

# **New Methods for the Characterization of Human Glycoproteins by LC-MS/MS techniques**

Thesis

Submitted in fulfillment of the requirements of the degree  
Doctor rer. nat. of the Department of Chemistry,  
Faculty of Sciences, University of Hamburg  
Fakultät für Mathematik, Informatik, Naturwissenschaften

by

Melissa Bärenfänger

Hamburg, 2019



This Thesis was prepared at the Institute of Organic Chemistry from October 2014 to March 2019. (managing director: Prof. Dr. Christian Stark)

I would like to thank Prof. Dr. Bernd Meyer for his continuous and motivating support during the work on my PhD thesis.

I would like to thank Prof. Dr. Dr. h. c. mult. Wittko Francke for being the second reviewer of this thesis.

First Reviewer: Prof. Dr. Bernd Meyer

Second Reviewer: Prof. Dr. Dr. h. c. mult. Wittko Francke

Date of defense: 07.02.2020

## Outline

Abbreviations.....	I
<b>1. SUMMARY .....</b>	<b>1</b>
<b>2. ZUSAMMENFASSUNG .....</b>	<b>4</b>
<b>3. INTRODUCTION.....</b>	<b>7</b>
3.1. Structure and Biological Relevance of Protein Glycosylation .....	7
3.2. Clinical Relevance of Glycosylation .....	9
3.3. Structural Analysis of Glycoproteins .....	11
3.3.1. Methods for Isolation of Proteins from Complex Body Fluids .....	12
3.3.2. Chromatographic Methods for the Analysis of Glycoproteins.....	13
3.3.3. Mass Spectrometric Techniques for Glycoprotein Analysis.....	14
3.3.3.1. Bottom-Up Proteomics .....	15
3.3.3.2. Top-Down Proteomics of Intact Glycoproteins.....	17
<b>4. OBJECTIVE.....</b>	<b>19</b>
<b>5. RESULTS AND DISCUSSION.....</b>	<b>21</b>
5.1. Protein Purification from Human Plasma .....	21
5.1.1. Anion Exchange Chromatography .....	21
5.1.2. Size Exclusion Chromatography .....	24
5.1.3. Reversed Phase Chromatography with Online ESI-MS Detection.....	26
5.1.4. Automated Peptide Identification of Proteins .....	29
5.2. Analysis of Human Plasma Ceruloplasmin .....	32
5.2.1. Analysis of Intact Ceruloplasmin .....	33
5.2.2. Analysis of Released Glycans by MS and NMR Techniques .....	36
5.2.3. Quantification of Glycopeptides with Different Peptide Length .....	40
5.2.2.1. Fragmentation of Large Glycopeptides .....	51
5.2.3. Analysis of Genetic Variants of Ceruloplasmin .....	55
5.3. Glycosylation of Human Immunoglobulin IgG1 .....	59
5.3.1. Relative Quantification of Glycopeptides from IgG1 .....	59
5.4. Analysis of Human Serum Transferrin .....	66
5.4.1. Analysis of Intact Transferrin by ESI-MS .....	66

5.5. Highly Glycosylated Alpha-Acid-Glycoprotein .....	73
5.5.1. Analysis of Intact AGP from Pooled Human Plasma.....	74
5.5.1.1. Analysis of Desialylated AGP for a Conclusive Interpretation of Fully Sialylated AGP	76
5.5.1.2. Interpretation of Fully Sialylated AGP Based on the Information of Desialylated AGP	79
5.5.2. Analysis of Desialylated AGP from Individual Plasma Samples .....	84
5.5.3. Analysis of Free Glycans of Alpha-Acid-Glycoprotein .....	89
5.6. Analysis of Human $\beta$ -2-Glycoprotein 1 .....	91
5.6.1. Genetic Variants of $\beta$ -2-Glycoprotein 1 .....	92
5.6.2. Glycosylation of $\beta$ -2-Glycoprotein 1 .....	98
5.7. Glycosylation of Plasma Proteins as Potential Biomarker.....	106
5.7.1. Unusual Glycosylation Patterns of Plasma Sample P .....	107
5.7.2. Changes in Glycosylation in Patients with Liver Cirrhosis and Hepatocellular Carcinoma .....	108
5.7.3. Conclusion and Future Perspective.....	118

## **6. EXPERIMENTAL PROCEDURES..... 119**

6.1. Protein Purification from Human Plasma.....	119
6.1.1. Anion Exchange Chromatography.....	121
6.1.2. Size Exclusion Chromatography .....	122
6.1.3. Purification of IgG using Protein A Affinity Columns .....	123
6.1.4. Determination of Protein Concentration .....	123
6.2. Polyacrylamide Gel Electrophoresis.....	123
6.3. MALDI MS of Ceruloplasmin .....	124
6.4. LC-MS Analysis of Intact Proteins .....	124
6.5. ESI-MS of Intact AGP by Direct Infusion .....	125
6.6. Enzymatic Digest of Glycoproteins.....	126
6.6.1. Tryptic Digest.....	126
6.6.2. Digest with Endoproteinase Arg-C .....	126
6.6.3. Limited Tryptic Digest .....	127
6.6.4. Neuraminidase Digest of Human AGP .....	127
6.6.5. PNGase Digest for the Analysis of Free Glycans.....	127
6.6.6. Tryptic In-Gel Digest .....	128
6.7. LC-ESI-MS/MS of Glycopeptides .....	129
6.8. LC-ESI-MS of PNGase Released Glycans.....	130
6.9. Online LC-ESI-MS and Offline $^1\text{H-NMR}$ Analysis of Free Glycans .....	131
6.10. Data Interpretation .....	132
6.10.1. Pre-Processing of MS Data using DataAnalysis 4.2 .....	132
6.10.2. Automated Peptide Identification using MASCOT MS/MS Ion Search .....	133
6.10.3. Calculation of Simulated Spectra .....	134

6.10.3.1. Calculation of a Simulated Spectrum for the Analysis of Fully Sialylated AGP .....	135
6.10.4. Integration of Extracted Ion Chromatograms .....	135
6.10.5. Interpretation of NMR Data and Correlation with MS Data .....	136
<b>7. HAZARDS.....</b>	<b>137</b>
<b>8. PUBLICATIONS .....</b>	<b>139</b>
<b>9. REFERENCES.....</b>	<b>140</b>
<b>10. APPENDIX.....</b>	<b>154</b>
10.1. Results for Mascot search after Tryptic Digest of Purified Ceruloplasmin.....	154
10.2. NMR Data of Free N-Glycans of Ceruloplasmin .....	157
10.3. Matlab Script for the Relative Quantification of Different Proteoforms .....	159
10.4. Relative Amounts of Glycopeptides from Ceruloplasmin.....	160
10.5. Alpha-Acid-Glycoprotein .....	164
10.6. $\beta$ -2-Glycoprotein 1 .....	175
<b>11. ACKNOWLEDGEMENTS .....</b>	<b>181</b>
<b>AFFIDAVIT .....</b>	<b>182</b>





## Abbreviations

AFP	Alpha-fetoprotein
AGP	Alpha-acid-glycoprotein
AGP-1	Alpha-acid-glycoprotein 1
AGP-2	Alpha-acid-glycoprotein 2
Arg-C	Endoproteinase Arg-C, Clostripain
B2GP1	$\beta$ -2-Glycoprotein 1
CDT	Carbohydrate Deficient Transferrin
CDG	Congenital Disorder of Glycosylation
CFG	Consortium for Functional Glycomics
CID	Collision Induced Dissociation
DEAE	Diethylaminoethyl
DHAP	2,5-Dihydroxybenzoic Acid
ECD	Electron Capture Dissociation
EDC	Extracted Delta Chromatogram
EIC	Extracted Ion Chromatogram
EPO	Erythropoietin
ESI	Electrospray Ionization
ETD	Electron Transfer Dissociation
FT-ICR	Fourier-transform Ion Cyclotron Resonance
HBV	Hepatitis B Virus
HCC	Hepatocellular Carcinoma
HILIC	Hydrophilic Interaction Liquid Chromatography
HPLC	High Performance Liquid Chromatography
IgG	Immunoglobulin G
ISCID	In-source Collision Induced Dissociation
LacNAc	N-Acetyl-D-Lactosamine
LC	Liquid Chromatography
LP	Linear Positive

## ABBREVIATIONS


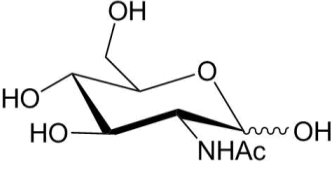

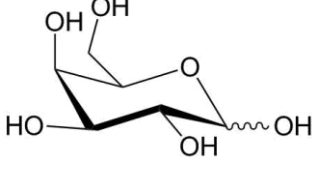

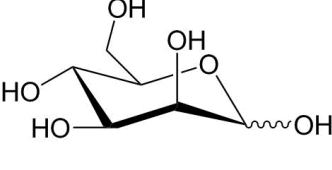

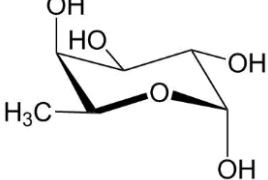

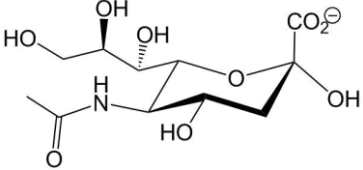
---

mAb	Monoclonal Antibody
MALDI	Matrix Assisted Laser Desorption Ionization
MS	Mass Spectrometry
n.d.	Not Detected
NESP	Novel Erythropoiesis Stimulation Protein
NMR	Nuclear Magnetic Resonance
NP	Normal Phased
PDB	Protein Data Bank
PGC	Porous Graphitized Carbon
pI	Isoelectric Point
PSA	Prostate Specific Antigen
PTM	Post-translational Modification
RP	Reversed Phased
R <sub>t</sub>	Retention Time
SDS	Sodium Dodecyl Sulfate
SPE	Solid Phase Extraction
TIC	Total Ion Chromatogram
TOF	Time of Flight

**Amino acids and their abbreviation**

Amino acid	Three letter code	Single letter code
alanine	Ala	A
arginine	Arg	R
asparagine	Asn	N
aspartic acid	Asp	D
cysteine	Cys	C
glutamine	Gln	Q
glutamic acid	Glu	E
glycine	Gly	G
histidine	His	H
isoleucine	Ile	I
leucine	Leu	L
lysine	Lys	K
methionine	Met	M
phenylalanine	Phe	F
proline	Pro	P
serine	Ser	S
threonine	Thr	T
tryptophan	Trp	W
tyrosine	Tyr	Y
valine	Val	V

**Monosaccharide units and their abbreviations**

Glycan	Abbreviation	CFG symbol	Structure
D-N-acetylglucosamine	GlcNAc (N)		
D-galactose	Gal (H)		
D-mannose	Man (H)		
L-fucose	Fuc (F)		
D-N-acetylneuraminic acid	Sia (S)		

## 1. Summary

The glycosylation of proteins has numerous important functions in nature. Next to the stability and solubility of proteins, glycosylation is relevant for cellular recognition processes and immune modulation. Changes in glycosylation reflect the physiological and pathological state of a cell or an organism. However, the complex mechanisms involved in these changes are not fully understood.

A sophisticated structural characterization of glycans is needed to understand their biological relevance, but also for the development and quality control of biopharmaceuticals or for the development of glycan-based biomarkers. Mass spectrometry has become a well-established technique for the structural analysis of protein modifications due to low detection limits and high mass accuracy. In combination with other analytical methods such as liquid chromatography, mass spectrometry has made it possible to analyze complex glycosylation despite their structural complexity and diversity. Nevertheless, the development of new analytical approaches and MS-based techniques as well as their validation is necessary to further improve the analysis of glycans. This applies in particular to the quantification of glycans, the detection of low abundant structures, and the implementation of robust and timely methods. The aim of this work was therefore to evaluate different mass spectrometric methods for the structural characterization of glycoproteins and to analyze the glycosylation of different plasma proteins from several plasma samples.

A highly efficient purification protocol was established to purify several plasma proteins simultaneously from only a few hundred microliters of human plasma. The plasma proteins ceruloplasmin, immunoglobulin IgG1, transferrin, alpha-acid-glycoprotein, and  $\beta$ -2-glycoprotein 1 were isolated with a combination of anion exchange chromatography, size exclusion chromatography, and reversed phased chromatography coupled to online MS detection. This approach allowed the purification of several proteins from minimal sample amount in a timely manner, enabling the analysis of plasma samples from multiple donors.

The glycosylation of the plasma proteins ceruloplasmin and immunoglobulin IgG1 was analyzed by ESI-MS as glycopeptides with different peptide lengths. By this,

the conventional bottom-up approach involving a tryptic digest was compared to a middle-down approach. It was shown that the quantification of glycopeptides is more precise on glycopeptides with enhanced peptide length, presumably due to better ionization properties of longer glycopeptides. For both proteins, immunoglobulin IgG1 and ceruloplasmin, the middle-down approach showed a higher abundance of highly sialylated structures. The use of a middle-down approach for the qualitative and quantitative analysis of glycopeptides even allowed the detection of previously unreported, tetraantennary glycan structures on ceruloplasmin.

Considering these results, alpha-acid-glycoprotein,  $\beta$ -2-glycoprotein 1 and transferrin were analyzed as intact glycoproteins. This allowed not only a more precise quantification of glycosylation but is also a rapid method for the detection of all proteoforms caused by other post-translational modifications or single-nucleotide polymorphism.

The analysis of alpha-acid-glycoprotein is especially challenging as it exhibits five glycosylation sites, presenting dominantly tri- and tetra-antennary glycans. Due to the high number of negatively charged sialic acids, alpha-acid-glycoprotein was analyzed as intact glycoprotein in negative ion mode, revealing the tremendous heterogeneity caused by glycosylation and variation in amino acid sequences. Using different enzymatic methods and bioinformatics tools, an unambiguous peak assignment was achieved successfully. By this, 90 different proteoforms caused by glycosylation were detected for the most abundant variant ORM1\*F1.

To evaluate the potential use of the analyzed glycoproteins as biomarkers, samples from patients diagnosed with liver cirrhosis and hepatocellular carcinoma were analyzed using the established protocols. The plasma proteins alpha-acid-glycoprotein, transferrin and  $\beta$ -2-glycoprotein 1 showed an increase in fucosylated structures; immunoglobulin IgG1 showed a decrease in galactosylated structures, which is in good agreement with previously reported data.

Additionally, the analysis of several glycoproteins from the same plasma samples allowed to compare the changes in glycosylation between different proteins. Alpha-acid-glycoprotein and IgG1 showed a more drastic change in glycosylation than transferrin and  $\beta$ -2-glycoprotein 1, leading to the conclusion that certain glycoproteins are more susceptible to changes in glycosylation than others. The

methods presented in this work allowed a straight-forward and accurate determination of changes in glycosylation and can be used to study the biological implications further on.

## 2. Zusammenfassung

Die Glykosylierung von Proteinen hat in der Natur zahlreiche wichtige Funktionen. Neben der Stabilität und Löslichkeit von Proteinen ist die Glykosylierung für zelluläre Erkennungsprozesse und die Immunmodulation relevant. Veränderungen in der Glykosylierung spiegeln den physiologischen und pathologischen Zustand einer Zelle oder eines Organismus wider. Die komplexen Mechanismen, die an diesen Veränderungen beteiligt sind, sind jedoch nicht vollständig verstanden.

Eine anspruchsvolle strukturelle Charakterisierung von Glykanen ist notwendig, um ihre biologische Relevanz zu verstehen, aber auch für die Entwicklung und Qualitätskontrolle von Biopharmazeutika oder für die Entwicklung von glykanbasierten Biomarkern. Die Massenspektrometrie hat sich aufgrund niedriger Nachweisgrenzen und hoher Massengenauigkeit als etablierte Technik zur Strukturanalyse von Proteinmodifikationen etabliert. In Kombination mit anderen Analysemethoden wie der Flüssigkeitschromatographie hat die Massenspektrometrie es ermöglicht, komplexe Glykosylierungen trotz ihrer strukturellen Komplexität und Vielfalt zu analysieren. Dennoch ist die Entwicklung neuer analytischer Ansätze und MS-basierter Techniken sowie deren Validierung notwendig, um die Analyse von Glykanen weiter zu verbessern. Dies gilt insbesondere für die Quantifizierung von Glykanen, die Detektion von niedrig vorkommenden Strukturen und die Implementierung robuster und schneller Methoden. Ziel dieser Arbeit war es daher, verschiedene massenspektrometrische Methoden zur strukturellen Charakterisierung von Glykoproteinen zu evaluieren und die Glykosylierung verschiedener Plasmaproteine aus mehreren Plasmaproben zu analysieren.

Ein hocheffizientes Reinigungsprotokoll wurde etabliert, um mehrere Plasmaproteine gleichzeitig aus nur wenigen hundert Mikrolitern menschlichem Plasma zu reinigen. Die Plasmaproteine Ceruloplasmin, Immunglobulin IgG1, Transferrin, *Alpha-acid-glycoprotein* und  $\beta$ -2-Glykoprotein 1 wurden mit einer Kombination aus Anionenaustauschchromatographie, Größenausschlusschromatographie und Umkehrphasenchromatographie, gekoppelt mit Online-MS-Detektion, isoliert. Dieser Ansatz ermöglichte eine zeitlich effiziente



Aufreinigung mehrerer Proteine aus minimaler Probenmenge und ermöglichte so die Analyse von Plasmaproben mehrerer Spender.

Die Glykosylierung der Plasmaproteine Ceruloplasmin und Immunglobulin IgG1 wurde mittels ESI-MS anhand von Glykopeptiden mit unterschiedlichen Peptidlängen analysiert. Dabei wurde der konventionelle Bottom-up-Ansatz mit tryptischen Verdau mit einem Middle-Down-Ansatz verglichen. Es wurde gezeigt, dass die Quantifizierung von Glykopeptiden mit erhöhter Peptidlänge genauer ist, vermutlich aufgrund besserer Ionisationseigenschaften längerer Glykopeptide. Für beide Proteine, Immunglobulin IgG1 und Ceruloplasmin, zeigte der Middle-Down-Ansatz einen höheren Anteil von hochsialylierten Strukturen. Die Verwendung eines Middle-Down-Ansatzes für die qualitative und quantitative Analyse von Glykopeptiden ermöglichte dadurch sogar den Nachweis von bisher nicht bekannten, tetraantennären Glykanstrukturen für Ceruloplasmin.

Unter Berücksichtigung dieser Ergebnisse wurden *Alpha-acid-glycoprotein*,  $\beta$ -2-Glykoprotein 1 und Transferrin als intakte Glykoproteine analysiert. Dies ermöglicht nicht nur eine genauere Quantifizierung der Glykosylierung, sondern ist auch eine schnelle Methode zum Nachweis aller Proteoformen, die durch andere posttranslationale Modifikationen oder Single-Nukleotid-Polymorphismus verursacht werden.

Die Analyse von *Alpha-acid-glycoprotein* zeigte sich als besonders anspruchsvoll, da es fünf Glykosylierungsstellen aufweist, die überwiegend tri- und tetraantennäre Glykane präsentieren. Aufgrund der hohen Anzahl an negativ geladenen Sialinsäuren wurde *Alpha-acid-glycoprotein* im Negativ-Ionen-Modus als intaktes Glykoprotein detektiert, was die enorme Heterogenität durch Glykosylierung und Variation der Aminosäuresequenzen aufzeigt. Mit verschiedenen enzymatischen Prozeduren und bioinformatischen Methoden wurde eine eindeutige Peak Zuordnung erfolgreich erreicht. Dabei wurden für die am häufigsten vorkommende Variante ORM1\*F1 90 verschiedene Proteoformen durch Glykosylierung nachgewiesen.

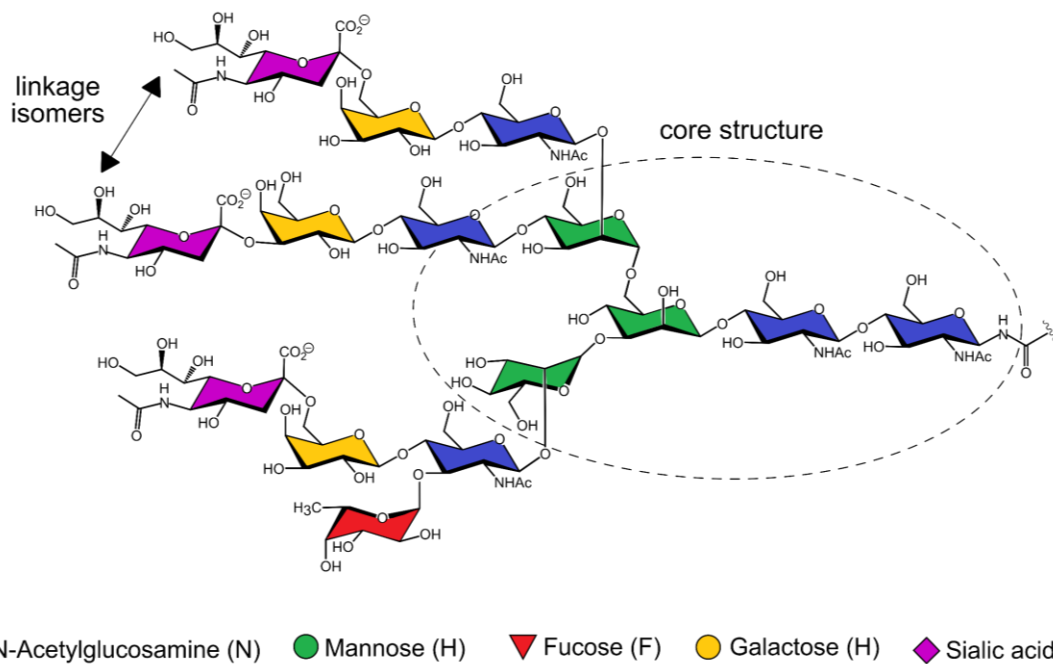
Um den möglichen Einsatz der analysierten Glykoproteine als Biomarker zu bewerten, wurden Proben von Patienten mit Leberzirrhose und hepatozellulärem Karzinom anhand der etablierten Protokolle untersucht. Die Plasmaproteine *Alpha-acid-glycoprotein*, Transferrin und  $\beta$ -2-Glykoprotein 1 zeigten einen Anstieg

der fucosylierten Strukturen; Immunglobulin IgG1 zeigte einen Rückgang der galactosylierten Strukturen, was in guter Übereinstimmung mit der Literatur ist. Zusätzlich ermöglichte die Analyse mehrerer Glykoproteine aus den gleichen Plasmaproben den Vergleich der Veränderungen in der Glykosylierung zwischen verschiedenen Proteinen. *Alpha-acid-glycoprotein* und IgG1 zeigten eine drastischere Veränderung der Glykosylierung als Transferrin und  $\beta$  2 Glykoprotein 1, was zu dem Schluss führt, dass bestimmte Glykoproteine anfälliger für Veränderungen der Glykosylierung sind als andere. Die in dieser Arbeit vorgestellten Methoden ermöglichten eine einfache und genaue Bestimmung der Veränderungen in der Glykosylierung und können verwendet werden, um die biologischen Auswirkungen weiter zu untersuchen.

### 3. Introduction

#### 3.1. Structure and Biological Relevance of Protein Glycosylation

The glycosylation of proteins describes the post-translational modification (PTM) of amino acid side chains with highly diverse carbohydrate moieties. This diversification is an abundant modification and is found in more than 80% of all proteins.[1] The complexity of glycosylation is caused by many different factors, starting with the attachment of glycans to different amino acid side chains. The most common types of glycosylation are N-glycosylation with glycan attachment to asparagine residues and O-glycosylation with glycan attachment to serine or threonine side chains. Other types of glycosylation have been observed as well like C-glycosylation on tryptophan.[2] For N-type glycosylation a consensus sequence, being Asn-X-Ser/Thr, is required. This glycosylation is divided into three subtypes, exhibiting a mutual core structure composed of a pentasaccharide with two N-acetylglucosamines (GlcNAc) and three mannose (Man) residues. The attachment of further mannose residues to this core structure leads to oligo- or high-mannose-type N-glycans. Complex-type N-glycans are further modified with N-acetylglucosamines. The hybrid-type is characterized by the presence of mannose residues as well as other monosaccharide moieties. The most abundant monosaccharides include hexoses (H) like glucose (Glc), mannose (Man) and galactose (Gal) as well as N-acetylhexosamines (N) like N-acetylglucosamines (GlcNAc) and N-acetyl-galactosamine (GalNAc). Pyranoses like fucose (F) and nonoses like sialic acid (S) are also common structures. The severe heterogeneity of glycosylation is therefore also a result of a variety of monosaccharides units forming these oligosaccharide structures. But even glycans with the same monosaccharide composition can be different in linkage, like the attachment of a fucosyl residue to an N-acetylglucosamine, which can either be 1→2, 1→3, 1→4, or 1→6 linked. Next to this difference in connectivity, configurational isomers in form of different anomers are present (Figure 1). Hence, glycosylation of a protein leads to many different proteoforms expressed in biological systems.



**Figure 1.** Schematic structure of an N-type glycan. The proximal N-acetylglucosamine is covalently bound to an asparagine side chain of a peptide by a glycosidic linkage. The mutual core structure (Man<sub>3</sub>GlcNAc<sub>2</sub>) is extended with further monosaccharide units that can differ in linkage position and anomericity. Monosaccharides are represented as a colored symbol according to CFG nomenclature (bottom).[3]

The diverse nature of glycosylation is not only described by its heterogeneity in structure, but also by its manifold biological functions. This includes the alteration of physical properties of proteins like solubility and electrostatic interactions influencing the pI of a protein. Glycosylation is also relevant for folding and oligomerization of proteins and their subunits, which in turn is the basis for their successful biosynthesis.[4-6] Equally important is the contribution to a large number of different recognition processes, ranging from cell-cell or cell-matrix recognition mediated by interactions between lectins and carbohydrates to immunomodulation and signaling processes.[7] Even clearance of circulating glycoproteins is mediated by their glycosylation through the asialoglycoprotein receptor, which recognizes and binds  $\beta$ -linked galactose residues. These endocytic receptors are dominantly expressed on liver tissue, leading to hepatic uptake and subsequent degradation of desialylated proteins.[8] The asialoglycoprotein receptor can also be targeted for the delivery of therapeutic

agents to hepatocytes for the possible treatment of hepatocellular carcinoma or other liver diseases.[9]

Another example for the biological significance of glycosylation is the human ABO blood group system. Red blood cells express glycolipids and glycoproteins which are acting as antigens. These specific glycan epitopes are built by glycosyltransferases, transferring either one N-acetylglucosamine (galactosyltransferase A, leading to blood type A) or one galactose (galactosyltransferase B, leading to blood type B) to the H-antigen composed of the disaccharide  $\alpha\text{Fuc}(1\rightarrow2)\beta\text{Gal}$ . The absence of galactosyltransferases leads to blood type 0, presenting only the H-antigen as glycan epitope. Each blood group also expresses antibodies against the epitopes that are not presented by its red blood cells. This immunogenic function is crucial for transfusion medicine since the transfusion of an incompatible blood type can lead to death. Even though the human ABO blood group system, as well as the structure of the glycan antigens, is well studied its physiological function and its evolutionary significance are mostly unknown.[10, 11]

### **3.2. Clinical Relevance of Glycosylation**

Changes of glycosylation are observed under physiological conditions like inflammation, alcoholism, pregnancy, and a variety of diseases.[12-15] This results in a great interest in glycans for various medical applications. One major interest is the development of therapeutic drugs. Small molecule drugs often target carbohydrate receptors or lectins. Such a lectin is for example the hemagglutinin of the influenza virus, which presents a promising target in the development of anti-influenza drugs by preventing virus entry into the host cell.[16] Other proteins like glycosyltransferases also present attractive possibilities for drug development. In patients with Alzheimer disease, the expression of O-GlcNAc transferase is upregulated and presents a target for the development of inhibitors for the treatment of Alzheimer disease.[17-19]

A different class of therapeutics is recombinant proteins. One well-established drug is erythropoietin (EPO), a highly sialylated glycoprotein that is used to increase the

number of red blood cells in the treatment of anemia. To reduce its clearance and therefore enhance its efficacy in vivo, two additional N-glycosylation sites were introduced, leading to an increase in overall sialylation. This new drug, called novel erythropoiesis stimulation protein (NESP), resulted in an up to 3-fold greater serum half-life and consequently less frequent dosing for the patient.[20, 21]

Glycoengineering and monitoring glycosylation as a critical quality attribute is also an important part for the development of therapeutic monoclonal antibodies (mAbs). This growing class of biopharmaceuticals has emerged over the last three to four decades and is now the majority of recombinant proteins in clinical use.[22] In early 1980s, murine mAbs were introduced but showed a lack of efficacy, rapid clearance, and immunogenic effects in patients. The immune response does not only lead to an ineffective drug but can cause anaphylaxis due to an immediate allergic reaction. Advancements in genetic engineering have led to the development of humanized or fully human antibodies, increasing the efficacy and reducing the immune response. Nevertheless, hypersensitive reactions have also been reported on humanized mAbs like cetuximab, leading in some cases to death.[23, 24] The allergic response to cetuximab is caused by the expression of the antigenic epitope Gal- $\alpha$ 1,3-Gal.[25] Other identified antigenic epitopes are for example N-glycolyneuraminic acid (NeuGc), core-xylose, and  $\alpha$ 1,3-core-fucose.[26] These structures are not expressed in humans and thus provoke an immune response. However, glycosylation does not only cause negative immunogenic effects but can also enhance clinical efficacy. Fucosylated glycans, as well as glycans terminated with mannose or N-acetylglucosamine, show a general decrease in mAb efficacy. Different parameters in the production of mAbs are responsible for the expression of specific glycan structures. A major factor is the choice of the host cell line. Commonly used cell lines are mammalian cell lines, which can be altered to lack glycosyltransferases like fucosyltransferase 8 to express afucosylated glycan structures and enhance efficacy. Glycosylation is also influenced by growing conditions like glucose content, pH and temperature.[27] Even though an homogenous, well-defined glycosylation of antibodies is favored, an unambiguous glycan analysis is necessary for the development of biopharmaceutical mAbs and for batch-to-batch quality control.

The glycosylation of proteins can also be used as a clinical marker for various diseases. These includes rheumatoid arthritis, long-term alcohol abuse, and several cancer diseases.[28-33] The majority of protein biomarkers are based on changes in concentration of certain proteins. However, detecting changes in protein glycosylation can lead to a more accurate and conclusive biomarker. One example for this is the serum prostate-specific antigen (PSA), a 28 kDa glycoprotein that shows an increase in concentration in patients with prostate cancer. In 1986, PSA was approved by the U.S. Food and Drug Administration as a clinical biomarker detecting elevated serum levels. In the following years, prostate cancer incidences doubled in the U.S., mainly because of false positive diagnosed patients. Changes in PSA concentration are also provoked by increasing age, weight, and several medications, which led to high false positive rates of PSA.[34, 35] In order to enhance the specificity of PSA, several studies monitored the glycosylation of PSA and found an altered fucosylation and additional  $\alpha$ -(2,3)-linked sialic acids in patients with prostate cancer.[32, 36] Developing a glycan-based biomarker from these results seems to be a promising approach to enhance specificity and lower false positive rates. Similar findings are observed for alpha-fetoprotein (AFP) in use as a biomarker for hepatocellular carcinoma. Next to changes in concentration, an increase in 1,6-fucosylation on AFP was observed.[37] Lectin assays for the detection of the increased fucosylation resulted in high specificity of nearly 92%. However, sensitivity was low with a sensitivity score of 37%.[38] This implies the need for highly sensitive technologies, detecting low abundant glycoproteins and permitting precise determination of glycosylation. Especially for the application as clinical biomarker, an accurate quantification of glycan structures is a necessity.

### **3.3. Structural Analysis of Glycoproteins**

The structural analysis of glycoproteins represents a challenging task. Their diversity due to post-translational modifications, as well as limited sample amount require highly sophisticated and sensitive methods. Especially advances in analytical instrumentation over the last few decades helped to meet the needs for

a qualitative and quantitative determination of glycan structures from limited sample amount. Mass spectrometry (MS) is one major tool for glycan analysis due to its high sensitivity, robustness, and accuracy. Although primarily used for compositional analysis of glycosylation, it has proven to be applicable for linkage determination and configuration.[39-41] Another possibility for a complete characterization of glycan structures is the use of nuclear magnetic resonance spectroscopy (NMR), which is slightly less sensitive. Nevertheless, it was shown that 15 pmol of a complex type N-glycan is sufficient for an unambiguous characterization of linkage and anomericity.[42] Prior to glycan analysis, isolation of glycoproteins from different sample materials like tissue or body fluids, is sometimes required. Among other things, chromatographic methods can be used for purification of glycoproteins or for separation of released glycans or glycopeptides and present a crucial area of expertise in glycan analysis.

### **3.3.1. Methods for Isolation of Proteins from Complex Body Fluids**

Analyzing glycoproteins or using a certain glycoprotein as a clinical marker often includes the isolation of the desired analyte from a complex matrix like body fluids. Blood plasma, as a rather non-invasive method of sample collection, is often used for clinical applications.[43] Human plasma proteins can originate from numerous tissues and blood cells, reflecting the physiological and pathological state. However, the concentration of plasma proteins can differ in a range of at least 10 orders of magnitude.[44] This makes the isolation and analysis of less abundant plasma protein quite challenging. Precipitation techniques present a good method for an initial enrichment of several proteins, going hand in hand with the inactivation of proteases or possible virus activity.[45] However, isolation of distinct proteins often requires extensive chromatographic methods. Glycoproteins can be separated by different physical characteristics like size or hydrodynamic radius, net charge, hydrophilic or hydrophobic interactions, or affinity.[46] In most cases, more than one method is used in a complementary approach to isolate distinct proteins. Additional steps in the purification result, however, in a decreased concentration of



the target protein and lead to time-consuming procedure. Choosing efficient purification methods is therefore essential for a successful isolation of proteins.[47] Ion exchange chromatography separates analytes by ionic interactions, which is of special interest for the isolation of proteins with either low or high pI values. The stationary phase carries either anionic or cationic charges, interacting with negatively or positively charged amino acid side chains of the protein. Post-translational modifications like negatively charged sialic acids can also contribute to the interactions. The analytes can be eluted by either changing the pH or increasing the ionic strength of the mobile phase by increasing the salt concentration.[48] Gel filtration or size exclusion chromatography separate proteins by their hydrodynamic radius which, among other things, correlates with the size or molecular weight of the protein. Simultaneous desalting can be achieved by choosing the buffer conditions of the mobile phase.[49]

### **3.3.2. Chromatographic Methods for the Analysis of Glycoproteins**

Chromatographic methods are not only used for protein purification but also to identify and characterize structural properties of glycoproteins. High-performance liquid chromatography (HPLC) presents a sophisticated instrumental method for the automated analysis of complex mixtures and is applicable for various types of analytes.[50] It can be especially helpful for the identification and characterization of glycoproteins. A common method in this area is the identification of proteins by peptide mapping. Therefore, proteins are enzymatically digested, increasing the complexity of the sample. Subsequent chromatographic separation of peptides provides insights about the amino acid sequence, disulfide formation, and other modifications. Peptides can be separated on reversed phase columns containing long hydrophobic alkyl chains, usually C18 material. The reversed phase character is optimal for the retention of hydrophobic peptides, as well as separating intact proteins when using shorter alkyl chains ranging from C2 to C8. Reducing the alkyl chains minimizes the effect of hydrophobic adsorption of the protein, enabling other column parameters like column length to influence

separation. In 2017, Shen *et al.*, were able to identify over 900 proteins from a cell lysate, using up to 200 cm long RP capillary columns, with online MS detection of the intact proteins.[51]

Hydrophilic compounds like some peptides, glycopeptides or released glycans, however, are not well suited for separation on hydrophobic material.[52] The hydrophilic interaction chromatography (HILIC) presents an alternative method for separation of polar compounds, showing distinct advantages over normal phase (NP) chromatography. The stationary phase of HILIC columns is composed of silica or polymer-based materials, modified with different polar functional groups. Hydroxyl groups in form of silanols or diols, as well as cyano groups or amino groups, can be applied to create a polar particle surface.[53] Another often used material for the separation of polar compounds is porous graphitic carbon (PGC), a crystalline material composed of graphitic ribbons. Compared to silica-based materials, PGC is stable to extreme pH values and high temperatures. Another major advantage is the separation of isomeric glycans, making it an ideal material for the separation and analysis of released glycans.[54-56]

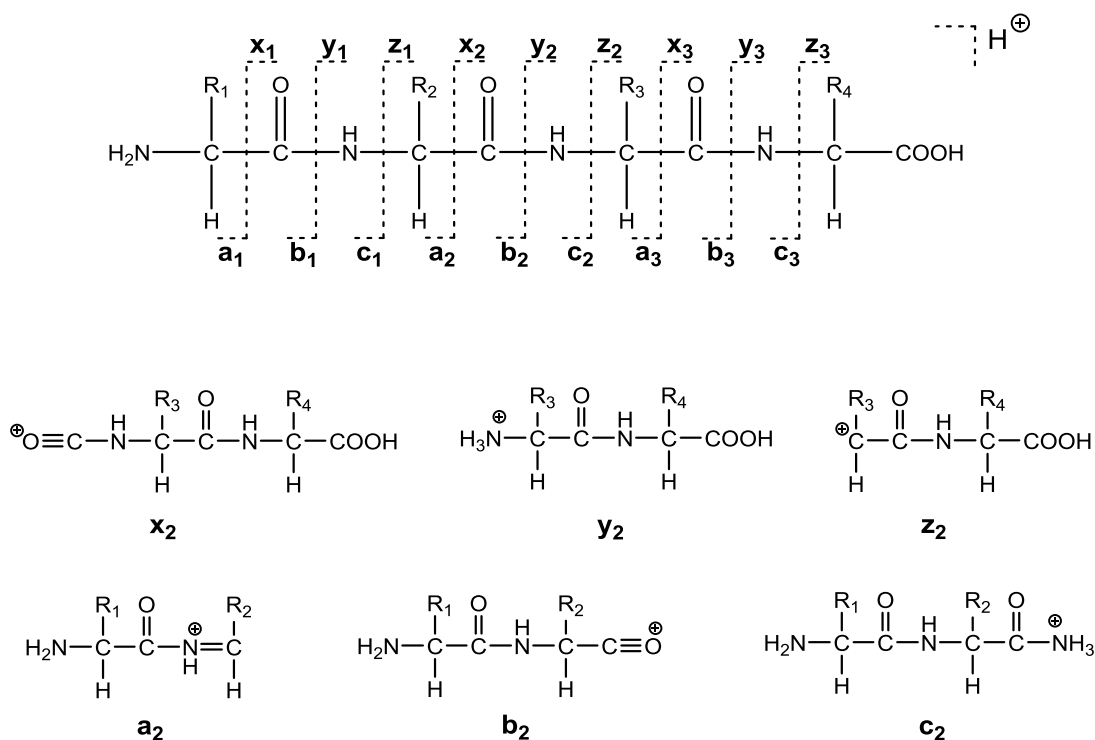
### **3.3.3. Mass Spectrometric Techniques for Glycoprotein Analysis**

Chromatographic methods, more specifically HPLC, for separation of peptides, proteins, or glycans as described above, can easily be connected to mass spectrometric detection. One way of online detection is the coupling of the HPLC flow to an electrospray ionization (ESI) source. ESI ionization presents a soft ionization method and is suitable for the ionization of biomolecules with high molecular masses as well as smaller peptides and glycopeptides. The analytes are detected as multiply charged ions. An alternative ionization technique for analyzing biomolecules with minimal fragmentation is the matrix-assisted laser desorption/ionization (MALDI) where analytes are crystallized with a matrix and ionized by a laser pulse. Compared to ESI, MALDI produces dominantly singly-charged ions.[57] Although soft ionization techniques with little fragmentation like ESI and MALDI are necessary for the analysis of glycoproteins, intentional fragmentation can also lead to structural information. Choosing the right ionization

technique, fragmentation mode and instrumental setup is highly dependent on the analyte of interest. As mentioned above, when analyzing proteins and their post-translational modifications different approaches for sample preparation are possible. A top-down approach describes the mass spectrometric analysis of intact proteins with minimal sample preparation. A more commonly used method is the bottom-up approach where proteins are enzymatically digested by different proteases. Subsequently, peptides can be analyzed with LC-MS. In between these two approaches lies the middle-down approach in which larger protein fragments are analyzed.

### **3.3.3.1. Bottom-Up Proteomics**

The commonly used approach for analyzing proteins and their post-translational modification is the bottom-up approach. The protein of interest is digested with enzymes like trypsin, chymotrypsin, or other proteases. The resulting peptides can be identified by LC-ESI-MS or MALDI-MS without fragmentation based on their mass. However, MS/MS fragmentation can give additional information about the peptide sequence and modifications. Collision-induced fragmentation (CID) describes the acceleration of protonated peptides in a vacuum chamber of the mass spectrometer. Collisions with a neutral gas like argon leads to bond disruption of the peptide backbone and generates mostly b- and y-ions (Figure 2).



**Figure 2.** Possible fragmentation pattern of a peptide with different fragmentation techniques. B- and y-ions are dominantly observed with CID fragmentation; z-, c- and y-ions are observed when electron-based fragmentation techniques like ETD or ECD are applied. The fragmentation of peptides can give information about amino acid sequence and modifications.

Modifications like phosphorylation or glycosylation are labile to this fragmentation mode and show characteristic fragmentation pattern. CID fragmentation is applicable for peptide ions that are doubly or triply-charged. Highly charged species are more difficult to sequence with this fragmentation mode.[58, 59]

In 1998, Zubarev *et al.* described a new method for peptide identification using electron-based fragmentation called electron capture dissociation (ECD).[60] The positively charged peptides or proteins can capture electrons and subsequently form radical cations, leading to specific cleavage of the peptide backbone. Based on this discovery, other techniques like electron transfer dissociation (ETD) have emerged to cover a variety of different applications. Compared to conventional CID fragmentation, electron based methods can be used for sequencing of larger peptides and proteins.[61] Post-translational modifications are retained, which can provide information about the position of these modifications.[62] The product ions

are dominantly z-, c- and y-ions, depending on the parameters used in electron-based fragmentation like electron energy or ion-electron irradiation time.[63]

MS/MS spectra of peptides obtained by different fragmentation techniques can be identified *de novo* or by automated database searches. This allows automated, high throughput protein identification of single proteins and protein mixtures.

### **3.3.3.2. Top-Down Proteomics of Intact Glycoproteins**

Top-down mass spectrometry describes the analysis of proteins without prior enzymatic or chemical treatment. Analyzing proteins and glycoproteins as intact biomolecules has several advantages over conventional proteomic approaches involving those treatments. Breaking down the protein of interest often results in a loss of information regarding post-translational modifications and amino acid sequence due to incomplete sequence coverage. Top-down MS is especially suited for the detection of proteoforms including genetic variants that show exchanges in amino acids and the analysis of glycosylation.[64-67] For an unambiguous assignment of proteoforms, a high-resolution instrument with high mass accuracy is required to detect small mass shifts between different proteoforms. The formation of a disulfide bridge leads for example to a mass shift of 2 Da and the difference between acetylation and trimethylation is only 0.036 Da. Fourier transform ion cyclotron resonance (FT-ICR) instruments fulfill the criteria of high resolution and mass accuracy and are often used for analyzing intact proteins. But also time-of-flight (TOF) instruments can be used to characterize proteins.[68]

Over the last few years, the mass limitations for native protein analysis were dramatically moved to higher masses. In 2008, Uetrecht *et al.* were able to ionize a 4 MDa hepatitis B virus (HBV) capsid with ESI-MS.[69, 70] This enables mass spectrometry to analyze not only small proteins but protein complexes and other large-scale biomolecules as intact molecules.

As mentioned above, fragmentation techniques for sequencing the amino acid backbone can be applied in the top-down analysis of proteins as well. Especially

electron-driven dissociation methods like electron capture dissociation (ECD) and electron transfer dissociation (ETD) have proven to be useful for the fragmentation of peptide backbones. Observed fragments are mostly c-, y- and z-ions (see Figure 2). Side chain modifications are preserved by this method. The use of fragmentation techniques in top-down mass spectrometry of proteins shows promising results for protein identification. Amino acid sequence coverage from MS/MS data of intact proteins can be over 90% by simultaneous detection of post-translational modifications.[71]

## 4. Objective

Glycosylation is one of the most important and abundant post-translational modification of proteins. Its biological function is, however, not fully understood. The biological relevance of glycosylation ranges from cellular signaling processes to protein folding and stability. But monitoring glycosylation is not only important to improve our understanding of this post-translational modification but also for the pharmaceutical industry. Analysis of glycosylation is crucial for e.g. quality control of biopharmaceuticals or the development of glycan-based biomarkers for a variety of disease like cancer or arthritis. However, analyzing the glycosylation of proteins is a challenging and time-consuming task due to their severe heterogeneity and structural complexity. The aim of this work was therefore focused on the development of new strategies for the rapid and unambiguous characterization of glycoproteins with HPLC, MS and NMR techniques.

This includes not only the structural characterization of post-translational modifications like glycosylation but also the detection of different protein variants caused by single nucleotide polymorphism (SNP). The characterization of different proteoforms of a specific protein can be a useful tool in the diagnosis and prognosis of many diseases, which leads to a great interest in the analysis of human serum proteins. To detect changes in glycosylation, different human plasma proteins should be purified and analyzed. These plasma proteins are alpha-acid-glycoprotein,  $\beta$ -2-glycoprotein 1, ceruloplasmin, immunoglobulin IgG1, and transferrin as changes in concentration or glycosylation due to photogenic processes have been observed on these proteins. Therefore, a purification strategy should be established that allows the simultaneous isolation of all plasma proteins from limited sample amount. Different chromatographic procedures, as well as precipitation techniques, should allow the isolation of proteins from a few microliters of human plasma. By analyzing the glycosylation of different proteins from the same plasma sample, it is possible to determine if changes in glycosylation only affect a specific protein or are a general feature that can be detected on more than one protein. The relative quantification of different glycan structures plays thereby a crucial role, which is why different methods for relative

quantification should be tested and evaluated in terms of accuracy. This includes bottom-up, middle-down and top-down approaches for the analysis of free glycans, glycopeptide and glycoproteins. LC-MS and MS/MS methods should be optimized for different analytes. Glycoproteins isolated from plasma of patients diagnosed with liver cirrhosis and hepatocellular carcinoma should be analyzed using the established methods and compared to the results obtained from healthy individuals. By this, the potential use of the analyzed proteins as glycan-based biomarker can be evaluated.



## 5. Results and Discussion

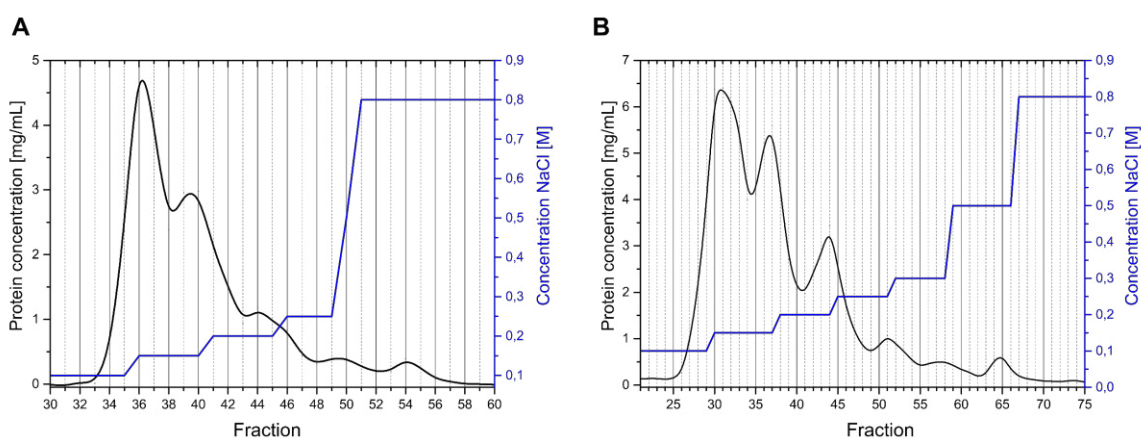
### 5.1. Protein Purification from Human Plasma

To minimize sample amount, a method for the simultaneous isolation of different plasma proteins from the same sample was established. By this it possible to monitor the glycosylation from several plasma proteins from a few microliters of human plasma. A combination of anion exchange chromatography, size exclusion chromatography and reversed phased chromatography was used to purify alpha-acid-glycoprotein (AGP),  $\beta$ -2-glycoprotein 1 (B2GP1), ceruloplasmin (CP), immunoglobulin IgG1 (IgG1), and transferrin (TF) from several plasma samples. A complete list of all analyzed plasma samples can be found in the experimental procedures, Table 16 and Table 17.

#### 5.1.1. Anion Exchange Chromatography

For the isolation of several glycoproteins from human plasma, an anion exchange chromatography was tested as a first step. This procedure separates proteins based on their pI through ionic interaction between a positively charged stationary phase and negative charges on the protein surface. A DEAE (diethylaminoethyl) sepharose column material was chosen as stationary phase. The pH was adjusted to 5.50 by equilibrating the column material with 0.05 M sodium acetate buffer. Proteins with a pI lower than 5.50 present negative charges and interact with the stationary phase. Unbound proteins with a pI higher than 5.50 can be eluted with sodium acetate buffer. This approach was used for the separation of low pI proteins which include ceruloplasmin (pI = 4.4) and alpha-acid-glycoprotein (pI 2.8 to 3.8), as well as abundant serum albumin (pI = 4.9).[72-74] Proteins with a higher pI like fibrinogen, immunoglobulin 1 (IgG1), and  $\beta$ -2-glycoprotein 1 are found in the flow through. To elute bound protein, an increasing sodium chloride gradient was applied. To optimize separation of bound proteins, in this case ceruloplasmin and alpha-acid-glycoprotein, different salt gradients were tested. Therefore 8 mL equilibrated DEAE material were loaded to

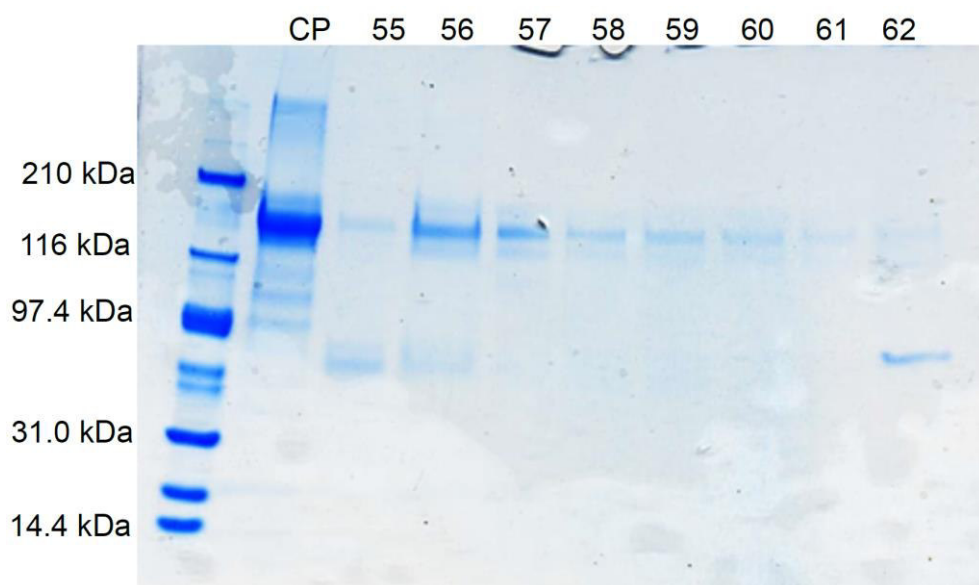
a column with a diameter of 1.5 cm and a column height of 4.5 cm. 1.5 mL of crude plasma was applied to this column and unbound proteins were eluted with 15 mL of 0.05 M sodium acetate buffer (pH 5.50). Subsequently, increasing salt gradients were applied to the column to elute bound proteins. Elution of proteins was monitored by measuring an absorption spectrum at 280 nm, using a *Nanodrop* spectrophotometer. According to the Lambert-Beer law, the relative absorption was used to calculate a protein concentration.[75] For that, the extinction coefficient of human albumin was used. Figure 3 shows the elution behavior of the bound proteins after applying different salt gradients. The size of each fraction was 0.5 mL.



**Figure 3.** Anion exchange chromatography with two different sodium chloride gradients from human plasma (**A** and **B**). Unbound proteins were washed off with 0.05 M sodium acetate buffer (pH 5.50; not shown in chromatogram). After applying different sodium chloride concentrations (in 0.05 M sodium acetate buffer, pH 5.50), bound proteins were eluted and collected. Applying more steps in the gradient (**B**) resulted in a slightly better separation of proteins.

The initial attempt of isolating plasma proteins by anion exchange chromatography led to the separation of high pI proteins like abundant fibrinogen and immunoglobulins. Low pI proteins like ceruloplasmin and alpha-acid-glycoprotein were eluted with different salt concentrations. Even though abundant albumin was dominantly found in the later fractions, it was also found in the flow through as contaminant. The application of the step gradient shown in Figure 3, panel B, however, showed promising results for the isolation of human ceruloplasmin. Figure 4 shows the SDS-PAGE gel of fractions 55 to 62 from the AIEX purification

shown in Figure 3, panel B. Lanes for fraction 56 and 57 show dominantly ceruloplasmin (compared to lane “CP” for 5  $\mu$ g of purchased ceruloplasmin). For all fractions, a double band is visible, which corresponds to ceruloplasmin (most intense, upper band) and proteoform of ceruloplasmin with a shortened N-terminus. This phenomenon is discussed in section 5.2.1. *Analysis of Intact Ceruloplasmin* in more detail.



**Figure 4.** SDS-PAGE Gel of different fraction after AIEX containing human ceruloplasmin. Heat-denatured samples (95 °C, 5 min) were applied to a precast gel (Biorad, 8-16% TGX Gel, 15 well). Running gel buffer was composed of 0.25 M Tris HCl, 2 M glycine and 0.1% SDS. Molecular masses of proteins were determined using a protein ladder (Amresco Wide range (K494), lane 1). Lane CP: 5  $\mu$ g of purchased ceruloplasmin. Lane 55 to 62 correspond to fraction after AIEX (see Figure 3, panel B).

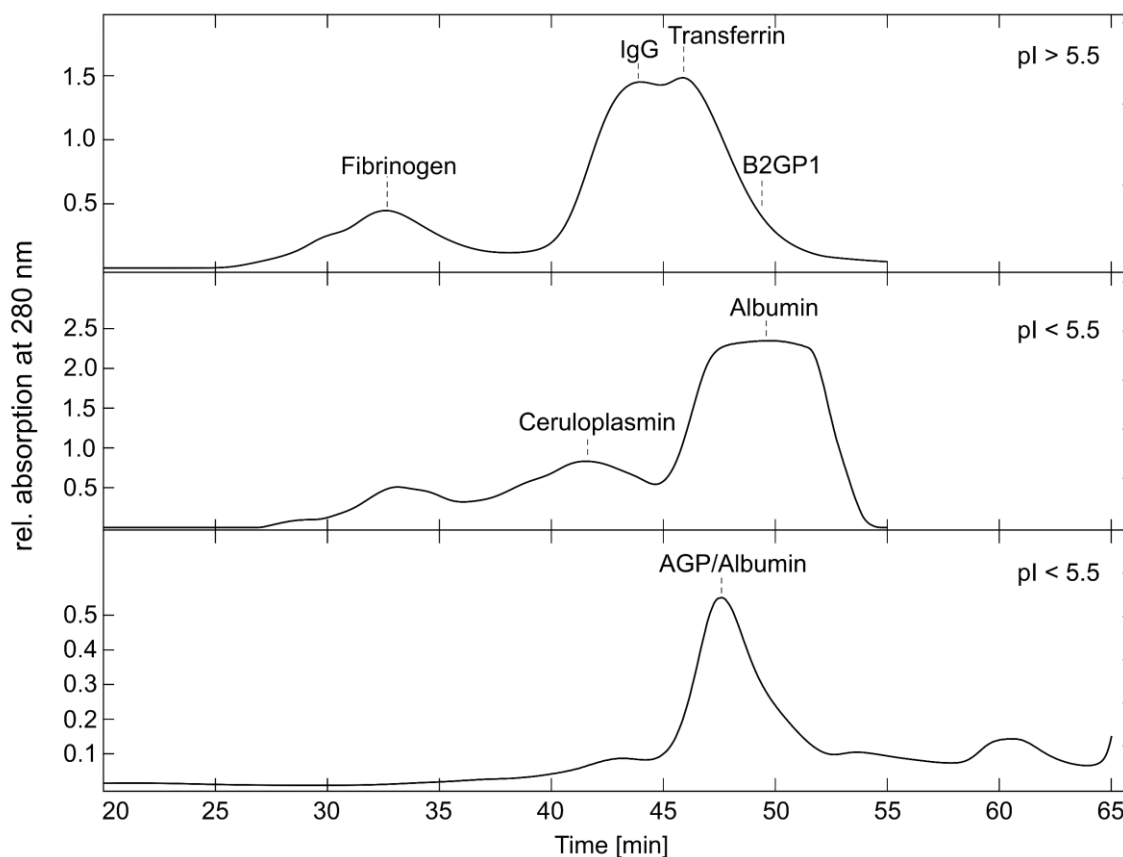
Even though this procedure yielded human ceruloplasmin in very high purity in a single step, this procedure was not optimized for the parallel purification of immunoglobulin IgG1, alpha-acid-glycoprotein and  $\beta$ -2-glycoprotein 1. The application of a manual step gradient for initial AIEX chromatography seemed too time consuming for the purification of several plasma samples. Therefore, the procedure was adjusted to a more robust and efficient protocol.

The sample amount was reduced to 500  $\mu$ L of human plasma and the volume of the column material was reduced to 2 mL. The column dimensions were adjusted to a diameter of 0.7 cm and a column height of 5 cm. Instead of applying a step

gradient to the AEX chromatography, unbound proteins were washed off with 0.05 M sodium acetate, pH 5.50. This flow through contained immunoglobulins, transferrin and  $\beta$ -2-glycoprotein 1. The bound proteins, including ceruloplasmin and alpha-acid-glycoprotein, were then eluted with 0.05 M sodium acetate buffer, pH 5.50 with 0.8 M sodium chloride and collected in fractions of 2 mL. After elution of ceruloplasmin detected by a slightly green color of the fractions, fractions containing alpha-acid-glycoprotein were collected.

### **5.1.2. Size Exclusion Chromatography**

After fractionating of plasma samples using anion exchange chromatography, a size exclusion purification was performed. This was not only done to further purify the samples but also as a desalting step after ion exchange chromatography. All fractions of the AEX chromatography (the flow through, fractions containing ceruloplasmin and fractions containing alpha-acid-glycoprotein) were purified using a size exclusion column. The column was operated at a flow rate of 300  $\mu$ L/min and fractions of 0.5 min were collected. Figure 5 shows an exemplary chromatogram for the three different fractions after size exclusion with UV detection.



**Figure 5.** LC chromatogram after size exclusion chromatography from three different fractions after anion exchange chromatography. The flow rate was set to 300  $\mu\text{L}/\text{min}$  and fractions in a time interval of 30 s were collected. **Top panel:** SEC chromatogram of the fraction obtained from the flow through of the AIEX purification step ( $p\text{I} > 5.5$ ). Major components were identified by ESI-MS as fibrinogen, transferrin, and B2GP1 as intact glycoprotein and IgGs after tryptic digest by a conventional bottom up approach. **Middle panel:** SEC chromatogram of AIEX fractions containing ceruloplasmin. Albumin was detected as major impurity. **Bottom panel:** SEC chromatogram of AIEX fractions containing low pI proteins like alpha-acid-glycoprotein and albumin with similar retention times.

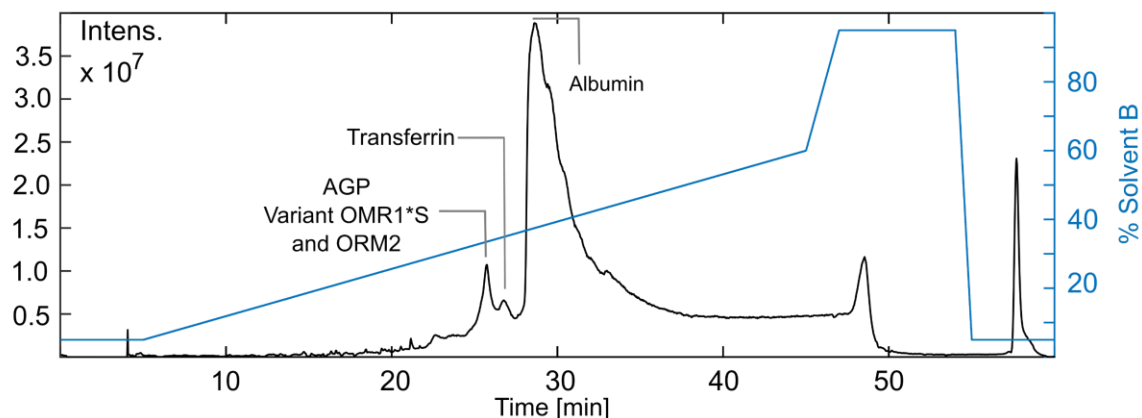
Different fractions containing alpha-acid-glycoprotein,  $\beta$ -2-glycoprotein 1, ceruloplasmin, immunoglobulin IgG1, and transferrin were collected and pooled separately.

Pooled fractions containing ceruloplasmin were once more subjected to size exclusion purification using the same parameters. Subsequently, fractions containing ceruloplasmin were reduced in volume and analyzed using a middle-down approach and a bottom-up approach described in section 5.2. *Analysis of Human Plasma Ceruloplasmin.*

The volume of the enriched protein samples (alpha-acid-glycoprotein,  $\beta$ -2-glycoprotein 1, immunoglobulin IgG1 and transferrin) was reduced using a centrifugal vacuum concentrator. Further purification steps varied between the different protein samples.

