



Universiteit  
Leiden  
The Netherlands

## **Glyco(proteo)mic workflows for cancer biomarker discovery**

Moran, A.B.

### **Citation**

Moran, A. B. (2023, November 1). *Glyco(proteo)mic workflows for cancer biomarker discovery*. Retrieved from <https://hdl.handle.net/1887/3655862>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3655862>

**Note:** To cite this publication please use the final published version (if applicable).



# Glyco(proteo)mic Workflows for Cancer Biomarker Discovery

Alan Moran



# **Glyco(proteo)mic Workflows for Cancer Biomarker Discovery**

Alan Moran

ISBN: 978-94-6469-630-1

©2023 Alan Moran. All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author or the journals holding the copyrights of the published manuscripts. All published material was reprinted with permission.

The work presented in this thesis was performed at the Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands.

This work was supported by the European Union's Horizon 2020 Research and Innovation Program GlySign, grant number 722095.

Cover design: Stephen Ledwidge

Printed by: Proefschriftmaken | [www.proefschriftmaken.nl](http://www.proefschriftmaken.nl)

# **Glyco(proteo)mic Workflows for Cancer Biomarker Discovery**

**Proefschrift**

ter verkrijging van  
de graad van doctor aan de Universiteit Leiden,  
op gezag van rector magnificus prof.dr.ir. H. Bijl,  
volgens besluit van het college voor promoties  
te verdedigen op woensdag 1 november 2023  
klokke 11:15 uur

door

**Alan Moran**

geboren te Dublin, Ierland

in 1993

**Promotor:**

Prof. Dr. M. Wuhrer

**Co-promotor:**

Dr. G.S.M Lageveen-Kammeijer

*University of Groningen, Groningen Research Institute of Pharmacy, Analytical Biochemistry*

**Leden van de Promotie Commissie:**

Prof. Dr. C.H. Hokke

Prof. Dr. C.M. Cobbaert

Prof. Dr. G.W. Somsen

*VU University Amsterdam, Faculty of Science, Bioanalytical Chemistry*

Dr. K.R. Reiding

*Utrecht University, Pharmaceutical Sciences, Biomolecular Mass Spectrometry and Proteomics*

*“Do not go gentle into that good night.  
Rage, rage against the dying of the light.”*

– Dylan Thomas, 1951





# Table of Contents

<b>Chapter 1</b>	Introduction	9
<b>Chapter 2</b>	Profiling the Proteoforms of Urinary Prostate-Specific Antigen by Capillary Electrophoresis – Mass Spectrometry	37
<b>Chapter 3</b>	Software-Assisted Data Processing Workflow for Intact Glycoprotein Mass Spectrometry	65
<b>Chapter 4</b>	Sialic Acid Derivatization of Fluorescently Labeled <i>N</i> -Glycans Allows Linkage Differentiation by RPLC-FD-MS	91
<b>Chapter 5</b>	Serum <i>N</i> -Glycosylation RPLC-FD-MS Assay to Assess Colorectal Cancer Surgical Interventions	119
<b>Chapter 6</b>	Discussion and Perspectives	149
<b>Appendices</b>	List of Abbreviations	172
	English Summary	176
	Nederlandse Samenvatting	178
	Curriculum Vitae	180
	PhD Portfolio	182
	List of Publications	184
	Acknowledgements	186



# Chapter 1

Introduction

## Introduction

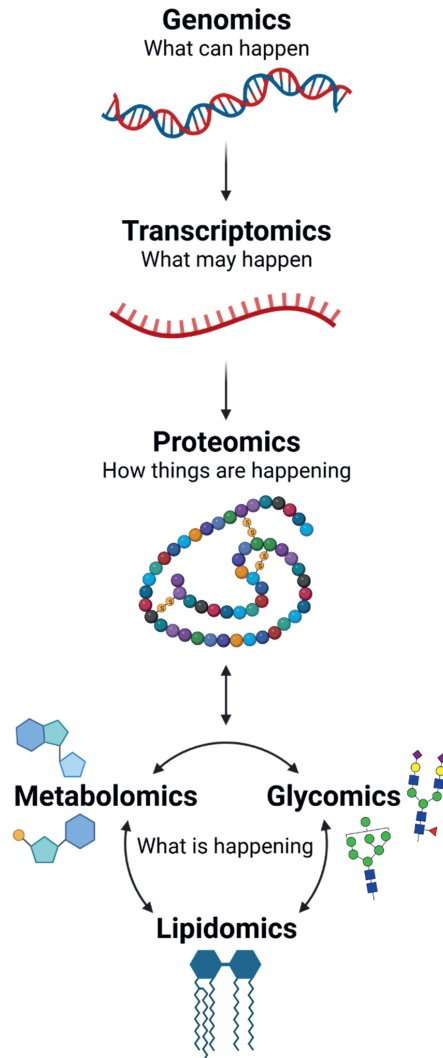
Cancer describes a vast set of diseases in which multiple alterations occur to normal healthy cells in order to transform them towards malignant derivatives.<sup>1</sup> These aberrations can have numerous and variable origins, occur in different types of cells as well as various genetic and metabolic pathways.<sup>2</sup> As a result, the presentation of the cancer phenotype can be diverse and, due to this complexity, the same clinical intervention may not work equally for different patients who seemingly suffer from the same condition.<sup>3</sup> As a result, a personalized approach to medicine may lead to improved outcomes for patients. In this sense, personalized medicine aims to use our understanding of the molecular pathways<sup>4</sup> manipulated by cancer in order to prescribe the right intervention at the right time to the right patient.<sup>5</sup>

## Biomarker Discovery

Disease complexity may be resolved by following a set of biological signatures or “biomarkers” that are correlated with the disease and thus provide an individual perspective of the disease for each patient. A biomarker may be described as a biological parameter that is objectively measured to give information regarding biological, pathological or therapeutic processes.<sup>6</sup> To provide insights into these complex biological processes and interactions, the presence, absence, increase, or decrease of a specific biomarker, can be monitored. As a result, biomarkers may be used for clinical applications such as screening, diagnosis, prognosis, as well as the prediction of response to treatment and monitoring disease progression.<sup>7</sup> Thus, biomarkers have great importance and huge potential for improving patient health by means of developing non-invasive testing, greater clinical accuracy, and personalized medicine.

The discovery and development of biomarkers for clinical use is not limited to a single field and, as shown in **Scheme 1**, encompasses research into the expression of genes (genome), RNAs (transcriptome), proteins (proteome), and metabolites (metabolome)<sup>8</sup> as well as lipids (lipidome) and glycans (glycome). Although integrating multiple fields would provide a comprehensive overview of the biological system of interest,<sup>8,9</sup> the availability of resources, time and expertise pertains that researchers must often focus on a specific area. As physiological changes caused by disease are

often reflected in the proteome, this field represents an attractive opportunity for biomarker discovery and development.<sup>10</sup> In relation to this, proteins demonstrate great versatility and flexibility to carry out the functions of the cell, in contrast to the genome which is generally constant.



Scheme 1. Overview of the 'omics' fields. Created with BioRender.com.

## Challenges for Biomarker Research

The chief requirement of a clinical biomarker is that it must “improve patient outcomes”.<sup>11</sup> Moreover, it must be objectively measured using a suitable and analytically validated assay, which must also demonstrate sufficient performance in important areas such as clinical sensitivity and specificity, as well as positive and negative predictive value.<sup>12,13</sup> Thus, failure to meet these key performance metrics leads to challenges for biomarker testing whereby it is questionable whether patient outcomes are improved or not. A well-known example in this regard is the case of the prostate-specific antigen (PSA) test, an Food and Drug Administration (FDA)-approved test for early detection of prostate cancer (PCa).<sup>14</sup> The test is based primarily on the premise that the development of PCa is associated with the disruption of the basement membrane and basal cells of the prostate. As a result, an elevated PSA concentration is observed in serum after the protein “leaks” into the circulatory system.<sup>15</sup> However, it has been reported that the PSA test has a sensitivity of 21% and a specificity of 91% for detection of any PCa when a cut-off of 4.0 ng/mL is applied.<sup>16</sup> Although the sensitivity for high-grade cancer (Gleason score  $\geq 8$ ) was reported as 51% at this cut-off value.<sup>16</sup> As a result, most PCa will not be detected in patients (false-negative) using a PSA test alone. In addition, it has been shown that high-grade cancers are found amongst men with PSA levels  $< 4.0$  ng/mL.<sup>17</sup> Thus, the test is unable to differentiate malignant and benign forms of the disease. Furthermore, an elevated PSA concentration in serum may also be the result of other factors and conditions, such as benign prostate hyperplasia (BPH).<sup>18</sup> Thus, when the cut-off is lowered to 3.0 ng/mL (32% sensitivity and 85% specificity for detection of any PCa), the likelihood of a false-positive diagnosis as well as overdiagnosis increases.<sup>16</sup> Overall, the poor performance of this biomarker in these important areas leads to underdiagnosis (false-negatives),<sup>16</sup> whereby the malignant disease is not detected early enough in order to trigger an effective clinical intervention, and overdiagnosis (false-positives).<sup>19</sup> As a result, patients with a benign or non-cancerous condition undergo an invasive and unnecessary biopsy procedure.

A single biomarker approach such as the PSA test is an appealing strategy as concentration cut-off values are straightforward to implement.<sup>20</sup> Evidently, however, there are pitfalls associated with this method which may arise due to a failure to take protein complexity into account. Proteome complexity mainly arises due to post-

transcriptional and post-translational events such as alternative splicing<sup>21</sup> as well as post-translational modifications (PTMs)<sup>22</sup> which may generate multiple proteoforms of the same protein that is transcribed from a single gene.<sup>23,24</sup> Thus, single biomarker assays, including many assays using antibody-based detection, generally do not distinguish between various proteoforms which may lead to inaccurate results.<sup>25</sup>

### Opportunities for Discovery

The challenges posed by proteoforms for clinical testing also present future opportunities for biomarker discovery and development. It has been postulated that improvements in clinical test performance will be achieved when proteoforms are incorporated into clinical assays.<sup>23,26</sup> For example, pathophysiological changes associated with cancer cell development result in changes to various proteoforms, including PTMs such as glycosylation, in addition to differences in protein expression levels. Glycosylation is a critical PTM as correct glycosylation is required for protein folding, interactions between receptors and ligands, cellular signaling and recognition, and immune response.<sup>27</sup> Thus, oncogenesis is associated with aberrant glycosylation as the expression of regulatory enzymes such as glycosyltransferases as well as the availability of monosaccharide residues is impacted.<sup>28–33</sup> As a result, the interaction between the onset of cancer and glycosylation also represents potential opportunities for biomarker discovery and development.<sup>30,31</sup> In fact, the majority of FDA-approved cancer biomarkers currently implemented in clinical practice are glycoproteins,<sup>34</sup> although the glycosylation-specific features are often not taken into account during clinical testing.<sup>35</sup> In any case, this has given rise to the development of glycosylation-focused “omics” fields, namely the glycoproteome and glycome,<sup>36</sup> that may be explored in their own right as biomarkers of disease.

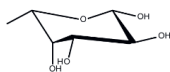
## Glycosylation

Glycosylation refers to the process of modifying a carrier protein or lipid with carbohydrate molecules. In this process, sugar building blocks, also known as monosaccharides, are used to form larger and more complex carbohydrate structures, better known as oligosaccharides or glycans. These structures consist of a series of monosaccharides covalently linked together via glycosidic bonds. As shown in **Figure 1A**, there are seven monosaccharides that are commonly found throughout the human

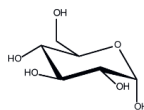
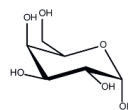
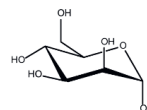


**A****Deoxyhexose (F)**

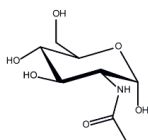
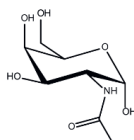
M: 146.0579 Da


  
 α-L-fucose  
 (Fuc)
**Hexose (H)**

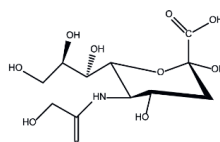
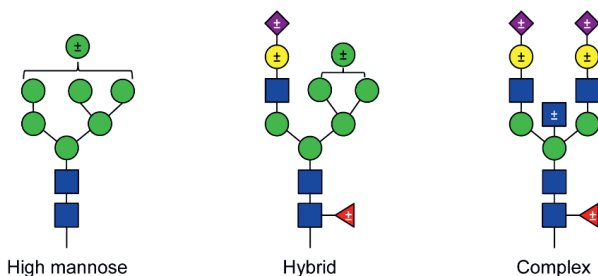
M: 162.0528 Da


  
 α-D-glucose  
 (Glc)

  
 α-D-galactose  
 (Gal)

  
 α-D-mannose  
 (Man)
**N-acetylhexosamine (N)**

M: 203.0794 Da


  
 α-N-acetyl-D-glucosamine  
 (GlcNAc)

  
 α-N-acetyl-D-galactosamine  
 (GalNAc)
**N-acetylneuraminic acid (S)**

M: 291.0954 Da


  
 α-N-acetyl-D-neuraminic acid  
 (Neu5Ac)
**B**

**Figure 1. Most common monosaccharides found in humans.** (A) Monosaccharides are shown in the alpha ( $\alpha$ ) conformation. Abbreviations for each group of sugars that share the same monoisotopic mass (M) as well as each individual sugar residue are shown. In this case, fucose (F) and sialic acid (S) are the primary deoxyhexose and *N*-acetylneuraminic acid found in humans, respectively. (B) The three glycan types, high mannose, hybrid and complex are shown. The symbol “ $\pm$ ” denotes that monosaccharides may be further added or subtracted from the structure. The colored symbol for each monosaccharide is depicted according to the Consortium of Functional Glycomics.<sup>102</sup>

body as part of different glycoconjugates whereby the term “conjugate” refers to another compound, often a protein or a lipid, that is decorated with glycan structures. In the case of proteins, the glycans may be *N*- or *O*-linked. The former refers to the attachment of the glycan to the nitrogen atom of an asparagine (Asn) when the motif Asn-X-serine (Ser)/threonine (Thr) is found in the amino acid sequence whereby “X”

is any amino acid except proline (Pro). An *O*-linked glycan refers to glycans that are bound to the protein via the oxygen atom of Ser or Thr. The sequences of Asn-X-Ser/Thr and Ser/Thr are referred to as the glycosylation sites of the protein and there may be multiple *N*- and *O*-linked glycosylation sites present on a single protein. Furthermore, different glycosylation sites may also have different occupation levels. Thus, therein lies the complexity that is found on glycoproteins as macroheterogeneity (multiple glycosylation sites) and microheterogeneity (different glycans found on the same glycosylation site) contribute to the formation of different glycoforms of the same protein.<sup>37</sup>

In eukaryotes, glycosylation begins in the endoplasmic reticulum (ER). In order to produce *N*-linked glycoproteins, first a lipid-linked oligosaccharide is built by linking multiple mannoses and glucoses to a GlcNAc that is priorly attached to the lipid. Following this, the glycan portion is transferred to a polypeptide chain whereby it ensures proper folding during post-translation modification of the protein before the entire glycoconjugate is translocated to the Golgi apparatus. Depending on the location of the glycoprotein in the Golgi apparatus (*cis*-, *medial*-, or *trans*-) as well as the presence of different glycosyltransferases, exoglycosidases, and endoglycosidases, **Figure 1B** shows that different glycan structures may be produced (high mannose, hybrid, and complex) before the glycoprotein is secreted by the cell.<sup>37</sup>

Glycans are structurally more complex than other macromolecules such as proteins owing to their non-linear production and branching nature. For example, polysaccharides are built by linking a monosaccharide to any available hydroxyl oxygen of another sugar via its anomeric carbon. Furthermore, glycosidic linkages may be present in an alpha ( $\alpha$ ) or beta ( $\beta$ ) configuration. As a result, this gives rise to structures that may contain different compositions (monosaccharides), linkages, configurations ( $\alpha$ - or  $\beta$ -), and branching. Thus, the structural diversity found within glycans gives rise to their numerous functions and roles in biological systems, including molecular structure and organization, as well as intrinsic and extrinsic signaling and recognition.<sup>37</sup> For instance, glycans with the same composition may have different linkages or structures (isomers) that affect their function. Sialic acids may mask or allow access to an underlying galactose for binding depending on whether they are  $\alpha$ 2,6- or  $\alpha$ 2,3-linked.<sup>38</sup> For example,  $\alpha$ 2,6-linked sialylation allows interaction with the asialoglycoprotein receptor whereas the  $\alpha$ 2,3-linked variant blocks

this interaction.<sup>39</sup> In addition, a galactosylated  $\alpha 6$ -antenna, rather than the  $\alpha 3$ -antenna, increases monoclonal antibody Fc $\gamma$ RIIIA binding affinity.<sup>40</sup> Finally, it has been shown that core-fucosylation plays a role in regulating the immune system<sup>41–43</sup> whereas antennary fucosylation may function more in mediating host–microbe interactions.<sup>44,45</sup>

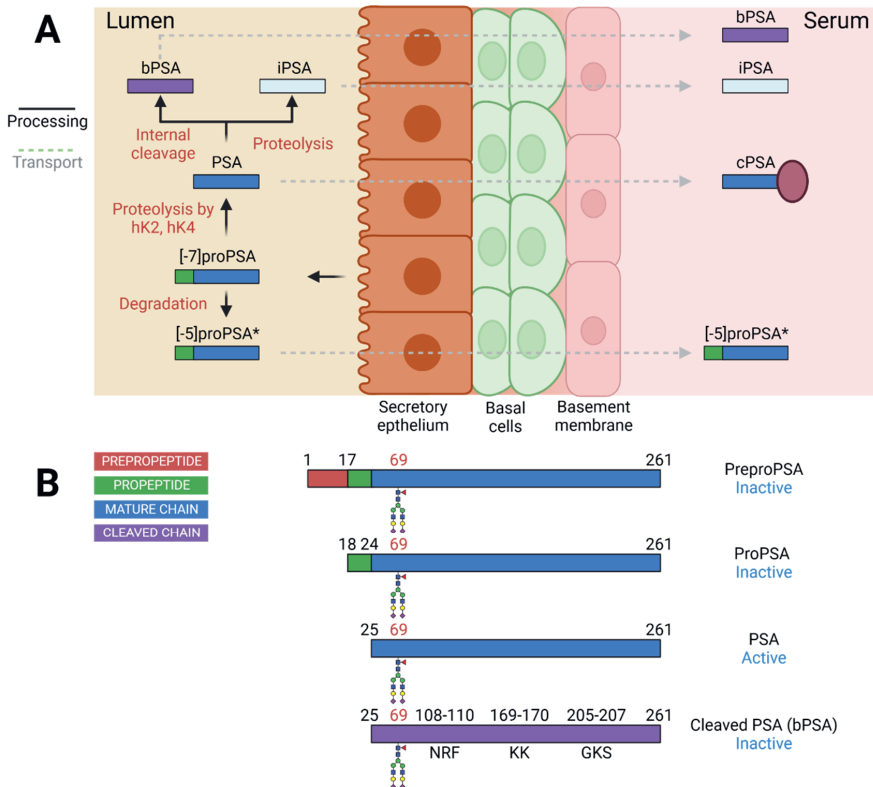
## Glycoproteins as Cancer Biomarkers

The importance of glycosylation is evident in the malignant and physiological changes that result in aberrant glycosylation profiles during cancer progression. It has been shown that not only abnormal glycosylation patterns are generated, but that glycosylation actively contributes to tumor proliferation, metastasis and angiogenesis.<sup>46,47</sup> As a result, aberrant glycosylation is continually being recognized as a hallmark of cancer. Undoubtedly there are a plethora of cancer-associated physiological alterations that affect cellular glycosyltransferases and glycosylation profiles, and vice versa. However, although glycosylation shows strong disease indications and greater promise for biomarker development, the glycosylation of human proteins and their clinical validity have not yet been thoroughly defined.<sup>35</sup> As a result, characteristic glycosylation signatures in cancer still require further investigation and, importantly, suitable methodology to measure these changes must be developed.

### PSA Proteoforms as PCa Biomarkers

PSA is a glycoprotein with a single *N*-glycosylation (Asn<sub>69</sub>) and, although the PSA test underperforms in relation to PCa screening and diagnosis, its various glycoforms have been associated with differentiating specific groups of PCa patients. In relation to this, specific PSA glycosylation signatures such as an increase in both  $\alpha 2,3$ -sialylation<sup>48</sup> and fucosylation<sup>49</sup> were correlated with the distinction between PCa and non-PCa groups, as well as normal and tumor-derived cells, respectively. Furthermore, other possibilities for biomarker development may be found by examining the disruption of the biological processing of PSA into its various proteoforms during prostate disease. **Figure 2** summarizes the PSA proteoforms that are found in serum during normal bioprocessing of the protein in healthy cells. The prepropeptide form of PSA is produced in epithelial cells where it is co-translationally cleaved at Ala<sub>17</sub> into [-7]proPSA PSA,<sup>50,51</sup> an inactive precursor form that contains a 7-amino acid long

propeptide, proPSA ([-7]proPSA).<sup>52</sup> Cleavage of [-7]proPSA at Arg<sub>24</sub> in the lumen results in the active form of the protein (PSA), which is normally secreted by the epithelial cells into semen. However, in cases of PCa, the epithelial lining of the prostate becomes disrupted, resulting in the entry of active PSA and inactive [-7]proPSA, as well as its constituent degradation products ([-5]proPSA, [-4]proPSA, [-



**Figure 2. Normal bioprocessing of PSA.** (A) The prepropeptide form of PSA is cotranslationally cleaved into [-7]proPSA.<sup>50,51</sup> [-7]proPSA may also undergo further degradation which results in truncated proPSA forms, including [-5]proPSA. In the lumen,<sup>50-52</sup> [-7]proPSA undergoes proteolysis by hK2 and hK4 to obtain PSA. This active form of the protein is secreted by the prostate gland into seminal fluid, however it may also undergo further processing in the lumen to inactivate the protein, yielding internally cleaved PSA (bPSA) and inactive PSA (iPSA).<sup>50,51</sup> A small portion of each proteoform also enters into blood circulation whereby active PSA forms complexes with protease inhibitors which produces complexed PSA (cPSA).<sup>50,51</sup> Asterisk (\*) refers to other proPSA forms such as [-4], [-3], [-2], and [-1]proPSA. (B) PSA proteoforms include preproPSA with signal peptide, [-7]proPSA, active PSA, and cleaved bPSA.<sup>50</sup> Amino acid positions are highlighted along the top. The N-glycosylation site is highlighted in red and the most abundant glycan (H5N4F1S2) is shown. The most prominent cleavage sites and associated amino acids are shown for bPSA.<sup>50,51</sup> Created with BioRender.com.

2]proPSA, [-1]proPSA) into the bloodstream.<sup>50–52</sup> Conversely, active PSA may also be inactivated by undergoing internal cleavage at specific amino acid positions (R<sub>109</sub> – F<sub>110</sub>, K<sub>169</sub> – K<sub>170</sub>, K<sub>206</sub> – S<sub>207</sub>)<sup>53–55</sup> to produce inactive cleaved PSA<sup>54</sup>, also known as benign PSA (bPSA)<sup>56</sup>. Interestingly, bPSA appears to be enriched in the transition zone of the prostate in hyperplastic BPH tissue which is thought to be a highly proteolytic environment.<sup>56</sup> Thus, the impact that disease has on the biological processing of PSA gives a clear example of how specific proteoforms may be investigated in order to improve the clinical performance of a diagnostic test when testing for PCa (proPSA) or BPH (bPSA). Despite these developments, comprehensive studies of PSA proteoforms, especially in body fluids other than serum, are lacking. Although PSA variants have been examined with some beneficial outcomes,<sup>51</sup> these tests are still dependent on protein concentration, rather than a proteoform profile. As a result, further studies into the clinical relevance of PSA proteoforms must still be carried out.

### **CRC N-Glycosylation Signatures**

In relation to colorectal cancer (CRC), aberrant *N*-glycosylation signatures may be found during CRC progression. For example, a study that used CRC stem cells showed an increase in lectin binding for branched tri- and tetra-antennary structures, which was accompanied by an increase in the expression of *N*-acetylglucosaminyltransferase V (MGAT5),<sup>57</sup> the enzyme responsible for  $\beta$ 1,6-branching.<sup>58</sup> Interestingly, MGAT5 overexpression is associated with an increase in malignancy and tumor growth rate in human colon carcinoma cells.<sup>59</sup> In addition, upregulation of ST6 beta-galactoside  $\alpha$ -2,6-sialyltransferase 1 (ST6GAL1)<sup>60,61</sup> and a concurrent increase in expression of  $\alpha$ 2,6-sialylation is reported in CRC.<sup>62</sup> Here, sialylation modifies  $\beta$ 1 integrin conformation, resulting in a protective effect against apoptosis in CRC cells.<sup>63</sup> Additionally, the aforementioned glycosylation signatures may also be investigated as biomarkers of the disease. For example, Vroome *et al.* showed an increase in tri- and tetra-antennary *N*-glycan structures as well as an overall elevation in  $\alpha$ 2,6-sialylation when comparing the total serum *N*-glycome of CRC patients with healthy controls.<sup>64</sup> Furthermore, the authors illustrated the discrimination of cases from controls and CRC survival was also predicted.<sup>64</sup> Another study focused more specifically on the glycosylation patterns of the current CRC biomarker, carcinoembryonic antigen (CEA).<sup>65</sup> CEA is a highly glycosylated protein

and it is expected that the assessment of its heterogeneous and abundant glycosylation could result in biomarkers that better reflect cancer progression.<sup>66</sup> In this case, Zhao *et al.* examined CEA from tumor tissues in comparison with paired tumor-adjacent normal tissues using a lectin microarray.<sup>65</sup> Greater levels of fucose and mannose were observed whilst there was a decrease in *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), galactose as well as branched and bisecting *N*-glycans on CEA.<sup>65</sup> Evidently, glycosylation is involved throughout numerous cancer processes in CRC, however further investigations are required in order to determine the most suitable approach (total serum *versus* specific glycoprotein), glycosylation signature, as well as analytical methodology that will enable translation of the results into the clinics.

The discovery and development of glycoproteins as biomarkers for clinical use requires suitable analytical methodology for this purpose. In this case, the technique must demonstrate sufficient sensitivity in order to detect low abundant analytes as well as specificity so that associations of specific analytes with the disease may be determined. In relation to the latter, antigen-binding approaches using antibody or lectin arrays are fraught with the potential for non-specific binding.<sup>25,35</sup> For example, Zhao *et al.* reported that *Aleuria aurantia* lectin (AAL) shows specificity for  $\alpha$ 1-2 (antennary) fucose<sup>65</sup> whereas other studies describe specificity towards  $\alpha$ 1-6 (core) fucose,<sup>67</sup> highlighting the ambiguity that may arise from such approaches. In contrast, measurands used in clinical tests must be clearly defined.<sup>68</sup> Furthermore, it is important to develop a proper understanding of the disease biology in order to develop biomarkers that can accurately monitor disease processes. Thus, it is important to build the required specificity into the biomarker test throughout the discovery and development period. In this sense, mass spectrometry (MS) is a technique that is ideally suited for this purpose as it allows specific identification via unique precursor mass and fragmentation patterns as well as direct quantification of proteins.<sup>35,69</sup>

## Glyco(proteo)mics

### Mass Spectrometry

MS describes a range of techniques wherein gas-phase ions of compounds are produced followed by the measurement of these ions as mass-to-charge ratios ( $m/z$ ). The process of ionization produces gas-phase ions in the source, often via matrix-assisted laser desorption ionization (MALDI) or electrospray ionization (ESI). This is followed by mass separation of the ions in the mass analyzer under a vacuum. This is a critical function of MS as the ions are directly related to the nature, structure, and composition of the precursor molecule. Several types of mass analyzers have been developed for this purpose and, in many cases, they are combined in order to enhance the selectivity of the measurement. For example, quadrupole time-of-flight (QTOF) instruments implement a mass range pre-selection via the quadrupole which is followed by the separation of the ions inside a flight tube. The  $m/z$  of the ion is determined by the time taken to reach the detector, which is directly related to its mass. Finally, the data output from the detector is given as a mass spectrum, which may then be interpreted manually or using appropriate software in order to yield important information regarding the molecule of interest.<sup>70</sup>

The power of MS can be further enhanced by adding an extra separation dimension prior to mass detection. This may be achieved via the coupling of an online separation technique before the introduction of the sample into the mass spectrometer. With regard to the analysis of glycoproteins, two separation techniques have demonstrated a wide range of applications when hyphenated with MS: capillary electrophoresis (CE)–ESI–MS<sup>66,71–77</sup> and liquid chromatography (LC)–ESI–MS.<sup>78–84</sup>

### Capillary Electrophoresis–Electrospray Ionization–Mass Spectrometry

Electrophoresis refers to the separation of compounds in solution following the application of voltage in order to generate an electric field.<sup>85</sup> Thus, analytes will migrate towards the cathode or anode depending on whether they are positively or negatively charged, respectively. In the case of capillary zone electrophoresis (CZE), the capillary is filled with a buffer known as the background electrolyte (BGE) in order to facilitate separation.<sup>86</sup> Depending on the charge of the analytes and their hydrodynamic size, each analyte displays its own electrophoretic mobility. In this case, an analytes charge

is determined based on its pKa as well as the pH, ionic strength, and concentration of the BGE.<sup>87</sup> However, the migration of each analyte within the capillary depends on the combined influence of its electrophoretic mobility as well as the electroosmotic flow (EOF) of the system. The EOF is generated by applying a voltage to the capillary and an interaction between cations, anions, and the capillary wall will occur, resulting in a bulk flow of the solution. In relation to this, the strength of the EOF may be influenced by controlling parameters such as the properties of the BGE, the coating present on the capillary internal wall, and the applied voltage.<sup>87</sup> Moreover, the resulting charge of the compound of interest and direction of the EOF must also be carefully considered to ensure that the analyte migrates efficiently towards the detector.

CE is regarded as an analytical technique which generates excellent peak resolution and high peak capacity, mainly because of the sharp peaks that are produced due to the flow profile that is EOF driven.<sup>87</sup> As a result, the hyphenation of CE with MS via ESI is a powerful approach which achieves high sensitivity and selectivity by the online separation of analytes based on their properties inside the capillary and used BGE. In the context of glycoproteomics, CE-ESI-MS has demonstrated efficient separation of sialylated glycans due to the negative charge that sialic acids carry.<sup>88</sup> Furthermore, isomeric separation can also be achieved as it was shown that sialic acids with different linkages are resolved as they have different pKas,<sup>88</sup> and likely also different hydrodynamic volumes, although the latter observation was not investigated. In addition, co-migrating compounds with different masses may be determined when coupled with MS detection. Overall, CE-ESI-MS has been applied across multiple areas in the glycomics and glycoproteomics fields, including the analysis of released glycans,<sup>71–73</sup> glycopeptides,<sup>74,75</sup> and intact glycoproteins.<sup>76,77</sup>

### **Liquid Chromatography–Electrospray Ionization–Mass Spectrometry**

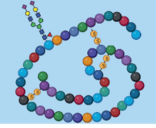

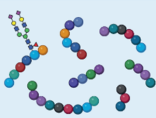

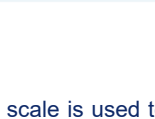
Chromatography achieves separation based on the interaction of compounds with the stationary and mobile phases. In this regard, the chemistry that is observed is largely dependent on the chromatographic system and the properties of the analyte inside that system. In relation to this, there are several chromatographic principles that are mainly applied in glycomic and glycoproteomics research, including high-pH/performance anion-exchange chromatography (HPAEC),<sup>89</sup> porous graphitized carbon (PGC)-LC,<sup>78,83</sup> reversed-phase (RP)-LC,<sup>79,80</sup> and hydrophilic interaction-LC



(HILIC).<sup>81,90</sup> In regard to HPAEC, glycans are separated following the application of a high pH which results in the deprotonation of monosaccharide hydroxyl groups.<sup>91</sup> In particular, this technique has been employed with pulsed-anomeric detection (PAD) for the analysis of individual monosaccharides following glycan hydrolysis.<sup>89</sup> Typically, this separation approach is difficult to couple with MS due to the incompatibility of the solvents, however hyphenation has been achieved via implementing an online desalting step prior to MS detection.<sup>82</sup> In general, the latter LC approaches (PGC, RP, HILIC) are more easily coupled with MS than HPAEC and have demonstrated numerous applications within the field of glycomics and glycoproteomics.<sup>78–81,83,84,92</sup> For example, hydrophobic and polar interactions influence PGC separation, as a result this technique demonstrates excellent separation of glycan isomers.<sup>92,93</sup> In the context of RPLC, a nonpolar stationary phase is employed and a noncovalent interaction is formed between the stationary phase and the analytes' nonpolar residues. Analytes elute in order of increasing hydrophobicity as their retention is dependent on the competitive solubilization between the stationary and mobile phases.<sup>91</sup> In the context of *N*-glycan analysis, elution is largely dominated by non-carbohydrate moieties at the reducing-end of the glycan.<sup>91</sup> Finally, HILIC is also a popular technique for separating glycoconjugates due to their hydrophilic nature and, as a result, separation is mainly determined by the glycans themselves. In this case, it has been reported that an analyte partitioning effect is created due to the presence of organic solvent in the mobile phase and the water-enriched layer at the polar stationary phase.<sup>94,95</sup> Thus, analyte elution is strongly influenced by the polar associations of the carbohydrate residues with the stationary phase, as well as other interactions which include hydrogen bonding, ion exchange and dipole-dipole interactions.<sup>94</sup> Therefore, HILIC-MS has been an important technique for the efficient separation and analysis of various types of glycoconjugates.

### **Top to Bottom MS Approaches**

There are several approaches for analyzing glycoproteins by MS which range from top-down to bottom-up techniques (**Figure 3**). The approach for analyzing glycoproteins depends on the research question and level of detail that is required. For example, intact glycoprotein analysis involves the analysis of the entire glycoconjugate, with little sample preparation and, as a result, fewer modifications are introduced to the molecule during sample processing.<sup>96</sup> The separation may be

	Analysis time	Sample purity	Protein specificity		Glycan heterogeneity	
			Proteform	Site		
Intact protein / Top-down	+	+++	+++	-	+	
Middle-up / down	++	+++	+++	+	+	
Bottom-up	+++	+	++	+++	++	
Released glycan	+++	+	-	-	+++	
	-	+	++	+++		
	None	Low	Medium	High		

**Figure 3. Top to bottom glyco(proteo)mic approaches.** An ordinal scale is used to compare the four approaches across five categories, ranging from none to high. Analysis time refers to total time to perform sample preparation, measurement and data analysis. *Created with BioRender.com.*

performed under denaturing or native conditions with the aim of assessing the intact mass of different proteoforms as well as protein integrity and complexes when native conditions are employed.<sup>97</sup> Importantly, it should be noted that top-down analysis includes the performance of MS/MS fragmentation on the protein in order to gain sequence information.<sup>97</sup> However, intact protein or top-down analysis requires mainly isolated proteins and these techniques does not allow site-specific information to be gained. Furthermore, sensitivity is a challenge due to the dispersion of the signal over the charge envelope of the protein. Native top-down MS generates proteins with lower charge states,<sup>98</sup> however the aqueous solvent lacks the organic solvent and low pH that is often favorable for protein solubility and ion desolvation.<sup>98</sup>

Subunit analysis may also be carried out by following a middle-up or middle-down approach. Protein subunits may be generated by reducing protein disulfide bonds or enzymatic digestion, such as IdeS digestion of antibodies.<sup>81</sup> Similarly to the top-down approach mentioned above, the term “middle-down” is used when MS/MS

experiments are carried out during the measurement.<sup>97</sup> Several benefits are gained from following a protein subunit approach. For instance, the molecular weight of the analyte is lowered and, therefore, isotopic resolution may be achieved which allows the accurate mass to be determined. In addition, modifications may be assigned to specific subunits of the protein.<sup>99</sup> Despite these advantages, enzymes that cleave the protein into large polypeptides are generally not available for most proteins.<sup>98</sup>

Bottom-up analysis involves the enzymatic digestion of the protein and, unlike middle-up/down analysis, the protein is digested into peptides rather than subunits. Moreover, there is a large number of enzymes to choose from in order to generate smaller peptides, although trypsin is the most commonly applied.<sup>100</sup> This approach generates great detail about the glycoprotein of interest, including the identification of *N*-glycosylation sites and glycosylation site-occupancy, provided the peptide backbone is unique. Although glycopeptide analysis provides information regarding both the peptide and the glycan, the technique generates highly complex spectra. In addition, little information regarding the peptide sequence is gained from applying collisional-induced dissociation (CID), a commonly used technique for glycosidic bond fragmentation, and other fragmentation techniques should be utilized in this case.<sup>100</sup>

Finally, *N*-glycans may also be released entirely from the glycoprotein via enzymes such as an endoglycosidases. In this case, the complete range of glycan microheterogeneity may be assessed. This approach has several advantages, including the absence of any protein interference which produces spectra with less complexity. However, as a result, information regarding the carrier protein is lost. Furthermore, glycans possess poor ionization efficiency and do not contain a natural chromophore. In this case, sample preparation of glycans often involves reducing-end labeling with a fluorophore-containing compound in order to enable fluorescence detection.<sup>101</sup>

As demonstrated in **Figure 3**, each approach shows that there are several factors that should be considered when determining which tactic to follow, including the sample purity, time required for performing sample preparation and measurements, the modification of interest, and the complexity of the generated data. Alternatively, to gain a comprehensive picture, multiple approaches can be applied. For example, in cases where the assessment of different glycoprotein proteoforms via a top-down approach

may be the primary objective, bottom-up analysis would provide the necessary complimentary information regarding glycosylation site identification and occupancy determination. In addition, released glycan analysis can provide the full overview of glycan microheterogeneity and abundance. Moreover, subunit analysis provides further support regarding modifications to the protein sequence. Thus, a multilevel approach generates the most complete overview of the glycoprotein.

## Scope

The aim of this thesis is to explore and develop MS workflows coupled with online separation for the discovery of glyco(proteo)mic signatures in cancer. In this case, two different approaches are taken in order to explore PCa and CRC, both in terms of sample processing and separation mode.

In the first part of this thesis, we focus on a specific glycoprotein, namely PSA, and perform intact protein analysis in order to assess the potential of its proteoforms for stratifying PCa patients. **Chapter 2** describes the development of a CE method in combination with MS detection in order to profile urinary intact PSA. In addition, the performance and application of the method is validated via intermediate and repeatability measurements as well as a small patient study. In **Chapter 3**, the data processing workflow is further developed, overcoming one of the main challenges that was identified in Chapter 2. Importantly, this chapter focuses on the same data set as Chapter 2 in order to validate the findings by comparing extracted ion and deconvolution quantification approaches. New insights are also obtained as the profiles of urinary and seminal PSA are compared, and further support for proteoform assignment is achieved via middle-up and bottom-up analyses.

In the second part of this thesis, the potential of total serum *N*-glycosylation as a biomarker for CRC is evaluated. **Chapter 4** describes the development and validation of a RPLC method in order to separate fluorescently labeled and sialic acid linkage derivatized *N*-glycans. Importantly, specific isomeric structures are differentiated by retention time and mass differences. Following this, the application of the method is demonstrated using a cohort of pre- and post-operative samples from CRC patient sera in **Chapter 5**. The clinical results are corroborated by comparison of the results with previous MALDI-MS measurements of the same cohort. Additional information is also gained by the separation of isomeric structures that was previously not observed.

Finally, a general discussion is provided in **Chapter 6**. Here, the analysis of isomeric *N*-glycans by RPLC is further examined. In addition, the technical challenges encountered during this thesis and potential improvements to the methodology are discussed. Finally, future perspectives regarding biomarker and method translation are described.

## References

- (1) Hanahan, D.; Weinberg, R. A. The Hallmarks of Cancer. *Cell*. **2000**, *100* (1). DOI: 10.1016/S0092-8674(00)81683-9.
- (2) Fouad, Y. A.; Aanei, C. Revisiting the Hallmarks of Cancer. *American Journal of Cancer Research*. **2017**, *7* (5).
- (3) Karapetis, C. S.; Khambata-Ford, S.; Jonker, D. J.; O'Callaghan, C. J.; Tu, D.; Tebbutt, N. C.; Simes, R. J.; Chalchal, H.; Shapiro, J. D.; Robitaille, S.; Price, T. J.; Shepherd, L.; Au, H.-J.; Langer, C.; Moore, M. J.; Zalcborg, J. R. K-Ras Mutations and Benefit from Cetuximab in Advanced Colorectal Cancer. *New England Journal of Medicine*. **2008**, *359* (17). DOI: 10.1056/nejmoa0804385.
- (4) Chan, I. S.; Ginsburg, G. S. Personalized Medicine: Progress and Promise. *Annual Review of Genomics and Human Genetics*. **2011**, *12*. DOI: 10.1146/annurev-genom-082410-101446.
- (5) Abrahams, E. Right Drug-Right Patient-Right Time: Personalized Medicine Coalition. *Clinical and Translational Science*. **2008**, *1* (1). DOI: 10.1111/j.1752-8062.2008.00003.x.
- (6) Atkinson, A. J.; Colburn, W. A.; DeGruttola, V. G.; DeMets, D. L.; Downing, G. J.; Hoth, D. F.; Oates, J. A.; Peck, C. C.; Schooley, R. T.; Spilker, B. A.; Woodcock, J.; Zeger, S. L. Biomarkers and Surrogate Endpoints: Preferred Definitions and Conceptual Framework. *Clinical Pharmacology and Therapeutics*. **2001**, *69* (3). DOI: 10.1067/mcp.2001.113989.
- (7) Borrebaeck, C. A. K. Precision Diagnostics: Moving towards Protein Biomarker Signatures of Clinical Utility in Cancer. *Nature Reviews Cancer*. **2017**, *17* (3). DOI: 10.1038/nrc.2016.153.
- (8) Ilyin, S. E.; Belkowski, S. M.; Plata-Salamán, C. R. Biomarker Discovery and Validation: Technologies and Integrative Approaches. *Trends in Biotechnology*. **2004**, *22* (8). DOI: 10.1016/j.tibtech.2004.06.005.
- (9) Nishizuka, S. et al. Diagnostic Markers That Distinguish Colon and Ovarian Adenocarcinomas: Identification by Genomic, Proteomic, and Tissue Array Profiling. *Cancer Research*. **2003**, *63* (17). DOI: 10.1097/00125480-200403000-00009.
- (10) Zhan, X.; Long, Y.; Lu, M. Exploration of Variations in Proteome and Metabolome for Predictive Diagnostics and Personalized Treatment Algorithms: Innovative Approach and Examples for Potential Clinical Application. *Journal of Proteomics*. **2018**, *188*. DOI: 10.1016/j.jprot.2017.08.020.
- (11) Holland, R. L. What Makes a Good Biomarker? *Advances in Precision Medicine*. **2016**, *1* (1). DOI: 10.18063/apm.2016.01.007.
- (12) Shreffler, J.; Huecker, M. R. *Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios*; **2020**.
- (13) Lalkhen, A. G.; McCluskey, A. Clinical Tests: Sensitivity and Specificity. *Continuing Education in Anaesthesia, Critical Care and Pain*. **2008**, *8* (6). DOI: 10.1093/bjaceaccp/mkn041.

- (14) De Angelis, G.; Rittenhouse, H. G.; Mikolajczyk, S. D.; Blair Shamel, L.; Semjonow, A. Twenty Years of PSA: From Prostate Antigen to Tumor Marker. *Reviews in urology*. **2007**, *9* (3).
- (15) Lilja, H.; Ulmert, D.; Vickers, A. J. Prostate-Specific Antigen and Prostate Cancer: Prediction, Detection and Monitoring. *Nature Reviews Cancer*. **2008**, *8* (4). DOI: 10.1038/nrc2351.
- (16) Wolf, A. M. D.; Wender, R. C.; Etzioni, R. B.; Thompson, I. M.; D'Amico, A. V.; Volk, R. J.; Brooks, D. D.; Dash, C.; Guessous, I.; Andrews, K.; DeSantis, C.; Smith, R. A. American Cancer Society Guideline for the Early Detection of Prostate Cancer: Update 2010. *CA: A Cancer Journal for Clinicians*. **2010**, *60* (2). DOI: 10.3322/caac.20066.
- (17) Thompson, I. M.; Pauler, D. K.; Goodman, P. J.; Tangen, C. M.; Lucia, M. S.; Parnes, H. L.; Minasian, L. M.; Ford, L. G.; Lippman, S. M.; Crawford, E. D.; Crowley, J. J.; Coltman, C. A. Prevalence of Prostate Cancer among Men with a Prostate-Specific Antigen Level  $\leq 4.0$  Ng per Milliliter. *New England Journal of Medicine*. **2004**, *350* (22). DOI: 10.1056/nejmoa031918.
- (18) Monda, J. M.; Barry, M. J.; Oesterling, J. E. Prostate Specific Antigen Cannot Distinguish Stage T1a (A1) Prostate Cancer from Benign Prostatic Hyperplasia. *Journal of Urology*. **1994**, *151* (5). DOI: 10.1016/S0022-5347(17)35234-5.
- (19) Carter, H. B. Prostate Cancers in Men with Low PSA Levels — Must We Find Them? *New England Journal of Medicine*. **2004**, *350* (22). DOI: 10.1056/nejme048003.
- (20) van der Burgt, Y. E. M. Protein Biomarker Discovery Is Still Relevant and Has Entered a New Phase. *EBioMedicine*. **2019**, *43*. DOI: 10.1016/j.ebiom.2019.04.026.
- (21) Blencowe, B. J. Alternative Splicing: New Insights from Global Analyses. *Cell*. **2006**, *126* (1). DOI: 10.1016/j.cell.2006.06.023.
- (22) Prabakaran, S.; Lippens, G.; Steen, H.; Gunawardena, J. Post-Translational Modification: Nature's Escape from Genetic Imprisonment and the Basis for Dynamic Information Encoding. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*. **2012**, *4* (6). DOI: 10.1002/wsbm.1185.
- (23) van der Burgt, Y. E. M.; Cobbaert, C. M. Proteoform Analysis to Fulfill Unmet Clinical Needs and Reach Global Standardization of Protein Measurands in Clinical Chemistry Proteomics. *Clinics in Laboratory Medicine*. **2018**, *38* (3). DOI: 10.1016/j.cll.2018.05.001.
- (24) Kelleher, N. L. A Cell-Based Approach to the Human Proteome Project. *Journal of the American Society for Mass Spectrometry*. **2012**, *23* (10). DOI: 10.1007/s13361-012-0469-9.
- (25) Hoofnagle, A. N.; Wener, M. H. The Fundamental Flaws of Immunoassays and Potential Solutions Using Tandem Mass Spectrometry. *Journal of Immunological Methods*. **2009**, *347* (1–2). DOI: 10.1016/j.jim.2009.06.003.
- (26) Kelleher, N. L.; Thomas, P. M.; Ntai, I.; Compton, P. D.; Leduc, R. D. Deep and Quantitative Top-down Proteomics in Clinical and Translational Research. *Expert Review of Proteomics*. **2014**, *11* (6). DOI: 10.1586/14789450.2014.976559.
- (27) Lauc, G.; Rudan, I.; Campbell, H.; Rudd, P. M. Complex Genetic Regulation of Protein Glycosylation. *Molecular BioSystems*. **2010**, *6* (2). DOI: 10.1039/b910377e.

- (28) Hakomori, S. Glycosylation Defining Cancer Malignancy: New Wine in an Old Bottle. *Proceedings of the National Academy of Sciences of the United States of America*. **2002**, *99* (16). DOI: 10.1073/pnas.172380699.
- (29) Silva, M. L. S. Cancer Serum Biomarkers Based on Aberrant Post-Translational Modifications of Glycoproteins: Clinical Value and Discovery Strategies. *Biochimica et Biophysica Acta - Reviews on Cancer*. **2015**, *1856* (2). DOI: 10.1016/j.bbcan.2015.07.002.
- (30) Reis, C. A.; Osorio, H.; Silva, L.; Gomes, C.; David, L. Alterations in Glycosylation as Biomarkers for Cancer Detection. *Journal of Clinical Pathology*. **2010**, *63* (4). DOI: 10.1136/jcp.2009.071035.
- (31) Padler-Karavani, V. Aiming at the Sweet Side of Cancer: Aberrant Glycosylation as Possible Target for Personalized-Medicine. *Cancer Letters*. **2014**, *352* (1). DOI: 10.1016/j.canlet.2013.10.005.
- (32) Vajaria, B. N.; Patel, P. S. Glycosylation: A Hallmark of Cancer? *Glycoconjugate Journal*. **2017**, *34* (2). DOI: 10.1007/s10719-016-9755-2.
- (33) Munkley, J.; Elliott, D. J. Hallmarks of Glycosylation in Cancer. *Oncotarget*. **2016**, *7* (23). DOI: 10.18632/oncotarget.8155.
- (34) Kirwan, A.; Utratna, M.; O'Dwyer, M. E.; Joshi, L.; Kilcoyne, M. Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics. *BioMed Research International*. **2015**. DOI: 10.1155/2015/490531.
- (35) de Haan, N.; Wuhrer, M.; Ruhaak, L. R. Mass Spectrometry in Clinical Glycomics: The Path from Biomarker Identification to Clinical Implementation. *Clinical Mass Spectrometry*. **2020**, *18*. DOI: 10.1016/j.clinms.2020.08.001.
- (36) Hizal, D. B.; Wolozny, D.; Colao, J.; Jacobson, E.; Tian, Y.; Krag, S. S.; Betenbaugh, M. J.; Zhang, H. Glycoproteomic and Glycomic Databases. *Clinical Proteomics*. **2014**, *11* (1). DOI: 10.1186/1559-0275-11-15.
- (37) Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Darvill, A. G.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H. *Essentials of Glycobiology, Third Edition*; **2017**.
- (38) Schauer, R. Sialic Acids: Fascinating Sugars in Higher Animals and Man. *Zoology (Jena)*. **2004**, *107* (1), 49–64. DOI: 10.1016/j.zool.2003.10.002.
- (39) Park, E. I.; Mi, Y.; Unverzagt, C.; Gabius, H. J.; Baenziger, J. U. The Asialoglycoprotein Receptor Clears Glycoconjugates Terminating with Sialic Acid $\alpha$ 2,6GalNAc. *Proceedings of the National Academy of Sciences of the United States of America*. **2005**, *102* (47). DOI: 10.1073/pnas.0508537102.
- (40) Hatfield, G.; Tepliakova, L.; Gingras, G.; Stalker, A.; Li, X.; Aubin, Y.; Tam, R. Y. Specific Location of Galactosylation in an Afucosylated Antiviral Monoclonal Antibody Affects Its Fc $\gamma$ R1IIIA Binding Affinity. *Frontiers in Immunology*. **2022**, *13*, 972168. DOI: 10.3389/fimmu.2022.972168.
- (41) Huhn, C.; Selman, M. H. J.; Ruhaak, L. R.; Deelder, A. M.; Wuhrer, M. IgG Glycosylation Analysis. *Proteomics*. **2009**, *9* (4), 882–913. DOI: <https://doi.org/10.1002/pmic.200800715>.



