growing advances in molecular technologies and the increasing knowledge in biology of cancer (Henry and Hayes, 2012).

Regarding the Nacional Cancer Institute (NCI), the concept of biomarker can be described 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", for example, cancer. There is a wide variety of biomarkers, including, among others: proteins, nucleic acids, antibodies and peptides (Henry and Hayes, 2012). Most recently, scientists have started to look at the outcome of a collection of alterations (e.g., gene expression, proteomic and metabolomic signatures), as potential cancer biomarkers (My Cancer, 2017). The use of these biological markers provides a huge advantage in the field of oncology, due to their many potential applications. These include the estimation of risk of developing cancer, screening, differential diagnosis, determination of prognosis, prediction of response to treatment and monitoring the progression of disease (Henry and Hayes, 2012). The fact that biomarkers could be used as a tool for early detection of cancer, is what most encourages scientists to focus on developing new approaches for biomarker improvement and discovery. Certainly, the costs in health care could be significantly reduced by identifying the tumor at its early stages as well as providing noninvasive, sensitive and reliable assays (Sivastrava and Gopal-Sivastrava, 2002). However, its development and further application in clinical uses presents a big challenge. For instance, benign illnesses can also be the cause by which the concentration of certain biomarkers might be altered in some body fluids (National Cancer Institute, 2017). Therefore, they also lead to an increase in the number of false-positive results and overtreatments (Sivastrava and Gopal-Sivastrava, 2002). Moreover, not all individuals with the same type of cancer will show high levels for a specific biomarker (National Cancer Institute, 2017). For these reasons, and because of the critical role they play at all stages of cancer, it is of great importance that biomarkers undergo rigorous evaluation prior to being applied in a clinical context (Henry and Hayes, 2012).

1.2 Glycomics and cancer

Triggered by the purpose of improving and discovering biological markers for the early detection of cancer, many improvements have been achieved in the discipline of proteomics, analytical and computational technologies (Pan et al., 2011). Thus, providing useful tools to increase the specificity and sensitivity that every reliable biomarker should have (Sivastrava and Gopal-Sivastrava, 2002). Currently, several scientists have started to develop a particular interest for the rapid emerging **glycomic** and **glycoproteomic** approaches, which are considered to be of high and clinical relevance (Pan et al., 2011). Indeed, these fields provide crucial information about the enrollment that **glycans** (carbohydrate chains) play in the biology of cells, both in physiological and pathological conditions, such as cancer (Varki A., 2017).

1.2.1 Glycobiology in cancer

Glycobiology is an arising area of biochemistry which encompasses the multiple functions of sugars attached to proteins and lipids (Taylor and Drickamer, 2011). Carbohydrates play an important role in the assembly of complex multicellular organs and organisms. Indeed, interactions between cells and the surrounding matrix are required to ensure an appropriate development of the human body (Varki et al., 2009).

According to Nilsson J. *et al.*, **glycosylation** is "a non-template driven enzymatic process that allows for glycan attachment, trimming, chain elongation and branching as well as glycan derivatization". The end products are known as **glycoconjugates** and are basically **glycosphingolipids, proteoglycans/glycosaminoglycans (GAGs), glycosylphosphatidylinositol** and **glycoproteins**, which are the most abundant (Ho et al., 2016). In fact, it is estimated that more than 50% of all the proteins of the human body are glycosylated (Zhang and Williamson, 2005). For instance, many membrane-bound and secretory proteins produced in mammalian cells have covalently linked glycans that enclose various structures (Pan et al., 2011). What is more, glycoproteins are known to participate in key regulatory events for cellular and organ homeostasis (Meretier et al., 2016).

Glycosylation in proteins is a common form of protein post-translational modification (PTM), by which carbohydrate chains have been attached covalently to the polypeptide side-chain (Geyer and Geyer, 2006). The glycosylation pattern of a determinate glycoprotein is highly specific at each glycosylation site and usually constant for cell type and physiological condition (Wittmann, 2008). This post-translational modification is extremely important, as it determines the processing, distribution, metabolism and biological functions of most protein molecules. In particular, glycans facilitate the appropriate folding of proteins, confer protease resistance and solubility, and also serve as ligands for carbohydrate binding proteins (Geyer and Geyer, 2006). Nevertheless, the glycosylation form of a protein can be notably altered due to changes in cellular pathways as well as in processes resulting from diseases (Varki et al., 2009). Specifically, alterations in glycoproteins can occur because the glycosylation pattern of the glycan moiety is altered or that glycosylation sites are either hypo, hyper or newly glycosylated (Pan et al., 2011).

Previous studies in the field of cancer have proved the existence of a strong correlation between the aberrant glycoproteins and this specific illness. Namely, because of the associations found between cancer and the differential expressions of the enzymes involved in glycosylation, i.e. **glycosidases** and **glycosyltransferases**, aberrant glycosylation is being considered as a hallmark of cancer (Wittmann, 2008). In fact, alterations in glycan processing can interfere with normal molecular functions, for instance, in cell-cell recognition, cell signaling and adhesion. Thus, leading to the acquisition of malignant features such as invasion and metastasis (Meretier et al., 2016). Consequently, glycoproteins are becoming potential targets for biomarker improvement and discovery (Wittmann, 2008). **[Figure 1]**

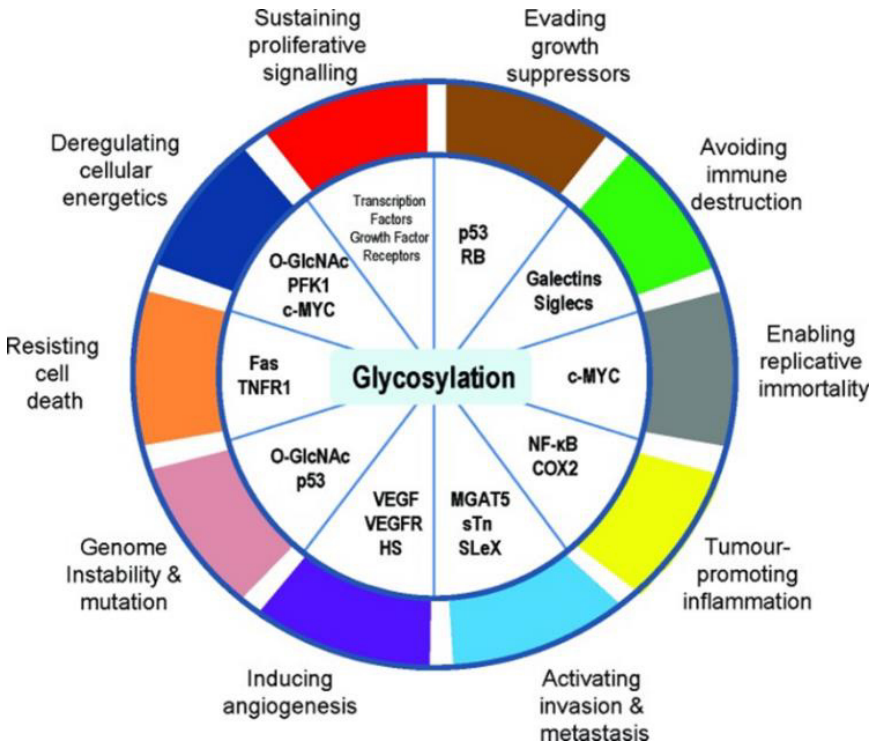


Figure 1. Hallmarks of glycosylation in cancer. Jennifer Munkley1 and David J. Elliott1. Oncotarget. 2016 June 7; 7(23): 35478–35489.

Glycomic and glycoproteomic analysis, however, are complicated due to the large diversity and heterogeneity of carbohydrates as well as the complexity of the linkage that bounds the glycans to the protein (Geyer and Geyer, 2006, Nilsson et al., 2012).

1.2.2 Glycoprotein N- and O-glycans

The process of glycosylation can take place at different amino acid residues in the polypeptide sequence (Varki et al., 2009). There are two main forms of protein glycosylation commonly known as N- and O-linked glycosylations (Nilsson et al., 2012). Usually, both types coexist in the same protein (Ho et al., 2016).

1) -N-glycosylation

The biosynthesis and maturation of N-glycan structures are defined by a sophisticated interplay of several glycosidases and glycosyltransferases (Ho et al., 2016). In N-glycosylation, the glycans are bound to the amide nitrogens of asparagine residues which occur within an **Asn-X-Ser/Thr** motif (X can be any amino acid in exception of Proline) (Nilsson et al., 2012, Pan et al., 2011). In animals, the sugar attached to the Asp residue is mainly an **N-acetylglucosamine (GlcNAc)** and the linkage is in **β configuration**, which can be abbreviated as **GlcNAcβ1-Asn**. In general, N-glycans contain a common pentasaccharide core structure as all of them share the same biosynthetic pathway. (Geyer and Geyer, 2006, Nilsson et al., 2012, Taylor and Drickamer, 2011). This structure consists of 2 Mannose and 3 GlcNAc units, but vary in the terminal elaborations that extend from this core (Taylor and Drickamer, 2011). Three classes of N-glycoproteins can be distinguished depending on the monosaccharides attached (Geyer and Geyer, 2006): **[Figure 2]**

A) Oligomannosidic or high mannose oligosaccharides: these are the simplest structures composed by 2 units of GlcNAc, 5 to 9 mannose units and sometimes can incorporate Galactose residues (Taylor and Drickamer, 2011).

B) Complexed-type: complex glycans are built on a core which consists of 3 mannose residues and 2 GlcNAc units linked to the glycoprotein. However, in many occasions re-elongation of the core structure can take place by adding more GlcNAc residues to the core, which is commonly referred as **branching** and is carried out by distinct GlcNAc-transferases such as **GlcNAc-transferase IV, V or VI**, among others. The addition of these GlcNAc units leads to a defined branch structure that can be classified depending on the number of **branches** or **antennas** that extend from the core. The most prevalent type is the **bi-antennary** complex glycan, although **tri-antennary** and **tetra-antennary** structures are also quite common. Furthermore, the branched structures are frequently extended by the addition of a single **Galactose** and a **Sialic acid** residue to one or both terminal GlcNAc units and it is catalyzed by a **galactosyltransferase** and the **sialyltransferases**, respectively. Besides the branching, other variations are seen in the terminal residues as well as on the core structure of different complex N-glycans. For instance, the glycans can also be modified by the addition of a **bisecting GlcNAc** residue to the 4^{th} position of the core mannose by **GlcNAc-transferase-III.** Moreover, a **core fucose** residue can be linked to the GlcNAc unit that is directly attached to the Asn residue of the protein (Taylor and Drickamer, 2011). Noteworthy, the complex N-glycans display the largest structural diversity (Meretier et al., 2016).

C) Hybrid-type: the resulting combination of the features of both oligomannosidic and complexed-type species (Meretier et al., 2016).

2) -O-glycosylation

In contrast to N-glycosylation, the groups of O-linked glycans are built on distinct protein glycan linkages, in which GalNAc, Man, xylose or Gal can be bound to the hydroxyl group on serine or threonine residues of the protein **[Figure 2].** Thus, indicating that O-linked glycans are notably diverse in both structure and function (Taylor and Drickamer, 2011). The frequency of O-glycosylation is relatively high, in particular, on the secreted or membrane bound mucins that are known to be rich in serine and threonine (Ho et al., 2016). The mucin-type of O-glycosylation is the most prevalent form and it consist of a large glycoprotein that carries many O-glycans which are α-glycosidically linked via **N-acetylgalactosamine (GalNAc)** to the amino acid residues (Varki et al., 2009). These type of glycans do not share a different core structure, although they encompass several different core regions with common motifs (Geyer and Geyer, 2006, Taylor and Drickamer, 2011).

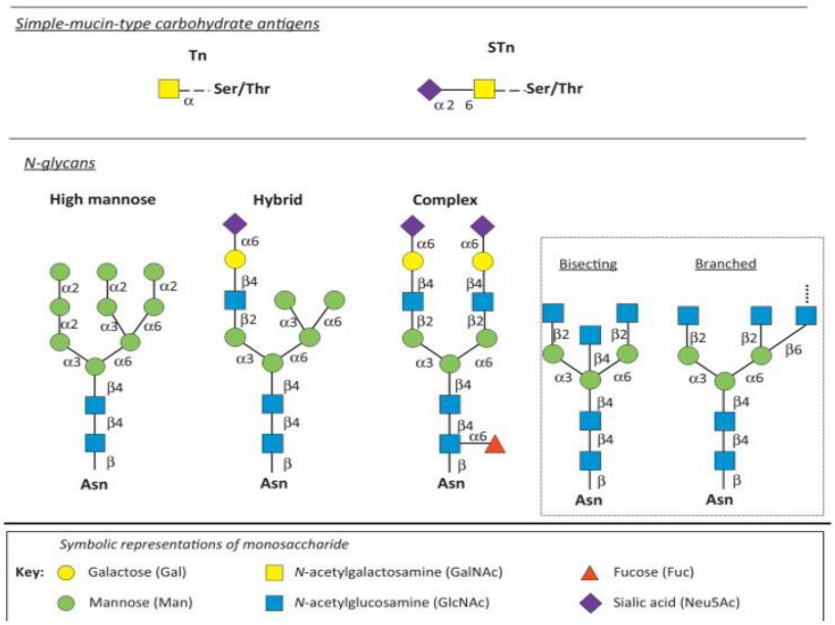


Figure 2. Example of the structures of simple-mucin-type carbohydrate antigens (Tn and its sialylated form STn), and the main N-glycosylation structures (High, mannose, Hybrid and Complex), as well as the two main terminal variations that occur in N-glycans (Bisecting and Branching). Pinho et al. Trends in Molecular Medicine. 2013 August 8; 19(11): 664 – 676.

