In-Depth Mass Spectrometry-Based Glycoproteomics: Advances in Sample Preparation, Measurement and Data-Analysis of Glycoproteins

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur
(Dr.-Ing.)

von: Dipl.-Ing. Marcus Hoffmann
geboren am: 21. Oktober 1984 in Aschersleben

genehmigt durch die Fakultät für Verfahrens- und Systemtechnik
der Otto-von-Guericke Universität Magdeburg

Promotionskommission:

Prof. Dr. rer. nat. habil. Helmut Weiß (Vorsitz)
Prof. Dr.-Ing. Udo Reichl (Gutachter)
Dr. rer. nat. Erdmann Rapp (Gutachter)
Prof. Dr. rer. nat. Manfred Wuhrer (Gutachter)

eingereicht am: 26. November 2019 vorgelegte Dissertation
Promotionskolloquium am: 29. Juni 2020

Abstract

Protein glycosylation is a universal and essential feature that is prominent in all domains of life (eukarya, bacteria, archaea, viruses). The co- and posttranslational attachment of glycans to specific amino acids may thereby profoundly alter the biophysical characteristics, and with this also the functioning of the proteins they are bound to. This may include structural and modulatory effects, such as proper protein folding or protection from proteolytic digestion, along with extrinsic and intrinsic recognition effects, such as recognition of antigens, protein trafficking and turnover, intra- and extracellular signaling and adhesion, as well as molecular mimicry. Protein glycosylation therefore plays an important role in many physiological but also pathophysiological processes. To understand the functional implications of the protein glycosylation and to take advantage of these insights for clinical diagnosis and prognosis, as well as for biopharmaceutical and vaccine development and quality control, the continuous development and improvement of glycoanalytical methods and workflows is required. This thesis describes the development of two mass spectrometry-based glycoproteomic workflows that enable the in-depth and site-specific analysis of *N*- and *O*-glycoproteins. Further, the results of applying of those workflows for the glycoproteomic analysis of clinically and biopharmaceutically relevant samples are presented.

In the first part of this thesis a glycoproteomic workflow, developed for the site-specific *O*-glycosylation analysis of human blood plasma glycoproteins, is introduced and the results of this analysis are presented and discussed. To this end pooled human blood plasma of healthy donors was proteolytically digested using an enzyme with a broad cleavage specificity (proteinase K), followed by a precipitation step, as well as a glycopeptide enrichment and fractionation step via hydrophilic interaction liquid chromatography (HILIC). The latter was optimized for intact *O*-glycopeptides carrying short mucin-type core-1 and 2 *O*-glycans, which represent the vast majority of *O*-glycans on human blood plasma proteins. Enriched *O*-glycopeptide fractions were subjected to mass spectrometric analysis using reversed-phase liquid chromatography coupled online to an ion trap (IT) mass spectrometer operated in positive-ion mode. Peptide identity and glycan composition were derived from low-energy collision-induced dissociation fragment ion spectra acquired in multistage mode. To pinpoint the *O*-glycosylation sites glycopeptides were fragmented using electron transfer dissociation. Acquired fragment ion spectra were annotated by database searches and manual assignment. Overall, 31 *O*-glycosylation sites and regions belonging to 22 proteins were identified, with the majority being acute-phase proteins. Strikingly, also 11 novel *O*-glycosylation sites and regions were identified. In total 23 of the 31 *O*-glycosylation sites/regions could be pinpointed. Interestingly, the use of proteinase K, which cleaves primarily after aliphatic, aromatic and hydrophobic amino acids, proved to be particularly beneficial in this context. The identified *O*-glycan compositions most probably correspond to mono- and disialylated core-1 mucin-type *O*-glycans (T-antigen). In summary, the developed workflow allows the identification and characterization of the major population of the human blood plasma *O*-glycoproteome.

Thereby, these results provide novel insights, which may help to unravel glycosylation-related structure-function relationships, e.g. during glycoproteomic biomarker discovery studies.

In the second part of the thesis a glycoproteomic workflow, combining lower and stepped collisional energy fragmentation on an Orbitrap mass spectrometer, is introduced for the in-depth and site-specific analysis of intact *N*- and *O*-glycopeptides. Further, using a set of four representative and biopharmaceutically-relevant glycoproteins (immunoglobulin gamma, fibrinogen, lactotransferrin, and ribonuclease B), the benefits and limitations of this workflow are highlighted. For these representative glycoproteins several new glycosylation sites and regions with their respective glycoforms were discovered and characterized – in addition to confirming already known glycosylation sites and glycoforms (elucidation of glycan micro- and macroheterogeneity). Moreover, for the enrichment of glycopeptides a modified and improved version of a cotton HILIC-based solid phase extraction protocol is described. For the unambiguous identification of *N*-glycopeptides the use of a conserved fragmentation signature $[M_{peptide}+H+^{0,2}X\ GlcNAc]^+$, that has rarely been employed in Orbitrap-based glycoproteomic analyses up to now, is proposed. This signature represents a valuable indicator for the determination of the correct peptide mass and is particularly important when analyzing glycopeptides obtained by proteases with broad or no cleavage specificity (e.g. glycopeptides generated by proteinase K digest). In this thesis, it is shown for the first time that this fragmentation signature can consistently be found across all *N*-glycopeptides analyzed, but not for *O*-glycopeptides. Moreover, we have systematically and comprehensively evaluated the use of the relative abundance of oxonium ions to retrieve glycan structure information, e.g. differentiation of hybrid- and high-mannose-type *N*-glycans, differentiation between antenna GlcNAc and bisecting GlcNAc, or differentiation between *N*-glycopeptides and mucin-type *O*-glycopeptides. These new findings may increase confidence and comprehensiveness in manual and software-assisted glycoproteomics. Overall, the developed analytical workflow along with the analyzed representative glycoproteins and the given insights into diagnostic glycopeptide fragment ions serve as a guide to in-depth glycoproteomic analysis for a broad range of glycoproteins. The workflow may therefore be beneficial to basic and clinical glycoproteomic research and diagnostics, but also to biopharmaceutical research and process development as well as quality control.

In summary, in this thesis two mass spectrometry-based glycoproteomic workflows were developed and applied which enable the in-depth and site-specific analysis of *N*- and *O*-glycoproteins. Thereby, several advances in sample preparation, measurement and data-analysis were achieved that increase the confidence, depth and comprehensiveness of glycoproteomic analyses.

Kurzfassung

Die Protein-Glykosylierung ist ein universelles und essentielles Merkmal, das in allen Organismen (Eukarya, Bakteria, Archaea), einschließlich Viren, zu finden ist. Die ko- und posttranslationale Bindung von Glykanen an bestimmte Aminosäuren kann dabei die biophysikalischen Eigenschaften und damit auch die Funktion der Proteine, an die sie gebunden sind, grundlegend verändern. Dies schließt strukturelle und modulatorische Effekte, wie z.B. die richtige Proteinfaltung oder den Schutz vor Proteasen, aber auch extrinsische und intrinsische Erkennungseffekte wie z.B. die Erkennung von Antigenen, sowie intra- und extrazelluläre Signalgebung, Adhäsion oder molekulare Mimikry mit ein. Die Glykosylierung von Proteinen spielt daher eine wesentliche Rolle in einer Vielzahl physiologischer sowie pathophysiologischer Prozesse. Um die funktionellen Auswirkungen der Proteinglykosylierung zu verstehen und diese Erkenntnisse für die klinische Diagnose und Prognose sowie für die Entwicklung von Biopharmazeutika und Impfstoffen und deren Qualitätskontrolle nutzen können, ist die kontinuierliche Entwicklung und Verbesserung glykoanalytischer Methoden und Arbeitsabläufe erforderlich. Diese Dissertation beschreibt die Entwicklung zweier Arbeitsabläufe für die detaillierte und ortsspezifische Glykoproteom-Analyse von *N*- und *O*-Glykoproteinen mittels Massenspektrometrie (MS). Ferner wird die Anwendung beider Arbeitsabläufe für die Glykoproteom-Analyse von klinisch- und biopharmazeutisch-relevanten Proben gezeigt.

Im ersten Teil der Arbeit wird die ortsspezifische *O*-Glykoproteom-Analyse von Glykoproteinen aus menschlichem Blutplasma beschrieben. Zu diesem Zweck wurde menschliches Blutplasma gesunder Spender unter Verwendung eines proteolytischen Enzyms mit breiter Spezifität (Proteinase K) proteolytisch verdaut, gefolgt von einem Präzipitationsschritt sowie einem Glykopeptid-Anreicherungs- und Fraktionierungsschritt mittels hydrophiler Interaktionsflüssigkeitschromatographie (HILIC). Letzterer wurde für intakte *O*-Glycopeptide mit *O*-Glykanen vom Mucin-Typ (Core 1 und 2) optimiert, welche die große Mehrheit der *O*-Glykane auf menschlichen Blutplasmaproteinen darstellen. Die *O*-Glykopeptid-angereicherten Fraktionen wurden einer massenspektrometrischen Analyse unterzogen. Die *O*-Glykopeptide wurden dabei unter Verwendung von Umkehrphasen-Flüssigkeitschromatographie aufgetrennt und anschließend mit einem Ionenfallen-Massenspektrometer, das im Positivionenmodus betrieben wurde, analysiert. Unter Verwendung von niedrigenergetischer kollisionsinduzierter Dissoziation wurden mehrstufige Fragmentionenspektren der *O*-Glykopeptide erzeugt. Diese Fragmentionenspektren ermöglichten es, die Identität des Peptids sowie die Zusammensetzung der dazugehörigen Glykane zu bestimmen. Um die *O*-Glykosylierungsstellen genau bestimmen zu können, wurden die *O*-Glykopeptide mittels Elektronentransferdissoziation fragmentiert. Zur Identifizierung der Glykopeptide wurden die erzeugten Fragmentionenspektren sowohl manuell als auch durch Datenbanksuchen annotiert und interpretiert. Insgesamt wurden 31 *O*-Glykosylierungsstellen bzw. *O*-

Glykosylierungsregionen identifiziert. Diese konnten 22 Blutplasmaglykoproteinen zugeordnet werden, wobei es sich bei der Mehrzahl der Glykoproteine um Akute-Phase-Proteine handelt. Interessanterweise konnten auch 11 neue *O*-Glykosylierungsstellen bzw. -regionen identifiziert werden. Insgesamt konnte bei 23 der 31 *O*-Glykosylierungsstellen bzw. -regionen auch die glykosylierte Aminosäure genau lokalisiert werden. In diesem Zusammenhang erwies sich die Verwendung von Proteinase K als besonders vorteilhaft, da diese sowohl nach aliphatischen, als auch aromatischen und hydrophoben Aminosäuren schneidet. Die Analyse der *O*-Glykan-Zusammensetzung der identifizierten *O*-Glykopeptide ergab sowohl mono- als auch disialylierte *O*-Glykane vom Mucin-Typ 1 (T-Antigen). Zusammenfassend lässt sich sagen, dass der entwickelte Arbeitsablauf die Identifizierung und Charakterisierung der Hauptpopulation des menschlichen Blutplasma-*O*-Glykoproteoms ermöglicht. Dabei liefern die erzielten Ergebnisse neuartige Erkenntnisse, die dazu beitragen können, glykosylierungsbezogene Struktur-Funktions-Beziehungen aufzuklären, z.B. im Rahmen von Glykoproteomstudien zur Biomarkerfindung.

Der zweite Teil der Arbeit umfasst die detaillierte und Glykosylierungsstellen-bezogene Analyse intakter *N*- und *O*-Glycopeptide unter Verwendung eines Orbitrap Massenspektrometers. Ein zentrales Element des dafür entwickelten Arbeitsablaufes ist die Verwendung unterschiedlicher Kollisionsenergien für die Fragmentierung von *N*- und *O*-Glykopeptiden (Anwendung ausschließlich niedriger Kollisionsenergie, sowie niedriger und hoher Kollisionsenergie in Kombination). Im Rahmen der Arbeit werden die Vorteile und Einschränkungen des entwickelten Arbeitsflusses anhand von vier repräsentativen und biopharmazeutisch-relevanten Glykoproteinen (Immunglobulin Gamma, Fibrinogen, Lactotransferrin und Ribonuclease B) beleuchtet. Für diese Glykoproteine konnten mehrere neue Glykosylierungsstellen und -regionen mit ihren jeweiligen Glykoformen identifiziert und charakterisiert werden – zusätzlich zur Bestätigung bereits bekannter Glykosylierungsstellen und Glykoformen (Aufklärung der Glykan-Mikro- und -Makroheterogenität). Des Weiteren wird in dieser Arbeit eine modifizierte und verbesserte Version der Glykopeptidanreicherung mittels baumwollbasierter HILIC-Festphasenextraktion beschrieben. Zur eindeutigen Identifizierung von *N*-Glykopeptiden wird im Rahmen der Arbeit die Verwendung einer konservierten Fragmentierungssignatur $[M_{peptid} + H + {}^{0,2}X\ GlcNAc]^+$ vorgeschlagen. Diese Signatur wurde bisher selten in Orbitrap-basierten Glykoproteom-Analysen eingesetzt. Sie stellt jedoch einen nützlichen Indikator für die Bestimmung der korrekten Peptidmasse dar, und ist daher besonders hilfreich bei der Analyse von Glykopeptiden, die durch Proteasen mit breiter oder keiner Spaltungsspezifität generiert worden (z. B. durch Proteinase-K-Verdau). In der vorliegenden Arbeit wird erstmalig gezeigt, dass diese Fragmentierungssignatur über alle analysierten *N*-Glykopeptide hinweg konsistent gefunden werden kann – nicht jedoch bei *O*-Glykopeptiden. Um noch detaillierte Informationen über die an den Glykopeptiden befindliche Glykanstruktur zu erhalten, wurde im Rahmen der Arbeit systematisch und umfassend die Verwendung der relativen Oxoniumionen-Häufigkeit untersucht. Die dabei gewonnenen Erkenntnisse ermöglichen unter anderem die genaue Unterscheidung zwischen

Hybrid- und High-Mannose-*N*-Glykanen, die Unterscheidung zwischen Antennen-GIcNAc und Bisecting-GIcNAc sowie die Unterscheidung zwischen *N*-Glykopeptiden und Mucin-Typ *O*-Glykopeptiden. Insgesamt können diese neuen Erkenntnisse den Detailgrad sowie die Konfidenz manueller als auch softwaregestützter Glykoproteom-Analysen erhöhen. Der entwickelte Arbeitsablauf dient, zusammen mit Erkenntnissen aus der Analyse der repräsentativen Glykoproteine und den Erkenntnissen zu diagnostischen Glykopeptid-Fragmentionen, als Grundlage für zukünftige Glykoproteomanalysen eines breiten Spektrums von Glykoproteinen. Der entwickelte Arbeitsablauf kann daher für die Grundlagen- und klinische Glykoproteomforschung und -diagnostik, aber auch für die biopharmazeutische Forschung und Entwicklung sowie die Qualitätskontrolle von Nutzen sein.

Zusammenfassend wurden in dieser Arbeit zwei MS-basierte Arbeitsablaufe entwickelt und angewendet, die detaillierte und ortsspezifische Glykoproteom-Analysen von *N*- und *O*-Glykoproteinen ermöglichen. Die dabei erzielten Verbesserungen im Rahmen der Probenvorbereitung, der Messungen sowie der Datenauswertung, führen zu einer erhöhten Konfidenz sowie einem höheren Detailgrad und Umfang von Glykoproteomanalysen.

Table of Contents

Abstract ..... **I**

Kurzfassung ..... **III**

List of Abbreviations ..... **IX**

1. Introduction **1**

2. Theoretical Background **8**

2.1. Protein Glycosylation 8

2.2. Biological Implications of *N*- and *O*-Glycosylation 21

2.3. Glycosylation in Human Health and Disease 26

2.4. Glycoengineering and Glycotherapeutics 27

2.5. Analysis of Glycoproteins 30

2.5.1. Monosaccharide Analysis 32

2.5.2. Analysis of Released Glycans 33

2.5.3. Analysis of Intact Glycoproteins 34

2.5.4. Analysis of Glycopeptides 35

3. Materials and Methods **45**

3.1. Human Blood Plasma *O*-Glycoproteomics 46

3.1.1. Proteolytic Digestion 46

3.1.2. Glycopeptide Enrichment and Fractionation via HILIC-HPLC 47

3.1.3. Nano-RP-LC-ESI-IT-MSn (CID, ETD) 47

3.1.4. Data Analysis 49

3.2. In-Depth *N*- and *O*-Glycoproteomics 52

3.2.1. Proteolytic Digestion via Filter-Aided Sample Preparation (FASP) 52

3.2.2. Glycopeptide Enrichment via Spin-Cotton-HILIC-SPE 53

3.2.3. Nano-RP-LC-ESI-OT-OT-MS/MS (HCD) 54

3.2.4. Glycopeptide Data Analysis Using glyXtool^{MS} 55

3.2.5. Site-Specific Relative Quantitation of *N*-Glycoforms 56

3.2.6. Relative Quantitation of *N*- and *O*-Glycopeptide Oxonium Ions 57

4. Human Blood Plasma *O*-Glycoproteomics **58**

4.1. Introduction 59

4.2. Results 61

4.2.1. Reproducibility of the Proteinase K Digest 61
4.2.2. Glycopeptide Enrichment and Fractionation via HILIC-HPLC 62
4.2.3. Determination of the Glycan Composition 63
4.2.4. Identification of the Peptide Moiety 64
4.2.5. Localization of the *O*-Glycosylation Sites 72
4.2.6. Identified Glycoproteins: Selected Examples 74
4.3. Discussion 77
4.3.1. Other *O*-Glycoproteomic Studies on Complex Biofluids 78
4.3.2. Proteinase K Digest 79
4.3.3. Glycopeptide Enrichment via HILIC 80
4.3.4. Identification of the O-Glycan Composition 81
4.3.5. Glycopeptide Identification 81
4.3.6. Pinpointing of O-Glycosylation Sites 83
4.3.7. Caveats of the Approach 84
4.4. Summary 84
5. In-Depth *N*- and *O*-Glycoproteomics 85
5.1. Introduction 86
5.2. Results and Discussion 86
5.2.1. Fragmentation of N- and O-Glycopeptides by HCD.low and HCD.step ... 86
5.2.2. Spin-Cotton-HILIC-SPE 91
5.2.3. Glycoproteomic Analysis of Immunoglobulin Gamma, Fibrinogen, Lactotransferrin, and Ribonuclease B 95
5.2.4. Quantification of Oxonium Ions and Its Potential for N- and O-Glycoproteomics 99
5.3. Summary 112
6. Conclusion and Outlook 113
6.1. Human Blood Plasma *O*-Glycoproteomics 113
6.2. In-Depth N- and O-Glycoproteomics 115
7. Bibliography 118
8. List of Contributions 134
9. List of Figures 138
10. List of Tables 139

11. List of Materials and Chemicals 139

12. List of Devices and Instruments 140

List of Abbreviations

2

2D-PAGE · two-dimensional gel electrophoresis

A

ABC · ammonium bicarbonate
ACN · acetonitrile
ADCC · antibody-dependent cellular cytotoxicity
AGC · automatic gain control
ALG · asparagine-linked glycosylation
Asn · asparagine

B

BPC · base peak chromatogram

C

C(G)E-LIF · capillary (gel) electrophoreses coupled to laser-induced fluorescence
C(Z)E · capillary (zone) electrophoresis
CDG · congenital disorders of glycosylation
CFG · Consortium for Functional Glycomics
CHO · Chinese hamster ovary
CID · collision-induced dissociation
COSMC · core-1 β3-Gal-T-specific molecular chaperone
CRISPR/Cas9 · clustered regularly interspaced short palindromic repeats/CRISPR-associated 9
CSF · cerebrospinal fluid
Cys · cysteine

D

Da · Dalton
Dol-P-Glc · dolichol-phosphate glucose
Dol-P-Man · dolichol-phosphate mannose
DTT · dithiothreitol

E

ECD · electron-capture dissociation
EIC / XIC · extracted ion chromatogram
EMA · European Medicines Agency
Endo D / Endo H · endo-ß-*N*-acetylglucosaminidases
ER · endoplasmic reticulum
ERLIC · electrostatic repulsion hydrophilic interaction chromatography
ESI · electrospray ionization
ETD · electron-transfer dissociation
EThcD · electron-transfer/higher-energy collisional dissociation

F

FA · formic acid
FASP · filter-aided sample preparation
Fc · fragment crystallizable
FDA · U.S. Food and Drug Administration
FDR · false discovery rate
Fib · fibrinogen
FLR · fluorescence detection
FTICR · Fourier transform ion cyclotron resonance
Fuc / dHEX · fucose

G

Gal · galactose
GalNAc · *N*-acetylgalactosamine
GalNAc-T · GalNAc-transferase isoenzyme
GDP-Man · guanosine diphosphate mannose
Glc · D-glucose
GlcA · glucuronic acid
GlcNAc · *N*-acetylglucosamine
GnT I · *N*-acetylglucosaminyltransferase I
GnT-III · β1,4-*N*-acetylglucosaminyltransferase III

H

HA · hemagglutinin
HCD · higher-energy collision dissociation
Hex · hexose
HexNAc · *N*-acetylhexosamine
HILIC · hydrophilic interaction liquid chromatography
HPAEC-PAD · high-performance anion exchange chromatography with pulsed amperometric detection
HPLC · high-performance liquid chromatography

I

IAA · iodoacetamide
ICC · ion charge control

IdeS · immunoglobulin-degrading enzyme of *Streptococcus pyogenes*
IgA · immunoglobulin A
IgD · immunoglobulin D
IgG · immunoglobulin gamma
IgM · immunoglobulin M
IM-MS · ion mobility-MS
IT · ion-trap
IUPAC · International Union of Pure and Applied Chemistry

K

KDN · 2-keto-3-deoxynononic acid

L

LacdiNAc / LDN · $GlcNAc_1GalNAc_1$
LacNAc · *N*-acetyllactosamine
LC · liquid chromatography, see also **HPLC** and **UPLC**
LTF · Lactotransferrin
LTQ · linear quadrupole ion trap

M

mAbs · monoclonal antibodies
Man · mannose
MIRAGE · minimum information required for a glycomics experiment
MRM · multiple reaction monitoring
MS · mass spectrometry
MS^2 · tandem mass spectrometry
MS^n · multistage mass spectrometry

N

NA · neuraminidase
nano-RP-LC-ESI-IT-MS^n · nano-reversed-phase liquid chromatography coupled to electrospray ionization multistage ion-trap mass spectrometry
nano-RP-LC-ESI-OT-MS^2 · nano-reversed-phase liquid chromatography coupled to electrospray ionization orbitrap tandem mass spectrometry
NCAMs · neural cell adhesion molecules
NCE · normalized collisional energy
NeuAc / NANA · *N*-acetylneuraminic acid
NeuGc / NGNA · *N*-glycolylneuraminic acid
NMR · capillary electrophoresis, liquid chromatography, nuclear magnetic resonance spectroscopy

O

OST · oligosaccharyltransferase
OT · orbitrap

P

PAD · pulsed amperometric detection
PBS · phosphate-buffered saline
PEG · polyethylene glycol
Pep · peptide
PGC-LC · porous graphitized carbon liquid chromatography
PNGase F · endoglycosidase/amidase peptide-*N*-Glycosidase F
Pro · proline
PTM · post-translational modification

Q

qTOF · quadrupole TOF

R

RNase B · ribonuclease B
RP · reversed-phase
RT · room temperature

S

SDS-PAGE · sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser · serine
SPE · *See*

T

TALEN · Transcription activator-like effector nucleases
TFA · trifluoroacetic acid
TFE · 2,2,2-trifluoroethanol
Thr · threonine
TIC · total ion chromatogram
TTBK2 · Tau-tubulin kinase 2

U

UDP-GlcNAc · uridine diphosphate *N*-acetylglucosamine
UPLC · ultra-performance liquid chromatography
UV · ultraviolet
UVPD · ultraviolet photodissociation

V

Val · valine
VNTRs · variable number tandem repeats

X

Xyl · xylose

1 Chapter One

Introduction

In the year 1970 – Ashwell and Morell were the first to identify a hepatic receptor as the key player for the selective clearance of exogenously administered blood plasma glycoproteins from the circulating blood stream of mammals [1]. This finding is now regarded as a milestone in glycobiology research and is remarkable for several reasons: Back then protein glycosylation, meaning the covalent glycosidic linkage of mono-, oligo- or polysaccharides – referred to as glycans – to specific amino acids along the peptide backbone, has been known for quite some time (1865: Eichwald found that mucins are glycosylated proteins) [2]. The biological implications of this protein modification, however, remained largely unknown. The general tenor, well established in the 1920s, was, that specific information is carried mainly by proteins – and protein bound glycans were considered lacking any interesting biological functions. Ashwell and Morell were among the first researchers who could demonstrate that glycans can act as recognition signals. The hepatic receptor they found recognizes specific glycan motifs present on glycoproteins – the initial step for the subsequent degradation of those glycoproteins in the liver [1]. More precisely, that the presence or absence of a single terminal monosaccharide of a glycan, namely sialic acid, determines the half-life of the administered glycoproteins in the circulation. The authors could prove that only those glycoproteins, whose sialic acids have been enzymatically removed prior to the injection, were bound by the receptor and effectively

cleared from the blood stream. The Ashwell-Morell receptor, originally termed asialoglycoprotein receptor, was the first ever described lectin – a protein that selectively binds to the glycan moiety of another protein. At this time, this finding together with the observation that glycans are dramatically altered in cancer cells [3, 4], marked a turning point in the perception of glycans within the scientific community. Glycans or generally spoken carbohydrates could no longer be considered as compounds merely related to structure, energy consumption or energy storage. Glycans rather have become a new member among the four fundamental classes of biomolecules (carbohydrates, proteins, lipids, and nucleic acids).

Since then, knowledge about protein glycosylation has expanded constantly, and soon it was realized that glycans are ubiquitous in nature and can be found in all domains of life (eukarya, bacteria, archaea), including viruses [5, 6]. Glycans can be found on intra- and extracellular proteins. The membrane of every cell in nature features membrane-bound or transmembrane glycoproteins. These glycoproteins along with glycolipids and proteoglycans represent the so-called glycocalyx – the outermost (apical) surface of every cell [7] (Figure 1).

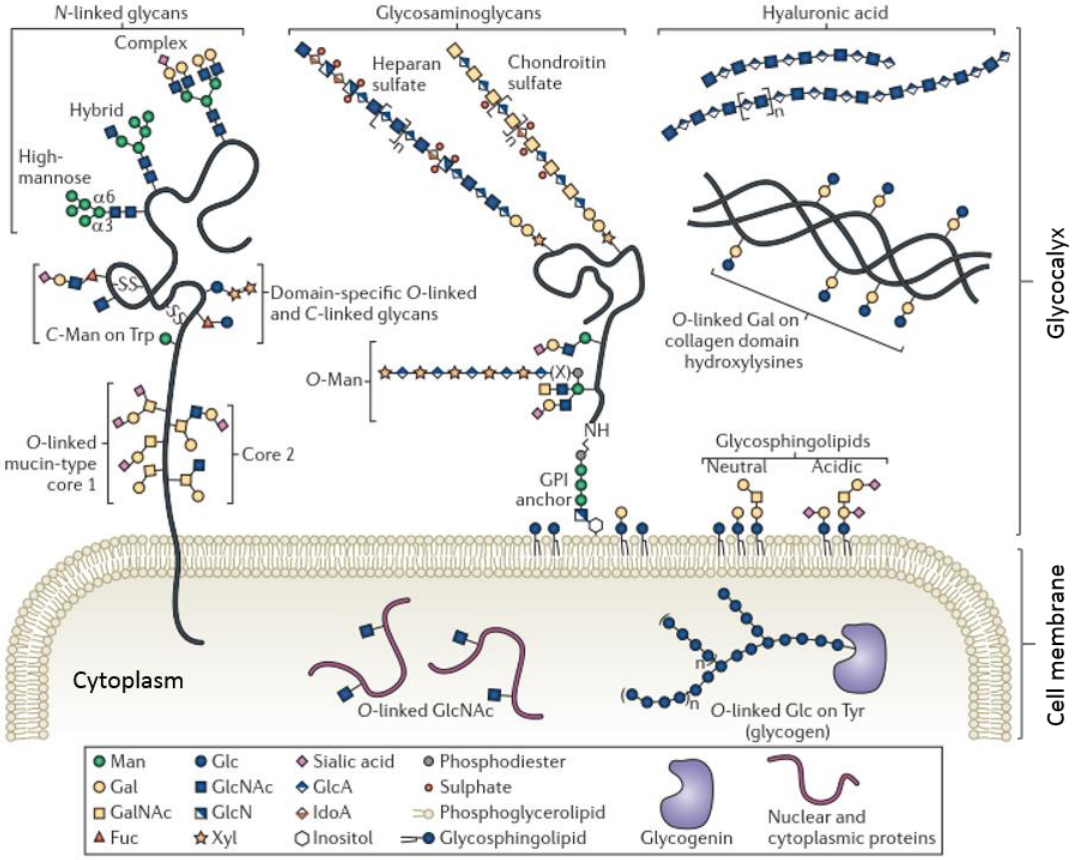


Figure 1: Classes of glycoconjugates present in vertebrates (taken from Moreman *et al.*, 2012 [8]).

From a cell's point of view, the glycocalyx is the first contact zone any external molecule, cell, or organism encounters; making it essential for cell-cell recognition, uptake and secretion, signaling, intercellular adhesion, protection, and stability. Also, most secreted proteins are known or assumed to be glycoproteins, including for instance enzymes, transporters or proteins involved in immune response such as immunoglobulins [7]. Starting in the late 1970s

advances in derivatization as well as chemical or enzymatic release of protein bound glycans, along with the development of nuclear magnetic resonance spectroscopy (^{1}H-NMR), mass spectrometry (MS), liquid chromatography (LC) and capillary electrophoresis (CE), enabled the analysis of the primary structure of glycans derived from single proteins [9, 10].

In contrast to nucleic acids and proteins, glycans were found to be far more complex. While the former have solely linear primary structures, glycosidic linkages between monosaccharides are much more diverse – leading to linear but also branched glycan structures (stereo- and regioisomers: α- and β-anomers, different linkage positions). In addition, glycans were found to be linked to specific amino acids within the protein sequence – a covalent linkage that can be established at the glycosylation sites via different atoms, leading to different glycosylation types (*N*, *O*, *C*, *S*) [11-15]. Glycosylation is a dynamic modification implying differences in glycosylation site occupancy and differences in glycans that can be attached to each potential glycosylation site (macro- and microheterogeneity) – ultimately leading to different protein glycoforms (Figure 2).

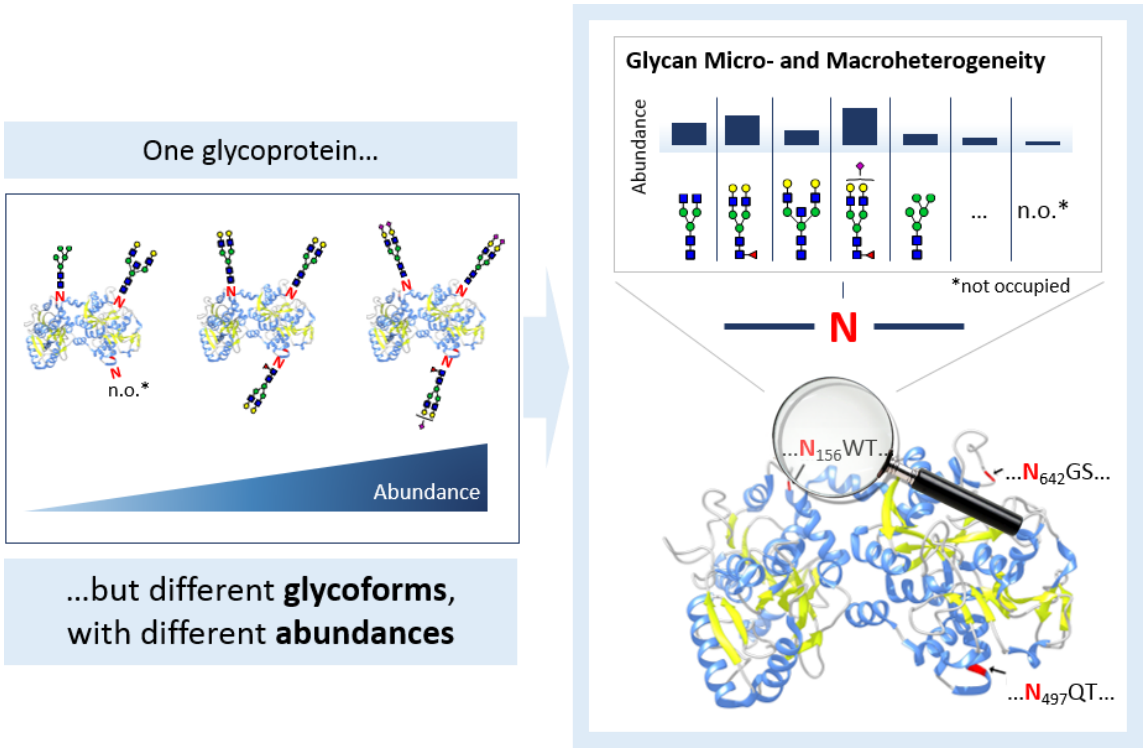


Figure 2: One glycoprotein may exhibit different glycoforms with different abundances, as exemplarily shown for an *N*-glycoprotein with different *N*-glycans being attached to three potential *N*-glycosylation sites. The attachment of different glycans to a potential glycosylation site and the occupancy of this site are referred to as glycan micro- and macroheterogeneity.

This glycan diversity holds an immense modulation and information potential that can significantly influence physico-chemical properties of individual proteins (e.g. size, charge state, hydrophilicity [7]), the interplay between proteins and other biomolecules (e.g. affinity [7]), and the overall phenotype of a cell or an organism (e.g. embryonic development [16]). Glycans, unlike the proteins they modify, are synthesized in a non-template-driven manner.

For mammalian cells, biosynthesis of the majority of known glycosylation types is a co/post-translational event taking place in the endoplasmic reticulum and/or the Golgi apparatus, and is catalyzed by a set of tightly-orchestrated enzymes, involving mainly glycosyltransferases and glycosidases [7]. In mammals, there are two major types of protein glycosylation namely *N*-glycosylation and mucin-type *O*-glycosylation (from now on referred to as *O*-glycosylation). *N*-glycans are linked to the nitrogen on the amino acid asparagine according to a conserved sequence motif Asn-X-Ser/Thr/Cys/Val [17] – where X can be any amino acid, except proline, followed by serine, threonine, cysteine, or valine. Attachment of *O*-glycans does not follow a certain sequence motif, instead *O*-glycans can be linked to the hydroxyl oxygen of virtually every present serine or threonine [7]. Though glycosylation is generally conserved among all domains of life, evolution has led to different repertoires of glycans, lectins, and glycan-modifying enzymes in many cases – evidenced for instance by glycosylation differences between human and bovine cells; or by different human blood groups (AB0 system) [7], being a result of glycan antigens linked to proteins present on the surface of erythrocytes [18]. Protein glycosylation is involved in virtually every cellular process and essential for all living organisms [6]. For many pathogens, for instance, glycan biosynthesis and recognition are crucial for their host specificity, life cycle, and survival [19]. This includes, among others, evasion of the host immune system by glycan mimicry; or the use of glycosidases and lectins to attack the host cell glycocalyx, to adhere to the membrane, and to finally enter the host cell [6]. It is therefore not surprising that several human infectious diseases are associated with altered protein glycosylation, which is either inherited or acquired [20, 21]. Also, for other types of human diseases including cancer, autoimmune diseases, neurological diseases and genetic diseases (e.g. congenital disorders of glycosylation), atypical protein glycosylation can be observed [22]. Analyzing glycans and glycoproteins thus holds an enormous potential to get further insights into a wide range of physiological but also pathophysiological states of a cell, tissue, organ or organism. The gained knowledge can then be used to understand and monitor the onset and progression of diseases, or to develop and improve therapeutics as well as clinical prevention and intervention strategies (e.g. vaccines). Production of several recombinant therapeutic proteins, including biosimilars, has already benefitted from such knowledge. The most prominent examples in this regard are therapeutic monoclonal antibodies (mAbs), antibody fragments, antibody fusion proteins, and antibody-drug conjugates. Besides insulin, recombinantly produced in *Escherichia coli*, mAb production is currently the biggest market driver on the biopharmaceutical industry with a global sales revenue of ~$94 billion in 2017 [23, 24], and an expected revenue of ~$138.6 billion in 2024 [25]. MAbs, due to their target specificity and effector functions, including receptor-blocking and cell cytotoxicity, have a wide spectrum of applications including treatment of different autoimmune diseases, such as Crohn's disease and rheumatoid arthritis, and treatment of different cancer types, such as lung and breast cancer [23]. A major impact on the potency, efficacy and immunogenicity of those therapeutic mAbs can be ascribed to *N*-glycans present in the fragment crystallizable (Fc)-gamma receptor region (constant region of the heavy chains, CH2) [26]. Depending on (I) the

cellular target, (II) the desired effect (e.g. increased cytotoxicity, or anti-inflammation), and (III) the production cell line and process conditions, the Fc-gamma *N*-glycosylation of mAbs needs to be adapted, i.e. glycoengineered, accordingly [27-29]. To reduce the risk of immunogenicity by non-human *N*-glycans, the majority of approved (humanized) mAbs are produced in mammalian cell lines (mainly Chinese hamster ovary cell lines) [30, 31]. However, also other systems for the production of (humanized) antibody fragments and (aglycosylated) antibodies have entered, or are expected to enter, the market including glycoengineered insect, plant, yeast, and bacterial cell lines, or cell-free protein expression systems [32, 33]. Nowadays, glycosylation of recombinant therapeutic proteins, in general, is considered a critical quality attribute, that is inspected by drug regulatory authorities [34-37] and that needs to be accounted for [38, 39].

Even though, glycosylation and glycoengineering have found their way into modern biotechnology and biomedical research, many open questions about protein glycosylation and its implications remain. Many facets of protein glycosylation are still not fully understood, and structure-function relationships identified for one protein or scenario cannot easily be translated to another. To understand glycosylation and its implications better, the constant development and improvement of analytical techniques [40-44] and bioinformatic tools [45, 46] is required. Within the Bio/Process Analytics group of the Bioprocess Engineering group at the Max Planck Institute in Magdeburg, methods and software tools for the high-throughput structural analysis of enzymatically released *N*-glycans derived from single glycoproteins or complex mixtures of glycoproteins have been developed that will be constantly improved and extended [47-53]. These techniques provide valuable qualitative and quantitative insights on the entirety of *N*-glycans of a sample under investigation – the so-called *N*-glycome.

One aim of the present work is to complement those glycome-centered techniques, by techniques that focus on the protein and its glycan moieties as an intact entity – analyzed in the form of glycopeptides – the product of a proteolytic digest of the glycoproteins under investigation. Mass spectrometry, alone or combined with liquid chromatography, is currently the analytical state-of-the-art platform for such glycoproteomic analyses [42, 44, 54]. The analysis of glycopeptides, and their corresponding fragment ion spectra, enables a site-specific mapping and characterization of all glycoforms present on a glycoprotein at a certain point in time. As an addition to glycomics or proteomics approaches, glycoproteomics delivers further insights into the structure-function relationship of glycans and their respective protein carriers in a certain biological context. Establishing such methods is sought to gain a deeper understanding of biological processes that involve glycoproteins (for instance: influenza virus propagation in cell culture).

Within this work, it is intended to establish MS-based glycoproteomic methods and workflows that enable the site-specific glycoproteomic analysis of single glycoproteins but also of complex mixtures of glycoproteins. Thereby, the determination of the micro- and macroheterogeneity of *N*- as well as *O*-glycosylated proteins is sought. The first step towards this objective was to

establish glycoproteomic methods centered on nano-reversed-phase liquid chromatography coupled to electrospray ionization ion-trap multistage mass spectrometry (nano-RP-LC-ESI-IT-MS^n). Additionally, methods to generate and enrich *N*- and *O*-glycopeptides needed to be established. Those methods were then evaluated with well-characterized glycoproteins, such as human immunoglobulin gamma. Ultimately, it was envisioned to apply the developed workflow to study the human blood plasma *O*-glycoproteome in an explorative manner. This project seeked to identify and characterize new *O*-linked glycans on human blood plasma proteins, and therefore to increase our yet limited knowledge about the human blood plasma *O*-glycoproteome.

Crucial to any MS-based glycoproteomic experiment is the unambiguous identification of the peptide moiety of the corresponding glycopeptide. This however cannot always be guaranteed because most MS analyses of intact glycopeptides primarily provide information on the glycan moiety, while sufficient information on the peptide sequence is sometimes difficult to obtain. With the advent of high-resolution MS, equipped with advanced fragmentation techniques, new opportunities for the analysis of glycopeptides arise. In the second part of this thesis it was therefore intended to develop a glycopeptide fragmentation regime on a hybrid linear ion trap/orbitrap high-resolution mass spectrometer, which enables the reliable and unambiguous identification of both, glycan and peptide moiety, for the glycoproteins under investigation. The developed method was validated using a set of selected and representative *N*- and *O*-glycoproteins (human glycoproteins: immunoglobulin gamma, fibrinogen, lactotransferrin; bovine glycoproteins: ribonuclease B).

Glycoproteomics is an emerging scientific field that adopted most of its procedures and approaches from proteomics. However, unlike proteomics, the analysis of glycopeptide spectra is still mainly performed manually. This is primarily due to the lack of reliable and freely available software programs which can cope with the complexity of the acquired datasets, and which provide an all-in-one solution for the automatic identification of glycopeptide fragment ion spectra. Therefore, the third objective of this work was to foster the in-house development of a software program that facilitates the semi-automatic, i.e. manually supervised, annotation and identification of *N*- and *O*-glycopeptide fragment ion spectra. For this purpose, glycoproteomic data, acquired and manually validated within this thesis, were provided along with insights into fragmentation characteristics of *N*- and *O*-glycopeptides. The proposed software not only intends to significantly reduce the hands-on time required for the data analysis; it should also increase the confidence, reproducibility, and the depth of glycoproteomic analyses.

Structure of the thesis

The present thesis is structured into six chapters. Following this first introductory chapter, chapter two outlines the basics of protein glycosylation, its importance and known implications. Further, it gives an overview on state-of-the-art techniques and methods to analyze protein glycosylation (Chapter 2: Theoretical Background). The subsequent chapter

then introduces and describes the analytical workflows that were established or developed during this work (Chapter 3: Material and Methods). Chapters four and five highlight results from two different glycoproteomic projects. Each of the two chapters features an introduction, results, and discussion section. In chapter four the results of an explorative site-specific *O*-glycoproteomic study of human blood plasma glycoproteins using nano-RP-LC-ESI-IT-MS^n are presented (Chapter 4: Human Blood Plasma *O*-Glycoproteomics). Chapter five describes a newly developed workflow for the in-depth analysis of *N*- and *O*-glycoproteins. The workflow is centered on nano-RP-LC-ESI-OT-MS^2 using a hybrid linear ion trap/orbitrap MS. Its capabilities, including the benefits of using diagnostic fragment ions for refined characterization of the glycan moiety, are demonstrated using a set of four selected *N*- and *O*-glycoproteins (Chapter 5: In-Depth *N*- and *O*-Glycoproteomics). The thesis closes with a general conclusion and an outlook (Chapters 6: Conclusion and Outlook).

