

**Development of highly sensitive and  
selective applications for glycoproteomics  
and clinical glycomics**

**Inaugural-Dissertation**

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, and  
Pharmacy of the Freie Universität Berlin

by

**Hannes Hinneburg**

2016

Prepared from Nov. 2011 until Sep. 2015

Under supervision of Dr. Daniel Kolarich at the

Max Planck Institute of Colloids and Interfaces

First reviewer : Dr. Daniel Kolarich

Second reviewer : Prof. Dr. Kevin Pagel

Date of defense : 15.12.2016

## Abstract

Glycan structure elucidation, either on released oligosaccharides or still attached to proteins/peptides, remains a challenging task mostly due to the complexity embedded in these key molecules. Uncovering the functional relevance of glycosylation in biological systems can be used for disease diagnosis and treatment. The aim of this work was to develop highly sensitive and selective clinical glycomics and glycoproteomics tools.

A method for the extraction of *N*- and *O*-glycans from histopathological tissue sections (formalin-fixed & paraffin-embedded and/or Hematoxylin & Eosin stained) was developed. The use of laser capture microdissection provided the opportunity to gain spatial insights into the glyco-epitope distribution in hepatocellular carcinoma and surrounding non-tumor tissue from as low as a few thousand cells. For the first time detailed linkage information on structural isomers (e.g. 2,3 vs. 2,6 linked sialic acids) and important glycan epitopes (e.g. such as Lewis X) could be determined from minimal amounts of clinical tissue specimens. This developed approach represents an enormous step forward in the quest to identify disease specific glycan signatures.

For the glycoproteomic investigations defined glycopeptide standards were synthesized. These were employed to systematically investigate glycopeptide fragmentation in quadrupole-time-of-flight mass spectrometer to improve the information content of tandem MS data for software-assisted glycopeptide analysis. Optimal analysis conditions were used in a glycoproteomic investigation of the entire set of human immunoglobulins.

The glycopeptide standards were also key compounds to explore the potential of ion mobility–mass spectrometry (IM-MS) for providing glycan structure information from isobaric glycopeptides. IM-MS was able to differentiate *N*-glycopeptide positions isomers. Even more importantly, IM-MS could be used to

differentiate  $\alpha$ 2-6 from  $\alpha$ 2-3 sialic acid (NeuAc) linkage isomers. This was achieved by IM-MS separation of oxonium ions (NeuAc-Gal-GlcNAc; 657 *m/z*) generated by collision induced dissociation of isolated precursor ions but not on the level of intact glycopeptides. Using this technology for the first time NeuAc linkages can be studied in a site-specific manner within a single experiment. In future this capacity will provide novel insights on the role of sialic acid in health and disease.

## Zusammenfassung

Die detaillierte Charakterisierung von Glykanen, in gelöster Form oder in Verbindung mit Proteinen/Peptiden, stellt aufgrund der ihnen innewohnenden umfassenden Komplexität eine enorme Herausforderung dar. Die Aufklärung ihrer biologischen Funktionen ist für den Kampf gegen Krankheiten von entscheidender Bedeutung. Das Ziel dieser Arbeit war die Entwicklung von hochsensitiven und selektiven Methoden zur Anwendung in der klinischen Glykomik und Glykoproteomik.

Es wurde eine Methode für die Extraktion von *N*- und *O*-glykanen aus klinischen Gewebeproben (Formalin-fixiert & Paraffin-eingebettet und/oder Hematoxylin & Eosin gefärbt) entwickelt. Mit Hilfe von Laser-Mikrodissektionen war es zudem möglich räumliche Glykanprofile von Gewebestücken von nur wenigen tausend Zellen (Leberzellkarzinom) und angrenzenden Bereichen zu erstellen. Zum ersten Mal konnten detaillierte Informationen über Glykan-Isomere (z.B. 2,3 und 2,6 verknüpfte Sialinsäuren) und wichtige Epitope (z.B. Lewis X) von kleinsten Mengen klinischen Materials gewonnen werden. Daher stellt diese Methode ein wertvolles Werkzeug bei der Suche nach geeigneten Glykan Biomarkern dar.

Für die glykoproteomischen Arbeiten wurden definierte Glykopeptidstandards hergestellt. Diese wurden dann benutzt um systematisch deren Fragmentierung in Quadrupol-Flugzeitmassenspektrometern zu untersuchen und den Informationsgehalt der entstandenen Fragment-Ion-Spektren für eine bessere computer-assistierte Datenanalyse zu erhöhen. Die optimalen Parameter wurden dann u.a. mittels diverser humaner Immunglobuline validiert.

Die synthetischen Glykopeptide wurden weiterhin verwendet, um das Potential der Ionen-Mobilitäts-Massenspektrometrie zur Glykopeptid Charakterisierung zu untersuchen. Es hat sich gezeigt, dass die untersuchten *N*-glykopeptid-Positionsisomere mittels der Methode unterscheidbar waren. Weiterhin konnten

isobare Glykopeptide, welche nur in ihrer terminalen Sialinsäureverknüpfung variierten ( $\alpha$ 2-6 bzw.  $\alpha$ 2-3), auf Fragment-Ebene unterschieden werden. Dies geschah an Hand von Oxonium-Ionen (NeuAc-Gal-GlcNAc;  $m/z$  657), welche zuvor durch kollisionsinduzierte Dissoziation der Vorläuferionen generiert wurden. Zukünftig ermöglicht die Methode die Bestimmung von Sialinsäureverknüpfungen auf Glykopeptiden in einem Experiment, was die Untersuchung von Sialiansäuren in einem bestimmten biologischen Kontext sehr vereinfachen wird.

## Acknowledgments

Finally, the years of effort performing my PhD, teaching me valuable life lessons, materializes into this piece of work. It would not have been possible without the effort and dedication of many friends and colleagues. There is no way to thank everybody contributing to this work whatsoever. Here, I can only pick a small selection of the most valuable to me throughout the years and I apologize to the unnamed equally important ones.

I would like to thank the people in the Department of Biomolecular Systems of the Max Planck Institute of Colloids and Interfaces for their everyday support and advice. Prof. Seeberger I thank for giving me the opportunity to work in his department.

Daniel I especially thank for his supervision, his incredibly supportive manner in all sorts of things and his positive attitude along with his gift to create a good and productive working atmosphere.

I learnt a lot regarding glycopeptide synthesis and about the challenges of being a chemist from Dr. Daniel Varón Silva, Dana and Maurice. I also would like to mention my dear PhD “siblings” Kathir, Uwe, Kathrin, Andreia and Falko as well as my fellows at the institute: Timo, Stephi, Julia, Magda, Andreas, Anika... Thanks for the joy I had during lunch breaks, coffee, beer, swimming at lakes etc.

Kathrin, working with you was always a great pleasure and our conversations left me with joy and new ideas. Your friendly attitude is only matched by your habit to get things right.

I am also grateful for the support by my collaboration partners; without their time and effort the work would not have been possible. Johanna and Kevin introduced me to ion mobility-MS, were never annoyed by my many questions, and helped me with a lot of constructive suggestions. Ulrike, Stuart, Wolfgang,

Peter and others from the Bruker team invested more than expected from a company and helped me immensely by allowing me to make use of all their skills and equipment, broadening my horizon wherever they could.

I appreciate the support I received from the whole team at Genos in Croatia. Gordon I thank for his constructive suggestions, the freedom to pursue my own scientific interests and for creating a pleasant working atmosphere and research environment. The team from the faculty I do not want to miss; you were always supportive and friendly! It was you who made my stay enjoyable and I appreciate your effort in teaching me new things. Petra and Vlatka were also incredibly supportive and always transferred some of their motivation to me.

Last but not least, I thank my friends and family who supported, encouraged and advised me through all the years, whenever it was required.



## List of Abbreviations

3-D	Three-dimensional
Bn	Benzyl
BnBr	Benzyl bromide
Boc	Tert-butyloxycarbonyl
C1GalT-1	Core 1 $\beta$ 1-3 galactosyltransferase
C2GnT1/3	Core 2 $\beta$ 1-6 <i>N</i> -acetylglucosaminyltransferase
C2GnT2	Core 4 $\beta$ 1-6 <i>N</i> -acetylglucosaminyltransferase
CCS	Collision cross section
CGE-LIF	Capillary gel electrophoresis with laser-induced fluorescence detection
CID	Collision-induced dissociation
Cosmc	C1GalT-1 specific molecular chaperone
CTP/CDP/CMP	Cytidine tri-/di-/monophosphate
DMF	Dimethylformamide
Dol-P	Dolicholphosphate
ER	Endoplasmic reticulum
ERMan-I	ER-resident $\alpha$ 1-2-mannosidases
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FFPE	Formalin-fixed and paraffin-embedded
Fmoc	Fluorenylmethyloxycarbonyl
Fmoc-OSu	Fluorenylmethoxycarbonyl succinimide
Fuc	Fucose
FucT	$\alpha$ 1-6-fucosyltransferase
GAGs	Glycosaminoglycans
Gal	Galactose
GalT	$\beta$ 1-4-galactosyltransferase
Glc	Glucose
GlcNAc	<i>N</i> -acetylglucosamine
GlcSidase	Glucosidase
GnT-I	<i>N</i> -acetylglucosaminyltransferase-I
GPI	Glycosylphosphatidylinositol
GTP/GDP	Guanosine tri-/diphosphate
H&E	Haematoxylin and eosin
HCD	Higher-energy collisional dissociation
HILIC	Hydrophilic interaction liquid chromatography
HPAE-PAD	High-performance anion-exchange chromatography with pulsed amperometric detection

---

Hyl	Hydroxylysine
Ig	Immunoglobulin
IgE	Immunoglobulin $\epsilon$
IM-MS	Ion mobility-mass spectrometry
LacNAc	<i>N</i> -acetylglucosamine
LC	Liquid chromatography
LCM	Laser capture microdissection
<i>m/z</i>	Mass to-charge
MALDI-TOF	Matrix-assisted laser desorption/ionization time of flight
Man	Mannose
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
O-GlcNAc	O-linked <i>N</i> -acetylglucosamine
OST	Oligosaccharyltransferase
PGC	Porous graphitized carbon
Pi	Orthophosphate
PP	Pyrophosphate
ppGNTs	Polypeptide- <i>N</i> -acetylgalactosaminyltransferases
RP HPLC	Reserved-phased high performance liquid chromatography
ST3-Gal	$\alpha$ 2-3-sialyltransferases
ST6	$\alpha$ 2-6-sialyltransferase
ST6GalNAcT	CMP-Neu5Ac:GalNAc $\alpha$ 2-6-sialyltransferases
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPPS	Solid phase peptide synthesis
TEA	Triethylamine
TFA	Trifluoroacetic acid
TGN	Trans Golgi network
UTP/UDP/UMP	Uridine tri-/di-/monophosphate
UGT	UDP-Glc:glycoprotein glucosyltransferase
UPLC-FLR	Ultraperformance liquid chromatography with fluorescence detection
VEGF	Vascular endothelial growth factor
Xyl	Xylose
$\beta$ 3GnT	Core 3 $\beta$ 1-3 <i>N</i> -acetylglucosaminyltransferase

## **Selected Articles Used in This Thesis**

**Hinneburg, H.;** Stavenhagen K.; Schweiger-Hufnagel U.; Pengelley S.; Jabs W.; Seeberger PH.; Silva D.; Wührer M.; Kolarich D. The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics. *Journal of the American Society for Mass Spectrometry* **27**, 507-519 (2016).

**Hinneburg, H.;** Hofmann J.; Struwe WB.; Thader A.; Altmann F.; Varón Silva D.; Seeberger PH.; Pagel K.; Kolarich D. Distinguishing *N*-acetylneuraminic acid linkage isomers on glycopeptides by ion mobility-mass spectrometry. *Chemical Communications* **52**, 4381-4384 (2016).

**Hinneburg, H.;** Korać P.; Schirmeister F.; Gasparov S.; Seeberger PH.; Zoldoš V.; Kolarich D. Unlocking cancer glycomes from histopathological formalin-fixed and paraffin-embedded (FFPE) tissue microdissections. In review at *Molecular and Cellular proteomics* (2016).

# Table of Contents

<b>1</b>	<b>INTRODUCTION .....</b>	<b>1</b>
<b>1.1</b>	<b>Glycan Structure Diversity: Monosaccharides and Linkages.....</b>	<b>2</b>
<b>1.2</b>	<b>Glycosylation .....</b>	<b>5</b>
1.2.1	<i>N</i> -glycans .....	9
1.2.1.1	<i>N</i> -glycan Biosynthesis .....	10
1.2.2	<i>O</i> -glycans .....	15
1.2.2.1	Biosynthesis of Mucin-type <i>O</i> -glycans .....	16
1.2.3	Glycosylation in Health and Disease .....	18
1.2.3.1	Protein Glycosylation in Cancer .....	20
1.2.3.2	Formalin-fixed and Paraffin-embedded Specimens as Sample Source.....	23
<b>1.3</b>	<b>Methods to Study Glycans and Glycoproteins .....</b>	<b>24</b>
1.3.1	Mass Spectrometry-based Techniques .....	27
1.3.1.1	Fragmentation Techniques .....	29
1.3.1.2	Ion Mobility-Mass Spectrometry .....	30
1.3.2	Glycan Analysis .....	32
1.3.2.1	Porous Graphitized Carbon-Liquid Chromatography.....	33
1.3.2.2	Negative-ion Mode .....	34
1.3.3	Mass Spectrometric Glycopeptide Analysis .....	34
1.3.3.1	Glycopeptides as Versatile Tools .....	36
<b>1.4</b>	<b>Objective .....</b>	<b>40</b>
<b>2</b>	<b>COLLECTION OF SELECTED SCIENTIFIC PUBLICATIONS.....</b>	<b>41</b>
<b>2.1</b>	<b>Unlocking Cancer Glycomes from Histopathological Formalin-fixed and Paraffin-embedded (FFPE) Tissue Microdissections .....</b>	<b>41</b>
2.1.1	Abstract .....	41
2.1.2	Personal Project Contribution .....	41
<b>2.2</b>	<b>The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics.....</b>	<b>42</b>
2.2.1	Abstract .....	42
2.2.2	Personal Project Contribution .....	43
<b>2.3</b>	<b>Distinguishing <i>N</i>-Acetylneuraminic Acid Linkage Isomers on Glycopeptides by Ion Mobility-Mass Spectrometry .....</b>	<b>44</b>
2.3.1	Abstract .....	44
2.3.2	Personal Project Contribution .....	45

---

<b>3</b>	<b>DISCUSSION.....</b>	<b>46</b>
<b>3.1</b>	<b>Glycomics from Formalin-fixed and Paraffin-embedded Tissue .....</b>	<b>46</b>
<b>3.2</b>	<b>Synthetic Glycopeptide Standards are Unique Tools for Glycoproteomic Method Development and Validation.....</b>	<b>51</b>
3.2.1	Simultaneous Glycan and Peptide Sequencing by Optimizing Collision-induced Dissociation Parameters.....	55
3.2.2	Ion Mobility-Mass Spectrometry for Glycopeptide Characterization .....	56
<b>4</b>	<b>REFERENCES .....</b>	<b>60</b>
<b>5</b>	<b>LIST OF PUBLICATIONS.....</b>	<b>73</b>
<b>6</b>	<b>SUPPLEMENTARY INFORMATION .....</b>	<b>74</b>

## 1 Introduction

Sugars are omnipresent. In fact, the polysaccharides cellulose<sup>1</sup> and chitin<sup>2</sup> are the two most abundant biopolymers on earth. Carbohydrates, proteins, nucleic acids and lipids – the four major biomolecule classes of life – have been studied by researchers for over a century. Historically, carbohydrates were mainly associated with sugar metabolism, energy transport/storage (e.g. starch, glycogen) and structural integrity of cell walls (e.g. cellulose, chitin). Furthermore, sugars such as ribose and deoxyribose are important components of nucleic acids, the fundamental building blocks of the genetic code.

More recently carbohydrates have also been identified to play essential roles in a variety of biological processes within single cells but also in multicellular organisms. Their distinct monosaccharides are connected via glycosidic linkages to form glycans. The term is often utilized synonymously with oligosaccharide (actually 2-20 monosaccharide units) or carbohydrate. Covalently attached to proteins and lipids (then referred to as glycoconjugates), they have been identified to extend the functionalities of their carriers. Glycoproteins and glycolipids incorporated in the plasma membrane of cells form most of the dense, complex layer surrounding the outer membrane of cells. This layer, called the glycocalyx, represents the interface for cell-cell communication and serves as identification tag for other cells but also as protection at the same time. It also represents the first area of contact when dealing with pathogenic intruders. In other words, the glycocalyx is a major zone of interaction where glycans significantly contribute to the communication between different parties (cells, host and pathogens, etc.). The glycan code (“language”) is, however, also exploited by viruses, bacteria and parasites that decorate their outer surfaces with sugar coats for different purposes such as invasion or protection against the immune system. Furthermore, glycosylation alterations on structural as well as quantitative level are often observed in diseases such as cancer and chronic

inflammation<sup>3, 4</sup>. Glycans are very complex biomolecules and constructed by monosaccharides that are typically not used by cells as an energy source. This favors their use as an universal communication tool across organisms and species<sup>5</sup>.

In order to understand the glycan language, the varying parts that reflect the letters, words and grammar need to be studied. Glycans come in different conjugated forms such as glycoproteins, glycolipids, proteoglycans, but also as free oligosaccharides (e.g. in human milk). Understanding their function also requires knowledge on glycan binding proteins such as lectins as well as glycosyltransferases and glycosidases that are mandatory for glycan biosynthesis and degradation. The research field of glycobiology is still in its infancy compared to proteomics (study of the entire protein pool of one single biological source) and genomics (study of the complete set of nucleic acids), probably because glycans belong to the most diverse molecules among the known biopolymers<sup>6, 7</sup>. They consist of several monosaccharides and are joined through different linkages in a non-linear, non-template driven fashion (unlike proteins and nucleic acids), which makes prediction of glycan structures virtually impossible.

Non-template glycan biosynthesis is influenced by external factors such as diet and the physiological conditions at a given time point. Moreover, the assembly is also affected by the temporal actions of specific glycan processing enzymes (proteins) in a spatial-depending manner (cell, tissue, etc.), which makes glycans also a valuable messenger for the global genetic and proteomic status of cells. These dynamics makes glycosylation an interesting disease research target since pathological conditions could be earlier detected, which provides more time for disease treatment<sup>5, 8, 9</sup>.

### **1.1 Glycan Structure Diversity: Monosaccharides and Linkages**

Glycans diversity arises from the complexity already implemented in their smallest units: monosaccharides. The abundance of monosaccharide building

blocks incorporated in glycan structures varies and mainly depends on the type of the investigated glycan and studied organism<sup>10-12</sup>.

Mammals use a set of different monosaccharides to assemble larger glycans (Figure 1). Cyclic monosaccharides, usually formed in solution and in equilibrium with the acyclic forms, cannot only exist as pyranose (six-membered ring) but also as furanose structures (five-membered ring). Furthermore, once the ring is closed the hydroxyl group at the anomeric carbon has a specific orientation and can be either in  $\alpha$  or  $\beta$  position, which in solution are constantly transformed from one state to another through the acyclic form (Figure 2). In contrast to amino acids or nucleic acids, monosaccharides, such as a single aldohexose, contain four chiral centers (CHOH groups) that can all be used to link onto the reducing end of another monosaccharide. In larger structures those aspects can result in a tremendous glycan complexity. During glycosidic bond formation the hydroxyl group at the anomeric carbon becomes permanently fixed in either of the two positions ( $\alpha$  or  $\beta$ ). Furthermore, each monosaccharide can be involved in more than one glycosidic linkage via its hydroxyl groups (Figure 3). Consequently, this can lead to highly complex, branched structures that differ significantly in their biological features<sup>10, 11</sup>. The diversity created through variations in linkage stereochemistry, regiochemistry, branching and monosaccharides can be further modulated by substituting mainly amino or hydroxyl groups with modifications such as sulfation, methylation, acylation and phosphorylation<sup>10, 13</sup>.



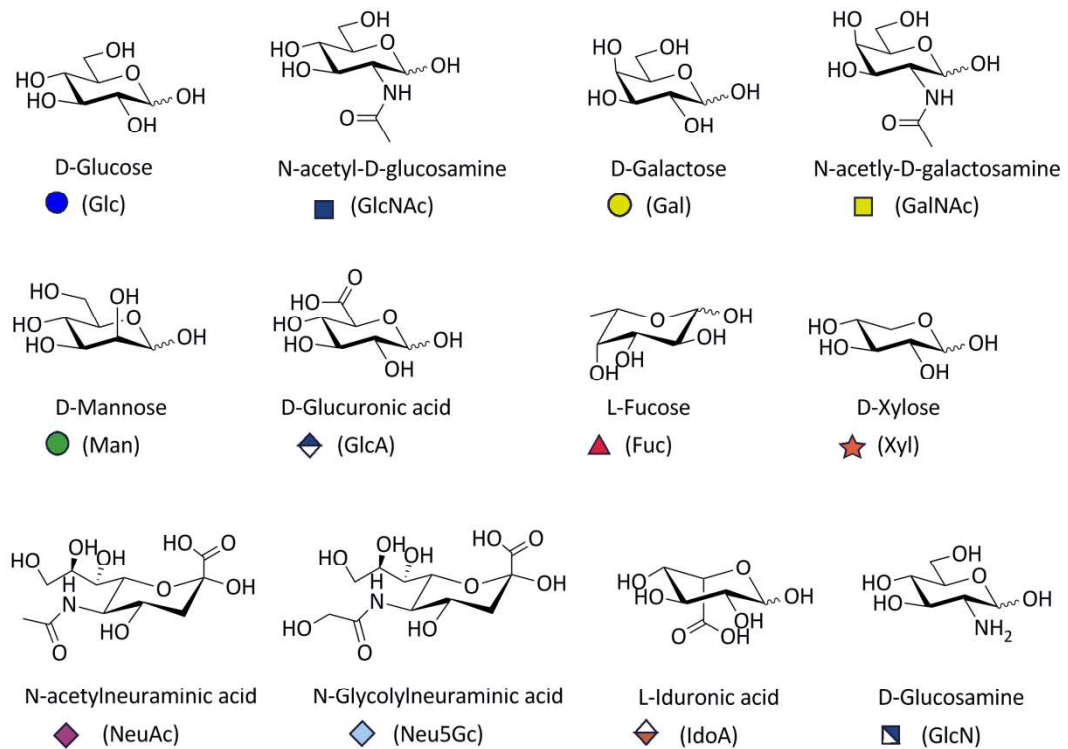


Figure 1: Common mammalian monosaccharide building blocks and respective symbol nomenclature. Abbreviations and symbols will be used throughout this work and are based on Varki *et al.*<sup>14</sup> (Essentials of Glycobiology Editors).

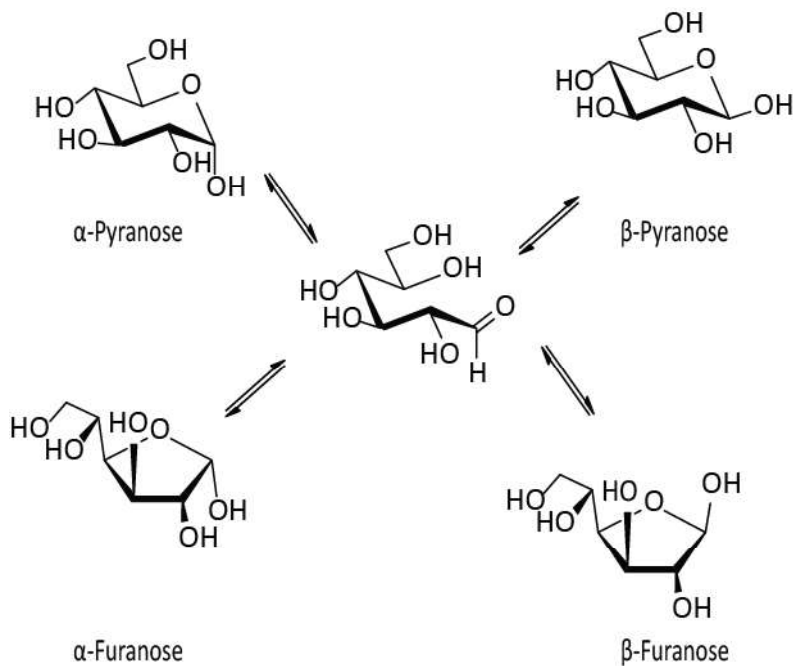


Figure 2: Transition of acyclic D-glucose to its cyclic forms. During this process either pyranose or furanose  $\alpha$  and  $\beta$  structures are observed.

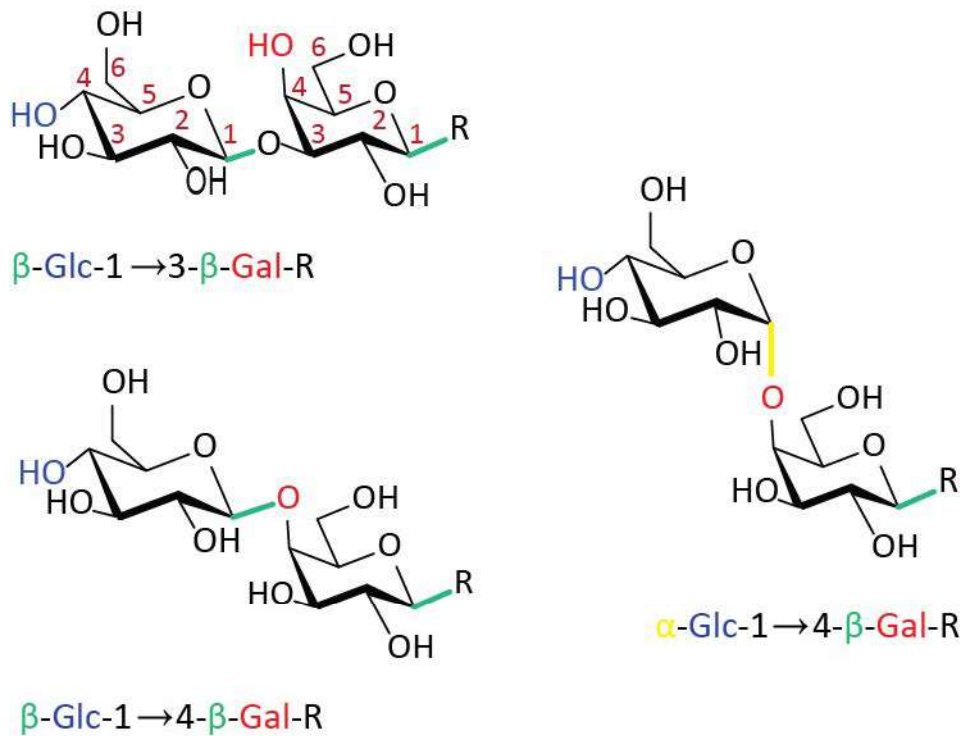


Figure 3: Complexity embedded in carbohydrates. The diversity of larger glycan structures is due to the connected monosaccharides (e.g. Glc vs. Gal, both differ only in one hydroxyl group orientation at the C4-position), the hydroxyl group involved in the connection (e.g. 1 $\rightarrow$ 3 vs. 1 $\rightarrow$ 4) and the stereochemistry of the glycosidic bond ( $\alpha$  vs.  $\beta$ ).

## 1.2 Glycosylation

Glycosylations are abundant post translational modifications with a huge biological relevance and are found on several glycoconjugates (Figure 4). The covalent attachment of glycans to proteins and lipids extends the functional repertoire of their carriers. Attached to proteins several types of glycosylation are differentiated based on the atom the glycan is connected to. It can be attached to the side chain amide nitrogen of Asn residues (*N*-glycans), to the side chain oxygen of mostly Ser/Thr residues (*O*-glycans) or in rarer cases (e.g. collagen) to hydroxylysines (Hyl). Linkages to the carbon atom of Trp (C-glycosidic bond) are less common and the biological relevance of C-glycosylation still needs to be identified<sup>10</sup>. S-glycosidic bonds<sup>15</sup> between sulfur atoms of Cys residues have been reported in bacteria, but are extremely rare. Furthermore, various other types of glycan linkages to amino acid side chains have been described in principle, however the eukaryotic *N*- and *O*-glycosylation are the best studied<sup>10, 16</sup>.

*N*-glycans share a common pentasaccharide core as the initial *N*-glycan precursor biosynthesis is well conserved among eukaryotic species. This highly facilitates structure analysis by mass spectrometric methods. Depending on the various extensions/modifications of this core pentasaccharide *N*-glycans are further classified into high mannose, hybrid and complex-type structures (Figure 5).

*O*-glycans are a more diverse group but also share certain common core motifs. Glycans starting with a GalNAc attached to the protein are termed mucin-type *O*-glycans. Over 20 different types of GalNAc transferases are known in the human genome to catalyze that reaction<sup>17</sup>. This initial GalNAc residue may be extended by different transferases to form eight different core structures, before the glycans can be further elongated. The assembly of *O*-glycans can also start with other sugars such as Gal (e.g. in collagen), Glc and Fuc residues. The latter ones have been first identified on epidermal growth factor-like domains in membrane proteins. Other sugars that can directly be attached to amino acid side chains in proteins are Xyl (see proteoglycans), Man or GlcNAc residues. Starting motifs with Man residues are common in yeasts but to date have only been identified in mammals on certain proteins such as  $\alpha$ -dystroglycan. *O*-linked GlcNAc (*O*-GlcNAc) glycosylation is distinct from the other forms of protein glycosylation. It exclusively targets proteins of the nucleus and cytoplasm and is highly dynamic. Linked to Ser or Thr it carries only a single monosaccharide (termed as monoglycosylation) and functionally shares more similarities to phosphorylation rather than glycosylation<sup>10, 16, 18</sup>.

Proteoglycans have been designated their own category since the protein backbone carries linear glycans of much larger dimension (up to thousands monosaccharides<sup>5</sup>) and biosynthetically, as well as functionally, they differ substantially from *N*- and *O*-glycans. The large, often sulfated carbohydrate structures on proteoglycans are called glycosaminoglycans (GAGs). Proteins can carry GAGs as well as *N*- and *O*-glycans. Some GAGs, such as hyaluron do exist in

free form and are not covalently attached to a protein or lipid. GAGs are connected to proteins via *O*-glycosidic linkages to Ser via an initial Xyl residue, which is part of a tetrasaccharide core sequence. This tetrasaccharide core is subsequently extended by disaccharide repeating units (basis for further classifications) to form linear polymer chains, similar to free polysaccharides. *N*-glycosidic linkages on proteoglycan (e.g. some *N*-asparaginyl-linked keratan sulfates<sup>19</sup>) constructs are also possible<sup>10, 20, 21</sup>.

GAGs can provide a binding reservoir for growth factors (e.g. transforming growth factor- $\beta$ ) and anti-coagulation agents (e.g. anti-thrombin) or pathogens<sup>5, 21</sup>. They can also generate chemokine gradients which influence the infiltration of neutrophils<sup>21</sup>. Furthermore, proteoglycans also play a role in developmental<sup>21</sup> (e.g. inhibition of neuronal attachment and neurite outgrowth) and physiological processes (e.g. regulation of cartilage development, growth and homeostasis)<sup>5, 21</sup>.

Glycolipids form another major class of glycoconjugates that is subdivided into glycolycerolipids, glycosphingolipids and glycosylphosphatidylinositols (GPIs). This differentiation is based on their hydrophobic lipid tail that is incorporated into membranes. Glycosphingolipids are the largest group among the glycolipids and make up the majority of glycans in the vertebrate brain<sup>10</sup>. Glc, and to a lesser extent Gal, residues are linked to the primary hydroxyl group of a ceramide. Both, Glc and Gal can be further extended to form neutral or acidic glycolipid structures. Glycolipids are heavily involved in signaling events and influence the structure of the plasma membrane by initiating receptor clustering (e.g. lipid rafts/micro-domains) and therefore influence their distribution. The dynamic spatial distribution can also facilitate multivalent binding events of glycans to their respective receptors<sup>5, 22</sup>. Glycolipids can also serve as ligands for pathogens (e.g. Simian virus) and toxins (e.g. cholera toxin)<sup>22</sup>.

GPIs fulfil a special role among the glycolipids and their currently known main function is to anchor proteins via a glycan bridge and a lipid tail into the outer leaflet of the plasma membrane. The process is also sometimes referred to as glypiation. GPI anchors have a diverse structure, especially in protozoa, but share a common glycan core. A phosphatidylinositol is bound to a GlcN followed by three Man residues. The last Man is then in turn linked through an ethanolamine phosphate to the C-terminus of a protein<sup>10, 23, 24</sup>. In eukaryotes they are thought to influence the physical localization<sup>25</sup> of proteins. Furthermore, they enable a clear separation of attached proteins to integrated plasma membrane proteins and allow high protein packing. They also seem to be involved in defense mechanisms against the host immune system in certain protozoa<sup>25</sup>.

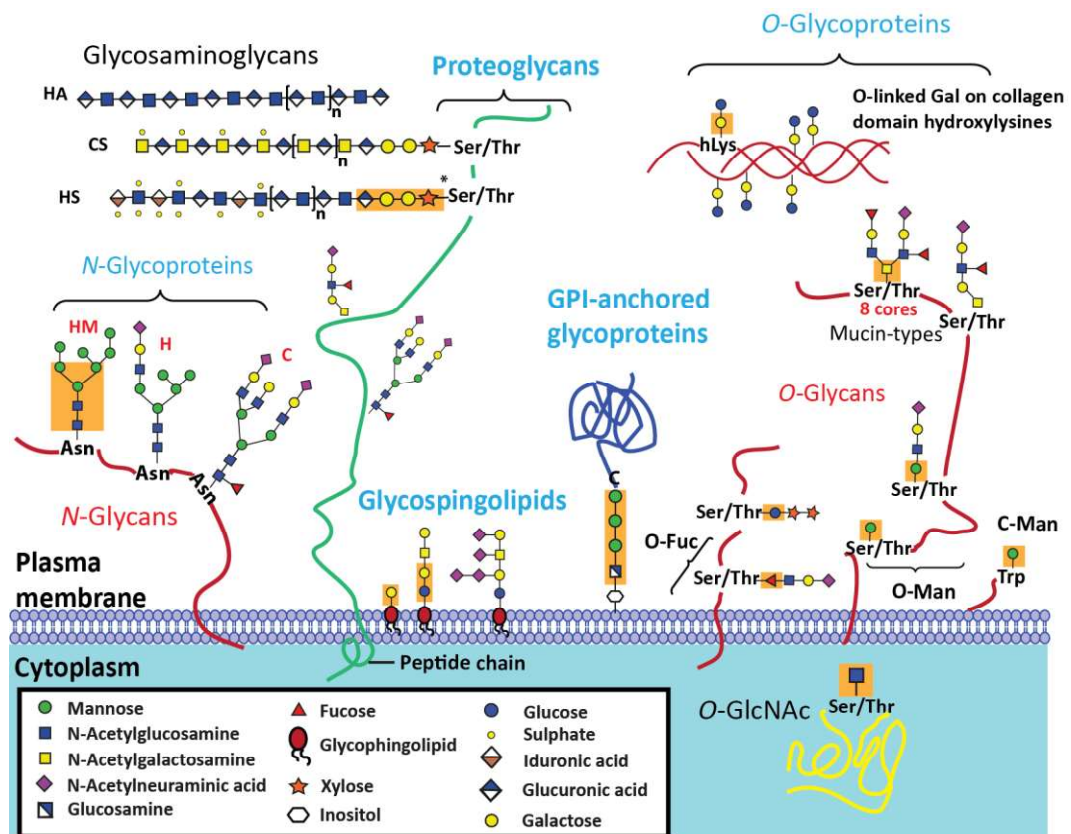


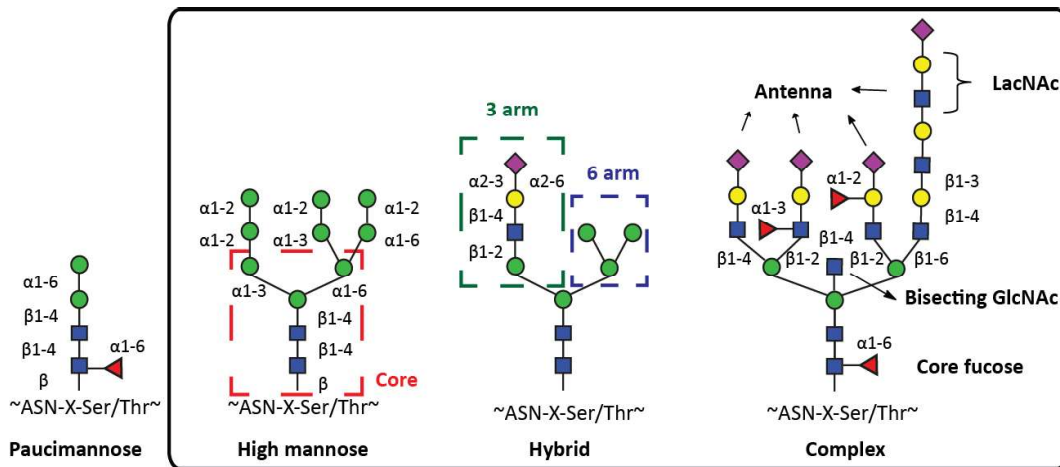
Figure 4: Common glycoconjugates found in mammalian cells. Glycans are attached in various ways to several biomolecules to form macromolecules such as glycoproteins, proteoglycans, glycolipids and GPI-anchored proteins. Common core structures/starting units of the glycan portion of the respective glycoconjugate are shown with orange background. Variations in *N*-glycan processing produce high mannose (HM), hybrid (H) and complex-type (C) structures. Note that not all core elements known for *O*-glycans are shown and that it has been reported that *O*-glycans can also be attached to Tyr in some cases<sup>26</sup>. *O*-glycans starting with Fuc or Glc residues are less common and can be found in extracellular domains of

Notch receptors. GAGs are also mainly attached via *O*-glycosidic linkages to Xyl residues but can also exist in free form (\*). HA: Hyaluronic acid, CS: Chondroitin sulfate, HS: Heparan sulfate.

In the following paragraphs only *N*- and mucin-type *O*-glycans are described in more detail as they have been a major research focus of the work described in this thesis.

### 1.2.1 *N*-glycans

*N*-glycans have been named that way since the glycan is attached via an amide linkage to the nitrogen of the amino acid side chain of Asn. However, not every Asn in a protein sequence can carry an *N*-glycan. *N*-glycosylation usually occurs at the consensus sequence of N-X-S/T, whereas X can be any amino acid except Pro. Less commonly, other motives such as N-X-C<sup>27, 28</sup> or N-X-V<sup>27, 28</sup> can be glycosylated. Also N-G motives have been postulated, however it is still not clear whether they are real glycosylation motives or just a deamidation artefact of the sample preparation<sup>27, 29 16</sup>. To date, however, no clear evidence for glycosylation of an N-G motive has been brought forward. Due to a universal biosynthetic pathway, *N*-glycans possess a common pentasaccharide core structure that is formed by two GlcNAc and three Man residues. Based on the extension made to this core element *N*-glycans are further differentiated. High mannose (oligomannose) structures only carry Man residues attached to the core element. Complex-type structures exhibit Gal, GlcNAc and Fuc residues, as well as sialic acids as the major components on the outer antenna. Hybrid structures are a mixed form where one antenna retained high mannose type features while the other has already been processed to exhibit complex type epitopes (Figure 5)<sup>10, 16</sup>. Another category are paucimannose structures<sup>30</sup>, which are truncated core structures. They are a common feature of insect and plant glycans but have recently also attracted more attention in human glycosylation studies<sup>10, 31, 32</sup>.



**Figure 5: Examples of human *N*-glycosylation.** The three major types of *N*-glycosylation are differentiated: High mannose, hybrid and complex-type structures (black box) share a common pentasaccharide core (red dashed line), which may be extended to form larger structures. Paucimannosic structures resemble truncated glycan versions<sup>30</sup>. Certain architectural elements for glycans are also illustrated. Extensions on the Man residues linked either in  $\alpha 1-6$  or  $\alpha 1-3$  position in biantennary structures are referred to as 6-arm and 3-arm, respectively. Additional attachment of GlcNAc residues to the outer Man residues can produce up to five<sup>33</sup> antennae. A GlcNAc residue attached to the base of the trimannosyl element is called bisecting GlcNAc. Combinations of Gal $\beta 1-4$ GlcNAc are termed *N*-acetyllactosamine (LacNAc). Fuc residues in antennae are important elements in binding motifs (see also Figure 8b). Fuc attached to the first GlcNAc of the core are referred to as core-fucose residues. Other monosaccharides constituting important elements in binding motifs (Figure 8b) and as terminal sugars on the non-reducing end are sialic acids such as NeuAc. In humans the residues are most commonly linked in  $\alpha 2-3$  or  $\alpha 2-6$  linkage<sup>10, 16, 33</sup>.

### 1.2.1.1 *N*-glycan Biosynthesis

The biosynthesis of glycans does not follow any template and depends on various factors such as organism, cell type, protein trafficking and localization<sup>10</sup>. Furthermore, the relative abundance of donors and acceptors as well as coordinated efforts of enzymes such as glycosyltransferases and glycosidases as well as their catalytic preferences influence glycan biosynthesis. Thus, not all glycoproteins assembled in this process are identical, giving every mature glycoprotein potentially slightly different characteristics. Glycans attached to the very same site within a polypeptide chain can vary (micro-heterogeneity) as well as the sites occupied by glycans (macro-heterogeneity)<sup>10, 16</sup>.