1.2.3 Common glycosylation alterations in cancer

Glycosylation is the most complex PTM of proteins and has shown to be involved in many physiological events, for instance, cell differentiation and trafficking, host-pathogen interaction as well as cell signaling. In cancer, altered glycan structures have demonstrated to mediate critical events in tumor pathogenesis and progression (Ho et al., 2016). Previous structural studies involving the detection of glycoproteins with specific antibodies, confirmed that the significant differences on antibody reactivity between a tumor and a normal tissue, were due to specific changes in the carbohydrate structures present in the cell surface (Taylor and Drickamer, 2011). The diversity of the glycans expressed on the cell surface is controlled in part by the expression, localization and the activity of the glycosyltransferases in a given cell. Consequently, minor alterations in the enzyme expression or function can notably affect the types of carbohydrate modifications in different ways. Moreover, the biological activity of the cell can be altered as well (Stowell et al., 2015). Tumor-associated glycan modifications encompass increased branching of N-glycans, higher density of O-glycans, generation of truncated versions of normal counterpart and generation of unusual glycans forms of terminal structures driving from sialyation and fucosylation (Ho et al., 2016). **[Figure 3]**

Two main structural features of the complex N-glycan type are the β1,6-branching and the bisecting-GlcNAc, which are catalyzed by the glycosyltransferases **GnT-V** and **GnT-III**, respectively (Ho et al., 2016). In some cancers the GnT-V has shown to be upregulated. As a consequence, it has led to an increase in the branching of the N-glycans as well as contributed to cancer cell invasion and metastases. On the other hand, the bisecting structures that are related to the suppression of cancer metastasis, are decreased during caner progression (Meretier et al., 2016).

Another common feature observed in cancer is the overexpression and exposure of truncated O-glycans. Increased expression of these shortened O-glycan structures, which include Tn antigen (GalNAcα1-Ser/Thr), Thomsen-Friedenreich T antigen (Galβ1-3GalNAcα1-Ser/Thr) and their

respective sialylated forms, STn (Neu5Acα2-6GalNAcα1-Ser/Thr) and ST (Neu5Acα2-3Galβ1-3GalNAcα1-Ser/Thr), have been associated to malignant transformation, cancer growth and metastatic ability (Meretier et al., 2016). Indeed, the impairment of the glycosyltransferases involved in the synthesis of core structures may result in increased expression of incomplete O-glycans (Ho et al., 2016). Furthermore, increased O-GlcNAcylation and the modification of proteins with O-GlcNAc have demonstrated to be enrolled in key regulatory roles in tumor cell signaling (Meretier et al., 2016).

In addition to alterations in core glycans, each of these carbohydrates can be further modified to generate unique terminal glycan motifs that may also undergo specific changes following neoplastic transformation. These changes can be due to sialyation or fucosylation (Stowell et al., 2015). Sialylation is a glycosylation feature which occurs in different linkages at the non-reducing ends of a glycan moiety (Kammeijer et al., 2017). An increase in the global sialylation, has been closely related to metastatic tumor due to the altered glycosyltransferases expression as well as to the increased branching of complex N-linked glycans. (Meretier et al., 2016 and Taylor and Drickamer, 2011). The major α2,3-silaylated antigens associated to cancer are, in fact, **Sialyl-Lewis x/a (SLe^a and SLe^x),** which are trisaccharide antigens expressed on many cell-surface glycoproteins. Moreover, glycans that have been added more fucose sugar units, such as **Lewis antigens [Lewis a/b (Le $^{a/b}$)** and **Lewis x/y (Le $^{x/y}$)],** have also shown to be related to cancer and inflammation (Ho et al., 2016).

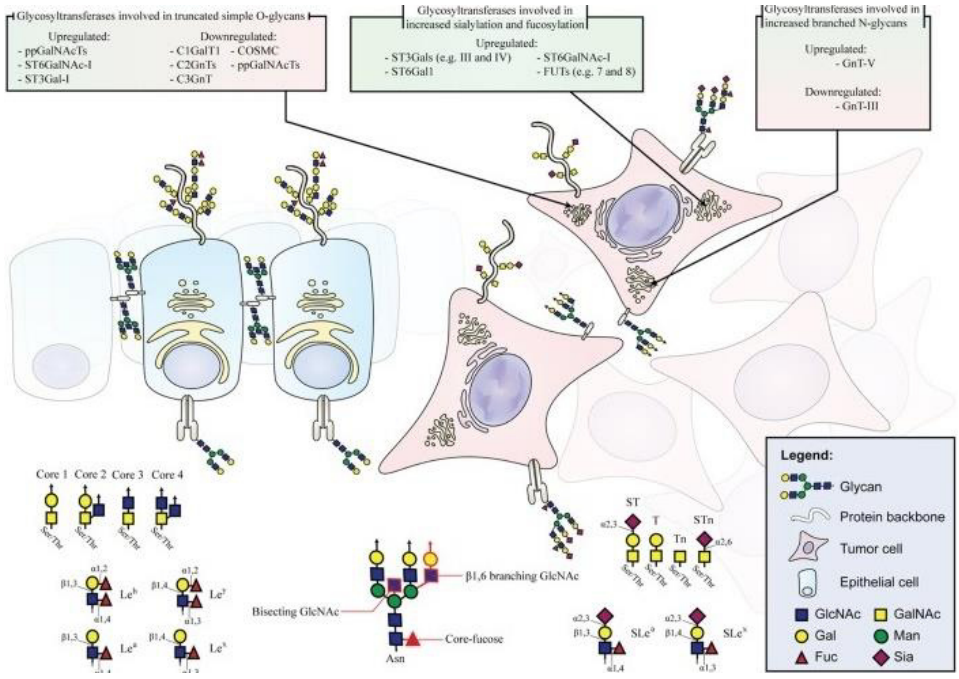


Figure 3. Principal glycan alterations in cancer. Depiction of a healthy epithelium, which displays cell-polarity, organized glycoprotein localization and normal glycosylation pattern, against malignant cells with miss-localized glycoproteins that present an aberrant expression of their glycan moieties. Designation of the different glycosyltransferases involved in truncated simple O-glycans, increased sialyation and fucosylation, as well as increased branched N-glycans is included as well. Meretier et al. Frontiers in Oncology. 2016 March 9; 6(55):1-19.

Taken together, these diverse alterations in the glycosylation of proteins may indicate that, by analyzing and comparing the glycosylation pattern of different glycoproteins driven from healthy and cancer patients, potential biomarkers of cancer could be improved and even discovered. In fact,

many serum tumor markers used currently are glycoproteins or altered tumor-associated carbohydrate antigens. Some of these biomarkers are included in **Table 1**.

Table 1. List of several FDA-approved cancer biomarkers currently used in clinical practice. The type of cancer, detection type and the year in which the markers were approved are listed as well. Prostate specific antigen is framed in blue. Modified from: Kirwan et al. BioMed Research International. 2015 May 31; 2015(ID 490531): 1-16

Marker	Full name	Cancer types	Detection method	Year of FDA approval
AFP	α-Fetoprotein	Liver Protein	concentrations and core fucosylation	1992/2008
PSA, Pro2PSA	Prostate-specific antigen	Prostate	Protein concentrations	1986/1994/ 2012
CA125 (MUC16)	Cancer antigen 125	Ovarian	Protein concentrations	1997/2011
CA15-3 (MUC1)	Cancer antigen 15-3	Breast	Monitoring therapy	1997
CA19-9	Carbohydrate antigen 19-9 or cancer antigen 19-9	Pancreatic, ovarian	Monitoring therapy	2002
Tg	Thyroglobulin	Thyroid	Protein concentrations	1997

Among the different glycoproteins used as cancer biomarkers, in this work I have focus on reviewing the different studies that have applied glycomic and glycoproteomic approaches to improve the prostate-specific antigen biomarker, extensively used for the diagnosis of prostate cancer (Gilgunn et al., 2013).

2. GLYCOSYLATION OF PROSTATE SPECIFIC ANTIGEN (PSA) AS A PROSTATE CANCER BIOMARKER

2.1 Biology of PSA

PSA stands for Prostate-Specific Antigen and still remains as the most important biomarker used as a screening test for the diagnosis of prostate cancer among male population, as well as for assessing responses to treatment (Schröder, 2009). PSA is also known as glandular kallikrein-3 (Kh-3) and is encoded by KLK3 gene located in chromosome 19. (Uniprot.org, 2017). This molecule is an androgen-regulated glycoprotein, more accurately, a serine protease produced almost uniquely by the prostatic epithelial cells. Among the extensive variety of prostate-driving proteins found in the seminal fluid, PSA is indeed the most abundant (Wittmann, 2008). Protein-specific antigen has various molecular functions, which are **endopeptidase**, **hydrolase** and **serine-type protease**. The main biological process in which this enzyme is involved is the dissolution of the seminal fluid coagulum (Uniprot.org, 2017).

In serum, the **total circulating PSA (tPSA)** can be found in multiple molecular forms. Approximately an 84% of tPSA is covalently linked to protease inhibitors, which have rapidly bound to the secreted active form of PSA, and is known as **complexed PSA (cPSA)** (Gilgunn et al., 2013). The most predominant enzymatically inactive complex consists of an active PSA molecule that is bound to the inhibitor **α-1-antichymotrypsin (PSA-ACT)**. On the other hand, a relatively smaller part of the tPSA circulates as **freePSA (fPSA).** The majority of freePSA exists as at least 3 different forms: **proPSA (pPSA), BPSA** and **inactive PSA (iPSA)**. Currently, BPSA and a truncated subform of proPSA **(i.e., [-2] proPSA)** are being used as indicators of prostate benign hyperplasia **(BPH)** and prostate cancer, respectively (Adhyam and Gupta, 2012, Filella and Foj, 2015).

2.2 PSA and prostate cancer

In the recent years, prostate cancer (PCa) has increased its incidence and mortality. Indeed, PCa has been the third leading cause of all cancer-related deaths in men in EU, and the second most common worldwide (Gilgunn et al., 2013, Pihikova et al., 2016). **[Table 2]**

Table 2. Cancer incidence and mortality Worldwide. Prostate cancer is remarked in red. IARC CancerBase No.11[Online]; assessed on 2015 January 16.

Rank	Cancer	New cases diagnosed in 2012 (1,000s)	Per cent of all cancers (excl. non-melanoma skin cancer)
1	Lung	1,242	16.7
2	Prostate	1,112	15.0
3	Colorectum	746	10.0
4	Stomach	631	8.5
5	Liver	554	7.5
6	Bladder	330	4.4
7	Oesophagus	323	4.3
8	Non-Hodgkin lymphoma	218	2.9
9	Kidney	214	2.9
10	Leukaemia	201	2.7

Prostate specific antigen is produced at high concentrations by both, normal and malignant prostatic epithelium. However, under physiological conditions the prostate architecture confines PSA to the gland and only minor amounts of PSA leak out into the circulation (Gilgunn et al., 2013, Stenman et al., 1999).

Unlike men who have a normal prostate, the amount of freePSA in serum is notably decreased in PCa, whereas levels of complexed PSA are shown to be higher in cancer, because of the disruption to the ductal lumen and the perturbation of the basement membrane. Consequently, the total amount of PSA is increased under these circumstances (Gilgunn et al., 2013). Indeed, the difference in the absolute value of PSA between cancer and physiological conditions, is what turns PSA into a considerably useful biomarker for determining the extent of prostate cancer and assessing the response to treatment. However, under certain biological conditions, mostly in benign prostate hyperplasia, the disruption of the prostate also takes place. Hence, despite the fact that it is also commonly used as screening method for early detection of cancer, PSA test appears to be somehow controversial (Adhyam and Gupta, 2012).

2.3 PSA serum assay: a controversial topic

The Prostate-specific antigen test was first approved by the US Food and Drug Administration (FDA) in 1986 (Report.nih.gov, 2017). Since then, the scenario of PCa screening and diagnosis completely changed. In fact, it was in the early 1990s when the first evidence for potential effectiveness in the early detection of prostate cancer came out (Schröder, 2009). Until now, no other test besides PSA has proved to be more effective in the identification of the initial stages of prostate cancer (Report.nih.gov, 2017). Indeed, its use as a screening test has been prevalent worldwide with rates, such as, 70% in the US and 20 to 40% in Europe. (Schröder, 2009).