### **5.1.3. Reversed Phase Chromatography with Online ESI-MS Detection**

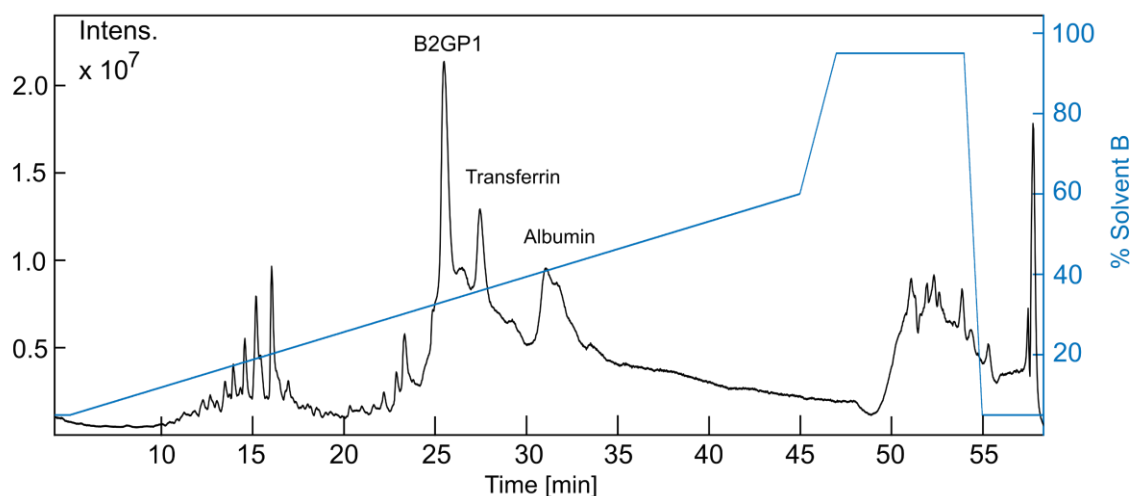
Alpha-Acid-Glycoprotein enriched from plasma samples was analyzed by ESI-MS as desialylated glycoprotein without further purification.  $\beta$ -2-Glycoprotein 1 and transferrin were analyzed as intact glycoproteins without prior enzymatic or another sample treatment. Even though the combination of anionic exchange chromatography and size exclusion did not yield protein samples of high purity, online LC-MS analysis showed different elution behavior for the proteins of interest and for other impurities. For LC-MS, a reversed phased column with C8 material was used, separating proteins by their hydrophobic character. The major plasma protein albumin was found in the samples containing alpha-acid-glycoprotein,  $\beta$ -2-glycoprotein 1, and transferrin. However, albumin eluted after the glycoproteins of interest, indicating a more hydrophobic character compared to the glycosylated proteins. Figure 6 shows an exemplary total ion chromatogram (TIC) for enriched and desialylated AGP from one plasma sample. Next to albumin, transferrin was found as a minor impurity. Even though AGP is an abundant plasma protein with a concentration of about 1 mg/mL, ion intensity is low in positive ion mode.[76] The high abundance of glycan residues might hamper ionization in positive ion mode even after desialylation. Therefore, the TIC cannot be used as an indication of relative abundance of different proteins in the sample.



**Figure 6.** Exemplary total ion chromatogram from one plasma sample after enrichment and desialylation of alpha-acid-glycoprotein (AGP). HPLC was performed on a RP-C8 column with online ESI-MS detection in positive ion mode. The gradient for solvent B: ACN + 0.1% FA is shown in blue. Next to major plasma protein albumin, transferrin was found as an impurity, showing a different elution behavior as desialylated AGP.

Desialylated AGP was analyzed as intact glycoprotein by summation of the LC peak in a time range from 25 to 26 min. As further discussed in the section 5.5.2. *Analysis of Desialylated AGP from Individual Plasma Samples*, different genetic variants of alpha-acid-glycoprotein-1 and alpha-acid-glycoprotein-2 showed a slightly different elution behavior.

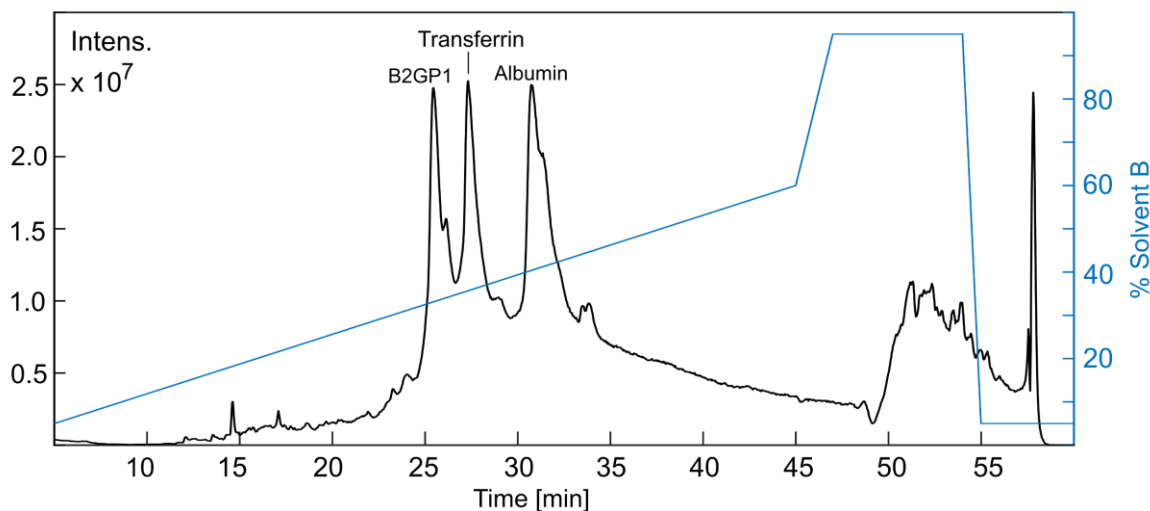
$\beta$ -2-Glycoprotein 1 is also a heavily glycosylated protein with four or five N-glycosylation sites. Therefore, it is not surprising that it showed a similar retention time as desialylated AGP on the RP-C8 column under similar conditions. Figure 7 shows the total ion chromatogram of enriched  $\beta$ -2-glycoprotein 1 with online ESI-MS detection in positive ion mode. Albumin and transferrin were found as impurities, showing different elution behavior.



**Figure 7.** Exemplary total ion chromatogram from one plasma sample after enrichment of B2GP1. HPLC was performed on a RP-C8 column with online ESI-MS detection in positive ion mode. The gradient for solvent B: ACN + 0.1% FA is shown in blue. Next to major plasma protein albumin, transferrin was found as an impurity, showing a different elution behavior as B2GP1.

Both,  $\beta$ -2-glycoprotein 1 and transferrin, were purified from the same size exclusion run but from different fractions. Consequently,  $\beta$ -2-glycoprotein 1 was also found as an impurity for enriched transferrin as well as albumin. Figure 8 shows the total ion chromatogram from one plasma sample for enriched transferrin with online ESI-MS detection in positive ion mode. Transferrin shows a retention time of 27 to 28 min. Even though the sample does not show a high purity, ESI-MS spectra of multiply charged transferrin were obtained without overlapping signals from other proteins. Serum transferrin is often analyzed as intact glycoprotein by top-down mass spectrometry since the loss of glycosylation is associated with alcohol abuse and other diseases.[77-81] The combination of anion exchange chromatography and size exclusion chromatography in combination with LC-MS presents a fast method for protein purification that can be applied to an already well-established glycan-based biomarker.





**Figure 8.** Exemplary total ion chromatogram from one plasma sample after enrichment of serum transferrin. HPLC was performed on a RP-C8 column with online ESI-MS detection in positive ion mode. The gradient for solvent B: ACN + 0.1% FA is shown in blue. Next to major plasma protein albumin, B2GP1 was found as an impurity, showing a different elution behavior as transferrin.

#### 5.1.4. Automated Peptide Identification of Proteins

Proteins that were not analyzed as intact proteins were identified by automated peptide identification using the free web version of the Mascot search engine.[82] This was the case for ceruloplasmin, which was purified with a combination of anion exchange chromatography and size exclusion chromatography, and for IgG1 with an additional purification step using *Protein A SpinTrap* columns. Protein A shows a high affinity for human IgG1, IgG2, and IgG4, presenting a good way of purifying IgG subclasses from other proteins.

Ceruloplasmin and IgG were digested using trypsin and analyzed by LC-ESI-MS using a C18 column. MS/MS spectra of peptides and glycopeptides were obtained by CID fragmentation and compound spectra were generated for MS and MS/MS spectra using DataAnalysis 4.2. The SwissProt database was searched for mammalian proteins using the Mascot MS/MS Ions Search. Table 1 shows the results for the Mascot search for purified ceruloplasmin and purified IgG (IgG1, IgG2, and IgG4) from sample P. A complete list for all samples can be found in the

appendix (10.1. Results for Mascot search after Tryptic Digest of Purified Ceruloplasmin).

**Table 1.** Exemplary results for automated peptide identification using Mascot MS/MS Ions Search for ceruloplasmin and IgG purified from plasma sample P.

Sample	Protein	Score	Mass [Da]
Ceruloplasmin purified from sample P	Ceruloplasmin	716	122128
	Complement C3	486	187030
	Vitronectin	144	54271
	Immunoglobulin gamma heavy chain	91	49298
	Apolipoprotein A-2	84	11168
	Prothrombin	77	69992
	Cationic trypsin (bovine)	69	25769
	Serum albumin	69	69321
	Apolipoprotein A-1	59	30759
	Alpha-1-antichymotrypsin	57	47621
Clusterin	44	52461	
IgG purified from sample P	Cationic trypsin (bovine)	859	25769
	Immunoglobulin gamma heavy chain	468	49298
	Immunoglobulin heavy constant gamma 2	156	35878
	Immunoglobulin kappa constant	133	11758

The SwissProt data bank was searched using the Mascot MS/MS ion search with the following parameters: enzyme = trypsin; taxonomy = *mammalia*; maximum missed cleavages = 1; no fixed modifications; variable modifications = pyro-glutamine; peptide tolerance = 0.02 Da; peptide charge = 2+, 3+ and 4+.[83]

As shown in Table 1, ceruloplasmin and IgG were successfully isolated from a few microliters of human plasma. For ceruloplasmin, the Mascot search identified several other proteins in the samples with a lower score. However, when analyzing glycopeptides the peptide moiety usually allows an unambiguous assignment of glycan structures to a specific glycosite of a glycoprotein. Therefore, analyzing glycopeptides of ceruloplasmin and IgG1 was possible even though other proteins were identified in the samples.

Analyzing free glycans released with PNGase F requires a higher purity to ensure that only glycan structures of a specific protein are analyzed. To analyze free glycans released from ceruloplasmin, one sample was further purified by repeating the size exclusion chromatography as described above. In total, the sample was

three times subjected to size exclusion chromatography. Table 2 shows the results for the Mascot MS/MS Ions Search for sample I which was once more purified by size exclusion chromatography.

**Table 2.** Mascot search results for ceruloplasmin from sample I. An additional size exclusion step was performed to increase purify for the analysis of free glycan.

Sample	Protein	Score	Mass [Da]
Ceruloplasmin purified from sample I	Ceruloplasmin	4064	122128
	Cationic trypsin (bovine)	798	25769
	Serum albumin	724	69321
	Glutamate receptor ionotropic	19	143661
	Apolipoprotein A-1	19	30759

The SwissProt data bank was searched using the Mascot MS/MS ion search with the following parameters: enzyme = trypsin; taxonomy = *mammalia*; maximum missed cleavages = 1; no fixed modifications; variable modifications = pyro-glutamine; peptide tolerance = 0.02 Da; peptide charge = 2+, 3+ and 4+.[83]

In sample I, ceruloplasmin is found with a much higher score. The bovine trypsin was used for proteolytic digest. Although albumin was identified as a minor impurity, it will not interfere with glycan analysis since albumin does not carry any N-glycosylation. Heavily glycosylated glutamate receptor ionotropic and apolipoprotein A-1 were found with a very low score so that no further purification of ceruloplasmin was done for the analysis of free glycans.

## 5.2. Analysis of Human Plasma Ceruloplasmin

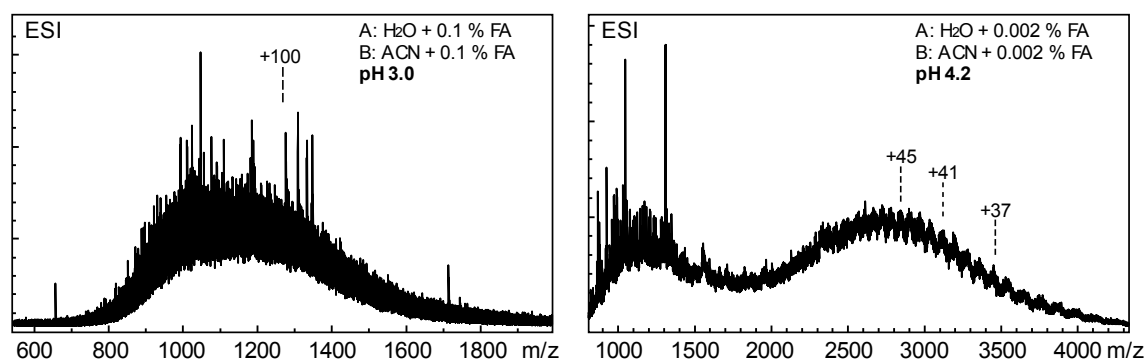
Ceruloplasmin is a 130 kDa glycoprotein with 1046 amino acids, mainly produced in liver cells. Its four glycosylation sites carry dominantly biantennary and triantennary, fully sialylated glycans with up to three fucosyl residues.[84] The N-glycans of ceruloplasmin are linked to asparagine 138 (<sup>138</sup>N), asparagine 358 (<sup>358</sup>N), asparagine 397 (<sup>397</sup>N), and asparagine 762 (<sup>762</sup>N).[84] Its main function is the transport of copper ions and oxidation of iron(II) to iron(III) ions, enabling iron transport through the cell membrane.[85] This allows an interplay of ceruloplasmin and transferrin, one of the main proteins for transport of iron(III) ions. The oxidation of iron(II) to iron (III) by ceruloplasmin is not only important for the transport of iron but also reduces the amount of highly damaging hydroxy radicals in cells. These hydroxyl radicals are formed by the Fenton reaction which occurs between iron(II) and hydrogen peroxide, forming hydroxy radicals.[86] Organic substrates like phenols and ascorbate can be oxidized by ceruloplasmin as well.[85]

Ceruloplasmin is a positive acute-phase protein with an average concentration of 300 µg/mL in plasma. Plasma levels of ceruloplasmin significantly decrease in patients diagnosed with Wilson disease to a level of 50 µg/mL, whereas an increase to 480 µg/mL was observed in patients diagnosed with breast cancer.[87, 88] However, many more diseases are associated with a change in ceruloplasmin concentration, questioning its diagnostic value.[89-91]

The quantitative analysis of the glycosylation of ceruloplasmin has been discussed as a diagnostic marker as well. Especially in patients with cancer diseases, a change in glycosylation was observed. This includes for example patients with alcohol-related hepatocellular carcinoma, who showed an increased core fucosylation of ceruloplasmin.[14] In patients with pancreatic cancer an increase in the expression of the Sialyl-Lewis X motive on ceruloplasmin was observed.[92] Hence, analyzing the glycosylation of ceruloplasmin is a promising approach for monitoring disease-related changes and for biomarker development. Therefore, the possibility of analyzing ceruloplasmin as intact glycoprotein was evaluated as well as the use of bottom-up and middle-down proteomics for a qualitative and quantitative analysis of the different proteoforms of ceruloplasmin.

### 5.2.1. Analysis of Intact Ceruloplasmin

With about 130 kDa, ceruloplasmin is a rather large glycoprotein. Its glycosylation, as well as the varying number of metal cations, lead to a heterogeneous structure with different proteoforms. Analyzing intact ceruloplasmin using ESI-MS did not yield resolvable spectra and only showed broad, overlapping charge states of highly charged ion species in positive ion mode as well as negative ion mode. Changes in ion source conditions like ISCID or drying temperature did not improve ionization. The increase in pH of the LC solvent from 3.0 to 4.2 in positive mode led to a shift of the ion distribution to higher masses (Figure 9). However, no interpretable spectrum was obtained. The lack of resolution and ion intensity as well as overlapping charge states interfered with the identification of different proteoforms.

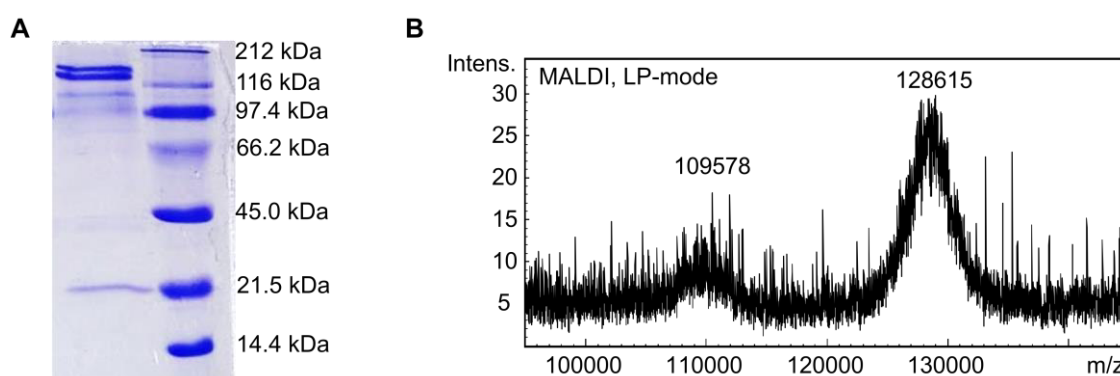


**Figure 9.** ESI-MS spectra of human plasma ceruloplasmin in positive ion mode. At pH 3.0 (left), a broad peak is observed presumably due to highly charged, overlapping ion species. At a mass of about 1280 m/z, the theoretical charge state is +100. Increasing the pH of the LC solvent to 4.2 shifted the ion distribution to higher masses (right). Peaks corresponding to distinct charge states are observed (e.g. +45, +41 and +37). However, no interpretable spectrum was obtained.

Compared to ESI, MALDI produces dominantly singly or doubly charged ions. By using MALDI-TOF MS, it was possible to obtain characteristic peaks to determine the intact mass of human serum ceruloplasmin. MS spectra, as well as SDS-PAGE gels of ceruloplasmin from different plasma samples, often showed two peaks or a double band, as shown in Figure 10 and as reported earlier.[93] MALDI-MS of intact human ceruloplasmin revealed a mass of 128 kDa for the heavier protein form and 110 kDa for the smaller protein form. However, the resolution of the used

MALDI-TOF mass spectrometer was insufficient to resolve different proteoforms, caused by PTMs like glycosylation or SNPs.

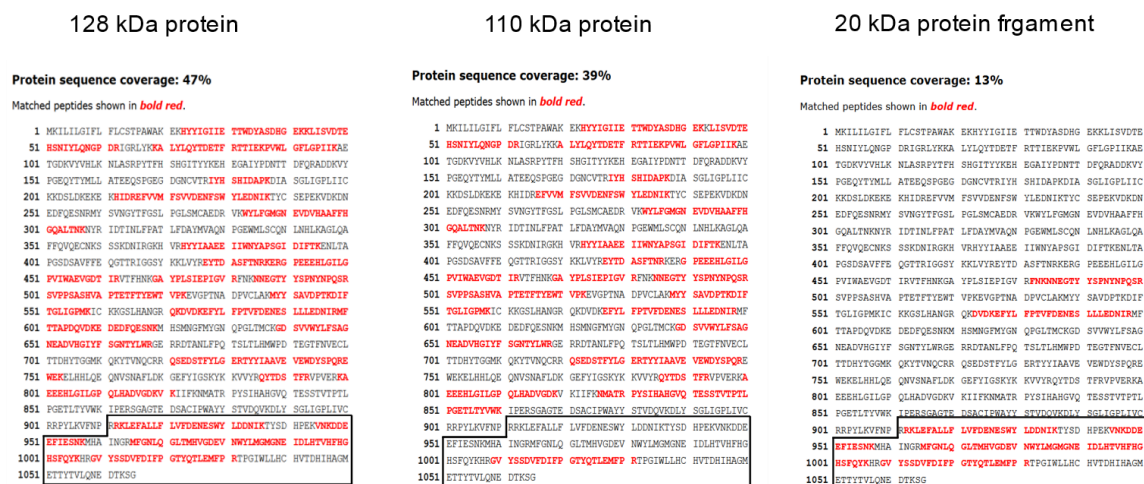
The proportion of the two bands (SDS-PAGE) or peaks (MALDI spectra) varied between different plasma samples (see Figure 4 and Figure 10). The origin of these two different forms has been discussed before. Sato *et al.* suggested a difference in glycosylation for the two forms.[94] This seems, however, unlikely because the loss of all glycosylation would lead to a mass of about 120 kDa instead of 110 kDa for intact ceruloplasmin.



**Figure 10.** Intact ceruloplasmin shows two distinct masses. **(A)** SDS-PAGE gel of denatured ceruloplasmin purified from human plasma. A double band at about 120 kDa is visible. A third band is present at about 20 kDa. **(B)** MALDI-MS spectra in linear positive mode (LP) of ceruloplasmin from human plasma. A dominant peak at 129 kDa and a second peak at 110 kDa are present, which is in good agreement with previously published results.[93]

Another explanation for the mass difference is a proteolytic cleavage of the 128 kDa fragment to generate the smaller form of about 110 kDa.[95, 96] To evaluate the two theories discussed in the literature concerning the origin of the different protein forms, an in-gel digest was performed. The upper band and lower band at 128 kDa and 110 kDa as well as the band at about 20 kDa were cut out separately (see Figure 10, panel A) and digested with trypsin. The resulting peptides and glycopeptides were separated on a RP C18 column coupled to online ESI-MS/MS detection. Resulting compound spectra were interpreted with automated peptide identification using the Mascot search engine as described in section 5.1.4. *Automated Peptide Identification of Proteins*

Figure 11 shows the sequence coverage of human ceruloplasmin for the different protein fragments after in-gel digest and LC-MS analysis.



**Figure 11.** Sequence coverage of human plasma ceruloplasmin after tryptic in-gel digest from three different protein fragments. Using the Mascot database search engine, a sequence coverage of 47% was determined for the protein form with a mass of about 128 kDa (by MALDI-MS). The smaller protein form with a size of 110 kDa (by MALDI-MS) showed a decreased sequence coverage of 39%, with less peptides found in the C terminal region (circled with a black box) of ceruloplasmin. The smallest protein fragment with a mass of 20 kDa (by SDS-PAGE) showed dominantly peptide matches for the C terminal region. The results indicate a proteolytic digest of the 128 kDa protein form, generating a 110 kDa protein form and a 20 kDa protein fragment.

The protein form with a mass of 128 kDa showed a sequence coverage of 47%. The smaller protein form with a mass of 110 kDa showed a sequence coverage of only 39%, missing peptide matches in the C terminal region compared to the 128 kDa protein form. The protein fragment at 20 kDa however, showed dominantly peptide matches in the C terminal region. The results indicate a proteolytic cleavage of the 128 kDa protein, generating the smaller protein of 110 kDa and the C terminal fragment of about 20 kDa. The manual analysis of the glycopeptides from the 128 kDa protein form and the 110 kDa protein form showed no differences in glycosylation.

### 5.2.2. Analysis of Released Glycans by MS and NMR Techniques

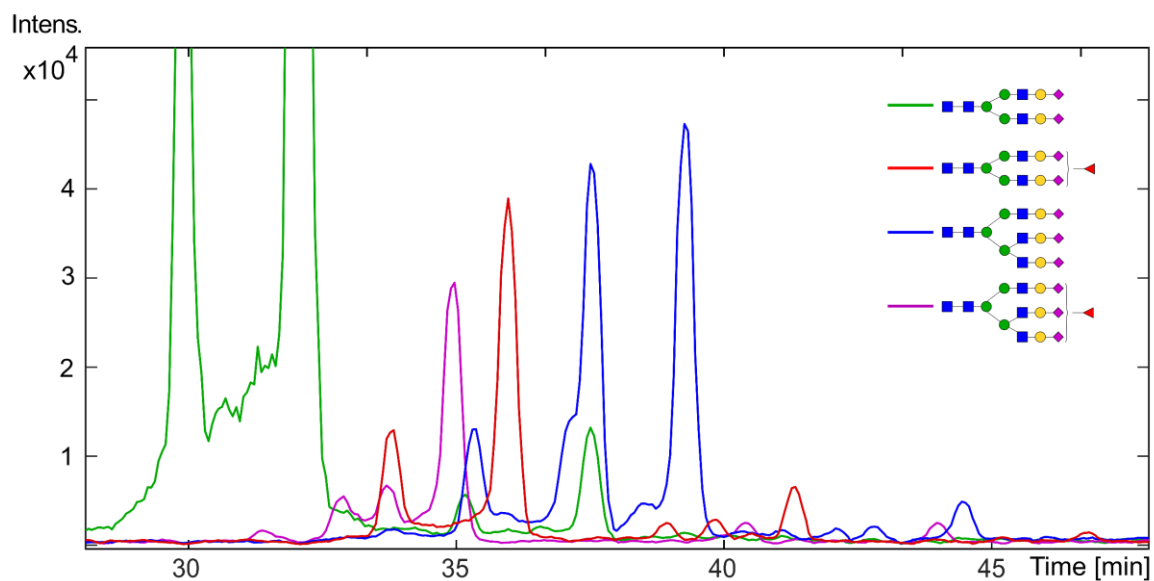
This chapter focuses on the characterization of released N-glycans from human ceruloplasmin using the combination of LC, MS and NMR data.

Structure determination of complex glycans is very challenging due to their variety of isomers. Nevertheless, linkage position and anomericity play an important role in many biological functions. A complete structural determination of glycans can be achieved using NMR spectroscopy. However, the heterogeneity of glycosylation leads to a complex mixture of glycans released from proteins. Therefore, the combination of different techniques (LC, MS, and NMR) is a highly promising approach for the identification of free glycans. The workflow includes the chromatographic separation of glycans with online MS analysis on an ESI-qTOF mass spectrometer. Simultaneously, the LC flow is split and fractionated for offline  $^1\text{H-NMR}$  analysis. The NMR analysis of pure glycan structures yields in the determination of connectivity and configuration and can be performed with minimal sample amount down to 15 pmol.[97] However, chromatographic separation of complex glycan mixtures is tremendously challenging, often resulting in overlapping peaks in the LC chromatogram. Using *three dimensional cross correlation* (3DCC), pure NMR spectra can be obtained from mixtures, including information from LC and MS analysis.[42, 98] The concept of 3DCC is the correlation of the extracted ion chromatograms (EIC) (*first dimension*) with the extracted delta chromatograms (EDC) (*second dimension*) over their shared retention time dimension (*third dimension*). The correlation can be done with a least square optimization (3DCC<sub>L</sub>), with correlation coefficients (3DCC<sub>C</sub>) or by singular value decomposition (3DCC<sub>S</sub>). By this, NMR peaks can be assigned to compounds with specific retention times and a pure NMR spectrum is obtained.

The glycosylation of human ceruloplasmin was analyzed using LC-MS and NMR techniques to identify not only the composition but also the configuration of the most abundant glycan structures. Therefore, purified ceruloplasmin was subsequently digested with trypsin and PNGase F. The released glycans were enriched and separated from peptides using C18 solid phase extraction. LC-MS analysis was performed on a PGC column with MS detection in positive ion mode.



When separating free glycans on a PGC column, the choice of the mobile phase is very important. Volatile solvents must be used when coupled to online MS analysis. Figure 12 shows the extracted ion chromatograms of the most abundant glycans, released from human ceruloplasmin. The solvents of the mobile phase were composed of A: H<sub>2</sub>O + 0.1% FA and B: ACN + 0.1% FA, resulting in an acidic pH of 3.0.

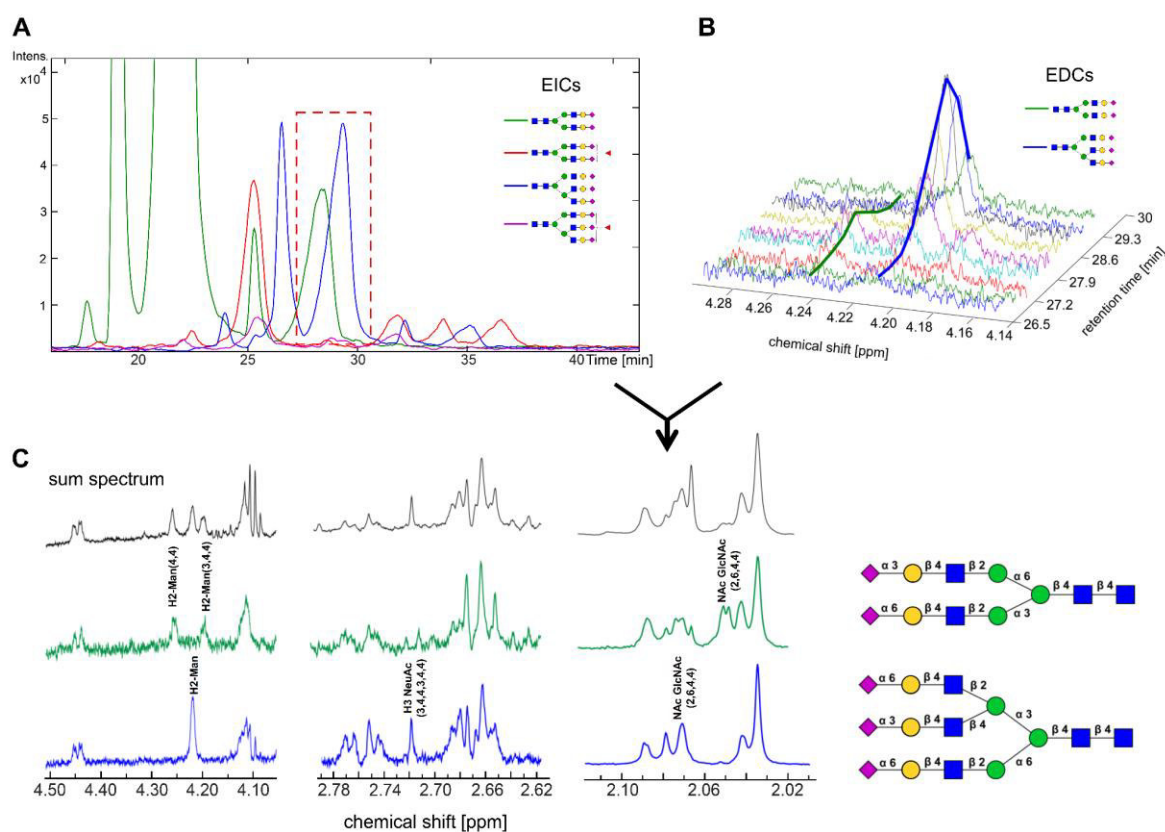


**Figure 12.** Extracted ion chromatogram of released glycans from ceruloplasmin at pH 3.0. A PGC column coupled to online ESI-MS detection in positive ion mode was used. Separation of alpha and beta anomers is observed.

The use of an acidic pH leads to a homogeneous protonation of the glycan compounds resulting in narrow elution peaks. However, separation of alpha and beta anomers is also observed due to mutarotation. To reduce this effect, 65 mM ammonia was added to the LC solvents instead of FA. This changes the pH to alkaline conditions and a peak broadening is observed (Figure 13, panel A).

For a complete structural assignment of different glycan compositions, partly separated glycans were analyzed by NMR. Therefore, the LC flow from PGC separation with the addition of ammonia was split for MS analysis (1/20) and fraction collection (19/20). Fractions were collected in a time interval of 20 s and dried using a centrifugal vacuum concentrator. The samples were resolved in 180  $\mu$ L of D<sub>2</sub>O and analyzed by <sup>1</sup>H-NMR. The most abundant glycan corresponds

to a fully sialylated diantennary structure as shown in Figure 12 and Figure 13 (A). No overlapping peaks were observed, and an unambiguous assignment was possible. NMR data showed that both sialic acids are 2,6-linked, which is in good agreement with previously published results.[99] A  $^1\text{H}$ -NMR spectrum with a complete peak assignment for the dominant glycan structure can be found in the appendix (10.2. NMR Data of Free N-Glycans of Ceruloplasmin). In contrast to the dominant structure, other glycan structures showed overlapping peaks during LC separation. This was the case for an isomer of a diantennary, fully sialylated structure and a triantennary, fully sialylated structure. To obtain pure NMR spectra for these two isomers 3DCC<sub>c</sub> was applied in the elution time interval. The procedure is visualized in Figure 13.



**Figure 13.** A three-dimensional cross-correlation was applied to calculate pure NMR for glycan structures who showed overlapping peaks during LC separation at a retention time from 26.5 min to 30 min. The EICs were correlated with the EDCs using 3DCC<sub>c</sub>. Pure NMR spectra were obtained by multiplication of the correlation coefficients with the sum NMR spectrum. The diantennary structure carries a 2,3-linked sialic acid at the 6-arm and a 2,6-linked neuraminic acid at the 3-arm. The triantennary structure carries two 2,6-linked and one 2,3-linked sialic acids.

The EICs were correlated with the EDCs over their shared time domain. The correlation was done using the Matlab implemented command *corrcoef* according to Pearson's linear correlation. Correlation coefficients between 0 and  $\pm 1$  are obtained. Negative correlation coefficients were set to 0. The correlation coefficients were multiplied with the NMR sum spectrum (sum of all NMR spectra in the retention time interval) and pure NMR spectra for both structures were obtained. The diantennary structure carries a 2,3-linked sialic acid at the 6-arm and a 2,6-linked neuraminic acid at the 3-arm. The triantennary structure carries two 2,6-linked and one 2,3-linked sialic acids. The results are in good agreement with previously published results.[99]

Due to the very high abundance of the diantennary structure with both sialic acids 2,6-linkes and a retention time of 22 min, it was not possible to assign more structures except the one described above. Further structures showed overlapping peaks and low signal-to-noise ratio.

### 5.2.3. Quantification of Glycopeptides with Different Peptide Length

Most parts in this chapter concerning the analysis of ceruloplasmin in a middle-down approach compared to a bottom-up approach have been published in the *glycoconjugate journal* and were obtained during the supervision of the bachelor thesis of M. Moritz.[83]

The glycosylation of proteins can be analyzed on different levels using mass spectrometry. Analyzing free glycans with mass spectrometry can provide information not only about glycan composition, but also linkage position and in some cases configuration.[39, 100] However, quantification of free glycans is quite challenging. Due to the lack of basic sites and due to their hydrophilicity, underivatized, free glycans often show difficulties in ionization in positive ion mode. Consequently, larger oligosaccharides are more difficult to ionize than smaller analytes. The increase of hydrophobicity by permethylation, peracetylation or other labeling techniques allows a better ionization and thus lowers the detection limits and increases the signal-to-noise ratio. However, labeling techniques require a certain level of expertise and can alter glycan composition due to loss of neuraminic acids or other glycan residues.[101] Analyzing glycopeptides has the advantage of increasing hydrophobicity by the peptide part, altering the ionization properties. Overall intensities are improved in positive ion mode, allowing the detection of less abundant glycans and yielding different results for quantitation. To investigate the influence of the peptide part for the quantification of glycans from glycopeptides, ceruloplasmin was analyzed using a common bottom-up approach and a middle-down approach. This was done to generate a set of glycopeptides with different peptide lengths. For the bottom-up approach, ceruloplasmin from seventeen different plasma samples was digested with trypsin. To generate glycopeptides with longer peptide parts in a middle-down approach, ceruloplasmin from the same plasma samples was digested with endoproteinase Arg-C or treated with trypsin using a limited digestion procedure. This limited procedure is described by the treatment of ceruloplasmin with trypsin for only a few minutes compared to a complete digestion, which is usually performed for several hours. The glycopeptides with different peptide parts were then analyzed by ESI-MS in positive

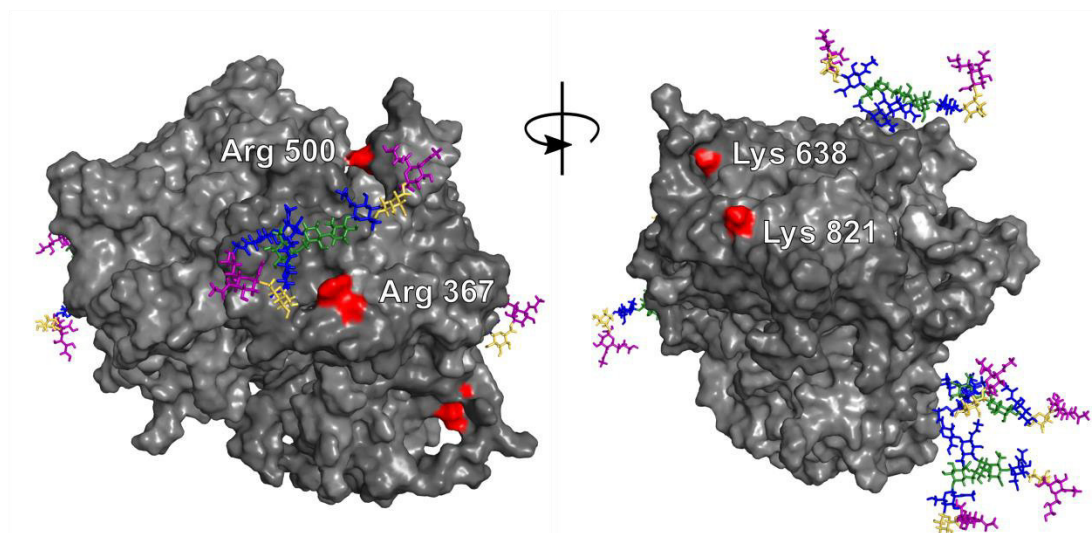
ion mode with online chromatographic separation on a C18 column. Table 3 describes the properties for the set of glycopeptides for the four glycosylation sites.

**Table 3.** Glycopeptides generated by different enzymatic procedures. For each glycosite, glycopeptides from a tryptic digest were compared to glycopeptides with enhanced peptide length.

Glyco site	Enzyme	Peptide sequence	Peptide mass <sup>a</sup>	GRAVY <sup>b</sup>	Retention time <sup>c</sup>	Most intense charge
138	Trypsin	<sup>129</sup> EHEGAIYDNTTDFQR <sup>144</sup>	1891.83	-1.519	22.0 min	+3
138	Arg-C	<sup>116</sup> PYTFHSHGITYYKEHEGAIYDNTTDFQR <sup>144</sup>	3486.59	-1.221	26.2 min	+6
358	Trypsin	<sup>346</sup> AGLQAFFQVQECKSSSK <sup>363</sup>	1970.95	-0.467	30.3 min	+3
358	Arg-C	<sup>311</sup> IDTINLFPATLFDAYMVAQNPGEWMLSCQNLNHLK AGLQAFFQVQECKSSSKDNIR <sup>367</sup>	6458.14	-0.219	36.2 min	+5
397	Trypsin	<sup>396</sup> ENLTAPGSDSAVFFEQGTTR <sup>415</sup>	2125.99	-0.545	33.3 min	+4
397	Limited digest with trypsin	<sup>368</sup> GKHVRHYIIAAEEIWNYPAPSGIDIFTKENTLAPGSD SAVFFEQGTTRIGGSYKLVYREYTDASFTNRKERGP EEEHLGILGPVIWAEVGDTRVTFHNKGAYPLSIEPIGV RFNKNNEGTYSPNYNPQSR <sup>500</sup>	15082.55	-0.654	31.7 min	+10 <sup>d</sup> (+8 to +23)
762	Trypsin	<sup>754</sup> ELHHLQEQNVSNFLDKGEFYIGSK <sup>778</sup>	2902.43	-0.736	31.6 min	+5
762	Limited digest with trypsin	<sup>639</sup> GDSVWVYLFAGNEADVHGIYFSGNTYLWRGERR DTANLFPQTSLLHMWPDTEGTFNVECLTTDHYTGG MKQKYTVNQCRQSEDFYLGERTYIIAAVEVEWD YSPQREWEKELHHLQEQNVSNFLDKGEFYIGSKYK KVYRQYTDSTFRVPERKAEEEHLGILGPQLHADVG DKVK <sup>821</sup>	21370.33	-0.773	31.4 min	+22 <sup>d</sup> (+8 to +24)

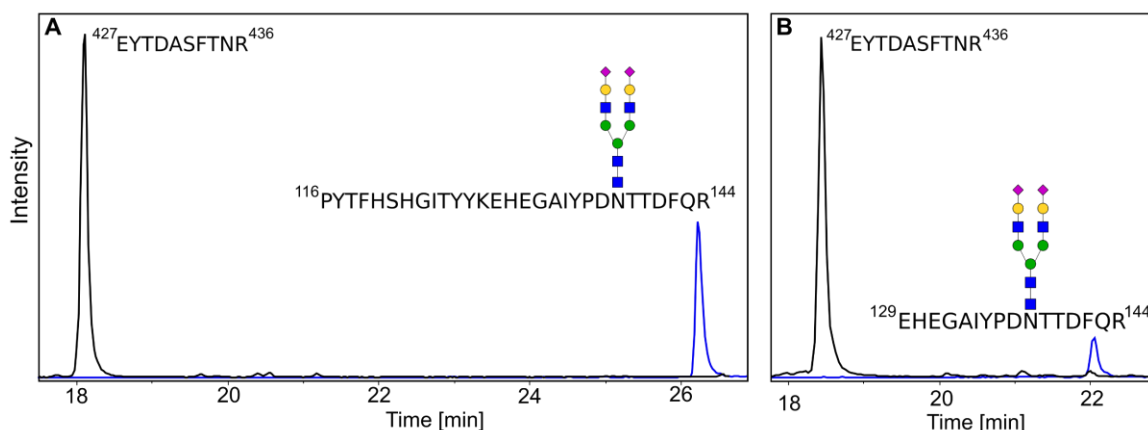
<sup>a</sup>Peptide mass states the monoisotopic mass. <sup>b</sup>Hydrophobicity index (GRAVY) for peptide part only. <sup>c</sup>On C18 column, refers to the most abundant glycopeptide. <sup>d</sup>Range of charge states is given in parenthesis.

The tryptic peptide for glycosite <sup>358</sup>N shows a missed cleavage near the glycosylation site (<sup>359</sup>K). Glycosylation may impede the action of trypsin during proteolysis. The shorter fully cleaved glycopeptides were detected with very low intensity and were not used for interpretation. The glycopeptides obtained by limited digestion with trypsin are predominantly formed by cleavage sites on the protein surface (Figure 14). This probably led to the reproducibility of glycopeptide fragments. The glycopeptides obtained in by limited digestion may also span more than one domain.



**Figure 14.** Cleavage sites for glycopeptides produced by limited digestion with trypsin. Shown are front and back side of ceruloplasmin with exemplary glycans at the four glycosites. The glycopeptides for glycosites <sup>397</sup>N (cleaved at arginine 367 and arginine 500) and glycosites <sup>762</sup>N (cleaved at lysine 638 and lysine 821) were obtained reproducibly for all samples, probably due to exposed cleavage sites on the surface of the protein (marked red). The protein structure was modified from PDB entry 2J5W.[102]

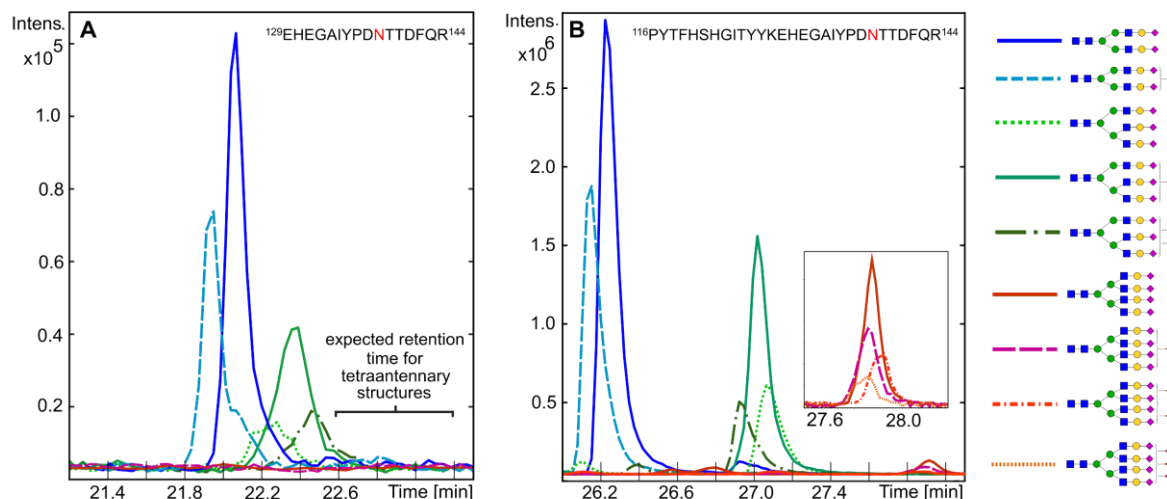
The stated GRAVY index in Table 3 calculates the hydrophobicity only based on the peptide parts and does not consider glycosylation.[103] However, the retention behavior on the RP column for the different glycopeptides also provides information about their hydrophobicity. In most cases, longer glycopeptides showed an increase in retention time, indicating a decrease in hydrophobicity. As expected, their relative intensities also increased with increasing peptide length and hence, decreasing hydrophobicity. Figure 15 shows the extracted ion chromatograms for the two glycopeptides of glycosite <sup>138</sup>N, either obtained by digestion with trypsin or with digestion with Arg-C. Both glycopeptides are modified with the same glycan residue, H5N4S2. Their intensities are normalized to the mutual glycopeptide <sup>427</sup>EYTDASFTNR<sup>436</sup>, which was present in both samples.



**Figure 15.** Comparison of glycopeptide intensity from digestion with Arg-C (panel **A**) and trypsin (panel **B**). Intensities were normalized to the mutual peptide  $^{427}\text{EYTDASFTNR}^{436}$ . EICs were generated for the protonated ion peak of the most intense charge state, for the Arg-C glycopeptide with  $z = +6$  and tryptic glycopeptide  $z = +3$ . The tryptic glycopeptide shows a significantly decreased intensity compared to the glycopeptide obtained with Arg-C digestion, presumably due to a decreased ionization efficiency of the peptide residue.

The longer glycopeptide obtained by digestion with Arg-C does not only show an increase in retention time, but also an increase in relative intensity. This was especially favorable for the detection of low abundant glycans as described further in this chapter. The most abundant charge state was +3 for the tryptic peptide and +6 for the peptide obtained by digestion with Arg-C. The increase in charge indicates that the longer peptide is more prone to protonation as it exhibits more amino acid side chains than the tryptic peptides. The increased number of positive charges and the more hydrophobic character of longer glycopeptides might be responsible for the increase in ion intensity in positive ion mode.

Next to a shift in retention time and an increase in ion intensity, a relative shift for glycopeptides with different glycan residues was observed. This was especially the case for the glycopeptides for glycosite  $^{138}\text{N}$  generated by digestion with Arg-C, where the different glycoforms were separated by their number of sialic acids. Figure 16 shows the chromatographic behavior of the tryptic glycopeptides and the glycopeptides from Arg-C digestion for glycosite  $^{138}\text{N}$ .



**Figure 16.** Extracted ion chromatograms for tryptic glycopeptides (A) and glycopeptides obtained by digestion with Arg-C (B) for glycosite  $^{138}\text{N}$ . Enhancing the peptide length improved chromatographic separation of different glycopeptides as well as ion intensity. Tetraantennary structures were only detected in the middle-down approach.

The EICs were generated for the monoisotopic mass of the most abundant charge state, which was +3 for the tryptic glycopeptides and +6 for the glycopeptides generated by digestion with Arg-C. The tryptic glycopeptides for glycosite  $^{138}\text{N}$  show overlapping peaks for all glycan residues (Figure 16, left) whereas the glycopeptides generated by digestion with Arg-C are separated by their number of sialic acids. Furthermore, increasing the peptide length by digestion with Arg-C leads to the detection of a greater number of glycopeptides. Comparing the results obtained with a limited tryptic digest to the glycosite specific analysis of human ceruloplasmin from tryptic glycopeptides reported by Harazono *et al.*, so far unreported, tetraantennary glycan structures for glycosite  $^{138}\text{N}$  were detected (see Table 3, Figure 16 and Figure 19).[84] One reason seems to be the increased ionization of glycopeptides with increased peptide length. This allowed the detection of low abundant glycopeptides, carrying fully sialylated tetraantennary glycan structures, on glycopeptides with enhanced peptide length.

Analyzing the glycopeptides of glycosite  $^{358}\text{N}$  obtained from a complete tryptic digest, only one glycopeptide was detected, carrying a non-fucosylated biantennary glycan structure (H5N4S2). Harazono *et al.* were able to detect low amounts of the fucosylated biantennary structure (H5N4S2F) as well as a

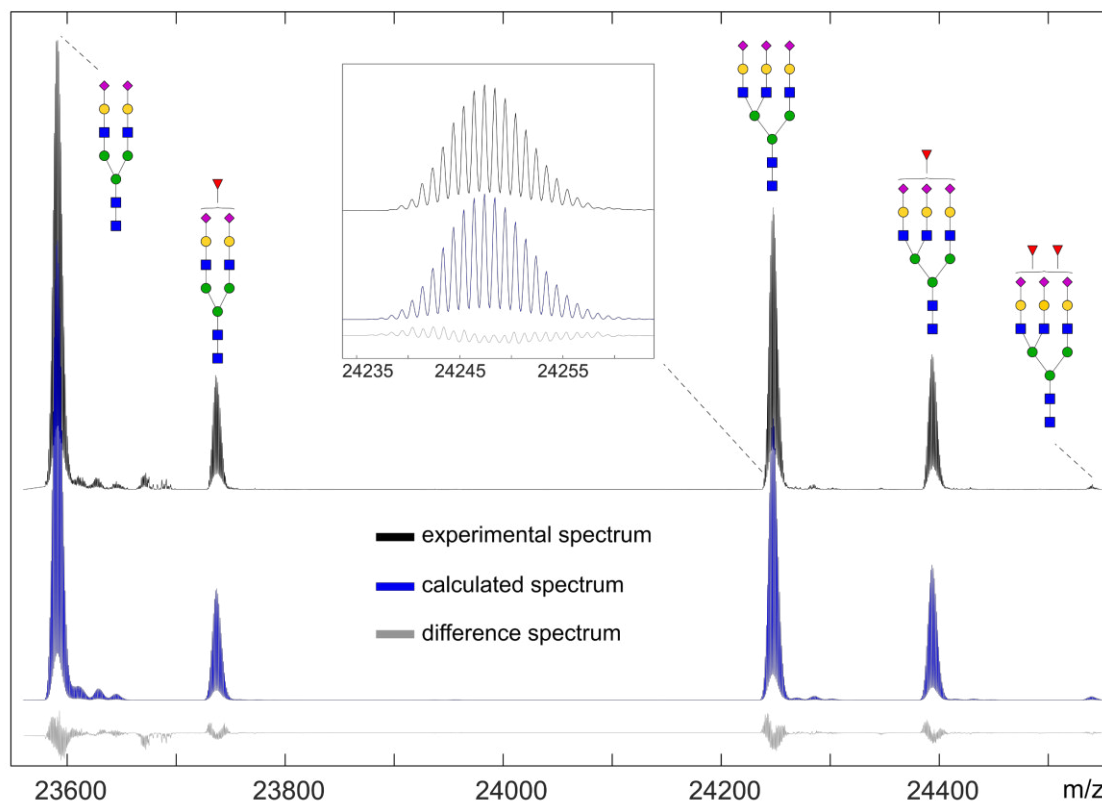


triantennary structure (H6N5S3). However, it was possible to detect more glycopeptides from digestion with Arg-C which produced glycopeptides with increased peptide length. In addition to the glycopeptides reported by Harazono *et al.*, a fucosylated form of the triantennary structure (H6N5S3F) was detected applying a middle-down approach.

For glycosite <sup>397</sup>N and <sup>762</sup>N, the same glycan structures as reported earlier were detected, although with different amounts (Figure 18).

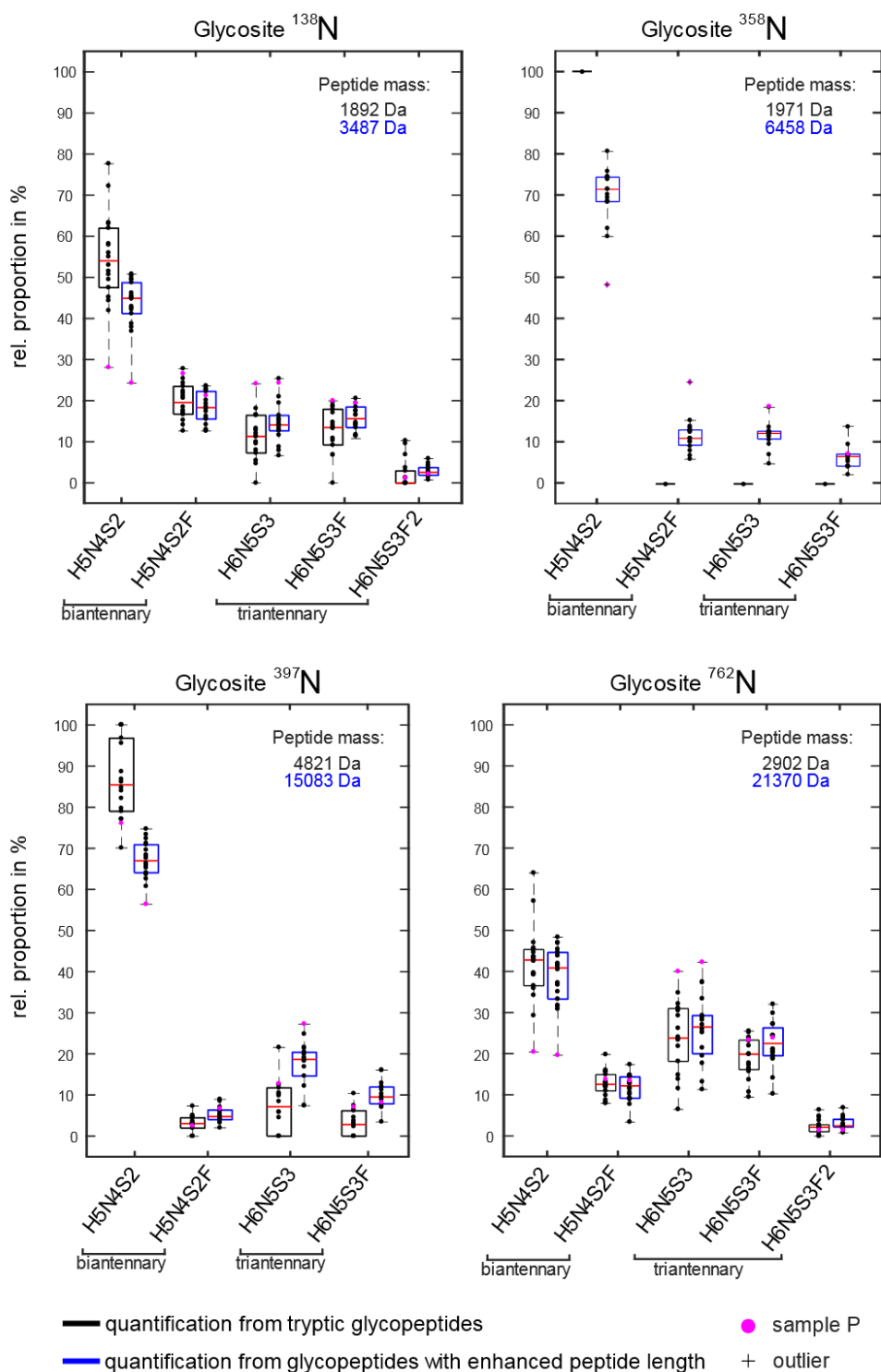
A quantification of glycopeptides for a given peptide sequence was performed by summation of all MS spectra using the peaks of all glycopeptides with the same peptide structure and thus covering a range of up to 2 min of the LC elution profile. Subsequent deconvolution with a maximum entropy algorithm yielded a spectrum with a theoretical charge of +1. This was done to include all observed charge states for glycan quantification. A calculated spectrum was fitted to the experimental data by a least squares minimization. The mass of the peptide backbone was considered constant while different glycan compositions could match the deconvoluted spectra. Possible sodium and potassium adducts were taken into account as well. By this, weighting coefficients for each glycoform were obtained, representing their relative amounts (see appendix script A1 for the Matlab based script). Normalization of all glycopeptides of a given peptide sequence to 100% yielded the percentages of the individual glycan structures that are attached to the respective peptide.

The deconvoluted spectrum (black) and the corresponding calculated spectrum (blue) for the glycopeptides obtained by limited tryptic digest for glycosite <sup>762</sup>N is shown in Figure 17. The difference spectrum is shown in gray. The expansion from 24235 to 24260 m/z demonstrates the isotopic resolution of the glycoprotein fragment. Isotopic resolution improves accurate peak assignment and data interpretation but is highly dependent on the resolving power of the mass spectrometer and the mass of the analyte. Consequently, analyzing native proteins in a top-down approach doesn't always lead to isotopically resolved spectra. This can compromise the detection of modifications like disulfide bridge formation, deamination or the distinction between one sialic acid and two fucosyl residues.[68]



**Figure 17.** Deconvoluted ESI-MS spectra for a glycopeptide generated by a limited tryptic digest for glycosite  $^{762}\text{N}$ . A calculated spectrum (blue) was fitted to the experimental spectrum (black) by a least squares minimization. This minimization yields weighting factors for each glycan composition present on the peptide backbone. A difference spectrum representing the error in the fit is shown in gray. The expansion from 24235 to 24260 m/z displays the isotopic resolution for the experimental spectrum (black) and the calculated spectrum (blue).

With a mass of about 130 kDa, ceruloplasmin represents a large glycoprotein for which a middle-down approach is an appealing alternative to a top-down analysis. Figure 18 shows the statistical analysis for the relative quantification of glycan compositions for each glycosite for the most abundant glycan compositions. The boxplot diagrams compare the results from the bottom-up approach with tryptic glycopeptides (black) and the middle-down approach (blue) for seventeen plasma samples from different individuals. A complete list of all found glycan structures for all seventeen samples can be found in the appendix (*10.4. Relative Amounts of Glycopeptides from Ceruloplasmin*).



**Figure 18.** Glycosite specific quantification for the most abundant glycan compositions from seventeen different plasma samples. Results from analysis using a bottom-up approach with tryptic glycopeptides are shown in black. Quantification utilizing a middle-down approach (either digestion with Arg-C or limited tryptic cleavage) is shown in blue. In all cases, the abundance of triantennary glycans is increased with increased peptide length. Values for sample P are marked in magenta.[83]

The boxplot diagrams show the variation of the detected glycan composition from seventeen different plasma samples. Using a bottom-up approach with tryptic glycopeptides resulted in an increased variation for most cases. It is also remarkable that triantennary structures are found in greater abundance using a middle-down approach. Accordingly, biantennary glycans are less abundant. One sample, plasma sample P, showed for all four glycosites decreased values for the glycan composition H5N4S2 and increased values for H6N5S3 (marked in magenta in Figure 18, not shown in [83]). The higher number of antennae leads to an increased amount of sialylation compared to the average values over 17 samples. For glycosite  $^{358}\text{N}$ , values for sample P are marked as outliers for the diantennary glycan compositions.

Table 4 summarizes the average values for the abundance of each glycan composition over all analyzed plasma samples.

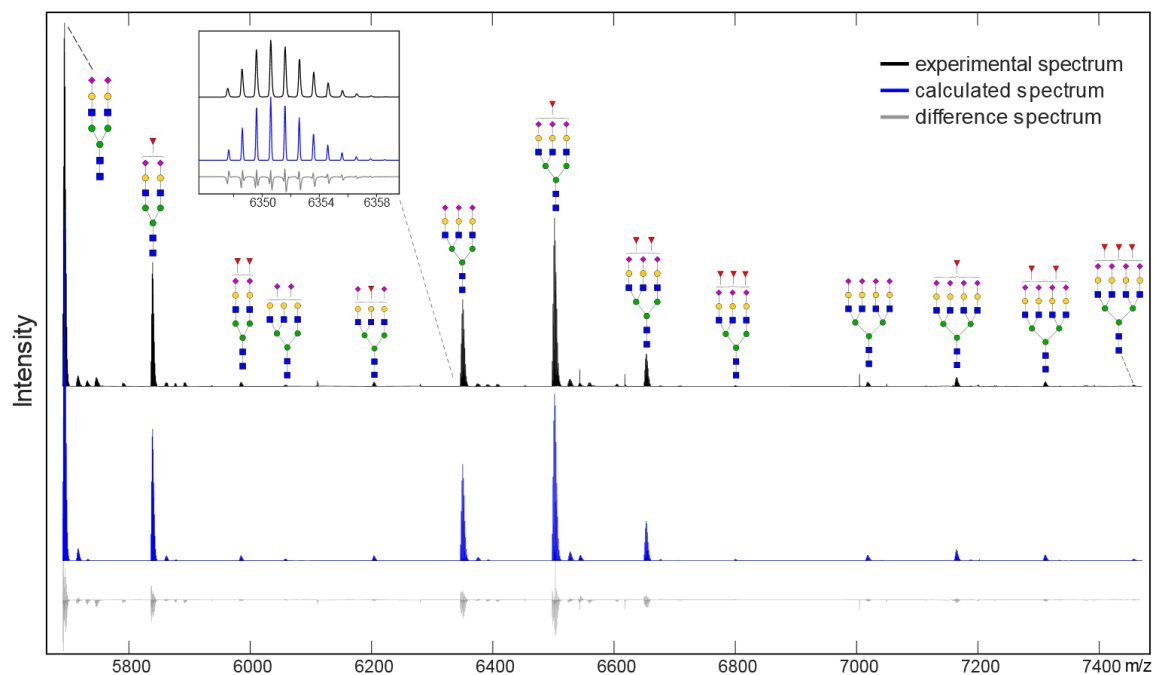
**Table 4.** Mean values for glycosite specific quantification over seventeen plasma samples, using a tryptic bottom-up and a middle-down approach.