*N*-glycan biosynthesis starts in contrast to the one of *O*-glycans as a co-translational process. It can be divided into three stages and happens in the endoplasmic reticulum (ER) and the Golgi apparatus (Figure 6-8):

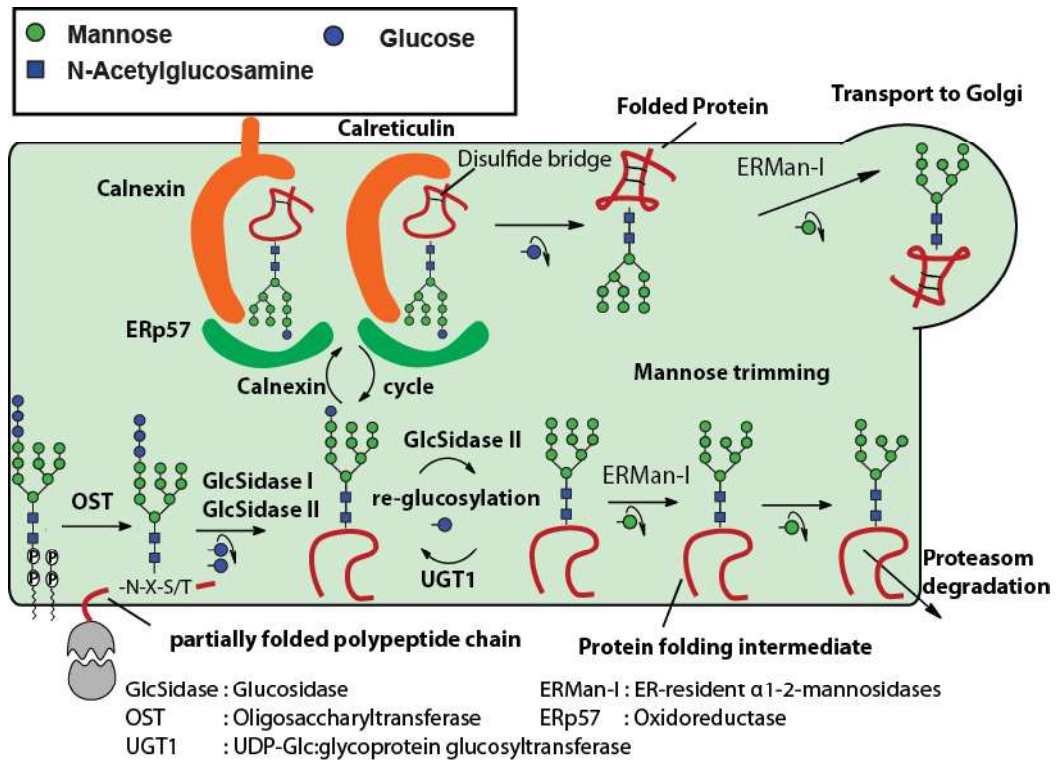
- 1) Formation of an oligosaccharide/dolicholpyrophosphate (Dol-P) precursor. *N*-glycan biosynthesis is initiated on the cytoplasmic side of the ER with the attachment of a uridine diphosphate- (UDP)-GlcNAc residue to the membrane-anchored lipid-like molecule dolichol phosphate. Via the stepwise addition of several nucleotide activated monosaccharide building blocks (Man) catalyzed by specific glycosyltransferases an  $\text{Man}_5\text{GlcNAc}_2$  intermediate oligosaccharide structure is formed on the lipid carrier. This whole construct is enzymatically re-orientated (“flipped”) into the ER lumen by a flippase. Subsequently four Man residues are sequentially added, followed by three Glc residues to form the donor  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$  *N*-glycan precursor that is the substrate for the next steps of protein *N*-glycosylation. The activated monosaccharide building blocks are all synthesized in the cytosol and need to be actively transported into the lumen of the ER. The ER, however, does not use specific transporter proteins for sugar nucleotide transport but imports these substrates after they are linked to a lipid anchor on the cytosol side before they are also flipped into the lumen<sup>10, 16, 34-37</sup>.
- 2) *En bloc*-transfer of the oligosaccharide precursor. The entire  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$  glycan is transferred co-translationally from the lipid donor to defined Asn residues of nascent polypeptide chains by the oligosaccharyltransferase (OST) enzyme complex while the protein is still being synthesis by the ribosome and imported into the ER by the translocon (see Figure 6 for details)<sup>10, 16, 35-38</sup>. Thus the initial steps of *N*-glycosylation are, strictly spoken, co-translational.
- 3) Further glycan processing occurs in the ER and Golgi cisternae. Following the successful glycosylation of the protein chain glucosidases remove the terminal Glc units from the  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$  *N*-glycan and enable the freshly synthesized glycoprotein to enter the calnexin/calreticulin cycle, where correct folding through lectin chaperones is catalyzed. Through an interplay



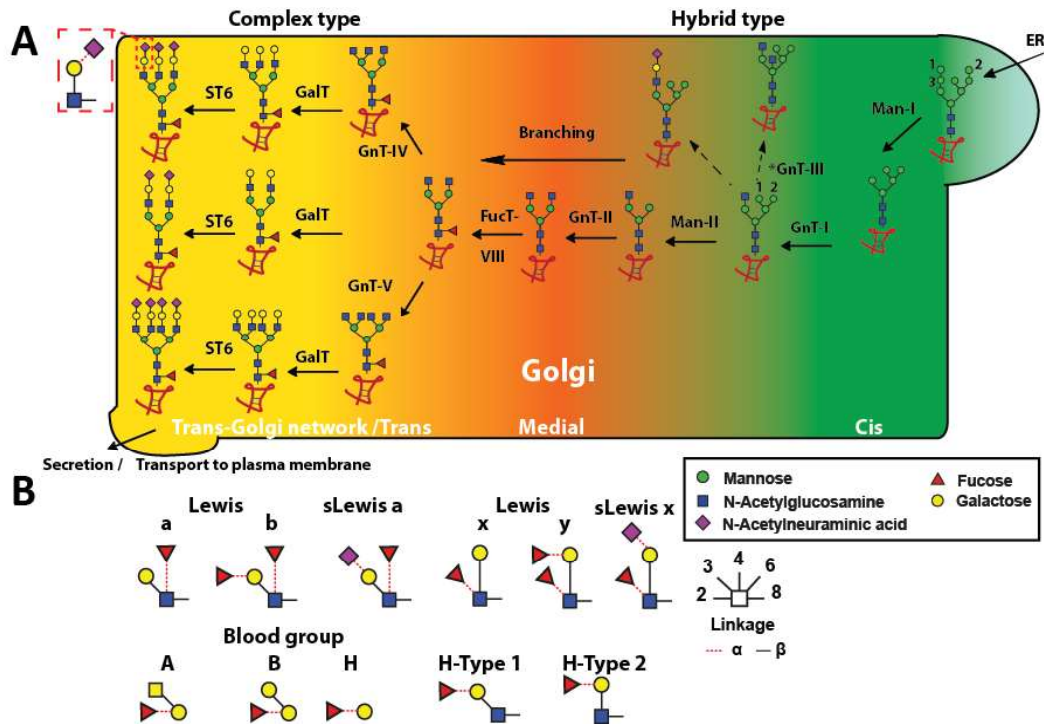
of de- and re-glucosylation the glycosylated polypeptide chains can re-enter the cycle but eventually get degraded in the proteasome when correct folding is not achieved (Figure 7)<sup>10, 16, 33-38</sup>.

Subject to correct glycoprotein folding further *N*-glycan processing takes place in the Golgi apparatus, where different steps are carried out in the different Golgi compartments, also known as cis, medial and trans *cisternae*. In contrast, the ER sugar nucleotide donors are transported via selective transporters embedded in the Golgi membrane. The exact processes in the Golgi are still under debate<sup>16</sup> but they result in the formation of the above described hybrid and complex-type structures that can contain several capping motifs (for further details refer to Figure 8).





**Figure 7: ER located N-glycan biosynthesis.** After the oligosaccharide transfer to the nascent or partially folded polypeptide chains, the glycan part is trimmed by various enzymes in the ER. Two Glc residues are removed through the actions of glycosidase I and II resulting in a  $\text{GlcMan}_9\text{GlcNAc}_2$  intermediate, an important ligand for the chaperone elements calnexin (membrane-bound) and calreticulin (soluble). Both support the correct folding of polypeptide structures in combination with recruited ERp57, a thiol oxidoreductase, which is responsible for disulfide bond formation. Removal of the last inner most Glc residue (glucosidase II), its reintroduction by UDP-Glc:glycoprotein glucosyltransferase1 (UGT1), as well as the iterative removal of the  $\alpha$ 1-2 linked Man residues (ER-resident  $\alpha$ 1-2-mannosidases [ERMan1] and others) determine the time window given to proteins to adopt to their proper three-dimensional (3-D) structure. Folding-defective polypeptides enter the proteasomal degradation pathway, but once they pass the quality checkpoints, correctly folded counterparts are distributed to their intra and extra cellular destinations, usually after the removal of the terminal  $\alpha$ 1-2Man (ERMan1) from the central arm to yield  $\text{Man}_8\text{GlcNAc}_2$ <sup>10, 16, 33</sup>.



**Figure 8: Golgi complex-located *N*-glycan biosynthesis. A) Pathways.** The ER-derived  $\text{Man}_8\text{GlcNAc}_2$  structure is processed by Man-I in the cis-Golgi, which removes  $\alpha 1-2$ Man residues to yield the key intermediate  $\text{Man}_5\text{GlcNAc}_2$ . Trimming is not always complete resulting in High mannose *N*-glycans on mature proteins, which can also arise from alternative routes (e.g. mannose 6-phosphate pathway for lysosomal enzymes). By consecutive action of *N*-acetylglucosaminyltransferase-I (GnT-I), Man-II and GnT-II hybrid and complex-type structures can be synthesized. Resulting structures might or might not be modified by a  $\alpha 1-6$ -fucosyltransferase (FUT8). In addition, the GnT-I product is also a substrate for GnT-III (\*can act also later), which leads to bisected structures. Additional GlcNAc modifications to initiate branching can also occur. Enzymes adding Gal residues ( $\beta 1-4$ -galactosyltransferase [GalT]), further GlcNAc residues (e.g. to form LacNAc motifs), sulfate and *N*-acetylneuraminic acids ( $\alpha 2-6$ -sialyltransferase [ST6]) direct the pathway to complex-type structures<sup>10, 16, 33, 38</sup>. The color gradient depicts the compartmentalization of the Golgi complex. Note that enzyme locations are only rough and that there is evidence that sequentially acting medial- and trans-Golgi enzymes transit as complexes (homo- and heteromers) through the Golgi environment in a pH-driven manner<sup>41, 42</sup>. Numbers on sugar residues indicate the order in which they are removed by the respective enzymes. Dashed arrows illustrate alternative pathways for hybrid structures assembled downstream the *cis* Golgi compartment. **B) Selection of terminal oligosaccharide structures and Lewis antigens.**

### 1.2.2 O-glycans

*O*-glycosylation is a modification occurring on the side chain hydroxyl-oxygen of mainly Ser and Thr, but in rare cases also on Tyr<sup>26</sup> and frequently in collagen on Hyl, with a carbohydrate via a glycosidic bond. In contrast to *N*-glycans, there is no universal consensus sequence motif known and potential glycosylation sites might be located in close proximity to each other. These aspects can impede the

analysis of those structures. However, *O*-glycans usually exhibit less antennas and mucin-type *O*-glycosylation does not contain Man residues<sup>5</sup>.

Mucin-type oligosaccharides represent the best studied *O*-glycan type in mammals. They are characterized by an  $\alpha$ -linked GalNAc residue attached to mainly exposed amino acid residues of Ser/Thr in Ser/Thr/Pro rich regions, depending on the specificity of the GalNAc-Transferases responsible for their attachment<sup>16, 17, 43-45</sup>.

*O*-GalNAc glycans are commonly found on large glycoproteins secreted on epithelia (mucins) but also in other body fluids (e.g. saliva, synovial fluid<sup>46</sup>) and occur also on other proteins. To date, eight core structure motifs (Figure 9) have been described that build the basis for further elongation<sup>43, 47</sup>.

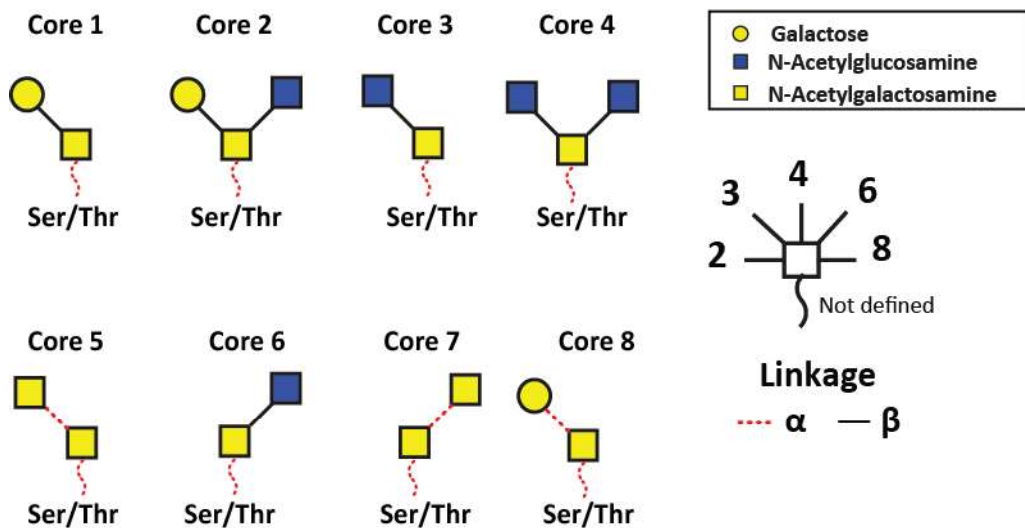


Figure 9: Core structure motifs of mucin *O*-glycans.

### 1.2.2.1 Biosynthesis of Mucin-type *O*-glycans

Mucin-type *O*-glycan synthesis (Figure 10) is initiated by polypeptide-*N*-acetylgalactosaminyltransferases (ppGNTs), a family of about 20 enzymes located in the ER/cis Golgi that catalyze the attachment of a GalNAc to newly synthesized glycoproteins. The motif present on the glycoproteins after this initial step is called Tn antigen, due to its association with cancer and other

diseases. Sequential addition of a Gal residue in  $\beta$ 1-3 linkage gives rise to the T antigen (Core 1) structure. The Tn and T antigen are precursors for other core structures which can then become further prolonged while attached to proteins during their transition through the Golgi complex (for details see Figure 10)<sup>17, 33, 43, 47, 48</sup>

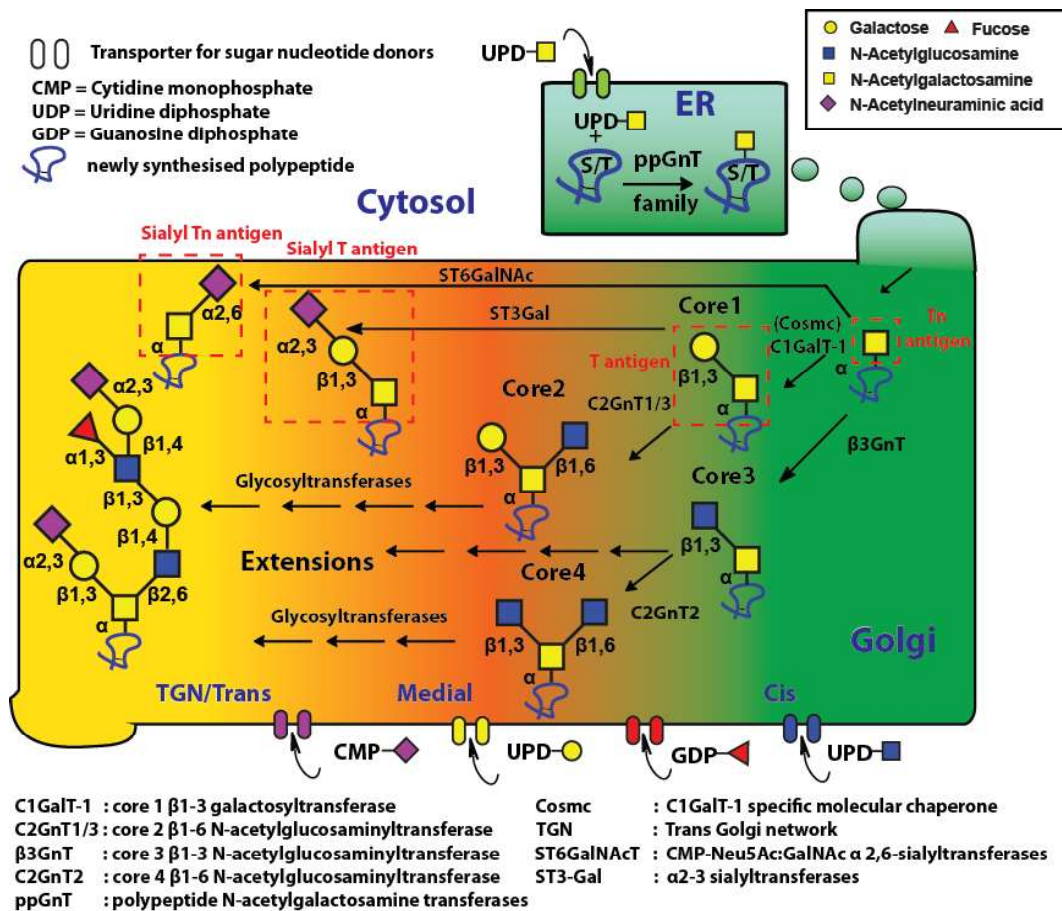


Figure 10: Common mucin-type O-glycan biosynthetic pathways in the ER and Golgi complex (exemplified for core 1–4). The synthesis starts with the action of the ppGnT enzyme family, whose subcellular localization determines GalNAc residue attachment and O-glycosylation densities. The Tn antigen motif then gets processed by C1GalT-1 (T synthase) after its activation by Cosmc, an ER located chaperone. Further GlcNAc residue addition to the so called T antigen by C2GnT1/3 yields the core 2 structure. Core 1 and the Tn antigen can be further extended using sialic acid building blocks, to form the sialyl T and sialyl Tn antigen by the action of ST3Gal or ST6GalNAc, respectively<sup>17, 43, 47, 48</sup>. Alternatively, the Tn antigen can also be processed by  $\beta$ 3GnT to give rise to the core 3 structure and downstream to the core 4 structure (C2GnT2). Further extension of all core structures (e.g. to form the already mentioned Lewis type antigens, the H-epitope or others of the ABO(H) blood groups) are realized by an array of glycosyltransferases and elongated structures can additionally be modified further by sulfation, methylation or acetylation<sup>17, 33, 47, 48</sup>. The color gradient depicts the compartmentalization of the Golgi complex (assignment of glycosylation locations is only rough). Note that the position of nucleotide sugar transporters is simplified and does not reflect the exact position within the Golgi complex.

### 1.2.3 Glycosylation in Health and Disease

Protein *N*- and *O*-glycans are common posttranslational modifications that expand proteome diversity. The majority of membrane and secreted proteins carries these modifications or are at least predicted to do so<sup>10, 49-51</sup>.

*N*-glycans play an essential role in the quality control mechanisms of protein folding as they support correct folding through presenting binding epitopes for lectin chaperones in the ER (see also Figure 6/7). They also influence transit, selective targeting and finally secretion of glycoproteins (trafficking) in processes occurring within the Golgi complex<sup>16, 35, 37</sup>.

The covalent attachment of oligosaccharide to proteins can also influence the stability of their carriers through modulating physico-chemical properties<sup>49, 52</sup> (e.g. pH denaturation, thermolysis, solubility,...), susceptibility to proteolytic degradation<sup>49, 52</sup> and the circulatory half-lives/serum clearance<sup>52, 53</sup>. The enzymatic degradation of glycoproteins can be hindered due to a protecting effect (steric hindrance) mediated by glycans around the peptide backbone near a site of glycosylation<sup>52</sup>. Missing "capping" sialic acids expose underlying epitopes such as Gal residues, indicating the need for recycling<sup>5</sup> of the respective glycoproteins, which are then sorted out by receptors in the liver and other organs<sup>54</sup>. That mechanism can decrease the abundance of partially desialylated coagulation determinates such as Von Willebrand factor generated during sepsis<sup>54</sup>.

The structural integrity and organization of tissues also depends partially on glycans, mostly on proteoglycans. Their possibility to absorb large amounts of water makes them crucial factors required to form gels and mucus within the body. Mucins in particular also serve as physical barriers that prevent bacterial invasions<sup>55, 56</sup>. Glycoconjugates can also be involved in storage of important biological molecules (e.g. growth factors) in the extracellular matrix<sup>57</sup> (see also 1.2, proteoglycans).

The ability of glycans to initiate binding events with other proteins such as lectins has been identified to play fundamental roles mediating biological processes at the glycocalyx interface, such as cell-cell/ cell-matrix interactions<sup>57-59</sup>, receptor recognition (signal transduction)<sup>3, 57, 60, 61</sup>, endocytosis<sup>3</sup>, immune modulation<sup>57, 61-63</sup> and fertilization<sup>64</sup>.

In this context the important role of glycan-binding proteins must be emphasized. The major classes of animal carbohydrate-binding proteins, which themselves have no enzymatic activity but bind their substrates in a highly specific manner are galectins and lectins (C-, P-, I-type [including siglecs]). Another important subclass belonging to the C-type lectin family is the selectin group (P-, E-and L-type). All mentioned proteins are mainly classified based on sequence and structural homology<sup>10</sup> and bind diverse glyco-epitopes.

Galectins are usually soluble, secreted proteins and can be further distinguished based on their architecture (amino acid sequence of their carbohydrate-recognition domain) into proto-, chimera- and tandem repeat-type galectins<sup>7, 10, 65</sup>. Sialic acid-binding immunoglobulin (Ig)-like lectins are called Siglecs and belong to Ig superfamily class of I-type lectins. They are known to discriminate between  $\alpha$ 2-3,  $\alpha$ 2-6 or  $\alpha$ 2-8 linked sialic acids<sup>10, 65</sup>. C-type lectins bind their substrates in a  $\text{Ca}^{2+}$ -dependent-manner and are like siglecs usually membrane-bound<sup>7, 10, 65</sup>. Selectins belong to the group of C-type lectins since they share the respective domain but can be further differentiated based on their cellular location. The three subsets of selectins can be found on platelets and endothelial cells (P-selectin), endothelial cells (E-selectin) and lymphocytes (L-selectin)<sup>10, 66</sup>. P-type lectins recognize Man-6-Phosphate and are crucial for the correct sorting of lysosomal proteins<sup>10</sup>.

Lectins and their ligands play important roles within multicellular organisms. Leukocyte rolling<sup>67</sup> is a fundamental biological event where circulating immune cells adhere to a damaged or activated endothelium through an interaction



involving sialic-acid-motif-recognizing (e.g. sialyl Lewis X) proteins (P-, E- and L-selectin). Without this interaction, leucocytes would be unable to target the areas where they are required to fulfill their duties.

In the case of Ig  $\epsilon$  (IgE) the presence of a particular single high mannose glycan on a specific site of the protein (Asn 394) is crucial for cross-linking mast cells, an initiating event in anaphylaxis<sup>68</sup>. Furthermore, glycans, especially sialic acids, often serve as binding partners for several pathogens, including bacteria and viruses (e.g. influenza virus A, B and C)<sup>69</sup> but are also used by the host to prevent infections or by pathogens to evade detection by the immune system<sup>70</sup>.

Quantitative and qualitative glycosylation changes are often linked to severe inherited genetic disorders such as congenital disorders of glycosylation (CDG)<sup>71</sup> or acquired syndromes such as inflammatory diseases<sup>3, 72, 73</sup>, autoimmunity<sup>74</sup> and cancer<sup>74-78</sup>.

#### 1.2.3.1 Protein Glycosylation in Cancer

Tumor development and progression are of particular interest to scientists not only because of the large patient numbers<sup>79</sup> but also due to widespread evidence that altered glycosylation has functional consequences for survival, invasion and metastasis. Tumor-associated glycosylation changes are in general characterized by an increase of  $\beta$ 1-6 GlcNAc branching, poly LacNAc residues, changes in sialylation and fucosylation. Often it is accompanied with elevated expression of Lewis type antigens but can also result in truncated glycans (e.g. Tn antigen). The appearance of new structures is rarely observed and changes often affect the relative abundance of individual structures. However, due to their importance in cancer development and progression, together with their shedding by tumor cells (e.g. into the circulation), changes in glycosylation are considered to be potential biomarkers that can be exploited for diagnostic purposes<sup>33, 75, 76, 80, 81</sup>.

Tumor cells use various mechanisms to evade the immune system, therapies or to compensate for their changed metabolic needs. It has been demonstrated that increased *O*-GlcNAcylation is associated with cancer, modulating tumor cell signaling<sup>82</sup>. Another example is associated to the tumor cells' requirements for larger amounts of nutrients, usually expressed by an increased blood vessel formation around tumor tissue mediated through vascular endothelial growth factor (VEGF). Some tumors counteract the available antibody based therapy to block VEGF signaling through a prevention of the VEGF interaction with its receptor. Abnormal tissues change their VEGF receptor glycosylation pattern towards more  $\beta$ 1-6GlcNAc branched *N*-glycans carrying low levels of  $\alpha$ 2-6 linked sialic acid residues along with an increased secretion of galectin-1. Thus, mimicking a binding of VEGF (replaced now by galectin) to its receptor and eventually triggering VEGF-like signaling<sup>77</sup>.

Glycans play essential roles in maintaining body functions. Protein glycosylation also changes in pathological events, often in a rapid and dynamic manner. Changes can be based on the metabolism, microenvironment and different expression levels for glycoprocessing enzymes and/or sugar nucleotide donors among others. As a result, glycans have drawn attention as potential biomarker candidates over the last few years<sup>8,9,83</sup>.

Most of the cancer biomarkers available today are glycoproteins, although very few are in clinical use. The entire characterization of a complex pool of glycoproteins using current technologies often requires a choice of focus on either of the chemical distinctive parts of this biomolecule. Furthermore, often low abundance proteins cannot be reliably detected in the presence of a complex matrix background where other (glyco)proteins are also present in several orders of magnitude larger concentration. Furthermore, altered glycan features are not always specific for one single disease state but could be the result of various causes, hampering marker selectivity. Additionally, inter-patient

variability poses additional challenges that need to be understood to translate glycosylation patterns into clinically valid biomarkers. However, several glycan biomarkers have been described over the last few years and are summarized in Table 1. Many of these do not function as a sole marker but are often used as companion biomarkers<sup>8, 83, 84</sup>. Thus, understanding the role glycans and glycoproteins play in complex biological and pathological processes is vital and requires sophisticated methods for their analysis (see 1.3).

Table 1: Glycan biomarkers for cancer diagnosis. Modified from Svarosky *et al.*<sup>83</sup>

Cancer type	Protein	Glycan biomarkers (associated glycan alterations)
Breast	MUC1 (carbohydrate- antigen 15-3)	Expression of truncated Tn, sialy-Tn or $\beta$ Gal-2-3- $\alpha$ GalNAc antigens
	Total serum glycome	$\uparrow$ sialylation, higher levels of sialyl Lewis X, significant changes in fucosylation, $\uparrow$ agalactosylated biantennary glycans
Colorectal	$\beta$ -haptoglobin	$\uparrow$ fucosylation
	Carcinoembryonic antigen and carbohydrate- antigen 19-9	High mannose structures (Hex5-9HexNAc2); complex-type glycans (e.g. Hex3-7HexNAc4-7Fuc0-2NeuAc0-4)
Liver	$\alpha$ -1-antitrypsin, $\alpha$ -fetoprotein	Core ( $\alpha$ 1-6) fucose
Lung	$\beta$ -haptoglobin	Expression of sialyl Lewis X, monoantennary glycans, $\uparrow$ sialylation
	Total serum glycome	$\uparrow$ sialyl Lewis X, $\downarrow$ biantennary core-fucosylated glycans
Prostate	$\beta$ -haptoglobin	$\uparrow$ fucosylation, sialylation, monosialyl tri-antennary structures
	Prostate specific antigen	$\uparrow$ $\alpha$ 2-3 sialic acid and $\downarrow$ core fucosylation; $\alpha$ 1-2fucose and $\beta$ -N-acetylgalactosaminylation
	Serum glycoproteins	Changes in high mannose and fucosylated biantennary complex N-linked glycans
Stomach	Total serum glycome	$\uparrow$ sialyl Lewis X
	IgGs	$\uparrow$ core-fucosylated agalactosyl biantennary glycans

The used biological starting material for biomarker discovery studies can already impact the outcome. Therefore, the choice of the right sample sources needs to be considered before analyses.

#### **1.2.3.2 Formalin-fixed and Paraffin-embedded Specimens as Sample Source**

In the investigation of patient tissue specimens, analyses of biomolecules from fresh or frozen samples, have been carried out for years. Recently, formalin-fixed and paraffin-embedded (FFPE) tissue samples have received more attention since they offer unique advantages. FFPE specimens are widely used in clinics by pathologists around the globe along with dyeing procedures such as haematoxylin and eosin (H&E) staining<sup>85-87</sup>. They allow long-term preservation (through cross-linking of proteins and macromolecules) of tissue, visualization and diagnosis of potential diseases. Consequently, samples are routinely obtained in high numbers and are usually linked to additional data such as patient, treatment and outcome information that are opening up the chance for retro-perspective biomarker studies<sup>85, 86, 88</sup>. Tissue amounts taken during biopsies from healthy or early disease state patients are small and cannot be increased significantly but are valuable sources required for diagnostic and prognostic biomarker research<sup>89</sup>. FFPE tissue can be handled at room temperature and can be further partitioned, minimizing the risk of contaminations/interference with other cell types or tissue other than the subject of interest<sup>89, 90</sup>. Laser-capture microdissection (LCM) of FFPE tissue has recently also been shown to be particularly useful for cell population specific enrichments<sup>90</sup> strengthening disease signature studies and investigating tissue microenvironment.

The proteomic analysis of FFPE tissue is connected with certain shortcomings such as protein modifications which hamper identification and reduced proteolytic digestion, but many of these shortcomings can be circumvented<sup>85, 86</sup>. In general, FFPE samples can be used as a sample source as long as analytes of interest can be extracted and show no interfering modifications and/or

degradation. FFPE tissue analysis has also been successfully demonstrated for various analytes including but not only restricted to proteins<sup>86</sup> nucleic acids<sup>91, 92</sup> and glycans<sup>93, 94</sup> (from large amounts). Latter analyses often focus only on either *N*- or *O*-glycans.

### 1.3 Methods to Study Glycans and Glycoproteins

Glycans play important roles in biological and disease-related events, yet a full understanding of their functions remains to be discovered. The dynamic nature of glycan expression, combined with its non-template based biosynthesis drives the need to develop sophisticated analytical tools for their sequencing. These attempts are hampered by the fact that protein oligosaccharide moieties can include several structural isomers for a single monosaccharide composition. The challenge to analyze the glycan portion becomes even more difficult if simultaneously the entire information is required about a specific glycoprotein. Not all potential glycosylation sites within a protein are necessarily occupied and each individual molecule of the same protein can carry different glycans at a single site of glycosylation, resulting in an array of different glycan structures that can be present at a specific site in the entire protein pool<sup>95</sup>.

*O*-glycan attachment sites are particularly challenging to analyze, since *O*-glycans are not attached to amino acids, part of a consensus peptide sequence<sup>47</sup> and numerous glycosylation sites can be located in very close proximity, giving rise to multiple glycosylated glycopeptides<sup>96</sup> that are obtained after protease treatment. Additionally, there is no enzyme known that has the ability to release all *O*-glycans from their carriers and the chemical release procedures come with various disadvantages. In case of  $\beta$ -elimination, which is usually done under reductive conditions to prevent peeling, the aldehyde group of the reducing-end GalNAc is converted to the alditol (reduced) form of the sugar. This is a problem for all analytical methods using chromophores or fluorophores for labeling the reducing end aldehyde form to be able to analyze trace amounts of glycans<sup>97</sup>.

Hydrazinolysis is another method to release glycans, but the needed chemical is highly toxic and the procedure causes certain acetyl groups to detach from the sugar. Therefore, a re-*N*-acetylation step is needed<sup>98</sup>.

Glycan structure elucidation is usually facilitated when they are released from their carrier proteins and can be accomplished by several techniques such as but not limited to nuclear magnetic resonance (NMR)<sup>99-101</sup>, capillary gel electrophoresis with laser-induced fluorescence detection (CGE-LIF)<sup>102, 103</sup> or mass spectrometry (MS)<sup>102, 104, 105</sup>. There is no universally applicable technique and the optimal technique for a given experiment depends on the required depth of information and the questions that need to be answered. Certain aspects of selected techniques are presented in Table 2.

Table 2: Comparison of methods for glycomic/glycoproteomic analyses, adapted from Huffman *et al.*<sup>106</sup>, Vanderschaeghe *et al.*<sup>107</sup> and Etxebarria *et al.*<sup>84</sup>

Method	Benefits	Drawbacks	Comments
UPLC-FLR	Sensitive*, high resolution, broad dynamic range	Medium throughput,	Can be coupled to MS
	Good isomer separation (neutral and charged glycans analyzed simultaneously)	Labeling required, retention time database needed	
	Very good relative quantification (reproducible retention times)		
CGE-LIF	Very sensitive (fmol range), very high resolution, broad dynamic range	Labeling required, retention time database needed	Coupling with MS not straightforward
	Very good isomer separation, very high-throughput, small sample volume,	Variable retention times possible (might be circumvented by using suitable internal standard)	
	good relative quantification		

<b>HPAEC-PAD</b>	Sensitive (nmol-pmol), broad dynamic range, label free, isomer separation	Time consuming, signal dependent on glycan nature, standards required	Less sensitive than UPLC, no coupling with MS (high salt content)
<b>MALDI-TOF-MS</b>	Sensitive (pmol-fmol), very high resolution, very high-throughput  Label free, no database needed	No isomer separation, isomer differentiation limited  Sensitive to contaminations (less than ESI), medium relative quantification (loss of labile residues possible, ionization issues)	Sialic acids usually require special treatment
<b>LC-ESI-MS</b>	Sensitive (pmol-fmol), very high resolution, good isomer separation, soft technique  Good relative quantification, label free, no retention time database needed (but beneficial)	Medium throughput, sensitive to contaminants  Complex data output, currently limited options for automation	
<b>Microarrays (Lectin)</b>	High throughput, multiplex format, parallel detection  pmol-fmol range  Screen of crude samples including cells and tissue (no glycan removal necessary)  Functional information might be gained	Relatively low sensitivity, no full structural assignment, no sufficiently well-characterized lectins available  Weak interactions might be missed, reproducibility issues  Label required	

**UPLC-FLR:** Ultraperformance liquid chromatography with fluorescence detection; **HPAE-PAD:** High-performance anion-exchange chromatography with pulsed amperometric detection; **MALDI-TOF-MS:** Matrix-assisted laser desorption/ionization time of flight MS; **LC-ESI-MS:** Liquid chromatography electrospray ionization MS. \* Sensitivity depends on used label.

However, not all techniques and aspects of glycan and glycoprotein analyses can be covered here and the following paragraphs are focusing on topics associated with MS-based analytical approaches that have been a major focus in this thesis.

### 1.3.1 Mass Spectrometry-based Techniques

A typical mass spectrometer consists of an ion source, a mass analyzer (e.g. quadrupole, ion trap) and a detector (e.g. electron multiplier). The most common types of ionization used nowadays are ESI and MALDI, which can produce net positive or net negative charged ions depending on the polarity of the applied voltage. Beside the mass-to-charge ( $m/z$ ) ratio of a compound it is also possible to obtain information on the molecular structure of an analyte by specific fragmentation experiments that can be carried out within the instrument. The fragmentation approach (also called tandem MS or MS/MS) gives rise to a very informative fragment or product-ion spectra that, for example in the case of a peptide, allow conclusions on its amino acid sequence. This technique has been crucial for proteomics as we know it today. In that context collision-induced dissociation ([CID], see 1.3.1.1) has become the most widespread dissociation method used<sup>97, 104, 108</sup>.

MS has become a versatile tool for glycoprotein characterization and is unique in terms of sensitivity, robustness and structure determination for high-throughput techniques. Due to the complex nature of glycans and the limitations of currently available analysis techniques a certain reduction of complexity is often required to achieve best results.

Depending on the scientific question the focus can be on the analysis of released glycans, on non-glycosylated peptide species including proteolysis products and enzymatically deglycosylated peptides or on intact glycopeptides (Figure 11). One of the first steps in MS-based techniques is usually the enrichment of glycoproteins via affinity or physicochemical properties-exploding strategies, such as lectine or hydrophilic interaction liquid chromatography (HILIC),



respectively. In most cases the enriched protein will be subjected to digestion with specific proteases, and alternatively glycans can also be released from their carrier proteins (see also 1.3.2). The combination with separation techniques such as reversed-phased high-performance liquid chromatography (RP HPLC), HILIC or porous graphitized carbon (PGC)-LC, in combination with different instrument types, ionizations techniques and fragmentation strategies MS has developed to be a powerful and indispensable analytical tool for life sciences. However, to date none of the approaches alone is able to provide all information important for detailed molecular characterization and understanding of protein glycosylation<sup>97, 109</sup>.

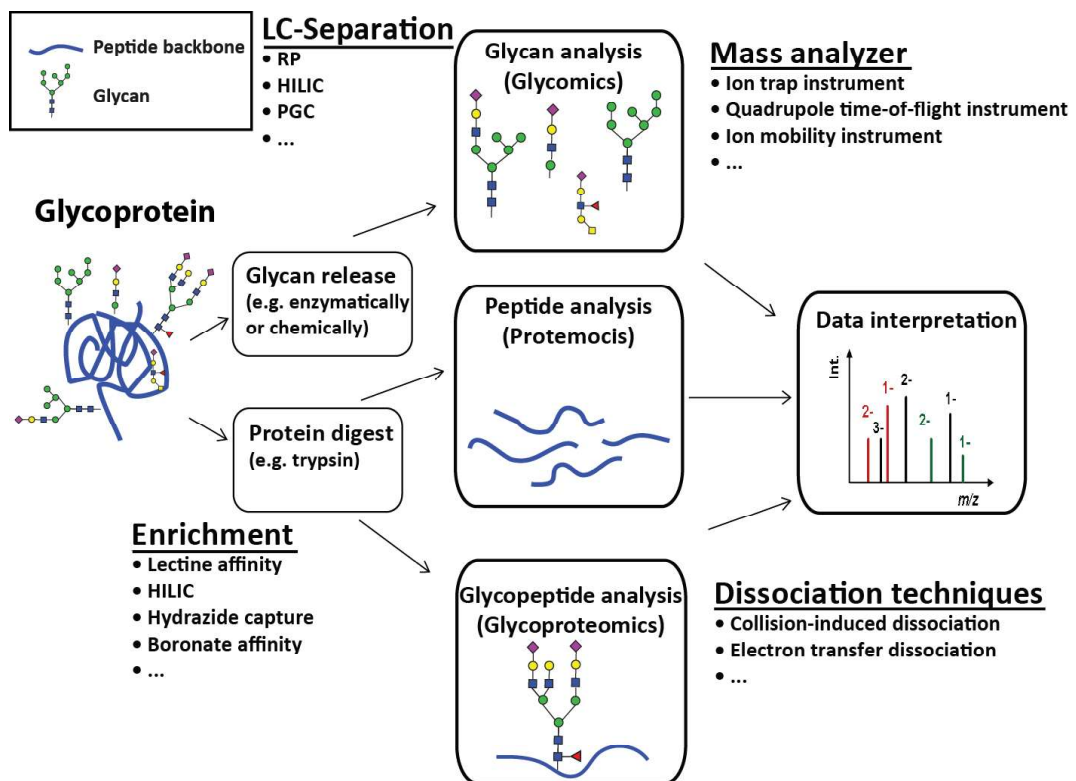
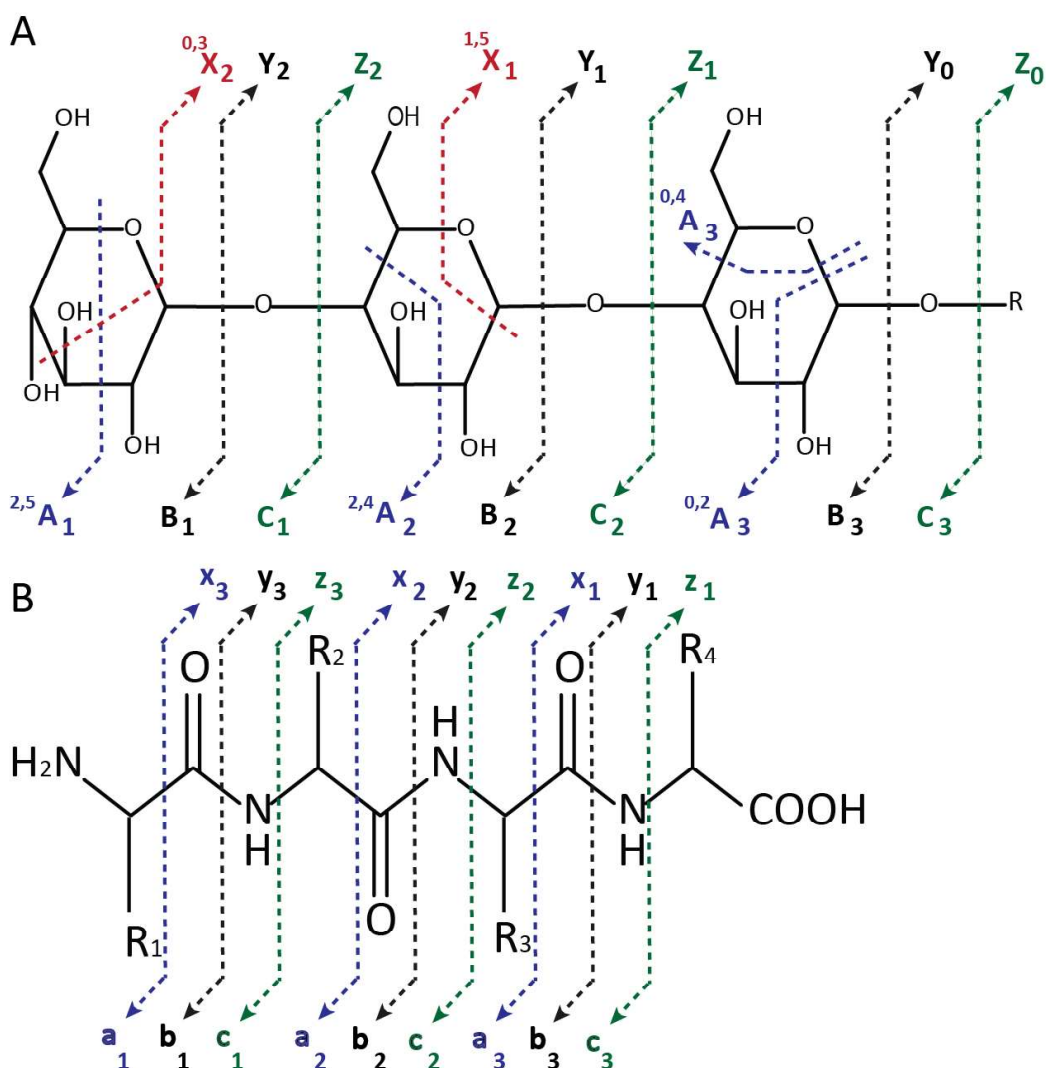


Figure 11: Schematic overview of glycoprotein characterization approaches. Different techniques (glycomics, proteomics and glycoproteomics) can be employed separately or in combination for glycoprotein characterization with numerous further methodical diversifications (e.g. use of various enrichment techniques [on protein/peptide level], labelling strategies, proteolysis, LC-separations, mass spectrometers, ionizations types, dissociation procedures etc.). Adapted from Kolli *et al.*<sup>109</sup>

### 1.3.1.1 Fragmentation Techniques

In CID, the translational energy a precursor acquires through collisions with a buffer gas is converted into internal energy causing the molecule to fragment. In “low-energy” CID, conducted at energies of around  $10^{-1}$ - $10^2$  eV, the precursor molecule is exposed to multiple collisions before dissociation. In “high-energy” CID ( $10^3$ - $10^4$  eV) one impact might be enough to surpass the dissociation threshold. Besides CID there are also other fragmentation techniques available that use different mechanisms for molecule dissociation. In electron transfer dissociation (ETD) peptide bond rupture occurs after the transfer of low-energy electrons onto multiple charged precursors. Analyte cations and the negatively charged reagent anions such as fluoranthene, generated by chemical ionization in a methane atmosphere, undergo gas-phase ion/ion reactions and need an appropriate delivery into the same space in time<sup>10, 95, 97</sup>.

Interestingly, in glycopeptide fragmentation CID favors fragmentation of the carbohydrate moiety, producing so called B- and Y-type ions, but no or little peptide backbone fragments that would be informative on the peptide sequence. The opposite is happening when ETD fragmentation is applied, giving preferably rise to c- and z -type peptide ions that are generated by N-C $\alpha$  bond scission while remaining the glycan intact, allowing unambiguous identification of oligosaccharide attachment sites<sup>95, 97</sup>. For ease of description and interpretation of product-ion spectra arising from peptide<sup>110-112</sup> and glycan<sup>113</sup> precursors certain ion nomenclatures have been introduced (Figure 12).



**Figure 12: Fragmentation nomenclatures. A) Carbohydrates.** Series of X, Y and Z ions describe fragments retaining the charge on the reducing terminus and are numbered starting also from the reducing end. Series of A, B and C ions are numbered starting at the non-reducing terminus; where also the charge is retained. A and X ions refer to cross-ring fragments, whereas B, C, Y and Z can be observed only with cleavage of the glycosidic bond. Superscript numbers indicate bond cleavages for cross-ring fragments. **B) Peptides.** Ions that retain the charge at the amino-terminal part are called either a, b or c ions. If the charge is retained at the carboxy-terminal then fragment ions are classified either x, y or z. Subscript numbers refer to the number of R groups these ions contain; for a, b and c ions the labelling starts at the amino terminus and for x, y and z ions at the carboxyl terminus. For glycopeptide fragmentation description both nomenclatures are combined. For differentiation between peptide and glycan fragments, all peptide fragments are marked in lower case letters and all glycan fragments are marked in upper case letters. Note: In glycopeptide fragmentation Y-type ions refer to the glycan fragments which are still attached to the peptide backbone.

### 1.3.1.2 Ion Mobility-Mass Spectrometry

Ion mobility based separation techniques are one relatively new development that has been introduced into mass spectrometry instruments. It provides an additional level of separation prior to the mass spectrometric analyses that is

based on the charge, size and shape of an analyte while traversing through a gas-filled ion mobility cell (Figure 13). This allows separation of isomeric/isobaric structures and thus individual detection and analysis by the mass spectrometric detector that otherwise might not be separable for the mass analyzer alone. In ion mobility-mass spectrometry (IM-MS) condensed ions undergo less collision with the inert neutral buffer gas compared to molecules that adopt a more extended shape in the gas phase. The mobility of an ion is related to the biophysical properties and overall shape of the molecule. The drift behavior can in turn be used to determine collision cross sections (CCSs), an instrument independent value<sup>114</sup>.

IM-MS has already been successfully employed to characterize isomeric carbohydrates and glycoconjugates and provided extensive structural insights, including linkage and anomeric information<sup>115-119</sup> and is a very promising technique suited for the in-depth analysis of complex glycoconjugates. Interestingly, in certain instrument settings ions can be fragmented before and after the mobility cell allowing either mobility separation of the parent or fragment ions.