Over the years, it has become clear that PSA offers various diagnostic advantages when PSA levels are higher than 10ng/mL serum. However, is the detection of PSA levels within a range of 2.5-10ng/mL in serum, also known as the *grey zone*, what results controversial. The main problem is that

there are many other factors affecting the normal PSA levels in sera besides cancer, such as: age, race, body mass index (BMI), additional medication and other benign conditions, for instance, prostatitis and BPH (Adhyam and Gupta, 2012). These, in turn, make it extremely difficult to certainly affirm whether an individual with a level of PSA within the grey zone, truly has prostate cancer. Actually, it is estimated that only 25% of the men who undergo biopsy due to the elevated PSA levels (above 4 ng/mL), have PCa (Filella and Foj, 2015). Moreover, PSA test is still not able to distinguish between aggressive and non-aggressive cancers (Gilgunn et al., 2013). Consequently, there is an increase in the number of individuals that are being treated when is not necessary (Filella and Foj, 2015). The over usage of aggressive treatments, not only raises the costs but also supposes a potential harm for the patient's health, including urinary incontinence and erectile dysfunction (Adhyam and Gupta, 2012).

Currently, the most accepted standard cut-off is 4.0 ngPSA/mL serum because it balances the trade-off between missing relevant cancers at a curable stage and avoiding detection of clinically insignificant disease. However, the American Cancer Society estimated that the sensitivity at this threshold value was 21% for detecting any prostate cancer and 51% for detecting high-grade cancers. In addition, some high-risk cancers must be dismissed as the PSA levels of some individuals can be lower than 4ngPSA/mL (Adhyam and Gupta, 2012).

Aiming to reduce the amount of false positives and increase PSA specificity and sensitivity, a lot of efforts have been put on studying different strategies that could improve the interpretation of PSA levels (Adhyam and Gupta, 2012). Some of them, consisting in the use of derived values, for example, PSA density, PSA velocity, the ratio Free PSA/Total PSA, Prostate cancer gene 3 (PCA3) and the PHI index (Filella and Foj, 2015). Certainly, these refinements have helped to detect more cancers, although data on increased survival has not been the one expected so far (Adhyam and Gupta, 2012).

2.4 Targeting the PSA glycosylation pattern

The fact that aberrant glycosylation is considered as a fundamental characteristic of tumor genesis has enhanced many researches to improve and develop new tools for a more detailed analysis of the glycome (Meany and Chan., 2011). In particular, the study of PSA glycosylation pattern has been and, still reminds, as the central stage to increase the specificity and sensitivity of this particular biomarker (Stenman et al., 1999). Indeed, by monitoring the cancer-associated glycoforms of PSA, a high progress could be achieved (Gilgunn et al., 2013).

Nevertheless, a complete glycoproteome analysis is quite complicated. In contrast to the biosynthesis of nucleic acids and proteins, the synthesis of glycans is a non-template-driving process, which encompasses the coordinated action of various enzymes and lacks any proofreading machinery. As a consequence, the structural diversity of the glycans becomes quite large (Geyer and Geyer, 2006, Nilsson et al., 2012). Therefore, to determine which glycomic and/or glycoproteomic approaches are the most suitable for glycan and glycoprotein characterization, some aspects about the glycosylation in PSA and the type of glycan linked, should be clarified in the first place.

2.4.1 PSA glycosylation in normal and malignant cells

The PSA extracted from the seminal fluid of healthy individuals has been characterized. The glycoprotein consists of a complex biantennary N-linked oligosaccharide, which is commonly di-sialylated rather than mono-silaylated, and incorporates frequently a core-fucose (White et al., 2009, Gilgunn et al., 2013). **[Figure 4]**

Studies focused on studying PSA glycans, which were driven from tumor sources significantly differ in their content of sialic acid (SA) and fucose. Moreover, analysis of several clinical samples associated to PCa identified more tri- and tetra-antennary glycan species (White et al., 2009).

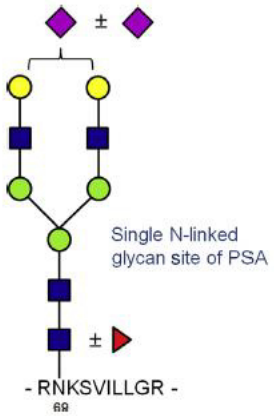


Figure 4. Most common structure of PSA. Glycans are represented as GlcNAc (blue square), Mannose (green circle), Gal (yellow circle), Sialic acid (purple diamond) and Fucose (red triangle). Drake et al. ScienceDirect. 2015 February 7; 126: 345-382.

3. OBJECTIVES

This project aims to:

- Review the principle glycomic and glycoproteomic strategies and methodologies used in the recent years to improve the specificity and sensitivity of PSA, in order to provide a more reliable test for the early detection of cancer.
- Introduce some bioinformatic concepts, namely, bioinformatic databases and tools, which are currently used and have a great impact on the analysis of protein glycosylation.

4. METHODOLOGY

In this work, I have used various databases and bioinformatic tools to gain more in-detail knowledge about PSA structure, glycosylation site and other information, which may help to improve the analysis of PSA glycosylation pattern. These bioinformatic sources include: UniProtKB, ExPASy (PeptideMass/Cutter and Compute pI/MW), MASCOT, NetNGlyc1.0 Server, NetOGlyc 4.0 Server, GlypID 2.0, Glycome DB and Glycan mass spectral DB**.**

On the other hand, in order to review the principle glycoproteomic strategies and methodologies used to improve PSA specificity and sensitivity, I have focused on a broad literature review about several studies related to the glycosylation and its role in cancer, tumor markers, prostate cancer, PSA and a vast number of set-ups that aimed to improve PSA as a biomarker, among others. Mainly, I have used the database PubMed, through which I was able to obtain a tremendous amount of papers, used few textbooks, such as Introduction to Glycobiology and Essentials of Glycobiology, as well as several webs (My Cancer, CancerBase and NIH, among others) mainly to obtain information about prostate cancer, statistical data and prostate specific antigen.

5. RESULTS AND DISCUSSION

5.1 Bioinformatics search

In the recent years, the field of **Bioinformatics** has had a tremendous impact in the research community, by providing methods and software tools to understand complex biological data. From the glycomics and glycoproteomics point of view, bioinformatics approaches play an important role in automated analysis and interpretation of the large amount of experimental data generated throughout the application of high through-put techniques (Encyclopedia Britannica, 2017). Moreover, by looking at different databases, information about the protein sequence, chemical and structural characteristics, and glycosylation sites, among other features, might be easily acquired and further used for defining certain criteria of the experimental design. Taking advantage of this and the fact that many of the bioinformatic tools are freely accessible, I used some databases and searched for few analytical tools, through which a lot of information about different features of PSA has been obtained and will be discussed hereafter.

One of the sources used in this work is the **Universal Protein resource web site**, known as **UniProtKB**. UniProt contains the protein sequence of the prostate-specific antigen, which is in turn needed to gain knowledge about its potential glycosylation sites and other structural information from other databases. As it can be observed in the image below, much more information related to PSA can be found in this web site. For instance, the principal functions that this glycoprotein displays, the biological processes in which it is involved and its interaction with other proteins as well as about the sequences of several isoforms of PSA. **[Figure 5]**

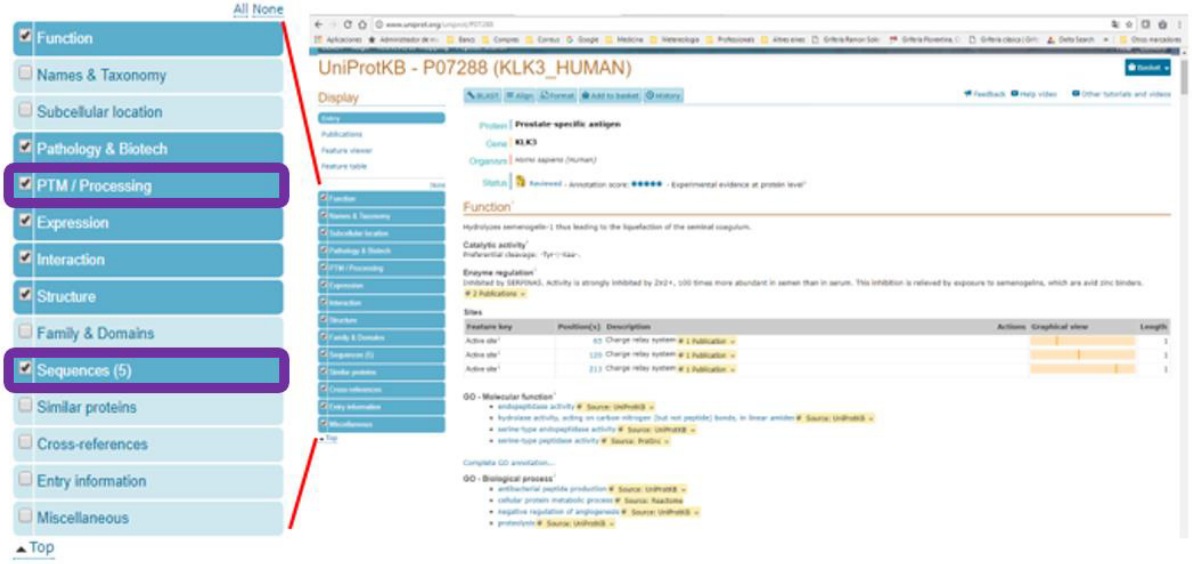
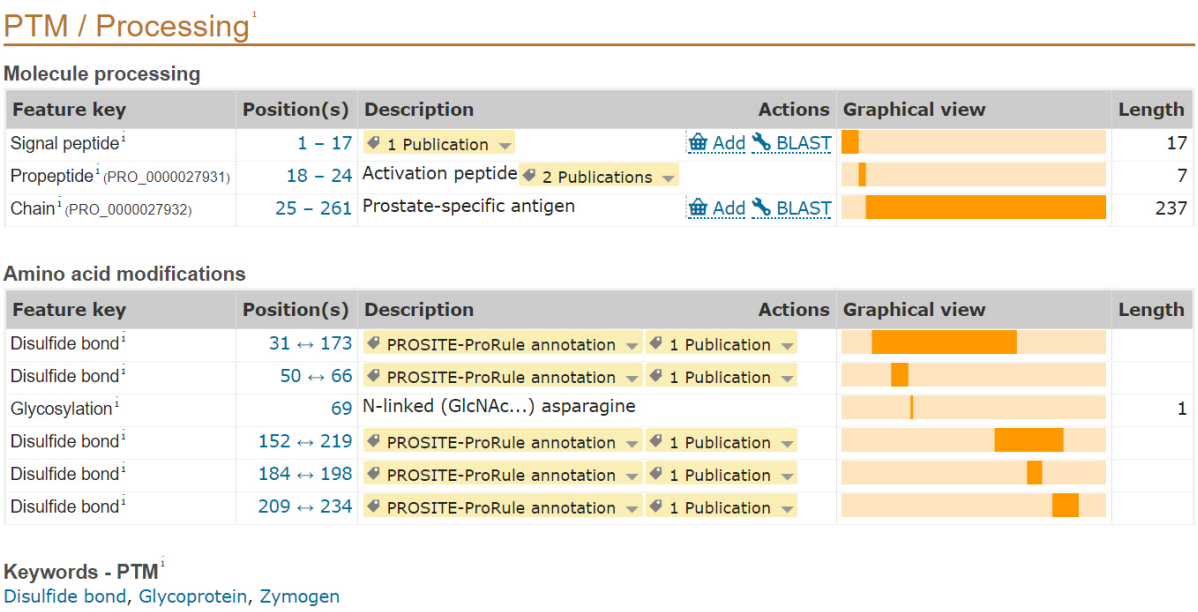


Figure 5. Screen shot of UniProtKB. It shows the entry number of PSA (P07288) and the different information regarding PSA function. The list of the different sections has been amplified and, in particular, the sections PTM/Processing and Sequences (5) have been circled in purple.

Aiming to earn more detailed information about PSA glycosylation, I went through the **PTM/Processing** and **Sequences** sections. **[Figure 6/7]**



PTM / Processing

Molecule processing

Feature key	Position(s)	Description	Actions	Graphical view	Length
Signal peptide	1 – 17	1 Publication	Add BLAST		17
Propeptide (PRO_0000027931)	18 – 24	Activation peptide 2 Publications			7
Chain (PRO_0000027932)	25 – 261	Prostate-specific antigen	Add BLAST		237

Amino acid modifications

Feature key	Position(s)	Description	Actions	Graphical view	Length
Disulfide bond	31 ↔ 173	PROSITE-ProRule annotation 1 Publication			
Disulfide bond	50 ↔ 66	PROSITE-ProRule annotation 1 Publication			
Glycosylation	69	N-linked (GlcNAc...) asparagine			1
Disulfide bond	152 ↔ 219	PROSITE-ProRule annotation 1 Publication			
Disulfide bond	184 ↔ 198	PROSITE-ProRule annotation 1 Publication			
Disulfide bond	209 ↔ 234	PROSITE-ProRule annotation 1 Publication			

Keywords - PTM

Disulfide bond, Glycoprotein, Zymogen

Figure 6. Screen shot of the PTM/Processing section found in PSA's UniProt entry. Several key features about the Molecule processing and Amino acid modifications are listed and described.