2 Chapter Two

Theoretical Background

This chapter lays the theoretical foundations for the research presented in this thesis. In the first part, background information on protein glycosylation is given, focusing mainly on mammalian protein *N*- and *O*-glycosylation. This includes structural characteristics, details on the biosynthesis, and known biological implications of *N*- and *O*-glycans. For the latter, emphasis is put on the importance of protein glycosylation for physiological and pathophysiological cellular processes, as well as for biopharmaceuticals. The second part outlines techniques and methods to analyze protein glycosylation. It is subdivided into: (I) the analysis of monosaccharides (II) the analysis of intact glycoproteins, (III) the analysis of glycans that were released from glycoproteins, (IV) and the analysis of glycopeptides derived from glycoproteins. Central to this thesis is the MS analysis of glycopeptides. This topic is thus given a particular focus throughout the second part of this chapter.

2.1. Protein Glycosylation

The covalent attachment of mono-, oligo- or polysaccharides to the polypeptide chain of proteins is termed glycosylation – a process ubiquitously found among all domains of life (eukarya, bacteria, archaea), including viruses. About 50% of all proteins are believed to be

glycosylated, making protein glycosylation the most common co/-posttranslational modification of proteins [55]. During protein glycosylation a covalent bond, termed glycosidic bond, is established between the side chain group of distinct amino acids and a carbohydrate. The resulting compound, termed glycoprotein, thus holds a carbohydrate moiety – the glycon – and a non-glycosylated peptide backbone – the aglycon. The carbohydrates, termed glycans, are complex biomolecules, whose structural diversity exceeds that of nucleic acids and proteins by far. While nucleic acids and proteins adopt linear primary structures, glycans can exhibit linear but also branched structures. The building blocks of glycans, but also of other important biomolecules, are monosaccharides (Figure 3).

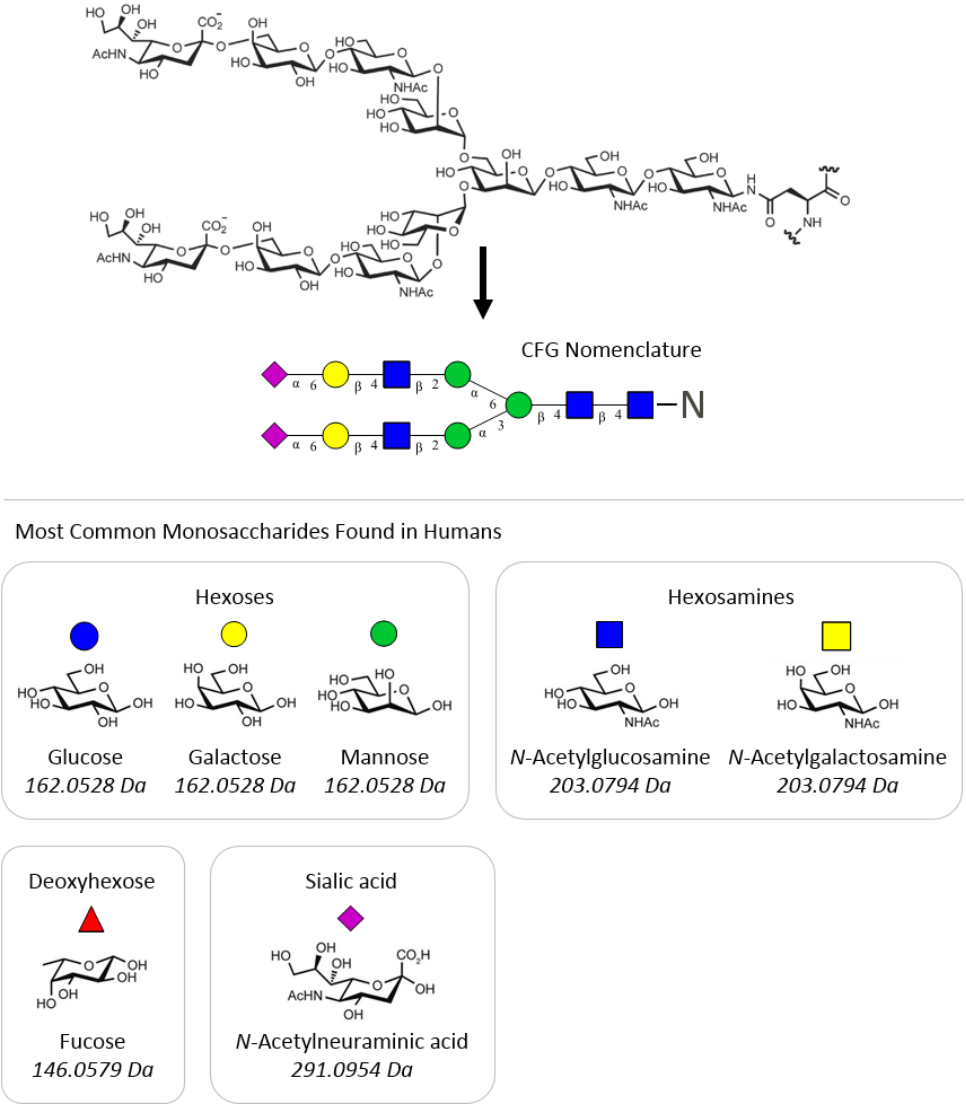


Figure 3: Graphical representation of glycan structures (*top*) and monosaccharides (*bottom*) according to the CFG nomenclature (Consortium for Functional Glycomics) [7]. Theoretical residue masses are shown in italics (modern MS instruments typically report up to four decimal places).

Naturally occurring monosaccharides can contain between three to seven carbon atoms. The most common constituent of glycans, however, are monosaccharides with six carbon atoms i.e. hexoses. Depending on the terminal group monosaccharides can be classified into aldoses

(characterized by aldehyde group) and ketoses (characterized by ketone group). Monosaccharides, or carbohydrates in general, can occur in a large variety of isomeric forms. A hexose, for instance, has four chiral centers – that is, four carbon atoms whose substituents can be arranged in two stereochemically different ways. Hence, there are in total 16 hexose isomers. The three hexoses glucose, galactose, and mannose, for instance, differ only in the stereochemical orientation of the hydroxyl group of the carbon atoms C-4 and C-2, respectively. Such stereoisomers are termed epimers, as they differ in the configuration of only one of their chiral centers. Epimers belong to the diastereomers, a subgroup of stereoisomers. The second subgroup of stereoisomers is enantiomers. A hexose, like glucose, has two enantiomers, i.e. two mirror configurations (or optical isomers): D-glucose and L-glucose. In glycans, hexoses are normally found in the D-configuration. It is therefore common to make this state implicit when describing glycans. Monosaccharides in glycans primarily exist in a cyclic form. The most common, and chemically most stable forms are five- and six-membered rings – termed furanoses and pyranoses, respectively. The cyclic configuration is established by a reaction between the C-5 hydroxyl group and either the C-1 aldehyde group present in aldoses, or the C-2 ketone group present in ketoses – eventually resulting in a hemiacetal or hemiketal, respectively. With this ring formation a new chiral center is built at position C-1. The stereoisomeric orientation of the hydroxyl group attached to this so-called anomeric carbon atom leads to two diastereomers: the α-anomer and the β-anomer. For the cyclic form of glucose, for instance, this results in four stereoisomers: α-D-glucose, α-L-glucose, β-D-glucose, and β-L-glucose. In glycans of vertebrates the most commonly found monosaccharides are D-glucose, and epimers of glucose, namely D-mannose and D-galactose. The repertoire of monosaccharides is further expanded by substitution of the C-2 hydroxyl group of glucose or galactose with an acetylated amino group. The resulting monosaccharides are *N*-acetyl-D-glucosamine or *N*-acetyl-D-galactosamine, respectively – two hexosamines. Two other commonly found monosaccharides of vertebrate glycans are L-fucose and *N*-acetylneuraminic acid. L-fucose is structurally related to galactose. Compared to D-galactose, though, fucose lacks a hydroxyl group at position C-6, and adopts an L- rather than a D-configuration. Due to the lack of the hydroxyl group, fucose is classified as a deoxyhexose. *N*-acetylneuraminic acid belongs to the large group of sialic acids (sometimes also referred to as neuraminic acids). Up to now over 50 naturally occurring derivatives of sialic acids were identified [56]. In humans *N*-acetylneuraminic acid (NeuAc or NANA) is the most commonly found sialic acid. In mammals other than human, sialic acids such as *N*-glycolylneuraminic acid (NeuGc or NGNA) and 2-keto-3-deoxynononic acid (KDN) can prevail. Sialic acids are characterized by a nine-carbon backbone and a C-1 carboxyl group attached to the anomeric C-2 carbon. The carboxyl group is deprotonated at physiological pH (pKa of 2.6) and confers a negative charge to sialic acids. Further diversity of sialic acid arises from naturally occurring substitutions of the hydroxyl groups with functional groups such as *O*-acetyl or *O*-sulfate groups.

Multimeric glycan structures are built by the covalent linkage of monosaccharides with each other. Such a linkage is established by a condensation reaction between the hemiacetal or

hemiketal group of a monosaccharide and a hydroxyl group of another monosaccharide. During this reaction the hemiacetal or hemiketal is converted into an acetal or ketal, respectively. The established bond is termed glycosidic bond. The disaccharide lactose, for instance, is formed through a glycosidic bond between the C1 of galactose in β-configuration and the 4-hydroxyl group of glucose. The resulting linkage is termed β1-4 linkage. During the formation of lactose, the glucose unit retains its hemiacetal group, i.e. its free anomeric carbon. The ability of this group to reduce inorganic ions such as Cu^{2+} leads to the designation reducing end of the disaccharide. Consequently, the galactose unit, constitutes the non-reducing end. The terms reducing end and non-reducing end, are not only used with disaccharides. Also, polysaccharides or glycans feature a reducing end, along with a single non-reducing end or with multiple non-reducing ends – depending on the presence of branched structures. However, there are also exceptions: for instance, the disaccharides sucrose and trehalose, and the polysaccharide starch are non-reducing sugars as they lack a free anomeric carbon. The multitude of possible glycosidic linkages that can be established between monosaccharides yields numerous possible carbohydrate structures. This complexity can be illustrated by comparing the disaccharides maltose and trehalose. Both disaccharides are composed of two glucose units, meaning they are identical with respect to the monosaccharide composition. Those two glucose units, however, are linked on different positions. While in trehalose the two glucose units are linked C1 to C1 with both links in the alpha configuration (glucose α(1,1)α glucose), in maltose the glycosidic bond is established between C1 and C4, again in alpha configuration (glucose α(1,4) glucose). This subtle structural difference results in two individual compounds, each with unique biochemical and biological properties: Trehalose is a non-reducing sugar that can by synthesized by bacteria, fungi, plants, and invertebrates. It serves as an energy source, but can also confer cellular protection against freezing and drying. Maltose, in contrast, is a reducing sugar. The plant-derived disaccharide is an intermediate of the starch catabolism and solely serves as a source of energy.

Comparing the two structural isomers trehalose and maltose stands representative for a general concept in carbohydrate biology: the identity of oligo- and polysaccharides, including glycans, is defined by the (I) monosaccharide composition (e.g. D-glucose + D-glucose), the (II) anomeric configuration (i.e. α or β), and (III) the connectivity (i.e. linkage positions) of each glycosidic bond. With these degrees of freedoms carbohydrates hold an enormous structural diversity – a diversity which far exceeds that of genes and peptides. As an example, a disaccharide composed of two D-glucose units has 11 potential isomers (regio- and stereoisomers) [57]; a linear trisaccharide composed of three D-glucose units already has 176 potential isomers. In contrast, a di- or tripeptide composed of just one amino acid (e.g. AA and AAA) has only one isomer each [58]. Thus, considering not only linear glycans, but also branched ones, the number of possible glycan structures seems nearly unlimited. Still, and despite we currently have only limited insights into the entirety of all naturally occurring glycans, it becomes clear that not all theoretically possible glycans are present in nature.

Considering only vertebrates, in fact, many evolutionary conserved glycan structures can be found across all taxa (there are nine monosaccharides commonly found vertebrate glycoconjugates) [7].

Glycobiology is a growing scientific field that was pioneered by chemists. To facilitate access to this new research field to scientists without a strong chemical background, a simplified graphical representation of glycans and oligosaccharides has been introduced. This condensed form of representation makes use of geometrical shapes to depict monosaccharides, it thus allows to interpret glycans at a glance and serves to understand glycans as biological entities rather than complex chemical structures (Figure 3). Depending on the depth of the glycan analysis the graphical representation of the glycans analyzed can be adapted accordingly. If, for instance, no information on the glycan linkages can be gained from the conducted measurement, the glycan is typically depicted without any linkage information. Yet, the monosaccharide identity (e.g. D-mannose) and the general topology of the glycans, meaning the glycan sequence without any linkage information (e.g. D-galactose + *N*-Acetyl-D-glucosamine) but including potential branching points, is often given anyways – even if the conducted analysis does not give enough evidences to derive such structural information. This assumption is usually made whenever the performed analysis yields only an indirect prove of the glycan identity – such as a chromatographic glycan retention time matching a validated database entry, or a glycan mass with corresponding fragment ions that matches a certain glycan structure. In any case this assumption can only be made for glycans whose structure has already been elucidated before. Again, it reflects the concept of considering glycans as entire biological entities governed by conserved biosynthesis pathways. Still remaining uncertainties during the annotation of glycan structures, such as terminal monosaccharides that can reside on either of two or more arms of a branched glycan structure, are handled by depicting the potential locations of those terminal monosaccharides with parenthesis. One of the most commonly used nomenclatures for the symbolic representation of glycans is the CFG nomenclature introduced in 1999 in the textbook "Essentials of Glycobiology, 1st edition" and revisited by the Consortium for Functional Glycomics (CFG). This nomenclature is inspired by monosaccharide symbols already introduced in 1978 by Kornfeld *et al*. [59]. In addition to the symbolic representation of monosaccharides, corresponding abbreviations were introduced: for instance, D-glucose is abbreviated with Glc. Notably, the designation of monosaccharide D- or L-enantiomer is usually made implicit with respect to the most common form, meaning that the term glucose usually referrers to D-glucose. Moreover, generic terms for a certain class of monosaccharides can be used: the term hexose (abbreviated wit Hex), for instance, combines the isomers glucose, galactose, and mannose. The use of these abbreviations and terms was recommended by the International Union of Pure and Applied Chemistry (IUPAC) in 1996 (http://www.sbcs.qmul.ac.uk/iupac/2carb/). The CFG nomenclature has received several iterations of renewal. The latest revision was introduced in 2015 [7]. Another widely used glycan nomenclature is the oxford-notation proposed by Harvey *et al*. in 2009 [60].

Throughout this thesis the revised CFG nomenclature introduced in "Essentials of Glycobiology, 2nd edition" is used [61] (Figure 3).

_N_-Glycosylation

There are several forms of protein glycosylation – differing in the employed monosaccharides, the attachment sites within the protein backbone, their biosynthesis as well as their biological implications. The most commonly found and hitherto best studied form is protein *N*-glycosylation (sometimes also termed *N*-linked glycosylation). *N*-glycosylation is conserved among all three domains of life, including viruses. As with all other known forms of protein glycosylation, the biosynthesis of *N*-glycans is catalyzed by a set of tightly-orchestrated enzymes. Unlike the biosynthesis of nucleic acids or peptides, glycosylation is a non-template driven process that relies on the coordinated interplay and availability of enzymes, glycosyl donors and acceptor substrates. Mainly involved are enzymes that catalyze the addition or removal of monosaccharides – so-called glycosyltransferases and glycosidases. Initially those enzymes were thought to have a strict donor and acceptor specificity, meaning that each glycosyltransferase can only add one specific monosaccharide in a specific linkage; e.g. the enzyme β1,4-galactosyltransferase only catalyzes the β-linked glycosidic bond created between C1 of D-galactose and C4 of D-glucose to build lactose (the same holding true for the hydrolysis of glycosidic bonds by glycosidases). This led to the "one enzyme-one linkage" hypothesis. This theory, however, was refuted: It is now known that (I) more than one enzyme can catalyze the formation or hydrolysis of a certain glycosidic bond; and that, in rare cases, (II) one enzyme can catalyze the formation of multiple types of glycosidic bonds. For example, the human fucosyltransferases III–VII, all attach fucose in α1-3 linkage to *N*-acetyllactosamine moieties on glycans. Fucosyltransferase III, however, can additionally also attach fucose in α1-4 linkage [7]. To establish a glycosidic bond between a monosaccharide and its acceptor, i.e. a nascent protein, or another monosaccharide, the activated form of this monosaccharide must be provided. Energetically-activated monosaccharides are referred to as nucleotide sugars, and include for instance uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc) and guanosine diphosphate mannose (GDP-Man). Pivotal to the biosynthesis of nucleotide sugars is the glucose metabolism, as glucose can be converted into all other monosaccharides. In some instances, the monosaccharide donor substrate can also feature a lipid moiety, such as dolichol-phosphate (Dol-P). This lipid moiety can be linked to mannose or glucose, to form dolichol-phosphate mannose and dolichol-phosphate glucose, respectively (Dol-P-Man, Dol-P-Glc).

In eukaryotes, two cellular compartments are involved in protein *N*-glycosylation: the endoplasmic reticulum (ER) and the Golgi apparatus (or simply Golgi). The biosynthesis of *N*-glycans is initiated at the cytosolic side of the ER membrane. During these steps nucleotide sugars, namely UDP-GlcNAc and GDP-Man are sequentially added to the lipid carrier Dol-P to build a lipid-linked *N*-glycan precursor. This process is facilitated by a set of glycosyl-transferases encoded by genes designated as ALG (asparagine-linked glycosylation). The first monosaccharides added are two GlcNAc residues. They represent the so-called chitobiose core

of the *N*-glycan. Further, five Man residues are added to this chitobiose core. The resulting structure is referred to as Man_5 *N*-glycan (similarly, an *N*-glycan with three Man residues is referred to as Man_3 *N*-glycan). The first three mannoses added represent the so-called trimannosyl core. Together with the chitobiose core, the trimannosyl core constitute a conserved structure present in all conventional *N*-glycan types (unconventional paucimannose-type *N*-glycans are an exception in this regard). The biosynthesis continues at the luminal side of the ER membrane. To this end, the *N*-glycan precursor is translocated across the membrane into the inside of the ER – a process that is still not fully understood. There, the *N*-glycan precursor is further extended by four Man residues and three Glc residues. The necessary monosaccharides are provided in the form of cytosolically synthesized Dol-P-Man and Dol-P-Glc. The preassembled *N*-glycan precursor $Glc_3Man_9GlcNAc_2$-P-P-Dol is then transferred to the unfolded nascent target protein. This process is catalyzed by oligosaccharyltransferase complexes (OST), which are part of the ER protein translocon (in metazoans there are two distinct OSTs with partially non-overlapping roles [62]). The addition of the *N*-glycan precursor to the protein acceptor site, by the OSTs, is enabled mainly by an N- to C-terminal sequence scanning mechanism, and can occur either co- or posttranlationally [62, 63]. The acceptor site for *N*-glycans is the amino acid asparagine; or more precisely the nitrogen atom of the amide group of the asparagine side-chain. A prerequisite to add an *N*-glycan to this glycosylation site is the presence of a conserved sequence motif, a so-called sequon. Typically, *N*-glycosylation occurs on the Asn-X-Ser/Thr sequon (canonical sequence motif); where X can be any amino acid except for proline; note that Thr is more common than Ser (throughout this thesis amino acids are represented by their respective one-letter or three-letter symbol, as recommended by the IUPAC [64]). In rare cases, however, *N*-glycosylation has also been observed on a Asn-X-Cys or Asn-X-Val sequon [17]. Still, the mere presence of an *N*-glycosylation sequon within a protein sequence does not necessarily mean that this site becomes *N*-glycosylated at the end. There are a few other requirements or constraints that also have to be considered: (I) only proteins that traverse the secretory pathway can become *N*-glycosylated, which only applies to secreted and membrane-bound proteins; (II) OSTs require an appropriate three-dimensional structure of the target protein to access potential *N*-glycosylation sites; (III) the presence of certain amino acids, such as proline, in the proximity of a sequon can reduce the *N*-glycosylation efficiency; (IV) the same holds true for other types of glycosylation in the proximity of an *N*-glycosylation sequon; (V) *N*-glycosylation sites close to the protein's C-terminus become less efficiently glycosylated; (VI) among the two main sequons, the Asn-X-Thr sequon becomes more efficiently *N*-glycosylated (Figure 4); (VII) and the *N*-glycosylation efficiency is impaired by the progression of protein folding [65, 66]. Once the preassembled *N*-glycan precursor is covalently linked to its acceptor site, the final steps of the protein folding will be initiated, accompanied by further *N*-glycan processing. During these steps the three terminal Glc residues of the *N*-glycan will be sequentially removed by means of two glycosidases. The removal, or if necessary, the re-addition of these Glc residues is part of the ER's protein folding quality control. This cyclic process is guided by two chaperones, calnexin and calreticulin, along with

several lectins, i.e. glycan-binding proteins. In the end, only correctly folded *N*-glycoproteins, i.e. *N*-glycoproteins whose *N*-glycan Glc residues were completely removed during the folding process (resulting in $Man_9GlcNAc_2$ *N*-glycan structures), will be transported further for secretion or integration into the cell membrane. Irrecoverably misfolded *N*-glycoproteins will be deglycosylated and degraded in the proteasome.

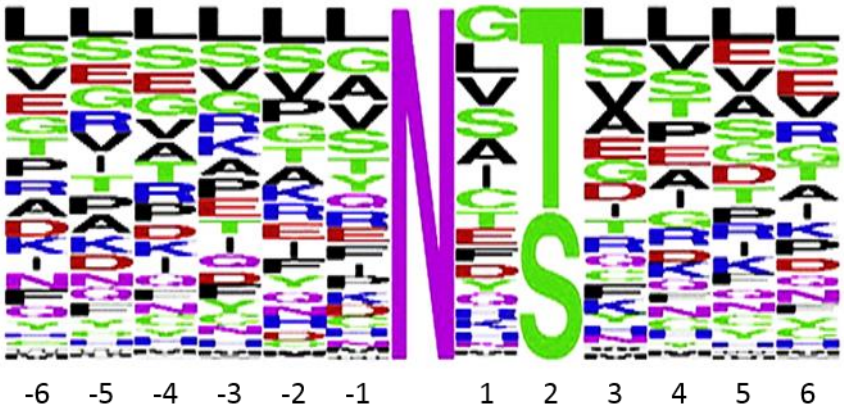


Figure 4: Frequency of occurrence of amino acids in the vicinity of *N*-glycosylation sites (sequence logo). Sequence logo generated by MS-based mapping of 6367 *N*-glycosylation sites (canonical recognition motif) on 2352 glycoproteins derived from mouse tissue and human blood plasma [Zielinska *et al.*, 2010 [17]].

Before the correctly folded *N*-glycoprotein traverses to the Golgi, one terminal Man residue is removed from the *N*-glycan by a mannosidase (resulting in a $Man_8GlcNAc_2$ *N*-glycan). In the Golgi a substantial remodeling of the *N*-glycan takes place while the protein makes its way from the *cis*- to the *trans*-Golgi. This *N*-glycan remodeling aims at generating a mature *N*-glycan structure by trimming and addition of monosaccharides – a process that is mediated by glycosidases, glycosyltransferases, and lectins. Thereby, monosaccharides added to glycans within the Golgi are generally provided in the form of nucleotide sugars. While in the ER, processes are streamlined to generate a single universal *N*-glycan precursor that can be transferred *en bloc* to dedicated protein acceptor sites, processes running in the Golgi are much less stringent and more diverse – a diversity that is reflected by the plethora of possible mature *N*-glycans that can be generated within this cellular compartment. Without going into the detail of their biosynthesis, there are three major, i.e. conventional *N*-glycan types that can be built during *N*-glycan processing in the Golgi – all having the chitobiose and trimannosyl core in common (Figure 5).

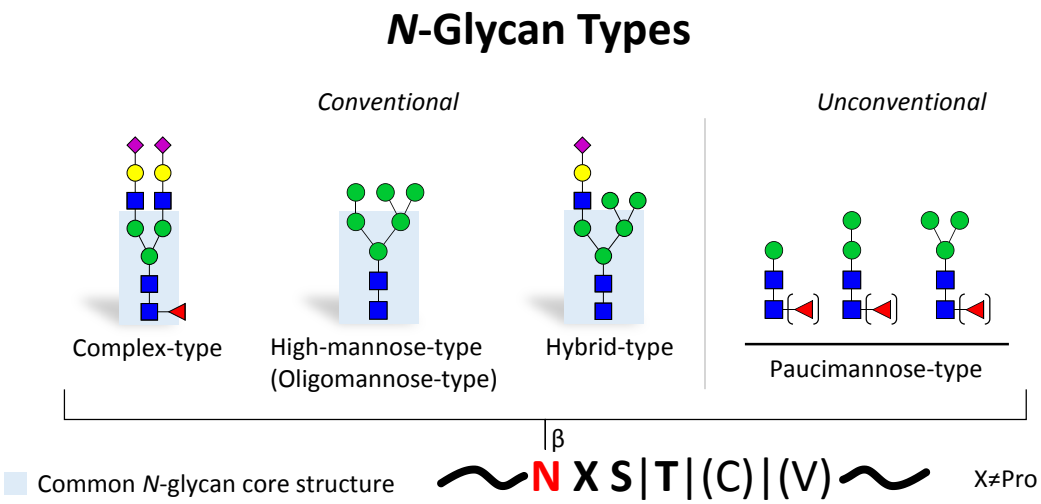


Figure 5: *N*-glycan types present in mammals. The three major *N*-glycan types are the complex-, high-mannose- and hybrid-type. The paucimannose-type is less frequently observed and therefore referred to as unconventional.

First, there are high-mannose-type *N*-glycans (also known as oligomannose *N*-glycans): those *N*-glycans feature two to six mannoses being attached to the trimannosyl core (Figure 5). The resulting *N*-glycans are often abbreviated as Man_5, Man_6, Man_7 etc. *N*-glycans, with the total number of mannose residues being expressed as subscript number. Second, there are complex-type *N*-glycans: those *N*-glycans are characterized by the sequential addition of different non-mannose monosaccharides to the mannose residues of the trimannosyl core, i.e. the α1-3 or α1-6 arm of the trimannosyl core (Figure 5 ,Figure 6).

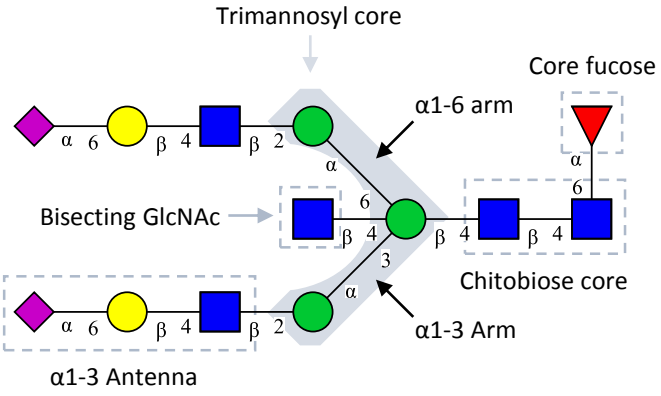


Figure 6: Exemplary *N*-glycan structure found in humans. Common structural elements, such as bisecting GlcNAc or core fucose are highlighted.

The linear substructures added to those mannose residues are termed antennae, or branches. Predominantly there are complex-type *N*-glycans with two antennae, so-called di-antennary *N*-glycans. However, mono- tri- and tetra-antennary *N*-glycans are also common. Complex-type *N*-glycans with more than four antennae have only been rarely documented [7]. The biosynthesis of an *N*-glycan antenna is initiated by addition of a GlcNAc (β1-2) residue to the trimannosyl core. The antennae can be further elongated by attachment of a galactose (Gal,

β1-4) residue to the GlcNAc residue, leading to a so-called *N*-acetyllactosamine residue (LacNAc). LacNAc residues are typically terminated by addition of a sialic acid to the Gal residue, as sialic acids are naturally occurring capping structures that cannot be further elongated (in humans α2-3- or α2-6-linked NeuAc). The only exception are so-called polysialic acid-containing *N*-glycans, where long chains of α2-8-linked NeuAc residues are linked to the terminal α2-3-linked NeuAc residue of the *N*-glycan antennae [67]. This type of sialylation, however, is limited to neural cell adhesion molecules (NCAMs) and very few other glycoproteins. Aside from this exception, the described GlcNAc-Gal-Sia motif – in full length or reduced (e.g. GlcNAc-Gal) – is the most common antenna motif found in complex-type *N*-glycans derived from mammalian glycoproteins [7]. Other motifs may include additional antenna extension and/or modification. One prominent example is the extension of the antenna LacNAc residue by LacNAc tandem repeats, i.e. addition of GlcNAc (β1-3) and Gal (typically β1-4, but also β1-3 possible) disaccharide residues. Such poly-LacNAc extensions preferentially occur on multiantennary *N*-glycans – particularly on the β1-6-linked GlcNAc-branch. Less frequently found is a GalNAc (β1-4) residue, instead of a Gal (β1-4) residue, being linked to the trimannosyl core GlcNAc, resulting in a so-called LacdiNAc (or LDN) residue. Similar to LacNAc extensions, also tandem repeats of LacdiNAc repeats, can extend an *N*-glycan antenna [68]. LacNAc and LacdiNAc residues can be further modified in various positions and linkages by additional residues such as Gal, GlcNAc, GalNAc, fucose (Fuc), sialic acid, glucuronic acid (GlcA), or sulfate. In this context poly-LacNAc and poly-LacdiNAc chains are thought to serve as linear extended scaffolds for the presentation of those terminal residues. The best-studied LacNAc motifs are the human blood group antigens. The presence of these antigens on secreted glycoproteins and glycolipids, on free glycans (i.e. oligosaccharides, such as milk oligosaccharides), and on glycoproteins and glycolipids populating the surface of erythrocytes as well as many epithelial and endothelial cells represents the basis of our entire transfusion, transplantation, and newborn medicine. The agglutinating effect caused by the immunological reaction of circulating antibodies against blood group antigens present on erythrocyte glycoproteins of incompatible donor blood was already described in the early 20th century by Landsteiner *et al.* (with this finding, the AB0 blood group classification was introduced, also known as ABO or ABH blood group system) [18]. However, it took more than fifty years to be able to elucidate the exact molecular structures of these blood group antigens [27] (Figure 7).

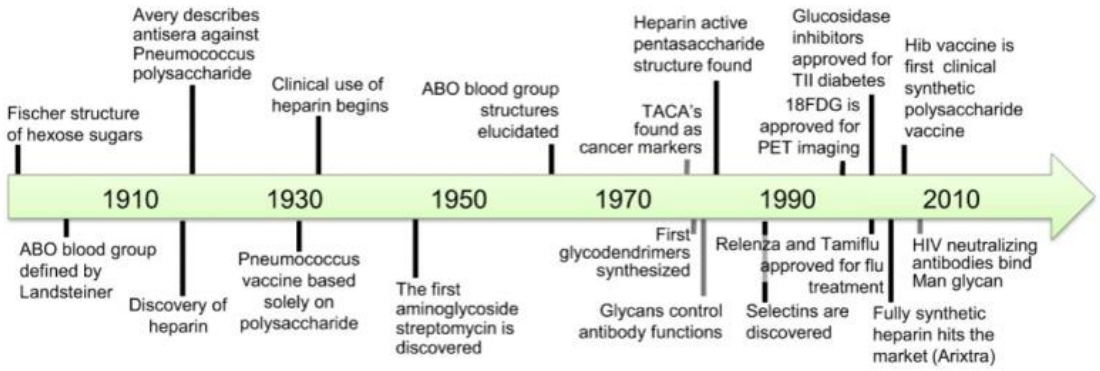


Figure 7: History of glycans in medicine: discoveries and milestones (taken from Hudak *et al.*, 2014 [27]).

Besides AB0 blood group antigens also other blood group antigens were identified on glycoproteins in the past. Among those are the "I" and "i" blood group antigens, the Sd^a blood group antigens, and the Lewis blood group antigens [69]. Another important LacNAc motif is the alpha-Gal antigen. This antigen is characterized by the addition of a Gal in α1-3-linkage to the β1-4-linked Gal of a LacNAc residue (Figure 8).

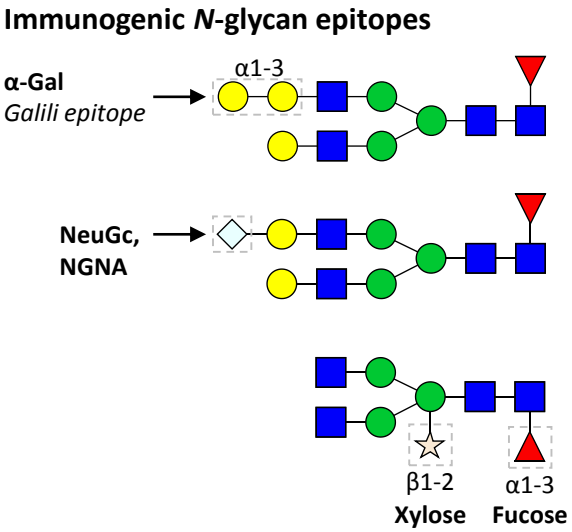


Figure 8: *N*-glycan epitopes causing immunogenic reactions in humans (modified from Jones, 2017 [70]).

The glycosyltransferase responsible for this reaction is expressed in New World primates and many non-primate mammals, but it is not expressed in humans and Old World primates. In humans the alpha-Gal antigen is immunogenic – similar to NeuGc and the plant-specific xylose and α1-3-fucose epitopes [71]. This circumstance, in turn, is crucial for dietary-related allergic reactions, xenotransplantation, as well as administration of biopharmaceuticals produced in non-human expression systems [26, 72]. LacNAc but also LacdiNAc residues can also be modified by addition of a sulfate group linked to the Gal, GalNAc, GlcA, or GlcNAc residue [73, 74]. A common modification of complex-type *N*-glycans is the addition of a fucose to the innermost and Asn-linked GlcNAc residue of the *N*-glycan chitobiose core (Figure 6). This fucose residue is referred to as core fucose – in contrast, fucoses attached to *N*-glycan antennae, as for example

with blood group antigens, are generally referred to as antenna fucoses. In vertebrates, *N*-glycan core fucoses are α1-6-linked; in plants, fungi and invertebrates core fucoses can also be linked in α1-3 position. Normally, core and antenna fucoses cannot be elongated. Another common modification of complex-type *N*-glycans is the addition of a GlcNAc residue in-between the α1-3 and α1-6 arm of the trimannosyl core. The GlcNAc residue is attached in β1-4 position and referred to as bisecting GlcNAc. A bisecting GlcNAc cannot be further elongated. In plants a xylose (Xyl; β1-2-linked), instead of a GlcNAc, can be linked to the β1-4-linked Man of the trimannosyl core. The third major *N*-glycan type is hybrid-type *N*-glycans: those *N*-glycans aggregate structural features of both complex-type and high-mannose-type *N*-glycans (Figure 5). Similar to complex-type *N*-glycans, hybrid-type *N*-glycans can carry core fucoses and/or a bisecting GlcNAc. Also, their antennae can be composed and modified in the same way as for complex-type *N*-glycans. However, hybrid-type *N*-glycans feature only one antenna – the one linked to α1-3 Man arm of the trimannosyl core. The remaining part of the hybrid-type *N*-glycan (α1-6 Man arm) resembles the high-mannose-type.

Mucin-type O-Glycans

Another form of protein glycosylation that is very common, and that has been studied intensively, is mucin-type *O*-glycosylation; also known as *O*-GalNAc-glycosylation, or simply *O*-glycosylation. The latter, however, is a generic term, that, strictly spoken, covers also other existing forms of protein *O*-glycosylation such as *O*-mannosylation, *O*-fucosylation, and *O*-GlcNAcylation [75]. Mucin-type *O*-glycosylation is very common among the animal kingdom, but seems to be absent in bacteria, yeast, and plants [7]. The biosynthesis of mucin-type *O*-glycans (for simplicity in this thesis referred to as *O*-glycans/*O*-glycosylation) differs significantly from that of *N*-glycans. Unlike, *N*-glycans, *O*-glycans are synthesized exclusively in the Golgi. This brings along structural differences, including available monosaccharides, as well as differences in the glycan acceptor sites. In contrast to *N*-glycans, the monosaccharide repertoire of *O*-glycans does not include Man, Glc, and Xyl. Moreover, *O*-glycans do not share one single common core structure. Instead, there are at least eight different core structures, of which the core structures 1-4 are most commonly observed (Figure 9).

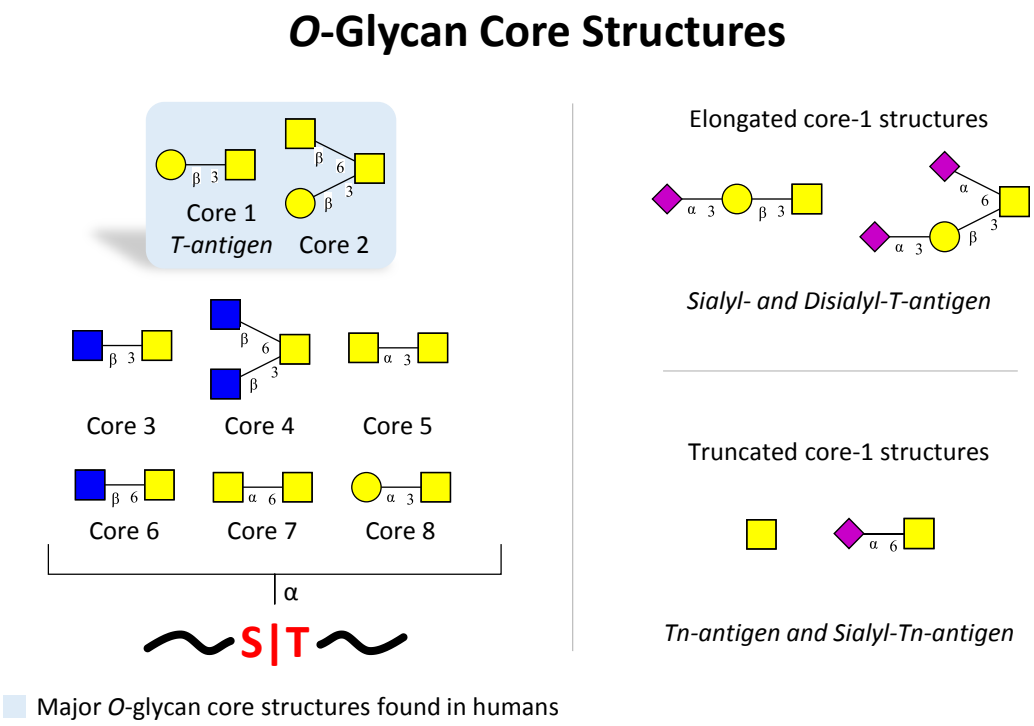


Figure 9: *O*-glycan core structures (*left*) and common core-1 antigen structures (*right*) present in humans.

Like *N*-glycans, *O*-glycans can adopt linear as well as branched structures, ranging in their length from one monosaccharide to more than 20. Common to all *O*-glycans is the α-linked GalNAc residue all structures are initiated with. This GalNAc residue is linked to the hydroxyl group oxygen atom of either the Ser or Thr side-chain. However, in contrast to *N*-glycans, the biosynthesis of *O*-glycans does not include the *en bloc* transfer of a preassembled immature glycan core structure. Instead the initial GalNAc residue is directly transferred to the respective acceptor sites, in the form of UDP-GalNAc. In humans this step can be executed by more than 20 different GalNAc-transferases (GalNAc-T isoforms) – each with different, but partly overlapping, substrate specificities, as well as tissue-dependent activities [76]. Strikingly, *O*-glycans will be attached solely posttranslational, i.e. to the already fully synthesized and folded proteins. Thereby the attachment of *O*-glycans to the peptide chain relies on the accessibility of the protein surface, but does not rely on a conserved consensus motif, i.e. virtually every available Ser or Thr can serve as potential *O*-glycan acceptor site (attachment to Tyr, hydroxylysine, or hydroxyproline is also possible, but has rarely been observed). However, *in vitro* assays, testing the GalNAc-T substrate specificities, revealed that Thr seems to be the preferred acceptor site (Figure 10).

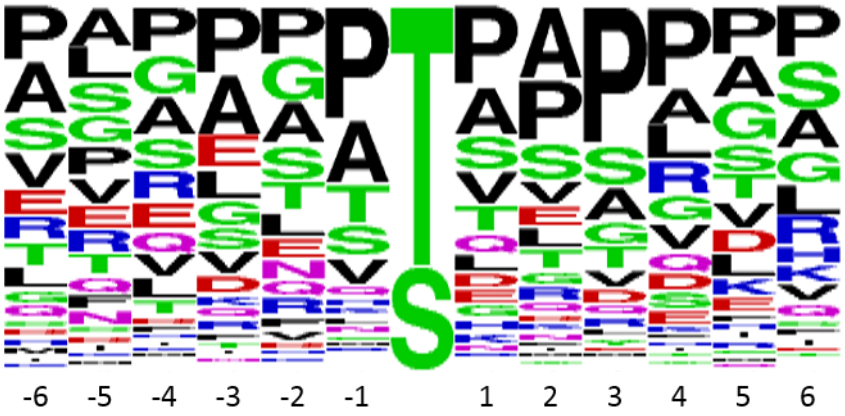


Figure 10: Frequency of occurrence of amino acids in the vicinity of *O*-glycosylation sites (sequence logo). Sequence logo generated by MS-based mapping of 124 *O*-glycosylation sites on 51 glycoproteins derived from bovine serum (Darula et al., 2012 [77]).

Despite there is no conserved consensus motif for *O*-glycan attachment, *O*-glycans are usually present in repeating domains rich in Pro, Thr, and Ser – so-called PTS domains, also known as "variable number tandem repeats" (VNTRs). Within those PTS domains, *O*-glycans typically exist as clusters adopting a bottle brush-like conformation. Upon attachment of the initial GalNAc residue, *O*-glycans can be elongated by other monosaccharides provided in the form of nucleotide sugars. The glycosyltransferases responsible for those reactions are, in parts, also involved in the *N*-glycan biosynthesis [7]. Consequently, linear and branched chains added to the different *O*-glycan core structures resembles those also found on *N*-glycan antennae or glycolipids. Thus, similar to *N*-glycans, *O*-glycans may also feature for instance LacNAc and LacdiNAc residues, as well as fucoses and terminal sialic acids. And with this, *O*-glycans may also exhibit antigens such as AB0 and Lewis blood group antigens. Among the four major *O*-glycan core structures, the core structures 1 and 2 are most ubiquitously found in humans, as they are present on many glycoproteins produced in many different cell types. The core structures 3 and 4, in contrast, are mainly present on proteins expressed in gastrointestinal and bronchial tissues. *O*-glycan biosynthesis is a complex process involving many tightly-regulated enzymes. Dysregulation of the *O*-glycan biosynthesis, for instance in tumor cells, can lead to incomplete elongation of *O*-glycans. Two commonly found tumor-associated *O*-glycan antigens are the Tn-antigen and the Thomsen-Friedenreich antigen (the latter is also known as TF- or T-antigen) (Figure 9). The Tn-antigen is formed by a single non-elongated GalNAc residue linked to either Ser or Thr, while the T-antigen is formed by an unsubstituted core-1 *O*-glycan (Ser/Thr-$GalNAc_1Gal_1$). Both antigens can also feature sialic acids, resulting in sialyl-Tn or sialyl-T-antigens, respectively.