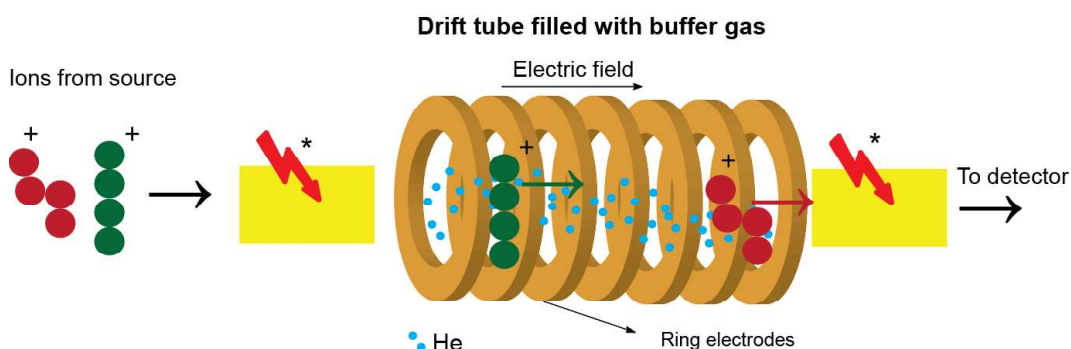


Figure 13: Ion separation occurring in an ion mobility cell. Ionized molecules travel along an electric field through a gas filled cell where they are separated. Separation is based on their charge, shape and size. Molecules with a more compact structure leave the cell faster. \*Depending on the type of instrumentation ions can also be fragmented before and/or after the ion mobility cell, providing additional possibility for glycoconjugate analyses.

### 1.3.2 Glycan Analysis

Analysis of protein bound glycans usually includes their release from the protein backbone. This can be achieved by either enzymatic or chemical techniques. Often the released glycans are subjected to chemical derivatization (e.g. with the fluorescent tag 2-aminobenzamide or permethylation) or modification (e.g. reduction) of the released glycans. Labeling and/or modification of glycans can be done to increase the detection sensitivity, stabilizing analytes or reduce signal complexity. Although this approach allows an in-depth structural oligosaccharide analysis any information about the site of attachment to the protein, as well as the protein the glycan is derived from, is lost unless single purified proteins are analyzed<sup>98, 120</sup>.

Specific enzymes exhibiting broad sensitivity towards the majority of known *N*-glycans are available for their release. Peptide-N-Glycosidase F is the most prominent enzyme applied for this purpose. It cleaves all *N*-glycans between the reducing-terminal GlcNAc residue and the Asn residues of glycoproteins unless they carry an  $\alpha$ 1-3 linked core fucose. *N*-glycans carrying this modification can be enzymatically released using PNGase A, which is of particular importance if non-mammalian species are studied<sup>10</sup>. Nevertheless, as this modification is not present in mammals PNGase F is usually the enzyme of choice for the analysis of mammalian glycoproteins. Next to these entire glycan releasing enzymes specific exoglycosidases are valuable tools that release monosaccharides from the non-reducing end of glycans. They are of particular interest for oligosaccharide sequencing purposes.

To date no broad-specific enzyme has been characterized that is able to release all types of *O*-glycan structures from glycoproteins. Thus, researchers usually apply chemical approaches such as reductive  $\beta$ -elimination or hydrazinolysis<sup>98, 120</sup> to release *O*-glycans from the respective proteins. Once glycans are released, various other analysis and separation techniques are frequently combined with

mass spectrometric detection for comprehensive glycan sequencing. Liquid chromatography is the most frequently coupled separation technique that is combined with MS<sup>98, 102, 104</sup>. Different stationary phases such as HILIC<sup>98, 102, 121</sup> and PGC<sup>46, 102, 105, 122</sup> can be online coupled with MS detection for the released glycans.

### 1.3.2.1 Porous Graphitized Carbon-Liquid Chromatography

Two properties of PGC are believed to influence the chromatographic behavior of analytes: 1) The planarity of the PGC surface and 2) PGC's sensitivity to the electronic density of the solutes.

The 3-D structure of a solute, its flexibility and ability to adapt to a planar conformation, influences its interaction and thus retention by PGC. The planarity possibly promotes solute-adsorbent contacts and affects the extent of dispersive interactions. Besides the steric aspects, electrostatic interactions also influence analyte retention. One explanation is based on the delocalized electrons in the graphite ribbons and the consequential high potential for polarizability at the surface, if polar compounds approach. The phenomenon is affected by the distance of the analyte to the surface and likely by the 3-D orientation of the polar functional groups. Another theory explains the phenomenon by electronic repartition, giving rise to an electron excess (negative charge) near the edges of graphite planes and an electron-deficient area in the center (positive to close to neutral charge)<sup>123</sup>.

However, PGC is a superior material for separation of isobaric glycans, even if the molecular interactions are not fully understood<sup>46, 105, 124</sup>. The absolute retention times of carbohydrates in PGC-LC are difficult to predict but are subjected to certain rules. In general, the higher the numbers of monosaccharides, antenna and negatively charged residues within a glycan, the stronger these glycans are retained by the PGC stationary phase. Core fucosylation also results in later elution times compared to non-fucosylated and antenna-fucosylated

analogues<sup>125</sup>. Furthermore, NeuAc linkage isomers can be well separated.  $\alpha$ 2,6-sialylated glycans elute well before their  $\alpha$ 2,3-sialylated counterparts, whereas topology isomers (same sialic acid on different antenna/arm) elute close to each other<sup>124, 126</sup>. The chromatographic behavior of glycans is also influenced by the presence or absence of a bisected GlcNAc. Bisected *N*-glycans elute earlier compared to more branched structures<sup>46</sup>. Noteworthy, higher column temperatures also increase the retention of oligosaccharides<sup>124</sup>.

### 1.3.2.2 Negative-ion Mode

Positively or negatively charged ions can be generated for subsequent analysis by mass spectrometric techniques. The combination of PGC-LC with tandem MS of negatively charged ions provides in-depth structural information, in particular on released glycans<sup>122</sup>. Product-ion spectra obtained by CID of negatively charged-ions contain detailed information on the structure of reduced glycans. Similar as described for the fragmentation of peptides, the different fragment ions are assigned to specific ion series that are assigned based on whether the ions are derived from the reducing or non-reducing end. Thus, these ions are assigned as B- and Y-type ions, but can also contain so called C-type glycosidic and A-type cross-ring fragments (Figure 12A). Antenna-specific D-type ions (reflect 6 antenna) and E-type ions (reflect 3 antenna) are also observed and are highly informative for glycan structure interpretation. The withdrawal of a proton from hydroxyl groups or acidic functions of a carbohydrate structure gives rise to those diagnostic fragments usually absent or with very low abundance when analyzed in positive-ion mode<sup>127-130</sup>. The typical fragments arising from glycan fragmentation in negative-ion mode can be used to elucidate *N*-<sup>120, 127-135</sup> and *O*-glycans<sup>135-138</sup> structures.

### 1.3.3 Mass Spectrometric Glycopeptide Analysis

Different analytical approaches need to be combined to gain a maximum of information on the primary structure of glycoproteins. In ideal glycoproteomics

studies proteins are identified based on the amino acid sequence information, along with the characterization of the attached glycans in a site specific manner. One approach to acquire combined information on glycans and proteins is the analysis of the attached glycans directly on the proteins (top down approach) or proteolytically generated peptides (bottom-up approach) without their prior release<sup>46, 97</sup> (see also Figure 11). Due to the diverse chemical properties of the peptide and the glycan moieties of glycopeptides the simultaneous and detailed molecular characterization of both parts comes with additional challenges but holds the possibility to also obtain site-specific information. In mass spectrometry one obstacle is to obtain good quality product-ion spectra that provide useful information on both moieties within a single tandem MS experiment<sup>97, 139</sup>. The benefits of a comprehensive glycopeptide approach is that the obtained additional information can be beneficial to increase the sensitivity and selectivity of glycoprotein biomarker candidates<sup>8, 140, 83</sup>.

Liver dysfunctions such as liver fibrosis, cirrhosis and hepatocellular carcinoma (HCC) are also characterized by distinguished glycan profiles of serum proteins. In HCC it has been shown that the diagnostic power of core  $\alpha$ 1-6 fucosylated  $\alpha$ -fetoprotein was increased compared to just taking  $\alpha$ -fetoprotein itself into account<sup>8, 118, 141</sup>. This example clearly shows the importance to include glycomics data in systems biology studies to gain comprehensive insights in biological functions and dysfunctions<sup>46, 140</sup>.

A thorough glycopeptide characterization includes the exploration of the amino acid sequence for subsequent protein identification but also the extensive elucidation of the carbohydrate structure along with its site of attachment. Conventional CID experiments yield carbohydrate B- and Y-type ions (Figure 12A) by rapture of the glycosidic bonds between monosaccharide units. Within a glycopeptide the energy barriers to produce sufficient peptide b- and y-type ions (Figure 12B) are higher, resulting in a very low abundance (if at all) of those ions



in glycopeptide fragment spectra<sup>97, 139, 142, 143</sup>. These fragments, however, are necessary for unambiguous protein identification and thus one option to extract a maximum of information of just product ions spectra is to optimize glycopeptide fragmentation conditions so that B- and Y-, but also b- and y-ions are generated in sufficient intensity.

Numerous approaches have been described for the structural characterization of glycopeptides using mass spectrometry, and a comprehensive overview can be found in the excellent reviews by Wuhrer *et al.*<sup>139</sup> and Alley *et al.*<sup>97</sup>.

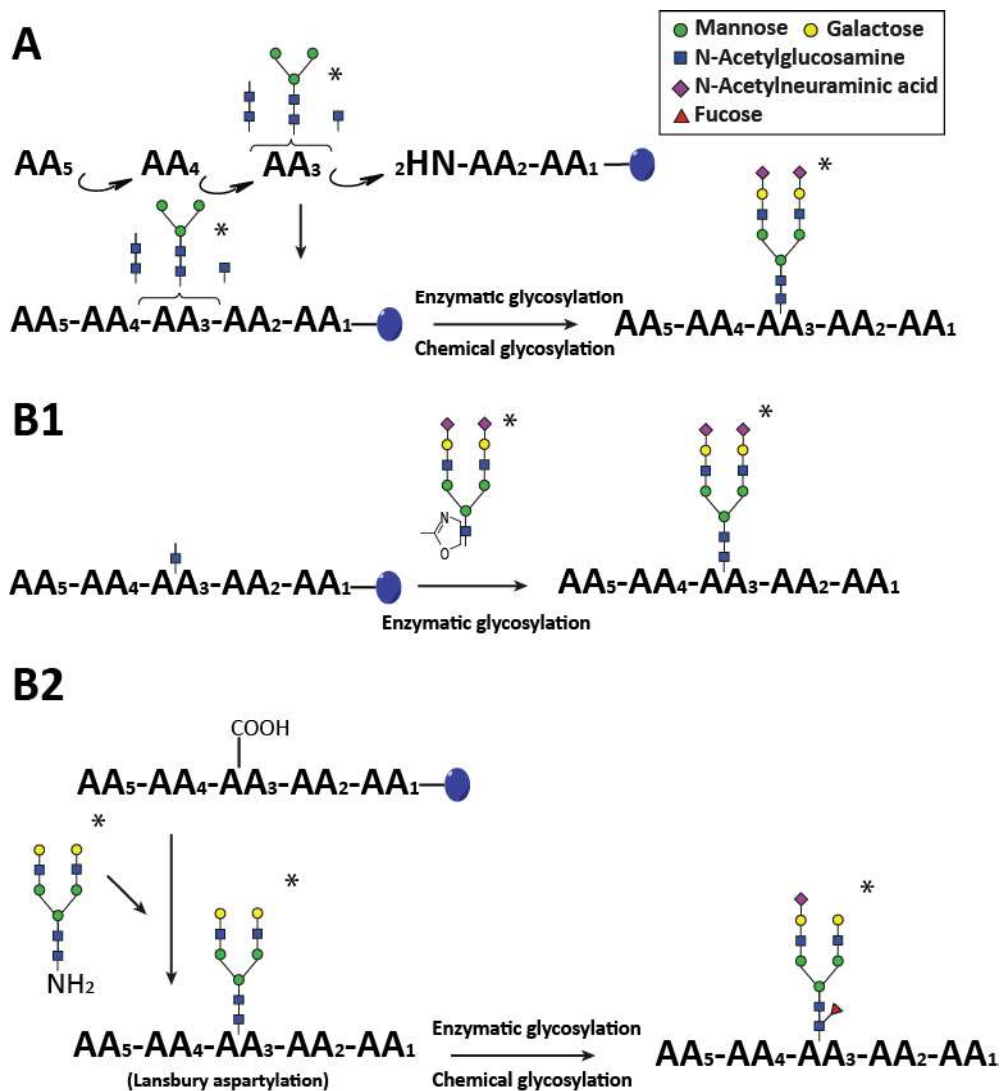
### 1.3.3.1 Glycopeptides as Versatile Tools

Model proteins and standards are crucial tools for method development and optimization of mass spectrometric techniques. Based on such standards structural characterization of glycopeptides can be improved and novel technologies developed<sup>143-145</sup>. However, preparation of sufficient amounts of glycopeptides from model glycoproteins is not an easy task and might not always provide the optimal compounds depending on the particular questions that need to be answered. Glycopeptides with defined peptide sequence, carrying large oligosaccharides such as sialylated, complex-type *N*-glycans are also not commercially available. Only a handful of groups synthesize them routinely. As the production of synthetic *N*-glycopeptides was one central theme of this thesis the following passage briefly describes the most important aspects of their synthesis.

#### Glycopeptide synthesis

Depending on the glycopeptide intended to be synthesized (peptide/glycan sequence) the process encompasses chemical, enzymatic, and molecular biological approaches used in combination or independently to produce these model compounds. In principle, two general strategies have been described: the stepwise synthesis and the convergent synthesis (Figure 14). In solid phase

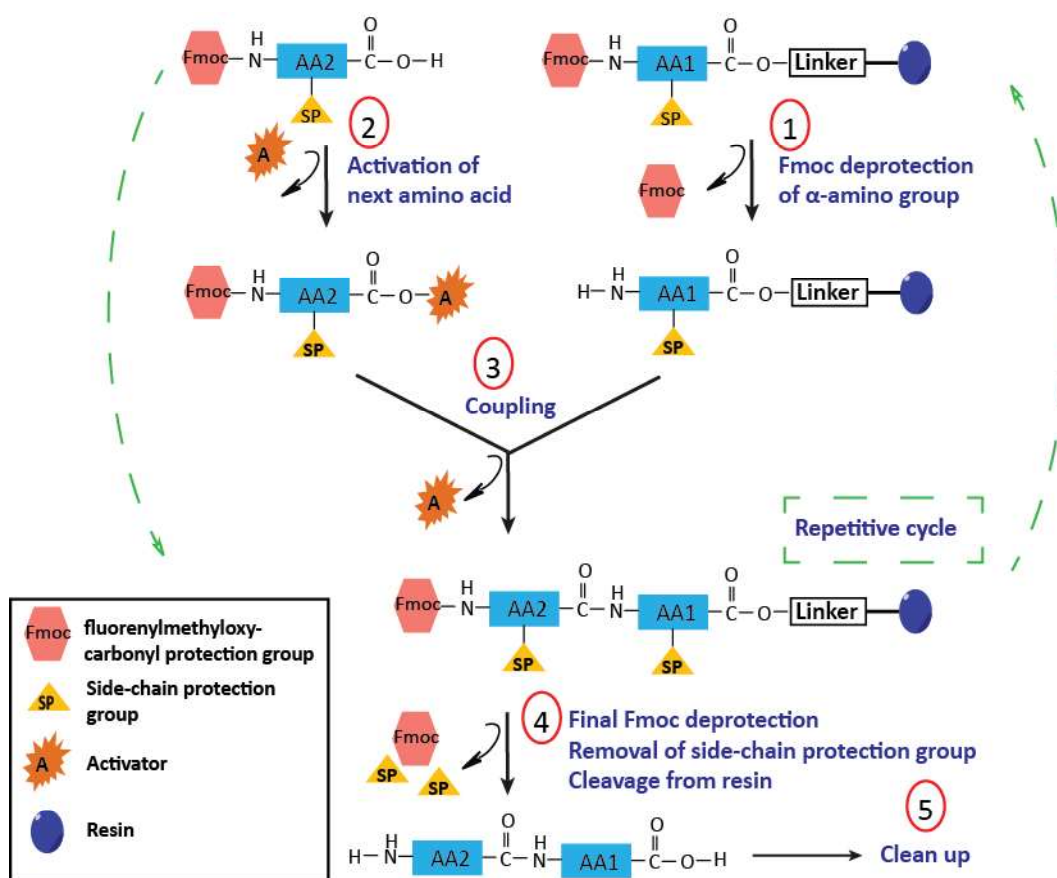
peptide synthesis (SPPS) traditionally the tert-butyloxycarbonyl (Boc) and the fluorenylmethyloxycarbonyl (Fmoc) methodologies are the ones most often used. In glycopeptide synthesis the Fmoc-based strategy (Figure 15) is favored due to the acid instability of glycosidic linkages. During this procedure hydroxyl groups of the carbohydrate portion are protected via acetyl or benzyl (Bn) groups<sup>146-149</sup>.



**Figure 14: Principle of glycopeptide synthesis strategies.** **A) Stepwise synthesis of glycopeptides.** In this linear assembly a glycosylamino acid is used directly in SPPS. The pre-formed glycosylamino acid may consist of a single monosaccharide, a disaccharide or elongated “mature” glycan etc. The structures can then be further modulated after SPPS. **B) Convergent synthesis of glycopeptides.** In this approach the complete peptide backbone is first synthesized either incorporated with a GlcNAc (**B1**) or as pure peptide backbone (**B2**). In the first coupling approach the amino sugar is elongated enzymatically (e.g. Endo- $\beta$ -N-acetylglucosaminidase family) by *en bloc* transfer of a large oligosaccharide. For the second coupling

strategy, the amino acid at the site of “future” attachment is incorporated with an orthogonal protecting group; enabling the selective deprotection only at this side chain function. The remaining protection groups stay intact until the final deprotection. In this protocol a glycosylamine is then merged to the peptide backbone. Following conjugation, glycan structures may be further modulated by glycosidases and/or glycosyltransferases. \*Only representative.

The synthesis strategy used by Yamamoto *et al.*<sup>150</sup> to produce a glycopeptide carrying a complex-type biantennary, sialylated *N*-glycan is particular appealing, since the total synthesis of the glycan moiety can be avoided. As it was also used in this work this approach is briefly described<sup>151, 152</sup>.

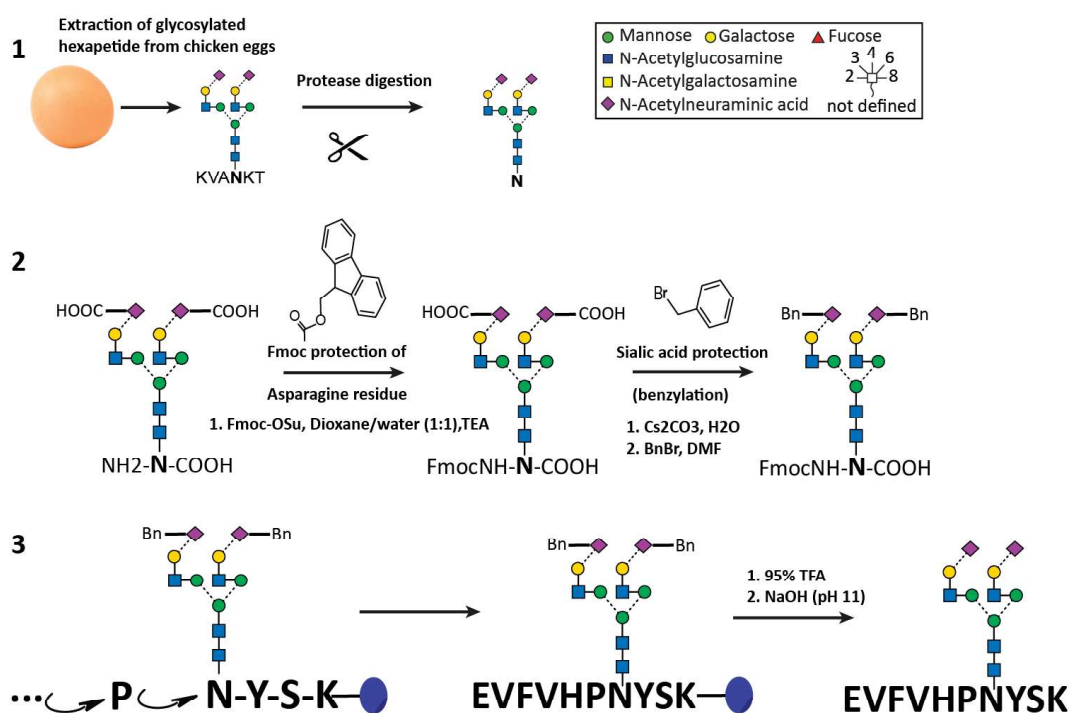


**Figure 15: Fmoc-based solid phase peptide synthesis.** For the synthesis protected amino acids (N-terminus and side chains) are used after activation. They are coupled to the Fmoc deprotected N-terminus of amino acids, already attached to other amino acids or via a linker to a resin. The cycle of Fmoc deprotection, activation and coupling goes on until the last amino acid of the polypeptide chain is connected. Final deprotection and cleavage from the solid support are performed simultaneously using acidic conditions yielding the desired unprotected polypeptide chain.

The whole workflow can be roughly divided into three parts: 1) The isolation of a glycan containing building block; 2) the protection of vulnerable functional

groups and 3) the SPPS that is performed in a stepwise manner using the fully protected building block (Figure 16).

After isolation of the sialylglycopeptide from chicken egg yolks, the six amino acid containing peptide backbone ( $\text{NH}_2\text{-Lys-Val-Ala-Asn-Lys-Thr-COOH}$ ) is fully digested by pronase E (actinase E) to obtain a single Asn residue carrying the glycan moiety. Subsequently the amine function is Fmoc protected, followed by a benzyl protection of the carboxylic acid groups present on the sialic acids using a selective esterification approach. The fully protected building block can now be incorporated into any peptide sequence using standard SPPS. Once the peptide is fully synthesized it is cleaved from the resin by trifluoroacetic acid (TFA) and the benzyl esters are hydrolyzed by sodium hydroxide treatment to yield the desired unprotected glycopeptide.



**Figure 16: Chemical synthesis of sialylglycopeptides.** The procedure can be divided into three parts: **1)** Preparation of the unprotected, biantennary sialylglycopeptide starting from egg yolk. **2)** Sequential introduction of the Fmoc and the Bn protection groups. **3)** SPPS in a stepwise manner (see also Figure 14A) using the biantennary building block, followed by final deprotection steps. DMF: Dimethylformamide, BnBr: Benzyl bromide, TEA: Triethylamine, Fmoc-OSu: Fluorenylmethoxycarbonyl succinimide.

## 1.4 Objective

- 1.) Development of a novel, highly sensitive and selective workflow for *N*- and *O*-glycomics from formalin-fixed and paraffin-embedded (FFPE) clinical tissue specimens.

Here we explored the limits of *N*- and *O*-glycan profiling from minimal amounts of FFPE tissue using a porous graphitized carbon LC-MS approach. In the course of the investigation we wanted to answer the questions; 1) whether clinical FFPE tissue sections can be used for glycan biomarker discovery studies; 2) whether common staining protocols are consistent with the profiling approach; and 3) whether micro-dissected tissue with very low cell numbers can be used and add any value to the profiling approach.

- 2.) Production of well-characterized glycopeptide standards of known structure and quantities for validation and development of novel mass spectrometric glycoproteomics techniques. Information embedded in glycopeptides was aimed to be unraveled by establishing new variants of mass spectrometric methods.
- 3.) Use of synthetic glycopeptides to develop novel mass spectrometry-based glycoproteomic techniques.
  - a) Simultaneous glycan and peptide sequencing by optimizing collision-induced dissociation parameters in one **single** tandem MS experiment using a Quadrupole-Time of Flight instrument. It was intended to increase product-ion spectra quality with this approach to facilitate consecutive software-assisted data analysis and pave the way for complex biomolecule analysis using the established settings.
  - b) Investigation of the potential of ion mobility-mass spectrometry in distinguishing glycopeptide position isomers and *N*-Acetylneuraminic acid linkage isomers of glycopeptides.

## 2 Collection of Selected Scientific Publications

### 2.1 Unlocking Cancer Glycomes from Histopathological Formalin-fixed and Paraffin-embedded (FFPE) Tissue Microdissections

#### 2.1.1 Abstract

“*N*- and *O*-glycans are attractive clinical biomarkers as glycosylation changes in response to diseases. The limited availability of defined clinical specimens impedes glyco-biomarker identification and validation in large patient cohorts. Formalin-fixed paraffin-embedded (FFPE) clinical specimens are the common form of sample preservation in clinical pathology, but qualitative and quantitative *N*- and *O*-glycomics of such samples has not been feasible to date. Here, we report a highly sensitive and glycan isomer selective method for simultaneous *N*- and *O*-glycomics from histopathological slides. As few as 2,000 cells isolated from FFPE tissue sections by laser capture microdissection were sufficient for in-depth histopathology-glycomics using porous graphitized carbon nanoLC ESI-MS/MS. *N*- and *O*-glycan profiles were similar between unstained and hematoxylin and eosin stained FFPE samples but differed slightly compared to fresh tissue. This method provides the key to unlock glyco-biomarker information from FFPE histopathological tissues archived in pathology laboratories worldwide.”

#### 2.1.2 Personal Project Contribution

In this project I developed, optimized and validated a highly sensitive clinical glycomics approach that now allows extraction and in-depth structural analysis of *N*- and *O*-glycans from very low amounts of cells isolated from FFPE-preserved clinical tissue specimens. Initial work was done on mouse tissue (not included in the paper). In addition, I was responsible for data organization and lead manuscript and figure preparation. Statistical analysis was supported by Falko Schirmeister. The liver specimens used in this project were prepared by my

collaboration partners (mainly Dr. Petra Korać) including FFPE tissue, staining (H&E) and laser capture microdissection.

## **2.2 The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics.**

### **2.2.1 Abstract**

“In-depth site-specific investigations of protein glycosylation are the basis for understanding the biological function of glycoproteins. Mass spectrometry-based *N*- and *O*-glycopeptide analyses enable determination of the glycosylation site, site occupancy, as well as glycan varieties present on a particular site. However, the depth of information is highly dependent on the applied analytical tools, including glycopeptide fragmentation regimes and automated data analysis. Here, we used a small set of synthetic disialylated, biantennary *N*-glycopeptides to systematically tune Q-TOF instrument parameters towards optimal energy stepping collision induced dissociation (CID) of glycopeptides. A linear dependency of *m/z*-ratio and optimal fragmentation energy was found, showing that with increasing *m/z*-ratio, more energy is required for glycopeptide fragmentation. Based on these optimized fragmentation parameters, a method combining lower- and higher-energy CID was developed, allowing the online acquisition of glycan and peptide-specific fragments within a single tandem MS experiment. We validated this method analyzing a set of human immunoglobulins (IgA1+2, sIgA, IgG1+2, IgE, IgD, IgM) as well as bovine fetuin. These optimized fragmentation parameters also enabled software-assisted glycopeptide assignment of both *N*- and *O*-glycopeptides including information about the most abundant glycan compositions, peptide sequence and putative structures. Twenty-six out of 30 *N*-glycopeptides and four out of five *O*-glycopeptides carrying >110 different glycoforms could be identified by this optimized LC-ESI tandem MS method with minimal user input. The Q-TOF based

glycopeptide analysis platform presented here opens the way to a range of different applications in glycoproteomics research as well as biopharmaceutical development and quality control.”<sup>96</sup>

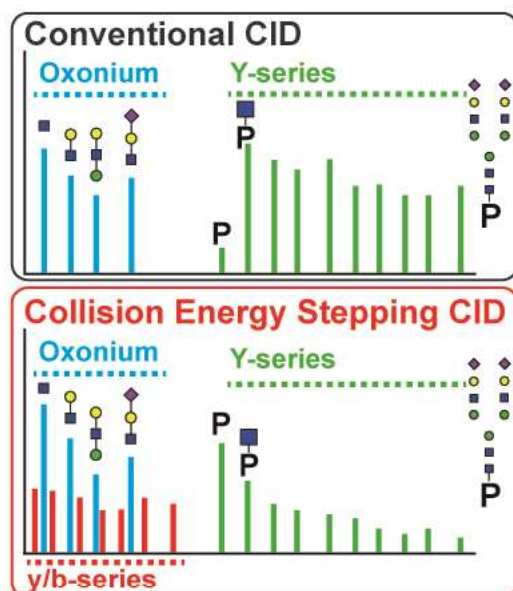


Figure 17: Graphical abstract for “The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics”<sup>96</sup>.

### 2.2.2 Personal Project Contribution

For this scientific work I first extracted large amount of building blocks from natural sources using established protocols<sup>151, 153, 154</sup>. Briefly, a hexapeptide carrying a disialylated, biantennary *N*-glycan was enriched from chicken egg yolk and subsequently proteolytically digested to a single asparagine carrying the glycan moiety. Subsequently, protection of the amino acid's N-terminus was achieved using a Fmoc group and the sialic acid residues were selectively protected by esterification through BnBr. In the course of this work the hexapeptide extraction procedures were also optimized to achieve higher yields. However, these steps were not included in the final manuscript. The fully protected building block was used to synthesize defined glycopeptides by SPPS. Intermediates as well as the final products were checked by various analytical tools (NMR, MS, HPLC...).



I also performed the laboratory work necessary for the preparation of Ig samples and partially (sodium dodecyl sulfate polyacrylamide gel electrophoresis [SDS-PAGE] onwards) for the human colon samples. The optimization of Q-TOF fragmentation conditions using the glycopeptides and samples I prepared was done in direct collaboration with partners at Bruker in Bremen.

Data analysis of the Ig sample set using the ProteinScape 4.0 (Bruker) data analysis software was mainly done by the author with some support by my colleagues Kathrin Stavenhagen and Dr. Kolarich.

The author participated in the discussion and conception of the study, wrote great parts of the manuscript and designed and edited multiple figures.

### **2.3 Distinguishing *N*-Acetylneuraminic Acid Linkage Isomers on Glycopeptides by Ion Mobility-Mass Spectrometry**

#### **2.3.1 Abstract**

“Differentiating the structure of isobaric glycopeptides represents a major challenge for mass spectrometry-based characterisation techniques. Here we show that the regiochemistry of the most common *N*-acetylneuraminic acid linkages of *N*-glycans can be identified in a site-specific manner from individual glycopeptides using ion mobility-mass spectrometry analysis of diagnostic fragment ions.”<sup>155</sup>

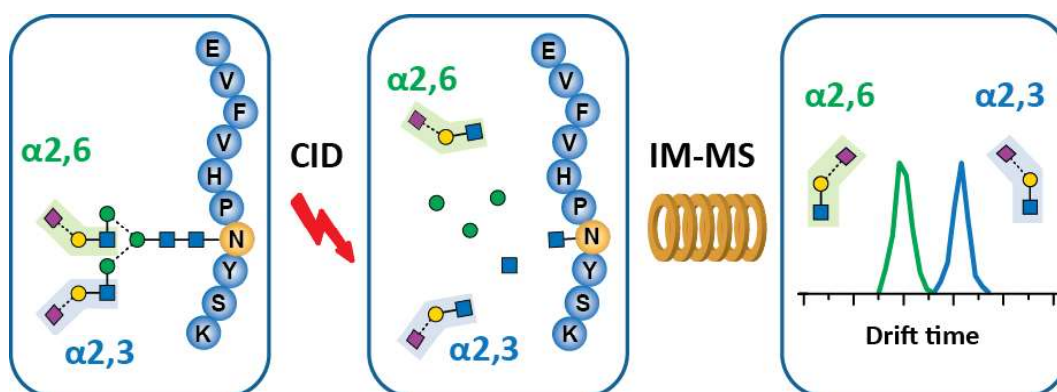


Figure 18: Graphical abstract for “Distinguishing *N*-Acetylneuraminic Acid Linkage Isomers on Glycopeptides by Ion Mobility-Mass Spectrometry”<sup>155</sup>.

### 2.3.2 Personal Project Contribution

For this study I used the glycosylated Asn-building blocks to produce a synthetic *N*-glycopeptide library similar to the one described in chapter 2.2.2. The glycopeptide library was, however, further extended by additional glycopeptides. Furthermore, the glycan moiety was also enzymatically modified by my collaborators in Prof. Altmann’s group to produce glycopeptides exhibiting different linkage isomers as well as different glycan epitopes. In the final manuscript not all produced glycopeptides were included as the focus was on differentiation of sialic acid linkage isomers.

I also performed the purification of human  $\alpha$ 1-protease inhibitor using SDS-PAGE, subsequent in-gel tryptic digestion and glycopeptide enrichment via offline HILIC (including a procedure optimization). The synthetic and enriched glycopeptides were analyzed by my cooperation partner Johanna Hofmann from Prof. Pagel’s group with my assistance.

Finally, the manuscript was developed in a co-authorship manner, with extensive contributions in writing and figure design from both parties.

### 3 Discussion

#### 3.1 Glycomics from Formalin-fixed and Paraffin-embedded Tissue

The ability to use FFPE-preserved tissues for glyco-biomarker studies opens numerous novel opportunities to investigate disease associated glycosylation patterns directly from tissues. FFPE tissue preservation is a routine procedure in clinics around the world and also provides the opportunity for long-term storage of sufficient numbers of patient samples. These specimens are usually also associated with comprehensive patient, treatment and outcome information. As long as analytes of interest are preserved during FFPE tissue preparation, specimens can be used for retro-perspective studies. Moreover, small amounts of tissue can be analyzed after further partitioning, minimizing the risk of contaminations with surrounding tissue and cells, which can hamper the detection of specific markers<sup>85, 86, 90</sup>.

Protein glycosylation plays a crucial role for biological functions in healthy cells and has also been reported to change in disease. The protein glycosylation patterns are complex and influenced by many factors that are not directly encoded via the genetic code. Therefore, disease specific alterations in glycosylation patterns can be more prominent than changes in the proteome as single changes can affect numerous proteins of a cell. Thus their molecular characterization is vital for a better understanding of biological and pathological processes within organisms<sup>5, 8</sup>. Released from their protein carriers they can be analyzed by different methodologies such as MS, LC or combinations of these to reveal structural information. Such glycan structure information is gathered by glycomics studies<sup>97, 102, 139</sup>.

In a study by Isailovic<sup>118</sup> and co-workers they compared the serum *N*-glycome of 28 liver cirrhosis and 25 liver cancer (HCC) patients to the one obtained from healthy individuals after permethylation of the released glycans. In their ESI-IM-MS approach they just identified 17 glycans that were present across all samples,

which is a relatively low number owing to the high-throughput set up they used. However, using nine glycan structures they were able to distinguish patients suffering from HCC from healthy individuals and the liver cirrhosis group. Most of the glycan structures were of the sialylated complex-type (6) and carried three (3) or two antenna (5). Interestingly, also a mannosylated structure (Man5) was among the nine distinguishing glycans.

Although serum *N*-glycan profiling can in principle deliver disease markers, analyses of disease tissue can provide a more direct view on the molecular mechanisms occurring in affected tissues. Turiak and colleagues<sup>156</sup> used murine FFPE liver tissue for *N*-glycan characterization and reported 14 compositions. For these analyses they used six sequential tissue slides (10  $\mu\text{m}$  thick) with a total area of around 2  $\text{mm}^2$  each ( $\sim 0.12 \text{ mm}^3$  tissue volume). Despite the labor-intensive workflow spatial data, at least for *N*-glycans, appeared obtainable.

Spatial glycan data can also be acquired using a MALDI-imaging technique. Powers *et al.*<sup>157</sup> investigated 16 cases of liver cancer by MALDI-imaging and identified 26 *N*-glycan compositions. Although MALDI-imaging is a powerful tool to deliver *N*-glycan profiles with spatial distribution information starting from minor amounts of specimens including FFPE tissue, this method has certain limitations. Sialic acids of native glycans are generally labile during MALDI ionization due to in-source fragmentation or post-source decay processes<sup>102</sup>. This, however, can be compensated by including additional derivatization steps for stabilizing these labile moieties. Moreover, Fuc residues have been reported to migrate while tandem MS applications (mainly seen in positive-ion mode). Also isomer differentiation is not readily achieved by this technique since they cannot be resolved by traditional tandem MS techniques. The analysis of low-mass analytes ( $m/z < 500 \text{ Da}$ ) is also challenging due to abundant matrix signals<sup>102, 120, 126</sup> and the presence of residual polymer from paraffin blocks can interfere with glycan measurements<sup>157</sup>. Another drawback is that the MALDI-

imaging approach can currently not be extended to the analysis of *O*-glycans due to the lack of suitable techniques to release these from the proteins without destroying the tissue and thus spatial resolution.<sup>158</sup>

Obtaining glycomics data from minute amounts of biological starting material remains a challenging task. This becomes even more prominent when *O*-glycans are the object of interest. In one of the first reports describing glycan analyses from FFPE tissues *N*- and *O*-glycans could be obtained from FFPE specimens, however milligram amounts of material were required. In addition, the yield of *O*-glycans was poor and the overall structural information was very limited<sup>93</sup>.

Satooma *et al.*<sup>159</sup> reported the release of protein-linked glycans via a nonreductive  $\beta$ -elimination approach (harboring the danger of possible peeling reactions<sup>160</sup>) from at least 1 mm<sup>3</sup> of tissue. Using an MALDI-TOF-MS approach and exoglycosidase sequencing they were able to detect differences between tumor and healthy tissue. However, they did not present any *O*-glycan data.

More recently Furukawa and colleagues<sup>94</sup> reported an *O*-glycomics MALDI-TOF-MS approach using microwave-assisted  $\beta$ -elimination in the presence of pyrazolone analogues prior analysis. The method is able to deliver profiles from FFPE material equal to four sections (10  $\mu$ m thick, unspecified surface area) or 10<sup>6</sup> cells in case other sample sources were used. Miura *et al.*<sup>161</sup> used a similar approach but had considerable amounts of *N*-glycan contaminants.

Although the extraction of protein bound oligosaccharides from FFPE tissue has in principle been demonstrated all these methods still suffer from certain drawbacks when biomarker screening starting from low amounts of clinical material is envisaged. Behind that background we developed a method for simultaneous profiling of *N*- and *O*-glycans with highest possible sensitivity from FFPE tissues, allowing now to analyze a low number of cells (1,000 - 5,000) that can also be obtained by LCM to increase the information on the spatial tissue

distribution of the glycans. The protocol is robust, enables reliable extraction of even labile structures (e.g. sialic acids and/or Fuc residues) and in-depth glycomics including differentiation and quantitation of glycan isomers.

In the entire study 86 *N*-glycan and 12 *O*-glycan structures were detected. The approach developed during my PhD works allowed us to obtain glycan profiles from almost 2 orders of magnitude less FFPE tissue (0,045 mm<sup>3</sup>) and even principle profiles could be determined from low as 1,000 cells isolated by LCM. Sections prepared according to standard pathological procedures, including staining (H&E), were evaluated to be suitable for glycome analysis. In a proof of principle study clear differences in *N*- and *O*-glycan patterns could be detected between tumor (HCC) and surrounding non-tumor tissue.

As this study was a proof of principle study a statistically relevant number of individual patients was not included that would have allowed making conclusions on HCC associated glycosylation. Interestingly, Isailovic<sup>118</sup> identified the high mannose structure Hex<sub>5</sub>HexNAc<sub>2</sub> (derived from serum proteins) with the most impact on the separation of samples of patients and healthy individuals. This very same structure was also found to be significantly increased in tumor tissue during our investigation, possibly due to the fact that tissue glycans enter the blood stream eventually. However, even with some correlations seen in both studies, it is possible that glycan changes picked up by serum analysis as conducted by Isailovic are not reflecting the same changes occurring directly in the tissue of interest.

Our results are also consistent with the findings reported from the MALDI-imaging approach mentioned earlier by Powers and colleagues<sup>157</sup> even for the small number of samples investigated. From the 26 *N*-glycan compositions detected in their study, 15 overlapped with the 44 compositions found in our study. This probably reflects inter-personal variations. The approach developed during my PhD, however, can in future studies be applied to investigate a larger

sample cohort to evaluate inter-person variability and validate HCC specific glycan signatures.

The developed method was extensively validated. With this capacity in hand it is now possible to obtain in-depth *N*- and *O*-glycan profiles from clinical FFPE tissue as low as a few thousand cells. Furthermore, it could be shown that routine staining procedures (H&E) did not affect analysis. In combination with LCM, the obtained glycomic profiles can also be correlated with histopathologically visible tissue features. This approach provides a valuable tool to the medical and glycobiology community allowing glycan profiling of enriched (potentially pure) cell population, increasing the chance to identify glycan biomarker signatures. As the starting material is FFPE-preserved tissue, this method is of course not limited to hepatic tissue but can be applied to analyze any biopsy material. The logical next steps include the collection of well-defined sample cohorts and the identification of disease associated glycosylation signatures that in future might provide the next generation of disease markers for personalized medicine.

### 3.2 Synthetic Glycopeptide Standards are Unique Tools for Glycoproteomic Method Development and Validation

The major challenge in glycoproteomic studies is the simultaneous acquisition of data that allows determination of the peptide sequence and provides information on the attached glycan structure. This becomes even trickier if knowledge on the monosaccharide connectivity is to be acquired in a site-specific manner. Glycopeptide analysis is further hindered by the higher background of non-glycosylated peptides in the proteolytic (glyco-)peptide mixture; decreased ionization efficiencies when compared to non-glycosylated counterparts<sup>162</sup>; the degree of occupancy of individual glycosylation sites as well as glycan variety present at a single site. In particular the latter increases the number of glycopeptide molecules exhibiting the same peptide sequence, which in itself already results in reduced signal intensities<sup>163</sup>.

A typical glycoproteomic workflow usually includes the following crucial steps: enrichment on protein and/or peptide level, proteolysis (for peptide analysis), multidimensional peptide separation, tandem mass-spectrometric analysis and software-assisted data interpretation<sup>97</sup>. Structural information on the glycan moiety can only be obtained to a certain degree by this typical approach and is lagging far behind from what is possible when released glycans are analysed (e.g. by PGC-LC ESI/MSMS)<sup>97, 139, 163</sup>. Beside the technical limitations and issues related to the physico-chemical characteristics of glycopeptides, software tools for mass spectrometric glycopeptide data analysis are also by far not as advanced, in particular when compared to what is already available and possible for pure proteomic data<sup>164, 165</sup>. Essential aspects in a glycoproteomic workflow important for improving in-depth glycopeptide analysis are focusing on the generation of more informative data within a mass spectrometer. This facilitates the work of software algorithms in the automated analyses of glycopeptide production spectra.



For this purpose Alley *et al.*<sup>143</sup> used tryptic (glyco-)peptides derived from various glycoproteins, including ribonuclease B, fetuin, horseradish peroxidase, and haptoglobin. They focused mainly on the optimization of ETD conditions for peptide backbone fragmentation, which provides the opportunity to preserve a PTM on the site of modification. This enables the unambiguous identification of the sites of attachment. Upon ETD fragmentation the glycan moieties remain intact, whereas in CID preferential cleavage of the glycosidic bonds through low energy pathways is observed. Although this CID / ETD combination approach in principle demonstrated the potential for characterization of model glycoproteins in one single tandem MS experiment, the authors also noted some limitations. ETD fragmentation efficiency is known to be rather low (~20% for the glycopeptides investigated in this particular study) and favors multiple protonated precursors ( $\geq 3+$ )<sup>166</sup>. Moreover, ETD appears to be limited to a certain  $m/z$  range (just suitable for precursors below  $m/z$  1400 in the chosen example). Thus the authors did not observe any peptide backbone fragmentation for the fetuin-derived glycopeptide RPTGEVYDIEIDTLETTCHVLDPTPLANCSVR carrying a trisialyated, triantennary complex glycan ( $m/z$  1634).

Likewise, Darula *et al.*<sup>167</sup> characterized tryptic *O*-glycopeptides from bovine serum after lectin enrichment (Jacalin) using CID and ETD. They also reported insufficient ETD fragmentation for precursors above  $m/z \sim 850$ . Interestingly, they tried to incorporate computational tools for data analysis, an aspect usually neglected in studies to improve glycopeptide identification by optimizing mass spectrometric processes. Data interpretation was facilitated but still required considerable degree of manual input. Prior knowledge on carbohydrate size and composition was necessary before database searches were initiated and could not be automatically extracted from the data. The approach was also just applicable for simple *O*-linked carbohydrate structures.