- (42) Pučić, M. et al. High Throughput Isolation and Glycosylation Analysis of IgG-Variability and Heritability of the IgG Glycome in Three Isolated Human Populations. *Molecular and Cellular Proteomics*. **2011**, *10* (10). DOI: 10.1074/mcp.M111.010090.
- (43) Hwang, H.; Jeong, H. K.; Lee, H. K.; Park, G. W.; Lee, J. Y.; Lee, S. Y.; Kang, Y. M.; An, H. J.; Kang, J. G.; Ko, J. H.; Kim, J. Y.; Yoo, J. S. Machine Learning Classifies Core and Outer Fucosylation of N-Glycoproteins Using Mass Spectrometry. *Scientific Reports*. **2020**, *10* (1). DOI: 10.1038/s41598-019-57274-1.
- (44) Schneider, M.; Al-Shareffi, E.; Haltiwanger, R. S. Biological Functions of Fucose in Mammals. *Glycobiology*. **2017**, *27* (7). DOI: 10.1093/glycob/cwx034.
- (45) Becker, D. J.; Lowe, J. B. Fucose: Biosynthesis and Biological Function in Mammals. *Glycobiology*. **2003**, *13* (7). DOI: 10.1093/glycob/cwg054.
- (46) Fuster, M. M.; Esko, J. D. The Sweet and Sour of Cancer: Glycans as Novel Therapeutic Targets. *Nature Reviews Cancer*. **2005**, *5* (7). DOI: 10.1038/nrc1649.
- (47) Pinho, S. S.; Reis, C. A. Glycosylation in Cancer: Mechanisms and Clinical Implications. *Nature Reviews Cancer*. **2015**, *15* (9). DOI: 10.1038/nrc3982.
- (48) Yoneyama, T.; Ohyama, C.; Hatakeyama, S.; Narita, S.; Habuchi, T.; Koie, T.; Mori, K.; Hidari, K. I. P. J.; Yamaguchi, M.; Suzuki, T.; Tobisawa, Y. Measurement of Aberrant Glycosylation of Prostate Specific Antigen Can Improve Specificity in Early Detection of Prostate Cancer. *Biochemical and Biophysical Research Communications*. **2014**. DOI: 10.1016/j.bbrc.2014.04.107.
- (49) Peracaula, R.; Tabarés, G.; Royle, L.; Harvey, D. J.; Dwek, R. A.; Rudd, P. M.; de Llorens, R. Altered Glycosylation Pattern Allows the Distinction between Prostate-Specific Antigen (PSA) from Normal and Tumor Origins. *Glycobiology*. **2003**, *13* (6). DOI: 10.1093/glycob/cwg041.
- (50) Balk, S. P.; Ko, Y. J.; Bublely, G. J. Biology of Prostate-Specific Antigen. *Journal of Clinical Oncology*. **2003**, *21* (2). DOI: 10.1200/JCO.2003.02.083.
- (51) Gilgunn, S.; Conroy, P. J.; Saldova, R.; Rudd, P. M.; O'Kennedy, R. J. Aberrant PSA Glycosylation - A Sweet Predictor of Prostate Cancer. *Nature Reviews Urology*. **2013**, *10* (2). DOI: 10.1038/nrurol.2012.258.
- (52) Mikolajczyk, S. D.; Grauer, L. S.; Millar, L. S.; Hill, T. M.; Kumar, A.; Rittenhouse, H. G.; Wolfert, R. L.; Saedi, M. S. A Precursor Form of PSA (PPSA) Is a Component of the Free PSA in Prostate Cancer Serum. *Urology*. **1997**, *50* (5). DOI: 10.1016/S0090-4295(97)00449-4.
- (53) Zhang, W. M.; Leinonen, J.; Kalkkinen, N.; Dowell, B.; Stenman, U. H. Purification and Characterization of Different Molecular Forms of Prostate-Specific Antigen in Human Seminal Fluid. *Clinical Chemistry*. **1995**, *41* (11). DOI: 10.1093/clinchem/41.11.1567.
- (54) Mattsson, J. M.; Valmu, L.; Laakkonen, P.; Stenman, U. H.; Koistinen, H. Structural Characterization and Anti-Angiogenic Properties of Prostate-Specific Antigen Isoforms in Seminal Fluid. *Prostate*. **2008**, *68* (9), 945–954. DOI: 10.1002/pros.20751.
- (55) Watt, K. W. K.; Lee, P. J.; M'Timkulu, T.; Chan, W. P.; Loor, R. Human Prostate-Specific Antigen: Structural and Functional Similarity with Serine Proteases. *Proceedings of the*

- National Academy of Sciences of the United States of America*. **1986**, 83 (10). DOI: 10.1073/pnas.83.10.3166.
- (56) Mikolajczyk, S. D.; Millar, L. S.; Wang, T. J.; Rittenhouse, H. G.; Wolfert, R. L.; Marks, L. S.; Song, W.; Wheeler, T. M.; Slawin, K. M. "BPSA," a Specific Molecular Form of Free Prostate-Specific Antigen, Is Found Predominantly in the Transition Zone of Patients with Nodular Benign Prostatic Hyperplasia. *Urology*. **2000**, 55 (1). DOI: 10.1016/S0090-4295(99)00372-6.
- (57) Guo, H.; Nagy, T.; Pierce, M. Post-Translational Glycoprotein Modifications Regulate Colon Cancer Stem Cells and Colon Adenoma Progression in Apcmin/+ Mice through Altered Wnt Receptor Signaling. *Journal of Biological Chemistry*. **2014**, 289 (45). DOI: 10.1074/jbc.M114.602680.
- (58) De Freitas-Junior, J. C. M.; Morgado-Díaz, J. A. The Role of N-Glycans in Colorectal Cancer Progression: Potential Biomarkers and Therapeutic Applications. *Oncotarget*. **2016**, 7 (15). DOI: 10.18632/oncotarget.6283.
- (59) Kim, Y. S.; Ahn, Y. H.; Song, K. J.; Kang, J. G.; Lee, J. H.; Jeon, S. K.; Kim, H. C.; Yoo, J. S.; Ko, J. H. Overexpression and  $\beta$ -1,6-N-Acetylglucosaminylation-Initiated Aberrant Glycosylation of TIMP-1: A "Double Whammy" Strategy in Colon Cancer Progression. *Journal of Biological Chemistry*. **2012**, 287 (39). DOI: 10.1074/jbc.M112.370064.
- (60) Dall'Olio, F.; Chiricolo, M.; Mariani, E.; Facchini, A. Biosynthesis of the Cancer-Related Sialyl-Alpha 2,6-Lactosaminy Epitope in Colon Cancer Cell Lines Expressing Beta-Galactoside Alpha 2,6-Sialyltransferase under a Constitutive Promoter. *European journal of biochemistry*. **2001**, 268 (22), 5876–5884. DOI: 10.1046/j.0014-2956.2001.02536.x.
- (61) Dall'Olio, F.; Malagolini, N.; Chiricolo, M. Beta-Galactoside Alpha2,6-Sialyltransferase and the Sialyl Alpha2,6-Galactosyl-Linkage in Tissues and Cell Lines. *Methods in molecular biology (Clifton, N.J.)*. **2006**, 347.
- (62) Park, J. J.; Lee, M. Increasing the  $\alpha$  2, 6 Sialylation of Glycoproteins May Contribute to Metastatic Spread and Therapeutic Resistance in Colorectal Cancer. *Gut and Liver*. **2013**, 7 (6). DOI: 10.5009/gnl.2013.7.6.629.
- (63) Zhuo, Y.; Chammas, R.; Bellis, S. L. Sialylation of Beta1 Integrins Blocks Cell Adhesion to Galectin-3 and Protects Cells against Galectin-3-Induced Apoptosis. *The Journal of biological chemistry*. **2008**, 283 (32), 22177–22185. DOI: 10.1074/jbc.M8000015200.
- (64) de Vroome, S. W.; Holst, S.; Gironde, M. R.; van der Burgt, Y. E. M.; Mesker, W. E.; Tollenaar, R. A. E. M.; Wuhrer, M. Serum N-Glycome Alterations in Colorectal Cancer Associate with Survival. *Oncotarget*. **2018**, 9 (55), 30610–30623. DOI: 10.18632/oncotarget.25753.
- (65) Zhao, Q.; Zhan, T.; Deng, Z.; Li, Q.; Liu, Y.; Yang, S.; Ji, D.; Li, Y. Glycan Analysis of Colorectal Cancer Samples Reveals Stage-Dependent Changes in CEA Glycosylation Patterns. *Clinical Proteomics*. **2018**, 15 (1). DOI: 10.1186/s12014-018-9182-4.
- (66) Pont, L.; Kuzyk, V.; Benavente, F.; Sanz-Nebot, V.; Mayboroda, O. A.; Wuhrer, M.; Lageveen-Kammeijer, G. S. M. Site-Specific N-Linked Glycosylation Analysis of Human Carcinoembryonic Antigen by Sheathless Capillary Electrophoresis-Tandem Mass Spectrometry. *Journal of Proteome Research*. **2021**, 20 (3). DOI: 10.1021/acs.jproteome.0c00875.

- (67) Kekki, H.; Peltola, M.; van Vliet, S.; Bangma, C.; van Kooyk, Y.; Pettersson, K. Improved Cancer Specificity in PSA Assay Using Aleuria Aurantia Lectin Coated Eu-Nanoparticles for Detection. *Clinical Biochemistry*. **2017**, *50* (1–2). DOI: 10.1016/j.clinbiochem.2016.06.015.
- (68) Smit, N. P. M.; Ruhaak, L. R.; Romijn, F. P. H. T. M.; Pieterse, M. M.; Van Der Burgt, Y. E. M.; Cobbaert, C. M. The Time Has Come for Quantitative Protein Mass Spectrometry Tests That Target Unmet Clinical Needs. *Journal of the American Society for Mass Spectrometry*. **2021**, *32* (3). DOI: 10.1021/jasms.0c00379.
- (69) Matthiesen, R.; Bunkenborg, J. Introduction to Mass Spectrometry-Based Proteomics. In *Methods in Molecular Biology*; **2013**; Vol. 1007. DOI: 10.1007/978-1-62703-392-3\_1.
- (70) de Hoffmann, E.; Stroobant, V. *Mass Spectrometry: Principles and Applications*, 3rd ed.; Wiley: Chichester, **2007**.
- (71) Marie, A.-L.; Ray, S.; Ivanov, A. R. Highly-Sensitive Label-Free Deep Profiling of N-Glycans Released from Biomedically-Relevant Samples. *Nature Communications*. **2023**, *14* (1), 1618. DOI: 10.1038/s41467-023-37365-4.
- (72) Marie, A. L.; Ray, S.; Lu, S.; Jones, J.; Ghiran, I.; Ivanov, A. R. High-Sensitivity Glycan Profiling of Blood-Derived Immunoglobulin G, Plasma, and Extracellular Vesicle Isolates with Capillary Zone Electrophoresis-Mass Spectrometry. *Analytical Chemistry*. **2021**, *93* (4). DOI: 10.1021/acs.analchem.0c03102.
- (73) Lageveen-Kammeijer, G. S. M.; de Haan, N.; Mohaupt, P.; Wagt, S.; Filius, M.; Nouta, J.; Falck, D.; Wuhrer, M. Highly Sensitive CE-ESI-MS Analysis of N-Glycans from Complex Biological Samples. *Nature Communications*. **2019**, *10* (1), 1–8. DOI: 10.1038/s41467-019-09910-7.
- (74) Giménez, E.; Ramos-Hernan, R.; Benavente, F.; Barbosa, J.; Sanz-Nebot, V. Analysis of Recombinant Human Erythropoietin Glycopeptides by Capillary Electrophoresis Electrospray-Time of Flight-Mass Spectrometry. *Analytica Chimica Acta*. **2012**, *709*. DOI: 10.1016/j.aca.2011.10.028.
- (75) Kammeijer, G. S. M.; Nouta, J.; De La Rosette, J. J. M. C. H.; De Reijke, T. M.; Wuhrer, M. An In-Depth Glycosylation Assay for Urinary Prostate-Specific Antigen. *Analytical Chemistry*. **2018**, *90* (7), 4414–4421. DOI: 10.1021/acs.analchem.7b04281.
- (76) Gstöttner, C.; Nicolardi, S.; Habegger, M.; Reusch, D.; Wuhrer, M.; Domínguez-Vega, E. Intact and Subunit-Specific Analysis of Bispecific Antibodies by Sheathless CE-MS. *Analytica Chimica Acta*. **2020**, *1134*. DOI: 10.1016/j.aca.2020.07.069.
- (77) Moran, A. B.; Domínguez-Vega, E.; Nouta, J.; Pongracz, T.; de Reijke, T. M.; Wuhrer, M.; Lageveen-Kammeijer, G. S. M. Profiling the Proteoforms of Urinary Prostate-Specific Antigen by Capillary Electrophoresis – Mass Spectrometry. *Journal of Proteomics*. **2021**, *238*. DOI: 10.1016/j.jprot.2021.104148.
- (78) Ashwood, C.; Pratt, B.; MacLean, B. X.; Gundry, R. L.; Packer, N. H. Standardization of PGC-LC-MS-Based Glycomics for Sample Specific Glycotyping. *Analyst*. **2019**, *144* (11), 3601–3612. DOI: 10.1039/c9an00486f.
- (79) Selman, M. H. J.; Derks, R. J. E.; Bondt, A.; Palmblad, M.; Schoenmaker, B.; Koeleman, C. A. M.; van de Geijn, F. E.; Dolhain, R. J. E. M.; Deelder, A. M.; Wuhrer, M. Fc Specific IgG Glycosylation Profiling by Robust Nano-Reverse Phase HPLC-MS Using a Sheath-Flow

- ESI Sprayer Interface. *Journal of Proteomics*. **2012**, 75 (4). DOI: 10.1016/j.jprot.2011.11.003.
- (80) Higel, F.; Demelbauer, U.; Seidl, A.; Friess, W.; Sorgel, F. Reversed-Phase Liquid-Chromatographic Mass Spectrometric N-Glycan Analysis of Biopharmaceuticals. *Anal Bioanal Chem*. **2013**, 405 (8), 2481–2493. DOI: 10.1007/s00216-012-6690-3.
- (81) Sénard, T.; Gargano, A. F. G.; Falck, D.; de Taeye, S. W.; Rispens, T.; Vidarsson, G.; Wuhrer, M.; Somsen, G. W.; Domínguez-Vega, E. MS-Based Allotype-Specific Analysis of Polyclonal IgG-Fc N-Glycosylation. *Frontiers in Immunology*. **2020**, 11, 2049. DOI: 10.3389/fimmu.2020.02049.
- (82) Bruggink, C.; Wuhrer, M.; Koeleman, C. A. M.; Barreto, V.; Liu, Y.; Pohl, C.; Ingendoh, A.; Hokke, C. H.; Deelder, A. M. Oligosaccharide Analysis by Capillary-Scale High-PH Anion-Exchange Chromatography with on-Line Ion-Trap Mass Spectrometry. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*. **2005**, 829 (1–2). DOI: 10.1016/j.jchromb.2005.10.006.
- (83) Madunić, K.; Zhang, T.; Mayboroda, O. A.; Holst, S.; Stavenhagen, K.; Jin, C.; Karlsson, N. G.; Lageveen-Kammeijer, G. S. M.; Wuhrer, M. Colorectal Cancer Cell Lines Show Striking Diversity of Their O-Glycome Reflecting the Cellular Differentiation Phenotype. *Cellular and Molecular Life Sciences*. **2021**, 78 (1), 337–350. DOI: 10.1007/s00018-020-03504-z.
- (84) Smith, J.; Millán-Martín, S.; Mittermayr, S.; Hilborne, V.; Davey, G.; Polom, K.; Roviello, F.; Bones, J. 2-Dimensional Ultra-High Performance Liquid Chromatography and DMT-MM Derivatization Paired with Tandem Mass Spectrometry for Comprehensive Serum N-Glycome Characterization. *Analytica Chimica Acta*. **2021**, 1179, 338840. DOI: 10.1016/j.aca.2021.338840.
- (85) Kuhn, R.; Hoffstetter-Kuhn, S. *Capillary Electrophoresis: Principles and Practice*; **1993**. DOI: 10.1007/978-3-642-78058-5.
- (86) Kammeijer, G. S. M. *Unravelling the Sugar-Coating of Prostate-Specific Antigen: Method Development and Its Application to Prostate Cancer Research*; Leiden University, **2019**.
- (87) Schmitt-Kopplin, P.; Fekete, A. The CE-Way of Thinking: “All Is Relative!” In *Methods in Molecular Biology*; **2016**; Vol. 1483. DOI: 10.1007/978-1-4939-6403-1\_1.
- (88) Kammeijer, G. S. M.; Jansen, B. C.; Kohler, I.; Heemskerk, A. A. M.; Mayboroda, O. A.; Hensbergen, P. J.; Schappler, J.; Wuhrer, M. Sialic Acid Linkage Differentiation of Glycopeptides Using Capillary Electrophoresis - Electrospray Ionization - Mass Spectrometry. *Scientific Reports*. **2017**, 7 (1). DOI: 10.1038/s41598-017-03838-y.
- (89) Rohrer, J. S.; Thayer, J.; Weitzhandler, M.; Avdalovic, N. Analysis of the N-Acetylneuraminic Acid and N-Glycolylneuraminic Acid Contents of Glycoproteins by High-PH Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC/PAD). *Glycobiology*. **1998**, 8 (1). DOI: 10.1093/glycob/8.1.35.
- (90) Ruhaak, L. R.; Huhn, C.; Waterreus, W. J.; de Boer, A. R.; Neususs, C.; Hokke, C. H.; Deelder, A. M.; Wuhrer, M. Hydrophilic Interaction Chromatography-Based High-Throughput Sample Preparation Method for N-Glycan Analysis from Total Human Plasma Glycoproteins. *Anal Chem*. **2008**, 80 (15), 6119–6126. DOI: 10.1021/ac800630x.
- (91) Vreeker, G. C.; Wuhrer, M. Reversed-Phase Separation Methods for Glycan Analysis. *Anal Bioanal Chem*. **2017**, 409 (2), 359–378. DOI: 10.1007/s00216-016-0073-0.

- (92) Madunić, K.; Wagt, S.; Zhang, T.; Wuhrer, M.; Lageveen-Kammeijer, G. S. M. Dopant-Enriched Nitrogen Gas for Enhanced Electrospray Ionization of Released Glycans in Negative Ion Mode. *Analytical Chemistry*. **2021**, *93* (18). DOI: 10.1021/acs.analchem.1c00023.
- (93) Ruhaak, L. R.; Deelder, A. M.; Wuhrer, M. Oligosaccharide Analysis by Graphitized Carbon Liquid Chromatography-Mass Spectrometry. *Analytical and Bioanalytical Chemistry*. **2009**, *394* (1). DOI: 10.1007/s00216-009-2664-5.
- (94) Zauner, G.; Deelder, A. M.; Wuhrer, M. Recent Advances in Hydrophilic Interaction Liquid Chromatography (HILIC) for Structural Glycomics. *Electrophoresis*. **2011**, *32* (24). DOI: 10.1002/elps.201100247.
- (95) Alpert, A. J. Hydrophilic-Interaction Chromatography for the Separation of Peptides, Nucleic Acids and Other Polar Compounds. *Journal of Chromatography A*. **1990**, *499* (C). DOI: 10.1016/S0021-9673(00)96972-3.
- (96) Camperi, J.; Pichon, V.; Delaunay, N. Separation Methods Hyphenated to Mass Spectrometry for the Characterization of the Protein Glycosylation at the Intact Level. *Journal of Pharmaceutical and Biomedical Analysis*. **2020**, *178*. DOI: 10.1016/j.jpba.2019.112921.
- (97) Lermyte, F.; Tsybin, Y. O.; O'Connor, P. B.; Loo, J. A. Top or Middle? Up or Down? Toward a Standard Lexicon for Protein Top-Down and Allied Mass Spectrometry Approaches. *Journal of the American Society for Mass Spectrometry*. **2019**, *30* (7). DOI: 10.1007/s13361-019-02201-x.
- (98) Yang, Y.; Franc, V.; Heck, A. J. R. Glycoproteomics: A Balance between High-Throughput and In-Depth Analysis. *Trends in Biotechnology*. **2017**, *35* (7), 598–609. DOI: 10.1016/j.tibtech.2017.04.010.
- (99) Lageveen-Kammeijer, G. S. M.; Kuster, B.; Reusch, D.; Wuhrer, M. High Sensitivity Glycomics in Biomedicine. *Mass Spectrometry Reviews*. **2022**, *41* (6). DOI: 10.1002/mas.21730.
- (100) Zhu, Z.; Desaire, H. Carbohydrates on Proteins: Site-Specific Glycosylation Analysis by Mass Spectrometry. *Annual Review of Analytical Chemistry*. **2015**, *8*. DOI: 10.1146/annurev-anchem-071114-040240.
- (101) Ruhaak, L. R.; Xu, G.; Li, Q.; Goonatilleke, E.; Lebrilla, C. B. Mass Spectrometry Approaches to Glycomic and Glycoproteomic Analyses. *Chemical Reviews*. **2018**, *118* (17). DOI: 10.1021/acs.chemrev.7b00732.
- (102) Varki, A. et al. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology*. **2015**, *25* (12), 1323–1324. DOI: 10.1093/glycob/cwv091.





# Chapter 2

## Profiling the Proteoforms of Urinary Prostate-Specific Antigen by Capillary Electrophoresis – Mass Spectrometry

*Alan B. Moran<sup>1</sup>, Elena Domínguez-Vega<sup>1</sup>, Jan Nouta<sup>1</sup>, Tamas Pongracz<sup>1</sup>, Theo M. de Reijke<sup>2</sup>, Manfred Wuhrer<sup>1</sup>, Guinevere S.M. Lageveen-Kammeijer<sup>1\*</sup>*

<sup>1</sup> *Leiden University Medical Center, Center for Proteomics and Metabolomics, 2300 RC Leiden, The Netherlands*

<sup>2</sup> *Amsterdam UMC, location Academic Medical Center, Department of Urology, University of Amsterdam, Meibergdreef 9 1105 AZ Amsterdam, The Netherlands*

*Reprinted (adapted) from Moran, A. B., Domínguez-Vega, E., Nouta, J., Pongracz, T., de Reijke, T. M., Wuhrer, M., & Lageveen-Kammeijer, G. S. M. (2021). Profiling the proteoforms of urinary prostate-specific antigen by capillary electrophoresis – mass spectrometry. Journal of Proteomics, 238, <https://doi.org/10.1016/j.jprot.2021.104148> under the Creative Commons CC-BY license.*





## Abstract

Early detection of prostate cancer may lead to the overdiagnosis and overtreatment of patients as well as missing significant cancers. The current diagnostic approach uses elevated serum concentrations of prostate-specific antigen (PSA) as an indicator of risk. However, this test has been widely criticized as it shows poor specificity and sensitivity. In order to improve early detection and diagnosis, several studies have investigated whether different PSA proteoforms are correlated to prostate cancer. Until now, studies and methodologies for the comprehensive characterization of PSA proteoforms from biofluids are scarce. For this purpose, we developed an intact protein assay to analyse PSA by capillary electrophoresis-electrospray ionization-mass spectrometry after affinity purification from patients' urine. Here, we determined six proteolytic cleavage variants. In regard to glycosylation, tri-, di-, mono- and non-sialylated complex-type *N*-glycans were found on non-cleaved PSA, as well as the non-glycosylated variant. The performance of the intact protein assay was assessed using a pooled sample, obtaining an inter-day variability of 15%. Furthermore, urinary patient samples were analysed by intact protein analysis and a bottom-up approach (glycopeptide analysis). This combined approach revealed complimentary information on both levels, demonstrating the benefit of using two orthogonal techniques to provide a thorough profile of urinary PSA.

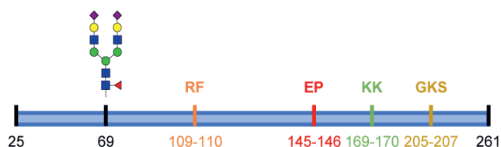
## Introduction

An elevated serum level of prostate-specific antigen (PSA) was the first FDA approved tumour marker for the early detection of prostate cancer (PCa).<sup>1,2</sup> However, since its approval, this serum-based test has come under increased scrutiny due to its lack of specificity and sensitivity.<sup>3</sup> For example, the test is unable to distinguish PCa from other prostate-related diseases, such as benign prostate hyperplasia (BPH),<sup>1</sup> as well as showing poor performance when differentiating aggressive from insignificant PCa. These issues are caused, in part, when the PSA concentration in serum is between 3 – 10 ng/mL (Netherlands)<sup>4</sup> or 4 – 10 ng/mL (United States)<sup>5</sup>, the so-called diagnostic 'grey zone'. Patients with elevated serum PSA concentrations may undergo biopsy for further confirmation, however, only 26% of cases with PSA levels in the diagnostic grey zone have PCa.<sup>5</sup> Furthermore, the biopsy procedure is accompanied by risk factors such as infection and hematuria.<sup>5</sup> In general, patients risk being over- or under-treated and, therefore, would benefit greatly from more specific and sensitive biomarkers.

Several studies have investigated whether different PSA proteoforms might improve or aid the current PSA test.<sup>6,7</sup> In serum and seminal plasma, several naturally occurring PSA proteoforms are cleaved at specific internal cleavage sites (**Figure 1**), resulting in enzymatically inactive variants.<sup>8–10</sup> A cleaved PSA form, containing a cleavage at Lys<sub>169</sub> and Lys<sub>206</sub>, is described as benign PSA (bPSA) due to its association with BPH.<sup>7</sup> Importantly, a feature of BPH is the overgrowth of cells located in the transition zone of the prostate,<sup>11</sup> precisely the location where this cleavage is hypothesised to occur.<sup>7</sup> Conversely, in PCa, PSA circumvents this process when the epithelial lining becomes disrupted, resulting in the protein entering the bloodstream.<sup>12</sup>

Investigations into PSA proteolytic cleavage variants have mainly been described in serum and seminal plasma; the proportion of cleaved PSA may account for 10 to 30% of total PSA in serum<sup>13–15</sup> whereas these proteoforms comprise 30% of PSA in seminal plasma.<sup>13,16</sup> However, another promising biofluid, urine, is relatively unexplored in this regard.<sup>17,18</sup> This is unfortunate since urine is likely to reflect the biological status of the prostate as there is evidence that PSA is directly secreted into urine via the periurethral glands.<sup>19,20</sup> Moreover, urine is relatively easy to obtain in high volumes with high

concentrations of PSA.<sup>21,22</sup> In this case, an investigation into urinary PSA is warranted in order to define the specific proteoforms that are present there.



**Figure 1: N-glycosylation site and internal cleavage sites present in PSA.** Amino acids are represented along the top whereas the sequence positions are represented along the bottom. Most abundant N-glycan (H5N4F1S2) is depicted on the N-glycosylation site (Asn<sub>69</sub>). Figure 1 adapted with permission<sup>13</sup> © (2003) American Society of Clinical Oncology. All rights reserved.

PSA glycosylation has been extensively studied as it has been reported to have potential diagnostic and prognostic value.<sup>18,23–25</sup> This protein contains a single N-glycosylation site present at Asn<sub>69</sub>, normally containing complex-type disialylated and core-fucosylated glycans. A second N-glycosylation site might be present when Asp<sub>102</sub> is replaced by Asn<sub>102</sub> due to a mutation in the gene encoding PSA, kallikrein 3

(KLK3),<sup>26</sup> yet this has only been observed in a single study.<sup>27</sup> In general, PSA glycosylation in urine, seminal plasma and serum has been well examined.<sup>9,27–29</sup> Sarrats *et al.* demonstrated a similar glycosylation pattern between cleaved and non-cleaved PSA variants in seminal plasma and serum, respectively.<sup>29</sup> In contrast, small discrepancies were found when non-cleaved PSA glycopeptides were compared to the total PSA glycosylation profile.<sup>28</sup> Undoubtedly, the relative contribution of cleaved PSA forms to its total glycosylation is not well understood and requires further investigation.

Charge-based separation of PSA has previously been employed on the basis of the different isoelectric points (pIs) of its proteoforms.<sup>17,29</sup> For example, five distinct proteoforms were reliably separated using two-dimensional electrophoresis (2-DE).<sup>29</sup> Here, however, only an estimated protein mass was determined using a molecular weight marker. Investigations that have applied mass spectrometry (MS) have provided a more accurate determination of the mass of PSA proteoforms.<sup>9</sup> In this case, separation was achieved using anion-exchange chromatography, although the proteoforms were analyzed offline by MS. Interestingly, capillary electrophoresis (CE) has been shown to be a suitable technique when hyphenated with electrospray ionization-MS (CE-ESI-MS) for online separation and identification of PSA proteoforms.<sup>30</sup>

The main aim of this study was to investigate the profile of PSA proteoforms present in urine. For this purpose, PSA was captured from a patient urine pool and analysed by CE-ESI-MS. In order to demonstrate the applicability of the protocol, the method was validated and further assessed using individual patient urine samples. Finally, an in-depth glycopeptide analysis was carried out to complement the glycosylation profile obtained at the intact protein level.

## Materials and Methods

### Chemicals and Reagents

Deionized water (MQ) was obtained with a Q-Gard 2 system (Millipore, Amsterdam, The Netherlands). Ammonium bicarbonate, potassium dihydrogen phosphate ( $\text{KH}_2\text{PO}_4$ ), sodium bicarbonate ( $\text{NaHCO}_3$ ), sodium chloride ( $\text{NaCl}$ ), sodium hydroxide ( $\text{NaOH}$ ), and sodium phosphate dibasic dihydrate ( $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$ ) were purchased from Merck (Darmstadt, Germany). Ammonium acetate, formic acid (FA), and water (LC-MS grade) were acquired from Fluka (Steinheim, Germany). Acetonitrile (LC-MS grade) was purchased from Biosolve (Valkenswaard, The Netherlands). Phosphate-buffered saline (PBS), five times concentrated (5x), was prepared from 0.16 M  $\text{Na}_2\text{HPO}_4$ , 0.02 M  $\text{KH}_2\text{PO}_4$ , 0.73 M  $\text{NaCl}$  at pH 7.2. 5x PBS was diluted with MQ to obtain 1x PBS, pH 7.6. DL-dithiothreitol (DTT), glacial acetic acid (HAc), hydrochloric acid (HCl), and iodoacetamide (IAA) were obtained from Sigma-Aldrich (Steinheim, Germany). Polyethylenimine (PEI) was obtained from Gelest (Morrisville, NC). Seminal PSA standard was acquired from Lee BioSolutions (St. Louis, MO). The protein test mixture, containing cytochrome C, lysozyme and ribonuclease A, was purchased from SCIEX (Brea, CA). Mass spectrometry-grade trypsin derived from bovine pancreas was purchased from Promega (Madison, WI). A synthetic peptide (LSEPAELTEAVK) was prepared in-house by FMoc solid phase peptide synthesis.

### Sample Collection

A waiver was obtained from the medical ethical committee (W16\_010#16.020) of the Amsterdam University Medical Center (AMC) and clinical samples were collected over a two year period. Patients who were suspected of PCa and presented with an elevated PSA serum concentration ( $> 3 \text{ ng/mL}$ ) donated urine prior to digital rectal examination (DRE) and prostate biopsy. Urinary samples were homogenised and

portioned into BD Vacutainer™ Plastic Urinalysis Tubes (Fisher Scientific, Loughborough, UK). The samples were stored at -80°C for a maximum of three months at the AMC before they were transferred to Leiden University Medical Center and stored at -80°C. Sample processing was carried out at the end of the collection period.

### **Experimental Design**

#### *Repeatability and Intermediate Precision Assessment*

In order to assess the repeatability and intermediate precision of the assay, an intra- and inter-day analysis was performed using a patient urine pool consisting of 27 randomly selected patients (> 180 mL), hereby referred to as the validation pool. This was divided into 20 mL portions and stored at -20°C. Three full technical replicates were processed and measured each day ( $n = 3$ ) over three days ( $n = 9$ ). Urine was prepared as detailed in the following sections.

#### *Method Application*

The applicability of the protocol on individual patients was assessed using a separate cohort of patient samples ( $n = 10$ ). Patient samples were randomized prior to sample treatment and 20 mL from each patient was processed over two batches. In addition, a positive control was created by pooling 5 mL from each of the 10 urine samples (50 mL) that was processed and measured in each batch. Urine was prepared according to the following sections.

### **Sample Processing**

Urine was processed as previously described by Kammeijer *et al.*<sup>28</sup> Briefly, samples were removed from storage, thawed at room temperature (RT) and vortexed to ensure a homogeneous mixture. Following this, the samples were centrifuged (500g, 5 min) and 20 mL of urine supernatant was removed. The protocol has been validated to process 20 mL of urine and, therefore, in cases whereby urine was available below 20 mL, these samples were added up with 1x PBS to reach a total volume of 20 mL. Following this, 5 mL 5x PBS was added to all samples (final volume of 25 mL).

### **PSA Capturing**

PSA was isolated and immunopurified from urine as previously described.<sup>28</sup> Briefly, anti-PSA nanobodies (QVQ, Utrecht, The Netherlands) were coupled in-house with

Sepharose beads (GE healthcare, Little Chalfont, UK). An overnight incubation at 4°C was performed with the addition of 60 µL of a 50% anti-PSA beads solution to the samples. Next, the samples were centrifuged (100g, 1 min), and the supernatant was removed to leave 500 µL of anti-PSA beads solution. This was followed by re-suspension and transfer to a 96-well polypropylene filter plate (Orochem, Naperville, IL), which was placed inside a vacuum manifold (Merck Millipore, Darmstadt, Germany). A vacuum was applied until there was no liquid remaining. Subsequently, the beads were washed with 600 µL 1x PBS followed by two washes using 600 µL of 50 mM ammonium bicarbonate. Following this, PSA was eluted using 200 µL of 100 mM FA containing a synthetic peptide (LSEPAELTEAVK; 45 fmol/µL). The plate was covered using an adhesive seal and placed on a plate shaker (max. rpm, 5 min). Finally, the plate was centrifuged (100g, 2 min) and the eluent was split into two portions for intact protein analysis (80%) or enzymatic digestion with trypsin (20%) for glycopeptide analysis. Each portion was concentrated to dryness using a vacuum centrifuge operating at 45°C (Eppendorf Concentrator 5301, Eppendorf). Samples were reconstituted in 3 µL of MQ (intact protein analysis) or 5 µL of 25 mM NaHCO<sub>3</sub> (glycopeptide analysis), and stored at -20°C until further processing.

### **In-solution Tryptic Digestion**

The portions set for glycopeptide analysis underwent reduction and alkylation prior to digestion with trypsin, as previously described.<sup>28</sup> Shortly, 1 µL of 12 mM DTT was added prior to incubation at 60°C for 30 min. Further incubation was performed with 1 µL of 42 mM IAA for 30 min in darkness at RT. Finally, 1 µL of 48 mM DTT was added to the samples which remained in brightness for 20 min. Following this, 1 µL of 0.15 mg/mL trypsin was added in 25 mM NaHCO<sub>3</sub> and overnight digestion was performed at 37°C.

### **Capillary Electrophoresis**

All experiments were performed on a CESI 8000 instrument (SCIEX) using bare-fused silica capillaries (91 cm x 30 µm i.d. x 150 µm o.d.) containing a porous tip. For intact protein analysis, capillaries were coated in-house using PEI and in accordance with a previously published protocol.<sup>31</sup> The temperature was set to 5°C and 15°C for the sample tray and capillary, respectively. The performance of the capillary was determined prior to performing measurements by assessing the migration time and

signal intensity of a protein test mixture (SCIEX), as described previously.<sup>31</sup> When deemed necessary, the capillary coating was removed using 1 M NaOH (100 psi for 60 min) and re-coated according to the same procedure as mentioned above. In this case, however, no re-coating was required. Prior to analysis, the separation capillary was rinsed (100 psi for 5 min) with the background electrolyte (BGE) consisting of 20% HAc (v/v, 3.49 M, pH 2.3). Following this, the conductive line was filled with BGE (100 psi for 4 min) and hydrodynamic injection of the sample (8.8 psi for 35 sec) was applied. Assuming a viscosity of aqueous BGE of 1.14 cP, 6.3% of the total capillary volume was filled which equals an injection of 41 nL. The instrument was operated in reverse polarity mode and a separation voltage of 20 kV was applied for 45 min. The capillary was prevented from drying out between measurements by maintaining a constant low flow with the continuous application of 10 psi to the BGE vial.

Electrophoretic separation of glycopeptides was carried out in line with the protocol from literature.<sup>28</sup> Briefly, a bare-fused silica (BFS) capillary was flushed with 0.1 M NaOH (2.5 min), LC-MS grade H<sub>2</sub>O (3 min), 0.1 M HCl (2.5 min), H<sub>2</sub>O (3 min), followed by 3 min with the BGE of 20% HAc (v/v, 3.49 M, pH 2.3). Here, the sampling tray was set to 10°C and the capillary temperature was maintained at 15°C. The sample (4 µL) was mixed with 2 µL of leading electrolyte (1.2 M ammonium acetate, pH 3.17) in a PCR tube and subsequently transferred to nanovials (SCIEX, Brea, CA). Hydrodynamic injection of the sample was performed at 25 psi for 24 sec (12.3% capillary volume / 79 nL). Following this, a BGE post plug was injected (0.5 psi for 25 sec).

### **Mass Spectrometry**

On-line coupling of the CE system was achieved with a maXis Impact UHR-QqTOF (Ultra-High Resolution Qq-Time-Of-Flight) MS (Bruker Daltonics GmbH, Bremen, Germany) via a sheathless CE-ESI-MS interface and a nano-electrospray source. All MS experiments were performed in positive ionization mode and specific parameters were used for intact protein and glycopeptide analysis. For intact protein analysis, the following settings were used: electrospray voltage, 1350 V; nitrogen drying gas, 1.2 L/min at 100°C; quadrupole ion energy, 5 eV; collision cell energy, 5 eV; transfer time, 120 µs; pre-pulse storage time, 20 µs. For glycopeptide analysis, the following parameters were implemented: electrospray voltage, 1100 V; nitrogen drying gas, 1.2

L/min at 150°C; quadrupole ion energy, 3 eV; collision cell energy, 7 eV; transfer time, 130  $\mu$ s; pre-pulse storage time, 15  $\mu$ s. Furthermore, dopant enriched (ca. 4% mole percent acetonitrile) nitrogen (DEN)-gas was employed during glycopeptide analysis as previously described.<sup>32</sup> Briefly, the DEN-gas was introduced by allowing dry gas into a nano-Booster (OT3118G002, Bruker Daltonics); the coaxial sheath flow of the DEN-gas around the ESI emitter was introduced via an in-house made polymer cone, surrounding the housing of the porous tip. The MS data was acquired with a spectral acquisition frequency of 1 Hz within the range  $m/z$  600 – 3000 or  $m/z$  200 – 2200 for intact protein and glycopeptide analysis, respectively.

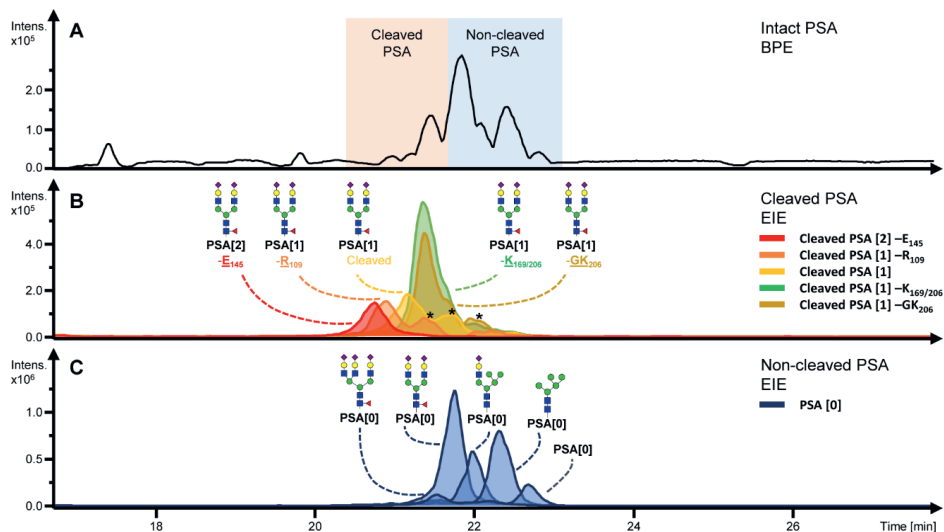
### Data Analysis

Bruker DataAnalysis version 5.0 was used to perform data analysis. For assignments of intact proteoforms, internal calibration was carried out using the top 10 charge states ( $[M+13H]^{13+}$  –  $[M+22H]^{22+}$ ) of the most abundant peak in combination with  $[M+1H]^{1+}$  and  $[M+2H]^{2+}$  of a synthetic peptide (LSEPAELTEAVK) as calibration points. Then, the spectra of interest were averaged, smoothed (Gaussian, 0.05 Da) and deconvoluted. Mass spectra deconvolution was performed using the maximum entropy algorithm. Data point spacing and instrument resolving power were set to  $m/z$  0.5 and  $10^4$  FWHM, respectively. Assignments of the peaks were made by employing a tolerance of  $\leq 25$  ppm. Theoretical  $m/z$  values were generated for assigned peaks using an in-house developed software (LaCyTools, version 1.1).<sup>33</sup> Extracted ion electropherograms (EIEs) were generated using the three most abundant charge states of assigned peaks with an extraction window of  $\pm m/z$  0.1. Electropherogram smoothing was applied to all EIEs (Gaussian, 2.0 points). Following this, peak areas were obtained via manual integration.

For glycopeptides, EIEs were created using the first three isotopes across three charge states ( $[M+H]^+$  –  $[M+3H]^{3+}$ ) with an extraction window of  $\pm m/z$  0.05. Glycopeptide structures, including differently linked sialic acid isomers, were assigned based on a mass tolerance of  $\leq 10$  ppm and previously published research.<sup>28</sup> Glycoforms were labeled as glycan net compositions that specifies number of hexoses (H), *N*-acetylglucosamines (N), fucoses (F) and *N*-acetyl neuraminic acids (S); these monosaccharide abbreviations will be further used throughout this manuscript. EIE



areas were obtained via manual integration and adjusted to represent the entire isotopic envelope.



**Figure 2: Intact protein analysis using CE-ESI-MS of PSA captured from a patient urinary pool.** (A) Base peak electropherogram (BPE) of intact urinary PSA. A variation in PSA proteoforms can be observed with and without internal cleavages. (B) The EIEs illustrate the most abundant glycoprotein present per proteoform. Proteoforms with different numbers of internal cleavages can be observed with the number of internal cleavages indicated in the square brackets: PSA[2] -E<sub>145</sub>, double cleaved PSA at E<sub>145</sub> and one other unidentified position; PSA[1] -R<sub>109</sub>, cleaved PSA at R<sub>109</sub>; PSA[1] -Cleaved, cleaved PSA at a unidentified position; PSA[1] -K<sub>169/206</sub>, cleaved PSA at either K<sub>169</sub> or K<sub>206</sub>; PSA[1] -GK<sub>206</sub>, cleaved PSA at GK<sub>206</sub> (C) For non-cleaved PSA the most abundant tri-, di- and mono-sialated form is illustrated as well as the most abundant high mannose type and the non-glycosylated form of the protein. Asterisk (\*) indicates overlapping *m/z* values in different electrophoretic peaks. Blue square: *N*-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: fucose, pink diamond: *N*-acetylneuraminic acid.

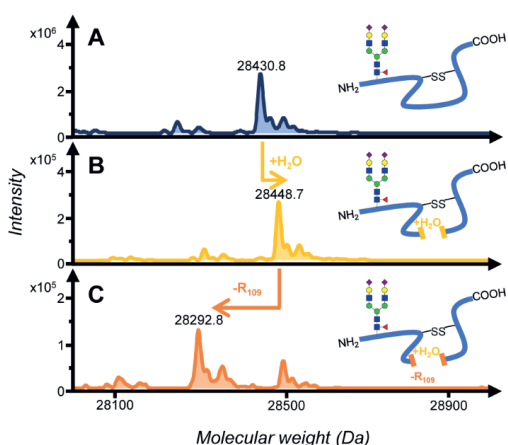
## Results

### Intact Urinary PSA

#### *PSA Proteoform Analysis by CE-ESI-MS*

Separation for intact protein analysis with CE-ESI-MS was optimised using a seminal PSA standard, as summarised in **Supporting Information, Table S-1**. Briefly, the capillary was positively coated using PEI, a polymer which has been reported to

reduce protein-capillary inner surface interactions.<sup>31</sup> The capillary was operated at 20 kV in reverse polarity mode in order for the electroosmotic flow to migrate towards the capillary outlet and the MS inlet. Separation of proteoforms was achieved by using 20% HAc as the BGE and the separation between the sialylated and non-sialylated glycoforms was increased by lowering the capillary temperature to 15°C. Capturing and purification was performed as previously described with no amendments to the original procedure<sup>28</sup> and PSA from urine was directly analysed by CE-ESI-MS.



**Figure 3: Assignment of proteolytic cleavage variants based on amino acid loss at the cleavage site. (A)** Deconvoluted mass spectrum of non-cleaved PSA accompanied by a graphical representation (right) of the protein and the most abundant *N*-glycan. **(B)** Hydrolysis of the peptide bond at an unidentified cleavage site increases the mass (18 Da) of the protein. The protein is inactivated due to this internal cleavage, however the structure remains intact as a result of disulfide bridges at the cysteine residues. **(C)** The internal cleavage site may be determined, in accordance with literature,<sup>9</sup> whereby hydrolysis of the peptide bond and loss of an amino acid at the cleavage site occurs. Blue square: *N*-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: fucose, pink diamond: *N*-acetylneuraminic acid.

The analysis of intact PSA from the validation pool revealed a number of different proteoforms; six proteolytic cleavage variants, which included cleaved and non-cleaved forms, as well as a variation in the composition of glycan structures attached to the protein (**Figure 2**). The tentative assignment of cleavage sites is demonstrated in **Figure 3**. A single cleavage, due to hydrolysis of the peptide bond between two amino acids at a cleavage site, is indicated by an 18 Da increase of the intact protein mass caused by the incorporation of a water molecule. Despite this cleavage, disulfide bridges keep the protein chains covalently connected. In the instance of a protein with a double cleavage, two peptide bonds undergo hydrolysis at two cleavage sites, respectively. This is indicated by a 36 Da increment of the protein mass due to the addition of a water molecule at

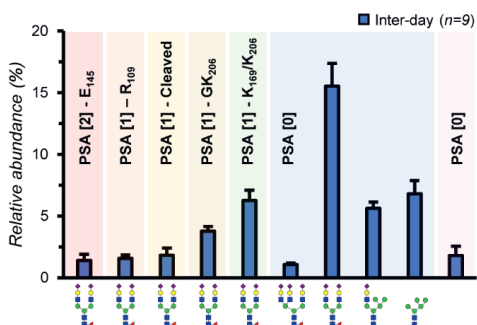
each cleavage site. Importantly, specific cleavage sites could be indicated when

apparent losses of adjacent amino acids were observed at the well-known cleavage sites (Arg<sub>109</sub>Phe, Glu<sub>145</sub>Pro, Lys<sub>169</sub>Lys, Gly<sub>205</sub>Lys<sub>206</sub>Ser).<sup>9</sup>

As illustrated in **Figure 2**, the double cleaved variant migrated first followed by the single cleaved and eventually the non-cleaved variants. Here, the most abundant glycoprotein present per proteoform is shown. The main form of PSA with two cleavages was assigned to a cleavage at position Glu<sub>145</sub>, due to the loss of Glu, whereas the second cleavage site remained unidentified, as only a second hydrolysis was indicated by an 18 Da increase of the intact protein mass and no amino acid loss was observed. The single cleavage sites were determined at Arg<sub>109</sub> and Lys<sub>206</sub> due to a single hydrolysis and loss of the adjacent amino acids there. Another single cleavage at Lys<sub>169</sub>/Lys<sub>206</sub> could not be further specified as loss of a lysine residue is possible at both sites. One variant migrating at 21.2 mins (**Figure 2B**) could not be assigned to a specific site as only a single hydrolysis was observed and no loss of an amino acid was detected. For non-cleaved PSA, the most abundant glycoforms on this proteoform are shown in **Figure 2C**. Here, the glycoforms with the highest number of sialic acids migrate first and the non-sialylated glycoforms migrated latest, followed by the non-glycosylated protein. Interestingly, the di-sialylated *N*-glycan structure, H5N4F1S2, was the dominant glycoform for all cleaved and non-cleaved PSA variants (**Figure 2** and **Supporting Information, Table S-2**). Overall, intact protein analysis revealed that PSA consists of a complex array of proteoforms, including cleaved and non-cleaved forms, as well as a high variation in glycosylation.

### *Repeatability and Intermediate Precision*

In order to validate and assess the applicability of the complete intact protein assay for the analysis of biological samples, the repeatability and intermediate precision was established via intra- ( $n = 3$ ) and inter-day ( $n = 3$ ) measurements (total  $n = 9$ ). The data processing protocol was also optimised by testing a number of data extraction parameters as summarised in **Supporting Information, Table S-1**. Similar results were obtained for the quantification via deconvolution or EIEs. Thus, it was decided to perform the quantification using EIEs, using the three most abundant charge states, as the Bruker Data Analysis software is more suited for this purpose. In addition, an extraction window of  $\pm m/z$  0.1 was used because it allowed the least overlap with other masses while maintaining the highest sensitivity. Various methods of internal



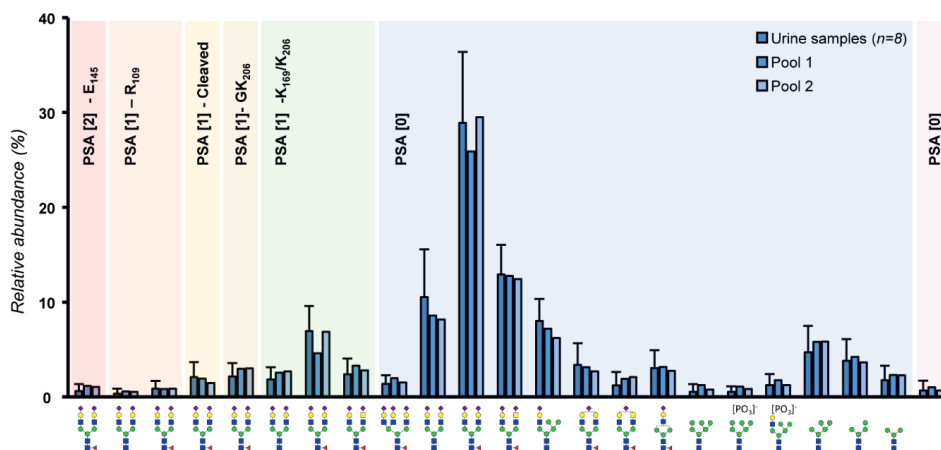
**Figure 4: Intermediate precision and repeatability of intact PSA assay measured by CE-ESI-MS.** Inter-day relative areas (%) are shown for the most abundant glycoprotein present per proteoform (normalized to the sum area of all assigned glycoprotein forms). Each bar represents the mean and the standard deviation ( $n = 9$ ) is represented by the error bars. Background colors denote different internal cleavage sites: the square brackets denote the number of cleavages present in that proteoform. The site of internal cleavage, in accordance with the literature<sup>9</sup> and observed amino acid loss, is illustrated using the position within the amino acid sequence. In green, internal cleavage may occur at either K<sub>169</sub> or K<sub>206</sub>. Blue square: N-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: fucose, pink diamond: N-acetylneuraminic acid.

calibration were tested, including internal calibration based only on the charge envelope of H5N4F1S2, or internal calibration using a combination of H5N4F1S2 and a synthetic peptide (LSEPAELTEAVK). The former method only covers  $m/z$  1200 – 2200, whereas the latter method covers a larger range of  $m/z$  ratios ( $m/z$  650 – 2200) and was eventually selected for further processing. In total, 32 proteoforms ( $\leq 25$  ppm) were identified in the validation pool, including six proteolytic cleavage variants and 21 N-glycan structures (Figure 4 and Supporting Information, Table S-2). In the end, an average *coefficient of variation* (CV) of 11% and 15%, including all 32 proteoforms, was found for the intra-day and inter-day analysis, respectively. The CVs shown here

represent the full assay performance, demonstrating the entire variation introduced from the capture procedure, the CE-ESI-MS measurement up to the manual data processing step.

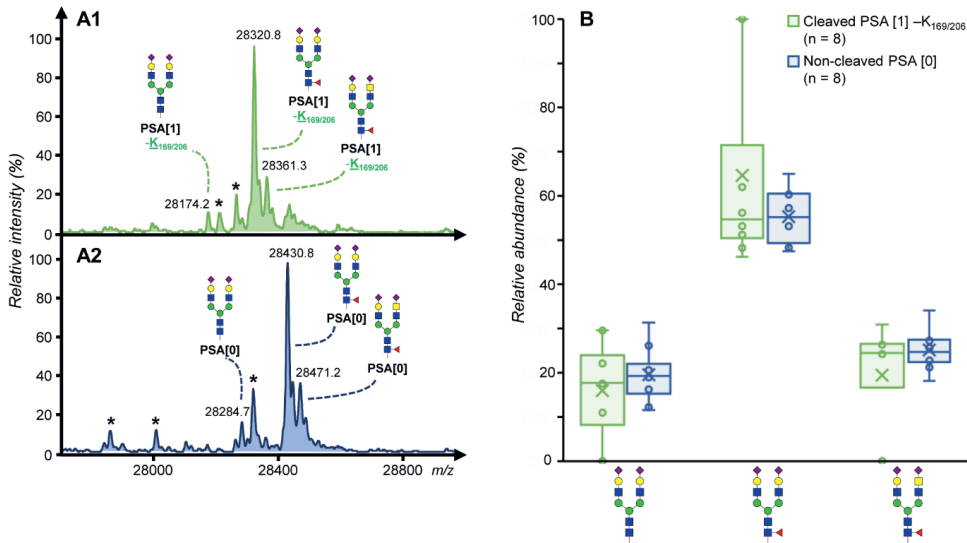
#### Assessment of PSA Proteoforms from Patient Urine

The presence and abundance of PSA proteoforms in biological samples (urine) was assessed using a small cohort of patients' urine samples ( $n = 10$ ). Eight patients provided a detectable signal for the intact protein analysis of urinary PSA; diagnoses for these patients was provided after completion of the measurement and data processing, resulting in a cohort of five non-PCa and three PCa patients. During the intra- and inter-day assessment, 32 proteoforms were identified. From this, 23 out of the 32 could be observed in the positive control with ppm errors  $\leq 25$  (Figure 5 and



**Figure 5: Intact protein assay for the analysis of patient urinary PSA with CE-ESI-MS.** All 23 identified proteoforms (ppm error  $\leq 25$ ) are represented as a relative (%) area (normalized to the sum of all 23 glycoprotein forms). A single bar represents the mean of patients' urine samples measurements ( $n = 8$ ), whereas both pools are shown, representing the positive control of each batch. The error bars represent the biological variation amongst the urine samples. Colors denote the different proteolytic cleavage variants of PSA; square brackets denote the number of cleavages present in that proteoform. The site of internal cleavage, in accordance with literature<sup>9</sup> and observed amino acid loss, is illustrated using the position within the amino acid sequence. Blue square: *N*-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: fucose, pink diamond: *N*-acetylneuraminic acid.

**Supporting Information, Figure S-1 and Table S-3).** In all samples the relative abundance of di-sialylated glycoforms was assessed on cleaved and non-cleaved PSA. No major differences were detected in the relative abundances of di-sialylated glycoforms present on the most abundant form of cleaved PSA, Lys<sub>169</sub>/Lys<sub>206</sub>, in comparison with non-cleaved PSA (**Figure 6**). Additionally, one patient (patient E) showed evidence of having an extra *N*-glycosylation site, as a result of the conversion of Asp<sub>102</sub> to Asn<sub>102</sub>, which was populated mainly with high-mannose type glycans (**Supporting Information, Table S-3**). Similar relative abundances were found for the positive control ( $n = 2$ ; **Figure 5**), which suggests that little bias was introduced into the analysis as a result of batch processing. Furthermore, the performance of the assay, as indicated by the positive control, is consistent with the performance previously validated by the intra- and inter-day study. Finally, an average CV of 48% ( $n = 8$ ) was found for the 10 most abundant proteoforms, which should represent the broad biological variation present within the entire cohort.