2.2. Biological Implications of *N*- and *O*-Glycosylation

Protein glycosylation is a universal characteristic that can be found in all living cells and that is essential to all life forms [6, 78]. The biosynthesis of glycans and their covalent attachment to specific sites of the protein backbone is a complex and energetically costly process that is not template-driven and that requires the tightly orchestrated interplay of several enzymes,

transporters, chaperones along with various metabolites. As such protein glycosylation is a dynamic process that differs, for instance, between cell types, cellular state (e.g. apoptosis) and metabolism. Compared to other classes of biomolecules, glycans appear to be less conserved and more exposed to evolutionary forces (Figure 11) [79].

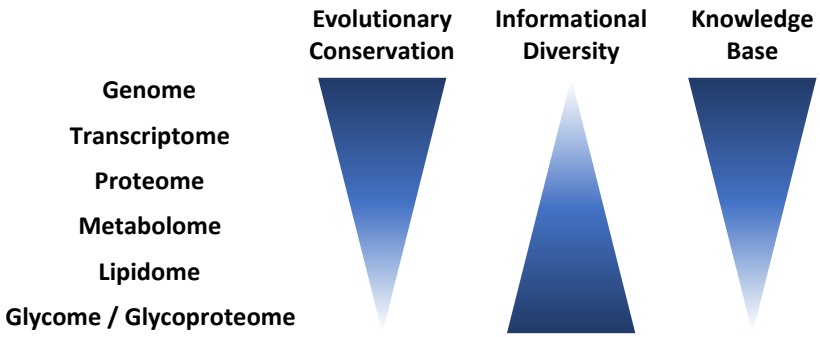


Figure 11: Evolutionary conservation, informational diversity, and knowledge base of the cellular building blocks (taken from Varki, 2017 [6]).

This evolutionary shaping contributed to the emergence of different "versions" of the glycosylation machinery among the different taxa – and made glycans one of the most complex and diverse groups of biomolecules [80, 81]. Inherent to this structural complexity and diversity is an enormous information content that glycans can transmit to their carrier proteins. The addition of glycans is therefore a way to alter attributes of a protein without the need to change the underlying genome. Protein glycosylation thus leads to an amplification of the functional diversity of proteins and allows the cell or organism to rapidly adapt to developmental or environmental changes [78]. This notion is best described by the fact that there is only one (static) genome but several (dynamic) proteomes – and even more (dynamic) glycoproteomes (the proteome predicts the phenotype, but the glyco(proteo)me is the phenotype) [82]. Protein glycosylation can add functional variability to a protein in different ways. First, a glycoprotein can feature more than one type of glycosylation, for instance both *N*- and *O*-glycosylation. Second, any glycosylation site can feature different glycans of a certain type of glycosylation, such as high-mannose-type and complex-type *N*-glycans – a characteristic referred to as micro-heterogeneity. Third, not every potential glycosylation site is occupied, i.e. there is also variability in the glycosylation site occupancy of a glycoprotein – a characteristic referred to as macroheterogeneity. The combination of all these factors leads to the existence of different glycoforms of a single protein (isoforms or proteoforms) (Figure 2).

N- and *O*-glycosylation are the most common forms of protein glycosylation in vertebrates [8]. The majority of secreted or membrane-bound proteins carry at least one of those two types of glycosylation [55]. *N*- and *O*-glycans are therefore believed to be involved in virtually every biological process, either directly or indirectly. Despite their ubiquity, our knowledge about *N*- and *O*-glycans and their biological implications is still limited. One explanation for this lack of knowledge owes to the fact that the functions of the protein and the glycan moiety can be independent from each other: On the one hand all glycoforms of a single protein can perform

the same function – meaning the glycan heterogeneity does not influence the function of the protein; and on the other hand, one single glycan structure can perform the same function, although it is attached to different proteins. In the first case, glycosylation may be only important for proper folding and secretion of the glycoprotein but does not affect for instance binding kinetics or binding selectivity. In the second case, certain glycan structures can serve as recognition motifs for the binding to lectins, thus mediating the concerted processing of different glycoproteins – e.g. effective clearance of blood plasma glycoproteins from the circulation. A second explanation for the difficulty to derive structure-function relationships between proteins and their attached glycans relates to the fact that many functions may only be evident in an organismal context rather than a cellular context, thus requiring appropriate gene-knockout mutants to unravel those functions. In the last two decades information on biological roles of *N*- and *O*-glycans has expanded significantly [8, 83]. It is now well-established that *N*- and *O*-glycans are involved in numerous biological processes such as protein folding and protein stability, cell-cell and cell-matrix interaction, immune response, intracellular targeting, fertilization, inflammation, embryonic development, and microbial infections (Table 1 in Varki *et al.* [6]). Generally, those biological roles can be categorized into three groups: (I) structural and modulatory roles, (II) intrinsic (intraspecies) recognition, and (III) extrinsic (interspecies) recognition – which, of course, may overlap, since structural properties may also affect recognition. Besides the most obvious structural effects of glycans on their carrier proteins – increasing the molecular weight, changing the shape and hydrodynamic volume and potentially changing the net charge state – several other structural effects have been elucidated. A prominent structural effect of glycosylation, particularly of *N*-glycosylation, is its involvement in the protein folding quality control in the lumen of the ER [84]. Preventing the *N*-glycan biosynthesis using the inhibitor tunicamycin leads to incorrect or incomplete protein folding of many glycoproteins. Without reaching their native conformation glycoproteins cannot pass the ER quality control and will ultimately be degraded. The hydrophilic character of glycans, supported by terminal sialic acids, confers yet another important structural feature of glycans – its ability to increase the water solubility of their carrier proteins. Exemplarily, only with the help of intense glycan decorations the high concentrations of proteins present in blood plasma can be realized (~50-70 mg/mL in human blood plasma)[6]. Another important structural function of glycans, conferred by steric hindrance and/or negative charges, is to shield the underlying peptide backbone from proteolytic digest – a mechanism not only important for intracellular protein processing (e.g. activation of premature proteins [85, 86]) but also for protection from pathogenic proteases [6].

Glycan structures can also modulate protein functions. The most prominent example in this context is the modulation of the immunoglobulin gamma (IgG) effector functions by the structural features of the complex-type *N*-glycans in the IgG Fc region (constant region of the heavy chains, CH2) (Figure 12).

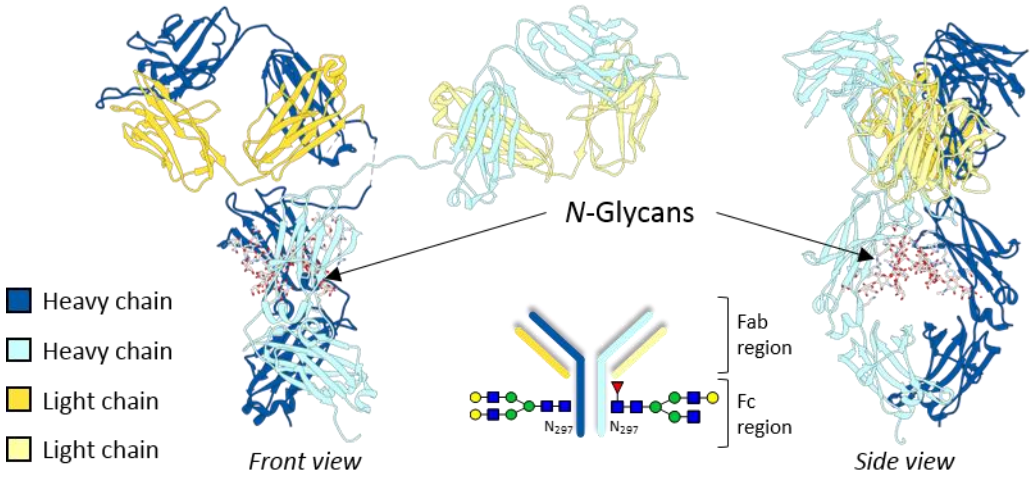


Figure 12: Modeled protein structure of human IgG. The Fc region *N*-glycosylation (site N_{297}) is highlighted accordingly.

Missing or incomplete galactosylation of IgG-Fc *N*-glycans, for instance, has been associated with pro-inflammatory immune responses [87], whereas complete sialylation has been ascribed anti-inflammatory effects (only α2,6-linked NeuAc)[88] (Figure 13). Another modulatory effect is linked to fucosylation. IgG antibodies whose Fc *N*-glycans lack α1,6-core-fucosylation exhibit a much higher antibody-dependent cellular cytotoxicity (ADCC) compared their core-fucosylated variants (up to 100-fold increase)[89] (Figure 13).

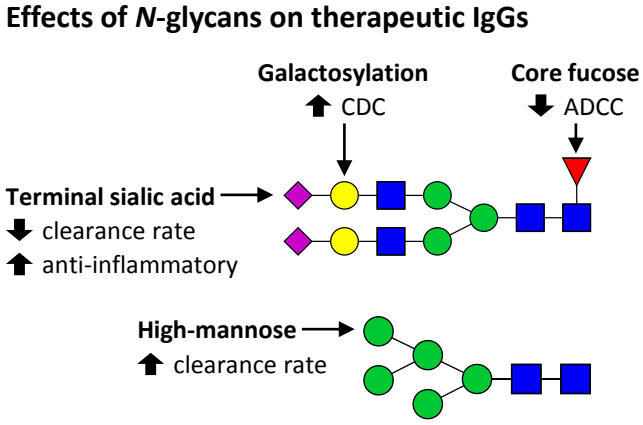


Figure 13: Effects of *N*-glycans on therapeutic IgGs (CDC, complement-dependent cytotoxicity. Modified from Jones, 2017 [70]).

Protein glycosylation can also affect intrinsic and extrinsic recognition – or in other words self- and non-self-recognition [6, 67]. Biosynthesis and adaptation of glycans as well as their recognition via lectins is therefore an inherent part of signaling and immunoregulatory processes (innate and adaptive immune response, tolerance vs inflammation), cell-cell and cell-matrix interactions (regulated cell growth vs tumorigenesis), homeostasis, and microbial-host interactions (tolerance vs immune response). A central element of cellular recognition processes is the glycocalyx – an extracellular layer of glycoconjugates and lectins embedded in the cell membrane of essentially all cells [6] (Figure 14).

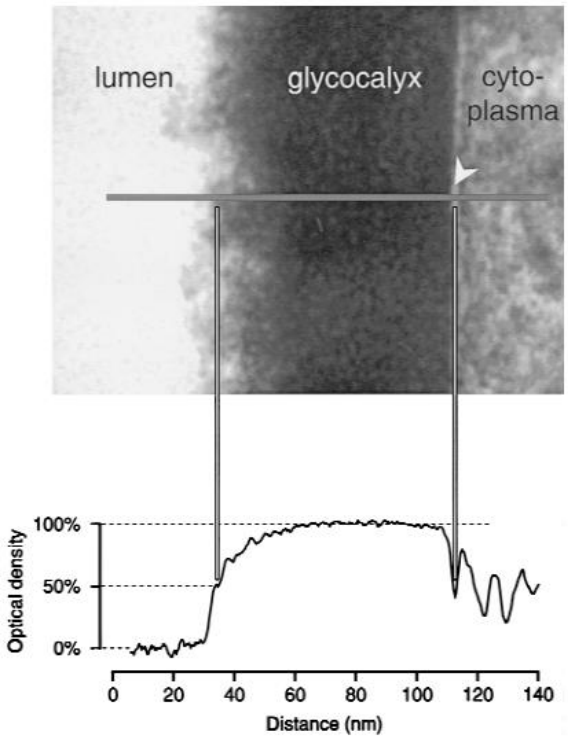


Figure 14: Thickness of the endothelial glycocalyx of murine cerebral capillaries measured by transmission electron microscopy (taken from Vogel et al., 2000) [90].

The glycocalyx acts as a barrier protecting the cell by shielding the underlying cell membrane, thus effectively preventing pathogens from adhering and entering the host cell. Moreover, the glycocalyx forms a lattice that together with the extracellular matrix is essential for cell adhesion, movement, and regulation. Many pathogens directly target the host cell glycocalyx. They have therefore evolved effective ways to adhere to the glycocalyx via lectins and to remodel or destroy it via cell surface-resident and secreted glycosidases [91]. A prominent example for such strategy is the influenza A virus. The virus features two antigens embedded in its lipid envelope that are essential for its life cycle: the glycoproteins hemagglutinin (HA) and neuraminidase (NA). HA is a lectin that specifically binds NeuAc residues on receptors present on vertebrate target cells [19]. Once attached, HA also facilitates the viral infection, i.e. the entry of the virus into the target cell via endocytosis. Binding of HA to the target cell and virus entry is supported by NA, as the viral neuraminidase, among others, inactivates NeuAc decoys, i.e. mucin glycoproteins, present in the mucus that covers the lung epithelial cells [92-94]. NA is also crucial for completing the influenza life cycle. NA removes NeuAc residues from progeny virus NA and HA glycoproteins and from glycoconjugates present on the surface of infected cells – thus preventing newly assembled viruses from binding to the cell surface of the infected host cell, and from aggregating with each other via HA-NeuAc binding. NA inhibitors are currently the most effective therapeutics to prevent and treat influenza A and B virus infections [95]. The influenza A virus example illustrates another important mechanism in the glycobiology between different species – host specificity and adaptation. Depending on the influenza A virus strain, the viral HA either binds to α2,3-linked NeuAc – present on glycoconjugates of avian lung epithelial cells, or to α2,6-linked NeuAc – present on glycoconjugates

of human lung epithelial cells. Both events seem to be dependent on a characteristic glycan topology [96]. Glycans are not only a fundamental part of the host cell defense system, also attacking pathogens make use of glycans for their own ends. For instance the glycocalyx of many pathogens can serve as a form of molecular mimicry, which enables pathogens to evade the host cell immune system [6].

2.3. Glycosylation in Human Health and Disease

In the human body, protein glycosylation is highly regulated and plays an important role in numerous processes. The glycan biosynthesis and glycan structure-function relationships are thereby influenced by both genetic and environmental factors. As a result, glycans and their direct recognition/interaction partners are not only involved in physiological conditions – they can also contribute to pathophysiological conditions [6, 20]. Altered glycosylation has been associated with various diseases and disorders. However, it is important to note that aberrant glycosylation does not necessarily have to be the cause of a disease; it can also be an outcome of a disease. Inborn defects in glycosylation (including protein glycosylation, but also glycosylation of lipids and proteoglycans) can affect multiple organs, leading to multifaceted symptoms such as mental retardation, muscular dystrophies, or cardiomyopathy. Such inherited defects in glycosylation are referred to as congenital disorders of glycosylation (CDG), a group of rare disorders that can have mild but also severe (mostly lethal) clinical manifestation [97-101]. Inborn or acquired malfunction of the glycosylation machinery or the expression of certain histo-blood group antigens can also favor or lead to the emergence of several diseases, including infectious diseases, cardiovascular diseases, autoimmune diseases, neurological diseases, as well as allergies [20-22, 102-107]. A defective glycocalyx on epithelial cells of the intestines (mucosa), for instance, can be exploited by pathogenic (e.g. *Helicobacter pylori*) but also non-pathogenic bacteria (e.g. opportunistic bacteria from the gut microbiota) – and can be the starting point of a gastro-intestinal infection [108-110]. Using their repertoire of adhesins, bacteria can adhere to the exposed membrane and can cause serious infections, which can ultimately lead to gastritis, gastric ulcers, and gastric cancers. Altered glycosylation, particularly of blood plasma proteins such as immunoglobulins, has also been linked to autoimmune diseases including inflammatory bowel diseases (e.g. Crohn's disease)[111], or rheumatoid arthritis [112, 113]. As an example, IgGs from rheumatoid arthritis patients exhibit a reduced or even no galactosylation in their Fc *N*-glycans – with the reduced degree of galactosylation being directly correlated with the severity of the disease [114]. Aberrant protein glycosylation is also associated with cancer development and metastasis [107, 115, 116]. A change in glycosylation (not randomly) is considered a universal hallmark of cancer cells, linking glycoconjugate expression to malignant transformation. Certain glycan changes are frequently associated with tumor progression and metastasis: e.g. (I) increased β1,6-GlcNAc branching of *N*-glycans, (II) expression of Tn and sialyl Tn *O*-glycan antigens, or (III) altered expression of blood-group antigens. Protein glycosylation plays a major role in many inherited and acquired pathophysiological conditions. The analysis of diseases-associated glycosylation

alterations therefore holds the potential (I) to unveil genetic and environmental factors relevant for pathogenesis, (II) to monitor the onset and progression of a disease, (III) to predict exacerbation or remission of a disease, (IV) to identify potential therapeutic targets, and (V) to develop intervention strategies.

2.4. Glycoengineering and Glycotherapeutics

Investigating glycosylation machineries across different taxa offers the possibility to gain a deeper understanding of evolutionary connections (phylogeny), such as conserved or diverged glycosylation pathways – and to unravel environmental adaptations and survival strategies, such as lectin-mediated pathogen-host interactions. The analysis of heterologous glycosylation machineries allows to dissect and compare glycosylation pathways of different species and to reveal homologous pathways that were unknown before. A prominent example in this regard is *O*-mannosylation [117, 118]. This type of *O*-glycosylation was first discovered in 1968 in yeast [119]. Thirty years later, in 1999, *O*-mannosylation was also described in humans [120]. Knowledge of the glycosylation machinery of different species can be used to perturb or modify single glycosylation pathways, and thus to generate glycoengineered cells or organisms. Perturbation of a glycosylation pathway, for instance by knocking out genes coding for a single glycosyltransferase, allows studying biological implications of this pathway, this glycosyltransferase, or the produced glycoconjugates (common techniques for genome editing are CRISPR/Cas9, TALENs, Zinc finger nucleases [121]). Such gene knockouts can be used, for instance, to generate cell lines and organisms that serve as disease models [122]. Additionally, by analyzing glycan micro- and macroheterogeneity of expressed glycoproteins, glycosylation perturbation experiments may also reveal structural implications of the affected glycosylation pathway [122]. In 2011 Steentoft *et al.* introduced SimpleCell, an innovative genome editing concept that allows generating glycoengineered cell lines with a simplified core-1 mucin-type *O*-glycosylation [123]. Zinc-finger nuclease-mediated knockout of the COSMC gene, a gene that codes for a chaperone, perturbs initial steps of the *O*-glycan biosynthesis and inhibits elongation of core-1 mucin-type *O*-glycans. The resulting glycoengineered SimpleCell lines exhibit only single GalNAc or $GalNAc_1NeuAc_1$ *O*-glycans that are attached to either serine or threonine. *O*-glycosylation, unlike *N*-glycosylation, does not feature a conserved sequence motif that indicates potentially occupied glycosylation sites. Identification of *O*-glycosylated proteins and their occupied *O*-glycosylation sites is therefore more challenging and less well advanced. By reducing the structural complexity of the *O*-glycoproteome, i.e. the entirety of all core-1 mucin-type *O*-glycosylated proteins of a cell, the SimpleCell approach can identify *O*-glycosylated proteins and occupied *O*-glycosylation sites. Another innovative glycoengineering approach is the GlycoDelete strategy introduced in 2014 [124]. With this *in vivo N*-glycan remodeling approach, cell lines can be generated that express recombinant glycoproteins with truncated yet homogenous *N*-glycan structures (Asn-$GlcNAc_1\pm Gal_1\pm NeuAc_1$). The first key element of this approach is the knock-out of the gene encoding for *N*-acetylglucosaminyl-transferase I (GnT I). This Golgi-resident glycosyltransferase initiates the buildup of complex-

type or hybrid-type *N*-glycans by addition of a β1,2-linked GlcNAc residue to the Man α1,3-arm of Man5-*N*-glycan intermediate structures. Lack of GnT I in GlycoDelete cells still produces correctly folded *N*-glycoproteins – the *N*-glycans, however, will remain in the Man5 stage. In the second step those Man5 *N*-glycans are cleaved at the chitobiose core by endoT, a recombinant fungal endo-β-*N*-acetylglucosaminidase that is targeted to the Golgi. Finally, the remaining and asparagine-linked GlcNAc residues are elongated by Golgi glycosyltransferases (galactosyltransferase and sialyltransferase) to generate the $GlcNAc_1Gal_1NeuAc_1$ trisaccharides and to complete the *N*-glycan remodeling. A prominent application of the GlycoDelete approach is the production of neutralizing therapeutic antibodies with improved and *N*-glycan-related characteristics (increased efficacy by reduced binding to Fc-γ-receptors; prolonged blood circulation, leading to administration of lower doses) [124]. Another application relates to the production of biopharmaceutical glycoproteins (glycotherapeutics) in non-human expression systems, such as plant cell lines [125]. Using GlycoDelete, plant-derived glycoproteins can be produced that do no feature plant-specific *N*-glycan modifications (β1,2-xylose, α1,3-fucoses) that may cause immunogenic reactions upon parenteral administration [125, 126]. The GlycoDelete approach is representative for a trend in biopharmaceutical industry to develop and produce so-called "biobetters" [127]. Biobetters belong to the group of biosimilars (or follow-on biologics), i.e. "generic" versions of previously licensed biopharmaceuticals/biologicals ("innovators" or "originators") that are similar in terms of quality, safety, as well as pharmacokinetics (including for instance half-life) and pharmacodynamics (including for instance potency, efficacy, and adverse effects). Compared to its originator, a biosimilar is produced from different cell clones and with different manufacturing processes. This can result in the biosimilar to exhibit differences in glycosylation and other microvariations, e.g. asparagine or glutamine deamidation, point mutations, or disulfide bridge formation [127]. Those differences are considered as critical quality attributes as they can affect the quality, safety, pharmacokinetics and pharmacodynamics of the product (manufacturing of biopharmaceuticals should therefore be in line with the "quality by design" framework) [128-130]. For approval by regulatory authorities (e.g. U.S. Food and Drug Administration – FDA, European Medicines Agency – EMA) manufacturers must proof interchangeability between the biosimilar and the originator and must reassure that differences do not have any clinical relevance. In 2013 the EMA approved the first biosimilar monoclonal antibody in Europe [131]. This biosimilar, sold under the brand names Remsima and Inflectra, is based on the innovator infliximab (brand name Remicade) which is used for the treatment of autoimmune disorders, such as rheumatoid arthritis and Crohn's disease. A biobetter, in contrast to a biosimilar, is considered as a completely new drug. It still has the same mode of action as the originator and biosimilar, e.g. targeting the same epitope, but it exhibits improved characteristics in terms of quality, safety, as well as pharmacokinetics and pharmacodynamics [127]. Those improved characteristics arise from the rational design of critical quality attributes, such as protein glycosylation [132] (also in combination with modifications such as the attachment of polyethylene glycol, i.e. glycoPEGylation [133, 134]). The production of biopharmaceuticals is

constantly evolving, with new products, or improved and/or more cost-efficient versions of already existing products, being approved each year. With this, biopharmaceuticals account for an increasing share of the revenues in the pharmaceutical industry. Most produced biopharmaceuticals are antibody-based therapeutics (including mAbs, but also multispecific antibodies, antibody fragments, single-domain antibodies, and antibody drug conjugates). The remainder accounts for growth factors, hormones, cytokines, enzymes, vaccine components, and others. In 2013, global sales of mAbs were nearly $75 billion US dollars, representing half of the total sales of all biopharmaceutical products (insulins, produced in *Escherichia coli,* are the next most lucrative product class) [24]. Glycoengineering plays a pivotal role in the success of current and future biopharmaceuticals, as it enables modulating or improving pharmacokinetics and pharmacodynamics of the product – essential requirements to produce biopharmaceuticals that are both potent and cost-efficient [29, 135-137]. A prominent example for successful glycoengineering is the development of FUT8-deficiet cell lines that produce IgG mAbs with lower or no α1,6-core-fucosylation (FUT8, α1,6-fucosyltransferase)[26, 138]. Such antibodies induce a much higher ADCC compared to core-fucosylated antibodies – a characteristic crucial for cancer treatment [139, 140]. An increase of ADCC can also be induced by overexpression of β1,4-*N*-acetylglucosaminyltransferase III (GnT-III), a glycosyltransferase that catalyzes the addition of a bisecting GlcNAc to the trimannosyl core of *N*-glycans – an event, which in turn inhibits the addition of a core-fucose [141]. Several glycoengineered IgG mAbs, with higher ADCC, are now in clinical studies and two of them have already been approved since 2012 (mogamulizumab [142], obinutuzumab [143, 144]). Another form of glycoengineering, besides genetic engineering, is metabolic glycoengineering. In this approach cells are cultivated in the presence of substrates (e.g. non-natural monosaccharides), cofactors (e.g. manganese) or enzymes (e.g. kifunensine [145], swainsonine [146]) that transiently alter the glycosylation machinery. As an example, cells cultivated in the presence of kifunensine, an α-mannosidase inhibitor [145], produce IgG antibodies that solely exhibit high-mannose-type *N*-glycans. Those antibodies were shown to induce a strong ADCC response, making this approach appealing for producing therapeutic antibodies [147]. Another metabolic glyco-engineering approach for promoting the biosynthesis of high-mannose-type *N*-glycans on IgG antibodies is to supplement the cell culture medium with manganese, while limiting glucose supply [148]. Several expression systems for the production of biopharmaceuticals have been established, with mammalian cell lines being the preferred ones [28, 149]. The most commonly used platform for the production of biopharmaceuticals is Chinese hamster ovary cell lines (CHO)[30]. The use of CHO cells for production of recombinant proteins, in particular mAbs, is well-established and has a longstanding history. With regard to production of glycoproteins, CHO cells are favorable as they add no, or only low levels of, immunogenic glycan epitopes to recombinant proteins (no immunogenic α-Gal residues; mainly α2,3-NeuAc, only trace amounts of immunogenic NeuGc; no bisecting-GlcNAc) [150]. In 2016 seven of the top 10 best-selling biopharmaceuticals were produced in CHO cells [151]. Other currently available mammalian production cell lines include, among others, human cell lines (e.g. HEK293,

PER.C6, HT-1080, and CAP), rodent cell lines (e.g. NS0, Sp2/0, BHK21), dog cell lines (MDCK), and monkey cell lines (Vero) [32, 152]. There is also an increasing variety of non-mammalian biopharmaceutical production cell lines ranging from bacteria (mainly *Escherichia coli*) [153], filamentous fungi (e.g. *Aspergillus niger*) [154, 155], yeast (mainly *Pichia pastoris*, *Saccharomyces cerevisiae*) [155-157], insect cell lines (e.g. Sf9 cells of *Spodoptera frugiperda*) [158, 159], plant cell lines (e.g. *Nicotiana benthamiana*, *Arabidobsis thaliana*) [160], to avian cell lines (e.g. cES cells)[161]. Further production platforms include transgenic plants [160] and animals [162], as well as cell-free expression systems (*in vitro* expression systems) [33, 163]. For the production of biopharmaceutical glycoproteins – dedicated to be administered to humans – the glycosylation of non-human production cell lines has to be either glycoengineered i.e. humanized to be compatible with the human immune system; or any immunogenic reaction, caused by the non-human glycosylation of the drug, has to be ruled out. Progress and current trends in the development of glycotherapeutics have been reviewed, among others, by Dalziel *et al.* [135], Kesik-Brodacka [151], and Beck *et al.* [164].

2.5. Analysis of Glycoproteins

The qualitative and quantitative analysis of glycosylated proteins is essential to unravel their biological implications and to fully understand a variety of biological processes. To comprehensively analyze a glycoprotein, it is necessary to identify, characterize, and quantify all proteoforms (glycoforms). This generally requires the in-depth analysis of both the protein moiety and the glycan moiety. The in-depth analysis of the protein moiety usually involves the identification of the protein based on its primary sequence, characterization of all post-translation modifications (including type, site, and stoichiometry of the modification), as well as determination of the protein abundance. Additional higher order structural analyses are possible, too (e.g. secondary/tertiary conformation analysis, or aggregation and sizing analysis). For an in-depth analysis of the glycan moiety, the glycoprotein has to be analyzed with respect to the glycan micro- and macroheterogeneity. Analysis of the glycan microheterogeneity involves identification and complete structural elucidation of all glycans present on the glycoprotein – in a site-specific and quantitative manner. The analysis is completed by investigation of the glycan macroheterogeneity, which involves the quantitative determination of the glycosylation site occupancy.

Several analytical methods are available to study glycoproteins. These methods can be generally categorized into: x-ray crystallographic, NMR spectroscopic, mass spectrometric, chromatographic, electrophoretic, and lectin-based methods. Chromatographic, electrophoretic, and lectin-based methods, when used stand-alone, have to be coupled with suitable detection methods, such as pulsed amperometric detection (PAD) or colorimetric detection including ultraviolet (UV) detection and fluorescence detection. Additionally, mass spectrometric, chromatographic, electrophoretic, and lectin-based methods may require derivatization of the sample, to enable, or to improve detection and separation of the analyte (e.g. fluorescent labeling[48, 52, 165, 166] or permethylation[166, 167] of the glycan moiety). Additionally, the depth and

confidence of glycoanalytical methods can be further increased by inclusion of endo- and exoglycosidase digests into the workflow (e.g. using α-galactoisidase, an exoglycosidase, to confirm or refute the presence of terminal α-galactoses on *N*-glycan structures). The aforementioned analytical methods can also be combined, resulting in so-called hyphenated methods. A common combination in this regard is the coupling of liquid chromatographic methods with mass spectrometric methods (LC-MS).

For the analysis of glycoproteins four general strategies evolved. Thereby, the focus can be either on (I) glycan monosaccharides, (II) released glycans, (III) intact glycoproteins, or (IV) glycopeptides (Figure 15). Depending on the focus a distinct level of information with respect to the glycosylation of a protein is provided, which thus also defines the fields of application of these strategies. Still, up to now, none of these strategies alone allows for a comprehensive analysis of a glycoprotein; only the combination of different strategies can fulfill this task.

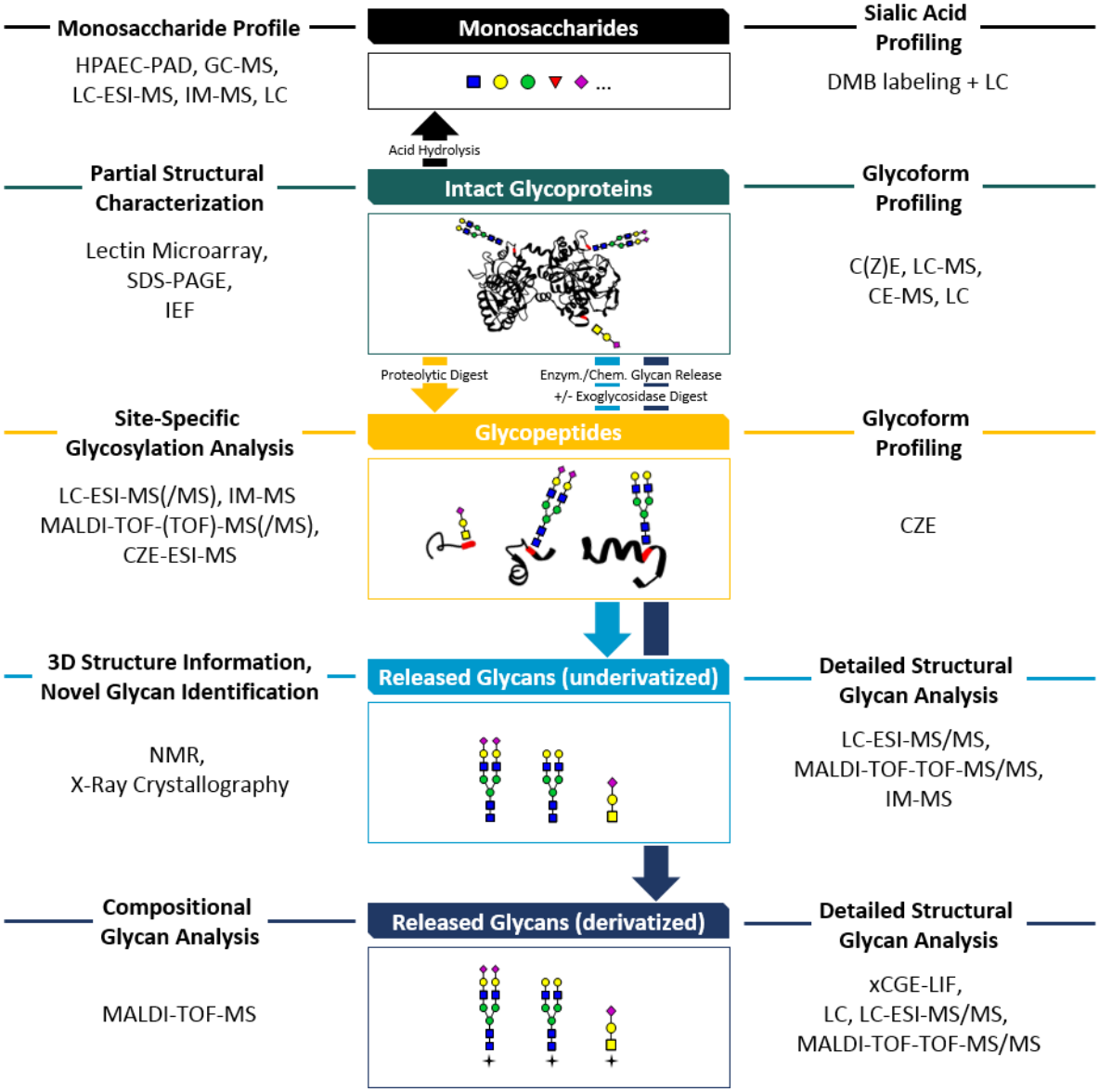


Figure 15: Overview of glycoanalytical approaches (modified from Marino et al., 2010 [168] and Jones, 2017 [70]).

2.5.1. Monosaccharide Analysis

Monosaccharide analysis allows determining the total monosaccharide composition of the entirety of all *N*- and/or *O*-glycans attached to a single protein or a complex mixture of proteins (e.g. proteins derived from a cell lysate). Such an analysis can provide both qualitative (including detection of structural isomers) and quantitative insights (relative as well as absolute values) into the monosaccharide composition. With the help of monosaccharide analysis, information about the type of glycosylation (*N*- vs *O*-glycosylation), as well as general characteristic of each type of glycosylation can be retrieved, including differentiation of high-mannose-type from complex-type *N*-glycans, or presence or absence of certain monosaccharides (e.g. absence of fucose on *N*-glycans derived from therapeutic IgGs). Monosaccharide analysis requires the *N*- and/or *O*-glycans to be released from the protein and decomposed into monosaccharides. This step is usually performed by acid hydrolysis (e.g. via

trifluoroacetic acid; for the analysis of acid-labile sialic acids milder conditions have to be used, e.g. acetic acid). The analysis of released monosaccharides can be conducted, for instance, with capillary electrophoreses coupled to laser-induced fluorescence (CE-LIF) [169], high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) [170], as well as liquid or gas chromatography coupled with mass spectrometry [171-173]. Monosaccharide analysis belongs to the glycocentric, or glycomic approaches (sometimes also referred to as glycocentric glycoproteomics), and provides limited, but valuable, information on the glycan moiety. However, it does not provide any information on the protein and its relation to the glycan moiety, i.e. protein identity, and glycan micro- and macroheterogeneity.

2.5.2. Analysis of Released Glycans

The second strategy is based on the analysis of released *N*- and *O*-glycans. It also belongs to the glycomics or glycocentric approaches and allows analyzing the entirety of all *N*- and/or *O*-glycans. Thereby, the analysis of released *N*- and *O*-glycans can provide both qualitative (including complete structural elucidation) and quantitative information (absolute and relative quantitation). An essential step is the non-destructive release of *N*- and/or *O*-glycans from the protein backbone – a step that can be performed enzymatically but also chemically. The enzymatic release of *N*-glycans is commonly achieved using the endoglycosidase/amidase peptide-*N*-Glycosidase F (PNGase F). The enzyme catalyzes the hydrolysis of the *N*-glycosidic bond between glycosylated asparagines and the innermost GlcNAc of complex-, high-mannose-, and hybrid-type *N*-glycans – unless the *N*-glycans feature an α1,3-core-fucosylation (which is common to plant and insect glycoproteins); for this type of *N*-glycans PNGase A can be used. With regard to *O*-glycans, to date, there is no single enzyme that can release *O*-linked glycans in general. The enzyme *O*-glycanase, for instance, is restricted to the release of core-1 *O*-glycans. Therefore, the preferred form of *O*-glycan release is a chemical release. A common method in this regard is hydrazinolysis, which enables the release of *O*-glycans, but also *N*-glycans, from the protein backbone in a selective manner (temperature-dependent release). Additionally, *N*- and *O*-glycans can also be chemically released by reductive or non-reductive β-elimination under alkaline conditions. Of note, chemical *O*-glycan release may suffer from "peeling", i.e. the sequential degradation of reducing end monosaccharide units by consecutive β-elimination reactions.

The analysis of released *N*- and *O*-glycans can be perfomed using various methods. Currently, NMR spectroscopy is still the only method that enables a complete structural elucidation of released *N*- and *O*-glycans without any *a priori* knowledge (*de novo* analysis of glycan structure) [174]. Other methods such as mass spectrometric, electrophoretic and chromatographic methods – alone or in combination – may also provide structural *N*- and *O*-glycan information, e.g. identification of structural isomers or determination of linkages between monosaccharides. However, interpretation of data acquired with those methods may require knowledge about *N*- and *O*-glycan structures and their biosynthetic pathways. Still, such so-called glycoprofiling methods are powerful and widely-used for the analysis of released *N*- and *O*-glycans, as they do

outperform NMR spectroscopy with regard to sensitivity and ease of use (including required expertise, analysis speed, and high-throughput capability). The most commonly used glycoprofiling methods are [50, 175]: multiplexed capillary gel electrophoresis with laser-induced fluorescence detection (xCGE-LIF)[48, 49, 52, 53], hydrophilic interaction liquid chromatography (HILIC) on an ultra-performance liquid chromatography (UPLC) system with fluorescence (FLR) detection (HILIC-UPLC-FLR) [176, 177], HILIC-UPLC-FLR coupled to electrospray ionization (tandem) mass spectrometry (HILIC-UPLC-FLR-ESI-MS(/MS)) [165], porous graphitized carbon liquid chromatography coupled to ESI-MS/MS (PGC-LC-ESI-MS/MS) [178], ion mobility-MS (IM-MS) [179], and MALDI-TOF-MS(/MS) [180, 181].

The analysis of released glycans provides in-depth qualitative and quantitative information on the glycan moiety, including even structural details. Recent developments for instance allow to determine the linkage of terminal NeuAc residues (α2-3- vs α2-6-linked) present on *N*-glycans by MS [180]. To this end, NeuAc residues were derivatized before the MS analysis, resulting in distinct mass shifts that enable discrimination of the respective linkage type. Glycoprofiling methods, such as xCGE-LIF or MALDI-TOF-MS, also allow the high-throughput analysis of multiple samples – a feature that makes these methods very powerful for the analysis of glycosylation-related alterations of large sample cohorts, such as clinical samples [48, 50, 182]. Still, the analysis of released glycans does not provide any information on the protein and its relation to the glycan moiety (unless a perfectly purified glycoprotein, harboring only one fully occupied glycosylation site and only one type of glycosylation, is analyzed, e.g. analysis of the IgG 1 Fc *N*-glycosylation).

2.5.3. Analysis of Intact Glycoproteins

The third strategy is based on the analysis of intact glycoproteins. It belongs to the top-down proteomics or proteocentric approaches (sometimes also referred to as proteocentric glycoproteomics) and allows analyzing the entirety of glycoforms of a protein population. The analysis of intact glycoproteins, also known as intact glycoprotein profiling, can therefore provide a holistic view on the type, abundance, and extent of glycosylation – including protein sequencing as well as determination of glycan micro- and macroheterogeneity (each with some limitations). However, intact glycoprotein profiling does not enable determination of the exact glycan structures. Intact glycoprotein profiling is usually performed by (ultra)high-resolution mass spectrometry, namely MALDI-MS and ESI-MS (both only MS^1-mode, i.e. no tandem MS is performed). While in MALDI-MS either a TOF or Fourier transform ion cyclotron resonance (FTICR) mass analyzer is used, in ESI-MS various mass analyzers including TOF, quadrupole TOF (qTOF), FTICR, and Orbitrap (OT) mass analyzers can be used. Recently, this repertoire of intact glycoprotein profiling techniques was extended by ion mobility (IM) MS [183] (usually implemented in an ESI-qTOF-MS). Depending on the complexity of the sample and/or depth of analysis, the aforementioned MS-based techniques can be used stand-alone or coupled to various LC (e.g. HILIC- or RP-LC) or capillary (zone) electrophoresis (C(Z)E) techniques (online- or offline-coupling, MALDI-MS only offline-coupling) [184, 185]. In

addition to MS-based techniques, CE-UV can be used for intact glycoprotein profiling. Intact glycoproteins are usually measured in their reduced state, i.e. after additional reduction using a reducing agent such as dithiothreitol (DTT). However, in recent studies also the analysis of intact glycoproteins in their native (non-reduced) state has also been demonstrated [186]. Still, despite its undeniable potential and benefits, the analysis of intact glycoproteins is currently relegated to a niche existence due to technical challenges. For instance, the majority of current-generation mass spectrometers does not offer enough resolution to reliably identify intact glycoproteins and to detect small post-translational modifications (PTMs) such as deamidation (+0.98402 Da), also correct deconvolution of mass spectra is an issue. The analysis of intact glycoproteins is therefore currently usually limited to purified proteins with a low number of different glycans and/or a low number of glycosylation sites (e.g. mAbs) [186-188]. Of note, glycoprotein profiling of mAbs can also be performed using a middle-down proteomics approach [189]. This approach is similar to the described top-down approach, however it reduces the complexity of the immunoglobulin molecule by analyzing only a subunit of the antibody – either the antigen-binding Fab region or the Fc region (Figure 12). This can be achieved by generating antibody Fab and Fc fragments using proteases that cleave in the hinge region of the antibody, such as papain or the immunoglobulin-degrading enzyme of *Streptococcus pyogenes* (IdeS).

2.5.4. Analysis of Glycopeptides

The fourth strategy is based on the analysis of glycopeptides generated from a proteolytic digestion of a single glycoprotein or a group of glycoproteins. It belongs to the bottom-up proteomics approaches yet combines insights and strategies from both, proteocentric and glycocentric glycoproteomic approaches. Based on this notion, the analysis of glycopeptides is sometimes referred to as mono- or polyproteic glycoproteomics, respectively. In general, though, the term glycoproteomics has established itself for this type of analysis, and hence will also be used in this meaning throughout the present work. As this thesis is focused on the glycoproteomic analysis of *N*- and *O*-glycopeptides, this type of analysis is described in more detail in the following.