Using complementary fragmentation techniques is one way to obtain glycopeptide structure information. CID, when tuned properly, also bears the potential to generate more informative structural data when different fragmentation energy levels are employed. Jebanathirajah *et al.*<sup>144</sup> used tryptic (glyco-)peptides derived from urokinase plasminogen activator receptor to tune collision energies towards the generation of “reporter” fragments (sugar oxonium ions such as Hex<sup>+</sup> at  $m/z$  163.060, HexNAc<sup>+</sup> at  $m/z$  204.084 and HexHexNAc<sup>+</sup> at  $m/z$  366.139) to screen for glycopeptides present in mixtures with mainly non-glycosylated peptides.

Irungu *et al.*<sup>168</sup> fragmented the virus-envelope-glycoprotein-derived glycopeptide SNITGLLLTR carrying a GlcNAc<sub>2</sub>Man<sub>9</sub> *N*-glycan by low-energy CID (ESI-MS/MS) and high-energy CID (MALDI-TOF/TOF MS) using different platforms. They demonstrated nicely the benefits of combining information of product-ion spectra generated by different collision energy levels. The high-energy processes yielded a peptide sequence, with minimal glycan moiety information, whereas for the low-energy processes mainly carbohydrate fragments and peptide + HexNAc fragments were observed. It should be pointed out, even when the results are in agreement with other findings, the different energy level used for CID are not the only parameters that influence glycopeptide fragmentation. Also charge state, charge carrier and overall composition have a major impact on which fragments are produced<sup>95</sup>. Although the study provided researchers with valuable glycoprotein information it was also apparent that data interpretation was labor-intensive. MS data analysis was mainly done manually and just supported by the web-based tool GlycoPep ID once a “MS/MS peak list of the glycopeptides in question” was established. It was also necessary to manually validate and verify all potential glycopeptide compositions suggested by the software.

Certain instrument configurations allow a higher-energy collisional dissociation (HCD), resulting in some desired peptide b- and y-ions as well as diagnostic oxonium ions, but on extend of glycopeptide Y-type ion series (glycan fragments still attached to the peptide). This diminishes site specific glycan characterization<sup>169</sup>.

This can partially be compensated by using HCD in combination with ETD<sup>170</sup> (e.g. HCD product ion-triggered ETD), however the issues associated with ETD fragmentation, as discussed above, remain. In order to make this approach realistically applicable the authors pointed out that development of suitable software tools is also urgently required.

Recently, Zeng *et al.*<sup>171</sup> reported also an integrated approach for characterization of glycopeptide microheterogeneity by combining HCD- and CID-MS/MS and MS<sup>3</sup> for the analysis of HILIC-enriched glycopeptides derived from standard glycoproteins. Here, HCD and CID-MS/MS spectra provide information about the glycan moiety and the Y<sub>1</sub> ions (peptide backbone with HexNAc attached). The Y<sub>1</sub>-ions could then also be selected in a data-dependent manner based on intensity for subsequent MS<sup>3</sup> analysis to allow peptide backbone identification. However, for the HCD-MS<sup>3</sup> set up a second LC run was necessary due to the fact that longer duty cycles were required for data acquisition, which was incompatible with a combined HCD- and CID-MS/MS set up.

In conclusion, in the past years researchers have invested a lot of effort to determine glycopeptide structures in great detail. Although in glycoproteomics great progress has been made in analytical (optimized parameters) and computational aspects, data analysis still remains a major bottleneck and lags behind the possibilities available for other –omics branches such as proteomics. According to recently published reports<sup>164, 165</sup> reviewing software solutions for glycoproteomics approaches there is not a single tool available that combines all the functions required for automated data analysis under one umbrella.

Important software features required for large scale glycoproteomics studies include the structural elucidation of *N*- and *O*-glycopeptides, automated use of MS/MS data, incorporated search engines with scoring (ranking) algorithms to evaluate putative structures, abilities for batch inputs, the ability for parallel detection of non-glycosylated peptides and their matching to known protein sequences by database searches.

### 3.2.1 Simultaneous Glycan and Peptide Sequencing by Optimizing Collision-induced Dissociation Parameters

Systematic tuning of Q-TOF collision energy parameters was achieved using well-defined synthetic glycopeptides and enabled acquisition of a maximum of information on site-specific structure information from *N*- and *O*- glycopeptides within a single tandem MS experiment<sup>96</sup>. The optimal collision energy parameters showed a linear dependency to precursor *m/z* values (charge state independent). The optimized procedure provided very good quality tandem MS spectra which facilitated their use in subsequent software-assisted data analysis with minimal user input. This step in particular is often regarded as a bottleneck in glycoproteomics studies<sup>164, 165</sup>. Using a sample set derived from different Igs (IgA, sIgA, IgG, IgE, IgM, IgD) in the optimized LC-MS set up this approach was validated to be suitable for large scale proteomics analyses.

In the course of the study it became clear that the unambiguous detection of the oligosaccharide attachment site is hindered if a tryptic glycopeptide contained more than one potential site of glycosylation, which is frequently found for *O*-glycopeptides. The obvious next steps include the combination of the collision energy stepping CID approach with ETD. If an ETD experiment could only be triggered once certain reporter ions are observed, it could facilitate acquiring site specific peptide- and glycan sequence information while saving analysis time.

To date the amount of information that can be obtained on glycan structures by glycoproteomics studies cannot be compared to what is being achieved by

dedicated glycomics approaches<sup>105</sup>. The synthetic *N*-glycopeptide libraries produced in the course of this thesis could also be used in future to investigate whether the collision energy stepping approach can be applied in negative-ion mode to produce more informative carbohydrate fragments<sup>120, 127</sup>.

In this context, Nishikaze *et al.*<sup>172</sup> analyzed singly-deprotonated *N*-linked glycopeptides by low-energy CID and demonstrated that valuable information regarding the oligosaccharide moiety could be directly obtained, including definition of certain structural glycan features. They also observed peptide backbone cleavage indicated by b- and y- but also c- and z- type ions. However, they noted that the glycan fragmentation was largely depending on the amino acid sequence (especially the length) and/or composition. Nonetheless, the study illustrates the potential of negative-ion glycopeptide fragmentation for deeper structure characterization within a single experiment. This could in future possibly be exploited by combinatorial approaches including collision energy stepping CID.

Despite the achievements made in the MS field to obtain in-depth information on glycopeptide structure characterization of the glycan, moiety beyond composition remains challenging. This data is often accessed through tandem MS analyses and/or exoglycosidase sequencing<sup>120</sup> of released glycans. Until recently the determination of the 3-D stereochemical assembly of glycoconjugates, including monosaccharide linkages was not possible using mass spectrometry as the only tool. That task becomes even more challenging when the glycans are attached to the protein/peptide backbone.

### 3.2.2 Ion Mobility-Mass Spectrometry for Glycopeptide Characterization

With the recent introduction of an ion mobility cell into a MS instrument platform a novel analytical technology has become available providing novel opportunities for the gas phase structure analysis of biomolecules. Researchers have used IM-MS to analyze isomeric/isobaric glycoconjugates such as glycans,

glycolipids, GAGs, glycoproteins and glycopeptides. A comprehensive overview on the current state of the art is provided by a recent review from Gary *et al.*<sup>119</sup>. The analysis of carbohydrates with this technology has been a major focus in many studies using this promising new approach. Hofmann *et al.*<sup>116</sup> demonstrated that the identification of carbohydrate anomers (stereoisomers) and linkage-isomers (regioisomers) attached to a linker is possible using IM-MS, whereas the deprotonated ions and chloride adducts showed the largest CCS differences. In the context of glycopeptide characterization only few publications are available. Li *et al.*<sup>173</sup> used IM-MS to analyze tryptic glycopeptides derived from human  $\alpha$ -1-acid glycoprotein and antithrombin III, showing that some isomeric compounds could be separated by their drift times, but the vast majority of molecules did not show any reasonable separation by IM-MS. Creese *et al.*<sup>174</sup> separated and identified (through reporter ions generated by ETD) two isomeric *O*-linked glycopeptides only differing in the site of attachment for the GalNAc residue. In similar fashion, Both *et al.*<sup>175</sup> measured isomeric glycopeptides carrying only a single monosaccharide and showed that to some extent separation and identification was possible with these compositional isomers. Based on their drift times peptides carrying GlcNAc or GalNAc separated better compared to anomeric linkage/compositional isomeric (peptides carrying  $\alpha$ -GalNAc vs.  $\beta$ -GlcNAc). It should be noted that drift time differences were only moderate and glycopeptide assignment based on arrival time distribution was not possible. Pure anomeric linkage isomers (e.g. peptides with  $\alpha$ - vs.  $\beta$ -GalNAc) could also not be separated. In the same work a mixture of isomeric glycopeptides carrying either one  $\alpha$ -GlcNAc or one  $\alpha$ -GalNAc at several potential glycosylation sites could only be partially separated by ion mobility, but their monosaccharide CID product-ions were distinguishable.

To further investigate the promising potential of IM-MS to be used for glycopeptide structure elucidation we used a set of well-defined synthetic glycopeptides<sup>155</sup>. The library consisted of glycopeptides (YGNVNETQNNSFK) with

an  $\alpha$ 2-6 disialylated, biantennary *N*-glycan occupying either one of the glycosylation sites (bold) either near the N- or C-terminus (position isomers). It also included glycopeptides carrying the glycan structure mentioned above at the very same site within the amino acid sequence (EVFVHPNYSK), but with the terminal NeuAc residues linked either in  $\alpha$ 2-3- or  $\alpha$ 2-6 position (linkage isomers). The gained knowledge was then applied to identify NeuAc linkage isomers within a classical glycoproteomics experiment on tryptically digested, offline HILIC-enriched glycopeptides of  $\alpha$ 1-proteinase inhibitor.

Similar to the results published by Creese *et al.*<sup>174</sup> the two regioisomers could be separated based on their drift times, however only when present as quadruply deprotonated ions, indicating a different shape in the gas phase was just adopted in this particular charge state. The investigated linkage isomers could not be separated under the conditions tested. These results are in accordance with earlier ones reported by Both *et al.*<sup>175</sup> who showed that the  $\alpha$ -GlcNAc- and  $\beta$ -GlcNAc glycopeptides (anomeric linkage isomers) were undistinguishable by their IM-MS setup. It seems that the variation of only a single monosaccharide differently attached to its adjacent sugar does not influence the drift behavior of an entire glycopeptide molecule significantly enough to adopt a different gas phase structure that would allow separation by IM-MS. A possible effect appears to be “diluted” by the contribution of the non-glycan portion of the molecule. However, in our work we could show that when the molecule was fragmented into smaller building blocks by CID prior the IM cell, distinctive drift times for a NeuAc-Gal-GlcNAc oxonium ion ( $m/z$  657) were detected supporting the results of Both *et al.*<sup>175</sup> and Hofmann *et al.*<sup>116</sup>. This data shows that oligosaccharide-only fragments with distinct linkages can be separated using IM-MS when fragments consist of a small number ( $\sim$  up to five) of monosaccharides.

The most valuable outcome of this study was that different isobaric glycopeptides generated after tryptic proteolysis could be clearly distinguished

based on their oligosaccharide product-ions carrying different sialic acids residues ( $\alpha$ 2-6 vs  $\alpha$ 2-3 NeuAc). As sialic acids are key molecules of intercellular interactions the ability to differentiate these sialic acid linkages in a site specific manner directly on glycopeptides will be of great importance to study the role of site specific glycosylation in biological processes<sup>176, 177</sup>. In future this approach will be applied within an online LC set up to investigate more complex samples and to overcome some challenges faced in monosaccharide linkage determination of glycopeptides. Shortly after our study was published Guttman *et al.*<sup>178</sup> also published a report that confirmed our results on entirely different glycoproteins, emphasizing that the developed method is broadly applicable for in-depth glycoproteomic investigations.

Recently Chen *et al.*<sup>179</sup> also published an approach aiming to identify NeuAc linkages of glycopeptides using a pseudo-MS<sup>3</sup> ion trap set up. However, the provided information and data is limited and does not allow conclusions on numerous experimental conditions. Enzyme and glycoprotein sources as well as certain methodical procedures (e.g. proteolytic digests, PGC-LC set up, glycan purification) have not been reported. It is also in most cases not clear which glycopeptides were actually analyzed since no sequence information,  $m/z$  values or MS spectra are shown, questioning the results and conclusions reported.



#### 4 References

1. Cummings, J.H. & Stephen, A.M. Carbohydrate terminology and classification. *Eur J Clin Nutr* **61**, S5-S18 (2005).
2. Tharanathan, R.N. & Kittur, F.S. Chitin--the undisputed biomolecule of great potential. *Crit. Rev. Food Sci. Nutr.* **43**, 61-87 (2003).
3. Ohtsubo, K. & Marth, J.D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **126**, 855-867 (2006).
4. Dube, D.H. & Bertozzi, C.R. Glycans in cancer and inflammation [mdash] potential for therapeutics and diagnostics. *Nat Rev Drug Discov* **4**, 477-488 (2005).
5. Springer, S.A. & Gagneux, P. Glycomics: revealing the dynamic ecology and evolution of sugar molecules. *Journal of Proteomics* **135**, 90-100 (2016).
6. Merry, A.H. & Merry, C.L. Glycoscience finally comes of age. *EMBO reports* **6**, 900-903 (2005).
7. Ghazarian, H., Idoni, B. & Oppenheimer, S.B. A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics. *Acta histochemica* **113**, 236-247 (2011).
8. Adamczyk, B., Tharmalingam, T. & Rudd, P.M. Glycans as cancer biomarkers. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1820**, 1347-1353 (2012).
9. Ruhaak, L.R., Miyamoto, S. & Lebrilla, C.B. Developments in the identification of glycan biomarkers for the detection of cancer. *Mol Cell Proteomics* **12**, 846-855 (2013).
10. in *Essentials of Glycobiology*. (eds. A. Varki et al.) (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY); 2009).
11. Marino, K., Bones, J., Kattla, J.J. & Rudd, P.M. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat. Chem. Biol.* **6**, 713-723 (2010).
12. Herget, S. et al. Statistical analysis of the Bacterial Carbohydrate Structure Data Base (BCSDB): characteristics and diversity of bacterial carbohydrates in comparison with mammalian glycans. *BMC Struct. Biol.* **8**, 35 (2008).
13. Muthana, S.M., Campbell, C.T. & Gildersleeve, J.C. Modifications of glycans: biological significance and therapeutic opportunities. *ACS chemical biology* **7**, 31-43 (2012).
14. Varki, A. et al. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25**, 1323-1324 (2015).
15. Stepper, J. et al. Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Lett.* **585**, 645-650 (2011).

16. Moremen, K.W., Tiemeyer, M. & Nairn, A.V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* **13**, 448-462 (2012).
17. Bennett, E.P. et al. Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **22**, 736-756 (2012).
18. Bond, M.R. & Hanover, J.A. A little sugar goes a long way: the cell biology of O-GlcNAc. *J. Cell Biol.* **208**, 869-880 (2015).
19. Barry, F.P. et al. N- and O-linked keratan sulfate on the hyaluronan binding region of aggrecan from mature and immature bovine cartilage. *J Biol Chem* **270**, 20516-20524 (1995).
20. Couchman, J.R. & Pataki, C.A. An introduction to proteoglycans and their localization. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **60**, 885-897 (2012).
21. Schaefer, L. & Schaefer, R.M. Proteoglycans: from structural compounds to signaling molecules. *Cell Tissue Res.* **339**, 237-246 (2010).
22. Aigal, S., Claudinon, J. & Römer, W. Plasma membrane reorganization: A glycolipid gateway for microbes. *Biochim. Biophys. Acta* **1853**, 858-871 (2015).
23. Durrant, L.G., Noble, P. & Spendlove, I. Immunology in the clinic review series; focus on cancer: glycolipids as targets for tumour immunotherapy. *Clin. Exp. Immunol.* **167**, 206-215 (2012).
24. Tsai, Y.H., Liu, X. & Seeberger, P.H. Chemical biology of glycosylphosphatidylinositol anchors. *Angewandte Chemie (International ed. in English)* **51**, 11438-11456 (2012).
25. McConville, M.J. & Ferguson, M.A. The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes. *Biochem. J.* **294 ( Pt 2)**, 305-324 (1993).
26. Halim, A. et al. Site-specific characterization of threonine, serine, and tyrosine glycosylations of amyloid precursor protein/amyloid beta-peptides in human cerebrospinal fluid. *Proc Natl Acad Sci U S A* **108**, 11848-11853 (2011).
27. Zielinska, D.F., Gnad, F., Wisniewski, J.R. & Mann, M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897-907 (2010).
28. Sun, S. & Zhang, H. Identification and Validation of Atypical N-Glycosylation Sites. *Analytical Chemistry* (2015).
29. Krokhin, O.V., Antonovici, M., Ens, W., Wilkins, J.A. & Standing, K.G. Deamidation of -Asn-Gly- sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. *Analytical Chemistry* **78**, 6645-6650 (2006).
30. Thaysen-Andersen, M., Packer, N.H. & Schulz, B.L. Maturing glycoproteomics technologies provide unique structural insights into the

- N-glycoproteome and its regulation in health and disease. *Mol Cell Proteomics* (2016).
31. Thaysen-Andersen, M. et al. Human Neutrophils Secrete Bioactive Paucimannosidic Proteins from Azurophilic Granules into Pathogen-Infected Sputum. *Journal of Biological Chemistry* **290**, 8789-8802 (2015).
  32. Sethi, M.K. & Fanayan, S. Mass Spectrometry-Based N-Glycomics of Colorectal Cancer. *International journal of molecular sciences* **16**, 29278-29304 (2015).
  33. Vasconcelos-Dos-Santos, A. et al. Biosynthetic Machinery Involved in Aberrant Glycosylation: Promising Targets for Developing of Drugs Against Cancer. *Frontiers in oncology* **5**, 138 (2015).
  34. Schwarz, F. & Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Curr. Opin. Struct. Biol.* **21**, 576-582 (2011).
  35. Aebi, M., Bernasconi, R., Clerc, S. & Molinari, M. N-glycan structures: recognition and processing in the ER. *Trends Biochem. Sci.* **35**, 74-82 (2010).
  36. Tannous, A., Pisoni, G.B., Hebert, D.N. & Molinari, M. N-linked sugar-regulated protein folding and quality control in the ER. *Semin. Cell Dev. Biol.* **41**, 79-89 (2015).
  37. Aebi, M. N-linked protein glycosylation in the ER. *Biochim Biophys Acta* **1833**, 2430-2437 (2013).
  38. Colley, K.J. Golgi localization of glycosyltransferases: more questions than answers. *Glycobiology* **7**, 1-13 (1997).
  39. Petrescu, A.J., Milac, A.L., Petrescu, S.M., Dwek, R.A. & Wormald, M.R. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **14**, 103-114 (2004).
  40. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K.P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350-355 (2011).
  41. Hassinen, A. et al. Functional organization of Golgi N- and O-glycosylation pathways involves pH-dependent complex formation that is impaired in cancer cells. *J Biol Chem* **286**, 38329-38340 (2011).
  42. Hassinen, A. & Kellokumpu, S. Organizational interplay of Golgi N-glycosyltransferases involves organelle microenvironment-dependent transitions between enzyme homo- and heteromers. *J Biol Chem* **289**, 26937-26948 (2014).
  43. Brockhausen I, S.H., Stanley P. O-GalNAc Glycans. In: Essentials of Glycobiology. 2nd edition. ,Varki A, Cummings RD, Esko JD, et al. . *Cold Spring Harbor Laboratory Press* (2009).
  44. Lodish H, B.A., Zipursky SL, et al. Section 17.7 Protein Glycosylation in the ER and Golgi Complex in Molecular Cell Biology. 4th edition. *W. H. Freeman* (2000).

45. Gerken, T.A. et al. Emerging paradigms for the initiation of mucin-type protein O-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. *J Biol Chem* **286**, 14493-14507 (2011).
46. Stavenhagen, K., Kolarich, D. & Wührer, M. Clinical Glycomics Employing Graphitized Carbon Liquid Chromatography-Mass Spectrometry. *Chromatographia* **78**, 307-320 (2015).
47. Van den Steen, P., Rudd, P.M., Dwek, R.A. & Opdenakker, G. Concepts and principles of O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol.* **33**, 151-208 (1998).
48. Ju, T., Aryal, R.P., Kudelka, M.R., Wang, Y. & Cummings, R.D. The Cosmc connection to the Tn antigen in cancer. *Cancer biomarkers : section A of Disease markers* **14**, 63-81 (2014).
49. Lis, H. & Sharon, N. Protein glycosylation. Structural and functional aspects. *Eur. J. Biochem.* **218**, 1-27 (1993).
50. Deeb, S.J., Cox, J., Schmidt-Supprian, M. & Mann, M. N-linked glycosylation enrichment for in-depth cell surface proteomics of diffuse large B-cell lymphoma subtypes. *Mol Cell Proteomics* **13**, 240-251 (2014).
51. Wollscheid, B. et al. Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat Biotech* **27**, 378-386 (2009).
52. SolÁ, R.J. & Griebenow, K. Effects of Glycosylation on the Stability of Protein Pharmaceuticals. *J. Pharm. Sci.* **98**, 1223-1245 (2009).
53. Stockert, R.J. The asialoglycoprotein receptor: relationships between structure, function, and expression. *Physiol. Rev.* **75**, 591-609 (1995).
54. Varki, A. Sialic acids in human health and disease. *Trends Mol. Med.* **14**, 351-360 (2008).
55. Hooper, L.V. & Macpherson, A.J. Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nat Rev Immunol* **10**, 159-169 (2010).
56. Johansson, M.E., Larsson, J.M. & Hansson, G.C. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4659-4665 (2011).
57. Varki A, L.J. Biological Roles of Glycans In: Essentials of Glycobiology. 2nd edition. ,Varki A, Cummings RD, Esko JD, et al. *Cold Spring Harbor Laboratory Press* (2009).
58. Kleene, R. & Schachner, M. Glycans and neural cell interactions. *Nat. Rev. Neurosci.* **5**, 195-208 (2004).
59. Sauerzapfe, B. et al. Chemo-enzymatic synthesis of poly-N-acetyllactosamine (poly-LacNAc) structures and their characterization for CGL2-galectin-mediated binding of ECM glycoproteins to biomaterial surfaces. *Glycoconjugate journal* **26**, 141-159 (2009).

60. Grigorian, A., Torossian, S. & Demetriou, M. T-cell growth, cell surface organization, and the galectin-glycoprotein lattice. *Immunol Rev* **230**, 232-246 (2009).
61. Kolarich, D., Lepenies, B. & Seeberger, P.H. Glycomics, glycoproteomics and the immune system. *Current opinion in chemical biology* **16**, 214-220 (2012).
62. van Kooyk, Y. & Rabinovich, G.A. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat Immunol* **9**, 593-601 (2008).
63. Johnson, J.L., Jones, M.B., Ryan, S.O. & Cobb, B.A. The regulatory power of glycans and their binding partners in immunity. *Trends Immunol.* **34**, 290-298 (2013).
64. Wassarman, P.M. Sperm receptors and fertilization in mammals. *The Mount Sinai journal of medicine, New York* **69**, 148-155 (2002).
65. Mascanfroni, I.D. et al. Endogenous lectins shape the function of dendritic cells and tailor adaptive immunity: mechanisms and biomedical applications. *Int. Immunopharmacol.* **11**, 833-841 (2011).
66. McEver, R.P. Selectins: lectins that initiate cell adhesion under flow. *Curr. Opin. Cell Biol.* **14**, 581-586 (2002).
67. Sperandio, M., Gleissner, C.A. & Ley, K. Glycosylation in immune cell trafficking. *Immunol Rev* **230**, 97-113 (2009).
68. Shade, K.T. et al. A single glycan on IgE is indispensable for initiation of anaphylaxis. *J. Exp. Med.* **212**, 457-467 (2015).
69. Matrosovich, M., Herrler, G. & Klenk, H.D. Sialic Acid Receptors of Viruses. *Topics in current chemistry* **367**, 1-28 (2015).
70. Varki, A. & Gagneux, P. Multifarious roles of sialic acids in immunity. *Ann. N. Y. Acad. Sci.* **1253**, 16-36 (2012).
71. Freeze HH, S.H. Genetic Disorders of Glycosylation. In: Essentials of Glycobiology. 2nd edition. ,Varki A, Cummings RD, Esko JD, et al. . *Cold Spring Harbor Laboratory Press* (2009).
72. Gornik, O. & Lauc, G. Glycosylation of serum proteins in inflammatory diseases. *Dis. Markers* **25**, 267-278 (2008).
73. Theodoratou, E. et al. The role of glycosylation in IBD. *Nat Rev Gastroenterol Hepatol* advance online publication (2014).
74. Rabinovich, G.A. & Croci, D.O. Regulatory circuits mediated by lectin-glycan interactions in autoimmunity and cancer. *Immunity* **36**, 322-335 (2012).
75. Stowell, S.R., Ju, T. & Cummings, R.D. Protein glycosylation in cancer. *Annu Rev Pathol* **10**, 473-510 (2015).
76. Christiansen, M.N. et al. Cell surface protein glycosylation in cancer. *Proteomics* **14**, 525-546 (2014).

77. Croci, Diego O. et al. Glycosylation-Dependent Lectin-Receptor Interactions Preserve Angiogenesis in Anti-VEGF Refractory Tumors. *Cell* **156**, 744-758 (2014).
78. Tsuboi, S. Tumor defense systems using O-glycans. *Biol. Pharm. Bull.* **35**, 1633-1636 (2012).
79. Torre, L.A. et al. Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians* **65**, 87-108 (2015).
80. Pinho, S.S. & Reis, C.A. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* **15**, 540-555 (2015).
81. Kim, Y.J. & Varki, A. Perspectives on the significance of altered glycosylation of glycoproteins in cancer. *Glycoconjugate journal* **14**, 569-576 (1997).
82. Ma, Z. & Vosseller, K. Cancer Metabolism and Elevated O-GlcNAc in Oncogenic Signaling. *Journal of Biological Chemistry* **289**, 34457-34465 (2014).
83. Svarovsky, S.A. & Joshi, L. Cancer glycan biomarkers and their detection - past, present and future. *Analytical Methods* **6**, 3918-3936 (2014).
84. Etxebarria, J. & Reichardt, N.-C. Methods for the absolute quantification of N-glycan biomarkers. *Biochimica et Biophysica Acta (BBA) - General Subjects*.
85. Magdeldin, S. & Yamamoto, T. Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues. *Proteomics* **12**, 1045-1058 (2012).
86. Ostasiewicz, P., Zielinska, D.F., Mann, M. & Wisniewski, J.R. Proteome, phosphoproteome, and N-glycoproteome are quantitatively preserved in formalin-fixed paraffin-embedded tissue and analyzable by high-resolution mass spectrometry. *J Proteome Res* **9**, 3688-3700 (2010).
87. Fischer, A.H., Jacobson, K.A., Rose, J. & Zeller, R. Hematoxylin and eosin staining of tissue and cell sections. *CSH protocols* **2008**, pdb.prot4986 (2008).
88. Berg, D., Malinowsky, K., Reischauer, B., Wolff, C. & Becker, K.F. Use of formalin-fixed and paraffin-embedded tissues for diagnosis and therapy in routine clinical settings. *Methods Mol Biol* **785**, 109-122 (2011).
89. Gräntzdörffer, I. et al. Comparison of different tissue sampling methods for protein extraction from formalin-fixed and paraffin-embedded tissue specimens. *Experimental and Molecular Pathology* **88**, 190-196 (2010).
90. De Marchi, T. et al. The advantage of laser-capture microdissection over whole tissue analysis in proteomic profiling studies. *Proteomics* **16**, 1474-1485 (2016).
91. Serth, J., Kuczyk, M.A., Paeslack, U., Lichtinghagen, R. & Jonas, U. Quantitation of DNA extracted after micropreparation of cells from frozen and formalin-fixed tissue sections. *The American journal of pathology* **156**, 1189-1196 (2000).

92. Roberts, L. et al. Identification of methods for use of formalin-fixed, paraffin-embedded tissue samples in RNA expression profiling. *Genomics* **94**, 341-348 (2009).
93. Dwek, M.V., Brooks, S.A., Streets, A.J., Harvey, D.J. & Leatham, A.J. Oligosaccharide release from frozen and paraffin-wax-embedded archival tissues. *Analytical biochemistry* **242**, 8-14 (1996).
94. Furukawa, J.-i. et al. Quantitative O-Glycomics by Microwave-Assisted  $\beta$ -Elimination in the Presence of Pyrazolone Analogues. *Analytical Chemistry* (2015).
95. Dodds, E.D. Gas-phase dissociation of glycosylated peptide ions. *Mass spectrometry reviews* **31**, 666-682 (2012).
96. Hinneburg, H. et al. The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics. *J Am Soc Mass Spectrom* **27**, 507-519 (2016).
97. Alley, W.R., Jr., Mann, B.F. & Novotny, M.V. High-sensitivity analytical approaches for the structural characterization of glycoproteins. *Chemical reviews* **113**, 2668-2732 (2013).
98. Ruhaak, L.R. et al. Glycan labeling strategies and their use in identification and quantification. *Anal Bioanal Chem* **397**, 3457-3481 (2010).
99. Manzi, A.E. et al. Exploring the glycan repertoire of genetically modified mice by isolation and profiling of the major glycan classes and nano-NMR analysis of glycan mixtures. *Glycobiology* **10**, 669-689 (2000).
100. Mulloy B, H.G., Stanley P. Structural Analysis of Glycans. In: Essentials of Glycobiology. 2nd edition. ,Varki A, Cummings RD, Esko JD, et al. *Cold Spring Harbor Laboratory Press* (2009).
101. Duus, J., Gottfredsen, C.H. & Bock, K. Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. *Chemical reviews* **100**, 4589-4614 (2000).
102. Wuhrer, M. Glycomics using mass spectrometry. *Glycoconjugate journal* **30**, 11-22 (2013).
103. Kottler, R. et al. Development of a high-throughput glycoanalysis method for the characterization of oligosaccharides in human milk utilizing multiplexed capillary gel electrophoresis with laser-induced fluorescence detection. *Electrophoresis* **34**, 2323-2336 (2013).
104. Morelle, W. & Michalski, J.-C. Analysis of protein glycosylation by mass spectrometry. *Nat. Protocols* **2**, 1585-1602 (2007).
105. Jensen, P.H., Karlsson, N.G., Kolarich, D. & Packer, N.H. Structural analysis of N- and O-glycans released from glycoproteins. *Nat Protoc* **7**, 1299-1310 (2012).
106. Huffman, J.E. et al. Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol Cell Proteomics* **13**, 1598-1610 (2014).

107. Vanderschaeghe, D., Festjens, N., Delanghe, J. & Callewaert, N. Glycome profiling using modern glycomics technology: technical aspects and applications. *Biol. Chem.* **391**, 149-161 (2010).
108. Glish, G.L. & Vachet, R.W. The basics of mass spectrometry in the twenty-first century. *Nat Rev Drug Discov* **2**, 140-150 (2003).
109. Kolli, V., Schumacher, K.N. & Dodds, E.D. Engaging challenges in glycoproteomics: recent advances in MS-based glycopeptide analysis. *Bioanalysis* **7**, 113-131 (2015).
110. Biemann, K. Mass spectrometry of peptides and proteins. *Annu. Rev. Biochem.* **61**, 977-1010 (1992).
111. Johnson, R.S., Martin, S.A., Biemann, K., Stults, J.T. & Watson, J.T. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Analytical Chemistry* **59**, 2621-2625 (1987).
112. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical mass spectrometry* **11**, 601 (1984).
113. Domon, B. & Costello, C.E. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate journal* **5**, 397-409.
114. Lanucara, F., Holman, S.W., Gray, C.J. & Eyers, C.E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat Chem* **6**, 281-294 (2014).
115. Harvey, D.J. et al. Travelling wave ion mobility and negative ion fragmentation for the structural determination of N-linked glycans. *Electrophoresis* **34**, 2368-2378 (2013).
116. Hofmann, J., Hahm, H.S., Seeberger, P.H. & Pagel, K. Identification of carbohydrate anomers using ion mobility-mass spectrometry. *Nature* **526**, 241-244 (2015).
117. Pagel, K. & Harvey, D.J. Ion Mobility–Mass Spectrometry of Complex Carbohydrates: Collision Cross Sections of Sodiated N-linked Glycans. *Analytical Chemistry* **85**, 5138-5145 (2013).
118. Isailovic, D. et al. Delineating Diseases by IMS-MS Profiling of Serum N-linked Glycans. *Journal of Proteome Research* **11**, 576-585 (2012).
119. Gray, C.J. et al. Applications of ion mobility mass spectrometry for high throughput, high resolution glycan analysis. *Biochim Biophys Acta* (2016).
120. Harvey, D.J. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass spectrometry reviews* **18**, 349-450 (1999).
121. Zauner, G., Deelder, A.M. & Wuhrer, M. Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics. *Electrophoresis* **32**, 3456-3466 (2011).



122. Ruhaak, L.R., Deelder, A.M. & Wührer, M. Oligosaccharide analysis by graphitized carbon liquid chromatography-mass spectrometry. *Anal Bioanal Chem* **394**, 163-174 (2009).
123. West, C., Elfakir, C. & Lafosse, M. Porous graphitic carbon: a versatile stationary phase for liquid chromatography. *Journal of chromatography. A* **1217**, 3201-3216 (2010).
124. Pabst, M. & Altmann, F. Influence of Electrosorption, Solvent, Temperature, and Ion Polarity on the Performance of LC-ESI-MS Using Graphitic Carbon for Acidic Oligosaccharides. *Analytical Chemistry* **80**, 7534-7542 (2008).
125. Melmer, M., Stangler, T., Premstaller, A. & Lindner, W. Comparison of hydrophilic-interaction, reversed-phase and porous graphitic carbon chromatography for glycan analysis. *Journal of chromatography. A* **1218**, 118-123 (2011).
126. Palmisano, G., Larsen, M.R., Packer, N.H. & Thaysen-Andersen, M. Structural analysis of glycoprotein sialylation -part II: LC-MS based detection. *Rsc Advances* **3**, 22706-22726 (2013).
127. Harvey, D.J. & Abrahams, J.L. Fragmentation and ion mobility properties of negative ions from N-linked carbohydrates: Part 7. Reduced glycans. *Rapid Commun Mass Spectrom* **30**, 627-634 (2016).
128. Harvey, D.J. Fragmentation of negative ions from carbohydrates: part 1. Use of nitrate and other anionic adducts for the production of negative ion electrospray spectra from N-linked carbohydrates. *J Am Soc Mass Spectrom* **16**, 622-630 (2005).
129. Harvey, D.J. Fragmentation of negative ions from carbohydrates: part 2. Fragmentation of high-mannose N-linked glycans. *J Am Soc Mass Spectrom* **16**, 631-646 (2005).
130. Harvey, D.J. Fragmentation of negative ions from carbohydrates: part 3. Fragmentation of hybrid and complex N-linked glycans. *J Am Soc Mass Spectrom* **16**, 647-659 (2005).
131. 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. *Analytical biochemistry* **376**, 44-60 (2008).
132. Harvey, D.J. et al. Fragmentation of negative ions from N-linked carbohydrates, part 4. Fragmentation of complex glycans lacking substitution on the 6-antenna. *J Mass Spectrom* **45**, 528-535 (2010).
133. Harvey, D.J. & Rudd, P.M. Fragmentation of negative ions from N-linked carbohydrates. Part 5: Anionic N-linked glycans. *International Journal of Mass Spectrometry* **305**, 120-130 (2011).
134. Harvey, D.J. et al. Fragmentation of negative ions from N-linked carbohydrates: part 6. Glycans containing one N-acetylglucosamine in the core. *Rapid Commun Mass Spectrom* **28**, 2008-2018 (2014).

135. Everest-Dass, A.V., Kolarich, D., Campbell, M.P. & Packer, N.H. Tandem mass spectra of glycan substructures enable the multistage mass spectrometric identification of determinants on oligosaccharides. *Rapid Commun Mass Spectrom* **27**, 931-939 (2013).
136. Thomsson, K.A. et al. Detailed O-glycomics of the Muc2 mucin from colon of wild-type, core 1- and core 3-transferase-deficient mice highlights differences compared with human MUC2. *Glycobiology* **22**, 1128-1139 (2012).
137. Schulz, B.L. et al. Glycosylation of sputum mucins is altered in cystic fibrosis patients. *Glycobiology* **17**, 698-712 (2007).
138. Schulz, B.L., Packer, N.H. & Karlsson, N.G. Small-Scale Analysis of O-Linked Oligosaccharides from Glycoproteins and Mucins Separated by Gel Electrophoresis. *Analytical Chemistry* **74**, 6088-6097 (2002).
139. Wuhrer, M., Catalina, M.I., Deelder, A.M. & Hokke, C.H. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **849**, 115-128 (2007).
140. Angata, T. et al. Integrated approach toward the discovery of glyco-biomarkers of inflammation-related diseases. *Ann. N. Y. Acad. Sci.* **1253**, 159-169 (2012).
141. Mehta, A., Herrera, H. & Block, T. Glycosylation and liver cancer. *Adv. Cancer Res.* **126**, 257-279 (2015).
142. Huddleston, M.J., Bean, M.F. & Carr, S.A. Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. *Analytical Chemistry* **65**, 877-884 (1993).
143. Alley, W.R., Jr., Mechref, Y. & Novotny, M.V. Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun Mass Spectrom* **23**, 161-170 (2009).
144. Jebanathirajah, J., Steen, H. & Roepstorff, P. Using optimized collision energies and high resolution, high accuracy fragment ion selection to improve glycopeptide detection by precursor ion scanning. *J Am Soc Mass Spectrom* **14**, 777-784 (2003).
145. Kolli, V. & Dodds, E.D. Energy-resolved collision-induced dissociation pathways of model N-linked glycopeptides: implications for capturing glycan connectivity and peptide sequence in a single experiment. *The Analyst* **139**, 2144-2153 (2014).
146. Unverzagt, C. & Kajihara, Y. Chemical assembly of N-glycoproteins: a refined toolbox to address a ubiquitous posttranslational modification. *Chemical Society reviews* **42**, 4408-4420 (2013).
147. Buskas, T., Ingale, S. & Boons, G.J. Glycopeptides as versatile tools for glycobiology. *Glycobiology* **16**, 113r-136r (2006).

148. Hojo, H. & Nakahara, Y. Recent progress in the field of glycopeptide synthesis. *Biopolymers* **88**, 308-324 (2007).
149. Mäde, V., Els-Heindl, S. & Beck-Sickinger, A.G. Automated solid-phase peptide synthesis to obtain therapeutic peptides. *Beilstein Journal of Organic Chemistry* **10**, 1197-1212 (2014).
150. Yamamoto, N. et al. Solid-Phase Synthesis of Sialylglycopeptides through Selective Esterification of the Sialic Acid Residues of an Asn-Linked Complex-Type Sialyloligosaccharide. *Angewandte Chemie International Edition* **42**, 2537-2540 (2003).
151. Kajihara, Y. et al. Prompt Chemoenzymatic Synthesis of Diverse Complex-Type Oligosaccharides and Its Application to the Solid-Phase Synthesis of a Glycopeptide with Asn-Linked Sialyl-undeca- and Asialo-nonasaccharides. *Chemistry – A European Journal* **10**, 971-985 (2004).
152. Kajihara, Y., Yamamoto, N., Okamoto, R., Hirano, K. & Murase, T. Chemical synthesis of homogeneous glycopeptides and glycoproteins. *Chemical record (New York, N.Y.)* **10**, 80-100 (2010).
153. Seko, A. et al. Occurrence of a sialylglycopeptide and free sialylglycans in hen's egg yolk. *Biochim Biophys Acta* **1335**, 23-32 (1997).
154. Piontek, C. et al. Semisynthesis of a Homogeneous Glycoprotein Enzyme: Ribonuclease C: Part 2. *Angewandte Chemie International Edition* **48**, 1941-1945 (2009).
155. Hinneburg, H. et al. Distinguishing N-acetylneuraminic acid linkage isomers on glycopeptides by ion mobility-mass spectrometry. *Chemical Communications* **52**, 4381-4384 (2016).
156. Turiák, L. et al. Workflow for Combined Proteomics and Glycomics Profiling from Histological Tissues. *Analytical Chemistry* **86**, 9670-9678 (2014).
157. Powers, T.W. et al. MALDI imaging mass spectrometry profiling of N-glycans in formalin-fixed paraffin embedded clinical tissue blocks and tissue microarrays. *PloS one* **9**, e106255 (2014).
158. Song, X. et al. Oxidative release of natural glycans for functional glycomics. *Nat Meth advance online publication* (2016).
159. Satomaa, T. et al. Analysis of the Human Cancer Glycome Identifies a Novel Group of Tumor-Associated N-Acetylglucosamine Glycan Antigens. *Cancer Research* **69**, 5811-5819 (2009).
160. Yu, G. et al. Effect and limitation of excess ammonium on the release of O-glycans in reducing forms from glycoproteins under mild alkaline conditions for glycomic and functional analysis. *Analytical Chemistry* **82**, 9534-9542 (2010).
161. Miura, Y. et al. Glycoblotting-Assisted O-Glycomics: Ammonium Carbamate Allows for Highly Efficient O-Glycan Release from Glycoproteins. *Analytical Chemistry* **82**, 10021-10029 (2010).

162. Stavenhagen, K. et al. Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *J Mass Spectrom* **48**, 627-639 (2013).
163. Stavenhagen, K., Plomp, R. & Wuhrer, M. Site-Specific Protein N- and O-Glycosylation Analysis by a C18-Porous Graphitized Carbon–Liquid Chromatography–Electrospray Ionization Mass Spectrometry Approach Using Pronase Treated Glycopeptides. *Analytical Chemistry* **87**, 11691-11699 (2015).
164. Dallas, D.C., Martin, W.F., Hua, S. & German, J.B. Automated glycopeptide analysis--review of current state and future directions. *Briefings in bioinformatics* **14**, 361-374 (2013).
165. Woodin, C.L., Maxon, M. & Desaire, H. Software for automated interpretation of mass spectrometry data from glycans and glycopeptides. *The Analyst* **138**, 2793-2803 (2013).
166. Kjeldsen, F., Giessing, A.M., Ingrell, C.R. & Jensen, O.N. Peptide sequencing and characterization of post-translational modifications by enhanced ion-charging and liquid chromatography electron-transfer dissociation tandem mass spectrometry. *Analytical Chemistry* **79**, 9243-9252 (2007).
167. Darula, Z., Chalkley, R.J., Baker, P., Burlingame, A.L. & Medzihradszky, K.F. Mass spectrometric analysis, automated identification and complete annotation of O-linked glycopeptides. *European journal of mass spectrometry (Chichester, England)* **16**, 421-428 (2010).
168. Irungu, J. et al. Comparison of HPLC/ESI-FTICR MS versus MALDI-TOF/TOF MS for glycopeptide analysis of a highly glycosylated HIV envelope glycoprotein. *J Am Soc Mass Spectrom* **19**, 1209-1220 (2008).
169. Scott, N.E. et al. Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of *Campylobacter jejuni*. *Mol Cell Proteomics* **10**, M000031-mcp000201 (2011).
170. Singh, C., Zampronio, C.G., Creese, A.J. & Cooper, H.J. Higher Energy Collision Dissociation (HCD) Product Ion-Triggered Electron Transfer Dissociation (ETD) Mass Spectrometry for the Analysis of N-Linked Glycoproteins. *Journal of Proteome Research* **11**, 4517-4525 (2012).
171. Zeng, W.F. et al. pGlyco: a pipeline for the identification of intact N-glycopeptides by using HCD- and CID-MS/MS and MS3. *Scientific reports* **6**, 25102 (2016).
172. Nishikaze, T., Kawabata, S.-i. & Tanaka, K. Fragmentation Characteristics of Deprotonated N-linked Glycopeptides: Influences of Amino Acid Composition and Sequence. *J. Am. Soc. Mass Spectrom.* **25**, 988-998 (2014).

173. Li, H., Bendiak, B., Siems, W.F., Gang, D.R. & Hill, H.H., Jr. Ion Mobility-Mass Correlation Trend Line Separation of Glycoprotein Digests without Deglycosylation. *International journal for ion mobility spectrometry : official publication of the International Society for Ion Mobility Spectrometry* **16**, 105-115 (2013).
174. Creese, A.J. & Cooper, H.J. Separation and Identification of Isomeric Glycopeptides by High Field Asymmetric Waveform Ion Mobility Spectrometry. *Analytical Chemistry* **84**, 2597-2601 (2012).
175. BothP et al. Discrimination of epimeric glycans and glycopeptides using IM-MS and its potential for carbohydrate sequencing. *Nat Chem* **6**, 65-74 (2014).
176. Anthony, R.M. et al. Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. *Science* **320**, 373-376 (2008).
177. Suzuki, Y. et al. Sialic acid species as a determinant of the host range of influenza A viruses. *J. Virol.* **74**, 11825-11831 (2000).
178. Guttman, M. & Lee, K.K. Site-Specific Mapping of Sialic Acid Linkage Isomers by Ion Mobility Spectrometry. *Analytical Chemistry* **88**, 5212-5217 (2016).
179. Chen, C.-H. et al. Rapid Identification of Terminal Sialic Acid Linkage Isomers by Pseudo-MS3 Mass Spectrometry. *Israel Journal of Chemistry* **55**, 412-422 (2015).