Glycan Composition	Site $^{138}\text{N}$		Site $^{358}\text{N}$		Site $^{397}\text{N}$		Site $^{762}\text{N}$	
	Bottom up	Middle down	Bottom up	Middle down	Bottom up	Middle down	Bottom up	Middle down
H5N4S2	53.8	43.1	100	69.5	86.4	67.6	42.2	38.6
H5N4NS2F	20.2	18.4	-	11.7	2.8	5.2	13.0	11.5
H5N4S2F2	<u>0.2</u>	<u>0.4</u>	-	-	-	-	-	-
H6N5S3	11.1	15.0	-	12.1	7.4	17.6	23.1	25.3
H6N5S3F	12.7	15.8	-	<u>6.7</u>	3.4	9.8	18.9	22.1
H6N5S3F2	2.1	2.9	-	-	-	-	2.3	2.6
H6N5S2	-	<u>0.9</u>	-	-	-	-	-	-
H6N5S2F	-	<u>0.9</u>	-	-	-	-	-	-
H7N6S4	-	<u>1.0</u>	-	-	-	-	-	-
H7N6S4F	-	<u>1.0</u>	-	-	-	-	-	-
H7N6S4F2	-	<u>0.3</u>	-	-	-	-	-	-
H7N6S4F3	-	<u>minor</u>	-	-	-	-	-	-

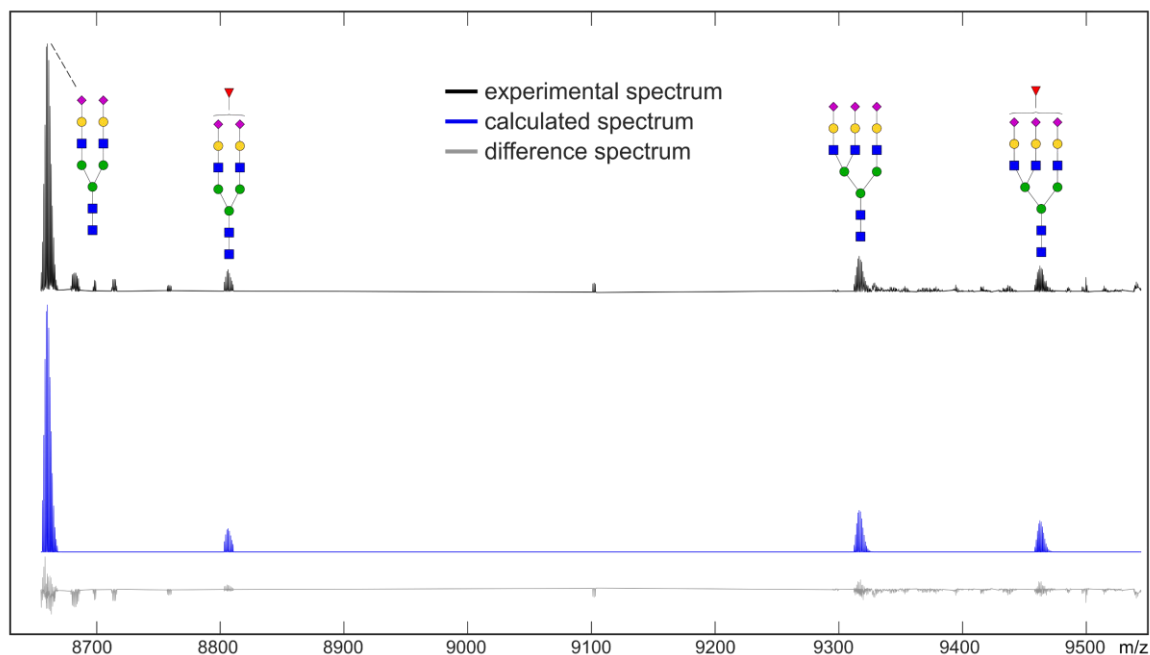
Minor glycan compositions were detected with less than 0.1%. Underlined percentages correspond to previously unreported glycan compositions.

The increase in amount of higher antennary glycan structures on glycopeptides with increased peptide length did not only lead to different results for quantification

but also to the detection of so far unreported glycan structures. It was possible to detect tetraantennary glycan structures for glycosite  $^{138}\text{N}$  in small amounts (see Figure 19 for an exemplary MS spectrum) as well as a fucosylated triantennary structure for glycosite  $^{358}\text{N}$  with an average abundance of 6.7% (see Figure 20).



**Figure 19.** Deconvoluted spectrum of glycopeptides obtained by digest with Arg-C for glycosite  $^{138}\text{N}$ . Experimental spectrum is shown in black, calculated spectrum in blue and difference spectrum in gray. Previously unreported tetraantennary structures with different amounts of fucosyl residues were detected only after digest with Arg-C but not for tryptic peptides.



**Figure 20.** Deconvoluted spectrum of glycopeptides obtained by digest with Arg-C for glycosite  $^{358}\text{N}$ . Experimental spectrum is shown in black, calculated spectrum in blue and difference spectrum in gray. A previously unreported fucosylated triantennary composition is detected in small amounts.

Glycosite  $^{358}\text{N}$  shows the biggest differences in glycan quantification from glycopeptides comparing short and longer peptide parts. Even though the most abundant charge increases only from +3 for the tryptic peptide (2A2N) to +5 for the glycopeptide generated by Arg-C digest, the shift in retention time (from 30.3 min for the tryptic glycopeptide to 36.2 min for the Arg-C glycopeptide) is remarkable. Glycopeptides with higher antennary glycans are more easily detected if the number of charges and potentially the hydrophobicity increases.

For glycosite  $^{762}\text{N}$ , the tryptic glycopeptides are already observed at a charge state of +5 (most intense charge state) and the GRAVY index for both, tryptic peptide and peptide obtained by limited digestion, are comparable. Accordingly, quantification for glycosite  $^{762}\text{N}$  shows a small variation between the bottom-up approach with tryptic glycopeptides and the middle-down approach with limited tryptic cleavage.

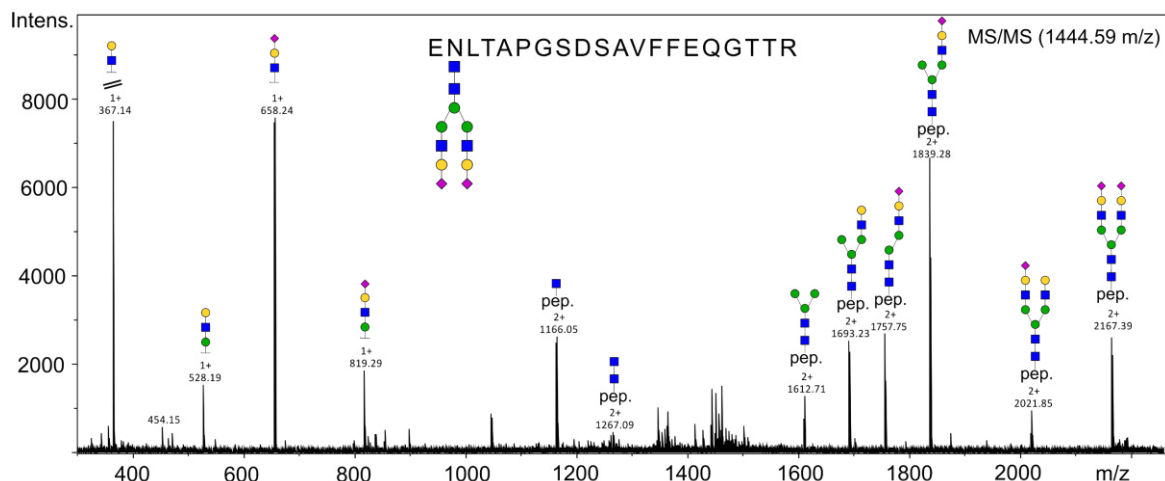
The tryptic peptides for glycosite  $^{397}\text{N}$  are even more hydrophobic than the peptides obtained by a middle-down approach as indicated by GRAVY index and retention time. Still, triantennary glycans are observed in greater abundance in the longer

glycopeptides (cf. Table 4). This indicates that the number of charges on the glycopeptide or the ease of protonation is important for the observation of highly sialylated glycan compositions.

Utilizing a bottom-up approach with a full tryptic digest showed a larger variation of the quantities of each glycopeptides species and accordingly of the glycan structures than the limited digest that produces larger peptides. Also, the bottom-up approach demonstrated significant weaknesses in the detection of less abundant, highly sialylated glycan structures.

#### **5.2.2.1. Fragmentation of Large Glycopeptides**

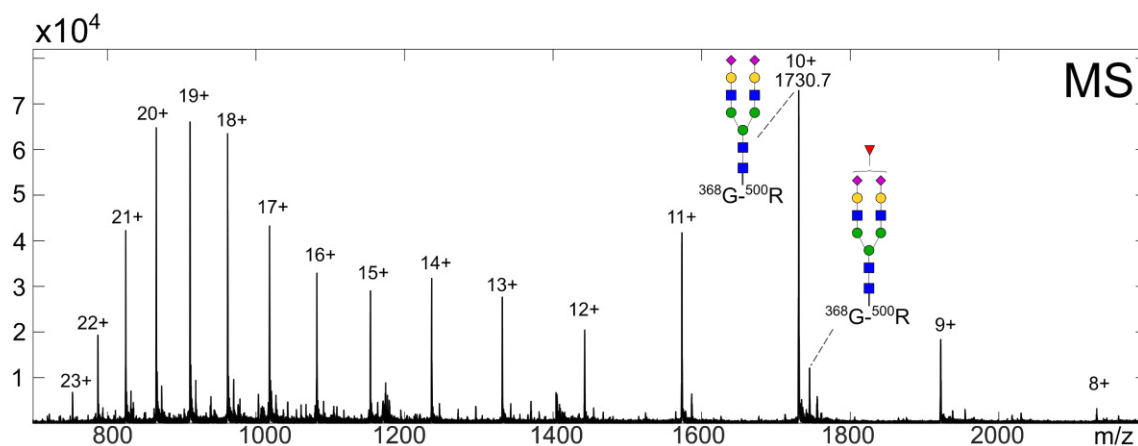
Tandem mass spectrometry has proven to be a powerful tool for the identification of peptides and glycopeptides. Fragmentation experiments of glycopeptides can not only confirm the amino acid sequence but also provide information about post-translational modifications (PTMs). Fragmentation techniques like electron capture dissociation (ECD) retain post-translational modifications, whereas collision-induced dissociation (CID) of glycopeptides yields a fragmentation of the peptide backbone and of the glycan residue.[104] Figure 21 shows the CID fragment spectrum of tryptic glycopeptide obtained at glycosite <sup>397</sup>N. The precursor ion was triply charged with a m/z value of 1444.59. Intense signals for fragmentation of the glycan residue are observed, either for oxonium ions of the glycan at lower m/z values or for the glycan with peptide residue at higher m/z values as doubly charged ions. Fragmentation of the peptide backbone is not observed.



**Figure 21.** CID fragment spectrum of tryptic glycopeptide at glycosite <sup>397</sup>N from the triply charged precursor  $m/z = 1444.59$ . Fragment ions show fragmentation of the glycan either with peptide residue at higher  $m/z$  values as doubly charged ions or for oxonium ions of the glycan at lower  $m/z$  values.

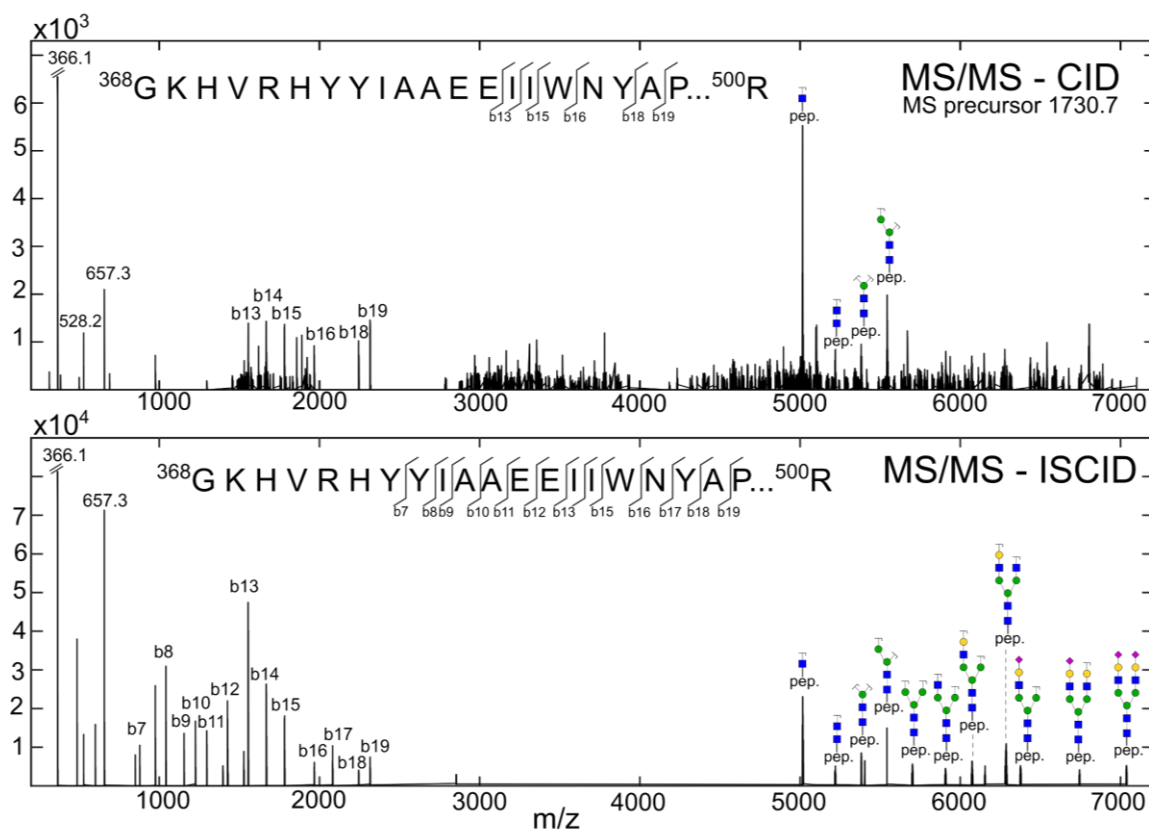
Even though CID fragmentation shows good results for doubly or triply charged glycopeptides as shown in Figure 21 it is not applicable for sequencing larger peptides or proteins from medium-length peptides.[58, 59] The generation of large polypeptides or glycopeptides in a middle-down approach leads to analytes with a higher number of charges displaying a charge envelope. Figure 22 shows the ESI MS spectrum for the glycopeptide from <sup>368</sup>G to <sup>500</sup>R with glycosylation site at <sup>397</sup>N. The most abundant glycopeptide corresponds to glycan moiety H5N4S2. The charge states show a multimodal distribution with maxima at  $z = +10$  and  $z = +19$ , which indicates that several denatured states of the protein are present and that the states differ in accessibility of residues that can be protonated. Different states of this glycopeptide could be caused by heat denaturation of the sample used to inactivate trypsin.





**Figure 22.** Charge envelope spectra for the glycopeptide from  $^{368}\text{G}$  to  $^{500}\text{R}$  with glycosylation site at  $^{397}\text{N}$  showing the most abundant glycan moiety as H5N4S2. Due to insufficient chromatographic separation, other glycan moieties are observed as well (e.g. H5N4S2F). Charge states from +8 to +23 are observed. The spectrum was generated by summation over the LC peak with a retention time of 31.7 min in a time range of  $\pm 0.5$  min.

Fragmentation patterns similar to CID are obtained with in-source CID (ISCID) fragmentation where fragment ions are generated in the intermediate pressure transmission region between the atmospheric pressure region and skimmer.[105, 106] The application of a cone voltage within the ESI source leads to fragmentation without prior precursor selection. Nevertheless, applying ISCID showed better results for the fragmentation of larger glycopeptides with increased peptide length than CID fragmentation. Figure 23 shows the deconvoluted MS/MS spectra from CID fragmentation and ISCID fragmentation for the glycopeptide from  $^{368}\text{G}$ - $^{500}\text{R}$  with glycosylation site  $^{397}\text{N}$ , respectively.



**Figure 23.** Deconvoluted fragmentation spectra for the glycopeptide from  $^{368}\text{G}$  to  $^{500}\text{R}$  with glycosylation site at  $^{397}\text{N}$ . The top panel shows the MS/MS spectra from CID fragmentation. Precursor ion for CID fragmentation was  $m/z = 1730.7$  with a charge state of +10. ISCID fragmentation spectrum without precursor selection is shown at the bottom. The spectrum at the bottom has an intensity which is 10 times higher than the MS/MS spectrum at the top.

In both, CID and ISCID fragment spectra, b-type peptide fragments N-terminal to the peptide backbone are observed. Next to characteristic glycan fragments at  $m/z = 366.1$  (HN), 528.2 (H<sub>2</sub>N) and 657.3 (HNS), glycan fragmentation occurred with a peptide residue ranging from  $^{391}\text{D}$  to  $^{433}\text{F}$  at higher masses. The ISCID spectra showed a significantly higher intensity and therefore an increased sequence coverage for the peptide residue. Fragment ions of the glycan moiety of the intermediate glycopeptide  $^{391}\text{D}$  to  $^{433}\text{F}$  were more abundant with ISCID fragmentation as well. This might favor structural identification of the glycan moiety, providing information not only about glycan composition but also about linkage position. A crucial difference between the two fragmentation techniques is the selection of a precursor ion for CID fragmentation. The precursor selection leads to the fragmentation of a single compound without interference of fragment ions

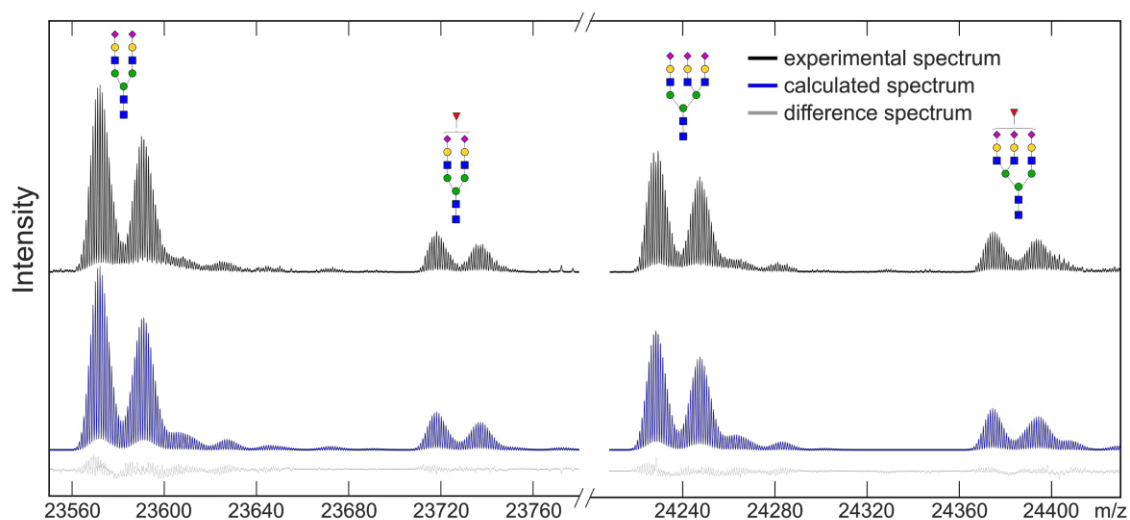
from other compounds with similar elution behavior. However, not selecting a precursor ion but performing a fragmentation over all observed charge states is beneficial for fragment ion intensity. This can enhance sequence coverage and identification of glycan moiety when analyzing large glycopeptides or glycoproteins. Recent advances in data-independent acquisition for glycopeptide identification already demonstrated the advantages of not selecting a precursor ion but simultaneous fragmentation over a broad  $m/z$  range.[107-109] With sophisticated algorithms used for data analysis, an unambiguous assignment of distinct fragment ions to a precursor is possible even without precursor selection.[110-112] Even though assignment of different glycan compositions to coeluting glycopeptides will be challenging, the use of ISCID fragmentation as an untargeted MS/MS acquisition seems to be an excellent method for fragmentation of multiply charged, big peptides and glycopeptides.

### **5.2.3. Analysis of Genetic Variants of Ceruloplasmin**

Two different genetic variants, or SNPs, were detected in one sample. One SNP is described by an amino acid change at position R793H, leading to a mass shift of -19 Da. The other identified SNP, D544E, leads to a mass shift of +14 Da. Both SNPs showed a heterozygous expression and could have been easily overlooked using only a bottom-up approach with automatic peptide identification. The detection of genetic variants is nevertheless an essential task since variant polymorphism is correlated to a variety of diseases. For ceruloplasmin, a connection between SNP variants and development of Parkinson's disease is hypothesized.[86, 113]

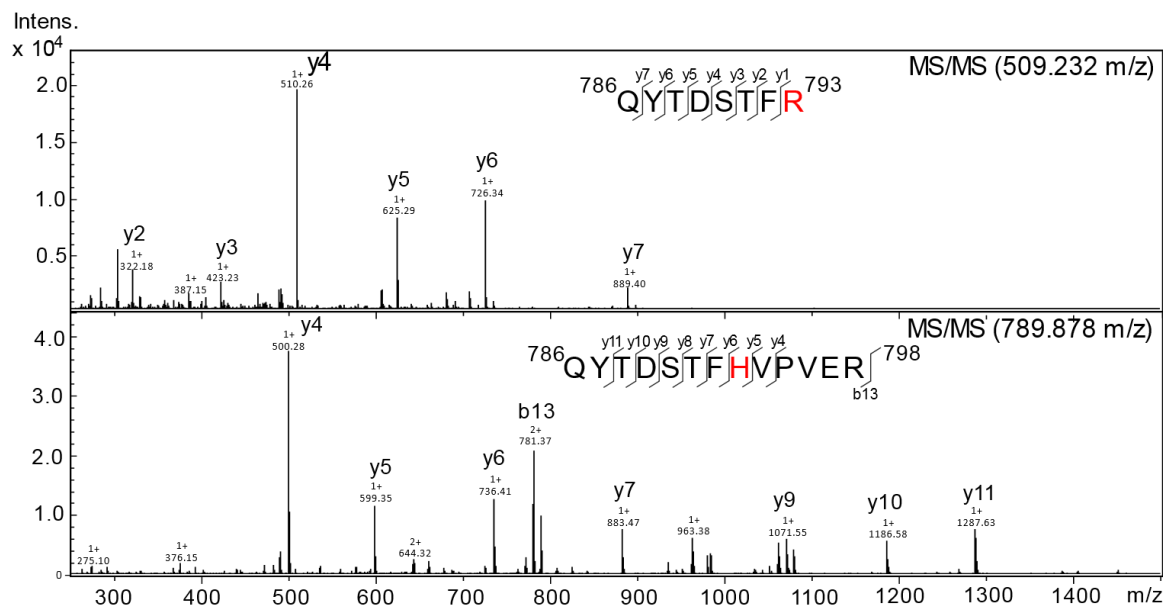
The SNP variant R793H was first observed after a limited tryptic digest as described in the previous section. After deconvolution of the protein fragment with a mass of about 24 kDa, including glycosylation site  $^{762}\text{N}$ , a double peak with a mass difference of 19 Da was observed for each proteoform. This indicated the presence of the SNP R793H. Surprisingly, the native variant and the SNP variant

did not occur in equal amounts which might originate in biological factors like different expression rates or clearance rates. Equally conceivable are reasons like reduced or increased protease activity due to steric alteration of the protein by the amino acid exchange and different ionization properties. Even though the amounts of genetic variants differed, no significant changes in glycosylation at glycosite <sup>762</sup>N were detected for the different variants as shown in Figure 24. However, changes in glycosylation may occur on one of the three other glycosylation sites.



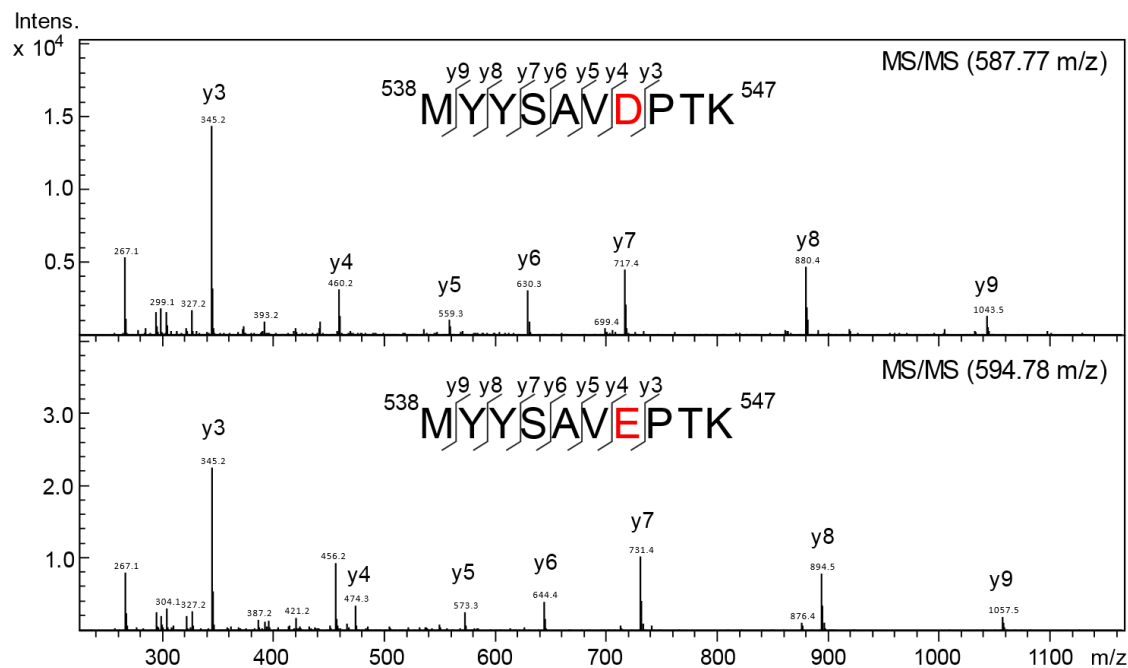
**Figure 24.** Deconvoluted spectrum for protein fragment <sup>639</sup>Gly to <sup>821</sup>Lys including one glycosite at <sup>762</sup>N after limited tryptic digest. Measured spectrum is shown in black, calculated spectrum in blue. The spectrum shows two signals for each of the four identified glycans due to an amino acid exchange at R793H. Hence, the donor carries a point mutation heterozygote. Glycosylation is not affected.

The SNP R793H was also confirmed on peptide level after a tryptic digest. Even though automated peptide identification based on MS/MS compositions using the Mascot search engine did not identify the peptide with the amino acid exchange, manual interpretation showed distinct fragmentation pattern for the peptide containing the amino acid exchange. The exchange from arginine to histidine leads to the loss of a tryptic cleavage site, forming a longer tryptic peptide. Figure 25 shows the MS/MS spectra of both peptides after CID fragmentation.



**Figure 25.** MS/MS CID spectra of tryptic peptides of human ceruloplasmin, confirming SNP R793H. Precursor ions are in a charge state of +2. The upper spectrum shows the fragment spectrum of the peptide without amino acid exchange. The exchange from arginine to histidine leads to the loss of a tryptic cleavage site. Therefore, a longer peptide is observed for the SNP peptide (bottom spectrum). Peaks are annotated with mass and charge state.

The second SNP, also detected heterozygous in the same sample, leads to the amino acid exchange D544E with a mass shift of +14 Da. Position 544 is not close to any glycosylation site, therefore no glycopeptides including glycosylation were found neither with tryptic peptides nor with peptides obtained by Arg-C digestion. The MS/MS fragment spectra, confirming the heterozygous expression of SNP D544E, are shown in Figure 26.



**Figure 26.** MS/MS CID spectra of tryptic peptides from <sup>538</sup>M to <sup>547</sup>K of human ceruloplasmin, confirming SNP D544E. The upper spectrum shows the fragment spectrum of the peptide without amino acid exchange. The bottom spectrum shows the MS/MS spectrum of the peptide with the amino acid exchange. Peaks are annotated with masses. All peaks show a charge state of +1.

As mentioned above, ceruloplasmin reduces the amount of highly toxic hydroxyl radicals by oxidation of iron(II) to iron(III). However, its ferroxidase activity is impaired when certain SNPs are expressed.[86] The accumulation of iron(II) is especially harmful to the brain, causing neurodegenerative diseases like Parkinson's disease. Ayton *et al.* found an 80% loss of ferroxidase activity of ceruloplasmin in the substantia nigra of patients with Parkinson's disease.[114]

As demonstrated, mass spectrometry represents a powerful tool for the detection of SNPs of ceruloplasmin and their connection to neurodegenerative diseases. The results show that a middle-down approach is favorable for the detection of SNP variants compared to a bottom-up approach with automated peptide search.

### 5.3. Glycosylation of Human Immunoglobulin IgG1

Results from this chapter were obtained during the practical training of M. Moritz.

Immunoglobulin IgG has a high serum concentration of 11 mg/mL in healthy individuals, making it the most abundant antibody with a molecular mass of roughly 150 kDa.[115] IgG antibodies are composed of two heavy and two light chains which are linked together via disulfide bridges. The variable loop in the Fab (fragment antigen binding) domain is responsible for antigen binding and can, therefore, mediate antigen neutralization.[116] About 15 to 20% of IgG antibodies carry one or more glycosylation sites in the Fab domain, resulting from somatic hypermutation during an antigen-specific immune response.[117] The Fc (fragment crystallizable) domain shows a conserved N-glycosylation site at <sup>297</sup>N on the heavy chain, observed in all IgG antibodies. Glycosylation of antibodies is not only crucial for stability, conformation and solubility but also for effector functions like pro- and anti-inflammatory activity and therefore mandatory for immune response.[118] Changes in IgG glycosylation of the Fc domain have been observed under several physiological conditions. This includes aging, inflammation, autoimmune diseases, several cancer diseases, and Parkinson's disease.[119-121] Hence, IgG glycosylation is studied as a potential biomarker for several diseases.[122-127]

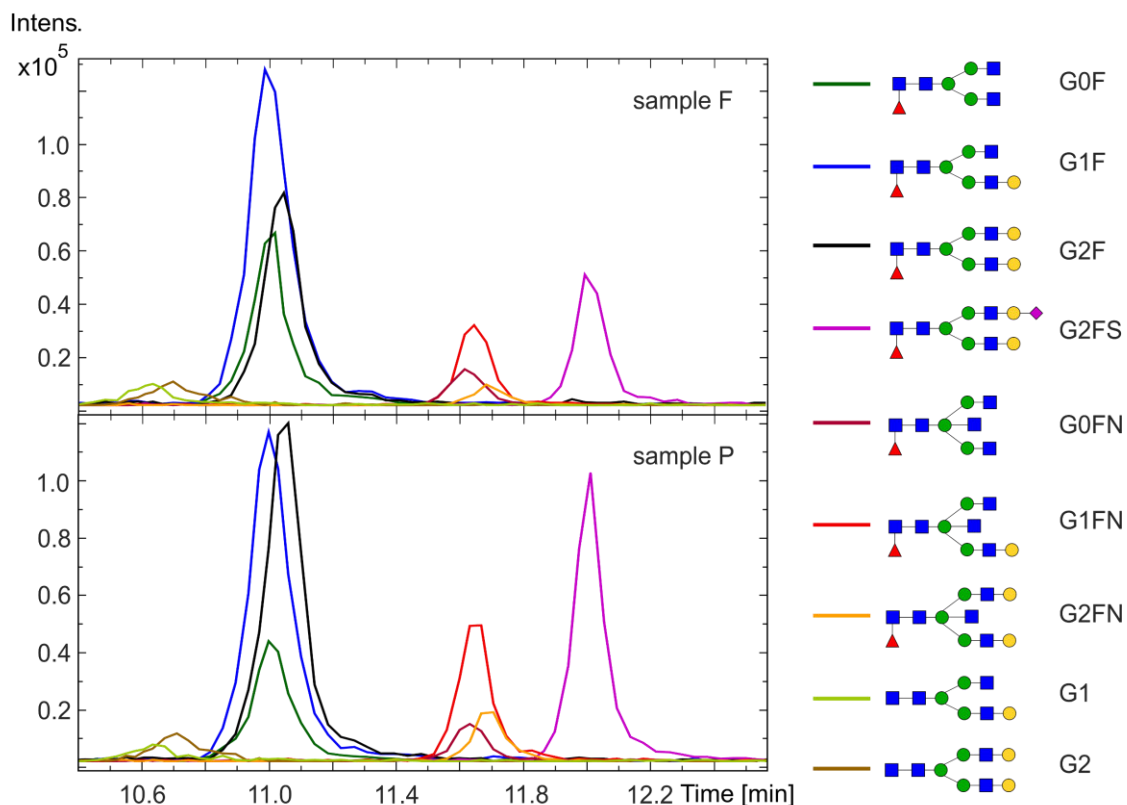
In this work IgG1, the most abundant subclass of IgG antibodies, was purified from seven plasma samples and digested with trypsin to analyze glycosylation on glycopeptide level in a bottom-up approach.

#### 5.3.1. Relative Quantification of Glycopeptides from IgG1

IgG1 was purified from 7 plasma samples using *Protein A Spin Trap* columns. Protein A shows a high affinity for the Fc region of IgG1, IgG2, and IgG4. IgG1 was not isolated from other subclasses since its natural concentration is much higher than other subclasses. Additionally, the peptide residue from glycopeptides allows an unambiguous assignment of glycan structures to specific IgG subclasses (except for subclass IgG2 and IgG3). After purification, the samples were digested

using trypsin and peptides and glycopeptides were analyzed by LC-ESI-MS in positive ion mode using a C18 RP column. The tryptic glycopeptides EEQYNSTYR with glycosite <sup>297</sup>N were detected in the charge states  $z = +2$  and  $+3$ . However, more intense peaks were observed for the tryptic glycopeptides TKPREEQYNSTYR which showed two missed cleavages and a slightly different retention time. An incomplete tryptic digest, as well as better ionization properties for the longer glycopeptide with two missed cleavages, might explain the increased intensity. The increased ion intensity of longer glycopeptides was also observed for glycopeptides from ceruloplasmin described in section 5.2.3. *Quantification of Glycopeptides with Different Peptide Length* and shown in Figure 16. Extracted ion chromatograms were created for the tryptic glycopeptides without missed cleavages for the charge states  $z = +2$  and  $+3$  and for the glycopeptides with two missed cleavages with a charge state of  $z = +4$ . The EICs were created for the monoisotopic peak with a mass difference of  $\pm 0.05$  m/z. Figure 27 shows the EICs for tryptic glycopeptides TKPREEQYNSTYR with two missed cleavages for plasma sample F and P. Sample P shows a higher amount of glycosylation which is expressed by increased intensities of G2F and G2FN. Also, the amount of sialylation is increased which is seen in structure G2FS. Sample F shows relative abundances of different glycan structures which are similar to the abundances found in the other five samples.





**Figure 27.** Extracted ion chromatograms for tryptic glycopeptide TKPREEQYNSTYR with glycosite  $^{297}\text{N}$  using a C18 RP column coupled to online ESI-MS detection. Relative amounts of different glycostructures were similar for six of the seven samples and are represented by sample F (top). Sample P (bottom) showed elevated levels of galactosylation seen in structure G2F or G2FN as well as sialylation seen in structure G2FS.

It is not surprising that galactosylation and sialylation are both increased in sample P since sialylation requires galactosylation. For IgG, increased levels of sialylation in the Fc domain are correlated with anti-inflammatory effects by reducing the antibody dependent cell-mediated cytotoxicity.[128, 129] However, increased levels of sialylation for sample P were also observed in the analysis of ceruloplasmin (see e.g. Figure 18). This indicates a general increase of glycosyltransferases like N-acetylglucosaminyltransferases, galactosyltransferases (IV and/or V) and sialyltransferases.

Since only seven plasma samples were analyzed, average values for glycan quantification were compared with previously published results. However, results from glycan quantification from glycopeptides are dependent on many parameters like instrumentational setup and on peptide length as shown for ceruloplasmin. Therefore, glycan quantification was done for different glycopeptides (EEQYNSTYR and TKPREEQYNSTYR) and for different charge states ( $z = +2$ ,  $z = +3$  and  $z = +4$ ). The EICs were generated for the monoisotopic peak and integrated in a retention time window of one minute using the Matlab command trapz (trapezoidal numerical integration). The sum of all values was set to 100%. Table 5 to Table 7 show the respective quantification results for glycopeptide EEQYNSTYR with a charge of +2 and +3 and for glycopeptide TKPREEQYNSTYR with a charge state of +4.

**Table 5.** Results for glycan quantification from glycopeptide EEQYNSTYR in a charge state of  $z = +2$ . Values are given in %.

Donor	G0F	G1F	G2F	G2FS	G0FN	G1FN	G2FN	G0	G1	G2
F	24.4	36.4	16.6	5.4	3.7	5.5	2.2	n.d.	3.1	2.7
H	19.8	33.7	19.9	7.5	5.3	6.8	1.9	n.d.	3.0	2.1
J	18.6	33.9	22.5	8.1	4.0	7.6	1.9	n.d.	1.9	1.5
N	24.8	34.4	14.7	5.4	5.4	6.9	2.6	n.d.	4.0	1.8
P	15.8	31.9	23.0	8.0	4.4	8.8	3.5	n.d.	2.1	2.5
AA	27.1	39.6	15.2	3.5	4.2	5.7	1.4	n.d.	1.9	1.5
AC	18.5	36.3	22.4	7.2	3.6	6.7	1.5	n.d.	2.3	1.3
<b>Average<sup>a</sup></b>	<b>21.3</b>	<b>35.2</b>	<b>19.2</b>	<b>6.4</b>	<b>4.4</b>	<b>6.9</b>	<b>2.2</b>	<b>n.d.</b>	<b>2.6</b>	<b>1.9</b>

<sup>a</sup> Average values are calculated without sample P.

**Table 6.** Results for glycan quantification from glycopeptide EEQYNSTYR in a charge state of  $z = +3$ . Values are given in %.

Donor	G0F	G1F	G2F	G2FS	G0FN	G1FN	G2FN	G0	G1	G2
F	14.3	34.3	20.2	11.9	4.3	7.6	3.1	n.d.	2.3	2.1
H	14.6	31.5	21.7	12.8	4.6	8.1	4.3	n.d.	1.3	1.1
J	12.0	30.7	23.4	14.1	3.8	9.6	3.1	0.5	1.4	1.3
N	14.8	32.2	19.8	11.5	4.9	9.4	3.4	n.d.	2.3	1.7
P	7.8	27.1	24.8	17.4	4.2	10.2	4.7	n.d.	2.0	1.8
AA	18.3	35.1	18.5	8.4	5.2	9.0	2.5	n.d.	1.8	1.2
AC	11.4	31.7	24.6	13.3	3.9	9.0	2.9	0.4	1.6	1.2
<b>Average<sup>a</sup></b>	<b>13.3</b>	<b>31.8</b>	<b>21.9</b>	<b>12.8</b>	<b>4.4</b>	<b>9.0</b>	<b>3.4</b>	<b>0.1</b>	<b>1.8</b>	<b>1.5</b>

<sup>a</sup> Average values are calculated without sample P.

**Table 7.** Results for glycan quantification from glycopeptide TKPREEQYNSTYR (two missed cleavages) in a charge state of  $z = +4$ . Values are given in %.

Donor	G0F	G1F	G2F	G2FS	G0FN	G1FN	G2FN	G0	G1	G2
F	15.8	35.1	20.8	10.1	3.1	6.7	1.7	0.9	2.8	3.0
H	13.0	32.0	23.4	11.9	4.3	9.6	2.3	n.d.	1.6	1.9
J	13.2	30.9	23.9	11.9	4.0	9.0	2.7	0.6	2.0	1.7
N	16.2	33.3	20.4	10.5	4.4	8.4	2.2	1.0	2.1	1.6
P	9.5	27.4	26.5	17.0	2.8	9.3	3.3	n.d.	1.8	2.5
AA	18.1	36.0	18.2	7.5	4.4	9.5	1.9	0.8	2.1	1.6
AC	12.6	33.5	23.8	12.0	3.4	9.0	2.5	0.5	1.5	1.2
<b>Average<sup>a</sup></b>	<b>14.8</b>	<b>33.5</b>	<b>21.7</b>	<b>10.7</b>	<b>3.9</b>	<b>8.7</b>	<b>2.2</b>	<b>0.7</b>	<b>2.0</b>	<b>1.8</b>

<sup>a</sup> Average values are calculated without sample P.

Quantification results from glycopeptides with a charge of +2 show an average degree of sialylation (G2FS) of 6.4% and of 8.0% for sample P. Results from glycopeptides with a charge of +3 and +4 however, show a much higher abundance for structure G2FS. The average value of G2FS for charge state +3 and +4 are 12.8% and 10.7% and for sample P 17.4% and 17.0% respectively. As already discussed for ceruloplasmin, glycan quantification from glycopeptides is highly dependent on the peptide part and as demonstrated here, also on the charge state. This can impair an accurate quantification and should be considered when analyzing glycans from glycopeptides. However, analyzing glycans from glycopeptides is a commonly used approach and has been published for Fc glycans of human IgG1. In 2017, Plomp *et al.* analyzed the Fc glycosylation of IgG subclasses with a sample cohort of 1826 individuals.[119] Glycan quantification was done for the glycopeptide EEQYNSTYR, observed in a charge state of +2 and +3 by nanoLC-ESI-qTOF-MS. Quantification was based on the average values of the intensity for the first three isotopic peaks for each charge state. Table 8 compares the quantification results from Plomp *et al.* with the results obtained for glycopeptides with a charge of +2, +3 and +4 as described above. Results do not include values from sample P. Degree of fucosylation, galactosylation, sialylation and bisecting structures were calculated according to Plomp *et al.*[119]

**Table 8.** Quantification results from Plomp *et al.* compared to results obtained for glycopeptides with a charge of +2, +3 and +4. Glycopeptides with a charge of +4 were observed with two missed cleavages.

	Plomp <i>et al.</i> [119] Peptide charge z = +2 and z = +3	Peptide charge z = +2	Peptide charge z = +3	Peptide charge z = +4
Fucosylation	91.2 ± 4.1%	95.5%	96.6%	95.5%
Galactosylation <sup>a</sup>	50.1 ± 6.4%	52.1%	60.6%	64.5%
Sialylation	6.8 ± 1.4%	6.4%	12.8%	10.7%
Bisecting	19.1 ± 3.3%	13.5%	16.8%	14.8%

<sup>a</sup> Percentage of galactosylation is calculated for the number of galactose units per antennae.

Despite the low number of samples, the results obtained from six plasma samples are in general in good agreement with the results published by Plomp *et al.* with a large-scale cohort of 1825 individuals for subclass IgG1. However, increased levels of sialylation were observed for glycopeptides with a charge state of +3 and +4. This is also the reason why values for galactosylation are increased as structure G2FS is also contributing to the amount of galactosylation. Negatively charge sialic acids might hamper ionization in low charge states in positive ion mode which would explain an increased relative abundance in higher charge states and on longer peptides with a higher number of protonated amino acid side chains. As also concluded for ceruloplasmin (see 5.2.3. *Quantification of Glycopeptides with Different Peptide Length*), this effect should be considered for the relative quantification of glycans from glycopeptides. Results from charge state +3 and +4 (with missed cleavages) are in good agreement and might represent a more accurate determination of relative quantification, especially for the amount of sialylation.

## 5.4. Analysis of Human Serum Transferrin

Transferrin is a glycoprotein with a mass of 80 kDa and two N-glycosylation sites, occupied with dominantly biantennary, fully sialylated complex type glycan structures. It exhibits two binding sites for Fe(III) ions as it plays the key role in the intercellular transport of iron.[130] Since transferrin is mainly produced in liver cells, changes in transferrin concentration as well as changes in glycosylation have been detected in patients with liver-related diseases like hepatocellular carcinoma.[131] Changes in glycosylation of serum transferrin are used as a diagnostic marker for long-term alcohol abuse where the concentration of carbohydrate deficient transferrin (CDT) is identified.[132] However, changes in glycosylation of transferrin have also been observed during pregnancy which may lead to false-positive results in testing for alcohol abuse.[12] But other clinical markers for e.g. colorectal cancer or pancreatic cancer based on transferrin are being developed as well.[133, 134] Furthermore, the glycosylation of transferrin can be used not only to diagnose congenital disorders of glycosylation (CDG) but also to differentiate between different CDG types.[135, 136]

The correlation of transferrin with many different physiological conditions like cancer diseases or pregnancy implies that an unambiguous method is needed to differentiate between structural changes in glycosylation and other modifications. Mass spectrometry presents a method to detect all present proteoforms which is useful in the differentiation between certain diseases. Therefore, transferrin has been analyzed by mass spectrometry extensively and presents a well-characterized protein including modifications like glycosylation (see Table 9).

### 5.4.1. Analysis of Intact Transferrin by ESI-MS

To determine different proteoforms of transferrin, the intact glycoprotein was analyzed by ESI-qTOF-MS as it has been described before in literature (see Table 9). However, analyzing transferrin from the same plasma samples as used for the analysis of alpha-acid-glycoprotein, ceruloplasmin,  $\beta$ -2-glycoprotein 1, and immunoglobulin IgG1 allows the analysis of glycosylation patterns among serum

proteins from the same individuals which is discussed in section 5.7. *Glycosylation of Plasma Proteins as Potential Biomarker.*

The dominant proteoform of transferrin in healthy individuals is a proteoform with the attachment of two biantennary, fully sialylated N-glycans (two times H5N4S2). The amino acid sequence consists of 38 cysteine residues which form 19 disulfide bridges. Denaturation of transferrin with DTT showed that only 15 disulfide bridges are susceptible for reduction.[137, 138] Based on the amino acid sequence (UniProtKB P02787, reviewed by Swiss-Prot), the chemical formula for two N-glycans (two times H5N4S2) and the presence of 19 disulfide bridges, a theoretical, most abundant mass of 79567.9 Da is obtained. However, slightly different masses have been published for the analysis of intact human serum transferrin after deconvolution. Table 9 shows a few examples of published masses for the dominant proteoform of human transferrin.

**Table 9.** Measured mass of the dominant proteoform of human transferrin (glycosylated with two times H5N4S2) from different publications and this work.

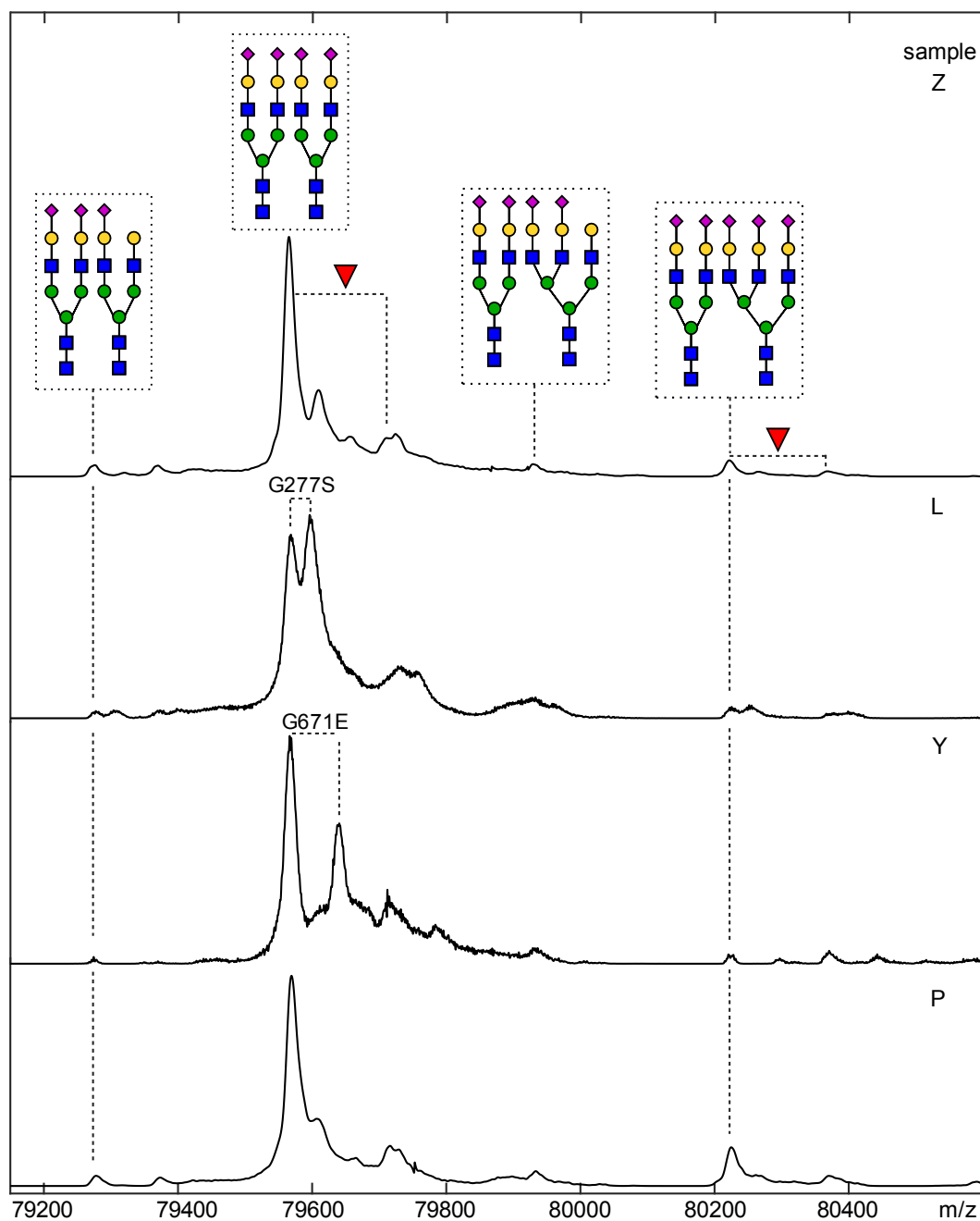
<b>Deconvoluted measured mass [Da]</b>	<b>Method</b>	<b>Year</b>	<b>Reference</b>
79561	API triple quadrupole LC-MS	2001	[78]
79566 / 79548 / 79561	ESI-qTOF	2005	[139]
79600 ± 72	MALDI-TOF	2005	[139]
79559	ESI-qTOF	2008	[81]
79558 / 79556	ESI-qTOF	2009	[80]
79554	ESI-qTOF	2014	[77]
79553.9	API/ESI- quadrupole orbitrap hybrid	2016	[79]
79555 [M+H] <sup>+</sup>	ESI-qTOF	2018	This work
<b>Calculated most abundant mass [Da]</b>			
79567.9			

In more recent publications since 2014 and in this work, a mass of 79554 Da on average has been described. Even though the mass difference between the calculated mass and the measured mass is roughly 12 Da, no further experiments have been performed to identify the origin of this mass difference. Top-down fragmentation using ETD or ECD might be useful in protein sequencing and identification amino acid modifications although the protein size and a high number of disulfide bridges are not beneficial for top-down sequencing of transferrin.[140] The listed publications after 2005 in Table 9 report similar intact masses for the dominant proteoform of the native amino acid sequence (UniProtKB P02787, reviewed by Swiss-Prot) of transferrin from human serum. Therefore, data interpretation was based on these findings and a shift of -12.4 Da was applied for the calculation of a theoretical spectrum to determine the relative abundances of different proteoforms. This relative quantification was performed similar to the procedure described for ceruloplasmin using a Matlab based script (see appendix, *10.3. Matlab Script for the Relative Quantification of Different Proteoforms*)

After purification from 25 plasma samples, transferrin was analyzed by LC-ESI-qTOF-MS in positive ion mode (see Figure 8 for a total ion chromatogram of the LC-MS run). An additional ISCID of 100 eV was applied to improve ionization of the 80 kDa glycoprotein. Next to the native protein variant with a dominant peak at 79555 kDa SNP variant G277S was detected in four samples and variant G671E was detected in one sample. Both SNPs were detected in heterozygous expression next to the native variant in each sample. The amino acid exchange from glycine to serine leads to a mass shift of +30 Da (G277S) and the amino acid exchange from glycine to glutamine leads to a mass shift of + 72 Da. Figure 28 shows the deconvoluted ESI-MS spectra from sample Z, L, Y, and P. The most abundant peak (79555 Da) in all samples corresponds to a proteoform with two diantennary, fully sialylated N-glycans. A peak at 80211 Da indicated the presence of a diantennary and triantennary, fully sialylated N-glycan. For both compositions, peaks indicating the addition of a fucosyl residue as well as the loss of one neuraminic acid are observed. Sample L shows the heterozygous expression of SNP variant G277S and sample Y shows the heterozygous expression of SNP variant G671E next to the native variant. The reported minor allele frequency for SNP G277S (rs1799899) is 7.4% in Europeans and for G671E (rs121918677)



0.3% in Europeans.[141] Even though only 25 samples were analyzed in this work, a fitting allele frequency of 8% for SNP G277S was found. For SNP G671E, which was only detected in one sample heterozygous, the calculated allele frequency is 2%.



**Figure 28.** Deconvoluted ESI-MS spectra of human serum transferrin isolated from four different plasma samples. The dominant peak in all samples at 79555 Da corresponds to a proteoform with two diantennary fully sialylated N-glycans. Additional glycan compositions are observed which indicate the loss of one neuraminic acid or the addition of a fucosyl residue. Peaks indicating the presence of an additional antennae are observed at higher masses (80211 Da), also with the addition of a fucosyl residue and with the loss of one neuraminic acid. Sample L shows a heterozygous expression of SNP G277S with a mass shift of +30 Da; Sample Y shows the heterozygous expression of SNP G671E with a mass shift of +72 Da. As also observed for ceruloplasmin, proteoforms of sample P shows an increased abundance of glycan structures with increased antennarity (peak at 80211 Da, bottom spectrum).

Most samples showed a glycosylation pattern as shown for sample Z with only minor variation between different plasma samples. However, sample P showed an increased expression of glycan compositions corresponding to a combination of di- and triantennary glycan structures, which is for example seen for the peak at 80211 Da. This glycan composition is found with a relative abundance of 3.3% ( $\pm 2.2\%$ ) over 24 samples but has a relative abundance of 11% in sample P. This feature of plasma sample P has already been observed for IgG and ceruloplasmin (see Figure 18 and Figure 27), where elevated amounts of branching and sialylation were detected.

Table 10 shows the results for relative quantification of all glycan compositions for all analyzed samples. Except for sample P, all samples showed similar glycosylation patterns. The most abundant proteoform with two diantennary, fully sialylated glycans was found with an average abundance of 65%. The degree of fucosylation was found to be 6.1% on average.

**Table 10.** Relative amounts of different proteoforms caused by glycosylation of human transferrin from 25 plasma samples.

Donor	2 biantennary glycans [%]				1 biantennary glycan, 1 triantennary glycan [%]			
	H10N8S4	H10N8S4F	H10N8S3	H10N8S3F	H11N9S5	H11N9S5F	H11N9S4	H11N9S4F
B, native	65.2	24.0	0.8	0.8	1.8	2.2	4.9	0.3
C, native	61.9	23.4	1.6	2.7	2.2	2.1	5.9	0.2
C, G277S	64.5	22.2	1.5	3.6	2.2	2.5	3.6	0.1
E, native	59.7	26.2	0.5	1.2	2.2	2.3	7.2	0.8
E, G277S	58.9	29.7	0.2	2.1	2.0	2.8	3.9	0.4
F, native	66.8	15.8	3.4	2.1	4.1	1.7	5.7	0.4
G, native	67.2	15.8	2.8	3.4	3.9	2.8	4.1	0.1
H, native	56.4	23.2	0.5	2.8	2.8	3.3	9.5	1.6
H G277S	57.9	24.5	0.5	3.6	3.0	3.8	5.9	0.7
J, native	68.7	15.6	1.8	2.6	4.3	2.2	4.3	0.4
L, native	64.7	18.4	1.7	3.2	3.0	1.4	7.4	0.1
L, G277S	66.3	18.2	2.0	3.9	3.4	1.8	4.3	0.0
M, native	67.7	16.2	2.3	3.3	4.0	2.2	4.0	0.2
N, native	67.3	16.2	2.4	3.3	3.6	2.1	4.6	0.5
O, native	66.0	16.7	2.8	2.7	3.2	2.0	5.6	1.0
P, native	62.0	14.3	2.6	2.1	11.0	3.2	4.7	0.1
Q, native	70.1	12.0	3.2	1.5	5.5	2.0	5.4	0.2
S, native	66.3	16.2	3.1	2.7	2.9	3.2	5.1	0.5
T, native	64.0	16.4	2.8	1.8	4.5	2.4	6.9	1.2
U, native	67.2	16.5	3.7	2.5	3.6	2.4	4.2	0.1
V, native	65.1	16.0	3.5	3.5	3.3	2.9	5.3	0.3
W, native	68.0	18.4	1.2	1.1	4.9	3.4	2.9	0.0
X, native	66.6	18.1	3.2	1.8	2.9	4.0	3.1	0.4
Y, native	65.7	23.5	0.8	0.6	1.7	2.9	4.9	0.0
Y, G671E	70.6	21.3	0.4	2.4	1.5	2.9	0.8	0.0
Z, native	67.9	15.8	3.0	2.9	4.4	1.6	3.9	0.5
AA, native	65.7	13.7	4.3	3.7	3.8	2.0	6.4	0.5
AB, native	69.4	15.2	4.1	2.8	3.5	1.9	3.0	0.1
AC, native	65.2	14.8	2.7	3.0	3.4	2.6	7.5	0.8
AD, native	64.5	17.5	1.9	1.9	3.8	4.5	5.4	0.5
<b>Average<sup>a</sup></b>	<b>65.4</b>	<b>18.7</b>	<b>2.2</b>	<b>2.5</b>	<b>3.3</b>	<b>2.5</b>	<b>5.0</b>	<b>0.4</b>
<b>STD<sup>b</sup></b>	<b>3.4</b>	<b>4.1</b>	<b>1.2</b>	<b>0.9</b>	<b>1.0</b>	<b>0.7</b>	<b>1.7</b>	<b>0.4</b>

<sup>a</sup>Average values are calculated without sample P. <sup>b</sup> Standard derivation of values without sample P.

## 5.5. Highly Glycosylated Alpha-Acid-Glycoprotein

Parts from this chapter have been published as research article in the *Journal of Proteome Research* entitled “Intact Human Alpha-Acid Glycoprotein (AGP) Analyzed by ESI-qTOF-MS: Simultaneous Determination of the Glycan Composition of Multiple Glycosylation Sites”. [142]

The plasma protein alpha-acid-glycoprotein (Orosomucoid) is involved in many biological mechanisms like transport of lipophilic compounds, anti-inflammatory reactions or immunomodulatory functions as one of the major positive acute-phase proteins. It is an abundant glycoprotein present at 0.6 to 1.2 mg/mL of blood plasma. [76] Nearly half of its mass originates from glycosylation of five N-glycosylation sites, leading to a 40 kDa glycoprotein with a very low pI ranging from 2.8 to 3.8, caused mainly by its high amount of sialic acid. [73, 143] It has been shown that its biological functions are highly dependent on the glycosylation presented on the proteins. A change in glycosylation due to physiological conditions like pregnancy or inflammation has been reported. [13, 144] An altered glycosylation of human AGP was also observed in a variety of cancer diseases like hepatocellular carcinoma, ovarian carcinoma, lung cancer or pancreatic cancer and other diseases as rheumatic arthritis. [15, 145-152] Observed changes include increased expression of the Lewis X motive, increased or decreased sialylation and branching. Accordingly, AGP is an interesting target for developing biomarkers in terms of early detection, prognosis or monitoring physiological conditions including cancer diseases. However, glycan analysis of AGP seems to be an enormous task because of the heterogeneity in glycosylation of each of five sites. In 1997, Sottani *et al.* performed a MALDI experiment for molecular weight determination of bovine alpha-1-acid glycoprotein, which resulted in broad and unresolved molecular ion peaks, demonstrating the heterogeneity of this protein. They also analyzed other highly sialylated proteins like erythropoietin but found ESI to be ineffective for molecular weight measurement and hypothesized that the oligosaccharides are likely to interfere with the multiply charged ion production. [153]

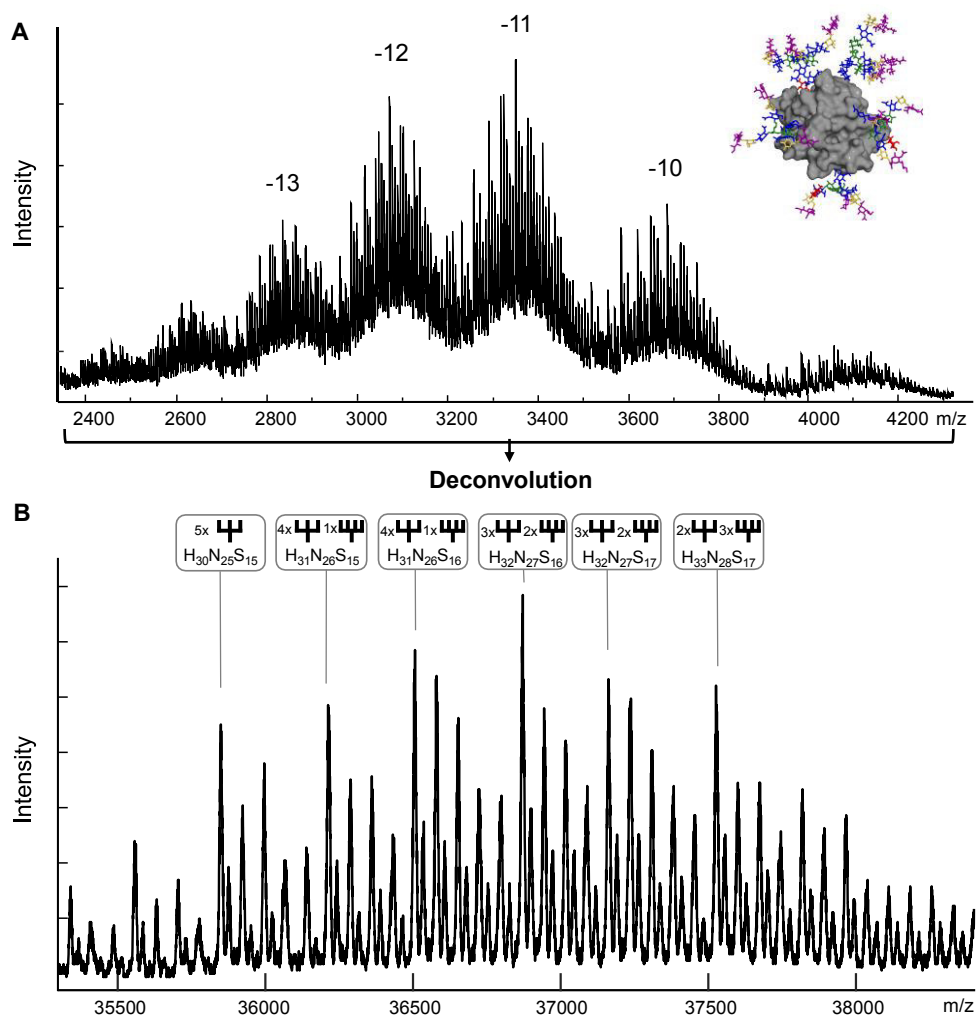
A few years later, Nagy *et al.* showed that AGP can be ionized by ESI in combination with an FT-ICR instrument. They observed protein species that

ionized as multiple sodium adducts and identified one proteoform as 24-fold sodium adduct in a charge state of +12. No overall glycosylation pattern was reported.[154] In 2010, Ongay *et al.* separated different proteoforms of intact human alpha-1-acid glycoprotein by capillary electrophoresis and were able to determine the glycosylation patterns by ESI-MS after separation of genetic variants and glycoforms.[155, 156] However, analyzing AGP as intact glycoprotein without prior separation of different proteoforms would be a more efficient approach with reduced instrumental setup and time requirement.