If little attention is paid to the figure above, it can be noticed that the whole molecule of PSA has a total length of 261 amino acids. Nonetheless, the real length of the polypeptide sequence is 237. Indeed, from the 1st until the 17th amino acid, the sequence corresponds to the peptide signal attached to the protein, whereas the sequence regarding the propeptide is found between the 18th and the 24th amino acid. Under physiological conditions, PSA is synthesized as a **proenzyme (proPSA)** of 244 amino acids by the secretory cells of the acini (i.e. prostate glands) and later secreted into the lumen. In this compartment, PSA goes through a post-translational modification that yields to a 28KDa mature active glycoprotein (Gilgunn et al., 2013, Schröder, 2009, Adhyam and Gupta, 2012). This modification comprises the removal of 7 residues (18-24) and is performed by the **human glandular kallikrein (hK-2)** (Balk et al., 2003). In some cases, a partially cleavage of the proPSA can occur and, consequently, leading to truncated forms of proPSA (Gilgunn et al., 2013, Adhyam and Gupta, 2012).

Regarding the post translational modifications, five different disulfide bonds have been identified in PSA as well as one N-glycosylation at the asparagine 69. Taken together, all this data results to be considerably useful when the protocols of an experimental set-up must be stablished, for example, to decide how a sample of PSA should be prepared or which purification methodologies have to be applied in order to analyze its sequence.

On the other hand, in the Sequences section the whole sequence of PSA and some of its isoforms and their respective sequences are given, as well as few tools like **PeptideMass/Cutter** and **Compute pI/MW [Figure 8, 9,10 and 11].** Each of these bioinformatic tools provide different information. The Compute pI/MW can be applied to know which are the values of the average molecular weight (MW) and the isoelectric point of various sequence fragments, which are already pre-determined or can be selected by the investigator itself. For instance, if the sequence of the mature form of PSA is chosen, the resulting values are given as **Figure 9** shows.

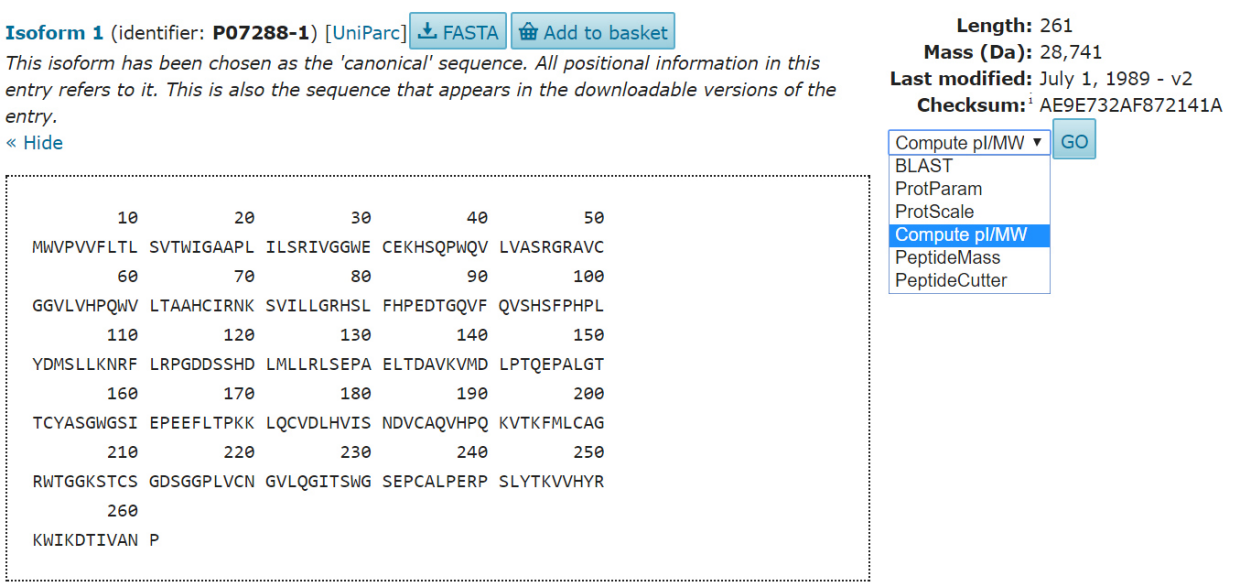


Figure 7. Screen shot of the Sequence section found in PSA's UniProt entry. The whole sequence, including the peptide signal, the pro-peptide and the polypeptide chain, is given for isoform 1. The length and the mass of the protein are designed as well. A list of few bioinformatic tools is displayed and Compute pI/MW is highlighted.

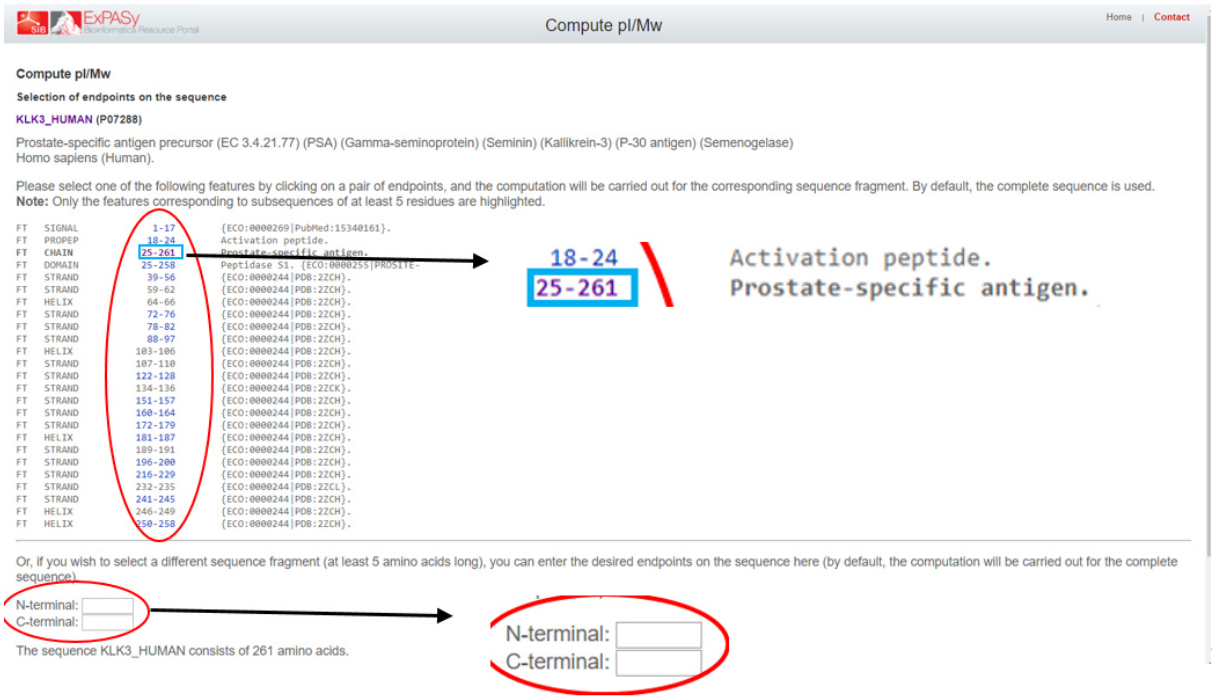


Figure 8. Screen shot of the Compute pI/MW tool. Different fragments of the sequence of isoform 1 are listed and circled in red. The sequence regarding the mature form of PSA (position 25 to 261) has been selected to obtain the values of the isoelectric point and molecular weight average of PSA. Also circled in red, there are the gaps in which the investigator itself can choose the first and last positions of the sequence that needs to be analyzed.

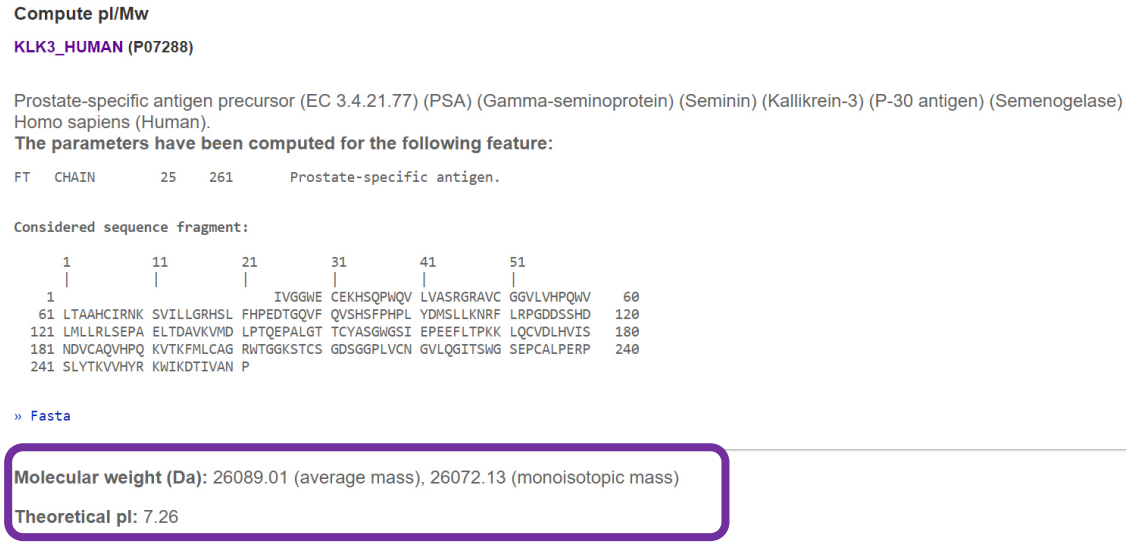


Figure 9. Screen shot about the Compute pI/MW tool, after selecting the prostate specific antigen sequence. The sequence is presented in FASTA format, the theoretical isoelectric point and the molecular weight values are given for PSA and circled in purple.

Otherwise, the PeptideMass/Cutter tools are used to perform an *in-silico* enzymatic digestion of one certain protein (in this case PSA has been used as an example), where the conditions in which the digestion is executed can be chosen. In the case of PeptideMass, for instance, the type of protease, treatment of the cysteines with an alkalylation step to block any free thiols to prevent disulfide bonds formation as well as the maximum and minimum molecular mass of the peptides, can be selected by the investigator as shown in the figure below. In addition, with PeptideMass the values regarding the pI and MW are also provided for the protein of interest. **[Figure 10]**

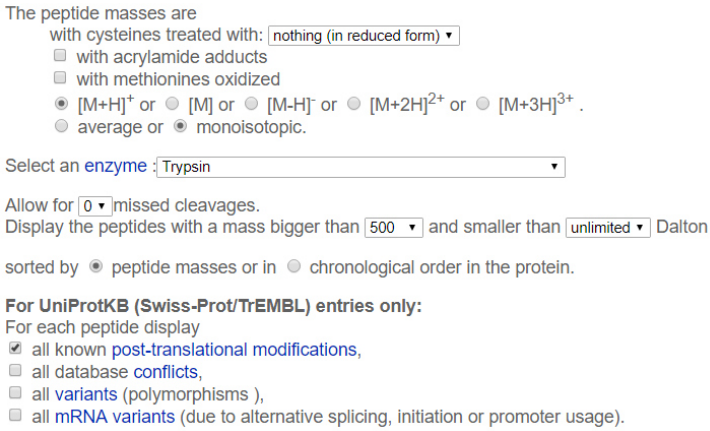


Figure 10. Screen shot of the PepideMass tool. Several options about the conditions, under which the enzymatic digestion will be carried out, are displayed and must be set before executing the in-silico digestion, basing on the experimental set-up.

As an example, I have used PeptideMass to carry out the *in-silico* digestion of PSA using the enzyme trypsin, which has been used in many studies where the glycopeptides of PSA or its glycan structures needed to be analyzed. The resulting digested peptides after the performance of the trypsin *in-silico* digestion of PSA are presented on **Figure 11**. PeptideMass is, thereafter, very useful for comparing the experimental data with the information given by the database, and matching the resulting peaks provided by mass spectrometry with the fragments and their respective molecular weights of the database. The PeptideCutter is very similar, although it gives information about the potential cleavage sites by a given enzyme and protein instead.

You have selected KLK3_HUMAN (P07288) from UniProtKB/Swiss-Prot:

Prostate-specific antigen precursor (EC 3.4.21.77) (PSA) (Gamma-seminoprotein) (Seminin) (Kallikrein-3) (P-30 antigen) (Semenogelase)
Signal and propep in positions 1-24 have been removed.

- Chain Prostate-specific antigen at positions 25 - 261 [Theoretical pI: 7.26 / Mw (average mass): 26089.01 / Mw (monoisotopic mass): 26072.13]

mass	position	#MC	modifications	peptide sequence
3966.8670	207-245	0		STCSGDSGGPLVCNGVLQGI TSWGSEPCALPERPSLYTK
3493.6997	78-107	0		HSLFHPEDTGQVFQVSHSFP HPLYDMSLLK
3467.6385	138-169	0		VMDLPTQEPALGTTCYASGW GSIEPEEFLTPK
2346.1693	171-191	0		LQCVDLHVISNDVCAQVHPQ K
2230.1735	48-68	0		AVCGGVLVHPQWVLTAAHCI R
1871.9432	110-125	0		FLRPGDDSSHDLMLLR
1407.7491	34-45	0		HSQPWQVLVASR
1272.6681	126-137	0		LSEPAELTDAVK
1020.4819	25-33	0		IVGGWECEK
797.3797	195-201	0		FMLCAGR
757.4930	71-77	0		SVILLGR
729.3777	255-261	0		DTIVANP
673.3780	246-250	0		VVHYR
548.2827	202-206	0		WTGGK

94.1% of sequence covered (you may modify the input parameters to display also peptides < 500 Da or > 100000000000 Da):

Figure 11. Screen shot after the in-silico digestion with PeptideMass tool. The different resulting fragments after the digestion of PSA sequence with trypsin and without cysteine treatment are listed.