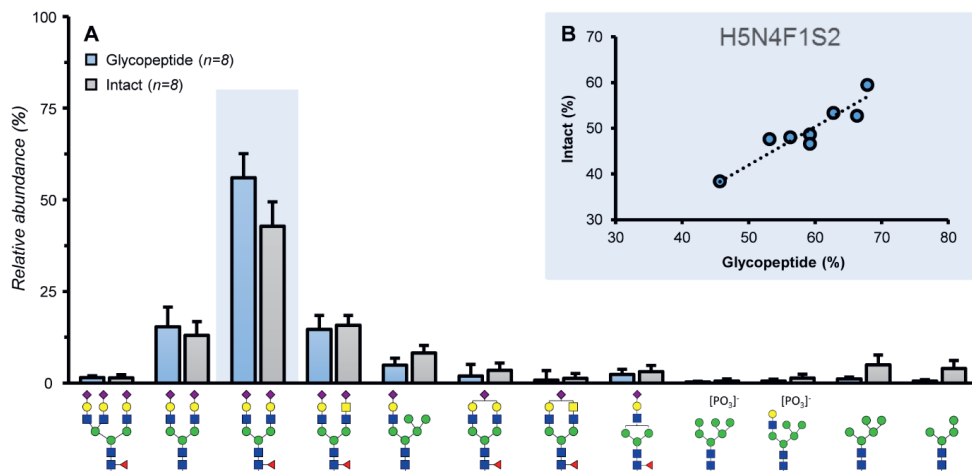


**Figure 6: Relative abundances of di-sialylated glycans on cleaved PSA -K<sub>169/206</sub> and non-cleaved PSA.** (A) Representative deconvoluted mass spectra of intact urinary PSA of (A1) cleaved PSA -K<sub>169/206</sub> that contains the same di-sialylated complex glycans as (A2) non-cleaved PSA. Asterisk (\*) indicates overlapping proteoforms in different electrophoretic peaks. (B) Relative abundance of H5N4S2, H5N4F1S2, and H4N5F1S2 (normalized to all 3 di-sialylated structures) in cleaved PSA -K<sub>169/206</sub> and non-cleaved PSA in the urine samples ( $n = 8$ ). Blue square: *N*-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: fucose, pink diamond: *N*-acetylneuraminic acid.

### Intact Protein and Glycopeptide Comparison

In order to support and expand the glycoform assignment performed at the intact protein level, the patient samples were additionally analysed by glycopeptide analysis. Importantly, the entire sample preparation was performed within the same protocol. Glycopeptide analysis detected glycoforms in all urine samples ( $n = 10$ ), whereas the intact protein approach only detected proteolytic cleavage variants and glycoforms in some ( $n = 8$ ). In total, 12 glycoforms could be identified by both intact protein and glycopeptide analysis (Figure 7A and Supporting Information, Table S-4). To test the correlation between both approaches, the variation in the most abundant glycoform, H5N4F1S2, was examined (normalized to the sum of the six most abundant glycoforms; Figure 7B). An association of  $R^2 = 0.86$  was found. Furthermore, glycopeptide analysis identified an additional 63 structures, whilst intact protein

analysis detected two structures (H6N2 and H3N2) that were not detected during glycopeptide analysis (**Supporting Information, Table S-3**). The identities of the two most abundant glycans present at the extra *N*-glycosylation site in patient E (H5N2 and H6N2) were confirmed by glycopeptide analysis. In total, the combined PSA analysis was able to identify 77 *N*-glycans, as well as two extra *N*-glycans on Asn<sub>102</sub>, and six proteolytic cleavage variants.



**Figure 7: Comparison of the intact protein and glycopeptide approaches for glycosylation quantification of urinary PSA from patients. (A)** All glycoforms detected across different cleaved and non-cleaved PSA forms were determined as a single glycoform for the purpose of comparing glycosylation profiles between intact and glycopeptide. In both cases, only glycoforms detected in both the intact protein and glycopeptide datasets (12) were used to calculate the relative abundances (normalized to all 12 glycoforms). Each bar represents the mean of patients' urine samples measurements ( $n = 8$ ) in both datasets, while error bars represent the standard deviation. **(B)** Relative abundance of H5N4F1S2 (normalized to top six most abundant glycoforms: H5N4S2, H5N4F1S2, H4N5F1S2, H6N3S1, H5N4F1S1, H4N3F1S1) detected in the glycopeptide patient dataset (x-axis) against the same glycoform detected during intact protein analysis of patients (y-axis). Equation of the regression line:  $y = 0.8387x$ ,  $R^2 = 0.8648$ . Blue square: *N*-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: fucose, pink diamond: *N*-acetylneuraminic acid.

## Discussion

Many proteoforms of the glycoprotein PSA are related to PCa or other prostate-related diseases, including cleaved and precursor forms, as well as aberrant glycosylation.<sup>6,7,24</sup> PSA proteoforms have been described in serum and seminal plasma, yet urinary PSA remains poorly understood in this sense. Thus, in order to assess the broad range of PSA proteoforms present in urine, we established and validated an intact protein assay for the analysis of urinary PSA.

### Proteoform Analysis by CE-ESI-MS

As illustrated by **Figure 2**, PSA proteoforms display a predictable migration pattern caused most likely by their different hydrodynamic volume and pI.<sup>16</sup> In previous studies, the pI of PSA proteoforms, such as non-glycosylated PSA (7.2) and di-sialylated PSA (6.8), was examined using two-dimensional electrophoresis.<sup>29</sup> However, *N*-glycan analysis revealed the same glycosylation pattern for di-sialylated PSA and proteoforms with lower pIs. Therefore, the difference in pIs cannot be attributed to sialylation and is likely due to different proteolytic cleavage variants.<sup>29</sup> Similarly, we observed the separation of these forms mainly due to proteolytic cleavages and the resulting loss of polar amino acids (Glu, Arg and Lys).<sup>17,34</sup> The cleavages at Arg<sub>109</sub>, Glu<sub>145</sub>, Lys<sub>169</sub>/Lys<sub>206</sub>, and Gly<sub>205</sub>Lys<sub>206</sub>, as determined by the loss of these amino acids and their corresponding mass, are in line with cleavage sites detected on PSA in seminal plasma and serum.<sup>9,35</sup> Cleavage at Lys<sub>206</sub> is associated with the formation of bPSA, although Mikolajczyk *et al.* originally described bPSA as a double cleavage variant with cleavages at Lys<sub>169</sub> and Lys<sub>206</sub>.<sup>7</sup> However, we did not detect this particular double cleavage variant during this study.

Despite this, bPSA has generally been characterised by SDS-PAGE following isolation using chromatographic based methods, which were unable to separate distinct cleaved PSA forms.<sup>7,36</sup> In addition, bPSA has been described as several cleaved PSA forms in the literature.<sup>13,37</sup> Thus, it is apparent that cleaved PSA requires further characterization as, for example, it is unclear whether bPSA consistently features cleavages at Lys<sub>169</sub> and Lys<sub>206</sub>, or whether other PSA cleavages are likewise associated with the development of BPH. While our study already provides more confidence for the assignment of the PSA proteolytic cleaved variants, compared to assignments based upon SDS-PAGE, further studies are needed to obtain an even



more detailed structural characterization (e.g. MS/MS) of the wide range of proteolytic cleaved variants.

PSA is secreted by the prostate and internal cleavage is thought to occur post-translationally by proteases in hyperplastic prostate tissue.<sup>38</sup> In contrast, glycosylation occurs intracellularly where it may be disrupted due to cancer-associated changes to the cell glycosylation machinery.<sup>39</sup> Although it is unclear whether protein glycosylation undergoes further matrix-associated modifications, it is unlikely that these processes are specifically related to the internal cleavage of PSA. This view is supported by the present study as the same di-sialylated species were detected both on cleaved and non-cleaved PSA (**Figure 4**). Furthermore, a similar relative abundance of these structures is detected in urine across various patients (**Figure 6** and **Supporting Information, Figure S-2**). As a result, it seems that the cleavage of PSA is independent of the glycosylation. Similarly, di-sialylated complex-type *N*-glycans have consistently been described as the main glycoform across cleaved PSA variants in seminal plasma and serum.<sup>9,29,40</sup> Presumably, glycan species that are less abundant are also present on cleaved PSA, yet have not been detected thus far and such an in-depth investigation is warranted in order to fully explore the glycan diversity of cleaved PSA.

### **Intact Data Processing and Quantification**

This method has demonstrated a similar performance to analyses that are more commonly performed with biopharmaceuticals. The intra- (11%) and inter-day (15%) CVs reported here are comparable with those described for the intact protein analysis of recombinant gonadotropin ( $\leq 10\%$ )<sup>41</sup> and antibody-derived therapeutics (8-11%).<sup>42</sup> Naturally, the analysis of standardised products represents only the measurement variation whereas analysis of clinical samples also includes error prone steps such as protein capturing and purification. Furthermore, the relative quantification of intact PSA proteoforms applied here is useful in this setting as the relative increase or decrease of specific proteoforms is likely to represent biological changes related to prostate disease.

The platform, however, is limited by data processing, which represents an analytical bottleneck for the investigation of clinical cohorts. Multiple protocols are presented in the literature with little consensus.<sup>43</sup> In this study, several data integration techniques

were tested (**Supporting Information, Table S-1**) and our findings revealed that integrating the EIE area, generated by the average mass of the three most abundant charge states per proteoform, provided reliable results in a reasonably time-efficient manner. Similar results and parameters have been reported previously, whereby quantification of intact proteins was performed using EIEs.<sup>44</sup> However, overlap of  $m/z$  signals causes broad and sometimes numerous EIE peaks. As a result, manual peak integration is required which greatly increases the processing time per sample. Additionally, the issue is further compounded by the lack of dedicated software.<sup>43</sup> Therefore, data processing presents the major bottleneck for the application of this method to larger clinical cohorts. In the case of PSA, cleaved and non-cleaved variants may be reliably separated and it seems that the expression and level of glycosylation is similar between these forms. Thus, future investigations could focus solely on cleaved PSA by performing *N*-glycan release prior to performing the intact protein measurement. This would reduce the complexity of the analysis whilst maintaining important information regarding the proteolytic cleavage variants.

### **Glycosylation Quantification**

The glycopeptide approach revealed 40 unique *N*-glycan compositions, or 75 structures in total including sialic acid isomers. Linkage determination of sialic acids is a crucial aspect as  $\alpha$ 2,3-sialylated PSA has been shown to differentiate PCa from non-PCa patients.<sup>24</sup> However, the intact protein approach was unable to determine differently linked sialylated PSA, and only a limited set of 12 glycoforms was covered by both methods, nonetheless, similar relative abundances were observed (**Figure 7**). Accordingly, relative quantification of H5N4F1S2 in the patient urine samples showed good correlation for the two approaches. However, some disparity was found in the relative quantification of the glycoforms H4N2 and H5N2. Moreover, the glycoforms, H6N2 and H3N2 were detected at the intact protein level but were not observed by the glycopeptide approach. This disparity in high-mannose form quantification might be related to their low abundance as the quantification of minor forms is often inconsistent for both approaches. Furthermore, different ionization efficiencies between mannose and sialylated species might also contribute to different relative quantification results by both techniques. Nonetheless, it remains unclear why specific high-mannose forms could not be detected during glycopeptide analysis and further research will be needed to determine whether this is a limitation of the method.

## Complementary Intact Protein and Glycopeptide Analysis

The glycopeptide approach, as described above, demonstrates advantages in sialic acid linkage determination and overall higher sensitivity as illustrated by the number of glycoforms detected. This is achieved as the same dipeptide backbone (Asn<sub>69</sub>Lys) is observed for cleaved and non-cleaved PSA after proteolytic digestion by trypsin, and represents a summed glycosylation profile of all proteoforms. However, specificity in relation to each proteolytic cleavage variant is lost. This is in contrast with the intact protein analysis, where six distinct proteolytic cleavage variants were detected, including non-glycosylated PSA for which the peptide is too small to be assessed by CE-ESI-MS. In addition, the intact protein analysis was able to determine the most abundant glycoforms per proteolytic cleavage variant (**Figure 6** and **Supporting Information, Figure S-2**). One of the most interesting findings was the presence of an extra *N*-glycosylation site on PSA of patient E which could be confirmed by the glycopeptide approach (**Supporting Information, Table S-5**). The identification in the glycopeptide approach had initially been hampered due to the co-migration of more intense tryptic peptide peaks (**Supporting Information, Figure S-3**). As in agreement with the intact protein analysis, high-mannose type glycans H5N2 and H6N2 were detected as the two most abundant glycans present at this site and the non-glycosylated variant was also not observed. Interestingly, the second *N*-glycosylation site has only been described in one other study, of which the reports of high-mannose structures are consistent.<sup>27</sup> Overall, this study demonstrates the complementarity between the two different approaches, which together form a highly suitable platform to study the proteolytic cleaved PSA variants, non-cleaved PSA and its glycosylation as each is required in order to perform full characterization of PSA, and in some cases, may be used to crosscheck results.

## Perspectives

Urine, serum and seminal plasma represent different biological reservoirs of PSA,<sup>12,21,45</sup> and PSA analysis from each of these biofluids comes with specific challenges as well as advantages for clinical use. During this study, the main focus was on urinary PSA, yet PSA proteoforms in other biofluids, particularly serum, also lack such an in-depth analysis. In order to apply the established assay to serum PSA, further development would be required as serum PSA is generally 5 to 50-fold lower

in concentration compared to urine.<sup>21</sup> For this purpose, it would be necessary to improve the sensitivity, capturing efficiency and detection of the assay. Furthermore, during this study, some patients revealed urinary PSA concentrations that were too low for intact protein analysis, therefore the analysis of urine following DRE might be more suitable as it would be expected to have a higher concentration of PSA.<sup>46</sup> Finally, improvements to intact protein sample and data throughput would be needed in order to obtain statistically significant results from a larger clinical cohort. Nevertheless, further exploration of PSA from urine, serum and seminal plasma should be continued as it represents a rich source of potential informative and non-invasive biomarkers for prostate cancer.

## Conclusions

In this study, we validated an intact protein assay for the determination of PSA proteoforms. The assay was applied to a small cohort of patients, in which six different proteolytic cleavage variants were detected. The proteolytic cleavage variants included double and single cleaved forms and non-cleaved forms of PSA that have also been described in seminal plasma and serum. Additionally, cleaved PSA showed a similar relative abundance of glycoforms as non-cleaved PSA. Glycosylation profiles obtained at the intact protein level were confirmed using the more in-depth glycopeptide approach, and both approaches in combination determined a second *N*-linked glycosylation site of PSA in one patient. Finally, it was illustrated that the developed protocol is able to assess cleaved PSA (intact protein analysis) and glycosylation (glycopeptide analysis) of individual patients in the same procedure, revealing the benefit of adding the intact protein level to the bottom-up approach.

## Data Availability

The raw mass spectrometric data files that support the findings of this study are available in the MassIVE repository and may be found with the data set identifier MSV000086699 <https://doi.org/10.25345/C5HF79>.

## Funding

This research was supported by Marie Skłodowska-Curie actions as part of European Union's Horizon 2020 Research and Innovation Program (GlySign, Grant No. 722095), Cure for Cancer Foundation and Astellas Pharma B.V. The authors declare no competing financial interest.

## Acknowledgements

We thank Sarah Azaabal for her early work on the intact protein CE-ESI-MS method. We would also like to thank Bas Jansen for his valuable guidance on using LaCyTools.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jprot.2021.104148>:

**Supplementary Figures:** **Figure S-1.** Intact PSA assay for the analysis of patient urinary PSA with CE-ESI-MS. **Figure S-2.** Relative abundances of di-sialylated glycans on cleaved PSA and non-cleaved PSA. **Figure S-3.** Tryptic PSA measured by CE-ESI-MS.

**Supplementary Tables:** **Table S-1.** Intact CE-ESI-MS and data processing parameters. **Table S-2.** Repeatability and intermediate precision CE-ESI-MS. **Table S-3.** Patient characteristics and intact protein assignments. **Table S-4.** Comparison of glycoform relative abundances in patients between intact and glycopeptide approaches. **Table S-5.** Patient characteristics and glycopeptide assignments.

## References

- (1) De Angelis, G.; Rittenhouse, H. G.; Mikolajczyk, S. D.; Blair Shamel, L.; Semjonow, A. Twenty Years of PSA: From Prostate Antigen to Tumor Marker. *Reviews in urology*. **2007**, *9* (3).
- (2) Kouriefs, C.; Sahoyl, M.; Grange, P.; Muir, G. Prostate Specific Antigen through the Years. *Archivio Italiano di Urologia e Andrologia*. **2009**, *81* (4), 195–198.
- (3) Virginia A Moyer; U.S. Preventive Services Task Force. Screening for Prostate Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals of Internal Medicine*. **2012**, *157* (2), 120–134. DOI: 10.7326/0003-4819-157-2-201207170-00459.

- (4) Schröder, F. H. et al. Screening and Prostate Cancer Mortality: Results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 Years of Follow-Up. *The Lancet*. **2014**, 384 (9959). DOI: 10.1016/S0140-6736(14)60525-0.
- (5) Ross, T.; Ahmed, K.; Raison, N.; Challacombe, B.; Dasgupta, P. Clarifying the PSA Grey Zone: The Management of Patients with a Borderline PSA. *International Journal of Clinical Practice*. **2016**, 70 (11). DOI: 10.1111/ijcp.12883.
- (6) Mikolajczyk, S. D.; Millar, L. S.; Wang, T. J.; Rittenhouse, H. G.; Marks, L. S.; Song, W.; Wheeler, T. M.; Slawin, K. M. A Precursor Form of Prostate-Specific Antigen Is More Highly Elevated in Prostate Cancer Compared with Benign Transition Zone Prostate Tissue. *Cancer Research*. **2000**, 60 (3).
- (7) Mikolajczyk, S. D.; Millar, L. S.; Wang, T. J.; Rittenhouse, H. G.; Wolfert, R. L.; Marks, L. S.; Song, W.; Wheeler, T. M.; Slawin, K. M. "BPSA," a Specific Molecular Form of Free Prostate-Specific Antigen, Is Found Predominantly in the Transition Zone of Patients with Nodular Benign Prostatic Hyperplasia. *Urology*. **2000**, 55 (1). DOI: 10.1016/S0090-4295(99)00372-6.
- (8) Watt, K. W. K.; Lee, P. J.; M'Timkulu, T.; Chan, W. P.; Loor, R. Human Prostate-Specific Antigen: Structural and Functional Similarity with Serine Proteases. *Proceedings of the National Academy of Sciences of the United States of America*. **1986**, 83 (10). DOI: 10.1073/pnas.83.10.3166.
- (9) Mattsson, J. M.; Valmu, L.; Laakkonen, P.; Stenman, U. H.; Koistinen, H. Structural Characterization and Anti-Angiogenic Properties of Prostate-Specific Antigen Isoforms in Seminal Fluid. *Prostate*. **2008**, 68 (9), 945–954. DOI: 10.1002/pros.20751.
- (10) Peltola, M. T.; Niemelä, P.; Väisänen, V.; Viitanen, T.; Alanen, K.; Nurmi, M.; Pettersson, K. Intact and Internally Cleaved Free Prostate-Specific Antigen in Patients with Prostate Cancer with Different Pathologic Stages and Grades. *Urology*. **2011**, 77 (4). DOI: 10.1016/j.urology.2010.11.007.
- (11) Untergasser, G.; Madersbacher, S.; Berger, P. Benign Prostatic Hyperplasia: Age-Related Tissue-Remodeling. *Experimental Gerontology*. **2005**, 40 (3). DOI: 10.1016/j.exger.2004.12.008.
- (12) Lilja, H.; Ulmert, D.; Vickers, A. J. Prostate-Specific Antigen and Prostate Cancer: Prediction, Detection and Monitoring. *Nature Reviews Cancer*. **2008**, 8 (4). DOI: 10.1038/nrc2351.
- (13) Balk, S. P.; Ko, Y. J.; Bubley, G. J. Biology of Prostate-Specific Antigen. *Journal of Clinical Oncology*. **2003**, 21 (2). DOI: 10.1200/JCO.2003.02.083.
- (14) Özen, H.; Sözen, S. PSA Isoforms in Prostate Cancer Detection. *European Urology Supplements*. **2006**, 5 (6), 495–499. DOI: <https://doi.org/10.1016/j.eursup.2006.02.017>.
- (15) Stenman, U. H.; Leinonen, J.; Alfthan, H.; Rannikko, S.; Tuhkanen, K.; Alfthan, O. A Complex between Prostate-Specific Antigen and Alpha 1-Antichymotrypsin Is the Major Form of Prostate-Specific Antigen in Serum of Patients with Prostatic Cancer: Assay of the Complex Improves Clinical Sensitivity for Cancer. *Cancer research*. **1991**, 51 (1), 222–226.
- (16) Zhang, W. M.; Leinonen, J.; Kalkkinen, N.; Dowell, B.; Stenman, U. H. Purification and Characterization of Different Molecular Forms of Prostate-Specific Antigen in Human Seminal Fluid. *Clinical Chemistry*. **1995**, 41 (11). DOI: 10.1093/clinchem/41.11.1567.

- (17) Barrabés, S.; Farina-Gomez, N.; Llop, E.; Puerta, A.; Diez-Masa, J. C.; Perry, A.; de Llorens, R.; de Frutos, M.; Peracaula, R. Comparative Analysis of Prostate-Specific Antigen by Two-Dimensional Gel Electrophoresis and Capillary Electrophoresis. *Electrophoresis*. **2017**, *38* (3–4), 408–416. DOI: 10.1002/elps.201600432.
- (18) Vermassen, T.; Speeckaert, M. M.; Lumen, N.; Rottey, S.; Delanghe, J. R. Glycosylation of Prostate Specific Antigen and Its Potential Diagnostic Applications. *Clinica chimica acta; international journal of clinical chemistry*. **2012**, *413* (19–20), 1500–1505. DOI: 10.1016/j.cca.2012.06.007.
- (19) Iwakiri, J.; Granbois, K.; Wehner, N.; Graves, H. C.; Stamey, T. An Analysis of Urinary Prostate Specific Antigen before and after Radical Prostatectomy: Evidence for Secretion of Prostate Specific Antigen by the Periurethral Glands. *The Journal of urology*. **1993**, *149* (4), 783–786. DOI: 10.1016/s0022-5347(17)36207-9.
- (20) Oesterling, J. E.; Tekchandani, A. H.; Martin, S. K.; Bergstralh, E. J.; Reichstein, E.; Diamandis, E. P.; Yemoto, C.; Stamey, T. A. The Periurethral Glands Do Not Significantly Influence the Serum Prostate Specific Antigen Concentration. *The Journal of urology*. **1996**, *155* (5), 1658–1660.
- (21) Bolduc, S.; Lacombe, L.; Naud, A.; Grégoire, M.; Fradet, Y.; Tremblay, R. R. Urinary PSA: A Potential Useful Marker When Serum PSA Is between 2.5 Ng/ml and 10 Ng/ml. *Journal of the Canadian Urological Association*. **2007**, *1* (4). DOI: 10.5489/cuaj.444.
- (22) Sato, I.; Sagi, M.; Ishiwari, A.; Nishijima, H.; Ito, E.; Mukai, T. Use of the “SMITEST” PSA Card to Identify the Presence of Prostate-Specific Antigen in Semen and Male Urine. *Forensic Science International*. **2002**, *127* (1–2). DOI: 10.1016/S0379-0738(02)00111-1.
- (23) Scott, E.; Munkley, J. Glycans as Biomarkers in Prostate Cancer. *International Journal of Molecular Sciences*. 2019. DOI: 10.3390/ijms20061389.
- (24) Yoneyama, T.; Ohyama, C.; Hatakeyama, S.; Narita, S.; Habuchi, T.; Koie, T.; Mori, K.; Hidari, K. I. P. J.; Yamaguchi, M.; Suzuki, T.; Tobisawa, Y. Measurement of Aberrant Glycosylation of Prostate Specific Antigen Can Improve Specificity in Early Detection of Prostate Cancer. *Biochemical and Biophysical Research Communications*. **2014**. DOI: 10.1016/j.bbrc.2014.04.107.
- (25) Leymarie, N. et al. Interlaboratory Study on Differential Analysis of Protein Glycosylation by Mass Spectrometry: The ABRF Glycoprotein Research Multi-Institutional Study 2012. *Molecular & cellular proteomics: MCP*. **2013**, *12* (10), 2935–2951. DOI: 10.1074/mcp.M113.030643.
- (26) Mononen, N.; Seppälä, E. H.; Duggal, P.; Autio, V.; Ikonen, T.; Ellonen, P.; Saharinen, J.; Saarela, J.; Vihinen, M.; Tammela, T. L. J.; Kallioniemi, O.; Bailey-Wilson, J. E.; Schleutker, J. Profiling Genetic Variation along the Androgen Biosynthesis and Metabolism Pathways Implicates Several Single Nucleotide Polymorphisms and Their Combinations as Prostate Cancer Risk Factors. *Cancer research*. **2006**, *66* (2), 743–747. DOI: 10.1158/0008-5472.CAN-05-1723.
- (27) Song, E.; Hu, Y.; Hussein, A.; Yu, C. Y.; Tang, H.; Mechref, Y. Characterization of the Glycosylation Site of Human PSA Prompted by Missense Mutation Using LC-MS/MS. *Journal of Proteome Research*. **2015**. DOI: 10.1021/acs.jproteome.5b00362.

- (28) Kammeijer, G. S. M.; Nouta, J.; De La Rosette, J. J. M. C. H.; De Reijke, T. M.; Wuhrer, M. An In-Depth Glycosylation Assay for Urinary Prostate-Specific Antigen. *Analytical Chemistry*. **2018**, *90* (7), 4414–4421. DOI: 10.1021/acs.analchem.7b04281.
- (29) Sarrats, A.; Saldova, R.; Comet, J.; O'Donoghue, N.; De Llorens, R.; Rudd, P. M.; Peracaula, R. Glycan Characterization of PSA 2-DE Subforms from Serum and Seminal Plasma. *OMICS A Journal of Integrative Biology*. **2010**. DOI: 10.1089/omi.2010.0050.
- (30) Santos, M. R.; Ratnayake, C. K.; Horn, D. M.; Karger, B. L.; Ivanov, A. R.; Viner, R. Separation and Analysis of Intact Prostate Specific Antigen (PSA) and its Proteoforms by CESI-MS Under Native and Denaturing Conditions. [https://sciex.jp/content/dam/SCIEX/pdf/tech-notes/ce1/CESI-MS\\_Intact\\_Native\\_Protein\\_Proteoforms\\_PSA.pdf](https://sciex.jp/content/dam/SCIEX/pdf/tech-notes/ce1/CESI-MS_Intact_Native_Protein_Proteoforms_PSA.pdf) (accessed 2023 -05 -21).
- (31) Santos, M. R.; Ratnayake, C. K.; Fonslow, B.; Guttman, A. A covalent , cationic polymer coating method for the CESI-MS analysis of intact proteins and polypeptides [www.sciex.com/content/dam/SCIEX/pdf/tech-notes/all/Covalent-cationic-polymer-coating-method-CESI-MS.pdf](http://www.sciex.com/content/dam/SCIEX/pdf/tech-notes/all/Covalent-cationic-polymer-coating-method-CESI-MS.pdf) (accessed 2022 -10 -29).
- (32) Kammeijer, G. S. M.; Kohler, I.; Jansen, B. C.; Hensbergen, P. J.; Mayboroda, O. A.; Falck, D.; Wuhrer, M. Dopant Enriched Nitrogen Gas Combined with Sheathless Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry for Improved Sensitivity and Repeatability in Glycopeptide Analysis. *Analytical Chemistry*. **2016**, *88* (11). DOI: 10.1021/acs.analchem.6b00479.
- (33) Jansen, B. C.; Falck, D.; De Haan, N.; Hipgrave Ederveen, A. L.; Razdorov, G.; Lauc, G.; Wuhrer, M. LaCyTools: A Targeted Liquid Chromatography-Mass Spectrometry Data Processing Package for Relative Quantitation of Glycopeptides. *Journal of Proteome Research*. **2016**, *15* (7). DOI: 10.1021/acs.jproteome.6b00171.
- (34) Wu, J. T.; Zhang, P.; Wang, T.; Wilson, L.; Astill, M. Evaluation of Free PSA Isoforms, PSA Complex Formation, and Specificity of Anti-PSA Antibodies by HPLC and PAGE-Immunoblotting Techniques. *Journal of clinical laboratory analysis*. **1995**, *9* (1), 1–14. DOI: 10.1002/jcla.1860090102.
- (35) Linton, H. J.; Marks, L. S.; Millar, L. S.; Knott, C. L.; Rittenhouse, H. G.; Mikolajczyk, S. D. Benign Prostate-Specific Antigen (BPSA) in Serum Is Increased in Benign Prostate Disease. *Clinical Chemistry*. **2003**, *49* (2), 253–259. DOI: 10.1373/49.2.253.
- (36) Wang, T. J.; Slawin, K. M.; Rittenhouse, H. G.; Millar, L. S.; Mikolajczyk, S. D. Benign Prostatic Hyperplasia-Associated Prostate-Specific Antigen (BPSA) Shows Unique Immunoreactivity with Anti-PSA Monoclonal Antibodies. *European Journal of Biochemistry*. **2000**, *267* (13). DOI: 10.1046/j.1432-1327.2000.01441.x.
- (37) Gilgunn, S.; Conroy, P. J.; Saldova, R.; Rudd, P. M.; O'Kennedy, R. J. Aberrant PSA Glycosylation - A Sweet Predictor of Prostate Cancer. *Nature Reviews Urology*. **2013**, *10* (2). DOI: 10.1038/nrurol.2012.258.
- (38) Mikolajczyk, S. D.; Marks, L. S.; Partin, A. W.; Rittenhouse, H. G. Free Prostate-Specific Antigen in Serum Is Becoming More Complex. *Urology*. **2002**, *59* (6). DOI: 10.1016/S0090-4295(01)01605-3.
- (39) Varki, A.; Kannagi, R.; Toole, B.; Stanley, P. Glycosylation Changes in Cancer.; Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita,



- T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., Seeberger, P. H., Eds.; Cold Spring Harbor (NY), **2015**; pp 597–609. DOI: 10.1101/glycobiology.3e.047.
- (40) Charrier, J. P.; Tournel, C.; Michel, S.; Comby, S.; Jolivet-Reynaud, C.; Passagot, J.; Dalbon, P.; Chautard, D.; Jolivet, M. Differential Diagnosis of Prostate Cancer and Benign Prostate Hyperplasia Using Two-Dimensional Electrophoresis. *Electrophoresis*. **2001**, *22* (9). DOI: 10.1002/1522-2683(200105)22:9<1861::AID-ELPS1861>3.0.CO;2-6.
- (41) Thakur, D.; Rejtar, T.; Karger, B. L.; Washburn, N. J.; Bosques, C. J.; Gunay, N. S.; Shriver, Z.; Venkataraman, G. Profiling the Glycoforms of the Intact Alpha Subunit of Recombinant Human Chorionic Gonadotropin by High-Resolution Capillary Electrophoresis-Mass Spectrometry. *Analytical chemistry*. **2009**, *81* (21), 8900–8907. DOI: 10.1021/ac901506p.
- (42) Haselberg, R.; De Vijlder, T.; Heukers, R.; Smit, M. J.; Romijn, E. P.; Somsen, G. W.; Domínguez-Vega, E. Heterogeneity Assessment of Antibody-Derived Therapeutics at the Intact and Middle-up Level by Low-Flow Sheathless Capillary Electrophoresis-Mass Spectrometry. *Analytica Chimica Acta*. **2018**, *1044*. DOI: 10.1016/j.aca.2018.08.024.
- (43) Van Den Broek, I.; Van Dongen, W. D. LC-MS-Based Quantification of Intact Proteins: Perspective for Clinical and Bioanalytical Applications. *Bioanalysis*. **2015**, *7* (15), 1943–1958. DOI: 10.4155/bio.15.113.
- (44) Qiu, X.; Kang, L.; Case, M.; Weng, N.; Jian, W. Quantitation of Intact Monoclonal Antibody in Biological Samples: Comparison of Different Data Processing Strategies. *Bioanalysis*. **2018**, *10* (13). DOI: 10.4155/bio-2018-0016.
- (45) Lilja, H.; Oldbring, J.; Rannevik, G.; Laurell, C. B. Seminal Vesicle-Secreted Proteins and Their Reactions during Gelation and Liquefaction of Human Semen. *The Journal of clinical investigation*. **1987**, *80* (2). DOI: 10.1172/JCI113070.
- (46) Vermassen, T.; Van Praet, C.; Vanderschaeghe, D.; Maenhout, T.; Lumen, N.; Callewaert, N.; Hoebeke, P.; Van Belle, S.; Rottey, S.; Delanghe, J. Capillary Electrophoresis of Urinary Prostate Glycoproteins Assists in the Diagnosis of Prostate Cancer. *Electrophoresis*. **2014**, *35* (7). DOI: 10.1002/elps.201300332.





# Chapter 3

## Software-Assisted Data Processing Workflow for Intact Glycoprotein Mass Spectrometry

Alan B. Moran<sup>1</sup>, Elena Domínguez-Vega<sup>1</sup>, Manfred Wuhrer<sup>1</sup>, Guinevere S.M. Lageveen-Kammeijer<sup>1,2\*</sup>

<sup>1</sup> Leiden University Medical Center, Center for Proteomics and Metabolomics, 2300 RC Leiden, The Netherlands

<sup>2</sup> University of Groningen, Department of Analytical Biochemistry, Groningen Research Institute of Pharmacy, Groningen, The Netherlands

Reprinted (adapted) from Moran, A. B., Domínguez-Vega, E., Wuhrer, M., & Lageveen-Kammeijer, G. S. M. (2023). Software-Assisted Data Processing Workflow for Intact Glycoprotein Mass Spectrometry. *Journal of Proteome Research*, 22(4), 1367–1376, <https://doi.org/10.1021/acs.jproteome.2c00762> under the Creative Commons CC-BY license.



## Abstract

Intact protein analysis by mass spectrometry is important for several applications such as assessing post-translational modifications and biotransformation. In particular, intact protein analysis allows the detection of proteoforms that are commonly missed by other approaches such as proteolytic digestion followed by bottom-up analysis. Two quantification methods are mainly used for intact protein data quantification, namely the extracted ion and deconvolution approaches. However, a consensus with regard to a single best practice for intact protein data processing is lacking. Furthermore, many data processing tools are not fit-for-purpose and, as a result, the analysis of intact proteins is laborious and lacks the throughput required to be implemented for the analysis of clinical cohorts. Therefore, in this study, we investigated the application of a software-assisted data analysis and processing workflow in order to streamline intact protein integration, annotation and quantification via deconvolution. In addition, the assessment of orthogonal datasets generated *via* middle-up and bottom-up analysis enabled the cross-validation of cleavage proteoform assignments present in seminal prostate-specific antigen (PSA). Furthermore, deconvolution quantification of PSA from patients' urine revealed results that were comparable with manually-performed quantification based on extracted ion electropherograms. Overall, the presented workflow allows fast and efficient processing of intact protein data. The raw data is available on MassIVE using the identifier MSV000086699.

## Introduction

The analysis of intact proteins by mass spectrometry (MS) involves the examination of the complete protein, including various post- and co-translational modifications.<sup>1</sup> During electrospray ionization (ESI)-MS, a protein may be analyzed under native or denaturing conditions.<sup>2</sup> The latter is referred to here as intact protein analysis, and involves the application of volatile MS-compatible solvents with a low pH in order to improve the solubility and ionization of the protein.<sup>2</sup> Furthermore, an online separation technique is often employed prior to the introduction of the protein into the MS. For example, reversed-phase liquid chromatography (RPLC)-MS is commonly employed to analyze biotherapeutics<sup>3,4</sup>, yet this approach has been reported to show poor peak resolution of proteoforms during intact protein analysis.<sup>1</sup> In contrast, hydrophilic interaction LC (HILIC)-MS has demonstrated efficient separation of protein glycoforms.<sup>5,6</sup> In addition, capillary electrophoresis (CE)-ESI-MS has been recognized as an excellent technique to investigate intact proteins as proteoforms may be separated based on their intrinsic properties, including also post-translational modifications.<sup>7,8</sup>

Intact protein analysis offers multiple advantages over solely performing a protease digestion and bottom-up investigation. For example, minimal sample preparation is required, thus there is a smaller likelihood that modifications may be introduced to the protein and less time is needed for sample processing. In addition, different proteoforms of the protein may also be observed during the analysis.<sup>9</sup> Despite this, a global approach that incorporates information from different levels, from top-down to bottom-up, is often required during protein analysis<sup>2</sup> and enables the analysis of the entire protein as well as cross-validation between the results.<sup>10</sup>

When MS is hyphenated with a separation technique, there are two predominant methods of performing data processing and quantification of intact protein spectra, namely either *via* extracted ion chromatogram (XIC) or mass deconvolution approaches. Both techniques have been well covered in recent reviews.<sup>1,11,12</sup> Briefly, the XIC method involves the determination of the area under the peak *via* selection of one or several  $m/z$  that generally cover the most abundant charge states of the protein. In order to maximize sensitivity and specificity, the selection of the charge states as

well as the width of the mass window are important parameters for this method, respectively.<sup>13</sup> In comparison, deconvolution employs an algorithm to convert the multiple charge states observed in an intact protein mass spectrum into a neutral spectrum that demonstrates the masses of the observed proteoforms. There are several algorithms available to perform this function with maximum entropy<sup>14</sup> being the most commonly employed by most data processing softwares,<sup>15</sup> although there is the emergence of more recently developed approaches such as parsimonious charge deconvolution.<sup>15</sup> Importantly, the input  $m/z$  range as well as the output range require optimization by the user to ensure a suitable number of charge states of the protein are included in the formula, while also aiming to reduce the production of any artifacts due to the data processing algorithm.<sup>1,11,15–17</sup> Overall, there is still no clear consensus in the field as to which is the most suitable technique to apply when performing intact protein data processing.<sup>1,17</sup> Undoubtedly, in order to develop a set of best data processing practices, there is a need for comparisons between the two approaches to be made within the same software, as well as software offered by different vendors.<sup>13,17</sup>

In general, several intact protein studies are mainly concerned with the absolute quantitation of biotherapeutics,<sup>13,16–19</sup> although the determination of proteoform relative abundance has also been applied for the quantification of drug-antibody ratios.<sup>20,21</sup> Despite this, the determination of best practices for intact protein data processing, based on these studies, does not encompass challenges faced in a biomarker discovery setting. For example, relative quantification is a suitable approach for the quantification of intact proteoforms in the clinical setting as differences between patient groups may be readily observed.<sup>1</sup> Furthermore, the approaches applied for assessing biotherapeutics<sup>16–19,22,23</sup> may not account for patient to patient variation that is observed in clinical assays.<sup>7</sup> Finally, it has been recognized that the throughput of data processing is one of the main challenges facing the intact protein analysis of clinical cohorts whereby large numbers of samples are required to derive statistically significant data.<sup>7</sup>

In this study, we sought to improve the throughput and efficiency of intact protein data processing by developing a software-assisted workflow. This approach was demonstrated using urinary PSA to compare the deconvolution (software-assisted) quantification results with the previously published extracted ion electropherogram

(XIE) data (manual),<sup>7</sup> both of which were generated by using two different software tools. In addition, we also further examined the proteoform profile of prostate-specific antigen (PSA) by performing orthogonal analyses of seminal PSA *via* intact protein, middle-up and bottom-up approaches, and compared this with the previously established profile of urinary PSA.<sup>7</sup>

## Experimental Section

### Sample Preparation

The sample preparation of urinary PSA, including sample collection, immunocapturing, and in-solution tryptic digestion, has previously been described.<sup>7,24</sup> Seminal PSA standard (Lee BioSolutions, St. Louis, MO) was prepared for intact protein analysis as follows: PSA was reconstituted (2.2  $\mu\text{g}/\mu\text{L}$ ) in LC-MS grade  $\text{H}_2\text{O}$  (Fluka, Steinheim, Germany) and buffer-exchange was carried out using centrifugal filters with a 10 kDa MWCO (Merck Life Science, Amsterdam, The Netherlands). This was performed by conditioning the filter with 500  $\mu\text{L}$   $\text{H}_2\text{O}$  followed by centrifugation (14,000  $\text{g} \times 5$  min). The filtrate was discarded and the sample (26  $\mu\text{g}$ ) was added to the filter. The volume was made up to 500  $\mu\text{L}$  in total with  $\text{H}_2\text{O}$ . Another centrifugation step was performed and the filtrate was discarded. Then, 250  $\mu\text{L}$   $\text{H}_2\text{O}$  was added, centrifugation was carried out (14,000  $\text{g} \times 5$  min), and the filtrate was removed. This was repeated three times in total. Finally, the sample was retrieved by inverting the filter into a fresh tube and centrifuging (4000  $\text{g} \times 5$  min).

The reduction (and alkylation) of seminal PSA for middle-up analysis was carried out with PSA prepared at a concentration of 2.2  $\mu\text{g}/\mu\text{L}$ . The sample (100  $\mu\text{g}$ ) was added up to 100  $\mu\text{L}$  with  $\text{H}_2\text{O}$  in an Eppendorf tube (1.5 mL). Then, 1  $\mu\text{L}$  of 200 mM DL-dithiothreitol (DTT, Sigma-Aldrich, Steinheim, Germany) was added at a final concentration of 2 mM. The sample was vortexed for one min and heated at 60°C for 30 min. Following this, 1.5  $\mu\text{L}$  of 400 mM iodoacetamide (IAA, Sigma-Aldrich) was added (final concentration of 6 mM). For the preparation of reduced samples without alkylation, the same volume of  $\text{H}_2\text{O}$  was added instead of IAA. The samples were incubated at room temperature (RT) in the dark for 60 min. DTT (200 mM) was added (3  $\mu\text{L}$ ) at a final concentration of 6 mM. This was followed by an incubation at RT in



brightness for 20 min. Finally, the samples were desalted by performing the buffer-exchange procedure as described above.

### **CE-ESI-MS**

The CE experiments were carried out using a CESI 8000 (Sciex, Brea, CA). All capillaries were sheathless bare-fused silica (BFS) with a porous tip (91 cm, 30  $\mu\text{m}$  i.d.  $\times$  150  $\mu\text{m}$  o.d.) and in the case of intact protein and middle-up analysis, capillaries were coated in-house with polyethylenimine (Gelest, Morrisville, NC)<sup>25</sup> as previously published.<sup>7</sup> Prior to the separation, background electrolyte (BGE) consisting of 20% glacial acetic acid (HAc, Sigma-Aldrich) was prepared (*v/v*, 3.49 M, pH 2.3) and was used to rinse (100 psi  $\times$  5 min) the separation line. Then, the conductive capillary was filled (100 psi  $\times$  4 min) with BGE and the sample was hydrodynamically injected. In the case of seminal PSA, an injection of 2.5 psi  $\times$  15 sec was applied (approximately 5 nL, 0.8% of the total capillary volume). Finally, separation was achieved by applying - 20 kV over 45 min with the capillary temperature set to 15°C.

As previously published,<sup>7,24</sup> the CE separation of PSA tryptic peptides for bottom-up was performed on non-coated BFS capillaries which were conditioned by applying 0.1 M NaOH  $\times$  2.5 min, then H<sub>2</sub>O  $\times$  3 min, 0.1 M HCl  $\times$  2.5 min and H<sub>2</sub>O  $\times$  3 min. Following this, the BGE was applied for 3 min. Digested seminal PSA standard was prepared at a concentration of 100 ng/ $\mu\text{L}$  and 6.7  $\mu\text{L}$  was mixed with 3.4  $\mu\text{L}$  of leading electrolyte, 1.2 M ammonium acetate, pH 3.39 (Fluka). Hydrodynamic injection was performed (1 psi  $\times$  60 s), corresponding to a volume of 8 nL (1.3% capillary volume). Then, an injection (0.5 psi  $\times$  25 s) of a BGE post plug was carried out. Following this, a separation voltage of 20 kV was performed for 80 min at 15°C.

The CESI 8000 was coupled with a maXis Impact Ultra-High Resolution QqTOF MS (Bruker Daltonics GmbH, Bremen, Germany) equipped with a nano-electrospray source. The MS acquisition parameters were previously published for the intact protein and middle-up approaches<sup>7</sup> as well as bottom-up analysis<sup>24,26</sup>. Importantly, in order to perform fragmentation of small peptides generated via internal cleavage of the protein followed by tryptic digestion, the following parameters were applied for a bottom-up approach using a concentrated sample of digested seminal PSA standard (100 ng/ $\mu\text{L}$ ): electrospray voltage, 1250 V; nitrogen drying gas, 1.2 L/min at 150°C;

quadrupole ion energy, 3 eV; collision cell energy, 7 eV; transfer time, 130  $\mu$ s; pre-pulse storage time, 15  $\mu$ s;  $m/z$  range,  $m/z$  150–3500.

## Data Processing

Seminal PSA data, generated by intact protein, middle-up and bottom-up analysis, was directly analyzed using Byos (v4.4, Protein Metrics, Cupertino, CA) in the Bruker DataAnalysis file format (.d). In addition, a software-assisted data processing workflow was applied to the urinary PSA intra- and interday ( $n = 9$ ), and patient ( $n = 8$ ) datasets. The workflow consisted mainly of three stages, firstly a 12-point internal mass calibration was performed in Bruker DataAnalysis (v5.0) using the  $m/z$  of the most abundant nine charge states from the most abundant PSA proteoform, active PSA containing H5N4F1S2 (28430.91 Da; 2187.9913<sup>13+</sup>, 2031.7782<sup>14+</sup>, 1896.3934<sup>15+</sup>, 1777.9318<sup>16+</sup>, 1673.4068<sup>17+</sup>, 1580.4958<sup>18+</sup>, 1497.3648<sup>19+</sup>, 1422.5469<sup>20+</sup>, 1354.8545<sup>21+</sup>). In addition, the  $m/z$  of internally spiked PSA (LSEPAELTEAVK; 1286.6837<sup>1+</sup>, 643.8454<sup>2+</sup>) and IgG (GPSVFPVAPSSK; 1172.6309<sup>1+</sup>) peptides (developed in-house by FMoc solid phase peptide synthesis) were also used in the mass calibration. Secondly, the data was converted to .mzXML format and migration time alignment was performed in LaCyTools (v2.01)<sup>27</sup> using abundant  $m/z$  values found in each sample (**Supporting Information, Table S1**). Finally, the aligned .mzXML datafiles were imported into Byos (v4.4) by Protein Metrics and the base peak electropherogram trace was used for automatic integration, annotation and quantification *via* deconvolution. The version of Byos in this work included a beta-release feature of the mass XIC function for visualization of XIC data. This feature has subsequently been officially released in Byos v4.5.

The parameters for manually generating and integrating XIEs using Bruker DataAnalysis (v5.0) were reported previously whereby the three most abundant charge states with an extraction window of  $\pm m/z$  0.1 were combined to generate an XIE (Gaussian smoothing, 2 points) for each proteoform.<sup>7</sup> In the current study, the following deconvolution settings were applied to intact protein and middle-up data: charge vector spacing, 0.6; smoothing sigma  $m/z$ , 0.02; spacing  $m/z$ , 0.04; mass smoothing sigma, 3; mass spacing, 0.5; minimum charge, 5; iteration maximum, 10. In the case of spectra with isotopic resolution, such as some of the fragments observed in the middle-up data, the following deconvolution parameters were used instead:

charge vector spacing, 0.5; smoothing sigma  $m/z$ , 0.01; spacing  $m/z$ , 0.01; mass smoothing sigma, 0.1; mass spacing, 0.1; minimum charge, 3; iteration maximum, 10. In addition, a  $m/z$  input range / mass output range of  $m/z$  1000 – 3000 / 26000 – 30000 Da and  $m/z$  600 – 3000 / 1000 – 30000 Da was applied for the intact protein and middle-up data, respectfully. However, the input and output ranges for the middle-up deconvolution settings were also further modified per peak to enable the search for fragments of different sizes and abundances. Furthermore, the integration windows used for the generation of deconvoluted spectra in urinary PSA may be found in **Supporting Information, Table S2**. In addition, bottom-up data was examined using the following processing parameters: minimum MS2 score, 15; maximum precursor  $m/z$  error,  $\pm 20$  ppm; maximum fragment  $m/z$  error,  $\pm 20$  ppm; missed cleavages, 2; fixed modification, carbamidomethyl. Importantly, fully specific and N- and C-term ragged searches were applied in order to search for peptides that have amino acid loss due to naturally occurring internal cleavage of the protein.

### Intact Proteoform Assignments

In order to automatically annotate the observed masses in the deconvoluted spectra, a delta mass list including glycan masses previously determined by MS/MS using a bottom-up approach<sup>24</sup> as well as expected cleavage variants and internal amino acid loss,<sup>7</sup> was generated (**Supporting Information, Table S3**). In the case of mass calibrated data a mass error cut-off was applied ( $\pm 25$  ppm). Furthermore, annotations were only considered if they followed the expected order of migration as shown in **Supporting Information, Table S2**. In addition, cleavage sites were specified if corresponding fragments and peptides were found in the middle-up and bottom-up datasets, respectively (**Supporting Information, Table S4**). Notably, the annotation of fragments in the middle-up results was further supported by the comparison of the reduced *versus* reduced and alkylated masses. Importantly, the mass was expected to increase by the corresponding number of cysteines (+ 57.05 Da per cysteine) that were present in each fragment due to the alkylation step.

## Results & Discussion

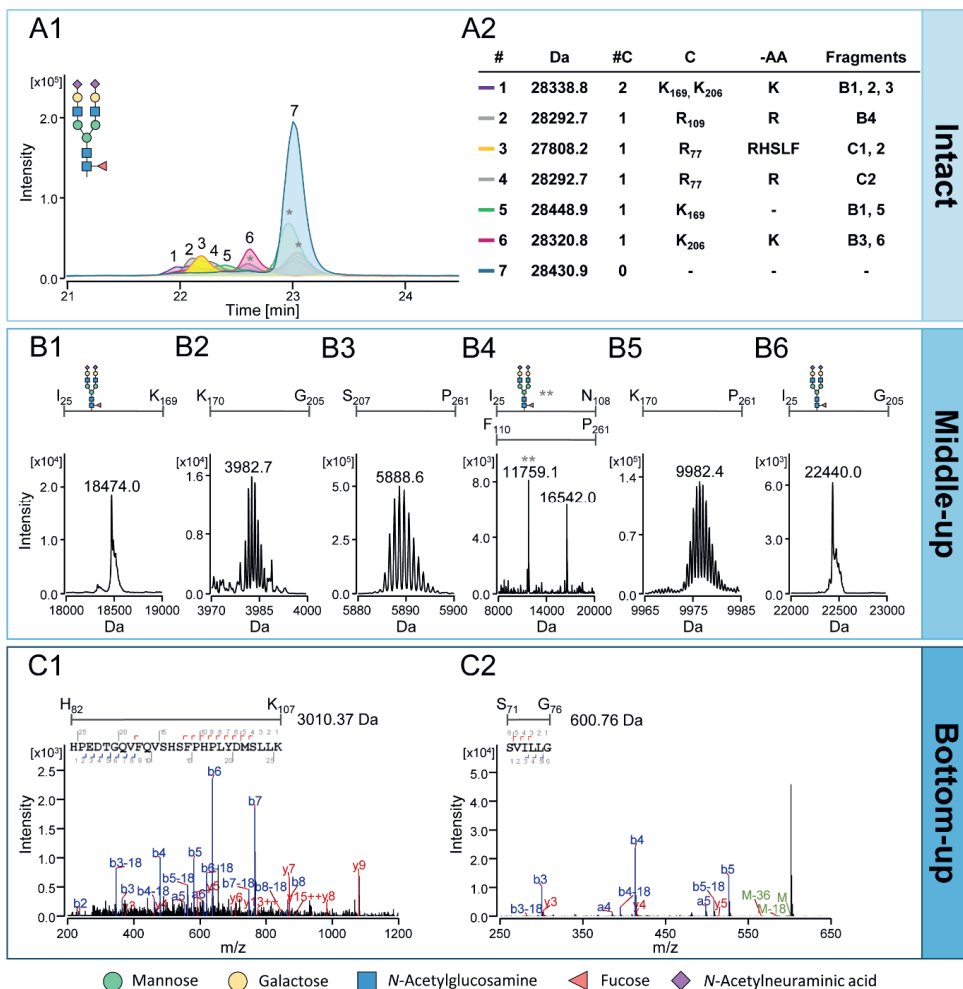
The current study further characterizes cleaved proteoforms and glycoforms in seminal PSA via CE-ESI-MS by assessing orthogonal data generated by intact

protein, middle-up and bottom-up approaches. These assignments were then compared with the previously established profile of urinary PSA.<sup>7</sup> Furthermore, the software-assisted workflow for intact protein data processing was developed by implementing three software tools to perform mass calibration (DataAnalysis), migration time alignment (LaCyTools), and deconvolution quantification (Byos). Finally, the developed workflow was applied in order to analyze intact urinary PSA proteoforms, incorporating new assignment information from the aforementioned orthogonal datasets, and performing a comparison with previously published quantification results.<sup>7</sup>

### Orthogonal Data Analysis of Seminal PSA

In **Figure 1.A1**, seminal PSA proteoforms with cleavages at various cleavage sites migrate first, as shown by peaks 1 – 6. Notably, the cleaved proteoforms with the most abundant glycoform, H5N4F1S2, are shown in **Figure 1.A1**. This is then followed by non-cleaved PSA whereby these proteoforms migrate in order of decreasing sialic acid content, from tri- to non-sialylated (**Supporting Information, Figure S1**). The electrophoretic profile of intact seminal PSA was similar to the urinary PSA profile as previously demonstrated.<sup>7</sup> The cleavage site and number of cleavages of PSA proteoforms are determined based on the amino acid loss from the internal sequence and the total number of water molecule (+18 Da) additions, respectively.<sup>7</sup> The intact mass gives the sum of all modifications to the protein and, therefore, it is useful to further dissect the nature and site of modifications *via* orthogonal approaches to support intact protein assignments.