Among the four general approaches to analyze the protein glycosylation (I-IV), the glycopeptide-based analysis is currently the most informative, yet also the most challenging approach. It is the only approach that provides detailed and site-specific information on the glycosylation of a protein, including glycan micro- and macroheterogeneity, while also providing detailed information on the protein itself, including its identity as well as other present modifications such as phosphorylation [42, 190]. Unlike other approaches, this information can also be retrieved from glycoproteins present in a complex matrix such as human blood plasma, and it can be obtained from glycoproteins harboring multiple glycosylation sites and multiple types of glycosylation [191]. In fact, glycoproteomics is at present the only approach capable to detect and to site-specifically assign (theoretically) all known types of glycosylation. With current glycomics approaches *N*-glycans from a wide range of protein sources can be

identified and quantified with high reliability due to the existence and availability of the enzymes PNGase F and PNGase A. The glycomic analysis of other types of protein glycosylation such as mucin-type *O*-glycosylation, *O*-mannosylation or *C*-mannosylation is still not fully mature or even possible, due to the lack of an appropriate enzyme for the selective release of the glycans (chemical release methods do exist, though these may induce unwanted side-reactions). The glycoproteomic analysis is based on the generation of glycopeptides, followed by the qualitative and (relative) quantitative analysis of these glycopeptides. Thereby, the ultimate goal is to reconstruct the intact glycoprotein with all its glycoforms, i.e. to cover all types of glycosylation, all glycoforms, and all glycosylation sites. This endeavor requires the generation of appropriate glycopeptides covering all glycosylation sites of interest (e.g. all potential *N*-glycosylation sites), the detection of these glycopeptides including the glycan micro- and macroheterogeneity (relative quantitation), and finally the unambiguous identification of the generated glycopeptides. The latter involves the identification and characterization of two structurally very diverse entities of the glycopeptide – the peptide moiety and the glycan moiety. This structural diversity must be taken into account during sample preparation, measurement, and data analysis. Identification of the peptide moiety is usually based on classical bottom-up proteomics strategies: that is, sequencing of amino acids taking into account of natural and artificial amino acid modifications, followed by a database search to identify the corresponding protein. In this context, also pinpointing the occupied glycosylation site is required. Identification and characterization of the glycan moiety is based on glycomics strategies and seeks to elucidate the structure of the attached glycan to the best extend possible; a full structural elucidation of the glycan moiety solely based on state-of-the-art glycoproteomic methods is currently not possible. Therefore, usually only the glycan composition can be determined.

Mass spectrometry serves as the core technology platform in glycoproteomics [44, 103]. The analysis of glycopeptides relies on soft MS ionization techniques – primarily MALDI and ESI. Independent of the ionization technique, glycopeptides can be ionized in positive-, as well as negative-ion mode (usually leading to protonated or deprotonated molecular ions, i.e. $[M+z\times H]^z$ or $[M-z\times H]^z$ respectively; z equals the charge state of the molecular ion (M)). Ionization of glycopeptides in positive-ion mode is, thereby, the established and predominately used mode. Despite showing enormous potential for the analysis of glycopeptides, negative-ion mode ionization of glycopeptides has only rarely been adopted in current glycoproteomic approaches up to now [192]. For the analysis of glycopeptides a wide range of different mass analyzers can be combined with each of the two soft MS ionization techniques. These days the most commonly used MS instrument configurations in glycoproteomics are: MALDI-TOF(/TOF)-MS, MALDI-FTICR-MS, ESI-ion trap (IT)-MS, ESI-linear quadrupole ion trap (LTQ)-MS, ESI-qTOF-MS, ESI-triple quadrupole (QqQ)-MS, ESI-OT-MS, ESI-LTQ-OT-MS, as well as ESI-IM-qTOF-MS. Thereby, each of these instrument setups has its advantages and disadvantages with regard to sensitivity, resolution, mass accuracy, scan speed, covered mass-over-charge (m/z) range, versatility, costs etc. Similar to the MS-based analysis of glycans

and intact glycoproteins, also glycopeptides can be analyzed solely by MS or by online- or offline-coupled methods, such as LC-MS or in rarer cases also CE-MS. Depending on the complexity of the analyte and/or the depth of analysis either of these two general approaches is applied. A prominent example for the first case is the analysis of tryptic IgG Fc *N*-glycopeptides by MALDI-TOF-MS or MALDI-FTICR-MS (a type of glycoproteomic analysis referred to as glycopeptide profiling) [193-195]. Due to its high sensitivity and selectivity and its high-throughput capability MALDI-MS-based glycopeptide profiling is a powerful technique to analyze the Fc *N*-glycosylation of IgGs, also including therapeutic mAbs (relative *N*-glycan quantitation). The approach is comparable to MALDI-MS-based *N*-glycan profiling of IgGs. However, the *N*-glycopeptide-based approach adds further confidence in the analysis of the IgG Fc *N*-glycosylation by maintaining information of the protein identity in the form of the peptide moiety that is still linked to the glycan. With the help of unique *N*-glycopeptides, this approach enables to selectively identify and quantify *N*-glycans present in the Fc region of IgG, while neglecting *N*-glycopeptides derived from the Fab region or other proteins present in the sample (on purpose or not) – a characteristic that IgG *N*-glycan profiling by itself does not have. However, for the majority of glycoproteomic analyses the second case – hyphenated MS strategies – are applied. Among those, C18 nano-reversed-phase liquid chromatography coupled online to electrospray tandem mass spectrometry (nano-RP-LC-ESI-MS/MS) has become the predominantly used method (Figure 16).

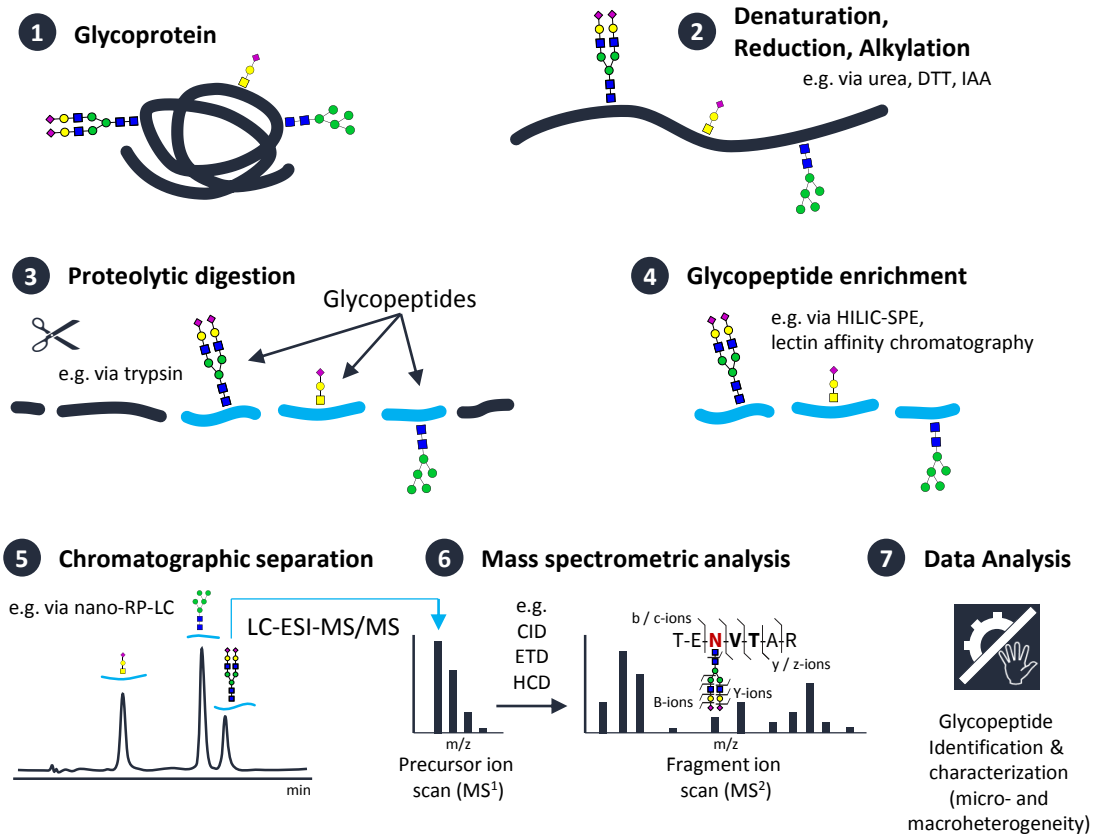


Figure 16: Generic workflow for the LC-ESI-MS/MS-based analysis of *N*- and *O*-glycoproteins.

Other setups such as nanoPGC-ESI-MS/MS or nanoHILIC-ESI-MS/MS are less frequently applied. C18 RP-LC for the separation of glycopeptides prior to ESI-MS/MS analysis is derived from bottom-up proteomics, where many of today's glycoproteomic approaches were borrowed from. The use of C18 as a stationary phase enables to retain and separate the vast majority of peptides and glycopeptides based on their hydrophobicity – even though glycopeptides exhibit additional hydrophilicity due to their glycan moiety. At the same time the employed buffers during RP separation are well-suited for the subsequent mass spectrometric analysis of the eluting (glyco)peptides in positive-ion mode. The retention of glycopeptides depends on the hydrophobic character of the peptide (defined by the length of the peptide, the type of amino acids present, as well as present amino acid modifications), and depends on the number and the type of monosaccharides of its glycan(s). Comparing the C18 chromatographic retention behavior of a glycopeptide and its non- or de-glycosylated counterpart, the glycopeptide usually elutes earlier due to the increased hydrophilicity conferred by the additional glycan moiety. This characteristic can be used to optimize RP-LC elution gradients to favor retention and separation of glycopeptides over non-glycosylated ones. When comparing the C18 chromatographic retention behavior of different glycoforms of a particular glycopeptide, e.g. different *N*-glycoforms of an *N*-glycopeptide, glycopeptides usually elute earlier with increasing number of monosaccharides forming the glycan moiety. However, a special case in this regard is the presence of sialic acid residues, e.g. NeuAc: due to its negative charge, glycopeptides with sialylated glycans elute later compared to their non-sialylated counterparts on a C18 column [196].

The MS analysis of glycopeptides can be carried out on multiple levels: On MS^1 level, solely the *m/z* values, charge state and the abundance of each glycopeptide (or any other detected molecular ion) will be determined. The output of this analysis is termed MS^1 or precursor ion spectrum (the analysis itself is also termed precursor ion scan or survey scan). In general, every molecular ion detected by MS, produces multiple signals that are associated with each other, the so-called isotope pattern (also known as isotope cluster or isotope distribution). The isotope pattern directly reflects the elemental composition of the molecular ion, i.e. it allows to determine the molecular formula of the molecular ion. A useful principle in this context is, for instance, the so-called nitrogen-rule: this rule allows to determine whether the number of present nitrogen atoms in an organic molecule is an odd number or an even number [197]. According to this rule, the number of nitrogen atoms is an odd number if the nominal mass of the molecule is an odd number – and, *vice versa* in the case of even numbers. The rule can be applied for every molecule that features exclusively hydrogen, carbon, nitrogen, oxygen, silicon, phosphorus, sulfur, and halogens. From the isotope pattern the *m/z* value of the monoisotopic peak and the charge-state of the molecular ion can be obtained, which in turn allows to determine the mass of the molecular ion (i.e. also mass of a glycopeptide). Critical instrument parameters in this context are the resolution and the mass accuracy achieved by the mass spectrometer: the higher the resolution and the mass accuracy, the more accurate the mass of the molecular ion can be determined (and the more reliable the molecular ion can be

identified and differentiated from other molecules). Determining the exact mass of the glycopeptide on MS^1 level may already be sufficient to reliably identify the glycopeptide in some instances, e.g. during IgG Fc *N*-glycopeptide profiling. The acquired mass of the glycopeptide is the sum of the mass of the peptide moiety plus the mass of the glycan moiety. Both values can be calculated *in silico* based on the theoretical masses of the building blocks, i.e. the amino acids and monosaccharides – in conjunction with the peptide sequence, amino acid modifications, and the glycan composition (all three need to be known). To identify the glycopeptide, the theoretical glycopeptide mass is compared to the measured mass, while considering the mass error of the mass spectrometer (for increased confidence the theoretical isotope pattern and/or the nitrogen-rule can be included, too). Identifying and quantifying a peptide or glycopeptide in this way is termed (glyco)peptide mass fingerprinting. This term is usually linked to MALDI-TOF-MS analyses. If the same type of analysis is performed via LC-ESI-MS the term (glyco)peptide mapping is used, as the chromatographic peak shape of the peptides and glycopeptides will also be included in the analysis. Glycopeptides can not only be analyzed on MS^1 level. If more in-depth information on the structure or composition of the glycopeptides is required, the glycopeptide can be subjected to analyses on MS^2 (MS/MS) level or higher within the mass spectrometer – a process termed tandem MS or multi-stage MS, respectively. During MS^2 analyses a defined amount of energy is transferred to the glycopeptide in the gas-phase. Depending on the amount and type of transferred energy the glycopeptide dissociates (or fragments) into larger or smaller subsets. The outcome of the fragmentation, thereby, also depends on the strength of the chemical bonds between and within the building blocks. The energy required for the fragmentation can be provided as collisional, vibrational, photochemical, chemical, or thermal energy (and combinations thereof). For an MS^2 analysis of a glycopeptide, first, a precursor-ion scan is conducted; subsequently, the precursor mass is manually or automatically selected and eventually fragmented. The corresponding spectrum is termed MS^2 fragment ion spectrum (or fragment ion scan). If required, further MS^n analyses can be conducted similarly; i.e. fragmentation of molecular ions detected in the MS^2 fragment ion spectrum produces MS^3 fragment ion spectra.

The sum of all signals detected in a precursor or fragment ion spectrum over the entire chromatographic retention time is represented by the total ion chromatogram (TIC). If only a selected m/z value, e.g. the m/z value of a particular fragment ion, is extracted from the TIC, the resulting chromatogram is termed extracted ion chromatogram (EIC, or XIC). These days, the most frequently used fragmentation techniques for glycopeptide analysis are low-energy collision-induced dissociation (CID), higher-energy collision dissociation (HCD), electron-capture or electron-transfer dissociation (ECD or ETD), and electron-transfer/higher-energy collisional dissociation (EThcD). MS^2 fragmentation of glycopeptides using CID primarily yields fragment ions derived from glycosidic cleavages within the glycan moiety. The resulting fragment ions are so-called B- and Y-ions (Figure 17).

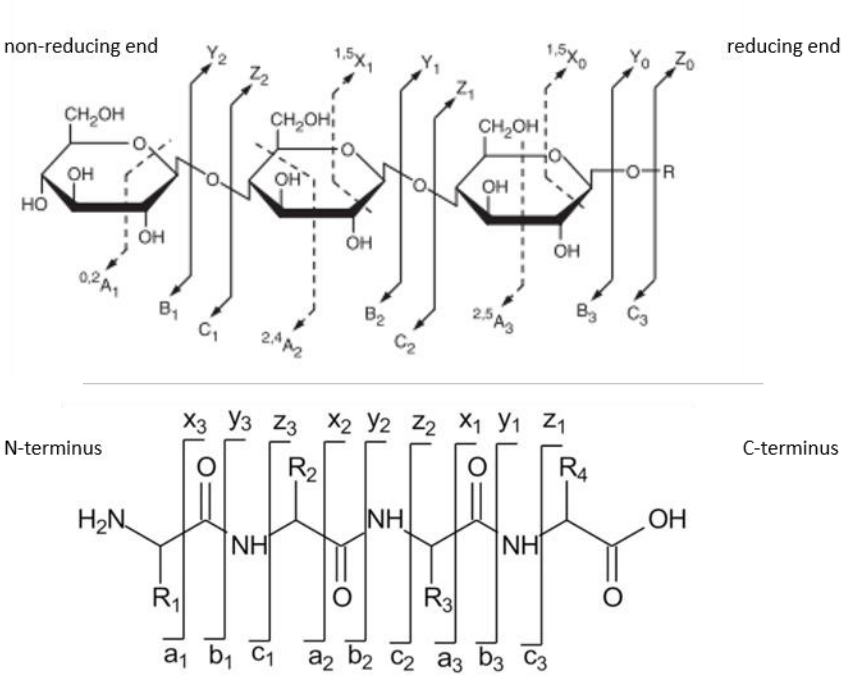


Figure 17: Nomenclature for ions generated during fragmentation of glycans (*top*) (Domon and Costello, 1988 [198]) and peptides (*bottom*) (Biemann *et al.*, 1990 [199]).

B-ions indicate fragment ions derived from the nonreducing end(s) of the glycan. These ions are also known as oxonium ions. Y-ions are derived from the reducing end of the glycan and contain the intact peptide moiety or subsets of it. The peptide backbone, in contrast, usually remains intact under lower-energy CID fragmentation conditions. Hence, in CID MS^2 fragment ion spectra of glycopeptides, fragment ions derived from the peptide backbone, i.e. b- and y-ions, are normally underrepresented (Figure 17). CID MS^2 fragmentation of glycopeptides thus usually only provides sufficient information on the glycan moiety (usually restricted to the glycan composition, though); whilst unambiguous identification of the peptide moiety can only rarely be achieved. To still obtain information on the peptide identity via CID fragmentation, CID MS^3 experiments can be conducted on fragment ions corresponding to the peptide mass, or peptide plus HexNAc – normally, at least one of these two fragment ions can be detected with high reliability in low-energy CID MS^2 fragment ion spectra of glycopeptides. The resulting CID MS^3 fragment ion spectra then may allow sequencing of the peptide moiety via corresponding b- and y-ions. Unlike, the name suggests, HCD fragmentation allows to apply both – high, but also low – normalized collisional energy (NCE). Low-energy HCD MS^2 fragmentation of glycopeptides yields fragment ion spectra that resemble those obtained by CID: that is, resulting fragment ion spectra are mainly populated with glycan-derived B- and Y-ions. High-energy HCD, in contrast, primarily yields peptide-derived b- and y-ions and glycan-derived oxonium ions, but usually fewer glycan-derived Y-ions. Also glycan cross-ring fragment ions (A- and X-ions) can be detected via high-energy HCD. Varying the HCD NCE values, can also be applied sequentially during the same fragmentation duty cycle. This, so-called stepped-energy HCD fragmentation, produces hybrid fragment ion spectra that contain information acquired at different NCE values [200, 201]. An alternative MS^2 fragmentation

technique is ETD or ECD, respectively. During this type of fragmentation, the glycan moiety bound to the modified amino acid largely remains intact. The observed fragment ions, termed c- and z-ions, are produced by cleavage of the peptide N–Cα bond and thus may provide peptide sequence information (Figure 17). Maybe even more important, those c- and z-ions may also allow to pinpoint the occupied glycosylation site(s) – a feature that is particularly crucial for the analysis of glycopeptides that do not contain a conserved sequence motif for the glycan attachment, e.g. mucin-type *O*-glycopeptides. EThcD combines ETD with HCD [202]. Upon detection of glycan-specific oxonium ions in HCD MS^2 fragment ion spectra, ETD fragmentation on the very same glycopeptide precursor ion is triggered to retrieve corresponding peptide sequence information on the glycopeptide and to pinpoint the occupied glycosylation site(s). An emerging fragmentation technique in the field of glycoproteomics is ultraviolet photodissociation (UVPD) [203, 204]. Similar to stepped HCD, UVPD may generate glycan- as well as peptide-derived fragment ions, thus providing information on both entities.

The generation of peptides and glycopeptides form the intact glycoprotein can be done enzymatically as well as chemically. The enzyme trypsin is the most frequently used protease in glycoproteomics (Figure 16). Using trypsin to generate (glyco)peptides has become the gold standard in bottom-up proteomics. Since glycoproteomics evolved from proteomics, trypsin is also adopted in the majority of glycoproteomic workflows. Trypsin cleaves specifically c-terminal to the amino acids R and K (even if these amino acids are followed by proline) [205]. Peptides and glycopeptides generated by a tryptic digest have an average length of 14 amino acid, a length that is well-suited for C18-based chromatographic separations (the average peptide number has been deduced from an *in silico* proteolytic digest of human proteins listed in the UniProt database [206]). This reasonable peptide length, along with the presence of at least two positive charges – present at the N-terminus and C-terminal R or K that enhances the ionization process in the positive ionization mode – render tryptic peptides highly amenable to (LC-)MS measurements (higher charge states, i.e. $\geq z = 2$, are favorable for the detection and fragmentation of peptides). Still, at times, the peptide moiety of tryptic glycopeptides can be too long and can harbor more than one potential glycosylation site, rendering the identification and characterization of those glycopeptides difficult, if not impossible. This problem can be circumvented, by using proteases with none or broad cleavage specificity, such as pronase E and proteinase K. Those proteases can generate short and distinct glycopeptides – even in densely glycosylated regions as found in mucin-type *O*-glycosylated glycoproteins [207]. The use of those enzymes can thus enable the identification and characterization of glycosylation sites and respective glycoforms that otherwise would have evaded the analysis. Due to the broad peptide cleavage specificity, this advantage however comes at the expense of increased ambiguity and high search times during software-assisted peptide identification. Besides enzymes, also chemicals, such as cyanogen bromide, can be used for the proteolytic digest. However, those chemicals are usually very toxic and therefore not commonly used in glycoproteomics. The proteolytic digest itself can be carried out in-solution, in-gel, or filter-assisted. Thereby, the proteases can be present in-solution or can be bead-immobilized; though

the latter has only rarely been used up to now. During an in-solution digest all reactions take place in one vessel – including reactions such as urea-mediated denaturation of the protein that can also affect the efficacy of the proteolytic enzyme. Additionally, an in-solution digest normally requires desalting of the generated peptides prior to the MS analysis. An in-gel digest avoids these aggravating circumstances, since the protein to be digested is embedded in an acrylamide gel matrix, and hence each reaction can be carried out individually with intermitted washing steps. In addition, gel-based separation techniques, such as sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) or two-dimensional gel electrophoresis (2D-PAGE), performed prior to a proteolytic digest, allow to separate mixtures of proteins, and to selectively digest a single protein, protein isoform, or subdomain of a protein (e.g. IgG light chain). However, the in-gel digest also faces some downsides compared to an in-solution digest: first, the amount of protein that can be separated is a limiting factor; further, the digest efficacy and the peptide recovery are affected by the gel matrix. A third approach often employed for the proteolytic digest in glycoproteomics is the filter-aided sample preparation (FASP) [208]. This approach combines the benefits of an in-gel digest – purification of the target protein(s) by removal of the background matrix – with those of an in-solution digest – high amount of protein sample and ease of use.

An essential step in many glycoproteomic workflows is the selective enrichment of glycopeptides prior to the MS analysis (Figure 16). This additional step is required to overcome sensitivity issues caused by the lower ionization efficiency and lower abundance of glycopeptides compared to their non-glycosylated counterparts; the latter being the result of the glycan micro- and macroheterogeneity. For the glycopeptide enrichment a range of LC- and solid phase extraction (SPE)-based methods can be employed [209], e.g. lectin affinity chromatography [77], HILIC [210], or SPE-based capturing via hydrazide chemistry [211].

Glycoproteomic analyses, as well as proteomic analyses in general, can be grouped into explorative and targeted analyses. Explorative approaches are optimized for comprehensiveness and sensitivity, and can therefore only be applied to a limited number of samples. Targeted approaches, in contrast, can be applied to a large number of samples as they are tailored for the sensitive and accurate analysis of only a limited number of features of each sample. Consequently though, it is impossible to optimize all three features – sensitivity, comprehensiveness, and throughput – at the same time during proteomics or glycoproteomics experiments (an issue referred to as "the proteomics conflict", Figure 18).

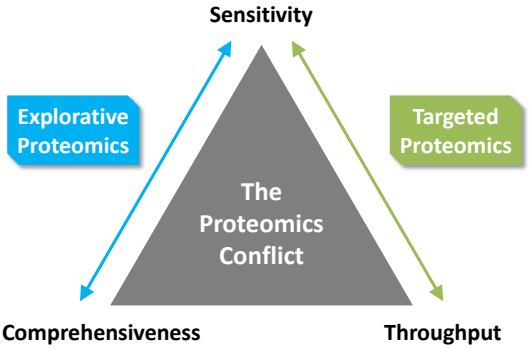


Figure 18: The proteomics conflict. It is impossible to optimize sensitivity, comprehensiveness and throughput simultaneously (taken from [212]).

Depending on the application or scientific question both or only either of the two general approaches can applied. As an example, results obtained from an explorative study of a representative sample of a particular cohort, can pave the way for a targeted high-throughput assay applicable to the entire cohort.

In many glycoproteomic studies, quantitative information on the glycan micro- and macroheterogeneity are required. However, an accurate label-free quantitation of glycopeptides by MS is extremely challenging [213]. This can be mainly addressed to different ionization efficiencies between glycopeptides and their corresponding non-glycosylated counterparts, as well as to different ionization efficiencies between glycoforms of the very same glycopeptide, e.g. between a sialylated *N*-glycopeptide and its non-sialylated counterpart [213]. Moreover, also the kinetic of the proteolytic enzyme, i.e. the digestion efficiency and specificity, is a crucial factor for MS-based glycopeptide quantitation. A promising approach for the absolute quantitation of glycopeptides is the use of multiple reaction monitoring (MRM) during LC-MS/MS [214, 215]. This approach can also be applied for the analysis of monosaccharides and released glycans, and is usually conducted on a triple quadrupole MS. MRM offers high sensitivity and selectivity, and allows the quantification of trace-amounts of glycopeptides that can be even present in a complex matrix such as blood plasma. However, establishing a reliable and valid MRM-assay is very challenging and time-consuming, and therefore not very common, yet. It is important to note that glycopeptide enrichment strategies compromise or even preclude accurate glycopeptide quantitation (glycan macroheterogeneity), due to the depletion of non-glycosylated peptides.

The analysis of glycoproteomics data, i.e. annotation and interpretation of glycopeptide precursor and fragment ion spectra, is usually challenging and time-consuming (Figure 16). However, with the advent of specialized software tools, powerful search engines, and databases, the data analyses became less cumbersome and error-prone in recent years [45, 216-220]. For example, software tools such as Proteinscape and Byonic now render the analysis and interpretation of *N*- and *O*-glycopeptide more reliable, more comparable, and less time-consuming [221]. Still, in many cases, it is recommended or required to manually validate or re-analyze the results suggested by these software packages, e.g. to avoid misinterpretation of the

data by false-positive identifications. Moreover, in some instances it is advisable or necessary to complement the acquired glycoproteomic results with results obtained from an orthogonal glycomics approach. Usually, only by combining glycomics and glycoproteomics approaches a full picture of the glycosylation of a glycoprotein can be obtained [222-224].

3 Chapter Three

Materials and Methods

3.1. Human Blood Plasma *O*-Glycoproteomics

In the following, information on the experimental procedures for Chapter 4 "Human Blood Plasma *O*-Glycoproteomics" are given. Please note, parts of this section are taken from the original publication Hoffmann *et al.* (2016) [225].

All chemicals and solvents were of the highest purity available. Purified water used for sample preparation and HILIC fractionation was freshly prepared using a Milli-Q water purification system (referred to as "Milli-Q water", 18.2 MΩ·cm^{-1} at 25°C, Total Organic Carbon 3 ppb; Merck Millipore, Darmstadt, Germany). For preparation of LC-MS solvents, ultrapure water was used, which was freshly prepared using the same system but equipped with an additional LC-Pak polisher (referred to as "Milli-Q water MS"; LC-Pak Polisher, Merck Millipore, Darmstadt, Germany).

3.1.1. Proteolytic Digestion

Human blood plasma (pooled sample, derived from 20 healthy donors) was purchased from Affinity Biologicals Inc. (VisuCon-F, frozen normal control blood, FRNCP0125; Ancaster, ON, Canada). To 25 µL of the sample (about 2 mg protein), 25 µL 100 mM ammonium bicarbonate$_{(aq)}$ (NH_4HCO_3, pH 8.0) (Sigma Aldrich, Steinheim, Germany) was added to obtain a final concentration of 50 mM $NH_4HCO_{3(aq)}$ (pH 8.0). Disulfide bonds were reduced by addition of 6.25 µL 100 mM 1,4-dithiothreitol (DTT; Sigma Aldrich) dissolved in 50 mM $NH_4HCO_{3(aq)}$ (pH 8.0), to a final concentration of 10 mM DTT. The sample was incubated for 45 min at 60°C, and subsequently allowed to cool to room temperature (RT). Cystein alkylation was achieved by addition of 12.5 µL 100 mM iodoacetamide (IAA; Sigma Aldrich) dissolved in 50 mM $NH_4HCO_{3(aq)}$ (pH 8.0) to a final concentration of 16.67 mM IAA. The sample was incubated at RT for 20 min under light exclusion. The alkylation reaction was quenched by addition of 2.5 µL 100 mM DTT dissolved in 50 mM $NH_4HCO_{3(aq)}$ (pH 8.0), followed by addition of 3.75 µL 50 mM $NH_4HCO_{3(aq)}$ (pH 8.0), before placing the sample under a fluorescent lamp (gas-discharge lamp) for 15 min to decompose the light-sensitive IAA. By adding 169 µL 50 mM $NH_4HCO_{3(aq)}$ (pH 8.0), the sample was brought to a final volume of 250 µL.

Proteinase K Digestion

Proteolytic digestion was achieved by addition of proteinase K (Sigma Aldrich), a serine protease with a broad specificity that cleaves primarily after aliphatic, aromatic and hydrophobic amino acids. The pooled blood plasma sample (about 2 mg protein in 250 µL buffer) was supplemented with 66 µg proteinase K dissolved in 122 µL 50 mM $NH_4HCO_{3(aq)}$ (pH 8.0) to obtain a final enzyme to protein ratio of 1:30 (w/w, 0.033 mg enzyme per mg protein). The sample was incubated for 16 h at 37°C with gentle agitation (200 rpm).

Acetonitrile Precipitation

For post-digestion cleanup the sample was precipitated using acetonitrile (ACN; Sigma Aldrich). To this end, four volumes of ACN were added and the sample was centrifuged for

10 min at 2,880 × g (Centrifuge 5804 R; Eppendorf, Hamburg, Germany). The supernatant was transferred and dried by vacuum centrifugation (RVC 2-33 CDplus, ALPHA 2-4 LDplus; Martin Christ GmbH, Osterode am Harz, Germany).

3.1.2. Glycopeptide Enrichment and Fractionation via HILIC-HPLC

The dried proteinase K digest was resuspended in 500 µL 80% ACN in 50 mM $NH_4HCO_{3(aq)}$ (v/v, pH 8.0) and subsequently centrifuged for 10 min at 20,238 × g to remove any particles (Centrifuge 5424; Eppendorf). The supernatant, containing about 2 mg peptides and glycopeptides, was subjected to HILIC-HPLC (UltiMate 3000 Nano HPLC-System: Thermo Scientific/Dionex, Dreieich, Germany; HILIC Column: ACQUITY UPLC BEH HILIC Column, 130Å, 1.7 µm, 2.1 mm X 100 mm; Waters, Manchester, UK) for fractionation and glycopeptide enrichment.

The HPLC system was operated using a binary gradient of 100% ACN (v/v; solvent A) and 50 mM ammonium $formate_{(aq)}$ (NH_4FA, pH 4.4; solvent B, Sigma Aldrich). After sample injection (500 µL) 20% solvent B was applied isocratically for 5 min, followed by a linear gradient to 50% solvent B within 25 min, both using a constant flow rate of 250 µL/min. Subsequently, a linear gradient went to 90% solvent B within 1 min, while reducing the flow rate to 150 µL/min. To wash the column solvent B was kept at 90% for 9 min. Column re-equilibration was achieved by isocratic elution with 20% solvent B for 20 min; (the flow rate was increased to 250 µL/min after 10 min). During the separation, the column temperature was kept constant at 40°C. The elution profile was monitored by UV absorption at 214 nm. Fractions were collected every two minutes from 0 min to 34 min. The fractions were dried by vacuum centrifugation and reconstituted in 50 µL Milli-Q water.

3.1.3. Nano-RP-LC-ESI-IT-MSn (CID, ETD)

HILIC fractions were analyzed by reversed-phase nano-LC-MS^n using an Ultimate3000 nanoHPLC system (Thermo Scientific/Dionex) coupled online to an ion trap mass spectrometer (AmaZon ETD, Bruker Daltonics, Bremen, Germany). Within the first two minutes after sample injection, (glyco)peptides were loaded isocratically on a C18 µ-precolumn (Acclaim PepMap100, C18, 5 µm, 100 Å, 300 µm i.d. x 5 mm; Thermo Scientific/Dionex). During this pre-concentration and desalting step, "loading pump solvent 1" (98% Milli-Q water MS, 2% ACN, 0.05% trifluoroacetic acid (TFA; Sigma Aldrich)) was used at a flow rate of 7 µL/min. Subsequently, the C18 µ-precolumn was switched in line with the C18 nano-separation column (Acclaim PepMap RSLC, C18, 2 µm, 100 Å, 75 µm i.d. x 15 cm; Thermo Scientific/Dionex) for gradient elution. Here, the following solvents were used at a constant flow rate of 300 nL/min: "A" (98% Milli-Q water MS, 2% ACN, 0.1% formic acid (FA; Sigma Aldrich)); "B" (10% Milli-Q water MS, 10% 2,2,2-trifluoroethanol (TFE; Merck, Darmstadt, Germany), 80% ACN, 0.1% FA (Sigma Aldrich)). A binary gradient was applied as follows: 4% B for 2 min; linear gradient to 30% B within 30 min; isocratic washing step at 90% B for 5 min, finally 20 min re-equilibration at 4% B. After 42 min the C18 µ-

precolumn was switched back into loading-pump flow, to be washed for 3 min at 100% "loading pump solvent 2" (20% Milli-Q water MS, 80% ACN, 0.05% TFA (Sigma Aldrich)), and eventually to be re-equilibrated for 15 min at 100% "loading pump solvent 1", both at 7 µL/min flow rate (Figure 19). In this setup the gradient applied for on the C18 nano-separation column was optimized for the separation of glycopeptides, and therefore differs from a gradient normally applied for the separation of peptides in classical proteomics experiments (Figure 19, B [%] Nano Pump, hatched area). With 30% B after 33 min, instead of 50% B, the gradient is more shallow. That way, the reduced ACN concentration enables a better separation of the more hydrophilic and therefore earlier eluting glycopeptides.

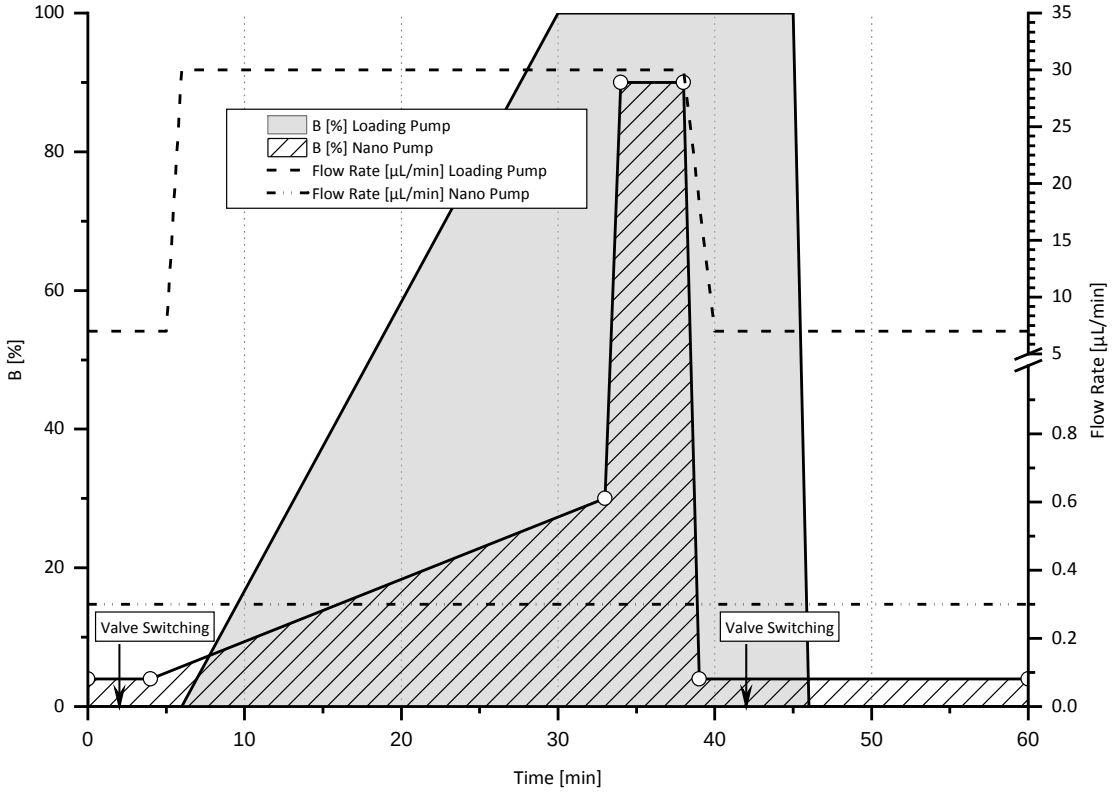


Figure 19: Separation of *O*-glycopeptides via nano-RP-LC: applied solvent gradients (%B), flow rates, and valve switching times.

The ion trap mass spectrometer was interfaced with a nanoFlow ESI Sprayer (Bruker Daltonics) and was operated in positive ion mode. For electrospray ionization the following parameters were used: capillary voltage (-4,500 V), end plate offset (-500 V), N_2 dry gas (5 L/min), nebulizer (8 psi), dry gas temperature (220°C). The (glyco)peptides were fragmented via CID using multistage fragmentation (CID-MS^2, CID-MS^3 experiments) and ETD-MS^2. Fluoranthene radical anions were generated by negative-mode chemical ionization using methane as mediator.

CID experiments were carried out using the following precursor scan settings: precursor scan mass range (*m/z* 100-2,500); ion charge control (ICC) target (300,000); maximum accumulation time (200 ms); averages (5); rolling averaging (off); target mass for smart

parameter settings (*m/z* 850). CID-MS² experiments were conducted using a data-dependent fragmentation routine. The top four most intense precursor ions, in the range of *m/z* 500-1,500, were subjected to CID fragmentation in the ion trap mass analyzer (MS/MS fragmentation amplitude 1.20 V). The relative intensity threshold for fragmentation was set to 5%. Singly charged ions were excluded and selected precursors were actively excluded for 0.15 min after acquiring two fragment ion spectra. Charge state preference was set to "none". Recorded scan range, ICC target and maximum accumulation time were the same as for the precursor scan. In CID-MS³ experiments precursor selection and fragmentation was applied manually. The fragmentation amplitude was set to 1.20 V. The recorded scan range was set individually with respect to the *m/z* of the precursor. ICC target and maximum accumulation time were the same as for the precursor scan. In both CID-MS² and CID-MS³ experiments the following CID parameters were used: cut-off selection (default); smart fragmentation (on); start amplitude (30%); end amplitude (200%); reaction time (40 ms). All CID experiments were carried out using the enhanced resolution mode. For CID-MS² measurements, 1 µL of each HILIC fraction was injected. For CID-MS³ measurements, 5 µL were used, respectively.

ETD experiments were carried out using the following precursor scan settings: precursor scan mass range (*m/z* 400-2,500); ICC target (200,000); maximal accumulation time (50 ms); averages (5); rolling averaging (on, number: 1); target mass for smart parameter settings (*m/z* 850), enhanced resolution mode. Fragment spectra were acquired using a data-dependent fragmentation routine in the ultra scan mode. The top three most intense precursor ions were subjected to ETD fragmentation in the ion trap mass analyzer. The relative intensity threshold for fragmentation was set to 1%. Singly charged ions were excluded and selected precursors were actively excluded for 0.15 min after acquiring two fragment ion spectra. Charge state preference was set to "none". Fragment ions between *m/z* 100-3,000 were detected. ICC target was set to 400,000 and maximal accumulation time was set to 100 ms. The following parameters were used for the EDT reagent: ICC target (500,000); max. accu. time (10 ms); Remove <= *m/z* 210 (On); Max. ETD Precursor (*m/z* 1,200), cut-off (*m/z* 160); reaction time (160 ms); smart decomposition (auto). For ETD measurements, 1 µL of each HILIC fraction was injected.

All MS parameters were tested and optimized using *N*-glycopeptides derived from human IgG (sample preparation according to Selman *et al.* [226]) and *O*-glycopeptides from erythropoietin (Protea Biosciences, Morgantown, WV, USA) (data not shown).

3.1.4. Data Analysis

Two separate strategies were followed for the spectra analysis. The first approach focused exclusively on the identification of non-glycosylated peptides, while the second approach aimed for the characterization and identification of the glycopeptides.

I) Analysis of Non-Glycosylated Peptides

Fragment spectra (MS^2) acquired with CID and ETD were searched for non-glycosylated peptides. To this end, spectra were processed in DataAnalysis software 4.0 (Bruker Daltonics) using a built-in function for MS^n spectra processing ("processautomsn"; compound detection: standard settings). Processed spectra were imported into ProteinScape 3.1 (Bruker Daltonics) and were searched against a UniProtKB/Swiss-Prot database (SwissProt 51.6; 257964 sequences; 93947433 residues; downloaded February, 2013) using MASCOT version 2.2.04 (Matrix Science, London, UK). The following search parameters were applied: taxonomy (human); enzyme (none); fixed modifications (carbamidomethylation of cysteine residues); variable modifications (deamidation of asparagine and/or glutamine; methionine oxidation); precursor ion mass tolerance (±0.3 Da, with # ^{13}C=1; monoisotopic mass); fragment ion mass tolerance (CID: ±0.5 Da; ETD: ±1.3 Da); preferred charge state ($2^+/3^+$); peptide decoy search (1% FDR). Proteins and peptides with a MASCOT ion score higher than 50 (for proteins) and 25 (for peptides) were accepted, respectively.

II) Analysis of Glycopeptides

CID and ETD fragment ion spectra (MS^2/MS^3) were manually analyzed assisted by the DataAnalysis software 4.0 (Bruker Daltonics) without any pre-processing. Fragmentation of glycopeptides using low-energy CID almost exclusively yields fragment ions derived from the glycan moiety. This allows filtering of CID-MS^2 spectra for the presence of low-molecular weight fragment ions derived from the non-reducing end of the glycan [198] (B-ions, oxonium ions; $[M+H]^+$; e.g. Hex *m/z* 163.06; NeuAc -H_2O *m/z* 274.09; NeuAc *m/z* 292.10; Hex_1NeuAc_1 *m/z* 454.16; $HexNAc_1Hex_1NeuAc_1$ *m/z* 657.24; tolerance: *m/z* ±0.3) using dedicated EICs. In addition to this CID-MS^2 glycopeptide spectra feature multiply charged fragment ions (Y-ions) that show characteristic mono(oligo)-saccharide mass differences caused by the consecutive fragmentation of the glycan moiety down to the de-glycosylated peptide. Both features were used to deduce the glycan composition along with the putative peptide mass in CID-MS^2 glycopeptide spectra. To identify the peptide moiety, the putative peptide mass was used to trigger manual CID-MS^3 fragmentation in a separate run. In rare cases, the peptide mass with an additional HexNAc had to be used for CID-MS^3 fragmentation. CID-MS^3 fragment ion spectra were exported to BioTools software 3.2 (Bruker Daltonics). Subsequent peptide identification was conducted using MASCOT. The spectra were searched against a UniProtKB/Swiss-Prot database using the following parameters: taxonomy (human); enzyme (none); fixed modifications (carbamidomethylation of cysteine residues); variable modifications (deamidation of asparagine and/or glutamine; methionine oxidation); precursor ion mass tolerance (±0.3 Da, with # ^{13}C=1; monoisotopic mass); fragment ion mass tolerance (CID: ±0.35 Da); preferred charge state ($2^+/3^+$); MASCOT significance threshold (0.05); maximum number of reported hits: 10.

Peptides with a MASCOT ion score greater than 20 were accepted; in very rare cases also lower scored peptides were accepted. Peptide identification was supported by the presence of

a glycosylation consensus motif within the putative peptide sequence (*N*-glycosylation: N,X,S/T; *O*-glycosylation: S/T). Furthermore, knowledge derived from public databases (UniProtKB and UniCarbKB) on already described *N*-/*O*-glycosylation sites within the putative peptide sequence or within the entire protein was used to validate a peptide/protein hit. ETD-MS² fragment ion spectra of identified and characterized glycopeptides were annotated manually with respect to the presence of glycan fragment ions (Y-ions). Subsequently, the spectra were exported to BioTools software 3.2 (Bruker Daltonics) to identify the glycosylation site(s). The peptide sequences, proposed by CID-MS³ measurements, were modified *in silico* with the corresponding glycan compositions inferred from CID-MS², taking in account all the potential glycosylation sites. Fragment ions (c- and z-type ions) derived from these *in silico* glycopeptide sequences were then matched to their counterparts in the measured ETD-MS² spectra. The accuracy of this annotation was validated using the BioTools score along with manual inspection of the respective spectra. The entire glycopeptide data analysis workflow is briefly summarized in Figure 20.