Comparably to PeptideMass, databases such as **MASCOT** are very practical when proteins are analyzed by mass spectrometry techniques. In fact, with this tool researchers can identify the targeted protein from the resulting peptide fragments obtained.

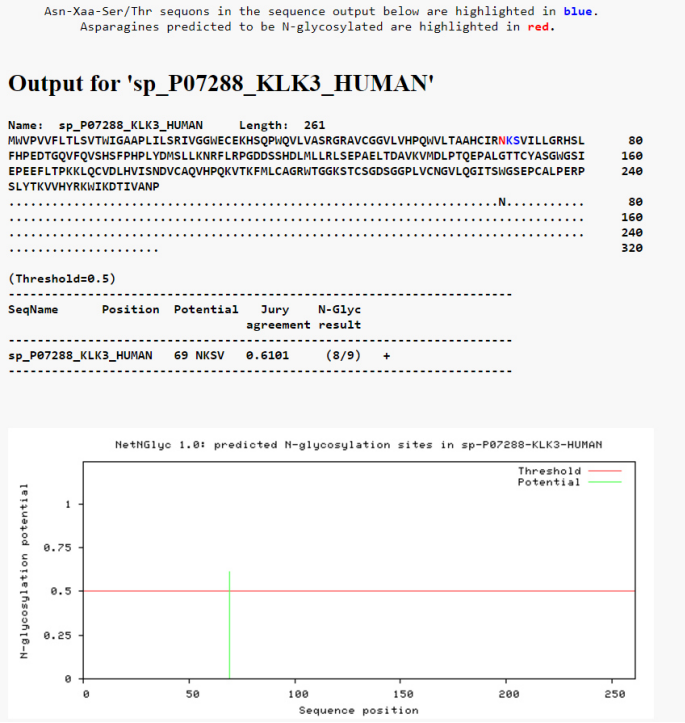


Figure 12. Screen shot after introducing the sequence of isoform 1 in FASTA format at NetNGlyc1.0 Server. One glycosylation site is spotted at Asn 69 with approximately 61% of potential.

Other databases that can be helpful in terms of glycoproteomic research are the **NetNGlyc1.0 Server** and the **NetOGlyc4.0 Server**, because they predict the potential N-glycosylation and O-glycosilation sites, respectively, for a given protein, by just introducing the protein sequence in FASTA format.

In this case, I used both servers to find out the highly potential N- and O-glycosilation sites of PSA. From NetNGlyc server only one N-glycosylaton site was spotted at Asn 69 with 61% of potential **[Figure 12].** On the contrary, no O-glycosylation site was determined, as all potential sites could not reach the threshold, which is required to be considered glycosylated.

Besides the different databases already mentioned, many other web sites such as **GlypID 2.0** can also be helpful, due to its ability to identify glycopeptides from their high energy C-trap and collision induced dissociations as well as the glycans, after the application of tandem mass spectrometry. On the other hand, databases such as Glycome DB and Glycan mass spectral DB may also be efficient, by providing valuable information about the different glycan structures for a given glycoprotein.

Apart from databases, data analysis softwares are also commonly used to enterprise data to identify patterns and stablish relationships (Google+, 2017). Few Data Analyzers have been used in two of the approaches that aimed to analyze PSA glycosylation pattern and which are later discussed in section 6.2. In particular, the softwares **Data Analysis 4.2** and **FlexAnalysis**.

Overall, bioinformatics has highly influenced the scientist-community, as it has helped to analyze and interpret a tremendous amount of data and, therefore, has significantly increased the overall speed average for analyzing biological compounds and obtaining reliable data (Encyclopedia Britannica, 2017). Certainly, bioinformatics is becoming of great importance for the glyco-since community since it has contributed to identify many glycan structures and glycoproteins from complex biological samples. Hence, bioinformatics is a key factor to improve and develop biomarkers of cancer, such as PSA (Baycin Hizal et al., 2014).

5.2 Approaches to study PSA glycosylation

The amount of studies regarding the characterization of PSA has helped to simplify the analysis of this specific glycoprotein, by providing a defined target. Namely, because PSA just presents one N-glycosylation site, it is easier to discern whether a determined approach may be adequate for the structural characterization of the aberrant glycoprotein (White et al., 2009). However, this analysis seems to be quite challenging due to the broad structural diversity of oligosaccharides. This issue is, in fact, encouraging the glyco-science community to keep on developing and applying relevant methods for a detailed characterization of the glycoconjugates. Unfortunately, no single method is yet available for the complete analysis of any glycoconjugate, including the glycosylation pattern in prostate cancer (Nilsson et al., 2012).

The following section consists of an overview of the principle methods that have been used to determine PSA glycosylation pattern.

5.2.1 Methods used to determine PSA glycosylation

In the last decades, several novel techniques have been developed to analyze PSA glycosylation features, such as sialyation, core fucosylation and detailed structure of carbohydrate chain. Among the different techniques used by the glyco-scientist community, there are five general methods that are being predominantly applied in this field. These include: **lectin-based detection methods**, **mass spectrometry (MS)**, **2-dimensional electrophoresis (2-DE)**, **capillary electrophoresis (CE)** and **High-Performance Liquid Chromatography (HPLC)** (Vermassen et al., 2012, Kattla et al., 2011)**.**

1) Lectin based-detection methods

Lectins are plant, fungi, bacterial and animal proteins, which are part of the family of **carbohydrate-binding proteins (CBP)** and do not display any enzymatic properties (Vermassen et al., 2012). Lectins are able to recognize and reversibly bind to mono- or oligosaccharide moieties through complementary sugar-binding sites on polypeptide sequences with high specificity and affinity, without any biochemical modification. Lectins are classified in different groups depending on their affinity towards individual carbohydrate units (Damborský et al., 2017). Some lectin species, however, can selectively bind to various glycan epitopes (Geyer and Geyer, 2006). **[Table 3]**

Many studies focused on analyzing aberrant PSA glycosylation pattern have taken advantage of the fact that, lectins attached to appropriate matrices such as agarose, membranes or magnetic beads; allow the isolation, fractionation as well as the elucidation of glycoproteins according to their glycan structures (Geyer and Geyer, 2006). Specifically, lectins have been mostly used to enrich PSA samples (Meany et al., 2009). Indeed, by using lectins with a wide range of specificities, entire pools of the target glycoprotein can be recovered (Geyer and Geyer, 2006). The principle lectin-based methods developed so far are the **lectin microarray**, **lectin column chromatography**, **lectin affinity electrophoresis** and **enzyme-linked lectin assay (ELLA)** or **glycoprotein lectin-based immunosorbent assay (GLIA)** (Vermassen et al., 2012).

On the table below, there is a list of some of the commercially available lectins that have been used in several lectin-based detection methods for PSA glycoprofiling and the carbohydrate moieties that they recognize (Damborský et al., 2017).

Table 3. List of commercially available lectins suitable for PSA glycoanalysis and the carbohydrate moieties they recognize. Modified from: Damborský et al. BioNanoScience. 2017 February 23.

Lectin	***Carbohydrate moiety***
Aleuria aurantia agglutinin (AAA)	Fuc-α1-6/3GlcNAc
Canavalia ensiformis (Con A)	α-Man and α-Glc
Hipperastrum hybrid (HHL)	α-Man
Lens culinaris/Lentil lectin (LCA)	Fuc-α1,6-GlcNAc-N-Asn, α-Man and α-Fuc
Lotus tetragonolobus agglutinin (LTA, LTL)	α -Fuc
Lycopersicon esculentum (LEL)	$(GlcNAc)_{2-4}$
Maackia amurensis I, II (MAL, MAA)	Galβ4GlcNAc (I) / Neu5Acα3GalβGalNAc (II)
Narcissus pseudonarcissus (NPA, NPL)	α-Man
Phaseolus vulgaris leucoagglutinin (PHA-L)	GalNAcα1-3(Fuc-α1-2)Galβ Gal(β1,4), GlcNAc(β1,2) and Man(α1,6)
Sambcus nigra agglutinin (SNA)	Neu5Acα2-6Gal/GalNAc
Trichosanthes japonica	β-GalNAc- and Fucα1-2Gal
Ulex europaeus (UEA-1)	α-Fuc
Wheat germ (WGA)	β-GalNAc- and Neu5Ac

Fuc fucose, *Gal* galactose, *GalNAc* N-acetil galactosamine, *Glc* glucose, *GlcNAc* N-acetyl glucosamine, *Man* mannose, *Neu5Ac* N-acetylneuraminic acid

The methodologies that use lectins have several advantages. In fact, many lectins have high capacity to isolate and/or characterize the glycoproteins based on protein glycoforms as well as on glycan structures, and they are cheaper compared to other methods (Geyer and Geyer, 2006). Nonetheless, single-site binding of many lectins is known to be low (Kd in the micromolar range). Therefore, these lectins require multivalent binding to carbohydrates to achieve a high avidity (Nilsson et al., 2012). In particular, analysis of low abundant glycans by lectin microarrays is quite problematic because labelling of the sample is needed and, consequently, it can negatively affect the robustness of the

analysis (Pihikova et al., 2016). In addition, many affinity-based methods require a prior purification of the glycoprotein of interest by using antibodies or by applying other protein purification techniques, because lectins do not interact with protein backbones and, therefore, do not recognize the protein in which the sugar chain is attached to (Varki et al., 2009). Thus, leading to the detection of other glycoproteins that may share a common epitope with the target glycosylated protein. Alternatively, lectins can be used directly to detect those glycans that were already captured by antibodies coated on a plate, by forming a sandwich, as in the case of ELLA (Vermassen et al., 2012).

2) Mass spectrometry-based methods

Alternatively to lectin-based detection methods in cancer research, there are also the **Mass spectrometry-based methods (MS)** (Vermassen et al., 2012)**.** MS techniques, together with advanced separation methodologies like electrophoresis gels or chromatographic strategies, offer a generic approach to achieve a high-through put analysis of aberrant protein glycosylation patterns driven from complex mixtures (Damborský et al., 2017). Mainly due to its sensitivity and selectivity, mass spectrometry has been one of the most powerful tools for identifying the glycoproteins, analyzing the glycosylation sites and elucidating the glycan structures (Vermassen et al., 2012). Most of these techniques, however, have been applied to characterize the glycopeptides and/or glycans rather than the intact glycoproteins. Because of the extensive microheterogeneity of the carbohydrate moiety and the mixed information obtained from the fragment ion spectra, direct analysis of the complete glycoprotein is still technically challenging. In addition, glycosylated proteins are less efficiently ionized than proteins or even glycopeptides (Geyer and Geyer, 2006).

Regarding PSA, numerous MS-based strategies dealing with PSA glycosylation pattern in PCa patients, BPH and healthy individuals have been carried out aiming to comprehend the association between aberrant glycan structures and tumorigenesis (Damborský et al., 2017). Some of the most common used techniques is the **matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF),** through which glycopeptides and glycans can be easily analyzed and sequenced due to the formation of characteristic fragment ions. Besides MALDI, the **electrospray ionization (ESI)** is also often used as the ionization source in glycomic characterization and is frequently applied after **nanoflow-liquid chromatography** or **collision induced dissociation (CID)** (Wuhrer et al., 2007). In spite of this, ESI mass spectra leads to the formation of multiply charged ions, thus, making the assignment of carbohydrate structures more difficult. Furthermore, information on the peptide sequence and the attachment sites is rarely revealed (Geyer and Geyer, 2006, Wuhrer et al., 2007).

Throughout MS techniques, an in-depth qualitative and quantitative glycoproteomic/glycan analysis of PSA has been achieved (Pan et al., 2011). Namely, information about the glycan composition, the sialylation and fucosylation patterns of PSA driven from different health conditions and the identification of the vast amount of PSA glycoforms has been revealed. Taken together, this reported data suggest that MS methods have great potential in cancer biomarker research as they can provide many detailed information of the changes related to tumorigenesis, compared to other strategies such as ELLA. However, some issues regarding the sample pre-treatments, purification techniques, time-consuming data interpretation and highly skilled operators are still unsolved (Damborský et al., 2017). For instance, the fact that multiple-step sample preparation is needed for glycan analysis, negatively affects the quantification accuracy as well as limits the analysis of large numbers of

samples in clinical studies to obtain data with sufficient statistical significance (Vermassen et al., 2012).

3) 2-Dimensional electrophoresis

Separation techniques such as sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and 2-dimensional gel electrophoresis (2DE), have proved to be quite efficient. Usually these are applied as the first step for isolation or characterization of glycoproteins, which can be further coupled to MS-based methods or other techniques, such as HPLC (Geyer and Geyer, 2006).

The 2-DE technique is based on the separation of proteins, driven from complex biological samples, in two steps. In the first place, proteins are separated according to their isoelectric points (pI) by isoelectric focusing (IEF). On the other hand, the second-dimension consists of an SDS-PAGE polyacrylamide gel that separates the proteins on the basis of their molecular weights (MW) (Sarrats et al., 2010).