### **5.5.1. Analysis of Intact AGP from Pooled Human Plasma**

AGP from pooled human plasma was used for the analysis of intact AGP described in this section. The purified protein was purchased as lyophilizate, containing AGP-1 and AGP-2.

AGP was analyzed as intact glycoprotein by using direct injection of a sample in acetonitrile/water 1:1 into an ESI-qTOF mass spectrometer. In positive ion mode, no resolvable spectrum was observed by direct infusion or after LC-MS separation on a C8 column. However, in negative ion mode, well-resolved peaks corresponding to charge states from -9 to -15 were detected. The large number of peaks observed for each charge state indicates the presence of multiple different glycans. Deconvolution of the spectrum yields the hypothetical species with a charge of zero. The deconvoluted spectrum also revealed severe heterogeneity of the glycans of human AGP. The most abundant peaks correspond mainly to a combination of tri- and tetraantennary glycan structures at each glycosylation site (c.f. Figure 29, antennarity is visualized as comic icons).



**Figure 29.** ESI-qTOF-MS spectrum of the highly sialylated alpha-acid-glycoprotein from pooled plasma in negative ion mode. **(A)** Charge states of the ESI-MS spectrum range from -9 to -15. **(B)** The deconvoluted spectrum shows the severe glycan heterogeneity with the most abundant glycan composition of 32 hexoses, 27 N-acetylglucosamines and 16 sialic acids ( $H_{32}N_{27}S_{16}$ ), which can be realized as three triantennary and two tetraantennary glycan structures (cf. discussion in text). Figure was modified from reference [142].

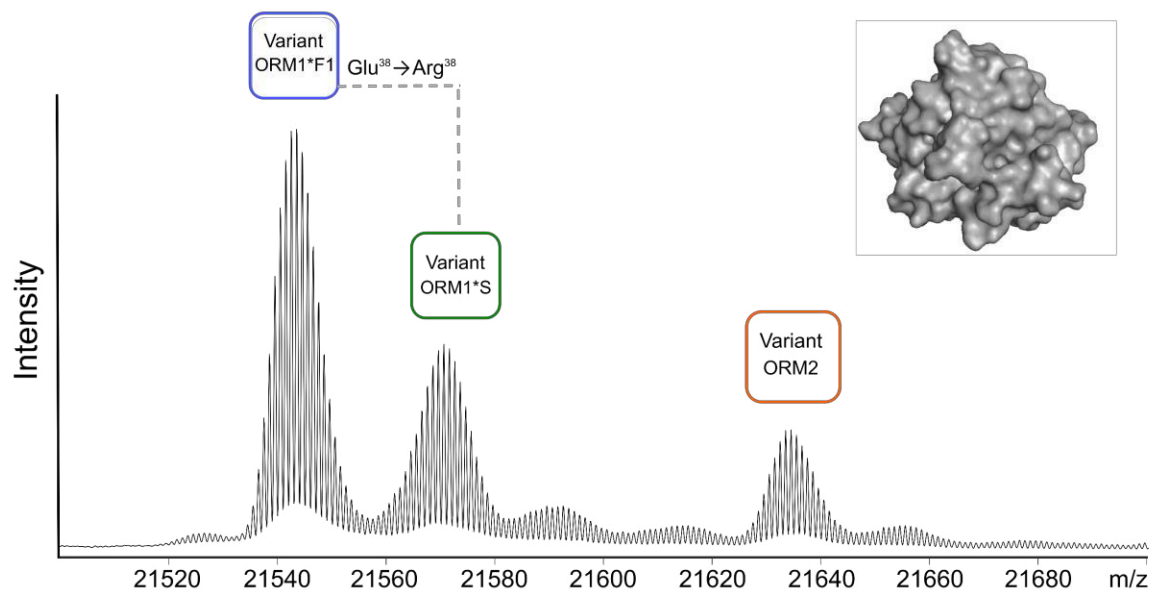
Besides the complexity of the spectrum due to the glycosylation pattern, AGP consists of different genetic variants, mainly, ORM1\*F1 and ORM1\*S differing in one amino acid exchanged from glutamine to arginine at position 38 for AGP-1 and one variant for AGP-2 (ORM2) (cf. discussion below).[76, 157] In addition to the high number of glycoforms and genetic variants there is a mass degeneracy between the presence of two fucosyl residues and that of one sialic acid, which differ only by one in their respective mass. To overcome this problem and precisely

determine glycan composition, AGP was treated with sialidase to remove all sialic acids. As a result, the number of fucosyl residues can be determined without overlapping peak intensities. This information can then be used as a starting point for interpretation of the fully sialylated glycoprotein spectrum.

#### **5.5.1.1. Analysis of Desialylated AGP for a Conclusive Interpretation of Fully Sialylated AGP**

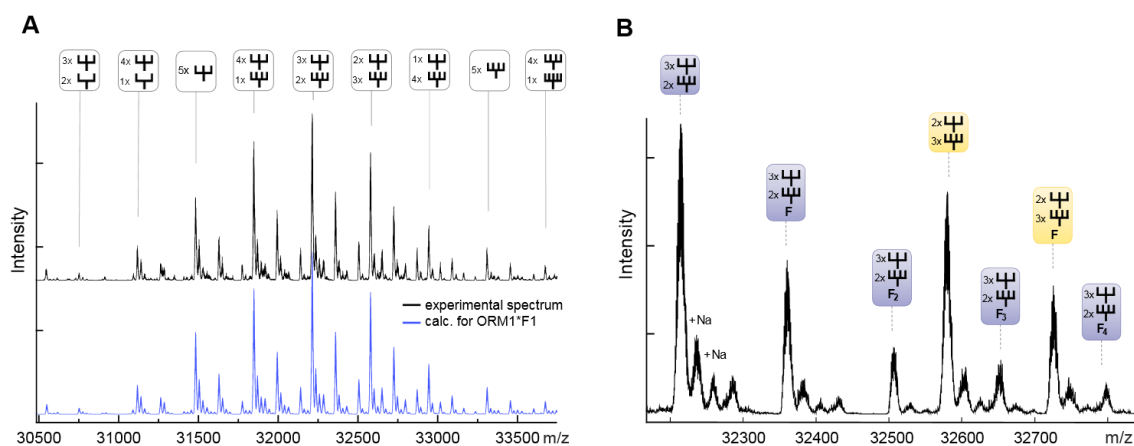
After treatment with non-specific sialidase, the desialylated glycoprotein was subjected to a chromatographic separation over a C8 reversed phase column to separate the different genetic variants with online MS measurement on an ESI-qTOF instrument. In this case, the positive ion mode showed better intensities for the desialylated glycoprotein probably because of the removal of many negative charges. The AGP sample from pooled plasma showed dominantly the genetic variants ORM1\*F1 (Figure 31) and ORM1\*S for AGP-1, corresponding to the most abundant genetic variants as well as one variant for AGP-2 (ORM2). Observed charge states were in the range from +11 to +17 for all genetic variants. The most intense charge state was observed as +13. Subsequent deconvolution over all observed charge states with a maximum entropy algorithm led to a spectrum with a theoretical charge of +1. A theoretical spectrum obtained from the optimization of all genetic protein variants and all possible oligosaccharide structures was then fitted to the deconvoluted spectrum like the procedure described for ceruloplasmin (see section 5.2.3. *Quantification of Glycopeptides with Different Peptide Length*). The protein variants ORM1\*F1 and ORM1\*S of AGP-1 correspond to 201 amino acid long sequences without the signal peptide 1-18 and the formation of two disulfide bridges. Both variants show the formation of pyroglutamic acid at the amino-terminal glutamine at position 19 as described elsewhere.[158] The chemical formula for the protein backbone results in C<sub>966</sub>H<sub>1465</sub>N<sub>251</sub>O<sub>299</sub>S<sub>5</sub> for variant ORM1\*F1 and C<sub>967</sub>H<sub>1469</sub>N<sub>253</sub>O<sub>298</sub>S<sub>5</sub> for variant ORM1\*S. The variant ORM2 (AGP-2) consist of 201 amino acids as well but shows a slightly higher mass with a chemical formula of C<sub>967</sub>H<sub>1459</sub>N<sub>254</sub>O<sub>298</sub>S<sub>7</sub>. Figure 30 shows the deconvoluted ESI-MS spectrum of the different variants, confirming the amino acid sequence.





**Figure 30.** Deconvoluted ESI-MS spectrum of fully deglycosylated and reduced AGP-1 (ORM1) and AGP-2 (ORM2). AGP-1 shows two dominant signals corresponding to variant ORM1\*F1 and ORM1\*F1 with slightly different abundances. The two variants differ in one amino acid (Q38R), leading to a mass difference of 28 Da. Variant ORM2 is found in lower abundances which is in good agreement with literature.[157]

Figure 31 shows the experimental spectrum of the de-sialylated protein for variant ORM1\*F1, compared to the calculated isotopic distribution pattern based on the amino acid sequence and various glycan compositions. This calculated pattern was obtained using the Matlab based script (appendix, section 10.3. *Matlab Script for the Relative Quantification of Different Proteoforms*) as described before.



**Figure 31.** (A) Deconvoluted spectrum of the desialylated variant ORM1\*F1 in black from pooled plasma, showing more than 30 glycan compositions and their fucosylated analogs. The calculated spectrum is shown in blue. Glycan compositions over the five glycosylation sites range from 14 antennas to 22 antennas, representing biantennary and triantennary structures to tetraantennary and pentaantennary structures, respectively. (B) Excerpt of A with a mass range from 32200 to 32800 m/z. Each of these glycans is also found modified with different amounts of fucosyl residues. Also observed are adducts of sodium. Glycan structures with the same number of antennas but more than two fucosyl residues interleave with glycan structures with a higher number of antennas as shown for the structures marked in blue and yellow. Figure was modified from reference [142].

The most abundant glycan composition of the de-sialylated AGP-1 contains 32 hexoses and 27 N-acetylglucosamines (H32N27) without any fucoses (F). Glycan compositions were interpreted as bi-, tri-, tetra- and pentaantennary structures since NMR data of free glycans from human AGP showed mainly highly branched glycans without repeating N-acetyllactosamine (LacNAc) motifs.[159] However, the presence of repeating LacNAc motifs must also be considered since only glycan compositions are determined. Considering five N-glycosylation sites, the most abundant composition can be realized as three triantennary and two tetraantennary glycan structures, which is in good agreement with reported data.[151, 160, 161]

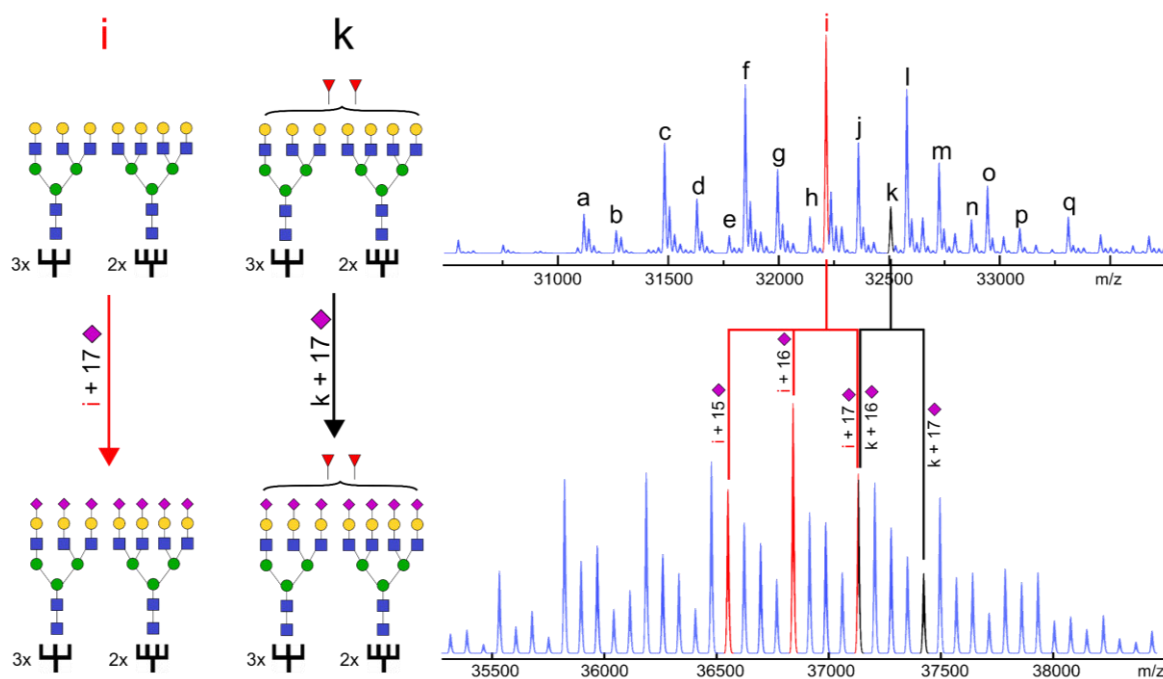
The spectrum of the desialylated AGP-1 of variant ORM1\*F1 from pooled human plasma shows a variety of different glycan compositions. The degree of fucosylation ranges from unfucosylated species which make up ~50% of all glycoproteins, over one fucose per protein at ~25%, two fucoses at ~10% and up to a composition with four fucoses attached to the glycan structures. It is shown that the intensities decrease with higher amount of fucosyl residues.

Desialylated AGP-1 of variant ORM1\*S from pooled plasma showed a broader distribution of glycoforms and signals that can be interpreted as either agalacto forms or as bisecting N-acetylglucosamines (see appendix, section 10.5. *Alpha-Acid-Glycoprotein*). These results were also confirmed by ESI-MS analysis of the PNGase F released glycans from AGP (see section 5.5.3. *Analysis of Free Glycans of Alpha-Acid-Glycoprotein*). Agalacto glycan structures with an increased amount of fucosylation on AGP were previously found from ascitic fluid from patients with stomach cancer. Studies on the immunomodulating effects of AGP showed that asialo-agalacto forms of AGP with increased fucosylation possessed higher immunosuppressive activity in terms of lymphocyte proliferation.[162, 163] Additionally, glycan compositions corresponding to agalacto or bisecting N-acetylglucosamines have also been found in low amounts on AGP from human serum for biantennary structures without assignment to a specific genetic variant.[145]

#### **5.5.1.2. Interpretation of Fully Sialylated AGP Based on the Information of Desialylated AGP**

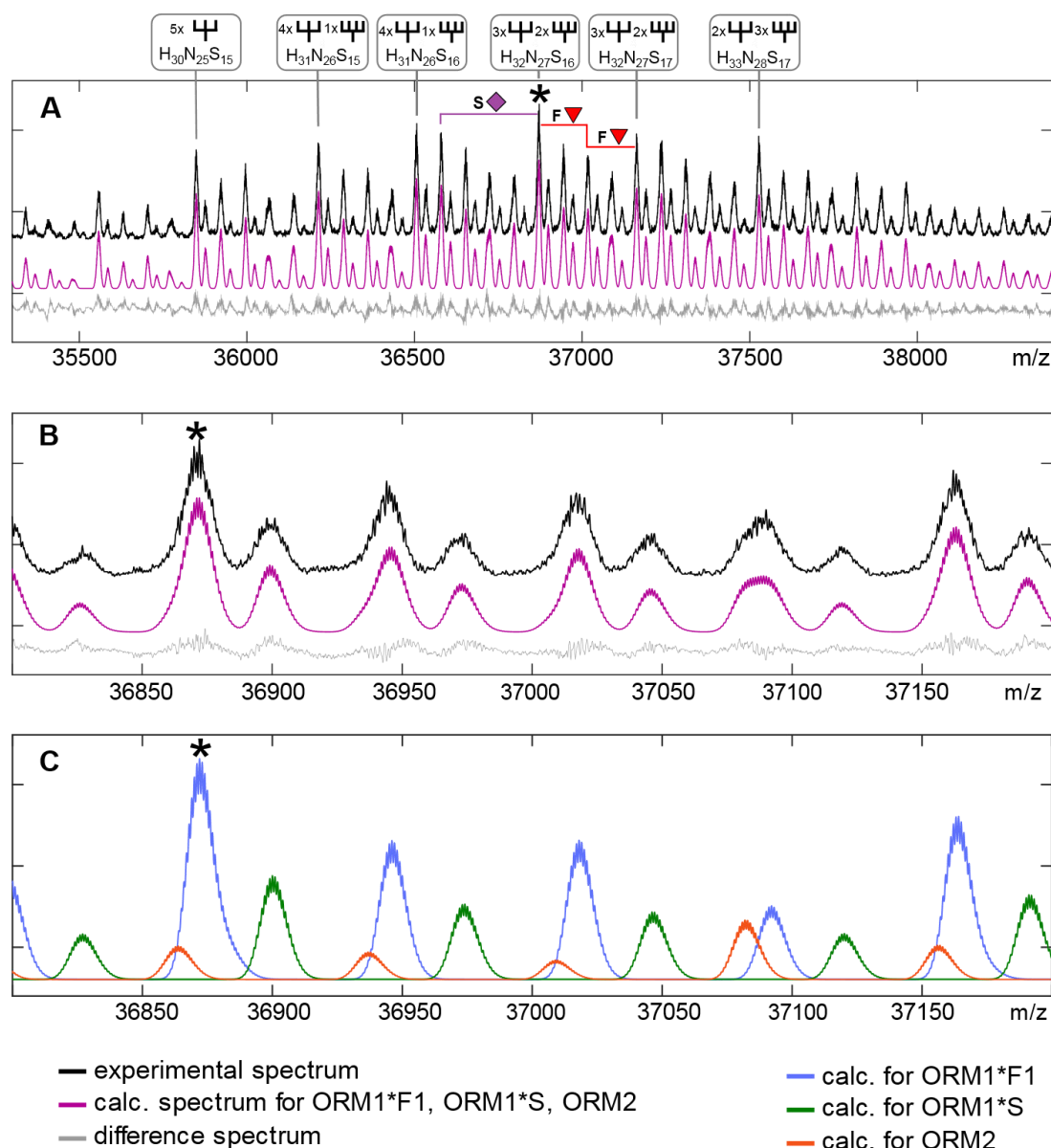
The fully sialylated protein from pooled human plasma was measured by ESI-MS in negative ion mode since the positive ion mode showed low resolution with a low signal to noise ratio. Ionization properties were much improved in negative ion mode due to a high number of acidic groups of the sialic acids, leading to the very low  $pI \sim 3.3$  of the protein. Additionally, no adduct formation was observed in negative ion mode as the glycoprotein variants ionize as  $(M - xH)^{-}$ . For analysis of the intact protein, the lyophilized protein was dissolved in equal volumes of acetonitrile and water and without chromatographic separation subjected to the ESI source, using a syringe pump. The MS spectrum of the variants ORM1\*F1, ORM1\*S and ORM2 displaying their complex heterogeneity was obtained within a few minutes. After deconvolution over the most abundant charge states from -9 to -15, the signal-to-noise ratio increased from 13 in the unprocessed spectrum to 60 for the deconvoluted spectrum calculated for the most intense signal respectively. Different glycoprotein forms could be identified in the range of about 35.5 to

38.5 kDa. In addition, a single nucleotide polymorphism (SNP) was evident with variant ORM1\*S showing peaks with lower intensity and a mass difference of +28 Da, due to the amino acid exchange Gln38Arg. The difference in intensity is consistent with the reported allele frequency of 0.61 for variant ORM1\*F1 compared to 0.39 for variant ORM1\*S in white population.[76] A matching frequency of 0.70 for variant ORM1\*F1 and 0.30 for variant ORM1\*S was detected. AGP-2 (ORM2) was detected in a much lower amount with a five times lower intensity compared to the sum of both variants for AGP-1. The reported proportion for AGP-1 and AGP-2 is 3:1.[157] The complex spectrum contains overlapping peaks, due to the fact that two fucoses have a mass higher by only 1 Da compared to one sialic acid. A few other compositions also show degeneracies like  $H_{30}N_{25}FS_{15}$  and  $H_{32}N_{27}S_{13}$ , which have a mass difference of only 2 Da. For a conclusive interpretation of the glycan distribution, the proportion of the fucosylated structures determined in the spectrum of the desialylated AGP forms were used to calculate the degree of sialylation. Like the interpretation of the desialylated spectrum, possible mass distributions for different glycoprotein variants and glycoprotein species were calculated, based on the amino acid sequence and various glycan compositions, respectively. The mass distributions were then fitted to the experimental spectrum using a least square algorithm. This minimization was performed with the addition of constraints to include the proportions of fucosylated structures observed in the spectrum of the desialylated glycoprotein. A linear equality constraints matrix ( $A_{eq}$ ) and a linear equality constraints vector ( $b_{eq}$ ) were defined using the proportions of the fucosylated structures. For variant ORM1\*F1, 28 desialylated glycan composition and their relative proportions were identified by ESI-MS (see previous chapter). The theoretical addition of 12 to 19 sialic acids to the desialylated compositions leads to 232 possible glycan compositions for variant ORM1\*F1. For all three variants (ORM1\*F1, ORM1\*S and ORM2), a total of 728 possible proteoforms were considered by the linear equality constraints matrix. However, biologically irrelevant structures were excluded by setting their upper bounds to 0, lowering the number of possible structures to 565. Structures that were excluded are heavily sialylated glycans with more than a complete termination of all antenna with sialic acids. The procedure is illustrated in Figure 32.



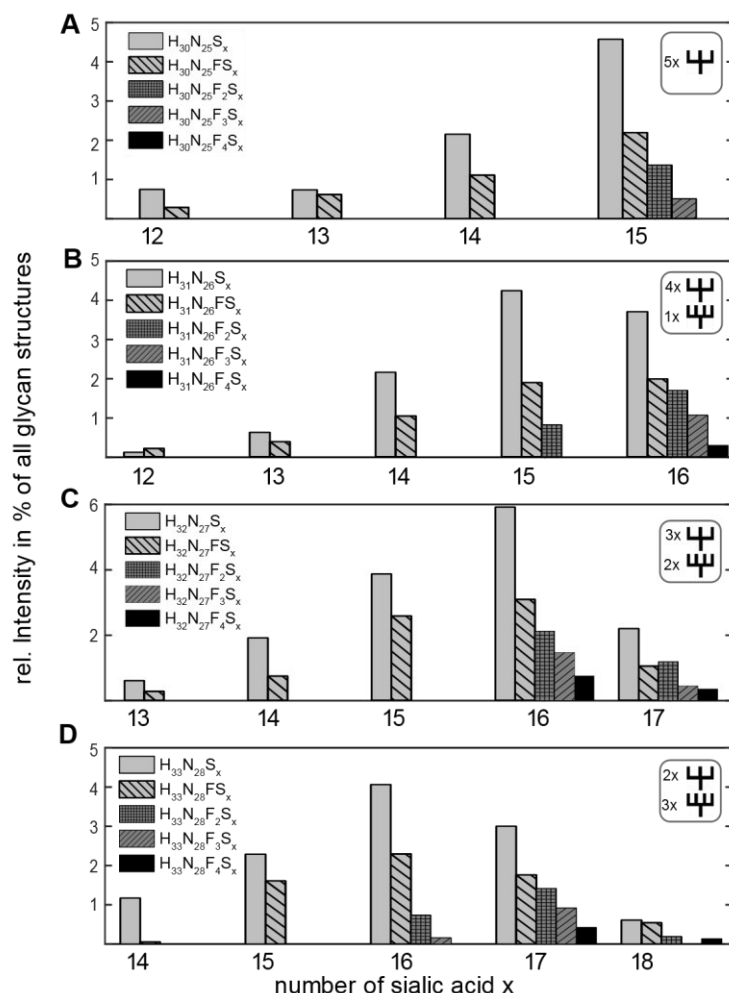
**Figure 32.** Calculation of a fully sialylated spectrum (bottom), using the proportions of desialylated glycan compositions obtained from the spectrum of the enzymatically desialylated forms (top). The intensities of the desialylated compositions (a to q) were related and used as a linear equality constraints matrix for the calculation of the fully sialylated spectrum. As shown for composition i and k, different sialylated forms are found in the fully sialylated spectrum, in part overlapping. Figure was modified from reference [142].

The deconvoluted spectrum (Figure 33) of the fully sialylated protein shows that the most abundant glycan composition (marked with an asterisk) consists of 32 hexoses, 27 N-acetylglucosamines and 16 sialic acids ( $H_{32}N_{27}S_{16}$ ), being consistent with the results for the desialylated glycan compositions. In total, 90 different glycan composition over five N-glycosylation sites were identified for variant ORM1\*F1 and 101 for variant ORM1\*S. AGP-2 was detected in a much lower amount and 64 glycan compositions were identified. A complete list of all glycan composition for all genetic variants can be found in the appendix (10.5. *Alpha-Acid-Glycoprotein*).



**Figure 33.** ESI-qTOF-MS spectrum of the highly sialylated alpha-acid-glycoprotein in negative ion mode. **A)** Experimental deconvoluted spectrum (top), calculated spectrum (middle) and difference of the two (bottom). **B)** expansion of A) over 500 mass units. **C)** assignment of the section shown in B) to the individual variants. The deconvoluted spectrum (black, panels A and B) shows the severe glycan heterogeneity of variant ORM1\*F1, ORM1\*S and ORM2 with the most abundant glycan composition (marked with an asterisk) of 32 hexoses, 27 N-acetylglucosamines and 16 sialic acids (H32N27S16). The distances of added fucosyl residues and sialic acid residues are exemplified for this peak. For a complete interpretation of glycan compositions, the proportions of the desialylated glycoforms were used to calculate the fully sialylated spectrum for variant ORM1\*F1 (blue), ORM1\*S (green) and ORM2 (red). The green spectrum is shifted by +28 Da to higher masses, whereas the peaks for variants ORM1\*F1 and ORM2 overlap (panel C). The purple spectrum (panel A and B) show the calculated spectrum as sum over all detected variants. The gray spectrum represents the difference between the experimental spectrum and the calculated spectrum.[142]

The percentages for the most abundant glycan compositions (84% of all found glycan structures), in total 58 tri- and tetraantennary structures for variant ORM1\*F1, are displayed in Figure 34, sorted by increasing number of LacNAc motifs.



**Figure 34.** Distribution of sialic acid for the 58 most abundant glycan structures for variant ORM1\*F1, representing over 84% of all glycan structures. Panel A to D differ in increasing number of LacNAc motifs. The amount of sialylation seems to increase with higher amount of fucosylation and the degree of sialylation seems to decrease with increasing size of the glycan structure. Figure was modified from [142].

Panel A displays the distribution of sialic acids for the five triantennary structures with zero to four fucosyl residues and 12 to 15 sialic acids, showing that the most abundant structures are fully sialylated. This could be visualized such that all triantennary structures carry a sialic acid on each branch, yielding in total 15 sialic acids for five N-glycosylation sites. Considering all shown distributions, the amount of sialylation increases slightly with the degree of fucosylation. For example, the unfucosylated structure  $H_{31}N_{26}S_x$  (panel B) shows the highest intensity for  $x=15$  sialic acids, whereas the fucosylated structures  $H_{31}N_{26}F_{1-4}S_x$  have the highest

intensities for  $x=16$  sialic acids. This indicates a correlation between sialylation and fucosylation, which is in agreement with the assignment of Sialyl-Lewis X structures from NMR spectra.[159] Additionally, the degree of sialylation seems to decrease somewhat with increasing size of the glycan structures. This is demonstrated by the most intense glycan composition H32N27S16 (panel C) and H33N28S16 (panel D), both carrying 16 sialic acids. Overall, the results are in good agreement with the data reported by Ongay *et al.* who analyzed intact AGP by capillary electrophoresis ESI-MS.[155] Using ESI-MS in negative ion mode without prior chromatographic separation, a slightly greater amount of higher antennary glycans carrying up to 19 sialic acids (for example H36N31S19 or H37N32S19) was found. However, these differences can be explained by naturally occurring variations in glycosylation from different samples or the use of different MS techniques.

### **5.5.2. Analysis of Desialylated AGP from Individual Plasma Samples**

To analyze individual differences in glycosylation, AGP was purified from several plasma samples by a combination of AIEX, size exclusion and RP-chromatography as described in section 5.1. *Protein Purification from Human Plasma*. Analyzing intact AGP from individual plasma samples was not possible due to insufficient protein purification. Other proteins, as well as residual buffer salts, hampered ionization in negative ion mode. Since the analysis of intact AGP in negative ion mode from pooled plasma showed an almost complete termination with sialic acids for all antennas, it was assumed that no additional information would be obtained from the analysis of fully sialylated AGP from individual plasma samples. Therefore, AGP purified from plasma samples was analyzed after desialylation by ESI-MS in positive ion mode. Desialylation was achieved by incubation of the samples with sialidase. As described above, changes in fucosylation or antennarity can be detected easily after desialylation, providing a powerful tool for monitoring disease-specific changes like increase expression of Sialyl-Lewis X structures (see section 5.7. *Glycosylation of Plasma Proteins as Potential Biomarker*).



Unlike the purchased AGP sample from pooled plasma, AGP samples isolated from plasma of single individuals showed differences in their genetic makeup. Table 11 summarizes the desialylated variants found in the different samples. Most of the donors were heterozygous for AGP-1 and expressed variant ORM1\*F1 and ORM1\*S, only samples O, P and V showed a homozygous expression for either variant ORM1\*S or ORM1\*F1. Variant ORM1\*F2 was only found in sample N and Q. In all samples, only one genetic variant was found for AGP-2, which is in good agreement with reported data.[157]

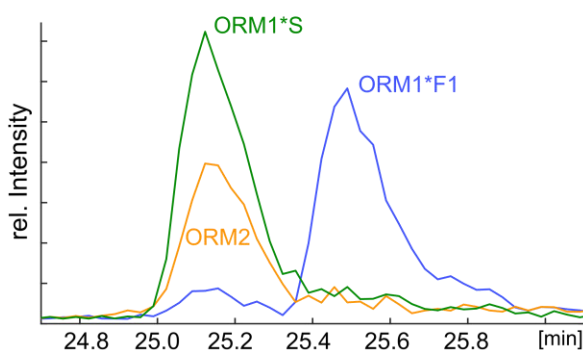
**Table 11.** Analyzed plasma samples and genetic variants of AGP-1 and AGP-2.

	Sample	Variant ORM1 (AGP-1)			Variant ORM2 (AGP-2)
		*F1	*S	*F2	
AGP-1	Pooled human plasma	x	x	-	x
L	Female, 51 years	x	x	-	x
N	Male, 21 years	-	x	x	x
O	Male, 54 years	x	-	-	x
P	Female, 28 years	-	x	-	x
Q	Male, 24 years	x	-	x	x
R	Male, 55 years	x	x	-	x
T	Male, 47 years	x	x	-	x
V	Male, 51 years	x	-	-	x
Z	Male, 38 years	-	x	-	x
AA	Male, 27 years	x	x	-	x
AD	Male, 46 years	x	x	-	x
Most abundant mass of peptide backbone (reduced) [Da]		21542.6	21571.7	21574.6	21633.5

The samples were analyzed by LC-ESI-MS in positive ion mode, using a C8 column. A total ion chromatogram for sample P can be found in the section 5.1.3. *Reversed Phase Chromatography with Online ESI-MS Detection*, showing albumin as major impurity but with a different retention time than desialylated AGP such that it is well separated during the C8 chromatography. It was possible to achieve a partial chromatographic separation of the genetic variants on a reversed phase column.

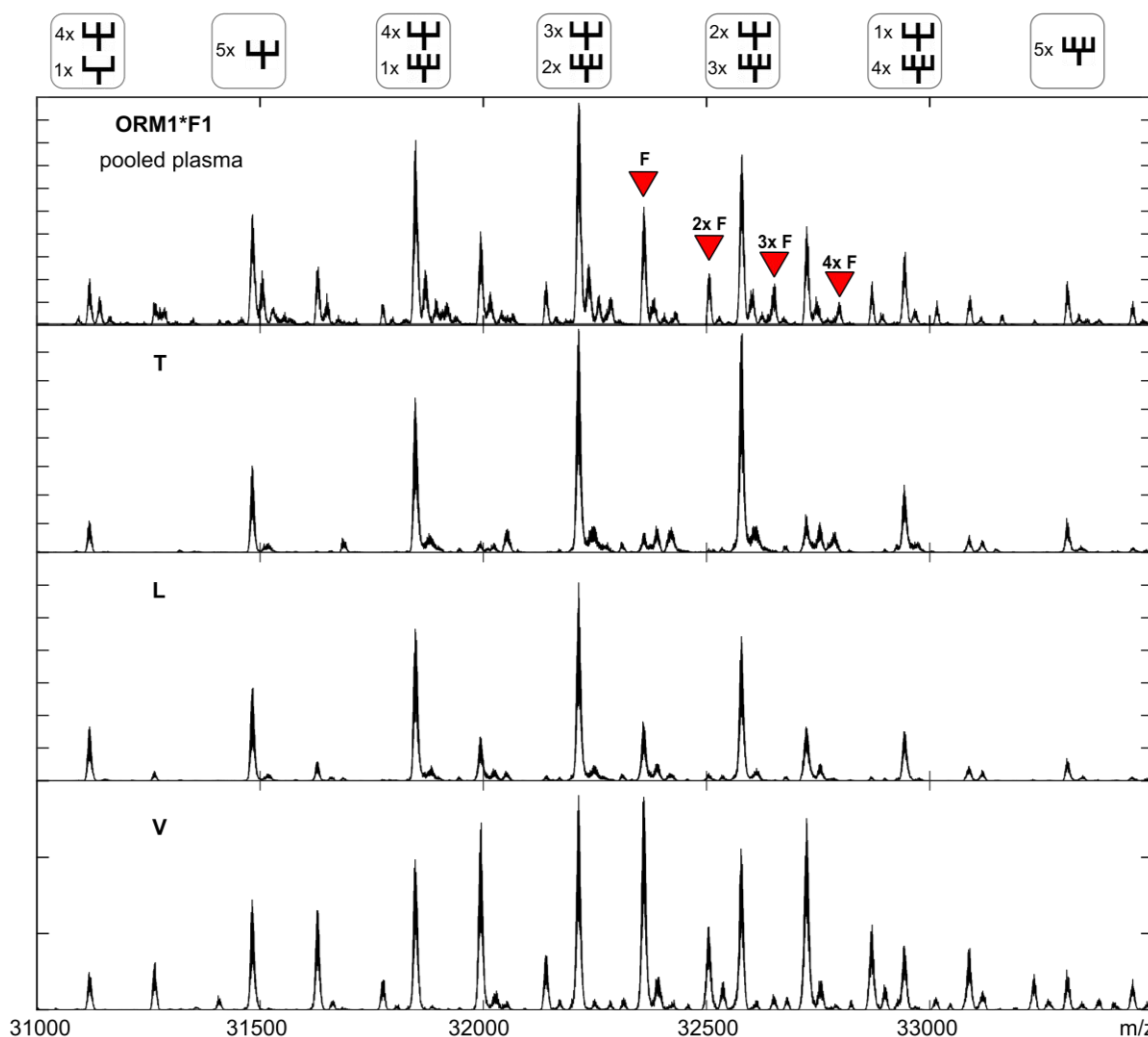
Figure 35 shows the extracted ion chromatogram from sample T for the different variants. Variant ORM1\*S and ORM2 both eluted at a retention time of 25.1 min.

However, no overlapping peaks were observed after deconvolution but rather an interleaved signal pattern of variant ORM1\*S and ORM2, due to the mass difference of the two variants. Variant ORM1\*F1 and ORM1\*F2 eluted at a retention time of 25.5 min, also showing an interleaved signal pattern without overlapping signals.



**Figure 35.** Extracted ion chromatogram of different genetic variants of desialylated AGP from sample T. Separation was performed on a RP-C8 column with ESI-MS detection. Variant ORM1\*S and ORM2 showed similar elution behavior as well as variants ORM1\*F1 and ORM1\*F2. However, variants with the same elution time showed interleaved signals without overlapping due to their differences in mass.[142]

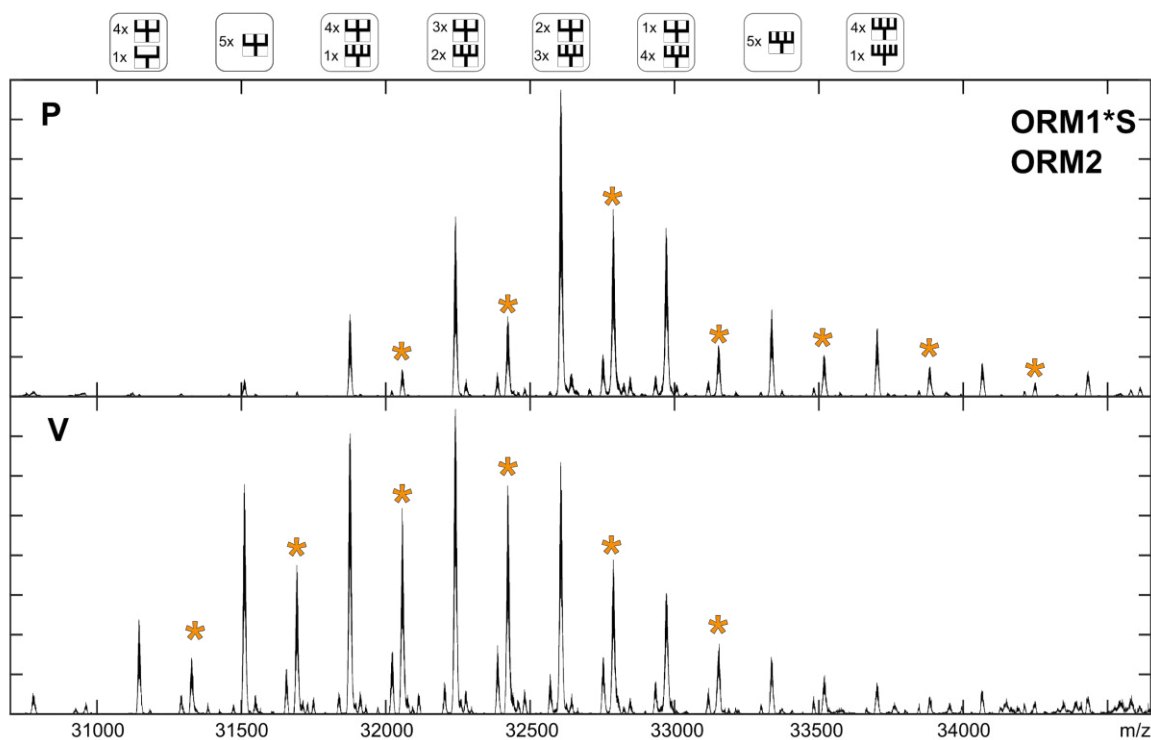
The spectra of the desialylated glycoprotein from single individuals showed differences in the degree of fucosylation and number of antennae. Figure 36 shows exemplary spectra of desialylated AGP for variant ORM1\*F1 and the corresponding AGP sample from pooled plasma as reference. Three of the samples showed a significant lower amount of fucosylated structures compared to the sample from pooled plasma. In contrast, sample V shows a higher expression of fucosylated structures. It seems possible that the amount of fucosylation shows a great variation in different individuals and that the spectrum from pooled plasma is a good representation of an average value. However, no data about the sample size from the purchased AGP from pooled plasma or any medical information about the purified plasma samples from out-dated blood bank samples is available.



**Figure 36.** Exemplary spectra of the de-sialylated AGP-1 for variant ORM1\*F1 from different samples. The spectrum of the sample from pooled plasma show a decreasing intensity with increasing number of fucosyl residues (marked in red for the most abundant composition H32N27). Samples from single individuals show differences in fucosylation. Especially sample V shows a dramatically higher amount of fucosylation. Each peak is accompanied by sodium adducts. Figure was modified from [142].

As mentioned above, genetic variant ORM1\*S and ORM2 were not separated chromatographically. Therefore, data interpretation was done for both variants from the same spectrum, except for samples in which variant ORM1\*S was not present. Genetic variant ORM2 showed in all samples a mass shift of +119 Da compared to the calculated mass for ORM2. This shift was observed in the desialylated spectra but not after enzymatic treatment under reducing conditions (see Figure 30). This mass shift can be explained by a cysteinylolation of one of the cysteine residues of the peptide backbone of AGP-2. AGP-1 as well as AGP-2

each form two disulfide bridges. AGP-2, however, consists of five cysteine residues compared to four residues in AGP-1. Thus, one unpaired cysteine is present and can explain the mass difference in AGP-2. Like the results for variant ORM1\*F1, sample V showed an increased fucosylation for ORM2. It was also remarkable that sample P showed a significantly greater amount of high antennary structures for both variants ORM1\*S and ORM2 (Figure 37).

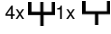

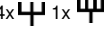
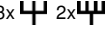
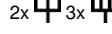
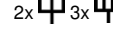

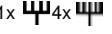


**Figure 37.** Spectra of the de-sialylated AGP for variant ORM1\*S and ORM2 for sample P (top) and V (bottom). Most abundant signals for ORM2 are marked with an orange asterisk with a mass shift of + 182 Da (mass difference of +63 and one cysteinylation of +119). Sample P shows more structures with higher antennary glycans or repeat motifs, whereas sample V shows a slightly higher degree of fucosylated glycans.[142]

The data show that the glycosylation pattern can vary in individuals, especially fucosylation and number of antennae. No significant differences in glycosylation from different genetic variants were detected. Table 12 summarizes the average values of the most abundant glycan structures found in the samples from single individuals. The values represent over 90% of all glycan compositions. A complete list of all glycan compositions for desialylated AGP, including the purchased

sample, can be found in the appendix (10.5. *Alpha-Acid-Glycoprotein*) for variant ORM1\*F1, ORM1\*F2, ORM1\*s and ORM2, respectively.

**Table 12.** Average abundance of glycan compositions found in all eleven samples. Values are given in %.

Variant	4x 		5x 		4x 		3x 			2x 			2x 		5x 	1x 	
	H29N24	F0 F1	H30N25	F0 F1	H31N26	F0 F1	H32N27	F0 F1 F2	H33N28	F0 F1 F2	H34N29	F0 F1	H35N30	F0	H36N31	F0 F1	
<b>ORM1*F1</b>	3.2	1.2	7.3	2.9	12.0	4.9	14.9	6.2	2.7	12.4	5.6	2.0	5.1	1.7	3.5	1.8	0.6
<b>ORM1*F2</b>	4.0	0.5	7.8	2.2	13.3	3.6	16.6	4.7	1.4	16.2	4.7	1.4	6.1	1.9	3.5	2.5	0.6
<b>ORM1*S</b>	3.0	0.6	8.0	2.2	12.4	3.1	17.5	4.2	1.1	16.7	3.8	1.1	6.8	1.5	3.2	2.2	0.5
<b>ORM2</b>	2.6	0.9	7.3	1.9	11.6	3.8	18.2	5.6	2.2	16.3	4.8	1.7	5.7	1.5	2.9	1.3	0.2

Values represent over 88% of all glycan compositions for each variant. A complete list can be found in the appendix (10.5. *Alpha-Acid-Glycoprotein*).

Analyzing desialylated AGP-1 and AGP-2 did not only reveal individual differences in glycosylation as the degree of fucosylation or number of antennae. The procedure presented here offers an easy method for analyzing genetic makeup as well as other post-translational modification like cysteinylolation. Improved ionization properties compared to fully sialylated AGP were observed in positive ion mode and a partial separation of genetic variants on a standard reversed phased column was achieved.

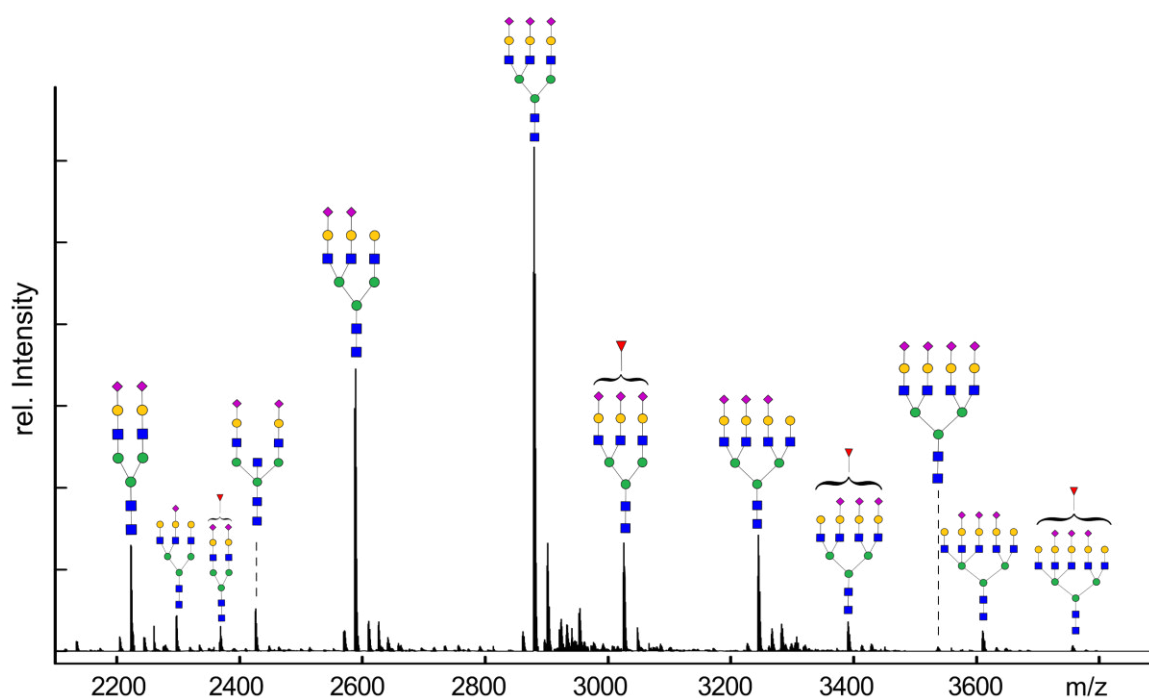
The fully sialylated AGP in negative ion mode was not analyzed because the purification method did not yield in a salt-free sample which drastically reduced ion intensity in negative ion mode.

### 5.5.3. Analysis of Free Glycans of Alpha-Acid-Glycoprotein

To analyze released glycans AGP from pooled human plasma was treated with endoglycosidase F (PNGase F) under reducing conditions. A C8-RP column coupled to an ESI-qTOF-MS was used to separate the free glycans from the deglycosylated protein. The free glycans showed similar retention behavior and

eluted at a time interval from 8 to 12 minutes. The mass spectra showed glycans with a charge of  $z=+2$  and  $z=+3$ . After summation of the mass spectra over this time interval a deconvolution was performed. Figure 38 shows the ESI-MS spectrum of the PNGase F released glycans of AGP from pooled plasma.

The most abundant signal corresponds to a fully sialylated triantennary glycan structure. This is in good agreement with previously reported data. In contrast, the amount of highly sialylated structures with a high number of antennae seems to be underrepresented compared to the results of intact AGP measured in negative ion mode. This finding can be explained by reduced ionization efficiency of highly sialylated glycan structures in positive ion mode with ESI-MS and possibly by a reduced activity of the enzyme to high antennary structures.[164, 165] Analyzing the intact glycoprotein in negative ion mode seems therefore favorable for quantifying highly sialylated glycan structures.



**Figure 38.** Deconvoluted ESI-MS spectrum of PNGase F released glycans of AGP from pooled human plasma. The dominant glycan composition corresponds to a fully sialylated triantennary glycan without fucosyl residues. Highly sialylated compositions are observed in lower quantity due to both reduced ionization and enzyme activity.[142]

## 5.6. Analysis of Human $\beta$ -2-Glycoprotein 1

Most parts in this chapter are currently under revision for publication in the journal *BBA Proteins and Proteomics*. [166]

The plasma protein  $\beta$ -2-glycoprotein 1 (B2GP1) is a 50 kDa glycoprotein with four or five occupied N-glycosylation sites, also referred to as apolipoprotein H. [167-169] The physiological functions of  $\beta$ -2-glycoprotein 1 include the interactions with heparin, binding to platelets and triggering the blood coagulation cascade. [170-172]  $\beta$ -2-Glycoprotein 1 plays also a key role in the development of antiphospholipid syndrome (APS). Patients diagnosed with the APS showed increased levels of anti- $\beta$ -2-glycoprotein 1 antibodies associating B2GP1 with APS as most important antigen. [169] This widespread autoimmune disease, primarily observed in women, can cause thrombosis, strokes, and miscarriages in pregnant women. [173, 174]

The interaction of  $\beta$ -2-glycoprotein 1 and its auto antibodies is triggered by conformational changes of  $\beta$ -2-glycoprotein 1. The most abundant conformation of  $\beta$ -2-glycoprotein 1 is a J-shaped conformation. However, circular and S-shaped conformations have been postulated shielding the antibody binding epitope. [175-177] The glycosylation of  $\beta$ -2-glycoprotein 1 plays thereby a crucial role in covering the antigenic binding epitope of G40-R43. Adhesion of B2GP1 to a negatively charged surface like anionic phospholipids can lead to conformational changes and exposure of the antigen binding epitope. The subsequent binding of anti- $\beta$ -2-glycoprotein 1 antibodies can lead to symptoms like thrombosis and pregnancy morbidity. [178] The role of glycosylation of  $\beta$ -2-glycoprotein 1 in this mechanism has been studied by Dupuy d'Angeac *et al.*, who showed that the biotinylation of glycan residues of  $\beta$ -2-glycoprotein 1 leads to antibody recognition. [179]

In 2009, Kondo *et al.* showed that patients diagnosed with APS when compared to healthy individuals show a different glycosylation pattern of  $\beta$ -2-glycoprotein 1. [180] In patients diagnosed with APS, glycosite  $^{143}\text{N}$  of  $\beta$ -2-glycoprotein 1, which is close to the arginine-rich antibody binding epitope, showed a decreased sialylation. This might alter the electrostatic properties in this

region due to the reduced number of negative charges from sialic acids and enhance antibody interaction.

However, antibody recognition is not only determined by the glycosylation pattern but also by the amino acid sequence. Different allelic variants are known for  $\beta$ -2-glycoprotein 1, leading to 150 different single nucleotide polymorphisms (SNP). Out of these 150 SNPs, eight are in the coding region.[181] One of the most abundant SNP is V247L with an allele frequency of 0.765 for valine and 0.235 for leucine.[182] APS patients showed a prevalence for the homozygous expression of valine, indicating a higher risk of suffering from APS. [183, 184] Furthermore, increased levels of oxidized  $\beta$ -2-glycoprotein 1 were found in patients suffering from APS and  $\beta$ -2-glycoprotein 1 concentration was proposed as diagnostic marker.[185, 186]

The biological and clinical relevance of different genetic variants and post-translational modifications like glycosylation and oxidation of  $\beta$ -2-glycoprotein 1 in terms of antibody recognition and disease development elucidate the need of a fast and unambiguous method for structural analysis. Therefore,  $\beta$ -2-glycoprotein 1 was analyzed as intact glycoprotein by ESI-MS to detect different genetic variants and post-translational modifications.

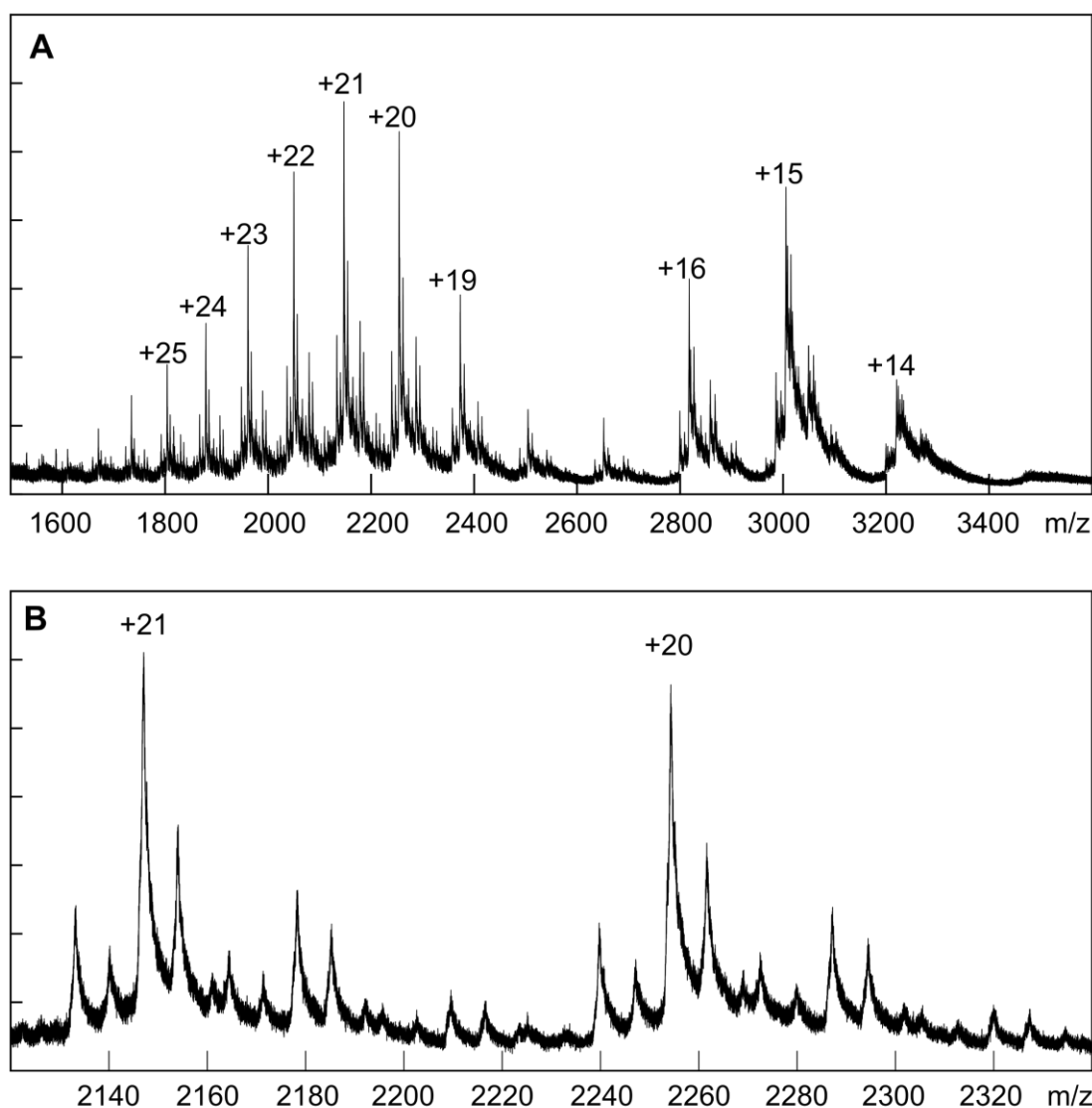
### **5.6.1. Genetic Variants of $\beta$ -2-Glycoprotein 1**

$\beta$ -2-glycoprotein 1 from 16 plasma samples was analyzed as intact glycoprotein to identify different genetic variants and glycosylation pattern simultaneously. This way sample preparation is reduced compared to conventional bottom-up approaches involving proteolytic or another digest and different proteoforms can be detected easily.[65-67] Analyzing the intact glycoprotein is also advantageous for the relative quantification of glycan structures. As demonstrated before, quantification of glycan structures is more precise using top-down mass spectrometry compared to bottom-up analysis of free glycans or glycopeptides.[83, 187, 188]



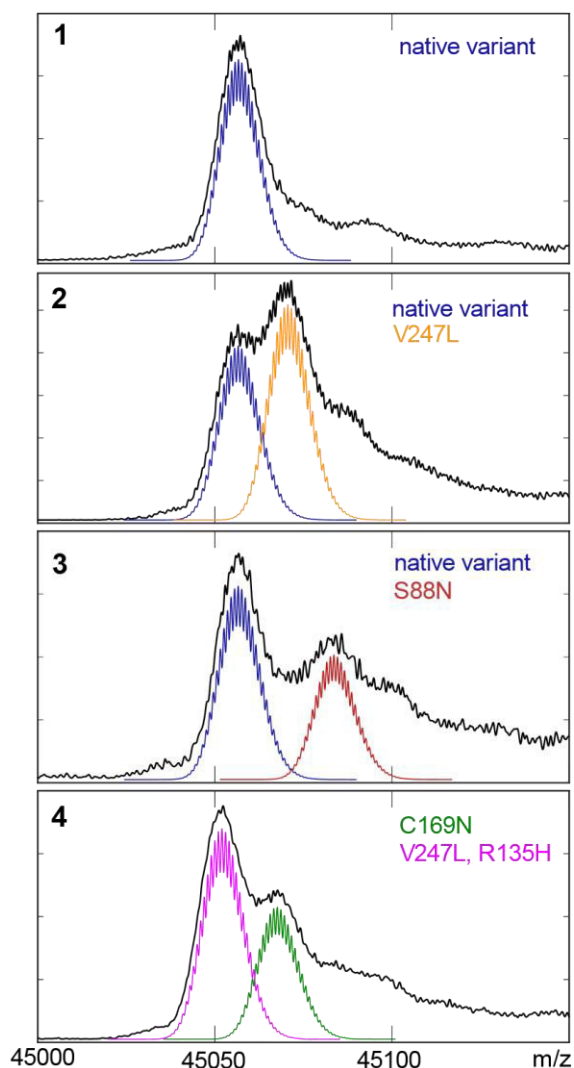
The ESI-MS spectra of intact  $\beta$ -2-glycoprotein 1 showed highly charged ion species ranging from +14 to +25 as shown in Figure 39, panel A. The distribution of charge states shows two maxima at a charge of +21 and +15, indicating different conformational states of B2GP1.[189-191] Since the experiments were not performed under native conditions for ESI-MS measurement (for example LC solvent with a pH of 3.0), partial or complete denaturation might have occurred. However, using ESI-MS under native conditions might provide additional insight into protein conformation. Applying ion mobility techniques like drift tube ion mobility or high-field asymmetric waveform ion mobility in combination with mass spectrometry has proven to be most useful for the analysis of protein conformation.[192, 193] This might enable the distinction between J-shaped, S-shaped and circular  $\beta$ -2-glycoprotein 1.

Panel B in Figure 39 shows the expansion of the spectrum in panel A. Different peaks corresponding to different proteoforms can be identified in every charge state. To simplify data interpretation and enhance signal-to-noise ratio, the spectra were deconvoluted using a maximum entropy algorithm.[194] By this, a spectrum with a virtual charge of +1 was obtained from all observed charge states and used further on for qualitative and quantitative determination of all proteoforms.



**Figure 39.** Exemplary ESI-MS spectrum of B2GP1 from human plasma. The donor is homozygous for the native gene. **(A)** The charge envelope shows two maxima at a charge of +21 and +15, indicating different conformational states of B2GP1. **(B)** Expansion of A. For each charge state, different peaks corresponding to different proteoforms caused by glycans are detected.[166]

Different SNPs are known for  $\beta$ -2-glycoprotein 1, leading to an amino acid exchange in the protein sequence. These amino acid exchanges show specific shifts in mass as for example a shift of +14 Da for SNP variant V247L. The dominant glycan on  $\beta$ -2-glycoprotein 1 is known to be a biantennary, fully sialylated structure (H5N4S2), which accounts for about 58% of all glycan structures. [180, 195] To identify the different genetic variants present in the samples, the isotopic distribution for different amino acid sequences and an occupation with the glycan composition of H5N4S2 on all four glycosylation sites was simulated. Since the protein was not reduced, 11 disulfide bridges were considered for the chemical formula.[196] Figure 40 shows the most abundant peak for four samples, corresponding to a proteoform with the glycan composition H5N4S2 attached to all four glycosylation sites (black). The simulated peaks for different genetic variants are shown in different colors.