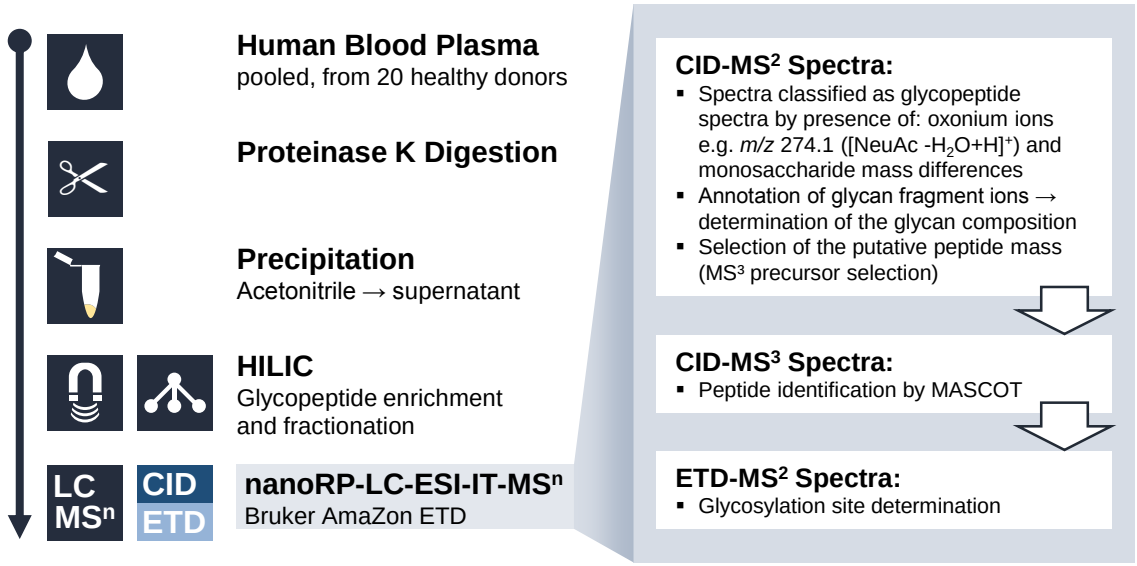


Figure 20: *O*-glycoproteomic workflow for the analysis of human blood plasma glycoproteins.

All mass spectrometry raw data as well as (glyco)peptide identifications and spectra annotations have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the MassIVE repository with the dataset identifier PXD002315 or MSV000079141, respectively *(ftp://MSV000079141@massive.ucsd.edu)*.

3.2. In-Depth *N*- and *O*-Glycoproteomics

In the following information on the experimental procedures for Chapter 5 "In-Depth *N*- and *O*-Glycoproteomics" are given. Please note, parts of this section are taken from the original publication Hoffmann *et al.* (2018) [227].

All chemicals and solvents were of the highest purity available. Purified water used for sample preparation and for LC-MS measurements was prepared as described before. Aqueous solutions are indicated with (aq).

3.2.1. Proteolytic Digestion via Filter-Aided Sample Preparation (FASP)

Commercially available standard glycoproteins of different origin were proteolytically digested using a modified version of the FASP method introduced by Wisniewski *et al.* [208]. Fibrinogen (Fib), IgG – both derived from human plasma, Lactotransferrin (LTF) – derived from human milk, and ribonuclease B (RNase B) – derived from bovine pancreas, were purchased from Sigma-Aldrich (# F3879; # I4506-10MG; # L4894-5MG; R7884-100MG; Steinheim, Germany). Briefly, 50 µg of each protein were dissolved in phosphate-buffered saline (PBS) pH 7.4, transferred to a centrifugal filter unit (Nanosep® Omega™ with polyethersulfone membrane, molecular weight cut-off 10 kDa; # OD010c34, PALL Life Sciences, Ann Arbor, MI, USA), and centrifuged for 10 min at 14,000 x g at RT (holds true for all subsequent centrifugation steps, Heraeus™ Fresco™ 17 Microcentrifuge, 24 x 1.5/2.0 mL rotor, Thermo Scientific, Waltham, MA, USA). The flow-through was discarded. Subsequently, samples were re-dissolved in 200 µL of urea buffer$_{(Tris\text{-}HCl)}$ (8 M urea in 0.1 M Tris-HCl$_{(aq)}$ pH 8.5; # A1049, # A3452, AppliChem, Darmstadt, Germany), incubated at RT for 5 min while constantly shaking at 600 rpm on a Thermomixer comfort (Eppendorf, Hamburg, Germany), and centrifuged again. Cysteine disulfide bonds were reduced with DTT (# D5545-5G, Sigma-Aldrich), and subsequent carbamidomethylation of -SH groups (thiol) was carried out using IAA (# I1149-25G, Sigma-Aldrich). DTT (24.68 mg) and IAA (40.7 mg) were dissolved in 50 mM ammonium bicarbonate$_{(aq)}$ (ABC buffer; # 09830-500G, Sigma-Aldrich) and diluted tenfold with urea buffer$_{(Tris\text{-}HCl)}$ to a final concentration of 40 mM DTT and 55 mM IAA, respectively. To each filter unit, 100 µL of 40 mM DTT were added followed by shaking at RT for 1 min and 600 rpm. After incubation at 56°C for 20 min with gentle agitation (300 rpm), the samples were centrifuged and the flow-through was discarded. Subsequently, 100 µL of 55 mM IAA were added to each filter. Protected from light, samples were shaken at RT for 1 min and 600 rpm, and allowed to incubate for 20 min at RT. After centrifugation, the flow-through was discarded. Each filter unit was washed three times with urea buffer$_{(Tris\text{-}HCl)}$ and then three times with ABC buffer$_{(aq)}$ according to the following scheme: addition of 100 µL of buffer, incubation for 5 min at RT while constantly shaking at 600 rpm, centrifugation, discarding the flow-through. Finally, the filter units were transferred into new 2 mL microcentrifuge tubes. Proteins were proteolytically digested either with trypsin (Sequencing Grade Modified Trypsin; # V5111, Promega, Madison, WI, USA) using an enzyme to protein ratio of 1:30 (w/w, 0.033 µg enzyme per µg protein, in total 1.67 µg

enzyme), or proteinase K (# A4392, AppliChem) using an enzyme to protein ratio of 1:10 (w/w corresponding to 0.1 µg enzyme per µg protein, in total 5 µg enzyme). Before addition to the filter units, both enzymes were brought to a final volume of 100 µL by dilution with ABC buffer$_{(aq)}$ + 1 mM calcium chloride ($CaCl_{2(ABC\ buffer)}$) + 5% (v/v) $ACN_{(aq)}$ ($CaCl_2$, # A4689, AppliChem; ACN, LC-MS CHROMASOLV®, # 34967-1L, Sigma-Aldrich). Samples were shaken at RT for 1 min and 600 rpm before incubation overnight at 37°C and 350 rpm using a temperature-controlled incubator (Titramax 1000 + Inkubator 1000, both Heidolph, Schwabach, Germany). Digests were collected by centrifugation. Filter units were washed twice, first using 50 µL ABC buffer$_{(aq)}$ + 5% (v/v) $ACN_{(aq)}$, then using 50 µL Milli-Q water; in between samples were centrifuged. The flow through was kept along with the digest for subsequent drying by vacuum centrifugation (RVC 2–33 CDplus, ALPHA 2–4 LDplus, Martin Christ GmbH, Osterode am Harz, Germany). The dried digests were reconstituted in 50 µL 0.1% (v/v) $TFA_{(aq)}$ (# 28904, Thermo Fisher Scientific, Waltham, MA, USA) to a final concentration of 1 µg/µL. Samples were shaken at RT for 10 min and 1,000 rpm, and finally stored at -20°C.

3.2.2. Glycopeptide Enrichment via Spin-Cotton-HILIC-SPE

Glycopeptides were enriched using a modified version of the cotton-HILIC-SPE protocol introduced by Selman *et al.* [226]. In brief, cotton-HILIC microtips were prepared by filling 20 µL pipette tips to the 10 µL marking with cotton wool derived from commercially available cotton pads (100% cotton) (**Figure 21**).

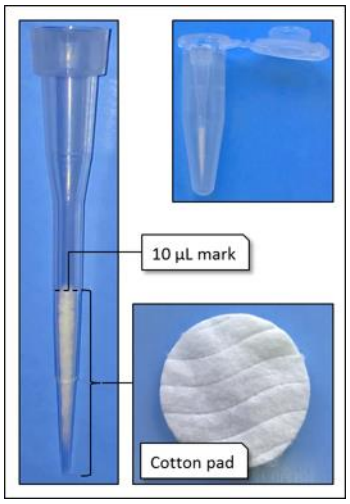


Figure 21: Spin-Cotton-HILIC-SPE. **Left:** 20 µL pipette tip filled with cotton wool taken from a commercially available cotton pad. **Top right:** Cotton-HILIC pipette tip in 1.5 mL microcentrifuge tube.

In contrast to the original protocol, centrifugation – instead of pipetting up and down – was used during washing, loading, and elution steps; i.e. liquid was pipetted on top of the cotton wool stationary phase, the cotton-HILIC microtip was placed in an 1.5 mL microcentrifuge tube, and was centrifuged for 1 min at 2,400 x g (Heraeus™ Fresco™ 17 Microcentrifuge, 24 x 1.5/2.0 mL rotor, Thermo Scientific). Cotton-HILIC microtips were washed and equilibrated three times with 20 µL Milli-Q water and 20 µL 83% (v/v) $ACN_{(aq)}$ + 0.1% (v/v)

$TFA_{(aq)}$, respectively. Each cotton-HILIC microtip was loaded with 10 µg of protein digest (dried by vacuum centrifugation and resuspended in 10 µL 83% (v/v) $ACN_{(aq)}$ + 0.1% (v/v) $TFA_{(aq)}$). For proper adsorption, the protein digest was loaded three times. The final flow-through was collected and referred to as "glycopeptide-depleted fraction". Cotton-HILIC microtips were washed three times with 20 µL 83% (v/v) $ACN_{(aq)}$ + 0.1% (v/v) $TFA_{(aq)}$. The flow-through was collected and referred to as "wash fraction". Glycopeptides were eluted three times with 20 µL of Milli-Q water each. Along with the other two fractions, the eluate – referred to as "glycopeptide-enriched fraction" – was dried, and resuspended in 20 µL of 0.1% (v/v) $TFA_{(aq)}$ while shaking at RT for 10 min and 1,000 rpm.

3.2.3. Nano-RP-LC-ESI-OT-OT-MS/MS (HCD)

HILIC-enriched (glyco)peptides were analyzed by nano-reversed-phase liquid chromatography using an UltiMate 3000 RSLCnano system (Thermo Scientific) coupled online to a linear ion trap-orbitrap hybrid mass spectrometer (nano-RP-LC-ESI-OT-OT-MS/MS; LTQ Orbitrap Elite hybrid mass spectrometer, Thermo Scientific). For each measurement, an equivalent of 4 µg protein digest were injected. Within the first four minutes after sample injection, (glyco)peptides were loaded isocratically on a C18 nano pre-column (Nano Trap Column, packed with Acclaim PepMap100 C18, 5 µm, 100Å, 100 µm i.d. x 2 cm, # 164564, Thermo Scientific). During this pre-concentration and desalting step, 98% (v/v) Milli-Q water MS, 2% (v/v) ACN, 0.05% (v/v) TFA ("loading pump solvent") was used at a flow rate of 7 µL/min. Subsequently, the C18 nano pre-column was switched in line with the C18 nano separation column (Acclaim PepMap RSLC C18, 2 µm, 100Å, 75 µm i.d. x 25 cm, # 164536, Thermo Scientific) for gradient elution. Here, the following solvents were used at a constant flow rate of 300 nL/min: "A" (98% (v/v) Milli-Q water MS, 2% (v/v) ACN, 0.1% (v/v) FA [# 56302-10X1ML-F, Sigma Aldrich]); "B" (10% (v/v) Milli-Q water MS, 10% (v/v) 2,2,2-TFE [# 808259, Merck, Darmstadt, Germany], 80% (v/v) ACN, 0.1% (v/v) FA). A binary gradient was applied as follows: 4% B for 4 min, linear gradient to 30% B for another 29 min, linear gradient to 34% B within 1 min, isocratic washing at 90% B for 4 min, finally 22 min re-equilibration at 4% B. After 42 min the pre-column was switched back into loading-pump flow to be re-equilibrated for 18 min at 100% "loading pump solvent" at 7 µL/min flow rate. The column oven temperature was kept constant at 40°C.

The mass spectrometer was interfaced with a nanoelectrospray source (Nanospray Flex™ ion source, NSI, Thermo Scientific) operated in positive ion mode. The source was equipped with nanoelectrospray emitters (SilicaTip, FS360-20-10-D-20, New Objective, Cambridge, USA). For electrospray ionization, the following parameters were used: source voltage (2.7 kV); capillary temperature (275°C); sheath gas flow, aux gas flow, and sweep gas flow were set to zero arbitrary units. Full scan spectra (MS^1) were acquired in the orbitrap mass analyzer (OT): ions between m/z 300-2,000 were recorded at a resolution of 30,000 (defined at m/z 400) using profile mode; automatic gain control (AGC) target was set to $1x10^6$; and maximum injection time was set to 500 ms. For enhanced mass accuracy, internal real-time mass

calibration was enabled using a background polysiloxane peak at *m/z* 371.1012 ($[M+H]^+$) as lock mass. Fragment ion spectra (MS^2) were obtained by data-dependent acquisition: the five most intense precursor ions with a charge state ≥2 and signal intensity ≥500 counts were subjected to HCD fragmentation (data-dependent acquisition). HCD spectra were acquired in the OT using profile mode: mass range was set to "normal"; AGC target was set to $1x10^6$, and maximum injection time was set to 500 ms. Three HCD fragmentation regimes were used, differing in applied normalized collisional energies (NCE): *HCD.low* with NCE of 20, *HCD.high* with NCE of 50, and *HCD.step* with stepped NCE of 35 (width: 15%, steps: 2). An isolation width of 4 *m/z* units was used; charge state screening as well as monoisotopic precursor selection was enabled; target ions selected for MS^2 were dynamically excluded using the following settings: exclusion size list (500), exclusion duration (5 s), repeat count (1), and repeat duration (30 s). The number of micro scans was set to one.

3.2.4. Glycopeptide Data Analysis Using glyXtool^{MS}

glyXtool^{MS} is an OpenMS- and python™-based software pipeline that was developed in-house for the semi-automated analysis of glycopeptide mass spectrometry data (Pioch *et al.* [228]) . Briefly, acquired spectra were converted to *.mzML format and subsequently processed using a dedicated analysis workflow in OpenMS. For each glycoprotein, the respective protein sequence (UniProtKB/Swiss-Prot database; Swiss-Prot 51.6; 257964 sequences; 93947433 residues; downloaded February, 2013) along with appropriate *N*- or *O*-glycan composition databases (in-house; *N*-glycan compositions: 378; core-1 & -2 mucin-type *O*-glycan compositions: 8; for mucin-type *O*-glycans only the major compositions relevant for human plasma proteins were considered: non-, mono- and disialylated glycoforms, no further modifications/elongations [229]) were implemented in the workflow. The following parameters were used for the in silico proteolytic digest: number of missed cleavages (2; only relevant for tryptic digests); cysteine carbamidomethylation as fixed modification; methionine oxidation and asparagine/glutamine deamidation as variable modifications. Fragment ion spectra were automatically classified as glycopeptide spectra based on the presence of oxonium ions (B-ions) and glycan-derived neutral loss fragment ions (Y-ions) by glyXtool^{MS}. Automatic annotation of the glycan and peptide moiety was manually validated using the glyXtool^{MS} Evaluator. A mass tolerance of ±10 ppm and ±20 ppm was accepted for precursor ions and fragment ions, respectively. For the annotation of the glycan moiety the following abbreviations/symbols are used: *N*-acetylglucosamine (GlcNAc, HexNAc, N, "blue square"), mannose (Man, Hex, H, "green circle"), fucose (Fuc, DHex, F, "red triangle"), *N*-acetylneuraminic acid (NeuAc, NANA, Sa, "purple diamond"), *N*-glycolylneuraminic acid (NeuGc, NGNA, Ng, "light blue diamond"), peptide (Pep). Further details on glyXtool^{MS} are given in Pioch *et al.* [228]. In addition to the glycopeptide-focused analysis via glyXtool^{MS}, conventional protein identification was also performed. To this end the fragment ion spectra of the untreated proteolytic digests (without HILIC enrichment) were searched against the UniProtKB/Swiss-Prot database using MASCOT 2.5.1.0 (Matrix Science, London, UK) and Proteome Discoverer 1.4.1.14 (Thermo Scientific, Waltham, MA, USA). The applied search parameters

and results can be found in supplemental table 1 in Hoffmann *et al.*, 2016 [225]. To standardize analyses, the MIRAGE consortium (minimum information required for a glycomics experiment) has specified reporting guidelines for collecting, sharing, integrating, and interpreting mass spectrometry-based glycomics and glycoproteomics data [230]. Data reported in this thesis is in agreement with these guidelines. All *N*- and *O*-glycopeptide mass spectrometry raw data (HILIC-SPE fractions) have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the MassIVE repository (ftp://massive.ucsd.edu/MSV000082137).

The entire sample preparation and data analysis workflow that was developed during this study is depicted in Figure 22.

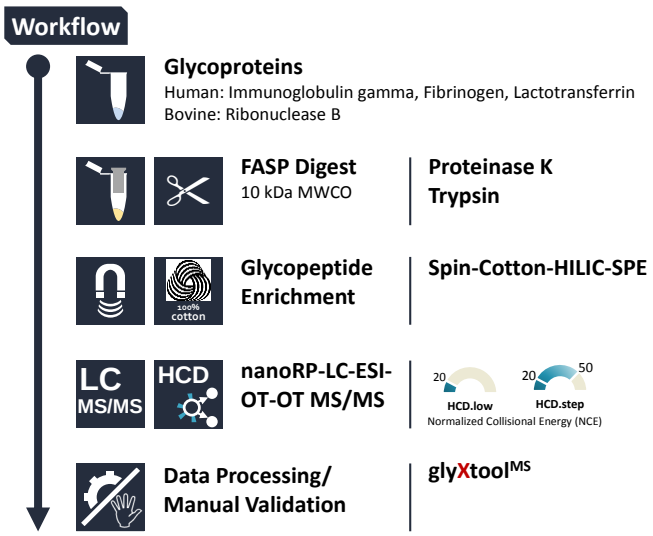


Figure 22: Glycoproteomic workflow developed for the in-depth analysis of *N*- and *O*-glycosylated proteins. A set of representative glycoproteins was proteolytically digested either with trypsin or proteinase K using a FASP approach. Resulting peptides were subjected to spin-cotton-HILIC-SPE to enrich *N*- and *O*-glycopeptides. HCD fragment ion spectra using fixed and stepped normalized collisional energy (HCD.low, HCD.step) were acquired from glycopeptide-enriched HILIC fractions. For data processing and annotations, glyXtool^{MS} was used; data validation was done manually.

3.2.5. Site-Specific Relative Quantitation of *N*-Glycoforms

To determine the site-specific relative abundance of IgG, Fib, and RNase B *N*-glycoforms (microheterogeneity), the summed peak intensities of the corresponding glycopeptide precursor ion isotopic pattern along the peak elution window were used (referred to as peak intensity). Thereby, peak integration boundaries of the respective extracted ion chromatograms were set automatically by glyXtool^{MS}. For glycopeptides registered with different precursor charge states, the respective peak intensities were summed. Generally, peak intensities were only considered for manually validated *HCD.low* or *HCD.step* fragment glycopeptide ion spectra; in case glycopeptides were identified using both methods an average value was used. The relative site-specific glycoform abundance was calculated by dividing the individual peak intensities by the sum of all peak intensities covering the same glycosylation site.

3.2.6. Relative Quantitation of *N*- and *O*-Glycopeptide Oxonium Ions

The relative oxonium ion abundance was calculated by dividing the individual oxonium ion peak intensities (monoisotopic peak) by the sum of all oxonium ion peak intensities detected in the fragment ion spectrum (relative abundance). For glycopeptides with more than one corresponding fragment ion spectrum (scans), the average relative abundance is given. Error bars indicate the standard deviation of the relative oxonium ion abundance for glycopeptides with more than two scans.

4 Chapter Four

Human Blood Plasma *O*-Glycoproteomics

4.1. Introduction

Parts of this chapter are taken from the original publication Hoffmann *et al.* (2016) [225].

Blood plasma harbors arguably the most complex yet also the most informative proteome present in the human body [231]. A significant impact on its clinical relevance and diagnostic potential is attributed to the features and functions of a plethora of proteins (60-80 mg protein/mL), covering a concentration range of more than ten orders of magnitude [232]. The majority (about 99% of these proteins) are classical blood plasma proteins, like albumins, (immuno)globulins, clotting factors, and proteins of the complement system; however, also a lower abundant but – no less meaningful – fraction of proteins is present that comprises a multitude of cytokines as well as tissue leakage proteins (Figure 23).

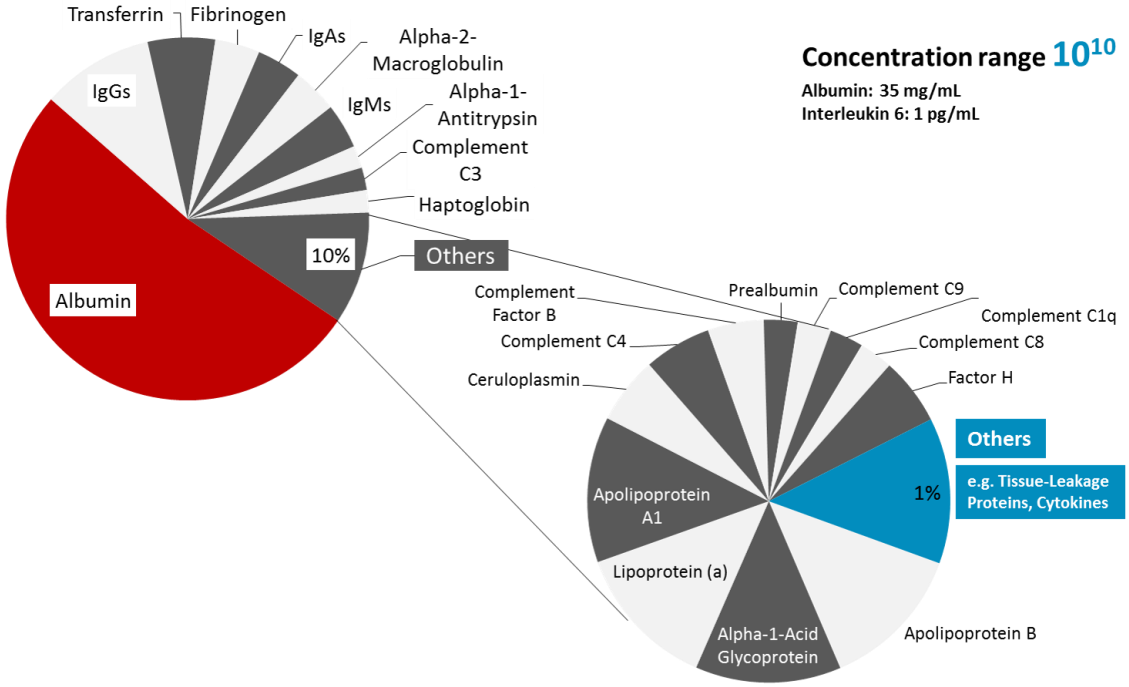


Figure 23: Relative proportion of major human blood plasma proteins (adapted from Putnam [233]).

Several clinical studies could show that qualitative and quantitative alterations of these proteins (and peptides) – analyzed individually or in their entirety as a proteome (or peptidome) – can directly reflect pathophysiological states, and can serve as biomarkers for the onset and progression of a number of diseases [234-236]. In recent years, the focus of in-depth analyses of the human blood plasma proteome has evolved from the identification and quantification of the entire proteome (or peptidome) [237-241] towards the analysis of subproteomes like the interactome [242], phosphoproteome [243, 244] or the glycoproteome [245]. The latter has received particular interest in recent years, since the majority of blood plasma proteins is *N*- and/or *O*-glycosylated [232]. While the comprehensive analysis of the *N*-glycoproteome is already quite advanced [246], even in complex samples like human blood plasma [17, 247], similar analyses of the *O*-glycoproteome - though arguably equally important and relevant - are still lagging behind. The most ubiquitously

found and functionally relevant form of *O*-glycosylation, as shown by a number of *O*-glycan-related (clinical) studies [116, 248-252], is the mucin-type *O*-glycosyation (*O*-GalNAc), in particular the core-1 and core-2 types [207, 253]. The predominantly clustered occurrence of mucin-type *O*-glycans on proteins is described to confer overall stability and proteolytic protection [254]. Apart from this global impact, recent studies could link the presence of *O*-glycans in the proximity of regulatory domains to proteolysis events involved in protein maturation (proprotein-convertase (PC)-processing) [255]. To better understand these protective and regulatory capabilities and to move the mucin-type *O*-glycoproteome from form to function, comprehensive site-specific *O*-glycosylation analyses are required.

One of the main obstacles in site-specific mucin-type *O*-glycosylation analyses relates to the lack of a predictable *O*-glycan consensus-motif within the peptide backbone as it can be found for *N*-glycans [256]. The attachment of the *N*-acetylgalactosamine monosaccharide to the hydroxyl group of either serine or threonine, but also to tyrosine or hydroxylysine, is governed by a family of 20 distinct polypeptide GalNAc-transferase isoenzymes (GalNAc-Ts) with different but partially overlapping peptide specificities and tissue expression patterns. This dynamic regulation, in turn, contributes to the complexity of the mucin-type *O*-glycoproteome. However, previous studies could show that mucin-type *O*-glycans are primarily attached to serine or threonine in regions with a high content of serine, threonine and proline (S/T-X-X-P, S/T-P and P-S/T) [257, 258]. As *O*-glycosylation is a postfolding event, taking place in the Golgi apparatus, the attachment is depended on protein surface accessibility and is thus predominantly found in coil, turn and linker regions [259]. Additional confounding factors during mucin-type *O*-glycosylation analyses are the clustered occurrence of *O*-glycans and the lack of a universal endo-*O*-glycosidase that enables the release of intact *O*-glycans from the proteins; though, chemical *O*-glycan release methods do exist [256].

Mass spectrometry has proven to be the core technique in site-specific *N*- and *O*-glycosylation analyses. A generic *O*-glycoproteomic workflow usually starts with the isolation, enrichment or pre-fractionation of a single glycoprotein or a group of glycoproteins. In subsequent steps, (glyco)peptides are generated by proteolytic digestion primarily using specific proteases like trypsin. Apart from this, also broad- and non-specific proteases like proteinase K or pronase E were employed successfully in recent years [260-262]. Essential to nearly every glycoproteomic approach is the removal of high-abundant and interfering non-glycosylated peptides by selective enrichment of the usually lower abundant glycopeptides. The repertoire of glycopeptide enrichment and separation techniques covers different solid phase extraction and chromatography-based methods such as HILIC [177, 263]. The most frequently used setup for the measurement of enriched (glyco)peptides is LC-ESI-MS/MS. Recent advances in instrumentation, in particular the development of ETD/ECD [264, 265], and high resolution orbital mass analyzers, have paved the way for the mapping of thousands of occupied *N*- and *O*-glycosylation sites as recently shown [17, 255]. Combined workflows using ETD/ECD fragmentation along with (multi-stage, MS^n) fragmentation with high- and/or low collisional induced dissociation energy

(HCD/CID) can provide compositional (structural) information on the glycan moiety as well as information on the peptide sequence and the glycosylation site [266, 267]. Recent advances in mass spectrometry driven *O*-glycoproteomics have been reviewed in detail elsewhere [42, 190]. Owing to the amount and complexity of *O*-glycoproteomic data, a number of bioinformatics tools for the prediction of mucin-type *O*-glycosylation sites [255] as well as for the database assisted interpretation and annotation of glycan and glycopeptide fragment ion spectra have been developed [268, 269]. Moreover, reporting guidelines for collecting, sharing, integrating, and interpreting mass spectrometry-based glycomics data have been specified by the MIRAGE consortium (minimum information required for a glycomics experiment) [270, 271].

The aim of this project was to develop a glycoproteomic workflow that allows the explorative non-targeted analysis of *O*-glycosylated human blood plasma proteins, which are known to carry mainly short mono- and disialylated mucin-type core-1 and -2 *O*-glycans. To achieve this, we have combined *O*-glycopeptide selective offline-HILIC fractionation of proteinase K digested peptides with nano-reversed-phase liquid chromatography coupled online to multistage ion-trap mass spectrometry (nano-RP-LC-ESI-IT-MS: CID-MS^2/-MS^3, ETD-MS^2) (Figure 20). Overall, the developed glycoproteomics approach should enable the identification of the peptide moiety as well as a characterization and localization of the *O*-glycosylation sites with the characterization of the corresponding *O*-glycans (Figure 24).

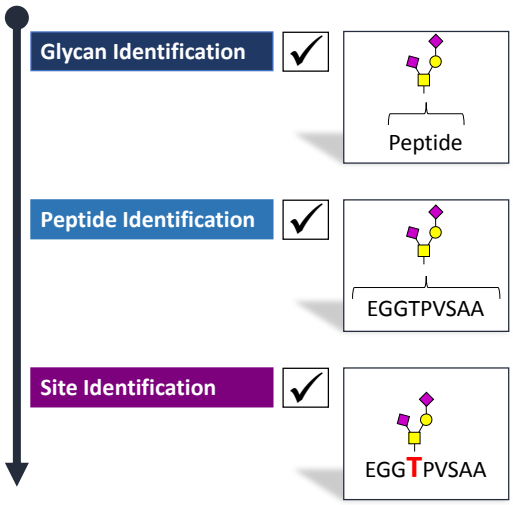


Figure 24: Comprehensive identification and characterization of mucin-type *O*-glycopeptides.

4.2. Results

4.2.1. Reproducibility of the Proteinase K Digest

Previous studies on single glycoproteins could show the successful application of proteinase K in the context of *N*- and *O*-glycoproteomics [260, 272-275]. However, its application on complex samples, like human blood plasma, has not been described so far. Here, we have employed proteinase K to generate (glyco)peptides from the entire (glyco)proteome of a pooled human blood plasma sample that was derived from 20 healthy donors. To assess the reproducibility of such a digest, five

independent proteinase K treated blood plasma samples (technical replicates) were measured with nano-RP-LC-ESI-IT-MS/MS in preliminary experiments. A comparison of the resulting base peak chromatograms revealed a high reproducibility of these digests (visual judgement), as shown in Figure 25.

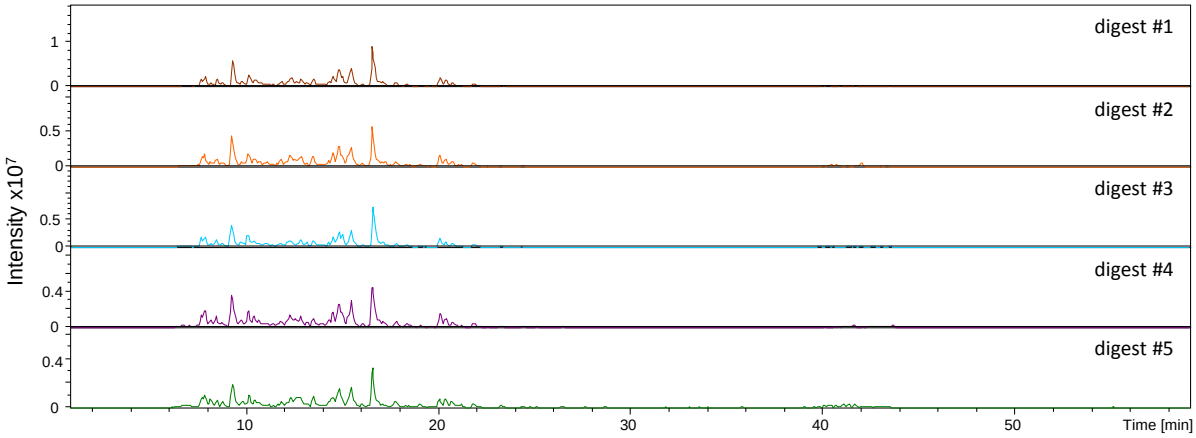


Figure 25: Base peak chromatograms (BPC) of five proteinase K digests measured with nano-RP-LC-ESI-IT-MS² show a high reproducibility of the proteolytic digest.

4.2.2. Glycopeptide Enrichment and Fractionation via HILIC-HPLC

The HILIC-HPLC fractionation carried out in the present study was optimized for the enrichment of *O*-glycosylated peptides (similar to a report from Zauner *et al.* [260]). In total 17 HILIC fractions were collected and analyzed by nano-RP-LC-ESI-IT-MS² (CID). The acquired fragment ion spectra were manually screened for the presence of *N*- and *O*-glycopeptides – relying on the detection of diagnostic oxonium ions (B-ions, e.g. $HexNAc_1Hex_1NeuAc_1$; *m/z* 657.24) and characteristic mono(oligo)-saccharide neutral loss fragment ions (Y-ions). Glycopeptides were detected in five HILIC fractions (#13-#17) (Figure 26).

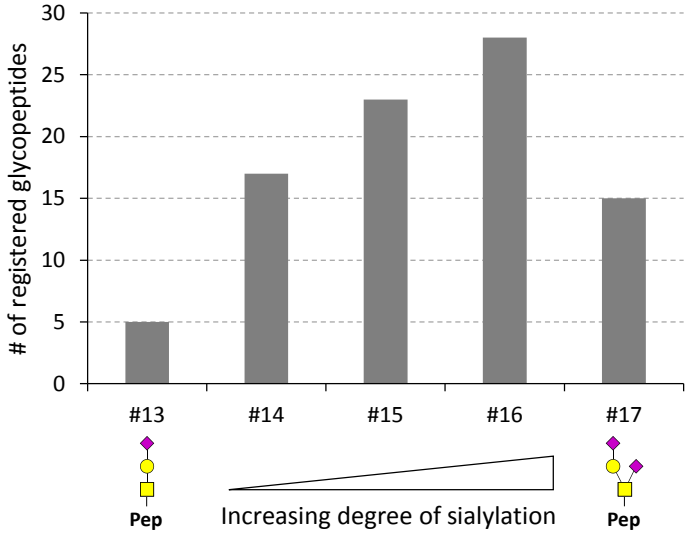


Figure 26: Number of detected glycopeptides in the HILIC fractions #13-#17.

Glycopeptides eluted in the range of 8.5-31 min from the reversed-phase column and clusters of glycopeptides were registered between 8.5-12.5 min, 14-18 min, 19-20.5 min, 22.5-23.5 min and 25-31 min (exemplarily shown for fraction #15, Figure 27).

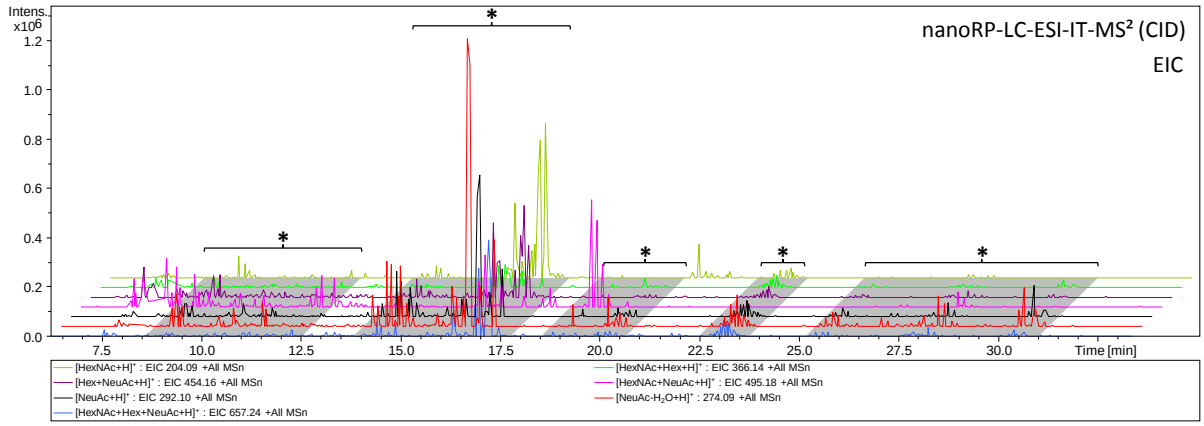


Figure 27: Extracted ion chromatograms (EICs) of diagnostic glycan oxonium ions (e. g. [HexNAc+Hex+NeuAc+H]$^+$: EIC m/z 657.24) reveal the clustered elution of *O*-glycopeptides (*) on a C18 reversed-phase column. EICs of HILIC fraction #15 are shown as an example.

4.2.3. Determination of the Glycan Composition

CID-MS² spectra were carefully inspected and manually annotated with respect to the glycan composition. Major signals in these spectra resulted from consecutive neutral losses (singly and doubly charged species) of the monosaccharides Hex, HexNAc and NeuAc from the intact glycopeptide. Most of the time, the applied collision energy induced the complete fragmentation of the glycan moiety while leaving the de-glycosylated peptide moiety intact. These fragment ions along with corresponding oxonium ions, allowed inferring the glycan composition and the putative peptide mass (Figure 28, Top).

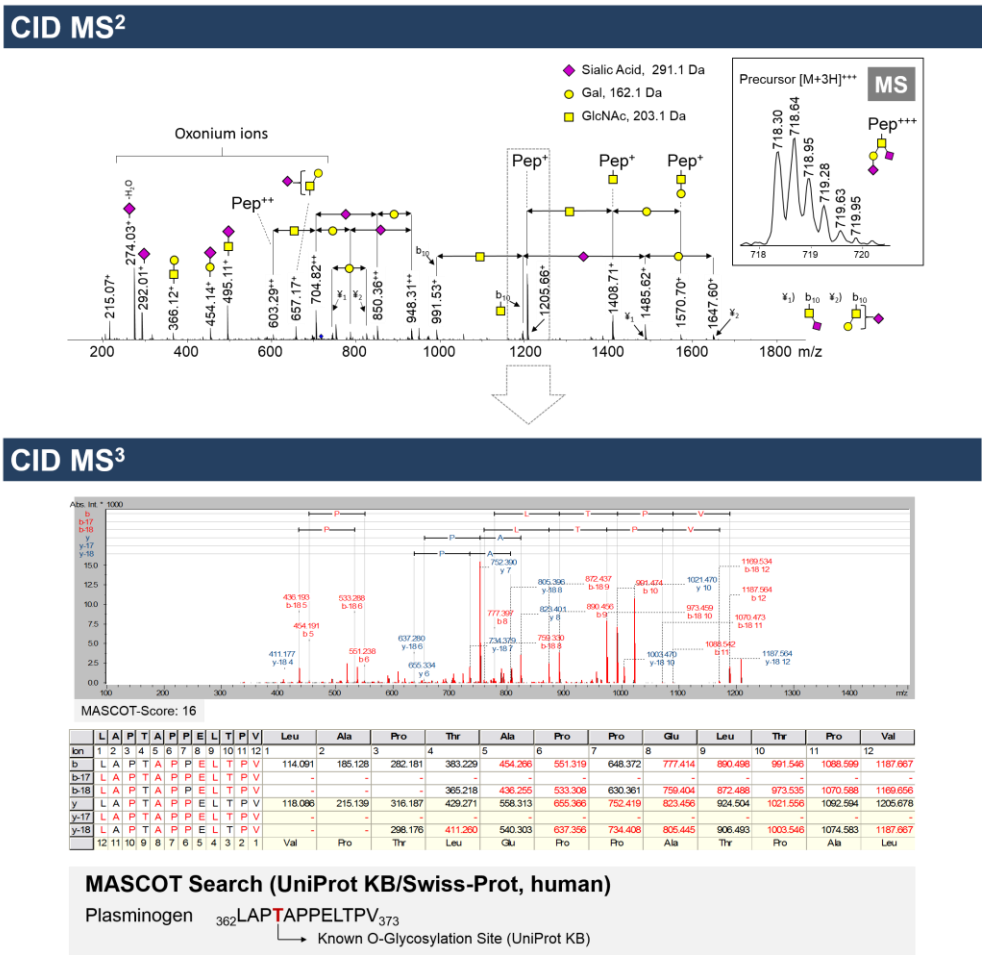


Figure 28: Exemplary CID fragment ion spectra of a proteinase K-generated plasminogen *O*-glycopeptide ($_{362}$LAP**T**APPEL**T**PV$_{373}$) measured with nanoRP-LC-ESI MSn (positive ion mode, CID). **Top:** For the given *O*-glycopeptide the CID-MS^2 spectrum is shown together with its corresponding precursor ion *m/z* 718.30 $[M+3H]^{3+}$ (inset). The spectrum allows the elucidation of the *O*-glycan composition (here disialylated T-antigen). In addition, also some internal glycopeptide fragments have been detected (e.g. b10+HexNAc). **Bottom:** The putative peptide mass (*m/z* 1205.66 $[M+H]^+$, dashed box) of the given *O*-glycopeptide was subjected to CID-MS^3 fragmentation. The peptide was identified via MASCOT protein database search (Score: 16, UniProt KB/Swiss-Prot, human).

Detailed analysis revealed that exclusively mucin-type core-1 mono- and disialylated *O*-linked glycopeptides ((di)sialyl-T-antigen) were present. For the glycan annotation, a mass error of ±0.3 Da was accepted. This parameter was justified as the observed mass errors were about 0.07 Da (median value). In total, 88 *O*-glycopeptides were detected and characterized with respect to their glycan composition. The registered glycopeptides covered an *m/z*-range of 507-945 (average *m/z* 728) and were either doubly (55 peptides) or triply charged (33 peptides).

4.2.4. Identification of the Peptide Moiety

To complement the deduced glycan composition with peptide sequence information, CID-MS^3 experiments were conducted on putative peptide masses, which were derived from CID-MS^2 spectra (Figure 28, bottom). In separate LC-MS runs the selected peptide precursor masses (predominantly singly charged) were used to trigger manual CID-MS^3 fragmentation. In rare cases peptide+HexNAc was selected for fragmentation, due to low signal intensity of the peptide

species in the MS^2 spectra. CID-MS^3 spectra were searched against the human subset of the highly-curated and non-redundant protein database UniProtKB/Swiss-Prot. Notably, also in some CID-MS^2 spectra b- and y-ions derived from peptide backbone cleavages were detected that enabled peptide identification (e.g. supplemental Figure S(I)5: alpha-2-HS-glycoprotein m/z 623.23^{3+}). For 88 detected glycopeptides, 60 corresponding peptides could be identified unambiguously (Table 1, Table 2).

Table 1: Site-specific *O*-glycan composition of identified human blood plasma glycoproteins. Glycoproteins are listed with their UniprotKB accession number as well as the number of identified glycopeptides. *O*-glycosylated sites or regions are indexed with respect to the attached *O*-glycans (mono- and/or disialylated mucin-type core 1 *O*-glycans). *O*-glycosylation sites in bold have been pinpointed within this study. Previously unknown sites und regions are indicated by underlining. Curled brackets mark regions with several possible *O*-glycosylation sites. Superscript numbers indicate literature references. For every protein the number of registered, previously known, as well as new *O*-glycosylation sites and regions are given. For underlined proteins, glycosylated as well as non-glycosylated peptides were identified (supplemental Table S1 in Hoffmann *et al.*, 2016 [225]). In addition, previously reported plasma concentrations are given. HexNAc (*N*-acetylhexosamine), Hex (hexose), NeuAc (*N*-acetylneuraminic acid, sialic acid).