Until now, 2-DE gels have been preferred for the evaluation of sialic acid variation in PSA, which had been driven from different physiological conditions (i.e., cancer and BPH patients). Namely, because sialic acid is negatively charged, it can modify the PSA's isoelectric point when its content in the N-glycan chain is increased or decreased. Therefore, 2DE gels may seem very useful tools for an efficient isolation and further analysis of the different sialylated isoforms of PSA (Sarrats et al., 2010, Vermassen et al., 2012).

However, a significant disadvantage of this method is the frequent inadequate representation of membrane glycoproteins. This is mainly due to the low solubilizing power of the detergents currently used for hydrophobic proteins. Furthermore, most of these proteins are weakly stained because of the elevated content of carbohydrates (Geyer and Geyer, 2006).

4) Capillary electrophoresis

Throughout capillary electrophoresis (CE) the charged analytes are separated according to their migration velocity in an electric field suited across the ends of a capillary column (Chem.libretexts.org, 2017).

Few complementary CE strategies compatible with online derivatization through amino acid moieties of PSA have been developed, mainly by using ultra violet (UV) quantification (Vermassen et al., 2012). Most of these strategies analyze the intact glycoprotein, although some studies have applied the CE-technique to analyze the glycopeptides and less commonly the glycans (Geyer and Geyer, 2006). CE methods are based on the use of different coatings, electric field polarities as well as capillary lengths. Thus, allowing the separation of more than 8 different isoforms of the glycoprotein (Vermassen et al., 2012). Depending on the targeted isoforms, the CE method applied can be chosen, such as decamethonium bromide coating or morpholine coatings, which are used for distinguishing PSA isoforms of lowest and highest pI respectively. In some occasions, however, the elution order can be reversed basing on the type of coating applied (Garrido-Medina et al., 2011).

The advantages provided by CE are high separation efficiency and speed of analysis, being able to analyze the specific candidates within a period of 8 to 17 minutes (Vermassen et al., 2012). Moreover, charged oligosaccharides, such as sialylated glycans, can be directly isolated due to their electric charge/molecular size ratio (Geyer and Geyer, 2006). On the other hand, for neutral glycan mixtures

the capillary separation is not as efficient as for charged species, because usually these molecules tend to co-migrate together as their charge and size ratio is practically the same (Kammeijer et al., 2017). Therefore, addition of fluorophores is required to increase the charge of the species. Indeed, a higher charge results in faster analyses and better resolution (Geyer and Geyer, 2006).

Although CE has a relatively high power of resolution to comparatively monitor the glycoform populations of, for instance, PSA isoforms, it does not allow any conclusions about the nature of glycan chains attached (Geyer and Geyer, 2006). Depending on the type of coating used, CE cannot be coupled with MS either. Furthermore, CE requires higher concentrations of sample to work efficiently and this, in fact, may be limited when the analyte levels in the blood are low (Kammeijer et al., 2017). For instance, in the case of most PCa patients, whose levers are within the *grey zone* (Kammeijer et al., 2017).

5) High-Performance Liquid Chromatography (HPLC)

One of the most common and robust methods to quantify the glycans, once they have been released from the prostate-specific antigen, is the HPLC coupled with fluorescence detection. This technique allows the identification and preliminary assignment of specific fluorophore-labeled carbohydrate moieties based on their retention time. As a standard HPLC profile of a glycan mixture, several glycan peaks containing different sugars are obtained and, by looking at the percentage area of each glycan peak, the number of carbohydrate chains can be elucidated for each one of the peaks. There are diverse HPLC systems and separation principles, which can be used to distinguish glycans based on charge and size. These include: **hydrophilic interaction liquid chromatography (HILIC), reversed-phase HPLC (RP-HPLC), weak anion exchange chromatography (WAX) and high-pH anion exchange with fluorescence detection (HPAE-FD)** or **with pulsed amperometric detection (HPAE-PDA)** (Kattla et al., 2011).

Taken together, HPLC methods allow the analysis of glycans according to their structure, size, composition, topology, and branching. Noteworthy, a key advantage of all HPLC-based techniques is that they are remarkably versatile and may be used preparatively for follow-on analyses as well as after exoglycosydase digestions. In spite of this, the major drawbacks of this method are that few glycans co-elute and, for a complete characterization of the sequence, orthogonal technologies are required, such as MS (Kattla et al., 2011).

Despite all the advantages provided by the recent developments in glycoproteomic and glycomic methodologies, such as the high sensitivity of the MS techniques or the different approaches that take advantage of the specific affinity of the lectins against the glycan units; there is still no single method that can efficiently analyze the glycosylation in PSA. Hence, in order to achieve an in-depth, comprehensive identification of the aberrant glycoprotein and further quantitative detection of specific glycosylation alterations, a concerted approach drawing from several techniques is required (Pan et al., 2011). These concerted approaches can be classified depending on the target in which the authors have focused on, including the analysis of the intact glycoproteins or the characterization of the glycoprotein after enzymatic treatment. Namely, the detection of the glycopeptides after trypsin treatment or the glycans released from PSA by using **Peptide-N-Glycosidase F (PNGase F)** (Damborský et al., 2017). **[Figure 13]**

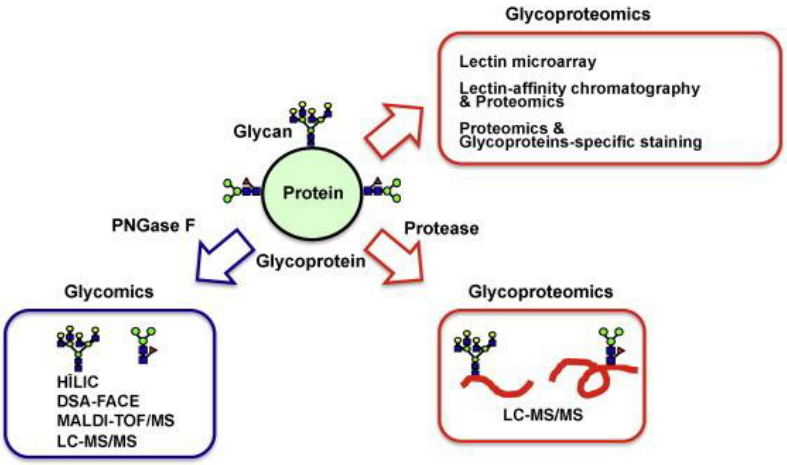


Figure 13. Schematic depiction of several glycomic and glycoproteomic methods to analyze PSA glycosylation. Specifically, the intact protein, the glycopeptides or the released glycans by the enzyme PNGase F. Commonly lectins are used for analyzing the intact glycoprotein, whereas HPLCs and mass spectrometry approaches are often applied for characterizing the glycopeptides or the glycans. Miura and Tamao. ScienceDirect. 2016 August; 1860(8): 1608-1614

5.2.2 Analysis of the glycoproteins

Glyoproteomics is the study of the glycosylated proteins in a biological system (Pan et al., 2011). The characterization of the glycoprotein profile can be performed by analyzing either the intact glycosylated proteins or the glycopeptides (Geyer and Geyer, 2006). Thus, allowing the elucidation of the glycan structures at individual sites, as well as the identification and quantification of glycoproteins (Gilgunn et al., 2013, Kattla et al., 2011). The most common workflow through which glycoproteins are analyzed is, indeed, **bottom-up analysis** (i.e, s**hotgun proteomics**). In general, glycoproteomic pipeline consist of different steps. These include glycoprotein/glycopeptide enrichment and/or multidimensional protein separation, tandem mass-spectrometric analysis and bioinformatics data interpretation (Meretier et al., 2016, Pan et al., 2011).

1) Analysis of the intact PSA

In a case study performed by Dwek and co-workers, the analysis of PSA glycosylation was assessed through the application of two lectin-based detection methods. The principal objective of this set-up was to demonstrate whether the fucosylation levels of PSA were significantly different between PCa and BPH patients, whose serum PSA levels were within the grey zone range (4-10ng/mL). Thus, pinpointing that by detecting alterations in the fucosylation pattern, an element of specificity would be added to PSA test (Dwek et at., 2010). In general, the experimental design consisted of two steps:

A) <u>Glycosylation assessment through the binding of UEA-1 lectin to immuno-purified PSA</u>

In the first place, samples of serum PSA were obtained from PCa and BPH individuals, whom had been pathologically confirmed previously. Secondly, serum PSA samples were immuno-purified by affinity chromatography, in which the polyclonal rabbit anti-human PSA antibody was used as the ligand (Dwek et at., 2010). This purification step is very important because, although lectins have a high specificity for glycans, they are not specific for the proteins in which the carbohydrates are bound to (Varki et al., 2009). Namely, if the lectins were used in a complex sample without previous isolation of PSA, the glycans that would be analyzed could drive from different proteins as well.

The resulting PSA fractions obtained through the chromatography were separated using two SDS-PAGE. Once purified, the PSA samples were transferred to nitrocellulose by Western blotting. In one case, the membrane was probed with polyclonal rabbit anti-human PSA antibody, whereas the other membrane was probed with biotinylated *Ulex Europaeus* lectin (UEA-1) that is specific for α1,2-linked fucose. The authors used two different probes in order to compare the binding between UEA-1 and anti-PSA antibody. In both cases, the addition of streptavidin-horse radish peroxidase to the probes was used for the detection and semi-quantification of the PSA samples. Eventually, the probing with UEA-1 lectin demonstrated that the fucosylation of PSA was increased in the PCa serum samples (Dwek et at., 2010). It is worth to mention, however, that UEA-1 may not be the best option to analyze PSA as fucosylation in this glycoprotein is always α1,6 fucose. In contrast, α1,2-linked fucoses are found on the external parts of PSA and the quantity it is rather low in PSA (EY Laboratories, Inc., 2017).

B) Validation with *enzyme-linked immunosorbent lectin assay (ELLA).*

ELLA is an interesting approach to study the different structures (Vermassen et al., 2012). The test principle of this method is derived from that of the ELISA sandwich technique, but lectins are used instead of antibodies (Gabius, 2017). In ELLA, the targeted glycoprotein is first captured from the complex sample onto antibody coated plate and further incubated with biotin/peroxidase labeled lectins that will detect the glycosylation pattern of a certain protein (Vermassen et al., 2012). In contrast to lectin affinity chromatography, the immunoabsorbant assay has the advantage that it can selectively analyze the glycans of a determined glycoprotein without prior purification (Gabius, 2017).

In the experimental design of Dwek and co-workers, the ELLA was used to detect and, therefore, validate the fucosylation of freePSA and totalPSA by using a new set of serum samples. In this case, an increase in the fucosylation of fPSA was observed in the PCa samples compared with fPSA from individuals with BPH. Moreover, the investigators reported that ELLA together with colorimetric labelling was 92% specific and 69% sensitive for PCa over BPH. Thus, suggesting that ELLA test may lead to a more specific and sensitive test for PCa based on fucosylation changes of fPSA. Taken together, these data support the results obtained with the affinity purified PSA described previously (Dwek et at., 2010). Besides the high sensitivity that ELLA assay presents, a drawback of this method is that as lectins are just specific for glycans and, antibodies may form some unspecific linkages with other proteins besides PSA, it cannot be certain that changes in glycosylation levels may drive from the alteration in PSA glycans or other proteins are influencing the results (Varki et al., 2009). Furthermore, ELLA assay does not provide in-detail information about the glycan structure as MS techniques and either the sample can be recovered.

2) Analysis of the intact glycopeptides

Most recently, Kammeijer and co-workers have developed a high-resolution separation platform for the analysis of the tryptic glycopeptides from PSA, through the application of capillary electrophoresis-electrospray ionization-mass spectrometry (CE-ESI-MS). With this approach, the authors aimed to determine whether the CE-ESI-MS strategy is a potential tool for discerning between the different sialic acid linked isomers. As previously mention, sialylation is one of the glycan features which has shown to be altered in prostate cancer and, in PSA, can occur in different linkage positions, i.e., α2,3 and α2,6.

Previous to the analysis of the sialylated glycopeptides by CE-ESI-MS, the PSA samples were treated with DTT to reduce the disulfide bridges and digested with the protease trypsin to break down the glycoprotein into peptide fragments. Afterwards, the glycopeptides were separated by capillary electrophoresis, which is a very efficient tool to separate the different isoforms of sialic acid linked glycopeptides. Whereas conventional MS techniques might have issues separating the differentially linked isomers, due to identical molecular formulas and analogous fragmentation patterns, CE is able to achieve this baseline separation owing to a difference in the electrophoretic motilities of the glycopeptides, which is in turn highly influenced by the degree of sialylation. As sialic acid is negatively charged and the migration of the analytes depends on their charge-to-size ratio, the more sialylated the glycopeptide is, the higher the retention time will be. Therefore, it is expected that the non-sialylated isoforms migrate first and the di-sialylated species the last. Moreover, the position of the sialic acid linkage also influences the migratory velocity. In fact, the variants with α2,3-linked sialic acid have shown to migrate slower than the α2,6-linked sialic acid glycopeptides (Kammeijer et al., 2017) **[Figure 14]**.