**Figure 40.** Excerpt of deconvoluted ESI-MS spectra of human B2GP1 from four different plasma samples (1 to 4) in black. Shown is the most abundant peak corresponding to a proteoform with biantennary, fully sialylated glycans on all four glycosylation sites. **1:** Donor N shows the native homozygous variant. The simulated peak for the native variant is shown in blue. The main peak shows broadening at higher masses which can be explained by sodium adducts and oxidation. **2:** Donor B shows the native variant as well as the SNP variant V247L (simulated in yellow). **3:** Donor D shows the native variant as well as SNP variant S88N (simulated in red). **4:** Donor E shows the variant C169N (simulated in green) and a variant with two SNPs, namely V247L and R135H (simulated in magenta).[166]

As shown in Figure 40, different intensities are observed for different variants, presumably due to different concentrations of  $\beta$ -2-glycoprotein 1 variants in plasma as reported earlier.[197] As mentioned previously, explanations like differences in ionization properties or an altered abundance of proteoforms due to the purification procedure cannot be excluded. The native variant was detected in 14 out of 16 samples either homozygous or heterozygous. Three of the samples showed variant V247L which is less than the reported allele frequency of 23% but can be explained by the limited number of samples that were analyzed. Table 13 summarizes the variants found in the samples.

**Table 13.** Genetic variants of B2GP1 found in the 16 analyzed samples.[166]

Donor	native variant	V247L	S88N	R135H	C169N
B	x	x	-	-	-
C	x	x	-	-	-
E	x	-	-	-	-
G	x	-	x	-	-
M*	-	x	-	x	x
O	-	-	-	-	x
P	x	-	-	-	-
S	x	-	x	-	-
T	x	-	-	-	-
U	x	-	-	-	-
V	x	-	-	-	-
W	x	-	-	-	-
Y	x	-	-	-	-
Z	x	-	-	-	-
AB	x	-	-	-	-
AC	x	-	-	-	-

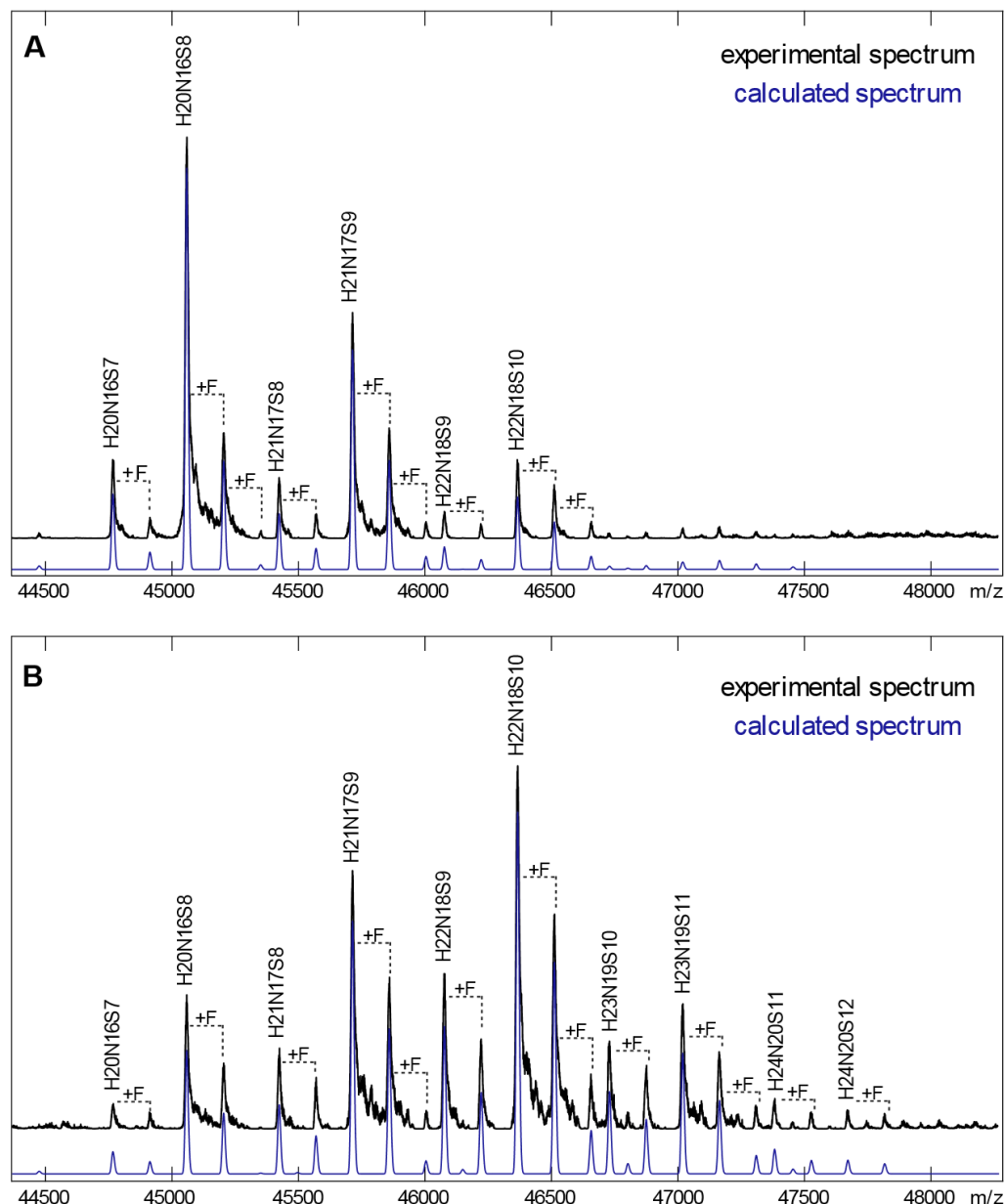
\*Donor M showed a heterozygous expression for variant C169N and a variant with two SNPs, V247L and R135H.

Variants V247L, S88N and R135H correspond to ddSNP entry rs4581, rs1801693 and rs8178847 respectively. [182, 198, 199] Their minor allele frequencies are 0.28, 0.03 and 0.07, respectively.[181] Variant C169N does not correspond to a ddSNP entry but has been reported as amino acid sequence by Lozier *et al.*[168] By analyzing  $\beta$ -2-glycoprotein 1 on intact protein level, sequence variants are easy to detect without genetic information. Commonly used bottom-up techniques can be used to detect genetic variants on peptide level, but automated peptide identification often lacks the ability to detect SNPs as already demonstrated for human ceruloplasmin.[83] Additionally, post-translational modifications like glycosylation interfere with automated peptide identification, which is especially the case for heavily glycosylated  $\beta$ -2-glycoprotein 1.

### 5.6.2. Glycosylation of $\beta$ -2-Glycoprotein 1

$\beta$ -2-glycoprotein 1 carries four or five N-glycosylation sites. Four reported sites are on asparagine 143, 164, 174, and 234.[181, 197, 198, 200] Lozier *et al.* reported a fifth glycosylation site at <sup>169</sup>N.[168] Analyzing the intact glycoprotein results in a compositional identification of different glycan structures without glycosite specific information. However, distinct mass differences between proteoforms with different glycosylation can give information about glycan antennarity or repeat structures as well as the degree of sialylation and fucosylation. Since the degree of sialylation of  $\beta$ -2-glycoprotein 1 is associated with a higher risk of APS and an increased expression of the Sialyl-Lewis X motive has been observed on many glycoproteins in patients with different cancer diseases, analyzing the intact glycoprotein can be a helpful tool for biomarker development as fast and reliable method.[180, 201] 15 of the samples showed a proteoform that can be explained by a glycan composition of H20N16S8 as most intense signal. On four glycosylation sites, this composition can be explained by all biantennary, fully sialylated structures (H5N4S2). With a mass distance of +657 Da, a glycan composition of H21N17S9 can be found which can be explained by the addition of one sialylated antenna. Next to mass distances corresponding to the addition of one or more antennae, fucosylated species with a mass shift of +146 Da as well as the loss of one neuraminic acid with a difference of -291 Da were observed. Since previous reports on the glycosylation of  $\beta$ -2-glycoprotein 1 described dominantly biantennary and triantennary complex type N-glycans, visualization of the glycan composition was based on these findings.[168, 195] Figure 41 shows two exemplary, deconvoluted ESI-MS spectra of intact  $\beta$ -2-glycoprotein 1 from two different individuals. As mentioned above, almost all samples showed a proteoform with H20N16S8 as most intense peak (panel A, Figure 41). Only one sample showed a different glycosylation pattern in terms of antennarity. Panel B in Figure 41 shows the deconvoluted ESI-MS spectrum of  $\beta$ -2-glycoprotein 1 from donor P. The most intense peak corresponds to a proteoform with H22N18S10 which can be visualized as for example two biantennary and two triantennary, fully sialylated glycan structures. As also observed for alpha-acid-glycoprotein (Figure 37), ceruloplasmin (Figure 18) and immunoglobulin IgG1 (Figure 27), glycosylation patterns are shifted to structures

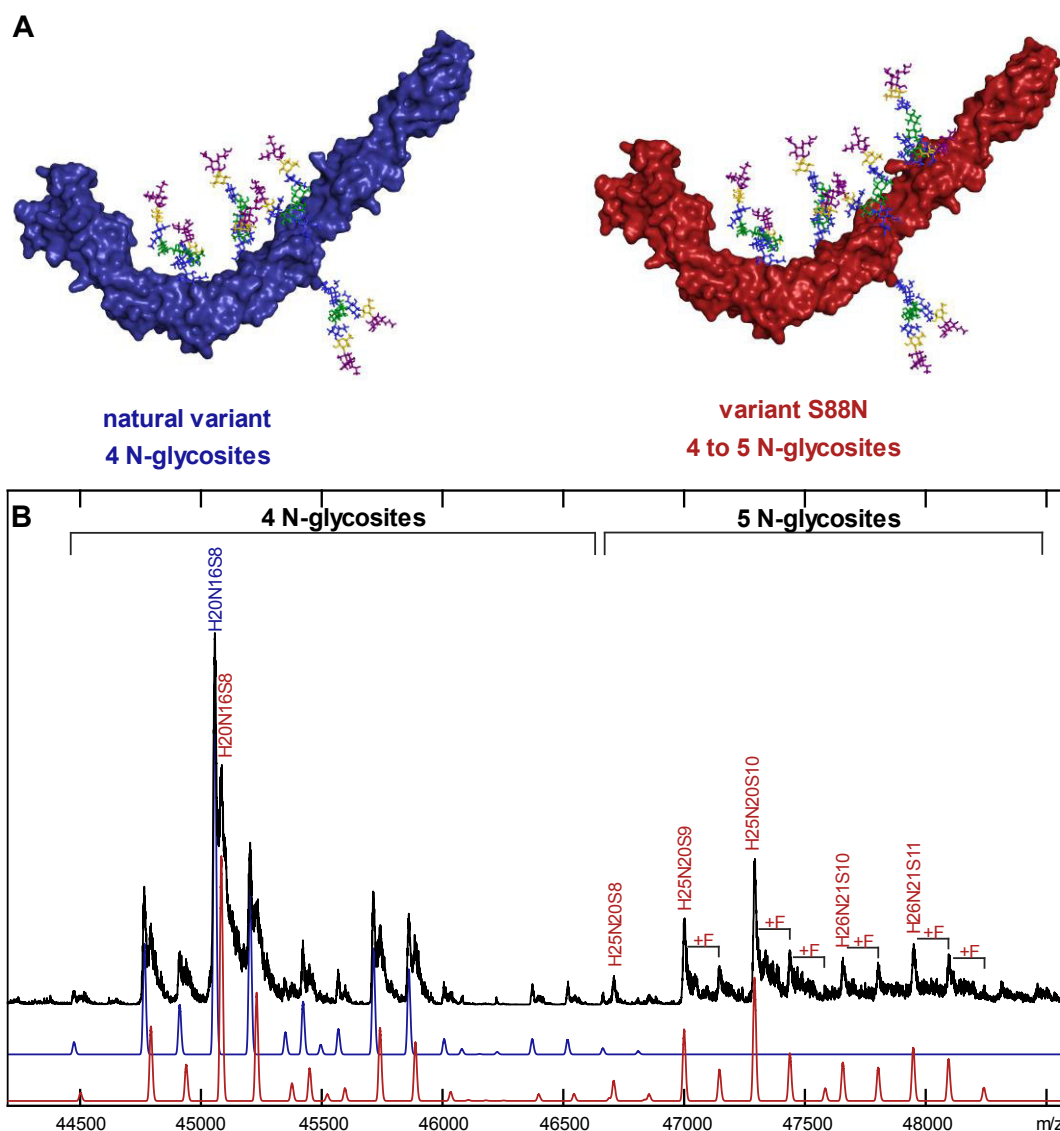
with increased antennarity or repeats, which was only observed in samples from this donor. No signals corresponding to O-glycosylation were detected in any of the samples as reported by Gambino *et al.*[202]



**Figure 41.** Deconvoluted ESI-MS spectra of human B2GP1 from two different donors. The experimental spectra are shown in black, the simulated spectra are shown in blue. Most of the analyzed samples showed a glycosylation pattern like donor AB displayed in panel **A** with the most intense signal being a proteoform with H20N16S8, which can be explained by four fully sialylated, biantennary glycan structures. Fucosylated species with a mass difference of +146 Da are marked with +F. B2GP1 from donor P showed a general increase in antennarity (panel **B**). The most intense peak corresponds to a proteoform with H22N18S10, which can be realized as two biantennary and two triantennary, fully sialylated glycan structures.[166]

Among others, a SNP variant with a mass difference of +27 Da was observed. This mass difference can be explained by the amino acid exchange from serine to asparagine as reported for position 88 (S88N, Figure 40 panel C).[198] The SNP S88N (rs1801693) leads to the formation of an additional consensus sequence for N-glycosylation, i.e. Asn-Phe-Ser. Two of the samples showed this SNP and in consequence exhibit signals in addition to the normal glycosylation pattern shown in Figure 41, panel A. These additional signals have a mass of about 2200 Da higher corresponding to a fifth glycosylation site. This variant showed in a ratio of about 60:40 a species with four glycans and a species with five glycans, respectively Figure 42 (panel B). The results indicate a partial occupation of this additional glycosite with complex-type N-glycans with the most abundant glycan composition being H25N20S10 if all five glycosylation sites are occupied. This composition can be visualized by five biantennary, fully sialylated glycans. At these higher masses, no peaks for the native variant are observed and only signals for the SNP variant are detected.

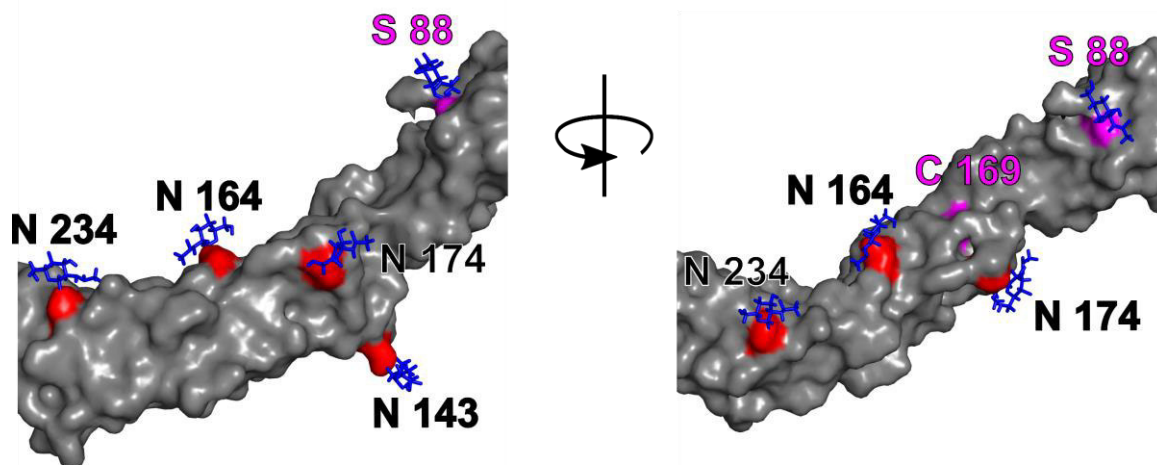




**Figure 42.** Number of N-glycosylation sites of B2GP1. **A:** Protein model for human B2GP1 with either four or five N-glycosylation sites. Protein model is modified from PDB entry 1C1Z and shows a J-shaped form. However, B2GP1 can undergo conformational changes for antibody binding and is also present in low amounts in S-shaped and circular conformation.[203, 204] Glycan residues are shown as sticks in colors corresponding to CFG nomenclature.[3] **B:** Deconvoluted ESI-MS spectrum of two variants of human B2GP1. One variant shows the native amino acid sequence with four glycosylation sites (blue protein in panel **A** and simulated blue spectrum in panel **B**). Variant S88N (red protein in panel **A** and simulated red spectrum in panel **B**) with a mass shift of +27 Da shows signals for four glycosylation sites as well as signal for five occupied glycosylation sites at higher masses.[166]

This SNP has so far not been reported to carry an additional glycan. Furthermore, no signals corresponding to a fifth glycosylation for variant C169N as reported by Lozier *et al.* were detected.

To identify asparagine residues that are accessible for enzymatic glycosylation, the accessibility at the surface for asparagine residues that are potentially glycosylated, i.e. that are part of a consensus sequence, was determined. B2GP1 shows in its native form four consensus sequences for N-glycosylation (<sup>143</sup>N, <sup>164</sup>N, <sup>174</sup>N, and <sup>234</sup>N). SNP variants S88N and C169N both create additional consensus sequences. The surface area for the six residues was determined using the open source Python script “*findSurfaceResidues*” by Jason Vertrees, applied to PDB structure 1c1z which shows a J-shaped conformation for human  $\beta$ -2-glycoprotein 1.[205] Exposed surface areas were determined as 46 Å<sup>2</sup>, 32 Å<sup>2</sup>, 29 Å<sup>2</sup> and 32 Å<sup>2</sup> for N-glycosylation sites <sup>143</sup>N, <sup>164</sup>N, <sup>174</sup>N, and <sup>234</sup>N, respectively. Serine at position 88 showed a similar size of its surface area of 33 Å<sup>2</sup>, whereas cysteine at position 169 showed a significantly less exposed surface area of 3.3 Å<sup>2</sup>. This indicates that position 88 is more assessable for enzymatic glycosylation than position 169 as visualized in Figure 43. The calculation was based on the native cysteine and serine residues. However, an amino acid exchange to asparagine might change the 3D structure.



**Figure 43.** Surface areas of the four glycosites N143, N164, N174 and N234 in red. Potential glycosites for SNP variant S88N and C169N are shown in magenta. The first N-acetylglucosamine of the attached N-glycans is shown in blue for the four reported glycosites and for SNP variant S88N. Position C169 is in this 3D structure not accessible for enzymatic glycosylation. Structure of human B2GP1 is in J-shaped conformation from PDB entry 1c1z.[166]

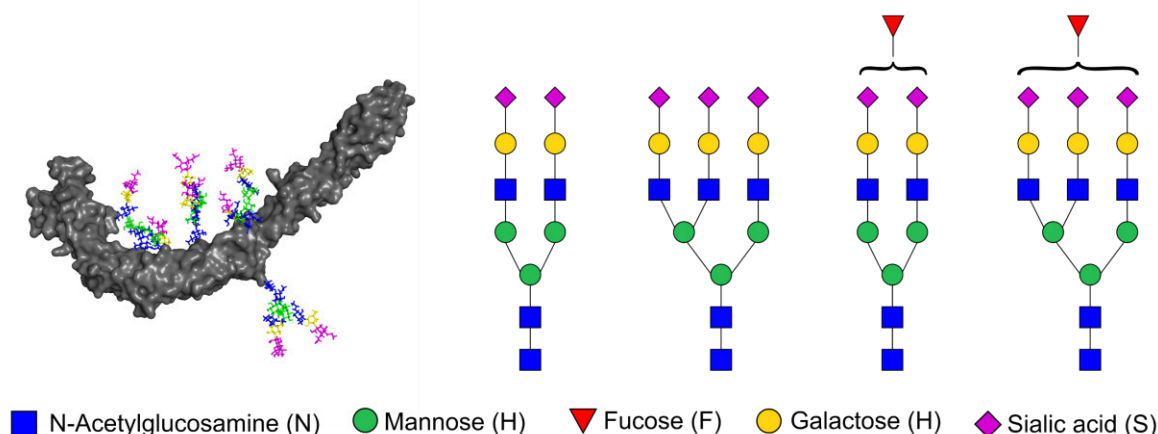
As mentioned above, most of the samples showed similar glycosylation patterns with exception of  $\beta$ -2-glycoprotein 1 from donor P with increased antennarity and the SNP variant S88N with a fifth, partial occupied glycosylation site observed in heterozygous expression in two samples. Table 14 summarizes the proportions for the most abundant glycan compositions (over 90%) with exception of sample P and SNP variant S88N, which showed different glycan compositions. ESI-MS spectra for all analyzed samples and a complete list of all detected glycan compositions including their relative proportions can be found in the Appendix, section 10.6.  *$\beta$ -2-Glycoprotein 1*.

Table 14. Relative proportions of glycan compositions found on human B2GP1 for different variants from different donors.[166]

Donor	Age	4 biantennary glycans [%]				3 biantennary glycans, 1 triantennary glycan [%]				2 biantennary glycans, 2 triantennary glycans [%]	
		H20N16S8	H20N16S8F	H20N16S7	H20N16S7F	H21N17S9	H21N17S9F	H21N17S8	H21N17S8F	H22N18S10	H22N18S10F
B, native	60	40.3	17.2	7.3	3.2	8.2	8.1	3.0	2.0	0.9	1.2
B, V247L	60	39.6	17.5	7.4	3.5	8.0	8.0	2.8	2.1	0.9	1.2
C, native	34	36.6	9.3	10.2	2.1	13.3	8.3	4.4	1.7	3.6	3.2
C, V247L	34	34.8	10.3	10.4	3.1	12.4	8.4	4.3	2.3	3.2	3.0
E, native	26	29.2	9.4	10.7	3.4	12.2	9.1	5.5	3.7	3.1	2.8
G, native	58	34.5	15.5	10.2	4.6	9.9	8.0	4.9	2.4	1.4	1.4
M, C169N	45	30.5	12.7	5.0	2.2	17.4	10.3	3.7	2.0	5.4	3.8
M, V247L, R145H	45	29.5	12.1	5.5	2.1	17.0	10.1	3.9	2.1	5.0	3.6
O, native	54	28.5	13.9	5.3	2.6	14.0	9.7	3.4	2.2	5.3	4.2
S, native	46	32.3	12.9	10.9	4.2	9.9	9.2	5.7	3.6	1.6	2.0
T, native	47	26.5	7.4	8.4	1.8	15.5	7.4	6.7	2.2	7.2	4.2
U, native	28	24.3	6.8	4.9	1.5	16.8	9.3	5.1	2.4	7.8	5.5
V, native	52	28.3	10.7	7.7	3.0	12.5	9.9	4.8	3.2	3.9	3.5
W, native	23	27.1	10.5	5.8	2.2	13.4	10.3	4.1	2.7	4.6	4.6
Y, native	40	25.5	10.0	4.2	1.6	12.2	12.0	3.4	2.9	4.4	5.2
Z, native	39	31.6	8.6	6.8	1.7	17.7	6.9	5.2	1.7	7.4	3.3
AB, native	47	32.5	8.8	6.1	1.4	17.9	8.9	4.5	1.7	6.0	3.9
AC, native	67	32.0	10.9	9.9	3.9	12.0	7.9	5.2	3.0	3.2	2.5
Average <sup>a</sup>	<b>44</b>	<b>31.3</b>	<b>11.4</b>	<b>7.6</b>	<b>2.7</b>	<b>13.4</b>	<b>9.0</b>	<b>4.5</b>	<b>2.4</b>	<b>4.2</b>	<b>3.3</b>
STD <sup>b</sup>	<b>13</b>	<b>4.5</b>	<b>3.1</b>	<b>2.3</b>	<b>1.0</b>	<b>3.2</b>	<b>1.2</b>	<b>1.0</b>	<b>0.6</b>	<b>2.1</b>	<b>1.3</b>
R <sup>c</sup>		<b>0.50</b>	<b>0.63</b>	<b>0.20</b>	<b>0.52</b>	<b>-0.40</b>	<b>-0.31</b>	<b>-0.15</b>	<b>-0.10</b>	<b>-0.43</b>	<b>-0.56</b>

<sup>a</sup>Values represent over 90% of all detected glycan compositions. Samples P, G (S88N), and S (S88N) were excluded from this table and average values and can be found in section 10.6.  *$\beta$ -2-Glycoprotein 1*, table A16. Further compositions include species with two fucosyl residues and additional antennae in low amounts. <sup>b</sup>Standard derivation of average values. <sup>c</sup>Correlation coefficient R of glycan compositions with age.

Even though a compositional analysis of glycan structures cannot differentiate between e.g. two biantennary and two triantennary glycans compared to three biantennary and one tetraantennary glycan, the results are in good agreement with previously reported data, describing dominantly biantennary and triantennary glycans.[171, 180, 202] The majority of the identified glycan composition, can be described as a combination of biantennary and triantennary glycan structures with or without fucosyl residues as shown in Figure 44. This is also true for variant S88N (samples G and S) with a partially occupied, fifth glycosylation site. Fully sialylated biantennary and triantennary structures were detected as most abundant peaks. However, signals indicating an incomplete sialylation for one site with less than maximal sialic acids were also detected. Statistically sialylation is about 20% less than the maximal amount of sialylation (cf. Table 14). This finding could also be interpreted in terms of the presence of repeat structures which would reduce the number of sialic acid residues.



**Figure 44.** Glycan composition of human B2GP1 can be described by a combination of bi- and triantennary glycans with or without fucosyl residues. The degree of fucosylation is on average 35% on all biantennary glycans and increases with increasing number of antennae. Glycan compositions indicating the loss of a neuraminic acid or repeat structures were detected as well in lower abundance.

The results show that the degree of fucosylation is on average 35% for all biantennary glycans and increases with increasing number of antennae. While slight differences in relative abundances are detected between the samples from

different donors, no different glycosylation patterns for different genetic variants were detected within one sample (for example native variant and variant V247L from donor B). Correlation of the detected glycan composition with the age of the donors (R values in Table 14) showed a decrease in higher antennarity glycans (H21N17 and H22N18), represented by a negative **R** value. At the same time, glycans with the composition, which can be realized as all diantennary structures, increase with R values from 0.20 for H20N16S7 to 0.63 for H20N16S8F. For these structures, fucosylated species show a higher increase with age. Increased fucosylation of plasma proteins with age has also been reported earlier.[206]

## 5.7. Glycosylation of Plasma Proteins as Potential Biomarker

The glycosylation of plasma proteins can be used as diagnostic biomarker for different types of diseases since glycosylation can change due to different physiological states. Various publications have focused on correlation between glycosylation of a specific protein or the whole plasma N-glycome with a certain disease (Table 15). Especially for cancer diseases, various correlation studies showed an increased fucosylation of many different proteins. But other diseases like arthritis or liver cirrhosis show an expression of altered glycosylation patterns as well. It is, however, questionable whether these changes are specific to a certain disease or a more general observation for a variety of diseases. Table 15 list reported changes in glycosylation for ceruloplasmin, IgG, transferrin, alpha-acid-glycoprotein and  $\beta$ -2-glycoprotein 1 from only a few publications.

**Table 15.** Reported changes in glycosylation for different plasma proteins due to different physiological conditions like disease and inflammation.

Protein	Changes in glycosylation
Ceruloplasmin	<b>Increased fucosylation (e.g. Lewis X) in:</b> hepatocellular carcinoma [14], pancreatic cancer [92], ovarian cancer [207]
IgG	<b>Increased levels of agalacto glycans in:</b> ovarian cancer [124, 149, 208], gastric cancer, liver cancer, lung cancer, colorectal cancer, esophageal cancer, pancreatic cancer, prostate cancer [123], cholangiocarcinoma metastasis [124], hepatocellular carcinoma and cirrhosis [125], rheumatoid arthritis [126], systemic lupus erythematosus, Crohn's disease [127]
Transferrin	<b>Increased fucosylation (e.g. Lewis X) in:</b> hepatocellular carcinoma [131], pancreatic cancer [134] <b>carbohydrate deficient transferrin in:</b> pregnancy and alcoholism [12]
AGP	<b>Increased fucosylation (e.g. Lewis X) in:</b> hepatocellular carcinoma and cirrhosis [146], ovarian cancer [149], lung cancer [147], breast cancer [209], colon cancer [210], pancreatic cancer [211], bladder cancer [212], inflammatory conditions [15]
B2GP1	<b>Decreased sialylation in</b> antiphospholipid syndrome (APS) [180]

An increased fucosylation, often Lewis X or Sialyl-Lewis X, has been reported for ceruloplasmin, transferrin, alpha-acid-glycoprotein and  $\beta$ -2-glycoprotein 1. Immunoglobulin G shows elevated levels of agalacto glycans in several cancer diseases but also in inflammatory diseases like rheumatoid arthritis or lupus and Crohn's disease. The data suggest, that up or down regulation of glycosyltransferases occur under certain physiological conditions which affects several glycoproteins. However, most publications focus on a specific target protein for correlation studies. In this work, five different plasma proteins were analyzed to see if changes in glycosylation are related to a specific protein or can be observed as general feature of plasma proteins. This affects biomarker development and leads to a better understanding about glycosylation and its changes under various physiological conditions.

### **5.7.1. Unusual Glycosylation Patterns of Plasma Sample P**

As demonstrated for plasma sample P, an increased expression of highly branched and highly sialylated structures were found not only for one protein but for all five analyzed proteins. However, the extent of this change varies between the proteins analyzed. For IgG1, the amount of sialylated structure G2FS increased from 11% on average to 17%. Transferrin showed an increased expression for glycan composition H11N9S5 (one biantennary and one triantennary structure) of 11% compared to 3.3%. Even though alpha-acid-glycoprotein was only analyzed as desialylated protein, an increased expression of highly branched or repeating LacNAc motifs was observed. Glycan composition corresponding to five tetraantennary structures increased from 3.8% on average to 8.7% for sample P. For B2GP1, compositions with 20 sialic acids increased from 3.8% on average to 8.7% for sample P. Ceruloplasmin showed for all analyzed glycosites increased levels of triantennary glycans and reduced values for diantennary glycans. As already mentioned above, involved glycosyltransferase are N-acetylglucosaminyltransferases, galactosyltransferases (IV and/or V) and sialyltransferases which lead to increased branching, galactosylation and

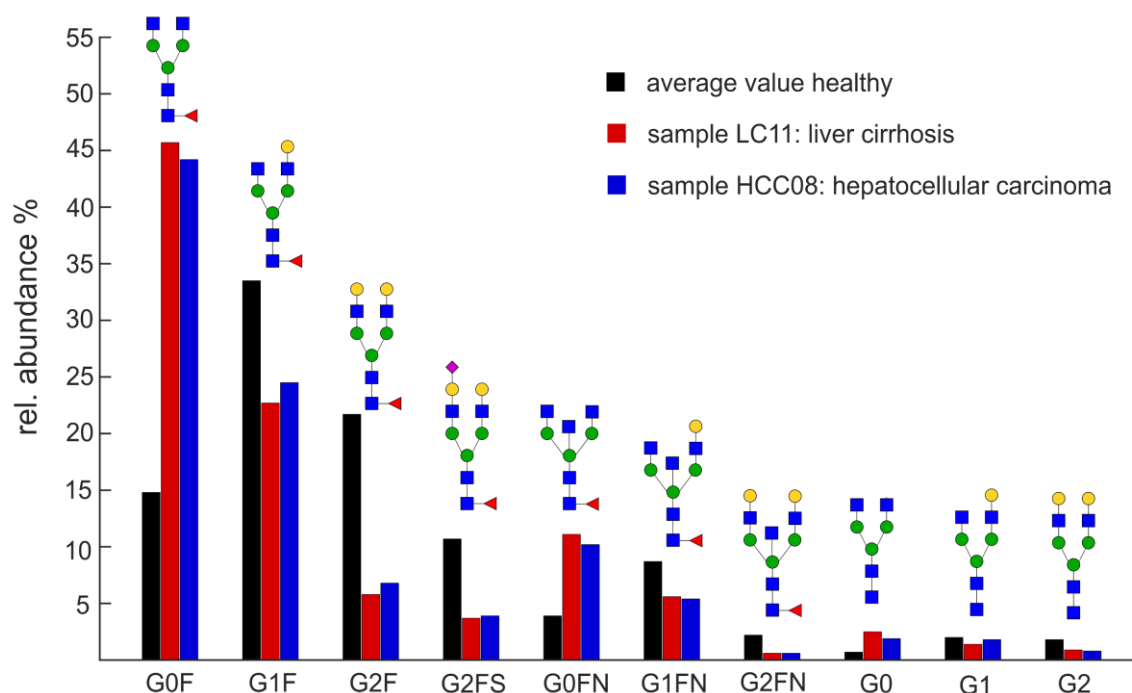
sialylation. Increased sialylation of proteins leads to a reduced affinity to the asialoglycoprotein receptor. Consequently, highly sialylated proteins are less susceptible to endocytosis and show lower clearance rates. For IgG, anti-inflammatory effects are related with an increased expression of sialylated structures. Therefore, sialylation might act as switch between pro-inflammatory and anti-inflammatory activity of IgGs.[129] But also for AGP, a correlation between number of antennae and therefore sialylation and inflammatory conditions was found. Next to an increase in fucosylation, an increased expression of diantennary structures was observed in patients during 'short-term' acute inflammation.[213] Increased levels of diantennary structures lead to a decreased overall sialylation and branching under acute inflammatory conditions. However, in 'long-term' chronic inflammation an increase in tri- and tetraantennary structures have been found on AGP as well as increased branching of other plasma proteins.[214, 215] The fact that increased sialylation of IgG is associated with anti-inflammatory effects and increased branching is associated with inflammation contrasts with the unusual glycosylation patterns in plasma sample P, where both IgG sialylation and increased branching of other plasma proteins are found. It remains unclear which physiological or genetic conditions are responsible for this unique glycosylation observed on several plasma proteins.

### **5.7.2. Changes in Glycosylation in Patients with Liver Cirrhosis and Hepatocellular Carcinoma**

Glycan-based biomarkers for cancer detection are promising class of biomarkers as for example shown for fucosylation of alpha-fetoprotein.[38] In a variety of cancer diseases, increased fucosylation and sialylation is observed on plasma proteins as listed in Table 15. To evaluate the potential use of different plasma proteins as cancer biomarker IgG1, alpha-acid-glycoprotein,  $\beta$ -2-glycoprotein 1 and transferrin from patients diagnosed with hepatocellular carcinoma and liver cirrhosis were analyzed and compared to the values obtained from healthy individuals. Samples from patients were previously analyzed by Nagel *et al.*, who



found altered phosphorylation and glycosylation of fibrinogen in patients diagnosed with liver cirrhosis and hepatocellular carcinoma.[216] Sample names were adopted from this study. An overview of the analyzed samples can be found in the experimental procedures, Table 17. Figure 45 shows the relative quantification of glycopeptides from IgG1 of one patient diagnosed with hepatocellular carcinoma and one patient diagnosed with liver cirrhosis compared to the average values obtained from six healthy individuals. Relative quantification was based on glycopeptides with a charge of +4 as described in section 5.3.1. *Relative Quantification of Glycopeptides from IgG1.*



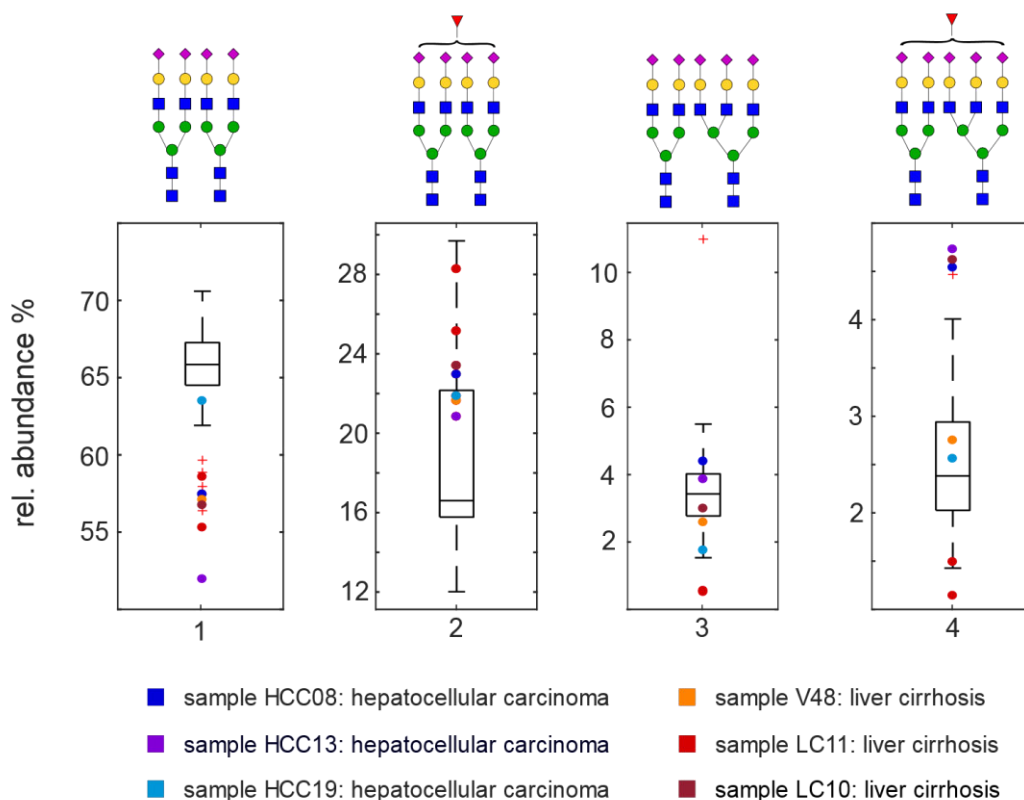
**Figure 45.** Changes in glycosylation of IgG1 observed in one patient diagnosed with liver cirrhosis (red, sample 43) and one patient diagnosed with hepatocellular carcinoma (blue, sample 11) compared to the average values found in six samples from healthy volunteers. Quantification was based on glycopeptides with a peptide charge of +4 as described in section 5.3.1. A decrease of galactosylated structures is observed for example in structure G0F for sample 43 and 11, which is in good agreement with literature (Table 15).

As already described in literature (c.f. Table 15), a tremendous decrease of galactosylated structures is observed in patients diagnosed with liver cirrhosis and

hepatocellular carcinoma. This immense difference would allow the development of a highly sensitive biomarker. However, low galactosylation and sialylation is also related to a state of inflammation and can lead to false-positive results and thus to a low specificity of a potential biomarker. This would also interfere with the differentiation between patients diagnosed with liver cirrhosis and hepatocellular carcinoma.

Changes in the glycosylation of transferrin are used as biomarker for long-term alcohol abuse where the concentration of carbohydrate deficient transferrin is increased. An increase in fucosylation of transferrin glycans is also associated with cancer diseases like hepatocellular carcinoma.[131] As described in section 5.4.1. *Analysis of Intact Transferrin*, transferrin shows a very homogeneous glycosylation with two biantennary, fully sialylated glycans (2x H5N4S2) as dominant modification. The variation among 25 analyzed plasma samples was also minor except for sample P (see section 5.4. *Analysis of Human Serum Transferrin*).

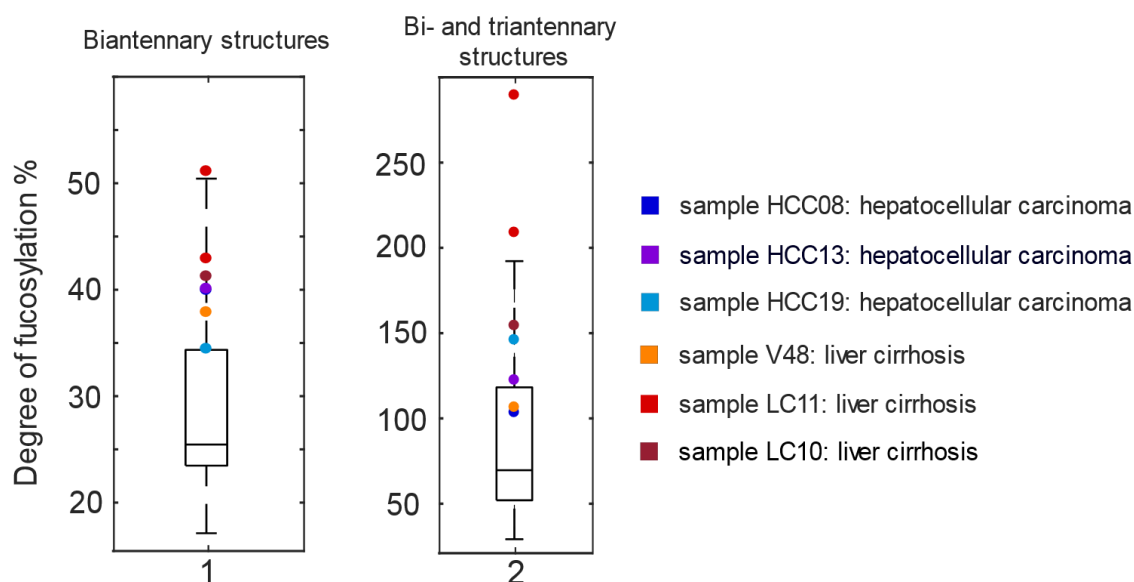
Transferrin was isolated from three plasma samples from patients diagnosed with hepatocellular carcinoma and from three samples from patients diagnosed with liver cirrhosis. Glycan analysis and relative quantification was done on the intact glycoprotein as described in section 5.4.1. *Analysis of Intact Transferrin* and compared to the values obtained from 25 healthy individuals. Figure 46 shows boxplot diagrams for the four most abundant proteoforms from 25 healthy individuals with overlaid data points from patients diagnosed with cirrhosis and hepatocellular carcinoma. Sample LC11 showed the heterozygous expression of SNP variant G277S next to the native variant. Values for both variants are shown separately.



**Figure 46.** Relative abundance of the four dominant proteoforms of human transferrin illustrated as boxplot diagram with a sample size of 25. Outliers are marked with a red plus. Relative abundances detected in samples from patients diagnosed with liver cirrhosis and hepatocellular carcinoma are overlaid. A general decrease in non-fucosylated, biantennary structures as well as an increased in fucosylated, biantennary structures is observed. Two data points are shown for sample LC11 which correspond to the natural variant and the SNP variant G277S.

As shown in Figure 46, decreased values of non-fucosylated, diantennary structures are observed and an increase in fucosylated diantennary structures. These findings are in good agreement with literature.[131] However, no trend for the differentiation between liver cirrhosis and hepatocellular carcinoma is observed. Compositions corresponding to bi- and triantennary structures don't show a distinction between healthy and diseased state which can be caused by their relatively low abundance. Compared to IgG, differences between samples from healthy volunteers and diseased patients are low which results in low sensitivity. It seems that the glycosylation of transferrin is less prone to changes as for example the glycosylation of IgG1. There is also a large variation of the data points from diseased patients as seen for the dominant proteoform. Values range

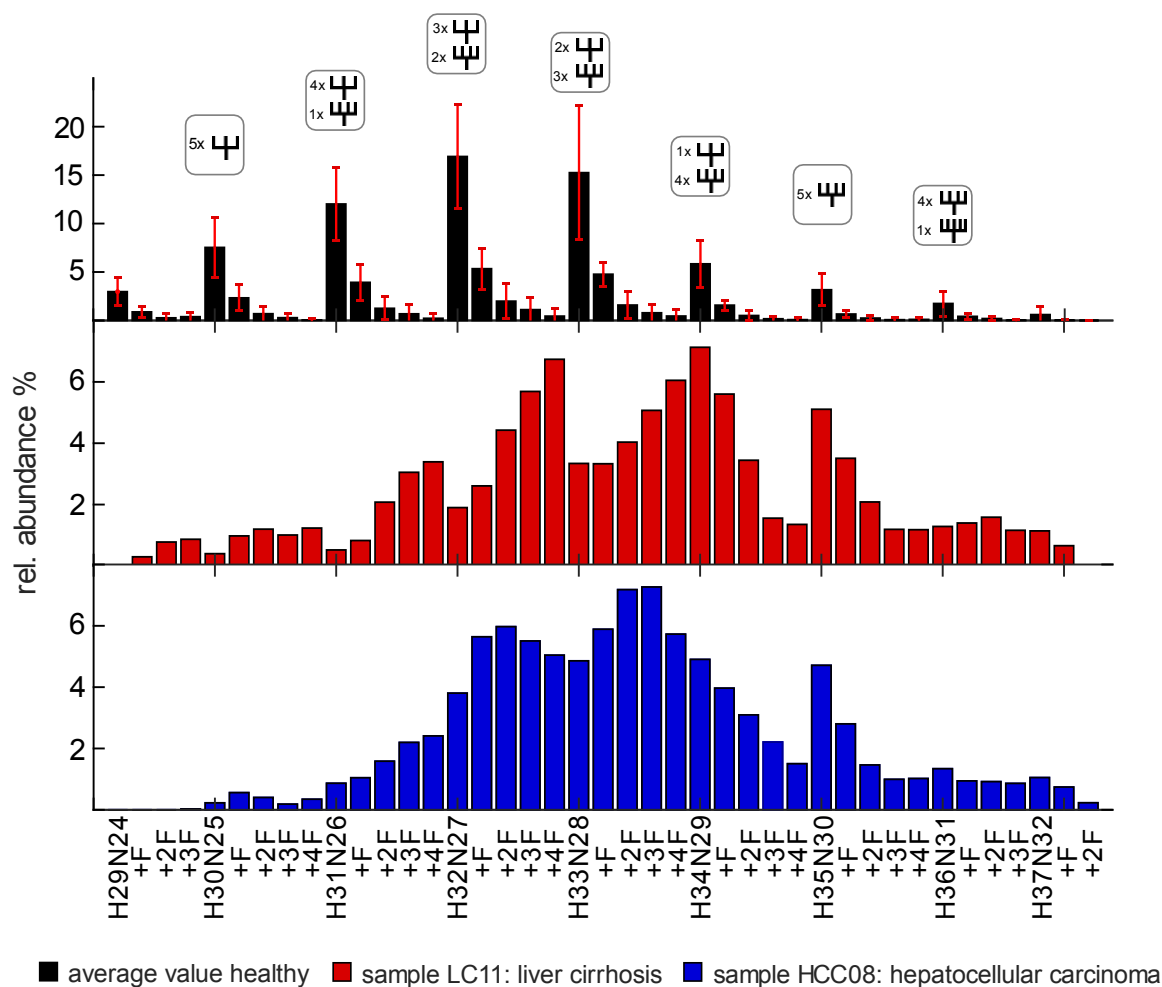
from 52% in sample HCC13 to 63.5% in sample HCC19 which is close to the average value from healthy individuals. To find a more distinct way of discrimination between the samples, the degree of fucosylation was determined by calculating the quotient of non-fucosylated and fucosylated structures for all biantennary structures and for bi- and triantennary structures. The average degree of fucosylation of all biantennary structures in healthy individuals is for example 29%, resulting from the quotient of the two proteoforms 1 and 2 shown in Figure 46. Figure 47 shows the boxplot diagrams for the degree of fucosylation for all biantennary structures and for bi- and triantennary structures from 25 healthy individuals with overlaid values from six samples from diseased patients.



**Figure 47.** Degree of fucosylation for all biantennary structures (2x H5N4S2, boxplot diagram 1) and for all bi- and triantennary structure (1x H5N4S2 and 1x H6N5S3, boxplot diagram 2) as boxplot diagram over 25 samples from healthy individuals. Data points from patients diagnosed with liver cirrhosis and hepatocellular carcinoma are overlaid. Samples from diseased patients show an increase in fucosylated structures, being in good agreement with previously reported data.[131] Two data points are shown for sample LC11, originating from a heterozygous expression of the natural variant and SNP variant G277S.

The degree of fucosylation between healthy individuals and patients diagnosed with liver cirrhosis and hepatocellular carcinoma shows a clear trend for the increased expression of fucosylated structures. Again, no trend is observed for the distinction between liver cirrhosis and hepatocellular carcinoma.

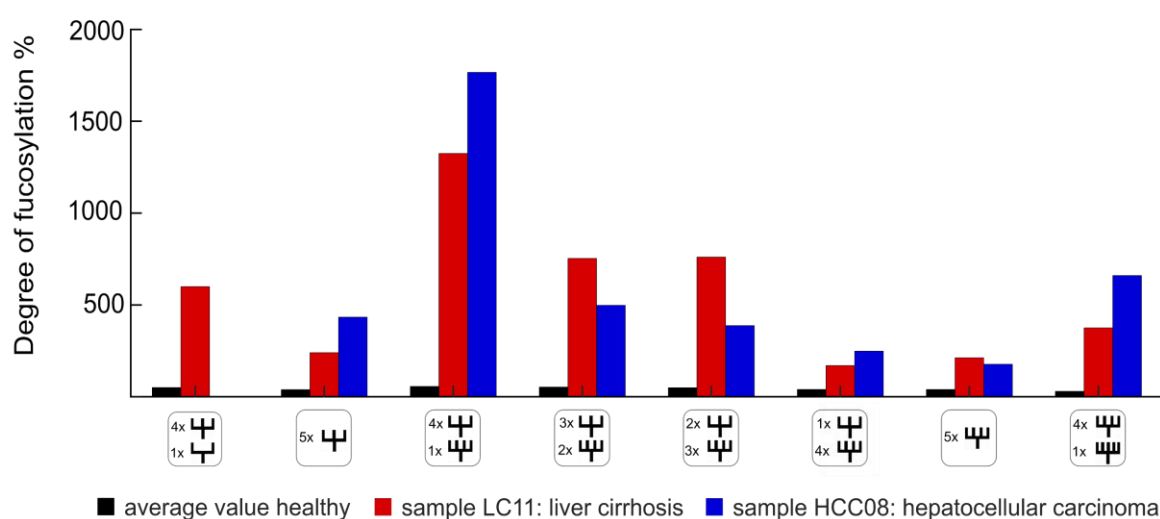
Alpha-acid-glycoprotein is a positive acute-phase protein. The increase in concentration of AGP due to an inflammatory state increases the total sialylation of the N-glycome from plasma proteins since high antennary structures (e.g. tetraantennary and pentaantennary structures) are dominantly expressed on AGP.[195] The expression of highly branched structures at five glycosylation sites results in a high number of potential positions for fucosylation. To monitor the increase in fucosylation of intact AGP after desialylation, one sample from a patient diagnosed with liver cirrhosis and one sample from a patient diagnosed with hepatocellular carcinoma were analyzed as described in section 5.5.2. *Analysis of Desialylated AGP from Individual Plasma Samples.* Figure 48 shows the relative abundances of glycan compositions detected for variant ORM2 of 11 healthy volunteers on average compared to the relative abundances in two samples from patients with liver diseases. Even though values are only displayed for desialylated variant ORM2, the variants ORM1\*F1, ORM1\*F2 and ORM1\*S show similar results. A complete list of the relative abundances of all variants can be found in the appendix, section 10.5. *Alpha-Acid-Glycoprotein.*



**Figure 48.** Relative abundances of glycan compositions from desialylated AGP (variant ORM2) over five glycosylation sites. Average values are obtained from 11 individuals (top). Values obtained from one patient diagnosed with liver cirrhosis (middle) and from one patient diagnosed with hepatocellular carcinoma (bottom) show a tremendous increase in fucosylated compositions as well as a shift to highly branched structures or structures with repeating LacNAc motifs. Similar results are obtained for AGP variants OMR1\*F1, OMR1\*F2 and OMR1\*S and can be found in the appendix 10.5. Alpha-Acid-Glycoprotein, Table A 9, Table A 10, Table A 11, respectively.

Analyzing desialylated AGP as intact glycoprotein allows the determination of glycan compositions over all five glycosylation sites at once. As shown in Figure 48, a tremendous increase in fucosylation is observed in samples from patients diagnosed with liver cirrhosis and hepatocellular carcinoma. Compositions with up to four fucosyl residues are observed in abundances as high as non-fucosylated compositions. However, considering five glycosylation sites, the number of fucosyl residues per N-glycan is much lower and averages zero to one fucosyl residue.

Nevertheless, the overall fucosylation increases from 32% in healthy individuals (average values for variant ORM2) up to 82% for sample LC11 and 80% for sample HCC08. Figure 49 compares the non-fucosylated structures to the sum of the fucosylated structures for different antennarity. The drastic increase in fucosylation is seen for both samples, LC11 and HCC08. However, antennarity does not correlated with fucosylation, which at first sight appears counterintuitive, as an increase number in antennarity or repeat units presents more possible positions for fucosylation. Other biochemical factors or steric hindrance for fructosyltransferases might influence the glycosylation pattern.



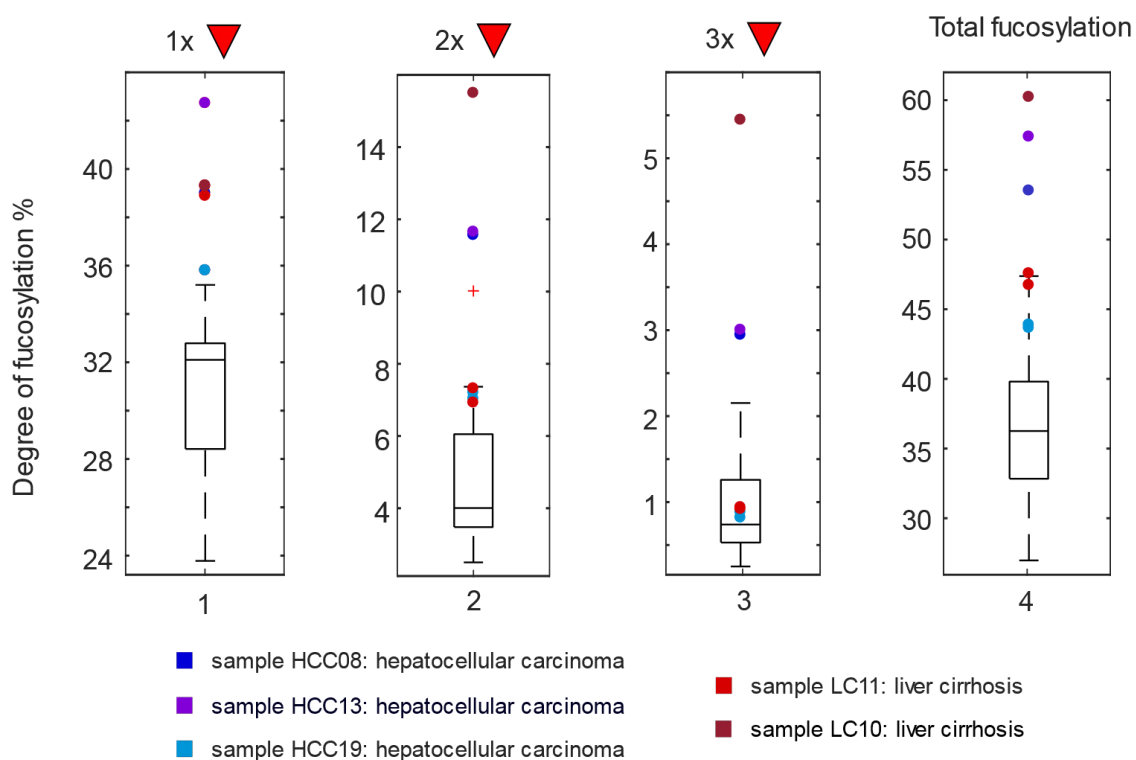
**Figure 49.** Degree of fucosylation by antennarity. Values obtained from patients diagnosed with liver cirrhosis or hepatocellular carcinoma show a tremendously higher amount of fucosylated structures compared to non-fucosylated structures. The highest degree of fucosylation is observed for composition H31N26 (e.g. realized by 4 triantennary and one tetraantennary structure). Sample LC11 shows a 13.3-fold higher abundance (1330%) of fucosylated structures compared to non-fucosylated structures. Sample HCC08 shows an even higher abundance with a 17.7-fold (1770%) increased abundance.

Analyzing the intact glycoprotein seems to be advantageous for the determination of overall changes in glycosylation like degree of fucosylation even if no glycosite specific information is obtained. However, as also reported for IgG1, increased values in fucosylation and sialylation of AGP are also observed in a variety of other cancer diseases as well as under inflammatory conditions, questioning the specificity of a potential biomarker. Both samples (sample HCC08 and LC11) show

the same trend towards the expressing of dominantly fucosylated structures, making it difficult to distinguish between patients diagnosed with liver cirrhosis and with hepatocellular carcinoma.

The glycosylation of  $\beta$ -2-glycoprotein 1 is not as well studied as for other plasma proteins like IgG, transferrin, ceruloplasmin and alpha-acid-glycoprotein. Analysis of glycopeptides from two of four to five glycosites showed a decrease of sialylation on glycosite 143N in patients diagnosed with APS.[180] However, no studies have been performed on the overall glycosylation or changes in glycosylation associated to other diseases or physiological conditions. In this work,  $\beta$ -2-glycoprotein 1 was analyzed as intact glycoprotein by ESI-MS, allowing the determination of glycan compositions from all four to five glycosites. To see if changes in glycosylation also occur on  $\beta$ -2-glycoprotein 1 during a diseased state, plasma samples from patients diagnosed with liver cirrhosis and hepatocellular carcinoma were analyzed as described in section 5.6.2. *Glycosylation of  $\beta$ -2-Glycoprotein 1*. As also observed for transferrin and alpha-acid-glycoprotein, an increase in fucosylated structures was observed. Figure 50 shows boxplot diagrams for the relative abundances of compositions carrying one, two or three fucosyl residues obtained from 15 healthy volunteers with overlaid values from five samples from diseases patients.





**Figure 50.** Boxplot diagram for the relative abundance of all compositions of B2GP1 with one, two and three fucosyl residues as well as total fucosylation from samples of 15 healthy volunteers. Values obtained from samples of patients diagnosed with liver cirrhosis and hepatocellular carcinoma are overlaid. An increase in fucosylation is observed, especially for compositions carrying one fucosyl residue over four glycosylation sites.

As shown in Figure 50, an increase in fucosylation is observed in patients diagnosed with liver cirrhosis and hepatocellular carcinoma. The increase in fucosylation is especially high for all mono-fucosylated compositions but also observed for compositions with two fucosyl residues and for the degree of total fucosylation. A complete list for the relative abundance of all glycan compositions can be found in the appendix, section 10.6. *B-2-Glycoprotein 1*, Table A 18. Again, no trend is observed for the differentiation between liver cirrhosis and hepatocellular carcinoma.

### 5.7.3. Conclusion and Future Perspective

Even though only a very small number of samples from patients with liver cirrhosis and hepatocellular carcinoma were analyzed, a general trend was observed for all analyzed proteins. It was shown that changes in glycosylation are observed not only for one specific protein but for all proteins. Except for IgG1, a general increase in fucosylation was observed on alpha-acid-glycoprotein,  $\beta$ -2-glycoprotein 1 and transferrin for patients diagnosed with liver cirrhosis and hepatocellular carcinoma. This is in good agreement with previously published results from Nagel *et al.*, who detected, among other things, an increase in fucosylation on fibrinogen using the same plasma samples. This indicates an increased expression of fucosyltransferases which not only affects a specific protein but can be observed on a variety of proteins.

That changes in glycosylation can affect more than one protein but a variety of proteins was also observed for plasma sample P. The unusual glycosylation pattern of plasma sample P, which showed an increase in antennarity and therefore sialylation, was observed on analyzed plasma proteins as well.

However, some proteins showed a more tremendous change in glycosylation than others. Transferrin showed only minor changes of increased fucosylation in patients with liver diseases or antennarity for sample P compared to alpha-acid-glycoprotein, where drastic changes were observed. One reason might be the number of glycosylation sites which is two for transferrin and five for alpha-acid-glycoprotein. But it also seems that specific proteins are more susceptible for changes in glycosylation than others. Steric effects and the accessibility of glycosyltransferases to the N-glycan might affect the heterogeneity and modulation of glycosylation. Additionally, physiological reasons like turnover rates might influence the relative abundances of different proteoforms in the circulatory system as well. Glycoproteins that are more susceptible for changes in glycosylation would result in a more sensitive biomarker for e.g. early cancer detection. However, a combination of different glycoproteins would probably enhance not only sensitivity but also specificity. In addition to finding a correlation between a certain physiological condition and altered glycosylation of a specific protein, the causal relationship should also be analyzed.

## 6. Experimental Procedures

### 6.1. Protein Purification from Human Plasma

Plasma proteins, namely alpha-acid-glycoprotein,  $\beta$ -2-glycoprotein 1, ceruloplasmin, immunoglobulin IgG, and transferrin were purified from several plasma samples using different chromatographic methods. Plasma samples from healthy volunteers were purchased from the *Zentralinstitut für Transfusionsmedizin GmbH*. Table 16 shows an overview of the analyzed samples.

**Table 16.** Sample information from out-dated blood bank plasma of healthy individuals. Samples were purchased from the *Zentralinstitut für Transfusionsmedizin GmbH*.

Sample	Age	Sex	Blood type
A	62	male	B
B	60	male	O
C	34	male	O
D	53	male	O
E	26	male	B
F	19	male	unknown
G	58	male	A
H	60	male	A
I	29	female	A
J	21	female	O
K	48	female	A
L	51	female	AB
M	45	male	B
N	21	male	O
O	54	male	A
P	28	female	A
Q	24	male	B
R	55	male	B
S	46	male	A
T	47	male	B
U	27	female	unknown
V	51	male	unknown
W	23	male	unknown
X	32	male	unknown
Y	39	male	unknown
Z	38	male	unknown
AA	27	male	unknown
AB	46	male	unknown
AC	66	male	unknown
AD	46	male	unknown

Plasma samples from patients diagnosed with hepatocellular carcinoma or liver cirrhosis were received from the university clinic *Hamburg Eppendorf UKE* in cooperation with Prof. Dr. Ansgar Lohse, Dr. Henning Wege and Dr. Inez Gil-Ibanez. Table 17 shows an overview of the analyzed samples.