## 5 List of Publications

Stavenhagen, K.; **Hinneburg, H.**; Thaysen-Andersen M.; Hartmann L.; Varón Silva D.; Fuchser J.; Kaspar S.; Rapp E.; Seeberger PH.; Kolarich D. Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *Journal of Mass Spectrometry* **48**, 627-639 (2013).

**Hinneburg, H.**; Stavenhagen K.; Schweiger-Hufnagel U.; Pengelley S.; Jabs W.; Seeberger PH.; Silva D.; Wuhrer M.; Kolarich D. The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics. *Journal of the American Society for Mass Spectrometry* **27**, 507-519 (2016). \*

**Hinneburg, H.**; Hofmann J.; Struwe WB.; Thader A.; Altmann F.; Varón Silva D.; Seeberger PH.; Pagel K.; Kolarich D. Distinguishing *N*-acetylneuraminic acid linkage isomers on glycopeptides by ion mobility-mass spectrometry. *Chemical Communications* **52**, 4381-4384 (2016). \*

**Hinneburg, H.**; Korać P.; Schirmeister F.; Gasparov S.; Seeberger PH.; Zoldoš V.; Kolarich D. Unlocking cancer glycomes from histopathological formalin-fixed and paraffin-embedded (FFPE) tissue microdissections. In review at *Molecular and Cellular proteomics* (2016). +

**Hinneburg, H.**; Schirmeister F.; Korać P.; Kolarich D.; *N*- and *O*-Glycomics from Minor Amounts of Formalin-Fixed, Paraffin-Embedded Tissue Samples. *Methods in Molecular Biology - "High throughput glycomics and glycoproteomics"* (in press) (2016); Editors: Manfred Wuhrer and Gordan Lauc.

Stavenhagen K.; **Hinneburg, H.**; Kolarich D.; Wuhrer M.; Site-Specific *N*- and *O*-Glycopeptide Analysis Using an Integrated C18-PGCLC-ESI-QTOF-MS/MS Approach. *Methods in Molecular Biology- "High throughput glycomics and glycoproteomics"* (in press) (2016); Editors: Manfred Wuhrer and Gordan Lauc.

## 6 Supplementary Information

The three selected papers described in this work are attached as published (\*), submitted (+) version (see also Chapter 5) or can be looked up under

DOI: 10.1002/anie.201700000

## **Unlocking cancer glycomes from histopathological formalin-fixed and paraffin-embedded (FFPE) tissue microdissections**

Hannes Hinneburg<sup>1,2</sup>, Petra Korać<sup>3</sup>, Falko Schirmeister<sup>1,2</sup>, Slavko Gasparov<sup>4,5</sup>, Peter H. Seeberger<sup>1,2</sup>, Vlatka Zoldoš<sup>3</sup>, Daniel Kolarich<sup>1\*¶</sup>

1 Max Planck Institute of Colloids and Interfaces, Department of Biomolecular Systems, 14424 Potsdam, Germany

2 Freie Universität Berlin, Department of Biology, Chemistry, Pharmacy, Institute of Chemistry and Biochemistry, 14195 Berlin, Germany

3 Faculty of Science, Department of Biology, Division of Molecular Biology, University of Zagreb, Zagreb, Croatia

4 Institute for Pathology and Cytology, University Hospital Merkur, Zagreb, Croatia

5 Department of Pathology, Medical School Zagreb, University of Zagreb, Zagreb, Croatia

\*Corresponding Author: Dr. Daniel Kolarich

Max Planck Institute of Colloids and Interfaces, Department of Biomolecular Systems, 14424 Potsdam, Germany

Email: daniel.kolarich@mpikg.mpg.de

Telephone: +49-30-838 59306; Fax: +49-30-838 459306

¶Present address: Institute for Glycomics, Gold Coast Campus, Griffith University, Queensland, 4222, Australia, Email: d.kolarich@griffith.edu.au, Telephone: +61-7-5552 7026



Keywords: clinical glycomics, *N*-glycans, *O*-glycans, FFPE, hepatocellular carcinoma, mass spectrometry, microdissection, porous graphitized carbon, histopathology

### **Abbreviations**

AUC	Area under the Curve
BPC	Base Peak Chromatogram
EIC	Extracted Ion Chromatogram
FFPE	Formalin-fixed and paraffin-embedded
H&E	Hematoxylin and Eosin
HCC	Hepatocellular Carcinoma
LC	Liquid Chromatography
LCM	Laser Capture Microdissection
MS/MS	Tandem Mass Spectrometry
PGC	Porous Graphitized Carbon
SPS	Smart Parameter Setting
ICC	Ion Charge Control

## Abstract

*N*- and *O*-glycans are attractive clinical biomarkers as glycosylation changes in response to diseases. The limited availability of defined clinical specimens impedes glyco-biomarker identification and validation in large patient cohorts. Formalin-fixed paraffin-embedded (FFPE) clinical specimens are the common form of sample preservation in clinical pathology, but qualitative and quantitative *N*- and *O*-glycomics of such samples has not been feasible to date. Here, we report a highly sensitive and glycan isomer selective method for simultaneous *N*- and *O*-glycomics from histopathological slides. As few as 2,000 cells isolated from FFPE tissue sections by laser capture microdissection were sufficient for in-depth histopathology-glycomics using porous graphitized carbon nanoLC ESI-MS/MS. *N*- and *O*-glycan profiles were similar between unstained and hematoxylin and eosin stained FFPE samples but differed slightly compared to fresh tissue. This method provides the key to unlock glyco-biomarker information from FFPE histopathological tissues archived in pathology laboratories worldwide.

## Introduction

Cell surface and body fluid proteins are extensively decorated with specific glycans (1). Glycoproteins and glycolipids are part of the plasma membrane glycocalyx, the first interface for intercellular interactions (2). As such, the glycome and glycoproteome are key elements of cellular communication (3, 4). In many disease states biological processes change to express different glycans and glycoproteins on the cell surface (5, 6). Dynamic changes of the glycome are a result of age, an organism's physiological state (homeostasis or disease) and various other environmental or intrinsic factors (7-9). As glycans integrate environmental and indirectly genetic factors and are closely associated with complex diseases, they have been established as predictive and prognostic markers for diseases such as cancer and chronic inflammation (6, 10-13).

Most efforts to identify glyco-biomarkers focused on body fluids due to the ease of accessibility (6, 14, 15). To better understand the molecular events responsible for disease onset and progression, analysis of specific glycosylation signatures directly from tissue specimens is fundamental (16-18). However, tissues are often highly heterogeneous mixtures of different cell types that hamper the detection of specific markers. The collection of significant numbers of well-defined clinical tissue specimens for high quality biomarker research is challenging due to ethical and technical obstacles.

A widely applied standard pathology procedure is the preservation of tissue samples by formalin fixation followed by subsequent paraffin-embedding (FFPE), along with e.g. hematoxylin and eosin (H&E) staining (19) for tumor diagnosis and long-term conservation. Many FFPE preserved tissue specimens are available in clinical

centers around the world and are often associated with comprehensive clinical data sets. FFPE tissues can be handled at ambient temperatures and various dissection methods such as laser capture microdissection can be applied in order to minimize contaminations by other tissues and/or cells. This makes them a highly valuable source for any disease related research, biomarker discovery or retrospective studies, as long as the compounds of interest are qualitatively and quantitatively preserved and can likewise be extracted for further analyses (20-22).

Proteins (20, 21, 23-25) or nucleic acids (23, 26, 27) are now frequently recovered from FFPE specimens even though they can be modified and/or degraded during the conservation process. The analysis of metabolites (28), glycosaminoglycans (29), as well as *N*- (25, 30-33) and *O*-glycans (32-34) has been described. To date, all analyses aimed at comprehensive clinical tissue *N*- and *O*-glycomics i) suffered from poor sensitivity (32), ii) could not be applied to both *N*- and *O*-glycans (25, 30, 31, 34), iii) fall short of reliable *O*-glycan extraction (33), iv) suffered from the differentiation of isobaric *N*- and *O*-glycans(30, 32, 33) or reliable analysis of intact sialylated glycans (30, 31). Thus, a robust method for the sensitive and selective extraction and in-depth analysis of FFPE tissue derived *N*- and *O*-glycans is required.

Here we present a highly sensitive approach to individually and sequentially extract and analyze structurally preserved *N*- and *O*-glycans from as few as 1,000 cells isolated from FFPE tissue specimens. Laser capture microdissection (LCM) ensures contact free picking-up of cell groups or even a single cell for easy subsequent manipulation. LCM isolated tissues were analyzed along with whole tissue sections using porous graphitized carbon (PGC) nano scale liquid chromatography (nanoLC) coupled with online ESI tandem mass spectrometric (MS/MS) detection (35, 36). In-

depth glycan structure information on fucose and sialic acid linkages was obtained within a single experiment providing a maximum of information from a minimal amount of clinical material. This PGC nanoLC-ESI MS/MS based glycomics approach applied onto LCM-isolated cells derived from FFPE tissue specimens is opening novel avenues for investigating the role of protein glycosylation in health and disease.

## **Experimental Procedures**

If not otherwise stated all materials were purchased in the highest possible quality from Sigma-Aldrich (St. Louis, MO, USA). Peptide-*N*-Glycosidase F (PNGase F) was obtained from Roche Diagnostics GmbH (Mannheim, Germany). Water was used after purification with a Milli Q-8 direct system (Merck KGaA, Darmstadt, Germany). Chloroform was from VWR international (Fontenay-sous-Bois, France). Human liver tissue was derived from three different patients suffering from liver cirrhosis or hepatocellular carcinoma obtaining a liver transplant according to the Milan criteria. This study was approved by the Ethics committee of "*Klinioka Bolnica Merkur*" For reporting the glycomics experiments the MIRAGE guidelines are followed (37).

### **Formalin fixation and paraffin-embedding**

Liver tissue was divided into several parts and either immediately frozen at -80°C or fixed with formalin (Formalin solution, neutral buffered, 10%, Biognost) and embedded in paraffin (Paraffin wax, Sakura) following routine procedures. Briefly, tissue went through 10% formalin (24 h), 70% ethanol (2 x 1 h), 95% ethanol (2 x 1 h), absolute ethanol (2 x 1 h), xylene (2 x 1 h at 37°C) and liquid paraffin (3 x 1 h at

60°C). Paraffin tissue blocks were cut into 2-10 µm thick sections using a standard sliding manual microtome (Microm). 2-4 µm thick sections were mounted onto glass slides (Microscope slides, Vitrognost). H&E staining was done according to routine protocols (19). After deparaffinization with xylene, and rehydration through decreasing series of ethanol, slides were washed and hematoxylin was applied. After subsequent washes and eosin treatment slides were dehydrated with increasing ethanol series and briefly air-dried. The specimens were used within 3 months after initial preparation.

### **Laser capture microdissection**

Laser capture microdissection was done as follows: stained sections mounted on membrane slides (MembraneSlide 1.0 PEN, Carl Zeiss Microscopy) were analyzed under the microscope which is part of a laser microdissection system (Palm Zeiss MicroBeam, Axiovert 200M with PALM RoboSoftware, Carl Zeiss Microscopy) and selected cells were counted, marked and cut using PALM RoboSoftware according to manufacturer's manual.

Laser pressure catapulting method was performed using the protocol described in the manufacturer's manual and isolated tissue parts were collected in adhesive caps of 500 µL collection tubes (Sample AdhesiveCap 500 opaque (D) PCR Tube, Carl Zeiss Microscopy).

### **Protein extraction/ Antigen retrieval from FFPE tissues**

#### *Unmounted sample set*

Protein extraction of around 50 unmounted FFPE tissue sections was done as described previously (21). Briefly, 2-10  $\mu\text{m}$  thick FFPE sections were transferred into 1.5 mL sample tubes, washed twice in xylene for 5 min under mild agitation, followed by two washes with absolute ethanol for 5 min before being allowed to dry. Deparaffinized and washed samples were combined and homogenized on ice in a 0.1 M Tris-HCl, pH 8.0; 0.1 M DTT buffer using a table top Branson sonifier B-12 sonicator (Branson Sonic Power, Danbury, CT; output control: 1.5; 3x 10-30 s). SDS was added to a final concentration of 4% (w/v) and samples were incubated at 99°C for 60 min under mild agitation and were allowed to cool down to room temperature afterwards. Subsequently, the samples were centrifuged for 20 min at 2,000 rcf to remove non-soluble material. Proteins were precipitated from supernatant according to the method used by Wessel and Flügge (38). Briefly, 200  $\mu\text{L}$  supernatant were mixed with 800  $\mu\text{L}$  methanol, followed by the addition of 200  $\mu\text{L}$  chloroform and 600  $\mu\text{L}$  water and vigorous mixing. The samples were centrifuged at 14,000 rcf for 5 min and the upper aqueous phase was carefully removed (the interphase contained precipitated proteins). Additional 600  $\mu\text{L}$  methanol were added, liquids mixed and then centrifuged for 10 min at 14,000 rcf. The supernatant was removed and the final pellet was dissolved in a solution containing 6 M urea and 2 M thiourea prior dot blotting onto a PVDF membrane (described below) or stored at -20°C for further use.

#### Mounted sample set

Microscope slides carrying H&E stained and unstained FFPE tissue sections (2-4  $\mu\text{m}$ ) were washed with xylene (3x 2 min) followed by absolute ethanol (3x 2 min), scraped off with a needle/razor blade and then transferred from the slide into a sample tube containing 4% SDS in 0.1 M Tris-HCl, pH 8.0 and 0.1 M DTT. The

suspension was sonicated on ice (see above) and then further processed as described for the unmounted sections.

#### Laser capture microdissected sample sets

Tubes containing unstained, already deparaffinized washed cells (~1,000; ~5,000 and ~10,000), as well as H&E stained counter parts were incubated upside down to ensure capturing of microdissected samples usually located at the lid of the tube after the dissection procedure, for 30 min in 4% SDS in 0.1 M Tris-HCl, pH 8.0, 0.1 M DTT. Subsequently the samples were incubated in the same buffer for 60 min at 99°C. Samples were then further processed as described for the unmounted sections.

A second set of laser capture microdissected tissue containing approx. 2,000 cells per microdissection were obtained from three subsequent FFPE tissue sections of either hepatocellular carcinoma or surrounding non-tumor tissue. Microdissected samples were treated as described above.

#### Frozen sample sets

Protein extraction of frozen tissues was done as described earlier (39). Briefly, liver tissue blocks were homogenized on ice using a IKA "T10 basic" homogenizer (Staufen, Germany) in lysis buffer (pH 7.4) containing 50 mM Tris-HCl, 0.1 M sodium chloride, 1 mM ethylenediaminetetraacetic acid and protease inhibitor cocktail (Promega, Mannheim, Germany) and incubated for 20 min. The solution was then sonicated on ice using a tabletop Branson sonifier B-12 sonicator (Branson Sonic Power, Danbury, CT; output control: 1.5; 3x 10-30 s) and centrifuged at 2,000 rcf for



20 min at room temperature. Supernatants were spotted directly onto PVDF membranes (see below) or stored at -20°C.

### **PVDF dot blotting and *N*- and *O*-glycan release**

Extracted proteins were dot-blotted (frozen/FFPE tissue) onto a 0.2 µm pore size PVDF membrane (Millipore, Tullagreen, Ireland) and *N*- and *O*-glycans were sequentially released as described in detail previously (35, 36).

To assure that comparable protein amounts were used for the glycan release, dilutions series (1:2; 1:4; 1:8; 1:16; 1:20; 1:32) of proteins extracted from FFPE tissue sections (unmounted sections only) were prepared and spotted onto the membrane (data not shown). Spots exhibiting similar direct-blue 71 staining intensities compared to the fresh-frozen proteins were considered to contain comparable amounts of protein immobilized on the PVDF membranes and were used for subsequent glycan release.

### **Analysis of released *N*- and *O*-glycans using porous graphitized carbon liquid chromatography electrospray ionization mass spectrometry**

Following a final offline carbon desalting step as described previously (35) samples were dissolved in 10 µL MilliQ-water and 3 µL were injected for each LC-MS run. The PGC nanoLC setup was as follows: a PGC precolumn (HYPERCARB 5 µm, 30x0.32 mm, Thermo Fisher Scientific, Waltham, MA, USA) and a PGC separation column (HYPERCARB 3 µm, 100x0.075 mm, Thermo Fisher Scientific) were installed in Ultimate 3000 UHPLC system (Dionex, Germering, Germany). The precolumn was equilibrated in buffer A (10 mM ammonium bicarbonate) and released glycans were loaded onto the precolumn for 5 min at 6 µL/min flow rate

prior valve switching. The equilibration conditions for the separation column were 3% buffer B (60% ACN in 10 mM ammonium bicarbonate). After loading the released glycans onto the trapping column, a linear gradient was established using buffer B as follows: reduced *N*-glycans were separated using a linear gradient from 3% buffer B to 15.8% in one minute before increasing to 40.3% buffer B over 49 minutes. *O*-glycans were analyzed using the same setup but applying a linear gradient from 2% to 35% buffer B over 33 min. The flow rate was set to 0.8  $\mu\text{L}/\text{min}$  and the columns were held at a constant temperature of 40°C. The LC was directly coupled to an amaZon speed ETD ion trap mass spectrometer equipped with a CaptiveSpray source (Bruker, Bremen, Germany) for online detection of glycans in negative ion mode. MS spectra were acquired in UltraScan mode within an *m/z* range of 380 – 1,800; smart parameter setting (SPS) was set to *m/z* 900; ion charge control (ICC) to 40,000 and max. acquisition time to 200 ms.

MS/MS spectra were generated using collision induced dissociation over an *m/z* range from 100 – 2,500 on the three most abundant precursors applying an isolation width of *m/z* 3. The fragmentation cut-off was set to 27% with 100% fragmentation amplitude using the Enhanced SmartFrag option from 30-120% in 32 ms. ICC was set to 150,000.

### **Structure determination and relative glycan quantitation**

*N*- and *O*-glycan structures were identified by manual screening of spectra and annotated, including isomeric structure elucidation, based on PGC retention time (35, 40-42), negative-ion mode fragmentation behavior (43-55) and biosynthesis pathways known from literature (56, 57). Nomenclature is based on Varki *et al.* (58). Representative MS/MS spectra for each identified glycan structure (if applicable) can

be generated from the information given ( $m/z$ , intensity) in the Supplementary Material. Mass lists of fragment spectra were generated from raw data using DataAnalysis Version 4.1 (Bruker, Bremen, Germany) with an absolute intensity threshold of 50 (no rel. threshold used), a S/N ratio of 1 and a peak width [FWHM ( $m/z$ )] of 0.1. Besides manual annotation, glycan product ion spectra were assigned using ProteinScape 4.0 (Bruker) using the GlycomeDB database or Glycoworkbench 2.1 (59). For ProteinScape searches, the following parameters were selected: Glycan type: *N*-glycan; Taxonomy: *Homo sapiens*; Derivatization: UND (underivatized); Reducing end: reduced; Ions: H<sup>+</sup> up to 3 charges; Charge -: 1-3; MS tolerance ( $m/z$ ): 0.6 Da; MS/MS tolerance ( $m/z$ ): 0.6 Da; Mass type: monoisotopic; Fragmentation type CID: CID a b c x y z, crossrings  $\leq 1$ , cleavages  $\leq 2$ . Thresholds were set at Score:  $\geq 5\%$ , Fragmentation coverage:  $\geq 5\%$  and Intensity coverage:  $\geq 5\%$ . ProteinScape glycan assignments were accepted if the score was  $\geq 10$ . All assigned spectra and identified peaks are listed in the supplementary material.

Relative abundances of individual glycan structures were determined using the area under the curve (AUC) from the respective individual chromatographic traces (extracted ion chromatograms [EIC]) of the corresponding monoisotopic precursors using QuantAnalysis software (Version 2.1, Bruker, Bremen, Germany). The integration limits of all AUCs were checked and, if needed, adjusted manually. The sum of all detected and quantified glycan structures within a sample corresponded to 100% and the individual relative amounts were determined from the individual AUC values. Relative abundances were also calculated in relation to the highest peak when necessary. Note that a relative abundance of 1% was set as feasible threshold to take structures in consideration for quantitative changes in the course of this study.

## **Experimental Design and Statistical Rationale**

We developed, optimized and downscaled an analytical FFPE glycomics workflow for *N*- and *O*-glycans using hepatic specimens obtained from patients diagnosed with hepatocellular carcinoma (HCC) and selected for liver transplantation. One part of the tissue was immediately fractionated and frozen in liquid nitrogen (in the text referred to as *frozen tissue*) whereas the other part was subsequently FFPE-preserved (referred to as *FFPE tissue*). Glycans were released from proteins dot-blotted onto a PVDF membrane after extracted from tissue. Using clinical H&E stained and unstained tissue sections, the entire analytical workflow was developed and validated, minimal sample requirements evaluated and potential unintended artefacts induced by FFPE preservation, staining or antigen retrieval identified. The protocol was further optimized for minimal amounts of FFPE tissue (1,000; 5000; 10,000 cells), selected by LCMs. The optimized protocol was then used with approximately 2,000 hepatocytes from HCC tissue and surrounding non-tumor tissue (Figure 1).

Statistical analyses of the relative abundances of the *N*- and *O*-glycans from different origins (e.g. frozen/FFPE or tumor/non-tumor) are based on a linear model. Different sample sources were tested for equal suitability in glycomics analyses. Samples from different disease states were investigated for differences in relative abundances of glycans. The fold changes of the relative abundance of the compared samples and their respective *p*-values were plotted in the shape of Volcano plots (see Supplementary Material). Oligosaccharide structures with a mean relative abundance of less than 0.3% were not plotted if that threshold was not reached in either of the compared groups. This was done for clarity and potential issues arising from not accurate quantitation of very small peak areas. The *p*-values were corrected

for multiple testing using the method of Benjamini and Hochberg (60). Changes in relative glycan abundances with a  $p$ -value of less than 0.01 were regarded to be statistically significant. Unsupervised hierarchical cluster analysis was applied for data visualization in the form of heatmaps. The logarithmic relative abundance values were used to calculate the Euclidean distance matrix and complete-linkage clustering in the sample dimension. Additionally, for plotting the values of each glycan structure were scaled to have mean zero and standard deviation one (Z-score normalization), except in Figure 4B. For Figure 2A a second stage of clustering was applied in the glycan dimension for each of the five main structural groups separately. The free software for statistical computing and graphics “R” (61) was used as data analysis environment. The code is available upon request.

## Results

Frozen and hepatic FFPE clinical tissue was used for extraction on *N*- and *O*-glycans. The analytical FFPE glycomics workflow was developed for whole tissue sections, further optimized for minimal sample amounts, including laser-derived microdissections, evaluated and potential unintended artefacts induced by sample preparation identified. Chloroform-methanol precipitation (38) prior glycoprotein immobilization on PVDF membranes (35) was found to be beneficial for the subsequent *N*- and *O*-glycan release as well as for PGC nanoLC-ESI MS/MS glycomics. This step removes excess SDS and other contaminant substances resulting in clearly reduced background contamination levels, better data quality and a clear sensitivity increase (data not shown). Comprehensive *N*- and *O*-glycome profiles were generally obtained from tissue slides down to 3  $\mu\text{m}$  in thickness and a surface area of approximately 15  $\text{mm}^2$ .

### **Easy cleavable glycoepitopes remain intact during preservation and extraction**

Once a highly sensitive *N*- and *O*-glycomics protocol for FFPE tissue material was established we compared the glycomes obtained from the different sources. The *N*-glycan base peak chromatograms (BPCs) acquired from representative FFPE and frozen tissue specimens appeared equivalent (Supplementary Material 1a). Among the ten most abundant compositions nine were identical between the two sample sources (Figure 2B). Sialylated, complex biantennary (~66%) and fucosylated structures (~40%) were the most abundant *N*-glycans (Figure 2C, Supplementary

Material 1g). In total 74 different *N*-glycan structures (derived from 40 compositions) were detected in both, FFPE and frozen tissues (details in Supplementary Material 1b). The distribution of  $\alpha$ 2,6: $\alpha$ 2,3 linked NeuAc residues on complex *N*-glycans was roughly 2:1. Fucose residues were mainly attached to the *N*-glycan core (35%) with just ~5% of the hepatic *N*-glycome exhibiting Lewis type fucosylation (Figure 2C, Supplementary Material 1g). The same glyco-epitopes were detected in similar amounts in both, FFPE and frozen samples, indicating that the preservation process and the sample preparation required for antigen retrieval did not affect any more readily cleavable glycoepitopes containing sialic acid or fucose residues (Supplementary Material 1b).

The hepatic *O*-glycome consisted of just eight different *O*-glycans present in six compositions (Figure 2D, Supplementary Material 2a). These were exclusively of the core 1 and core 2 type with up to two NeuAc residues. In summary, these data confirmed that the FFPE preservation process and the antigen retrieval did not introduce any qualitative changes on the detectable *N*- and *O*-glycome.

### **Subtle quantitative glycome differences are present between frozen and FFPE specimens**

Quantitative changes in the glycome are a major parameter determined in any disease related glycomics study. Therefore, we tested whether the quantitative glycome was similar between frozen and FFPE specimens. The detected *N*-glycans were categorized into global structure categories (Figure 2C) and their relative quantitative distribution assessed by an unsupervised hierarchical cluster analysis

(Figure 2A). These data indicated that mostly the levels of *N*-glycans carrying  $\alpha$ 2,3 linked NeuAc and/or fucose residues were slightly elevated in the FFPE tissue compared to frozen tissue (Figure 2A/C). In particular, complex sialylated core fucosylated *N*-glycans (e.g. IDs 30c, 33d and 35d) showed higher abundances in FFPE tissue samples, whereas  $\alpha$ 2,6 NeuAc carrying *N*-glycans were slightly elevated in frozen tissue (Figure 2A/C; Supplementary Material 1g/h). The unsupervised cluster analysis also indicated differences between frozen and FFPE tissue for some paucimannosidic and complex neutral type *N*-glycans, however, further statistically relevant differences were confirmed only for structures 5b and 25 (Figure 2A/B, see also Supplementary Figure 1). These data emphasize that a direct comparison between frozen and FFPE derived starting material should be avoided as the tissue source already results in minor, though statistically relevant glycome differences, which might lead to false positive results in differential studies. Despite these no other global trends indicating destruction or discrimination of particular more readily cleavable monosaccharides such as fucose or sialic acid residues were observed. Furthermore, no unintended chemical modifications could be detected, further supporting the principal usability of FFPE conserved specimens for glyco-biomarker studies.

Some quantitative differences between the different sample sources were also found for the *O*-glycome. The singly and doubly sialylated core 2 type structures 1a ( $\text{Hex}_2\text{HexNAc}_2\text{NeuAc}$ ) and 4b ( $\text{Hex}_2\text{HexNAc}_2\text{NeuAc}_2$ ) were more prominent in FFPE samples while the levels of structures 7 ( $\text{HexHexNAcNeuAc}_2$ ) and 6b ( $\text{HexHexNAcNeuAc}$ ) were less abundant in FFPE samples. Other structures present above a relative abundance of 1% did not show any detectable quantitative changes (Figure 2D). Based on these results we concluded that for disease signature



evaluation a direct comparison between frozen and FFPE tissues should be avoided as different starting materials obviously mimic statistically relevant signatures in clinical glycomics studies.

### **Glass slide mounted and H&E stained tissue specimens provide suitable sources for clinical *N*- and *O*-glycomics**

The mounting of FFPE tissue sections onto glass slides and H&E tissue staining represent standard procedures in every pathology laboratory. We evaluated whether FFPE tissue sections that underwent these procedures are also suitable for clinical *N*- and *O*-glycomics. PGC nanoLC-ESI MS/MS glycomics performed on mounted FFPE preserved hepatic tissue sections (both, H&E stained and unstained) resulted in the detection of 77 *N*-glycan and 9 *O*-glycan structures, which was in good agreement with the results obtained for the unmounted material (Figure 3A-C, Supplementary Material 3a/4a). Statistics indicated a significant difference for paucimannose type structures between untreated and H&E stained slides (Figure 3B), however, as these four structures amount only ~2% of relative intensity, the minor individual tissue section differences could also cause the detected differences. The similar results obtained for essentially all *N*- and *O*-glycans from H&E stained and unstained tissue sections indicated that the staining method did not affect glycan integrity and antigen retrieval (Figure 3A-C, Supplementary Material 3a/d and Table 4a/d). These data also clearly demonstrated that glycomics information can be obtained from the very same material used for pathological examinations.

### **One thousand cells are sufficient for composition glycomics**

LCM provides a unique possibility to selectively isolate areas of interest for subsequent glycomics studies. However, as this approach inevitably yields just small amounts of tissue, we evaluated the minimum number of cells required to still obtain a basic glycan profile. Approximately 1,000, 5,000 and 10,000 hepatocytes were isolated from unstained and H&E stained FFPE tissue sections for glycan profiling (Figure 4A/B). In order to remain comparable with the analyses described above, also just 1/3 of the sample was actually used for a single PGC nanoLC-ESI MS/MS analysis despite the low amount of starting material.

We succeeded in obtaining basic *N*- and *O*-glycan profiles from as low as 1,000 cells (Figure 4A/B, Supplementary Material 3k and 4g). The detected signal intensities were low and just 25 out of 71 detected *N*-glycans met our stringent quantitation criteria. MS/MS based identification was successful for 20 out of 71 structures (38 and 50 out of 71 for 5,000 and 10,000 cells, respectively, Supplementary Material 3k), with the residual being assigned based on retention time, precursor mass and hepatic glycome data acquired during this project. Four out of eight *O*-glycans could also be quantified from the 1,000 cells starting material (Supplementary Material 4g). Even though glycan profiles were obtained from as low as 1,000 cells, this amount was considered to be insufficient for reliable quantitative clinical glycomics. In particular, lower abundant structures could not be reliably quantified when starting from 1,000 cells, making it very difficult to reliably detect subtle glycome changes. The use of higher cell numbers ( $\geq 2,000$ ) provided in general reliable quantitative glycome profiling data (Figure 4A/B). These results encouraged us to further pursue the LCM route and validate the usability of microdissection to complement histopathology with glycomics using hepatic cellular carcinoma tissue sections in a proof of principle study.

## **Spatial insights into hepatocellular carcinoma *N*- and *O*-glycome using laser-capture microdissection**

Associating the glycome with microscopically visible tissue features provides a key requirement to exactly determine disease associated glycan signatures. Tumor tissue micro-heterogeneity, however, represents a major obstacle in tissue preparation for glycan profiling as the presence of non-tumor related cells can mask tumor specific glycosylation features (22). LCM can help overcome this obstacle as it provides the possibility to specifically select tissue areas of interest that are also cross-validated by classical pathology diagnoses. In addition, this technique provides the opportunity to gain access to non-tumor tissue parts from the very same patient reducing the factor of individual glycosylation heterogeneity. We applied this technique to gain further insights into the individual glycosylation signatures of HCC.

*N*- and *O*-glycan profiles were determined from approximately 2,000 cells isolated by the LCM method (Figure 5A). The glycome of non-tumor tissue was comparable to the ones obtained for different liver samples described above (Supplementary Material 10). Significant global glycome changes were observed in the isolated HCC cells for various *N*-glycans of the paucimannose and core fucosylated type, but also for hybrid as well as neutral and  $\alpha$ 2,6 sialylated *N*-glycans (Figure 5B-D). Though high mannose type glycans appeared unaltered as a group, the levels of individual structures (Man8, Man7 and Man6, IDs: 8, 7a, 6b, respectively) were significantly decreased, while Man5 (ID: 5b) showed an opposite trend (Figure 5C, and Supplementary Material 5b). The most striking changes were detected for paucimannosidic structures such as Hex<sub>2</sub>HexNAc<sub>2</sub>Fuc (ID: 2) and Hex<sub>3</sub>HexNAc<sub>2</sub>Fuc (ID: 3) and the biantennary, doubly  $\alpha$ 2,6 sialylated structure (ID: 30a, Figure 5B/C).

The O-glycome obtained from the very same microdissected tumor tissue specimens showed more complex core 2 type structures carrying Lewis X and sialyl Lewis X glyco-epitopes that shifted towards shorter core 1 type O-glycans in the non-cancerous surrounding cells (Figure 5E). Similar core 2 O-glycans have previously been described in numerous tumor types and have been linked to increased tumor survival and metastasis by modulating the immune response (62, 63). As the hepatic O-glycome is significantly less complex compared to the N-glycome, in future O-glycans might present a useful and simpler alternative in the context of HCC tissue differentiation.

## Discussion

In principle glycans have been isolated from FFPE tissue sections already previously (31-33). Despite these earlier reports we were, however, not able to find satisfying answers to numerous crucial questions. I) Are all glyco-epitopes quantitatively and qualitatively preserved during the preservation and antigen-retrieval procedures? II) What is the minimum sample amount required to obtain a representative glycome? III) Is glycan integrity and recovery still given even after H&E staining?

To address these questions, we developed and applied a novel PGC nanoLC-ESI MS/MS glycomics approach for qualitative and relative quantitative N- and O-glycomics of clinical FFPE tissue sections and even LCM-enriched tissue specimens. Our results showed that unstained or H&E stained FFPE tissue sections represent well suited sources for clinical glycomics. In a proof of principle experiment we analyzed N- and O-glycans from hepatic tumor and surrounding non-tumor tissue

derived from the very same tissue section and patient. Using laser capture microdissection patient-derived individual glycome variations were reduced and the results provided a more precise view on the specific and individual cancer tissue glyco-signatures as tumor and control samples were also derived from spatially close areas. This approach allowed an in-depth characterization of the hepatic *N*- and *O*-glycome from single tissue specimens including differentiation and relative quantitation of structure isomers (Figure 6).

Our results clearly emphasize that similar tissue specimen treatment is imperative for glyco-biomarker studies. Minor though significant differences were detected when the glycome of frozen starting material was compared to the FFPE-derived profile. Inevitably the different starting conditions required different initial tissue homogenization and protein extraction procedures. Thus most likely specific proteins were lost and/or insoluble in either one of the procedures, subsequently resulting in the lack of specific glycoproteins. The FFPE sections were also derived from a spatially narrower tissue fraction compared to the frozen material, thus a differential spatial glycan distribution as reported previously by MALDI imaging experiments (31, 64) could also contribute to the observed quantitative changes. Nevertheless, in the initial method development several larger tissue sections were combined, making spatial distribution differences less a likely option for the observed differences.

Turiak *et al.* also reported differences in the number and relative abundances of *N*-glycans released from frozen and FFPE murine liver tissue (13 vs. 14 compositions) (25). In their interesting integrated omics approach they identify proteins and profile glycosaminoglycans and *N*-glycans from the same FFPE-slide. Nevertheless, in their workflow reliable detection and quantitation of released *N*-glycans required six

sequential sections (10  $\mu\text{m}$  thick,  $\sim 2 \text{ mm}^2$  in size) to be combined, whereas we consistently identified 63 *N*-glycans and 7 *O*-glycans from 5,000 cells.

In comparison to a recent *N*-glycan MALDI-imaging study on HCC including two patients overall just 17 of the reported 33 *N*-glycans qualitatively overlapped with our results (64). Whereas in our data paucimannosidic *N*-glycans in particular showed the largest differences (Figure 5B-D), Powers *et al.*(64) reported the largest differences for high mannose type and selected complex, non sialylated *N*-glycans. Due to the *m/z* cut-off used in their study, paucimannosidic *N*-glycans eluded their detection. Nevertheless, the observed differences are possibly also a reflection of individual patient variability. The unavoidable influence of this factor on such clinical glycomics studies is further supported by the fact that Powers *et al.* reported contradicting results for some high mannose structures in an earlier, similar study conducted on 16 patients (65). Whereas a decrease of Man8 in HCC tissues is supported by data presented in this work and by two independent reports from Powers *et al.* (65) and Nie *et al.* (66), the follow up study from Powers and co-workers reported an increase in Man8 in HCC tissues (64). This example clearly shows that future biomarker oriented studies require both large sample cohorts and orthogonal approaches to differentiate individual patient glycan heterogeneity from possible method artefacts and real cancer signatures. This will be crucial to identify glycan signatures with the potential to support histopathology and also future personalized cancer diagnostics or therapies (6).

The here presented liver *O*-glycome showed a low structure diversity. To date just very few data is available on the liver tissue *O*-glycome, and our results agree well with the data provided by the Consortium of functional glycomics that also reported a

similar low *O*-glycan structure diversity in human liver (see glycan structure database for human tissue on: [www.functionalglycomics.org](http://www.functionalglycomics.org)).

The here presented method is generally applicable to any tissue type and opens novel opportunities to study protein *N*- and *O*-glycosylation including isobaric structure differentiation directly from histopathologically examined tissue specimens orthogonal to MALDI imaging techniques. Furthermore, our results unlock the potential of FFPE tissue as a valuable and easily accessible source for future clinical glyco-biomarker studies.

## Acknowledgements

We would like to thank Dr. Ursula Neu for fruitful discussion of the manuscript. We acknowledge support by the Max Planck Society and European Union (Seventh Framework Program "Glycoproteomics" project, grant number PCIG09-GA-2011-293847, IBD-BIOM project, grant number 305479).

## References

1. Moremen, K. W., Tiemeyer, M., and Nairn, A. V. (2012) Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* 13, 448-462
2. Held, W., and Mariuzza, R. A. (2011) Cis-trans interactions of cell surface receptors: biological roles and structural basis. *Cellular and molecular life sciences : CMLS* 68, 3469-3478
3. Croci, D. O., Cerliani, J. P., Dalotto-Moreno, T., Mendez-Huergo, S. P., Mascanfroni, I. D., Dergan-Dylon, S., Toscano, M. A., Caramelo, J. J., Garcia-Vallejo, J. J., Ouyang, J., Mesri, E. A., Junttila, M. R., Bais, C., Shipp, M. A., Salatino, M., and Rabinovich, G. A. (2014) Glycosylation-dependent lectin-receptor interactions preserve angiogenesis in anti-VEGF refractory tumors. *Cell* 156, 744-758
4. Marth, J. D., and Grewal, P. K. (2008) Mammalian glycosylation in immunity. *Nat Rev Immunol* 8, 874-887
5. Stowell, S. R., Ju, T., and Cummings, R. D. (2015) Protein glycosylation in cancer. *Annual review of pathology* 10, 473-510
6. Almeida, A., and Kolarich, D. (2016) The promise of protein glycosylation for personalised medicine. *Biochim Biophys Acta* 1860, 1583-1595



7. Kristic, J., Vuckovic, F., Menni, C., Klaric, L., Keser, T., Beceheli, I., Pucic-Bakovic, M., Novokmet, M., Mangino, M., Thaqi, K., Rudan, P., Novokmet, N., Sarac, J., Missoni, S., Kolcic, I., Polasek, O., Rudan, I., Campbell, H., Hayward, C., Aulchenko, Y., Valdes, A., Wilson, J. F., Gornik, O., Primorac, D., Zoldos, V., Spector, T., and Lauc, G. (2014) Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol Sci Med Sci* 69, 779-789
8. Knezevic, A., Gornik, O., Polasek, O., Pucic, M., Redzic, I., Novokmet, M., Rudd, P. M., Wright, A. F., Campbell, H., Rudan, I., and Lauc, G. (2010) Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. *Glycobiology* 20, 959-969
9. Pucic, M., Knezevic, A., Vidic, J., Adamczyk, B., Novokmet, M., Polasek, O., Gornik, O., Supraha-Goreta, S., Wormald, M. R., Redzic, I., Campbell, H., Wright, A., Hastie, N. D., Wilson, J. F., Rudan, I., Wuhrer, M., Rudd, P. M., Josic, D., and Lauc, G. (2011) High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. *Mol Cell Proteomics* 10, M111 010090
10. Fuster, M. M., and Esko, J. D. (2005) The sweet and sour of cancer: glycans as novel therapeutic targets. *Nature reviews. Cancer* 5, 526-542
11. Dube, D. H., and Bertozzi, C. R. (2005) Glycans in cancer and inflammation [mdash] potential for therapeutics and diagnostics. *Nat Rev Drug Discov* 4, 477-488
12. Christiansen, M. N., Chik, J., Lee, L., Anugraham, M., Abrahams, J. L., and Packer, N. H. (2014) Cell surface protein glycosylation in cancer. *Proteomics* 14, 525-546
13. Miwa, H. E., Song, Y., Alvarez, R., Cummings, R. D., and Stanley, P. (2012) The bisecting GlcNAc in cell growth control and tumor progression. *Glycoconj J* 29, 609-618
14. Adamczyk, B., Tharmalingam, T., and Rudd, P. M. (2012) Glycans as cancer biomarkers. *Biochim Biophys Acta* 1820, 1347-1353

15. Wuhrer, M. (2007) Glycosylation profiling in clinical proteomics: heading for glycan biomarkers. *Expert Rev Proteomics* 4, 135-136
16. Theodoratou, E., Campbell, H., Ventham, N. T., Kolarich, D., Pucic-Bakovic, M., Zoldos, V., Fernandes, D., Pemberton, I. K., Rudan, I., Kennedy, N. A., Wuhrer, M., Nimmo, E., Annese, V., McGovern, D. P., Satsangi, J., and Lauc, G. (2014) The role of glycosylation in IBD. *Nature reviews. Gastroenterology & hepatology* 11, 588-600
17. Abbott, K. L., Lim, J. M., Wells, L., Benigno, B. B., McDonald, J. F., and Pierce, M. (2010) Identification of candidate biomarkers with cancer-specific glycosylation in the tissue and serum of endometrioid ovarian cancer patients by glycoproteomic analysis. *Proteomics* 10, 470-481
18. Chen, C. L., Chung, T., Wu, C. C., Ng, K. F., Yu, J. S., Tsai, C. H., Chang, Y. S., Liang, Y., Tsui, K. H., and Chen, Y. T. (2015) Comparative Tissue Proteomics of Microdissected Specimens Reveals Novel Candidate Biomarkers of Bladder Cancer. *Mol Cell Proteomics* 14, 2466-2478
19. Fischer, A. H., Jacobson, K. A., Rose, J., and Zeller, R. (2008) Hematoxylin and eosin staining of tissue and cell sections. *CSH Protoc* 2008, pdb prot4986
20. Magdeldin, S., and Yamamoto, T. (2012) Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues. *Proteomics* 12, 1045-1058
21. Ostasiewicz, P., Zielinska, D. F., Mann, M., and Wisniewski, J. R. (2010) Proteome, phosphoproteome, and N-glycoproteome are quantitatively preserved in formalin-fixed paraffin-embedded tissue and analyzable by high-resolution mass spectrometry. *Journal of proteome research* 9, 3688-3700
22. de Gramont, A., Watson, S., Ellis, L. M., Rodon, J., Tabernero, J., de Gramont, A., and Hamilton, S. R. (2015) Pragmatic issues in biomarker evaluation for targeted therapies in cancer. *Nat Rev Clin Oncol* 12, 197-212
23. Klopffleisch, R., Weiss, A. T., and Gruber, A. D. (2011) Excavation of a buried treasure--DNA, mRNA, miRNA and protein analysis in formalin fixed, paraffin embedded tissues. *Histology and histopathology* 26, 797-810

24. Casadonte, R., and Caprioli, R. M. (2011) Proteomic analysis of formalin-fixed paraffin-embedded tissue by MALDI imaging mass spectrometry. *Nat Protoc* 6, 1695-1709
25. Turiak, L., Shao, C., Meng, L., Khatri, K., Leymarie, N., Wang, Q., Pantazopoulos, H., Leon, D. R., and Zaia, J. (2014) Workflow for combined proteomics and glycomics profiling from histological tissues. *Anal Chem* 86, 9670-9678
26. Serth, J., Kuczyk, M. A., Paeslack, U., Lichtinghagen, R., and Jonas, U. (2000) Quantitation of DNA extracted after micropreparation of cells from frozen and formalin-fixed tissue sections. *Am J Pathol* 156, 1189-1196
27. Roberts, L., Bowers, J., Sensinger, K., Lisowski, A., Getts, R., and Anderson, M. G. (2009) Identification of methods for use of formalin-fixed, paraffin-embedded tissue samples in RNA expression profiling. *Genomics* 94, 341-348
28. Kelly, A. D., Breitkopf, S. B., Yuan, M., Goldsmith, J., Spentzos, D., and Asara, J. M. (2011) Metabolomic profiling from formalin-fixed, paraffin-embedded tumor tissue using targeted LC/MS/MS: application in sarcoma. *PLoS One* 6, e25357
29. van Wijk, X. M., Vallen, M. J., van de Westerlo, E. M., Oosterhof, A., Hao, W., Versteeg, E. M., Raben, J., Wismans, R. G., Smetsers, T. F., Dijkman, H. B., Schalkwijk, J., and van Kuppevelt, T. H. (2012) Extraction and structural analysis of glycosaminoglycans from formalin-fixed, paraffin-embedded tissues. *Glycobiology* 22, 1666-1672
30. Toghi Eshghi, S., Yang, S., Wang, X., Shah, P., Li, X., and Zhang, H. (2014) Imaging of N-linked glycans from formalin-fixed paraffin-embedded tissue sections using MALDI mass spectrometry. *ACS Chem Biol* 9, 2149-2156
31. Gustafsson, O. J., Briggs, M. T., Condina, M. R., Winderbaum, L. J., Pelzing, M., McColl, S. R., Everest-Dass, A. V., Packer, N. H., and Hoffmann, P. (2015) MALDI imaging mass spectrometry of N-linked glycans on formalin-fixed paraffin-embedded murine kidney. *Anal Bioanal Chem* 407, 2127-2139