Once the analytes were isolated, they were ionized leading to different ions and further detected by the Mass Analyzer. Moreover, an exosialidase step was carried out to evaluate the migration behavior of the sialylated isomers. Eventually, CE-ESI-MS data was analyzed with DataAnalysis 4.2 in order to match the glycopeptides and the resulting compositions, based on the exact mass, selectivity and relative intensities. The outcome of this set-up was the detection of 75 different glycopeptides with one single N-linked glycosylation site in the position R-N69K-S (Kammeijer et al., 2017).

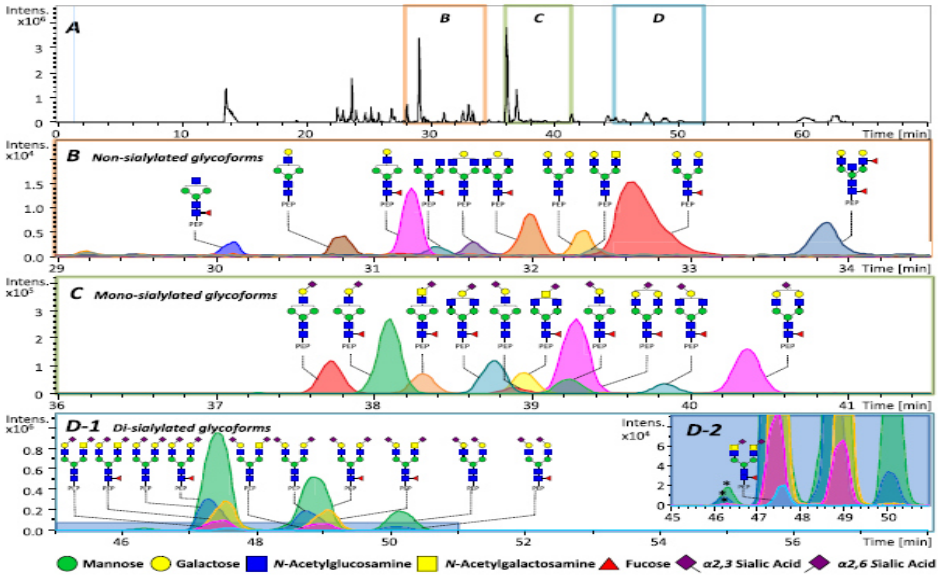


Figure 14. CE-ESI-MS analysis of PSA tryptic glycopeptides(A). Three distinct clusters were observed (B, C and D). (B) cluster contains only the non-sialylated forms. (C) presents the mono-sialylated and (D) the di-sialylated glycopeptides. Overall, a total amount of 75 glycopeptides were identified from the tryptic digestion of a standard sample of PSA. Kammeijer et al. CIENTIFIC REPORTS. 2017 June 16; 7(1): 1-8.

Three main peptide glycoforms were observed after the analysis of a standard PSA sample by CE-ESI-MS, as shown in **Figure 14**. Namely, non-sialylated, mono-sialylated and di-sialylated glycoforms. From section A, it can be determined that mono-sialylated isoforms were more predominant compared to non-sialylated and di-sialylated glycoforms. With respect to cluster B, the most prevalent isoforms have a Fucose-core α1,6- and either present one or two residues of N-Galactose at the terminal part. On the other hand, the isoforms that have one α1,6-fucose and either

α2,3 or α2,6 attached to a Gal-GlcNAc structure, are the major forms of mono-sialylated type, whereas the isoform with two sialylated glycans with α2,6 configuration at the reducing ends, is the most abundant di-sialylated glycoform.

The advantages CE-ESI-MS approach offers are, in fact, that no ambiguous and time-consuming sample preparation and derivatization is required to obtain an accurate and direct analysis of complex glycoproteins, as well as the high ability of this method to discriminate between the different linkages, independently of the glycopeptide structure. Indeed, this technique is more efficient and sensitive compared to nano-LC-ESI-MS and other conventional MS /MS approaches, which cannot deal with the separation of the differential linked isomers due to similar fragmentation patterns. Thereafter, CE-ESI-MS seems a very promising tool to improve sensitivity and selectivity of biomarkers and for biomarker discovery. Nonetheless, non-sialylated species might have more problems to separate their isomers as some of them might co-migrate, and the amount of sample needed for this approach (1ng) is quite complex to retrieve from serum of PCa patients because most of them have PSA concentrations lower than 10ngPSA/mL (Kammeijer et al., 2017). In addition, it cannot be determined that this approach will be that effective when analyzing PSA from a complex biological sample, as a sample of standard PSA was used in this case.

5.2.3 Analysis of the free glycans

The field in charge of studying the repertoire of glycans is known as glycomics. One glycomic approach designed for the characterization of prostate specific antigen, in particular focusing on sialylated glycan linkages, was also performed by Kammeijer and co-workers. The experimental set-up consisted in the analysis of the glycans after they were released from PSA and derivatized, through the application of MALDI-TOF-MS. Certainly, this technique was used as an orthogonal study to differentiate α2,3- from α2,6- linked sialic acids and to compare the results with the ones obtained previously, by the analysis of the glycopeptides throughout the CE-ESI-MS approach (Kammeijer et al., 2017).

As a general overview, this study consisted of several steps, which include: A) tryptic digestion of PSA, B) N-glycan release, C) derivatization of N-glycans by Ethyl esterification, D) glycan enrichment by HILIC SPE and, eventually, E) analysis of the derivatized N-glycans by MALDI-TOF-MS (Kammeijer et al., 2017) **[Figure 15].**

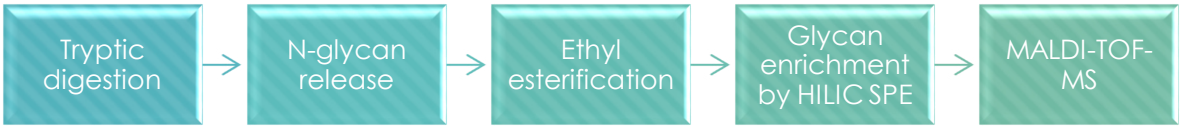


Figure 15. Scheme of the experimental set-up performed by Kammeijer and co-workers.

A) Tryptic digestion

A PSA sample already purified was digested with trypsin, as described in the analysis of the glycopeptides (Kammeijer et al., 2017).

B) Analysis of released N-linked glycans

The tryptic glycopeptide fragments were treated with PNGase F, which is an enzyme that cleaves between the innermost GlcNAc and asparagine residues of high mannose, hybrid, and complex oligosaccharides from N-linked glycoproteins (Biolabs, 2017). As a consequence for the cleavage

between GlcNAc and asparagine, the Asn is turned into an Asp residue. Thereafter, the resulting mass difference of one Dalton may be subsequently used to assess the attachment site of the glycan by mass spectrometry (Geyer and Geyer, 2006).

C) Ethyl esterification

The Ethyl esterification was applied for the chemical derivatization of the sialic acid carboxyl groups. (Kammeijer et al., 2017). This chemical derivatization leads to mass shifts allowing a direct discrimination of particular sialic acid linkages in MS spectrum. Glycans with α -(2,6)-linked sialic acid gave place to a mass increment of 28 Da per sialic acid in the MS spectrum, as a response to the ethyl esterification reaction. On the other hand, derivatization of the α -(2,3)-linked sialic acid resulted in 18 Da mass loss per each sialic acid. Therefore, the distinct linkages could be determined due to the differences on the molecular weights (Pihikova et al., 2016). **[Figure 16]**

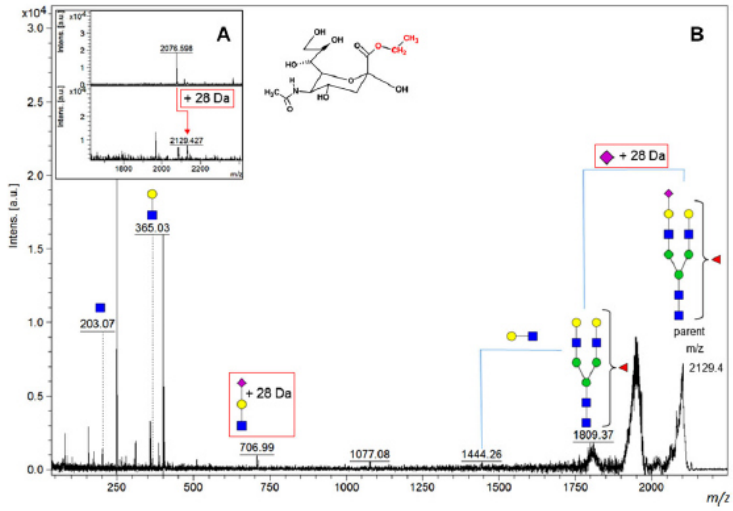


Figure 16. Mass increment of 28 Da caused by ethyl esterification of an α -(2,6)-linked sialic acid N-glycan. The charge-size ratio (m/z) without derivatization has a value of 2076.6, [M − H]. In contrast, the value of m/z after the ethyl esterification is 2129.4 [M + Na]+, therefore, giving place to a 28Da glycan **(A)** MALDI TOF/TOF spectrum of the respective structure after the derivatization.**(B)** The loss of ethyl ester of sialic acid resulting in the signal at m/z 1809.4, proved the selectivity of derivatization and the presence of α-(2,6)-linked sialic acid. Pihikova et al. Proteomics. 2016 April 13; 16(24): 3085-3095.

D) Glycan enrichment by HILIC SPE

In order to concentrate the N-glycans a cotton wool hydrophilic interaction liquid chromatography solid phase extraction (SPE) was carried out (Kammeijer et al., 2017).

E) Determination of sialic acid linkage

The ethylated released N-glycans were analyzed with MALDI TOF-MS. The resulting spectra were interpreted by the database FlexAnalysis. Eventually, 37 PNGase F released N-glycans were detected by the application of this approach, being the structure that is α2,6- di-sialylated and has a Fuc- α1,6, the most predominant isoform. Moreover, Fuc – α1,6-SA- α2,3, α2,6 glycoform is also quite abundant (Kammeijer et al., 2017).

In contrast to MALDI-TOF-MS glycoprofiling, the analysis with CE-ESI-MS is less time consuming and simpler in terms of sample preparation. Moreover, the CE-ESI-MS platform has the ability to detect significantly higher quantity of differentially linked sialylated isoforms, as well as presents a

high sensitivity of detection with low limit of detection (Kammeijer et al., 2017)., as shown in **Figure 17.**

Table S-1. Theoretical and observed glycopeptide and glycan mass during the CE–ESI-MS and MALDI-TOF-MS analysis of PSA with the corresponding deviation and ppm error (<10). The "PEP" label illustrates the tryptic peptide sequence $N_{69}K$ to which the glycan is attached.

			CESI-MS									MALDI-TOF-MS			
Glycan Species			$[M+2H]^{2+}$	Observed	deviation	PPM error	$[M+3H]^{3+}$	Observed	deviation	PPM error	MS2*	$[M+Na]^{+}$	Observed	deviation	PPM error
	Monosialylated/ Sulphated														
12	H3N4F1 $[SO_3]^-$		893.327	893.326	-0.001	-1.01	595.888	ND				1565.53	ND		
13	$H4N3S_{2,3}1$		905.355	905.353	-0.001	-1.50	603.906	ND				1571.53	1571.51	-0.02	-15.27
14	$H4N3S_{2,6}1$		905.355	905.354	-0.001	-0.94	603.906	ND			X	1617.58	1617.56	-0.02	-10.51
15	H3N5 $[SO_3]^-$		921.838	921.837	-0.001	-0.98	614.895	ND				1622.51	ND		
16	$H3N4S_{2,6}1$		925.868	925.867	-0.001	-1.33	617.581	ND			X	1658.60	1658.60	0.00	-0.60
17	$H4N3F1S_{2,3}1$		978.384	978.382	-0.001	-1.44	652.592	ND			X	1717.59	1717.59	0.00	0.00
18	$H4N3F1S_{2,6}1$		978.384	978.382	-0.001	-1.34	652.592	ND			X	1763.63	1763.64	0.00	2.84
19	$H5N3S_{2,3}1$		986.381	986.381	0.000	-0.47	657.923	ND				1733.59	ND		

Figure 17. This table shows several mono-sialylated glycan structures of a standard of PSA, which have been detected by performing CE-ESI-MS and MALDI-TOF-MS. It can be appreciated that with MALDI-TOF-MS, less structures were identified and the accuracy is lower in MALD-TOF-MS than in CE-ESI-MS. Kammeijer et al. CIENTIFIC REPORTS. 2017 June 16; 7(1): 1-8.