PSA with double-cleavage was observed migrating in peak 1 (22.0 min) in **Figure 1.A1**. The mass 28338.8 Da was tentatively assigned as having cleavages at K<sub>169</sub> and K<sub>206</sub>, and the loss of one K residue. In the previous study with urinary PSA, the proteoform with the mass 28338.8 Da was assigned as double-cleavage variant at the site E<sub>145</sub>.<sup>7</sup> However, this assignment is revised in the current study based on new evidence. For example, cleavage at K<sub>169</sub> and K<sub>206</sub> is supported by the fragments **B1 – 3** in **Figure 1** and, importantly, fragment **B2** (3982.7 Da) is the result of a double cleavage at K<sub>169</sub> and K<sub>206</sub>, and the loss of K<sub>206</sub>. In addition, the peptide V<sub>138</sub> – F<sub>165</sub> with amino acid loss of LTPK<sub>169</sub> was observed in the bottom-up analysis (**Supporting Information, Table S4**). Finally, the loss of a positively charged K decreases the net



**Figure 1. Analysis of seminal PSA via orthogonal approaches; (A) intact protein, (B) middle-up, and (C) bottom-up. (A1)** Electrophoretic profile of seminal PSA with XIE peaks 1 – 7. Only proteoforms with the most abundant glycoform H5N4F1S2 are shown. Asterisk (\*) denotes XIEs from overlapping *m/z* that are present in the charge envelopes of several different proteoforms. The table in **(A2)** shows the underlying proteoforms including the peak number (#), intact mass (Da), number of cleavages (#C), cleavage site (C), amino acid loss (-AA), and masses (Fragments) that support the assignment from the middle- and bottom-up approaches. PSA fragments found during middle-up analysis are shown in **(B1 – B6)** whereby the first and last residue of the fragment as well as the glycoform are represented above the deconvoluted spectra. The average mass is illustrated except when isotopic resolution is achieved, in which case the mass of the most abundant isotope is demonstrated. In **B4**, two fragments are shown, I<sub>25</sub> – N<sub>108</sub>, containing H5N4F1S2 (11759.1 Da) as illustrated by the double asterisk (\*\*), and F<sub>110</sub> – P<sub>261</sub> (16542.0 Da). Two PSA tryptic peptides that were found as a result of prior internal cleavage of the protein at and loss of R<sub>77</sub> are shown **(C1 and C2)**.

positive charge of the proteoform which is expected to decrease the migration time of cleaved PSA. In contrast, the loss of a negatively charged E would increase the net positive charge of the protein, thus increasing its migration time. Moreover, no evidence for fragments or peptides associated with cleavage at E<sub>145</sub> could be found by the middle-up or bottom-up approaches. Overall, these results suggest that the double cleavage proteoforms observed in seminal and urinary PSA are due to cleavages at K<sub>169</sub> and K<sub>206</sub>. Significantly, PSA with a cleavage at K<sub>169</sub> and K<sub>206</sub> is also referred to as benign PSA (bPSA) due to its association with the development of Benign Prostate Hyperplasia (BPH).<sup>28,29</sup> Thus, the revision of this assignment as a result of new evidence offered by orthogonal approaches enables this intact PSA assay to be used to identify and monitor the abundance of bPSA.

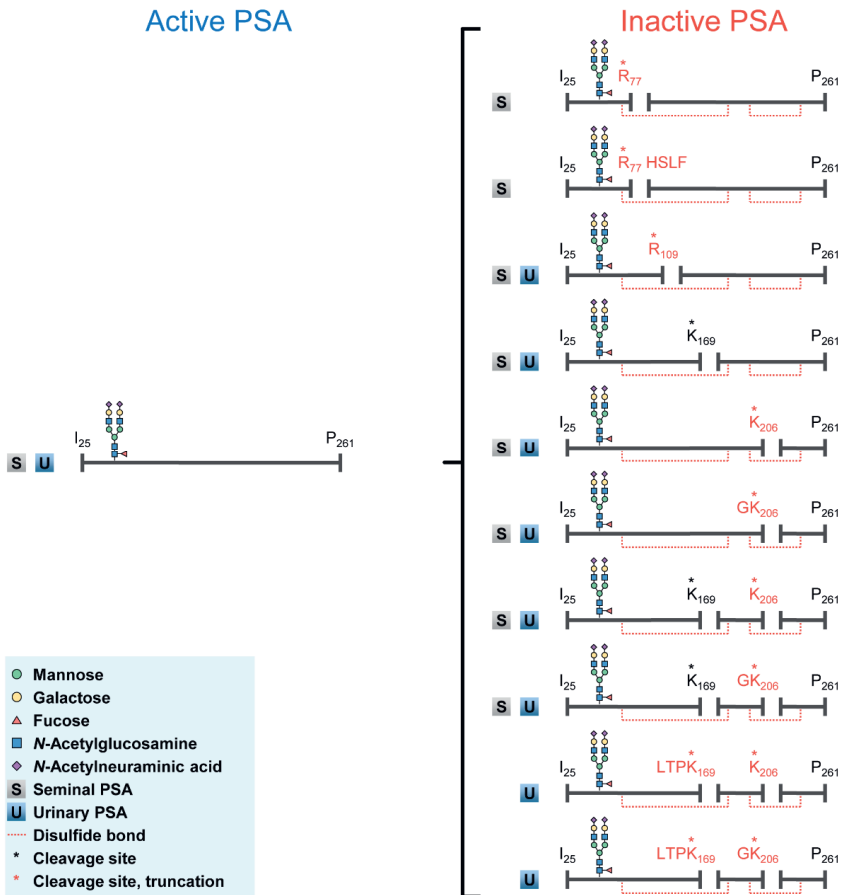
Two single-cleavage isoforms with the loss of R (28292.7 Da) were detected at peaks 2 and 4 in **Figure 1.A1**. In addition, cleaved PSA at R<sub>77</sub> and loss of RSLF (27808.2 Da) was observed in peak 3 in the intact protein profile. Further investigation by middle-up detected the fragments I<sub>25</sub> – G<sub>76</sub>, H5N4F1S2 (8184.0 Da) and H<sub>78</sub> – P<sub>261</sub> (20689.4 Da) which correspond to cleavage at and loss of R<sub>77</sub>. In addition, reduced seminal PSA showed that the fragment H<sub>82</sub> – P<sub>261</sub> with the loss of HSLF<sub>81</sub> (20204.9 Da) migrated earlier than H<sub>78</sub> – P<sub>261</sub>, most likely due to the loss of the positively charged H residue (**Supporting Information, Table S4 and Figure S2**). This was reflected in the intact protein profile as 27808.2 Da (peak 3) migrates before 28292.7 Da (peak 4), corresponding to cleavage at the site R<sub>77</sub>, with and without loss of HSLF, respectively. Cleavage at this site is also supported by bottom-up analysis as peptides with loss of HSLF<sub>81</sub> (**C1**) and R<sub>77</sub> (**C2**) were found, next to non-truncated forms (**Supporting Information, Figure S3**). Additionally, a 10.7 kDa fragment was observed in the seminal PSA standard which may correspond to the fragment S<sub>79</sub> – V<sub>174</sub> (**Supporting Information, Table S5**). This fragment may also be due to cleavage at R<sub>77</sub>, although sequence confirmation is required in order to confirm this as other peptides within the PSA sequence, such as T<sub>150</sub> – Y<sub>249</sub>, also correspond to this mass. Interestingly, the 10.7 kDa fragment was not found in urinary PSA<sup>7</sup> and it should be further explored whether this fragment is specific to seminal PSA and whether it is a degradation product following cleavage at R<sub>77</sub>. Importantly, this is the first study that reports on a cleavage at R<sub>77</sub> of PSA in any matrix.

Fragments were also found by middle-up analysis that demonstrate cleavage at R<sub>109</sub>. For example, the fragments found in **Figure 1.B4** correspond to cleavage at this site with loss of R (**Supporting Information, Table S4**). Thus, it may be determined that the mass of 28292.7 Da in peak 2 belongs to the PSA proteoform with a cleavage at R<sub>109</sub>. The mass 28292.7 Da is also observed in urinary PSA<sup>7</sup> and based on the relative migration time (relative to the most abundant peak), likely the R<sub>109</sub> cleavage variant is also observed in urinary PSA. Despite this, the relative migration times reported in **Supporting Information, Table S4** show some variation and due to the close migration times of the 28292.7 Da isomers in seminal PSA, further investigation by a middle-up approach is required for urinary PSA in order to confirm whether this proteoform is due to cleavage at R<sub>109</sub> or R<sub>77</sub>.

The single cleavage PSA variant with no amino acid loss (28448.9 Da) was detected under peak 5 in **Figure 1.A1**. The cleavage site may not be determined in the intact profile due to the absence of any amino acid loss. However, the middle-up fragments **B1** and **B5** illustrate that cleavage occurs at K<sub>169</sub> and this is the only cleavage site whereby no amino acid loss is observed in the middle-up analysis. Thus, the mass 28448.9 Da may be inferred as having a cleavage at the K<sub>169</sub> site. However, no tryptic peptides for this cleavage site could be found by the bottom-up approach as K is also the cleavage site targeted by trypsin. Thus, further investigations should examine this by generating peptides with alternate enzymes such as Arg-C.

Peak 6 (22.7 min) in **Figure 1.A1** shows PSA with a single cleavage and loss of K (28320.8 Da), likely due to cleavage at K<sub>206</sub>. In addition, the mass 28263.7 Da also elutes at 22.7 min and is assigned as a cleavage at K<sub>206</sub> with loss of GK, as it is expected that further loss of a non-charged amino acid at the same cleavage site would not result in any shift to the migration time of this proteoform (**Supporting Information, Table S4**). Furthermore, cleavage at K<sub>206</sub> is also supported by new evidence provided by the middle-up approach. The fragments **B3** (5888.6 Da) and **B6** (22440.0 Da) are shown in **Figure 1**, which correspond to PSA with cleavage at K<sub>206</sub> and loss of K. Furthermore, the tryptic peptide WTGG<sub>205</sub> was found (**Supporting Information, Table S4** and **Figure S3**) which contains the loss of K<sub>206</sub>. Although the expected fragment or peptide (I<sub>25</sub>...G<sub>204</sub>; H5N4F1S2 and WTG<sub>204</sub>, respectively) containing the loss of GK was not observed, **Supporting Information, Table S4** shows that a fragment arising due to double-cleavage (K<sub>170</sub>...G<sub>204</sub>) was observed that

contains the loss of GK<sub>206</sub>. To summarize, **Scheme 1** shows that PSA contains a fascinating diversity of proteoforms, which includes non-cleaved (active) and cleaved (inactive)<sup>30</sup> PSA found in seminal and urinary PSA during this study.



**Scheme 1. Overview of the cleaved proteoforms found in seminal and urinary PSA.** Non-cleaved, active PSA undergoes internal cleavage which inactivates the protein. Cleaved PSA proteoforms arise at one (single cleavage) or two (double cleavage) of four cleavage sites, as well as further truncated variants. Notably, PSA contains five disulfide bonds in total whereas the red dotted lines are shown here to represent how the overall structure of the protein is kept intact following cleavage *via* the disulfide bonds. Ten cleaved proteoforms are observed in total across seminal and urinary PSA, in addition to the glycoforms for each proteoform (**Supporting Information Tables S6** and **S7**). Only the most abundant glycoform, H5N4F1S2, is illustrated here. The legend may be found in the blue box.



## Data Processing of Intact Proteoforms

In the previous study we performed a “manual” approach for quantification using the DataAnalysis software. In this case, the large charge envelope of the intact proteoforms resulted in overlapping  $m/z$  signals and broad XIE peaks. Thus, it was necessary to manually integrate and deconvolute each XIE peak in order to determine peak areas and masses, respectively. In the current study, we present a workflow primarily using Byos in order to perform “software-assisted” intact protein data processing and deconvoluted quantification of peak intensities. In order to perform the comparison, both approaches were used to assess two datasets: an intra- ( $n = 3$ ) and interday ( $n = 9$ ) study which demonstrated the intermediate precision and repeatability of the intact urinary PSA assay using a patient urinary pool, and the measurement of individual patient samples ( $n = 8$ ).<sup>7</sup> Notably, the proteoforms for several intact masses determined in intact urinary PSA are revised in **Supporting Information, Table S4** based on the evidence obtained from the orthogonal analysis of seminal PSA.

In **Table 1**, the annotation of proteoforms in the intra- and interday dataset shows that implementing the software-assisted workflow in the current study resulted in the assignment of 35 proteoforms, including nine unique masses that were not observed when manual processing was performed. However, these unique masses were mainly very low abundant proteoforms with relative abundances  $< 1\%$  (**Supporting Information, Table S6**). In comparison, manual processing determined 32 proteoforms, including six unique masses. Additionally, 26 proteoforms were detected by both methods. The analysis of the individual patient data showed that 23 proteoforms were quantified in total by both data processing techniques. However, four unique proteoforms were determined each by the software-assisted and the manual approach, respectively, and 19 masses were quantified by both approaches.

The software-assisted workflow facilitates automatic proteoform assignment by the implementation of a delta mass list within the processing method (**Supporting Information, Table S3**). For example, the range of possible *N*-glycan structures was previously determined by the bottom-up approach. In addition, cleavage sites and amino acid loss were confirmed *via* intact protein and middle-up analysis in order to provide a library of potential proteoforms to perform a targeted search against as well as automatic annotation of the masses observed in the intact protein profile.

Assignments could then be confirmed using a mass error threshold as well as the expected migration position based on amino acid loss and the number of negatively charged sialic acids present on the proteoform. However, future studies could also focus on confirming glycoform structures directly in the intact protein spectrum by MS/MS experiments, as has previously been shown.<sup>31</sup>

The batch processing of both datasets was enabled by the migration time alignment step which was performed using LaCyTools prior to importing the data into Byos for assignment and quantification. Thus, expected assignments and integration times were verified in a reference file which was then applied to the entire batch of samples. This is illustrated in **Supporting Information, Figure S4.A** whereby the same integration window (22.09 – 22.80 min) was used across each sample to integrate and extract deconvoluted spectra of mono-sialylated species. However, as shown in **Supporting Information, Tables S6 and S7**, some masses were detected previously by the manual approach, and not by software-assisted annotation. The manual approach utilized smaller integration windows in order to reduce noise within the spectra and perform proteoform annotation followed by manual XIE of the assignments. In this study, integration windows that covered the beginning and end of the XIE peak were used in the software-assisted approach. However, this may also result in the integration of more noise which can affect the mass accuracy for the assignment of some low abundant species (< 4%, **Supporting Information, Table S6**). An example is provided in **Supporting Information, Figure S4.B** whereby the application of smaller integration windows resulted in the annotation of H5N4F1S1. Despite this, shorter integration windows may also result in greater variability of the extracted deconvoluted spectrum between measurements. **Supporting Information, Figure S4.A** shows that there are small shifts in the peak apex which will have a greater effect on the deconvolution spectrum when using narrower integration windows. This is further demonstrated in **Supporting Information, Figure S4.C** whereby integration windows based on integrating the full peak or FWHM of the peak were compared, resulting in average relative standard deviations (RSDs) of 15% and 22%, respectively. Overall, the application of the migration time alignment step and full peak integration improved the throughput and reproducibility of both the data processing and data analysis.

Additionally, quantification *via* maximum entropy using the DataAnalysis software was compared with the current workflow which employs parsimonious deconvolution. **Supporting Information, Figure S5** shows that both approaches resulted in similar relative abundances of selected proteoforms and average RSDs of 21% and 14% for maximum entropy and parsimonious deconvolution, respectively. However, it should be noted that shorter integration windows were used for the maximum entropy approach which may also contribute to the higher RSDs, as previously mentioned. Importantly, a migration time alignment step could not be performed prior to applying the maximum entropy approach due to the software accepting only a single datafile type. Furthermore, integration windows were manually entered in order to extract mass spectra to be used for deconvolution. Thus, although maximum entropy and parsimonious deconvolution gave similar quantification results, the implementation of electropherogram alignment in combination with automatic integration windows resulted in faster and more efficient data processing using the current workflow.

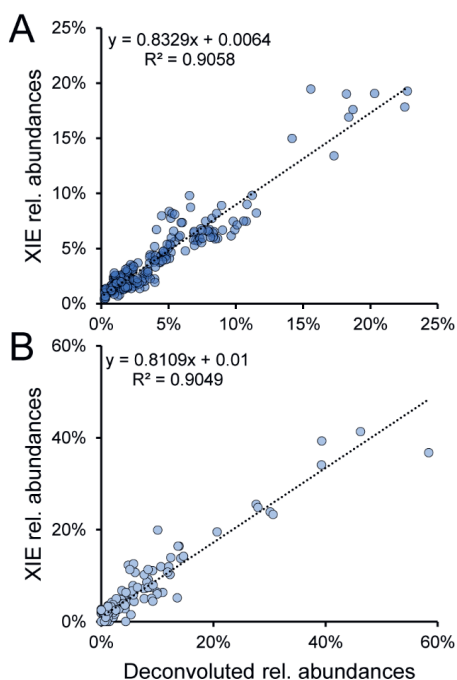
A PSA sequence variant with an additional *N*-glycosylation site was previously reported in urinary PSA from a single patient.<sup>7</sup> In this study, we explored the data by creating a delta mass list of all possible glycoform combinations based on glycans detected by the bottom-up technique in combination with the altered amino acid sequence whereby Asp<sub>102</sub> is replaced by Asn. Similar to the approach mentioned above, this allowed us to perform automatic annotation of an additional 12 unique proteoforms (28 in total; **Supporting Information, Table S8**). However, further validation is required by proteolytic digestion in order to determine the glycans present on the peptide containing the additional *N*-glycosylation site. Additionally, we also observed multiple peaks in the electropherogram of urinary PSA that likely belong to PSA peptides (**Supporting Information, Table S5**) that may be the result of co-capturing degraded PSA, or the degradation of PSA following capturing.

**Table 1. Comparison of mass assignments and quantification between the software-assisted (SA) workflow and the manual approach for the intra- and interday, and patient datasets.** Mass assignments refers to assignment of proteoforms to deconvoluted masses that were found within the mass error threshold ( $\pm 25$  ppm), as well as demonstrating an expected migration time. “Unique” masses are proteoform assignments that were only determined by one data processing method whereas “common” refers to proteoforms determined by both techniques. Intra- and interday quantification results are not applicable (n/a) for the patient study. For a full list of assignments and results, see **Supporting Information, Tables S7 - 8.**

Dataset	Method	Mass Assignments			Quantification (RSD)			
		Total	Unique	Common	Intra-day 1	Intra-day 2	Intra-day 3	Inter-day
Intra and inter-day	SA	35	9	26	13%	9%	12%	19%
	Manual	32	6		12%	8%	12%	15%
Patients	SA	23	4	19	n/a	n/a	n/a	n/a
	Manual	23	4		n/a	n/a	n/a	n/a

### Intact Proteoform Quantification

The RSD of the intra- and interday (**Table 1**) is 12% and 19% by deconvolution as part of the software-assisted workflow, and 11% and 15% by XIE in the manual approach, respectively. The precision determined by both processing methods is within the 20% acceptance criteria applied for other intact protein assays.<sup>32,33</sup> However, it should be noted that this acceptance criteria generally refers to the absolute quantitation of protein concentration rather than relative abundance. Interestingly, re-normalization to the 26 common proteoforms determined by both methods results in intra- and interday RSDs of 9% and 15% (deconvolution), and 12% and 16% (XIE), respectively. Thus, the slightly higher RSDs for the total number of analytes recorded by deconvoluted data processing is likely due to the additional low abundant proteoforms that were detected by this technique. These results are similar to previous studies that determined similar RSDs between both quantification approaches when performing absolute protein quantitation.<sup>11,13,16</sup> However, Lanshoeft *et al.* reported that greater precision was achieved when XIC areas were used rather than XIC or deconvoluted peak intensities.<sup>18</sup> In addition, Kellie *et al.* demonstrated that protein quantitation was more accurate by the XIC approach at the lower limit of quantification (LLOQ).<sup>17</sup>



**Figure 2. Linear regression plot of relative abundances for common proteoforms quantified by deconvolution and XIE methods. (A)** Intra- and interday ( $n = 9$ ) dataset. There are 26 common proteoforms detected by both processing methods. **(B)** Patient ( $n = 8$ ) dataset. There are 19 common proteoforms detected by both processing methods. Relative abundances determined by XIE quantification is represented on the y-axis and relative abundances determined *via* deconvoluted quantification is shown on the x-axis. The equation of the trendline and  $R^2$  are displayed.

The correlation between the two data processing methods is demonstrated in **Figure 2** for the intra- and interday, as well as the patient datasets. This shows the association of the results following re-normalization to the common proteoforms determined by both methods. Interestingly, this results in a  $R^2$  value of 0.91 and 0.90 for the intra- and interday and patient datasets, respectively. This was also investigated when the most abundant proteoform was omitted from the analysis (**Supporting Information, Figure S6**), which resulted in  $R^2$  values of 0.83 (intra- and interday) and 0.78 (patients), demonstrating that both processing techniques result in a sufficiently similar quantification of the data. Notably, quantification based on peak intensity has demonstrated greater performance than peak area for the deconvoluted approach.<sup>19</sup> In the case of quantitative assays using XIEs, peak areas are more commonly reported.<sup>19</sup> Thus, the use of peak intensities (deconvolution) in comparison with peak areas (XIE) may also introduce some discrepancy between the reported abundance values.

Data processing throughput is an important metric to consider when evaluating new tools for performing intact protein analysis of clinical samples. In this study, we demonstrated the processing of two datasets for which the majority of the data processing steps were software-assisted. For example, the mass calibration and time

alignment steps were carried out in half a day whilst automatic data integration and processing were performed in approximately one hour for nine samples. Following this, the most abundant mass in each electrophoretic peak was assessed in order to verify the processing had been performed correctly and, as a result of the complex proteoform profiles, the data was exported into spreadsheet format for further analysis. Thus, the largest hands-on time was due to the data analysis which took approximately one day. Overall, the full data processing and analysis was conducted in two days for each dataset. In contrast, the manual data processing of these datasets was previously performed over several weeks.<sup>7</sup> In this case, due to the broad XIE peaks as previously mentioned, manual peak integration and deconvolution was performed in order to obtain peak areas and masses, respectively. Thus, this took considerable more hands-on time as this was performed for every analyte in each sample. Furthermore, throughput in terms of proteoform assignment and quantification is an important feature when performing biomarker discovery studies and should be considered when validating methods for intact protein data processing. For example, in this study, the relative abundances of 23 heterogeneous proteoforms were quantified in the patient dataset ( $n = 8$ ), resulting in the processing of 127 analytes in total. In comparison, Lanshoeft *et al.* performed absolute protein quantitation of deglycosylated hlgG1A and [13C]-hlgG1A spiked into rat serum ( $n = 24$ ) resulting in the processing of 48 analytes in total.<sup>18</sup> Thus, undoubtedly there are different factors to consider when processing clinical samples or biotherapeutics, such as proteoform complexity, and the number of analytes and samples for analysis. In general, these results are similar to previous studies where it was reported that intact data processing could be streamlined *via* the inclusion of a deconvolution quantification step<sup>19</sup> and, as a result, similar workflows may be applied in the future for the investigation of larger clinical cohorts using intact protein mass spectrometry.

A one-size-fits-all approach has still not yet been defined for intact protein data processing.<sup>1,12,13,17</sup> However, several studies prefer the XIC approach<sup>7,9,16,23</sup> as it remains closer to the raw data and is less prone to the generation of artefacts that may occur due to the inclusion of an extra processing step such as deconvolution.<sup>1,12</sup> Despite this, as previously mentioned, automation of the majority of the data processing method is required in order to enhance throughput and facilitate the analysis of a greater number of samples. Thus, the XIC approach is currently less

amenable to automated processing of intact proteins due to the generation of broad or poorly resolved peaks as a result of overlapping  $m/z$  values in the charge envelopes of different proteoforms. As a result, the extracted ion peak must be manually verified and integrated.<sup>7</sup> In contrast, only a sufficient mass difference is required in the deconvoluted spectrum in order to perform annotation and quantification simultaneously.<sup>19,20,22</sup> For example, Wu *et al.* has recently demonstrated a promising and universally available software that is suitable for identification, deconvoluted quantification, and batch processing of top-down proteomics data.<sup>34</sup> Thus, deconvolution may be a more suitable approach in order to facilitate greater data processing throughput, and the performance of this method should be further verified in comparison with performing XIC quantification.

Some studies have demonstrated efficient data processing of intact proteins using XICs. For example, Kellie *et al.* performed deconvoluted mass assignment and XIC quantification in a semi-automated workflow.<sup>9</sup> However, this approach was applied to proteins up to 10 kDa whereby isotopic resolution was achieved, this resulted in the extraction of specific ion chromatograms. In contrast, such resolution is not commonly observed for proteins that are greater than 25 kDa, resulting in a greater overlap between  $m/z$  signals. This results in the use of the average masses of several charge states, rather than isotopic masses, to generate XICs.<sup>1</sup> Thus, future work should also focus on the development of such tools which are suitable for larger proteins, as well as enabling a facile user interface that would facilitate greater implementation within the field. Additionally, an improvement in the resolution of larger proteins would also facilitate greater selectivity when performing XIE and XIC quantification.

## Perspectives

The application of vendor-specific software impedes harmonization of practices for intact data processing, particularly in the case of deconvolution as these tools differ in important parameters used for the generation a deconvoluted spectrum.<sup>12,13,17</sup> Thus, the development of tools that are capable of handling multiple data formats, such as the data processing steps presented in this study, is an important development towards developing consistent practices in intact protein data processing. In addition, the performance of a  $m/z$ -based migration time alignment, as demonstrated here, may also facilitate automated processing of extracted ion peaks by defining the same

integration window in all samples. As a result, in terms of throughput and accuracy, this would enable a fairer comparison between the XIC and deconvolution approaches to be performed.

## Conclusions

In this study, we have built upon our previous work with new data that further support the assignment of cleaved proteoforms in seminal and urinary PSA, including the finding of a potential new cleavage site in seminal PSA. Undoubtedly, a greater understanding of the proteoform profile of PSA from these biological matrices has been achieved which will inform future studies regarding this protein. In addition, we have demonstrated a software-assisted workflow for the annotation and quantification of intact urinary PSA from a small cohort of patients. Importantly, a migration time alignment pre-processing step was performed which allowed the same integration parameters to be used across all samples and, as a result, fast and efficient quantification via deconvolution was achieved. Moreover, the similarity between our results and the extracted ion quantification method was demonstrated. Overall, this work will support the implementation of intact protein data analysis in the biomarker discovery setting.

3

## Acknowledgements

The authors wish to thank Protein Metrics for providing access to the beta-release version of the Byos software. In addition, the authors are grateful to Wei Wang and Irina Dragan for performing some of the measurements.

## Funding

This research was funded by the European Union's Horizon 2020 Research and Innovation Program (GlySign, Grant No. 722095).

## Conflict Of Interest

The authors do not declare any conflict of interest.



## Data Availability

The raw data from the intra- and interday and patient studies is deposited in the MassIVE repository, and may be located using the dataset identifier MSV000086699.

## Supporting Information

**Supporting Information 1: Figure S1.** Profile of intact seminal PSA analyzed by CE-ESI-MS. **Figure S2.** Effect of histidine loss on migration time. **Figure S3.** MS/MS spectra of diagnostic peptides for cleavage variants in seminal PSA. **Figure S4.** Impact of integration windows on annotation and quantification via deconvolution. **Figure S5.** Maximum entropy *versus* parsimonious deconvolution. **Figure S6.** Linear regression plots excluding the most abundant proteoform.

**Supporting Information 2: Table S1.** Migration time alignment using LaCyTools. **Table S2.** Integration windows for deconvoluted peak generation in Byos. **Table S3.** Delta mass in Byos containing expected glycoforms, and cleavage variants with internal amino acid loss. **Table S4.** Overview of orthogonal datasets supporting proteoform assignments in PSA. **Table S5.** Assignment of peptide signals discovered in intact urinary PSA electropherogram(s). **Table S6.** Comparison of urinary PSA quantification results on intra- and interday dataset via deconvolution and XIE. **Table S7.** Comparison of urinary PSA quantification results on patient dataset via deconvolution and XIE. **Table S8.** Urinary PSA sequence variant [D>N] glycoform assignment.

## References

- (1) van den Broek, I.; van Dongen, W. D. LC-MS-Based Quantification of Intact Proteins: Perspective for Clinical and Bioanalytical Applications. *Bioanalysis*. **2015**, *7* (15), 1943–1958. DOI: 10.4155/bio.15.113.
- (2) Yang, Y.; Franc, V.; Heck, A. J. R. Glycoproteomics: A Balance between High-Throughput and In-Depth Analysis. *Trends in Biotechnology*. **2017**, *35* (7), 598–609. DOI: 10.1016/j.tibtech.2017.04.010.
- (3) Ruan, Q.; Ji, Q. C.; Arnold, M. E.; Humphreys, W. G.; Zhu, M. Strategy and Its Implications of Protein Bioanalysis Utilizing High-Resolution Mass Spectrometric Detection of Intact Protein. *Analytical Chemistry*. **2011**, *83* (23), 8937–8944. DOI: 10.1021/ac201540t.

- (4) Ramagiri, S.; Garofolo, F. Large Molecule Bioanalysis Using Q-TOF without Predigestion and Its Data Processing Challenges. *Bioanalysis*. **2012**, *4* (5), 529–540. DOI: 10.4155/bio.12.10.
- (5) Sénard, T.; Gargano, A. F. G.; Falck, D.; de Taeye, S. W.; Rispens, T.; Vidarsson, G.; Wuhrer, M.; Somsen, G. W.; Domínguez-Vega, E. MS-Based Allotype-Specific Analysis of Polyclonal IgG-Fc N-Glycosylation. *Frontiers in Immunology*. **2020**, *11*, 2049. DOI: 10.3389/fimmu.2020.02049.
- (6) Domínguez-Vega, E.; Tengattini, S.; Peintner, C.; van Angeren, J.; Temporini, C.; Haselberg, R.; Massolini, G.; Somsen, G. W. High-Resolution Glycoform Profiling of Intact Therapeutic Proteins by Hydrophilic Interaction Chromatography-Mass Spectrometry. *Talanta*. **2018**, *184*. DOI: 10.1016/j.talanta.2018.03.015.
- (7) Moran, A. B.; Domínguez-Vega, E.; Nouta, J.; Pongracz, T.; de Reijke, T. M.; Wuhrer, M.; Lageveen-Kammeijer, G. S. M. Profiling the Proteoforms of Urinary Prostate-Specific Antigen by Capillary Electrophoresis – Mass Spectrometry. *Journal of Proteomics*. **2021**, *238*. DOI: 10.1016/j.jprot.2021.104148.
- (8) Gstöttner, C.; Nicolardi, S.; Habegger, M.; Reusch, D.; Wuhrer, M.; Domínguez-Vega, E. Intact and Subunit-Specific Analysis of Bispecific Antibodies by Sheathless CE-MS. *Analytica Chimica Acta*. **2020**, *1134*. DOI: 10.1016/j.aca.2020.07.069.
- (9) Kellie, J. F.; Higgs, R. E.; Ryder, J. W.; Major, A.; Beach, T. G.; Adler, C. H.; Merchant, K.; Knierman, M. D. Quantitative Measurement of Intact Alpha-Synuclein Proteoforms from Post-Mortem Control and Parkinson's Disease Brain Tissue by Intact Protein Mass Spectrometry. *Scientific Reports*. **2014**, *4* (1), 1–10. DOI: 10.1038/srep05797.
- (10) Yang, Y.; Liu, F.; Franc, V.; Halim, L. A.; Schellekens, H.; Heck, A. J. R. Hybrid Mass Spectrometry Approaches in Glycoprotein Analysis and Their Usage in Scoring Biosimilarity. *Nature Communications*. **2016**, *7*. DOI: 10.1038/ncomms13397.
- (11) Kang, L.; Weng, N.; Jian, W. LC–MS Bioanalysis of Intact Proteins and Peptides. *Biomedical Chromatography*. **2020**, *34* (1). DOI: 10.1002/bmc.4633.
- (12) Kellie, J. F.; Tran, J. C.; Jian, W.; Jones, B.; Mehl, J. T.; Ge, Y.; Henion, J.; Bateman, K. P. Intact Protein Mass Spectrometry for Therapeutic Protein Quantitation, Pharmacokinetics, and Biotransformation in Preclinical and Clinical Studies: An Industry Perspective. *Journal of the American Society for Mass Spectrometry*. **2021**, *32* (8). DOI: 10.1021/jasms.0c00270.
- (13) Qiu, X.; Kang, L.; Case, M.; Weng, N.; Jian, W. Quantitation of Intact Monoclonal Antibody in Biological Samples: Comparison of Different Data Processing Strategies. *Bioanalysis*. **2018**, *10* (13). DOI: 10.4155/bio-2018-0016.
- (14) Berger, A. L.; della Pietra, V. J.; della Pietra, S. A. A Maximum Entropy Approach to Natural Language Processing. *Computational Linguistics*. **1996**, *22* (1), 39–68.
- (15) Bern, M.; Caval, T.; Kil, Y. J.; Tang, W.; Becker, C.; Carlson, E.; Kletter, D.; Sen, K. I.; Galy, N.; Hagemans, D.; Franc, V.; Heck, A. J. R. Parsimonious Charge Deconvolution for Native Mass Spectrometry. *Journal of Proteome Research*. **2018**, *17* (3). DOI: 10.1021/acs.jproteome.7b00839.
- (16) Zhang, L.; Vasicek, L. A.; Hsieh, S.; Zhang, S.; Bateman, K. P.; Henion, J. Top-down LC-MS Quantitation of Intact Denatured and Native Monoclonal Antibodies in Biological Samples. *Bioanalysis*. **2018**, *10* (13). DOI: 10.4155/bio-2017-0282.

- (17) Kellie, J. F.; Kehler, J. R.; Karlinsey, M. Z.; Summerfield, S. G. Toward Best Practices in Data Processing and Analysis for Intact Biotherapeutics by MS in Quantitative Bioanalysis. *Bioanalysis*. **2017**, *9* (23). DOI: 10.4155/bio-2017-0179.
- (18) Lanshoeft, C.; Cianférani, S.; Heudi, O. Generic Hybrid Ligand Binding Assay Liquid Chromatography High-Resolution Mass Spectrometry-Based Workflow for Multiplexed Human Immunoglobulin G1 Quantification at the Intact Protein Level: Application to Preclinical Pharmacokinetic Studies. *Analytical Chemistry*. **2017**, *89* (4). DOI: 10.1021/acs.analchem.6b04997.
- (19) Jian, W.; Kang, L.; Burton, L.; Weng, N. A Workflow for Absolute Quantitation of Large Therapeutic Proteins in Biological Samples at Intact Level Using LC-HRMS. *Bioanalysis*. **2016**, *8* (16). DOI: 10.4155/bio-2016-0096.
- (20) Xu, K.; Liu, L.; Dere, R.; Mai, E.; Erickson, R.; Hendricks, A.; Lin, K.; Junutula, J. R.; Kaur, S. Characterization of the Drug-to-Antibody Ratio Distribution for Antibody-Drug Conjugates in Plasma/Serum. *Bioanalysis*. **2013**, *5* (9). DOI: 10.4155/bio.13.66.
- (21) Xu, K.; Liu, L.; Saad, O. M.; Baudys, J.; Williams, L.; Leipold, D.; Shen, B.; Raab, H.; Junutula, J. R.; Kim, A.; Kaur, S. Characterization of Intact Antibody-Drug Conjugates from Plasma/Serum in Vivo by Affinity Capture Capillary Liquid Chromatography-Mass Spectrometry. *Analytical Biochemistry*. **2011**, *412* (1). DOI: 10.1016/j.ab.2011.01.004.
- (22) Jin, W.; Burton, L.; Moore, I. LC-HRMS Quantitation of Intact Antibody Drug Conjugate Trastuzumab Emtansine from Rat Plasma. *Bioanalysis*. **2018**, *10* (11). DOI: 10.4155/bio-2018-0003.
- (23) Kellie, J. F.; Kehler, J. R.; Mencken, T. J.; Snell, R. J.; Hottenstein, C. S. A Whole-Molecule Immunocapture LC-MS Approach for the in Vivo Quantitation of Biotherapeutics. *Bioanalysis*. **2016**, *8* (20). DOI: 10.4155/bio-2016-0180.
- (24) Kammeijer, G. S. M.; Nouta, J.; de La Rosette, J. J. M. C. H.; de Reijke, T. M.; Wuhrer, M. An In-Depth Glycosylation Assay for Urinary Prostate-Specific Antigen. *Analytical Chemistry*. **2018**, *90* (7), 4414–4421. DOI: 10.1021/acs.analchem.7b04281.
- (25) Santos, M. R.; Ratnayake, C. K.; Fonslow, B.; Guttman, A. A covalent , cationic polymer coating method for the CESI-MS analysis of intact proteins and polypeptides [www.sciex.com/content/dam/SCIEX/pdf/tech-notes/all/Covalent-cationic-polymer-coating-method-CESI-MS.pdf](http://www.sciex.com/content/dam/SCIEX/pdf/tech-notes/all/Covalent-cationic-polymer-coating-method-CESI-MS.pdf) (accessed 2022 -10 -29).
- (26) Kammeijer, G. S. M.; Kohler, I.; Jansen, B. C.; Hensbergen, P. J.; Mayboroda, O. A.; Falck, D.; Wuhrer, M. Dopant Enriched Nitrogen Gas Combined with Sheathless Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry for Improved Sensitivity and Repeatability in Glycopeptide Analysis. *Analytical Chemistry*. **2016**, *88* (11). DOI: 10.1021/acs.analchem.6b00479.
- (27) Jansen, B. C.; Falck, D.; de Haan, N.; Hipgrave Ederveen, A. L.; Razdorov, G.; Lauc, G.; Wuhrer, M. LaCyTools: A Targeted Liquid Chromatography-Mass Spectrometry Data Processing Package for Relative Quantitation of Glycopeptides. *Journal of Proteome Research*. **2016**, *15* (7). DOI: 10.1021/acs.jproteome.6b00171.
- (28) Mikolajczyk, S. D.; Millar, L. S.; Wang, T. J.; Rittenhouse, H. G.; Wolfert, R. L.; Marks, L. S.; Song, W.; Wheeler, T. M.; Slawin, K. M. "BPSA," a Specific Molecular Form of Free Prostate-Specific Antigen, Is Found Predominantly in the Transition Zone of Patients with

- Nodular Benign Prostatic Hyperplasia. *Urology*. **2000**, *55* (1). DOI: 10.1016/S0090-4295(99)00372-6.
- (29) Linton, H. J.; Marks, L. S.; Millar, L. S.; Knott, C. L.; Rittenhouse, H. G.; Mikolajczyk, S. D. Benign Prostate-Specific Antigen (BPSA) in Serum Is Increased in Benign Prostate Disease. *Clinical Chemistry*. **2003**, *49* (2), 253–259. DOI: 10.1373/49.2.253.
- (30) Mattsson, J. M.; Valmu, L.; Laakkonen, P.; Stenman, U. H.; Koistinen, H. Structural Characterization and Anti-Angiogenic Properties of Prostate-Specific Antigen Isoforms in Seminal Fluid. *Prostate*. **2008**, *68* (9), 945–954. DOI: 10.1002/pros.20751.
- (31) Behnken, H. N.; Ruthenbeck, A.; Schulz, J. M.; Meyer, B. Glycan Analysis of Prostate Specific Antigen (PSA) Directly from the Intact Glycoprotein by HR-ESI/TOF-MS. *Journal of Proteome Research*. **2014**, *13* (2). DOI: 10.1021/pr400999y.
- (32) Duggan, J. X.; Vazvaei, F.; Jenkins, R. Bioanalytical Method Validation Considerations for LC-MS/MS Assays of Therapeutic Proteins. *Bioanalysis*. **2015**, *7* (11), 1389–1395. DOI: 10.4155/bio.15.69.
- (33) Knutsson, M.; Schmidt, R.; Timmerman, P. LC-MS/MS of Large Molecules in a Regulated Bioanalytical Environment - Which Acceptance Criteria to Apply? *Bioanalysis*. **2013**, *5* (18), 2211–2214. DOI: 10.4155/bio.13.193.
- (34) Wu, Z.; Roberts, D. S.; Melby, J. A.; Wenger, K.; Wetzel, M.; Gu, Y.; Ramanathan, S. G.; Bayne, E. F.; Liu, X.; Sun, R.; Ong, I. M.; McIlwain, S. J.; Ge, Y. MASH Explorer: A Universal Software Environment for Top-Down Proteomics. *Journal of Proteome Research*. **2020**, *19* (9). DOI: 10.1021/acs.jproteome.0c00469.



# Chapter 4

## Sialic Acid Derivatization of Fluorescently Labeled *N*-Glycans Allows Linkage Differentiation by RPLC-FD-MS

Alan B. Moran<sup>1,2\*</sup>, Richard A. Gardner<sup>2</sup>, Manfred Wuhrer<sup>1</sup>, Guinevere S.M. Lageveen-Kammeijer<sup>1</sup>, Daniel I.R. Spencer<sup>2</sup>

<sup>1</sup> Leiden University Medical Center, Center for Proteomics and Metabolomics, 2300 RC Leiden, The Netherlands

<sup>2</sup> Ludger Ltd., Culham Science Centre, OX14 3EB Abingdon, United Kingdom

*Reprinted (adapted) with permission from Moran, A. B., Gardner, R. A., Wuhrer, M., Lageveen-Kammeijer, G. S. M., & Spencer, D. I. R. (2022). Sialic Acid Derivatization of Fluorescently Labeled N-Glycans Allows Linkage Differentiation by Reversed-Phase Liquid Chromatography–Fluorescence Detection–Mass Spectrometry. Analytical Chemistry, 94(18), 6639–6648. <https://doi.org/10.1021/acs.analchem.1c02610>. Copyright 2023 American Chemical Society.*



## Abstract

Sialic acids have diverse biological roles, ranging from promoting up to preventing protein and cellular recognition in health and disease. The various functions of these monosaccharides are owed, in part, to linkage variants and, as a result, linkage-specific analysis of sialic acids is an important aspect of glycomic studies. This has been addressed by derivatization strategies using matrix-assisted laser desorption/ionization mass spectrometry (MS), or sialidase digestion arrays followed by liquid chromatography (LC)-MS. Despite this, these approaches are unable to simultaneously provide unambiguous assignment of sialic acid linkages and assess further isomeric glycan features within a single measurement. Thus, for the first time, we present the combination of procainamide fluorescent labeling with sialic acid linkage-specific derivatization via ethyl esterification and amidation for the analysis of released plasma *N*-glycans using reversed phase (RP)LC-fluorescence detection (FD)-MS. As a result,  $\alpha$ 2,3- and  $\alpha$ 2,6- sialylated *N*-glycans - with the same mass prior to derivatization - are differentiated based on retention time, precursor mass and fragmentation spectra and additional sialylated isomers were also separated. Furthermore, improved glycan coverage and protocol precision were found *via* the novel application using a combined FD-MS quantification approach. Overall, this platform achieved unambiguous assignment of *N*-glycan sialic acid linkages within a single RPLC-FD-MS measurement and, by improving their retention on RPLC, this technique can be used for future investigations of released *N*-glycans as an additional or orthogonal method to current analytical approaches.

## Introduction

Sialic acids are important monosaccharides that play a role in a wide range of biological processes.<sup>1</sup> Often found as the terminal residue on *N*- and *O*-linked glycans as well as glycolipids,<sup>2</sup> sialic acids act as mediators during biological recognition.<sup>1</sup> This includes processes such as protein and cell binding as well as host-pathogen interactions.<sup>3</sup> Another important role for sialic acids is their masking effect.<sup>4</sup> For example,  $\alpha$ 2,3-linked sialic acids allow the underlying galactose to be accessed by specific lectins, whereas  $\alpha$ 2,6-linked sialic acids may act as inhibitors of such processes.<sup>5</sup> These are important functions during numerous healthy and disease states, including cancer metastasis and tumor cell survival.<sup>2</sup>

Due to their critical role in biology, linkage-specific analysis of sialic acids is an important facet of glycomic investigations. However, the labile nature and negative charge of these monosaccharides presents several challenges for mass spectrometry (MS)-based analyses.<sup>6</sup> Matrix-assisted laser desorption/ionization mass spectrometry (MALDI)-MS approaches have largely overcome these issues by employing linkage-specific sialic acid derivatization,<sup>7,8</sup> which has the effect of stabilizing sialic acids and neutralizing their charge to ensure a more homogenous ionization.<sup>8</sup> Despite this, MALDI-MS methods generally lack an online separation component which is useful to assess further structural aspects that may differentiate glycan isomers. For this purpose, liquid chromatography (LC)-MS techniques, such as porous graphitized carbon (PGC)-LC, reversed phase (RP)-LC, and hydrophilic interaction-LC (HILIC), are often more suitable approaches. PGC-LC is a powerful approach for in-depth structural differentiation of glycans,<sup>9</sup> however, it is only applied in a few laboratories due to its complexity.<sup>9,10</sup> Although RPLC is a more widely used technique, it is often insufficient to separate several glycan species, particularly sialylated *N*-glycans, as the separation of *N*-glycans is largely influenced by the reducing-end label.<sup>11</sup> As a result, HILIC is most often the LC method of choice for released *N*-glycan analysis.<sup>12,13</sup> Despite this, unequivocal sialic acid linkage assignment may not be achieved within a single measurement and further analysis using sialidase enzymes is required.<sup>14</sup>

Unambiguous linkage assignment and online separation were achieved when sialic acid derivatization was combined with reducing-end fluorescent labeling using 2-



aminobenzamide (2-AB).<sup>15,16</sup> This allowed measurement by HILIC-MS while also producing linkage-specific mass shifts and fragmentation spectra. However, the DMT-MM (4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methyl morpholinium chloride) derivatization procedure involves harsher reaction conditions in comparison with more recently developed protocols.<sup>6,17</sup> In addition, the nature of such chemical modifications to sialic acids increases their hydrophobicity and reduces their retention time when analyzed using HILIC.<sup>15</sup> This is problematic for the analysis of complex samples, particularly when using fluorescence detection, as a diverse range of glycan species elute along the entire profile.<sup>18</sup> In this regard, it may be more suitable to analyze glycans with enhanced hydrophobicity by RPLC. This was previously demonstrated on RPLC-MS using 2-aminopyridine (2-PA) or Girard's P reagent labeled glycans, in combination with sialic acid linkage-specific derivatization via two-step alkylamidation or deuterated aniline amidation, respectively.<sup>19,20</sup> Interestingly, both of these studies employed a charge-based fractionation step prior to sialic acid derivatization that allowed in-depth structural characterization to be performed specifically on sialylated fractions. Despite this, the fractionation employed in these studies results in the loss of information regarding non-sialylated species, and may hinder the analysis of large numbers of samples as well as result in sample loss.

This study aimed to develop and validate a platform for sialic acid-linkage specific differentiation of fluorescently-labeled released *N*-glycans from a complex sample. As a result, ethyl esterification and amidation (EEA) was performed on procainamide-labeled *N*-glycans from plasma, which allowed the *N*-glycans to be effectively analyzed using RPLC-FD-MS. In addition, the method was developed on a robot which allowed it to be qualified using a large number of replicates ( $n = 50$ ). Finally, this research also sought to explore the complementarity of the newly developed EEA and RPLC-FD-MS platform with the current gold-standard method for released *N*-glycan analysis, HILIC-FD-MS.

## Experimental Section

### Materials

Lyophilized human plasma [P9523] (5 mL), formic acid (FA), methanol, hydrate 1-hydroxybenzotriazole (HOBt), and ammonia (28% NH<sub>3</sub>) were purchased from Sigma-Aldrich (Dorset, UK). Acetonitrile (CH<sub>3</sub>CN; Romil, 190 SPS for UV/gradient quality) and ethanol (EtOH) were acquired from Charlton Scientific (Charlton, UK). De-ionized water (H<sub>2</sub>O) was obtained using a Sartorius arium comfort (Goettingen, Germany) with 18.2 MΩ resistivity and 1-ethyl-3-(3-(dimethylamino) propyl) carbodiimide (EDC) was purchased from Fluorochem (Hadfield, UK). PNGase F storage buffer, composed of 50 mM sodium chloride (NaCl), 5 mM ethylenediaminetetraacetic acid (EDTA), and 20 mM tris-hydrochloric acid (Tris-HCl, pH 7.5), was purchased from New England Biolabs (Hitchin, UK). *N*-glycan A2G2S2 standard [CN-A2-20U], the PNGase F *N*-Glycan release kit [LZ-rPNGASEF-96], Protein Binding Membrane (PBM) plate [LC-PBM-96], 2 mL 96-well collection plate [LP-COLLPLATE-2ML-96], procainamide labeling kit [LT-KPROC-96], HILIC clean-up plate [LC-PROC-96] and ammonium formate solution [LS-N-BUFFX40] were purchased from Ludger Ltd. (Abingdon, UK). The 120 μL skirted 96-well PCR plate [4ti-0960/C], 300 μL non-skirted 96-well PCR plate [4ti-0710/C], 1.2 mL 96 well deepwell plate, foil pierce seal [4ti-0531], and peel seal [4ti-0521] were purchased from 4titude Ltd, (Surrey, UK). HPLC vials [186002639] were purchased from Waters Ltd., (Borehamwood, UK). The human milk oligosaccharide standards, sialyllacto-*N*-tetraose c (LST-C) [SLN506] and sialyllacto-*N*-tetraose a (LST-A) [SLN503], were purchased from Dextra (Reading, UK).

### *N*-glycan Analysis

Commercial lyophilized human plasma was reconstituted in 5 mL H<sub>2</sub>O, at a final concentration of 1 mg/mL. Preparation of plasma *N*-glycans was carried out in line with previously published procedures using a Hamilton Microlab STARlet liquid-handling robot.<sup>21</sup> The experimental procedures for performing PNGase F *N*-glycan release and procainamide labeling are included in **Supporting Information 1, Sections S1.1 and S1.2.**

### Sialic Acid Ethyl Esterification And Amidation

The lyophilized released and procainamide-labeled *N*-glycan samples were reconstituted in 15  $\mu\text{L}$   $\text{H}_2\text{O}$ . The ethyl esterification and amidation (EEA) protocol was performed as previously described<sup>17</sup> and was automated on the Hamilton Starlet. Briefly, the ethyl esterification reagent was prepared (250 mM EDC and 250 mM HOBt dissolved in EtOH) and 60  $\mu\text{L}$  was added per well in a 300  $\mu\text{L}$  96-well PCR plate. Following this, 3  $\mu\text{L}$  of the concentrated procainamide-labeled *N*-glycans was added to the reagent, then the plate was sealed with a foil pierce seal and incubated for 60 min at 37 °C. Following this, 12  $\mu\text{L}$  of 28%  $\text{NH}_3$  was added to the samples before the plate was re-sealed and incubated for another 60 min at 37 °C. A volume of 225  $\mu\text{L}$   $\text{CH}_3\text{CN}$  was added to the plate bringing the final volume in each sample well up to 300  $\mu\text{L}$ . A HILIC clean-up plate was placed on the vacuum manifold and prepared with successive washes of 200  $\mu\text{L}$  of 70% EtOH/  $\text{H}_2\text{O}$  (v/v), 200  $\mu\text{L}$  of  $\text{H}_2\text{O}$  and 200  $\mu\text{L}$  of  $\text{CH}_3\text{CN}$ . Then, 100  $\mu\text{L}$   $\text{CH}_3\text{CN}$  was added to each well of the clean-up plate followed by 100  $\mu\text{L}$  of the derivatized and labeled sample. The samples were eluted under gravity for 5 min before a vacuum was applied. This step was repeated two more times until the entire 300  $\mu\text{L}$  of the derivatized and labeled sample was transferred to the clean-up plate. The plate was blotted briefly onto a paper towel in order to remove excess  $\text{CH}_3\text{CN}$  before being placed back on the vacuum manifold. Following this, a 96-well 2 mL collection plate was placed inside the vacuum manifold and 100  $\mu\text{L}$   $\text{H}_2\text{O}$  was added to the samples. To start the sample elution a vacuum was used for about 5 sec, followed by further elution under gravity. After 15 min, a vacuum was applied to elute the entire sample into the collection plate. This step was repeated in order to elute the samples in a final volume of 200  $\mu\text{L}$ . The remaining concentrated sample (12  $\mu\text{L}$ ) and the derivatized procainamide-labeled samples were stored at -20°C until further analysis.

## RPLC-FD-MS

Samples for RPLC-FD-MS were prepared by adding 95  $\mu\text{L}$  of the derivatized procainamide-labelled *N*-glycans and 5  $\mu\text{L}$   $\text{CH}_3\text{CN}$  (5%) to a 1.2 mL deepwell plate and injecting 20  $\mu\text{L}$  onto an Ultimate 3000 UHPLC system (Thermo Scientific, Hampshire, UK). An ACE excel 2 C18-PFP, 150 x 2.1 mm column (ACE Ltd., Aberdeen, UK) was used and the column temperature was set to 60 °C. The fluorescence detector ( $\lambda_{\text{ex}} = 310 \text{ nm}$   $\lambda_{\text{em}} = 370 \text{ nm}$ ) sensitivity was set to 8 and bulb power was set to 'high'. A separation gradient was employed using solvent A (50 mM ammonium formate) and solvent B (10%  $\text{CH}_3\text{CN}$ ; 0.1% FA (v/v)): 0 to 26.5 min, 15 to 95% solvent B; 26.5 to 30.5 min, 95% B; 30.5 to 32.5 min, 95 to 15% B; 32.5 to 35.1 min, 15% B, all gradient steps were performed with a flow rate of 0.4 mL/min.

MS analysis was performed via coupling the UHPLC to an amaZon Speed ETD MS (Bruker Daltonics GmbH, Bremen, Germany) using electrospray ionization (ESI). The instrument was operated in positive ionization mode with enhanced resolution scanning. The mass range ( $m/z$  600 – 1600) was scanned with a target mass set to  $m/z$  900. In addition, the following parameters were employed: source temperature 250 °C; gas flow 10 L/min; capillary voltage 4500 V; ion charge control (ICC) target 200,000; max accumulation time 50.00 ms. Furthermore, MS/MS spectra were obtained via collision-induced dissociation according to these conditions: ICC target MS(n) 200,000; max accumulation time MS(n) 40.00 ms; number of precursors ions selected 3; release after 0.2 min; MS(n) scan range selection, scale to precursor; absolute and relative signal threshold for automatic MS(n), 25,000 and 5%, respectively. Information regarding HILIC-FD-MS measurements and the intermediate precision and repeatability study may be found in **Supporting Information 1, Sections S1.3** and **S1.4**, respectively.