32. Dwek, M. V., Brooks, S. A., Streets, A. J., Harvey, D. J., and Leathem, A. J. (1996) Oligosaccharide release from frozen and paraffin-wax-embedded archival tissues. *Anal Biochem* 242, 8-14
33. Satomaa, T., Heiskanen, A., Leonardsson, I., Angstrom, J., Olonen, A., Blomqvist, M., Salovuori, N., Haglund, C., Teneberg, S., Natunen, J., Carpen, O., and Saarinen, J. (2009) Analysis of the human cancer glycome identifies a novel group of tumor-associated N-acetylglucosamine glycan antigens. *Cancer Res* 69, 5811-5819
34. Furukawa, J., Piao, J., Yoshida, Y., Okada, K., Yokota, I., Higashino, K., Sakairi, N., and Shinohara, Y. (2015) Quantitative O-Glycomics by Microwave-Assisted beta-Elimination in the Presence of Pyrazolone Analogues. *Anal Chem* 87, 7524-7528
35. Jensen, P. H., Karlsson, N. G., Kolarich, D., and Packer, N. H. (2012) Structural analysis of N- and O-glycans released from glycoproteins. *Nat Protoc* 7, 1299-1310
36. Kolarich, D., Windwarder, M., Alagesan, K., and Altmann, F. (2015) Isomer-Specific Analysis of Released N-Glycans by LC-ESI MS/MS with Porous Graphitized Carbon. *Methods Mol Biol* 1321, 427-435
37. York, W. S., Agravat, S., Aoki-Kinoshita, K. F., McBride, R., Campbell, M. P., Costello, C. E., Dell, A., Feizi, T., Haslam, S. M., Karlsson, N., Khoo, K. H., Kolarich, D., Liu, Y., Novotny, M., Packer, N. H., Paulson, J. C., Rapp, E., Ranzinger, R., Rudd, P. M., Smith, D. F., Struwe, W. B., Tiemeyer, M., Wells, L., Zaia, J., and Kettner, C. (2014) MIRAGE: the minimum information required for a glycomics experiment. *Glycobiology* 24, 402-406
38. Wessel, D., and Flugge, U. I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* 138, 141-143

39. Lee, A., Kolarich, D., Haynes, P. A., Jensen, P. H., Baker, M. S., and Packer, N. H. (2009) Rat liver membrane glycoproteome: enrichment by phase partitioning and glycoprotein capture. *Journal of proteome research* 8, 770-781
40. Stavenhagen, K., Kolarich, D., and Wührer, M. (2015) Clinical Glycomics Employing Graphitized Carbon Liquid Chromatography-Mass Spectrometry. *Chromatographia* 78, 307-320
41. Pabst, M., and Altmann, F. (2008) Influence of Electrosorption, Solvent, Temperature, and Ion Polarity on the Performance of LC-ESI-MS Using Graphitic Carbon for Acidic Oligosaccharides. *Analytical Chemistry* 80, 7534-7542
42. Palmisano, G., Larsen, M. R., Packer, N. H., and Thaysen-Andersen, M. (2013) Structural analysis of glycoprotein sialylation -part II: LC-MS based detection. *Rsc Advances* 3, 22706-22726
43. Harvey, D. J. (1999) Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass spectrometry reviews* 18, 349-450
44. Harvey, D. J. (2005) Fragmentation of negative ions from carbohydrates: part 1. Use of nitrate and other anionic adducts for the production of negative ion electrospray spectra from N-linked carbohydrates. *J Am Soc Mass Spectrom* 16, 622-630
45. Harvey, D. J. (2005) Fragmentation of negative ions from carbohydrates: part 2. Fragmentation of high-mannose N-linked glycans. *J Am Soc Mass Spectrom* 16, 631-646
46. Harvey, D. J. (2005) Fragmentation of negative ions from carbohydrates: part 3. Fragmentation of hybrid and complex N-linked glycans. *J Am Soc Mass Spectrom* 16, 647-659
47. Harvey, D. J., Jaeken, J., Butler, M., Armitage, A. J., Rudd, P. M., and Dwek, R. A. (2010) Fragmentation of negative ions from N-linked carbohydrates, part 4. Fragmentation of complex glycans lacking substitution on the 6-antenna. *J Mass Spectrom* 45, 528-535

48. Harvey, D. J., and Rudd, P. M. (2011) Fragmentation of negative ions from N-linked carbohydrates. Part 5: Anionic N-linked glycans. *International Journal of Mass Spectrometry* 305, 120-130
49. Harvey, D. J., Edgeworth, M., Krishna, B. A., Bonomelli, C., Allman, S. A., Crispin, M., and Scrivens, J. H. (2014) Fragmentation of negative ions from N-linked carbohydrates: part 6. Glycans containing one N-acetylglucosamine in the core. *Rapid communications in mass spectrometry : RCM* 28, 2008-2018
50. Harvey, D. J., and Abrahams, J. L. (2016) Fragmentation and ion mobility properties of negative ions from N-linked carbohydrates: Part 7. Reduced glycans. *Rapid communications in mass spectrometry : RCM* 30, 627-634
51. Harvey, D. J., Royle, L., Radcliffe, C. M., Rudd, P. M., and Dwek, R. A. (2008) Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry. *Analytical biochemistry* 376, 44-60
52. Everest-Dass, A. V., Kolarich, D., Campbell, M. P., and Packer, N. H. (2013) Tandem mass spectra of glycan substructures enable the multistage mass spectrometric identification of determinants on oligosaccharides. *Rapid communications in mass spectrometry : RCM* 27, 931-939
53. Thomsson, K. A., Holmen-Larsson, J. M., Angstrom, J., Johansson, M. E., Xia, L., and Hansson, G. C. (2012) Detailed O-glycomics of the Muc2 mucin from colon of wild-type, core 1- and core 3-transferase-deficient mice highlights differences compared with human MUC2. *Glycobiology* 22, 1128-1139
54. Schulz, B. L., Packer, N. H., and Karlsson, N. G. (2002) Small-Scale Analysis of O-Linked Oligosaccharides from Glycoproteins and Mucins Separated by Gel Electrophoresis. *Analytical Chemistry* 74, 6088-6097
55. Schulz, B. L., Sloane, A. J., Robinson, L. J., Prasad, S. S., Lindner, R. A., Robinson, M., Bye, P. T., Nielson, D. W., Harry, J. L., Packer, N. H., and Karlsson, N. G. (2007) Glycosylation of sputum mucins is altered in cystic fibrosis patients. *Glycobiology* 17, 698-712

56. Trombetta, E. S. (2003) The contribution of N-glycans and their processing in the endoplasmic reticulum to glycoprotein biosynthesis. *Glycobiology* 13, 77r-91r
57. Varki A, C. R., Esko JD, et al., editors. (2009) In: Essentials of Glycobiology. 2nd edition. Chapter 8 and Chapter 9. *Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press*
58. Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J. J., Schnaar, R. L., Freeze, H. H., Marth, J. D., Bertozzi, C. R., Etzler, M. E., Frank, M., Vliegenthart, J. F., Lutteke, T., Perez, S., Bolton, E., Rudd, P., Paulson, J., Kanehisa, M., Toukach, P., Aoki-Kinoshita, K. F., Dell, A., Narimatsu, H., York, W., Taniguchi, N., and Kornfeld, S. (2015) Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* 25, 1323-1324
59. Ceroni, A., Maass, K., Geyer, H., Geyer, R., Dell, A., and Haslam, S. M. (2008) GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *Journal of proteome research* 7, 1650-1659
60. Benjamini, Y., and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57, 289-300
61. Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D. (2013) R Development Core Team (2012) nlme: linear and nonlinear mixed effects models. R package version 3.1-103. *R Foundation for Statistical Computing, Vienna*
62. Tsuboi, S. (2012) Tumor defense systems using O-glycans. *Biol Pharm Bull* 35, 1633-1636
63. Suzuki, Y., Sutoh, M., Hatakeyama, S., Mori, K., Yamamoto, H., Koie, T., Saitoh, H., Yamaya, K., Funyu, T., Habuchi, T., Arai, Y., Fukuda, M., Ohyama, C., and Tsuboi, S. (2012) MUC1 carrying core 2 O-glycans functions as a molecular shield against NK cell attack, promoting bladder tumor metastasis. *International journal of oncology* 40, 1831-1838

64. Powers, T. W., Holst, S., Wuhrer, M., Mehta, A. S., and Drake, R. R. (2015) Two-Dimensional N-Glycan Distribution Mapping of Hepatocellular Carcinoma Tissues by MALDI-Imaging Mass Spectrometry. *Biomolecules* 5, 2554-2572
65. Powers, T. W., Neely, B. A., Shao, Y., Tang, H., Troyer, D. A., Mehta, A. S., Haab, B. B., and Drake, R. R. (2014) MALDI imaging mass spectrometry profiling of N-glycans in formalin-fixed paraffin embedded clinical tissue blocks and tissue microarrays. *PLoS One* 9, e106255
66. Nie, H., Liu, X., Zhang, Y., Li, T., Zhan, C., Huo, W., He, A., Yao, Y., Jin, Y., Qu, Y., Sun, X. L., and Li, Y. (2015) Specific N-glycans of Hepatocellular Carcinoma Cell Surface and the Abnormal Increase of Core-alpha-1, 6-fucosylated Triantennary Glycan via N-acetylglucosaminyltransferases-IVa Regulation. *Sci Rep* 5, 16007



## Figure Legends

### **Figure 1: Overview on the FFPE-tissue workflow for clinical *N*- and *O*-glycomics.**

Entire tissue blocks, slide-mounted tissue sections before or after H&E staining or micro-dissected tissue specimens provide clinical tissue sources. After antigen retrieval (glyco)proteins are immobilized onto PVDF membranes and *N*- and *O*-glycans are sequentially released for porous graphitized carbon nanoLC-ESI MS/MS based glycomics.

### **Figure 2: Comparison of *N*- and *O*-glycans released from unmounted FFPE tissue sections and matching frozen human hepatic tissue.**

**A:** Heat map obtained after unsupervised clustering of the detected *N*-glycans generated from various technical replicates of the two sample sources. Detected *N*-glycan structures (the ten most abundant compositions are colored, see also Supplementary Material 7a for an overview) are represented by structure ID numbers on the top side of the heat map. The structure numbering refers to the respective IDs described in Supplementary Material 7a. Asterisks above numbers indicate the *p*-value range obtained for statistically relevant differences between the starting materials [ $\leq 0.01$  (\*\*) or  $\leq 0.05$  (\*)]. The heat map is sorted according to different structure classes (dendrograms at the bottom of the heat map), whereas the different color boxes on the bottom provide additional structural information on the individual *N*-glycans (e.g. NeuAc or fucose linkage, presence of bisecting GlcNAc).

The samples cluster according to sample origin (frozen and FFPE) due to differences in the relative intensity of mainly  $\alpha 2,3/\alpha 2,6$  linked NeuAc residues.

**B:** Structures are depicted for the ten most abundant *N*-glycan compositions (red label). Structure 19 and 25 (purple label) were the 10<sup>th</sup> most abundant composition in one sample source.

**C:** Category comparison of the hepatic *N*-glycome obtained from frozen or FFPE starting material. Error bars represent the standard deviation determined from 5 technical replicates. The number of structures in each category is indicated above bars.

**D:** Comparison of the hepatic *O*-glycome obtained from frozen or FFPE starting material. The structure numbering refers to the respective IDs described in Supplementary Material 8a.

**Figure 3: Comparison of the hepatic *N*- and *O*-glycome obtained from unstained and hematoxylin and eosin (H&E) stained FFPE tissue sections.**

**A:** Heat map obtained after unsupervised clustering of the detected *N*-glycans, sorted according to *N*-glycan categories. The differences in relative *N*-glycan abundances indicated by the clustering analyses between the H&E stained and untreated FFPE sections are smaller than the methodological and individual slide specific differences as indicated by the fact that no clear sample source derived clustering was obtained. Detected *N*-glycan structures are represented by structure ID numbers (for details see also Supplementary Material 7a) on top of the heat map. The heat map is sorted according to different structure classes (bottom of the heat

map), whereas the different color boxes on the bottom provide additional structural information on the individual *N*-glycans (e.g. NeuAc or fucose linkage, presence of bisecting GlcNAc).

**B:** Category comparison of the hepatic *N*-glycome obtained from unstained (blue) or H&E stained (red) FFPE tissue sections. Error bars represent the standard deviation determined from 4 technical replicates. The number of structures in each category is indicated above bars. The asterisks indicate the *p*-value range obtained for statistically relevant differences between the starting materials [ $\leq 0.05$  (\*)].

**C:** Comparison of the hepatic *O*-glycome obtained from unstained (blue) or H&E stained (red) FFPE tissue sections. Error bars represent the standard deviation determined from the analysis of 5 technical replicates.

**Figure 4: *N*-glycome determined from hepatocytes (1,000; 5,000 and 10,000 cells) microdissected from H&E stained FFPE tissue slides and analyzed by porous graphitized carbon nanoLC-ESI MS/MS glycomics.**

**A:** *N*-glycan base peak chromatograms (BPC) obtained from 1000; 5,000 and 10,000 cells isolated from H&E stained FFPE tissue sections by laser capture microdissection. The intensity for all BPCs was normalized to the most intense sample (10,000 cells). The major *N*-glycan structures have been indicated at their respective retention time in the BPC. The asterisk indicates a non-glycan contamination signal.

**B:** Heat map obtained after unsupervised hierarchical clustering of the *N*-glycans detected from 1,000; 5,000 and 10,000 cells compared to the ones obtained from

entire FFPE tissue sections and sorted according to major structure categories. Detected *N*-glycan structures are represented by structure ID numbers (for details see also Supplementary Material 7a) on top of the heat map. The heat map is sorted according to different structure classes (bottom of the heat map), whereas the different color boxes on the bottom provide additional structural information on the individual *N*-glycans (e.g. NeuAc or fucose linkage, presence of bisecting GlcNAc). Despite the fact that with lower cell numbers less *N*-glycans were detectable (indicated by the white boxes), no cell number associated clustering occurred, indicating that the overall profile remains comparable independent of the cell number and that the microdissection itself does not alter the glycan profile.

**Figure 5: *N*- and *O*-glycome of hepatocellular carcinoma (HCC) and non-cancer hepatic tissue (NC).**

**A:** Representative image of FFPE preserved tissue sections of HCC used for spatial isolation of HCC cells and surrounding non-cancerous tissue by laser capture microdissection. The respective areas used for the analyses are labelled as indicated.

**B:** Overview on the *N*-glycan structures exhibiting the largest expression level changes between HCC and NC tissue including their structure IDs as listed in Supplementary Material 7a. The number of asterisks indicate the *p*-value range obtained for statistically relevant differences between the starting materials [ $\leq 0.01$  (\*\*) or  $\leq 0.05$  (\*)].

**C:** Heat map obtained after unsupervised hierarchical clustering of the *N*-glycans detected from ~2,000 cells of HCC and NC hepatic tissue and sorted according to

major structure categories. Detected *N*-glycan structures are represented by structure ID numbers (for details see also Supplementary Material 7a) on top of the heat map. The heat map is sorted according to different structure classes (bottom of the heat map), whereas the different color boxes on the bottom provide additional structural information on the individual *N*-glycans (e.g. NeuAc or fucose linkage, presence of bisecting GlcNAc). Based on the *N*-glycome, HCC and NC tissue cluster in distinct classes. Paucimannose type and  $\alpha$ 2,6 NeuAc carrying *N*-glycans show the highest alterations between these two tissue types.

**D:** Category comparison of the HCC and NC *N*-glycomes. The number of structures in each category is indicated above bars. Error bars represent standard deviation determined from 3 technical replicates.

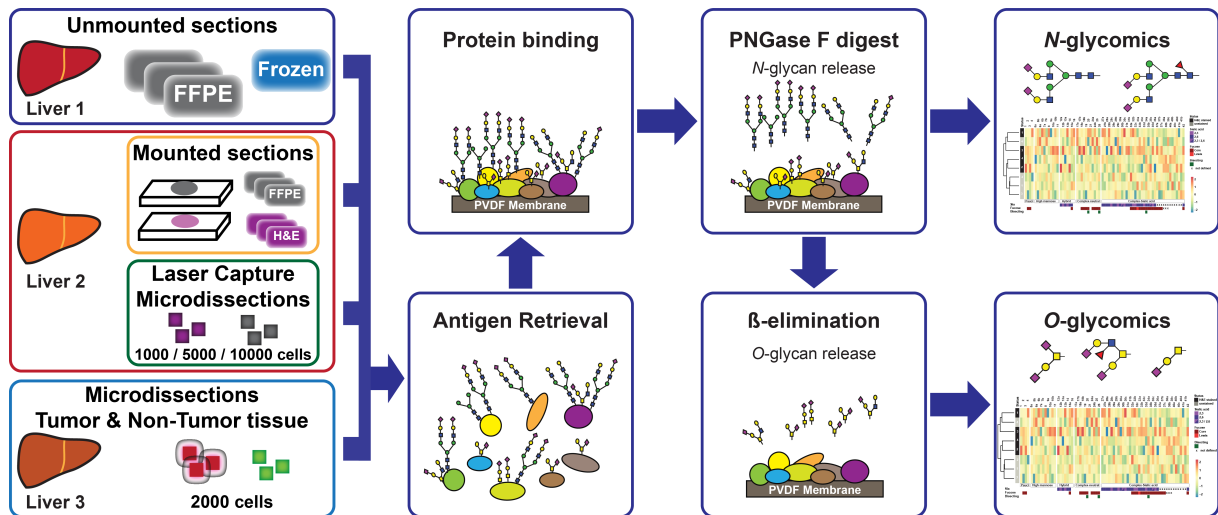
**E:** Comparison of HCC and NC *O*-glycomes obtained from the very same material the *N*-glycomes were obtained. Sialyl Lewis X epitopes present on core 2 type *O*-glycans show a significant increase in HCC tissue, whereas core 1 type *O*-glycan levels are reduced.

**Figure 6: Example for structure identification using porous graphitized carbon nano LC-ESI MS/MS.**

**A:** Base peak chromatogram (BPC, red trace) representing the *N*-glycome obtained from FFPE preserved hepatic tissue. Extracted ion chromatogram (EIC, black trace) of an example *N*-glycan (Hex<sub>5</sub>HexNAc<sub>4</sub>NeuAcFuc,  $[M-2H]^{2-} = 1038.9$  Da) that is present in five different structure isomers.

**B:** Individual product-ion spectra of the five Hex<sub>5</sub>HexNAC<sub>4</sub>NeuAcFuc isomers enabling differentiation and relative quantitation of the various *N*-glycan isomers.

## Figures



**Figure 1: Overview on the FFPE-tissue workflow for clinical *N*- and *O*-glycomics.**

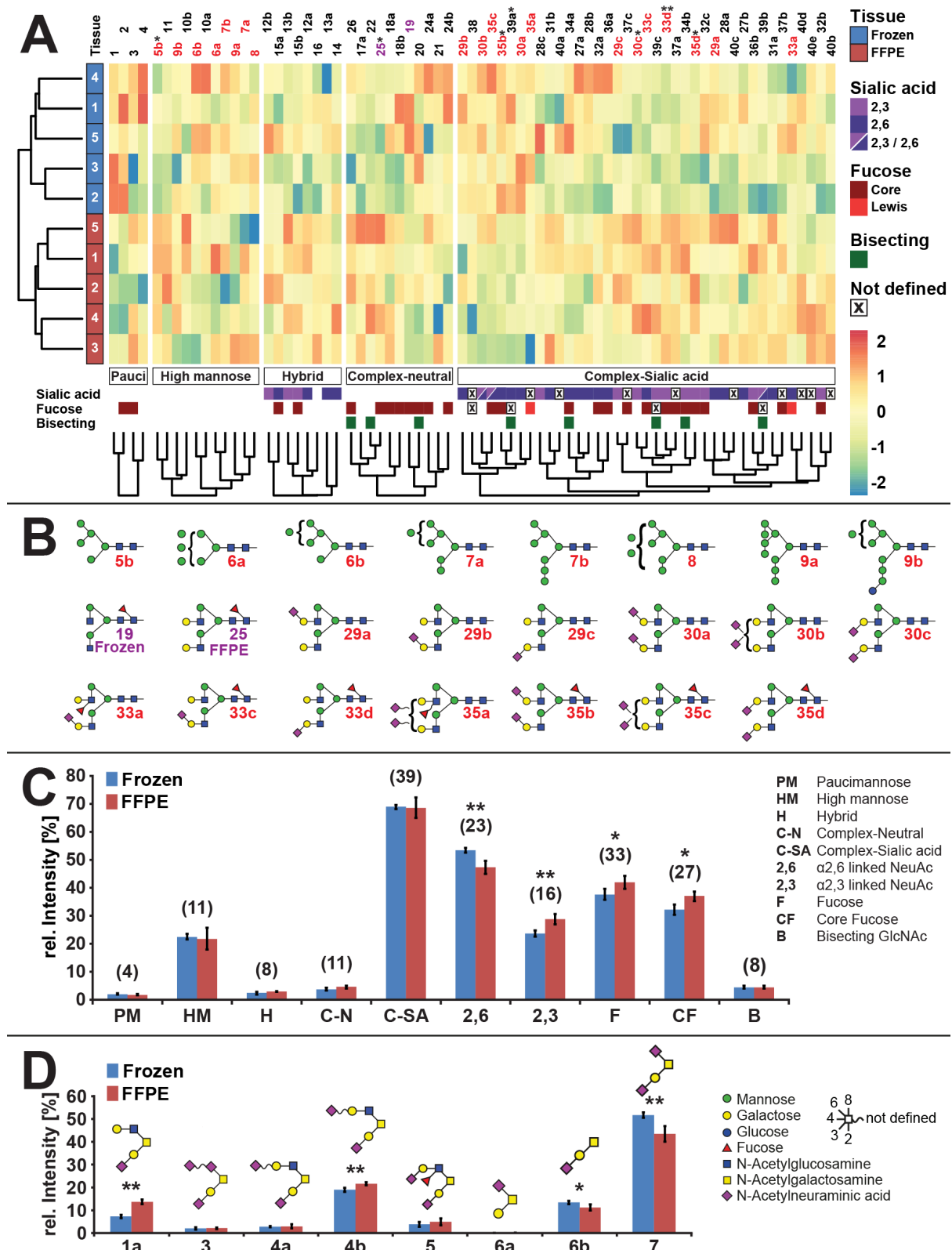
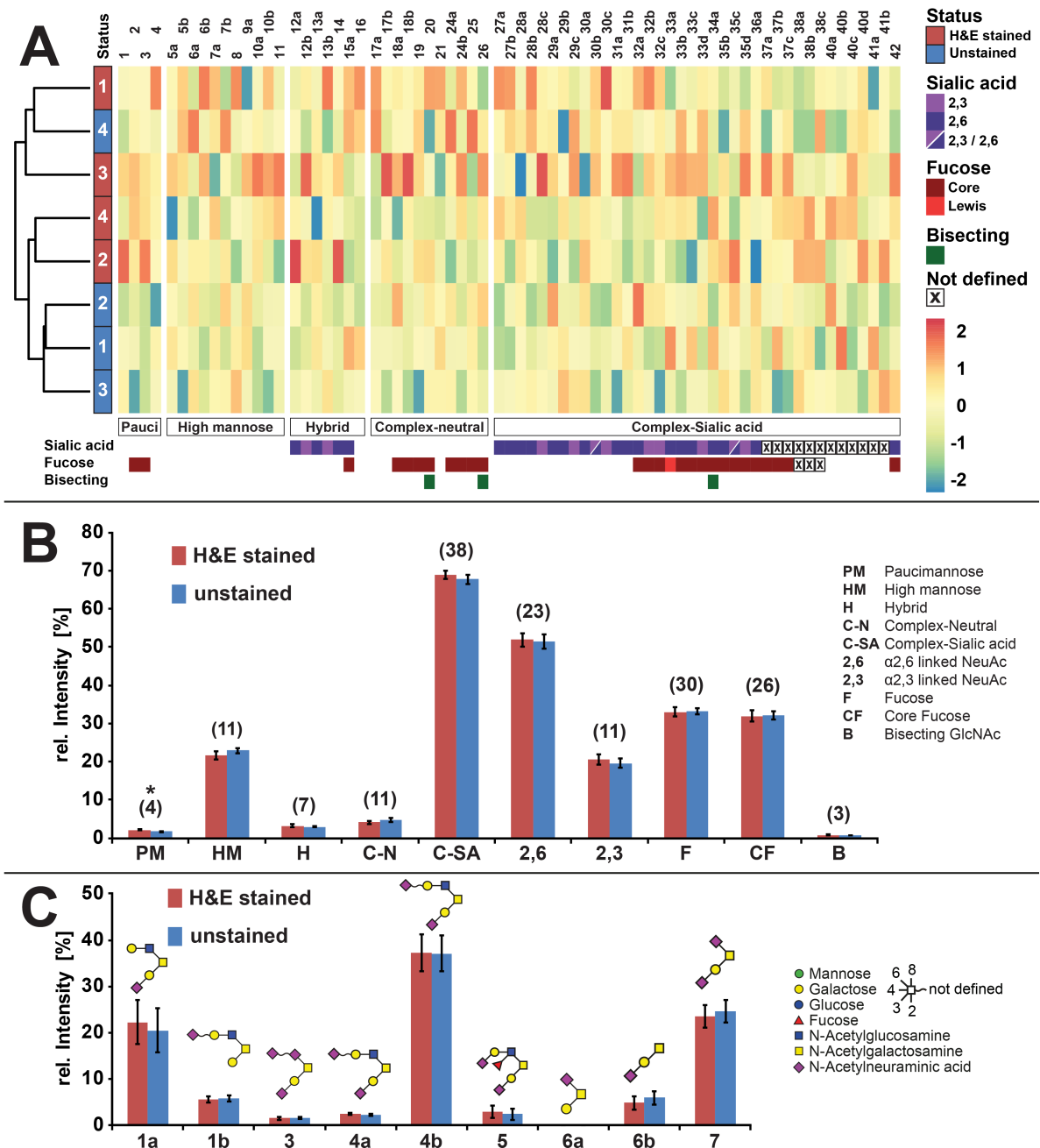
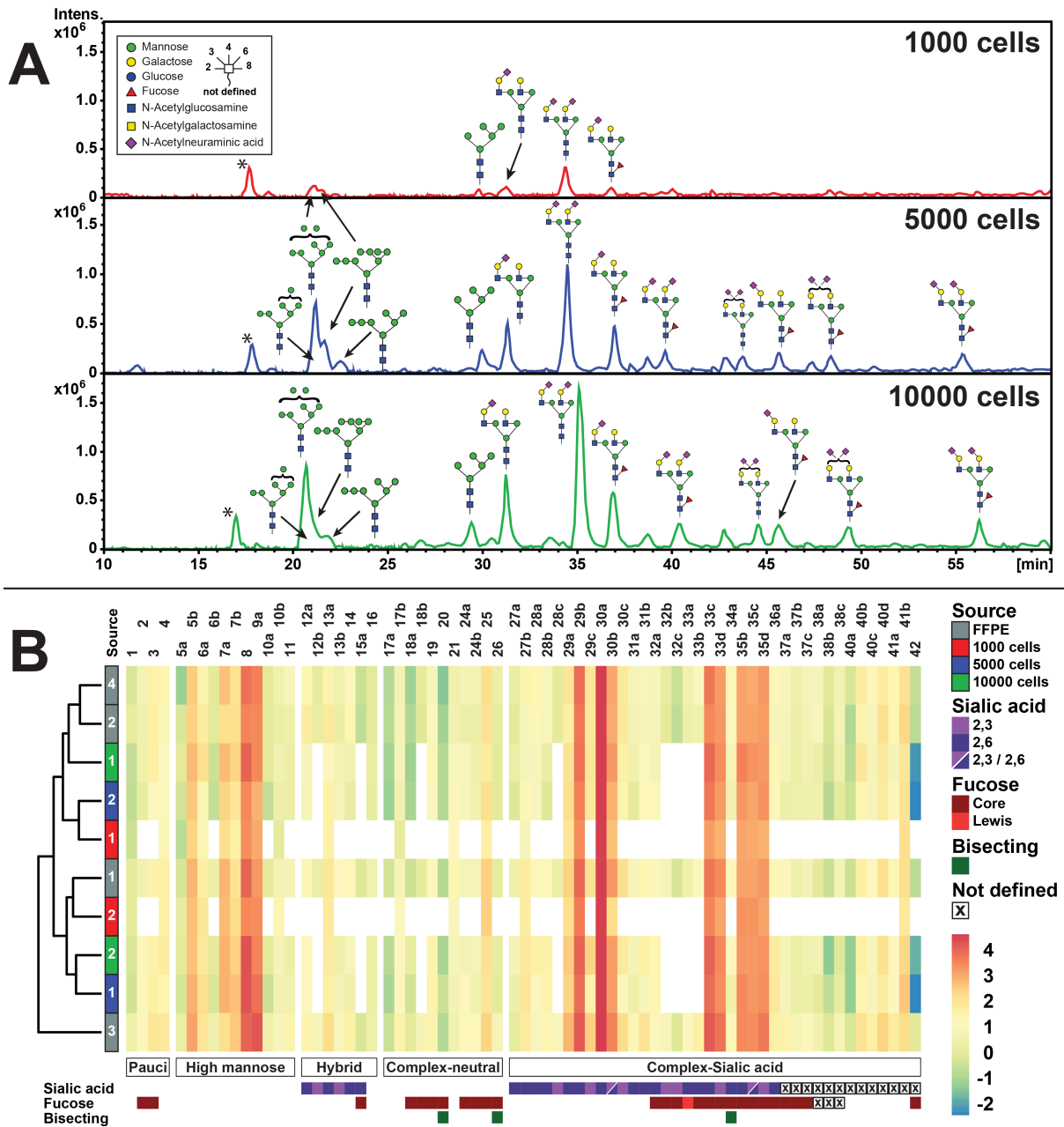


Figure 2: Comparison of *N*- and *O*-glycans released from unmounted FFPE tissue sections and matching frozen human hepatic tissue.

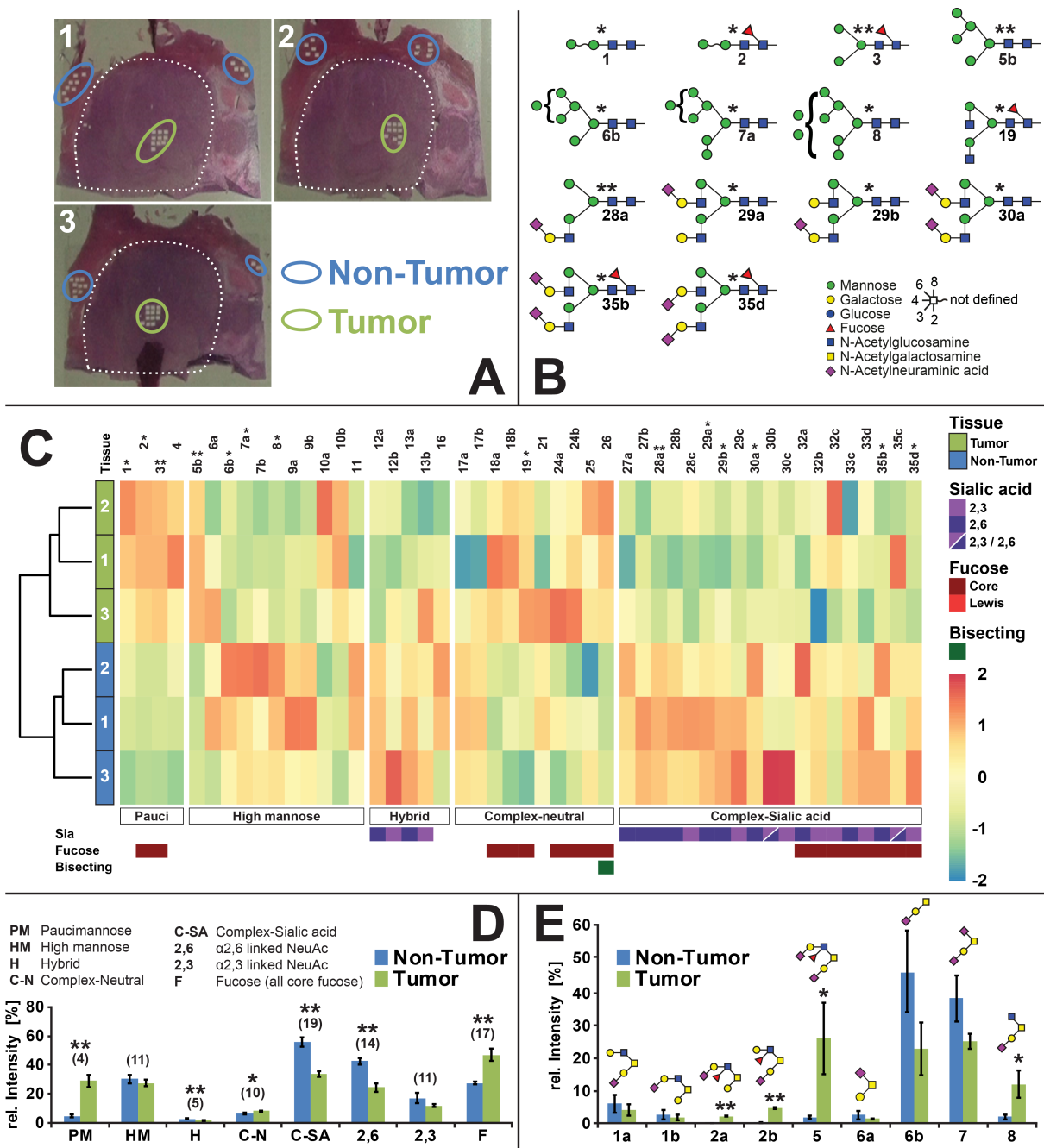


**Figure 3: Comparison of the hepatic *N*- and *O*-glycome obtained from unstained and hematoxylin and eosin (H&E) stained FFPE tissue sections.**





**Figure 4: N-glycome determined from hepatocytes (1,000; 5,000 and 10,000 cells) microdissected from H&E stained FFPE tissue slides and analyzed by porous graphitized carbon nanoLC-ESI MS/MS glycomics.**



**Figure 5: N- and O-glycome of hepatocellular carcinoma (HCC) and non-cancer hepatic tissue (NC).**

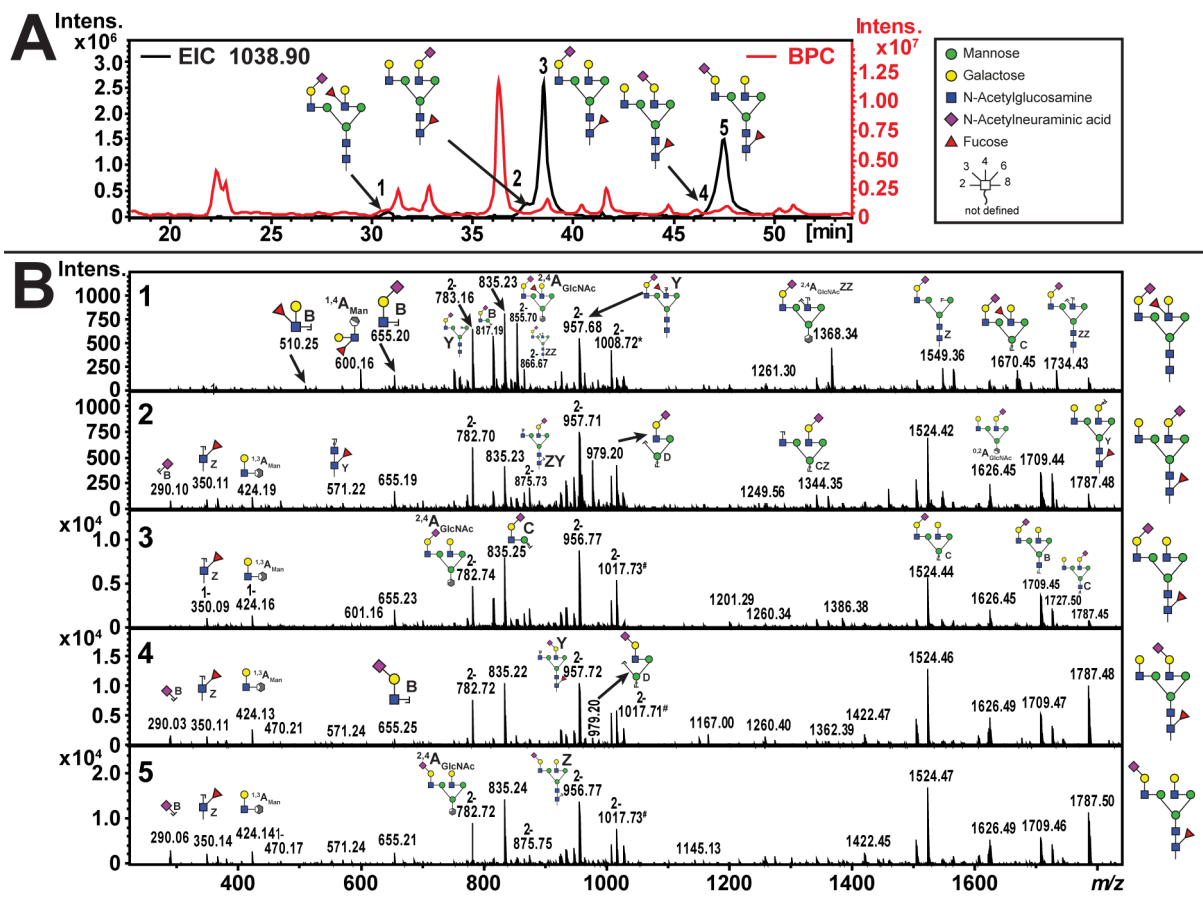



Figure 6: Example for structure identification using porous graphitized carbon nano LC-ESI MS/MS.

## RESEARCH ARTICLE

# The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics

Hannes Hinneburg,<sup>1,2</sup> Kathrin Stavenhagen,<sup>3</sup> Ulrike Schweiger-Hufnagel,<sup>4</sup> Stuart Pengelley,<sup>4</sup> Wolfgang Jabs,<sup>4</sup> Peter H. Seeberger,<sup>1,2</sup> Daniel Varón Silva,<sup>1</sup> Manfred Wuhrer,<sup>3,5</sup> Daniel Kolarich<sup>1</sup> 

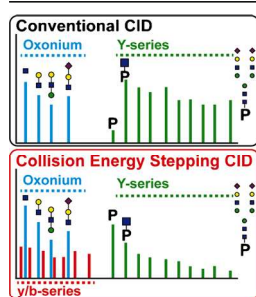
<sup>1</sup>Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, 14424, Potsdam, Germany

<sup>2</sup>Institute of Chemistry and Biochemistry, Freie Universität Berlin, Arnimallee 22, 14195, Berlin, Germany

<sup>3</sup>Division of BioAnalytical Chemistry, VU University Amsterdam, Amsterdam, The Netherlands

<sup>4</sup>Bruker Daltonik GmbH, Bremen, Germany

<sup>5</sup>Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands



**Abstract.** In-depth site-specific investigations of protein glycosylation are the basis for understanding the biological function of glycoproteins. Mass spectrometry-based N- and O-glycopeptide analyses enable determination of the glycosylation site, site occupancy, as well as glycan varieties present on a particular site. However, the depth of information is highly dependent on the applied analytical tools, including glycopeptide fragmentation regimes and automated data analysis. Here, we used a small set of synthetic disialylated, biantennary N-glycopeptides to systematically tune Q-TOF instrument parameters towards optimal energy stepping collision induced dissociation (CID) of glycopeptides. A linear dependency of  $m/z$ -ratio and optimal fragmentation energy was found, showing that with increasing  $m/z$ -ratio, more energy

is required for glycopeptide fragmentation. Based on these optimized fragmentation parameters, a method combining lower- and higher-energy CID was developed, allowing the online acquisition of glycan and peptide-specific fragments within a single tandem MS experiment. We validated this method analyzing a set of human immunoglobulins (IgA1+2, sIgA, IgG1+2, IgE, IgD, IgM) as well as bovine fetuin. These optimized fragmentation parameters also enabled software-assisted glycopeptide assignment of both N- and O-glycopeptides including information about the most abundant glycan compositions, peptide sequence and putative structures. Twenty-six out of 30 N-glycopeptides and four out of five O-glycopeptides carrying >110 different glycoforms could be identified by this optimized LC-ESI tandem MS method with minimal user input. The Q-TOF based glycopeptide analysis platform presented here opens the way to a range of different applications in glycoproteomics research as well as biopharmaceutical development and quality control.

**Keywords:** Glycoproteomics, Immunoglobulin, Q-TOF, Collision energy stepping CID, Synthetic glycopeptides, Glycopeptide, N-glycan, O-glycan

Received: 29 July 2015/Revised: 6 November 2015/Accepted: 9 November 2015/Published Online: 4 January 2016

## Introduction

Protein glycosylation is one of the most common post-translational modifications [1]. The vast majority of

membrane and secreted proteins are known or predicted to be N- and O-glycosylated [1–4]. Glycoproteins represent key molecules in many important biological processes such as cell adhesion, endocytosis, receptor activation, signal transduction, molecular trafficking, and clearance, as well as in diseases, including cancer [5, 6]. In-depth approaches to determine site-specific protein glycosylation have become indispensable tools for functional analyses of these complex biomolecules [7–9]. Specific glycans present on individual sites of a protein

**Electronic supplementary material** The online version of this article (doi:10.1007/s13361-015-1308-6) contains supplementary material, which is available to authorized users.

Correspondence to: Daniel Kolarich; e-mail: daniel.kolarich@mpikg.mpg.de

have been shown to be crucial for influencing the physico-chemical and functional properties of their respective protein carriers. The presence or absence of a single core fucose attached to biantennary complex-type structures of the crystallizable fragment (Fc) domain of immunoglobulin G (IgG) influences the interaction of the antibody with its Fc $\gamma$ -receptor, leading to a modulation of the antibody-dependent cellular cytotoxicity [10]. In IgE, the site-specific presence of an oligomannose-type N-glycan has been shown to be necessary for initiating anaphylaxis [11]. These examples illustrate that knowledge about site-specific glycosylation is an important prerequisite for studying the functional impact of protein glycosylation.

Glycoproteomic approaches, with focus on the analysis of N- and O-glycopeptides, can give detailed information on the type of structures present at a given site of a specific protein. The high sensitivity and selectivity of modern mass spectrometers in combination with different ionization methods, fragmentation techniques, and mass analyzers have made mass spectrometry the method of choice in glycoproteomics. However, analysis of glycopeptides can still be hindered by their low abundance in the entire peptide pool after proteolytic digestion because of their microheterogeneity (glycan variety attached to one glycosylation site) and macroheterogeneity (site occupancy). Furthermore, they have a general tendency to be less well ionized compared with non-glycosylated peptides [12], which can be compensated but not avoided by dedicated enrichment steps such as lectin affinity chromatography or hydrophilic interaction liquid chromatography (HILIC) [13, 14].

Glycopeptide fragmentation via tandem mass spectrometry can be achieved under standard collision induced dissociation (CID) conditions, preferentially yielding glycan fragments by cleaving the glycosidic bonds between carbohydrate units (B- and Y-ions), but rarely providing sufficient peptide b- and y-type ions (cleavage of peptide bond) for unambiguous peptide identification [15]. Higher-energy CID (HCD), in contrast, mainly results in b/y-type peptide ions as well as glycan oxonium ions and fewer Y-type ions from fragmentation of the glycosidic linkage [16]. Alternatively, electron transfer dissociation (ETD) and electron capture dissociation (ECD) keep the glycan portion on the modified amino acid mainly intact and the observed c- and z-ions produced by the N-C $\alpha$  bond cleavage can provide amino acid sequence information complementary to CID fragmentation [13, 14]. However, because of their size, the majority of glycopeptides are usually detected in the  $m/z$  range >900 Da, thus impeding with ETD experiments. Though Alley et al. reported that ETD was able to provide useful spectra for glycopeptide precursors below  $m/z$  1400 [17], our experience shows that best results are obtained for highly charged precursors ( $\geq +3$ ) in an  $m/z$  range <850 Da (personal observation).