Through the application of ELLA, CE-ESI-MS and MALDI-TOF-MS approaches, significant differences in the fucosylation of PSA were identified when samples from PCa patients and BPH individuals were compared, as well as many glycan isoforms could be elucidated through the analysis of a standard PSA sample, respectively. In general, ELLA has proved to be quite efficient in glycoprotein analysis, in particular, for the identification of meaningful changes in the glycosylation pattern of PSA without a prior purification step, as shown in the study of Dwek and co-workers. The high sensitivity offered by this assay is caused by the use of antibodies and lectins for the capturing and detection of the glycoprotein. Indeed, antibodies and lectins can recognize with high specificity epitopes from the polypeptide sequence and glycan moieties, respectively. However, as lectins are not specific for a given protein and the fact that antibodies can also bind nonspecifically to other polypeptides, the viability of the results may be relatively affected. In addition, lectin-based methods are not able to provide all the structural information given by the performance of CE-ESI-MS and MALDI-TOF-MS strategies, among other MS techniques. In fact, 75 isoforms from standard PSA were elucidated through the application of CE-ESI-MS platform and 37 with MALDI-TOF-MS. Furthermore, information about the relative abundance of each type of glycoform was also revealed. In regard to the differentially linked sialylated PSA isoforms, the most prevalent presented an α1,6-fucose and one or two sialic acid residues at the non-reducing ends of the glycan moiety, which were both α2,6- in the case of the di-sialylated species, whereas for the mono-sialylated glycoforms the linkages could be either α2,6- or α2,3-, according to the results of the CE-ESI-MS approach. On the other hand, the most abundant glycoforms detected by MALDI-TOF-MS were the α2,6- di-sialylated and the α2,6, α2,3-di-sialylated forms, both with an α1,6-fucosa. Comparing these techniques, it is certain that CE-ESI-MS platform is more efficient than MALDI-TOF-MS, mainly because it allows the baseline separation of many differentially linked glycans that, in fact, cannot be distinguished for the conventional MS techniques, such as MALDI-TOF-MS. Despite of this, capillary electrophoresis presents few co-migration problems as well as requires a larger amount of sample to afford

significant results. In contrast to ELLA assay, CE-ESI-MS and MALDI-TOF require prior immuno-purification of the PSA as they cannot distinguish the glycoproteins with such specificity. Apart from these approaches, many other strategies have been used for the glycoprofiling of PSA. For instance, in one study carried out by Sarrats and co-workers, five different sialylated subforms of healthy seminal PSA and 4 glycoforms from serum PSA driven from PCa patients, were reported and compared. In particular, the authors found that the most abundant PSA subform (N-acetyl-lactosamine with one or two α2,3-linked sialic acids), carried a higher proportion of α2,3-lined sialic acid in prostate cancer. Thus, indicating that PSA sialylation could be used as a biomarker for discerning PCa patients from BPH individuals. These results were obtained through the performance of a 2-DE gel coupled with HPLC and exoglycosidase digestions, in combination with fluorophore labeling and MS techniques. Besides the efficiency of this approach, 2-DE gels have rather low solubilizing power and issues to stain hydrophobic molecules, as well as are significantly slower than capillary electrophoretic techniques in analyzing PSA isoforms. Moreover, HPLC methods can also have few limitations, in particular, the co-elution of certain glycan structures that, in turn, make the identification less accurate.

Giving a close look to the different techniques reviewed in this work, it can be noticed that advances in the development for more sensitive and accurate analytical tools have been achieved. Besides the many advantages that the glycomic and glycoproteomic approaches offer, greater endeavors are required to overcome the limitations that these strategies present, for improving cancer biomarkers.

6. ETHICS AND SUSTAINABILITY

Currently, a focus toward biomarker research and strategies for the analysis of the glycosylation in biomarkers has become a substantial part of biomedical research with tremendous impact on health and quality of life (Damborský et al., 2017). In fact, biomarkers are potential tools for a more accurate diagnosis and prognosis of cancer and, most important, they are an alternative to the invasive methods used in cancer screening (Sivastrava and Gopal-Sivastrava, 2002). Nevertheless, before the analysis of PSA, several aspects about the experimental set-up should be considered. Thus, preventing the misuse of PSA samples as this must be taken from human beings. Noteworthy, the extraction of samples must be approved by an ethic committee. In addition, the methods should be chosen very carefully to not make improper use of high-priced chemicals and other material, which are quite expensive in research and noxious for the environment. Waste management should be considered as well.

7. CONCLUSIONS

The paper examined the methodologies and strategies applied for the analysis of the glycosylation pattern of PSA with a broad literature review and database analysis. The following conclusions can be drawn:

- Biomarkers have become of great importance due to their potential to detect cancer at its early stages, by only looking at body-fluid samples and avoiding invasive techniques. Nonetheless, there is an urgent need to improve and discover other biomarkers to obtain more sensitive and reliable assays.
- Prostate specific antigen test still is the most prevalent in our society for early detection of prostate cancer. However, endeavors are being carried out to increase the specificity and

sensitivity of PSA. Most recently, scientists have started to examine the glycosylation pattern, specifically sialylation and core-fucosylation, of PSA that seems to have a tremendous potential for the improvement of PSA as a biomarker.

- Bioinformatics, such as UniProt, MASCOT and Net(N/O)Glyc databases, significantly influence the speed and the efficiency of the analysis of PSA and provide many structural data about its glycosylation pattern in both physiological and pathological conditions.
- Lectin base-methods, MS techniques, capillary electrophoresis, multi-dimensional gels and high liquid performance chromatography are the principal methods used for the analysis of PSA glycosylation. However, to achieve a higher sensitivity, these techniques must be coupled together.

8. BIBLIOGRAHY

Adhyam, M. and Gupta, A. (2012). A Review on the Clinical Utility of PSA in Cancer Prostate. Indian Journal of Surgical Oncology, [online] 3(2), pp.120-129. Available at: http://www.ncbi.nml.nih.gov /pcm/articles/PMC3392481/pdf/13193_2012_Article_142.pdf

Balk, S., (2003). Biology of Prostate-Specific Antigen. Journal of Clinical Oncology, [online] 21(2), pp.383-391. Available at: http://ascopubs.org/doi/full/10.1200/JCO.2003.02.083

Baycin Hizal, D., Wolozny, D., Colao, J., Jacobson, E., Tian, Y., Krag, S., Betenbaugh, M. and Zhang, H. (2014). Glycoproteomic and glycomic databases. Clinical Proteomics, [online] 11(1), p.15. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3996109/

Biolabs, N. (2017). PNGase F | NEB. [online] Neb.com. Available at: https://www.neb.com/products/p0704-pngase-f [Accessed 10 Jul. 2017].

Chem.libretexts.org. (2017). Capillary Electrophoresis - Chemistry LibreTexts. [online] Available at: https://chem.libretexts.org/Core/Analytical_Chemistry/Instrumental_Analysis/Capillary_Electrophoresis

Damborský, P., Damborská, D., Belický, Š., Tkáč, J. and Katrlík, J. (2017). Sweet Strategies in Prostate Cancer Biomarker Research: Focus on a Prostate Specific Antigen. BioNanoScience. [online] Available at: https://www.researchgate.net/publication/313963780_Sweet_Strategies_in_Prostate_Cancer_Biomarker_Research_Focus_on_a_Prostate_Specific_Antigen

Drake, R., Jones, E., Powers, T. and Nyalwidhe, J. (2015). Altered Glycosylation in Prostate Cancer. Advances in Cancer Research, [online] 126, pp.345-382. Available at: http://dx.doi.org/10.1016/bs.acr.2014.12.001

Dwek, M., Jenks, A. and Leathem, A. (2010). A sensitive assay to measure biomarker glycosylation demonstrates increased fucosylation of prostate specific antigen (PSA) in patients with prostate cancer compared with benign prostatic hyperplasia. Clinica Chimica Acta, [online] 411(23-24), pp.1935-1939. Available at: https://www.ncbi.nlm.nih.gov/pubmed/20708609

Encyclopedia Britannica. (2017). bioinformatics | science. [online] Available at: https://www.britannica.com/science/bioinformatics [Accessed 10 Jul. 2017].

EY Laboratories, Inc. (2017). Ulex europaeus - EY Laboratories, Inc. [online] Available at: http://eylabs.com/ulex-europaeus/#.WWtHt9Tyi71 [Accessed 13 Jul. 2017].

Gabius, H. (2017). Lectins and Glycobiology. [ebook] München: Sigrun Gabius, p.141. Available at: https://books.google.es/books

Garrido-Medina, R., Díez-Masa, J. and de Frutos, M. (2011). CE methods for analysis of isoforms of prostate-specific antigen compatible with online derivatization for LIF detection. ELECTROPHORESIS, [online] 32(15), pp.2036-2043. Available at: http://www.electrophoresis-journal.com

Geyer, H. and Geyer, R. (2006). Strategies for analysis of glycoprotein glycosylation. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, [online] 1764(12), pp.1853-1869. Available at: https://www.ncbi.nlm.nih.gov/pubmed/17134948

Gilgunn, S., Conroy, P., Saldova, R., Rudd, P. and O'Kennedy, R. (2013). Aberrant PSA glycosylation—a sweet predictor of prostate cancer. Nature Reviews Urology, [online] 10(2), pp.99-107. Available at: http://www.nature.com/nrurol

Google+., T. (2017). Data Analysis Software and Tools | Information Builders. [online] Informationbuilders.com. Available at: http://www.informationbuilders.com/data-analysis [Accessed 11 Jul. 2017].

Henry, N. and Hayes, D. (2012). Cancer biomarkers. Molecular Oncology, [online] 6(2), pp.140-146. Available at: http://www.sciencedirect.com/science/article/pii/S1574789112000117

Kammeijer, G., Jansen, B., Kohler, I., Heemskerk, A., Mayboroda, O., Hensbergen, P., Schappler, J. and Wuhrer, M. (2017). Sialic acid linkage differentiation of glycopeptides using capillary electrophoresis – electrospray ionization – mass spectrometry. Scientific Reports, [online] 7(1), pp.1-8. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5473812/

Kattla JJ, Struwe WB, Doherty M, Adamczyk B, Saldova R, Rudd PM, and Campbell MP (2011) Biologics | Protein Glycosylation. In: MurrayMoo-Young (ed.), Comprehensive Biotechnology, Second Edition, volume 3, pp. 467-486. Elsevier.

Meany, D. and Chan, D. (2011). Aberrant glycosylation associated with enzymes as cancer biomarkers. Clinical Proteomics, [online] 8(1), p.7. Available at: http://clinicalproteomicsjournal.com/content/8/1/7

Meany, D., Zhang, Z., Sokoll, L., Zhang, H. and Chan, D. (2009). Glycoproteomics for Prostate Cancer Detection: Changes in Serum PSA Glycosylation Patterns. Journal of Proteome Research, [online] 8(2), pp.613-619. Available at: https://www.ncbi.nlm.nih.gov/pubmed/19035787

My Cancer. (2017). What Are Biomarkers? - My Cancer. [online] Available at: http://www.mycancer.com/resources/what-are-biomarkers/ [Accessed 30 Apr. 2017]

National Cancer Institute. (2017). Improving the Search for Biomarkers of Early Cancer. [online] Available at: https://www.cancer.gov/research/areas/diagnosis/improving-the-search-for-biomarkers-of-early-cancer [Accessed 1 May 2017].

Nilsson, J., Halim, A., Grahn, A. and Larson, G. (2012). Targeting the glycoproteome. Glycoconjugate Journal, [online] 30(2), pp.119-136. Available at: https://www.ncbi.nlm.nih.gov/pubmed/22886069

Sivastrava, S. and Gopal-Sivastrava, R. (2002). Biomarkers in Cancer Screening: A Public Health Perspective. [ebook] Bethesda: JN THE JOURNAL OF NUTRITION, pp.2471S-2475S. Available at: https://www.ncbi.nlm.nih.gov/pubmed/12163714

Stowell, S., Ju, T. and Cummings, R. (2015). Protein Glycosylation in Cancer. Annual Review of Pathology: Mechanisms of Disease, [online] 10(1), pp.473-510. Available at: https://www.ncbi.nlm.nih.gov/pubmed/25621663

Pan, S., Chen, R., Aebersold, R. and Brentnall, T. (2010). Mass Spectrometry Based Glycoproteomics—From a Proteomics Perspective. Molecular & Cellular Proteomics, [online] 10(1), pp. R110.003251. Available at: https://www.ncbi.nlm.nih.gov/pubmed/20736408

Uniprot.org. (2017). UniProt. [online] Available at: http://www.uniprot.org/ [Accessed 26 Apr. 2017]

Vermassen, T., Speeckaert, M., Lumen, N., Rottey, S. and Delanghe, J. (2012). Glycosylation of prostate specific antigen and its potential diagnostic applications. Clinica Chimica Acta, [online] 413(19-20), pp.1500-1505. Available at: https://www.ncbi.nlm.nih.gov/pubmed/?term=Glycosylation+of+prostate+specific+antigen+and+its+potential+diagnostic+applications

Varki A, Cummings RD, Esko JD, et al., editors. Essentials of Glycobiology. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1908/

White, K., Rodemich, L., Nyalwidhe, J., Comunale, M., Clements, M., Lance, R., Schellhammer, P., Mehta, A., Semmes, O. and Drake, R. (2009). Glycomic Characterization of Prostate-Specific Antigen and Prostatic Acid Phosphatase in Prostate Cancer and Benign Disease Seminal Plasma Fluids. Journal of Proteome Research, [online] 8(2), pp.620-630. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2651839/

Wittmann, V. (2008). Glycoproteins: Occurrence and Significance. [ebook] Heidelberg, pp.1735-1770. Available at: https://link.springer.com/referenceworkentry/10.1007%2F978-3-540-30429-6_43

Zhang, S., & Williamson, B. L. (2005). Characterization of Protein Glycosylation Using Chip-Based Nanoelectrospray with Precursor Ion Scanning Quadrupole Linear Ion Trap Mass Spectrometry. Journal of Biomolecular Techniques: JBT, 16(3), 209–219