## *N*-glycan Assignments

RPLC peaks were screened using the Bruker DataAnalysis software (version 5.0) and structures were manually assigned. Assignments of structures were made based on their exact mass, fragmentation pattern, and retention order. Important diagnostic ions for assignment are included by **Supporting Information 2, Table S1**, as well as the analysis and scan number of MS/MS spectra. Furthermore, a comprehensive overview of *N*-glycans reported across several studies in plasma has previously been published

by Lageveen-Kammeijer *et al.*<sup>17</sup> Thus, structures that were assigned in our study were compared against the plasma *N*-glycans present in this overview. HILIC peaks were assigned in conjunction with elution positions reported in the literature<sup>13</sup> as well as by searching the GlycomeDB database using Bruker Proteinscape (version 4.0). In this instance, the database search was narrowed by employing the following parameters: assignment score ( $\geq 30$ ), *N*-glycan fragment coverage ( $\geq 20\%$ ), CID classification depth ( $\geq 3$ ), and accurate mass ( $\pm 100$  ppm). *N*-glycan compositions are illustrated according to the Consortium for Functional Glycomics (CFG) notation:<sup>22</sup> *N*-acetylglucosamine (N; blue square), fucose (F; red triangle), galactose (H; yellow circle), mannose (H; green circle), *N*-acetylneuraminic acid (S; purple diamond).

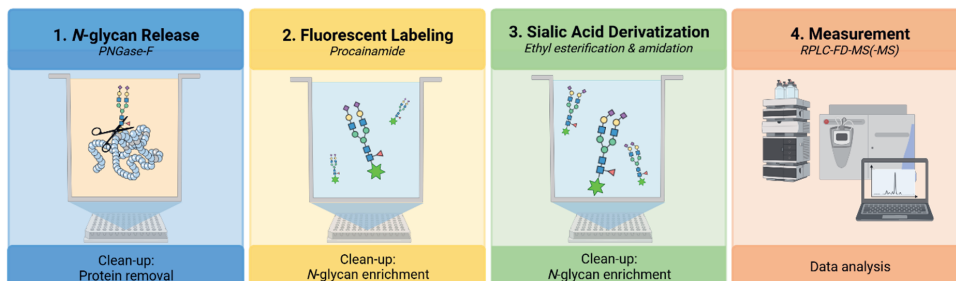
### Fluorescence Detection-Mass Spectrometry (FD-MS) Quantification

After FD and MS curation, as described in **Supporting Information 1, Sections S1.5** and **S1.6**, respectively, only the fluorescent peaks that contained at least one *N*-glycan that passed, were considered for further analysis. This was followed by determining the proportion of *N*-glycan compositions in each fluorescent peak by calculating the local relative abundance of compositions eluting under the same chromatographic peak using MS signal intensities. Following this, the FD-MS signal was derived by multiplying the proportion of each *N*-glycan composition by the fluorescent signal of the peak in which it is eluting.

## Results & Discussion

For the first time, we present the combination of procainamide fluorescent labeling with a sialic acid linkage-specific derivatization step *via* EEA for the analysis of *N*-glycans by RPLC-FD-MS (**Figure 1**). Procainamide was selected as the fluorescent label of choice because it is a well-established amination reagent, and EEA as the derivatization strategy as it has been widely applied, well-developed,<sup>17,23</sup> may be performed under relatively mild conditions, and promotes the formation of stable sialic acid derivatives during the reaction.<sup>6,23</sup> Several parameters were investigated in order to develop the protocol, which are summarized in **Supporting Information 1, Section S2.1** and **Supporting Information 2, Table S2**. In addition, the intermediate precision and repeatability of the complete sample preparation protocol was determined and the separation of the EEA-derivatized *N*-glycans on RPLC was evaluated, as well as

several quantification approaches. Finally, the complementarity of RPLC- and HILIC-FD-MS platforms was also assessed.



**Figure 1. Semi-automated workflow for the *N*-glycan release, procainamide labeling, sialic acid derivatization and RPLC-FD-MS measurement.** The workflow was completed on a Hamilton Microlab STARlet liquid-handling robot allowing 96 samples to be processed simultaneously. *N*-glycans were released from plasma proteins during an overnight digestion (37 °C). Following protein removal, *N*-glycans were fluorescently labeled using procainamide and enriched via a HILIC clean-up plate. Sialic acids were derivatized by ethyl esterification and amidation and enrichment was repeated using the HILIC clean-up plate. Fluorescently labeled and derivatized sialic acid *N*-glycans were measured using RPLC-FD-MS.

### Sialic Acid Differentiation by RPLC-FD-MS

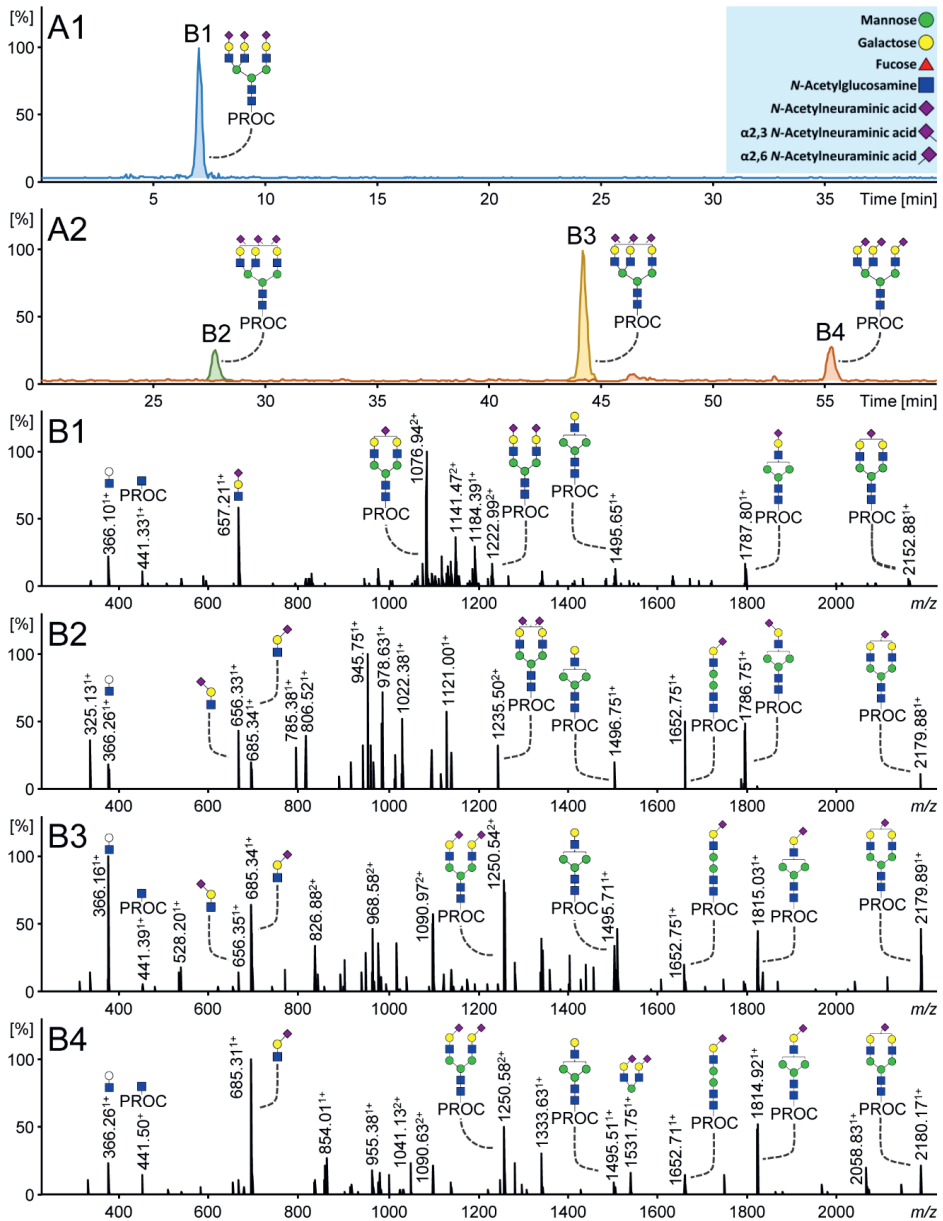
Procainamide-labeled sialylated *N*-glycans showed short elution times on RPLC and no separation of sialic acid linkage isomers. An example is provided in **Figure 2.A1**, where the *N*-glycan, H6N5S3, elutes at seven min as a single peak. Based on the RPLC separation, there is no evidence to suggest that this analyte consists of multiple linkage species. In general, the RPLC profile of procainamide-labeled *N*-glycans (**Supporting Information 1, Figure S1.A1**) showed an elution pattern similar to 2-AB and 2-AA labeled *N*-glycans whereby sialylated structures are poorly separated.<sup>24</sup> In comparison, labels such as 2-PA<sup>25</sup> or Rapifluor-MS<sup>26</sup> generally demonstrated an enhanced separation with sialylated species eluting according to an increasing number of sialic acid residues. Interestingly, isomer separation of procainamide labeled glycans was detected using RPLC for glycans with incomplete antenna sialylation, as shown by H5N4S1 in **Supporting Information 1, Figure S1.A2**. This is consistent with findings obtained with the aforementioned fluorescent labels.<sup>24–26</sup> With regard to this, studies have described the different contribution of antennae to the retention time on RPLC.<sup>25,27</sup> Thus, likely positional isomers are separated based upon which arm is occupied ( $\alpha 3$  versus  $\alpha 6$ ). Overall, procainamide has demonstrated enhanced fluorescence in comparison with 2-AB<sup>28</sup>, 2-AA and 2-PA, as well as a

greater positive ionization (MALDI and ESI) response than 2-AB<sup>28</sup> and 2-AA.<sup>29</sup> In addition, although RapiFluor-MS allows faster sample preparation and has shown increased ionization efficiency,<sup>26</sup> procainamide offers the advantage of being a widely available chemical that may be purchased in bulk or as part of specific glycan labeling kits.

In order to enable the separation of fluorescently-labeled sialylated *N*-glycans on RPLC, EEA was employed prior to measurement in order to enhance glycan hydrophobicity. The derivatization of sialic acids (**Figure 2.A2**) not only improves their retention, but also allows the resolution of three distinct sialic acid-linkage isomers, H6N5S<sub>2,3</sub>2S<sub>2,6</sub>1, H6N5S<sub>2,3</sub>1S<sub>2,6</sub>2, and H6N5S<sub>2,6</sub>3, due to the different chemical derivatization of differently linked sialic acids. Notably, **Figure 2** also shows that the position of the sialic acid on the galactose remains ambiguous and further topological isomers may exist for the tri-sialylated species with mixed linkages at **Figure 2.B2** and **B3**. Importantly, ethyl esterified  $\alpha$ 2,6-sialylated species showed greater retention and were separated from their amidated  $\alpha$ 2,3-sialic acid counterparts in order of increasing  $\alpha$ 2,6-sialic acid content. Additionally, EEA differently modifies the mass of sialylated *N*-glycans depending on the linkage and number of sialic acids present;  $\alpha$ 2,6-linked sialic acids gain 28.02 Da and  $\alpha$ 2,3-linked sialic acids lose 0.98 Da.<sup>17</sup> This is illustrated in **Figure 2.A1** whereby  $m/z$  1033.73 is used to generate an extracted ion chromatogram (EIC) for H6N5S3 [M+2H]<sup>2+</sup>, whereas three distinct precursor  $m/z$  are used for each isomer in order to generate EICs ([M+2H]<sup>2+</sup>) in **Figure 2.A2**;  $m/z$  1042.42 (H6N5S<sub>2,3</sub>2S<sub>2,6</sub>1), 1052.09 (H6N5S<sub>2,3</sub>1S<sub>2,6</sub>2), and 1061.77 (H6N5S<sub>2,6</sub>3). Previously, similar improvements to RPLC separation and sialic acid linkage differentiation have also been shown using other labeling and derivatization techniques.<sup>19,20</sup> Overall, these results suggest that the performance of fluorescently labeled glycans on RPLC may be generally enhanced by including a sialic acid derivatization step as the combination of these techniques is required for efficient separation and differentiation on RPLC.

The linkage assignment of derivatized sialylated *N*-glycans can be further supported by diagnostic ions in the MS/MS spectra. In contrast with **Figure 2.B1** whereby  $m/z$  657.21<sup>1+</sup> indicates a sialylated antenna with unspecified linkage, **Figures 2.B2 – B4** show informative B-ions with theoretical  $m/z$  656.25<sup>1+</sup> and 685.26<sup>1+</sup>, which indicate

Sialic Acid Derivatization of Fluorescently Labeled *N*-Glycans Allows Linkage Differentiation by RPLC-FD-MS



**Figure 2. Procainamide labeled trisialylated plasma *N*-glycans measured by RPLC-FD-MS. (A1)** H6N5S3 ( $m/z$  1033.73) with procainamide labeling and without sialic acid linkage-specific derivatization. **(A2)** H6N5S3 is separated into three distinct isomers following procainamide labeling and sialic acid linkage-specific derivatization: H6N5S<sub>2,3</sub>S<sub>2,6</sub>1, H6N5S<sub>2,3</sub>1S<sub>2,6</sub>2, and H6N5S<sub>2,6</sub>3 ( $m/z$  1042.42, 1052.09, and 1061.77, respectively). **B1 – B4** show the corresponding MS/MS spectra. Specific ions such as B-ions 656.25 and 685.26 confirm the type of sialic acid linkage(s) present. In the case of multiple charge states of a single fragment, only one charge state is annotated, such as  $m/z$  2179.89 [ $M+H$ ]<sup>1+</sup> ( $m/z$  1090.97 [ $M+2H$ ]<sup>2+</sup>) and  $m/z$  1652.75 [ $M+H$ ]<sup>1+</sup> ( $m/z$  826.88 [ $M+2H$ ]<sup>2+</sup>). Monosaccharide annotation is provided in the blue box.



amidated  $\alpha$ 2,3- and esterified  $\alpha$ 2,6-antennae, respectively. In addition, several Y-ions provide further support for the assignments. For example, **Figure 2.B2** shows a fragment at  $m/z$  1787.80<sup>1+</sup>, indicating the composition H5N4S<sub>2,3</sub>1, resulting from the loss of two sialylated antennae. A similar fragment at  $m/z$  1815.03<sup>1+</sup> with an  $\alpha$ 2,6-sialic acid is shown in **Figure 2.B3**. Another B-ion with two  $\alpha$ 2,6-linked sialylated antennae attached to a mannose is detected at  $m/z$  1531.75<sup>1+</sup> in **Figure 2.B4**. Furthermore, no ions corresponding to  $\alpha$ 2,3-linked sialic acids are detected in **Figure 2.B4** as the *N*-glycan solely contains  $\alpha$ 2,6-sialylated species. Thus, consistent with previous studies that employed sialic acid linkage-specific derivatization,<sup>8,17,23</sup> the present method allows the unambiguous assignment of the sialic acid linkages present in plasma *N*-glycans using RPLC-MS.

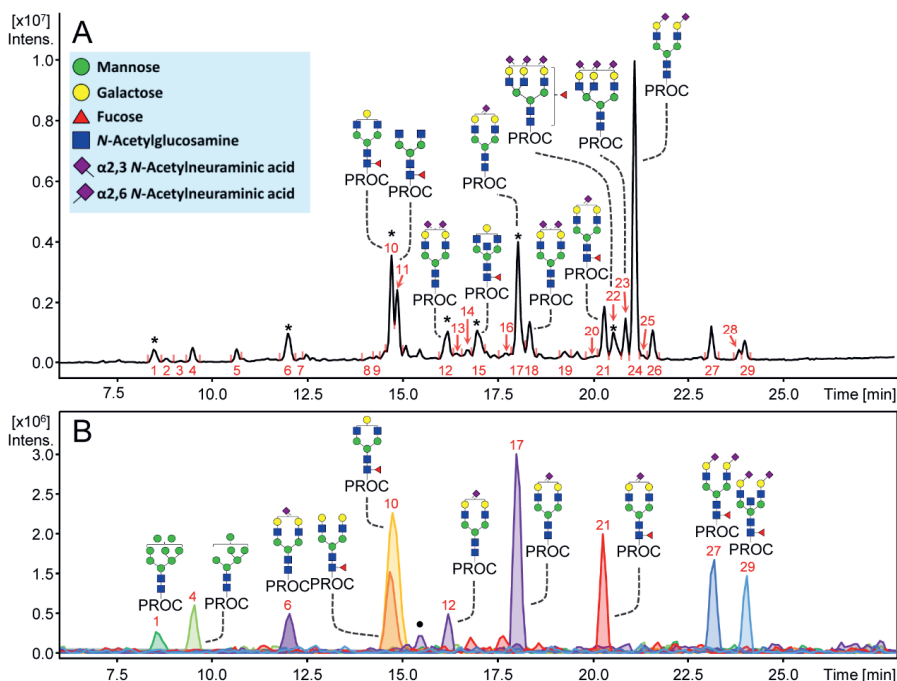
Interestingly, **Figure 2** illustrates a greater relative abundance of the derivatized sialylated oxonium ions, especially  $m/z$  685.31<sup>1+</sup> (**Figure 2.B4**) shows a 100% relative abundance for the *N*-glycan that is fully occupied with  $\alpha$ 2,6-linked sialic acids. This is in comparison with the non-derivatized ion  $m/z$  657.21<sup>1+</sup> (60%; **Figure 2.B1**). Furthermore, derivatized Y-ions such as  $m/z$  1235.50<sup>2+</sup> and 1250.54<sup>2+</sup> in **Figure 2.B2** and **2.B3**, respectively, also appear to have a greater relative abundance than their non-derivatized counterpart  $m/z$  1222.99<sup>2+</sup> (**Figure 2.B1**). Similar results were also obtained when Suzuki *et al.* compared non-derivatized and derivatized glycans from human alpha-1-acid glycoprotein ( $\alpha$ 1-AGP).<sup>19</sup> Thus, this is likely due to the greater stability that derivatized sialic acids exhibit during fragmentation following derivatization. However, an improvement in ionization efficiencies due to derivatization can also not be discounted. Nonetheless, the increase in relative abundance of fragment ions in the MS/MS spectra greatly enhances the identification of glycan structures. Further studies should focus on comparing derivatized and non-derivatized complex glycan standards whilst controlling for the injection amounts in order to explore this phenomenon further.

Separation of positional linkage isomers was achieved on RPLC following the combination of the procainamide labeling and EEA derivatization approaches. This is illustrated in **Supporting Information 1, Figure S1.B2** whereby two positional isomers are detected for each of the  $\alpha$ 2,3 and  $\alpha$ 2,6-linkage variants of H5N4S1. A comparison of the sum of the relative abundance of all isomers for H5N4S1 in the non-derivatized and derivatized profiles is provided in **Supporting Information 1, Figure**

**S1.C3.** It is notable that a similar total relative abundance for H5N4S1 was achieved for both profiles. This indicates that the positional isomers, determined in the non-derivatized chromatogram, were also captured in the derivatized profile, yet more information is achieved with the latter as it resolves both antenna occupation and sialic acid linkage. Interestingly, the separation time between H5N4S1 isomers in **Supporting Information 1, Figure S1.A2** is approximately 3 min whereas **Supporting Information 2, Figure S1.B2** shows that there is a greater separation time (7 min) between H5N4S1 positional isomers with the  $\alpha$ 2,6 linkage. However, this was less apparent for positional isomers with the  $\alpha$ 2,3-linkage linkage, further indicating the different influence that the ethyl esterification and amidation modifications have in RPLC. Additionally, **Supporting Information 1, Figure S2.A** shows that two isomers of H5N4S<sub>2,3</sub>1S<sub>2,6</sub>1 were detected. Moreover, this phenomenon has also been observed in previous studies.<sup>19,20</sup> Suzuki *et al.* demonstrated that triantennary glycan isomers were resolved based on the attachment position of *N*-acetyl- and *N*-glycolylneuraminic acids, with these monosaccharides being attached to either galactose or *N*-acetylglucosamine (GlcNAc) residues in fetuin.<sup>19</sup> However, similar to Jin *et al.*,<sup>20</sup> here we observed that structural isomers likely containing the same sialic acid linkage(s) attached to a galactose are separated based upon which arm is occupied ( $\alpha$ 3 *versus*  $\alpha$ 6). Despite this, no conclusions could be drawn from the RT as well as MS/MS spectra (**Supporting Information 1, Figure S2.B and C**) regarding which antenna occupancy gives rise to the later or earlier elution. Therefore, further research is required using well-defined standards in order to characterize this feature. Importantly, isomer separation of sialylated glycan structures has not previously been recognized as a strength of RPLC.<sup>18</sup> However, the results achieved by this investigation, among others, shows that both positional and sialic acid linkage isomers may be determined through the combined effect of fluorescent labeling and sialic acid derivatization.

### Plasma *N*-glycan RPLC Profile

Multiple *N*-glycans structures, including sialylated and non-sialylated structures, were resolved on RPLC following procainamide-labeling and EEA. In **Figure 3.A**, 29 fluorescent peaks were determined after data curation and the 10 most abundant peaks were assigned with the most abundant *N*-glycan based on MS detection. While some peaks are clearly dominated by a certain glycan composition, other feature



**Figure 3.** Released *N*-glycans from plasma on RPLC following procainamide fluorescent labeling and sialic acid derivatization. **(A)** *N*-glycan assignments of the 10 most abundant peaks are shown. In cases where multiple *N*-glycans elute under the same peak, the most abundant *N*-glycan is represented. Asterisk (\*) represents peaks with multiple eluting *N*-glycans. **(B)** The effect of monosaccharides on RT is illustrated. This is highlighted by extracted ion chromatogram of specific *N*-glycans (from left to right): H8N2, H6N2, H5N4S<sub>2,3</sub>1, H5N4F1, H4N4F1, H5N4S<sub>2,6</sub>1 (isomer 1), H5N4S<sub>2,6</sub>1 (isomer 2), H5N4F1S<sub>2,6</sub>1, H5N4F1S<sub>2,6</sub>2, H5N5F1S<sub>2,6</sub>2. Monosaccharide annotation is provided in the blue box. Symbol (•) denotes overlapping *m/z* of H5N4S<sub>2,6</sub>1 with a non-assigned analyte with the same *m/z*. Peak numbers are illustrated in red. For the full list of assignments see **Supporting Information 2, Table S3**.

several co-eluting *N*-glycans (**Supporting Information 2, Table S3**). The specific influence of different monosaccharides on retention is highlighted in **Figure 3.B**. A decreasing number of mannoses in the *N*-glycan composition is associated with higher hydrophobicity.<sup>13</sup> This is exemplified by the high-mannose structures that are eluting first in the profile and the EICs of *N*-glycans H8N2 and H6N2 demonstrate that elution occurs in order of decreasing hexose (mannose) numbers. We also observed the separation of some high-mannose isomers, similar to Chen *et al.*<sup>30</sup> This is illustrated by the composition H7N2 which is found in peaks 2 and 3 (**Supporting Information 2, Table S3**). Complex non-sialylated glycans also elute in order of decreasing hexose (galactose) numbers. This is shown in **Figure 3.A** whereby H4N4F1 (peak 10) elutes

before H3N4F1 (peak 11). However, it seems a similar influence on retention is observed when the antenna is partially or fully occupied with galactose as H5N4F1 co-elutes with H4N4F1 in peak 10 (**Figure 3.B**).

A diverse number of separated sialylated *N*-glycan species are shown in the plasma profile on RPLC (**Figure 3.A**), including positional isomers and linkage variants (**Figure 3.B**), as previously mentioned. In contrast, only a single peak was detected here for other isomeric species such as H4N4F1 (G1F), which is known to have two positional isomers in plasma.<sup>13</sup> For example, Higel *et al.* showed that the H4N4F1 isomers may be resolved using 2-AA.<sup>24</sup> Thus, the current method may be useful for the investigation of positional sialylated isomers, however other isomers which lack sialic acids and do not contain sufficient hydrophobic differences may not be separated and remain a challenge for this analysis.

In addition to sialic acid linkage, core fucosylation has a large influence on the RT of *N*-glycans observed here. The effect of core fucosylation is demonstrated by H5N4F1S<sub>2,6</sub>1 (peak 21) which elutes later than H5N4S<sub>2,6</sub>1 (peaks 12 and 17) in **Figure 3.B**. Furthermore, H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>2 appears to lack a diagnostic Y-ion for core-fucosylation ( $m/z$  587.33; N1F1-Proc), suggesting that the fucose is located on the antenna of this glycan. Thus, the influence of antennary fucosylation on RT is demonstrated by H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>2 which elutes earlier than H6N5S<sub>2,3</sub>1S<sub>2,6</sub>2 (**Figure 3.A**). This suggests that the addition of an antennary fucose decreases the RT. These results are similar to previous findings whereby it was determined that antennary fucosylation may decrease or have a negligible influence on RT,<sup>11,27</sup> depending on which antennae is fucosylated. Furthermore, the analysis of 2-PA labeled and derivatized glycans from  $\alpha$ 1-AGP showed that a triantennary *N*-glycan containing the sialyl-Lewis<sup>x</sup> antigen eluted before its non-fucosylated counterpart.<sup>19</sup> Importantly, antennary- and core-fucosylation confer different functions in biological systems<sup>31,32</sup> and, therefore, techniques that distinguish them are required.<sup>33</sup>

An increasing number of GlcNAc residues results in longer RTs. However, it is difficult to define the true effect of an increasing number of antennae as this is generally accompanied by increasing glycan size due to capping by a galactose and sialic acid, the latter of which has already been shown to have a large influence on retention. In addition, it should be noted that the second isomer of the *N*-glycan H5N4S<sub>2,3</sub>1S<sub>2,6</sub>1

(peak 18) elutes later than H6N5S<sub>2,3</sub>1S<sub>2,6</sub>1 (**Figure 3.A** and **Supporting Information 2, Table S3**, peak 14), despite having less antennae but the same number and linkage of sialic acids. Thus, this further highlights the large influence that derivatized sialylated positional isomers have on RT. In any case, **Figure 3.A** shows that *N*-glycans with different numbers of antennae may be separated. Furthermore, the influence of a bisecting GlcNAc is shown in **Figure 3.B**, as H5N5F1S<sub>2,6</sub>2 elutes later than H5N4F1S<sub>2,6</sub>2 (peaks at 22.5 – 25.0 minutes). Despite this, several other likely-bisected *N*-glycan species were unable to be confirmed by MS/MS (**Supporting Information 2, Table S1**). However, previous research using RPLC has shown the separation of *N*-glycans containing bisecting GlcNAc from those structures without bisection. Thus, this feature is an important aspect of this method as conventional released *N*-glycan analysis often requires exoglycosidase enzymes in order to confirm bisection, and should be further explored.

One of the main advantages of RPLC is its wide utilization and application.<sup>11</sup> As a result, it is a well-developed technique and there are a large number of applications that may be implemented. For example, separation parameters can be optimized per application as a wide variety of columns are available that vary in terms of stationary phase as well as length and particle size, and may be obtained from various manufacturers.<sup>11</sup> Furthermore, nano-flow techniques may be implemented on RPLC systems in order to achieve highly-sensitive analyses.<sup>34</sup> Moreover, as described here and in previous studies,<sup>24,30</sup> glycan analysis by RPLC is suitable for the investigation of various glycan species, such as high-mannose isomers as well as core- and antennary-fucosylated, and bisected *N*-glycans. In addition, recent advancements in RP stationary phases have also improved the separation of sialylated *N*-glycans labeled with 2-AA.<sup>35</sup> Despite this, sialic acid linkage and further isomer information have remained a challenge for such applications. However, recent research has shown that RPLC-MS may now be implemented for the analysis of derivatized sialylated *N*-glycans, allowing linkage-specific and isomeric information to be obtained.<sup>19,20</sup> In the case of Suzuki *et al.*, they demonstrated an in-depth approach, employing a fractionation step followed by RPLC-MS, whereas our study utilized semi-automated sample preparation and a shorter gradient time (80 vs. 35 mins, respectively) in order to enable more and faster analyses. Nonetheless, the application of different labeling and derivatization techniques in both studies demonstrate the

potential of this approach as the combination of fluorescent labels, derivatization strategies and RPLC systems may be explored to further enhance glycomic studies.

### Method Validation

The performance of the EEA and RPLC-FD-MS platform was validated and compared with the gold standard method for released *N*-glycan analysis, namely, HILIC-FD-MS. Crucially, for the HILIC-FD-MS platform, *N*-glycans were subjected only to the standard *N*-glycan protocol, whereby measurement is carried out following fluorescent labeling and clean-up. In addition, three quantification approaches were tested using FD, MS and a combination of these two approaches via FD-MS across both platforms. In **Table 1**, it is shown that the RPLC-FD-MS platform resulted in the assignment of 29 fluorescent peaks (FD) and 39 *N*-glycan compositions (MS). Similarly, 27 fluorescent peaks and 41 *N*-glycan compositions were detected by the HILIC-FD-MS platform. In comparison, previous research has determined 117 and 167 *N*-glycans in serum and plasma by HILIC-MS<sup>36,37</sup> and capillary electrophoresis (CE)-MS,<sup>17</sup> respectively. Importantly, these studies employed high resolution and high sensitivity MS. Moreover, Lageveen-Kammeijer *et al.* also performed linkage-specific sialic acid derivatization prior to measurement by CE-MS, a highly sensitive technique as it operates at a nano-flow level. Nonetheless, CE-MS is still not widely available in most laboratories and often lacks in repeatability (capillary to capillary), long separation times (> 80 min) and expertise. It is expected that the sensitivity of the developed platform reported here could be further improved via coupling with high-sensitive MS instruments as well as employing a nano-flow column.

The fraction of sialic acid linkage determined structures is represented in **Table 1**. In this case, 23 sialylated structures were determined by the RPLC-FD-MS platform and all structures could be linkage-specified by their precursor mass, and most were also confirmed via MS/MS. In comparison, 24 sialylated *N*-glycans were determined in the HILIC-FD-MS profile and 18 could be assigned with linkage information, in accordance with elution positions reported in the literature.<sup>13</sup> However, in order to provide experimental results of sialic acid linkages with the HILIC-FD-MS setup, further studies are required which involve sialidase enzyme treatments. In contrast, the combination

of derivatizing the procainamide-labeled glycans using EEA followed by RPLC-FD-MS analysis allows direct and unambiguous assignments of sialic acid-linkages.

**Table 1. Performance measures of the RPLC- and HILIC-FD-MS platforms.** Features determined by these platforms include the number [#] of FD peaks and *N*-glycan compositions. In addition, the fraction [a/b] of sialic acid-linkage detected *N*-glycans is provided (*N*-glycans with sialic acid-linkage determined/total number of sialylated *N*-glycans). Sialic acid linkages were determined directly by the EEA and RPLC-FD-MS protocol whereas assignments for HILIC-FD-MS were made in accordance with elution positions reported in the literature.<sup>13</sup> Three quantification approaches are displayed: FD and MS, or the combination of these two platforms via a third quantification approach, FD-MS. Furthermore, the inter-day variation for the 10 most abundant as well as all detected *N*-glycans is provided for all three quantification approaches of each platform. Quantification of fluorescent peaks and *N*-glycan compositions determined by MS was performed using HappyTools and LaCyTools, respectively, as detailed in **Supporting Information 1, Sections S1.5 and S1.6**. Assignments were: <sup>x</sup> confirmed by diagnostic ions in MS/MS or <sup>y</sup> made in accordance with the literature.<sup>13</sup> For the full list of assignments see **Supporting Information 2, Tables S1, S3, and S4**.

Features	RPLC			HILIC		
	FD	MS	FD-MS	FD	MS	FD-MS
FD peaks [#]	29	N/A	29	27	N/A	27
<i>N</i> -glycans [#]	N/A	39	39	N/A	41	41
Sialic acid linkage [a/b]	N/A	22/23 <sup>x</sup>	22/23 <sup>x</sup>	N/A	18/24 <sup>y</sup>	18/24 <sup>y</sup>

	S.D.	RSD	S.D.	RSD	S.D.	RSD	S.D.	RSD	S.D.	RSD	S.D.	RSD
<b>Median % (top 10)</b>	0.3	5.2	0.5	12.3	0.3	5.4	0.1	1.9	0.2	5.7	0.1	1.8
<b>Median % (total)</b>	0.2	7.3	0.2	18.6	0.2	11.4	0.0	2.7	0.1	6.1	0.0	4.4

The intermediate precision and repeatability of the EEA and RPLC-FD-MS platform was obtained via intra- ( $n = 3$ ) and inter-day measurements (total  $n = 50$ ). The inter-day relative standard deviation (RSD) of the 10 most abundant *N*-glycans (**Table 1**) revealed that FD had the best performance (5.2%), followed by FD-MS (5.4%), and MS (12.3%). **Table 1** also shows that similar results for the three quantification approaches were obtained for the HILIC-FD-MS platform when an intra-day experiment ( $n = 1$ ; total  $n = 3$ ) was performed. While FD showed the highest measure of performance, it provided the lowest coverage of features. This is in contrast with MS-only, as more structures could be quantified, however, with a less precise quantification than FD. Therefore, by combining and implementing the best of the two approaches, FD-MS resulted in an increase in coverage of specific features with a higher precision than solely using MS. This is illustrated in **Table 1** whereby 29 peaks were determined by RPLC-FD, whereas RPLC-FD-MS enabled the direct

quantification of 39 *N*-glycan species. Similarly, HILIC-FD detected 27 peaks while HILIC-FD-MS covered 41 *N*-glycan species. Thus, a combination of FD and MS via FD-MS quantification allowed improvements for both separation platforms in regard to *N*-glycan coverage and quantification precision.

Overall, the HILIC-FD-MS platform showed higher precision than the RPLC-FD-MS method across each of the quantification approaches. Nonetheless, with regard to the 10 most abundant *N*-glycans measured by RPLC-FD-MS, both quantification approaches incorporating FD (FD and FD-MS) showed RSDs below 10% and employing MS-only quantification resulted in an RSD below 15%. It should be noted that an overall increase in RSDs is observed when calculated for the total number of FD peaks and *N*-glycan compositions across both platforms. However, as shown in **Supporting Information 2, Table S3**, low abundant glycans display higher RSDs when determined by RPLC-FD-MS in comparison with its HILIC-FD-MS counterpart. The increase in variation of the EEA and RPLC-FD-MS setup is likely to be due to the two additional sample processing steps, including a chemical derivatization and HILIC-based clean-up. Thus, further research could focus on improving the procedure by determining whether only a single clean-up step could be performed following *N*-glycan labeling and derivatization.

### Platform Complementarity

The analysis of *N*-glycans is challenging and normally a single method is unable to capture many of the structural differences that exist between different species. However, the implementation of orthogonal methods allows in-depth and complimentary information to be obtained. In this study, we examined the complementarity between the newly developed RPLC-FD-MS platform and the gold standard method, HILIC-FD-MS. The similarities and differences between these two methods are highlighted in **Supporting Information 2, Table S3**. As mentioned previously, H4N4F1 was detected as a single structure by RPLC-FD-MS whereas two isomers were determined by the HILIC-FD-MS approach. Furthermore, two isomers are shown for both H5N4S<sub>2,6</sub>1 and H5N4S<sub>2,3</sub>1S<sub>2,6</sub>1 when analyzed by RPLC-FD-MS, whereas single structures are determined for both of these compositions by HILIC-FD-MS. Thus, a sum of both isomers may be determined by one platform whereas the alternate method may separate the small structural differences between the isomers.



In some cases, *N*-glycans were only detected by one platform. This is shown in **Supporting Information 2, Table S3** where two isomers of H6N5S<sub>2,3</sub>1S<sub>2,6</sub>1 were analyzed by RPLC-FD-MS whereas this structure was not detected by HILIC-FD-MS. Overall, 11 and 13 unique structures were quantified by RPLC-FD-MS and HILIC-FD-MS, respectively (**Supporting Information 1, Figure S3.A**). Despite this, these unique structures only account for 5% and 8% of total areas determined by these platforms, respectively (**Supporting Information 1, Figure S3.B**). In contrast, there are 28 *N*-glycans that were determined by both methods (**Supporting Information 1, Figure S3.A**), covering 95% of the total area for RPLC-FD-MS and 92% of the total area for HILIC-FD-MS. This shows that both platforms cover the majority of detected *N*-glycans and, as a result, provide important orthogonal information regarding the sample.

The association between the relative abundances of *N*-glycans determined by both platforms was investigated. This was performed by examining the 28 overlapping structures only (re-normalized to the sum of the total area of these compositions). In the case of isomeric species, their relative abundances were summed in order to perform the comparison between both platforms (**Supporting Information 1, Figure S4**). Importantly, H5N4S<sub>2,6</sub>1 and H5N4S<sub>2,6</sub>2, were not plotted as these two structures are much more abundant in plasma than other *N*-glycans and may result in an overestimation of association between the two platforms. Similar relative abundances were determined for overlapping structures ( $R^2 = 0.78$ ), however some discrepancies in quantification may arise from summing isomer signals for the purposes of the comparison, as well as slight ionization biases due to differential derivatization<sup>38</sup>. Nonetheless, this study shows that both platforms may be applied to the same sample and used as complementary approaches.

## Perspectives

The protocol presented here represents an important development for the application of RPLC-MS to analyze released *N*-glycans, enabling the elucidation of sialic acid linkage-specificity. Nevertheless, further developments should be carried out in order to further explore and exploit the capabilities of this technique. For example, the effects of isomers on separation should be defined using well-established *N*-glycan standards. In addition, further identification of by-products related to labeling or

derivatization should be performed. Finally, the combination of various derivatization and labeling strategies could also be explored.

## Conclusions

The developed platform allows released sialylated *N*-glycans to be efficiently analyzed using RPLC-FD-MS(/MS) and the procedure is compatible with, and complementary to, the standard *N*-glycan processing protocol. Thus, the platform is applicable whereby unambiguous sialic acid linkage assignment is required from a single measurement, in addition to information regarding specific types of *N*-glycan isomers. The investigation of isomeric species such as these is not a common application of RPLC techniques. Thus, this approach allows greater access to a platform that is already well-developed, widely available and easily applicable for the linkage-specific analysis of sialylated *N*-glycans.

## Acknowledgements

The authors wish to thank Bas Jansen for his support on the use HappyTools, as well as Jennifer Hendel and Paulina Urbanowicz for providing valuable input while carrying out experiments. The Table of Contents and Figure 1 were created with BioRender.com.

## Funding

This research was funded by the European Union's Horizon 2020 Research and Innovation Program (GlySign, Grant No. 722095).

## Conflict Of Interest

Daniel I.R. Spencer and Richard A. Gardner are employed by Ludger Ltd., a company that provides commercial glycoanalytical products and services.

## Data Availability

The data with regard to the intra- and inter-day validation are available using the identifier GPST000190 at the GlycoPost repository.<sup>39</sup> All other data, such as that

relating to the method development, are available from the corresponding author upon request.

## Supporting Information

**Supporting Information 1 (PDF):** **Section S1**, additional experimental section; **Section S2**, additional results section regarding method development; **Section S3**, Supporting Information Figures, **Figure S1**. Non-derivatized versus derivatized procainamide-labeled plasma *N*-glycans on RPLC (70 min gradient), **Figure S2**. Sialic acid linkage-specific MS and MS/MS spectra, **Figure S3**. Unique and overlapping *N*-glycan features detected by the RPLC-FD-MS and HILIC-FD-MS platforms, **Figure S4**. Comparison of relative abundances of 26 overlapping *N*-glycans detected between the RPLCFD-MS and HILIC-FD-MS platforms, **Figure S5**. Adduct ion formation during MS measurement, **Figure S6**. Derivatized and procainamide-labeled plasma *N*-glycan profile on RPLC; **Section S4**, Supporting Information References.

**Supporting Information 2 (XLSX):** Supporting Information Tables, **Table S1**: Diagnostic ions obtained by RPLC-FD-MS(-MS), **Table S2**: Method development of the EEA and RPLC-FD-MS protocol, **Table S3**: Overview of the glycan structures determined by the RPLC- and HILIC-FD-MS, **Table S4**: Overview of the glycan structures determined by the HILC-FD-MS, **Table S5**: HappyTools peak extraction and calibration parameters, **Table S6**: LaCyTools composition extraction and calibration parameters, **Table S7**: The 10 most abundant peaks in the 70 vs. 35 min gradient, **Table S8**: Derivatization efficiency of the EEA and RPLC-FD-MS protocol.

## References

- (1) Schauer, R.; Kamerling, J. P. Exploration of the Sialic Acid World. *Adv Carbohydr Chem Biochem* **2018**, *75*, 1–213. DOI: 10.1016/bs.accb.2018.09.001.
- (2) Zhang, Z.; Wuhrer, M.; Holst, S. Serum Sialylation Changes in Cancer. *Glycoconj J* **2018**, *35* (2), 139–160. DOI: 10.1007/s10719-018-9820-0.
- (3) Lehmann, F.; Tiralongo, E.; Tiralongo, J. Sialic Acid-Specific Lectins: Occurrence, Specificity and Function. *Cell Mol Life Sci* **2006**, *63* (12), 1331–1354. DOI: 10.1007/s00018-005-5589-y.
- (4) Schauer, R. Sialic Acids: Fascinating Sugars in Higher Animals and Man. *Zool.* **2004**, *107* (1), 49–64. DOI: 10.1016/j.zool.2003.10.002.

- (5) Schultz, M. J.; Swindall, A. F.; Bellis, S. L. Regulation of the Metastatic Cell Phenotype by Sialylated Glycans. *Cancer Metastasis Rev* **2012**, *31* (3–4), 501–518. DOI: 10.1007/s10555-012-9359-7.
- (6) de Haan, N.; Yang, S.; Cipollo, J.; Wuhrer, M. Glycomics Studies Using Sialic Acid Derivatization and Mass Spectrometry. *Nat. Rev. Chem.* **2020**, *4* (5), 229–242. DOI: 10.1038/s41570-020-0174-3.
- (7) Wheeler, S. F.; Domann, P.; Harvey, D. J. Derivatization of Sialic Acids for Stabilization in Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry and Concomitant Differentiation of Alpha(2 --> 3)- and Alpha(2 --> 6)-Isomers. *Rapid Commun Mass Spectrom* **2009**, *23* (2), 303–312. DOI: 10.1002/rcm.3867.
- (8) Reiding, K. R.; Blank, D.; Kuijper, D. M.; Deelder, A. M.; Wuhrer, M. High-Throughput Profiling of Protein N-Glycosylation by MALDI-TOF-MS Employing Linkage-Specific Sialic Acid Esterification. *Anal. Chem.* **2014**, *86* (12), 5784–5793. DOI: 10.1021/ac500335t.
- (9) Madunić, K.; Zhang, T.; Mayboroda, O. A.; Holst, S.; Stavenhagen, K.; Jin, C.; Karlsson, N. G.; Lageveen-Kammeijer, G. S. M.; Wuhrer, M. Colorectal Cancer Cell Lines Show Striking Diversity of Their O-Glycome Reflecting the Cellular Differentiation Phenotype. *Cell. Mol. Life Sci.* **2021**, *78* (1), 337–350. DOI: 10.1007/s00018-020-03504-z.
- (10) Ashwood, C.; Pratt, B.; MacLean, B. X.; Gundry, R. L.; Packer, N. H. Standardization of PGC-LC-MS-Based Glycomics for Sample Specific Glycotyping. *Analyst* **2019**, *144* (11), 3601–3612. DOI: 10.1039/c9an00486f.
- (11) Vreeker, G. C.; Wuhrer, M. Reversed-Phase Separation Methods for Glycan Analysis. *Anal Bioanal Chem* **2017**, *409* (2), 359–378. DOI: 10.1007/s00216-016-0073-0.
- (12) Ruhaak, L. R.; Huhn, C.; Waterreus, W. J.; de Boer, A. R.; Neususs, C.; Hokke, C. H.; Deelder, A. M.; Wuhrer, M. Hydrophilic Interaction Chromatography-Based High-Throughput Sample Preparation Method for N-Glycan Analysis from Total Human Plasma Glycoproteins. *Anal Chem* **2008**, *80* (15), 6119–6126. DOI: 10.1021/ac800630x.
- (13) Reiding, K. R.; Bondt, A.; Hennig, R.; Gardner, R. A.; O'Flaherty, R.; Trbojevic-Akmacic, I.; Shubhakar, A.; Hazes, J. M. W.; Reichl, U.; Fernandes, D. L.; Pucic-Bakovic, M.; Rapp, E.; Spencer, D. I. R.; Dolhain, R.; Rudd, P. M.; Lauc, G.; Wuhrer, M. High-Throughput Serum N-Glycomics: Method Comparison and Application to Study Rheumatoid Arthritis and Pregnancy-Associated Changes. *Mol Cell Proteomics* **2019**, *18* (1), 3–15. DOI: 10.1074/mcp.RA117.000454.
- (14) Jensen, P. H.; Karlsson, N. G.; Kolarich, D.; Packer, N. H. Structural Analysis of N- and O-Glycans Released from Glycoproteins. *Nat Protoc* **2012**, *7* (7), 1299–1310. DOI: 10.1038/nprot.2012.063.
- (15) Tousi, F.; Bones, J.; Hancock, W. S.; Hincapie, M. Differential Chemical Derivatization Integrated with Chromatographic Separation for Analysis of Isomeric Sialylated N -Glycans: A Nano-Hydrophilic Interaction Liquid Chromatography-MS Platform. *Anal. Chem.* **2013**, *85*

- (17), 8421–8428. DOI: 10.1021/ac4018007.
- (16) Smith, J.; Millán-Martín, S.; Mittermayr, S.; Hilborne, V.; Davey, G.; Polom, K.; Roviello, F.; Bones, J. 2-Dimensional Ultra-High Performance Liquid Chromatography and DMT-MM Derivatization Paired with Tandem Mass Spectrometry for Comprehensive Serum N-Glycome Characterization. *Anal. Chim. Acta* **2021**, *1179*, 338840. DOI: 10.1016/j.aca.2021.338840.
- (17) Lageveen-Kammeijer, G. S. M.; de Haan, N.; Mohaupt, P.; Wagt, S.; Filius, M.; Nouta, J.; Falck, D.; Wuhrer, M. Highly Sensitive CE-ESI-MS Analysis of N-Glycans from Complex Biological Samples. *Nat. Commun.* **2019**, *10* (1), 1–8. DOI: 10.1038/s41467-019-09910-7.
- (18) Veillon, L.; Huang, Y.; Peng, W.; Dong, X.; Cho, B. G.; Mechref, Y. Characterization of Isomeric Glycan Structures by LC-MS/MS. *Electrophoresis*. Wiley-VCH Verlag September 1, 2017, pp 2100–2114. DOI: 10.1002/elps.201700042.
- (19) Suzuki, N.; Abe, T.; Natsuka, S. Quantitative LC-MS and MS/MS Analysis of Sialylated Glycans Modified by Linkage-Specific Alkylamidation. *Anal. Biochem.* **2019**, *567*, 117–127. DOI: 10.1016/j.ab.2018.11.014.
- (20) Jin, W.; Wang, C.; Yang, M.; Wei, M.; Huang, L.; Wang, Z. Glycoqueuing: Isomer-Specific Quantification for Sialylation-Focused Glycomics. *Anal. Chem.* **2019**, *91* (16), 10492–10500. DOI: 10.1021/acs.analchem.9b01393.
- (21) Ventham, N. T.; Gardner, R. A.; Kennedy, N. A.; Shubhakar, A.; Kalla, R.; Nimmo, E. R.; Consortium, I.-B.; Fernandes, D. L.; Satsangi, J.; Spencer, D. I. Changes to Serum Sample Tube and Processing Methodology Does Not Cause Intra-Individual [Corrected] Variation in Automated Whole Serum N-Glycan Profiling in Health and Disease. *PLoS One* **2015**, *10* (4), e0123028. DOI: 10.1371/journal.pone.0123028.
- (22) Varki, A. et al. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **2015**, *25* (12), 1323–1324. DOI: 10.1093/glycob/cwv091.
- (23) Holst, S.; Heijs, B.; de Haan, N.; van Zeijl, R. J.; Briaire-de Bruijn, I. H.; van Pelt, G. W.; Mehta, A. S.; Angel, P. M.; Mesker, W. E.; Tollenaar, R. A.; Drake, R. R.; Bovee, J. V.; McDonnell, L. A.; Wuhrer, M. Linkage-Specific in Situ Sialic Acid Derivatization for N-Glycan Mass Spectrometry Imaging of Formalin-Fixed Paraffin-Embedded Tissues. *Anal Chem* **2016**, *88* (11), 5904–5913. DOI: 10.1021/acs.analchem.6b00819.
- (24) Higel, F.; Demelbauer, U.; Seidl, A.; Friess, W.; Sorgel, F. Reversed-Phase Liquid-Chromatographic Mass Spectrometric N-Glycan Analysis of Biopharmaceuticals. *Anal Bioanal Chem* **2013**, *405* (8), 2481–2493. DOI: 10.1007/s00216-012-6690-3.
- (25) Natsuka, S.; Masuda, M.; Sumiyoshi, W.; Nakakita, S. Improved Method for Drawing of a Glycan Map, and the First Page of Glycan Atlas, Which Is a Compilation of Glycan Maps for a Whole Organism. *PLoS One* **2014**, *9* (7), e102219. DOI: 10.1371/journal.pone.0102219.
- (26) Zhou, S.; Veillon, L.; Dong, X.; Huang, Y.; Mechref, Y. Direct Comparison of Derivatization

- Strategies for LC-MS/MS Analysis of N-Glycans. *Analyst* **2017**, *142* (23), 4446–4455. DOI: 10.1039/C7AN01262D.
- (27) Tomiya, N.; Takahashi, N. Contribution of Component Monosaccharides to the Coordinates of Neutral and Sialyl Pyridylaminated N-Glycans on a Two-Dimensional Sugar Map. *Anal Biochem* **1998**, *264* (2), 204–210. DOI: 10.1006/abio.1998.2849.
- (28) Kozak, R. P.; Tortosa, C. B.; Fernandes, D. L.; Spencer, D. I. R. Comparison of Procainamide and 2-Aminobenzamide Labeling for Profiling and Identification of Glycans by Liquid Chromatography with Fluorescence Detection Coupled to Electrospray Ionization–Mass Spectrometry. *Anal. Biochem.* **2015**, *486*, 38–40. DOI: 10.1016/j.ab.2015.06.006.
- (29) Pabst, M.; Kolarich, D.; Pöttl, G.; Dalik, T.; Lubec, G.; Hofinger, A.; Altmann, F. Comparison of Fluorescent Labels for Oligosaccharides and Introduction of a New Postlabeling Purification Method. *Anal. Biochem.* **2009**, *384* (2), 263–273. DOI: 10.1016/j.ab.2008.09.041.
- (30) Chen, X.; Flynn, G. C. Analysis of N-Glycans from Recombinant Immunoglobulin G by on-Line Reversed-Phase High-Performance Liquid Chromatography/Mass Spectrometry. *Anal Biochem* **2007**, *370* (2), 147–161. DOI: 10.1016/j.ab.2007.08.012.
- (31) Demus, D.; Jansen, B. C.; Gardner, R. A.; Urbanowicz, P. A.; Wu, H.; Stambuk, T.; Juszczak, A.; Medvidovic, E. P.; Juge, N.; Gornik, O.; Owen, K. R.; Spencer, D. I. R. Interlaboratory Evaluation of Plasma N-Glycan Antennary Fucosylation as a Clinical Biomarker for HNF1A-MODY Using Liquid Chromatography Methods. *Glycoconj J* **2021**, *38* (3), 375–386. DOI: 10.1007/s10719-021-09992-w.
- (32) Testa, R.; Vanhooren, V.; Bonfigli, A. R.; Boemi, M.; Olivieri, F.; Ceriello, A.; Genovese, S.; Spazzafumo, L.; Borelli, V.; Bacalini, M. G.; Salvioli, S.; Garagnani, P.; Dewaele, S.; Libert, C.; Franceschi, C. N-Glycomic Changes in Serum Proteins in Type 2 Diabetes Mellitus Correlate with Complications and with Metabolic Syndrome Parameters. *PLoS One* **2015**, *10* (3), e0119983. DOI: 10.1371/journal.pone.0119983.
- (33) Rebello, O. D.; Nicolardi, S.; Lageveen-Kammeijer, G. S. M.; Nouta, J.; Gardner, R. A.; Mesker, W. E.; Tollenaar, R.; Spencer, D. I. R.; Wuhrer, M.; Falck, D. A Matrix-Assisted Laser Desorption/Ionization-Mass Spectrometry Assay for the Relative Quantitation of Antennary Fucosylated N-Glycans in Human Plasma. *Front Chem* **2020**, *8*, 138. DOI: 10.3389/fchem.2020.00138.
- (34) Wuhrer, M.; Koeleman, C. A.; Deelder, A. M. Two-Dimensional HPLC Separation with Reverse-Phase-Nano-LC-MS/MS for the Characterization of Glycan Pools after Labeling with 2-Aminobenzamide. *Methods Mol Biol* **2009**, *534*, 79–91. DOI: 10.1007/978-1-59745-022-5\_6.
- (35) Wilhelm, J. G.; Dehling, M.; Higel, F. High-Selectivity Profiling of Released and Labeled N-Glycans via Polar-Embedded Reversed-Phase Chromatography. *Anal. Bioanal. Chem.*

- 2019**, *411* (3), 735–743. DOI: 10.1007/s00216-018-1495-7.
- (36) Royle, L.; Campbell, M. P.; Radcliffe, C. M.; White, D. M.; Harvey, D. J.; Abrahams, J. L.; Kim, Y. G.; Henry, G. W.; Shadick, N. A.; Weinblatt, M. E.; Lee, D. M.; Rudd, P. M.; Dwek, R. A. HPLC-Based Analysis of Serum N-Glycans on a 96-Well Plate Platform with Dedicated Database Software. *Anal Biochem* **2008**, *376* (1), 1–12. DOI: 10.1016/j.ab.2007.12.012.
- (37) Harvey, D. J.; Royle, L.; Radcliffe, C. M.; Rudd, P. M.; Dwek, R. A. Structural and Quantitative Analysis of N-Linked Glycans by Matrix-Assisted Laser Desorption Ionization and Negative Ion Nanospray Mass Spectrometry. *Anal Biochem* **2008**, *376* (1), 44–60. DOI: 10.1016/j.ab.2008.01.025.
- (38) Yang, S.; Jankowska, E.; Kosikova, M.; Xie, H.; Cipollo, J. Solid-Phase Chemical Modification for Sialic Acid Linkage Analysis: Application to Glycoproteins of Host Cells Used in Influenza Virus Propagation. *Anal Chem* **2017**, *89* (17), 9508–9517. DOI: 10.1021/acs.analchem.7b02514.
- (39) Watanabe, Y.; Aoki-Kinoshita, K. F.; Ishihama, Y.; Okuda, S. GlycoPOST Realizes FAIR Principles for Glycomics Mass Spectrometry Data. *Nucleic Acids Res.* **2021**, *49* (D1), D1523–D1528. DOI: 10.1093/nar/gkaa1012.