Simultaneous acquisition of MS spectra at lower and higher collision energies (collision energy stepping CID) represents an attractive approach for selectively yielding fragment ions

covering both the glycan and the peptide moieties of glycopeptides [18–21]. Nevertheless, in case higher dissociation energies are to be applied, adjusting the optimal fragmentation conditions is a prerequisite to ensure optimal fragmentation results [19], especially for the peptide part. In that context, synthetic and highly defined glycopeptide standards represent a unique and ideal tool to systematically determine the optimal conditions for glycopeptide fragmentation.

Successful large scale and software-assisted data analysis of glycopeptide product-ion spectra is not only instrument-dependent but also requires dedicated software tools. Despite the fact that great achievements have been made in this field within the last years, recent reviews on available software packages also pointed out some shortcomings due to the lack of desirable functions for comprehensive and automated glycopeptide analysis [22, 23]. Important features, such as elucidation of N- and O-glycans, matching of peptides to known protein sequences, scoring (ranking) of potential glycan and peptide moieties, parallel detection of non-glycosylated peptides, usage of product-ion data, or the option for batch inputs, have hitherto not been realized within a single software tool [22, 23].

Here we used synthetic and thus well-defined N-glycopeptides for the systematic optimization of CID energy parameters on quadrupole-time of flight (Q-TOF) instruments to obtain maximum information on both the glycan and peptide moiety within a single tandem MS experiment. Synthetic glycopeptides carrying biantennary, disialylated N-glycans were used to elucidate optimal collision energies for both the glycan and the peptide fraction. The optimized parameters were subsequently validated in a LC-ESI tandem MS online setup using tryptic peptides derived from the entire panel of human immunoglobulins (Igs) and bovine fetuin. Glycopeptide fragment data obtained from these optimized collision energy stepping CID conditions enabled software-assisted N- and O-glycopeptide data analysis in a semi-automated manner, including glycopeptide classification, compositional glycan prediction, and peptide moiety identification.

## Experimental

### *Material and Methods*

If not otherwise stated, all materials were purchased in high quality from Sigma-Aldrich (St. Louis, MO, USA). Trypsin (sequencing grade) was obtained from Roche Diagnostic GmbH (Mannheim, Germany). Water was used after purification with a Milli Q-8 direct system (Merck KGaA, Darmstadt, Germany). IgA (plasma), sIgA (human colostrum), IgD (plasma), IgE (myeloma plasma, lambda), IgG1 (myeloma plasma, kappa), IgG2 (myeloma plasma, kappa), and IgM (plasma) were obtained from Athens Research & Technology (Athens, GA, USA). The amino acid numbering applied for all proteins analyzed in this study is based on the respective UniProtKB entries.

### Glycopeptide Design and Purification

A tryptic N-glycosylated peptide sequence present in human protein C (entry P04070, <sub>284</sub>EVFVHPNYSK<sub>293</sub>) [24] was selected to design a small panel of synthetic N-glycopeptide standards. Besides the naturally occurring peptide sequence (termed GP-M), variations of the sequence were produced with the glycosylated amino acid either moved towards the N-terminus (GP-N) or the C-terminus (GP-C) of the peptide sequence. All glycopeptides were synthesized carrying a disialylated, biantennary N-glycan (Table 1). Glycopeptide synthesis and purification were performed as described previously [12].

### Systematic Optimization of Collision Energy Parameters for Glycopeptide Fragmentation

Synthetic glycopeptides were dissolved in 50% acetonitrile (ACN) containing 0.1% formic acid (FA) and used for direct infusion experiments (500 fmol/ $\mu$ L) on a Q-TOF impact II (Compass 1.9, otofControl 4.0) interfaced with an electrospray ionization (ESI) Apollo source (both Bruker, Bremen, Germany). Data was acquired using a modified version of the standard Instant Expertise method in which the product-ion spectra rate was 4–16 Hz (depending upon precursor intensity) and the number of precursors selected for fragmentation is adjusted automatically to retain a MS-tandem MS duty cycle of 3 s. For this experiment, precursors were manually selected but the tandem MS spectra rate was automatically determined as above. MS spectra were acquired at 2 Hz and precursors were isolated with a width of 3–5 Da depending on  $m/z$  values. Collision energies were increased from 10 to 140 eV in steps of 10 eV; for each collision energy data was acquired for 2 min in the range 150–2300  $m/z$ .

### Further Optimization of Collision Energies on the Basis of LC-Separated Tryptic Glycopeptides

Correlations between collision energies and  $m/z$  values of glycopeptides were investigated by mixing synthetic peptides with glycopeptides derived from colon biopsies of patients with

ulcerative colitis. Samples were taken with informed patient consent, ethical approval no. 39/2001. Detailed information on sample preparation can be found in the Supplementary Information.

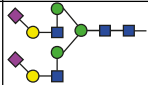
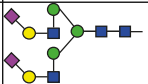

Tryptic and synthetic peptide (500 fmol) mixtures, dissolved in 0.1% FA, were trapped on a C18 pre-column (Acclaim PepMap RSLC Nano-Trap column; 3  $\mu$ m, 100  $\text{\AA}$ , 75  $\mu$ m  $\times$  20 mm, Thermo Fisher Scientific, Waltham, MA, USA) and separated on a C18 analytical column (Acclaim PepMap RSLC column; 2  $\mu$ m, 100  $\text{\AA}$ , 75  $\mu$ m  $\times$  150 mm, Thermo Fisher Scientific) using a linear gradient from 2% buffer B (100% ACN, 0.1% FA) to 50% in 30 min, with buffer A containing 0.1% FA. The flow rate was set to 300 nL/min and oven temperature to 40  $^{\circ}$ C. The impact II was interfaced with the CaptiveSpray nanoBooster source (Bruker). MS and tandem MS data were acquired at 1 Hz from 50–3000  $m/z$  with precursor selection in the range of 650–2000  $m/z$ . In each MS cycle, one MS spectrum was followed by product-ion spectra of the three most intense precursors using an isolated width of 5–7.5 Da. The collision energy was set to a different value in each of 13 separate acquisitions, increasing from 10 to 130 eV in steps of 10 eV.

### Application of Optimized Fragmentation Parameters to Standard Glycoproteins

The optimized parameter settings determined for the synthetic glycopeptides were applied in the analysis of tryptic (glyco-) peptides obtained from bovine fetuin and various human Igs. Purified IgA, sIgA, IgD, IgE, IgG1, IgG2, and IgM were separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Supplementary Figure 3) and the respective heavy chain bands in-gel-digested with trypsin, as described previously [25]. Fetuin was digested in solution as described elsewhere [26].

**Setup One** Tryptic digests (1:7 dilution, dissolved in 2% ACN, 1  $\mu$ L) were analyzed using the experimental setup described above with some modifications. The LC gradient was

**Table 1.** Synthetic N-Glycopeptides Used for Optimizing Q-TOF Fragmentation Conditions

Name	Sequence	Glycan	[M + H] <sup>+</sup>	[M + 2H] <sup>2+</sup>	[M + 3H] <sup>3+</sup>	[M + 4H] <sup>4+</sup>
GP-N	ENYSV <b>F</b> VHPK		3424.3830	1712.6952	1142.1325	856.8512
GP-M	EVFVHPN <b>Y</b> SK		3424.3830	1712.6952	1142.1325	856.8512
GP-C	EVFVHPY <b>S</b> NK		3424.3830	1712.6952	1142.1325	856.8512

■ N-Acetylglucosamine, ● Mannose, ● Galactose, ◆ N-Acetylneuraminic acid



set from 5% to 50% solvent B in 60 min. Data was acquired using another modified version of the standard Instant Expertise method, which had a MS-tandem MS duty cycle of 3.5 s. MS was fixed to 2 Hz and the product-ion spectra rate was variable in the range 1.5–4 Hz depending upon precursor intensity. Precursors were automatically selected in the range of 650 to 3000  $m/z$  and fragmented with an isolation width of 8–10 Da depending on  $m/z$  values. Collision energies were increased linearly in a  $m/z$  dependent manner from 55 eV at  $m/z$  700 to 124 eV at  $m/z$  1800, for all charge states. These values were applied to 80% of the TOF summations used for each spectrum; for the remaining 20% the collision energy was halved. In this mode of operation, ion counts from each subdivision are summed together.

*Setup Two* Tryptic digests were also analyzed on an additional liquid chromatography-electrospray ionization-quadrupole-time of flight-mass spectrometry (LC-ESI-Q-TOF-MS) system. In detail, LC-QTOF-tandem MS analysis on a nano reverse phase (RP) column was performed on a maXis HD Q-TOF mass spectrometer equipped with a CaptiveSpray nanoBooster source (both Bruker) coupled to a Ultimate 3000 nano ultra-performance liquid chromatography system (Thermo Scientific, Breda, The Netherlands). The mass spectrometer and the LC were controlled by Hystar 3.2 (Bruker).

One microliter of the tryptic digest (1:7 dilution, dissolved in water) was loaded onto a C18  $\mu$ -pre column (PepMap100; 300  $\mu\text{m} \times 5 \text{ mm}$ , 5  $\mu\text{m}$ , 100  $\text{\AA}$ , Thermo Scientific) with 10  $\mu\text{L}/\text{min}$  of 99% water/ 1% ACN/ 0.05% TFA for 5 min. (Glyco-) peptides were separated on a C18 analytical column (Acclaim PepMap RSLC; 75  $\mu\text{m} \times 15 \text{ cm}$ , 2  $\mu\text{m}$ , 100  $\text{\AA}$ , Thermo Scientific, Breda) and elution was performed at a flow rate of 700 nL/min with buffer A [water containing 0.1% FA (v/v)] and buffer B [80% acetonitrile/20% water containing 0.1% FA (v/v)]. A linear gradient of 3%–40% buffer B in 15 min was applied followed by column washing and reconditioning.

The CaptiveSpray nanoBooster was operated with acetonitrile-enriched gas (0.2 bar) and 3 L/min dry gas at 150 °C and a capillary voltage of 1200 V. MS spectra were acquired within a mass range of  $m/z$  50–2800. As before, basic stepping mode was applied for the tandem MS collision energy (100%–50%) each 80% and 20% of the time, respectively, and collision energies were set as a linear curve in a  $m/z$  dependent manner ranging from 55 eV at  $m/z$  700 to 124 eV at  $m/z$  1800, for all charge states. In this setup, product-ion spectra were generated from the three most abundant precursors in a range of  $m/z$  550–2800 with an isolation width of 8–10 Da depending on  $m/z$  values. MS was performed at a spectra rate of 1 Hz, tandem MS at 0.5 to 2 Hz dependent upon precursor intensity.

### *Software Assisted Glycopeptide Data Analysis*

Software-assisted glycopeptide data analysis DataAnalysis 4.4 was used for the generation of peak lists in the .XML format,

which were then further analyzed using the bioinformatics platform ProteinScape 4.0 (Bruker). The classification step, which filters for glycopeptide spectra and determines the respective masses of the peptide and glycan moieties, was performed as described previously based on the presence of oxonium ions and monosaccharide distances [27].

The exact parameter settings used for the classification as well as for the peptide and glycopeptide searches are listed in Supplementary Figure 4. GlycoQuest, the glycan search engine integrated in ProteinScape, was used to search potential corresponding glycan compositions within the glycan structure database GlycomeDB ([www.glycome-db.org](http://www.glycome-db.org)), using the glycan masses determined in the classification step. ProteinScape 4.0 automatically replaced the experimentally determined precursor mass of respective glycopeptides with the calculated peptide mass for subsequent mascot analyses. This enabled the peptide moiety to be identified using an in-house Mascot Server version 2.3 or 2.4 (Matrix Science Ltd., UK). The collision energy optimum for peptides is the maximum of the peptide intensity coverage plotted against the applied collision energies. The peptide intensity coverage itself is a value that is the sum of the intensities of all fragments explained by peptide fragmentation divided by the sum of the intensities of all product-ion spectra signals, and thus provides a good comparative value for evaluating both presence and intensity of peptide specific fragments. The GlycoQuest score, which combines the glycan fragmentation coverage and the glycan intensity coverage, was used for determining the collision energy optimum for glycan fragmentation.

The collision energy optimum for peptides is the maximum of the peptide intensity coverage plotted against the applied collision energies. The peptide intensity coverage itself is a value that is the sum of the intensities of all fragments explained by peptide fragmentation divided by the sum of the intensities of all product-ion spectra signals, and thus provides a good comparative value for evaluating both presence and intensity of peptide specific fragments. The GlycoQuest score was used for determining the collision energy optimum for glycan fragmentation.

## Results and Discussion

### *Optimization of Glycopeptide Fragmentation Using Synthetic N-Glycopeptides*

A panel of defined N-glycopeptides was synthesized to systematically elucidate the optimal fragmentation conditions for glycopeptides in Q-TOF instruments. A maximum on peptide sequence as well as glycan composition data should be simultaneously acquired on the chemically distinctive fractions while minimizing dwelling and acquisition times within the instrument. The optimal settings were elucidated using the peptide intensity coverage score as an indicator for optimal peptide fragmentation conditions, whereas the GlycoQuest score indicated when a maximum of information could be

obtained for the glycan moieties. This was then confirmed by manual validation of the spectra.

The collision energies applied to fragment the glycopeptides were increased in 10 eV steps from 10 to 140 eV. These analyses revealed distinguished fragmentation optima for the glycan and the peptide moieties (Figure 1). Independent from the precursor charge, higher energies were required to obtain optimal dissociation of the peptide moieties. Around 30 eV provided the highest GlycoQuest score for the quadruply charged GP-M, indicating this value to be the optimal collision energy to obtain a maximum of information on the glycan moiety. For the same glycopeptide, around 65 eV were required to achieve the highest peptide intensity coverage. The data showed that the peptide moieties had a somewhat narrower fragmentation energy optimum compared with the glycan part (Figure 1a, b). We also confirmed glycopeptide fragmentation energies to be  $m/z$ -dependent, and lower collision energies were required for compounds with a lower  $m/z$  [19, 28, 29].

Interestingly, although the collision energies required for fragmentation of the peptide moiety for all three synthetic N-glycopeptides overlapped with each other, the collision energy optimum for the peptide moiety of GP-N was lower compared with the other two. In addition, the intensity coverage determined for GP-N was generally reduced over the entire collision energy range of the analyzed charge state (Figure 1c). The specific GP-N amino acid sequence as well as the position of

the N-glycan present are possible explanations for the observed phenomenon, indicating that either the peptide sequence, the N-terminal proximity, or both influence the specific fragmentation behavior of this glycopeptide. In contrast, the glycan-specific collision energy optima were comparable for all three precursors, indicating that the specific amino acid sequence appears to have no or just negligible influence on the fragmentation of the N-glycans (Figure 1d). Exemplary fragment spectra for the doubly charged GP-M recorded under different collision energies (70, 100, 120 eV) are included in the Supplementary Figure 1.

### Further Optimization of Collision Energies Using LC-Separated Tryptic Glycopeptides

The collision energies and instrument settings were further optimized for a larger set of glycopeptides during a representative LC-ESI tandem MS experiment. The synthetic glycopeptide GP-M was mixed into a HILIC-enriched (glyco-)peptide fraction obtained after tryptic in-gel protein digestion of a randomly chosen SDS-PAGE band from human colon biopsy samples (Supplementary Figure 2). Overall, 16 different glycopeptide species ranging over four different charge states derived from four different proteins were successfully identified in addition to the synthetic glycopeptide GP-M (Supplementary Table 1). A range of different collision energies from 10 to 130 eV (10 eV steps) was applied to determine

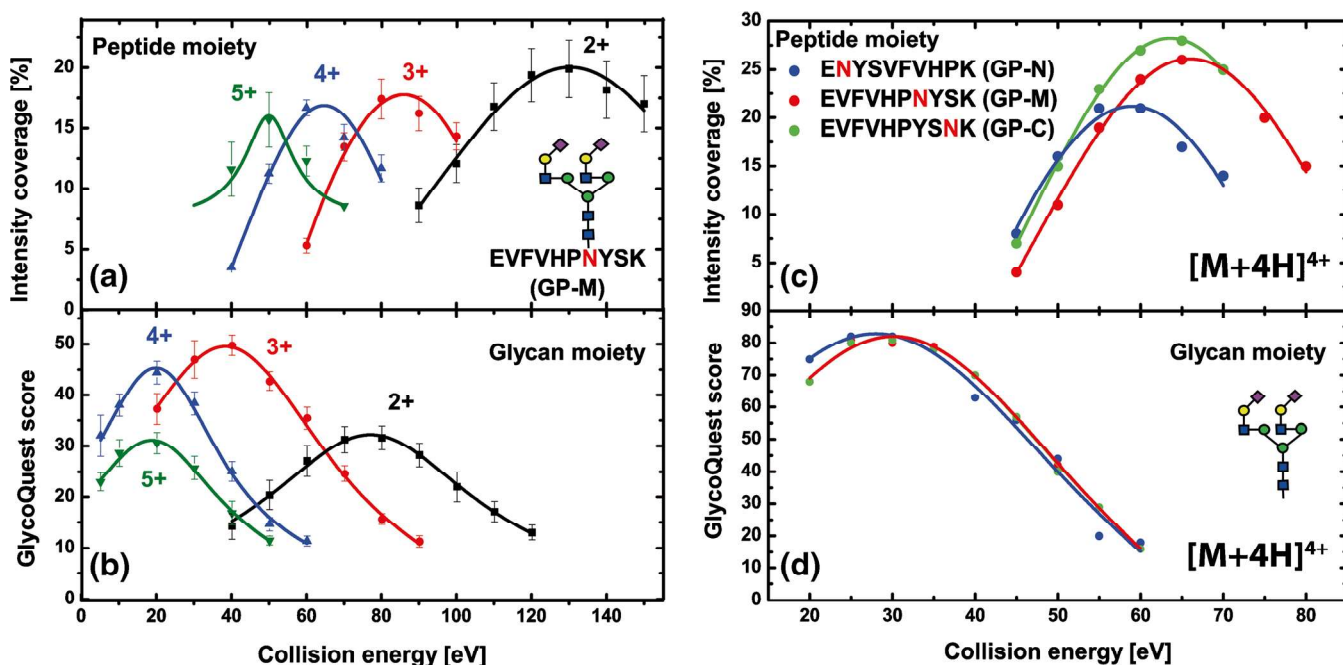
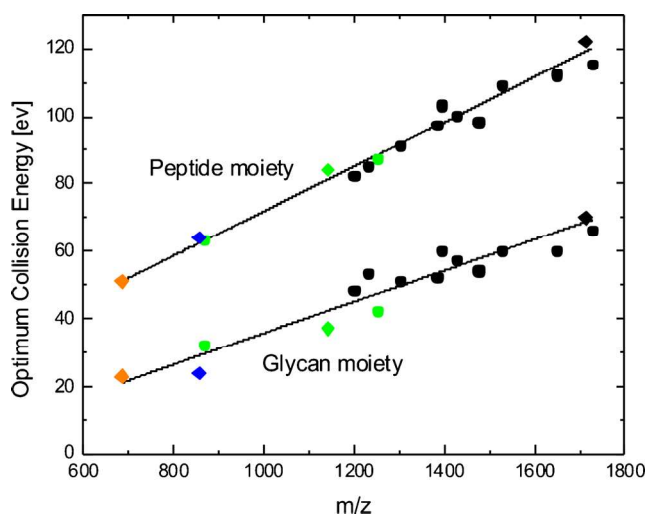


Figure 1. Collision energy optima for synthetic N-glycopeptides. For the peptide part, the intensity coverage was plotted versus the applied collision energy, and for the glycan part, the GlycoQuest score was used. From these plots, the respective optima were determined (refer to text for further details). For GP-M (left), the optimal collision energies were determined for charge states 2+, 3+, 4+, and 5+ (a) peptide part; (b) glycan part). Error bars represent the standard deviation determined from an average of approximately 180 individual product-ion spectra. Right: determination of the optimal collision energies for all three synthetic glycopeptides [charge 4+ (c) peptide part; (d) glycan part]



the optimal information output based on the GlycoQuest Score and peptide intensity coverage. The best balance was obtained when 80% of the fragmentation time was allocated to the higher energy CID conditions for acquiring peptide specific fragments. Lower energy conditions were applied for the remaining 20% fragmentation time to acquire glycan-specific fragments induced by glycosidic bond cleavages. Depending on the individual experimental requirements, more time can be allocated to the lower energy fragmentation conditions if more glyco-related fragments are desired and vice versa.

A linear dependency of the collision energy optimum and the  $m/z$  ratio was observed when the collision energies were stepwise incremented. This further confirmed the general trend that precursors with higher  $m/z$  ratios require higher collision energies for optimal fragmentation (Figure 2). Depending on the particular peptide composition and the distribution of certain amino acids within a peptide sequence, the availability of readily mobile protons may vary, resulting in slightly deviated optimized collision energies. This is best exemplified by our data obtained for glycopeptide GP-N (Figure 1c), but has also been described previously [19, 30]. Nonetheless, the optimized collision energies initially determined on the synthetic N-glycopeptides could be further refined using glycopeptides with various peptide backbones and glycan structures. Based on these results, a collision energy method was employed using 55 eV at  $m/z$  700 to 124 eV at  $m/z$  1800 as high-energy values for 80% of the fragmentation time before the energy values were halved for the remaining 20%.



**Figure 2.** Correlation between precursor  $m/z$  and optimal collision energies. Synthetic glycopeptides (diamonds) were spiked into a mixture of glycopeptides (circles) enriched from a tryptic digest derived from a complex sample and analyzed via C18-RP-LC-ESI-Q-TOF tandem MS. For the  $[M + 5H]^{5+}$  species and partially for the  $[M + 4H]^{4+}$  species of the synthetic glycopeptides, values were obtained additionally by direct infusion. The optimal collision energies for peptide backbone and glycan moiety were determined based on GlycoQuest Score and peptide intensity coverage.  $[M + 5H]^{5+}$  species are indicated in orange,  $[M + 4H]^{4+}$  in blue,  $[M + 3H]^{3+}$  in green, and  $[M + 2H]^{2+}$  in black

### *Method Application: Analysis of Tryptic (Glyco-) Peptides Derived from Various Human Immunoglobulins and Bovine Fetuin*

With the optimized collision energy method at hand, we set out to benchmark the method using tryptic (glyco-)peptides obtained from various human Igs (IgA1+2, sIgA [secretory component, joining chain, IgA1+2], IgD, IgE, IgG1+2, and IgM) as well as bovine fetuin. These glycoproteins contain both N- and O-glycopeptides and are well described (references in Table 2), which makes them prime model samples for a systematic evaluation of the optimized collision energy settings as well as the automated glycopeptide data analysis software embedded in ProteinScape 4.0.

The 10 glycoproteins investigated here have been previously reported to contain 33 N- and 18 O-glycosylation sites (Table 2). In silico digestion with trypsin (not taking missed cleavages into account) predicted 30 N-glycopeptides and five O-glycopeptides resulting from these glycoproteins. The individual tryptic digests were subjected to nanoRP-LC-ESI-Q-TOF tandem MS using the optimized method and without any further glycopeptide enrichment. Analysis of the obtained data with ProteinScape 4.0 revealed peptide sequences and glycan compositions of 21 tryptic N- and two O-glycopeptides (Table 2). The obtained spectra contained both features, B- and Y-ions resulting from the glycan fragmentation of lower-energy CID as well as b- and y-ions of the peptide portion attributable to the enhanced energy (Figure 3). ProteinScape also correctly classified seven additional spectra as glycopeptides, which could, however, not be positively associated with any protein by the software algorithm. Manual inspection of the spectra confirmed the initial glycopeptide classification, but also revealed the reason why they could not be automatically assigned. Most of these glyco-classified but unassigned spectra derived from glycopeptides with multiple sites of glycosylation. This holds true for the tryptic IgA1-peptide His89-Arg126, which can contain up to five O-glycosylation sites (Ser105, Ser111, Ser113, Ser119, Ser121). Additionally, this glycopeptide also shows a prominent fragmentation of the peptide backbone ( $b_{11}, y_{27} + \text{glycan}$ ) with retention of (parts of the) glycan moieties (Figure 4, Table 2), which prohibited the algorithm from determining a correct peptide mass of this highly complex glycopeptide.

IgD is the second glycoprotein in this panel for which up to seven sites of O-glycosylation have been reported earlier (Ser109, Ser110, Thr113, Thr126, Thr127, Thr131, and Thr132) [31, 32]. Similar to the IgA1 hinge-region glycopeptide (Figure 4), the detected tryptic glycopeptide Ala106-Arg137 (containing two missed cleavages) did not exhibit a fragmentation pattern, which resulted in a “peptide-only” or “peptide + HexNAc” fragment; instead a “peptide + Hex<sub>2</sub>HexNAc<sub>3</sub>NeuAc” was the smallest Y-type ion detected (Supplementary Figure 5). Thus, it was impossible for the ProteinScape algorithm to determine the right peptide mass value needed for automated glycopeptide assignment of this multiply glycosylated peptide. It is likely that an automated

**Table 2.** Overview of theoretical and automatically identified (by PS 4.0) human Igs and bovine fetuin glycosylation sites

Theoretical sites/ tryptic peptide	Detected		Peptide sequence (smallest detected or theoretical possible peptide portion indicated)	GF	Detected compositions (Hex,HexNAc,NeuAc,dHex)	Ref.
	PS4	Man				
<b>IgA (serum derived)</b>						
<b>IgA1 (P01876); IGHA1_HUMAN</b>						
N144	X		LSLHRPALEDLLLSEANLTCTLTGLR	2	5,4,1,0 / 5,5,1,0	[40–44]
N340	X		LAGKPTHVNVSVVMAEVDGTCY (unspecific cleavage)	3	5,4,1,1 / 5,4,0,1 / 5,5,1,1	[41, 44]
O105, O111, O113, O119, O121	-	X	HYTNPSQDVTVPCVPSTPPTPSPSTPPTPSPSCCHPR	9	3,4,1,0 / 4,4,1,0 / 3,5,1,0 4,5,1,0 / 4,5,2,0 / 4,4,2,0 3,4,2,0 / 4,4,3, 0 / 5,5,3,0	[37, 44, 45]
<b>IgA2 (P01877); IGHA2_HUMAN</b>						
N47	-	-	VFPLSLDSTPQDGNVVVACLVOGFFPQE PLSVTWSESGQNV TAR	n.d.		[46]
N92	X		HYTNPSQDVTVPCVPPPPPCCHPR	n.d.		[36]
N131	X		LSLHRPALEDLLLSEANLTCTLTGLR	2	5,5,1,0 / 5,4,1,0	[41, 46]
N205	X		TPLTANITK	3	5,4,1,1 / 5,5,1,1 / 5,5,0,1	[40, 41, 46–48]
N327	-	-	MAGKPTHVNVSVVMAEVDGTCY			[41, 46, 48]
<b>sIgA (colostrum derived)</b>						
<b>Polymeric immunoglobulin receptor (P01833); PIGR_HUMAN</b>						
N83, N90	-	X	ANLTNFPENGTFFVNNIAQLSQDDSGR		10,8,1,4 / 10,8,0,6	[24, 37, 40, 46, 48, 49]
N35	-	X	GLSFDVSLEVSQGPGLLNDTK	2	5,4,1,2 / 5,4,0,2	[37, 46, 49]
N186	-	X	QIGLYPVLVIDSSGYVNPNTGR	2	5,4,1,1 / 5,4,1,2	[37, 46, 48, 49]
N421	-	X	LSLLEEPGNFTFVILNQLTISR	4	5,4,1,2 / 5,4,0,2 / 5,4,0,3 5,4,21	[24, 37, 40, 46, 48–50]
N469	X		VPGNVTAVLGETLK	2	5,4,2,0 / 5,4,1,1	[24, 37, 40, 41, 46, 48–50]
N499	X		WNNTGCQALPSQDEGPSK	3	5,4,1,1 / 5,4,0,1 / 5,4,1,0 (more manually detected)	[37, 46, 48, 49]
<b>Joining chain (P01591); IGJ_HUMAN</b>						
N71	X		ENISDPTSPLR	1	5,4,1,0	[24, 37, 40, 46, 48, 50]
<b>IgA1 (P01876); IGHA1_HUMAN</b>						
N144	X		LSLHRPALEDLLLSEANLTCTLTGLR	7	8,2,0,0 / 7,2,0,0 / 6,2,0,0 / 5,2,0,0 / 3,5,0,0 / 5,4,0,0 / 4,5,0,0	see above
N340	X		LAGKPTHVNVSVVMAEVDGTCY	1	6,2,0,0	see above
O105, O111, O113, O119, O121		X	HVKHYTNPSQDVTVPCVPSTPPTPSPSTPPTPSPSCCHPR	6	3,4,1,0 / 4,4,1,0 / 3,4,2,0 / 4,5,1,0 / 4,4,2,0 / 4,4,3, 0 (more manually detected)	see above
<b>IgA2 (P01877); IGHA2_HUMAN</b>						
N47	X		(VFPLSLDSTPQDGNVVVACLVOGFFPQEPL) SVTWSESGQNV TAR (unspecific cleavage)	1	3,5,0,1	see above
N92	X		HYTNPSQDVTVPCVPPPPPCCHPR	n.d.		see above
N131	X		LSLHRPALEDLLLSEANLTCTLTGLR	5	8,2,0,0 / 7,2,0,0 / 6,2,0,0 / 5,2,0,0 / 3,5,0,0	see above
N205	X		TPLTANITK	6	5,5,0,1 / 5,5,1,1 / 4,5,0,1 / 4,5,1,1 / 4,4,1,1 / 3,5,0,1	see above
N327	-	-	MAGKPTHVNVSVVMAEVDGTCY	n.d.		see above
<b>IgG</b>						
<b>IgG1 (P01857); IGHG1_HUMAN</b>						
N180	X		EEQYNSTYR	7		[41, 50, 51]

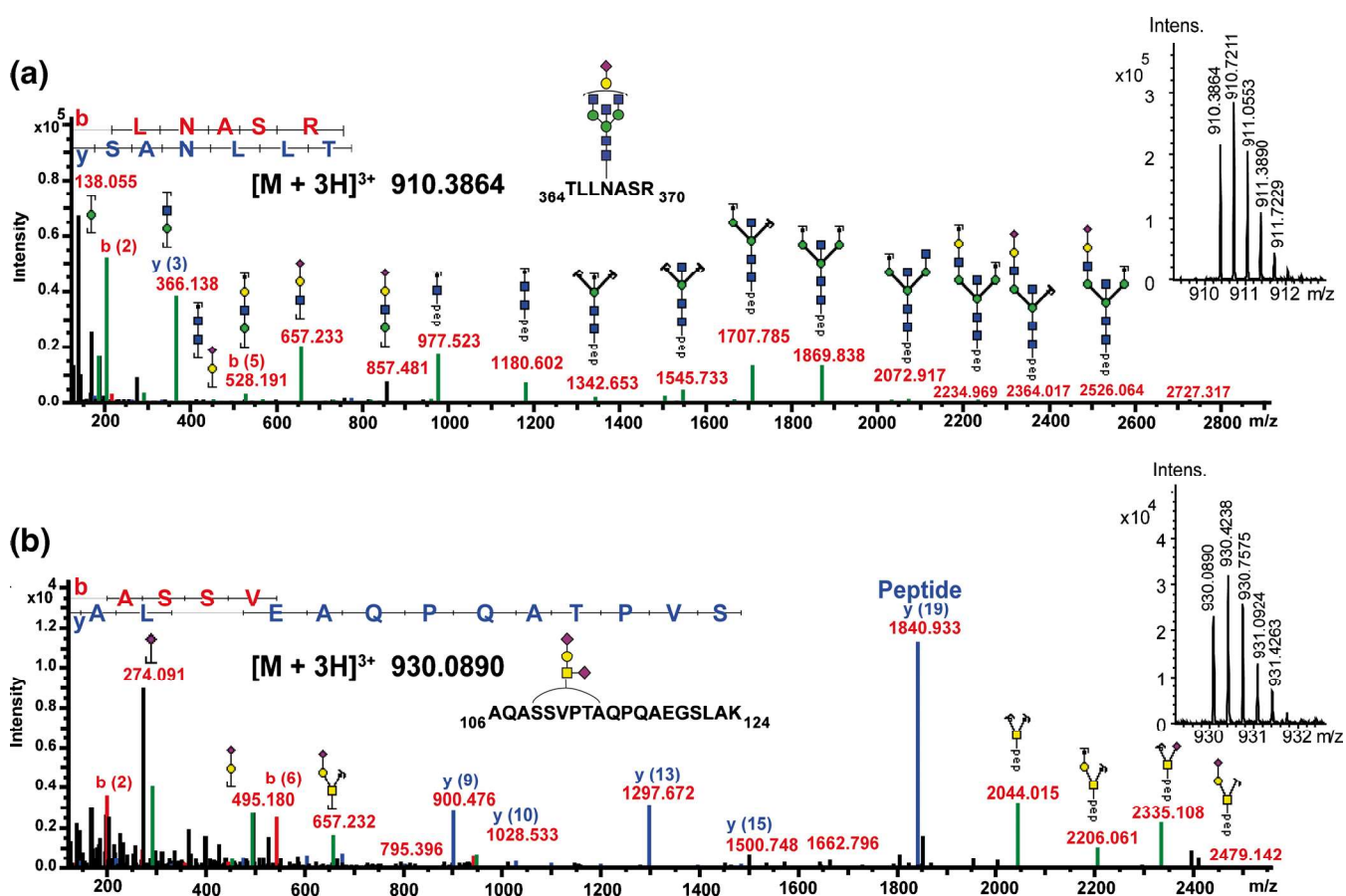
Table 2 (continued)

Theoretical sites/ tryptic peptide	Detected		Peptide sequence (smallest detected or theoretical possible peptide portion indicated)	GF	Detected compositions (Hex,HexNAc,NeuAc,dHex)	Ref.
	PS4	Man				
<b>IgG2 (P01859); IGHG2_HUMAN</b> N176	X		EEQFNSTFR	5	5,4,0,1 / 5,4,1,1 / 4,5,0,1 4,4,0,1 / 3,5,0,1 / 3,4,0,1 4,4,0,0	[40, 41, 50]
<b>IgM (P01871); IGHM_HUMAN</b> N46	X		YKNSDISSTR (1 missed cleavage)	13	6,3,1,1 / 6,3,1,0 / 6,2,0,0 5,5,0,1 / 5,5,1,1 / 5,5,1,0 5,4,0,1 / 5,4,2,1 / 5,4,1,1 5,4,2,0 / 5,4,1,0 / 4,3,0,1 4,3,1,1	[41, 47, 50, 52]
N209	X		GLTFQQNASSMCVPDQDTAIR	4	5,5,0,1 / 5,5,1,1 / 5,4,2,1 5,4,1,1	[41, 52]
N272, N279	-	-	THTNISESHPNATFSAVGEASICEDDWSNGER			[52, 53]
N439	X		STGKPTLYNVSLVMSDTAGTCY	1	6,2,0,0	[52, 53]
<b>IgE (P01854); IGHE_HUMAN</b> N21, N49	-	-	NIPSNATSVTLGCLATGYFPEPVMVTWDTGSLNGTTMTL PATLTLGSHYATISLLTVSGAWAK	n.d.		[31, 35, 54]
N99	X		VAHTPSSTDWVDNK	3	5,5,2,1 / 5,5,1,1 / 5,5,0,1	[31, 35, 54]
N146	-	-	ILQSSCDGGGHFPPTIQLLCLV SGYTPGTINITWLEDGQVMDVDLSTASTTQE GELASTQSELTLSQK	n.d.		[31, 35]
N252	X		GTVNLTWSR	1	5,5,0,0	[31, 35, 54]
N275	X		NGTLTVTSTLPGVTR	8	9,2,0,0 / 8,2,0,0 / 7,2,0,0 6,2,0,0 / 5,2,0,0 / 5,4,0,1 / 5,4,0,0 / 5,3,0,0	[31, 35]
<b>IgD (P01880); IGHD_HUMAN</b> N225	-	X	VPTGGVEEGLLERHSNGSQSHSR (1 missed cleavage)	3	10,2,0,0 / 9,2,0,0 / 8,2,0,0	[24, 31, 55]
N316	X		EVNTSGFAPARPPPQGSTTFWAWSVLR <sup>a</sup>	n.d.	<sup>a</sup>	[24, 31, 50, 55]
N367	X		TLLNASR	5	5,4,2,1 / 5,5,0,1 5,5,2,1 / 5,5,2,0 / 4,5,0,0	[31, 55]
O109, O110, O113	X		AQASSVPTAQQAEGSLAK	1	1,2,2,0	[31, 32]
O109, O110, O113, O126, O127, O131, O132	-	X	AQASSVPTAQQAEGSLAKATTAPATTRNTGR (2 missed cleavages)	3	5,5,6,0 / 5,5,7,0 / 5,5,5,0	[31, 32]
<b>Fetuin (P012763); FETUA_BOVIN</b> N99	X		RPTGEVYDIEIDTLETTCHVLDPTPLANCSVR	1	6,5,3,0	[39, 51, 56, 57]
N156	X		LCPDCPLLAPLNSR	2	6,5,3,0 / 5,4,2,0	[39, 51, 56, 57]
N176	X		VVHAVEVALATFNAESNGSYLQLVEISR	n.d.	<sup>b</sup>	[39, 51, 56, 57]
O271, O280, O282, O296	-	-	VTCTLFQTPVIPQPQPDGAEAEAPSAVPDA AGPTPSAAGPPVASVVVGPSVVAVPLPLHR	n.d.		[39, 58]
O334, O341	X		TPIVGQSPGPPVVR	1	1,1,1,0	[39, 58]

n.d. = not detected; PS4 = glycopeptides identified by ProteinScape 4.0 GlycoQuest and Mascot; Man. = manually identified; Ref. = literature reference, glycoforms (GF) = corresponds to the number of different glycoforms detected and identified by product-ion spectra; red letters indicate potential glycosylation sites reported in literature within the peptide sequence; blue letters indicate missed cleavages by trypsin.

<sup>a</sup>Site N316 has been described to be modified with a carbohydrate previously [55]. A low intensity signal of the unglycosylated tryptic peptide was detected; however, due to the specific sequence and size of a peptide including a missed cleavage at R313 (followed by E314) it is highly possible that any glycosylated forms were not detected by the specific setup used in this study.

<sup>b</sup>This peptide was just detected in its unglycosylated form.



**Figure 3.** Representative spectra of automatically assigned N- and O-glycopeptides from IgD using the optimized Q-TOF collision energy stepping dissociation method. **(a)** tryptic N-glycopeptide TLLNASR (N367,  $[M + 3H]^{3+}$  at  $m/z$  910.386) carrying a Hex<sub>5</sub>HexNAc<sub>5</sub>NeuAc N-glycan. The specific glycopeptide spectrum provided sufficient information to assign it as a monosialylated diantennary N-glycan with a bisecting GlcNAc. **(b)** Tryptic O-glycopeptide AQASSVPTAQPQAEGSLAK (Ser109, Ser110, Thr113)  $[M + 3H]^{3+}$  at  $m/z$  930.089 with a disialylated core 1 type glycan attached

assignment of such complex product-ion spectra by currently available software tools will deliver similar results, though this was not further evaluated and tested in the course of this study. The remaining five glycosylation sites that were classified as glycopeptides but not identified by the software showed low-intensity peptide b- and y-ions and thus could not be assigned automatically.

The intrinsic nature of the residual, as yet undetected, glycopeptides from this sample set was most likely responsible for why they evaded identification or were not detected at all. The predicted tryptic glycopeptides from IgE (Asn17-Lys80 [N-glycan at N21 and N49]) and Ile116-Lys183 (N-glycan at N146), from bovine fetuin (Val246-Arg306 [O-glycans at Ser271, Thr280, Ser282, and Ser296]) and from IgA2 (Val8-Arg51 [N-glycan at N47]) are all large glycopeptides with a length of more than 40 amino acids. Under the applied analysis conditions these compounds are possibly too hydrophobic and therefore likely to be retained irreversibly on the C18 stationary phase.

In addition, multiply glycosylated peptides such as IgE Asn17-Lys80, IgM Thr269-Arg300, secretory component Ala82-Arg107, and fetuin Val246-Arg306 will also carry

various charges and exhibit a higher glyco-heterogeneity compared with glycopeptides with a single site of glycosylation. These factors will additionally contribute to expected overall lower signal intensities, which also makes it less likely that these signals will be selected for tandem MS experiments under data-dependent selection criteria. As demonstrated earlier, the use of alternative and/or multiple proteases resulting in smaller, less heterogeneous sets of distinguished glycopeptides represents one opportunity to cover these sites [25, 33–35].

In the course of this study, no glycopeptides or peptides covering N327 from IgA2 could be identified, which might be explained by incomplete site occupation or lower ionization efficiency of the glycopeptide. Potential alterations on the C-terminus of the IgA2 used in this study could also result in different peptide backbones that evaded detection/identification. In contradiction to other studies claiming N92 in IgA2 to be quantitatively N-glycosylated despite the presence of a Proline within the N-glycosylation sequon (<sup>89</sup>HYTNPSQDVTVPVPPPPPCCHPR<sup>113</sup>) [36], we were not able to detect the respective site being N-glycosylated at all. Nevertheless, the non-glycosylated tryptic peptide could clearly be detected and identified (Supplementary Figure 6),

which is also in line with previous in-depth glycoproteomic data on secretory IgA [37]. Even manual inspection for traces of glycosylation on this site did not, at least in the sample set analyzed, result in any detectable form of glycosylation on N92.

The optimized collision energy stepping CID conditions reported here were found to deliver solid glycan composition as well as peptide sequence information for precursor ions distributed over a wide  $m/z$  range. Compared with alternative

fragmentation techniques such as ETD, we found collision energy stepping CID to be more robust in delivering useful data within a data-dependent LC-ESI tandem MS experiment, in particular if precursors with high  $m/z$  values were selected. Nevertheless, collision energy stepping CID also efficiently fragmented the linkage between the oligosaccharide and the peptide, making it practically impossible to determine sites of glycosylation by this approach. Though this is of less significance for N-glycosylated peptides because of the well

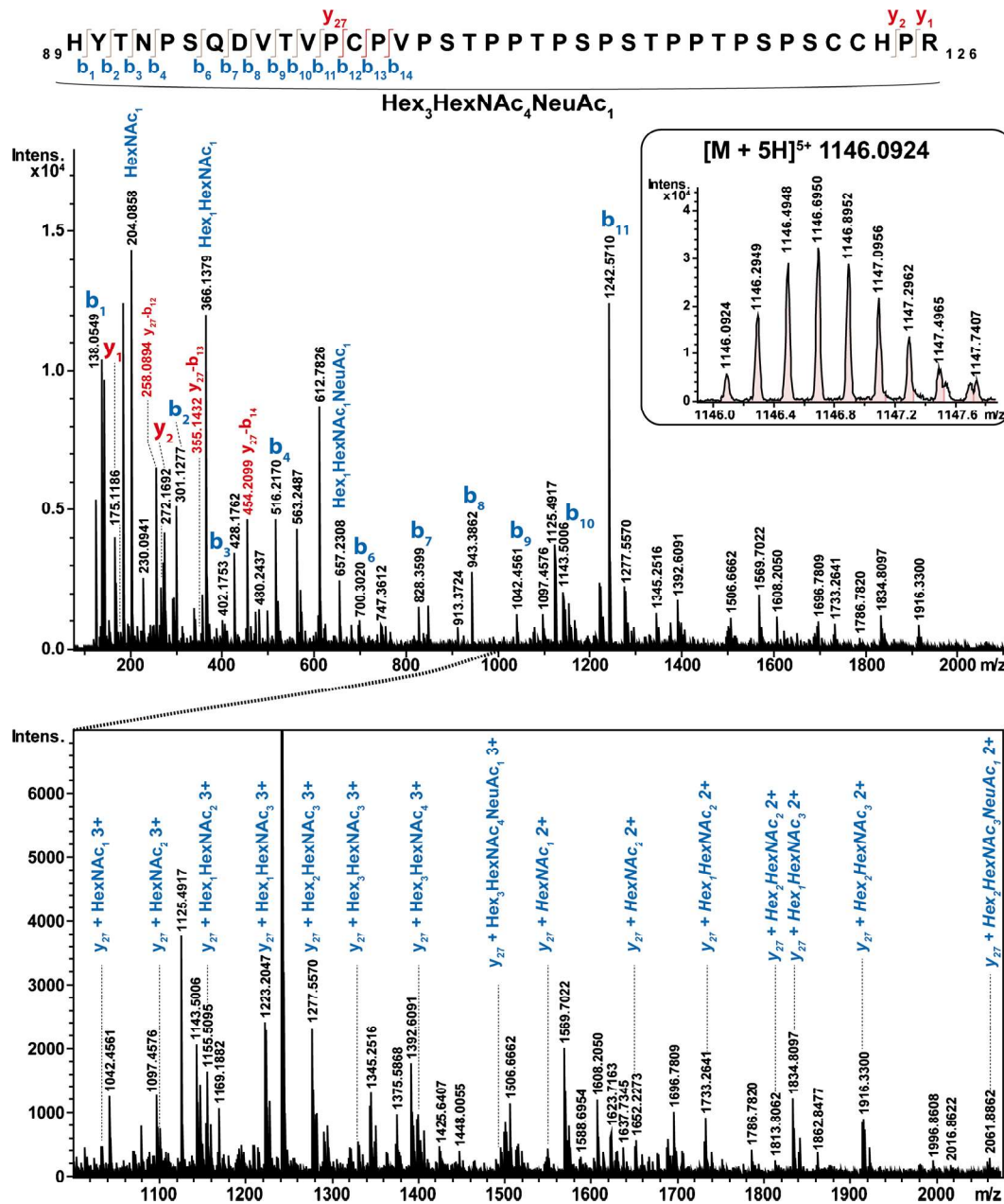


Figure 4. Product-ion spectra of a manually identified IgA1 O-glycopeptide using the optimized Q-TOF collision energy stepping dissociation method. The glycopeptide  $[M + 5H]^{5+}$  1146.0924 with 89His-Arg126 has several occupied O-glycosylation sites with a total composition of Hex<sub>3</sub>HexNAc<sub>4</sub>NeuAc. The glycopeptide could not be identified automatically because the peptide backbone also fragmented into  $b_{11}$  and  $y_{27}$  + glycan fragments. Diagnostic b-ion fragments of the  $y_{27}$  initial fragment are highlighted in red. Note the specific, proline-rich nature of the peptide sequence, resulting in fragments with larger sequence gaps

preserved glycosylation sequon, it poses certain drawbacks for site determination of post-translational modifications where the specific sites cannot be reliably predicted from the protein sequence (e.g., O-glycans or phosphorylation). However, collision energy stepping CID still provides a valuable first data set on peptide sequence and glycan composition even for multiply O-glycosylated peptides (Figure 4 and Supplementary Figure 5). Based on such results, alternative and targeted approaches can subsequently be applied where both sample preparation and acquisition parameters can be optimized to gain site attachment information using ETD [38, 39].