	Identified Proteins	Accession	# Identified Glycopeptides	Site-Specific *O*-Glycan Composition				Registered / Previously Known / New O-Glycosylation Sites (Regions)	Reported Plasma Concentrations [ng/mL]
				I: $HexNAc_1Hex_1NeuAc_1$	II: $HexNAc_1Hex_1NeuAc_2$	I & II	2x $HexNAc_1Hex_1NeuAc_1$		
1	Alpha-2-HS-glycoprotein	P02765		**T(256)[276] S(346)[258, 277]**	-	**T(270)[276]**	-	3 / 3 / 0	$6.1*10^5$ [236] / $8.2*10^4$ [238]
2	Apolipoprotein C-III	P02656	2	**T(94)[211, 258, 277, 278]**	-	-	-	1 / 1 / 0	$1.2*10^5$ [236] / $1.7*10^5$ [238]
3	Apolipoprotein-L domain-containing protein 1	Q96LR9	1	-	{T(133), T(135), S(136)}	-	-	1 / 0 / 1	-
4	Complement C4-B	P0C0L5	1	{S(1242), T(1244), S(1251)}[279]	-	-	-	1 / 1 / 0	$9.0*10^4$ [238]
5	Fibrinogen alpha chain	P02671	1	{S(524), T(525), T(528)}[274]	-	-	-	1 / 6 / 0	$2.7*10^6$ [236] / $1.3*10^5$ [238]
6	Fibrinogen beta chain	P02675	1	**S(58)[274]**	-	-	-	1 / 1 / 0	$9.7*10^5$ [240] / $1.3*10^5$ [238]
7	GTP-binding protein 1	O00178	1	{T(119), S(122)}	-	-	-	1 / 0 / 1	-
8	Hemopexin	P02790	1	{T(24)[211, 258, 279], T(29)[258], S(30)}	-	-	-	1 / 3 / 0	$7.5*10^5$ [239] / $1.8*10^5$ [238]
9	Immunoglobulin J Chain	P01591	2	**T(97)**	-	-	-	1 / 0 / 1	$3.5*10^4$ [240] / $5.6*10^3$ [238]
10	Insulin-like growth factor-binding protein 6	P24592	1	**T(126)[211, 258]**	-	-	-	1 / 5 / 0	$1.1*10^3$ [238]
11	Inter-alpha-trypsin inhibitor heavy chain H4	Q14624	4	**S(640)[280]** {T(722)[277], T(723)[277], T(725)[277], S(733)}	-	-	-	2 / 2 / 0	$1.8*10^5$ [240] / $4.2*10^4$ [238]
12	Inter-alpha-trypsin inhibitor heavy chain H2	P19823	3	**T(691)[279, 281, 282]**	-	-	**{S(673), T(675)}[277, 281, 282]**	3 / 4 / 0	$2.1*10^5$ [240] / $2.1*10^4$ [238]
13	Kininogen-1	P01042	6	{S(150), T(151)}[211, 258] **S(577)[283]**	**S(604)** **T(571)[283]**	**T(137)[211, 258]**	-	5 / 9 / 1	$7.0*10^4$ [240] / $2.8*10^4$ [238]
14	Legumain	Q99538	1	{T(305), S(307)}	-	-	-	1 / 0 / 1	-

Table 1, continued.

	Identified Proteins	Accession	# Identified Glycopeptides	Site-Specific *O*-Glycan Composition				Registered / Previously Known / New O-Glycosylation Sites (Regions)	Reported Plasma Concentrations [ng/mL]
				I: $HexNAc_1Hex_1NeuAc_1$	II: $HexNAc_1Hex_1NeuAc_2$	I & II	2x $HexNAc_1Hex_1NeuAc_1$		
15	Plasminogen	P00747	4	-	-	**T(365)**[284]	-	1 / 2 / 0	$1.4*10^5$ [239] / $2.5*10^4$ [238]
16	Protein AMBP	P02760	6	**T(24)**[285]	-	-	-	1 / 2 / 0	$6.0*10^4$ [240] / $4.8*10^4$ [238]
17	Protocadherin beta-11	Q9Y5F2	1	**S(24)**	-	-	-	1 / 0 / 1	-
18	Protocadherin-16	Q96JQ0	1	**T(592)**	-	-	-	1 / 0 / 1	-
19	Selenoprotein P	P49908	1	**T(236)**	-	-	-	1 / 0 / 1	$5.1*10^2$ [238]
20	Tau-tubulin kinase 2	Q6IQ55	1	**T(820)**	-	-	-	1 / 0 / 1	-
21	Tenascin-R	Q92752	1	-	**T(1288)**	-	-	1 / 2 / 1	-
22	Tenascin-X	P22105	1	-	**T(3586)**	-	-	1 / 0 / 1	$7.0*10^1$ [238]
23	not unambiguously identified	-		18x	10x	-	-	-	

Table 2: Detailed overview of all identified human blood plasma *O*-glycopeptides. For each *O*-glycopeptide, the corresponding glycoprotein including the UniProtKB accession number, the identified *O*-glycosylation site(s)/regions as well as the *O*-glycan composition are given, respectively. Likewise, the LC retention time, the mass of the intact glycopeptide precursor, the measured peptide mass as well as the corresponding error is listed. The peptide identification using CID-MS^3 was validated by the MASCOT peptide score and the Biotools score (CID). ETD-based determination of the *O*-glycosylation site(s) was validated by the Biotools score (CID) as well as the NetOGlyc 4.0 score.

	Identified Proteins	Accession	O-Glycosylation Site (Region)	# Identified Peptides	O-Glycan Composition	Retention Time [min]	Precursor Ion	Measured peptide mass $[M+H]^+$ [Da]	Error [Da]	Peptide Score (CID)	Biotools Score (CID)	Biotools Score (ETD)	NetOGlyc 4.0 Score
1	**Alpha-2-HS-glycoprotein**	P02765		19									
			T(256)	$_{252}$TQPVTSQPQPE$_{262}$	HexNAc(1)Hex(1)NeuAc(1)	14.1-14.3	$[934.36]^{2+}$	1211.55	-0.04	28	402	T(252): 12 / T(256): 19 / S(257): 18	0,81 / 0,90 / 0,87
						13.9-14.0	$[623.23]^{3+}$	1211.49	-0.11	42	46	T(252): 669 / T(256): 412 / S(257): 362	0,81 / 0,90 / 0,87
			T(270)	$_{267}$AVPTPV$_{272}$	HexNAc(1)Hex(1)NeuAc(1)	20.25-20.37	$[620.26]^{2+}$	583.33	-0.09	20	1	44	0.87
					HexNAc(1)Hex(1)NeuAc(2)	19.8-20.2	$[765.77]^{2+}$	583.27	-0.07	18	1	6	0.87
				$_{267}$AVPTPVVDPD$_{276}$	HexNAc(1)Hex(1)NeuAc(1)	22.1-22.2	$[833.3]^{2+}$	1009.43	-0.09	19	8	141	0.87
				$_{267}$AVPTPVVDPDAPPSPPL$_{283}$	HexNAc(1)Hex(1)NeuAc(1)	27.5-28.5	$[775.67]^{3+}$	1668.84	-0.04	54	20	T(270): 153 / S(280): 75	0,87 / 0,96
					HexNAc(1)Hex(1)NeuAc(2)	31.9-32.0	$[872.72]^{3+}$	1668.75	-0.13	59	10	T(270): 39 / S(280): 34	0,87 / 0,96
					HexNAc(1)Hex(1)NeuAc(2)	31.3-31.7	$[872.73]^{3+}$	1668.77	-0.12	65	15	T(270): 39 / S(280): 21	0,87 / 0,96
				$_{266}$EAVPTPV$_{272}$	HexNAc(1)Hex(1)NeuAc(1)	20.4-20.7	$[684.76]^{2+}$	712.35	-0.04	18	2	7	0.87
					HexNAc(1)Hex(1)NeuAc(2)	21.9-22.2	$[830.34]^{2+}$	712.32	-0.07	16	38	-	0.87
				$_{266}$EAVPTPVVDPDAPPSPPL$_{283}$	HexNAc(1)Hex(1)NeuAc(1)	31.5-31.7	$[818.68]^{3+}$	1797.75	-0.18	11	12	-	0,87 / 0,96
					HexNAc(1)Hex(1)NeuAc(2)	32.0	$[915.71]^{3+}$	1797.73	-0.10	46	23	-	0,87 / 0,96
				$_{265}$NEAVPTPV$_{272}$	HexNAc(1)Hex(1)NeuAc(1)	20.1-20.4	$[741.8]^{2+}$	826.38	-0.05	43	158	44	0.87
			S(346)	$_{341}$TVVQPSVG$_{348}$	HexNAc(1)Hex(1)NeuAc(1)	17.8-17.9	$[721.78]^{2+}$	786.39	-0.07	24	76	T(341): 533 / S(346): 187	0,93 / 0,94
					HexNAc(1)Hex(1)NeuAc(1)	16.7-16.8	$[721.76]^{2+}$	786.43	-0.01	25	149	-	0,93 / 0,94
				$_{341}$TVVQPSVGA$_{349}$	HexNAc(1)Hex(1)NeuAc(1)	17.3-17.4	$[757.3]^{2+}$	857.43	-0.05	20	36	-	0,93 / 0,94
				$_{342}$VQPSVGA$_{349}$	HexNAc(1)Hex(1)NeuAc(1)	14.0-14.1	$[657.2]^{2+}$	657.29	-0.11	20	15	-	0.94
				$_{342}$VVQPSVG$_{348}$	HexNAc(1)Hex(1)NeuAc(1)	15.3-15.7	$[671.26]^{2+}$	685.35	-0.03	44	13	-	0.94
				$_{342}$VVQPSVGA$_{349}$	HexNAc(1)Hex(1)NeuAc(1)	16.3-16.4	$[706.77]^{2+}$	756.37	-0.04	19	18	-	0.94
2	**Apolipoprotein C-III**	P02656		2									
			T(94)	$_{85}$WDLDPEVRPTSA$_{96}$	HexNAc(1)Hex(1)NeuAc(1)	25.2-25.4	$[681.25]^{3+}$	1385.57	-0.17	35	21	T(94): 46 / S(95): 40	0,51 / 0,51
					HexNAc(1)Hex(1)NeuAc(1)	24.6-25.1	$[681.35]^{3+}$	1385.71	0.04	46	15	T(94): 124 / S(95): 148	0,51 / 0,51
3	**Apolipoprotein-L domain-containing protein 1**	Q96LR9		1									
			T(133), T(135), S(136)	$_{131}$AVTITSDL$_{138}$	HexNAc(1)Hex(1)NeuAc(2)	24.3-24.5	$[883.82]^{2+}$	819.35	-0.09	31	10	T(133): 4 / T(135): 4 / S(136): 3	0,05 / 0,11 / 0,10
4	**Complement C4-B**	P0C0L5		1									
			S(1242), T(1244), S(1251)	$_{1241}$VSPTPAPRNPSDPMPQ$_{1256}$	HexNAc(1)Hex(1)NeuAc(1)	17.8-18.1	$[782.97]^{3+}$	1690.59	-0.23	34	30	S(1242): 36 / T(1244): 49 / S(1251): 145	0,96 / 0,94 / 0,51
5	**Fibrinogen alpha chain**	P02671		1									
			S(524), T(525), T(528)	$_{524}$STGKTFPG$_{531}$	HexNAc(1)Hex(1)NeuAc(1)	12.3-12.5	$[725.78]^{2+}$	794.34	-0.06	28	61	-	0,76 / 0,75 / 0,58
6	**Fibrinogen beta chain**	P02675		1									
			S(58)	$_{54}$EEAPSLRPAPPPIS$_{67}$	HexNAc(1)Hex(1)NeuAc(1)	23.3-23.5	$[706.27]^{3+}$	1460.67	-0.10	22	14	S(58): 34 / S(67): 13	0,95 / 0,86
7	**GTP-binding protein 1**	O00178		1									
			T(119), S(122)	$_{117}$YATVKSM$_{123}$	HexNAc(1)Hex(1)NeuAc(1)	15.4	$[728.75]^{2+}$	800.36	0.96	27	9	-	0,24 / 0,17
8	**Hemopexin**	P02790		1									
			T(24), T(29), S(30)	$_{24}$TPLPPTS$_{30}$	HexNAc(1)Hex(1)NeuAc(1)	14.4-14.6	$[684.73]^{2+}$	712.35	-0.09	19	8	-	0,51 / 0,83 / 0,96
9	**Immunoglobulin J Chain**	P01591		2									
			T(97)	$_{95}$DPTEV$_{99}$	HexNAc(1)Hex(1)NeuAc(1)	17.7-17.8	$[608.71]^{2+}$	560.28	0.02	27	15	-	0.11
					HexNAc(1)Hex(1)NeuAc(1)	18.6-18.8	$[608.72]^{2+}$	560.29	0.03	-	1	-	0.11
10	**Insulin-like growth factor-binding protein 6**	P24592		1									
			T(126)	$_{121}$KPQAGTA$_{127}$	HexNAc(1)Hex(1)NeuAc(1)	16.3-16.6	$[664.69]^{2+}$	672.30	-0.07	10	18	22	0.88

Table 2, continued.

	Identified Proteins	Accession	O-Glycosylation Site (Region)	# Identified Peptides	O-Glycan Composition	Retention Time [min]	Precursor Ion	Measured peptide mass $[M+H]^+$ [Da]	Error [Da]	Peptide Score (CID)	Biotools Score (CID)	Biotools Score (ETD)	NetOGlyc 4.0 Score
11	**Inter-alpha-trypsin inhibitor heavy chain H4**	Q14624		4									
			S(640)	$_{639}$ASFSPR$_{644}$	HexNAc(1)Hex(1)NeuAc(1)	11.0-11.3	$[660.72]^{2+}$	664.30	-0.07	16	51	S(640): 11 / S(642): 7	0,66 / 0,80
			T(722), T(723), T(725)	$_{722}$TTQTPAPI$_{729}$	HexNAc(1)Hex(1)NeuAc(1)	18.5-18.9	$[742.76]^{2+}$	828.42	-0.03	27	33	-	0,61 / 0,64 / 0,90
				$_{722}$TTQTPAPIQ$_{730}$	HexNAc(1)Hex(1)NeuAc(1)	15.3-15.8	$[806.77]^{2+}$	956.51	0.00	26	4	-	0,61 / 0,64 / 0,90
			T(722), T(723), T(725), S(733)	$_{722}$TTQTPAPIQAPS$_{733}$	HexNAc(1)Hex(1)NeuAc(1)	16.0-16.2	$[623.27]^{3+}$	1211.58	0.07	31	23	T(722): 152 / T(723): 34 / T(725): 19 / T(733): 5	0,61 / 0,64 / 0,90 / 0,58
12	**Inter-alpha-trypsin inhibitor heavy chain H2**	P19823		3									
			S(673), T(675)	$_{669}$WANPSPTPV$_{677}$	2x HexNAc(1)Hex(1)NeuAc(1)	21.0-21.1	$[760.92]^{3+}$	968.40	-0.08	30	1	-	0,79 / 0,88
			T(691)	$_{689}$ESTPPPHV$_{696}$	HexNAc(1)Hex(1)NeuAc(1)	12.5-12.7	$[760.24]^{2+}$	863.37	-0.06	32	65	S(690): 14 / T(691): 22	0,88 / 0,93
					HexNAc(1)Hex(1)NeuAc(1)	12.4-12.9	$[507.15]^{3+}$	863.34	-0.11	-	8	S(690): 65 / T(691): 84	0,88 / 0,93
13	**Kininogen-1**	P01042		6									
			T(137)	$_{132}$EGPVVTA$_{138}$	HexNAc(1)Hex(1)NeuAc(1)	16.6-16.7	$[664.7]^{2+}$	672.33	-0.03	-	43	-	0.51
				$_{132}$EGPVVTAQ$_{139}$	HexNAc(1)Hex(1)NeuAc(2)	17.4	$[874.28]^{2+}$	800.25	0.17	19	19	-	0.51
			S(150), T(151)	$_{146}$VHPISTQ$_{152}$	HexNAc(1)Hex(1)NeuAc(1)	11.0-11.2	$[719.23]^{2+}$	781.37	-0.02	24	29	-	0,30 / 0,34
			T(571)	$_{567}$DLIATM$_{572}$	HexNAc(1)Hex(1)NeuAc(2)	26.1-26.5	$[805.73]^{2+}$	663.34	-0.09	28	16	-	0.69
			S(577)	$_{573}$MPPISPAPIQ$_{583}$	HexNAc(1)Hex(1)NeuAc(1)	25.5-25.9	$[853.81]^{2+}$	1050.54	-0.10	40	11	22	0.87
			S(604)	$_{600}$FNPISDFPDTT$_{610}$	HexNAc(1)Hex(1)NeuAc(2)	31.8-32.2	$[734.26]^{3+}$	1253.49	-0.08	24	21	S(604): 44 / T(609): 22 / T(610): 14	0,64 / 0,69 / 0,51
14	**Legumain**	Q99538		1									
			T(305), S(307)	$_{301}$HLDLTPSPD$_{309}$	HexNAc(1)Hex(1)NeuAc(1)	14.0-14.7	$[550.86]^{3+}$	994.38	-0.10	17	9	T(305): 31 / S(307): 51	0,71 / 0,82
15	**Plasminogen**	P00747		4									
			T(365)	$_{362}$LAPTAPPELTPV$_{373}$	HexNAc(1)Hex(1)NeuAc(1)	27.9-28.8	$[931.4]^{2+}$	1205.64	-0.07	16	51	T(365): 115 / T(371): 173	0,98 / 0,45
					HexNAc(1)Hex(1)NeuAc(2)	28.2-28.5	$[718.3]^{3+}$	1205.63	-0.01	16	47	T(365): 150 / T(371): 36	0,98 / 0,45
				$_{363}$APTAPPELTPV$_{373}$	HexNAc(1)Hex(1)NeuAc(1)	25.6-26.1	$[583.58]^{3+}$	1092.56	-0.03	19	24	T(365): 65 / T(371): 199	0,98 / 0,45
					HexNAc(1)Hex(1)NeuAc(1)	25.8-26.2	$[874.88]^{2+}$	1092.60	0.01	24	326	T(365): 177 / T(371): 34	0,98 / 0,45
16	**Protein AMBP**	P02760		6									
			T(24)	$_{20}$GPVPTPPDNI$_{29}$	HexNAc(1)Hex(1)NeuAc(1)	25.4-25.7	$[831.84]^{2+}$	1006.51	-0.01	22	36	33	0.55
				$_{20}$GPVPTPPDNIQ$_{30}$	HexNAc(1)Hex(1)NeuAc(1)	22.8-23.7	$[895.87]^{2+}$	1134.58	0.00	8	196	34	0.55
					HexNAc(1)Hex(1)NeuAc(1)	24.1	$[895.85]^{2+}$	1134.48	-0.10	10	7	80	0.55
					HexNAc(1)Hex(1)NeuAc(1)	23.0-23.5	$[597.58]^{3+}$	1134.55	-0.03	11	45	117	0.55
				$_{20}$GPVPTPPDNIQV$_{31}$	HexNAc(1)Hex(1)NeuAc(1)	27.6-28.1	$[945.37]^{2+}$	1233.63	-0.02	12	32	-	0.55
					HexNAc(1)Hex(1)NeuAc(1)	27.5-28.5	$[630.57]^{3+}$	1233.58	-0.07	18	9	-	0.55
17	**Protocadherin beta-11**	Q9Y5F2		1									
			S(24)	$_{20}$LLGMSQ$_{25}$	HexNAc(1)Hex(1)NeuAc(1)	10.0-10.4	$[652.74]^{2+}$	648.35	-0.06	28	14	-	0.04
18	**Protocadherin-16**	Q96JQ0		1									
			T(592)	$_{589}$EGTQPG$_{594}$	HexNAc(1)Hex(1)NeuAc(1)	11.6-11.8	$[622.72]^{2+}$	588.26	0.00	29	44	-	0.05
19	**Selenoprotein P**	P49908		1									
			T(236)	$_{234}$APTHPAPPGLH$_{244}$	HexNAc(1)Hex(1)NeuAc(1)	13.8-14.2	$[584.24]^{3+}$	1094.51	-0.07	43	517	57	0.99
20	**Tau-tubulin kinase 2**	Q6IQ55		1									
			S(817), T(819), T(820)	$_{814}$KDHSATTEPL$_{823}$	HexNAc(1)Hex(1)NeuAc(1)	19.9-20.0	$[877.82]^{2+}$	1098.59	0.04	22	313	S(817): 72 / T(819): 285 / T(820): 399	0,77 / 0,94 / 0,90
21	**Tenascin-R**	Q92752		1									
			T(1288)	$_{1282}$DNDVAVTNC$_{1290}$	HexNAc(1)Hex(1)NeuAc(2)	23.0-23.2	$[652.93]^{3+}$	1009.47	0.10	18	27	15	0.42
22	**Tenascin-X**	P22105		1									
			T(3586)	$_{3586}$TPAPELAPEAP$_{3596}$	HexNAc(1)Hex(1)NeuAc(2)	26.8-27.4	$[680.61]^{3+}$	1092.53	-0.03	15	49	26	0.62
23	**not unambiguously identified**	-	-	28	HexNAc(1)Hex(1)NeuAc(1) (18x)	-	-	-	-	-	-	-	-
					HexNAc(1)Hex(1)NeuAc(2) (10x)	-	-	-	-	-	-	-	-

These 60 peptides could be linked to 22 different proteins, most of them being acute phase proteins. As the protein identification is based on a single peptide, validation of the potential peptide hits is of utmost importance. In particular, the protein inference problem [286], which is intrinsic to bottom-up proteomic approaches, had to be considered. To cope with this, peptide spectra were manually revised and only peptide hits with a MASCOT ion score of greater than 20 were considered; only in rare cases, and supported by other evidences, also lower scored peptides were accepted. Furthermore, peptide hits needed to exhibit at least one potential *O*-glycosylation site (Ser/Thr). If available, knowledge derived from public databases (UniProtKB and UniCarbKB) on already described *O*-glycosylation sites within the putative peptides or within the entire protein was used to support a potential hit.

The peptide identification was further corroborated by redundant identifications, that is the multiple occurrence of: (I) the same *O*-glycopeptide in different HILIC fractions, or (II) the same peptide but with a different glycan moiety, or (III) the identification of a peptide harboring the same glycosylation site, but differing in peptide length; the latter being attributed to the broad-specific proteolysis (e.g. alpha-2-HS-glycoprotein, $_{341}$TVVQP**S**[$HexNAc_1Hex_1NeuAc_1$]VG_{348} derived from HILIC fraction #13 and $_{342}$VVQP**S**[$HexNAc_1Hex_1NeuAc_1$]VG_{348} from fraction #14). In some cases, peptide identification was hampered or inconclusive. One of the main obstacles here was the frequent occurrence of prolines within the (glyco)peptide sequence, which is also described in literature. The cyclic structure of proline, gives rise to a high signal of the preceding y-ion but precludes in most cases the generation of a subsequent b-ion – thus introducing a sequence gap (referred to as the "proline effect") [287]. This in turn leads to incomplete peptide fragment ion series and the occurrence of dipeptide fragment ions (e.g. PS and SP), which may result in ambiguity in peptide identification. This effect is particularly critical for short peptide sequences, as usually obtained by a broad- or non-specific digest. The average peptide length of glycopeptides identified in this study was 10 amino acids (aa). This is significantly shorter than the average length of tryptic peptides (14 aa, based on an *in silico* digestion of the human UniProtKB database[206], Figure 29). All this – in conjunction with a non-specific peptide search – makes a reliable peptide identification challenging.

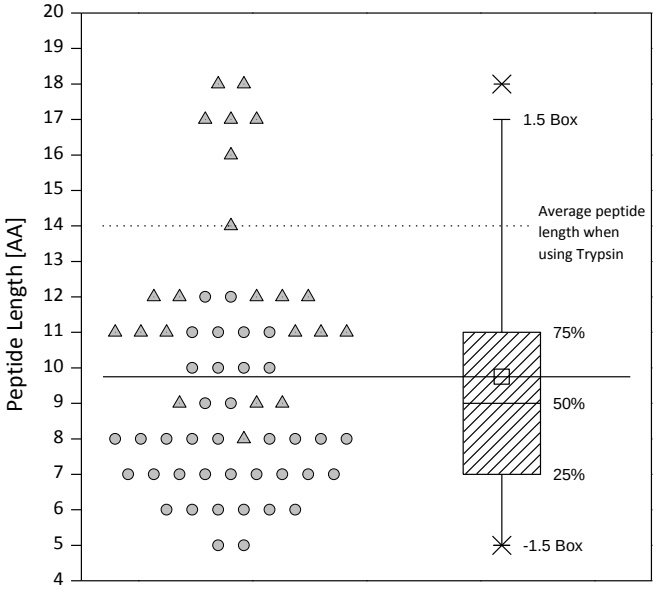


Figure 29: Box-Whisker-plot of the peptide length obtained with proteinase K (in total 60 peptides). The horizontal line within the box indicates the median, boundaries of the box mark the upper and lower quartile (Q1, Q2, Box=interquartile range (IQR)). The whiskers indicate Q1-(1.5xIQR) and Q3+(1.5xIQR). The mean value is illustrated by a square within the box. Outliers are marked with an "X". Data points plotted next to the Box-and-whisker plot are indexed with respect to the peptide charge state (circle=doubly charged, triangle=triply charged). Dotted line: average peptide length for a tryptic digest (based on an *in silico* digestion of the human Uniprot database, [206]).

To complement the identified *O*-glycopeptides with non-glycosylated peptides that are also present in blood plasma, CID und ETD fragment ion spectra of the corresponding HILIC fractions (#1-17) were searched against the human subset of the UniProtKB/Swiss-Prot protein database. In total 111 proteins were identified. CID and ETD spectra provided complementary results; 54 and 45 proteins were identified, respectively, and only 12 proteins were identified with both modes (Figure 30).

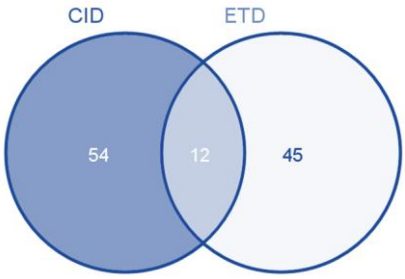


Figure 30: Number of human blood plasma proteins identified with CID and ETD.

Compared to ETD, significantly more peptides were identified with CID (321 vs 150), though. The majority of peptides were derived from immunoglobulins, serotransferrin, haptoglobin and serumalbumin (supplemental Table S1 in Hoffmann *et al.*, 2016 [225]). Notably, also non-glycosylated peptides corresponding to previously identified *O*-glycopeptides, e.g. of plasminogen and hemopexin, were identified (Table 1).

4.2.5. Localization of the *O*-Glycosylation Sites

To further characterize the identified *O*-glycopeptides, the corresponding *O*-glycosylation sites needed to be localized. In a few cases the use of proteinase K, already generated glycopeptides that exhibit only one possible *O*-glycosylation site, e.g. $_{132}$EGPVV**T**[$HexNAc_1Hex_1NeuAc_1$]A_{138} and $_{567}$DLIA**T**[$HexNAc_1Hex_1NeuAc_2$]M_{572} from kininogen-1 or $_{234}$AP**T**[$HexNAc_1Hex_1NeuAc_1$]HPAPPGLH_{244} from selenoprotein P. Noteworthy, in the first example a tryptic digest would have generated a peptide with a length of 43 aa ($_{119}$F**S**VA**T**Q**T**CQI**T**PA-EGPVV**T**AQYDCLGCVHPI**ST**Q**S**PDLEPIL**R**$_{161}$), harboring 8 potential *O*-glycosylation sites. This clearly illustrates a benefit of the proteinase K digest for the *O*-glycan site identification (Figure 31).

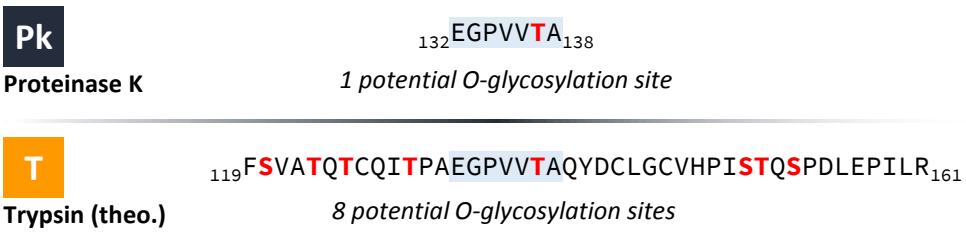
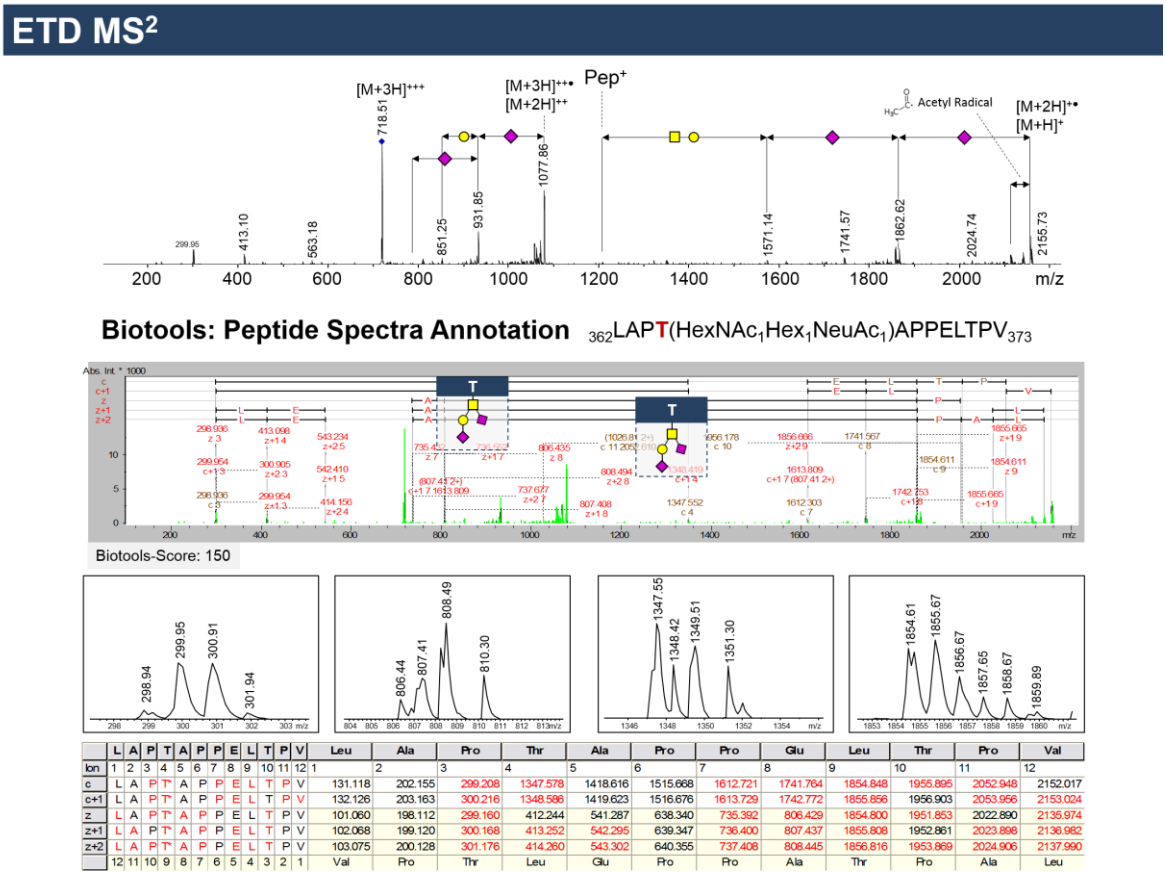


Figure 31: The use of proteinase K can facilitate identification of *O*-glycosylation sites.

When the *O*-glycosylation sites could not be inferred directly, glycopeptides were subjected to ETD fragmentation in a separate LC-MS run (Figure 32). The most prominent peaks in the acquired ETD glycopeptide spectra were the unfragmented precursor ions along with charge-reduced species; minor peaks were derived from c- and z-type peptide backbone cleavages. Furthermore, fragment ions indicating either the loss of 43.018 Da ($C_2H_3O^{\bullet}$) from the radical cationic species or 42.016 Da (C_2H_2O) from the even electron species $[M+H]^+$ were consistently detected. In the literature, this spectral feature was attributed to the loss of an acetyl-radical from the *N*-acetyl group of a HexNAc [288, 289]. This in turn can support the discrimination of ETD spectra derived from glycosylated and non-glycosylated species. Strikingly, and in contrast to the general mode of action of ETD, also fragmentations of the glycan moiety along the intact peptide backbone were observed, leading to a complete loss of the *O*-glycosylated Ser/Thr side-chain. Nevertheless, the resulting fragment ions enabled a verification of the glycan composition as well as the peptide mass. At first, ETD-generated glycopeptide spectra were searched against the human subset of the UniProtKB/Swiss-Prot database using MASCOT, under consideration of the *O*-glycan modification (theoretical glycan mass used as variable modification of Ser/Thr). However, this strategy failed due to the presence of intense signals in the ETD spectrum, which correspond to: (I) the precursor ion, (II) the charge reduced precursor ion, (III) acetyl radicals ions, (IV) or glycan fragment ions. These ions might be erroneously interpreted as peptide-derived fragment ions by the search engine, since ETD is supposed to solely produce peptide fragment ions while keeping fragile side-chain modifications, like the glycosylation, intact. To overcome this, glycopeptide spectra were exported to Bruker BioTools for manual spectra annotation. Here, the identified glycopeptides were built *in silico*, taking into account the corresponding *O*-glycan moieties as

well as all possible *O*-glycosylation sites. Subsequently, the resulting *in silico* fragment ions (c- and z-type ions) were matched to their counterparts in the measured ETD-MS² spectra. To evaluate the spectra annotation and to discern the correct *O*-glycosylation site, the BioTools spectra matching score along with manual inspection of the respective spectra were considered.



	L	A	P	T	A	P	P	E	L	T	P	V	Leu	Ala	Pro	Thr	Ala	Pro	Pro	Glu	Leu	Thr	Pro	Val
Ion	1	2	3	4	5	6	7	8	9	10	11	12	1		2		3		4		5		6		7		8		9		10		11		12	
c	L	A	P	T*	A	P	P	E	L	T	P	V	131.118	202.155	299.208	1347.578	1418.616	1515.668	1612.721	1741.764	1854.848	1955.895	2052.948	2152.017
c+1	L	A	P	T*	A	P	P	E	L	T	P	V	132.126	203.163	300.216	1348.586	1419.623	1516.676	1613.729	1742.772	1855.856	1956.903	2053.956	2153.024
z	L	A	P	T*	A	P	P	E	L	T	P	V	101.060	198.112	299.160	412.244	541.287	638.340	735.392	806.429	1854.800	1951.853	2022.890	2135.974
z+1	L	A	P	T*	A	P	P	E	L	T	P	V	102.068	199.120	300.168	413.252	542.295	639.347	736.400	807.437	1855.808	1952.861	2023.898	2136.982
z+2	L	A	P	T*	A	P	P	E	L	T	P	V	103.075	200.128	301.176	414.260	543.302	640.355	737.408	808.445	1856.816	1953.869	2024.906	2137.990
	12	11	10	9	8	7	6	5	4	3	2	1	Val		Pro		Thr		Leu		Glu		Pro		Pro		Ala		Thr		Pro		Ala		Leu	

Figure 32: Exemplary ETD fragment ion spectrum of a proteinase K-generated plasminogen *O*-glycopeptide ($_{362}$LAP**T**APPEL**T**PV$_{373}$) measured with nanoRP-LC-ESI MSn (positive ion mode, ETD). The *O*-glycosylation site (here Thr$_{365}$) was pinpointed by means of ETD (Biotools-Score: 150). Magnified regions show the isotope pattern of selected peptide fragment ions, confirming the annotation. In addition to peptide fragment ions also fragment ions derived from the glycan moiety were detected, allowing a verification of the glycan composition. Furthermore, a neutral loss of an acetyl radical from the intact *O*-glycopeptide was observed, which is typically seen in ETD spectra of glycopeptides.

Furthermore, public repositories, namely UniProtKB and UniCarbKB, were queried with respect to known *O*-glycosylation sites within the peptide in question. To further asses the validity of the *O*-glycosylation site annotation, the site occupancy was predicted using NetOGlyc – an online tool, based on machine-learning algorithms, which allows the prediction of mucin-type *O*-glycosylation sites [255]. For 36 of 60 identified glycopeptides, the quality of the corresponding ETD spectra was acceptable - in terms of signal intensity and the number of fragment ions. Overall, 31 *O*-glycosylation sites and regions were detected, of which 23 sites could be pinpointed (Table 1, Table 2). Strikingly, 11 previously unknown *O*-glycosylation sites and regions were registered, of which 8 sites could be pinpointed. Generally, *O*-glycosylation on threonine residues was observed more frequently than on serine (16x Thr, 7

x Ser). In accordance with literature, prolines were frequently found in close vicinity to the *O*-glycosylation site (positions n-1, n+1, n+3), e.g. $_{267}$AVP**T**[$HexNAc_1Hex_1NeuAc_1$]PV_{272}, $_{343}$VQP**S**[$HexNAc_1Hex_1NeuAc_1$]VGA_{349} from alpha-2-HS-glycoprotein [258]. In addition also prolines in position n+2 were found occasionally, e.g. $_{20}$GPVP**T**[$HexNAc_1Hex_1NeuAc_1$]$PPDNI_{29}$ from alpha-1 microglycoprotein (protein AMBP).

4.2.6. Identified Glycoproteins: Selected Examples

In the following, selected examples of identified *O*-glycopeptides are detailed that feature novel *O*-glycosylation sites or exhibit remarkable fragmentation characteristics.

Alpha-2-HS-glycoprotein

In this study, the majority of identified *O*-glycopeptides were derived from alpha-2-HS-glycoprotein, also known as fetuin-A. Fetuin-A is a negative acute phase glycoprotein that is highly abundant in fetal blood plasma. It is involved in transport of substances, such as calcium and features three *O*-glycosylation sites (T_{256}, T_{270}, S_{346}), which are decorated with sialylated mucin-type core-1 *O*-glycan structures [258, 276, 277]. In contrast to previous reports [258, 277], intact *O*-glycopeptides identified and characterized in the present study describe all three known fetuin-A *O*-glycosylation sites including the attached *O*-glycans (mono- and disialylated mucin-type core-1 *O*-glycans), respectively. By pinpointing *O*-glycosylation sites using ETD, the reported ETD Biotools scores can be misleading. This, for instance, holds true for the fetuin-A *O*-glycopeptide $_{252}$**T**QPV**TS**$QPQPE_{262}$ (m/z 623.23^{3+}) and its three potential *O*-glycosylation sites: T_{252} (669), T_{256} (412), T_{257} (362) (supplemental Figure S(I)5). According to the score values, T_{252} would be the occupied site; the presence of characteristic ETD fragment ions at m/z 344.01^{1+} (c_3), 1200.45^{1+} (c_5), 1287.47^{1+} (c_6), 1525.57^{1+} ($z+1_8$), 1751.51^{1+} ($z+2_{10}$), though, clearly indicates the occupancy of T_{256}, which is in agreement with literature findings. For the two other described fetuin-A *O*-glycosylation sites T_{270} and S_{346} ETD fragmentation was not mandatory, since corresponding *O*-glycopeptides were identified that solely harbor one *O*-glycosylation site (e.g. T_{270}: $_{267}$AVP**T**PV_{272}, S_{346}: $_{342}$VVQP**S**VG_{348}), respectively. Also of note, with respect to the peptide identification, b- and y-ions were detected in the CID-MS^2 fragment ion spectra of the fetuin-A *O*-glycopeptides $_{252}$TQPV**T**$SQPQPE_{262}$ (m/z 623.23^{3+}), $_{267}$AVP**T**$PVVDPDAPPSPPL_{283}$ (m/z 872.72^{3+}) and $_{266}$EAVP**T**$PVVDPDAPPSPPL_{283}$ (m/z 915.71^{3+}), which already permit the unambiguous peptide identification without consideration of CID-MS^3 spectra (supplemental Figures S(I)3-5). Furthermore, internal glycopeptide fragment ions resulting from concerted fragmentations along the peptide backbone and along the glycan moiety were detected in the same CID-MS^2 spectra – a low-energy CID glycopeptide fragmentation event that is rarely described in literature (e.g. $_{252}$TQPV**T**($HexNAc_1Hex_1NeuAc_1$)$SQPQPE_{262}$ → $_{252}$TQPV**T**(HexNAc)SQ_{258} m/z 945.45^{1+}) (supplemental Figure S(I)5).

Kininogen-1

The human KNG1 gene codes for two splicing variants of kininogen, namely low-molecular and high-molecular weight kininogen. The latter is involved in blood coagulation and the

assembly of the kallikrein-kinin system and was identified in the present study by six *O*-glycopeptides. Currently nine *O*-glycosylation sites/regions are described in literature for kininiogen-1 – presumably all being decorated with mucin-type core-1 or possibly core-8 *O*-glycans [211, 258, 279, 283]. Experimental glycoproteomic evidence on the macro and microheterogeneity of kininogen-1 is still missing, though. In the present study, four kininogen-1 *O*-glycosylation sites, including one novel site (S_{604}), could be pinpointed and described with respect to the composition of the attached *O*-glycans (Table 1). The identified *O*-glycopeptide $_{600}$FNPI**S**DFPD**TT**$_{610}$ (m/z 734.26^{3+}) carries a disialylated T-antigen and harbors three potential *O*-glycosylation sites. ETD analysis implies the occupancy of S_{604}, due to the presence of a signal at m/z 1638.30^{1+}, corresponding to a c_6 ion (supplemental Figure S(I)4). Also of note, in previous studies the use of trypsin did not allow to pinpoint occupied *O*-glycosylation sites in the region aa119-161 [211, 258]. Proteinase K, however, generated two distinct *O*-glycopeptides ($_{132}$EGPVV**T**A$_{138}$ m/z 664.70^{2+}, $_{146}$VHPI**ST**Q$_{152}$ m/z 719.23^{2+}) that allowed pinpointing the site T_{137} and the region S_{150}/T_{151}. For the latter, unfortunately, the ETD spectrum quality did not allow localizing the exact site. The peptide $_{132}$EGPVV**T**A$_{138}$ (m/z 664.70^{2+}) could not be identified correctly by MASCOT database search, due to missing fragment ions. However, the peptide could be identified via manual *de novo* annotation supported by mass tag-assisted *de novo* sequencing ([283.0 Da]VVTA) using the tool MS-Homology (http://prospector.ucsf.edu/prospector) (supplemental Figure S(I)2). The peptide identity was further verified by the identification of the glycosylated peptide $_{132}$EGPVV**T**AQ$_{138}$ (m/z 874.28^{2+}) in a subsequent HILIC fraction (supplemental Figure S(I)5).