# Chapter 5

## Serum N-Glycosylation RPLC-FD-MS Assay to Assess Colorectal Cancer Surgical Interventions

Alan B. Moran<sup>1,2</sup>, Georgia Elgood-Hunt<sup>2</sup>, Yuri E.M. van der Burgt<sup>1</sup>, Manfred Wuhrer<sup>1</sup>, Wilma E. Mesker<sup>3</sup>, Rob A.E.M. Tollenaar<sup>3</sup>, Daniel I.R. Spencer<sup>2</sup> and Guinevere S.M. Lageveen-Kammeijer<sup>1,4</sup> \*

<sup>1</sup> Leiden University Medical Center, Center for Proteomics and Metabolomics, 2300 RC Leiden, The Netherlands

<sup>2</sup> Ludger Ltd., Culham Science Centre, OX14 3EB Abingdon, United Kingdom

<sup>3</sup> Leiden University Medical Center, Department of Surgery, 2300 RC Leiden, The Netherlands

<sup>4</sup> University of Groningen, Department of Analytical Biochemistry, Groningen Research Institute of Pharmacy, Groningen, The Netherlands

Reprinted (adapted) from Moran, A. B., Elgood-Hunt, E., van der Burgt Y. E. M., Wuhrer, M., Mesker W. E., Tollenaar, R. A. E. M., Spencer, D. I. R., Lageveen-Kammeijer, G. S. M. Serum N-glycosylation RPLC-FD-MS Assay to Assess Colorectal Cancer Surgical Interventions. *Biomolecules*. **2023**, 13 (6), 896, <https://doi.org/10.3390/biom13060896> under the Creative Commons CC-BY license.



## Abstract

A newly developed analytical strategy was applied to profile the total serum *N*-glycome of 64 colorectal cancer (CRC) patients before and after surgical intervention. In this cohort it was previously found that serum *N*-glycome alterations in CRC associated with patient survival. Here, fluorescent labeling of serum *N*-glycans was applied using procainamide and followed by sialic acid derivatization specific for  $\alpha$ 2,6- and  $\alpha$ 2,3-linkage types via ethyl esterification and amidation, respectively. This strategy allowed efficient separation of specific positional isomers on reversed-phase liquid chromatography–fluorescence detection–mass spectrometry (RPLC-FD-MS) and complemented the previous glycomics data based on matrix-assisted laser desorption/ionization (MALDI)-MS that did not include such separations. The results from comparing pre-operative CRC to post-operative samples were in agreement with studies that identified a decrease of di-antennary structures with core fucosylation and an increase in sialylated tri- and tetra-antennary *N*-glycans in CRC patient sera. Pre-operative abundances of *N*-glycans showed good performance for the classification of adenocarcinoma and led to the revisit of the previous MALDI-MS dataset with regard to histological and clinical data. This strategy has potential to monitor patient profiles before, during and after clinical events such as treatment, therapy or surgery and should also be further explored.

## Introduction

Glycans drastically affect the function of a glycoprotein and contribute to the malignant phenotype of cancer cells by promoting proliferation, metastasis, and immunosuppression.<sup>1–3</sup> In this regard, aberrant glycosylation profiles add to understanding pathological steps of tumor development and progression and have become a hallmark of cancer.<sup>4,5</sup> For example, in the case of colorectal cancer (CRC), aberrant *N*-glycosylation with regard to tumor microenvironment is increasingly studied in CRC tissue samples using imaging mass spectrometry (MS).<sup>6–9</sup> Other glycomics studies also reported associations of *N*-glycome alterations with survival and tumor stage in CRC.<sup>10–12</sup> Overall, there is evidence to support that earlier diagnosis of cancer increases success of curative treatment and long-term survival in general and has motivated various omics-fields to explore retrospectively collected serum samples from cancer patients.<sup>13</sup>

Previously we have applied matrix-assisted laser desorption/ionization (MALDI)-MS as a readout for total serum *N*-glycome analysis from CRC patients where it was found that serum *N*-glycome alterations in CRC associated with patient survival.<sup>14</sup> Although MALDI-MS is a fast detection technique, it does not include any separation of positional isomers.<sup>15</sup> In contrast, recently we developed a platform whereby plasma *N*-glycans are fluorescently labeled using procainamide followed by sialic acid linkage-specific derivatization via ethyl esterification and amidation which allows efficient separation of sialylated and positional isomers on reversed-phase liquid chromatography–fluorescence detection–mass spectrometry (RPLC-FD-MS).<sup>16</sup> The performance of this assay was reported for plasma samples, but this method is also suitable for serum-based specimen. In particular, positional isomers which are related to galactose occupation on the glycan arm ( $\alpha 3$  versus  $\alpha 6$ ) as well as fucose location (antennary versus core) were separated. Thus, we reported this effective analytical strategy by benchmarking the performance of a complex mixture of released plasma *N*-glycans against the gold standard; hydrophilic interaction liquid chromatography (HILIC)-FD-MS.<sup>16</sup> We demonstrated sufficient complementarity between both platforms as an overlap up to 95% of the total relative area was demonstrated with the HILIC-FD-MS approach. Moreover, the repeatability and intermediate precision of the RPLC-FD-MS platform were determined at a median interday ( $n = 50$ ) relative standard deviation (RSD) of 5.4% for the 10 most abundant *N*-glycans. The overall

throughput, compared to a HILIC-approach for released *N*-glycan analysis, was improved as the sample preparation protocol is executed on a liquid handling robot and a 35 minute separation gradient is applied.

In the current work we re-visit the previous MALDI-MS-based serum glycomics study in order to demonstrate the clinical validity of the developed method,<sup>14</sup> and apply the RPLC-FD-MS platform on a sub-set of the samples, namely pre- and post-operative serum samples obtained from 64 CRC patients. In particular, we aim to uncover relevant serum *N*-glycomic signatures in the context of CRC, specifically related to sialic acid linkages as well as positional isomeric structures. The results will be compared with those from the MALDI-MS-based measurements of the pre- and post-operative samples and furthermore discussed in the context of histological type and cases versus controls differences.<sup>14</sup>

## Materials and Methods

### Study Design

Pre- and post-operative serum samples were collected from 64 CRC patients during the period of October 2002 and March 2013 using a standardized protocol by the LUMC Surgical Oncology Biobank, resulting in 128 samples in total. Post-operative samples were defined as “cancer-free” based on a post-operative wash out period (45 days), absence of recurrence, and a follow-up assessment after at least 1393 days. The cohort characteristics, following data curation steps that are described in the following sections, are shown in **Table 1**. The procedure for sample collection and storage of pre- and post-operative patient serum samples was previously reported.<sup>14</sup> Serum tubes (BD Vacutainer SST II plastic serum tubes, BD, United Kingdom) were used to draw blood before being centrifuged within a maximum of four hours. Both pre-operative and post-operative samples were stored for 30 days or less at - 20 °C before biobank collection and were then stored at - 80 °C. Aliquoting was carried out using a Microlab STAR line pipetting robot (Hamilton, Switzerland) into sterile 2D barcoded V-bottom 500 µL tubes (Thermo Scientific, Hudson), prior to this samples were thawed on ice and the aliquots were stored at -80 °C. Prior to sample preparation, the samples were stored at -20 °C before being thawed in order to begin sample processing.

**Table 1. Clinical characteristics of the pre- and post-operative CRC cohort.** Samples were taken before and after surgery from the same patient. A total of 64 patients were included in the analysis following data curation. Two measurements per patient were carried out in order to process the pre- and post-operative samples. Full metadata was collected for each patient in three categories of the disease: stage (1 – 4), local extent of tumor assessed pathologically (pT1 – 4), and histological type (Signet-ring cell carcinoma, >50% signet-ring cells; adenocarcinoma, and mucinous adenocarcinoma, >50% mucinous). Standard deviation (s.d.).

Pre- and post-operative CRC cohort (n = 64)		
Female sex, n (%)		33 (51.6)
Age in years, mean (s.d.)		64.2 (12.7)
Stage, n (%)	1	13 (20.3)
	2	31 (48.4)
	3	16 (25.0)
	4	4 (6.3)
pT, n (%)	pT1	7 (10.9)
	pT2	8 (12.5)
	pT3	42 (65.6)
	pT4	7 (10.9)
Histological type, n (%)	Signet-ring cell carcinoma	3 (4.7)
	Adenocarcinoma	54 (84.4)
	Mucinous adenocarcinoma	7 (10.9)

The entire workflow, including PNGase-F release, procainamide fluorescent labeling, ethyl esterification and amidation derivatization, and RPLC-FD-MS measurement, has previously been described<sup>16</sup> and is briefly explained in the following sections. The sample preparation protocol was performed using a Hamilton Microlab STARlet liquid-handling robot.

### N-glycan Release and Fluorescent Labeling

The entire cohort of pre- and post-operative serum samples were contained in two plates. Serum (5 µL) from each sample was manually and randomly added to a 96-well PCR plate (4ti-0960/C, 4titude Ltd., United Kingdom). Importantly, the samples were distributed over two plates in total and the paired pre- and post-operative samples from the same patient were contained on the same 96-well PCR plate. Commercial plasma (P9523, Sigma-Aldrich, United Kingdom) with a concentration of 1 mg/mL (n = 26, 13 per plate) was included as a positive control, deionized water (n

= 6, 18.2 M $\Omega$  resistivity, Sartorius Arium Comfort, Germany) was processed as a negative control on plate two. This was followed by the addition of 4  $\mu$ L deionized water and 1  $\mu$ L of a 10X denaturing solution obtained from an *N*-glycan release kit [LZ-rPNGASEF-96, Ludger Ltd., United Kingdom). Following this, incubation was performed at 100 °C for 10 min. The plate was removed from the oven and allowed to cool to room temperature before deionized water (6  $\mu$ L), reaction buffer (2  $\mu$ L, 500 mM sodium phosphate, pH 7.5), and 10% NP-40 (2  $\mu$ L) were added to each well. The PNGase F solution was diluted 1:1 (*v/v*) with PNGase F storage buffer, which consisted of 50 mM sodium chloride, 5 mM ethylenediaminetetraacetic acid, and 20 mM tris-hydrochloric acid (150  $\mu$ L, New England Biolabs, United Kingdom). Then, 2  $\mu$ L was added to each sample and the plate was incubated (37 °C) overnight. The next day, a 5% formic acid (Sigma-Aldrich) / deionized water solution (*v/v*; 5  $\mu$ L) was added to each well and incubation was performed at room temperature for 40 min.

A clean-up procedure was performed using a Protein Binding Membrane (PBM) plate [LC-PBM-96, Ludger Ltd.]. The plate was prepared with successive washes of methanol (100  $\mu$ L) and deionized water (300  $\mu$ L) followed by the addition of sample to the PBM plate. The PCR plate was washed with deionized water (90  $\mu$ L) which was also added to the PBM plate. Following this, a 2 mL collection plate [LP-COLLPLATE-2ML-96, Ludger Ltd.] was placed underneath the PBM plate and elution was carried out via centrifugation at 800 rcf (22°C) for 3 min. An additional 100  $\mu$ L deionized water was added to each well and another centrifugation step was performed. The samples were transferred to a 300  $\mu$ L 96-well PCR plate [4ti-0710/C, 4titude Ltd.] and dried down using a vacuum centrifuge.

Procainamide labeling solution [LT-KPROC-96, Ludger Ltd.] was prepared according to the vendor's guidelines. The labeling solution (20  $\mu$ L) was added to each well of dried released *N*-glycans and incubated (65 °C) for 60 min. After 10 min, the plate was removed, vortexed and centrifugated before being placed back in the oven for the remaining incubation time. A HILIC clean-up plate [LC-PROC-96, Ludger Ltd.] was prepared by washing with solutions of 70% ethanol (200  $\mu$ L, Sigma-Aldrich) / deionized water (*v/v*), deionized water (200  $\mu$ L) and acetonitrile (CH<sub>3</sub>CN, 200  $\mu$ L, Romil Ltd., United Kingdom). The sample was added to the plate followed by the addition of CH<sub>3</sub>CN (230  $\mu$ L) to each well. A vacuum was applied before three further CH<sub>3</sub>CN washes (200  $\mu$ L) were carried out. The samples were eluted in 100  $\mu$ L of deionized

water into a 96-well 2 mL collection plate [LP-COLLPLATE-2 ML-96, Ludger Ltd.]. Following this, a second elution was carried out using deionized water which resulted in a final volume of 200  $\mu$ L.

### Sialic Acid Derivatization

The derivatization technique has been reported in the literature<sup>17,18</sup> and the protocol was followed as previously described.<sup>16</sup> The released and labeled *N*-glycans were transferred (120  $\mu$ L) to a 300  $\mu$ L 96-well PCR plate and dried down using a vacuum centrifuge. Following this, the samples were reconstituted in deionized water (15  $\mu$ L) and 3  $\mu$ L was added to 60  $\mu$ L of ethyl esterification solution, which consisted of 250 mM 1-ethyl-3-(3-(dimethylamino) propyl) carbodiimide (EDC, Fluorochem, United Kingdom) and 250 mM hydrate 1-hydroxybenzotriazole (HOBT, Sigma-Aldrich) dissolved in ethanol. Incubation (37 °C) was performed for 60 min before 28% ammonia (12  $\mu$ L, Sigma-Aldrich) was added to the samples and the plate was incubated for a further 60 min at 37 °C. Following this, the volume in each well was brought up to 300  $\mu$ L by the addition of 225  $\mu$ L of CH<sub>3</sub>CN. A clean-up was performed using the HILIC clean-up plate procedure described above and may also be found in the literature<sup>16</sup>. The fluorescently labeled and derivatized *N*-glycans were eluted in 200  $\mu$ L deionized water and stored at -20°C until further analysis was performed.

### Reversed-Phase Liquid Chromatography–Fluorescence Detection–Mass Spectrometry

A single workup of each sample was performed followed by single injection on the RPLC-MS-platform. The measurements for each plate were carried out separately and additional deionized water blanks ( $n = 6$ ) were included during the measurement of plate one. Samples (95  $\mu$ L) were added up to 100  $\mu$ L with CH<sub>3</sub>CN so that the final concentration was 5% CH<sub>3</sub>CN (*v/v*). A 150 x 2.1 mm, ACE excel 2 C18-PFP column (ACE Ltd., United Kingdom) was prepared on an Ultimate 3000 UHPLC system (Thermo Scientific, United Kingdom) and a 20  $\mu$ L injection was performed. The column temperature was set to 60 °C, and the fluorescence detector ( $\lambda_{\text{ex}} = 310$  nm  $\lambda_{\text{em}} = 370$  nm) bulb power was set to 'high' whilst the sensitivity was set to eight. A separation gradient was carried out using a flow rate of 0.4 mL/min and 50 mM ammonium formate as solvent A and 10% CH<sub>3</sub>CN, 0.1% formic acid (*v/v*) as solvent B. The gradient consisted of the following steps: 0 – 26.5 min (15% – 95% solvent B), 26.5 –



30.5 min (95% solvent B), 30.5 – 32.5 min (95% – 15% solvent B), 32.5 to 35.1 min (15% solvent B). A representative chromatogram with numbered peaks is shown in **Supplementary Figure S1**.

Electrospray ionization–MS was carried out on an amaZon Speed ETD MS (Bruker Daltonics GmbH, Germany). The instrument parameters were as follows: ionization mode, positive mode with enhanced resolution scanning; mass range,  $m/z$  600 – 1600 and  $m/z$  900 set as the target mass; capillary voltage, 4.5 kV; source temperature, 250 °C; gas flow, 10 L/min; max accumulation time, 50.00 ms; ion charge control (ICC) target, 200,000. MS/MS spectra were obtained using helium collision-induced dissociation under the following conditions: max accumulation time MS(n): 40.00 ms; ICC target MS(n): 200,000; precursor ion selection: 3 precursors which are released after 0.2 min; MS(n) isolation width: 4  $m/z$ ; scan range selection MS(n), scale to precursor; automatic MS(n) absolute signal threshold: 25,000; automatic MS(n) relative signal threshold: 5%.

## Data Processing

The assignment of serum *N*-glycans was performed based on our previous study.<sup>16</sup> Briefly, this was carried out based on exact mass and retention order as well as diagnostic ions obtained in the MS/MS spectra. As a result, *N*-glycans with differently linked sialic acids are mass differentiated following sialic acid derivatization.<sup>16</sup> In addition, isomers with the same mass but different retention time (RT) are denoted as “isomer 1” (iso1) and “isomer 2” (iso2) based on their relative retention order. Notably, several *N*-glycans were not detected in our previous work yet were identified in this study. In this case, the diagnostic fragment ions which support the assignment of these structures are included in **Supplementary Table S1**. *N*-glycan compositions are displayed according to the notation recommended by the Consortium for Functional Glycomics:<sup>19</sup> *N*-acetylglucosamine (N; blue square), mannose (H; green circle), galactose (H; yellow circle), fucose (F; red triangle), *N*-acetylneuraminic acid (S; purple diamond).

Fluorescence detection (FD) data was exported (.txt) using Chromeleon (version 7.2) and imported into HappyTools (version 0.1-beta1, build 190115a)<sup>20</sup> for processing. The processing parameters for this software have previously been reported<sup>16</sup>. Briefly, these parameters were as follows: datapoints (100), baseline function order (1),

background window (1), peak edge type (sigma), peak edge value (2), minimum number of calibrants (4), minimum S/N (27). Importantly, baseline correction was applied and parameters for peak calibration and integration are included in **Supplementary Table S2**. Furthermore, a peak integration window (70% FWHM) was implemented which allowed sufficient peak area quantification while also minimizing peak overlap. In addition, previously optimized quality control parameters<sup>16</sup> were applied in order to perform data curation of fluorescent chromatogram peaks. For example, the maximum RT deviation was  $\pm 3$  sec and a S/N > 9 was required. With regard to the latter, a more lenient S/N criteria in comparison with analytical validation guidelines (S/N > 10)<sup>21</sup> was implemented. Furthermore, peaks eluting at the expected RT were required to be greater than the average intensity plus nine times the standard deviation (s.d.) of the same integration window in the blank and negative control measurements. A single measurement from the analysis did not meet the quality control requirements and, therefore, both paired measurements were removed from the study, resulting in a final sample size of  $n = 64$  (128 measurements).

Raw RPLC-MS data acquired in the Bruker data format (.d) were converted into the .mzXML format and imported in LaCyTools (version 2.0, build 200723).<sup>22</sup> Glycan compositions used for spectra alignment and extraction are included in **Supplementary Table S3**. Further processing parameters used for LaCyTools have previously been described.<sup>16</sup> Importantly, the area for each charge state of each glycan composition was integrated with at least 95% coverage of the theoretical isotopic distribution. Data curation parameters were also applied, including setting the  $m/z$  tolerance was set to  $\pm 100$  ppm and a S/N > 9 was used. Furthermore, the average isotopic pattern quality for each measurement was assessed for the  $[M+2H]^{2+}$ ,  $[M+3H]^{3+}$ , and  $[M+4H]^{4+}$  charge states meeting the above criteria. In this regard, the threshold was set to the average of all charge states plus one s.d. As a result, the isotopic pattern quality threshold was set to a deviation  $\leq 30\%$  from the theoretical isotopic pattern and this cut-off was applied to assess each charge state separately for every analyte. Finally, curated  $[M+2H]^{2+}$ ,  $[M+3H]^{3+}$  and  $[M+4H]^{4+}$  ions were individually quantified, followed by summation to provide a total area of all charge states for each glycan followed by adjustment to 100% in order to cover the complete isotopic envelope.

Only fluorescent peaks which contained at least one *N*-glycan that passed the MS curation parameters were considered for further processing. A third quantification approach, termed as FD-MS, was determined via the crossover of FD and MS data.<sup>16</sup> This was performed by calculating the local relative proportion of *N*-glycan peaks eluting under the same chromatographic peak based on their MS intensities. Then, the fluorescent signal was multiplied by the proportion of each *N*-glycan eluting underneath that chromatographic peak in order to derive an FD-MS signal for each *N*-glycan composition. Furthermore, the FD-MS data for each *N*-glycan composition were used to calculate “derived glycosylation traits”.<sup>14</sup> In this case, direct *N*-glycan traits refers to individual *N*-glycan compositions whereas derived traits describe common structural features shared amongst individual *N*-glycans and are based upon the biosynthetic pathway.<sup>23</sup> These include characteristics such as high-mannose, hybrid-type and complex-type as well as sialylation, galactosylation, and fucosylation. The formula for each calculated derived trait is included in **Supplementary Table S4**. Importantly, the abundances of isomers of the same *N*-glycan composition were summed in order to calculate derived glycosylation traits.

### Statistical Analysis

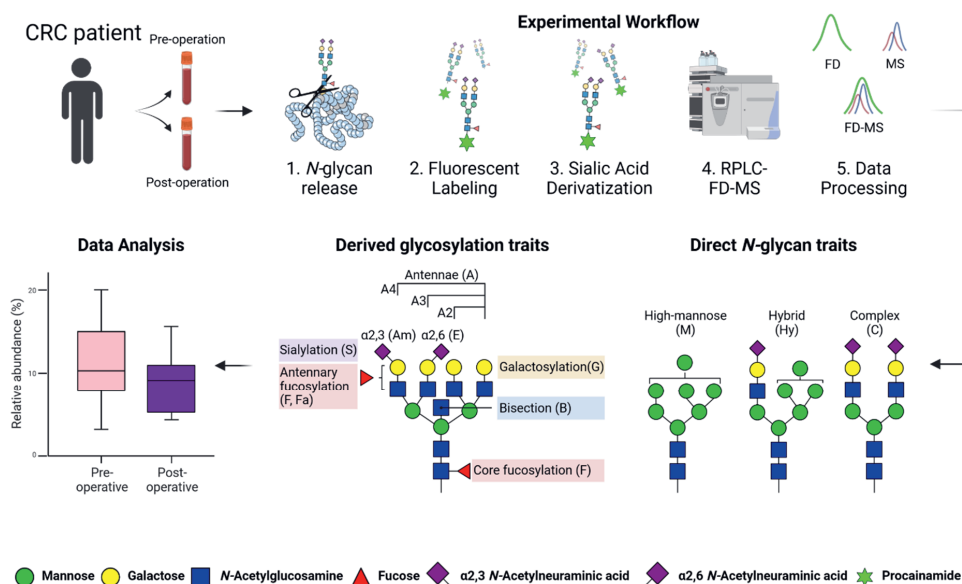
Principal component analysis (PCA) was performed in order to determine the presence of batch effects within the data. Confounders such as age, sex, and sample plate were explored and the batch effect in all three datasets (FD, MS, FD-MS) due to sample plate was corrected using linear regression based on the positive control samples.<sup>24</sup> Direct and derived traits were tested for significance based on the results of the Shapiro-wilk test, a dependent sample T-test or Wilcoxon signed-rank test in order to determine prognostic capability. A significance threshold ( $p$ -value < 0.05) with Bonferroni correction was applied. Univariate analysis was performed in python (Python 3, Create Space, Scotts Valley, CA).

Direct and derived *N*-glycan traits were considered for further analysis when significant differences between pre- and post-operative samples ( $p$ -value < 0.05 with Bonferroni correction) were observed in the current study, as well as being previously validated by the MALDI-MS study.<sup>14</sup> Associations between those specific direct and derived *N*-glycan traits and disease metadata were explored. This included three categories of the disease, namely stage (1 – 4), local extent of tumor assessed pathologically (pT1

– 4), and histological type (Signet-ring cell carcinoma, >50% signet-ring cells; adenocarcinoma, and mucinous adenocarcinoma, >50% mucinous). Thus, to determine whether there was a significant difference of each trait prior to surgery across these categories, a Kruskal-Wallis test was performed with Bonferroni correction, followed by post-hoc analysis looking at the pairwise differences of each subcategory via Dunn test. To gain an understanding of how well a trait is distinguished by a certain category, multiple models were tested and the model with the highest area under the curve (AUC) is reported, while adjusting for age and sex. In this case, the AUC was calculated by comparing one versus rest, in a 5-fold cross validation.

## Results & Discussion

*N*-glycomic profiles of pre- and post-operative serum samples from CRC patients were explored using a newly developed workflow (Figure 1).<sup>16</sup> In this approach, we



**Figure 1. Analysis of pre- and post-operative *N*-glycans in CRC by RPLC-FD-MS.** Serum samples were taken pre- and post-operation from the same patient. Each sample underwent fluorescent labeling and sialic acid derivatization *N*-glycan workflow followed by RPLC-FD-MS measurement. Three quantification approaches were used for data processing, namely FD, MS, and a combination of these two approaches via FD-MS. Direct *N*-glycan and derived glycosylation traits were determined in order to perform data analysis and comparison of the two groups. Adapted with permission from Moran, A.B., *et al.*, 2022. *Analytical Chemistry*, 94(18), pp.6639-6648 <sup>16</sup>. Copyright 2022 American Chemical Society.

performed fluorescent procainamide labeling and derivatization of  $\alpha$ 2,6- (ethyl esterification) and  $\alpha$ 2,3- (amidation) linked sialic acids followed by RPLC-FD-MS analysis in order to assess serum *N*-glycosylation changes following surgery. In addition to common *N*-glycan features, this method allowed us to perform sialic acid linkage-specific and positional isomer analysis. As shown in **Figure 1**, we investigated direct *N*-glycan traits as well as derived glycosylation traits the latter of which describes global glycosylation characteristics based on the biosynthetic pathway such as sialylation, galactosylation, bisection, fucosylation and antennarity. Unless mentioned otherwise, we refer to FD-MS quantification results throughout this study as we previously demonstrated that this technique achieves a more precise relative quantification and greater coverage of glycan structures than solely performing MS or FD quantification.<sup>16</sup> Furthermore, we compared our results with those of a previous study which analyzed the same cohort of patients using a MALDI-MS approach.<sup>14</sup> In this study, we include a comparison of the same 64 patients analyzed by both methods. Therefore, this allowed us to further validate those results using different ionization, detection and quantification parameters. Finally, we compared the clinical performance of the three quantification techniques that were applied in our study, namely FD, MS and the recently described FD-MS quantification approach.<sup>16</sup>

### **Serum *N*-glycosylation Analysis of Pre- and Post-operative CRC Samples**

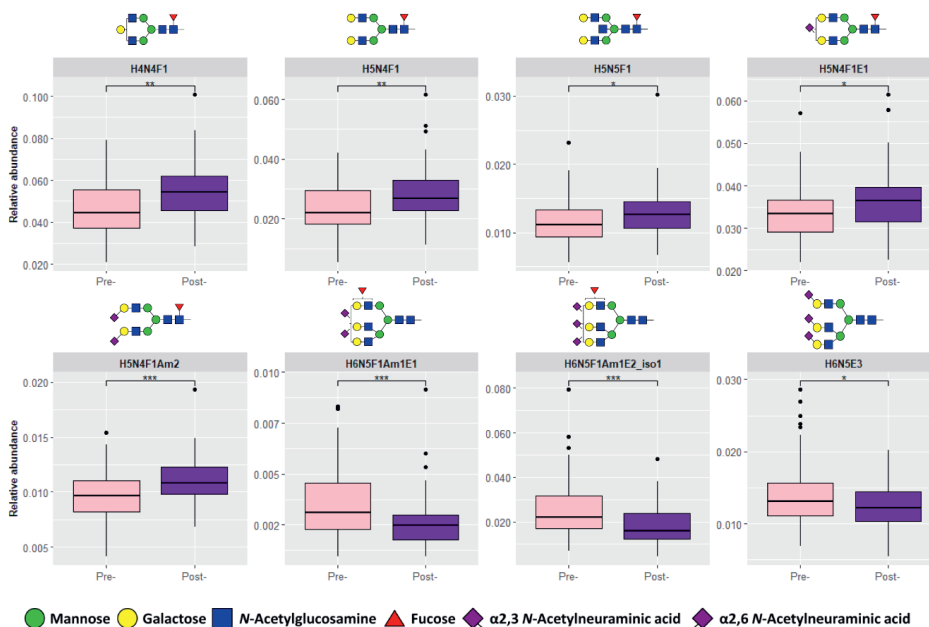
A total of 48 *N*-glycan structures were detected in this study following data curation (**Supplementary Table S5**). Univariate testing was performed in order to assess whether there were any significant differences in the relative abundances of *N*-glycans in pre- and post-operative samples, which resulted in 14 significant *N*-glycans in the FD-MS dataset. **Table 2** shows that 20 direct *N*-glycan traits were previously validated by de Vroome *et al.* using a MALDI-MS approach.<sup>14</sup> Importantly, that study included discovery and validation cohorts using a case/control setup, as well as the same pre- and post-operative CRC cohort that is also assessed by our workflow. From these, 11 *N*-glycans were found in the current study, eight of which were significant discriminators between the pre- and post-operative groups.

**Table 2. Comparison of RPLC-FD-MS and MALDI-MS results.** Asterisk (\*) denotes significant N-glycans that were found by the RPLC-FD-MS approach. CRC alteration was calculated as the pre-operative relative abundance divided by the post-operative relative abundance. All results may be found in **Supplementary Table S6**. Abbreviations: T: total, C: Common, n/a: not applicable.

Approach	All N-glycans					Validated N-glycans				
	Count			CRC Alteration		Count		CRC Alteration		
	T	C	Positional isomers	Same	Other	T	C	Significant #	Same	Other
RPLC-FD-MS	48		14	30	9	n/a		8	8	0
		39		(77%)	(23%)		11		(100%)	(0%)
MALDI-MS	83		0			20		n/a		

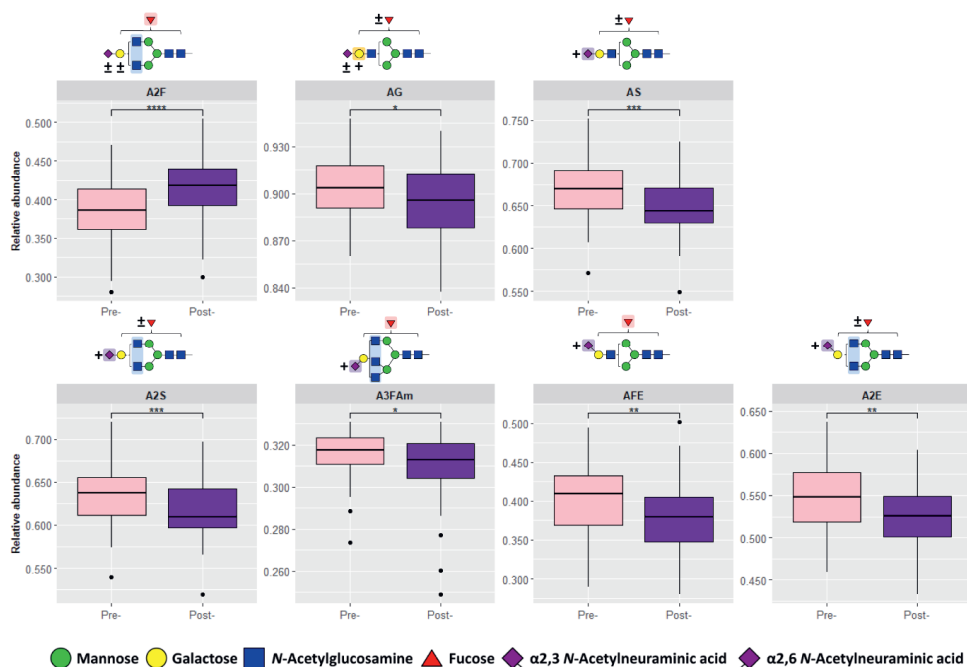
The alteration of these eight significant N-glycans in CRC is displayed in **Figure 2**, which shows a lower pre-operative expression of H4N4F1, H5N4F1, H5N4F1S<sub>2,6</sub>1, and H5N4F1S<sub>2,3</sub>2. These results are in line with previous reports that di-antennary structures with core fucosylation are decreased in CRC.<sup>14,25–27</sup> In addition, we also observed that H5N5F1 has a lower abundance in pre-operative CRC, which is supported by previous findings whereby bisected N-glycan species were also found to be decreased in serum as well as tissues from CRC patients.<sup>14,25,28</sup> In contrast, a higher abundance of larger structures including H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>1, H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>2 (isomer 1) and H6N5S<sub>2,6</sub>3 in pre-operative CRC is observed. This is in agreement with results which showed that mainly sialylated tri- and tetra-antennary N-glycans were found to be increased in CRC.<sup>14,27</sup> Importantly, the same trend in CRC is observed for these structures by the MALDI-MS approach (**Table 2**).

We investigated derived glycosylation traits that describe global features of glycosylation based on the biosynthetic pathway, using the 18 derived traits previously validated by the MALDI-MS study. In this case, we observed 14 common derived traits, of which seven were significant differentiators between pre- and post-operative samples (**Supplementary Table S7**). As shown in **Figure 3**, fucosylated di-antennary glycans (A2F) have a lower expression in pre-operative CRC which is similar to the results of some of the individual direct N-glycan traits mentioned above. In addition,  $\alpha$ 2,6-sialylated fucosylated glycans (AFE) and  $\alpha$ 2,6-sialylated di-antennary glycans (A2E) have a greater abundance in pre-operative CRC. In relation to this, it has been



**Figure 2. Significant direct N-glycan traits.** Eight significant direct N-glycan traits were found between pre-operative (pre-,  $n = 64$ ) and post-operative (post-,  $n = 64$ ) samples that were also previously validated:<sup>14</sup> H4N4F1, H5N4F1, H5N5F1, H5N4F1S<sub>2,6</sub>1, H5N4F1S<sub>2,3</sub>2, H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>1, H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>2 (isomer 1), and H6N5S<sub>2,6</sub>3. Asterisks denote significance with  $p$ -value < 0.05 (\*), 0.01 (\*\*), 0.001 (\*\*\*). Abbreviations: hexose (H), N-acetylhexosamine (N), fucose (F), amidated  $\alpha$ 2,3-linked N-acetylneuraminic acid (Am) and ethyl esterified  $\alpha$ 2,6-linked N-acetylneuraminic acid (E), isomer (iso).

reported that  $\alpha$ 2,6-sialylation is increased in CRC.<sup>29</sup> Moreover, sialylation per antenna across all glycan species (AS) and sialylated di-antennary N-glycans (A2S) are increased in pre-operative CRC, possibly due the greater abundance of galactosylation per antenna (AG) which would increase the number of available galactose moieties for sialic acid occupation. Furthermore, an increase in  $\alpha$ 2,3-sialylation is observed specifically in fucosylated tri-antennary glycans (A3FAm) in pre-operative CRC. As shown in **Supplementary Table S4**, this derived trait is composed of H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>1 and H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>2 (both isomers summed). **Supplementary Table S1** shows that antennary fucosylation was observed in the case of H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>1 and H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>2 (isomer 1). Thus, this observation supports previous findings that CRC is associated with an increase in sialyl-Lewis epitopes.<sup>30,31</sup> Moreover, Rebello *et al.* analyzed a subset of the same pre- and post-operative cohort



**Figure 3. Significant derived glycosylation traits.** Seven significant derived glycosylation traits were found between pre-operative (pre-,  $n = 64$ ) and post-operative (post-,  $n = 64$ ) samples that were also previously validated.<sup>14</sup> A2F (fucosylated diantennary glycans), AG (galactosylation per antenna), AS (sialic acid per antenna), A2S (sialic acids per diantennary glycan), A3FAm ( $\alpha$ 2,3-linked sialic acids per fucosylated triantennary glycan), AFE ( $\alpha$ 2,6-linked sialic acids per fucosylated glycan), and A2E ( $\alpha$ 2,6-linked sialic acids per diantennary glycan). Asterisks denote significance with  $p$ -value  $< 0.05$  (\*),  $0.01$  (\*\*),  $0.001$  (\*\*\*),  $0.0001$  (\*\*\*\*). Monosaccharide ranges are specified via plus-minus ( $\pm$ ; 0 – 4) and plus (+; 1 – 4) symbols. Abbreviations: antennae (A), fucose (F), galactosylation (G), *N*-acetylneuraminic acid (S), amidated  $\alpha$ 2,3-linked *N*-acetylneuraminic acid (Am) and ethyl esterified  $\alpha$ 2,6-linked *N*-acetylneuraminic acid (E).

using an antennary fucosylation MALDI-MS assay and uncovered that sialyl-Lewis X and Lewis X antigens were significantly increased in pre-operative samples.<sup>32</sup> Thus, our FD-MS results are consistent with those reported by the MALDI-MS<sup>14</sup> approach, indicating the same direction of these seven derived traits in CRC.

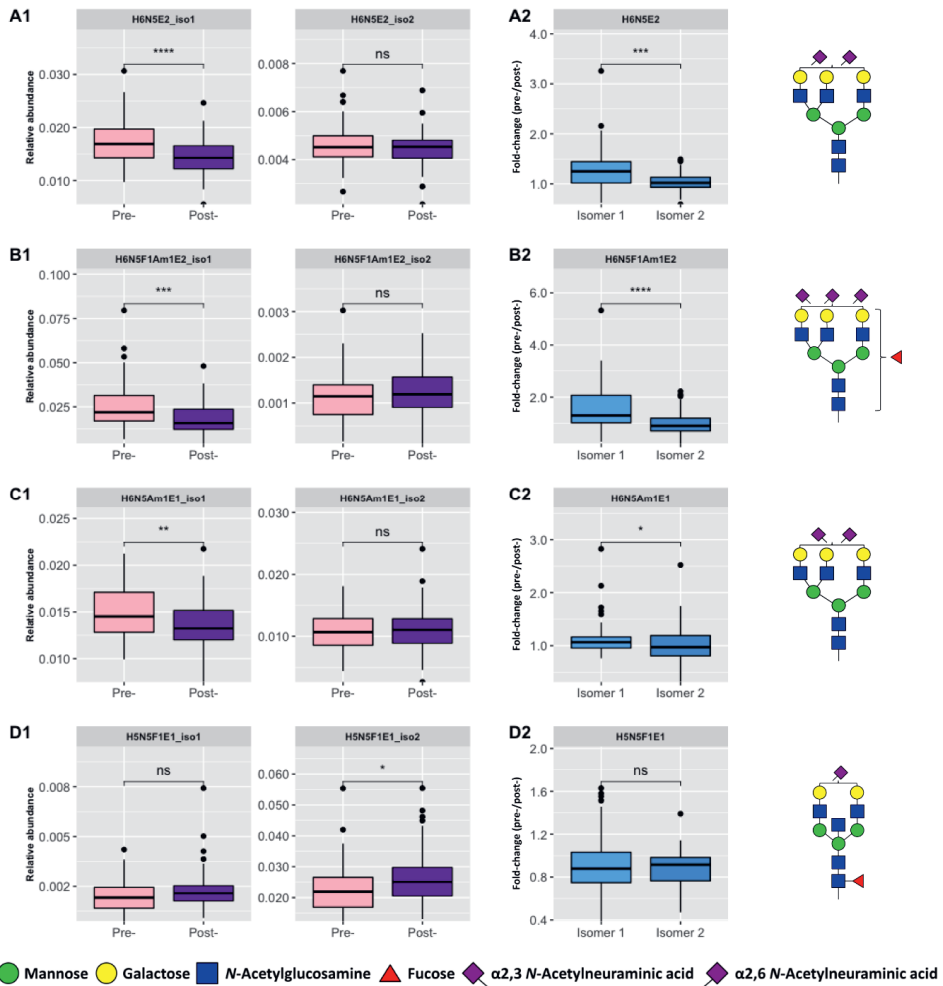
The analysis of the serum *N*-glycome in the pre- and post-operative CRC cohort by MALDI-MS previously resulted in detection of 83 *N*-glycan compositions (Table 2).<sup>14</sup> In comparison, as previously mentioned, we observed a total of 48 *N*-glycan structures in the current study. Whereas an ion trap was utilized in the current workflow it is notable that de Vroome *et al.* employed a Bruker ultrafleXtreme equipped with a TOF mass analyzer operated in reflection mode, which generally demonstrates greater



performance in areas such as mass resolution and mass accuracy.<sup>33</sup> For example, the mass resolution of low, medium and high abundant *N*-glycans measured by the ion trap MS was investigated. In this case, a mass resolution of 6500, 5290 and 6540 was obtained for H4N5 ( $m/z$  951.40), H6N5S<sub>2,3</sub>1S<sub>2,6</sub>2 (isomer 2,  $m/z$  1052.43), and H5N4S<sub>2,6</sub>2 ( $m/z$  834.01), respectively. In relation to this, the isotopic profile of H6N5S<sub>2,3</sub>1S<sub>2,6</sub>2 (isomer 2) is shown in **Supplementary Figure S1**. Takei *et al.* analyzed the serum *N*-glycome using a previous iteration of the Bruker MALDI instrument, the ultraflex III, which resulted in the quantification of 34 *N*-glycans.<sup>25</sup> Thus, it is likely the current workflow could be improved by coupling with a mass spectrometer with higher resolving power and greater sensitivity. Moreover, further improvements in sensitivity could be gained by increasing the injection amount or implementing a nanoflow column.

### Positional Isomers in CRC *N*-glycome

Our workflow demonstrated the efficient chromatographic separation of positional isomers using RPLC following fluorescent labeling and sialic acid derivatization.<sup>16</sup> This is evident in **Table 2** which shows that seven *N*-glycan compositions with the same  $m/z$  were detected as positional isomers, resulting in the assignment of 14 structures in total. We further investigated the structural components which may lead to different retention times for positional isomeric structures. To perform this, we applied our workflow in order to examine IgG *N*-glycans, for which H4N4F1 (G1F,  $m/z$  922.89) is a well-known composition that consists of two isomers.<sup>34</sup> As shown by **Supplementary Figure S2**, we observed a longer retention time for H4N4F1 with  $\alpha$ 3-antenna galactose occupation using a 70 min gradient, indicating that structural variations may lead to different retention times for certain positional isomeric *N*-glycans. Despite this, there is a complex interplay of influences on *N*-glycan retention during RPLC separation, including sialic acid-linkage dependent derivatization, galactose occupation by sialic acids, bisection, number of hexose residues, and fucosylation.<sup>16</sup> As a result, specific assignment of structures remains challenging for *N*-glycans, in particular for those with two antennae or more. Nonetheless, in the case of relevant structures related to a disease, an exoglycosidase approach could be followed in order to characterize specific isomeric features of interest. Following this, this method could then be applied in order to investigate those structures within a clinical context.



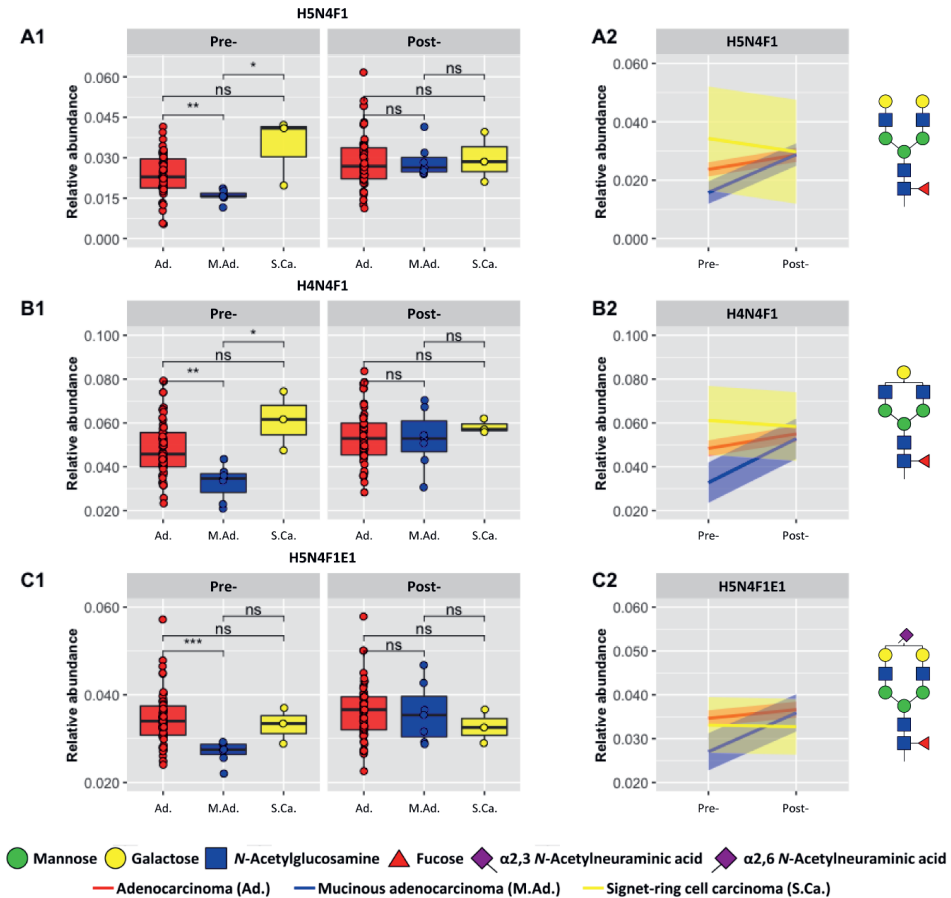
**Figure 4. Positional isomeric N-glycans that show significant differences in CRC.** Four positional isomers with significant differences were observed, as denoted by each row. **(1)** Boxplots of relative abundances observed in pre- ( $n = 64$ ) and post-operative ( $n = 64$ ) samples for the each isomeric form. **(2)** Fold-change in CRC (pre-/post-,  $n = 64$ ) for each isomer whereby an increase ( $>1$ ) or a decrease ( $<1$ ) may be observed. Asterisks denote significance with  $p$ -value  $< 0.05$  (\*),  $0.01$  (\*\*),  $0.001$  (\*\*\*),  $0.0001$  (\*\*\*\*). Abbreviations: Not significant (ns), hexose (H), N-acetylhexosamine (N), fucose (F), amidated  $\alpha 2,3$ -linked N-acetylneuraminic acid (Am) and ethyl esterified  $\alpha 2,6$ -linked N-acetylneuraminic acid (E).

In our study, we revealed that specific isomers are significantly associated with CRC. For example, **Figure 4 (A1 – D1)** showed four compositions where only one of the isomeric forms was a significant differentiator between the pre- and post-operative groups, including a previously validated composition, H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>2. Moreover, **Figure 4 (A2 – D2)** also shows that there were significant differences between the

isomers in CRC fold-change (pre-/post-) for three of these compositions. Previously we showed that antennary fucosylation results in an earlier RT than core fucosylation<sup>16</sup> and, expectedly, **Supplementary Table S5** shows that H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>2 (isomer 1) eluted earlier than H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>2 (isomer 2). In addition, the former assignment is supported by the presence of a fragment ion with  $m/z$  512.20 (**Supplementary Table S1**) which is indicative of antennary fucosylation whereas the latter showed the presence of a core fucose fragment ion ( $m/z$  587.33). In contrast, MALDI-MS may only differentiate mass differences between *N*-glycans. In this case, the location of the fucose residue on H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>2 was not specified by de Vroome *et al.*<sup>14</sup> whereas likely this composition is a mixture of core and antennary fucosylated forms when detected solely based on  $m/z$ . Evidence for fucose positional isomers has been provided by Rebello *et al.* who found a reduction in the relative abundance of H6N5F1S<sub>2,3</sub>1S<sub>2,6</sub>2 following core fucosidase digestion.<sup>32</sup> Thus, here we observed both isomer forms of this *N*-glycan and we illustrated that only the antennary fucosylated form (isomer 1) is significantly different between the pre- and post-operative groups. As a result, our workflow allows the assessment of the specific association of antennary fucosylation with CRC.

### Investigation of Pre- and Post-operative Histological Type

A Kruskal-Wallis test was performed in order to determine whether the pre-operative CRC abundances of the direct and derived *N*-glycan traits shown in **Figures 2 and 3** were significant across three categories of the disease, namely stage, pT and histological type. With regard to this, H5N4F1, H4N4F1 and H5N4F1S<sub>2,6</sub>1 demonstrated significant differences for histological type (**Figure 5**). As shown in **Figure 5 (A1 – C1)**, it can be seen that differences in the pre-operative abundances of these *N*-glycans are no longer observed following surgery (post-operative). In addition, **Figure 5 (A2 – C2)** further demonstrates that the relative abundances converge in the post-operative measurements. Thus, differences in histological type are no longer readily observed following surgery.



**Figure 5. Differences in histological type between pre- and post-operative CRC samples.** Three *N*-glycans differentiated significantly based upon histological type, adenocarcinoma ( $n = 57$ ), mucinous adenocarcinoma ( $n = 7$ ) and signet-ring cell carcinoma ( $n = 3$ ), as denoted by each row: row A (H5N4F1), row B (H4N4F1) and row C (H5N4F1S<sub>2,6</sub>1). (1) Boxplots of relative abundances observed in pre- and post-operative samples. (2) Trend observed following surgery, from pre- to post-operative samples from the same patients. The 95% CI is shown as colored bands around each line. Asterisks denote significance with  $p$ -value  $< 0.05$  (\*),  $0.01$  (\*\*),  $0.001$  (\*\*\*),  $0.0001$  (\*\*\*\*). Abbreviations: Not significant (ns), hexose (H), *N*-acetylhexosamine (N), fucose (F), amidated  $\alpha$ 2,3-linked *N*-acetylneuraminic acid (Am) and ethyl esterified  $\alpha$ 2,6-linked *N*-acetylneuraminic acid (E).

In order to further support these results, the *N*-glycans shown in **Figure 2** were investigated retrospectively in the MALDI-MS datasets, namely the pre- ( $n = 64$  pre-operative cases) and post-operative and discovery and validation studies ( $n = 185$  cases). This revealed that the same three *N*-glycans (H5N4F1, H4N4F1 and H5N4F1S<sub>2,6</sub>1) were also significant differentiators of histological type. As shown in

**Supplementary Figure S3**, a similar trend is observed whereby differences between the subcategories of histological type are no longer observed following surgery. In addition, **Supplementary Figure S4** shows that the results of H5N4F1 and H5N4F1S<sub>2,6</sub>1 were also replicated in cases from the MALDI-MS discovery and validation study. These findings are similar to the results previously demonstrated by de Vroome *et al.* whereby they showed similar *N*-glycan abundances between the post-operative samples and healthy controls,<sup>14</sup> signifying a positive response to surgery.

**Table 3. Comparison of AUCs for adenocarcinoma classification in pre-/post-operative CRC.** AUCs from the comparison of adenocarcinoma ( $n = 57$ ) versus signet-ring cell carcinoma ( $n = 3$ ) and mucinous adenocarcinoma ( $n = 7$ ) are shown. Three quantification approaches performed by this study (RPLC-FD, RPLC-MS, RPLC-FD-MS) are displayed, as well as data from a previous study performing MALDI-MS.<sup>14</sup> "Combination" refers to the combination of the significant features for histological type. All results may be found in **Supplementary Table S8**. Abbreviations: Not detected (n.d.), not applicable (n/a).

Peak	RPLC-FD	Direct <i>N</i> -glycan traits	RPLC-MS	RPLC-FD-MS	MALDI-MS
	AUC		AUC	AUC	AUC
10	0.79	H5N4F1	0.70	0.79	0.79
		H4N4F1	0.72	0.71	0.77
n.d.	n.d.	H5N4F1S <sub>2,6</sub> 1	0.70	0.80	0.76
n.d.	n.d.	Combination	0.68	0.83	0.77

The potential of the *N*-glycans H5N4F1, H4N4F1 and H5N4F1S<sub>2,6</sub>1 as prognostic markers for distinguishing between different histological types was evaluated using a receiver operating characteristic (ROC) curve. As a result of the small sample size, only the largest subcategory versus a combination of the other subcategories is reported. In the case of histological type, **Table 3** displays the classification of adenocarcinoma against signet-ring cell carcinoma and mucinous adenocarcinoma. The analysis of pre- and post-operative CRC samples by the RPLC-FD-MS approach achieved the highest AUC values (ranging from 0.71 to 0.83), and the model combining the three *N*-glycan traits showed the best performance (AUC 0.83). In comparison, the analysis of the same cohort using the MALDI-MS technique demonstrated a moderate performance (AUC 0.76 – 0.79). Although MALDI-MS showed significant differences for pT, its prognostic performance was poor (AUC 0.50 – 0.59; **Supplementary Table S8**). Similarly, the analysis of the MALDI-MS discovery and validation study (**Supplementary Table S8**) revealed significant differences yet poor performance for stratifying adenocarcinoma (AUC 0.56 – 0.60), pT3 (AUC 0.48

– 0.53), and stage 2 (AUC 0.49). The characteristics of the discovery and validation cohort may be found in **Supplementary Table S9**.

Histological type is an important prognostic factor in CRC, as different types of tumors may show different behavior and response to treatment. For example, signet-ring cell carcinoma is associated with poor differentiation (high-grade) and generally results in a worse outcome than adenocarcinomas.<sup>35–38</sup> In addition, microsatellite stable mucinous adenocarcinomas may demonstrate more aggressive behavior when detected at an advanced stage.<sup>35,39</sup> Undoubtedly, identifying the histological type of CRC can assist in guiding the treatment approach as well as predicting patient prognosis<sup>35</sup> and yet, interobserver variability can present challenges in this regard.<sup>40</sup> Thus, the results displayed by the current study demonstrate the utility of the serum *N*-glycome in aiding histological type classification. Moreover, we also illustrated that the response to surgery may be monitored by comparing pre- and post-operative *N*-glycan abundances from the same patient, thus promoting an individualized approach to treatment. This is an important result as the need and benefit of personalized medicine is continuously being recognized.<sup>41,42</sup> Despite this, it should be noted that this cohort consisted mainly of the most prominent histological type, adenocarcinoma<sup>43</sup> (**Table 1**), and as a result, it may be challenging to translate these results to subsequent studies. Nevertheless, this analysis shows the potential of the serum *N*-glycome as a prognostic marker and further validation should be performed, particularly with regard to histological type classification and response to surgery.