**Table 17.** Sample information from patients diagnosed with liver cirrhosis and hepatocellular carcinoma. Sample name was adapted from reference [216]. Samples were received from the *university clinic Hamburg Eppendorf UKE* in cooperation with Prof. Dr. Ansgar Lohse, Dr. Henning Wege and Dr. Inez Gil-Ibanez.

Sample	Sex	Age	Condition	HCC stage	Genesis	Prior therapy	Additional disease
HCC08	male	77	HCC/LC	Metastatic	Alcohol	alcohol	Diabetes mellitus
HCC13	male	73	HCC/LC	Outside Milan	HCV	TACE, ablation	SCC of the skin
HCC19	male	57	HCC/LC	Inside Milan	HCV	TACE	Multiple Myeloma
V48	female	53	LC	-	PSC	-	Ulcerative colitis
LC11	female	62	LC	-	NASH	-	CHD, Diabetes mellitus
LC10	female	44	LC	-	Alcohol	-	Atrial fibrillation

CHD = coronary heart disease; HCC = hepatocellular carcinoma, HCV = hepatitis B virus; LC = liver cirrhosis; NASH = non-alcoholic steatohepatitis; SCC = squamous-cell carcinoma; TACE = trans arterial chemoembolization; PSC = primary sclerosing cholangitis.

### 6.1.1. Anion Exchange Chromatography

As a first purification step, 500  $\mu$ L of plasma was applied to 1 mL of DEAE A-50 anion exchange *sepharose* (GE 17018001). The *sepharose* was first soaked with water overnight and equilibrated with 100 mL of 50 mM sodium acetate buffer (pH 5.5). The *sepharose* was used for flash chromatography with a column height of 5 cm and a column diameter of 0.7 cm. After application of 500  $\mu$ L of plasma to the equilibrated *sepharose*, unbound proteins were eluted with the same buffer and collected in a volume of about 8 mL. Unbound proteins were  $\beta$ -2-glycoprotein 1, IgG and transferrin. Bound proteins were eluted by applying 50 mM sodium acetate buffer (pH 5.5) with 0.8 M sodium chloride. First, fractions containing ceruloplasmin were collected in a volume of approximately 3 mL. The elution of ceruloplasmin was observed by a slightly green color. The color resulted from Cu(II) ions that are naturally attached to ceruloplasmin. Afterwards fractions containing alpha-acid-glycoprotein were collected in a volume of 2 mL.

Proteins of the pooled fractions containing  $\beta$ -2-glycoprotein 1, IgG and transferrin (8 mL) were precipitated with 70% (v/v) ethanol overnight at  $-20\text{ }^{\circ}\text{C}$ . After centrifugation, the protein pellet was washed with 2 mL of water and resolved in 1 mL of 50 mM Tris buffer (pH 8.0) containing 20 mM sodium chloride. Fractions containing ceruloplasmin or alpha-acid-glycoprotein were reduced to a volume of 1 mL.

### **6.1.2. Size Exclusion Chromatography**

All three fractions were applied to a size exclusion column (Superdex 200 Increase, GE 28990944) operated by a Hitachi HPLC system (LaChrom Elite). 500  $\mu\text{L}$  of sample solution with a protein concentration of approximately 2 mg/mL was injected. The column was operated at room temperature with a flow rate of 300  $\mu\text{L}/\text{min}$ . The elution buffer was composed of 50 mM Tris buffer (pH 8.0) containing 20 mM sodium chloride. Elution of proteins were monitored by UV detection at 280 nm and fraction of 150  $\mu\text{L}$  were collected in 96 well plates (Fraction Collector SF-3120, Advantec). Fractions containing alpha-acid-glycoprotein,  $\beta$ -2-glycoprotein 1, ceruloplasmin, immunoglobulin IgG1 and transferrin were pooled separately. Fractions of  $\beta$ -2-glycoprotein 1 and transferrin were reduced in volume and analyzed as intact glycoprotein by LC-ESI-MS. Fractions containing ceruloplasmin were once more purified by size exclusion chromatography as described above and analyzed after enzymatic digest by LC-ESI-MS and NMR. Fractions containing alpha-acid-glycoprotein were desialylated and analyzed by LC-ESI-MS.

### **6.1.3. Purification of IgG using Protein A Affinity Columns**

Fractions, obtained after size exclusion chromatography, containing immunoglobulins were further purified using Protein A HP SpinTrap columns (GE, 28903132) according to the manufacturer.

The storage solution of the SpinTrap column was removed by centrifugation and the medium was equilibrated with 600  $\mu$ L of binding buffer (20 mM  $\text{Na}_3\text{PO}_4$ , pH 7.0). The binding buffer was removed by centrifugation for 30 s at 100 x g. 600  $\mu$ L of sample solution containing 25  $\mu$ g of protein were applied and incubated for 4 min. Unbound proteins were eluted by centrifugation for 30 s at 100 x g. 600  $\mu$ L of binding buffer were added and the wash step repeated. Elution of IgG was done by applying 400  $\mu$ L of elution buffer (0.1 M glycine HCl, pH 2.7) and centrifugation for 30 s at 100 x g. The samples were collected in tubes containing 30  $\mu$ L of neutralization buffer (1 M Tris-HCl, pH 9.0). The elution step was once repeated as described above with the addition of another 30  $\mu$ L of neutralization buffer.

### **6.1.4. Determination of Protein Concentration**

Protein concentration was determined using a spectral photometer (Nanodrop 2000, ThermoFisher) based on the absorption at 280 nm. For protein mixtures, the extinction coefficient of human serum albumin was used to calculate protein concentration according to the Lambert-Beer law. The extinction coefficient of human IgG1 was used for purified IgG1.

## **6.2. Polyacrylamide Gel Electrophoresis**

Purification of proteins was monitored by SDS-PAGE. 5  $\mu$ g of lyophilized protein was dissolved in 10  $\mu$ L probe buffer (125 mM Tris-HCl, 4% (w/v) SDS, 25% (w/v) glycerol, 0.025% (w/v) 2,4,4,6-tetrabromo-2,5-cyclohexadienone and 100 mM DTT) and heat denatured at 95  $^{\circ}$ C for 5 min. The sample solution was loaded to a

polyacrylamide precast gel (8-16%, Mini-Protean TGX Precast Protein Gels, Bio-Rad) in an electrophoresis chamber (Mini-Protean Tetra Cell System, Bio-Rad). The running gel buffer was composed of 0.25 M Tris HCl, 2 M glycine and 0.1% SDS. Molecular masses of proteins were determined by loading a protein standard to the gel (Wide Range, Amresco K494). The gel was run for 60 min at a voltage of 140 V. The gel was stained with an aqueous Coomassie solution (40% (v/v) ethanol, 10% (v/v) acetic acid and 0.1% Coomassie) for 2 hours.

### **6.3. MALDI MS of Ceruloplasmin**

MS experiments using MALDI as ion source were performed on a MALDI TOF-TOF (UltrafleXtreme, Bruker) in linear positive ion mode. Protein samples in aquatic buffer solution with a concentration of about 1 mg/mL were used for MALDI-TOF analysis. Therefore, 1  $\mu$ L of protein solution was mixed with 1  $\mu$ L of 2% TFA and 1  $\mu$ L of DHAP (2,4-dihydroxyacetophenon) solution (0.12 M 2,4-dihydroxyacetophenon and 0.80 M di-ammonium hydrogencitrate in ethanol). 1  $\mu$ L of the homogenously mixed samples was applied to a ground steel target (Bruker) and dried at room temperature as crystallization took place. MALDI experiments were performed in reflector mode in a high mass range of  $m/z = 20000$  to  $m/z = 200000$ . A prior calibration was performed using the peptide calibration standard II (Bruker).

### **6.4. LC-MS Analysis of Intact Proteins**

Intact proteins (desialylated AGP, fully deglycosylated AGP, transferrin and B2GP1) were analyzed without prior digest with LC-MS on a high resolution ESI-qTOF (maXis 4G, Bruker) in positive ion mode. For each sample, a few micrograms of protein in 100  $\mu$ L distilled water were subjected to a RP C8 column (2.1 mm x 250 mm, Aeris Widepore) operated by a UHPLC system (Phenomenex, UPLC



Ultimate 3000, Dionex). The column was operated at 60 °C and a flow rate of 250  $\mu$ L/min with a gradient flow. Mobile phases were composed of solvent A: 0.1% formic acid in water and B: 0.1% formic acid in acetonitrile. The gradient was set as follows: 0 min, 5% B; 5 min, 5% B; 45 min, 60% B; 47 min, 95% B; 54 min, 95% B; 55 min, 5% B; 60 min, 5% B. The MS instrument was calibrated using a phosphazene mix (Agilent, G1969-85000) in enhanced quadratic mode. The dry gas flow of the ion source was set to 8 L/min with a drying temperature of 200 °C and a nebulizer pressure of 3.9 bar. Different sample amounts, ISCID voltages as well as mass ranges were applied for different proteins as listed in Table 18.

**Table 18.** MS acquisition parameters for ESI-qTOF-MS measurement of desialylated AGP, B2GP1 and transferrin in positive ion mode.

Protein	sample amount	ISCID	mass range [m/z]
desialylated AGP	7 $\mu$ g	20 eV	300 to 3500
fully deglycosylated AGP	7 $\mu$ g	20 eV	300 to 3500
B2GP1	10 $\mu$ g	60 eV	600 to 4500
Transferrin	10 $\mu$ g	100 eV	600 to 4500

### 6.5. ESI-MS of Intact AGP by Direct Infusion

Fully sialylated AGP from pooled human plasma (lyophilizate, Sigma G9885) was analyzed in negative ion mode by ESI-MS. Sample preparation for MS analysis involved dissolving the protein in a 1:1 mixture of deionized water and HPLC grade acetonitrile to give a concentration of 0.5  $\mu$ g/ $\mu$ L. This solution was injected via a syringe pump (KD Scientific) at 40  $\mu$ L/min directly into the MS ion source. MS experiments were performed on an ESI/qTOF (maXis 4G, Bruker) in negative ion mode for native AGP. The ion source was set to a dry gas flow of 4 L/min, dry temperature of 180 °C and a nebulizer pressure of 0.7 bar. An additional in-source CID voltage was applied with -190 eV. The spectra time was 1.0 s with a spectral width of m/z 1000 - 5000. [142]

## 6.6. Enzymatic Digest of Glycoproteins

### 6.6.1. Tryptic Digest

For generating tryptic peptides and glycopeptides, 15 µg of purified and lyophilized glycoprotein were dissolved in 10 µL of 6 M urea with 15 mM DTT. Protein denaturation was performed by incubating the sample at 60 °C for 20 min. After cooling to room temperature, the sample was diluted with 90 µL of 50 mM Tris buffer (pH 8.0) containing 20 mM sodium chloride. After addition of 0.25 µg bovine Trypsin (Sigma 1700002, diluted in 1 mM hydrochloric acid to give a concentration of 0.5 mg/mL), the sample was incubated at 37 °C for 24 hours. Inactivation of trypsin was done by incubation at 95 °C for 5 min.[83]

### 6.6.2. Digest with Endoproteinase Arg-C

Glycopeptides of ceruloplasmin were analyzed as tryptic peptides in a standard bottom-up approach and as glycopeptides with enhanced peptide length in a middle-down approach. Glycopeptides with enhanced peptide length were either generated by digestion with Arg-C or by limited tryptic digestion as described in the following section. For the digestion with Arg-C, 15 µg of purified ceruloplasmin were dissolved in 88 µL of 50 mM Tris-HCl buffer (pH 7.9) containing 5 mM CaCl<sub>2</sub> and 2 mM EDTA. Arg-C protease (Promega V1881) was dissolved in the same buffer to give a concentration of 0.025 µg/µL. 2 µL of enzyme solution were added to the sample to give an enzyme to protein ratio of 1:300. 10 µL of an activation buffer containing 50 mM Tris-HCl, 50 mM DTT and 2 mM EDTA (pH 7.9) were added and the sample was incubated at 37 °C for 24 hours. Inactivation of Arg-C was done by incubation at 95 °C for 5 min.[83]

### 6.6.3. Limited Tryptic Digest

To generate glycopeptides with a larger peptide residue a limited tryptic digest was performed. Therefore, 15 µg of purified and lyophilized ceruloplasmin from different plasma samples were dissolved in 97 µL 100 mM glycine buffer (pH 11.4) without prior denaturation. After addition of 0.5 µg bovine Trypsin (Sigma 1700002, diluted in 1 mM hydrochloric acid to give a concentration of 0.5 mg/mL), the sample was incubated at 24 °C for 24 min. Inactivation of trypsin was done by freezing the sample in liquid nitrogen for 10 min and subsequent incubation at 95 °C with the addition of 2 µL of a 100 mM DTT solution for 5 min.[83]

### 6.6.4. Neuraminidase Digest of Human AGP

Purchased AGP and AGP purified from several plasma samples were analyzed as fully sialylated and after desialylation by LC-MS. 30 µg of purchased AGP (lyophilizate, Sigma-Aldrich, G9885) or 30 µg of purified AGP were dissolved in 15 µL of 0.1 M sodium acetate buffer (2 mM CaCl<sub>2</sub>, 1% TritonX, pH 5.0). Enzymatic digestion was initialized by addition of 0.1 u of sialidase from *clostridium perfringens* (Sigma-Aldrich, N2876) and performed at 37 °C overnight.[142]

### 6.6.5. PNGase Digest for the Analysis of Free Glycans

Free glycans of alpha acid glycoprotein and ceruloplasmin were analyzed by ESI-MS. A complete deglycosylation of glycoproteins was achieved under reducing conditions with endoglycosidase PNGase F according to manufacturer (Promega, V4831). For the release of N-glycans from lyophilized alpha-acid-glycoprotein, 30 µg AGP were dissolved in 12 µL water. After addition of 1 µL 5% SDS and 1 µL 1 M 1,4-Dithiothreitol in aqueous solution, the sample was heated to 95 °C for 5 minutes. The samples were cooled to room temperature and 2 µL of 0.5 M sodium

phosphate buffer (pH 7.5), 2  $\mu\text{L}$  of 10% TritonX and 1  $\mu\text{L}$  (10 u) of PNGase F were added. Digestion was performed by incubation at 37 °C overnight.[142]

Free glycans of ceruloplasmin were also analyzed using NMR. A tryptic digest was performed prior to PNGase digestion. Therefore, 2.65 mg of purified and lyophilized ceruloplasmin (20.4 nmol) from one plasma sample were dissolved in 50  $\mu\text{L}$  of 6 M Urea and 5  $\mu\text{L}$  of 100 mM DTT solution were added. The protein solution was denatured at 60 °C for 20 min. The solution was diluted to a volume of 1 mL with 5 mM potassium phosphate buffer (pH 7.5). For the tryptic digest, 5  $\mu\text{L}$  of trypsin solution (Sigma 1700002, diluted in 1 mM hydrochloric acid to give a concentration of 0.5 mg/mL) were added and the sample was incubated at 37 °C overnight. Inactivation of trypsin was done by heat denaturation at 90 °C for 15 min. The sample was cooled to room temperature and 15  $\mu\text{L}$  (150 u) of PNGase F were added. Deglycosylation of glycopeptides was performed at 37 °C overnight. Enrichment of glycans and separation of peptides was done using C18 solid phase extraction (Chromabond C18 polypropylene columns, 730014, Macherey-Nagel). The sample solution was applied to the C18 SPE column, equilibrated with  $\text{H}_2\text{O}$  + 0.01% FA. Free glycans were eluted with  $\text{H}_2\text{O}$  + 0.01% FA and reduced in volume. Peptides retained on the C18 material and were eluted with ACN + 0.01% FA.

#### **6.6.6. Tryptic In-Gel Digest**

Different protein variants of ceruloplasmin were analyzed after a tryptic in-gel digest and subsequent peptide identification by LC-ESI-MS. For tryptic in-gel digest, an SDS-gel was run as described in section 6.2. *Polyacrylamide Gel Electrophoresis*. In total, 5 lanes were loaded with 5  $\mu\text{g}$  of protein each. Protein bands were cut out and de-stained with 500  $\mu\text{L}$  of 200 mM ammonium hydrogen carbonate solution in 40% (v/v) acetonitrile at 37 °C for 20 min. The supernatant was removed, and the de-staining was repeated three times. The gel was dried under vacuum. 150  $\mu\text{L}$  of digest buffer (10  $\mu\text{g}/\text{mL}$  trypsin (Sigma 1700002) in 0.1 mM HCl, 36 mM ammonium hydrogen carbonate, 8% (v/v) acetonitrile) was added to the gel and proteolytic digest was performed at 37 °C overnight. The gel was removed from the peptide solution and extracted with 150  $\mu\text{L}$  of 0.1% (v/v) TFA in

50% (v/v) acetonitrile. This step was repeated three times. Peptide solutions were pooled, reduced in volume and analyzed by LC-ESI-MS.

### **6.7. LC-ESI-MS/MS of Glycopeptides**

Peptides and glycopeptides obtained by treatment with trypsin or endoproteinase Arg-C were analyzed by LC-ESI-MS for the analysis and relative quantification of glycopeptides as well as automated peptide identification using the Mascot search algorithm. LC-MS experiments were performed on an UHPLC (Ultimate 3000, Dionex) equipped with a RP C18 column (2.1 mm x 150 mm, Aeris Peptide 1.7u XB-C18, Phenomenex). Online MS experiments were performed on an ESI/qTOF (maXis 4G, Bruker). The samples containing 15 µg of proteolytic digest in a volume of 100 µL were subjected to the RP-C18 column with a prior calibration using a phosphazene mix (Tune Mix, Agilent). The column was operated at 55 °C at a flow rate of 250 µL/min with a mobile phase of solvent A: 0.1% formic acid in water and B: 0.1% formic acid in acetonitrile. The gradient was set as follows: 0 min, 2% B; 5 min, 2% B; 72 min, 50% B; 74 min, 90% B; 82 min, 90% B; 85 min, 2% B; 90 min, 2% B. MS spectra were acquired in positive ion mode and ESI source conditions with a dry gas flow of 9 L/min, drying temperature of 200 °C and a nebulizer pressure of 4.6 bar. Bigger glycopeptides obtained by limited tryptic cleavage were measured with an additional in-source CID voltage of 20 eV. The acquisition time for a MS spectrum was set to 1.0 s with a spectral width of m/z 225 – 3000.

Peptides and smaller glycopeptides obtained by tryptic digest or digest with endoproteinase Arg-C were analyzed in MS/MS mode with CID fragmentation. MS ions with a charge state of +1, +2 and +3 were selected for fragmentation. The cycle time of a MS precursor was set to 3.0 seconds.[83] Table 19 shows the collision energies for CID fragmentation depending on charge state and mass range.

**Table 19.** Collision energies for CID fragmentation of peptides and glycopeptides with a charge of +1, +2 and +3

Charge state	Isolation mass [m/z]	Isolation Width [m/z]	Collision energy [eV]
1	300	3.00	34
1	500	4.80	39
1	1000	6.00	52
1	2000	9.00	55
2	300	3.00	26
2	500	4.80	34
2	1000	6.00	40
2	2000	9.00	45
3	300	3.00	21
3	500	4.80	28
3	1000	6.00	36
3	2000	9.00	40

The three most abundant ions in MS mode were selected for CID fragmentation. Active exclusion of MS precursor was activated after acquisition of three fragment spectra. Active exclusion was released after 60 s.

Glycopeptides of ceruloplasmin with enhanced peptide length were not fragmented with CID fragmentation but with ISCID fragmentation. Next to an ISCID voltage of 20 eV for MS ionization, an ISCID of 160 eV was applied for pseudo MS/MS fragmentation. ISCID fragmentation is an untargeted approach without MS precursor selection.

### 6.8. LC-ESI-MS of PNGase Released Glycans

Free glycans of ceruloplasmin released by PNGase digestion were analyzed by LC-ESI-MS, using a PGC column (Hypercarb, 150 x 2.1 mm, 3  $\mu$ m particle size, Thermo Scientific). Experiments were performed on a UHPLC (Ultimate 3000, Dionex) coupled to online detection on an ESI/qTOF (maXis 4G, Bruker) in positive ion mode. 10 pmol of released glycans (estimated from protein concentration) were

subjected to LC-MS analysis. The column was operated at 40 °C at a flow rate of 150  $\mu$ L/min with a mobile phase of solvent A: 0.1% formic acid in water and B: 0.1% formic acid in acetonitrile. The gradient was set as follows: 0 min, 5% B; 5 min, 10% B; 60 min, 35% B; 62 min, 90% B; 66 min, 90% B; 68 min, 5% B; 75 min, 5% B. The MS instrument was calibrated using a phosphazene mix (Tune mix, Agilent) with enhanced quadratic calibration. ESI source conditions were set to a dry gas flow of 7 L/min, drying temperature of 180 °C and a nebulizer pressure of 2.0 bar. The acquisition time for a MS spectrum was set to 1.0 s with a spectral width of  $m/z$  100 – 2800. For MS/MS, the cycle time of a precursor was set to 3.0 seconds.

### **6.9. Online LC-ESI-MS and Offline $^1\text{H-NMR}$ Analysis of Free Glycans**

N-glycans released of human ceruloplasmin were analyzed by a combination of LC-MS and NMR techniques. Therefore, 2 nmol of PNGase released N-glycans (estimated from protein concentration) were subjected to UHPLC (Ultimate 3000, Dionex) separation using a PGC column (Hypercarb, 150 x 4.6 mm, 3  $\mu$ m particle size, Thermo Scientific) with online MS detection on an ESI/qTOF (maXis 4G, Bruker) in positive ion mode. The column was operated at 40 °C at a flow rate of 800  $\mu$ L/min with a mobile phase of solvent A: 65 mM ammonium formate, 10 mM  $\text{NH}_3$  in water and B: 10 mM  $\text{NH}_3$  in acetonitrile. The gradient was set as follows: 0 min, 5% B; 5 min, 10% B; 60 min, 35% B; 62 min, 90% B; 66 min, 90% B; 68 min, 5% B; 75 min, 5% B. After LC separation, the flow was splitted (QuickSplit Adjustable Flow Splitter, Analytical Scientific Instruments). 5% of the flow was measured by online ESI-MS and 95% were collected in fractions (Fraction Collector SF-3120, Advantec) with a fraction size of 20 s (250  $\mu$ L) for offline NMR analysis. The MS instrument was calibrated using a phosphazene mix (Tune mix, Agilent) with enhanced quadratic calibration. ESI source conditions were set to a dry gas flow of 5 L/min, drying temperature of 180 °C and a nebulizer pressure of 0.8 bar. The acquisition time for a MS spectrum was set to 1.0 s with a spectral width of  $m/z$  100 - 2800. For MS/MS, the cycle time of a precursor was set to 3.0 seconds.

For NMR analysis, samples were dried in vacuum, resolved in 180  $\mu$ L D<sub>2</sub>O with 1 mM TMS (tetramethylsilane) and transferred to 3 mm NMR tubes. <sup>1</sup>H-NMR experiments were performed on a 700 MHz NMR (Avance700, Bruker), equipped with a CryoProbe head. An excitation sculpting pulse sequence was used to suppress the water signal. The duration of the selective square pulse was set to 2 ms. All spectra were recorded with 6k scans, an acquisition time of 2.3 s and a relaxation delay of 1 s.

## **6.10. Data Interpretation**

### **6.10.1. Pre-Processing of MS Data using DataAnalysis 4.2**

All MS data were processed using DataAnalysis 4.2 (Bruker). The spectra were calibrated either internal or external using a phosphazene mix (Tune mix, Agilent) with enhanced quadratic calibration. Deconvolution of glycoprotein and glycopeptides was done using the built-in maximum entropy deconvolution (Bruker). Mass ranges for deconvolution were adapted to the protein mass. A shifted spectrum with a theoretical mass of +1 was generated. Table 20 shows the different parameters for maximum entropy deconvolution of analyzed glycoproteins and glycopeptides.



**Table 20.** Parameters for maximum entropy deconvolution (Bruker) for analyzed glycoproteins and glycopeptides.

Protein	Protein Mass	Mass range for deconvolution	Resolving Power
Fully sialylated AGP	37 kDa	30 to 50 kDa	10.000
Desialylated AGP	34 kDa	5 to 50 kDa	10.000
Fully deglycosylated AGP	21 kDa	15 to 30 kDa	3000
$\beta$ -2-Glycoprotein 1	46 kDa	40 to 60 kDa	10.000
Glycopeptides of ceruloplasmin	from 4 kDa to 22 kDa	0.5 to 50 kDa	10.000
Transferrin	79 kDa	70 to 90 kDa	10.000

Smaller analytes like free glycans, peptides and glycopeptides from tryptic digest of IgG1 were deconvoluted using the Peptides/Small Molecules deconvolution with an abundance cutoff of 0.05%. Deconvoluted spectra were exported as .xy file and further analyzed using Matlab (version R2017a, MathWorks). Extracted ion chromatograms were generated using DataAnalysis 4.2 with a mass width of  $\pm$ 0.05 m/z and exported as .xy file.

### 6.10.2. Automated Peptide Identification using MASCOT MS/MS Ion Search

Tryptic peptides were analyzed using LC-MS and subjected to an automatic peptide identification to identify enriched proteins and other impurities.[82] For the tryptic peptides, a Mascot search was performed using the free web server to monitor impurities. Therefore, 850 compounds were generated using the automated MS/MS search implemented in the software (DataAnalysis 4.2). The compound spectra were deconvoluted using the Peptides/Small Molecules deconvolution with an abundance cutoff of 0.05%. The deconvoluted compounds were exported as Mascot generic file (.mgf) and uploaded to the Mascot server. The SwissProt data bank was searched using the Mascot MS/MS ion search with

the following parameters: enzyme = trypsin; taxonomy = *mammalia*; maximum missed cleavages = 1; no fixed modifications; variable modifications = pyro-glutamine; variable modifications = carbamylation (N-terminal and on lysine); peptide tolerance = 0.02 Da (MS) and 0.02 (MS/MS); peptide charge = 2+, 3+ and 4+.[83]

### 6.10.3. Calculation of Simulated Spectra

Deconvoluted MS spectra were interpreted by simulation of a theoretical spectrum. This was done to support not only structural assignment but also for relative quantification of different proteoforms. The deconvoluted spectra were imported as .xy file into Matlab (version R2017a, MathWorks) with the command *dlmread*. Chemical formula for all potential proteoforms were calculated based on the amino acid composition including disulfide bridges as well as post-translational modifications like glycosylation or cysteinylation. Thus, a matrix containing all chemical formula was obtained in the form of [C H N O S] for elemental compositions. The isotopic mass distribution for all proteoforms was calculated using the Matlab command *isotopicdist*. A theoretical spectrum was calculated to fit the experimental spectrum. This was done by a least square minimization using the Matlab command *lsqlin*. The experimental spectrum was approximated by a linear combination of the mass distributions for all proteoforms by equation 1.

$$\min_x \frac{1}{2} \|C * x - d\|_2^2 \text{ such that } \begin{cases} lb \leq x, \\ Aeq * x = beq. \end{cases} \quad \text{Equation 1}$$

With:

$x$  = weighing coefficients of different proteoforms

$C$  = matrix for the isotopic distribution of all proteoforms

$d$  = vector of the experimental spectrum

$lb$  = lower bounds

$Aeq$  = liner equality constraint matrix (only used for fully sialylated AGP described in the next chapter)

$beq$  = liner equality constraint vector (only used for fully sialylated AGP described in the next chapter)

Lower bounds were always set to 0 to avoid negative weighting of proteoforms. The solution of the linear combination leads to weighing coefficients for each proteoform. These coefficients were then used for relative quantification of different proteoforms by normalization to 100%. The Matlab based script can be found in the appendix (*10.3. Matlab Script for the Relative Quantification of Different Proteoforms*).

#### **6.10.3.1. Calculation of a Simulated Spectrum for the Analysis of Fully Sialylated AGP**

The calculation of a theoretical spectrum for fully sialylated AGP was based on the quantification results of desialylated AGP to allow a conclusive determination of sialylation and fucosylation. To include the proportions of fucosylation from the analysis of the desialylated protein, a linear equality constraint matrix  $A_{eq}$  with a linear equality constraint vector  $b_{eq}$  was used for the minimization described in equation 1. For variant ORM1\*F1, 29 desialylated glycan composition were included, for ORM1\*S 33 desialylated compositions and for ORM2 28 desialylated glycan compositions. For each desialylated composition 12 to 19 neuraminic acids were considered, leading to a total number of 728 possible sialylated proteoforms over all three variants. However, structures with more than one sialic acid per antennae were excluded from the minimization by setting their upper bonds to 0. After minimization, different coefficients were obtained for the relative abundance of different proteoforms.[142]

#### **6.10.4. Integration of Extracted Ion Chromatograms**

Extracted ion chromatogram were integrated to reflect the relative abundance of tryptic glycopeptides from IgG1. The EICs were imported into Matlab (version R2017a, MathWorks) using the command *dlimread* and integrated in a time interval











of one minute according to their retention time. A trapezoidal method with unit spacing was used for integration by the Matlab command *trapz*. The default value of 1 for point spacing was used. Values for all integrated EICs were normalized to 100% and used as representation of relative abundances.

#### **6.10.5. Interpretation of NMR Data and Correlation with MS Data**











NMR spectra were processed using Topspin 3.2. (Bruker BioSpin). Spectra were calibrated on internal tetramethylsilane. Calibrated spectra were imported to Matlab (version R2017a, MathWorks) using the function *rbnmr* (by Nils Nyberg). EICs from LC-MS analysis were imported using the Matlab command *dlmread*. Data points of EICs were reduced to fit the data point spacing of extracted delta chromatogram (EDCs). Reduced data points were generated by forming the mean value in a certain time interval. The number of data points is thereby predefined by the number of acquired NMR spectra. Pearson correlation coefficients between EICs and EDCs were calculated using the Matlab command *corrcoeff*. The correlation coefficients reflect the linear dependence of an EDCs with an EICs over their shared time domain. To generate pure NMR of single compounds, the correlation coefficients for each chemical shift were multiplied by the sum over all NMR spectra.[98]

## 7. Hazards

Table 21. Hazard and precautionary statements according to the Globally Harmonized System of Classification and Labelling Chemicals (GHS).

Compound	Pictogram	Hazard Statement	Precaution Statement
Acetic acid		H226, H290, H314	P210, P280, P301+P331, P305+P351+P338, P308+P310
Acetonitrile $\geq$ 99.9%		H225, H332, H302, H312, H319	P210, P240, P302+P352, P305+P351+P338, P403+P233
Ammonia (aq., 25%)		H314, H335, H400	P280, P273, P301+P330+P331, P305+P351+P338, P309, P310
Ammonium formate		H315, H319, H335	P261, 305+P351+P338
Calcium chloride		H315, H319, H335	P261, P305+P351+P338
di-Ammonium hydrogen citrate		H319	P305+351+338
2,5-Dihydroxybenzoic acid		H302, H315, H319, H335	P261, P305+P351+P338
1,4-D/L-Dithiothreitol		H302, H315, H319	P302+P352, P305+P351+P338
Endoproteinase Arg-C		H315, H319, H334, H335	P264, P280, P284, P302+P352, P305+P351+P338, P312
Ethanol		H225, H319	P210, P240, P305+P351+P338, P403+P233

HAZARDS

Ethylenediaminetetraacetic acid		H319	P305+P351+P338
ES-TOF <i>Tuning Mix</i>		H225, H302+H332, H319	P210, P305+P351+P338
Formic acid (99-100%)		H226, H302, H314, H331	P210, P280, P303+P361+P353, P304+P340+P310, P305+P351+P338, P403+P233
Hydrogen chloride (37% solution)		H290, H314, H335	P260, P280, P303+P361+P353, P304+P340+P310, P305+P351+P338
Sodium dodecyl sulfate		H228, H302+H332, H315, H318, H335, H412	P210, P261, P280, P301+P312+P330, P305+P351+P338+P310, P370+P378
2,4,4,6-Tetrabromo-2,5-cyclohexadienone		H315, H319, H335	P261, P305+P351+P338
Trifluoroacetic acid		H290, H331, H314, H412	P260, P273, P280, P303+P361+P353, P305+P351+P338, P312
Tris(hydroxymethyl)amino methane		H315, H319, H335,	P261, P305+P351+P338
TritonX, 4-(1,1,3,3-Tetramethylbutyl)phenylpolyethylene glycol		H302+315, H318+410	P280+301, P312, P330+305, P351, P228, P310
Trypsin		H319, H335, H315, H334	

## 8. Publications

Results from this work have been published or are under revision for publication.

- I. M. Baerenfaenger, B. Meyer, Intact Human Alpha-Acid Glycoprotein Analyzed by ESI-qTOF-MS: Simultaneous Determination of the Glycan Composition of Multiple Glycosylation Sites, *Journal of Proteome Research*, 17 (2018) 3693-3703
- II. M. Baerenfaenger, M. Moritz, B. Meyer, Quantitation of Glycopeptides by ESI/MS - size of the peptide part strongly affects the relative proportions and allows discovery of new glycan compositions of Ceruloplasmin, *Glycoconjugate Journal*, 36 (2019) 13-26.
- III. M. Baerenfaenger, B. Meyer, Simultaneous characterization of SNPs and N-glycans from multiple glycosylation sites of intact  $\beta$ -2-glycoprotein-1 (B2GP1) by ESI-qTOF-MS, *BBA Proteins and Proteomics*, (2019) under revision.

## 9. References

- [1] R. Apweiler, H. Hermjakob, N. Sharon, On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database<sup>1</sup> Dedicated to Prof. Akira Kobata and Prof. Harry Schachter on the occasion of their 65th birthdays, *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1473 (1999) 4-8.
- [2] J. Hofsteenge, D.R. Mueller, T. de Beer, A. Loeffler, W.J. Richter, J.F.G. Vliegthart, New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us, *Biochemistry*, 33 (1994) 13524-13530.
- [3] A. Varki, R.D. Cummings, M. Aebi, N.H. Packer, P.H. Seeberger, J.D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, J.J. Prestegard, R.L. Schnaar, H.H. Freeze, J.D. Marth, C.R. Bertozzi, M.E. Etzler, M. Frank, J.F.G. Vliegthart, T. Lütkeke, S. Perez, E. Bolton, P. Rudd, J. Paulson, M. Kanehisa, P. Toukach, K.F. Aoki-Kinoshita, A. Dell, H. Narimatsu, W. York, N. Taniguchi, S. Kornfeld, Symbol Nomenclature for Graphical Representations of Glycans, *Glycobiology*, 25 (2015) 1323-1324.
- [4] N.G. Jayaprakash, A. Surolia, Role of glycosylation in nucleating protein folding and stability, *Biochemical Journal*, 474 (2017) 2333-2347.
- [5] D. Shental-Bechor, Y. Levy, Effect of glycosylation on protein folding: A close look at thermodynamic stabilization, *Proceedings of the National Academy of Sciences of the United States of America*, 105 (2008) 8256-8261.
- [6] C. Xu, D.T.W. Ng, Glycosylation-directed quality control of protein folding, *Nature Reviews Molecular Cell Biology*, 16 (2015) 742.
- [7] E. Maverakis, K. Kim, M. Shimoda, M.E. Gershwin, F. Patel, R. Wilken, S. Raychaudhuri, L.R. Ruhaak, C.B. Lebrilla, Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: A critical review, *Journal of Autoimmunity*, 57 (2015) 1-13.
- [8] A.G. Morell, G. Gregoriadis, I.H. Scheinberg, J. Hickman, G. Ashwell, The Role of Sialic Acid in Determining the Survival of Glycoproteins in the Circulation, *Journal of Biological Chemistry*, 246 (1971) 1461-1467.
- [9] X. Huang, J.-C. Leroux, B. Castagner, Well-Defined Multivalent Ligands for Hepatocytes Targeting via Asialoglycoprotein Receptor, *Bioconjugate Chemistry*, 28 (2017) 283-295.
- [10] W.M. Watkins, P. Greenwell, A.D. Yates, P.H. Johnson, Regulation of expression of carbohydrate blood group antigens, *Biochimie*, 70 (1988) 1597-1611.
- [11] F. Yamamoto, E. Cid, M. Yamamoto, N. Saitou, J. Bertranpetit, A. Blancher, An integrative evolution theory of histo-blood group ABO and related genes, *Scientific Reports*, 4 (2014) 6601.
- [12] N. Kenan, A. Larsson, O. Axelsson, A. Helander, Changes in transferrin glycosylation during pregnancy may lead to false-positive carbohydrate-deficient transferrin (CDT) results in testing for riskful alcohol consumption, *Clinica Chimica Acta*, 412 (2011) 129-133.
- [13] D. Biou, C. Bauvy, H. N'Guyen, P. Codogno, G. Durand, M. Aubery, Alterations of the glycan moiety of human  $\alpha$ -acid glycoprotein in late-term pregnancy, *Clinica Chimica Acta*, 204 (1991) 1-12.
- [14] H. Yin, Z. Lin, S. Nie, J. Wu, Z. Tan, J. Zhu, J. Dai, Z. Feng, J. Marrero, D.M. Lubman, Mass-Selected Site-Specific Core-Fucosylation of Ceruloplasmin in Alcohol-Related Hepatocellular Carcinoma, *Journal of Proteome Research*, 13 (2014) 2887-2896.
- [15] T.W. De Graaf, M.E. Van der Stelt, M.G. Anbergen, W. van Dijk, Inflammation-induced expression of sialyl Lewis X-containing glycan structures on  $\alpha$  1-acid glycoprotein (orosomucoid) in human sera., *Journal of Experimental Medicine*, 177 (1993) 657-666.
- [16] X. Shen, X. Zhang, S. Liu, Novel hemagglutinin-based influenza virus inhibitors, *Journal of Thoracic Disease*, 5 (2013) S149-S159.



- [17] M. Sharma, A.K. Jainarayanan, AchE-OGT dual inhibitors: Potential Partners in Handling Alzheimer's Disease, *bioRxiv*, (2018).
- [18] R.F. Ortiz-Meoz, J. Jiang, M.B. Lazarus, M. Orman, J. Janetzko, C. Fan, D.Y. Duveau, Z.-W. Tan, C.J. Thomas, S. Walker, A Small Molecule That Inhibits OGT Activity in Cells, *ACS Chemical Biology*, 10 (2015) 1392-1397.
- [19] W.B. Dias, G.W. Hart, O-GlcNAc modification in diabetes and Alzheimer's disease, *Molecular BioSystems*, 3 (2007) 766-772.
- [20] R.J. Darling, U. Kuchibhotla, W. Glaesner, R. Micanovic, D.R. Witcher, J.M. Beals, Glycosylation of Erythropoietin Affects Receptor Binding Kinetics: Role of Electrostatic Interactions, *Biochemistry*, 41 (2002) 14524-14531.
- [21] R.E. Smith Jr, I.A. Jaiyesimi, L.A. Meza, N.S. Tchekmedyian, D. Chan, H. Griffith, S. Brosman, R. Bukowski, M. Murdoch, M. Rarick, A. Saven, A.B. Colowick, A. Fleishman, U. Gayko, J. Glaspy, Novel erythropoiesis stimulating protein (NESP) for the treatment of anaemia of chronic disease associated with cancer, *British Journal Of Cancer*, 84 (2001) 24.
- [22] J.M. Reichert, C.J. Rosensweig, L.B. Faden, M.C. Dewitz, Monoclonal antibody successes in the clinic, *Nature Biotechnology*, 23 (2005) 1073.
- [23] M.C. Tronconi, F. Sclafani, L. Rimassa, C. Carnaghi, N. Personeni, A. Santoro, Fatal Infusion Reaction to Cetuximab: The Need for Predictive Risk Factors and Safer Patient Selection, *Journal of Clinical Oncology*, 29 (2011) e680-e681.
- [24] S. Hopps, P. Medina, S. Pant, R. Webb, M. Moorman, E. Borders, Cetuximab hypersensitivity infusion reactions: Incidence and risk factors, *Journal of Oncology Pharmacy Practice*, 19 (2013) 222-227.
- [25] C.H. Chung, B. Mirakhur, E. Chan, Q.-T. Le, J. Berlin, M. Morse, B.A. Murphy, S.M. Satinover, J. Hosen, D. Mauro, R.J. Slebos, Q. Zhou, D. Gold, T. Hatley, D.J. Hicklin, T.A.E. Platts-Mills, Cetuximab-Induced Anaphylaxis and IgE Specific for Galactose- $\alpha$ -1,3-Galactose, *The New England journal of medicine*, 358 (2008) 1109-1117.
- [26] v.B.M.M. C., B. Muriel, Minimizing immunogenicity of biopharmaceuticals by controlling critical quality attributes of proteins, *Biotechnology Journal*, 7 (2012) 1473-1484.
- [27] F. Cymer, H. Beck, A. Rohde, D. Reusch, Therapeutic monoclonal antibody N-glycosylation – Structure, function and therapeutic potential, *Biologicals*, 52 (2018) 1-11.
- [28] S. Gilgunn, P.J. Conroy, R. Saldoval, P.M. Rudd, R.J. O'Kennedy, Aberrant PSA glycosylation—a sweet predictor of prostate cancer, *Nature Reviews Urology*, 10 (2013) 99.
- [29] A. Ercan, J. Cui, D.E.W. Chatterton, K.D. Deane, M.M. Hazen, W. Brintnell, C.I. O'Donnell, L.A. Derber, M.E. Weinblatt, N.A. Shadick, D.A. Bell, E. Cairns, D.H. Solomon, V.M. Holers, P.M. Rudd, D.M. Lee, Aberrant IgG galactosylation precedes disease onset, correlates with disease activity, and is prevalent in autoantibodies in rheumatoid arthritis, *Arthritis & Rheumatism*, 62 (2010) 2239-2248.
- [30] S. Katja, B.J. E., I. Yedy, N. Onni, Sensitivity and Specificity of Carbohydrate-Deficient Transferrin as a Marker of Alcohol Abuse Are Significantly Influenced by Alterations in Serum Transferrin: Comparison of Two Methods, *Alcoholism: Clinical and Experimental Research*, 20 (1996) 449-454.
- [31] K. Golka, A. Wiese, Carbohydrate-Deficient Transferrin (CDT)-A Biomarker for Long-Term Alcohol Consumption, *Journal of Toxicology and Environmental Health, Part B*, 7 (2004) 319-337.
- [32] A. Kirwan, M. Utratna, M.E. O'Dwyer, L. Joshi, M. Kilcoyne, Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics, *BioMed Research International*, 2015 (2015) 490531.
- [33] C.A. Reis, H. Osorio, L. Silva, C. Gomes, L. David, Alterations in glycosylation as biomarkers for cancer detection, *Journal of Clinical Pathology*, 63 (2010) 322-329.

- [34] J.R. Prensner, M.A. Rubin, J.T. Wei, A.M. Chinnaiyan, Beyond PSA: The next generation of prostate cancer biomarkers, *Science translational medicine*, 4 (2012) 127rv123-127rv123.
- [35] M. Adhyam, A.K. Gupta, A Review on the Clinical Utility of PSA in Cancer Prostate, *Indian Journal of Surgical Oncology*, 3 (2012) 120-129.
- [36] Q.K. Li, L. Chen, M.-H. Ao, J.H. Chiu, Z. Zhang, H. Zhang, D.W. Chan, Serum Fucosylated Prostate-specific Antigen (PSA) Improves the Differentiation of Aggressive from Non-aggressive Prostate Cancers, *Theranostics*, 5 (2015) 267-276.
- [37] J. Lou, L. Zhang, S. Lv, C. Zhang, S. Jiang, Biomarkers for Hepatocellular Carcinoma, *Biomarkers in Cancer*, 9 (2017) 1-9.
- [38] R.K. Sterling, L. Jeffers, F. Gordon, A.P. Venook, K.R. Reddy, S. Satomura, F. Kanke, M.E. Schwartz, M. Sherman, Utility of Lens culinaris Agglutinin-Reactive Fraction of  $\alpha$ -Fetoprotein and Des-Gamma-Carboxy Prothrombin, Alone or in Combination, as Biomarkers for Hepatocellular Carcinoma, *Clinical Gastroenterology and Hepatology*, 7 (2009) 104-113.
- [39] J. Hofmann, H.S. Hahm, P.H. Seeberger, K. Pagel, Identification of carbohydrate anomers using ion mobility–mass spectrometry, *Nature*, 526 (2015) 241.
- [40] C.J. Gray, B. Schindler, L.G. Migas, M. Pičmanová, A.R. Allouche, A.P. Green, S. Mandal, M.S. Motawia, R. Sánchez-Pérez, N. Bjarnholt, B.L. Møller, A.M. Rijs, P.E. Barran, I. Compagnon, C.E. Eyers, S.L. Flitsch, Bottom-Up Elucidation of Glycosidic Bond Stereochemistry, *Analytical Chemistry*, 89 (2017) 4540-4549.
- [41] N.C. Polfer, J.J. Valle, D.T. Moore, J. Oomens, J.R. Eyler, B. Bendiak, Differentiation of Isomers by Wavelength-Tunable Infrared Multiple-Photon Dissociation-Mass Spectrometry: Application to Glucose-Containing Disaccharides, *Analytical Chemistry*, 78 (2006) 670-679.
- [42] M. Fellenberg, H.N. Behnken, T. Nagel, A. Wiegandt, M. Baerenfaenger, B. Meyer, Glycan analysis: scope and limitations of different techniques—a case for integrated use of LC-MS(/MS) and NMR techniques, *Analytical and Bioanalytical Chemistry*, 405 (2013) 7291-7305.
- [43] N.L. Anderson, N.G. Anderson, The Human Plasma Proteome: History, Character, and Diagnostic Prospects, *Molecular & Cellular Proteomics*, 1 (2002) 845-867.
- [44] A. Leigh, Candidate-based proteomics in the search for biomarkers of cardiovascular disease, *The Journal of Physiology*, 563 (2005) 23-60.
- [45] R.R. Burgess, Chapter 20 Protein Precipitation Techniques, in: R.R. Burgess, M.P. Deutscher (Eds.) *Methods in Enzymology*, Academic Press, Place Published, 2009, pp. 331-342.
- [46] O. Coskun, Separation techniques: Chromatography, *Northern Clinics of Istanbul*, 3 (2016) 156-160.
- [47] A.J. A., A.B. A., Protein purification using chromatography: selection of type, modelling and optimization of operating conditions, *Journal of Molecular Recognition*, 22 (2009) 65-76.
- [48] H.A. Sober, F.J. Gutter, M.M. Wyckoff, E.A. Peterson, Chromatography of Proteins. II. Fractionation of Serum Protein on Anion-exchange Cellulose, *Journal of the American Chemical Society*, 78 (1956) 756-763.
- [49] H. Lars, Gel-Filtration Chromatography, *Current Protocols in Molecular Biology*, 44 (1998) 10.19.11-10.19.12.
- [50] Chapter 1 The Origins and development of liquid chromatography, in: A. Fallon, R.F.G. Booth, L.D. Bell (Eds.) *Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier, Place Published, 1987, pp. 1-7.
- [51] Y. Shen, N. Tolić, P.D. Piehowski, A.K. Shukla, S. Kim, R. Zhao, Y. Qu, E. Robinson, R.D. Smith, L. Paša-Tolić, High-resolution ultrahigh-pressure long column reversed-phase liquid chromatography for top-down proteomics, *Journal of Chromatography A*, 1498 (2017) 99-110.

- [52] T. Yoshida, Peptide Separation in Normal Phase Liquid Chromatography, *Analytical Chemistry*, 69 (1997) 3038-3043.
- [53] B. Buszewski, S. Noga, Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique, *Analytical and Bioanalytical Chemistry*, 402 (2012) 231-247.
- [54] C. West, C. Elfakir, M. Lafosse, Porous graphitic carbon: A versatile stationary phase for liquid chromatography, *Journal of Chromatography A*, 1217 (2010) 3201-3216.
- [55] J.L. Abrahams, M.P. Campbell, N.H. Packer, Building a PGC-LC-MS N-glycan retention library and elution mapping resource, *Glycoconjugate Journal*, 35 (2018) 15-29.
- [56] L.R. Ruhaak, A.M. Deelder, M. Wuhrer, Oligosaccharide analysis by graphitized carbon liquid chromatography–mass spectrometry, *Analytical and Bioanalytical Chemistry*, 394 (2009) 163-174.
- [57] M. Scalf, M.S. Westphall, J. Krause, S.L. Kaufman, L.M. Smith, Controlling Charge States of Large Ions, *Science*, 283 (1999) 194-197.
- [58] A. Guthrie, N. Bandeira, Peptide identification by tandem mass spectrometry with alternate fragmentation modes, *Mol Cell Proteomics*, 11 (2012) 550-557.
- [59] Y. Huang, J.M. Triscari, G.C. Tseng, L. Pasa-Tolic, M.S. Lipton, R.D. Smith, V.H. Wysocki, Statistical Characterization of the Charge State and Residue Dependence of Low-Energy CID Peptide Dissociation Patterns, *Analytical Chemistry*, 77 (2005) 5800-5813.
- [60] R.A. Zubarev, N.L. Kelleher, F.W. McLafferty, Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process, *Journal of the American Chemical Society*, 120 (1998) 3265-3266.
- [61] N.L. Kelleher, Peer Reviewed: Top-Down Proteomics, *Analytical Chemistry*, 76 (2004) 196 A-203 A.
- [62] Y. Mechref, Use of CID/ETD Mass Spectrometry to Analyze Glycopeptides, *Current protocols in protein science*, 68 (2012) 12.11.11 - 12.11.11.
- [63] Y.O. Tsybin, M. Ramström, M. Witt, G. Baykut, P. Håkansson, Peptide and protein characterization by high-rate electron capture dissociation Fourier transform ion cyclotron resonance mass spectrometry, *Journal of Mass Spectrometry*, 39 (2004) 719-729.
- [64] W. Cui, H.W. Rohrs, M.L. Gross, Top-Down Mass Spectrometry: Recent Developments, Applications and Perspectives, *The Analyst*, 136 (2011) 3854-3864.
- [65] N.M. Karabacak, L. Li, A. Tiwari, L.J. Hayward, P. Hong, M.L. Easterling, J.N. Agar, Sensitive and Specific Identification of Wild Type and Variant Proteins from 8 to 669 kDa Using Top-down Mass Spectrometry, *Molecular & Cellular Proteomics*, 8 (2009) 846-856.
- [66] M.J. Roth, A.J. Forbes, M.T. Boyne, Y.-B. Kim, D.E. Robinson, N.L. Kelleher, Precise and Parallel Characterization of Coding Polymorphisms, Alternative Splicing, and Modifications in Human Proteins by Mass Spectrometry, *Molecular & Cellular Proteomics*, 4 (2005) 1002-1008.
- [67] T. Nagel, B. Meyer, Simultaneous characterization of sequence polymorphisms, glycosylation and phosphorylation of fibrinogen in a direct analysis by LC–MS, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1844 (2014) 2284-2289.
- [68] J.D. Tipton, J.C. Tran, A.D. Catherman, D.R. Ahlf, K.R. Durbin, N.L. Kelleher, Analysis of Intact Protein Isoforms by Mass Spectrometry, *The Journal of Biological Chemistry*, 286 (2011) 25451-25458.
- [69] C. Uetrecht, C. Versluis, N.R. Watts, W.H. Roos, G.J.L. Wuite, P.T. Wingfield, A.C. Steven, A.J.R. Heck, High-resolution mass spectrometry of viral assemblies: Molecular composition and stability of dimorphic hepatitis B virus capsids, *Proceedings of the National Academy of Sciences*, 105 (2008) 9216-9220.
- [70] C. Uetrecht, C. Versluis, N.R. Watts, P.T. Wingfield, A.C. Steven, A.J.R. Heck, Stability and Shape of Hepatitis B Virus Capsids In Vacuo, *Angewandte Chemie International Edition*, 47 (2008) 6247-6251.

- [71] A.J. Forbes, M.T. Mazur, H.M. Patel, C.T. Walsh, N.L. Kelleher, Toward efficient analysis of <70 kDa proteins with 100% sequence coverage, *PROTEOMICS*, 1 (2001) 927-933.
- [72] M.M.J. Oosthuizen, L. Nel, J.A. Myburgh, R.L. Crookes, Purification of undegraded ceruloplasmin from outdated human plasma, *Analytical Biochemistry*, 146 (1985) 1-6.
- [73] T. Fournier, N. Medjoubi-N, D. Porquet, Alpha-1-acid glycoprotein, *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1482 (2000) 157-171.
- [74] V.A. Gaevskaia, G. Azhitskiĭ, [Isoelectric fractions of healthy human serum albumin and their ability to bind bilirubin], *Ukr Biokhim Zh* (1978), 50 (1978) 735-738.
- [75] Beer, Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten, *Annalen der Physik*, 162 (1852) 78-88.
- [76] S. Colombo, T. Buclin, L.A. Décosterd, A. Telenti, H. Furrer, B.L. Lee, J. Biollaz, C.B. Eap, Orosomucoid ( $\alpha$ 1-acid glycoprotein) plasma concentration and genetic variants: Effects on human immunodeficiency virus protease inhibitor clearance and cellular accumulation, *Clinical Pharmacology & Therapeutics*, 80 (2006) 307-318.
- [77] I. Kohler, M. Augsburg, S. Rudaz, J. Schappler, New insights in carbohydrate-deficient transferrin analysis with capillary electrophoresis–mass spectrometry, *Forensic Science International*, 243 (2014) 14-22.
- [78] J.M. Lacey, H.R. Bergen, M.J. Magera, S. Naylor, J.F. O'Brien, Rapid Determination of Transferrin Isoforms by Immunoaffinity Liquid Chromatography and Electrospray Mass Spectrometry, *Clinical Chemistry*, 47 (2001) 513-518.
- [79] P. Bengtson, B.G. Ng, J. Jaeken, G. Matthijs, H.H. Freeze, E.A. Eklund, Serum transferrin carrying the xeno-tetrasaccharide NeuAc-Gal-GlcNAc<sub>2</sub> is a biomarker of ALG1-CDG, *Journal of Inherited Metabolic Disease*, 39 (2016) 107-114.
- [80] S. Murko, R. Milačič, B. Kralj, J. Ščančar, Convective Interaction Media Monolithic Chromatography with ICPMS and Ultraperformance Liquid Chromatography–Electrospray Ionization MS Detection: A Powerful Tool for Speciation of Aluminum in Human Serum at Normal Concentration Levels, *Analytical Chemistry*, 81 (2009) 4929-4936.
- [81] D. Esteban-Fernández, M. Montes-Bayón, E. Blanco González, M.M. Gómez Gómez, M.A. Palacios, A. Sanz-Medel, Atomic (HPLC-ICP-MS) and molecular mass spectrometry (ESI-Q-TOF) to study cis-platin interactions with serum proteins, *Journal of Analytical Atomic Spectrometry*, 23 (2008) 378-384.
- [82] D.N. Perkins, D.J.C. Pappin, D.M. Creasy, J.S. Cottrell, Probability-based protein identification by searching sequence databases using mass spectrometry data, *ELECTROPHORESIS*, 20 (1999) 3551-3567.
- [83] M. Baerenfaenger, M. Moritz, B. Meyer, Quantitation of Glycopeptides by ESI/MS - size of the peptide part strongly affects the relative proportions and allows discovery of new glycan compositions of Ceruloplasmin, *Glycoconjugate Journal*, 36 (2019) 13-26.
- [84] A. Harazono, N. Kawasaki, S. Itoh, N. Hashii, A. Ishii-Watabe, T. Kawanishi, T. Hayakawa, Site-specific N-glycosylation analysis of human plasma ceruloplasmin using liquid chromatography with electrospray ionization tandem mass spectrometry, *Analytical Biochemistry*, 348 (2006) 259-268.
- [85] N.E. Hellman, J.D. Gitlin, Ceruloplasmin Metabolism and function, *Annual Review of Nutrition*, 22 (2002) 439-458.
- [86] H. Hochstrasser, J. Tomiuk, U. Walter, S. Behnke, J. Spiegel, R. Krüger, G. Becker, O. Riess, D. Berg, Functional relevance of ceruloplasmin mutations in Parkinson's disease, *The FASEB Journal*, 19 (2005) 1851-1853.
- [87] R. Xu, Y.-f. Jiang, Y.-h. Zhang, X. Yang, The optimal threshold of serum ceruloplasmin in the diagnosis of Wilson's disease: A large hospital-based study, *PLoS ONE*, 13 (2018) e0190887.

- [88] Ç. Yenisey, M. Fadiloğlu, B. Önvural, Serum copper and ceruloplasmin concentrations in patients with primary breast cancer, *Biochemical Society Transactions*, 24 (1996) 321S-321S.
- [89] M. Adamczyk-Sowa, P. Sowa, S. Mucha, J. Zostawa, B. Mazur, M. Owczarek, K. Pierzchała, Changes in Serum Ceruloplasmin Levels Based on Immunomodulatory Treatments and Melatonin Supplementation in Multiple Sclerosis Patients, *Medical science monitor : international medical journal of experimental and clinical research*, 22 (2016) 2484-2491.
- [90] A.M. Abou-Seif Mosaad, A.-A. Youssef, Oxidative Stress and Male IGF-1, Gonadotropin and Related Hormones in Diabetic Patients, *Clinical Chemistry and Laboratory Medicine*, 2001, pp. 618.
- [91] L. Jin, J. Wang, L. Zhao, H. Jin, G. Fei, Y. Zhang, M. Zeng, C. Zhong, Decreased serum ceruloplasmin levels characteristically aggravate nigral iron deposition in Parkinson's disease, *Brain*, 134 (2011) 50-58.
- [92] M. Balmaña, A. Sarrats, E. Llop, S. Barrabés, R. Saldova, M.J. Ferri, J. Figueras, E. Fort, R. de Llorens, P.M. Rudd, R. Peracaula, Identification of potential pancreatic cancer serum markers: Increased sialyl-Lewis X on ceruloplasmin, *Clinica Chimica Acta*, 442 (2015) 56-62.
- [93] S. Boivin, M.h. Aouffen, A. Fournier, M.-A. Mateescu, Molecular Characterization of Human and Bovine Ceruloplasmin Using MALDI-TOF Mass Spectrometry, *Biochemical and Biophysical Research Communications*, 288 (2001) 1006-1010.
- [94] M. Sato, M.L. Schilsky, R.J. Stockert, A.G. Morell, I. Sternlieb, Detection of multiple forms of human ceruloplasmin. A novel Mr 200,000 form, *Journal of Biological Chemistry*, 265 (1990) 2533-2537.
- [95] K.A. Moshkov, S. Lakatos, J. Hajdu, P. Závodszy, S.A. Neifakh, Proteolysis of Human Ceruloplasmin, *European Journal of Biochemistry*, 94 (1979) 127-134.
- [96] E. Ehrenwald, P.L. Fox, Isolation of Nonlabile Human Ceruloplasmin by Chromatographic Removal of a Plasma Metalloproteinase, *Archives of Biochemistry and Biophysics*, 309 (1994) 392-395.
- [97] M. Fellenberg, A. Çoksezen, B. Meyer, Characterization of Picomole Amounts of Oligosaccharides from Glycoproteins by <sup>1</sup>H NMR Spectroscopy, *Angewandte Chemie International Edition*, 49 (2010) 2630-2633.
- [98] H.N. Behnken, M. Fellenberg, M.P. Koetzler, R. Jirmann, T. Nagel, B. Meyer, Resolving the problem of chromatographic overlap by 3D cross correlation (3DCC) processing of LC, MS and NMR data for characterization of complex glycan mixtures, *Analytical and Bioanalytical Chemistry*, 404 (2012) 1427-1437.
- [99] M. Endo, K. Suzuki, K. Schmid, B. Fournet, Y. Karamanos, J. Montreuil, L. Dorland, H. van Halbeek, J.F. Vliegthart, The structures and microheterogeneity of the carbohydrate chains of human plasma ceruloplasmin. A study employing 500-MHz <sup>1</sup>H-NMR spectroscopy, *Journal of Biological Chemistry*, 257 (1982) 8755-8760.
- [100] W. Gabryelski, K.L. Froese, Rapid and sensitive differentiation of anomers, linkage, and position isomers of disaccharides using High-Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS), *Journal of the American Society for Mass Spectrometry*, 14 (2003) 265-277.
- [101] F.-T.A. Chen, T.S. Dobashi, R.A. Evangelista, Quantitative analysis of sugar constituents of glycoproteins by capillary electrophoresis, *Glycobiology*, 8 (1998) 1045-1052.
- [102] I. Bento, C. Peixoto, V.N. Zaitsev, P.F. Lindley, Ceruloplasmin revisited: structural and functional roles of various metal cation-binding sites, *Acta Crystallographica Section D: Biological Crystallography*, 63 (2007) 240-248.
- [103] J. Kyte, R.F. Doolittle, A simple method for displaying the hydrophobic character of a protein, *Journal of Molecular Biology*, 157 (1982) 105-132.