Immunoglobulin J Chain

The immunoglobulin J chain (joining chain) participates in the effective di-/polymerization of either IgA or IgM and is essential for the secretion of these immunoglobulins into the mucosa. In literature, the J chain was reported to be *N*-glycosylated at N_{49} [279, 290, 291]; however, *O*-glycosylation has hitherto not been described for the molecule. Interestingly, two *O*-glycopeptides detected in HILIC fractions #13 and #14 might correspond to the J chain and suggest *O*-glycosylation at T_{97} ($_{95}$DP**T**EV$_{99}$ m/z 608.72^{2+}, m/z 608.71^{2+}) (supplemental Figures S(I)1 and 2). This potentially new *O*-glycosylation site is in close vicinity to a cysteine (C_{91}) that can form a disulfide-bridge to IgM molecules. Hence, one might speculate that an occupied *O*-glycosylation site in this region might function in the establishment/preservation of this inter-molecular bond. However, the number of present fragment ions in the corresponding CID-MS^3 spectra did not allow an unambiguous identification of the peptide, as evidenced by several potential peptide hits being equally scored by the search engine. Manual fragment ion spectra annotation, though, suggest the identification of immunoglobulin J chain – nevertheless, this identification deserves further validation. Both identified *O*-glycopeptides were found to be decorated with monosialylated T-antigens.

Inter-Alpha-Trypsin Inhibitor Heavy Chain H4

For the protease inhibitor inter-alpha-trypsin inhibitor heavy chain H4, two *O*-glycosylation sites/regions, S_{640} and $T_{722/723}$ have been described in literature [277, 280]. In agreement with recent findings by Chandler *et al.*, S_{640} was found to be *O*-glycosylated. The *O*-glycopeptide $_{639}$A**S**F**S**PR$_{644}$ (m/z 660.72^{2+}) harbors two potential *O*-glycosylation sites and the occupied site could be clearly inferred from the ETD spectra by the presence of a signal at m/z 490.22^{1+}, corresponding to a z+1$_4$ ion (supplemental Figure S(I)3). In contrast to Chandler *et al.*, but in agreement with Halim *et al.*, ETD data of the *O*-glycopeptide $_{722}$**TT**QTPAPIQAPS$_{733}$ (m/z 623.27^{3+}) suggested the occupancy of the sites $T_{722/723}$ [277, 280] (supplemental Figure S(I)3). Unfortunately, none of the two potential *O*-glycosylation sites could be clearly ruled out by the detected fragment ions. Both sites/regions S_{640} and $T_{722/723}$ were decorated with a monosialylated T-antigen. This contrasts previous findings by Chandler *et al.*, who also observed a disialylated T-antigen on S_{640}.

Inter-Alpha-Trypsin Inhibitor Heavy Chain H2

For the H2 heavy chain of the Inter-alpha-trypsin inhibitor, a c-terminal cluster of mono- and disialylated mucin-type core-1 *O*-glycans (T_{666}, S_{673}, T_{675} and T_{691}) has been described in literature [277, 279, 281, 282]. These previously reported *O*-glycosylated sites, except for the site T_{666}, could be confirmed by the present study solely with monosialylated T-antigens. ETD spectra of the *O*-glycopeptide $_{689}$ES**T**PPPHV$_{696}$ (m/z 507.15^{3+}/760.24^{2+}) enabled a clear identification of the occupied *O*-glycosylation site T_{691}. This finding is supported, in particular, by a signal detected in the doubly charged species at m/z 1287.41^{1+}, which corresponds to a z+1$_6$ ion (supplemental Figure S(I)4). Remarkably, the CID-MS² spectrum of the *O*-glycopeptide $_{669}$WANP**S**P**T**PV$_{677}$ (m/z 760.92^{3+}) revealed that both *O*-glycosylation sites, S_{673} and T_{675}, are occupied by a monosialylated T-antigen (supplemental Figure S(I)5). Moreover, the spectrum features signals indicating the presence of hexose rearrangement products that is the transfer of an additional hexose either to the glycan or the peptide moiety, as described earlier [292, 293]. The occurrence of these artifacts necessitates the careful interpretation of CID glycopeptide fragment ion spectra.

Tau-Tubulin Kinase 2

The Tau-tubulin kinase 2 (TTBK2) phosphorylates tau and tubulin, preferably in the nervous system. Aberrant TTBK2 activity was linked to the progression of the Alzheimer's disease [294, 295]. The protein resides primarily in the cytosol; however, Böhm *et al.* could also detect TTBK2 in a secreted form in human tears [296, 297]. Hitherto, no glycosylation of this protein has been described. CID-MS³ as well as ETD spectra of the *O*-glycopeptide $_{814}$KDHSAT**T**EPL$_{823}$ +HexNAc$_1$Hex$_1$NeuAc$_1$ (m/z 877.82^{2+}), though, suggest the *O*-glycosylation of T_{820}. ETD fragment ions at m/z 485.22^{1+}, 557.43^{1+}, and 1098.63^{1+}, corresponding to c$_4$, c+1$_5$ and z$_4$ ions, allowed discerning the exact glycosylation site. As TTBK2 is involved in ciliogenesis [298, 299], a process which requires the vesicle transport from the Golgi to the basal bodies and cilia, we speculate that TTBK2 might become *O*-glycosylated during this process.

Fibrinogen Alpha and Beta Chain

The blood clotting protein fibrinogen is known to be *N*-glycosylated at the β- and γ-chain. Interestingly, a recent study by Zauner *et al.* could also show *O*-glycosylated sites and regions, seven in total, within the molecule [274]. In the present study *O*-glycosylation of the fibrinogen alpha region aa524-528 could be confirmed; pinpointing the exact *O*-glycosylation site was not possible, though (supplemental Figure S(I)3, $_{524}$**ST**GK**T**FPG$_{531}$, *m/z* 725.78^{2+}). Nevertheless, *O*-glycosylation within the fibrinogen beta region aa58-67 could be confirmed and pinpointed. Here, the presence of the ETD fragment ions *m/z* 931.54^{1+}, 1300.54^{1+} and 1915.50^{1+}, corresponding to $z+1_9$, c_6 and c_{12} ions (supplemental Figure S(I)4), $_{54}$EEAP**S**LRPAPPPIS$_{67}$, *m/z* 706.27^{3+}) indicates *O*-glycosylation at the site S_{58}. This contrasts recent findings by Bai *et al.*, who reported the site S_{67} to be *O*-glycosylated, but not the site S_{58} [300]. In agreement with previous findings, both fibrinogen *O*-glycopeptides ($_{524}$**ST**GK**T**FPG$_{531}$, *m/z* 725.78^{2+}, $_{54}$EEAP**S**LRPAPPPIS$_{67}$, *m/z* 706.27^{3+}), detected in the present study, were found to be decorated with monosialylated T-antigens. Interestingly, the peptide $_{54}$EEAP**S**LRPAPPPIS$_{67}$ was also found in its non-glycosylated form (HILIC fractions #12-#15, CID, see supplemental Table S2 in Hoffmann *et al.* [225]), which suggests only a partial occupation of the *O*-glycosylation site S_{58}.

4.3. Discussion

Over the last few years, mass spectrometry-based glycoproteomics has experienced significant advances in terms of instrumentation, methodology and bioinformatics. This resulted in a variety of excellent glycoproteomic publications that highlight the merits of high resolution mass spectra, complementary fragmentation techniques, improved multi-dimensional glycopeptide enrichment and separation techniques as well as sophisticated software tools, as reviewed by Thaysen-Andersen and others [42, 190, 191]. However, despite these advances – and despite its enormous clinical and pharmaceutical relevance as well as diagnostic potential – our knowledge about the human blood plasma glycoproteome is still very limited. This holds particularly true for the human blood plasma *O*-glycoproteome. Here several important questions can be raised: (I) Which proteins are *O*-glycosylated? (II) Which *O*-glycans are attached to which sites? (III) Which alterations in terms of the *O*-glycan micro- and macroheterogeneity can be observed in a certain biological context? (IV) What are the medical, biological and biotechnological implications of *O*-glycosylation?

In the present study, we have developed and employed an analytical workflow that allows the explorative, non-targeted analysis of the human blood plasma *O*-glycoproteome in a site-specific manner. To this end intact human blood plasma *O*-glycopeptides, generated by a broad-specific proteolytic digest via proteinase K, were selectively enriched using HILIC fractionation in order to be analyzed by multistage nano-RP-LC-ESI-IT-MS using low-energy CID as well as ETD (CID-MS^2/MS^3, ETD-MS^2). This combined workflow was applied on a pooled blood plasma sample derived from 20 healthy donors and allowed the identification of 31 *O*-glycosylation sites in 22 proteins, including the detection of 11 previously unknown *O*-

glycosylation sites. We were able to pinpoint 23 *O*-glycosylation sites, of which 8 sites have been described for the first time. The identified *O*-glycan compositions most probably correspond to mono- and disialylated core-1 mucin-type *O*-glycans (T-antigen).

4.3.1. Other *O*-Glycoproteomic Studies on Complex Biofluids

In the past efforts have been made to investigate the *O*-glycoproteome of different complex biological samples. Halim *et al.*, for instance, analyzed the *O*-glycoproteome of cerebrospinal fluid (CSF) using a sialic-acid capture-and-release protocol [258]. This protocol is based on the sialic acid-specific hydrazide capturing of periodate oxidized glycoproteins. Upon tryptic digestion the protocol allows the acid hydrolysis of sialic acid glycosidic bonds to release and analyze (formerly) sialylated glycopeptides. To focus on *O*-glycosylations, the authors included a PNGase F sample pre-treatment step to remove *N*-glycans. The authors have used an automated CID-MS^2/-MS^3 spectra search protocol for glycopeptide identification (Peptide-GalNAc-Gal) and have employed ECD and ETD to pinpoint the glycosylation sites. In total they have identified 106 *O*-glycosylation sites and could pinpoint 67 of these. The identified CSF *O*-glycopeptides belong to 49 different proteins that were predominately decorated with structures corresponding to core-1 mucin-type *O*-glycans. In a previous study, the same group has also investigated the human urinary *N*-and *O*-glycoproteome using a sialic-acid capture-and-release protocol [277]. Unfortunately, the applied protocol does not allow the enrichment of non-sialylated glycoproteins nor does it give any information on the degree of sialylation of the attached *O*-glycan moieties. This limits the applicability of this procedure, as the degree of *O*-glycan sialylation is a crucial determinant in the pathogenesis of a number of diseases [251].

In another large-scale glycoproteomics study conducted by Hägglund *et al.* in 2007, human plasma proteins derived from Cohn fraction IV of a plasma fractionation were analyzed with respect to occupied *N*- and *O*-glycosylation sites [279]. The analyzed Cohn fraction is supposed to contain mainly α-globulins, like plasminogen and haptoglobin, and is depleted from γ-globulins and serum albumin. The authors have employed two different enzymatic deglycosylation strategies to pinpoint occupied *N*-glycosylation sites: (I) PNGase F + $H_2{}^{18}O$; (II) endo-ß-*N*-acetylglucosaminidases (Endo D and Endo H) + different exoglycosidases. These two strategies were applied on HILIC-enriched tryptic (glyco)peptides that were fractionated by strong cation exchange chromatography and eventually measured by LC-ESI-MS/MS using high-energy CID. The authors were able to identify 103 *N*-glycosylation sites as well as 23 *O*-glycosylation sites/regions derived from 61 and 11 human blood plasma proteins, respectively. Unfortunately, the occupied *O*-glycosylation sites could not be pinpointed and no information on the glycan moiety could be deduced.

In 2012 Darula *et al.* reported on the *O*-glycoproteomic analysis of bovine serum [77]. In this study, the authors have combined different protein- and peptide-level prefractionation and enrichment strategies including jacalin lectin affinity chromatography, mixed-mode chromatography, and electrostatic repulsion hydrophilic interaction chromatography (ERLIC)

to enrich tryptic mucin-type *O*-glycopeptides. After additional use of exoglycosidases to improve glycopeptide characterization, truncated glycopeptides were subjected to LC-ESI-MS/MS with HCD and ETD for automated peptide identification and glycosylation site determination. Overall, the authors could identify and pinpoint 124 glycosylation sites in 51 proteins, including many *O*-glycosylation sites that have not been described before – unfortunately, though, at the expense of the intact glycan structure.

In a recent publication from Bai *et al.* an analytical workflow was presented, which allows the mapping of mucin-type *O*-glycosylation sites on glycoproteins present in human blood plasma [300]. The authors have used jacalin lectin affinity chromatography to enrich tryptic *O*-glycopeptides (peptide+GalNAc), which were treated with PNGase F and different exo-glycosidases. In this study, 49 *O*-glycopeptides belonging to 36 human blood plasma glycoproteins were identified by LC-ESI-MS/MS (CID). Overall, the authors could assign 13 *O*-glycosylation sites unambiguously, of which 9 sites have not been described before.

4.3.2. Proteinase K Digest

The majority of large-scale glycoproteomic studies features trypsin for the generation of (glyco)peptides. Trypsin is the proteolytic gold standard in LC-MS/MS based peptide identification and quantification, as it reproducibly generates predictable peptides that can be readily retained on reversed-phase column that give enough fragment ions for an unambiguous peptide identification, in most cases. In terms of glycoproteomics, though, the cleavage specificity of trypsin can be a limiting factor for the identification and the localization of certain glycosylation sites, in particular for densely clustered *O*-glycosylation sites. Hence, the use of broad- and non-specific proteases, like pronase E or proteinase K was proposed, to reduce the number of non-glycosylated peptides and to make certain glycosylation sites analytically amenable [262]. Proteinase K, for instance, has been successfully used in a number of publications that are centered on the *O*-glycoproteomic analysis of single proteins; though, the use of proteinase K in large-scale glycoproteomic studies on complex samples has not been described so far. In the present study, we could show that proteinase K generates (glyco)-peptides from a complex sample, like human blood plasma, in a reproducible and non-random manner, which is in agreement with a report from Hua *et al.* [262]. By analyzing a total of 646 proteinase K-generated peptides from human blood plasma proteins, we observed that proteinase K cleaves primarily after aliphatic (i.e. L, G, A, V) and polar amino acids (i.e. S, Q, T), but also negatively-charge amino acids (i.e. E, D) (Figure 33).

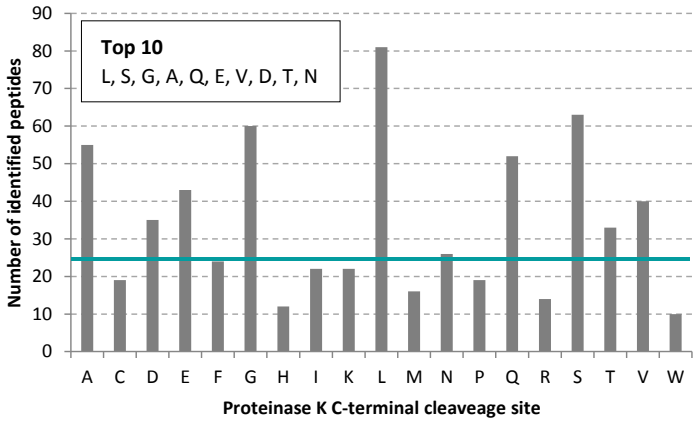


Figure 33: Preferred C-terminal cleavage sites of proteinase K (derived from the analysis of 644 proteinase K-generated peptides from human blood plasma proteins).

We could show that, most of the time, proteinase K generates shorter peptides compared to trypsin (Figure 29), and that proteinase K cleaves effectively in between densely *O*-glycosylated regions – thus, rendering the determination of the occupied *O*-glycosylation site(s) less difficult. In fact, we could show that proteinase K can generate *O*-glycopeptides that exhibit only one potential *O*-glycosylation site, thus allowing for an unambiguous localization of the occupied site. We could clearly show that some *O*-glycosylation sites could only be identified and pinpointed by the use of proteinase K, because tryptic peptides would have been too long and would have harbored too many potential *O*-glycosylation sites.

4.3.3. Glycopeptide Enrichment via HILIC

Glycopeptides are usually underrepresented in a peptide mixture, because of the glycan micro- and macroheterogeneity. In a tryptic digest of a typical glycoprotein only about 2% to 5% of the peptides are glycopeptides [301]. In addition, the ionization efficiency of glycopeptides is significantly lower compared to their non-glycosylated counterparts, thus making the efficient and selective enrichment of glycopeptides key to most glycoproteomics workflows. The use of HILIC-based glycopeptide enrichment methods has proven to be a vital tool in glycoproteomics due to their broad glycan specificity, reproducibility and compatibility with mass spectrometry. In a previous report by Zauner *et al.* it could be shown, that proteinase K-generated glycopeptides can be separated into earlier eluting *O*-glycopeptides and later eluting *N*-glycopeptides using HILIC [260]. Based on this publication we have employed HILIC for the selective enrichment and fractionation of human blood plasma *O*-glycopeptides. Here of particular importance is the removal of highly abundant non-glycosylated peptides derived from albumin and other major (glyco-)proteins. Careful manual inspection of CID-MS² fragment ion spectra of the acquired HILIC fractions revealed the efficient enrichment of glycopeptides – and indeed, the presence of solely mucin-type core-1 *O*-glycosylated glycopeptides. *N*-glycopeptides were not detected, as they were expected to be present in the late eluting HILIC wash fractions due to their generally higher hydrophilicity compared to the

most commonly found forms of mucin-type *O*-glycopeptides (non-, mono- and disialylated core-1 and -2 *O*-glycopeptides).

4.3.4. Identification of the O-Glycan Composition

For an automated glycopeptide spectra filtering and glycan fragment annotation the use of commercial software tools was considered, but turned out to be too error-prone in our case (data not shown). Hence, in the present work, we relied on manual annotation and interpretation of low-energy CID-MS² fragment ion spectra to elucidate the *O*-glycan composition – however, at the expense of throughput and the possibility to report false discovery rates. In total, we were able to characterize 88 *O*-glycopeptides with respect to their *O*-glycan composition. The detected *O*-glycan compositions most likely correspond to mucin-type core-1 mono- and disialylated *O*-glycans ((di)sialyl-T-antigen). In agreement with literature, glycopeptides carrying disialylated *O*-glycans, were found in later eluting HILIC fractions (#15-#17), as the additional sialic acid renders the molecule more hydrophilic (Figure 26). Mono- and disialylated glycoforms could be usually discriminated by the presence of distinct oxonium ions: while fragmentation of monosialylated *O*-glycans generated a characteristic oxonium ion at *m/z* 454.16 (Hex_1NeuAc_1), disialylated *O*-glycans gave rise to an additional intense peak at *m/z* 495.18 ($HexNAc_1NeuAc_1$) (supplemental Figure S(I)4, $_{266}$EAV-P**T**PVVDPDAPPSPPL$_{283}$, *m/z* 818.68^{3+}, $_{267}$AVP**T**PVVDPDAPPSPPL$_{283}$, *m/z* 872.73^{3+}). Furthermore, in disialylated species, characteristic fragment ions of the peptide+HexNAc+NeuAc were observed. In a few cases the glycan annotation was compromised by the presence of fragment ions corresponding to hexose rearrangement products [292, 293]. Generally, it is important to note that low-energy CID-MS² fragmentation of glycopeptides does usually not produce fragment ions that relate to the linkage of the attached monosaccharides. Therefore, validation of the inferred *O*-glycan structures using dedicated *O*-glycomics approaches, including for instance (reductive) beta-elimination or hydrazinolysis, is recommended. However, our findings are in good agreement with literature, as mono- and disialylated mucin-type core-1 *O*-glycans are known to be present on the majority of secreted blood plasma glycoproteins, produced by hepatic cells of healthy individuals [229]. Notably, a study on the plasma-derived von Willebrand factor could show that apart from mucin-type core-1 *O*-glycans (T-antigen), more complex *O*-glycan structures including ABH blood group antigen containing mucin-type core-2 ([GalNAcβ1-6-(Galβ1-3)-GalNAcα-*O*-Ser/Thr]), can be present on human blood plasma glycoproteins, too [302]. In the present work, analyzing the total human blood plasma *O*-glycoproteome, we could not detect any (glyco)peptides derived from von Willebrand factor nor could we find any indication for the presence of fucosylated (ABH blood group antigens) and/or LacNAc extended mucin-type core-2 *O*-glycans.

4.3.5. Glycopeptide Identification

Low-energy CID-MS² fragmentation of glycopeptides, as employed in the present work, almost exclusively generates fragment ions corresponding to the glycan moiety, while leaving the peptide backbone mainly intact. Thus, this type of fragmentation does usually not provide

any information on the sequence of the peptide backbone nor on the occupied glycosylation site. To identify the peptide, we have employed manual CID-MS³ fragmentation on the putative peptide mass, which has been inferred from the annotation of the corresponding CID-MS² spectra before. In a few cases, the signal of the putative peptide mass was too low to yield sufficient fragment ions. Consequently, the putative peptide+HexNAc ion was subjected to CID-MS³ fragmentation instead. We did not employ an automated CID-MS³ fragmentation procedure, e.g. fragmentation of the three most intense precursor ions in the CID-MS² spectrum, because we wanted to generate and sum up as many fragment ion spectra as possible from the selected putative peptide mass, to increase spectra quality and therefore the chance of successful peptide identification. By searching the acquired CID-MS³ fragment ion spectra against the human subset of the UniProtKB/Swiss-Prot protein database, a total of 60 peptides (of 88 detected *O*-glycopeptides) could be identified unambiguously. Notably, in a few cases, also peptide fragment ions present in CID-MS² spectra allowed an unambiguous peptide identification (supplemental Figure S(I)4, $_{267}$AVP**T**PVVDPDAPPSPPL$_{283}$, m/z 872.73^{3+}). Overall, the identified peptides belong to 22 different proteins – primarily acute phase proteins. This group of blood plasma proteins fulfills essential functions during inflammation (e.g. coagulation, anti-inflammatory and anti-pathogenic activity), and, accordingly, their expression is known to be either significantly up- or downregulated (positive and negative acute phase proteins) in this context. As a result, this group of proteins attracted a lot of attention as potential cancer biomarkers in recent years [236]. Noteworthy, the identified proteins span a concentration range of five orders of magnitude. Therefore, the applied approach seems to be suitable to also detect lower abundant proteins or peptides.

A group of *O*-glycosylated proteins that have frequently been identified in other large-scale glycoproteomic studies are coagulation factors [77, 258, 277, 279]. In our study, there is an indication for the presence of an *O*-glycosylated peptide derived from coagulation factor V (HILIC fraction #15, m/z 761.78^{2+}, $_{1453}$QI**S**PPPDL$_{1460}$+HexNAc$_1$Hex$_1$NeuAc$_1$, Table 2, supplemental Figure S(I)3). Interestingly, the detected coagulation factor V *O*-glycosylation site (S$_{1455}$) has not been described so far. Unfortunately, our data do not allow an unambiguous identification of this protein.

General remarks on Immunoglobulin *O*-Glycoproteomics

Another *O*-glycosylated protein that could not be identified in our study is Ig alpha-1 (IgA1). IgA1 is a high abundant human blood plasma glycoprotein that features a cluster of three to five mucin-*O*-glycans in the hinge region of the heavy chain [303]. This cluster harbors many prolines, hence corresponding proteinase K-generated peptides might have been not unambiguously identified (the tryptic IgA1 hinge region *O*-glycopeptide looks as follows: (K)$_{89}$HYTNPSQD-VTVPCPVP**S**TPPTP**S**P**S**TPPTP**S**P**S**CCHP**R**$_{126}$). Furthermore, due to the densely clustered *O*-glycans, a potential IgA1 *O*-glycopeptide carrying mucin-type *O*-glycans at each potential site, such as P**S**TPPTP**S**P**S**TPPTP**S**P**S**CC, might be too hydrophilic and consequently might have been among the (glyco)peptides present in the late eluting HILIC wash fraction. Worth

mentioning, in our study we could detect the IgA1 peptide $_{95}$QDVTVPCPVP**S**$_{105}$ in its non-glycosylated form (HILIC Fraction #11, CID, supplemental Table S2 in Hoffmann *et al.*, 2016 [225]). Therefore, the *O*-glycosylation site S_{105} seems to be only partially occupied. Surprisingly, human IgA1 *O*-glycopeptides have not been identified in any other large-scale glycoproteomic studies [77, 258, 277, 279, 300, 304]. However, there is a targeted glycoproteomic study from Takahashi *et al.* focusing on IgA1 *O*-glycosylation [303]. In this study, the authors analyzed human plasma derived IgA1 *O*-glycopeptides (tryptic and non-tryptic) with ESI-FT-ICR-MS/MS as well as ESI-LTQ-FT-MS/MS, both in online- and offline-mode. To pinpoint the *O*-glycosylation sites the authors have employed activated ion-electron capture dissociation (AI-ECD) and ETD. Another immunoglobulin that is reported to carry mucin-type *O*-glycans in the hinge region is IgD [305]. The plasma concentration of IgD is much lower than the concentration of IgA, IgG and IgM but higher than that of IgE (IgD represents 0.25 % of total plasma immunoglobulins). Apart from the study conducted by Takayasu *et al.* from 1982 [305] on truncated *O*-glycopeptides (peptide+GalNAc), at present no *O*-glycoproteomic data do exist for intact human IgD *O*-glycopeptides. Also of particular interest is a recent finding by Plomp *et al.*: using a targeted glycoproteomics approach these authors could demonstrate, for the first time, that IgG3 is partially *O*-glycosylated in its hinge region(mucin-type core-1 *O*-glycans) [306].

4.3.6. Pinpointing of O-Glycosylation Sites

Pinpointing the correct *O*-glycosylation sites is a crucial but very challenging task. Proteinase K, in this regard, proved to be beneficial as it can generate short glycopeptides, which exhibit only one potential *O*-glycosylation site. In case the occupied *O*-glycosylation site could not be inferred directly, we have employed ETD-MS² fragmentation. In first attempts of identifying the acquired ETD glycopeptide spectra, database-assisted peptide identification via MASCOT was tested – but turned out to be not successful. One reason for this is the presence of intense signals in the ETD-MS² spectrum, which do not correspond to peptide fragment ions (e.g. unfragmented precursor ions, glycan fragment ions), and which thus compromise automated peptide identification [307]. A possible solution for this is the (manual) removal of these additional *m/z*-values from the ETD-spectra before running the search algorithm. In the present study, however, this procedure did not improve the database-assisted peptide identification. Therefore, we relied on manual spectra annotation and interpretation using Bruker DataAnalysis, and Bruker Biotools as well as public repositories (UniProtKB and UniCarbKB). Furthermore, NetOGlyc 4.0 was employed to predict *O*-glycosylation sites and to support experimental findings. Predicted and experimentally determined *O*-glycosylation sites were mostly in good agreement for already known *O*-glycosylation sites – however, support for potentially novel sites could only be found in a few cases. A general shortcoming of glycopeptide enrichment methods is that they are biased towards glycosylated peptides, while underrepresenting potential corresponding aglyosylated counterparts. Hence, in the present study no conclusions with respect to the macro-heterogeneity of the glycoproteins (site-occupancy) can be drawn.

4.3.7. Caveats of the Approach

In contrast to tryptic (glyco)peptides, proteinase K-generated peptides and glycopeptides cannot be predicted due to the broad cleavage specificity of the enzyme. More importantly, though, is the reduced peptide length compared to a tryptic digest, as this can lead to an insufficient number of detected fragment ions to allow unambiguous peptide identifications. This problem can be even more intensified by the frequent occurrence of prolines within mucin-type *O*-glycopeptide sequences, as prolines can introduce additional sequence gaps during mass spectrometry-based peptide sequencing. Also important to note is the increased search space of the search engine due to the use of a non-specific enzyme, which results in an increased ambiguity with respect to the peptide identification (lower identification scores) and longer search times. A confounding factor that relates to the ETD analysis is the predominance of charge state 2^+ among the measured *O*-glycopeptide precursor ions, since ETD fragmentation is more efficient for precursor charge states greater than 2^+ [308]. The predominance of charge state 2^+ can be explained by a lack of ionizable/basic amino acids (lack of Arg, Lys, His) within the glycopeptides – a characteristic that can be linked to the broad-specific proteolytic digest by proteinase K [309]. Another caveat is related to the HILIC glycopeptide enrichment: this step was optimized to enrich *O*-glycopeptides carrying short mucin-type core-1 and -2 *O*-glycans, as they represent the vast majority of *O*-glycans on human blood plasma proteins [253]. Hence, *O*-glycopeptides carrying bigger and thus more hydrophilic *O*-glycans, such as *N*-acetyl-lactosamine (LacNAc) extended mucin-type core-2 *O*-glycans, or *O*-glycopeptides carrying multiple mucin-type *O*-glycans, might elute in the subsequent washing phase of the HILIC fractionation and as a consequence cannot be found during the analysis.

4.4. Summary

In summary, in the present study we have investigated the human blood plasma mucin-type *O*-glycoproteome of healthy individuals in an explorative and non-targeted manner. To this end, we have conducted a site-specific large-scale *O*-glycoproteomic analysis, which combines the use of a proteolytic enzyme with broad cleavage specificity, with HILIC enrichment/fractionation and subsequent multi-stage mass spectrometry measurement (nano-RP-LC-ESI-IT-MS^n) via CID and ETD. Centered on the characterization and identification of intact glycopeptides, we could demonstrate the in-depth *O*-glycoproteomic analysis of a number of important human blood plasma glycoproteins (mainly acute phase proteins), including alpha-2-HS-glycoprotein, fibrinogen, plasminogen and kininogen-1. Our results are in good agreement with previous findings by other research groups, but also add new aspects to the field, e.g. the identification of a couple of novel *O*-glycosylation site as well as the benefits and drawbacks of using proteinase K in large-scale mass spectrometric glycoproteomic studies.

5 Chapter Five

In-Depth *N*- and *O*-Glycoproteomics

5.1. Introduction

Please note, parts of this chapter are taken from the original publication Hoffmann *et al.* (2018) [227].

The reliable and unambiguous mass spectrometric identification and characterization of *N*- and *O*-glycopeptides is indispensable for in-depth glycoproteomic studies. This chapter describes the development of a glycoproteomic workflow for the analysis of *N*- and *O*-glycosylated proteins that allows the detailed characterization of the glycan moiety and the peptide backbone alike. The workflow is focused on the analysis of tryptic and non-tryptic glycopeptides that were enriched by spin-cotton-HILIC-SPE and measured by nano-RP-LC-ESI-OT-OT-MS/MS using HCD fragmentation (Figure 22). We demonstrate the workflow's efficiency, flexibility, merits and limitations with the help of four representative glycoproteins, comprising IgG, LTF, Fib, and RNase B. This set of glycoproteins covers a broad range of glycan and protein features relevant for many glycoproteomic studies, including different types of glycosylation (*N*- and *O*-glycosylation), different glycoforms (complex-, hybrid-, high-mannose-type *N*-glycans, mucin-type *O*-glycans; microheterogeneity), specific glycan features (core- vs antenna-fucosylation, bisecting *N*-acetylglucosamine, degree of sialylation) as well as multiple glycosylation sites (macroheterogeneity). Moreover, we highlight the importance of characteristic oxonium ion patterns that can be used to characterize the glycan moiety more precisely.

5.2. Results and Discussion

5.2.1. Fragmentation of N- and O-Glycopeptides by HCD.low and HCD.step

Keystone of our analysis workflow is the development of two Orbitrap Elite MS HCD fragmentation regimes (MS^2), namely *HCD.low* and *HCD.step* (Figure 34). HCD fragmentation is carried out in the nitrogen-filled HCD cell of the mass spectrometer. The applied normalized collisional energies during this process can be varied individually, such that for instance low- and high-energy-fragmentation can be carried out sequentially on the very same precursor ion. The readout of the resulting fragment ions is done in the Orbitrap mass analyzer – where fragment ion spectra can be acquired with high resolution and without compromising information by a low-mass cut-off. The latter is inherent to (linear quadrupole) ion trap mass analyzers usually used for CID fragmentation and leads to the absence of important fragment ions in the low-molecular range region (fragment ions with m/z values <30% that of the precursor ion are not stable in the ion trap, and therefore not present in the fragment ion spectrum). The low m/z range region of *N*- and *O*-glycopeptide fragment ion spectra is of particular interest for glycopeptide analyses, as it features diagnostic fragment ions derived from the glycan moiety of the glycopeptide, so-called oxonium ions. These singly-charged low-molecular weight fragment ions (B-ions, according to Domon and Costello [198]) can not only be used to discriminate a glycopeptide spectrum from a non-glycopeptide spectrum, they can also provide crucial information on the glycan moiety that is linked to the peptide backbone.

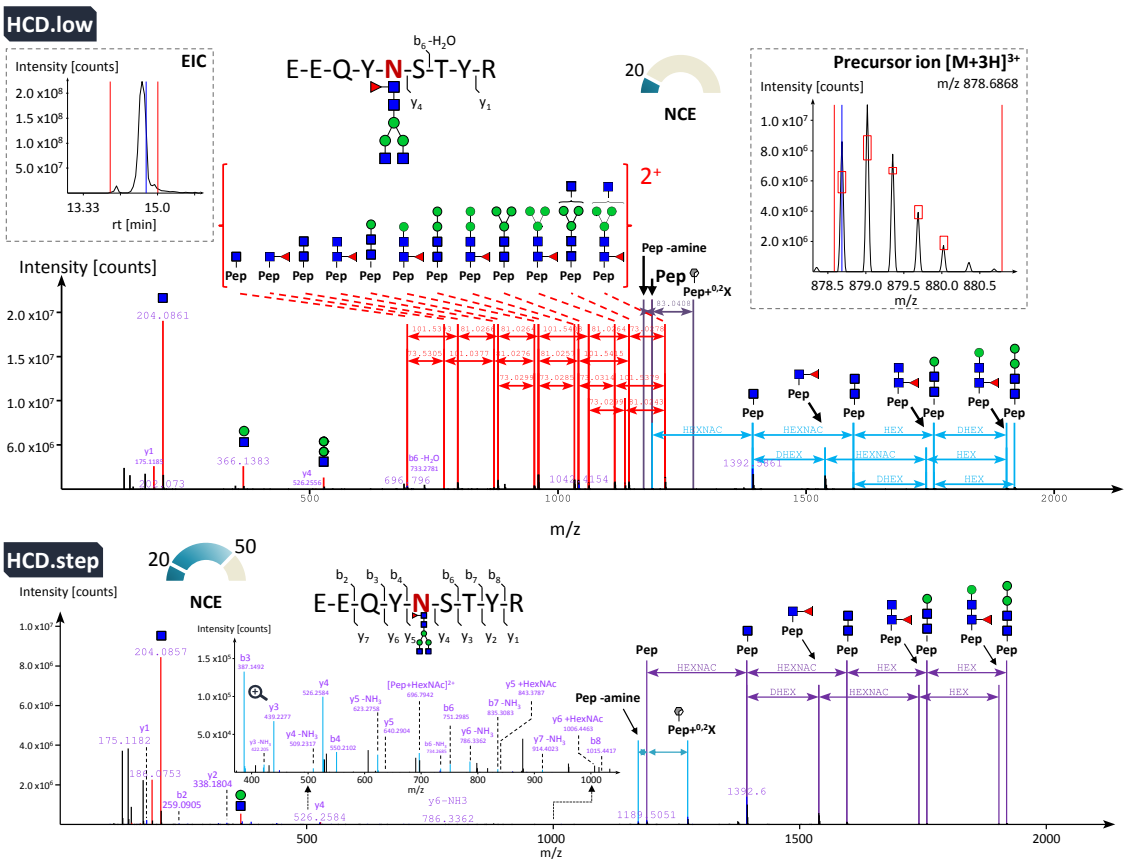


Figure 34: Glycopeptide fragmentation with *HCD.low* and *HCD.step*. HCD fragment ion spectra (MS^2) of a tryptic IgG 1 Fc *N*-glycopeptide $_{176}$EEQY**N**STYR$_{184}$ + $HexNAc_4Hex_3Fuc_1$ (*m/z* 878.687 $[M+3H]^{3+}$) acquired by nano-RP-LC-ESI-OT-OT-MS/MS using varying normalized collisional energies (NCE; positive ion mode, LTQ Orbitrap Elite hybrid mass spectrometer). **Top:** *HCD.low* – fragment ion spectrum acquired at low NCE (20): enables detailed analysis of the glycan moiety. Putative peptide mass can be deduced via a conserved fragmentation pattern: (I) $[M_{peptide} +H -NH_3]^+$, (II) $[M_{peptide} +H]^+$, (III) $[M_{peptide} +H +^{0,2}X\ GlcNAc]^+$, and (IV) $[M_{peptide} +H +GlcNAc]^+$. **Top (inset, left):** extracted ion chromatogram (EIC). **Top (inset, right):** isotopic pattern of the precursor ion. **Bottom:** *HCD.step* – fragment ion spectrum acquired using stepped NCE (20, 50): enables unambiguous identification peptide backbone. Again, the peptide mass can be deduced via a conserved fragmentation pattern.

During *HCD.low* fragmentation a constant NCE of 20 was used. At this low collisional energy, the resulting *N*- and *O*-glycopeptide fragment ion spectra resemble those acquired by CID fragmentation performed in an ion trap – yet at a higher resolution and without any low-mass cut-off. *HCD.low* fragment ion spectra were acquired using an *m/z* range of 100-2000. Similar to CID, *N*- and *O*-glycopeptide fragment ion spectra acquired by *HCD.low* are also mainly populated by fragment ions derived from the glycan moiety (B- and Y-ions), and only rarely feature fragment ions derived from the more stable peptide backbone (b- and y-ions) (Figure 34, top).

The most prominent glycan-derived signals in *HCD.low* glycopeptide spectra correspond to the successive neutral loss of monosaccharides from the non-reducing end towards the reducing end of the attached glycan (Y-ions, mostly singly, but also doubly charged). These fragment ions along with diagnostic oxonium ions are the basis for the determination of the glycan

composition. For *N*-glycopeptides the successive fragmentation of the glycan moiety usually leads to its complete loss – leaving the peptide still intact. Thereby, the signal corresponding to the peptide mass is often rather low, or even absent. Nevertheless, the putative peptide mass can still be determined with high reliability in most cases by means of another fragment ion: a main event during *HCD.low* and CID fragmentation is the cleavage of the glycosidic bond in between the chitobiose core, which results in an intense signal corresponding to the peptide with an additional *N*-acetylglucosamine (GlcNAc). This signal in turn allows to infer the putative peptide mass (Y_1 fragment ion) and thus to identify the peptide. Still, an unambiguous identification of the peptide is only possible with an appropriate number of peptide fragment ions.

For *O*-glycopeptides, particularly for the most commonly found mucin-type core-1 and core-2 *O*-glycopeptides, the glycan moiety comprises usually only a few monosaccharides (most commonly found glycoform: $GalNAc_1Gal_1NeuAc_{1/2}$) [207]. This normally results in less complex *HCD.low* fragment ion spectra. As for *N*-glycopeptides, the successive fragmentation of the glycan moiety of *O*-glycopeptides up to the de-glycosylated peptide mass can be observed. In contrast to *N*-glycopeptides, the signal corresponding to the peptide mass is usually rather intense. Additionally, *HCD.low* fragmentation of mucin-type *O*-glycopeptides is more likely to exhibit enough peptide-derived b- and y-ions to allow for an unambiguous identification. The latter can be linked to the smaller glycan moiety typically found for mucin-type *O*-glycopeptides.

Apart from *HCD.low* we have added a second fragmentation regime to our workflow. *HCD.step* was developed to complement *HCD.low* fragmentation, such that for all selected and fragmented *N*-and *O*-glycopeptides always enough peptide fragment ions are generated to allow for an unambiguous peptide identification, without sacrificing too much information of the glycan moiety. *HCD.step* features a two-step fragmentation duty cycle, where every selected precursor ion is fragmented at both, low- and high-collisional energy. As such, the resulting fragment ion spectra are hybrid spectra, combining fragment ions arising from the application of two different collisional energies. For the analysis of *N*- and *O*-glycopeptides this approach proved to be beneficial as it generates glycan-specific fragment ions and peptide backbone fragment ions alike. Similar to *HCD.low*, *HCD.step* also generates intense glycan-derived signals corresponding to oxonium ions and neutral loss fragment ions. The latter, though, are mostly present as singly charged species. Hence, at the given acquisition window of m/z 100-2000, the number of glycan-derived fragment ions that can be used to deduce the glycan composition is lower, compared to *HCD.low*. Still, *HCD.step* fragmentation generates enough Y-ions to reliably infer the putative peptide mass. Thereby the signal corresponding to the peptide mass is always present with decent intensity. Likewise, a high number of b- and y-ions are reliably generated – for both, *N*- and *O*-glycopeptides. The detection of these fragment ions along with the putative peptide mass are the basis for a reliable and unambiguous identification of the peptide moiety (Figure 34, bottom).

Further confidence in the correct peptide identification is provided by the presence of amino acid immonium ions in the low *m/z* range (immonium ions > 100 Da detected by *HCD.step*). These marker ions arise from internal fragmentation (a- and y-ions) of the peptide backbone, and can not only support peptide identification but can also be crucial for *de novo* sequencing [310, 311]. Another striking feature of *HCD.step* fragmentation is that it generates a conserved fragmentation pattern (I-IV) around the peptide mass of *N*-glycosylated peptides: (I) $[M_{peptide} +H -NH_3]^+$, (II) $[M_{peptide} +H]^+$, (III) $[M_{peptide} +H +^{0,2}X\ GlcNAc]^+$ (cross-ring fragment of the innermost GlcNAc), and (IV) $[M_{peptide} +H +GlcNAc]^+$ (Figure 35).

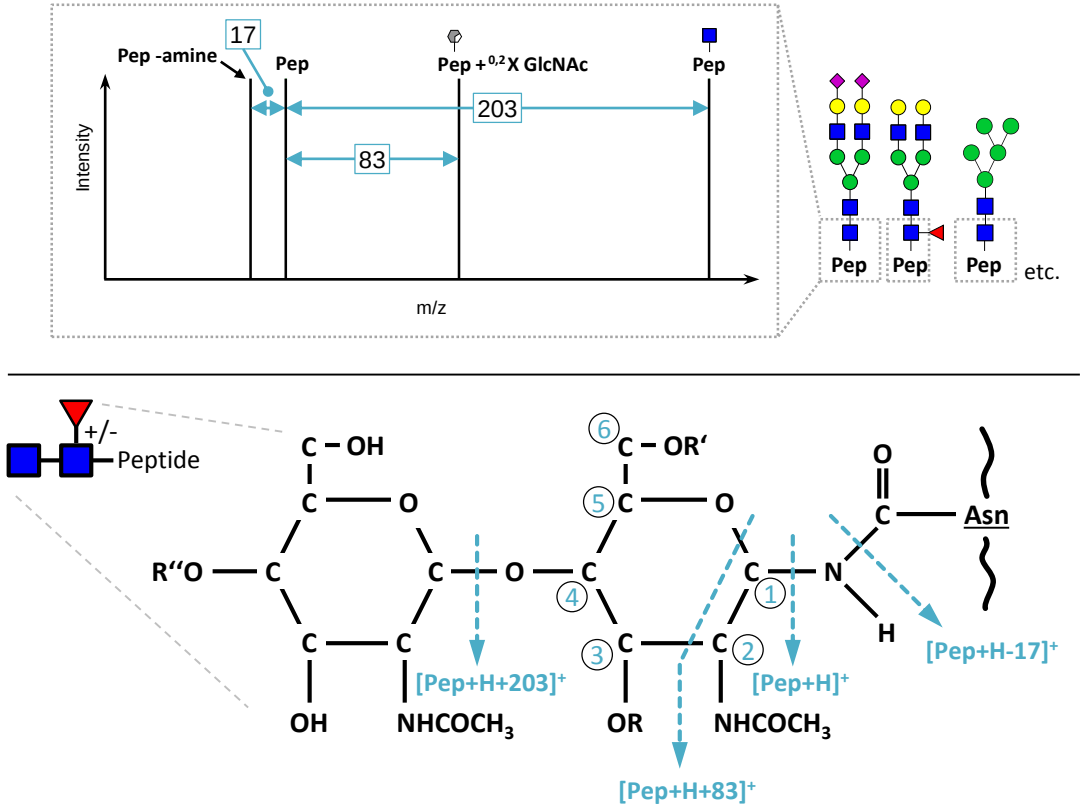


Figure 35: Conserved *N*-glycopeptide fragmentation pattern consistently obtained when using stepped normalized collisional energy fragmentation (*HCD.step*). **Top:** (I) $[M_{peptide} +H -NH_3]^+$, (II) $[M_{peptide} +H]^+$, (III) $[M_{peptide} +H +^{0,2}X\ GlcNAc]^+$, and (IV) $[M_{peptide} +H +GlcNAc]^+$. The fragmentation pattern occurs for core-fucosylated and non-core-fucosylated *N*-glycopeptides. **Bottom:** schematic representation of the fragmentations that occur at the innermost GlcNAc of the *N*-glycopeptide. R and R' = H or fucose. R''= glycan chain. Illustration modified from Wuhrer *et al.*, 2007 [44].