The glycoforms reported in this study on the respective sites of glycosylation represent only glycoforms that were selected and confirmed by product-ion spectra (Table 2); thus, the data does not represent the full variety of glycosylation present on specific sites. Several additional glycoforms were obvious from the MS1 data but are not reported since identification on MS-level and overall site-specific glycosylation heterogeneity were not the scope of this study. It should be noted that all identified glycopeptides were in line with previously published literature (see references in Table 2) and with currently accepted, repeatedly verified general knowledge on protein N-glycosylation.

## Conclusions

Overall, the optimization of CID energy parameters for Q-TOF instruments using synthetic N-glycopeptides yielded improved fragmentation for both the glycan and the peptide moiety. To obtain optimal fragmentation data, peptide moieties required generally higher energies. Twice the energy (compared with the glycan portion) was applied for 80% of the fragmentation time to obtain sufficient peptide bond cleavages. Under the conditions applied in this study, the most informative glycopeptide spectra were obtained when collision energies of 55 eV at 700  $m/z$  to 124 eV at 1800  $m/z$  were applied in a  $m/z$  dependent manner.

The applied fragmentation conditions are well suited to a broad range of glycopeptides. This is exemplified by data showing that the positioning of the glycan moiety on the peptide and/or slight sequence differences only lead to a small change in optimum collision energy for the peptide moiety. Furthermore, the peptide portion did not influence the optimal fragmentation conditions for the glycan part of the glycopeptides, at least for the ones analyzed in this study.

Glycopeptide fragmentation aspects on Q-TOF instruments has been the subject of investigation previously [19–21]. One of the earliest reports by Jebanathirajah et al. described glycan and peptide specific fragmentation aspects using synthetic N- and O-glycopeptides carrying a disaccharide [20]. The non-tryptic nature of these compounds, however, is likely to affect the fragmentation compared with tryptic glycopeptides. Recently, Kolli and Dodds reported a systematic in-depth investigation of enriched, tryptic glycopeptides from bovine ribonuclease B and *Erythrina cristagalli* lectin [19]. In this excellent study that evaluated glycan as well as peptide fragmentation aspects, both investigated glycopeptides carried

neutral N-glycans with a maximum of seven monosaccharide residues. Other reports mainly focused on the fragmentation characteristics of the glycan part only, without further dissociation of the peptide backbone [21]. The data presented here is, to the best of our knowledge, the first to systematically optimize fragmentation conditions of tryptic glycopeptides carrying complex biantennary, disialylated N-glycans and validate these optimized fragmentation parameters by performing a glycoproteomic investigation of the entire panel of human immunoglobulins (IgA, sIgA, IgD, IgE, IgG1, IgG2, and IgM) and bovine fetuin. Using the optimized MS parameters, most of the expected glycopeptides reported in literature could be identified. Peptide sequence as well as glycan composition data were obtained on a representative range of glycopeptides carrying N- as well as O-glycans.

Our results also highlight how bioinformatic tools strongly benefit from data acquired under optimized acquisition parameters. Software-assisted data analysis (ProteinScape 4.0) not only successfully identified N- and O-glycopeptides but also elucidated several glycoforms and made putative structure suggestions. The optimized methodology for glycopeptide analyses presented here using Q-TOF instruments and the associated computational tools will prove to be particularly useful in the fields of glycoproteomics research as well as biopharmaceutical development and quality control applications. The major challenges for glycopeptide characterization in complex biological sample matrices remain their lower ionization efficiency [12] and efficient and specific enrichment.

## Acknowledgments

Open access funding provided by Max Planck Society (or associated institution if applicable). The authors acknowledge support by the Max Planck Society and European Union (Seventh Framework Programme “Glycoproteomics” project, grant number PCIG09-GA-2011-293847, IBD-BIOM project, grant number 305479, and HighGlycan project, grant number 278535). The authors thank Dr. Daniel Baumgart and Dr. Peter Steinhagen for biopsy sample collection, and Anja Wiechmann for technical assistance.

## Open Access

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Lis, H., Sharon, N.: Protein glycosylation. Structural and functional aspects. *Eur. J. Biochem.* **218**, 1–27 (1993)
2. Deeb, S.J., Cox, J., Schmidt-Supprian, M., Mann, M.: N-linked glycosylation enrichment for in-depth cell surface proteomics of diffuse large B-cell lymphoma subtypes. *Mol. Cell. Proteom.* **13**, 240–251 (2014)

3. Wollscheid, B., Bausch-Fluck, D., Henderson, C., O'Brien, R., Bibel, M., Schiess, R.: Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat Biotech.* **27**, 378–386 (2009)
4. Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R.: *Essentials of Glycobiology*, 2nd edn. Cold Spring Harbor Laboratory Press, New York (2009)
5. Ohtsubo, K., Marth, J.D.: Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855–867 (2006)
6. Christiansen, M.N., Chik, J., Lee, L., Anugraham, M., Abrahams, J.L., Packer, N.H.: Cell surface protein glycosylation in cancer. *Proteomics* **14**, 525–546 (2014)
7. Sumer-Bayraktar, Z., Kolarich, D., Campbell, M.P., Ali, S., Packer, N.H., Thaysen-Andersen, M.: N-glycans modulate the function of human corticosteroid-binding globulin. *Mol. Cell. Proteom.* **10**, M111 009100 (2011)
8. Liu, Y.C., Yen, H.Y., Chen, C.Y., Chen, C.H., Cheng, P.F., Juan, Y.H.: Sialylation and fucosylation of epidermal growth factor receptor suppress its dimerization and activation in lung cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11332–11337 (2011)
9. Zauner, G., Selman, M.H., Bondt, A., Rombouts, Y., Blank, D., Deelder, A.M.: Glycoproteomic analysis of antibodies. *Mol. Cell. Proteom.* **12**, 856–865 (2013)
10. Huhn, C., Selman, M.H., Ruhaak, L.R., Deelder, A.M., Wührer, M.: IgG glycosylation analysis. *Proteomics* **9**, 882–913 (2009)
11. Shade, K.-T.C., Platzter, B., Washburn, N., Mani, V., Bartsch, Y.C., Conroy, M.: A single glycan on IgE is indispensable for initiation of anaphylaxis. *J. Experim. Med.* **212**, 457–467 (2015)
12. Stavenhagen, K., Hinneburg, H., Thaysen-Andersen, M., Hartmann, L., Varon Silva, D., Fuchser, J.: Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *J. Mass Spectrom.* **48**, 627–639 (2013)
13. Alley Jr., W.R., Mann, B.F., Novotny, M.V.: High-sensitivity analytical approaches for the structural characterization of glycoproteins. *Chem. Rev.* **113**, 2668–2732 (2013)
14. Wührer, M., Catalina, M.I., Deelder, A.M., Hokke, C.H.: Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **849**, 115–128 (2007)
15. Huddleston, M.J., Bean, M.F., Carr, S.A.: Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. *Anal. Chem.* **65**, 877–884 (1993)
16. Parker, B.L., Palmisano, G., Edwards, A.V., White, M.Y., Engholm-Keller, K., Lee, A.: Quantitative N-linked glycoproteomics of myocardial ischemia and reperfusion injury reveals early remodeling in the extracellular environment. *Mol. Cell. Proteom.* **10**, M110 006833 (2011)
17. Alley Jr., W.R., Mechref, Y., Novotny, M.V.: Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun. Mass Spectrom.* **23**, 161–170 (2009)
18. Dodds, E.D.: Gas-phase dissociation of glycosylated peptide ions. *Mass Spectrom. Rev.* **31**, 666–682 (2012)
19. Kolli, V., Dodds, E.D.: Energy-resolved collision-induced dissociation pathways of model N-linked glycopeptides: implications for capturing glycan connectivity and peptide sequence in a single experiment. *Analyst (Cambridge, U K)* **139**, 2144–2153 (2014)
20. Jebanathirajah, J., Steen, H., Roepstorff, P.: Using optimized collision energies and high resolution, high accuracy fragment ion selection to improve glycopeptide detection by precursor ion scanning. *J. Am. Soc. Mass Spectrom.* **14**, 777–784 (2003)
21. Vekey, K., Ozohanic, O., Toth, E., Jeko, A., Revesz, A., Krenyacz, J.: Fragmentation characteristics of glycopeptides. *Int. J. Mass Spectrom.* **345**, 71–79 (2013)
22. Dallas, D.C., Martin, W.F., Hua, S., German, J.B.: Automated glycopeptide analysis—review of current state and future directions. *Briefings Bioinformatic* **14**, 361–374 (2013)
23. Woodin, C.L., Maxon, M., Desaire, H.: Software for automated interpretation of mass spectrometry data from glycans and glycopeptides. *Analyst (Cambridge, U K)* **138**, 2793–2803 (2013)
24. Liu, T., Qian, W.J., Gritsenko, M.A., Camp, D.G., Monroe, M.E., Moore, R.J.: Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J. Proteome Res.* **4**, 2070–2080 (2005)
25. Kolarich, D., Jensen, P.H., Altmann, F., Packer, N.H.: Determination of site-specific glycan heterogeneity on glycoproteins. *Nat. Protoc.* **7**, 1285–1298 (2012)
26. Kolarich, D., Windwarder, M., Alagesan, K., Altmann, F.: Isomer-specific analysis of released N-glycans by LC-ESI MS/MS with porous graphitized carbon. *Methods Mol. Biol.* **1321**, 427–435 (2015)
27. Hufnagel, P., Resemann, A., Jabs, W., Marx, K., Schweiger-Hufnagel, U.: Automated detection and identification of N- and O-glycopeptides. *Discovering the subtleties of sugars*, June 10th–14th, 2013, Potsdam, Germany (2013).
28. Paizs, B., Suhai, S.: Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **24**, 508–548 (2005)
29. Cao, L., Tolic, N., Qu, Y., Meng, D., Zhao, R., Zhang, Q.: Characterization of intact N- and O-linked glycopeptides using higher energy collisional dissociation. *Anal. Biochem.* **452**, 96–102 (2014)
30. Godugu, B., Neta, P., Simon-Manso, Y., Stein, S.E.: Effect of N-terminal glutamic acid and glutamine on fragmentation of peptide ions. *J. Am. Soc. Mass Spectrom.* **21**, 1169–1176 (2010)
31. Arnold, J.N., Radcliffe, C.M., Wormald, M.R., Royle, L., Harvey, D.J., Crispin, M.: The glycosylation of human serum IgD and IgE and the accessibility of identified oligomannose structures for interaction with mannan-binding lectin. *J. Immunol.* **173**, 6831–6840 (2004)
32. Takayasu, T., Suzuki, S., Kametani, F., Takahashi, N., Shinoda, T., Okuyama, T.: Amino acid sequence of galactosamine-containing glycopeptides in the hinge region of a human immunoglobulin D. *Biochem. Biophys. Res. Commun.* **105**, 1066–1071 (1982)
33. Gauci, S., Helbig, A.O., Slijper, M., Krijgsveld, J., Heck, A.J., Mohammed, S.: Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. *Anal. Chem.* **81**, 4493–4501 (2009)
34. Choudhary, G., Wu, S.L., Shieh, P., Hancock, W.S.: Multiple enzymatic digestion for enhanced sequence coverage of proteins in complex proteomic mixtures using capillary LC with ion trap MS/MS. *J. Proteome Res.* **2**, 59–67 (2003)
35. Plomp, R., Hensbergen, P.J., Rombouts, Y., Zauner, G., Dragan, I., Koeleman, C.A.: Site-specific N-glycosylation analysis of human immunoglobulin E. *J. Proteome Res.* **13**, 536–546 (2014)
36. Huang, J., Guerrero, A., Parker, E., Strum, J.S., Smilowitz, J.T., German, J.B.: Site-specific glycosylation of secretory immunoglobulin A from human colostrum. *J. Proteome Res.* **14**, 1335 (2015)
37. Deshpande, N., Jensen, P.H., Packer, N.H., Kolarich, D.: GlycoSpectrumScan: fishing glycopeptides from MS spectra of protease digests of human colostrum sIgA. *J. Proteome Res.* **9**, 1063–1075 (2010)
38. Windwarder, M., Yelland, T., Djordjevic, S., Altmann, F.: Detailed characterization of the O-linked glycosylation of the neuropilin-1 c/MAM-domain. *Glycoconj. J.* **13**, 1106–1120 (2015)
39. Windwarder, M., Altmann, F.: Site-specific analysis of the O-glycosylation of bovine fetuin by electron-transfer dissociation mass spectrometry. *J. Proteom.* **108**, 258–268 (2014)
40. Kristiansen, T.Z., Bunkenborg, J., Gronborg, M., Molina, H., Thuluvath, P.J., Argani, P.: A proteomic analysis of human bile. *Mol. Cell. Proteom.* **3**, 715–728 (2004)
41. Jia, W., Lu, Z., Fu, Y., Wang, H.P., Wang, L.H., Chi, H.: A strategy for precise and large scale identification of core fucosylated glycoproteins. *Mol. Cell. Proteom.* **8**, 913–923 (2009)
42. Nilsson, J., Ruetschi, U., Halim, A., Hesse, C., Carlsohn, E., Brinkmalm, G.: Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat. Methods* **6**, 809–811 (2009)
43. Halim, A., Nilsson, J., Ruetschi, U., Hesse, C., Larson, G.: Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. *Mol. Cell. Proteom.* **11**, M111 013649 (2012)
44. Mattu, T.S., Pleass, R.J., Willis, A.C., Kilian, M., Wormald, M.R., Lellouch, A.C.: The glycosylation and structure of human serum IgA1, Fab, and Fc regions and the role of N-glycosylation on Fc-alpha receptor interactions. *J. Biol. Chem.* **273**, 2260–2272 (1998)
45. Wada, Y., Dell, A., Haslam, S.M., Tissot, B., Canis, K., Azadi, P.: Comparison of methods for profiling O-glycosylation: Human Proteome Organisation Human Disease Glycomics/Proteome Initiative multi-institutional study of IgA1. *Mol. Cell. Proteom.* **9**, 719–727 (2010)
46. Picariello, G., Ferranti, P., Mamone, G., Roepstorff, P., Addeo, F.: Identification of N-linked glycoproteins in human milk by hydrophilic interaction liquid chromatography and mass spectrometry. *Proteomics* **8**, 3833–3847 (2008)

47. Bunkenborg, J., Pilch, B.J., Podtelejnikov, A.V., Wisniewski, J.R.: Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics* **4**, 454–465 (2004)
48. Ramachandran, P., Boonthung, P., Xie, Y.M., Sondej, M., Wong, D.T., Loo, J.A.: Identification of N-linked glycoproteins in human saliva by glycoprotein capture and mass spectrometry. *J. Proteome Res.* **5**, 1493–1503 (2006)
49. Eiffert, H., Quentin, E., Decker, J., Hillemeir, S., Hufschmidt, M., Klingmuller, D.: The primary structure of human free secretory component and the arrangement of disulfide bonds. *Hoppe Seylers Z Physiol. Chem.* **365**, 1489–1495 (1984)
50. Chen, R., Jiang, X., Sun, D., Han, G., Wang, F., Ye, M.: Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry. *J. Proteome Res.* **8**, 651–661 (2009)
51. Thaysen-Andersen, M., Mysling, S., Hojrup, P.: Site-specific glycoprofiling of N-linked glycopeptides using MALDI-TOF MS: strong correlation between signal strength and glycoform quantities. *Anal. Chem.* **81**, 3933–3943 (2009)
52. Arnold, J.N., Wormald, M.R., Suter, D.M., Radcliffe, C.M., Harvey, D.J., Dwek, R.A.: Human serum IgM glycosylation: identification of glycoforms that can bind to mannan-binding lectin. *J. Biol. Chem.* **280**, 29080–29087 (2005)
53. Putnam, F.W., Florent, G., Paul, C., Shinoda, T., Shimizu, A.: Complete amino acid sequence of the Mu heavy chain of a human IgM immunoglobulin. *Science* **182**, 287–291 (1973)
54. Flanagan, J.G., Rabbitts, T.H.: The sequence of a human immunoglobulin epsilon heavy chain constant region gene, and evidence for three non-allelic genes. *EMBO J.* **1**, 655–660 (1982)
55. Takayasu, T., Takahashi, N., Shinoda, T.: Amino acid sequence and location of the three glycopeptides in the Fc region of human immunoglobulin D. *Biochem. Biophys. Res. Commun.* **97**, 635–641 (1980)
56. Yet, M.G., Chin, C.C., Wold, F.: The covalent structure of individual N-linked glycopeptides from ovomucoid and asialofetuin. *J. Biol. Chem.* **263**, 111–117 (1988)
57. Hagglund, P., Bunkenborg, J., Elortza, F., Jensen, O.N., Roepstorff, P.: A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J. Proteome Res.* **3**, 556–566 (2004)
58. Palmisano, G., Larsen, M.R., Packer, N.H., Thaysen-Andersen, M.: Structural analysis of glycoprotein sialylation – Part II: LC-MS based detection. *RSC Adv* **3**, 22706–22726 (2013)



## **Supplementary Material to:**

# **The art of destruction: Optimizing collision energies in quadrupole-time of flight (Q-TOF) instruments for glycopeptide based glycoproteomics**

Hannes Hinneburg<sup>1,2</sup>, Kathrin Stavenhagen<sup>3</sup>, Ulrike Schweiger-Hufnagel<sup>4</sup>, Stuart Pengelley<sup>4</sup>, Wolfgang Jabs<sup>4</sup>, Peter H. Seeberger<sup>1,2</sup>, Daniel Varón Silva<sup>1</sup>, Manfred Wuhrer<sup>3,5</sup>, Daniel Kolarich<sup>1\*</sup>

1 Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, 14424 Potsdam, Germany

2 Institute of Chemistry and Biochemistry, Freie Universität Berlin, Arnimallee 22, 14195 Berlin, Germany

3 Division of BioAnalytical Chemistry, VU University Amsterdam, Amsterdam, The Netherlands

4 Bruker Daltonik GmbH, Bremen, Germany

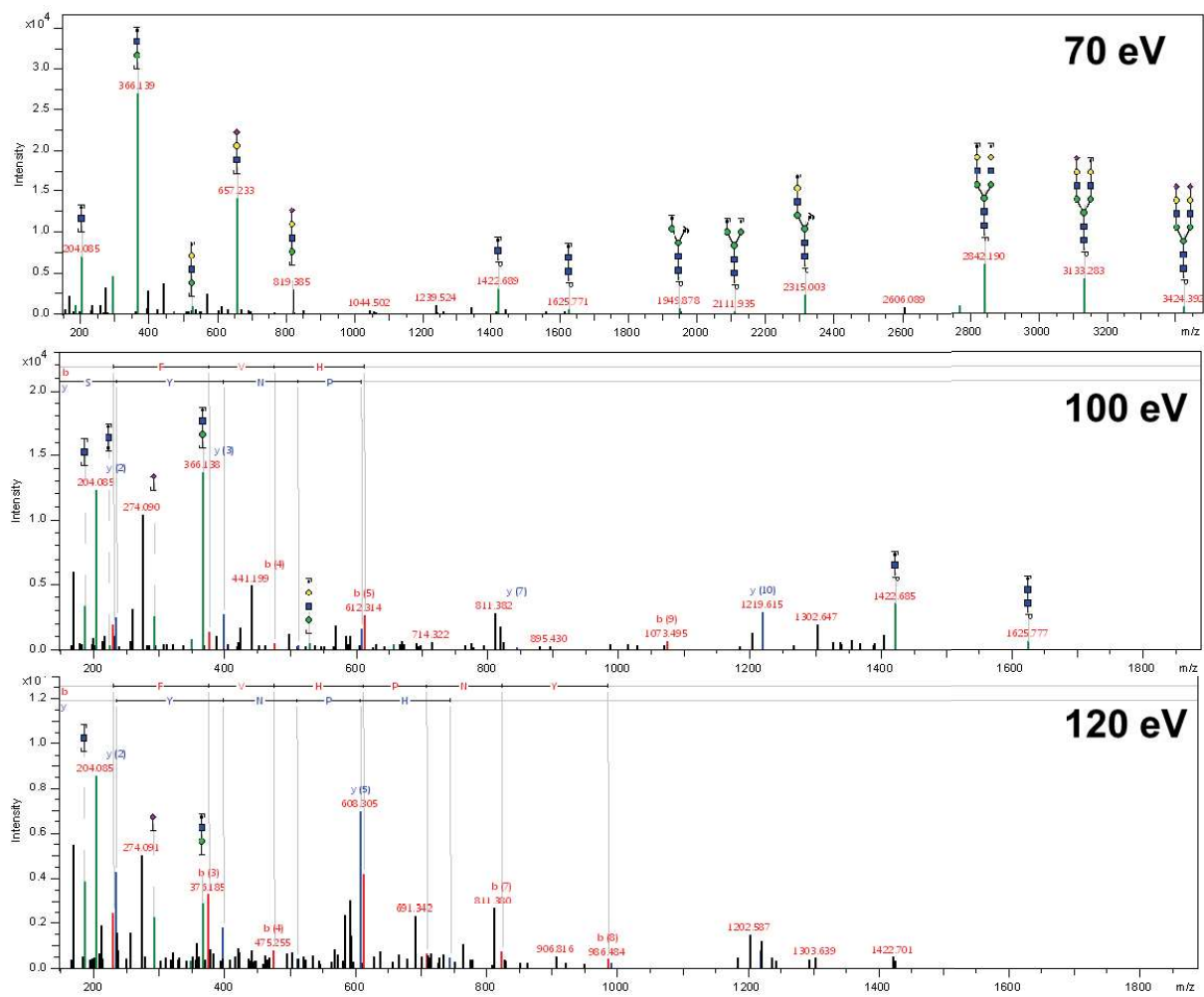
5 Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands

\* Address reprint requests to: Dr. Daniel Kolarich

Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, 14424 Potsdam, Germany

Telephone: +49-30-838 59306; Fax: +49-30-838 459306

Email: [daniel.kolarich@mpikg.mpg.de](mailto:daniel.kolarich@mpikg.mpg.de)



**Supplementary Figure 1: Fragment spectra of GP-M under different collision energies.** Fragment spectra obtained for GP-M ( $[M + 2H]^{2+}$ ) for collision energies of 70 eV, 100 eV and 120 eV.

### ***Human colon preparations (SDS PAGE; HILIC)***

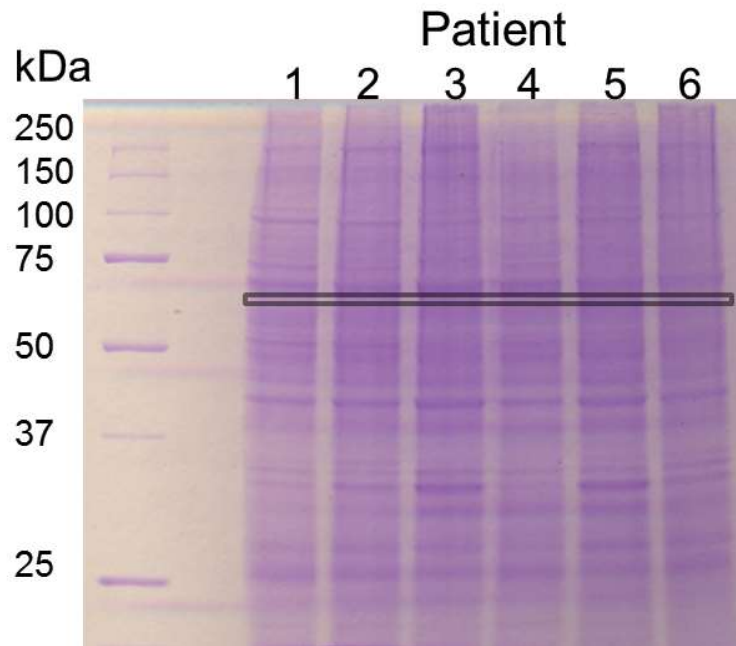
Membrane protein preparation of colon tissue was performed similar as described by Lee et al. [1]. In detail, the tissue pieces (~ 0.25 cm<sup>2</sup>) were mixed with 2 mL of lysis buffer (pH 7.4) containing 50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA and protease inhibitor cocktail (Roche Diagnostics GmbH, Mannheim, Germany, 1 tablet per 50 mL) and incubated for 1 h on ice. After incubation another 3 mL of dilution buffer (20 mM Tris-HCl, 0.1 M NaCl) were added and the tissue was homogenized on ice using a Polytron for 15 seconds (7 cycles). The homogenate was centrifuged at 2,000 g for 20 min at 4°C to remove any debris and nuclei. The supernatant containing the proteins was further processed by ultracentrifugation at 120,000 g for 90 min at 4°C to sediment the proteins. The pellet was dried in a vacuum concentrator and stored in the freezer for further use. 20 µg of protein (pellet) were dissolved in 40% glycerin, 0.25 mM Tris-HCl pH 6.8, 0.015 % bromophenol blue (w/v), 4% SDS (w/v), 50 mM dithiothreitol and incubated at 96°C for 5 min. Then iodoacetamide was added to a final concentration of 50 mM and samples were incubated at room temperature for 30 min in darkness. Samples were then used for SDS-PAGE.

For electrophoresis runs a consort EV265 power supply (Consort bvba, Turnhout, Belgium) and a Mini-PROTEAN® Tetra Cell (Biorad, Munich, Germany) as well as self-cast polyacrylamide gels (0.75 mm thick) and Laemmli [2] system was used. Stacking gels (pH 6.8) consisted of 5% polyacrylamide, separation gels (pH 8.8) of 10%. The acrylamide/bis ratio in both gels was 29:1. The running buffer contained 25 mM Tris, 192 mM glycine, 0.1% SDS (w/v), pH 8.3. Electrophoresis was carried out with 160 V for 75 min. As marker Precision Plus Protein™ Dual Color Standard (Biorad, Munich, Germany) was used.

Gel (Supplementary Figure 2) was stained for 5 min with 50% (v/v) aqueous methanol containing 0.25% (w/v) Coomassie brilliant blue R-250 (Serva Electrophoresis GmbH, Heidelberg, Germany); and 7% (v/v) acetic acid. Excess dye was removed by destaining with 50% (v/v) methanol containing 10% (v/v) acetic acid for at least 3 h. Bands (box Supplementary Figure 2) were excised and in-gel trypsin digested [3]. The extracted peptides were pooled (six patients), dried in a vacuum concentrator and stored at -20°C until further use.

Extracted (glyco-)peptides were solved in 50 µL 80% ACN, 1% TFA and used for HILIC enrichment with 8 mg PolyHYDROXYETHYL A, (12-µm, 100-Å) (PolyLC Inc., Columbia; USA) packed onto an Empore™ C8 membrane plug (3M Deutschland GmbH, Neuss, Germany) in a 10 µL Eppendorf tip (Eppendorf AG, Hamburg, Germany) using a 5417 R Eppendorf centrifuge (Eppendorf AG, Hamburg, Germany). Precondition was with 3x 50 µL methanol and 50 µL

water followed by equilibration with 3x 50  $\mu$ L 80% ACN, 1%TFA. After loading, column was washed 4x 50  $\mu$ L 80% ACN, 1%TFA before bound compounds were eluted with 50  $\mu$ L 0.1% TFA, 50  $\mu$ L 25 mM ammonium bicarbonate and 50  $\mu$ L 50% ACN. Fractions were combined, dried and dissolved in water or 2% ACN for MS analysis.



**Supplementary Figure 2: SDS-PAGE separated human colon samples from patients with ulcerative colitis. Box indicates bands which were used for HILIC enrichment.**

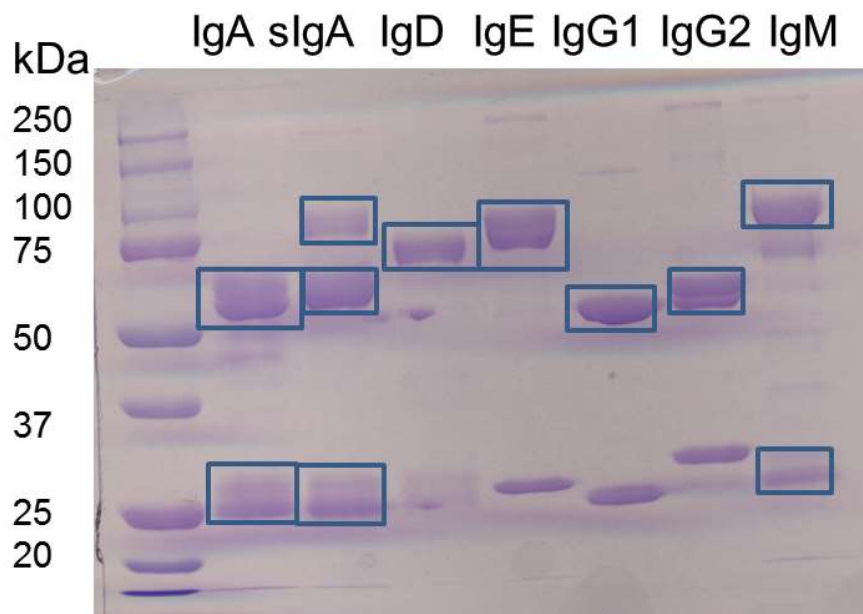
**Supplementary Table 1: Glycopeptides analyses from human colon samples for evaluation of optimized collision energies obtained in experiments with synthetic glycopeptides.**

<i>m/z</i>	<i>z</i>	[M+H] <sup>+</sup>	Peptide sequence	ProteinID	Composition (Hex, HexNAc, NeuAc, dHex)
685.682	5	3424.383	EVFVHPNYSK	PROC_HUMAN	5/4/2/0
856.851	4	3424.383	EVFVHPNYSK	PROC_HUMAN	5/4/2/0
869.395	3	2606.170	K.TPLTANITK.S	IGHA2_HUMAN	3/5/0/1
1142.133	3	3424.383	EVFVHPNYSK	PROC_HUMAN	5/4/2/0
1200.849	2	3600.533	R.TILVDNNTWNNTHISR.V	STT3A_HUMAN	8/2/0/0
1254.867	3	3762.586	R.TILVDNNTWNNTHISR.V	STT3A_HUMAN	9/2/0/0
1230.560	2	2460.112	K.TPLTANITK.S	IGHA2_HUMAN	3/5/0/0
1303.589	2	2606.170	K.TPLTANITK.S	IGHA2_HUMAN	3/5/0/1
1384.615	2	2768.223	K.TPLTANITK.S	IGHA2_HUMAN	4/5/0/1
1395.621	2	2790.234	R.NMSNQLGLLAVNQR.F	PERM_HUMAN	5/2/0/0
1428.623	2	2856.239	K.TPLTANITK.S	IGHA2_HUMAN	4/4/1/1
1476.647	2	2952.287	R.NMSNQLGLLAVNQR.F	PERM_HUMAN	6/2/0/0
1530.163	2	3059.318	K.TPLTANITK.S	IGHA2_HUMAN	4/5/1/1
1648.648	2	3296.288	K.EVFVHPNYS.K	PROC_HUMAN	5/4/2/0
1712.695	2	3424.38	EVFVHPNYSK	PROC_HUMAN	5/4/2/0
1728.173	2	3455.338	LGNWSAMPSCK	APOH_HUMAN	5/4/2/0

Hex: Hexose; NexNAc: N-Acetylhexosamine; NeuAc: N-acetylneuraminic acid; dHex: deoxy-hexose.

### Human Ig preparations

Approximately 5 µg **IgA** (plasma), **sIgA** (human colostrum), **IgD** (plasma), **IgE** (myeloma plasma, lambda), **IgG1** (myeloma plasma, kappa), **IgG2** (myeloma plasma, kappa) and **IgM** (plasma) were dissolved and reduced and alkylated as described above. Electrophoresis run, staining and in-gel tryptic digest was done as described above. Excised bands are indicated in Supplementary Figure 3.



Supplementary Figure 3: SDS-PAGE separated immunoglobulins. Boxes indicate bands cut for tryptic digestion.

### *Software assisted glycopeptide data analysis - Parameter settings*

For glycopeptide identification the two data sets were classified with two peptide mass off sets (H<sup>+</sup> and HexNAc + H<sup>+</sup>) before the glycan and protein search was done and results were combined. The parameters for the integrated protein data base searches (mascot used as search engine) as well as the parameters for the glycan searches can be seen in Supplementary Figure 4. Note that the mascot scores for acceptance (protein, peptide) were set to relatively low values which is necessary due to the search engine scoring algorithm which calculates low values if unassigned peaks with relative high intensities (glycan fragments!) occur in the MS/MS peptide spectrum.

Classify compounds as **glycopeptide**

Min. precursor m/z [Da]: 700.0

Signal-based

m/z signals: OxoniumIons CID pos

Tolerance [Da]: 0.01

Min. intensity coverage [%]: 10.0

Distance-based

m/z Distances

Distances

Min. no. of consecutive m/z distances required: 2

m/z Distance Pattern

Distance tolerance [Da]: 0.02

Ignore signals < 0.0 % of main peak

Calculate peptide Mr

Min. peptide mass [Da]: 500.0

Mascot

Database: SwissProt

Taxonomy: Mammalia (mammals)

Enzyme: semiTrypsin

Allow up to 2 missed cleavages

Modifications: Acetyl (N), Acetyl (N-term), Acetyl (Protein N-term), Amidated (C-term), Amidated (Protein C-term)

Carbamidomethyl (C):  Fixed  Variable

Oxidation (M):  Fixed  Variable

Peptide tol. ±: 20.0 ppm #<sup>13</sup>C: 1

MS/MS tol. ±: 0.1 Da

Significance threshold p <: 0.9

Peptide charge: 2+, 3+ and 4+

Instrument type CID: ESI-QUAD-TOF

Adjust FDR [%]: 5.0

Instrument type ETD: ETD-TRAP

Percolator

General Settings

Search engines:  Mascot

Protein Filter

Min. no. of significant peptide IDs: 1

Max. MW (kDa): 700.0

Generate a merged result for batch searches

Mascot

Ions score cut-off: 1.0

Peptide rank cut-off: 10

Min. peptide length: 5

Ions score threshold for significant peptide IDs:  Identity score calculated by search engine  Defined by user: 35.0

Protein Assessment

Accept

if Mascot Score > 13.0

otherwise

if

Peptide Assessment

For each peptide, accept:

top hit compound only

all compounds (SILC quantitation)

top hit for ETD and CID

Accept

if Mascot Score > 1.0

In case of MS/MS spectra matching peptides from more than one identified protein, accept:

peptide assigned to highest scoring protein

highest scoring peptide (rank 1 peptide)

GlycoQuest

Glycan type: N-glycan core basic

Database: GlycoDB

Modifications: Sulfate, N-sulfate, O-methyl, Phosphate

Denaturation: UNO

Ions: H+, K+, Li+

Neutral: Water, Ammonia, Sodium, Potassium, Hydrogen

Charge +: 1 to 6

Charge -: 1 to 6

#<sup>13</sup>C: 1

Fragmentation type CID: CID b.c.y.z

Threshold for glycan list compilation

Score >: 10.0

Fragmentation coverage [%] >: 0.0

Intensity coverage [%] >: 10.0

Taxonomy: All taxonomies

Composition: H 8-12, N 0-5, S 0-4, F 0-1, O

Reducing endl: Free end

H+ up to: 6

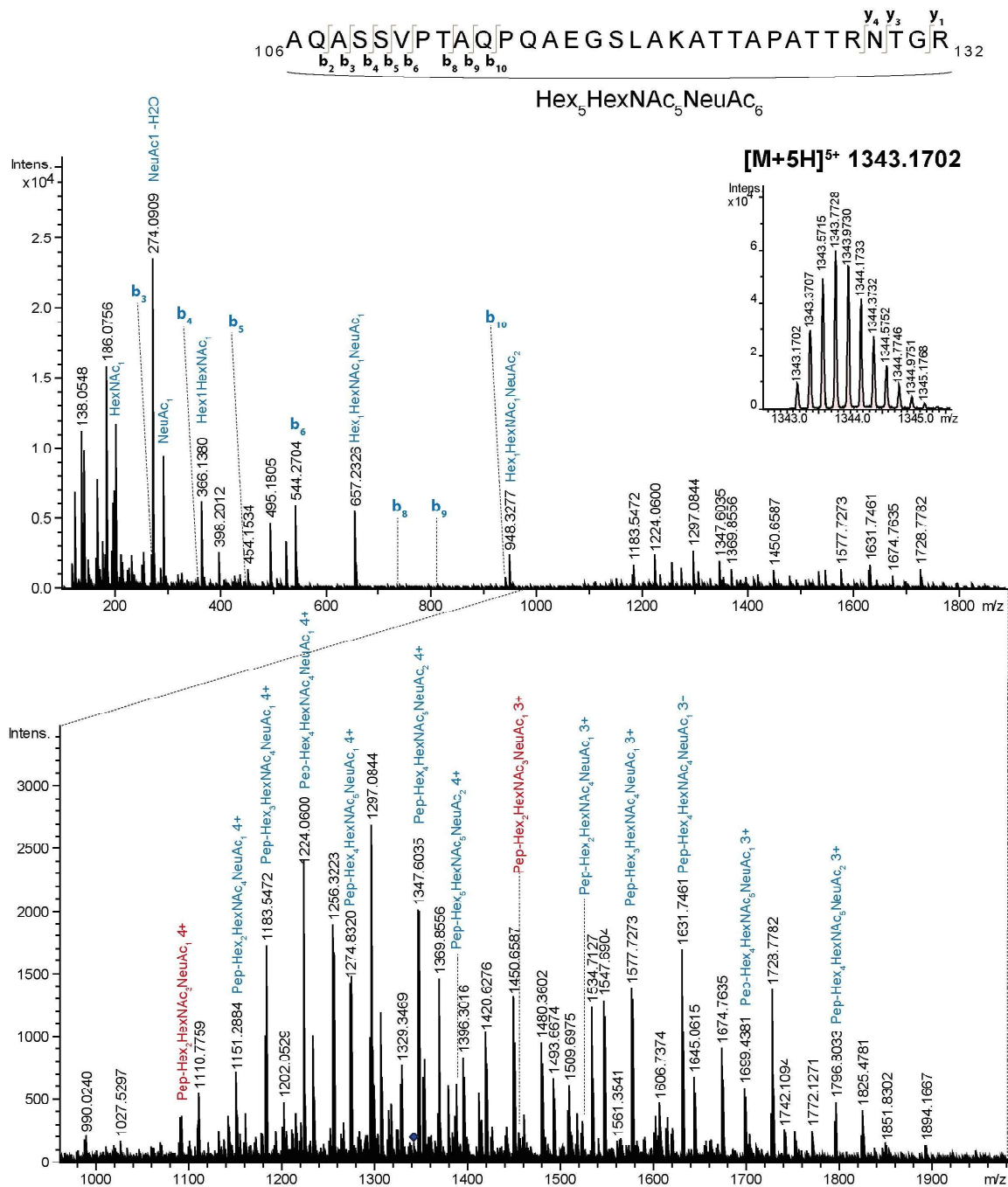
MS tolerance (m/z): 10.0 ppm

MS/MS tolerance (m/z): 0.01 Da

Monoisotopic  Average

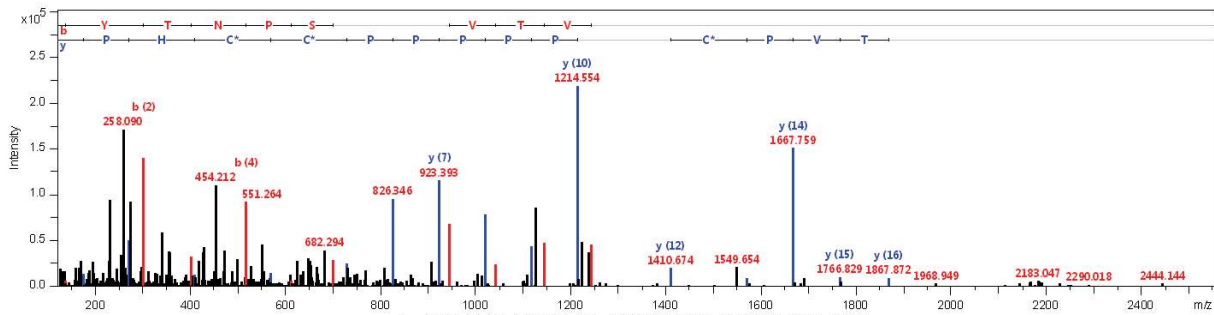
ETD: ETD

**Supplementary Figure 4: ProteinScape search parameters.** Parameter settings for classification (red box), protein identifications (blue box) and glycan searches (black box) using ProteinScape 4.0 for automated glycopeptide recognition.



**Supplementary Figure 5: MS/MS fragmentation spectra of a manually identified IgD O-glycopeptide.** The glycopeptide  $[M + 5H]^{5+}$  1343.1702 with 106Ala-Arg132 has several occupied O-glycosylation sites with a total composition of Hex<sub>5</sub>HexNAc<sub>5</sub>NeuAc<sub>6</sub>. The glycopeptide could not be identified automatically because the fragmentation of the Y-ion series of the glycan part was not complete down to the last attached HexNAc or the peptide backbone. Instead the smallest Y-ion still contains Hex<sub>2</sub>HexNAc<sub>3</sub>NeuAc (indicated in red).





**Supplementary Figure 6: MS/MS fragmentation spectrum of the tryptic peptide 89HYTNPSQDVTVPCVPPPPPCCHPR113 from IgA2.** This peptide was just detected in its unglycosylated form, and no indications for any traces of N- or O-glycosylation on this peptide could be detected. This is in line with current knowledge about the specificity of the oligosaccharyltransferase.

## References

1. Lee, A., Kolarich, D., Haynes, P.A., Jensen, P.H., Baker, M.S., Packer, N.H.: Rat liver membrane glycoproteome: enrichment by phase partitioning and glycoprotein capture. *Journal of proteome research*. **8**, 770-781 (2009)
2. Laemmli, U.K.: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. **227**, 680-685 (1970)
3. Kolarich, D., Jensen, P.H., Altmann, F., Packer, N.H.: Determination of site-specific glycan heterogeneity on glycoproteins. *Nat Protoc*. **7**, 1285-1298 (2012)