### Comparison of Clinical Performance of Three Quantification Approaches

In this study we employed three quantification approaches, FD, MS, and FD-MS, in order to analyze pre- and post-operative CRC. Previously we showed that FD-MS quantification, due to the detection of specific glycan masses, resulted in a greater coverage of glycan features than FD. Furthermore, as a result of incorporating the FD signal into the quantification of individual *N*-glycans, we also demonstrated that FD-MS achieved greater precision than MS quantification.<sup>16</sup> In the current work, we further compared the performance of these quantification approaches in a clinical setting. In **Supplementary Table S5**, the number of significant peaks or glycan structures is provided for the FD (14), MS (11), and FD-MS (14) approaches. In relation to this, the number of validated significant *N*-glycans for MS and FD-MS is seven and eight,

respectively. Furthermore, there were seven FD peaks (10, 11, 13, 17, 24, 25, 32) that were significant between the groups and also contained a previously validated *N*-glycan. Thus, it is shown here that the three quantification approaches perform similarly in terms of number of validated and significant features (FD peak or glycan structure). In addition, the three quantification approaches were also evaluated for their diagnostic performance over time in relation to adenocarcinoma classification via ROC curves whereby AUCs were also calculated (**Table 3**). This demonstrates a good performance for FD-MS and FD quantification which both fall into a similar AUC range (AUC 0.70 – 0.80). In comparison, MS quantification resulted in a fair performance (0.68 – 0.72). Thus, in comparison with FD, the FD-MS quantification approach showed a similar performance yet allowed the analysis of specific direct *N*-glycan traits in relation to CRC. The comparison of MS and FD-MS demonstrated that the latter showed improved diagnostic performance for adenocarcinoma classification.

Cancer-associated alterations in glycosylation have been well established in the literature.<sup>44–46</sup> Moreover, several studies have demonstrated the potential with regard to the application of glycomics for aiding and improving cancer diagnosis and prognosis.<sup>14,47,48</sup> Despite this, there is still a lack of glycomic tests that have been translated from biomedical research into clinical laboratories.<sup>49</sup> This is partly because of the challenges that carbohydrates themselves present for analytical method development, detection and quantification, as well as a lack of validation studies beyond the discovery phase.<sup>50</sup> In this study, we have demonstrated that MS-detection allows the underlying glycosylation alterations in CRC to be probed. As a result, a greater understanding of the disease may be developed whilst maintaining specificity with regard to the traits that can be targeted for diagnostic and prognostic testing. However, FD of labeled glycans shows promise for clinical development<sup>51</sup> as it may avoid common challenges with glycan analysis by MS, such as poor ionization. In this regard, we have demonstrated similar and satisfactory performance of FD and FD-MS quantification (**Table 3**). Moreover, FD-MS relative abundances more closely resemble FD relative abundances in comparison with MS-based quantification (**Supplementary Figure S5**). Thus, upon identification of the best-performing *N*-glycan traits, the FD-MS results could be more easily bridged with the FD platform, enabling efficient implementation within the clinical setting.

## Perspectives

Undoubtedly, as previously mentioned, the results presented here require further validation with a larger and more diverse cohort of CRC patients in order to evaluate the true prognostic potential of serum *N*-glycosylation. In this regard, the monitoring of patient profiles before, during and after clinical events such as treatment, therapy or surgery seems promising and should be further explored. In addition, clinical validation of FD-MS quantification and bridging to FD quantification should also be examined. In this regard, interferences that may affect MS and FD signals, such as different mobile phase conditions across the gradient as well as unidentified fluorescent compounds, should be defined and, if possible, eliminated. Overall, the presented workflow is suitable for future investigations into biomarker discovery as well as clinical validation and translation.

## Conclusions

In this study, we evaluated the clinical performance of our workflow which performs *N*-glycan fluorescent labeling and sialic acid derivatization followed by RPLC-FD-MS measurement. Under alternative ionization, detection and quantification conditions, we further supported previously validated direct *N*-glycan and derived glycosylation traits for differentiating pre-operative CRC in comparison with post-operative samples from the same patient. Furthermore, we observed that specific positional isomers were significant between the two groups which has important clinical implications in the context of this disease. In addition, pre-operative abundances of *N*-glycans showed good performance for the classification of adenocarcinoma and new clinical insights were gained as we illustrated that differences between histological types were abolished following surgery. Thus, serum *N*-glycosylation is a promising CRC prognostic marker that may be further investigated for performing disease classification and monitoring patient response to surgery.

## Supplementary Materials

**Supplementary figures (.pdf):** **Figure S1.** Representative profile of *N*-glycans measured by RPLC-FD-MS. **Figure S2.** Extracted ion chromatograms of H4N4F1 from IgG. **Figure S3.** Differences in histological type between pre- and post-operative



samples measured by MALDI-MS. **Figure S4.** Significant *N*-glycans for histological type in MALDI-MS discovery and validation study. **Figure S5.** Linear regression comparison.

**Supplementary tables (xlsx): Table S1.** Diagnostic ions obtained by RPLC-FD-MS(-MS). **Table S2.** HappyTools peak extraction and calibration parameters. **Table S3.** LaCyTools composition extraction and calibration parameters. **Table S4.** Formulas for calculating derived glycosylation traits. **Table S5.** Analysis of CRC cohort using three quantification methods. **Table S6.** Comparison of RPLC-FD-MS and MALDI-MS direct *N*-glycan trait results. **Table S7.** Comparison of RPLC-FD-MS and MALDI-MS derived glycosylation trait results. **Table S8.** Diagnostic performance of RPLC- (FD, MS, FD-MS) and MALDI-MS results for discriminating CRC disease categories. **Table S9.** Clinical characteristics of the discovery and validation CRC cohort.

## Author Contributions

Conceptualization, D.I.R.S., M.W., Y.E.M.v d B., G.S.M.L.K. and A.B.M.; methodology, A.B.M.; data curation, A.B.M.; formal analysis, A.B.M., G.E.H., Y.E.M.v d B., and G.S.M.L.K.; resources, W.E.M., R.A.E.M.T., D.I.R.S. and M.W.; writing—original draft preparation, A.B.M., G.S.M.L.K., G.E.H., Y.E.M.v d B., and M.W.; writing—review and editing, A.B.M, G.S.M.L.K., G.E.H., Y.E.M.v d B., M.W., W.E.M., R.A.E.M.T., and D.I.R.S.; visualization, A.B.M., and G.S.M.L.K.; supervision, G.S.M.L.K., Y.E.M.v d B., M.W., and D.I.R.S.; funding acquisition, G.S.M.L.K., D.I.R.S., and M.W. All authors have read and agreed to the published version of the manuscript.

## Funding

This research was funded by the European Union's Horizon 2020 Research and Innovation Program (GlySign, Grant No. 722095).

## Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki. Serum samples were collected at the Leiden University Medical Center (LUMC) Surgical Oncology Biobank between October 2002 and March 2013 according to a

standardized protocol. This study was approved by the Medical Ethics Committee of the LUMC and was performed in agreement with the Dutch law for medical research involving human subjects (see also Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands <http://www.federa.org/>).

## Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

## Data Availability Statement

The raw data is available from the authors upon request.

## Acknowledgments

The authors wish to acknowledge that Figure 1 was created using Biorender.com.

## Conflicts of Interest

The authors declare the following competing financial interest(s): Daniel I. R. Spencer and Georgia Elgood-Hunt are employees of Ludger Ltd., a company that provides commercial glycoanalytical products and services.

## References

- (1) Vajaria, B. N.; Patel, P. S. Glycosylation: A Hallmark of Cancer? *Glycoconjugate Journal*. **2017**, *34* (2). DOI: 10.1007/s10719-016-9755-2.
- (2) Silsirivanit, A. Glycosylation Markers in Cancer. In *Advances in Clinical Chemistry*; **2019**; Vol. 89. DOI: 10.1016/bs.acc.2018.12.005.
- (3) Ohtsubo, K.; Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell*. **2006**, *126* (5). DOI: 10.1016/j.cell.2006.08.019.
- (4) Munkley, J.; Elliott, D. J. Hallmarks of Glycosylation in Cancer. *Oncotarget*. **2016**, *7* (23). DOI: 10.18632/oncotarget.8155.
- (5) Pinho, S. S.; Reis, C. A. Glycosylation in Cancer: Mechanisms and Clinical Implications. *Nature Reviews Cancer*. **2015**, *15* (9). DOI: 10.1038/nrc3982.
- (6) Boyaval, F.; Dalebout, H.; Van Zeijl, R.; Wang, W.; Fariña-Sarasqueta, A.; Lageveen-Kammeijer, G. S. M.; Boonstra, J. J.; McDonnell, L. A.; Wuhler, M.; Morreau, H.; Heijs, B. High-Mannose N-Glycans as Malignant Progression Markers in Early-Stage Colorectal Cancer. *Cancers*. **2022**, *14* (6). DOI: 10.3390/cancers14061552.

- (7) Boyaval, F.; Van Zeijl, R.; Dalebout, H.; Holst, S.; Van Pelt, G.; Fariña-Sarasqueta, A.; Mesker, W.; Tollenaar, R.; Morreau, H.; Wuhrer, M.; Heijs, B. N-Glycomic Signature of Stage II Colorectal Cancer and Its Association with the Tumor Microenvironment. *Molecular and Cellular Proteomics*. **2021**, *20*. DOI: 10.1074/MCP.RA120.002215.
- (8) Drake, R. R.; Powers, T. W.; Jones, E. E.; Bruner, E.; Mehta, A. S.; Angel, P. M. MALDI Mass Spectrometry Imaging of N-Linked Glycans in Cancer Tissues. In *Advances in Cancer Research*; **2017**; Vol. 134. DOI: 10.1016/bs.acr.2016.11.009.
- (9) Houvast, R. D.; Vankemmelbeke, M.; Durrant, L. G.; Wuhrer, M.; Baart, V. M.; Kuppen, P. J. K.; de Geus-Oei, L. F.; Vahrmeijer, A. L.; Sier, C. F. M. Targeting Glycans and Heavily Glycosylated Proteins for Tumor Imaging. *Cancers*. **2020**, *12* (12). DOI: 10.3390/cancers12123870.
- (10) Holm, M.; Nummela, P.; Heiskanen, A.; Satomaa, T.; Kaprio, T.; Mustonen, H.; Ristimäki, A.; Haglund, C. N-Glycomic Profiling of Colorectal Cancer According to Tumor Stage and Location. *PLoS ONE*. **2020**, *15* (6 June). DOI: 10.1371/journal.pone.0234989.
- (11) Pan, Y.; Zhang, L.; Zhang, R.; Han, J.; Qin, W.; Gu, Y.; Sha, J.; Xu, X.; Feng, Y.; Ren, Z.; Dai, J.; Huang, B.; Ren, S.; Gu, J. Screening and Diagnosis of Colorectal Cancer and Advanced Adenoma by Bionic Glycome Method and Machine Learning. *Am J Cancer Res*. **2021**, *11* (6).
- (12) Bertok, T.; Bertokova, A.; Jane, E.; Hires, M.; Aguedo, J.; Potocarova, M.; Lukac, L.; Vikartovska, A.; Kasak, P.; Borsig, L.; Tkac, J. Identification of Whole-Serum Glycobiomarkers for Colorectal Carcinoma Using Reverse-Phase Lectin Microarray. *Frontiers in Oncology*. **2021**, *11*. DOI: 10.3389/fonc.2021.735338.
- (13) Crosby, D.; Lyons, N.; Greenwood, E.; Harrison, S.; Hiom, S.; Moffat, J.; Quallo, T.; Samuel, E.; Walker, I. A Roadmap for the Early Detection and Diagnosis of Cancer. *The Lancet Oncology*. **2020**, *21* (11). DOI: 10.1016/S1470-2045(20)30593-3.
- (14) de Vroome, S. W.; Holst, S.; Gironde, M. R.; van der Burgt, Y. E. M.; Mesker, W. E.; Tollenaar, R. A. E. M.; Wuhrer, M. Serum N-Glycome Alterations in Colorectal Cancer Associate with Survival. *Oncotarget*. **2018**, *9* (55), 30610–30623. DOI: 10.18632/oncotarget.25753.
- (15) Vreeker, G. C. M.; Nicolardi, S.; Bladergroen, M. R.; Van Der Plas, C. J.; Mesker, W. E.; Tollenaar, R. A. E. M.; Van Der Burgt, Y. E. M.; Wuhrer, M. Automated Plasma Glycomics with Linkage-Specific Sialic Acid Esterification and Ultrahigh Resolution MS. *Analytical Chemistry*. **2018**, *90* (20). DOI: 10.1021/acs.analchem.8b02391.
- (16) Moran, A. B.; Gardner, R. A.; Wuhrer, M.; Lageveen-Kammeijer, G. S. M.; Spencer, D. I. R. Sialic Acid Derivatization of Fluorescently Labeled N-Glycans Allows Linkage Differentiation by Reversed-Phase Liquid Chromatography–Fluorescence Detection–Mass Spectrometry. *Analytical Chemistry*. **2022**, *94* (18), 6639–6648. DOI: 10.1021/acs.analchem.1c02610.
- (17) Reiding, K. R.; Blank, D.; Kuijper, D. M.; Deelder, A. M.; Wuhrer, M. High-Throughput Profiling of Protein N-Glycosylation by MALDI-TOF-MS Employing Linkage-Specific Sialic Acid Esterification. *Analytical Chemistry*. **2014**, *86* (12), 5784–5793. DOI: 10.1021/ac500335t.
- (18) Lageveen-Kammeijer, G. S. M.; de Haan, N.; Mohaupt, P.; Wagt, S.; Filius, M.; Nouta, J.; Falck, D.; Wuhrer, M. Highly Sensitive CE-ESI-MS Analysis of N-Glycans from Complex

- Biological Samples. *Nature Communications*. **2019**, *10* (1), 1–8. DOI: 10.1038/s41467-019-09910-7.
- (19) Varki, A. et al. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology*. **2015**, *25* (12), 1323–1324. DOI: 10.1093/glycob/cwv091.
- (20) Jansen, B. C.; Hafkenscheid, L.; Bondt, A.; Gardner, R. A.; Hendel, J. L.; Wuhrer, M.; Spencer, D. I. R. HappyTools: A Software for High-Throughput HPLC Data Processing and Quantitation. *PLoS One*. **2018**, *13* (7), e0200280. DOI: 10.1371/journal.pone.0200280.
- (21) European Medicines Agency (EMA). *ICH Guidelines Q2(R2) on Validation of Analytical Procedures*; **2022**; Vol. 2.
- (22) Jansen, B. C.; Falck, D.; de Haan, N.; Hipgrave Ederveen, A. L.; Razdorov, G.; Lauc, G.; Wuhrer, M. LaCyTools: A Targeted Liquid Chromatography-Mass Spectrometry Data Processing Package for Relative Quantitation of Glycopeptides. *Journal of Proteome Research*. **2016**, *15* (7). DOI: 10.1021/acs.jproteome.6b00171.
- (23) Colley KJ, Varki A, K. T. Cellular Organization of Glycosylation In: *Essentials of Glycobiology*. **2017**.
- (24) Wehrens, R.; Hageman, J. A.; van Eeuwijk, F.; Kooke, R.; Flood, P. J.; Wijnker, E.; Keurentjes, J. J. B.; Lommen, A.; van Eekelen, H. D. L. M.; Hall, R. D.; Mumm, R.; de Vos, R. C. H. Improved Batch Correction in Untargeted MS-Based Metabolomics. *Metabolomics*. **2016**, *12* (5). DOI: 10.1007/s11306-016-1015-8.
- (25) Takei, D.; Harada, K.; Nouse, K.; Miyahara, K.; Dohi, C.; Matsushita, H.; Kinugasa, H.; Hiraoka, S.; Nishimura, S. I.; Okada, H. Clinical Utility of a Serum Glycome Analysis in Patients with Colorectal Cancer. *Journal of Gastroenterology and Hepatology (Australia)*. **2022**, *37* (4). DOI: 10.1111/jgh.15781.
- (26) Zhao, Y. P.; Ruan, C. P.; Wang, H.; Hu, Z. Q.; Fang, M.; Gu, X.; Ji, J.; Zhao, J. Y.; Gao, C. F. Identification and Assessment of New Biomarkers for Colorectal Cancer with Serum N-Glycan Profiling. *Cancer*. **2012**, *118* (3). DOI: 10.1002/cncr.26342.
- (27) Doherty, M.; Theodoratou, E.; Walsh, I.; Adamczyk, B.; Stöckmann, H.; Agakov, F.; Timofeeva, M.; Trbojević-Akmačić, I.; Vučković, F.; Duffy, F.; McManus, C. A.; Farrington, S. M.; Dunlop, M. G.; Perola, M.; Lauc, G.; Campbell, H.; Rudd, P. M. Plasma N-Glycans in Colorectal Cancer Risk. *Scientific Reports*. **2018**, *8* (1). DOI: 10.1038/s41598-018-26805-7.
- (28) Balog, C. I. A.; Stavenhagen, K.; Fung, W. L. J.; Koeleman, C. A.; McDonnell, L. A.; Verhoeven, A.; Mesker, W. E.; Tollenaar, R. A. E. M.; Deelder, A. M.; Wuhrer, M. N-Glycosylation of Colorectal Cancer Tissues. *Molecular & Cellular Proteomics*. **2012**, *11* (9). DOI: 10.1074/mcp.m111.011601.
- (29) Park, J. J.; Lee, M. Increasing the  $\alpha$  2, 6 Sialylation of Glycoproteins May Contribute to Metastatic Spread and Therapeutic Resistance in Colorectal Cancer. *Gut and Liver*. **2013**, *7* (6). DOI: 10.5009/gnl.2013.7.6.629.
- (30) Nakagoe, T.; Sawai, T.; Tsuji, T.; Jibiki, M. A.; Nanashima, A.; Yamaguchi, H.; Yasutake, T.; Kurosaki, N.; Ayabe, H.; Arisawa, K. Preoperative Serum Levels of Sialyl Lewis(a), Sialyl Lewis(x), and Sialyl Tn Antigens as Prognostic Markers after Curative Resection for Colorectal Cancer. *Cancer detection and prevention*. **2001**, *25* (3).

- (31) Miyoshi, E.; Moriwaki, K.; Nakagawa, T. Biological Function of Fucosylation in Cancer Biology. *Journal of Biochemistry*. **2008**, *143* (6). DOI: 10.1093/jb/mvn011.
- (32) Rebello, O. D.; Nicolardi, S.; Lageveen-Kammeijer, G. S. M.; Nouta, J.; Gardner, R. A.; Mesker, W. E.; Tollenaar, R. A. E. M.; Spencer, D. I. R.; Wuhrer, M.; Falck, D. A Matrix-Assisted Laser Desorption/Ionization—Mass Spectrometry Assay for the Relative Quantitation of Antennary Fucosylated N-Glycans in Human Plasma. *Frontiers in Chemistry*. **2020**, *8*. DOI: 10.3389/fchem.2020.00138.
- (33) de Hoffmann, E.; Stroobant, V. *Mass Spectrometry: Principles and Applications*, 3rd ed.; Wiley: Chichester, **2007**.
- (34) Keser, T.; Pavić, T.; Lauc, G.; Gornik, O. Comparison of 2-Aminobenzamide, Procainamide and RapiFluor-MS as Derivatizing Agents for High-Throughput HILIC-UPLC-FLR-MS N-Glycan Analysis. *Frontiers in Chemistry*. **2018**, *6*. DOI: 10.3389/fchem.2018.00324.
- (35) Fleming, M.; Ravula, S.; Tatishchev, S. F.; Wang, H. L. Colorectal Carcinoma: Pathologic Aspects. *Journal of Gastrointestinal Oncology*. **2012**, *3* (3). DOI: 10.3978/j.issn.2078-6891.2012.030.
- (36) Makino, T.; Tsujinaka, T.; Mishima, H.; Ikenaga, M.; Sawamura, T.; Nakamori, S.; Fujitani, K.; Hirao, M.; Kashiwazaki, M.; Masuda, N.; Takeda, M.; Mano, M. Primary Signet-Ring Cell Carcinoma of the Colon and Rectum: Report of Eight Cases and Review of 154 Japanese Cases. *Hepato-Gastroenterology*. **2006**, *53* (72).
- (37) Chen, J. S.; Hsieh, P. S.; Chiang, J. M.; Yeh, C. Y.; Tsai, W. S.; Tang, R.; Changchien, C. R.; Wu, R. C. Clinical Outcome of Signet Ring Cell Carcinoma and Mucinous Adenocarcinoma of the Colon. *Chang Gung Medical Journal*. **2010**, *33* (1).
- (38) Kang, H.; O'Connell, J. B.; Maggard, M. A.; Sack, J.; Ko, C. Y. A 10-Year Outcomes Evaluation of Mucinous and Signet-Ring Cell Carcinoma of the Colon and Rectum. *Diseases of the Colon and Rectum*. **2005**, *48* (6). DOI: 10.1007/s10350-004-0932-1.
- (39) Leopoldo, S.; Lorena, B.; Cinzia, A.; Gabriella, D. C.; Angela Luciana, B.; Renato, C.; Antonio, M.; Carlo, S.; Cristina, P.; Stefano, C.; Maurizio, T.; Luigi, R.; Cesare, B. Two Subtypes of Mucinous Adenocarcinoma of the Colorectum: Clinicopathological and Genetic Features. *Annals of Surgical Oncology*. **2008**, *15* (5). DOI: 10.1245/s10434-007-9757-1.
- (40) Compton, C. C.; Fielding, L. P.; Burgart, L. J.; Conley, B.; Cooper, H. S.; Hamilton, S. R.; Hammond, M. E. H.; Henson, D. E.; Hutter, R. V. P.; Nagle, R. B.; Nielsen, M. L.; Sargent, D. J.; Taylor, C. R.; Welton, M.; Willett, C. Prognostic Factors in Colorectal Cancer: College of American Pathologists Consensus Statement 1999. In *Archives of Pathology and Laboratory Medicine*; **2000**; Vol. 124.
- (41) Hanna-Sawires, R. G.; Schiphuis, J. H.; Wuhrer, M.; Vasen, H. F. A.; van Leerdam, M. E.; Bonsing, B. A.; Mesker, W. E.; van der Burgt, Y. E. M.; Tollenaar, R. A. E. M. Clinical Perspective on Proteomic and Glycomic Biomarkers for Diagnosis, Prognosis, and Prediction of Pancreatic Cancer. *International Journal of Molecular Sciences*. **2021**, *22* (5). DOI: 10.3390/ijms22052655.
- (42) Jackson, S. E.; Chester, J. D. Personalised Cancer Medicine. *International Journal of Cancer*. **2015**, *137* (2). DOI: 10.1002/ijc.28940.

- (43) Stewart, S. L.; Wike, J. M.; Kato, I.; Lewis, D. R.; Michaud, F. A Population-Based Study of Colorectal Cancer Histology in the United States, 1998-2001. *Cancer*. **2006**, *107* (SUPPL.). DOI: 10.1002/cncr.22010.
- (44) Mereiter, S.; Balmaña, M.; Campos, D.; Gomes, J.; Reis, C. A. Glycosylation in the Era of Cancer-Targeted Therapy: Where Are We Heading? *Cancer Cell*. **2019**, *36* (1). DOI: 10.1016/j.ccell.2019.06.006.
- (45) Lau, K. S.; Partridge, E. A.; Grigorian, A.; Silvescu, C. I.; Reinhold, V. N.; Demetriou, M.; Dennis, J. W. Complex N-Glycan Number and Degree of Branching Cooperate to Regulate Cell Proliferation and Differentiation. *Cell*. **2007**, *129* (1). DOI: 10.1016/j.cell.2007.01.049.
- (46) Hang, Q.; Isaji, T.; Hou, S.; Zhou, Y.; Fukuda, T.; Gu, J. N-Glycosylation of Integrin A5 Acts as a Switch for EGFR-Mediated Complex Formation of Integrin A5 $\beta$ 1 to A6 $\beta$ 4. *Scientific Reports*. **2016**, *6*. DOI: 10.1038/srep33507.
- (47) Yoneyama, T.; Ohyama, C.; Hatakeyama, S.; Narita, S.; Habuchi, T.; Koie, T.; Mori, K.; Hidari, K. I. P. J.; Yamaguchi, M.; Suzuki, T.; Tobisawa, Y. Measurement of Aberrant Glycosylation of Prostate Specific Antigen Can Improve Specificity in Early Detection of Prostate Cancer. *Biochemical and Biophysical Research Communications*. **2014**. DOI: 10.1016/j.bbrc.2014.04.107.
- (48) Vreeker, G. C. M.; Hanna-Sawires, R. G.; Mohammed, Y.; Bladergroen, M. R.; Nicolardi, S.; Dotz, V.; Nouta, J.; Bonsing, B. A.; Mesker, W. E.; van der Burgt, Y. E. M.; Wuhrer, M.; Tollenaar, R. A. E. M. Serum N-Glycome Analysis Reveals Pancreatic Cancer Disease Signatures. *Cancer Medicine*. **2020**, *9* (22). DOI: 10.1002/cam4.3439.
- (49) Almeida, A.; Kolarich, D. The Promise of Protein Glycosylation for Personalised Medicine. *Biochimica et Biophysica Acta - General Subjects*. **2016**, *1860* (8). DOI: 10.1016/j.bbagen.2016.03.012.
- (50) de Haan, N.; Wuhrer, M.; Ruhaak, L. R. Mass Spectrometry in Clinical Glycomics: The Path from Biomarker Identification to Clinical Implementation. *Clinical Mass Spectrometry*. **2020**, *18*. DOI: 10.1016/j.clinms.2020.08.001.
- (51) Colhoun, H. O.; Treacy, E. P.; MacMahon, M.; Rudd, P. M.; Fitzgibbon, M.; O'Flaherty, R.; Stepien, K. M. Validation of an Automated Ultraperformance Liquid Chromatography IgG N-Glycan Analytical Method Applicable to Classical Galactosaemia. *Annals of Clinical Biochemistry*. **2018**, *55* (5). DOI: 10.1177/0004563218762957.



# Chapter 6

Discussion and Perspectives



## Discussion and Perspectives

Glycosylation is highly amenable to surrounding influences in both healthy and disease microenvironments. As a result, this protein modification has great potential for the discovery and development of biomarkers. Thus, this thesis aimed to uncover unique glyco(proteo)mic signatures in prostate (PCa) and colorectal (CRC) cancer via the respective development of intact protein and released glycan MS workflows. Several advancements were made with regard to the methodology and new insights were gained regarding the respective samples. Nevertheless, several technical challenges were encountered for the developed platforms. In this chapter, potential improvements are discussed as well as future directions and necessary methodological developments for the described biomarkers in order to facilitate translation of the results beyond the discovery phase.

## Sialylated and Fucosylated Isomeric Glycans

Sialylation and fucosylation are important contributors to the development of cancer<sup>1</sup> and isomers of these glycosylation features play different roles during normal and metastatic cellular function. In **Chapter 4**, we demonstrated a method that allowed efficient isomer separation (sialylation and fucosylation) as well as mass differentiation (sialylation) using a C18 column on a RPLC-MS system. In this regard, linkage-specific derivatization of sialic acids was performed whereby  $\alpha$ 2,3- and  $\alpha$ 2,6-sialic acids underwent amidation ( $-0.98$  Da) and ethyl esterification ( $+28.02$  Da), respectively. Interestingly, it was observed that structures with an ethyl esterified  $\alpha$ 2,6-sialic acid showed a longer retention time than their amidated  $\alpha$ 2,3-sialic acid counterpart. This is likely due to the greater capacity of the amide group to form hydrogen bonds with water molecules<sup>2</sup> than the ester group, thus increasing its hydrophilicity.<sup>3-5</sup> In addition, the ester group increases the hydrophobicity of the molecule due to the propensity of its nonpolar carbonyl group to form nonpolar interactions. Additionally, fucoses are relatively apolar in comparison with other monosaccharides. Thus, a core fucose, which is located more closely to the fluorescent label than the antennary fucose, may affect the interaction of the glycan with the stationary phase, resulting in a longer retention time.<sup>6</sup> Importantly, the clinical relevance of isomer differentiation was highlighted in **Chapter 5** whereby significant differences in the relative abundances of

sialic acid-linkage and fucosylated isomers were observed between pre- and post-operative CRC patients.

There have been continuous technological developments which have also demonstrated efficient differentiation of sialylated and fucosylated isomeric species. For example, ion-mobility MS (IMS) has been applied to analyze released *N*-glycans.<sup>7-10</sup> In the study by Manz *et al.*, it was demonstrated that a fragment ion corresponding to the sialylated antenna from two chromatographic peaks containing H5N4S2 showed different drift times.<sup>10</sup> Consequently, the ratio of  $\alpha$ 2,3- to  $\alpha$ 2,6-linkage could be determined based on the extracted ion mobilograms. Furthermore, in addition to the MS/MS spectra, antennary fucosylated fragment ions were also confirmed using ion-mobility.<sup>10</sup> Those results were illustrated using HILIC yet, undoubtedly, it would be interesting to hyphenate IMS with the RPLC platform presented in **Chapter 4** in order to assess isomers with different branching positions. Notably, sialic acid derivatization is still required on this platform as we demonstrated that the derivatization procedure is important in order enable retention of these sugar residues using a C18 column. In addition, sialic acids are known to be very labile monosaccharides and during ionization sialic acid loss can be observed.<sup>11</sup> In this case, derivatization has been shown to have an important stabilizing effect on sialic acids.<sup>12,13</sup>

Another interesting approach is the investigation of ion ratios in order to identify isomers containing antenna with specifically linked sialic acids.<sup>14</sup> Wagt *et al.* built on the work of previous studies<sup>8,15</sup> with their analysis of H5N4S<sub>2,3</sub>1S<sub>2,6</sub>1 isomers in the total plasma *N*-glycome and fetuin. They observed that the ratio of specific Y-ions indicated a more prominent loss of the 3'-linked antenna compared to the 6'-linked antenna following fragmentation, regardless of the sialic acid linkage. Similarly, recent work by Maliepaard and co-workers reported that sialic acid linkages as well as 3'- and 6'-linked antennae could be determined based on the relative intensities of glycan oxonium ions in isomerically defined glycopeptide standards. In this case, the authors also noticed that 3'-linked saccharides already fragmented at lower collision energies than their 6'-linked counterpart.<sup>16</sup> Interestingly, we also observed different fragment Y-ion ratios of sialylated tri-antennary glycans in **Chapter 4**. Although this phenomenon was not explored by our study, the aforementioned findings suggest that the fragment ion ratios could be further investigated in order to examine antenna occupation by specifically-linked sialic acids.<sup>14,16</sup>

## Biomarker Discovery: Technical Challenges

To define the requirements for each stage of disease biomarker research and development, a set of guidelines have been developed for proteomic methods.<sup>17</sup> These tiers range from discovery (3) via validation (2) to clinical application (1). As a result, the field of glycomics can make use of such guidelines for the development of glycosylation-focused biomarkers. With regard to biomarker discovery, the requirements for tier 3 methods are less stringent and important parameters include the specificity of analyte identification as well as the repeatability of the measurement.<sup>17</sup> Naturally, development towards a tier 2 method requires even greater performance in these two areas as well as others, such as measurement precision. Finally, tier 1 methods require the highest degree of analytical validation, as well as reference standards, precision, and quantitative results. In relation to this, labeled internal standards are suggested for both tier 1 and tier 2 assays. However, it should also be noted that glycomics faces specific challenges, which hinders the field from meeting some of these criteria. For example, the development of isotopically-labeled internal standards for glycomics is problematic owing to the non-template driven biosynthesis of carbohydrates.<sup>18</sup> In this case, it may only be possible to produce specific isotope-labeled glycan standards which would present challenges in the case of a multipanel biomarker test. In addition, the sensitivity of the measurement may also affect meeting the aforementioned requirements. In this case, sufficient sensitivity must be achieved during an analytical measurement in order to reach an adequate signal-to-noise (S/N) ratio for quantification.<sup>19</sup> This ensures that the analyte signal may be differentiated from the surrounding noise which often results in more confident identifications via improved mass accuracy and isotopic pattern, as well as a more reliable quantification.

### Sensitivity

Sensitivity during denaturing intact protein MS analysis may be compromised as proteins become multiply charged during ESI, resulting in an analyte signal that is diluted over the protein charge envelope. Any inhibition to sensitivity can be problematic particularly in cases whereby the proteoform of interest has a low abundance. This was observed in **Chapter 2** as low abundant cleaved PSA proteoforms were not identified in some patients, possibly due to a lower amount of

captured prostate-specific antigen (PSA). This created difficulties in the comparisons between patients as the relative abundances were skewed when total area normalization was applied (data not shown). In order to overcome these challenges, there are several approaches that could be followed. As mentioned in **Chapter 2**, the analysis of urine following digital rectal examination is expected to contain a higher PSA concentration.<sup>20</sup> Furthermore, online concentration techniques in the field of CE could also be applied, including sample stacking and transient isotachopheresis.<sup>21</sup> Despite this, the distortion of the relative abundances highlights an important limitation with regard to performing relative quantification that could also present further challenges during subsequent development and translation of the method.<sup>22</sup> We also investigated normalization based upon the most abundant peak (data not shown), however this resulted in greater relative standard deviations. In a clinical context, absolute quantification using an internal standard is usually the preferred approach.<sup>18,23</sup> This was achieved by Jian *et al.* when performing intact protein analysis of a recombinant human monoclonal antibody (mAb) by using a stable-isotope labeled analogue.<sup>24</sup> In this case, the analogue protein contains amino acids (arginine and lysine) that are heavy isotope labeled (<sup>13</sup>C<sub>6</sub> and <sup>15</sup>N<sub>4</sub>). The analogue protein was spiked at known concentrations into the non-labeled sample prior to sample preparation and the resulting deconvoluted spectra shows the masses of both the non-labeled mAb and the internal standard. Thus, by comparing the peak intensity ratio of the most abundant peak from the non-labeled mAb and its analogue, absolute quantification was performed using the calibration curve that was constructed across the concentration range of the spiked internal standard. Importantly, the authors described that the variability during sample preparation and MS measurement is compensated by the internal standard. In order to apply this approach to our workflow (**Chapters 2 and 3**), a PSA standard could be cultured and extracted from an isotope label-enriched medium. Ultimately, the choice of quantification approach must be fit-for-purpose according to the stage of research that is being conducted and, most importantly, must be suitable for analyzing patient samples when conducting clinical research.

The released *N*-glycan RPLC-FD-MS workflow demonstrated in **Chapters 4 and 5** showed a lower number of *N*-glycan assignments and, subsequently, poorer sensitivity in comparison with CE-ESI-<sup>25</sup> and MALDI-MS<sup>26</sup> approaches. In this sense, coupling

the RPLC system with a high-performing MS would likely result in a greater number of *N*-glycan identifications. Furthermore, one of the advantages of RPLC is the potential to use nanoflow columns which would also increase the sensitivity. However, it should be noted that current fluorometers are not capable of performing detection at such low flow rates, thus this would greatly inhibit the advancements that were achieved with regard to the FD-MS quantification approach. A straightforward approach to further improve the sensitivity of this method would be by increasing the amount of *N*-glycans injected onto the column. For example, the developed workflow utilizes only 0.6  $\mu\text{L}$  of plasma for sialic acid derivatization and 57 nL of plasma released *N*-glycans are injected onto the column, equivalent to 1% of the starting material. Despite this, it is challenging to use a volume more than 3  $\mu\text{L}$ , containing 0.6  $\mu\text{L}$  of plasma, for the derivatization reaction as an increase in the sample volume requires a proportional increase in the other reaction components, including subsequent  $\text{CH}_3\text{CN}$  addition. Importantly,  $\text{CH}_3\text{CN}$  must reach a suitable concentration so that the glycans may bind to the HILIC plate during clean-up. In **Chapter 4**, this limit was already reached as the reaction volume (75  $\mu\text{L}$ ) was quenched with the addition of  $\text{CH}_3\text{CN}$  (225  $\mu\text{L}$ , 87.5%), resulting in a final volume (300  $\mu\text{L}$ ) that was at the maximum of the well capacity.

The use of glycan-specific beads during sample preparation would be an interesting approach to overcome volume constraints as well as multiple vacuum centrifuge steps. In addition, this method would theoretically allow the entire amount of plasma-released *N*-glycans to undergo fluorescent labeling and sialic acid derivatization. For example, Váradi *et al.* demonstrated a carboxyl coated magnetic beads protocol for efficient clean-up following PNGase F release and fluorescent labeling.<sup>27</sup> However, complications may arise during the sialic acid derivatization reaction due to the reactivity of the carboxyl groups on the beads with EDC and HOBt, similar to the derivatization reaction itself.<sup>28</sup> Another study reported by Nishikaze *et al.* illustrated a linkage-specific sialic acid derivatization approach using hydrazide beads which are chemoselective for aldehyde groups on the glycan reducing-end.<sup>29</sup> Interestingly, sialic acid derivatized *N*-glycans were eluted in water and reduced to dryness using a vacuum centrifuge which was followed by reducing-end labeling with 2-aminobenzoic acid (2-AA). Other studies reported direct labeling of the glycan reducing-end following release from the beads.<sup>30,31</sup> Therefore, it would be interesting to evaluate a procedure that allows on-bead linkage-specific sialic acid derivatization followed directly by

fluorescent labeling in order to achieve a true “one-pot” protocol. Overall, these modifications to the protocol would improve both the sensitivity and throughput, both of which are important parameters for further clinical translation.

## Future Directions for Glyco(proteo)mic Biomarkers

### Prostate-Specific Antigen

The PSA test measures the concentration of total PSA (tPSA) in serum, however tPSA consists of various proteoforms, each of which may be associated with different prostate diseases, including PCa.<sup>32</sup> With regard to other biological matrices, such as urine, less is known regarding the correlation of PSA proteoforms with PCa and other prostate conditions. For example, cleaved PSA proteoforms have not been described in previous reviews regarding urinary markers for PCa.<sup>33–36</sup> In addition, reports on urinary PSA mainly focus on variations of tPSA concentrations in urine and serum,<sup>37</sup> rather than assessing specific proteoforms of this protein. As previously mentioned, PSA glycosylation traits have been cited as a potential target for improving PCa diagnosis. However, a recent study assessing PSA glycosylation in urine did not show any differentiation between PCa and non-PCa groups.<sup>38</sup> As a result, the validity of urinary PSA proteoforms for patient stratification must still be investigated. For example, benign PSA (bPSA) proteoforms are associated with the development of benign prostate hyperplasia (BPH)<sup>39</sup> and, in this case, **Chapters 2** and **3** showed promising results as bPSA was detected in the urinary PSA profile. Interestingly, the clinical utility of bPSA was examined previously in sera using an immunoassay containing antibodies specific for cleaved bPSA (single Lys<sub>206</sub> or double Lys<sub>169</sub>, Lys<sub>206</sub> cleavage) whereby it was reported that bPSA was significantly increased in BPH.<sup>40,41</sup> However, a small sample size was used in **Chapter 2** and, therefore, further investigations on urinary PSA proteoforms should be performed as part of a larger study. The analysis of a larger cohort may then be combined with the data processing workflow presented in **Chapter 3** in order to derive clinically significant results.

Despite this, it is expected that BPH exists frequently in older men regardless of whether cancer is present<sup>40</sup> and, therefore, bPSA should be measured alongside other cancer-specific forms of PSA, such as proPSA.<sup>42</sup> It should be noted that PSA proteoforms such as complexed PSA and proPSA were not detected with the setup

used in **Chapter 2**. A possible explanation could be that complexed PSA is found in serum when mainly protease inhibitor alpha 1-antichymotrypsin forms a complex with intact active PSA.<sup>43</sup> Alpha 1-antichymotrypsin is a serum glycoprotein<sup>44</sup> and thus it may be less likely that it is present in urine in order to form complexes with PSA. With regard to proPSA proteoforms, it is possible that the sensitivity of the developed method in **Chapter 2** is currently not high enough in order to detect intact proPSA. Kitata *et al.* demonstrated that proPSA is present in urine, albeit at approximately 1% of the active PSA concentration.<sup>45</sup> Interestingly, that study also analyzed a small cohort of PCa patients and showed that a panel consisting of a combination of proPSA forms could distinguish PCa patients from healthy controls.<sup>45</sup> In addition, the authors reported that Lys-C digestion resulted in longer proPSA peptides (9 – 16 amino acids) which is more suitable for MS analysis using selected-reaction monitoring.<sup>45</sup> Thus, this digestion strategy could also be applied to our workflow (**Chapter 2**) in order to investigate whether the antibodies used in the immunocapture protocol are also specific for proPSA.

### **Total Serum N-glycosylation**

The potential of the total plasma/serum *N*-glycome as a cancer biomarker has been evaluated across multiple conditions, including investigations into lung,<sup>46</sup> breast,<sup>47,48</sup> stomach,<sup>49,50</sup> pancreatic,<sup>51</sup> prostate,<sup>52</sup> and colorectal<sup>26</sup> cancer.<sup>53</sup> Furthermore, there are several reports which perform characterization of the plasma/serum *N*-glycome, thus providing further in-depth information in relation to the sample.<sup>14,54,55</sup> Despite these promising results, it has been described that serum *N*-glycome profiles display large inter-individual differences<sup>56</sup> and are influenced by several factors including age,<sup>57</sup> sex,<sup>58</sup> and lifestyle.<sup>59</sup> For example, in **Chapter 5** we observed significant differences in the pre-operative relative abundances of specific *N*-glycans amongst different histological types in CRC. However, it could be argued which differences attribute to the progression of the disease in comparison with those differences that are already present between individuals in the general population. Thus, it is not yet clear to what extent inter-individual differences may confound the results and how such measurements can be integrated into current clinical pathways. In order to overcome this challenge, Hennig and co-workers describe that longitudinal analysis of the human plasma/serum *N*-glycome may be the best course to follow.<sup>53</sup> In this regard, Hennig *et al.* demonstrated that a significant difference in an individuals' plasma *N*-

glycome in relation to external events such as injury or illness could be observed over time.<sup>53</sup> Furthermore, the authors also showed that significant changes in one individuals' profile over time due to such events were not observed when compared with the profile of all healthy volunteers,<sup>53</sup> suggesting that the specificity of *N*-glycan biomarkers found in plasma and serum may be hindered in studies that measure only a single timepoint. Interestingly, our study in **Chapter 5** illustrated that pre-surgery differences in the relative abundances of specific *N*-glycans between histological types were eradicated when surgery was performed. In this sense, it would be interesting to further evaluate the change over time of the serum *N*-glycome as this could reveal the relative impact of events such as disease or surgery, which might otherwise be hidden due to already prevalent inter-individual differences. In this manner, a personal serum *N*-glycome “fingerprint” may be developed and used to monitor an individuals' response to internal or external events, thus further developing a personalized approach to medicine.

## Perspectives on Biomarker and Method Translation

A disparity exists between the large number of potential biomarkers that have been reported by “omics” techniques and the small proportion that find their way into clinical implementation.<sup>60</sup> This is partially due to poor study design which is reflected by the lack of reproducibility of results between studies.<sup>60,61</sup> Furthermore, with regard to the development of MS-based assays, the aforementioned guidelines<sup>17</sup> do not encompass some critical aspects of clinical research such as clinical validity and effectiveness.<sup>62</sup> In this sense, it is imperative to define the unmet clinical need early on in the development process so that outcomes related to clinical validity and effectiveness may also be defined.<sup>63</sup> In comparison, non-MS based tests have had relatively greater success with regard to clinical implementation. Immunoassays, such as enzyme-linked immunosorbent assays (ELISA), are the preferred technique in clinical laboratories<sup>64–66</sup> mainly due to their high sensitivity, throughput, speed, and they allow absolute quantification of the proteins of interest, provided an antibody is available. With regard to glycomics, a plate-based assay approach may also facilitate easier implementation within the clinics and various techniques including immunoassays,<sup>67</sup> lectin-<sup>68</sup> and glycosidase-based<sup>69</sup> assays are being investigated.



Although plate-based assays such as immunoassays are more established in the clinics, there are also significant challenges associated with the technique. For example, there is large variability between platforms, selectivity may be hampered by autoantibodies, and sensitivity is susceptible to any change that may affect protein-antibody binding.<sup>18</sup> Moreover, proteoform heterogeneity is generally not distinguished by antibodies.<sup>70</sup> As a result, method transfer between immunoassay and MS techniques can take considerable time and effort, possibly due to the lack of specificity in which the former detects proteoforms.<sup>70</sup> Such challenges may be overcome via the application of techniques with greater specificity that are capable of detection without antibodies. In this regard, online separation techniques, such as CE and LC, with optical detection show promise in relation to biomarker translation. Importantly, the biomarker profile should be previously characterized using MS. In any case, some advantages offered by optical detection techniques include the generation of readouts that are easier to interpret than mass spectra, absolute quantification in combination with a standard curve, and lower training requirements for operators. These technologies have already been implemented in order to measure hemoglobin proteins in diabetic<sup>71,72</sup> as well as thalassaemic<sup>73</sup> disorders, and glycomic applications are also under development.<sup>74,75</sup> In this sense, an approach incorporating optical detection could be taken in order to further translate the methods and results presented in this thesis.

As previously mentioned, the feasibility of patient differentiation using urinary PSA proteoforms (**Chapter 2**) still needs to be assessed. Based on the advancement of the data processing workflow that was achieved in **Chapter 3**, the method is now set for investigating a larger cohort of PCa patients. In this case, MS detection and quantification is still recommended as specificity with regard to proteoform detection and their association with various PCa groups must be examined. However, further translation of the method could be considered by transferring the technique onto a capillary zone electrophoresis-ultraviolet light (CZE-UV) platform, as previously mentioned. Farina-Gomez *et al.* analyzed standard seminal PSA and demonstrated the separation of ten electrophoretic peaks using CZE-UV in normal polarity mode with an average RSD of 3.7%.<sup>76</sup> However, in that study the authors did not report the analysis of any patient samples nor directly characterize the proteoforms under each peak. Thus, cross-validation between the results of the two platforms would be

required in order to ensure that electrophoretic peaks detected by UV are the same as the proteoforms determined by MS. Interestingly, a similar approach was followed in **Chapter 5** whereby online LC separation was followed by MS detection and it was observed that the FD-MS quantification of *N*-glycans closely resembled the FD quantification of chromatographic peaks. Thus, the chromatographic peaks could be used as a surrogate signal for the *N*-glycan structures during further translation of the method. Similarly, fluorescent *N*-glycan signals have been used to determine the association between serum *N*-glycans and inflammation and biological age<sup>77</sup> as well as treatment escalation in inflammatory bowel disease.<sup>78</sup> Notably, the findings from those studies have been translated into commercially available products.<sup>79,80</sup>

Nevertheless, MS remains the 'holy grail' of biomarker detection and quantification, owing to the unparalleled levels of sensitivity and specificity that may be achieved with this technique. In addition, MS allows multiplexed quantification of various proteins and isoform differentiation.<sup>70,81</sup> In this regard, multiple reaction monitoring (MRM) approaches on triple quadrupole MS-systems are the most well-known examples of MS clinical implementation.<sup>18,81–84</sup> Furthermore, the application of intact transferrin glycoprofiling for the detection and differentiation of congenital disorders of glycosylation has also been shown.<sup>18,85</sup> Despite these examples, MS techniques have not generally entered into routine application within clinical practice.<sup>86</sup> To summarize, there are still significant challenges to be expected when implementing glycomics as well as MS in the clinical laboratory. These challenges, as outlined above, include the absence of internal standards for glycomics, poor study design, and a lack of demonstrable clinical validity and effectiveness. In relation to glycomics standards, a possible solution could be the synthesis of standards where a panel of only a few biomarkers are being used. Although in the case of the total serum *N*-glycome, which reports on global features of glycosylation, this would hardly be conceivable. Furthermore, other hurdles include test result standardization,<sup>83,87</sup> as well as maintaining overall instrument performance.<sup>83</sup> However, in regard to the latter, newly released instruments are focusing more on instrument calibration strategies, user-friendly interfaces, and automatable solutions for data processing and reporting whilst maintaining a high level of sensitivity and resolution.<sup>88</sup> These developments should improve consistent instrument performance and uptake amongst non-specialist operators.

Overall, the full potential of MS in clinical laboratories has yet to be fully realized. To achieve this, continued communication and collaboration among key stakeholders, such as clinical and analytical chemists, clinicians, as well as instrument vendors, is necessary. Ultimately the aim must be the translation of the biomarker and not necessarily the analytical method itself. Thus, the method that was involved in the discovery of the biomarker may not be the same method that is implemented for routine clinical measurement. Nevertheless, transfer between different techniques can take considerable effort and the results may not always be replicated. Accordingly, we must continuously strive to improve our methodologies to meet the needs for both biomarker discovery and translation to the clinical laboratory.

## Concluding Remarks and Prospects

The workflows developed in this thesis followed an approach of demonstrating new insights regarding the biomarker profile of interest by applying newly improved methodologies. In both cases, the analytical performance was validated and the clinical validity of the analytical setup was also illustrated. Nevertheless, the presented approaches are at different stages of their respective lifecycles. For example, in relation to PCa and PSA, the approach described in **Chapters 2** and **3** must still undergo a true proof-of-concept study with a larger number of patient samples in order to determine whether urinary PSA proteoforms hold any association with the development of the disease. Furthermore, important questions must also still be addressed, including: *can proPSA be found in urine and, if so, why is it not detected using the CE-ESI-MS assay?* For example, the profile of seminal and serum PSA is already more established, yet the question remains: *What new insights could be gained from applying the in-depth analytical approach to study PSA from other biological fluids and what differences related to different disease conditions can be observed?* Following this, the workflow presented in **Chapters 4** and **5** is already at an advanced proof-of-concept stage and several serum *N*-glycomic signatures in CRC were corroborated. Despite this, further developments regarding the throughput of the sample preparation and sensitivity of the measurement are still required. Furthermore, the approach should be applied to a large cohort of cancer patients, primarily within a longitudinal study setup. It would be interesting to evaluate other types of samples with the method to further explore the structural and clinical relevance of various species of isomeric *N*-glycans. Finally, although this thesis adds important insights regarding the biomarkers of interest as well as methodological advancements, biomarker translation beyond the discovery phase is the main bottleneck with regard to biomarker implementation with the clinics. Thus, steps must continuously be taken in order to carry out well designed studies that address clearly defined unmet clinical needs which demonstrate obvious patient benefits.

## References

- (1) Pinho, S. S.; Reis, C. A. Glycosylation in Cancer: Mechanisms and Clinical Implications. *Nature Reviews Cancer*. **2015**, *15* (9). DOI: 10.1038/nrc3982.
- (2) Wang, C.; Ma, C. K. D.; Yeon, H.; Wang, X.; Gellman, S. H.; Abbott, N. L. Nonadditive Interactions Mediated by Water at Chemically Heterogeneous Surfaces: Nonionic Polar Groups and Hydrophobic Interactions. *Journal of the American Chemical Society*. **2017**, *139* (51). DOI: 10.1021/jacs.7b08367.
- (3) Bissantz, C.; Kuhn, B.; Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *Journal of Medicinal Chemistry*. 2010. DOI: 10.1021/jm100112j.
- (4) Grzybowski, B. A.; Ishchenko, A. V.; Dewitte, R. S.; Whitesides, G. M.; Shakhnovich, E. I. Development of a Knowledge-Based Potential for Crystals of Small Organic Molecules: Calculation of Energy Surfaces for C=O...H-N Hydrogen Bonds. *Journal of Physical Chemistry B*. **2000**, *104* (31). DOI: 10.1021/jp000644t.
- (5) Stjern Dahl, M.; Lundberg, D.; Chauhan, V.; Bordes, R.; Holmberg, K. Cleavable Surfactants: A Comparison between Ester, Amide, and Carbonate as the Weak Bond. *Journal of Surfactants and Detergents*. **2019**, *22* (5). DOI: 10.1002/jsde.12247.
- (6) Vreeker, G. C.; Wuhrer, M. Reversed-Phase Separation Methods for Glycan Analysis. *Anal Bioanal Chem*. **2017**, *409* (2), 359–378. DOI: 10.1007/s00216-016-0073-0.
- (7) Manz, C.; Grabarics, M.; Hoberg, F.; Pugini, M.; Stuckmann, A.; Struwe, W. B.; Pagel, K. Separation of Isomeric Glycans by Ion Mobility Spectrometry-the Impact of Fluorescent Labelling. *Analyst*. **2019**, *144* (17). DOI: 10.1039/c9an00937j.
- (8) Lane, C. S.; McManus, K.; Widdowson, P.; Flowers, S. A.; Powell, G.; Anderson, I.; Campbell, J. L. Separation of Sialylated Glycan Isomers by Differential Mobility Spectrometry. *Analytical Chemistry*. **2019**, *91* (15). DOI: 10.1021/acs.analchem.9b01595.
- (9) Barroso, A.; Giménez, E.; Konijnenberg, A.; Sancho, J.; Sanz-Nebot, V.; Sobott, F. Evaluation of Ion Mobility for the Separation of Glycoconjugate Isomers Due to Different Types of Sialic Acid Linkage, at the Intact Glycoprotein, Glycopeptide and Glycan Level. *Journal of Proteomics*. **2018**, *173*. DOI: 10.1016/j.jprot.2017.11.020.
- (10) Manz, C.; Mancera-Arteu, M.; Zappe, A.; Hanozin, E.; Polewski, L.; Giménez, E.; Sanz-Nebot, V.; Pagel, K. Determination of Sialic Acid Isomers from Released N-Glycans Using Ion Mobility Spectrometry. *Analytical Chemistry*. **2022**, *94* (39). DOI: 10.1021/acs.analchem.2c00783.
- (11) Nie, H.; Li, Y.; Sun, X. L. Recent Advances in Sialic Acid-Focused Glycomics. *Journal of Proteomics*. 2012. DOI: 10.1016/j.jprot.2012.03.050.
- (12) Miura, Y.; Shinohara, Y.; Furukawa, J. I.; Nagahori, N.; Nishimura, S. I. Rapid and Simple Solid-Phase Esterification of Sialic Acid Residues for Quantitative Glycomics by Mass Spectrometry. *Chemistry - A European Journal*. **2007**, *13* (17). DOI: 10.1002/chem.200601872.
- (13) Powell, A. K.; Harvey, D. J. Stabilization of Sialic Acids in N-Linked Oligosaccharides and Gangliosides for Analysis by Positive Ion Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Communications in Mass Spectrometry*. **1996**, *10* (9). DOI: 10.1002/(SICI)1097-0231(19960715)10:9<1027::AID-RCM634>3.0.CO;2-Y.