- [104] M. Wührer, M.I. Catalina, A.M. Deelder, C.H. Hokke, Glycoproteomics based on tandem mass spectrometry of glycopeptides, *Journal of Chromatography B*, 849 (2007) 115-128.
- [105] A.G. Harrison, Energy-resolved mass spectrometry: a comparison of quadrupole cell and cone-voltage collision-induced dissociation, *Rapid Communications in Mass Spectrometry*, 13 (1999) 1663-1670.
- [106] M. Hiroshi, T. Toshifumi, S. Yasutsugu, M. Takekiyo, Optimization of skimmer voltages of an electrospray ion source coupled with a magnetic sector instrument, *Rapid Communications in Mass Spectrometry*, 8 (1994) 205-210.
- [107] C.-H. Lin, C. Krisp, N.H. Packer, M.P. Molloy, Development of a data independent acquisition mass spectrometry workflow to enable glycopeptide analysis without predefined glycan compositional knowledge, *Journal of Proteomics*, 172 (2018) 68-75.
- [108] L.F. Zacchi, B.L. Schulz, SWATH-MS Glycoproteomics Reveals Consequences of Defects in the Glycosylation Machinery, *Molecular & Cellular Proteomics*, 15 (2016) 2435-2447.
- [109] K.Y.B. Yeo, P.K. Chrysanthopoulos, A.S. Nouwens, E. Marcellin, B.L. Schulz, High-performance targeted mass spectrometry with precision data-independent acquisition reveals site-specific glycosylation macroheterogeneity, *Analytical Biochemistry*, 510 (2016) 106-113.
- [110] C.-C. Tsou, D. Avtonomov, B. Larsen, M. Tucholska, H. Choi, A.-C. Gingras, A.I. Nesvizhskii, DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics, *Nature Methods*, 12 (2015) 258.
- [111] B. MacLean, D.M. Tomazela, N. Shulman, M. Chambers, G.L. Finney, B. Frewen, R. Kern, D.L. Tabb, D.C. Liebler, M.J. MacCoss, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments, *Bioinformatics*, 26 (2010) 966-968.
- [112] L.C. Gillet, P. Navarro, S. Tate, H. Röst, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis, *Molecular & Cellular Proteomics*, 11 (2012).
- [113] H. Hochstrasser, P. Bauer, U. Walter, S. Behnke, J. Spiegel, I. Csoti, B. Zeiler, A. Bornemann, J. Pahnke, G. Becker, O. Riess, D. Berg, Ceruloplasmin gene variations and substantia nigra hyperechogenicity in Parkinson disease, *Neurology*, 63 (2004) 1912-1917.
- [114] S. Ayton, P. Lei, J.A. Duce, B.X.W. Wong, A. Sedjahtera, P.A. Adlard, A.I. Bush, D.I. Finkelstein, Ceruloplasmin dysfunction and therapeutic potential for Parkinson disease, *Annals of Neurology*, 73 (2013) 554-559.
- [115] A. Gonzalez-Quintela, R. Alende, F. Gude, J. Campos, J. Rey, L.M. Meijide, C. Fernandez-Merino, C. Vidal, Serum levels of immunoglobulins (IgG, IgA, IgM) in a general adult population and their relationship with alcohol consumption, smoking and common metabolic abnormalities, *Clinical and experimental immunology*, 151 (2008) 42-50.
- [116] V. Irani, A.J. Guy, D. Andrew, J.G. Beeson, P.A. Ramsland, J.S. Richards, Molecular properties of human IgG subclasses and their implications for designing therapeutic monoclonal antibodies against infectious diseases, *Molecular Immunology*, 67 (2015) 171-182.
- [117] D. Dunn-Walters, L. Boursier, J. Spencer, Effect of somatic hypermutation on potential N-glycosylation sites in human immunoglobulin heavy chain variable regions, *Molecular Immunology*, 37 (2000) 107-113.
- [118] K. Zheng, C. Bantog, R. Bayer, The impact of glycosylation on monoclonal antibody conformation and stability, *mAbs*, 3 (2011) 568-576.
- [119] R. Plomp, L.R. Ruhaak, H.-W. Uh, K.R. Reiding, M. Selman, J.J. Houwing-Duistermaat, P.E. Slagboom, M. Beekman, M. Wührer, Subclass-specific IgG

- glycosylation is associated with markers of inflammation and metabolic health, *Scientific Reports*, 7 (2017) 12325.
- [120] A.C. Russell, M. Šimurina, M.T. Garcia, M. Novokmet, Y. Wang, I. Rudan, H. Campbell, G. Lauc, M.G. Thomas, W. Wang, The N-glycosylation of immunoglobulin G as a novel biomarker of Parkinson's disease, *Glycobiology*, 27 (2017) 501-510.
- [121] J. Krištić, F. Vučković, C. Menni, L. Klarić, T. Keser, I. Beceheli, M. Pučić-Baković, M. Novokmet, M. Mangino, K. Thaqi, P. Rudan, N. Novokmet, J. Šarac, S. Missoni, I. Kolčić, O. Polašek, I. Rudan, H. Campbell, C. Hayward, Y. Aulchenko, A. Valdes, J.F. Wilson, O. Gornik, D. Primorac, V. Zoldoš, T. Spector, G. Lauc, Glycans Are a Novel Biomarker of Chronological and Biological Ages, *The Journals of Gerontology: Series A*, 69 (2014) 779-789.
- [122] D. Zhang, B. Chen, Y. Wang, P. Xia, C. He, Y. Liu, R. Zhang, M. Zhang, Z. Li, Disease-specific IgG Fc N-glycosylation as personalized biomarkers to differentiate gastric cancer from benign gastric diseases, *Scientific Reports*, 6 (2016) 25957.
- [123] S. Ren, Z. Zhang, C. Xu, L. Guo, R. Lu, Y. Sun, J. Guo, R. Qin, W. Qin, J. Gu, Distribution of IgG galactosylation as a promising biomarker for cancer screening in multiple cancer types, *Cell Research*, 26 (2016) 963-966.
- [124] T.-T. Chang, H.-W. Tsai, C.-H. Ho, Fucosyl-Agalactosyl IgG<sub>1</sub> Induces Cholangiocarcinoma Metastasis and Early Recurrence by Activating Tumor-Associated Macrophage, *Cancers*, 10 (2018) 460.
- [125] I. Gudelj, G. Lauc, M. Pezer, Immunoglobulin G glycosylation in aging and diseases, *Cellular Immunology*, (2018).
- [126] I. Gudelj, P.P. Salo, I. Trbojević-Akmačić, M. Albers, D. Primorac, M. Perola, G. Lauc, Low galactosylation of IgG associates with higher risk for future diagnosis of rheumatoid arthritis during 10 years of follow-up, *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1864 (2018) 2034-2039.
- [127] M. Tomana, R.E. Schrohenloher, W.J. Koopman, G.S. Alarcán, W.A. Paul, Abnormal glycosylation of serum igg from patients with chronic inflammatory diseases, *Arthritis & Rheumatism*, 31 (1988) 333-338.
- [128] B.J. Scallon, S.H. Tam, S.G. McCarthy, A.N. Cai, T.S. Raju, Higher levels of sialylated Fc glycans in immunoglobulin G molecules can adversely impact functionality, *Molecular Immunology*, 44 (2007) 1524-1534.
- [129] Y. Kaneko, F. Nimmerjahn, J.V. Ravetch, Anti-Inflammatory Activity of Immunoglobulin G Resulting from Fc Sialylation, *Science*, 313 (2006) 670-673.
- [130] N. Dennis Chasteen, Human serotransferrin: structure and function, *Coordination Chemistry Reviews*, 22 (1977) 1-36.
- [131] K. Matsumoto, Y. Maeda, S. Kato, H. Yuki, Alteration of asparagine-linked glycosylation in serum transferrin of patients with hepatocellular carcinoma, *Clinica Chimica Acta*, 224 (1994) 1-8.
- [132] H. Walter, I. Hertling, N. Benda, B. König, K. Ramskogler, A. Riegler, B. Semler, A. Zoghalmi, O.M. Lesch, Sensitivity and specificity of carbohydrate-deficient transferrin in drinking experiments and different patients, *Alcohol*, 25 (2001) 189-194.
- [133] J.-q. Sheng, S.-r. Li, Z.-t. Wu, C.-h. Xia, X. Wu, J. Chen, J. Rao, Transferrin Dipstick as a Potential Novel Test for Colon Cancer Screening: A Comparative Study With Immuno Fecal Occult Blood Test, *Cancer Epidemiology Biomarkers & Prevention*, 18 (2009) 2182-2185.
- [134] E. Gruszewska, B. Cylwik, M. Gudowska, B. Kedra, M. Szmitkowski, L. Chrostek, Changes in Transferrin Isoforms in Pancreatic Cancer, *Annals of Clinical & Laboratory Science*, 46 (2016) 286-290.
- [135] D.J. Lefeber, E. Morava, J. Jaeken, How to find and diagnose a CDG due to defective N-glycosylation, *Journal of inherited metabolic disease*, 34 (2011) 849-852.

- [136] M. Van Scherpenzeel, E. Willems, D.J. Lefeber, Clinical diagnostics and therapy monitoring in the congenital disorders of glycosylation, *Glycoconjugate journal*, 33 (2016) 345-358.
- [137] M. Thevis, R.R.O. Loo, J.A. Loo, Mass spectrometric characterization of transferrins and their fragments derived by reduction of disulfide bonds, *Journal of the American Society for Mass Spectrometry*, 14 (2003) 635-647.
- [138] S. Wang, I.A. Kaltashov, Identification of reduction-susceptible disulfide bonds in transferrin by differential alkylation using O(16)/O(18) labeled iodoacetic acid, *Journal of the American Society for Mass Spectrometry*, 26 (2015) 800-807.
- [139] M.E. del Castillo Busto, M. Montes-Bayón, E. Blanco-González, J. Meija, A. Sanz-Medel, Strategies To Study Human Serum Transferrin Isoforms Using Integrated Liquid Chromatography ICPMS, MALDI-TOF, and ESI-Q-TOF Detection: Application to Chronic Alcohol Abuse, *Analytical Chemistry*, 77 (2005) 5615-5621.
- [140] L. Fornelli, J. Parra, R. Hartmer, C. Stoermer, M. Lubeck, Y.O. Tsybin, Top-down analysis of 30–80 kDa proteins by electron transfer dissociation time-of-flight mass spectrometry, *Analytical and Bioanalytical Chemistry*, 405 (2013) 8505-8514.
- [141] M. Lek, K.J. Karczewski, E.V. Minikel, K.E. Samocha, E. Banks, T. Fennell, A.H. O'Donnell-Luria, J.S. Ware, A.J. Hill, B.B. Cummings, T. Tukiainen, D.P. Birnbaum, J.A. Kosmicki, L.E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D.N. Cooper, N. DeFlaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M.I. Kurki, A.L. Moonshine, P. Natarajan, L. Orozco, G.M. Peloso, R. Poplin, M.A. Rivas, V. Ruano-Rubio, S.A. Rose, D.M. Ruderfer, K. Shakir, P.D. Stenson, C. Stevens, B.P. Thomas, G. Tiao, M.T. Tusie-Luna, B. Weisburd, H.-H. Won, D. Yu, D.M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J.C. Florez, S.B. Gabriel, G. Getz, S.J. Glatt, C.M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M.I. McCarthy, D. McGovern, R. McPherson, B.M. Neale, A. Palotie, S.M. Purcell, D. Saleheen, J.M. Scharf, P. Sklar, P.F. Sullivan, J. Tuomilehto, M.T. Tsuang, H.C. Watkins, J.G. Wilson, M.J. Daly, D.G. MacArthur, C. Exome Aggregation, Analysis of protein-coding genetic variation in 60,706 humans, *Nature*, 536 (2016) 285.
- [142] M. Baerenfaenger, B. Meyer, Intact Human Alpha-Acid Glycoprotein Analyzed by ESI-qTOF-MS: Simultaneous Determination of the Glycan Composition of Multiple Glycosylation Sites, *Journal of Proteome Research*, 17 (2018) 3693-3703.
- [143] S. Iijima, K. Shiba, M. Kimura, K. Nagai, T. Iwai, Changes of  $\alpha$ 1-acid glycoprotein microheterogeneity in acute inflammation stages analyzed by isoelectric focusing using serum obtained postoperatively, *Electrophoresis*, 21 (2000) 753-759.
- [144] I. Ryden, G. Skude, A. Lundblad, P. Pahlsson, Glycosylation of  $\alpha$ 1-acid glycoprotein in inflammatory disease: analysis by high-pH anion-exchange chromatography and concanavalin A crossed affinity immunoelectrophoresis, *Glycoconjugate Journal*, 14 (1997) 481-488.
- [145] J.Y. Lee, H.K. Lee, G.W. Park, H. Hwang, H.K. Jeong, K.N. Yun, E.S. Ji, K.H. Kim, J.S. Kim, J.W. Kim, S.H. Yun, C.-W. Choi, S.I. Kim, J.-S. Lim, S.-K. Jeong, Y.-K. Paik, S.-Y. Lee, J. Park, S.Y. Kim, Y.-J. Choi, Y.-I. Kim, J. Seo, J.-Y. Cho, M.J. Oh, N. Seo, H.J. An, J.Y. Kim, J.S. Yoo, Characterization of Site-Specific N-Glycopeptide Isoforms of  $\alpha$ -1-Acid Glycoprotein from an Interlaboratory Study Using LC-MS/MS, *Journal of Proteome Research*, 15 (2016) 4146-4164.
- [146] D. Zhang, J. Huang, D. Luo, X. Feng, Y. Liu, Y. Liu, Glycosylation change of alpha-1-acid glycoprotein as a serum biomarker for hepatocellular carcinoma and cirrhosis, *Biomarkers in Medicine*, 11 (2017) 423-430.
- [147] M. Ferens-Sieczkowska, E. Kratz, B. Kossowska, E. Passowicz-Muszynska, R. Jankowska, Comparison of haptoglobin and alpha<sub>1</sub>-acid glycoprotein glycosylation in the sera of small cell and non-small cell lung cancer patients., *Postepy Hig Med Dosw (Online)*, (2013) 828-836.



- [148] P. Hrycaj, M. Sobieska, S. Mackiewicz, W. Müller, Microheterogeneity of alpha 1 acid glycoprotein in rheumatoid arthritis: dependent on disease duration?, *Annals of the Rheumatic Diseases*, 52 (1993) 138-141.
- [149] R. Saldo, L. Royle, C.M. Radcliffe, U.M. Abd Hamid, R. Evans, J.N. Arnold, R.E. Banks, R. Hutson, D.J. Harvey, R. Antrobus, S.M. Petrescu, R.A. Dwek, P.M. Rudd, Ovarian Cancer is Associated with Changes in Glycosylation in Both Acute-Phase Proteins and IgG, *Glycobiology*, 17 (2007) 1344-1356.
- [150] T. Imre, T. Kremmer, K. Héberger, É. Molnár-Szöllősi, K. Ludányi, G. Pócsfalvi, A. Malorni, L. Drahos, K. Vékey, Mass spectrometric and linear discriminant analysis of N-glycans of human serum alpha-1-acid glycoprotein in cancer patients and healthy individuals, *Journal of Proteomics*, 71 (2008) 186-197.
- [151] A. Sarrats, R. Saldo, E. Pla, E. Fort, D.J. Harvey, W.B. Struwe, R. de Llorens, P.M. Rudd, R. Peracaula, Glycosylation of liver acute-phase proteins in pancreatic cancer and chronic pancreatitis, *PROTEOMICS – Clinical Applications*, 4 (2010) 432-448.
- [152] E. Giménez, M. Balmaña, J. Figueras, E. Fort, C.d. Bolós, V. Sanz-Nebot, R. Peracaula, A. Rizzi, Quantitative analysis of N-glycans from human alpha-acid-glycoprotein using stable isotope labeling and zwitterionic hydrophilic interaction capillary liquid chromatography electrospray mass spectrometry as tool for pancreatic disease diagnosis, *Analytica Chimica Acta*, 866 (2015) 59-68.
- [153] C. Sottani, M. Fiorentino, C. Minoia, Matrix Performance in Matrix-assisted Laser Desorption/Ionization for Molecular Weight Determination in Sialyl and Non-sialyl Oligosaccharide Proteins, *Rapid Communications in Mass Spectrometry*, 11 (1997) 907-913.
- [154] K. Nagy, K. Vékey, T. Imre, K. Ludányi, M.P. Barrow, P.J. Derrick, Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry of Human  $\alpha$ -1-Acid Glycoprotein, *Analytical Chemistry*, 76 (2004) 4998-5005.
- [155] S. Ongay, C. Neusüß, Isoform differentiation of intact AGP from human serum by capillary electrophoresis–mass spectrometry, *Analytical and Bioanalytical Chemistry*, 398 (2010) 845-855.
- [156] S. Ongay, C. Neusüß, S. Vaas, J.C. Díez-Masa, M. de Frutos, Evaluation of the effect of the immunopurification-based procedures on the CZE-UV and CZE-ESI-TOF-MS determination of isoforms of intact  $\alpha$ -1-acid glycoprotein from human serum, *ELECTROPHORESIS*, 31 (2010) 1796-1804.
- [157] I. Yuasa, K. Umetsu, U. Vogt, H. Nakamura, E. Nanba, N. Tamaki, Y. Irizawa, Human orosomucoid polymorphism: molecular basis of the three common ORM1 alleles, ORM1\*F1, ORM1\*F2, and ORM1\*S, *Human Genetics*, 99 (1997) 393-398.
- [158] T. Ikenaka, H. Bammerlin, H. Kaufmann, K. Schmid, The Amino-terminal Peptide of  $\alpha$ -1-Acid Glycoprotein, *Journal of Biological Chemistry*, (1966) 5560-5563.
- [159] B. Fournet, J. Montreuil, G. Strecker, L. Dorland, J. Haverkamp, J.F.G. Vliegenthart, J.P. Binette, K. Schmid, Determination of the primary structures of 16 asialo-carbohydrate units derived from human plasma  $\alpha$ -1-acid glycoprotein by 360-MHz proton NMR spectroscopy and permethylation analysis, *Biochemistry*, 17 (1978) 5206-5214.
- [160] T. Imre, G. Schlosser, G. Pócsfalvi, R. Siciliano, É. Molnár-Szöllősi, T. Kremmer, A. Malorni, K. Vékey, Glycosylation site analysis of human alpha-1-acid glycoprotein (AGP) by capillary liquid chromatography—electrospray mass spectrometry, *Journal of Mass Spectrometry*, 40 (2005) 1472-1483.
- [161] M.J. Treuheit, C.E. Costello, H.B. Halsall, Analysis of the five glycosylation sites of human alpha 1-acid glycoprotein, *Biochemical Journal*, 283 (1992) 105-112.
- [162] S. Shiian, V. Nasonov, N. Bovin, V. Aleshkin, L. Novikova, A. Liutov, Structure of carbohydrate chains of molecular forms of alpha1-acidic glycoprotein from ascitic fluid from patients with stomach cancer, *Bioorg Khim*, 20 (1994) 1125-1131.

- [163] S.D. Shiyan, N.V. Bovin, Carbohydrate composition and immunomodulatory activity of different glycoforms of  $\alpha$ 1-acid glycoprotein, *Glycoconjugate Journal*, 14 (1997) 631-638.
- [164] S.H. Walker, B.N. Papas, D.L. Comins, D.C. Muddiman, The Interplay of Permanent Charge and Hydrophobicity in the Electrospray Ionization of Glycans, *Analytical chemistry*, 82 (2010) 6636-6642.
- [165] D.J. Harvey, Quantitative aspects of the matrix-assisted laser desorption mass spectrometry of complex oligosaccharides, *Rapid Communications in Mass Spectrometry*, 7 (1993) 614-619.
- [166] M. Baerenfaenger, B. Meyer, Simultaneous characterization of SNPs and N-glycans from multiple glycosylation sites of intact  $\beta$ -2-glycoprotein-1 (B2GP1) by ESI-qTOF-MS, *BBA Proteins and Proteomics*, (2019) under revision.
- [167] H.E. Schultze, K. Heide, H. Haupt, Über ein bisher unbekanntes niedermolekulares  $\beta$ 2-Globulin des Humanserums, *Naturwissenschaften*, 48 (1961) 719-719.
- [168] J. Lozier, N. Takahashi, F.W. Putnam, Complete amino acid sequence of human plasma beta 2-glycoprotein I, *Proceedings of the National Academy of Sciences*, 81 (1984) 3640-3644.
- [169] P.G. De Groot, J.C.M. Meijers,  $\beta$ 2-Glycoprotein I: evolution, structure and function, *Journal of Thrombosis and Haemostasis*, 9 (2011) 1275-1284.
- [170] I. Schousboe, Inositolphospholipid-accelerated activation of prekallikrein by activated factor XII and its inhibition by  $\beta$ 2-glycoprotein I, *European Journal of Biochemistry*, 176 (1988) 629-636.
- [171] M.T. Walsh, H. Watzlawick, F.W. Putnam, K. Schmid, R. Brossmer, Effect of the carbohydrate moiety on the secondary structure of  $\beta$ 2-glycoprotein. I. Implications for the biosynthesis and folding of glycoproteins, *Biochemistry*, 29 (1990) 6250-6257.
- [172] A. Sato, K. Nakazawa, A. Sugawara, Y. Yamazaki, K. Ebina, The interaction of  $\beta$ 2-glycoprotein I with lysophosphatidic acid in platelet aggregation and blood clotting, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1866 (2018) 1232-1241.
- [173] D.A. Kandiah, A. Sali, Y. Sheng, E.J. Victoria, D.M. Marquis, S.M. Coutts, S.A. Krillis, Current Insights into the "Antiphospholipid" Syndrome: Clinical, Immunological, and Molecular Aspects, in: F.J. Dixon (Ed.) *Advances in Immunology*, Academic Press, Place Published, 1998, pp. 507-563.
- [174] V.X. Du, H. Kelchtermans, P.G. de Groot, B. de Laat, From antibody to clinical phenotype, the black box of the antiphospholipid syndrome: Pathogenic mechanisms of the antiphospholipid syndrome, *Thrombosis Research*, 132 (2013) 319-326.
- [175] Ç. Ađar, G.M.A. van Os, M. Mörgelein, R.R. Sprenger, J.A. Marquart, R.T. Urbanus, R.H.W.M. Derksen, J.C.M. Meijers, P.G. de Groot,  $\beta$ 2-Glycoprotein I can exist in 2 conformations: implications for our understanding of the antiphospholipid syndrome, *Blood*, 116 (2010) 1336-1343.
- [176] M. Hammel, M. Kriechbaum, A. Gries, G.M. Kostner, P. Laggner, R. Prassl, Solution Structure of Human and Bovine  $\beta$ 2-Glycoprotein I Revealed by Small-angle X-ray Scattering, *Journal of Molecular Biology*, 321 (2002) 85-97.
- [177] T. Koike, K. Ichikawa, H. Kasahara, T. Atsumi, A. Tsutsumi, E. Matsuura, Epitopes on  $\beta$ 2-GPI recognized by anticardiolipin antibodies, *Lupus*, 7 (1998) 14-17.
- [178] B. de Laat, R.H.W.M. Derksen, M. van Lummel, M.T.T. Pennings, P.G. de Groot, Pathogenic anti- $\beta$ 2-glycoprotein I antibodies recognize domain I of  $\beta$ 2-glycoprotein I only after a conformational change, *Blood*, 107 (2006) 1916-1924.
- [179] A. Dupuy d'Angeac, I. Stefas, H. Graafland, F. de Lamotte, M. Rucheton, C. Palais, A.-K. Eriksson, P. Bosc, C. Rosé, R. Chicheportiche, Biotinylation of glycan chains in  $\beta$ 2 glycoprotein I induces dimerization of the molecule and its detection by the human autoimmune anti-cardiolipin antibody EY2C9, *Biochemical Journal*, 393 (2006) 117-127.
- [180] A. Kondo, T. Miyamoto, O. Yonekawa, A.M. Giessing, E.C. Østerlund, O.N. Jensen, Glycopeptide profiling of beta-2-glycoprotein I by mass spectrometry reveals attenuated

- sialylation in patients with antiphospholipid syndrome, *Journal of Proteomics*, 73 (2009) 123-133.
- [181] Q. Chen, M.I. Kamboh, Complete DNA Sequence Variation in the Apolipoprotein H ( $\beta$ 2-glycoprotein I) Gene and Identification of Informative SNPs, *Annals of Human Genetics*, 70 (2006) 1-11.
- [182] A. Steinkasserer, C. Dörner, R. Würzner, R.B. Sim, Human  $\beta$ 2-glycoprotein I: molecular analysis of DNA and amino acid polymorphism, *Human Genetics*, 91 (1993) 401-402.
- [183] A.-J. Chamorro, M. Marcos, J.-A. Mirón-Canelo, R. Cervera, G. Espinosa, Val247Leu beta2-glycoprotein-I allelic variant is associated with antiphospholipid syndrome: Systematic review and meta-analysis, *Autoimmunity Reviews*, 11 (2012) 705-712.
- [184] S. Yasuda, T. Atsumi, E. Matsuura, K. Kaihara, D. Yamamoto, K. Ichikawa, T. Koike, Significance of valine/leucine247 polymorphism of  $\beta$ 2-glycoprotein I in antiphospholipid syndrome: Increased reactivity of anti- $\beta$ 2-glycoprotein I autoantibodies to the valine247  $\beta$ 2-glycoprotein I variant, *Arthritis & Rheumatism*, 52 (2005) 212-218.
- [185] F. El-Assaad, S.A. Krilis, B. Giannakopoulos, Posttranslational forms of beta 2-glycoprotein I in the pathogenesis of the antiphospholipid syndrome, *Thrombosis Journal*, 14 (2016) 20.
- [186] Y. Ioannou, J.-Y. Zhang, M. Qi, L. Gao, J.C. Qi, D.-M. Yu, H. Lau, A.D. Sturgess, P.G. Vlachoyiannopoulos, H.M. Moutsopoulos, A. Rahman, C. Pericleous, T. Atsumi, T. Koike, S. Heritier, B. Giannakopoulos, S.A. Krilis, Novel assays of thrombogenic pathogenicity in the antiphospholipid syndrome based on the detection of molecular oxidative modification of the major autoantigen  $\beta$ 2-glycoprotein I, *Arthritis & Rheumatism*, 63 (2011) 2774-2782.
- [187] N. Leymarie, P.J. Griffin, K. Jonscher, D. Kolarich, R. Orlando, M. McComb, J. Zaia, J. Aguilan, W.R. Alley, F. Altmann, L.E. Ball, L. Basumallick, C.R. Bazemore-Walker, H. Behnken, M.A. Blank, K.J. Brown, S.-C. Bunz, C.W. Cairo, J.F. Cipollo, R. Daneshfar, H. Desaire, R.R. Drake, E.P. Go, R. Goldman, C. Gruber, A. Halim, Y. Hathout, P.J. Hensbergen, D.M. Horn, D. Hurum, W. Jabs, G. Larson, M. Ly, B.F. Mann, K. Marx, Y. Mechref, B. Meyer, U. Möginger, C. Neusüß, J. Nilsson, M.V. Novotny, J.O. Nyalwidhe, N.H. Packer, P. Pompach, B. Reiz, A. Resemann, J.S. Rohrer, A. Ruthenbeck, M. Sanda, J.M. Schulz, U. Schweiger-Hufnagel, C. Sihlbom, E. Song, G.O. Staples, D. Suckau, H. Tang, M. Thaysen-Andersen, R.I. Viner, Y. An, L. Valmu, Y. Wada, M. Watson, M. Windwarder, R. Whittall, M. Wuhrer, Y. Zhu, C. Zou, Interlaboratory Study on Differential Analysis of Protein Glycosylation by Mass Spectrometry: The ABRF Glycoprotein Research Multi-Institutional Study 2012, *Mol Cell Proteomics*, 12 (2013) 2935-2951.
- [188] M. Baerenfaenger, M. Moritz, B. Meyer, Quantitation of Glycopeptides by ESI/MS - Size of the Peptide Part Strongly Affects the Relative Amounts and Allows Discovery of New Glycan Structures of Ceruloplasmin, submitted, (2018).
- [189] I.A. Kaltashov, A. Mohimen, Electrospray ionization mass spectrometry can provide estimates of protein surface areas in solution, *Analytical chemistry*, 77 (2005) 5370-5379.
- [190] L. Konermann, D.J. Douglas, Equilibrium unfolding of proteins monitored by electrospray ionization mass spectrometry: distinguishing two-state from multi-state transitions, *Rapid Communications in Mass Spectrometry*, 12 (1998) 435-442.
- [191] J. Li, C. Santambrogio, S. Brocca, G. Rossetti, P. Carloni, R. Grandori, Conformational effects in protein electrospray-ionization mass spectrometry, *Mass Spectrometry Reviews*, 35 (2016) 111-122.
- [192] J. Seo, W. Hoffmann, S. Warnke, M.T. Bowers, K. Pagel, G. von Helden, Retention of Native Protein Structures in the Absence of Solvent: A Coupled Ion Mobility and Spectroscopic Study, *Angewandte Chemie International Edition*, 55 (2016) 14173-14176.
- [193] C. Uetrecht, R.J. Rose, E. van Duijn, K. Lorenzen, A.J.R. Heck, Ion mobility mass spectrometry of proteins and protein assemblies, *Chemical Society Reviews*, 39 (2010) 1633-1655.

- [194] A.G. Ferrige, M.J. Seddon, S. Jarvis, J. Skilling, R. Aplin, Maximum entropy deconvolution in electrospray mass spectrometry, *Rapid Communications in Mass Spectrometry*, 5 (1991) 374-377.
- [195] F. Clerc, K.R. Reiding, B.C. Jansen, G.S.M. Kammeijer, A. Bondt, M. Wuhrer, Human plasma protein N-glycosylation, *Glycoconjugate Journal*, 33 (2016) 309-343.
- [196] A. Steinkasserer, C. Estaller, E.H. Weiss, R.B. Sim, A.J. Day, Complete nucleotide and deduced amino acid sequence of human beta 2-glycoprotein I, *Biochemical Journal*, 277 (1991) 387-391.
- [197] H. Mehdi, C.E. Aston, D.K. Sanghera, R.F. Hamman, M.I. Kamboh, Genetic variation in the apolipoprotein H ( $\beta$ 2-glycoprotein I) gene affects plasma apolipoprotein H concentrations, *Human Genetics*, 105 (1999) 63-71.
- [198] D.K. Sanghera, T. Kristensen, R.F. Hamman, M.I. Kamboh, Molecular basis of the apolipoprotein H ( $\beta$ 2-glycoprotein I) protein polymorphism, *Human Genetics*, 100 (1997) 57-62.
- [199] K.A. Mather, A. Thalamuthu, C. Oldmeadow, F. Song, N.J. Armstrong, A. Poljak, E.G. Holliday, M. McEvoy, J.B. Kwok, A.A. Assareh, S. Reppermund, N.A. Kochan, T. Lee, D. Ames, M.J. Wright, J.N. Trollor, P.W. Schofield, H. Brodaty, R.J. Scott, P.R. Schofield, J.R. Attia, P.S. Sachdev, Genome-wide significant results identified for plasma apolipoprotein H levels in middle-aged and older adults, *Scientific Reports*, 6 (2016) 23675.
- [200] T. Kristensen, I. Schousboe, E. Boel, E.M. Mulvihill, R.R. Hansen, K.B. Møller, N.P.H. Møller, L. Sottrup-Jensen, Molecular cloning and mammalian expression of human  $\beta$ 2-glycoprotein I cDNA, *FEBS Letters*, 289 (1991) 183-186.
- [201] J.W. Dennis, M. Granovsky, C.E. Warren, Glycoprotein glycosylation and cancer progression, *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1473 (1999) 21-34.
- [202] R. Gambino, G. Ruiu, G. Pagano, M. Cassader, Qualitative Analysis of the Carbohydrate Composition of Apolipoprotein H, *Journal of Protein Chemistry*, 16 (1997) 205-212.
- [203] Y. Ho, K. Ahuja, H. Körner, M. Adams,  $\beta$ 2GP1, Anti- $\beta$ 2GP1 Antibodies and Platelets: Key Players in the Antiphospholipid Syndrome, *Antibodies*, 5 (2016) 12.
- [204] R. Schwarzenbacher, K. Zeth, K. Diederichs, A. Gries, G.M. Kostner, P. Laggner, R. Prassl, Crystal structure of human  $\beta$ 2-glycoprotein I: implications for phospholipid binding and the antiphospholipid syndrome, *The EMBO Journal*, 18 (1999) 6228-6239.
- [205] J. Vertrees, FindSurfaceResidues, <https://pymolwiki.org/index.php/FindSurfaceResidues>, 2018/20/10.
- [206] K.R. Reiding, L.R. Ruhaak, H.-W. Uh, S. el Bouhaddani, E.B. van den Akker, R. Plomp, L.A. McDonnell, J.J. Houwing-Duistermaat, P.E. Slagboom, M. Beekman, M. Wuhrer, Human plasma N-glycosylation as analyzed by MALDI-FTICR-MS associates with markers of inflammation and metabolic health, *Molecular & Cellular Proteomics*, (2016).
- [207] M. Sogabe, H. Nozaki, N. Tanaka, T. Kubota, H. Kaji, A. Kuno, A. Togayachi, M. Gotoh, H. Nakanishi, T. Nakanishi, M. Mikami, N. Suzuki, K. Kiguchi, Y. Ikehara, H. Narimatsu, Novel Glycobiomarker for Ovarian Cancer That Detects Clear Cell Carcinoma, *Journal of Proteome Research*, 13 (2014) 1624-1635.
- [208] Y. Qian, Y. Wang, X. Zhang, L. Zhou, Z. Zhang, J. Xu, Y. Ruan, S. Ren, C. Xu, J. Gu, Quantitative Analysis of Serum IgG Galactosylation Assists Differential Diagnosis of Ovarian Cancer, *Journal of Proteome Research*, 12 (2013) 4046-4055.
- [209] K.F. Doak, Jodi Cruickshank, Yvonne; Smith, Kevin D, Alterations in the glycosylation pattern of alpha-1-glycoprotein may be diagnostic for the detection of breast cancer and/or reducing the effect of chemotherapy in vivo. , Nova Science Publishers, Inc., (2014) 143-171.

- [210] V.C.S. M.V. Croce, E. Lacunza and A. Segal-Eiras, Alpha1-acid glycoprotein (AGP): a possible carrier of sialyl lewis X (slewis X) antigen in colorectal carcinoma, *Cellular and Molecular Biology*, 20 (2005) 91-97.
- [211] M. Balmaña, E. Giménez, A. Puerta, E. Llop, J. Figueras, E. Fort, V. Sanz-Nebot, C. de Bolós, A. Rizzi, S. Barrabés, M. de Frutos, R. Peracaula, Increased  $\alpha$ 1-3 fucosylation of  $\alpha$ -1-acid glycoprotein (AGP) in pancreatic cancer, *Journal of Proteomics*, 132 (2016) 144-154.
- [212] O. Sara, M.Á.P. J., N. Christian, d.F. Mercedes, Statistical evaluation of CZE-UV and CZE-ESI-MS data of intact  $\alpha$ -1-acid glycoprotein isoforms for their use as potential biomarkers in bladder cancer, *ELECTROPHORESIS*, 31 (2010) 3314-3325.
- [213] W. Van Dijk, E.C. Havenaar, E.C.M. Brinkman-Van Der Linden,  $\alpha$ 1-Acid glycoprotein (orosomucoid): pathophysiological changes in glycosylation in relation to its function, *Glycoconjugate Journal*, 12 (1995) 227-233.
- [214] J.N. Arnold, R. Saldova, U.M.A. Hamid, P.M. Rudd, Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation, *PROTEOMICS*, 8 (2008) 3284-3293.
- [215] R. Peracaula, A. Sarrats, P.M. Rudd, Liver proteins as sensor of human malignancies and inflammation, *PROTEOMICS – Clinical Applications*, 4 (2010) 426-431.
- [216] T. Nagel, F. Klaus, I.G. Ibanez, H. Wege, A. Lohse, B. Meyer, Fast and facile analysis of glycosylation and phosphorylation of fibrinogen from human plasma—correlation with liver cancer and liver cirrhosis, *Analytical and Bioanalytical Chemistry*, 410 (2018) 7965-7977.

## 10. Appendix

### 10.1. Results for Mascot search after Tryptic Digest of Purified Ceruloplasmin

Table A 1. Purified ceruloplasmin was digested with trypsin and analyzed by LC-ESI-MS/MS. An automated peptide identification was performed using the Mascot search engine.

Sample	Accession	Score	Mass
A	Ceruloplasmin	253	122128
	Apolipoprotein A-1	156	30759
	Apolipoprotein A-2	89	11168
	Immunoglobulin kappa light chain	39	23364
C	Cationic trypsin (bovine)	86	25769
	Ceruloplasmin	43	122128
	Immunoglobulin kappa light chain	37	23364
	Prothrombin	30	69992
	Apolipoprotein A-2	22	11168
D	Ceruloplasmin	344	122128
	Apolipoprotein A-1	204	30759
	Apolipoprotein A-2	133	11168
	Immunoglobulin kappa light chain	40	23364
E	Serum albumin	1075	69321
	Ceruloplasmin	535	122128
	Haptoglobin	380	45177
	Cationic trypsin (bovine)	341	25769
	Apolipoprotein A-1	322	30759
	Immunoglobulin gamma heavy chain	184	49298
	Immunoglobulin lambda light chain	74	22816
	Vitronectin	60	54271
F	Apolipoprotein A-1	289	30759
	Ceruloplasmin	112	122128
	Cationic trypsin (bovine)	28	25769
H	Ceruloplasmin	296	122128
	Apolipoprotein A-1	181	30759
	Apolipoprotein A-2	90	11168
	Cationic trypsin (bovine)	86	25769
	Inter-alpha-trypsin inhibitor heavy chain	49	103293
I	Ceruloplasmin	4524	122128
	Cationic trypsin (bovine)	1072	25769
	Serum albumin	836	69321
	G patch domain-containing protein	29	58182
J	Ceruloplasmin	688	122128
	Cationic trypsin (bovine)	66	25769

APPENDIX

	Apolipoprotein A-1	24	30759
K	Ceruloplasmin	215	122128
	Apolipoprotein A-1	80	30759
	Alpha-1-antichymotrypsin	70	47621
	Cationic trypsin (bovine)	43	25769
	Immunoglobulin lambda light chain	17	22816
L	Ceruloplasmin	732	122128
	Inter-alpha-trypsin inhibitor heavy chain	301	103293
	Apolipoprotein A-1	151	30759
	Cationic trypsin (bovine)	83	25769
	Clusterin	26	52461
M	Ceruloplasmin	3453	122128
	Complement C3	1684	187030
	Complement C4-A	1434	192664
	Alpha-1-antichymotrypsin	859	47621
	Serum paraoxonase/arylesterase	572	39706
	Apolipoprotein A-1	500	30759
	Cationic trypsin (bovine)	398	25769
	Carboxypeptidase N subunit 2	360	60518
	Plasma protease C1 inhibitor	246	55119
	Alpha-1-Acid-Glycoprotein 2	255	23588
	Immunoglobulin kappa constant	192	11758
	Heparin cofactor 2	186	57034
	Inter-alpha-trypsin inhibitor heavy chain	158	103293
	Complement component C9	155	63133
	Transthyretin	151	15877
	Vitronectin	135	54271
	Immunoglobulin lambda constant	133	11287
	Complement C1s subcomponent	116	76635
Alpha-1-Acid-Glycoprotein 1	113	23497	
N	Apolipoprotein A-1	229	30759
	Ceruloplasmin	225	122128
	Cationic trypsin (bovine)	199	25769
	Apolipoprotein A-2	76	11168
P	Ceruloplasmin	1308	122128
	Complement C3	652	187030
	Vitronectin	125	54271
	Apolipoprotein A-2	114	11168
	Alpha-1-antichymotrypsin	101	47621
	Cationic trypsin (bovine)	94	25769
	Apolipoprotein A-1	91	30759
	Prothrombin	90	69992
	Serum albumin	68	69321
	Immunoglobulin gamma heavy chain	67	49298
	Kininogen-1	47	71912
	Clusterin	39	52461
	Q	Ceruloplasmin	381
Cationic trypsin (bovine)		173	25769
Inter-alpha-trypsin inhibitor heavy chain		74	103293
Actin-binding LIM protein		20	77752

APPENDIX

R	Cationic trypsin (bovine)	44	25769
	Ceruloplasmin	39	122128
S	Ceruloplasmin	517	122128
	Cationic trypsin (bovine)	465	25769
T	Ceruloplasmin	519	122128
	Cationic trypsin (bovine)	131	25769
	Apolipoprotein A-1	130	30759
	Inter-alpha-trypsin inhibitor heavy chain	67	103293
	Clusterin	38	52461
	Apolipoprotein A-2	25	11168

The SwissProt data bank was searched using the Mascot MS/MS ion search with the following parameters: enzyme = trypsin; taxonomy = *mammalia*; maximum missed cleavages = 1; no fixed modifications; variable modifications = pyro-glutamine, carbamyl (lysine), carbamyl (N-terminus); peptide tolerance = 0.02 Da; peptide charge = 2+, 3+ and 4+.



## 10.2. NMR Data of Free N-Glycans of Ceruloplasmin

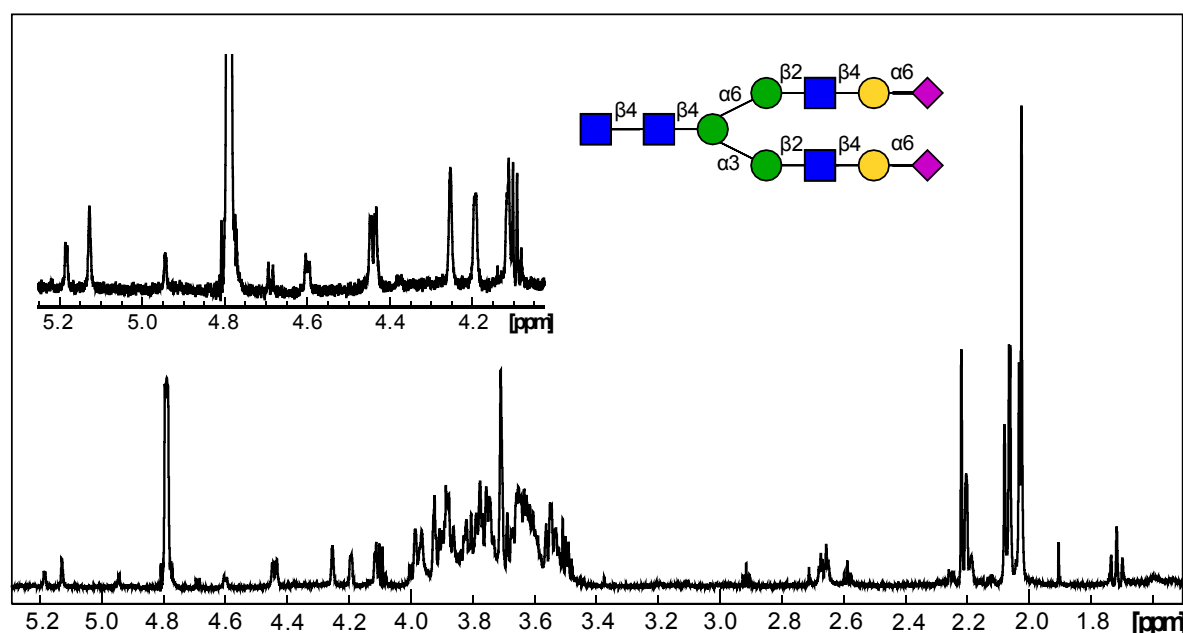


Figure A 1.  $^1\text{H}$ -NMR-Spectrum of the most abundant glycan structure of human ceruloplasmin purified from one plasma sample. The structure is a fully sialylated, diantennary glycan with both sialic acids 2,6-linked.

Table A 2.  $^1\text{H}$ -NMR-Data for the most abundant glycan structure of human ceruloplasmin. See Figure A 1.

Chem. shift [ppm]	Multiplicity	J [Hz]	Glycan moiety	Proton	Linkage
5.184	d	3.1	b-D-GlcNAc	H1 axial	
5.128	s		a-D-Man	H1	3,4,4
4.943	s		a-D-Man	H1	6,4,4
4.600	d	8.0	b-D-GlcNAc	H1	2,6,4,4 and 2,3,4,4
4.440	d	8.2	b-D-Gal	H1	4,2,3,4,4 and 4,2,6,4,4
4.253	s		b-D-Man	H2	4,4
4.193	s		a-D-Man	H2	3,4,4
4.114			a-D-Man	H2	6,4,4
2.674	dd	13.5, 4.5	a-D-NeuAc	H3 equatorial	6,4,2,6,4,4
2.666	dd	13.5, 4.4	a-D-NeuAc	H3 equatorial	6,4,2,6,4,4
2.083	s		b-D-GlcNAc	CH <sub>3</sub> NAc	4
2.080	s		b-D-GlcNAc	CH <sub>3</sub> NAc	4
2.065	s		b-D-GlcNAc	CH <sub>3</sub> NAc	2,3,4,4
2.061	s		b-D-GlcNAc	CH <sub>3</sub> NAc	2,6,4,4
2.033	s		b-D-GlcNAc	CH <sub>3</sub> NAc	
2.024	s		a-D-NeuAc	CH <sub>3</sub> NAc	6,4,2,6,4,4 and 6,4,2,3,4,4

APPENDIX

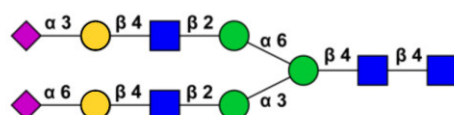


Table A 3. Calculated <sup>1</sup>H-NMR-data for the structure shown above. See Figure 13.

Chem. shift [ppm]	Multiplicity	J [Hz]	Glycan moiety	Proton	Linkage
5.191	d	2.9	b-D-GlcNAc	H1 axial	
5.133	s		a-D-Man	H1	3,4,4
4.440	d	8.0	b-D-Gal	H1	4,2,3,4,4
4.253	s		b-D-Man	H2	4,4
4.195	s		a-D-Man/ b-D-Gal	H2	3,4,4
4.114-4.130	m		a-D-Man	H2/H3	6,4,4/4,2,6,4,4
2.757	dd	13.2, 4.6	a-D-NeuAc	H3 equatorial	3,4,2,6,4,4
2.668	dd	13.0, 4.5	a-D-NeuAc	H3 equatorial	6,4,2,6,4,4
2.088	s		b-D-GlcNAc	CH <sub>3</sub> NAc	4
2.068	s		b-D-GlcNAc	CH <sub>3</sub> NAc	2,3,4,4
2.044	s		b-D-GlcNAc	CH <sub>3</sub> NAc	2,6,4,4
2.031	s		a-D-NeuAc	CH <sub>3</sub> NAc	

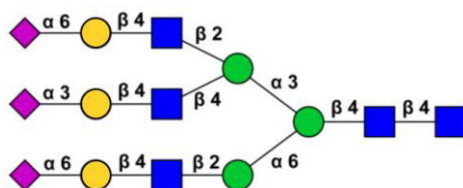


Table A 4. Calculated <sup>1</sup>H-NMR-data for the structure shown above. See Figure 13.

Chem. shift [ppm]	Multiplicity	J [Hz]	Glycan moiety	Proton	Linkage
5.188	d	3.0	b-D-GlcNAc	H1 axial	
5.132	s		a-D-Man	H1	3,4,4
4.933	s		a-D-Man	H1	6,4,4
4.444	d	7.9	b-D-Gal	H1	4,2,3,4,4 and 4,2,6,4,4
4.221	s		b-D-Man/a-D-Man	H2	4,4/3,4,4
4.114	s		a-D-Man	H2	6,4,4
2.760	dd	13.2, 4.6	a-D-NeuAc	H3 equatorial	3,4,4,3,4,4
2.669	dd	13.0, 4.5	a-D-NeuAc	H3 equatorial	6,4,2,6,4,4 and 6,4,2,3,4,4
2.082	s		b-D-GlcNAc	CH <sub>3</sub> NAc	4
2.072	s		b-D-GlcNAc	CH <sub>3</sub> NAc	4,3,4,4
2.067	s		b-D-GlcNAc	CH <sub>3</sub> NAc	2,6,4,4 and 2,3,4,4
2.031	s		a-D-NeuAc	CH <sub>3</sub> NAc	

### 10.3. Matlab Script for the Relative Quantification of Different Proteoforms

Script A 1. Matlab script for calculation of a simulated spectrum as linear combination of all proteoforms for deconvoluted spectra. A weighting coefficient [coeff] is obtained for each proteoform. This coefficient represents the relative amount for each proteoform.[83]

```
% specify experimental input data file
C = dlmread('XX.xy');
yyc = C(:,2);

% define all possible proteoforms with chemical formula [C H N O S]
proteoforms = []

% define resolution and line width
resolution = 0.4
line_width = 0.43;

gauss_min = 0.001;
lw = 1/(2.5*line_width^2);
gauss_width = sqrt(-log(gauss_min)/lw);

MD = zeros(length (proteoforms (:,1)),1,2);

for i = 1:1:length (proteoforms (:,1))
    for k = 1:1:5
        tmp(k) = proteoforms(i,k);
    end
    % calculates the isotopic distribution for all proteoforms
    [MDtmp infotmp xtmp] = isotopicdist([tmp(1) tmp(2) tmp(3) tmp(4)
tmp(5)], 'Resolution', resolution);
    for k = 1:1:size(MDtmp,1)
        for j = 1:1:2
            MD(i,k,j) = MDtmp(k,j);
        end
    end
end
end
all_MD = zeros(length (C),length (proteoforms (:,1)));
for n = length (proteoforms (:,1)):-1:1
    [max_unused xlen] = max(MD(n,:,1));
    for i = 1:1:length (C)
        if ((C(i,1)+gauss_width >= MD(n,1,1)) && ...
            (C(i,1)-gauss_width <= MD(n,xlen,1)))
            for m = 1:1:xlen
                dx = C(i,1) - (MD(n,m,1));
                all_MD(i-1+1,n) = ...
                    all_MD(i-1+1,n) + ...
                    MD(n,m,2) * exp(-lw*dx^2);
            end
        end
    end
end
% finds solution to linear combination (coeff x proteoforms = experimental
spectrum)
[coeff] = lsqlin(all_MD,yyc,[],[],[],[],0,[],[], optimset, 'MaxIter')
```

### 10.4. Relative Amounts of Glycopeptides from Ceruloplasmin

Table A 5. Relative amounts and retention time (C18 RP column) of glycopeptides of ceruloplasmin for glycosite <sup>138</sup>N in % from 17 different plasma samples. Results from a tryptic bottom-up approach are compared to results from glycopeptides with enhanced peptide length (middle-down approach). “-” stands for not detected, however, structures might be present and were detected with a different method.

Glycan moiety	Method	Rt [min]	A	C	D	E	F	H	I	J	K	L	M	N	P	Q	R	S	T	Average
H5N4S2	Bottom up	22.0	63.0	77.6	63.3	58.2	72.2	41.9	57.9	51.5	53.0	55.0	45.2	44.3	28.1	56.0	49.5	50.7	47.5	53.8
H5N4S2	Middle down	26.3	48.7	45.2	49.6	42.2	50.8	36.9	45.1	42.9	49.4	44.8	37.9	41.2	24.3	46.2	38.7	45.6	42.6	43.1
H5N4NS2F	Bottom up	21.9	25.4	22.4	17.4	17.6	27.8	23.5	16.6	21.8	17.2	18.5	24.3	20.6	26.6	12.6	15.3	21.2	14.1	20.2
H5N4NS2F	Middle down	26.2	22.8	12.6	15.5	15.7	17.4	23.6	14.2	18.6	18.1	19.3	23.4	21.3	22.3	12.8	22.7	20.2	13.0	18.4
H5N4S2F2	Bottom up	21.8	-	-	-	-	-	-	0.8	-	-	-	2	-	-	-	-	-	-	0.2
H5N4S2F2	Middle down	26.1	0.9	0.4	0.3	0.6	0.3	1.1	0.1	0.2	0.2	0.3	0.5	0.4	0.2	0.2	0.2	0.9	0.3	0.4
H5N4S1	Middle down	25.5	-	-	-	-	-	0.2	-	-	-	0.7	-	-	0.4	-	-	-	-	0.1
H5N4NS1F	Middle down	25.3	-	-	-	0.1	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0
H6N5S3	Bottom up	22.2	4.7	-	10.0	5.5	-	8.0	16.6	12.1	16.5	12.8	11.4	11.3	24.1	18.1	16.5	7.3	13.3	11.1
H6N5S3	Middle down	27.0	6.7	13.5	14.8	11.7	14.1	8.8	25.4	12.7	15.7	15.3	14.2	13.4	24.3	21.0	16.4	7.9	19.5	15.0
H6N5S3F	Bottom up	22.1	6.8	-	9.2	18.7	-	19.6	6.9	10.9	13.4	13.6	17.1	14.1	20	13.3	18.8	17.9	14.8	12.7
H6N5S3F	Middle down	27.0	14.2	20.2	13.5	20.6	11.5	19.4	10.8	11.8	11.5	14.7	17.6	16.5	19.4	14.4	16.6	17.5	18.5	15.8
H6N5S3F2	Bottom up	22.1	-	-	-	-	-	6.9	1.2	3.7	-	-	-	9.6	1.2	-	-	2.9	10.3	2.1
H6N5S3F2	Middle down	26.9	3.7	3.0	1.9	4.9	1.6	5.9	0.8	1.9	1.6	2.1	4.2	3.4	2.7	1.7	2.9	4.2	2.5	2.9
H6N5S3F3	Middle down	26.9	0.1	-	-	0.2	-	-	-	-	-	-	-	-	-	-	-	0.1	-	0.0
H6N5S2	Middle down	26.2	0.2	0.5	0.6	0.2	0.6	0.4	0.8	5.5	0.8	0.7	0.3	0.6	1.1	0.4	0.6	0.6	0.7	0.9
H6N5S2F	Middle down	26.1	0.6	0.8	0.4	0.6	0.6	0.9	0.3	4.7	0.6	0.7	0.6	0.9	0.9	0.3	0.8	1.3	0.7	0.9
H7N6S4	Middle down	28.0	0.4	1.2	1.5	0.7	1.5	0.6	1.8	0.8	1.2	0.6	0.5	0.8	2.3	1.5	0.8	0.4	1.2	1.0
H7N6S4F	Middle down	27.9	0.9	1.8	1.4	1.5	1.3	1.4	0.7	0.7	0.8	0.6	0.6	1.0	1.7	1.0	0.2	0.9	1.0	1.0
H7N6S4F2	Middle down	27.9	0.6	0.8	0.4	0.8	0.4	0.8	0.1	0.2	0.2	0.1	0.2	0.3	0.3	0.2	-	0.4	0.2	0.3
H7N6S4F3	Middle down	27.9	0.1	0.1	0.1	0.2	0.1	-	-	-	-	-	-	0.1	-	0.3	-	0.1	-	0.1

Table A 6. Relative amounts and retention time (C18 RP column) of glycopeptides of ceruloplasmin for glycosite <sup>358</sup>N in % from 17 different plasma samples. Results from a tryptic bottom-up approach are compared to results from glycopeptides with enhanced peptide length (middle-down approach).

Glycan moiety	Method	Rt [min]	A	C	D	E	F	H	I	J	K	L	M	N	P	Q	R	S	T	Average
H5N4S2	Bottom up	30.3	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
H5N4S2	Middle down	36.2	75.7	59.9	74.4	71.4	71.4	61.9	80.6	74.5	74.0	68.5	71.0	68.3	47.9	69.3	68.6	73.8	70.1	69.5
H5N4NS2F	Middle down	36.1	13.8	13.0	10.1	6.7	9.0	10.9	7.9	10.8	10.0	13.5	10.0	12.7	24.5	5.8	12.8	15.2	12.4	11.7
H6N5S3	Middle down	36.7	4.7	13.3	11.5	12.6	12.6	13.6	9.5	10.5	12.0	11.3	12.4	12.1	19.8	18.4	12.9	6.9	12.3	12.1
H6N5S3F	Middle down	36.7	5.8	13.8	4.1	9.4	7.1	13.7	2.1	4.2	3.9	6.7	6.6	6.9	7.8	6.5	5.8	4.1	5.2	6.7

Table A 7. Relative amounts and retention time (C18 RP column) of glycopeptides of ceruloplasmin for glycosite <sup>397</sup>N in % from 17 different plasma samples. Results from a tryptic bottom-up approach are compared to results from glycopeptides with enhanced peptide length (middle-down approach). “-” stands for not detected, however, structures might be present and were detected with a different method.

Glycan moiety	Method	Rt [min]	A	C	D	E	F	H	I	J	K	L	M	N	P	Q	R	S	T	Average
H5N4S2	Bottom up	33.3	86.1	96.8	82.1	100	100	79.7	86.8	100	100	70.1	88.6	77.1	77.1	84.0	79.0	76.1	84.7	86.4
H5N4S2	Middle down	31.7	73.3	70.8	68.3	65.2	71.1	65.7	72.3	63.6	67.5	66.3	64	62.5	56.3	66.1	69.5	74.6	68.1	67.4
H5N4NS2F	Bottom up	33.2	4.8	3.2	2.6	-	-	3.0	7.3	-	-	1.9	3.2	5.0	3.0	2.5	5.0	3.7	3.1	2.8
H5N4NS2F	Middle down	31.7	9.0	4.0	4.4	4.2	4.7	6.1	4.3	5.5	5.0	6.3	3.7	7.1	8.5	2.0	3.3	6.7	3.3	5.2
H5N4NS2F2	Middle down	31.7	-	-	-	-	-	-	0.1	-	-	-	-	-	-	-	-	-	-	0.0
H6N5S3	Bottom up	34.1	4.5	-	12.6	-	-	10.2	5.9	-	-	21.6	5.8	11.7	12.5	10.5	12.8	9.9	8.4	7.4
H6N5S3	Middle down	32.0	7.5	13.2	18.3	14.6	16.9	12.2	19.9	21.4	20.4	18.4	20.4	18.3	27.2	24.8	18.9	7.4	19.0	17.6
H6N5S3F	Bottom up	34.1	4.6	-	2.7	-	-	7.1	-	-	-	6.4	2.3	6.2	7.4	3.0	3.2	10.4	3.7	3.4
H6N5S3F	Middle down	31.9	10.3	11.9	8.9	16.0	7.3	16.1	3.5	9.5	7.1	9.0	11.9	12.1	7.9	7.1	8.2	11.3	9.6	9.8

Table A 8. Relative amounts and retention time (C18 RP column) of glycopeptides of ceruloplasmin for glycosite <sup>762</sup>N in % from 17 different plasma samples. Results from a tryptic bottom-up approach are compared to results from glycopeptides with enhanced peptide length (middle-down approach). “-“ stands for not detected, however, structures might be present and were detected with a different method.

Glycan moiety	Method	Rt [min]	A	C	D	E	F	H	I	J	K	L	M	N	P	Q	R	S	T	Average
H5N4S2	Bottom up	31.6	63.9	44.6	57.1	47.0	39.4	39.2	45.3	43.6	36.5	45.7	35.9	39.6	20.4	42.5	29.3	43.5	43.1	42.2
H5N4S2	Middle down	31.4	47.9	43.4	40.8	36.8	41.2	30.4	44.5	44.2	46.6	40.0	31.3	34.7	19.2	36.3	32.8	45	41.4	38.6
H5N4NS2F	Bottom up	31.5	19.8	11.0	13.0	15.6	11.9	15.7	9.9	13.2	12.2	12.1	14.9	16.0	13.8	8.2	12.8	12.3	8.7	13.0
H5N4NS2F	Middle down	31.4	16.9	8.8	8.5	10.2	8.7	14.3	9.7	14.4	11.0	12.3	12.9	13.0	12.4	7.3	13.9	14.3	7.2	11.5
H5N4NS2F2	Bottom up	31.5	-	0.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0
H6N5S3	Bottom up	32.1	6.5	19.4	18.2	11.6	31.0	14.8	31.1	24.1	34.8	23.6	23.4	21.8	40.0	32.1	29.3	13.8	26.2	23.6
H6N5S3	Middle down	31.5	10.9	19.6	28.9	19	28.6	17.3	32	21.0	28.3	26.6	25.5	24.7	41.8	36.9	27.7	12.7	28.1	25.3
H6N5S3F	Bottom up	32.0	9.5	22.6	10.7	23.1	17.7	23.9	13.7	17.2	16.4	16.1	22	19.9	23.3	15.7	24	25.5	19.9	18.9
H6N5S3F	Middle down	31.4	20.7	25.9	20.1	29.4	19.8	31.6	12.8	18.5	13.7	19.1	26.8	24.0	24.0	18.3	23.5	24.0	21.4	22.0
H6N5S3F2	Bottom up	31.9	0.3	1.9	1.1	2.7	-	6.4	-	2.0	-	2.5	3.8	2.6	2.5	1.4	4.6	4.8	2.1	2.3
H6N5S3F2	Middle down	31.4	3.6	2.4	1.7	4.6	1.8	6.4	1.0	1.9	0.4	1.9	3.5	3.6	2.5	1.2	2.0	4.1	1.9	2.6

## 10.5. Alpha-Acid-Glycoprotein

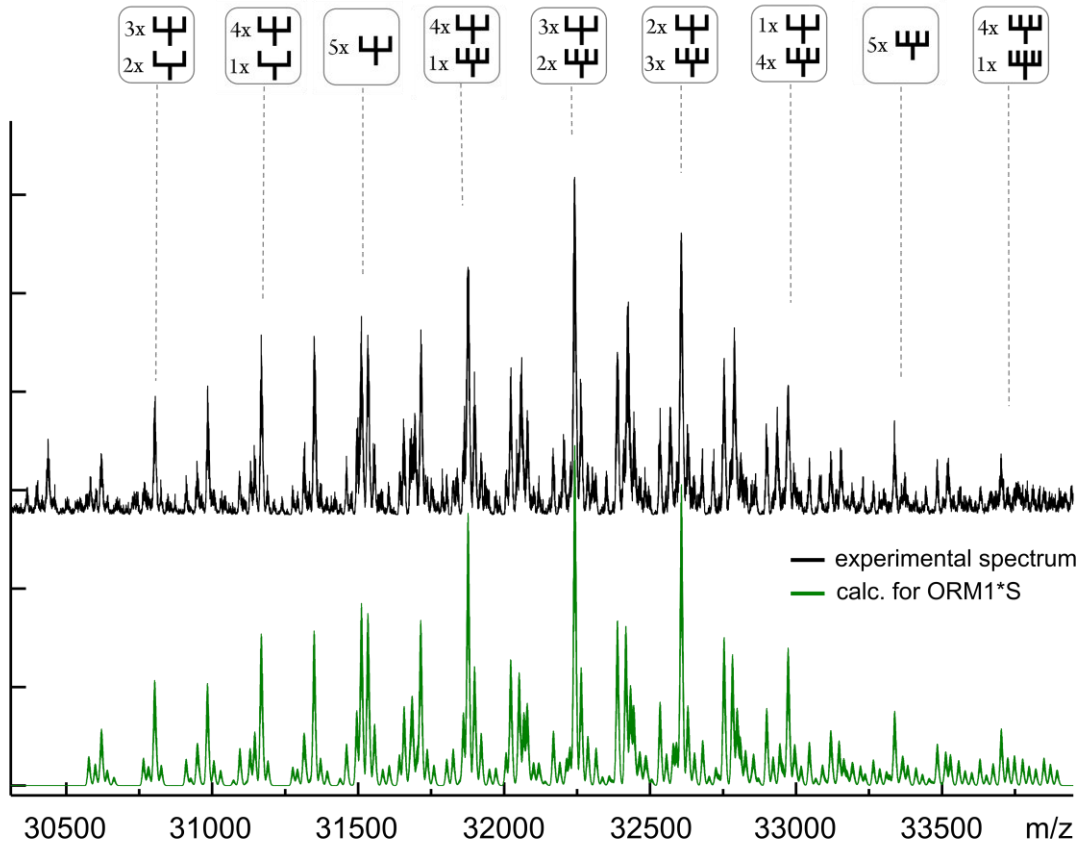


Figure A 2. Deconvoluted ESI-MS spectrum of de-sialylated AGP for variant ORM1\*S from pooled human plasma in black. Calculated spectrum in green. Up to two sodium adducts were considered for calculation. Variant ORM1\*S shows a slightly different glycosylation pattern than variant ORM1\*F1. Additional signals corresponding to a loss of galactose residues equivalent to bisected structures are found.[142]



APPENDIX

Table A 9. Relative amounts of glycan compositions over all five glycosylation sites of desialylated AGP, variant ORM1\*F1. Results from one sample of pooled plasma, eight samples from healthy volunteers and their average values as well as results from one sample of a patient diagnosed with hepatocellular carcinoma (HCC08) are shown. Values are given in %.