- (14) Wagt, S.; De Haan, N.; Wang, W.; Zhang, T.; Wuhrer, M.; Lageveen-Kammeijer, G. S. M. N-Glycan Isomer Differentiation by Zero Flow Capillary Electrophoresis Coupled to Mass Spectrometry. *Analytical Chemistry*. **2022**, *94* (38). DOI: 10.1021/acs.analchem.2c02840.
- (15) Zhou, S.; Huang, Y.; Dong, X.; Peng, W.; Veillon, L.; Kitagawa, D. A. S.; Aquino, A. J. A.; Mechref, Y. Isomeric Separation of Permethylated Glycans by Porous Graphitic Carbon (PGC)-LC-MS/MS at High Temperatures. *Analytical Chemistry*. **2017**, *89* (12). DOI: 10.1021/acs.analchem.7b00747.
- (16) Maliepaard, J. C. L.; Damen, J. M. A.; Boons, G.-J. P. H.; Reiding, K. R. Glycoproteomics-Compatible MS/MS-Based Quantification of Glycopeptide Isomers. *bioRxiv*. **2023**. DOI: 10.1101/2023.01.31.526390.
- (17) Carr, S. A. et al. Targeted Peptide Measurements in Biology and Medicine: Best Practices for Mass Spectrometry-Based Assay Development Using a Fit-for-Purpose Approach. *Molecular and Cellular Proteomics*. **2014**, *13* (3). DOI: 10.1074/mcp.M113.036095.
- (18) de Haan, N.; Wuhrer, M.; Ruhaak, L. R. Mass Spectrometry in Clinical Glycomics: The Path from Biomarker Identification to Clinical Implementation. *Clinical Mass Spectrometry*. **2020**, *18*. DOI: 10.1016/j.clinms.2020.08.001.
- (19) European Medicines Agency (EMA). *ICH Guidelines Q2(R2) on Validation of Analytical Procedures*; **2022**; Vol. 2.
- (20) Vermassen, T.; Van Praet, C.; Vanderschaeghe, D.; Maenhout, T.; Lumen, N.; Callewaert, N.; Hoebeke, P.; Van Belle, S.; Rottey, S.; Delanghe, J. Capillary Electrophoresis of Urinary Prostate Glycoproteins Assists in the Diagnosis of Prostate Cancer. *Electrophoresis*. **2014**, *35* (7). DOI: 10.1002/elps.201300332.
- (21) Mikšik, I. Coupling of CE-MS for Protein and Peptide Analysis. *Journal of Separation Science*. **2019**, *42* (1). DOI: 10.1002/jssc.201800817.
- (22) Uh, H. W.; Klaric, L.; Ugrina, I.; Lauc, G.; Smilde, A. K.; Houwing-Duistermaat, J. J. Choosing Proper Normalization Is Essential for Discovery of Sparse Glycan Biomarkers. *Molecular Omics*. **2020**, *16* (3). DOI: 10.1039/c9mo00174c.
- (23) Etxebarria, J.; Reichardt, N. C. Methods for the Absolute Quantification of N-Glycan Biomarkers. *Biochimica et Biophysica Acta - General Subjects*. **2016**, *1860* (8). DOI: 10.1016/j.bbagen.2016.03.003.
- (24) Jian, W.; Kang, L.; Burton, L.; Weng, N. A Workflow for Absolute Quantitation of Large Therapeutic Proteins in Biological Samples at Intact Level Using LC-HRMS. *Bioanalysis*. **2016**, *8* (16). DOI: 10.4155/bio-2016-0096.
- (25) Marie, A.-L.; Ray, S.; Ivanov, A. R. Highly-Sensitive Label-Free Deep Profiling of N-Glycans Released from Biomedically-Relevant Samples. *Nature Communications*. **2023**, *14* (1), 1618. DOI: 10.1038/s41467-023-37365-4.
- (26) de Vroome, S. W.; Holst, S.; Gironde, M. R.; van der Burgt, Y. E. M.; Mesker, W. E.; Tollenaar, R. A. E. M.; Wuhrer, M. Serum N-Glycome Alterations in Colorectal Cancer Associate with Survival. *Oncotarget*. **2018**, *9* (55), 30610–30623. DOI: 10.18632/oncotarget.25753.

- (27) Váradi, C.; Lew, C.; Guttman, A. Rapid Magnetic Bead Based Sample Preparation for Automated and High Throughput N-Glycan Analysis of Therapeutic Antibodies. *Analytical Chemistry*. **2014**, *86* (12). DOI: 10.1021/ac501573g.
- (28) Reiding, K. R.; Blank, D.; Kuijper, D. M.; Deelder, A. M.; Wuhrer, M. High-Throughput Profiling of Protein N-Glycosylation by MALDI-TOF-MS Employing Linkage-Specific Sialic Acid Esterification. *Analytical Chemistry*. **2014**, *86* (12), 5784–5793. DOI: 10.1021/ac500335t.
- (29) Nishikaze, T.; Tsumoto, H.; Sekiya, S.; Iwamoto, S.; Miura, Y.; Tanaka, K. Differentiation of Sialyl Linkage Isomers by One-Pot Sialic Acid Derivatization for Mass Spectrometry-Based Glycan Profiling. *Analytical Chemistry*. **2017**, *89* (4), 2353–2360. DOI: 10.1021/acs.analchem.6b04150.
- (30) Shimaoka, H.; Kuramoto, H.; Furukawa, J. I.; Miura, Y.; Kuroguchi, M.; Kita, Y.; Hinou, H.; Shinohara, Y.; Nishimura, S. I. One-Pot Solid-Phase Glycoblotting and Probing by Transoximization for High-Throughput Glycomics and Glycoproteomics. *Chemistry - A European Journal*. **2007**, *13* (6). DOI: 10.1002/chem.200601613.
- (31) Furukawa, J. I.; Shinohara, Y.; Kuramoto, H.; Miura, Y.; Shimaoka, H.; Kuroguchi, M.; Nakano, M.; Nishimura, S. I. Comprehensive Approach to Structural and Functional Glycomics Based on Chemoselective Glycoblotting and Sequential Tag Conversion. *Analytical Chemistry*. **2008**, *80* (4). DOI: 10.1021/ac702124d.
- (32) Stephan, C.; Jung, K.; Lein, M.; Sinha, P.; Schnorr, D.; Loening, S. A. Molecular Forms of Prostate-Specific Antigen and Human Kallikrein 2 as Promising Tools for Early Diagnosis of Prostate Cancer. *Cancer Epidemiology Biomarkers and Prevention*. **2000**, *9* (11).
- (33) Jamaspishvili, T.; Kral, M.; Khomeriki, I.; Student, V.; Kolar, Z.; Bouchal, J. Urine Markers in Monitoring for Prostate Cancer. *Prostate Cancer and Prostatic Diseases*. **2010**, *13* (1). DOI: 10.1038/pcan.2009.31.
- (34) Hessels, D.; Schalken, J. A. Urinary Biomarkers for Prostate Cancer: A Review. *Asian Journal of Andrology*. **2013**, *15* (3). DOI: 10.1038/aja.2013.6.
- (35) Stephan, C.; Ralla, B.; Jung, K. Prostate-Specific Antigen and Other Serum and Urine Markers in Prostate Cancer. *Biochimica et Biophysica Acta - Reviews on Cancer*. **2014**, *1846* (1). DOI: 10.1016/j.bbcan.2014.04.001.
- (36) Roobol, M. J.; Haese, A.; Bjartell, A. Tumour Markers in Prostate Cancer III: Biomarkers in Urine. *Acta Oncologica*. **2011**, *50* (SUPPL. 1). DOI: 10.3109/0284186X.2010.524935.
- (37) Irani, J.; Salomon, L.; Soulié, M.; Zlotta, A.; De La Taille, A.; Doré, B.; Millet, C. Urinary/Serum Prostate-Specific Antigen Ratio: Comparison with Free/Total Serum Prostate-Specific Antigen Ratio in Improving Prostate Cancer Detection. *Urology*. **2005**, *65* (3). DOI: 10.1016/j.urology.2004.10.003.
- (38) Kammeijer, G. S. M.; Nouta, J.; De La Rosette, J. J. M. C. H.; De Reijke, T. M.; Wuhrer, M. An In-Depth Glycosylation Assay for Urinary Prostate-Specific Antigen. *Analytical Chemistry*. **2018**, *90* (7), 4414–4421. DOI: 10.1021/acs.analchem.7b04281.
- (39) Mikolajczyk, S. D.; Millar, L. S.; Wang, T. J.; Rittenhouse, H. G.; Wolfert, R. L.; Marks, L. S.; Song, W.; Wheeler, T. M.; Slawin, K. M. “BPSA,” a Specific Molecular Form of Free Prostate-Specific Antigen, Is Found Predominantly in the Transition Zone of Patients with

- Nodular Benign Prostatic Hyperplasia. *Urology*. **2000**, *55* (1). DOI: 10.1016/S0090-4295(99)00372-6.
- (40) Linton, H. J.; Marks, L. S.; Millar, L. S.; Knott, C. L.; Rittenhouse, H. G.; Mikolajczyk, S. D. Benign Prostate-Specific Antigen (BPSA) in Serum Is Increased in Benign Prostate Disease. *Clinical Chemistry*. **2003**, *49* (2), 253–259. DOI: 10.1373/49.2.253.
- (41) Gilgunn, S.; Conroy, P. J.; Saldova, R.; Rudd, P. M.; O’Kennedy, R. J. Aberrant PSA Glycosylation - A Sweet Predictor of Prostate Cancer. *Nature Reviews Urology*. **2013**, *10* (2). DOI: 10.1038/nrurol.2012.258.
- (42) Mikolajczyk, S. D.; Rittenhouse, H. G. Pro PSA: A More Cancer Specific Form of Prostate Specific Antigen for the Early Detection of Prostate Cancer. *Keio Journal of Medicine*. **2003**, *52* (2). DOI: 10.2302/kjm.52.86.
- (43) Balk, S. P.; Ko, Y. J.; Bubley, G. J. Biology of Prostate-Specific Antigen. *Journal of Clinical Oncology*. **2003**, *21* (2). DOI: 10.1200/JCO.2003.02.083.
- (44) Kalsheker, N.; Morgan, K.; Chappell, S. Proteinase Inhibitors: Antichymotrypsin. In *Encyclopedia of Respiratory Medicine, Four-Volume Set*; **2006**. DOI: 10.1016/B0-12-370879-6/00327-6.
- (45) Kitata, R. B.; Hu, L. Y.; Lin, T. T.; Nicora, C. D.; Fillmore, T. L.; Nie, S.; Hudson, R. D.; Liu, T.; Leach, R. J.; Liu, A. Y.; Qian, W. J.; Shi, T. Targeted Mass Spectrometry Assays for Specific Quantification of Urinary ProPSA Isoforms. *Journal of Proteome Research*. **2023**, *22* (3). DOI: 10.1021/acs.jproteome.2c00745.
- (46) Arnold, J. N.; Saldova, R.; Galligan, M. C.; Murphy, T. B.; Mimura-Kimura, Y.; Telford, J. E.; Godwin, A. K.; Rudd, P. M. Novel Glycan Biomarkers for the Detection of Lung Cancer. *Journal of Proteome Research*. **2011**, *10* (4). DOI: 10.1021/pr101034t.
- (47) Kyselova, Z.; Mechref, Y.; Kang, P.; Goetz, J. A.; Dobrolecki, L. E.; Sledge, G. W.; Schnaper, L.; Hickey, R. J.; Malkas, L. H.; Novotny, M. V. Breast Cancer Diagnosis and Prognosis through Quantitative Measurements of Serum Glycan Profiles. *Clinical Chemistry*. **2008**, *54* (7). DOI: 10.1373/clinchem.2007.087148.
- (48) Abd Hamid, U. M.; Royle, L.; Saldova, R.; Radcliffe, C. M.; Harvey, D. J.; Storr, S. J.; Pardo, M.; Antrobus, R.; Chapman, C. J.; Zitzmann, N.; Robertson, J. F.; Dwek, R. A.; Rudd, P. M. A Strategy to Reveal Potential Glycan Markers from Serum Glycoproteins Associated with Breast Cancer Progression. *Glycobiology*. **2008**, *18* (12). DOI: 10.1093/glycob/cwn095.
- (49) Ruhaak, L. R.; Barkauskas, D. A.; Torres, J.; Cooke, C. L.; Wu, L. D.; Stroble, C.; Ozcan, S.; Williams, C. C.; Camorlinga, M.; Rocke, D. M.; Lebrilla, C. B.; Solnick, J. V. The Serum Immunoglobulin G Glycosylation Signature of Gastric Cancer. *EuPA Open Proteomics*. **2015**, *6*. DOI: 10.1016/j.euprot.2014.11.002.
- (50) Bones, J.; Byrne, J. C.; Odonoghue, N.; McManus, C.; Scaife, C.; Boissin, H.; Nastase, A.; Rudd, P. M. Glycomic and Glycoproteomic Analysis of Serum from Patients with Stomach Cancer Reveals Potential Markers Arising from Host Defense Response Mechanisms. *Journal of Proteome Research*. **2011**, *10* (3). DOI: 10.1021/pr101036b.
- (51) Vreeker, G. C. M.; Hanna-Sawires, R. G.; Mohammed, Y.; Bladergroen, M. R.; Nicolardi, S.; Dotz, V.; Nouta, J.; Bonsing, B. A.; Mesker, W. E.; van der Burgt, Y. E. M.; Wuhrer, M.; Tollenaar, R. A. E. M. Serum N-Glycome Analysis Reveals Pancreatic Cancer Disease Signatures. *Cancer Medicine*. **2020**, *9* (22). DOI: 10.1002/cam4.3439.



- (52) Saldova, R.; Fan, Y.; Fitzpatrick, J. M.; Watson, R. W. G.; Rudd, P. M. Core Fucosylation and A2-3 Sialylation in Serum N-Glycome Is Significantly Increased in Prostate Cancer Comparing to Benign Prostate Hyperplasia. *Glycobiology*. **2011**, *21* (2). DOI: 10.1093/glycob/cwq147.
- (53) Hennig, R.; Cajic, S.; Borowiak, M.; Hoffmann, M.; Kottler, R.; Reichl, U.; Rapp, E. Towards Personalized Diagnostics via Longitudinal Study of the Human Plasma N-Glycome. *Biochimica et Biophysica Acta - General Subjects*. **2016**, *1860* (8). DOI: 10.1016/j.bbagen.2016.03.035.
- (54) Lageveen-Kammeijer, G. S. M.; de Haan, N.; Mohaupt, P.; Wagt, S.; Filius, M.; Nouta, J.; Falck, D.; Wuhrer, M. Highly Sensitive CE-ESI-MS Analysis of N-Glycans from Complex Biological Samples. *Nature Communications*. **2019**, *10* (1), 1–8. DOI: 10.1038/s41467-019-09910-7.
- (55) Smith, J.; Millán-Martín, S.; Mittermayr, S.; Hilborne, V.; Davey, G.; Polom, K.; Roviello, F.; Bones, J. 2-Dimensional Ultra-High Performance Liquid Chromatography and DMT-MM Derivatization Paired with Tandem Mass Spectrometry for Comprehensive Serum N-Glycome Characterization. *Analytica Chimica Acta*. **2021**, *1179*, 338840. DOI: 10.1016/j.aca.2021.338840.
- (56) Knežević, A.; Polašek, O.; Gornik, O.; Rudan, I.; Campbell, H.; Hayward, C.; Wright, A.; Kolčić, I.; O'Donoghue, N.; Bones, J.; Rudd, P. M.; Lauc, G. Variability, Heritability and Environmental Determinants of Human Plasma n-Glycome. *Journal of Proteome Research*. **2009**, *8* (2). DOI: 10.1021/pr800737u.
- (57) Ruhaak, L. R.; Uh, H. W.; Beekman, M.; Hokke, C. H.; Westendorp, R. G. J.; Houwing-Duistermaat, J.; Wuhrer, M.; Deelder, A. M.; Slagboom, P. E. Plasma Protein N-Glycan Profiles Are Associated with Calendar Age, Familial Longevity and Health. *Journal of Proteome Research*. **2011**, *10* (4). DOI: 10.1021/pr1009959.
- (58) Knežević, A.; Gornik, O.; Polašek, O.; Pučić, M.; Redžić, I.; Novokmet, M.; Rudd, P. M.; Wright, A. F.; Campbell, H.; Rudan, I.; Lauc, G. Effects of Aging, Body Mass Index, Plasma Lipid Profiles, and Smoking on Human Plasma N-Glycans. *Glycobiology*. **2010**, *20* (8). DOI: 10.1093/glycob/cwq051.
- (59) Almeida, A.; Kolarich, D. The Promise of Protein Glycosylation for Personalised Medicine. *Biochimica et Biophysica Acta - General Subjects*. **2016**, *1860* (8). DOI: 10.1016/j.bbagen.2016.03.012.
- (60) Ransohoff, D. F. How to Improve Reliability and Efficiency of Research about Molecular Markers: Roles of Phases, Guidelines, and Study Design. *Journal of Clinical Epidemiology*. **2007**, *60* (12). DOI: 10.1016/j.jclinepi.2007.04.020.
- (61) van der Burgt, Y. E. M. Protein Biomarker Discovery Is Still Relevant and Has Entered a New Phase. *EBioMedicine*. **2019**, *43*. DOI: 10.1016/j.ebiom.2019.04.026.
- (62) Horvath, A. R.; Lord, S. J.; StJohn, A.; Sandberg, S.; Cobbaert, C. M.; Lorenz, S.; Monaghan, P. J.; Verhagen-Kamerbeek, W. D. J.; Ebert, C.; Bossuyt, P. M. M. From Biomarkers to Medical Tests: The Changing Landscape of Test Evaluation. *Clinica Chimica Acta*. **2014**. DOI: 10.1016/j.cca.2013.09.018.
- (63) Monaghan, P. J.; Lord, S. J.; St John, A.; Sandberg, S.; Cobbaert, C. M.; Lennartz, L.; Verhagen-Kamerbeek, W. D. J.; Ebert, C.; Bossuyt, P. M. M.; Horvath, A. R. Biomarker

- Development Targeting Unmet Clinical Needs. *Clinica Chimica Acta*. **2016**, *460*. DOI: 10.1016/j.cca.2016.06.037.
- (64) Human Carcino Embryonic Antigen ELISA Kit (CEA) [www.abcam.com/products/elisa/human-carcino-embryonic-antigen-elisa-kit-cea-ab99992.html](http://www.abcam.com/products/elisa/human-carcino-embryonic-antigen-elisa-kit-cea-ab99992.html) (accessed 2023 -04 -11).
- (65) Wu, T. L.; Sun, Y. C.; Chang, P. Y.; Tsao, K. C.; Sun, C. F.; Wu, J. T. Establishment of ELISA on 384-Well Microplate for AFP, CEA, CA 19-9, CA 15-3, CA 125, and PSA-ACT: Higher Sensitivity and Lower Reagent Cost. *Journal of Clinical Laboratory Analysis*. **2003**, *17* (6). DOI: 10.1002/jcla.10104.
- (66) Bon, G. G.; Kenemans, P.; Verstraeten, R.; Van Kamp, G. J.; Hilgers, J. Serum Tumor Marker Immunoassays in Gynecologic Oncology: Establishment of Reference Values. *American Journal of Obstetrics and Gynecology*. **1996**, *174* (1). DOI: 10.1016/S0002-9378(96)70381-2.
- (67) Endt, K.; Goepfert, J.; Omlin, A.; Athanasiou, A.; Tennstedt, P.; Guenther, A.; Rainisio, M.; Engeler, D. S.; Steuber, T.; Gillissen, S.; Joos, T.; Schiess, R. Development and Clinical Testing of Individual Immunoassays for the Quantification of Serum Glycoproteins to Diagnose Prostate Cancer. *PLoS ONE*. **2017**, *12* (8). DOI: 10.1371/journal.pone.0181557.
- (68) Hatano, K.; Yoneyama, T.; Hatakeyama, S.; Tomiyama, E.; Tsuchiya, M.; Nishimoto, M.; Yoshimura, K.; Miyoshi, E.; Uemura, H.; Ohyama, C.; Nonomura, N.; Fujita, K. Simultaneous Analysis of Serum A2,3-Linked Sialylation and Core-Type Fucosylation of Prostate-Specific Antigen for the Detection of High-Grade Prostate Cancer. *British Journal of Cancer*. **2022**, *126* (5). DOI: 10.1038/s41416-021-01637-x.
- (69) Rebello, O. D.; Gardner, R. A.; Urbanowicz, P. A.; Bolam, D. N.; Crouch, L. I.; Falck, D.; Spencer, D. I. R. A Novel Glycosidase Plate-Based Assay for the Quantification of Galactosylation and Sialylation on Human IgG. *Glycoconjugate Journal*. **2020**, *37* (6). DOI: 10.1007/s10719-020-09953-9.
- (70) van der Burgt, Y. E. M.; Cobbaert, C. M. Proteoform Analysis to Fulfill Unmet Clinical Needs and Reach Global Standardization of Protein Measurands in Clinical Chemistry Proteomics. *Clinics in Laboratory Medicine*. **2018**, *38* (3). DOI: 10.1016/j.cll.2018.05.001.
- (71) Urrechaga, E. High-Resolution HbA1c Separation and Hemoglobinopathy Detection With Capillary Electrophoresis. *American Journal of Clinical Pathology*. **2012**, *138* (3), 448–456. DOI: 10.1309/AJCPVYW9QZ9EVFXI.
- (72) Weykamp, C.; John, W. G.; Mosca, A. A Review of the Challenge in Measuring Hemoglobin A1c. In *Journal of Diabetes Science and Technology*; **2009**; Vol. 3. DOI: 10.1177/193229680900300306.
- (73) Landin, B.; Sisowath, C.; Strålfors, A. Factors Affecting the Evaluation and Use of a Hemoglobin A2 Method – Lot-to-Lot Variation, Long-Term Deviation and Carry-Over. *Clinica Chimica Acta*. **2023**, 117332. DOI: <https://doi.org/10.1016/j.cca.2023.117332>.
- (74) Colhoun, H. O.; Treacy, E. P.; MacMahon, M.; Rudd, P. M.; Fitzgibbon, M.; O'Flaherty, R.; Stepien, K. M. Validation of an Automated Ultrapformance Liquid Chromatography IgG N-Glycan Analytical Method Applicable to Classical Galactosaemia. *Annals of Clinical Biochemistry*. **2018**, *55* (5). DOI: 10.1177/0004563218762957.

- (75) Callewaert, N.; Van Vlierberghe, H.; Van Hecke, A.; Laroy, W.; Delanghe, J.; Contreras, R. Noninvasive Diagnosis of Liver Cirrhosis Using DNA Sequencer-Based Total Serum Protein Glycomics. *Nature Medicine*. **2004**, *10* (4). DOI: 10.1038/nm1006.
- (76) Farina-Gomez, N.; Navarro-Calderon, D.; Puerta, A.; Gonzalez, M.; Diez-Masa, J. C.; de Frutos, M. Capillary Electrophoresis Analysis of Prostate-Specific Antigen (PSA). In *Methods in Molecular Biology*; **2019**; Vol. 1972. DOI: 10.1007/978-1-4939-9213-3\_16.
- (77) Krištić, J. et al. Glycans Are a Novel Biomarker of Chronological and Biological Ages. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*. **2014**, *69* (7). DOI: 10.1093/gerona/glt190.
- (78) Shubhakar, A.; Jansen, B. C.; Adams, A. T.; Reiding, K. R.; Ventham, N. T.; Kalla, R.; Bergemalm, D.; Urbanowicz, P. A.; Gardner, R. A.; Consortium, I.-B.; Halfvarson, J.; Satsangi, J.; Fernandes, D. L.; Spencer, D. I. R. Serum N-Glycomic Biomarkers Predict Treatment Escalation in Inflammatory Bowel Disease. *Journal of Crohn's and Colitis*. **2023**. DOI: 10.1093/ecco-jcc/jjad012.
- (79) GlycanAge [www.glycanage.com](http://www.glycanage.com) (accessed 2023 -04 -11).
- (80) GlyHealth Index [www.avenna.com/products/glyhealth-index](http://www.avenna.com/products/glyhealth-index) (accessed 2023 -04 -11).
- (81) Van Den Broek, I.; Romijn, F. P. H. T. M.; Nouta, J.; Van Der Laarse, A.; Drijfhout, J. W.; Smit, N. P. M.; Van Der Burgt, Y. E. M.; Cobbaert, C. M. Automated Multiplex LC-MS/MS Assay for Quantifying Serum Apolipoproteins A-I, B, C-I, C-II, C-III, and E with Qualitative Apolipoprotein E Phenotyping. *Clinical Chemistry*. **2016**, *62* (1). DOI: 10.1373/clinchem.2015.246702.
- (82) Kushnir, M. M.; Rockwood, A. L.; Roberts, W. L.; Abraham, D.; Hoofnagle, A. N.; Meikle, A. W. Measurement of Thyroglobulin by Liquid Chromatography - Tandem Mass Spectrometry in Serum and Plasma in the Presence of Antithyroglobulin Autoantibodies. *Clinical Chemistry*. **2013**, *59* (6). DOI: 10.1373/clinchem.2012.195594.
- (83) Smit, N. P. M.; Ruhaak, L. R.; Romijn, F. P. H. T. M.; Pieterse, M. M.; Van Der Burgt, Y. E. M.; Cobbaert, C. M. The Time Has Come for Quantitative Protein Mass Spectrometry Tests That Target Unmet Clinical Needs. *Journal of the American Society for Mass Spectrometry*. **2021**, *32* (3). DOI: 10.1021/jasms.0c00379.
- (84) Ruhaak, L. R. The Use of Multiple Reaction Monitoring on QQQ-MS for the Analysis of Protein- and Site-Specific Glycosylation Patterns in Serum. In *Methods in Molecular Biology*; **2017**; Vol. 1503. DOI: 10.1007/978-1-4939-6493-2\_6.
- (85) Van Scherpenzeel, M.; Steenbergen, G.; Morava, E.; Wevers, R. A.; Lefeber, D. J. High-Resolution Mass Spectrometry Glycoprofiling of Intact Transferrin for Diagnosis and Subtype Identification in the Congenital Disorders of Glycosylation. *Translational Research*. **2015**, *166* (6). DOI: 10.1016/j.trsl.2015.07.005.
- (86) Lathrop, J. T.; Jeffery, D. A.; Shea, Y. R.; Scholl, P. F.; Chan, M. M. US Food and Drug Administration Perspectives on Clinical Mass Spectrometry. *Clinical Chemistry*. **2016**, *62* (1). DOI: 10.1373/clinchem.2015.244731.
- (87) Henry, N. L.; Hayes, D. F. Cancer Biomarkers. *Molecular Oncology*. **2012**, *6* (2), 140–146. DOI: <https://doi.org/10.1016/j.molonc.2012.01.010>.

- (88) Waters Corporation. Increasing LC-MS accessibility for streamlined biotherapeutic analysis [www.waters.com/webassets/cms/library/docs/increasing\\_lc-ms\\_accessibility\\_for\\_streamlined\\_biotherapeutic\\_analysis\\_application\\_ebook.pdf](http://www.waters.com/webassets/cms/library/docs/increasing_lc-ms_accessibility_for_streamlined_biotherapeutic_analysis_application_ebook.pdf) (accessed 2023 -04 -11).



# Appendices

*List of Abbreviations*

*English Summary*

*Nederlandse Samenvatting*

*Curriculum Vitae*

*PhD Portfolio*

*List of Publications*

*Acknowledgements*

## List of Abbreviations

<b>2-AA</b>	2-Aminobenzoic acid	<b>DTT</b>	Dithiothreitol
<b>2-AB</b>	2 Aminobenzamide	<b>E</b>	Ethyl esterified $\alpha$ 2,6-linked <i>N</i> -acetylneuraminic acid
<b>2-PA</b>	2-Aminopyridine	<b>EDC</b>	1-ethyl-3-(3-(dimethylamino) propyl) carbodiimide
<b>AAL</b>	Aleuria aurantia lectin	<b>EDTA</b>	Ethylenediaminetetraacetic acid
<b>AGP</b>	Alpha-1-acid glycoprotein	<b>EEA</b>	Ethyl esterification and amidation
<b>Ala</b>	Alanine (A)	<b>EIC</b>	Extracted ion chromatogram
<b>Am</b>	Amidated $\alpha$ 2,3-linked <i>N</i> -acetylneuraminic acid	<b>EIE</b>	Extracted ion electropherogram
<b>Arg</b>	Arginine (R)	<b>ELISA</b>	Enzyme-linked immunosorbent assay
<b>Asn</b>	Asparagine (N)	<b>EOF</b>	Electroosmotic flow
<b>AUC</b>	Area under the curve	<b>ER</b>	Endoplasmic reticulum
<b>BFS</b>	Bare fused silica	<b>ESI</b>	Electrospray ionization
<b>BGE</b>	Background electrolyte	<b>ETD</b>	Electron-transfer dissociation
<b>BPH</b>	Benign prostate hyperplasia	<b>EtOH</b>	Ethanol
<b>BPSA</b>	Benign prostate-specific antigen	<b>F</b>	Fucose
<b>CE</b>	Capillary electrophoresis	<b>FA</b>	Formic acid
<b>CEA</b>	Carcinoembryonic antigen	<b>Fc<math>\gamma</math>RIIIA</b>	Fc-gamma receptor IIIA
<b>CESI</b>	Capillary electrophoresis-electrospray ionization	<b>FDA</b>	Food and Drug Administration
<b>CID</b>	Collision-induced dissociation	<b>FLR</b>	Fluorescence
<b>CRC</b>	Colorectal cancer	<b>FWHM</b>	Full width half maximum
<b>CV</b>	Coefficient of variation	<b>GaINAc</b>	<i>N</i> -acetylgalactosamine
<b>CZE</b>	Capillary zone electrophoresis	<b>GlcNAc</b>	<i>N</i> -acetylglucosamine
<b>DEN</b>	Dopant enriched nitrogen	<b>Glu</b>	Glutamate (E)
<b>DMT-MM</b>	4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methyl morpholinium chloride		
<b>DRE</b>	Digital rectal examination		

<b>Gly</b>	Glycine (G)	<b>N</b>	<i>N</i> -acetylhexosamine
<b>H</b>	Hexose	<b>NaCl</b>	Sodium chloride
<b>HAc</b>	Glacial acetic acid	<b>NaOH</b>	Sodium hydroxide
<b>HCl</b>	Hydrochloric acid	<b>o.d</b>	Outer diameter
<b>HILIC</b>	Hydrophilic interaction liquid chromatography	<b>PBM</b>	Protein Binding Membrane
<b>HOBt</b>	Hydrate 1- hydroxybenzotriazole	<b>PBS</b>	Phosphate buffered saline
<b>HPAEC-</b>	High-pH/performance	<b>PCa</b>	Prostate cancer
<b>PAD</b>	anion-exchange chromatography-pulsed anomeric detection	<b>PCA</b>	Principal component analysis
<b>i.d</b>	Internal diameter	<b>PCR</b>	Polymerase chain reaction
<b>IAA</b>	Iodoacetamide	<b>PEI</b>	Polyethylenimine
<b>ICC</b>	Ion charge control	<b>PGC</b>	Porous graphitized carbon
<b>IMS</b>	Ion-mobility mass spectrometry	<b>pH</b>	Potential of Hydrogen
<b>KLK</b>	Kallikrein	<b>Phe</b>	Phenylalanine (F)
<b>LC</b>	Liquid chromatography	<b>PNGase F</b>	Peptide- <i>N</i> -glycosidase F
<b>LLOQ</b>	Lower limit of quantification	<b>Pro</b>	Proline (P)
<b>LST</b>	Sialyllacto- <i>N</i> -tetraose	<b>PROC</b>	Procainamide
<b>Lys</b>	Lysine (K)	<b>proPSA</b>	Precursor protein of prostate-specific antigen
<b><i>m/z</i></b>	Mass to charge ratio	<b>PSA</b>	Prostate-specific antigen
<b>MALDI</b>	Matrix-assisted laser desorption ionization	<b>PTM</b>	Post-translational modification
<b>MGAT</b>	<i>N</i> -acetylglucosaminyl transferase	<b>QTOF</b>	Quadrupole time-of-flight
<b>MQ</b>	Milli-Q deionized water	<b>RNA</b>	Ribonucleic acid
<b>MRM</b>	Multiple reaction monitoring	<b>ROC</b>	Receiver operating characteristic
<b>MS</b>	Mass spectrometry	<b>RPLC</b>	Reversed-phase liquid chromatography
<b>MS/MS</b>	Tandem mass spectrometry	<b>RSD</b>	Relative standard deviation
		<b>RT</b>	Retention time
		<b>S</b>	Sialic acid
		<b>S/N</b>	Signal-to-noise ratio



## Appendices

<b>SDS-</b>	Sodium dodecyl-sulfate	<b>UPLC</b>	Ultra performance liquid
<b>PAGE</b>	polyacrylamide gel		chromatography
	electrophoresis	<b>UV</b>	Ultraviolet
<b>Ser</b>	Serine (S)	<b>XIC</b>	Extracted ion
<b>Thr</b>	Threonine (T)		chromatogram
<b>UHPLC</b>	Ultra-high performance	<b>XIE</b>	Extracted ion
	liquid chromatography		electropherogram
<b>UHRMS</b>	Ultra-high resolution mass		
	spectrometry		



## English Summary

In prostate (PCa) and colorectal (CRC) cancer, there is a need to improve patient stratification techniques that aid diagnostic and prognostic decision-making. To fulfill this unmet clinical need, the measurement of disease-related biological parameters known as “biomarkers” from biofluids is an approach with the potential to develop non-invasive tests as well as achieve greater clinical accuracy and personalized medicine. Thus, the aim of this thesis was to develop a better understanding of biomarkers relevant to PCa and CRC as well as advancing analytical methodology and achieving methodological advancements for the purpose of biomarker discovery. In this regard, prostate-specific antigen (PSA) proteoform (PCa) and serum *N*-glycosylation (CRC) profiles were investigated using capillary electrophoresis-electrospray ionization-mass spectrometry (CE-ESI-MS) and reversed phase-liquid chromatography (RPLC)-MS, respectively.

In *part one* of this thesis (**Chapters 2 and 3**), the intact proteoform profile of both urinary and seminal prostate-specific antigen (PSA) was explored using CE-ESI-MS. **Chapter 2** focused on the development of the CE-ESI-MS method in order to explore urinary PSA proteoforms, in particular as these forms had previously not been well defined. In addition, previous studies often focused on further defining the glycosylation – rather than the proteoform – profile whereas both were equally assessed here. Furthermore, proteoforms had largely been described solely as “pI isoforms” whereas the assignment of cleaved proteoforms is provided in **Chapter 2**. Additionally, **Chapter 3** addressed some of the challenges associated with the data processing of intact proteoforms, namely semi-automatic proteoform annotation and quantification, thereby enhancing the throughput and accuracy of the workflow. Interestingly, in-depth bottom-up and middle-up approaches were integrated into the analysis in order to further support proteoform assignment in seminal and urinary PSA. Furthermore, it was also shown that extracted ion electropherogram and deconvolution approaches yielded similar quantification results with respect to the relative abundance of PSA proteoforms. Overall, the method is now poised for the analysis of a larger number of patient samples and our results may inform future studies regarding this protein’s proteoforms as well as any related clinical implications thereof.

In *part two* (**Chapters 4 and 5**), the potential of serum *N*-glycosylation as a biomarker for CRC was investigated. In this sense, important developments to the methodology were made in **Chapter 4**. Here, fluorescent labeling of total plasma released *N*-glycans via procainamide was combined with sialic acid-specific derivatization via ethyl esterification and amidation. This allowed the retention of all *N*-glycan species by RPLC-MS and, in particular,  $\alpha$ 2,3- and  $\alpha$ 2,6-sialylated *N*-glycans were differentiated based upon retention time, precursor mass, and fragmentation spectra, and additional sialylated isomers were resolved. Other isomeric species such as antennary and core-fucosylated *N*-glycans were also separated. In addition, a new quantification approach was derived by combining the fluorescent detection signal with *N*-glycan ratios determined by MS. This improved feature coverage as well as the precision of the method in comparison with performing either quantification approach alone. Finally, the performance of the method was benchmarked against the gold-standard method for released glycan analysis, namely HILIC-MS.

**Chapter 5** demonstrated the application of the developed method in order to analyze serum *N*-glycosylation from pre- and post-operative CRC. In this chapter, previous findings with regard to specific glycosylation patterns in CRC were corroborated. In addition, new results were obtained as it was demonstrated that specific *N*-glycan isomers, such as a trisialylated structure with antennary fucosylation, are implicated in the disease. It was also illustrated that pre-operative abundances of serum *N*-glycans may differentiate adenocarcinoma from other histological types and, in particular, differences between histological types were eradicated following surgery. Thus, these results promote the use of serum *N*-glycosylation signatures in order to monitor patient profiles in response to clinical interventions such as surgery.

In **Chapter 6**, the challenges and limitations of the presented work were discussed. In addition, future directions with regard to the biomarkers assessed in this thesis, namely PSA proteoforms and total serum *N*-glycosylation, were explored. Importantly, perspectives were provided on the translation of the results and methods described by this thesis to the clinics.

## Nederlandse Samenvatting

Zowel in het geval van prostaatkanker (PCa) als bij dikkedarmkanker (CRC) is er een klinische behoefte aan verbeterde methoden voor patiëntstratificatie met betrekking tot diagnose en prognose. Het meten van ziekte-gerelateerde parameters, dat wil zeggen “biomarkers”, in lichaamsvloeistoffen is een manier om aan de hand van een niet-invasieve test met verbeterde nauwkeurigheid te komen tot “personalized medicine” (zorg-op-maat). Het doel van het onderzoek dat is beschreven in dit proefschrift was het verkrijgen van meer inzicht in biomarkers die relevant zijn in PCa danwel CRC en daarnaast het ontwikkelen van analytische methoden ten behoeve van ontdekking van nieuwe biomarkers. Voor PCa en CRC werden respectievelijk profielen van prostaat-specifiek antigeen (PSA) en serum *N*-glycosylering bestudeerd met behulp van capillaire elektroforese electrospray ionisatie massaspectrometrie (CE-ESI-MS) en “reversed phase” vloeistofchromatografie (RPLC)-MS.

In het *eerste deel* van dit proefschrift (**Hoofdstukken 2 en 3**) zijn intacte proteovormprofielen van PSA uit urine en uit semen onderzocht aan de hand van CE-ESI-MS. In **Hoofdstuk 2** is een CE-ESI-MS methode beschreven die meting van PSA-proteovormen laat zien, waarvan structuren eerder slecht gedefinieerd waren. Terwijl in deze eerdere studies de focus vaak lag op glycosylering wordt in dit hoofdstuk zowel de glycosylering als het proteovormprofiel van PSA beschreven. Proteovormen die eerder werden gerapporteerd als “pI isovormen” zijn in **Hoofdstuk 2** in detail gekarakteriseerd als specifiek geknipte proteovormen. Vervolgens is in **Hoofdstuk 3** beschreven welke stappen genomen dienen te worden bij het verwerken van data van intacte proteovormen, en daarbij is een semi-geautomatiseerde kwantificering voorgesteld die de snelheid en nauwkeurigheid van de strategie vergroot. Bij deze analyses zijn gedetailleerde “bottom-up” en ‘middle-up” benaderingen geïntegreerd, waarmee toekenningen van PSA proteovormen verder bevestigd werden. Tevens is vastgesteld dat verschillende integratiestrategieën (electropherogram versus deconvolutie) vergelijkbare resultaten opleverden bij de bepaling van relatieve kwantificering van PSA proteovormen. Samenvattend is deze methode nu gereed voor de analyse van grotere patiëntaantallen en kunnen onze resultaten dienen als startpunt voor toekomstige studies aan PSA proteovormen.

In het *tweede deel* van dit proefschrift (**Hoofdstukken 4 en 5**) wordt het potentieel van serum *N*-glycosylering als biomarker voor CRC bestudeerd. Hiervoor waren belangrijke ontwikkelingen noodzakelijk zoals beschreven in **Hoofdstuk 4**, bestaande uit koppeling van glycaanstructuren met een fluorescent label (procainamide) gecombineerd met siaalzuurderivatisering. Hierdoor kunnen *N*-glycanen geanalyseerd worden met behulp van RPLC-MS en kunnen  $\alpha$ 2,3- en  $\alpha$ 2,6-gesialyleerde isomeren van elkaar gescheiden worden, evenals specifieke antennaire en core-gefucosyleerde structuren. De inzet van een fluorescent label maakt hierbij een nieuwe kwantificeringsmethode mogelijk voor het bepalen van de verschillende glycanen. Er wordt aangetoond dat de precisie van de methode verbetert en het aantal geanalyseerde structuren toeneemt in vergelijking tot wanneer slechts MS voor detectie wordt ingezet. Tenslotte worden de prestaties van de methode vergeleken met de gouden standaard voor glycaananalyse, namelijk HILIC-MS.

In **Hoofdstuk 5** wordt deze methode ingezet voor de analyse van serum *N*-glycosyleringsprofielen van CRC patiënten vóór en na chirurgische ingreep, en de resultaten worden vergeleken (en bevestigd) door eerder werk aan ditzelfde cohort. Nieuwe resultaten laten zien dat specifieke structuren zoals drievoudig gesialyleerde glycanen met antenne-fucosylering correleren met CRC. Daarnaast wordt beschreven dat specifieke structuren correleren met verschillen in histologie van de tumor en dat deze verschillen post-operatief verdwijnen. Geconcludeerd wordt dat deze resultaten een sterke aanmoediging om serum *N*-glycosyleringsprofielen in te zetten bij het volgen van patiënten tijdens therapie.

In **Hoofdstuk 6** worden tenslotte de uitdagingen en beperkingen van het hier beschreven onderzoek bediscussieerd. De richting van toekomstig onderzoek wordt beschreven, met name over PSA proteovormen en serum *N*-glycosyleringsprofielen. Belangrijk hierbij is dat de vooruitzichten voor translatie van resultaten en methoden naar de kliniek in ogenschouw worden genomen.

## Curriculum Vitae

Alan Bernard Moran was born on April 22<sup>nd</sup> 1993 in Dublin, Ireland. He attended the course of Biomedical and Biomolecular Sciences at University College Dublin (UCD) where he obtained his Bachelor of Science (BSc) degree, majoring in Genetics, in 2015. As part of his final year, Alan worked at the UCD Conway Institute of Biomolecular and Biomedical Research in the laboratory of Dr. Peadar Ó Gaora under the supervision of Dr. Lisa Rogers where he focused on improving bioinformatic workflows for the detection of positive selection in verotoxigenic *E. coli*. Following this, Alan obtained a diploma in Pharmaceutical Technology (Bioprocessing and Healthcare) from the Technological University Dublin in 2016. In the same year, Alan became employed at the National Institute for Bioprocessing Research and Training (NIBRT) as an Analytical Scientist in the Contract Research team. It was at NIBRT, under the supervision of Dr. Brian Morrissey and Patrick Jennings, where he was first introduced to the wonders of glycobiology and mass spectrometry (MS).

Motivated by his newfound curiosity in glycomics and MS, Alan moved to Leiden, The Netherlands in 2017 as a PhD candidate and Marie Curie Early Stage Researcher. He joined the Leiden University Medical Center and Center for Proteomics and Metabolomics under the supervision of Prof. Dr. Manfred Wuhrer and Dr. Guinevere Lageveen-Kammeijer. Here, he focused on developing a capillary electrophoresis-MS approach for detecting and quantifying intact prostate-specific antigen (PSA) proteoforms which could serve as new biomarkers aiming at prostate cancer patient stratification. Furthermore, as part of the European Industrial Doctorate, Alan spent 18 months at Ludger Ltd., UK in the laboratory of Dr. Daniel Spencer with the support of Dr. Richard Gardner. Here, Alan worked on a released *N*-glycan and sialic acid derivatization assay followed by LC-MS in order to identify specific glycomic signatures in the sera from colorectal cancer patients.

As part of the Horizons 2020 Marie Skłodowska-Curie Actions consortium, GlySign, Alan also had the opportunity to take part in a multidisciplinary and intersectoral training programme. He worked alongside the other GlySign researchers in order to organize two workshops aimed at sharing their experience on the topic of “industrial-

academic mobility” with Bachelors and Masters students. Furthermore, between 2018 – 2019, Alan presented his research at three international conferences and symposia.

Since 2021, Alan has been working at Janssen Vaccines and Prevention, Leiden in the Bacterial Vaccines department as part of the Analytical Development group headed by Dr. Chakkumkal Anish. As an Associate Scientist, Alan is contributing the knowledge of glycomics and MS that he gained during his PhD in order to develop new vaccines to fight antimicrobial resistance.



## PhD Portfolio

### Mandatory Courses

- PhD introductory meeting 2017
- Basic methods and reasoning in biostatistics 2018
- BROK course – regulations for conducting clinical research in the Netherlands 2020

### Additional Courses and Workshops

- Time and self-management for PhDs (*Leiden University, The Netherlands*) 2017
- Personal Effectiveness (*GlyCoCan workshop, LUMC, The Netherlands*) 2018
- High-throughput data processing of MALDI-TOF-MS data (*GlyCoCan workshop, LUMC, The Netherlands*) 2018
- Storytelling and Stagecraft for Scientists (*GlyCoCan workshop, LUMC, The Netherlands*) 2018
- Business Process and Quality Management (*GlySign workshop, Ludger Ltd., UK*) 2018
- Biologicals, Biotechnology & Quality Management (*GlySign workshop, LUMC, The Netherlands*) 2018
- Fundamentals of capillary electrophoresis (*LUMC, The Netherlands*) 2018
- Glycan analysis workshop (*Ludger Ltd., UK*) 2019
- Intellectual property rights and technology transfer (*GlySign workshop, Genos Ltd., Croatia*) 2019
- Industrial-academic mobility workshop (*GlySign workshop, Genos Ltd., Croatia*) 2019
- Statistics for Research workshop (*Oxford Brookes University, UK*) 2020
- Industrial-academic mobility workshop (*GlySign workshop, LUMC, The Netherlands*) 2020
- Ace The Deck presentation training (*www.acethedeck.com*) 2020

### Congress Attendance and Oral or Poster Presentations

- 12th Jenner Glycobiology and Medicine Symposium (*Dubrovnik, Croatia*) 2017
- 11th Mass Spectrometry School in Biotechnology and Medicine (MSBM) – **Poster presentation** (*Dubrovnik, Croatia*) 2017
- NVMS spring meeting (*Amsterdam, The Netherlands*) 2017
- NVMS fall meeting (*Delft, The Netherlands*) 2017

- NLab discussion (*LUMC, Leiden, The Netherlands*) 2017
- CHAINS conference; **Poster presentation** (*Veldhoven, The Netherlands*) 2017
- Mass Spectrometry: Applications to the Clinical Lab (MSACL) – **Oral presentation** (*Salzburg, Austria*) 2018
- Global CESI-MS symposium – **Oral presentation** (*Leiden, The Netherlands*) 2019
- EuroCarb XX – **Oral presentation** (*Leiden, The Netherlands*) 2019
- Glycoanalytics webinar, European Glycoscience Community – **Oral presentation (invitation)** (*www.euroglyco.com*) 2021
- Protein Metrics Webinar – **Oral presentation (invitation)** (*www.proteinmetrics.com*) 2023
- Analytical Glycosciences webinar, European Glycoscience Community – **Oral presentation (invitation)** (*www.euroglyco.com*) 2023

### **Awards and Grants**

- MSACL young investigator travel grant 2018

## List of Publications

- (1) **Moran, A. B.**; Domínguez-Vega, E.; Nouta, J.; Pongracz, T.; de Reijke, T. M.; Wuhrer, M.; Lageveen-Kammeijer, G. S. M. Profiling the Proteoforms of Urinary Prostate-Specific Antigen by Capillary Electrophoresis – Mass Spectrometry. *Journal of Proteomics*. **2021**, 238. DOI: 10.1016/j.jprot.2021.104148.
- (2) Ward, S. E.; Sullivan, J. M. O.; **Moran, A. B.**; Spencer, D. I. R.; Gardner, R. A.; Sharma, J.; Fazavana, J.; Monopoli, M.; McKinnon, T. A. J.; Chion, A.; Haberichter, S.; Donnell, J. S. O. Sialylation on O-Linked Glycans Protects von Willebrand Factor from Macrophage Galactose Lectin-Mediated Clearance. *Haematologica*. **2022**, 107 (3). DOI: 10.3324/haematol.2020.274720.
- (3) **Moran, A. B.**; Gardner, R. A.; Wuhrer, M.; Lageveen-Kammeijer, G. S. M.; Spencer, D. I. R. Sialic Acid Derivatization of Fluorescently Labeled N-Glycans Allows Linkage Differentiation by Reversed-Phase Liquid Chromatography–Fluorescence Detection–Mass Spectrometry. *Analytical Chemistry*. **2022**, 94 (18), 6639–6648. DOI: 10.1021/acs.analchem.1c02610.
- (4) Melo Diaz, J. M.; **Moran, A. B.**; Peel, S. R.; Hendel, J. L.; Spencer, D. I. R. Egg Yolk Sialylglycopeptide: Purification, Isolation and Characterization of N-Glycans from Minor Glycopeptide Species. *Organic and Biomolecular Chemistry*. **2022**, 20 (24). DOI: 10.1039/d2ob00615d.
- (5) **Moran, A. B.**; Domínguez-Vega, E.; Wuhrer, M.; Lageveen-Kammeijer, G. S. M. Software-Assisted Data Processing Workflow for Intact Glycoprotein Mass Spectrometry. *Journal of Proteome Research*. **2023**, 22 (4), 1367–1376. DOI: 10.1021/acs.jproteome.2c00762.
- (6) **Moran, A. B.**; Elgood-Hunt, E.; van der Burgt Y. E. M.; Wuhrer, M.; Mesker W. E.; Tollenaar, R. A. E. M.; Spencer, D. I. R.; Lageveen-Kammeijer, G. S. M. Serum N-glycosylation RPLC-FD-MS Assay to Assess Colorectal Cancer Surgical Interventions. *Biomolecules*. **2023**, 13 (6), 896. DOI: 10.3390/biom13060896.



## Acknowledgements

A PhD is a journey filled with many ups and downs. At moments it felt like climbing the highest mountain peak, watching the sun rise. Yet at times it felt more like being in the middle of a terrible storm. Fortunately, I received fantastic support and I am thankful to everybody that offered me a word of advice and encouragement.

From my days at NIBRT, I am grateful to Brian and Paddy who took me on as a fresh-faced graduate and sparked my interest in glycobiology and mass spectrometry. To Roisin, thank you for your advice about pursuing a PhD.

I was lucky to have several supervisors from whom I learned different and essential lessons for being a scientist. I would like to thank my promoter Prof. Dr. Manfred Wuhrer for giving me the opportunity to complete a PhD in the CPM. You greatly impacted my scientific and personal growth and always took on the role which I needed; teacher, coach, and also motivator. To my co-promotor Dr. Guinevere Lageveen-Kammeijer, I truly appreciate your optimism and dedication. You guided me through every step of the way and now I can stand on my own two feet thanks to your leadership. I am also grateful to Dr. Daniel Spencer – Daniel, your forward-thinking greatly helped the completion of this thesis and your compassion was particularly helpful during times of self-doubt. I would also like to thank Dr. Theo de Reijke for great discussions about the application of our work.

At the CPM I met great colleagues and friends (after they learned how to understand my Irish accent!). In particular, I would like to thank Yuri, who was always available to help me since day one, including the Dutch summary for this thesis. To Simone, with your endless kindness and humor, I learned not to take myself too seriously. To Elena, for always providing valuable feedback and Jan, who helped me in the lab countless times. I am also thankful to Christoph, Di, Sander, Steffen, Wei, and Wenjun, as well as Agnes, Bram, Marco, Noortje, and Suzanne for the great support in the CPM.

I acquired another great set of colleagues during my time at Ludger Ltd. For this I would like to say a special thanks to Richard for answering my endless questions and Helen for her inspiring workshops. I would also like to mention Andy, Bas, Carol, Christel, Conchi, Javi, Jen, Manuela, Max, Millie, Paulina, and Rad for the great memories at Ludger.

To the entire GlySign team, including David, Manu, Viktoria, and, of course, Thomas, Iwona, Osmond, Elham, and Daniel, thank you all for the fun and intense workshops. Working with such an amazing and diverse group of people taught me how to appreciate the unique traits of others. I am also thankful to Michel and indeed the entire BacVac department, writing this thesis while also working was the greatest challenge that I undertook during the PhD, however I received amazing support for which I am eternally grateful. I am especially thankful to Anish who inspired and believed in me, and the AD team – Alta, Christina, Diane, Elise, Emma, Hanna, Joram, Lydia, Malti, the two Marcos, Miranda, Nikkie, Rebekka, Remco, and Sander, I laughed and learned every day working with each of you.

To my friends in Leiden – Alessio, Ana, Andy, Elena, Fanny, Ieva, Kate, Leria, Maja, Marti, Michele, Mohan, Roc, Stefan, Tamas, Victor, and Wouter, we have made an international family and every word written here was worth it for your friendship. I am grateful for every Spontaneous Beer, dinner, and adventure, and I look forward to the many more to come. To my Amigos in Oxford – Alvaro, Ash, Carlos, Dani, David, Isabella, and Rosie, Oxford became another home away from home thanks to you. Lastly, I would like to thank the Lads from Ireland – Brian, Dan, Daneil, Darragh, Dylan, Fiachra, and Gary, you gave me the courage to take on new challenges and each of your successes has inspired me. A final word to both of my paranymphs – Alessio, you were always there to give another perspective and I could always count on you, and Dylan, my oldest friend who encouraged me to move abroad.

To Eva, you were my supporter, coach, motivator, collaborator, and unofficial (and strictest!) supervisor. The challenge of writing the thesis while also working full-time seemed almost impossible, but I owe this achievement to your love and guidance. Together, I believe that we can accomplish anything. T'estimo.

Finally, I would like to thank my family for all the warm and welcoming visits home. I would especially like to thank my cousin Stephen, who helped with the design of this thesis, as well as my aunt Patricia and brother(-in-law) Paddy for their support. To Karen, Paul, and Conor, a younger brother could not ask for better role models. And lastly, I am grateful to my parents, I owe everything to you and the encouragement you gave me to believe in myself.