Glycan composition	Pooled plasma	AD	L	R	V	Q	T	AA	O	Average	HCC08
H29N24	2.6	2.9	5.1	4.1	1.7	3.6	2.9	2.9	2.4	3.1	-
H29N24F1	1.5	-	0.8	0.7	2.1	1.3	1.7	1.4	1.6	1.2	-
H29N24F2	-	-	-	-	0.5	0.4	1.5	-	1.5	0.4	-
H29N24F3	-	-	-	-	-	-	1.1	-	0.9	0.5	-
H30N25	7.3	8.3	9.1	9.4	5.0	8.9	5.8	6.8	4.9	7.3	0.7
H30N25F	3.7	0.1	1.9	1.8	4.7	4.0	3.1	3.7	4.0	3.0	0.5
H30N25F2	1.2	-	0.1	0.1	1.4	1.1	1.5	1.9	2.3	1.1	0.3
H30N25F3	0.4	-	-	-	-	-	-	-	1.4	0.2	0.1
H30N25F4	0.3	-	-	-	-	-	-	-	0.4	0.1	1.1
H31N26	10.8	16.1	16.0	15.8	7.5	14.5	9.1	9.8	7.3	11.9	1.2
H31N26F	5.5	1.0	4.6	4.2	8.7	5.6	4.7	5.2	5.3	5.0	1.4
H31N26F2	2.5	-	0.4	0.2	2.7	1.9	2.2	1.6	3.2	1.6	2.5
H31N26F3	1.1	0.3	0.2	0.4	0.3	-	-	-	2.5	0.5	2.2
H31N26F4	0.3	-	-	-	0.2	-	-	-	1.4	0.2	3.6
H32N27	14.5	23.6	19.5	16.7	10	17.2	12.8	10.0	9.3	14.9	6.3
H32N27F	7.7	2.0	6.3	4.9	10.7	6.4	6.7	5.7	6.5	6.3	5.1
H32N27F2	3.3	0.2	0.7	0.4	4.3	2.6	2.9	5.8	4.8	2.8	5.0
H32N27F3	1.9	0.2	0.1	0.2	0.6	-	2.4	3.4	2.8	1.3	5.0
H32N27F4	1.1	-	-	0.1	0.1	-	1.9	1.4	1.4	0.7	4.1
H33N28	11.1	23.5	14.9	14.1	7.9	11.4	10.3	10.3	7.0	12.3	5.2
H33N28F	6.3	3.9	6.4	5.1	9.7	3.6	6.1	4.8	5.3	5.7	6.0
H33N28F2	2.4	-	0.3	0.3	4.2	1.8	2.6	3.1	3.3	2.0	6.5
H33N28F3	1.1	-	-	-	0.5	-	1.7	1.9	2.1	0.8	5.2
H33N28F4	0.6	-	-	-	-	-	1.8	1.6	1.3	0.6	4.8
H34N29	4.5	7.3	5.6	6.7	3.3	4.7	4.9	5.1	3.3	5.1	3.9
H34N29F	1.7	1.6	1.5	1.8	3.1	1.1	2.0	1.0	1.5	1.7	3.2
H34N29F2	0.3	-	0.1	0.1	1.6	0.6	1.5	-	1.3	0.6	3.0
H34N29F3	-	-	-	-	0.4	-	-	-	0.9	0.1	1.9
H34N29F4	-	-	-	-	-	-	-	-	0.4	-	4.9
H35N30	2.6	3.2	2.2	4	1.7	3.5	4.2	6.6	2.4	3.4	3.6
H35N30F	1.3	0.4	0.4	0.9	1.3	0.3	0.9	1.5	0.9	0.9	1.6
H35N30F2	0.5	0.1	0.1	0.1	0.7	0.1	0.4	-	0.4	0.3	1.0
H35N30F3	-	-	-	-	-	-	-	-	0.4	-	1.1
H35N30F4	-	-	-	-	-	-	-	-	0.5	0.1	1.6
H36N31	1.1	1.7	1.4	2.8	1.1	0.9	2.0	3.1	1.1	1.7	1.3
H36N31F	0.3	0.4	0.4	1	0.8	-	0.7	0.9	0.6	0.6	1.1
H36N31F2	-	0.1	0.2	0.3	0.4	-	0.6	0.3	0.3	0.2	0.9
H36N31F3	-	-	-	-	-	-	-	-	-	-	1.2
H37N32	0.5	0.3	0.3	1.3	0.2	0.6	-	-	-	0.4	1.1

APPENDIX

Table A 10. Relative amounts of glycan compositions over all five glycosylation sites of desialylated AGP, variant ORM1\*S. Results from one sample of pooled plasma, eight samples from healthy volunteers and their average values as well as results from one sample of a patients diagnosed with hepatocellular carcinoma (HCC08) and one patient diagnosed with liver cirrhosis (LC11) are shown. Values are given in %.

Glycan composition	Pooled plasma	AD	L	N	P	R	Z	T	AA	Average	LC 11	HCC 08
H29N24	1.5	2.6	3.5	6.4	0.1	4	3.9	2.2	2.7	3.0	0.3	-
H29N24F1	0.5	-	0.8	0.3	0.2	0.8	1.0	1.0	1.2	0.6	1.2	-
H29N24F2	0.2	-	-	-	-	-	0.2	0.7	0.2	0.1	1.2	-
H29N24F3	-	-	-	-	-	-	-	0.6	-	0.3	0.4	-
H30N25	5.2	7.8	11.7	12.2	1.5	14.1	8.0	5.9	5.6	8.0	1.0	0.7
H30N25F	2.3	0.2	2.7	1.6	0.1	2.9	2.6	3.5	4.3	2.2	1.8	0.6
H30N25F2	0.8	-	0.1	-	-	-	0.7	1.1	2.1	0.5	1.6	0.2
H30N25F3	0.5	-	-	-	-	-	0.8	1.1	1.3	0.4	1.3	0.1
H30N25F4	0.1	-	-	-	-	-	-	-	-	-	0.9	1.3
H31N26	7.8	13.5	16.8	16.5	6.8	17.2	12.3	10.7	10.1	12.4	0.8	1.3
H31N26F	3.7	0.5	4.2	1.7	0.5	3.5	3.1	5.4	5.6	3.1	2.0	1.6
H31N26F2	1.6	-	-	-	-	-	0.8	1.6	2.9	0.8	3.2	2.2
H31N26F3	1.2	-	-	-	-	-	0.5	-	2.1	0.4	3.3	2.5
H31N26F4	0.6	-	-	-	-	-	-	-	-	0.1	1.8	3.4
H32N27	9.9	27.5	21.6	20.9	15.3	20	14.6	14.5	13.4	17.5	2.3	5.9
H32N27F	4.8	1.9	5.2	2.3	2.0	2.9	4.1	7.7	7.0	4.2	3.3	6.9
H32N27F2	2.4	-	-	-	-	-	1.4	3.4	3.0	1.1	5.0	5.2
H32N27F3	1.4	-	-	-	-	-	1.2	2.7	2.6	0.9	6.0	4.3
H32N27F4	0.9	-	-	-	-	-	-	-	-	0.1	3.5	5.1
H33N28	8.7	26.8	15.6	19.5	28.7	14.5	12.5	12.7	11.7	16.7	3.4	5.4
H33N28F	4.4	2.5	3.8	2.4	3.7	2.1	3.9	6.2	5.5	3.8	3.5	7.2
H33N28F2	2.3	0.1	0.1	0.1	0.2	0.1	1.5	2.8	3.0	1.1	3.8	7.0
H33N28F3	1.3	-	-	-	0.2	-	1.2	1.3	1.9	0.7	5.6	6.0
H33N28F4	0.6	-	-	-	-	-	1.1	-	-	0.2	6.5	5.0
H34N29	3.9	7.8	4.8	6.2	15.6	5.6	6.2	5.6	5.2	6.8	5.5	4.3
H34N29F	1.7	1.4	0.8	0.7	1.5	1.4	1.7	1.9	2.2	1.5	3.9	3.8
H34N29F2	0.8	-	-	-	-	-	0.8	1.2	1.1	0.4	2.4	2.4
H34N29F3	0.6	-	-	-	-	-	0.7	0.5	0.3	0.2	1.5	1.3
H34N29F4	-	-	-	-	-	-	0.6	-	-	0.1	5.4	4.9
H35N30	2.2	3.1	2.3	2.7	7.8	3.2	3.4	2.5	1.9	3.2	4.5	2.6
H35N30F	1.3	0.7	0.3	0.1	0.7	0.9	0.7	0.5	-	0.6	2.3	1.4
H35N30F2	0.8	-	-	-	-	0.1	0.4	0.1	-	0.2	1.2	1.1
H35N30F3	-	-	-	-	-	-	0.1	-	-	-	1.3	1.0
H35N30F4	-	-	-	-	-	-	-	-	-	-	1.4	1.8
H36N31	1.7	1.6	1.3	2.0	5.6	1.9	2.0	1.7	2.2	2.2	1.6	1.0
H36N31F	0.8	0.4	0.1	0.2	0.5	0.5	0.7	0.5	0.5	0.5	1.3	0.7
H36N31F2	-	0.3	0.1	-	0.2	0.1	0.5	0.3	0.4	0.2	1.0	0.5
H36N31F3	-	-	-	-	0.1	-	0.4	-	-	0.1	0.9	0.9
H37N32	1.5	0.7	0.2	0.7	3.2	1.9	1.1	-	-	1.0	1.0	0.7
H37N32F2	-	-	-	-	-	-	0.3	-	-	-	-	-

APPENDIX

Table A 11. Relative amounts of glycan compositions over all five glycosylation sites of desialylated AGP, variant ORM1\*F2. Results for two plasma samples from healthy donors are shown. Values are given in %.

Glycan composition	N	Q
H29N24	4.8	3.2
H29N24F1	0.4	0.6
H29N24F2	-	0.1
H29N24F3	-	-
H30N25	8.8	6.8
H30N25F	0.8	3.6
H30N25F2	0.1	0.8
H30N25F3	-	-
H30N25F4	-	-
H31N26	13.9	12.6
H31N26F	1.7	5.4
H31N26F2	0.1	1.9
H31N26F3	-	1.6
H31N26F4	-	0.9
H32N27	17.9	15.3
H32N27F	2.7	6.7
H32N27F2	0.1	2.7
H32N27F3	-	1.3
H32N27F4	-	0.6
H33N28	20.9	11.6
H33N28F	4.4	5.0
H33N28F2	0.1	2.0
H33N28F3	-	0.8
H33N28F4	-	0.7
H34N29	8.3	3.9
H34N29F	1.6	2.2
H34N29F2	0.1	1.1
H34N29F3	-	0.3
H34N29F4	-	-
H35N30	4.5	2.5
H35N30F	0.6	0.5
H35N30F2	0.1	0.1
H35N30F3	-	-
H35N30F4	-	-
H36N31	3.4	1.7
H36N31F	0.7	0.5
H36N31F2	0.5	0.3
H36N31F3	-	-
H37N32	1.1	-
H37N32F2	-	-

APPENDIX

Table A 12. Relative amounts of glycan compositions over all five glycosylation sites of desialylated AGP, variant ORM2. Results from eleven samples from healthy volunteers and their average values as well as results from one sample of a patients diagnosed with hepatocellular carcinoma (HCC08) and one patient diagnosed with liver cirrhosis (LC11) are shown. Values are given in %.

Glycan composition	AD	L	N	P	R	V	Z	Q	T	AA	O	LC 11	HCC 08
H29N24	1.4	3.7	5.3	-	4.6	1.0	2.8	3.9	2.0	2.2	2.0	0.2	-
H29N24F1	-	0.8	0.2	-	0.7	1.7	0.6	1.9	0.9	0.9	1.7	0.3	-
H29N24F2	-	-	-	-	-	-	0.2	0.5	0.5	0.8	1.0	0.5	0.1
H29N24F3	-	-	-	-	-	-	0.2	-	0.4	0.7	0.8	0.4	0.7
H30N25	7.7	9.9	11.4	0.7	13.7	3.6	7.8	9.4	5.6	6.5	4.1	1.0	0.3
H30N25F	-	1.8	0.6	-	1.8	3.0	1.6	3.3	2.5	2.9	3.1	0.6	0.1
H30N25F2	-	-	-	-	-	0.3	0.4	1.1	1.2	1.6	1.7	0.4	-
H30N25F3	-	-	-	-	-	-	0.3	-	0.6	1.0	1.3	1.2	0.9
H30N25F4	-	-	-	-	-	-	-	-	-	-	0.9	0.2	0.3
H31N26	13.3	17.5	17.7	5.3	19.1	5.5	11.0	12.5	9.3	10.3	6.1	0.8	0.6
H31N26F	0.4	3.7	1.4	0.1	2.5	10.9	2.7	5.6	4.6	5.0	4.9	2.2	1.8
H31N26F2	-	0.1	-	-	-	2.6	1.3	2.1	2.7	3.1	3.3	2.9	1.9
H31N26F3	-	-	-	-	-	0.2	1.2	2.5	2.3	2.2	2.7	3.5	2.5
H31N26F4	-	-	-	-	-	0.3	-	1.7	-	-	1.6	2.0	4.4
H32N27	30.2	24.7	25.8	15.7	21.7	13	16.3	15.7	14.1	13.4	9.3	3.0	4.7
H32N27F	2.1	6.3	2.5	0.9	3.1	14	4.6	6.3	7.2	7.2	7.2	5.5	5.9
H32N27F2	0.3	0.2	0.1	0.4	0.1	4.6	2.4	3.1	4.3	3.7	4.7	6.4	6.3
H32N27F3	-	-	-	0.2	-	0.7	1.9	2.1	2.7	3.0	3.3	7.5	5.8
H32N27F4	-	-	-	-	-	0.2	-	1.3	2.7	-	2.1	3.2	5.4
H33N28	27.3	16.8	20.8	35.6	15.4	10.3	14.4	9.8	11.4	9.9	7.4	3.3	7.1
H33N28F	3.9	4.0	2.2	4	3.2	9.3	4.3	4.6	6.0	5.7	5.5	4.5	8.3
H33N28F2	-	-	-	0.1	-	3.8	2.6	2.0	3.0	3.6	3.8	6.3	8.3
H33N28F3	-	-	-	-	-	0.4	1.5	1.2	2.0	2.7	2.7	6.5	6.0
H33N28F4	-	-	-	-	-	-	-	1.1	1.8	1.8	1.8	7.8	4.9
H34N29	7.0	5.3	6.1	11.1	7.2	2.5	6.6	3.5	5.0	4.8	3.4	5.7	3.7
H34N29F	1.5	0.9	0.5	0.8	0.6	2.5	2.1	1.2	2.1	2.4	2.0	2.9	2.4
H34N29F2	-	-	-	-	0.1	1.3	0.6	0.5	1.0	0.6	1.2	0.7	1.2
H34N29F3	-	-	-	-	-	0.2	0.2	0.2	0.1	0.2	0.8	1.3	1.3
H34N29F4	-	-	-	-	-	0.4	-	-	-	-	1.1	4.8	4.3
H35N30	2.3	1.7	2.4	8.4	3.4	2.1	3.8	1.6	1.9	2.0	2.2	2.5	2.2
H35N30F	0.1	-	-	0.4	0.4	1.8	0.7	0.4	0.6	1.0	1.0	1.9	1.3
H35N30F2	0.1	-	-	-	0.1	0.9	0.5	0.3	0.5	0.5	0.8	1.2	0.9
H35N30F3	-	-	-	-	-	0.3	0.5	-	0.4	-	0.7	1.1	1.0
H35N30F4	-	-	-	-	-	-	-	-	-	-	0.7	1.1	0.7
H36N31	0.5	0.1	0.8	5.7	1.3	1.2	1.6	0.8	0.5	0.5	1.0	1.2	0.5
H36N31F	-	-	0.1	-	0.1	0.6	1.3	-	-	-	0.5	1.9	1.0
H36N31F2	-	-	-	-	-	0.7	0.6	-	-	-	-	1.3	1.2
H36N31F3	-	-	-	-	-	-	0.2	-	-	-	-	1.3	1.1
H37N32	0.4	0.3	0.5	2.6	-	-	1.3	-	-	-	-	0.3	0.5
H37N32F2	-	-	-	0.4	-	-	-	-	-	-	-	-	-

APPENDIX

Table A 13. Complete list of all found sialylated glycan compositions for variant ORM1\*F1 for pooled human plasma.[142]

<b>Glycan Composition</b>	<b>Sialic acids</b>	<b>Rel. abundance [%]</b>
H <sub>29</sub> N <sub>24</sub>	S <sub>12</sub>	0.87
	S <sub>13</sub>	0.88
	S <sub>14</sub>	0.94
H <sub>29</sub> N <sub>24</sub> F <sub>1</sub>	S <sub>12</sub>	0.85
	S <sub>13</sub>	0.36
	S <sub>14</sub>	0.35
H <sub>30</sub> N <sub>25</sub>	S <sub>12</sub>	0.67
	S <sub>13</sub>	0.66
	S <sub>14</sub>	1.93
	S <sub>15</sub>	4.11
H <sub>30</sub> N <sub>25</sub> F	S <sub>12</sub>	0.25
	S <sub>13</sub>	0.55
	S <sub>14</sub>	0.99
	S <sub>15</sub>	1.97
H <sub>30</sub> N <sub>25</sub> F <sub>2</sub>	S <sub>15</sub>	1.22
H <sub>30</sub> N <sub>25</sub> F <sub>3</sub>	S <sub>15</sub>	0.45
H <sub>30</sub> N <sub>25</sub> F <sub>4</sub>	S <sub>15</sub>	0.32
H <sub>31</sub> N <sub>26</sub>	S <sub>12</sub>	0.11
	S <sub>13</sub>	0.62
	S <sub>14</sub>	2.16
	S <sub>15</sub>	4.25
	S <sub>16</sub>	3.71
	S <sub>12</sub>	0.21
H <sub>31</sub> N <sub>26</sub> F	S <sub>13</sub>	0.38
	S <sub>14</sub>	1.04
	S <sub>15</sub>	1.90
	S <sub>16</sub>	1.99
	S <sub>15</sub>	0.82
H <sub>31</sub> N <sub>26</sub> F <sub>2</sub>	S <sub>16</sub>	1.70
	S <sub>16</sub>	1.06
H <sub>31</sub> N <sub>26</sub> F <sub>3</sub>	S <sub>16</sub>	0.31
H <sub>32</sub> N <sub>27</sub>	S <sub>13</sub>	0.60
	S <sub>14</sub>	1.92
	S <sub>15</sub>	3.89
	S <sub>16</sub>	5.95
	S <sub>17</sub>	2.21
H <sub>32</sub> N <sub>27</sub> F	S <sub>13</sub>	0.27
	S <sub>14</sub>	0.74
	S <sub>15</sub>	2.60
	S <sub>16</sub>	3.10
	S <sub>17</sub>	1.05
H <sub>32</sub> N <sub>27</sub> F <sub>2</sub>	S <sub>16</sub>	2.13
	S <sub>17</sub>	1.19
H <sub>32</sub> N <sub>27</sub> F <sub>3</sub>	S <sub>16</sub>	1.47
	S <sub>17</sub>	0.44
H <sub>32</sub> N <sub>27</sub> F <sub>4</sub>	S <sub>16</sub>	0.75
	S <sub>17</sub>	0.35
H <sub>33</sub> N <sub>28</sub>	S <sub>14</sub>	1.18
	S <sub>15</sub>	2.31
	S <sub>16</sub>	4.10
	S <sub>17</sub>	3.03
	S <sub>18</sub>	0.62

## Table continued

H <sub>33</sub> N <sub>28</sub> F	S <sub>14</sub>	0.06
	S <sub>15</sub>	1.63
	S <sub>16</sub>	2.31
	S <sub>17</sub>	1.78
	S <sub>18</sub>	0.55
H <sub>33</sub> N <sub>28</sub> F <sub>2</sub>	S <sub>16</sub>	0.75
	S <sub>17</sub>	1.43
	S <sub>18</sub>	0.20
H <sub>33</sub> N <sub>28</sub> F <sub>3</sub>	S <sub>16</sub>	0.16
	S <sub>17</sub>	0.93
H <sub>33</sub> N <sub>28</sub> F <sub>4</sub>	S <sub>17</sub>	0.43
	S <sub>18</sub>	0.14
H <sub>34</sub> N <sub>29</sub>	S <sub>15</sub>	0.50
	S <sub>16</sub>	1.39
	S <sub>17</sub>	1.71
	S <sub>18</sub>	0.51
	S <sub>19</sub>	0.47
H <sub>34</sub> N <sub>29</sub> F	S <sub>16</sub>	0.63
	S <sub>17</sub>	0.80
	S <sub>18</sub>	0.26
H <sub>34</sub> N <sub>29</sub> F <sub>2</sub>	S <sub>17</sub>	0.07
	S <sub>18</sub>	0.07
	S <sub>19</sub>	0.12
H <sub>35</sub> N <sub>30</sub>	S <sub>16</sub>	0.50
	S <sub>17</sub>	0.81
	S <sub>18</sub>	0.66
	S <sub>19</sub>	0.64
H <sub>35</sub> N <sub>30</sub> F	S <sub>16</sub>	0.26
	S <sub>17</sub>	0.14
	S <sub>18</sub>	0.34
	S <sub>19</sub>	0.54
H <sub>35</sub> N <sub>30</sub> F <sub>2</sub>	S <sub>16</sub>	0.11
	S <sub>17</sub>	0.12
	S <sub>18</sub>	0.11
	S <sub>19</sub>	0.22
H <sub>36</sub> N <sub>31</sub>	S <sub>16</sub>	0.11
	S <sub>17</sub>	0.40
	S <sub>18</sub>	0.61

Table A 14. Complete list of all found sialylated glycan compositions for variant ORM1\*S for pooled human plasma.[142]

<b>Glycan Composition</b>	<b>Sialic acids</b>	<b>Rel. abundance [%]</b>
H <sub>29</sub> N <sub>24</sub>	S <sub>12</sub>	0.15
	S <sub>13</sub>	1.13
	S <sub>14</sub>	0.68
H <sub>29</sub> N <sub>24</sub> F	S <sub>13</sub>	0.64
H <sub>29</sub> N <sub>24</sub> F <sub>2</sub>	S <sub>14</sub>	0.27
	S <sub>12</sub>	0.18
H <sub>30</sub> N <sub>25</sub>	S <sub>13</sub>	0.70
	S <sub>14</sub>	2.06
	S <sub>15</sub>	3.75
	S <sub>12</sub>	0.21
H <sub>30</sub> N <sub>25</sub> F	S <sub>13</sub>	0.31
	S <sub>14</sub>	0.67
	S <sub>15</sub>	1.79
H <sub>30</sub> N <sub>25</sub> F <sub>2</sub>	S <sub>13</sub>	0.11
	S <sub>14</sub>	0.21
	S <sub>15</sub>	0.66
H <sub>30</sub> N <sub>25</sub> F <sub>3</sub>	S <sub>15</sub>	0.62
H <sub>30</sub> N <sub>25</sub> F <sub>4</sub>	S <sub>15</sub>	0.21
	S <sub>12</sub>	0.45
	S <sub>13</sub>	0.88
	S <sub>14</sub>	1.10
	S <sub>15</sub>	3.43
	S <sub>16</sub>	4.31
H <sub>31</sub> N <sub>26</sub> F	S <sub>13</sub>	0.05
	S <sub>14</sub>	0.11
	S <sub>15</sub>	1.43
	S <sub>16</sub>	3.14
H <sub>31</sub> N <sub>26</sub> F <sub>2</sub>	S <sub>15</sub>	0.53
	S <sub>16</sub>	1.52
H <sub>31</sub> N <sub>26</sub> F <sub>3</sub>	S <sub>16</sub>	1.41
H <sub>31</sub> N <sub>26</sub> F <sub>4</sub>	S <sub>16</sub>	0.70
	S <sub>14</sub>	0.56
	S <sub>15</sub>	3.32
	S <sub>16</sub>	5.52
H <sub>32</sub> N <sub>27</sub>	S <sub>16</sub>	5.52
	S <sub>17</sub>	3.32
	S <sub>14</sub>	0.36
H <sub>32</sub> N <sub>27</sub> F	S <sub>15</sub>	1.87
	S <sub>16</sub>	3.05
	S <sub>17</sub>	0.94
	S <sub>14</sub>	0.56
H <sub>32</sub> N <sub>27</sub> F <sub>2</sub>	S <sub>15</sub>	1.23
	S <sub>16</sub>	1.08
	S <sub>17</sub>	0.26
	S <sub>14</sub>	0.26
H <sub>32</sub> N <sub>27</sub> F <sub>3</sub>	S <sub>15</sub>	0.53
	S <sub>16</sub>	0.92
	S <sub>14</sub>	0.18
H <sub>32</sub> N <sub>27</sub> F <sub>4</sub>	S <sub>15</sub>	0.21
	S <sub>16</sub>	0.80
	S <sub>17</sub>	0.04
	S <sub>15</sub>	2.20
H <sub>33</sub> N <sub>28</sub>	S <sub>16</sub>	4.79
	S <sub>17</sub>	4.30

Table continued

H <sub>33</sub> N <sub>28</sub> F	S <sub>15</sub>	1.14
	S <sub>16</sub>	1.82
	S <sub>17</sub>	2.58
	S <sub>18</sub>	0.07
H <sub>33</sub> N <sub>28</sub> F <sub>2</sub>	S <sub>16</sub>	1.00
	S <sub>17</sub>	1.92
H <sub>33</sub> N <sub>28</sub> F <sub>3</sub>	S <sub>15</sub>	0.18
	S <sub>16</sub>	0.21
	S <sub>17</sub>	0.92
H <sub>33</sub> N <sub>28</sub> F <sub>4</sub>	S <sub>18</sub>	0.33
	S <sub>16</sub>	0.31
	S <sub>17</sub>	0.35
H <sub>34</sub> N <sub>29</sub>	S <sub>18</sub>	0.07
	S <sub>15</sub>	0.60
	S <sub>16</sub>	2.45
	S <sub>17</sub>	1.56
H <sub>34</sub> N <sub>29</sub> F	S <sub>18</sub>	0.26
	S <sub>19</sub>	0.11
	S <sub>16</sub>	1.25
H <sub>34</sub> N <sub>29</sub> F <sub>2</sub>	S <sub>17</sub>	0.28
	S <sub>18</sub>	0.40
	S <sub>19</sub>	0.22
H <sub>34</sub> N <sub>29</sub> F <sub>3</sub>	S <sub>16</sub>	0.09
	S <sub>17</sub>	0.44
	S <sub>18</sub>	0.37
	S <sub>19</sub>	0.09
H <sub>35</sub> N <sub>30</sub>	S <sub>15</sub>	0.06
	S <sub>16</sub>	0.24
	S <sub>17</sub>	0.04
H <sub>35</sub> N <sub>30</sub> F	S <sub>18</sub>	0.35
	S <sub>16</sub>	1.41
	S <sub>17</sub>	0.85
	S <sub>18</sub>	0.57
H <sub>35</sub> N <sub>30</sub> F <sub>2</sub>	S <sub>15</sub>	0.26
	S <sub>16</sub>	0.84
	S <sub>17</sub>	0.35
H <sub>36</sub> N <sub>31</sub>	S <sub>18</sub>	0.14
	S <sub>16</sub>	0.04
	S <sub>17</sub>	0.07
H <sub>36</sub> N <sub>31</sub> F	S <sub>18</sub>	0.92
	S <sub>19</sub>	1.41
H <sub>37</sub> N <sub>32</sub>	S <sub>17</sub>	0.76
	S <sub>18</sub>	0.53
	S <sub>19</sub>	0.53
H <sub>37</sub> N <sub>32</sub>	S <sub>16</sub>	0.25
	S <sub>17</sub>	0.78
	S <sub>18</sub>	0.43
	S <sub>19</sub>	0.43



Table A 15. Complete list of all found sialylated glycan compositions for variant ORM2 for pooled human plasma.[142]

<b>Glycan Composition</b>	<b>Sialic acids</b>	<b>Rel. abundance [%]</b>
H <sub>30</sub> N <sub>25</sub>	S <sub>12</sub>	0.03
	S <sub>13</sub>	0.96
	S <sub>14</sub>	1.89
	S <sub>15</sub>	3.27
H <sub>30</sub> N <sub>25</sub> F	S <sub>12</sub>	0.37
	S <sub>13</sub>	0.52
	S <sub>14</sub>	0.34
	S <sub>15</sub>	0.81
H <sub>31</sub> N <sub>26</sub>	S <sub>12</sub>	1.06
	S <sub>13</sub>	1.18
	S <sub>14</sub>	2.93
	S <sub>15</sub>	4.21
	S <sub>16</sub>	5.06
H <sub>31</sub> N <sub>26</sub> F	S <sub>14</sub>	1.46
	S <sub>15</sub>	2.57
	S <sub>16</sub>	2.01
H <sub>31</sub> N <sub>26</sub> F <sub>2</sub>	S <sub>16</sub>	1.80
H <sub>32</sub> N <sub>27</sub>	S <sub>12</sub>	1.42
	S <sub>13</sub>	2.16
	S <sub>14</sub>	2.76
	S <sub>15</sub>	3.33
	S <sub>16</sub>	4.89
	S <sub>17</sub>	4.78
H <sub>32</sub> N <sub>27</sub> F	S <sub>15</sub>	2.55
	S <sub>16</sub>	1.47
	S <sub>17</sub>	1.13
H <sub>32</sub> N <sub>27</sub> F <sub>2</sub>	S <sub>17</sub>	2.29
H <sub>32</sub> N <sub>27</sub> F <sub>3</sub>	S <sub>15</sub>	1.75
H <sub>32</sub> N <sub>27</sub> F <sub>4</sub>	S <sub>17</sub>	0.88
H <sub>33</sub> N <sub>28</sub>	S <sub>14</sub>	1.15
	S <sub>15</sub>	3.19
	S <sub>16</sub>	3.03
	S <sub>17</sub>	4.17
	S <sub>18</sub>	2.47
H <sub>33</sub> N <sub>28</sub> F	S <sub>15</sub>	1.56
	S <sub>16</sub>	2.34
	S <sub>17</sub>	1.10
H <sub>33</sub> N <sub>28</sub> F <sub>2</sub>	S <sub>18</sub>	1.33
H <sub>33</sub> N <sub>28</sub> F <sub>3</sub>	S <sub>16</sub>	0.92
H <sub>34</sub> N <sub>29</sub>	S <sub>15</sub>	1.44
	S <sub>16</sub>	1.69
	S <sub>17</sub>	0.56
	S <sub>18</sub>	0.19
H <sub>34</sub> N <sub>29</sub> F	S <sub>15</sub>	0.13
	S <sub>16</sub>	0.39
	S <sub>17</sub>	0.37
	S <sub>18</sub>	0.51
H <sub>34</sub> N <sub>29</sub> F <sub>2</sub>	S <sub>17</sub>	0.87
H <sub>34</sub> N <sub>29</sub> F <sub>3</sub>	S <sub>17</sub>	0.62
H <sub>35</sub> N <sub>30</sub>	S <sub>16</sub>	1.48
	S <sub>17</sub>	1.39

## APPENDIX

---

H <sub>35</sub> N <sub>30</sub> F	S <sub>17</sub>	0.09
	S <sub>18</sub>	0.41
	S <sub>19</sub>	0.93
H <sub>35</sub> N <sub>30</sub> F <sub>2</sub>	S <sub>16</sub>	0.30
	S <sub>17</sub>	0.62
H <sub>35</sub> N <sub>30</sub> F <sub>3</sub>	S <sub>16</sub>	0.55
	S <sub>17</sub>	0.08
H <sub>36</sub> N <sub>31</sub>	S <sub>16</sub>	0.27
	S <sub>17</sub>	0.92
	S <sub>18</sub>	0.13
H <sub>36</sub> N <sub>31</sub> F	S <sub>16</sub>	1.00
H <sub>37</sub> N <sub>32</sub> F	S <sub>15</sub>	0.72
	S <sub>16</sub>	0.83

### 10.6. $\beta$ -2-Glycoprotein 1

$\beta$ -2-Glycoprotein 1 was analyzed as intact glycoprotein by LC-ESI-MS. Quantification of the different proteoforms was based on the deconvoluted spectra shown below.

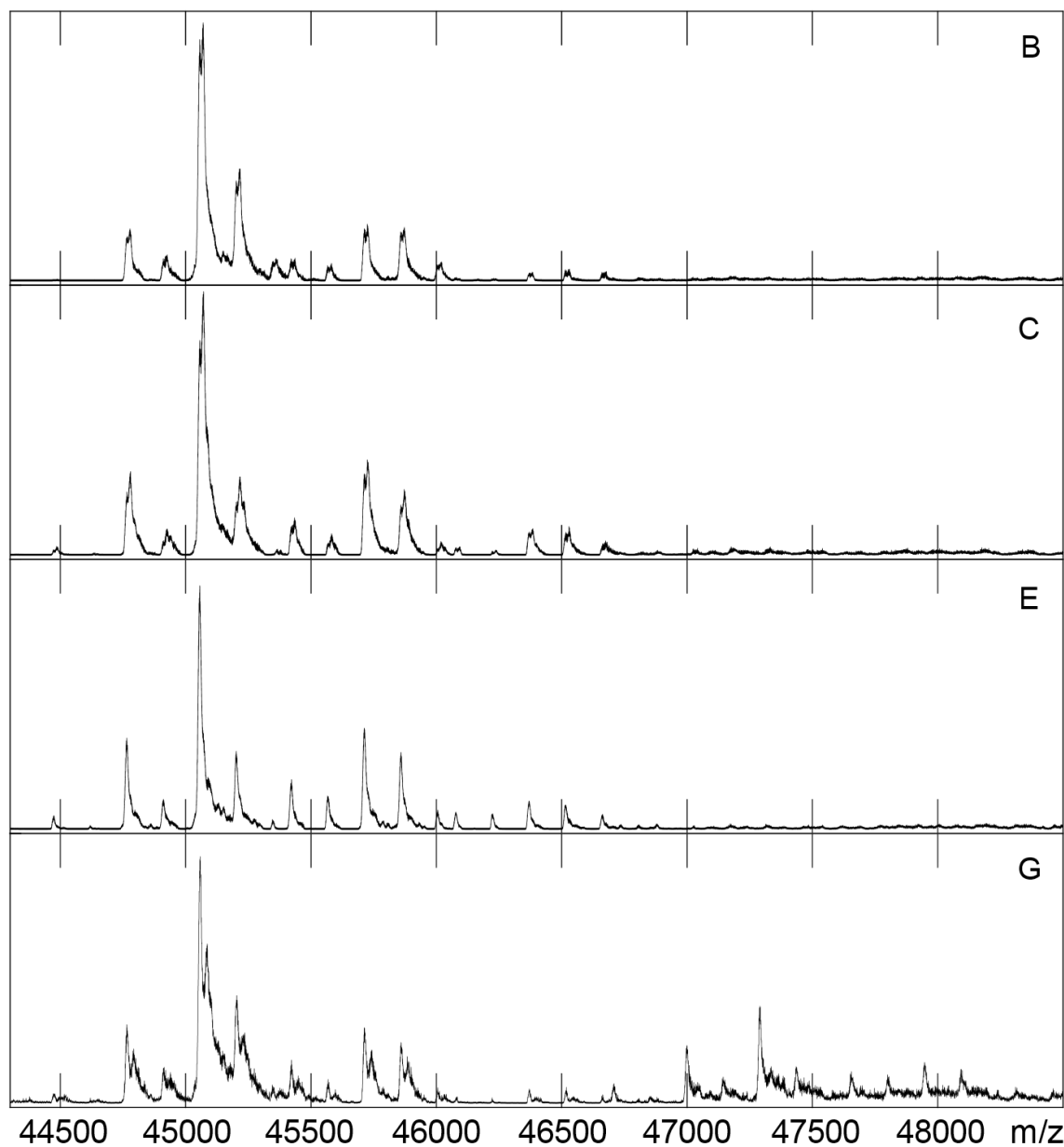


Figure A 3. Deconvoluted ESI-MS spectra of human B2GP1 for sample B, C, E, and G. Donor B and C are heterozygous for the native variant and variant V247L. Donor E is homozygous for the native variant and donor G is heterozygous for the native variant and variant S88N with a partially occupied fifth glycosylation site.

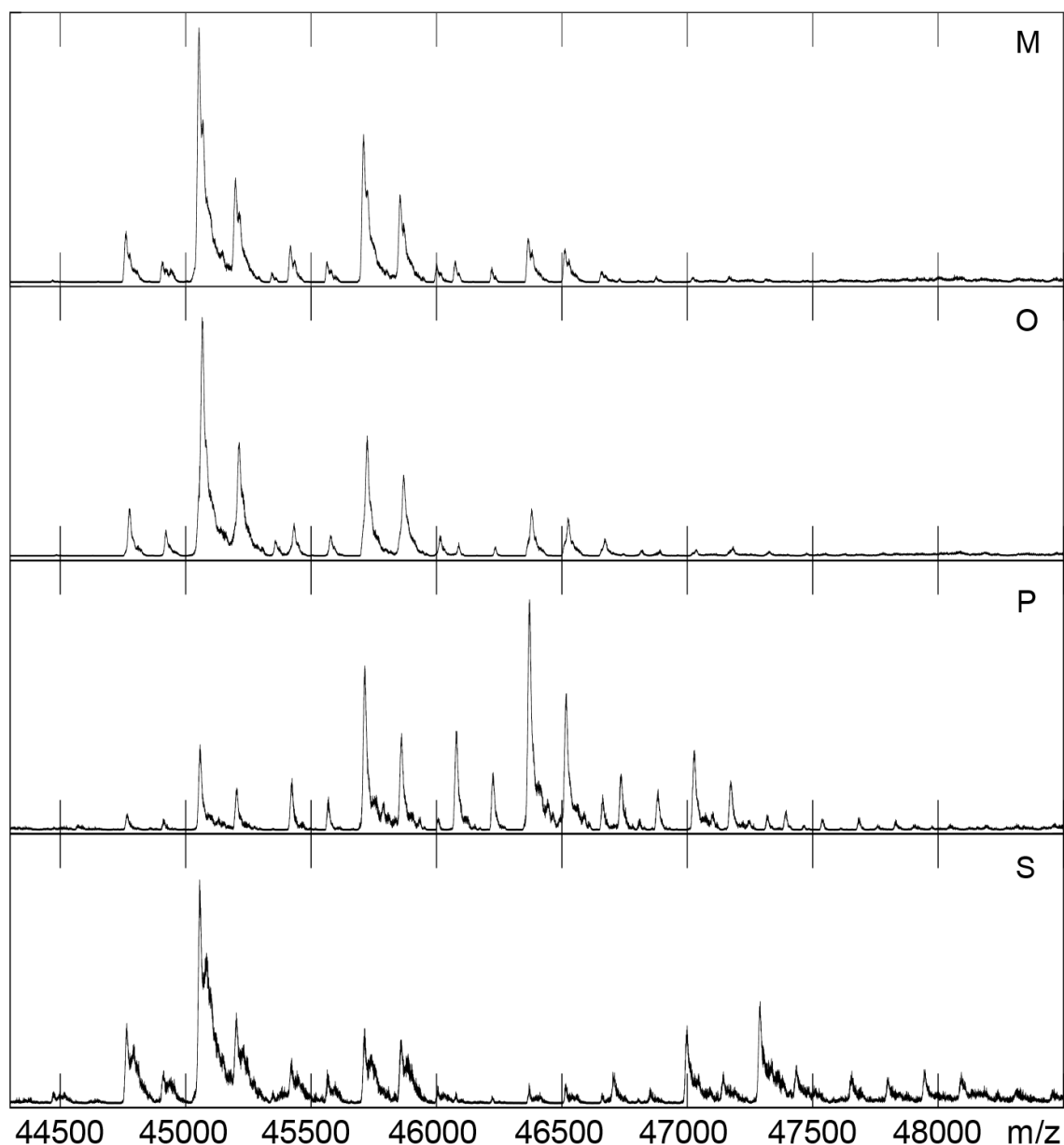


Figure A 4. Deconvoluted ESI-MS spectra of human B2GP1 for sample M, O, P, and S. Donor M shows a heterozygous expression for variant C169N and a variant with two SNPs, V247L and R135H. Donor O is homozygous for variant C169N. Donor P is homozygous for the native variant and shows a different glycosylation pattern with highly sialylated glycan structures. Donor S is heterozygous for the native variant and variant S88N with a partially occupied fifth glycosylation site.

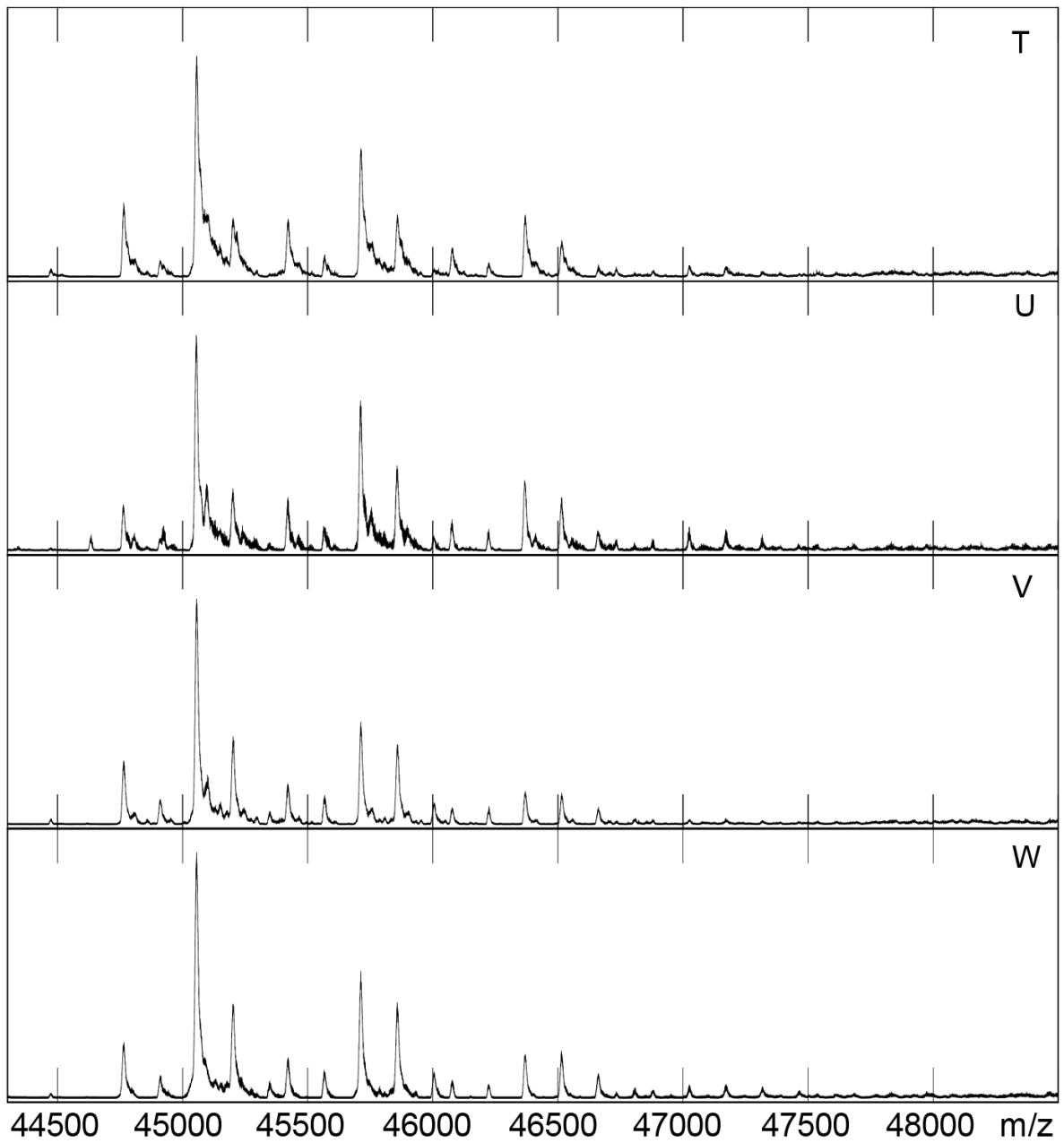


Figure A 5. Deconvoluted ESI-MS spectra of human B2GP1 for sample T, U, V, and W. All donors are homozygous for the native variant.

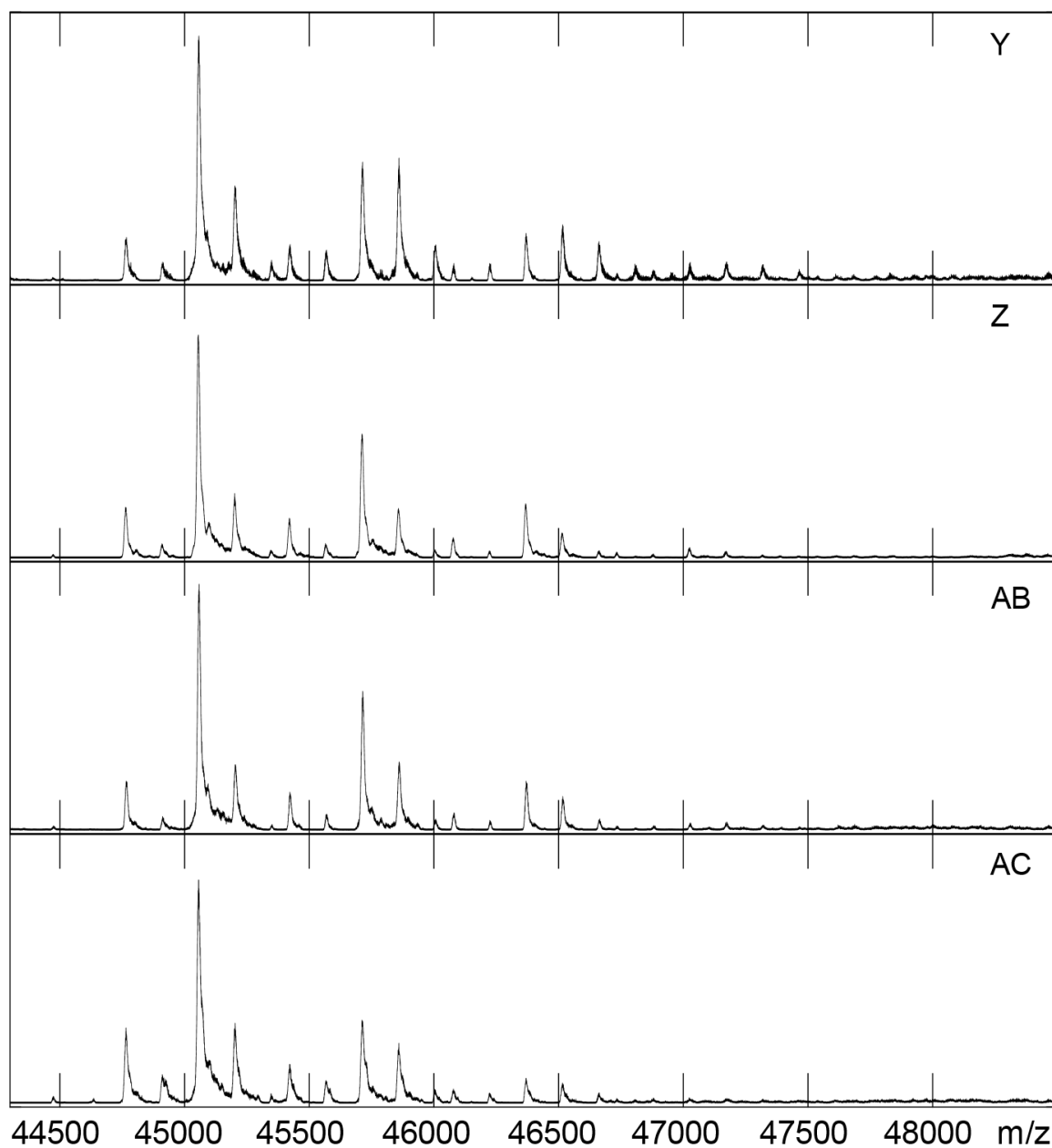


Figure A 6. Deconvoluted ESI-MS spectra of human B2GP1 for sample Y, Z, AB, and AC. All donors are homozygous for the native variant.

APPENDIX

Table A 16. Relative abundance of glycan compositions of  $\beta$ -2-glycoprotein 1 from different human plasma samples over all glycosylation sites. Sample G and sample S have a partially occupied fifth glycosylation site. These compositions can be found in Table A 17.

Glycan composition	B		C		E		G		M		O		P		S		T		U		V		W		Y		Z		AB		AC	
	native [%]	V247L [%]	native [%]	V247L [%]	native [%]	native [%]	S88N [%]	C169N [%]	V247L R145H [%]	native [%]	native [%]	native [%]	S88N [%]	native [%]	native [%]	native [%]	S88N [%]	native [%]	native [%]	native [%]	native [%]	native [%]	native [%]	native [%]	native [%]	native [%]	native [%]	native [%]	native [%]	native [%]	native [%]	
H20N16S6	-	-	0.6	0.8	1.2	1.1	0.7	0.1	0.0	0.1	0.1	1.2	0.9	0.7	0.2	0.5	0.3	0.2	0.2	0.2	0.3	0.6										
H20N16S7	7.3	7.4	10.2	10.4	10.7	10.2	6.2	5.0	5.5	5.3	1.1	10.9	6.7	8.4	4.9	7.7	5.8	4.2	6.8	6.1	9.9											
H20N16S7F	3.2	3.5	2.1	3.1	3.4	4.6	3.0	2.2	2.1	2.6	0.6	4.2	2.7	1.8	1.5	3.0	2.2	1.6	1.7	1.4	3.9											
H20N16S8	40.3	39.6	36.6	34.8	29.2	34.5	20.2	30.5	29.5	28.5	6.0	32.3	19.1	26.5	24.3	28.3	27.1	25.5	31.6	32.5	32.0											
H20N16S8F	17.2	17.5	9.3	10.3	9.4	15.5	9.0	12.7	12.1	13.9	3.0	12.9	6.9	7.4	6.8	10.7	10.5	10.0	8.6	8.8	10.9											
H20N16S8F2	2.9	3.1	0.1	0.4	0.8	2.1	1.5	0.6	0.8	1.6	-	1.1	1.2	0.2	0.5	1.3	1.5	1.6	0.8	0.4	0.8											
H20N16S8F3	0.1	0.2	-	-	-	0.9	0.5	-	-	-	0.1	1.1	0.6	0.6	0.2	0.1	-	-	0.3	-	0.1											
H21N17S8	3.0	2.8	4.4	4.3	5.5	4.9	2.7	3.7	3.9	3.4	3.4	5.7	3.1	6.7	5.1	4.8	4.1	3.4	5.2	4.5	5.2											
H21N17S8F	2.0	2.1	1.7	2.3	3.7	2.4	1.0	2.0	2.1	2.2	1.9	3.6	1.9	2.2	2.4	3.2	2.7	2.9	1.7	1.7	3.0											
H21N17S9	8.2	8.0	13.3	12.4	12.2	9.9	6.1	17.4	17.0	14.0	12.5	9.9	5.8	15.5	16.8	12.5	13.4	12.2	17.7	17.9	12.0											
H21N17S9F	8.1	8.0	8.3	8.4	9.1	8.0	4.9	10.3	10.1	9.7	7.2	9.2	5.3	7.4	9.3	9.9	10.3	12.0	6.9	8.9	7.9											
H21N17S9F2	2.4	2.6	0.9	1.4	1.9	1.4	0.7	1.5	1.7	2.1	0.6	1.6	1.0	0.8	1.4	2.5	2.6	3.5	0.9	1.1	1.5											
H21N17S9F3	-	0.1	-	-	-	0.1	0.1	-	-	-	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.1	-	-	0.1											
H22N18S9	0.3	0.2	0.9	0.7	1.8	0.5	0.1	1.4	2.0	1.0	7.3	0.9	0.2	3.3	2.8	1.8	1.5	1.1	2.5	1.9	1.7											
H22N18S9F	0.1	0.1	0.3	0.4	1.6	0.2	-	0.8	1.2	0.8	4.0	0.5	0.1	1.4	1.7	1.6	1.2	1.3	0.6	0.8	1.1											
H22N18S10	0.9	0.9	3.6	3.2	3.1	1.4	0.6	5.4	5.0	5.3	18.0	1.6	0.7	7.2	7.8	3.9	4.6	4.4	7.4	6.0	3.2											
H22N18S10F	1.2	1.2	3.2	3.0	2.8	1.4	0.6	3.8	3.6	4.2	10.5	2.0	0.8	4.2	5.5	3.5	4.6	5.2	3.3	3.9	2.5											
H22N18S10F2	0.9	0.9	1.3	1.4	1.5	0.6	0.3	1.1	1.1	1.9	2.2	0.8	0.5	1.0	2.1	1.8	2.4	3.6	0.7	1.1	1.1											
H22N18S10F3	0.3	0.2	0.2	0.2	0.3	0.3	0.1	0.1	0.1	0.5	0.5	0.3	0.1	0.3	0.4	0.5	0.7	1.2	0.1	0.1	0.3											
H23N19S10	0.1	-	0.3	0.1	0.3	-	-	0.1	0.3	0.2	4.1	-	-	0.8	0.8	0.3	0.3	0.4	0.5	0.3	0.3											
H23N19S10F	0.2	0.1	0.4	0.3	0.4	-	-	0.3	0.5	0.4	2.7	-	-	0.6	0.7	0.4	0.6	0.7	0.3	0.3	0.4											
H23N19S11	0.2	0.2	0.5	0.4	0.2	-	-	0.3	0.4	0.6	6.0	-	-	1.1	1.8	0.5	1.0	1.3	1.1	0.6	0.5											
H23N19S11F	0.5	0.5	0.8	0.8	0.4	-	-	0.4	0.5	0.9	3.7	-	-	1.1	1.6	0.5	1.2	1.6	0.7	0.7	0.5											
H23N19S11F2	0.4	0.4	0.7	0.6	0.4	-	-	0.3	0.3	0.4	0.9	-	-	0.5	0.9	0.3	0.9	1.3	0.2	0.5	0.3											
H23N19S11F3	0.2	0.2	0.4	0.4	0.2	-	-	0.2	0.1	0.2	0.2	-	-	0.3	0.4	0.2	0.6	0.8	0.1	0.2	0.3											
H24N20S11	-	-	-	-	-	-	-	-	-	-	1.2	-	-	-	-	-	-	-	-	-	-											
H24N20S11F	-	-	-	-	-	-	-	-	-	-	0.7	-	-	-	-	-	-	-	-	-	-											
H24N20S12	-	-	-	-	-	-	-	-	-	-	0.7	-	-	-	-	-	-	-	-	-	-											
H24N20S12F	-	-	-	-	-	-	-	-	-	-	0.5	-	-	-	-	-	-	-	-	-	-											

APPENDIX

Table A 17. Additional glycan compositions of  $\beta$ -2-glycoprotein 1 for variant S88N. The SNP leads to a fifth, partially occupied glycosite.

Glycan composition	<b>G</b>	<b>S</b>
	S88N [%]	S88N [%]
H25N20S8	1.7	2.8
H25N20S8F	0.6	1.0
H25N20S9	6.0	7.8
H25N20S9F	2.7	2.8
H25N20S10	10.4	10.7
H25N20S10F	4.0	3.8
H25N20S10F2	1.1	0.4
H26N21S10	3.3	3.0
H26N21S10F	2.8	2.5
H26N21S11	4.5	3.4
H26N21S11F	3.6	3.0
H26N21S11F2	1.1	0.9

Table A 18. Relative abundance of glycan compositions of  $\beta$ -2-glycoprotein 1 from plasma samples of patients diagnosed with liver cirrhosis and hepatocellular carcinoma.

Glycan composition	<b>HCC08</b>	<b>HCC13</b>	<b>HCC19</b>	<b>HCC19</b>	<b>LC10</b>	<b>LC11</b>	<b>LC11</b>
	native [%]	V247L [%]	native [%]	C188N [%]	native [%]	native [%]	V266L [%]
H20N16S6	0.1	0.1	0.2	0.2	0.5	0.2	0.4
H20N16S7	3.5	1.6	6.5	6.6	2.6	7.9	8.6
H20N16S7F	2.9	2.5	3.7	4.0	1.9	7.1	7.6
H20N16S8	23.4	12.2	32.9	33.0	15.3	32.2	32.5
H20N16S8F	16.1	6.3	14.3	13.9	8.8	19.2	18.6
H20N16S8F2	4.1	0.7	1.8	1.6	1.5	3.7	3.5
H20N16S8F3	0.7	-	-	-	0.2	0.2	0.2
H21N17S8	3.1	2.2	2.8	2.7	2.4	3.5	3.3
H21N17S8F	3.0	4.6	2.7	2.9	3.1	3.7	3.8
H21N17S9	9.8	9.3	10.2	10.3	9.4	6.8	6.9
H21N17S9F	9.9	9.9	10.6	10.5	12.0	7.2	7.1
H21N17S9F2	4.0	3.8	2.9	2.8	5.0	2.6	2.5
H21N17S9F3	0.7	0.3	-	-	0.8	0.1	-
H22N18S9	1.6	3.5	1.0	0.8	1.7	0.6	0.5
H22N18S9F	1.4	5.2	0.6	0.7	2.4	0.4	0.4
H22N18S10	3.5	7.8	2.1	2.2	4.7	0.7	0.6
H22N18S10F	3.7	8.3	3.1	3.0	7.2	0.8	0.7
H22N18S10F2	2.5	5.1	1.9	1.9	6.1	0.6	0.5
H22N18S10F3	1.1	1.6	0.4	0.4	2.5	0.4	0.3
H23N19S10	0.5	2.9	0.2	0.1	1.3	0.1	0.1
H23N19S10F	0.8	2.9	0.2	0.1	1.4	0.4	0.3
H23N19S11	1.0	3.0	0.3	0.3	2.0	0.4	0.3
H23N19S11F	1.1	3.1	0.7	0.7	2.4	0.5	0.4
H23N19S11F2	0.9	2.1	0.7	0.6	2.8	0.4	0.4
H23N19S11F3	0.5	1.1	0.4	0.4	2.0	0.3	0.3



## 11. Acknowledgements

I would like to thank my parents for their unconditional support and love throughout my whole life.

I am grateful to all former and present members of the AK Meyer as we had great time together. I learned a lot from each and everyone, especially Henning who introduced me to mass spectrometry.

I would like to thank my boyfriend Christian who supported me during all my stages of my PhD. Even during the last, difficult months.

I am tremendously grateful to my lifelong friends who stood by me not only during my PhD time but for so many years: Jennifer, Julia, Leevke, Sabrina, and Simone.

Special thanks to Anna-Lena and Claas who made the past years fun. *#Nicht ohne mein Team!*

For their male perspective and friendship, I would like to thank Dawid and Jim.

NMR support was always prompt and with creative solutions, especially with the 700er. Thank you, Thomas, Moritz, Jim, Mirco, and Claas!

I would like to thank all my student interns: Knut, Stefan, Theresa, Tim, and above all Manuela, who completed two internships highly committed and motivated.

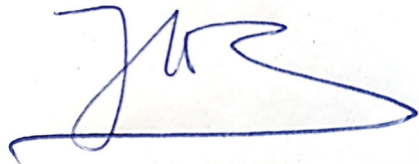
For proof reading, I would like to thank Anna-Lena, Christian, Jim, Sandra, and Simone.

## Affidavit

### Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbstständig angefertigt und keine anderen als die von mir angegebenen Hilfsmittel und Quellen verwendet habe. Ich versichere weiterhin, dass die vorliegende Dissertation weder in gleicher noch in veränderter Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Hamburg,

A handwritten signature in blue ink, consisting of a stylized 'M' followed by a large, sweeping flourish that ends in a horizontal line.

---

Melissa Bärenfänger