This *N*-glycopeptide fragmentation pattern has already been reported for MALDI-TOF/TOF analyses by Wuhrer *et al.* [44]. Recently this pattern was also identified for Orbitrap HCD fragmentation by Dong *et al.* [312] and Stadlmann *et al.* [313]. However, the detection of this fragment pattern among different *N*-glycosylation types was not systematically evaluated up to now. In our study, this pattern was consistently found in every *N*-glycopeptide fragment ion spectrum acquired by *HCD.step* – for complex-, high-mannose-, and hybrid-type *N*-glycopeptides alike. For *HCD.low* fragmentation the pattern was also found – not consistently and only at low intensity, though. Interestingly, the pattern was not observed for mucin-type *O*-glycopeptides. Therefore, using this conserved fragmentation pattern not only provides

additional confidence in the annotation of the peptide mass, and with this, also more confidence regarding sequencing of the peptide backbone and the glycan moiety; it also allows for a more reliable differentiation between *N*-glycopeptides and mucin-type *O*-glycopeptides. Of note, using solely high-energy collisional fragmentation (*HCD.high*) on *N*- and *O*-glycopeptides results in fragment ion spectra of low intensity that predominantly feature oxonium and immonium ions. Thereby only a very few to none peptide backbone fragment ions (water-loss b- and y-ions), and no glycan-derived Y-ions, can be registered (Figure 36).

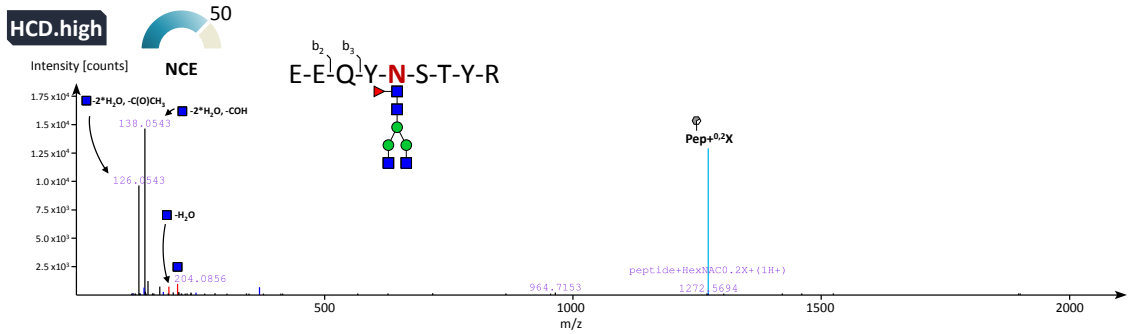


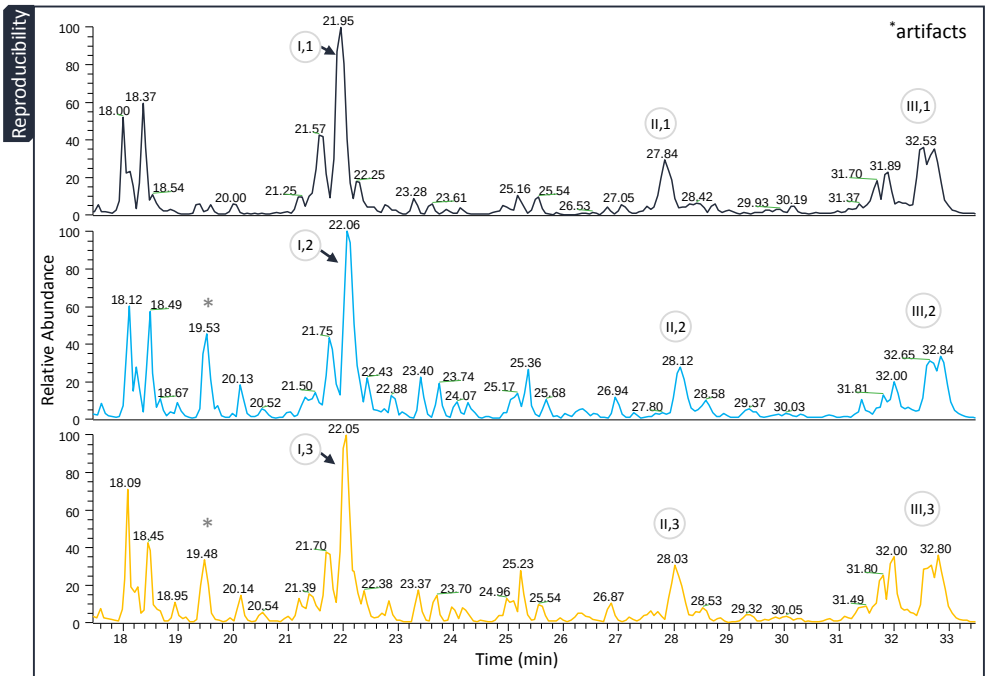
Figure 36: Glycopeptide fragmentation with *HCD.high*. HCD fragment ion spectrum (MS^2) of a tryptic IgG 1 Fc *N*-glycopeptide $_{176}$EEQY**N**STYR$_{184}$ + HexNAc4Hex3Fuc1 (m/z 878.687 $[M+3H]^{3+}$) acquired by nano-RP-LC-ESI-OT-OT-MS/MS using high normalized collisional energies (NCE (50); positive ion mode, LTQ Orbitrap Elite hybrid mass spectrometer). The fragment ion spectrum exhibits only very few fragment ions, predominantly oxonium and immonium ions. Peptide-derived b- and y-ions, as well as glycan-derived Y-ions are largely absent.

A general caveat of *HCD.step* compared to *HCD.low* is, that due to the acquisition with two different collisional energies, the duty cycle of *HCD.step* is slightly longer. Depending on the complexity of the sample, this can lead to a decrease in the number of acquired fragment ion spectra. For tryptic IgG glycopeptides, for example, 40% less fragment ion spectra were acquired by *HCD.step* compared to *HCD.low* (*HCD.low*: 3690; *HCD.step*: 2281 spectra). Still, for the majority of glycopeptides identified in our study we were able to acquire an *HCD.step* as well as *HCD.low* fragment ion spectrum (63 %). Another aspect that needs to be considered when using *HCD.step* fragmentation is that internal glycopeptide fragment ions can occur for peptide sequences which contain proline: for a tryptic *N*-glycopeptide ($_{23}$CANLVPVPITNATLDR$_{38}$ +HexNAc$_4$Hex$_5$NeuAc$_2$) derived from alpha-1-acid glycoprotein 2 we have noticed internal glycopeptide cleavages N-terminal to proline (Supplemental Figure S(II)1, subsections S(II)1-101 – 110). Accordingly, two truncated *N*-glycopeptides ("y-ions", $_{28}$PVPITNATLDR$_{38}$ +HexNAc$_2$, $_{30}$PITNATLDR$_{38}$ +HexNAc$_2$) were detected by the corresponding conserved *N*-glycopeptide fragmentation patterns, including the respective peptide masses and peptide fragment ions. This cleavage pattern might be explained by the cyclic structure of proline and the ability to introduce a kink in alpha helices [314]. Careful evaluation of this fragmentation behavior is important to avoid spectra misinterpretation.

5.2.2. Spin-Cotton-HILIC-SPE

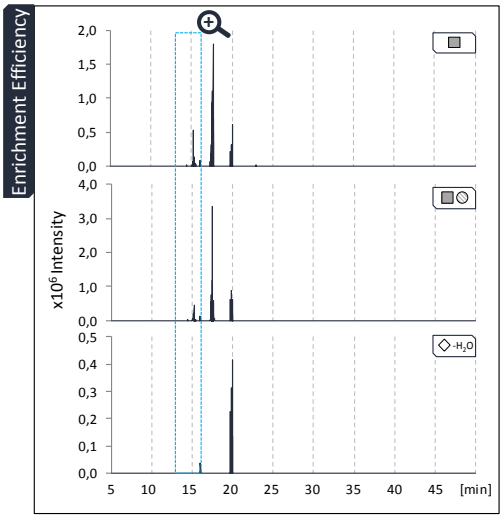
As a result of the glycan micro- and macroheterogeneity, glycopeptides are usually of lower abundance compared to their non-glycosylated counterparts. This, along with their inherent lower ionization efficiency, necessitates the enrichment of glycopeptides in many glycoproteomic studies. In our workflow we have implemented a modified version of the cotton-HILIC-SPE for the selective enrichment of *N*- and *O*-glycopeptides, published by Selman *et al.* [226]. The modified spin-cotton-HILIC-SPE comprises the following two differences to the original protocol: (I) the amount of cotton filled in the pipette tips was increased by a factor of 12 (before: 0.5 mg; new: 6 mg cotton wool), and (II) centrifugation, instead of pipetting up and down, was employed throughout the entire procedure (equilibration, loading, washing, and elution). While the former (I) allows for a higher loading capacity – at least 10 µg (glyco)peptide digest instead of 3.3 µg can be treated without any overloading effects – the latter (II) renders the workflow less cumbersome and labor-intensive, and at the same time allows a higher throughput (several samples can be prepared in parallel, depending on the centrifuge employed). Reproducibility and glycopeptide enrichment efficiency of the new spin-cotton-HILIC-SPE was assessed using tryptic peptides derived from LTF and IgG (Figure 37, A-E).

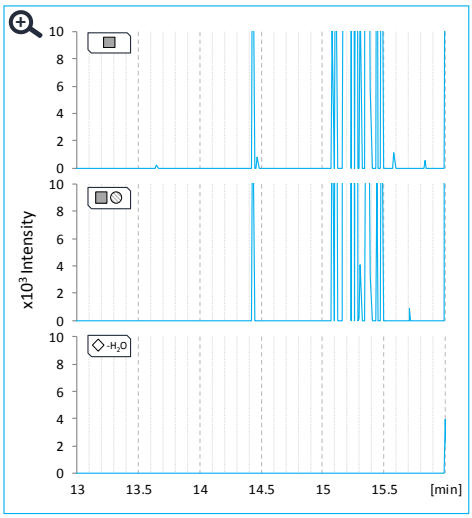
A Spin-Cotton-HILIC-SPE Triplicates, BPC (MS^1)

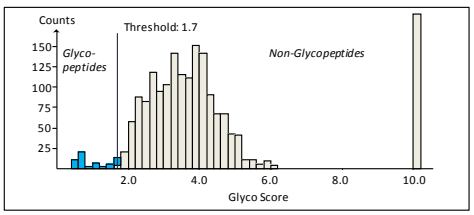


Peak	I	II	III
Base Peak [*m/z*]	894.6926 $[M+3H]^{3+}$	991.7254 $[M+3H]^{3+}$	1206.5107 $[M+2H]^{2+}$
Peak Area [AU, arbitrary units]	I,1: 3.03 x10^8 I,2: 2.38 x10^8 I,3: 2.80 x10^8	II,1: 8.75 x10^7 II,2: 6.84 x10^7 II,3: 1.00 x10^8	III,1: 9.03 x10^7 III,2: 7.20 x10^7 III,3: 9.07 x10^7
Peak Area, Mean [AU, arbitrary units]	2.74 x10^8	8.53 x10^7	8.43 x10^7
Peak Area, Coefficient of Variation [%]	11.94	18.72	12.68
Average Coefficient of Variation [%]	14.45		

B IgG 2 | Trypsin | untreated | *HCD.step*





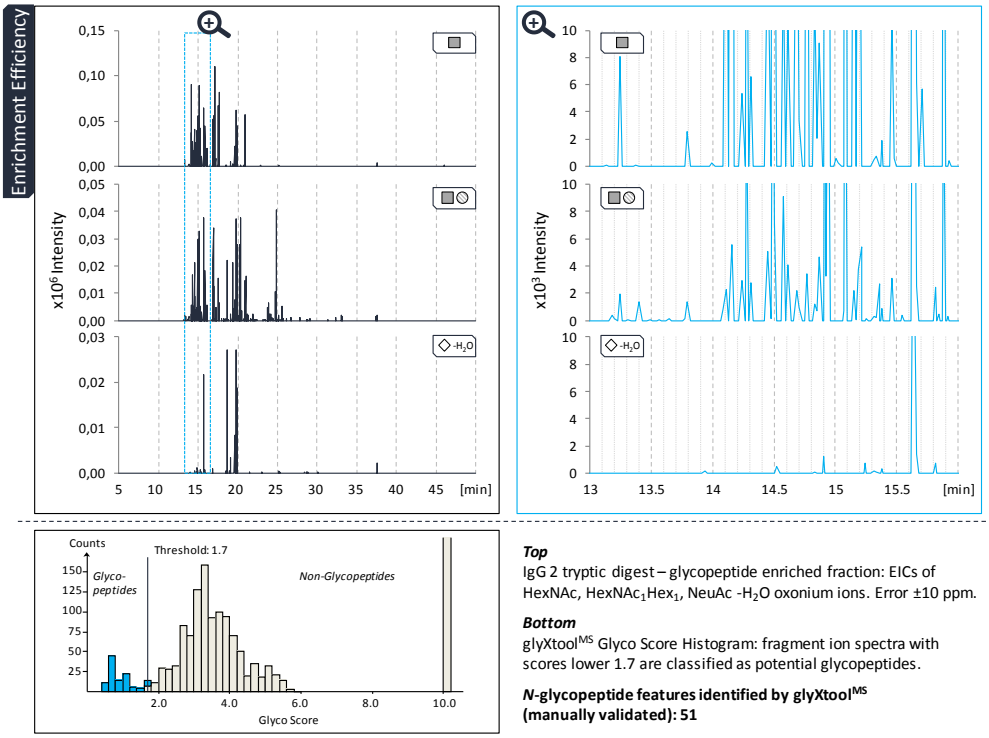


Top
IgG 2 tryptic digest – untreated: EICs of HexNAc, $HexNAc_1Hex_1$, NeuAc $-H_2O$ oxonium ions. Error ±10 ppm.

Bottom
glyXtool^{MS} Glyco Score Histogram: fragment ion spectra with scores lower 1.7 are classified as potential glycopeptides.

***N*-glycopeptide features identified by glyXtool^{MS} (manually validated): 25**

C IgG 2 | Trypsin |
Spin-Cotton-HILIC-SPE glycopeptide-enriched fraction | *HCD.step*

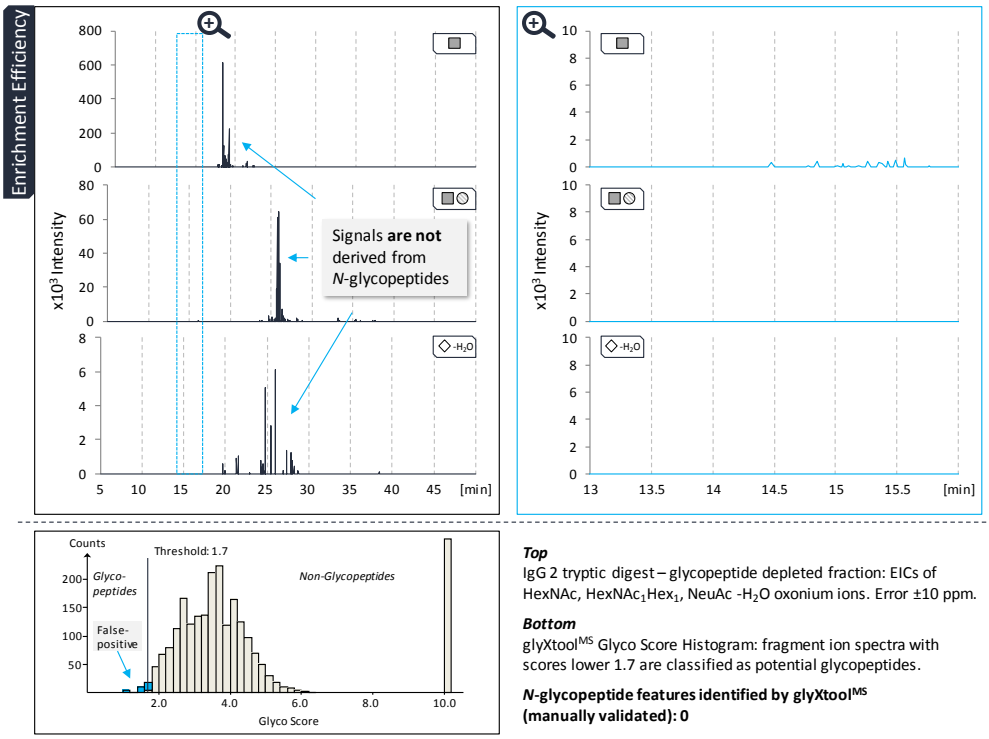


Top
IgG 2 tryptic digest – glycopeptide enriched fraction: EICs of HexNAc, $HexNAc_1Hex_1$, NeuAc $-H_2O$ oxonium ions. Error ±10 ppm.

Bottom
glyXtool^{MS} Glyco Score Histogram: fragment ion spectra with scores lower 1.7 are classified as potential glycopeptides.

***N*-glycopeptide features identified by glyXtool^{MS} (manually validated): 51**

D IgG 2 | Trypsin |
Spin-Cotton-HILIC-SPE glycopeptide-depleted fraction | *HCD.step*



Top
IgG 2 tryptic digest – glycopeptide depleted fraction: EICs of HexNAc, $HexNAc_1Hex_1$, NeuAc $-H_2O$ oxonium ions. Error ±10 ppm.

Bottom
glyXtool^{MS} Glyco Score Histogram: fragment ion spectra with scores lower 1.7 are classified as potential glycopeptides.

***N*-glycopeptide features identified by glyXtool^{MS} (manually validated): 0**

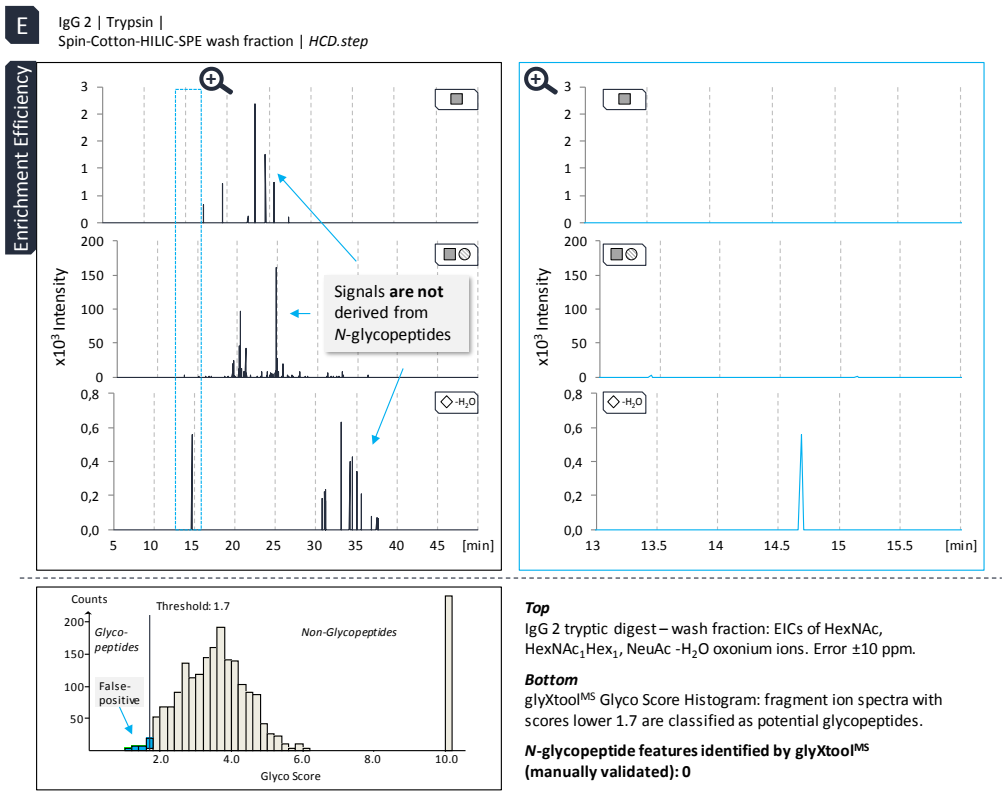


Figure 37: Glycopeptide enrichment via cotton-HILIC-SPE: Evaluation of the reproducibility and enrichment efficiency. **A:** Comparison of the BPCs (MS^1) of three independent cotton-HILIC-SPE preparations shows a high reproducibility of the procedure (LTF glycopeptides, glycopeptide-enriched fraction, technical replicates). **B-E:** Evaluation of the glycopeptide enrichment among untreated and cotton-HILIC-SPE preparations (glycopeptide-enriched, glycopeptide-depleted, and wash fractions) of tryptic IgG *N*-glycopeptides (HCD.step MS^2) demonstrates a high glycopeptide enrichment efficiency of the cotton-HILIC-SPE (51 IgG *N*-glycopeptides with enrichment, and only 25 IgG *N*-glycopeptides without enrichment). The evaluation is based on the comparison of EICs for three oxonium ions (HexNAc, $HexNAc_1Hex_1$, NeuAc) as well as manual spectra identification and validation in glyXtool^{MS}. The asterisk(*) in **A:** indicates the presence of artifacts.

Comparing the base peak chromatograms (BPCs MS^1) of three LTF glycopeptide spin-cotton-HILIC-SPE preparations shows a high reproducibility (average coefficient of variation: 14.45% for the peak areas of three representative chromatographic peaks, Figure 37 A). The glycopeptide enrichment efficiency was evaluated by analyzing the *glycopeptide-depleted fraction* (loading), the *wash fraction*, and the *glycopeptide-enriched fraction* (elution) of spin-cotton-HILIC-SPE treated IgG glycopeptides. To this end, the *HCD.low* fragment ion spectra of the aforementioned HILIC fractions were screened for the presence of glycopeptides, both manually and software-assisted (glyXtool^{MS}). Consistently, both analyses revealed a lossless enrichment of glycopeptides in the *glycopeptide-enriched fraction* (elution). I.e., no glycopeptides were detected in the *glycopeptide-depleted fraction* or the *wash fraction*. Compared to the unenriched IgG sample, the number of identified *N*-glycopeptides could be doubled, further highlighting the efficiency of the enrichment step (number of identified *N*-glycopeptides using glyXtool^{MS}: 25 (unenriched), 51 (spin-cotton-HILIC-SPE enriched)). Analyses of the glycopeptide enrichment efficiency of all other proteins investigated in this study also resulted in an efficient and lossless enrichment of *N*-and *O*-glycopeptides (data not shown).

5.2.3. Glycoproteomic Analysis of Immunoglobulin Gamma, Fibrinogen, Lactotransferrin, and Ribonuclease B

In the following, results of the glycoproteomic analyses of the four representative proteins, IgG, LTF, Fib, and RNase B, are shown in detail. Using the FASP approach, tryptic as well as proteinase K-generated (glyco)peptides of these proteins were treated by spin-cotton-HILIC-SPE, and the resulting HILIC fractions were subjected to nano-RP-LC-ESI-OT-OT-MS/MS analysis. Thereby the FASP approach combines the benefits of an in-gel digest – purification of the target protein(s) by removal of the background matrix – with those of an in-solution digest – high amount of protein sample and ease of use. Each sample was measured using both fragmentation regimes: *HCD.low* and *HCD.step*. The acquired fragment ion spectra were then analyzed using glyXtool^{MS}. The presence of abundant oxonium ion signals (Supplemental Figure S(II)2, top) together with glycan-derived neutral loss fragment ions allowed for a reliable discrimination of glycosylated and non-glycosylated fragment ion spectra by glyXtool^{MS}.

For each glycopeptide fragment ion spectrum, annotations and identifications suggested by the software were validated manually (occasionally, besides glyXtool^{MS}, also the software tools MS-Product and FindPept were used; http://prospector.ucsf.edu/prospector/ and https://-web.expasy.org/findpept/, respectively). In addition, the extracted ion chromatograms (EIC) as well as the isotopic patterns of the precursor ion were checked for each glycopeptide. The following twelve criteria were applied for accepting a glycopeptide identification: (I) EIC: no overlapping or split peak, (II) isotopic pattern: correctly assigned and non-overlapping, (III) presence of ions corresponding to the peptide mass and/or peptide +HexNAc, (IV) presence of the conserved *N*-glycopeptide fragmentation pattern (applies only to *N*-glycopeptides; fragmentation pattern needs to be present with *HCD.step*), (V) unambiguous identification of the glycan moiety, based on Y-ions, (VI) glycan features deduced from Y-ions, such as sialylation, need to agree with detected oxonium ions, (VII) unambiguous identification of the peptide moiety, based on a-, b-, and y-ions (b- and y-ions also with loss of water or ammonia), (VIII) presence of immonium ions to support peptide identification, (IX) accepted mass error: precursor ion (±10 ppm), fragment ions (±20 ppm), (X) check retention time for plausibility: retention times of different glycoforms attached to the very same peptide should not differ too much (<10 min) - sialylated glycopeptides have significantly longer retention times compared to nonsialylated glycopeptides, (XI) different charge states of the same glycopeptide should have nearly the same retention time, and (XII) if present, identifications, derived from both *HCD.low* and *HCD.step*, should be in line with each other. In rare cases, peptide identifications suggested by the software, were inconclusive. In this case manual *de novo* sequencing was performed to account for potential amino acid modifications or protein sequences not considered during the initial protein identification search. During every glycopeptide validation step, special attention was paid to ambiguous and conflicting masses (Supplemental Figure S(II)2, bottom). One prominent example in this context, is the mass of a carbamido-methylation (e.g. on cysteine), which equals the mass of glycine [315]. Awareness of this

ambiguity is particularly relevant for the analysis of (glyco)peptides generated by broad- or non-specific proteolysis. Examples for ambiguous identifications of IgG *N*-glycopeptides are given in Supplemental Figures S(II)3-1 – 2.

Human Immunoglobulin Gamma *N*-Glycoproteomics (Fc-glycosylation)

Human IgG comprises four subclasses (IgG 1, 2, 3, and 4). Our analysis allowed to identify 87 *N*-glycopeptides derived from the constant CH2 region of the heavy chain of IgG (Fc part, fragment crystallizable part; *N*-glycosylation sites: IgG 1 | N_{180}, IgG 2 | N_{176}, IgG 3 | N_{227}, N_{322}, IgG 4 | N_{177}) (Supplemental Figure S(II)1, subsections S(II)1-3 – 26). In agreement with literature, the majority of identified *N*-glycopeptides (84) is derived from IgG 1 and 2 – in human adult serum both have a relative abundance of 60% and 32%, respectively (total concentration of IgG about 10 g/L) [316]. A profiling of these IgG 1 and 2 *N*-glycopeptides, revealed predominantly core-fucosylated diantennary complex-type *N*-glycans with different degrees of galactosylation, with and without a terminal NeuAc and/or a bisecting-GlcNAc (Figure 38).

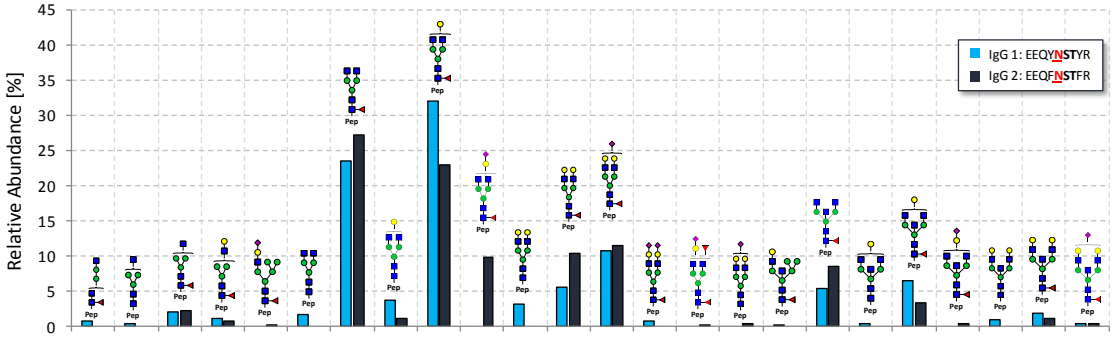


Figure 38: Relative abundance of human IgG 1 and 2 Fc glycoforms (microheterogeneity): relative abundances were derived from the peak intensities of precursor ion extracted ion chromatograms of tryptic IgG 1 and 2 *N*-glycopeptides (IgG 1: $_{176}$(R)EEQY**N**STYR(V)$_{184}$, IgG 2: $_{172}$(R)EEQF**N**STFR(V)$_{180}$). Peptide modifications as well as different precursor charge states were considered.

The three main glycoforms, for both IgG 1 and IgG 2, were $HexNAc_4Hex_3Fuc_1$, $HexNAc_4Hex_4Fuc_1$, and $HexNAc_4Hex_5Fuc_1NeuAc_1$ - also known as G0F, G1F, G2FS (IgG 1|2: 23|27%; 32|23%; 11|11%). This is in good agreement with previously reported data [50]. For IgG 4 only three *N*-glycopeptides corresponding to two glycoforms (G0F, G1F; 72|28%) were identified (relative abundance of IgG 4 in human adult serum about 4% [316]). For IgG 3, no *N*-glycopeptides were identified. In total, glycan compositions corresponding to 24 different *N*-glycoforms were registered for IgG 1, 2, and 4 (Supplemental Table 2 in Hoffmann *et al.*, 2018 [317]). Since the tryptic digest already generated *N*-glycopeptides that could be readily analyzed by LC-MS/MS, meaning the generated peptide moieties were neither too short nor too long, the proteinase K digest was not considered for the glycoproteomic analysis here.

Human Fibrinogen *N*- and *O*-Glycoproteomics

Human fibrinogen is a hexamer comprising two identical heterotrimers connected by a central nodule. Each heterotrimer comprises an α-, β-, and γ-chain. The protein is known to be *N*-

and *O*-glycosylated [196, 225, 274, 318]. The α-chain is reported to solely carry mono- and disialylated mucin-type-1 *O*-glycans (T-antigen), while the γ-chain is reported to carry solely diantennary non-, mono- and disialylated complex-type *N*-glycans [274]. The β-chain, in contrast, is known to carry both types of glycosylation [274]. Fibrinogen features several heavily *O*-glycosylated regions that cannot be analyzed by trypsin alone due to the lack of sufficient tryptic cleavages sites. Previous studies have shown, that the broad, yet reproducible, cleavage specificity of proteinase K yields complementary results to trypsin - making heavily *O*-glycosylated regions analytically amenable, too [225, 274]. In the present study, both the tryptic and the proteinase K digest allowed a reliable identification and characterization of *N*-glycopeptides covering the already known *N*-glycosylation sites present on the β-, and γ-chain (β-chain: N_{394}, γ-chain: N_{78}). Additionally, some lower-abundant *N*-glycopeptides derived from the *N*-glycosylation site N686 present on the α-chain were detected. This *N*-glycosylation site has been reported to be occupied before [319], but, to the best of our knowledge, the attached *N*-glycans have not been described so far (Supplemental Figure S(II)1, subsections S(II)1-27 – 31; Supplemental Figure S(II)4-1 – 4-2; Supplemental Table 3 in Hoffmann *et al.*, 2018 [317]). For the tryptic digest, a total of 38 *N*-glycopeptides were identified. Analysis of these *N*-glycopeptides revealed for the α- and γ-chain exclusively diantennary mono- and disialylated complex-type *N*-glycans without a core-fucose ($HexNAc_4Hex_5Fuc_1NeuAc_{1/2}$ (G2FS1/2)). The β-chain, in contrast, also features non-sialylated, core-fucosylated, and glycoforms with a bisecting GlcNAc - though at much lower relative abundance. Using proteinase K, a total of 15 *N*-glycopeptides were registered. In contrast to the tryptic digest, only the two major glycoforms $HexNAc_4Hex_4Fuc_1NeuAc_{1/2}$ (G1FS1/2) were detected (α-chain: only monosialylated; γ-chain: only disialylated; β-chain: both). It is worth mentioning that also some *N*-glycopeptides of human alpha-1-antitrypsin were detected in the fibrinogen sample, pointing to a potential contamination of this supposedly pure protein sample. The tryptic digest allowed identification of 20 fibrinogen *O*-glycopeptides. Slightly more, 26 *O*-glycopeptides, were identified using proteinase K (Supplemental Figure S(II)1, subsections S(II)1-32 – 60; Supplemental Figure S(II)4-1 – 2; Supplemental Table 3 in Hoffmann *et al.*, 2018 [317]). In both cases, non-, mono- and disialylated mucin-type core-1 *O*-glycans (T-antigen) were detected for these glycopeptides. The fibrinogen alpha chain appears to be heavily *O*-glycosylated: using trypsin, 18 *O*-glycopeptides were detected - covering six different *O*-glycosylation regions including one newly discovered region (S_{609}, S_{616}, S_{618}, S_{619}). The proteinase K digest allowed to identify 18 α-chain *O*-glycopeptides, and yielded, for the most part, complementary results with respect to the covered *O*-glycosylation sites/regions. In total eight *O*-glycosylation sites/regions were detected, of which three regions (I: S_{325}, S_{326}; II: S_{356}, S_{357}, S_{359}; III: T_{499}, T_{501}) and three sites (I: T_{499}, II: T_{522}; III: S_{534}) have not been reported so far. For the β-chain, both trypsin- and proteinase K-generated *O*-glycopeptides covering the already known *O*-glycosylation region (S_{58}, S_{67}). In agreement with previous reports, no *O*-glycopeptides were detected on the fibrinogen γ-chain.

As a side note, two contaminant *O*-glycopeptides derived from von Willebrand factor (disialylated mucin-type core-1 *O*-glycan, S_{1263}) were also detected in the fibrinogen sample. Those two *O*-glycopeptides cover an *O*-glycosylation site that has just recently been reported by Solecka *et al.* [320].

Human Lactotransferrin *N*-Glycoproteomics

Human lactotransferrin is a globular glycoprotein that is reported to be solely *N*-glycosylated. The protein harbors three potential *N*-glycosylation sites (N_{156}, N_{497}, N_{642}), of which only the first and the second are reported to be glycosylated [321, 322]. In this study, a total of 73 tryptic lactotransferrin *N*-glycopeptides were detected – for the first time, covering all three potential *N*-glycosylation sites (Supplemental Figure S(II)1, subsections S(II)1-61 – 90; Supplemental Table 4 in Hoffmann *et al.*, 2018 [317]). The first *N*-glycosylation site, N_{156}, features mono- and disialylated core-fucosylated complex- and hybrid-type *N*-glycans that can also have additional antenna fucoses and/or LacNAc extensions (LacNAc = $GlcNAc_1Gal_1$). The two major glycoforms are $HexNAc_4Hex_5Fuc_1NeuAc_{1/2}$ (G2FS1/2). Discrimination between a LacNAc extended antenna and the presence of an additional antenna is based on a diagnostic oxonium ion that corresponds to two LacNac units plus a fucose (LacNac extended antenna plus antenna fucose, $HexNAc_2Hex_2Fuc_1$). Detected by *HCD.low*, this oxonium ion was only present in lactotransferrin *N*-glycopeptides presumed to feature a LacNAc extension with an antenna fucose. Moreover, this oxonium ion is absent in *HCD.low N*-glycopeptide fragment ion spectra of IgG, Fib, and RNase B – all of which are reported to not feature any LacNAc extensions with an antenna fucose (the $HexNAc_2Hex_2Fuc_1$ fragment ion is also not present in fragment ion spectra of multi-antennary glycoproteins, like alpha-1-acid glycoprotein 1). Unfortunately, LacNAc extensions that lack an antenna fucose could not be linked to a diagnostic oxonium ion: a potential candidate would have been an oxonium ion with two LacNAc units ($HexNAc_2Hex_2$); this oxonium ion, however, is also present in *N*-glycopeptide fragment ion spectra featuring diantennary *N*-glycans that lack a LacNAc extension. The second *N*-glycosylation site, N_{497}, features non-, mono-, and disialylated complex-type *N*-glycans,that are mainly core-fucosylated, but also non-fucosylated. The two major glycoforms are $HexNAc_4Hex_5Fuc_1NeuAc_{1/2}$ (G2FS1/2). As for site N_{156}, glycoforms with additional antenna fucoses were also detected. Interestingly, though, no LacNAc extended glycoforms were registered. Surprisingly, we also detected *N*-glycans on the third *N*-glycosylation site, N_{642} (not described before). This site features mono- and disialylated core-fucosylated complex-type *N*-glycans, that can also have additional antenna fucoses and/or LacNAc extensions. All identified glycoforms exhibited similar abundance; hence, there is no single predominant glycoform on this *N*-glycosylation site. For LTF, no *O*-glycopeptides were detected, which is in agreement with common knowledge.

Bovine Ribonuclease B *N*-Glycoproteomics

RNase B is a globular glycoprotein that is solely *N*-glycosylated. It harbors only one *N*-glycosylation site, which is occupied by high-mannose-type *N*-glycans [323, 324]. In this study, a

total of six tryptic *N*-glycopeptides of RNase B was identified (Supplemental Figure S(II)1, subsections S(II)1-91 – 97; Supplemental Table 5 in Hoffmann *et al.*, 2018 [317]). Three different high-mannose-type glycoforms were allocated to these *N*-glycopeptides ($HexNAc_2Hex_{5,6,8}$ – also known as Man5, Man6, Man8). Detection of these glycoforms is in good agreement with literature reports, as these are the major ones present one RNase B [323, 324] (Figure 39). The two lower-abundant glycoforms, Man7 and Man9, were not detected in the present work.

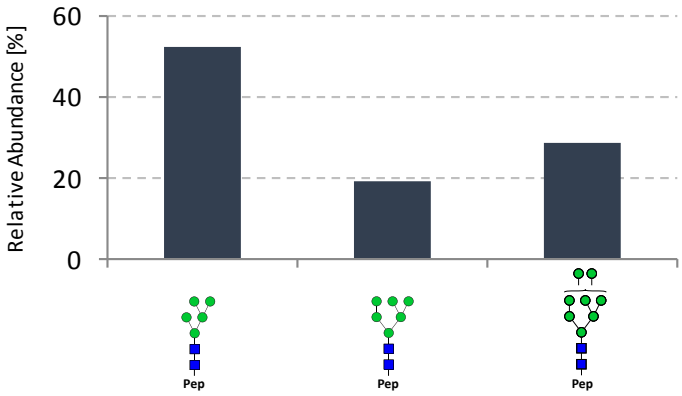


Figure 39: Relative abundance of bovine Rnase B *N*-glycoforms (microheterogeneity): relative abundances were derived from the peak intensities of precursor ion extracted ion chromatograms of tryptic Rnase B *N*-glycopeptides ($_{59}$(R)**N**LTK(D)$_{63}$). Peptide modifications as well as different precursor charge states were considered.

5.2.4. Quantification of Oxonium Ions and Its Potential for N- and O-Glycoproteomics

For the analysis of *N*- and *O*-glycopeptides, oxonium ions represent important diagnostic markers. Derived from the glycan moiety, these singly-charged low-molecular weight fragment ions (B-ions) can not only be used to discriminate a glycopeptide spectrum from a non-glycopeptide spectrum, they can also provide crucial information on the glycan moiety that is linked to the peptide backbone (Supplemental Figure S(II)2-1, top). The presence, absence, and relative abundance of oxonium ions can serve as a fingerprint characteristic for a certain *N*- or *O*-glycoform present on any peptide. To evaluate the diagnostic characteristics of oxonium ions, glycopeptide fragment ion spectra of a set of *N*- and/or *O*-glycosylated proteins comprising IgG, LTF, Fib, and RNase B, were analyzed. For these four glycoproteins the relative abundance of the detected oxonium ions were investigated with respect to the type of glycosylation, the glycoform and certain glycan features, the peptide moiety, the precursor ion charge state, and the impact of the collisional energy that was applied for the fragmentation of the glycopeptides.

Complex-Type N-glycopeptides: General Observations

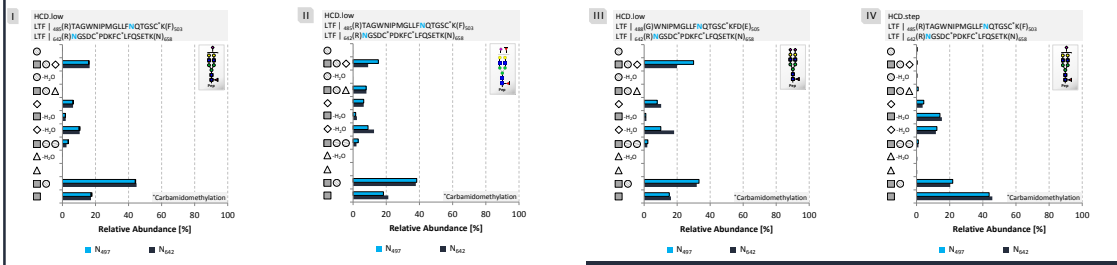
In the following, characteristics of oxonium ions produced by fragmentation of complex-type *N*-glycopeptides via *HCD.low* and *HCD.step* will be described. The first noticeable characteristic is that relative oxonium ion abundance do not differ between different glycoproteins or *N*-

glycosylation sites when considering the same *N*-glycoform and the same collisional energy, as evidenced by comparison of IgG 1 and 2 as well as LTF and IgG 2 (Figure 40 Parts A and B). The produced oxonium ion patterns thus seem to be conserved for specific complex-type *N*-glycoforms independent of the peptide backbone. In most cases, the charge state of the *N*-glycopeptide precursor ion also seems to have only a minor influence on the produced oxonium ions. Using *HCD.low*, a slight increase in the relative abundance of HexNAc, HexNAc -H_2O, NeuAc, and NeuAc -H_2O oxonium ions was observed at higher charge states for some glycoforms (Figure 40 Part C, NeuAc/NeuAc -H_2O not shown); with *HCD.step*, the influence of the precursor ion charge state seems negligible. In general, relative abundance of oxonium ions produced by *HCD.low* and *HCD.step* differ significantly from each other when comparing the very same complex-type *N*-glycopeptides (Figure 40 Part D). With increasing collisional energy, the relative abundance of di- and trisaccharide oxonium ions, such as $HexNAc_1Hex_1$, $HexNAc_1Hex_2$ or $HexNAc_1Hex_1NeuAc_1$ decreases – some di- and trisaccharide oxonium ions even disappear due to decomposition into mono- and disaccharide oxonium ions.

Immunoglobulin gamma: complex-type N-glycopeptides
Part A
IgG 1 vs IgG 2
Lactotransferrin vs Immunoglobulin gamma: complex-type N-glycopeptides
Part B
Immunoglobulin gamma: complex-type N-glycopeptides
Part C
Charge state of the precursor
Immunoglobulin gamma: complex-type N-glycopeptides
Part D
HCD.low vs HCD.step
Relative Abundance [%]
IgG 1
IgG 2
LTF
HCD.low
HCD.step

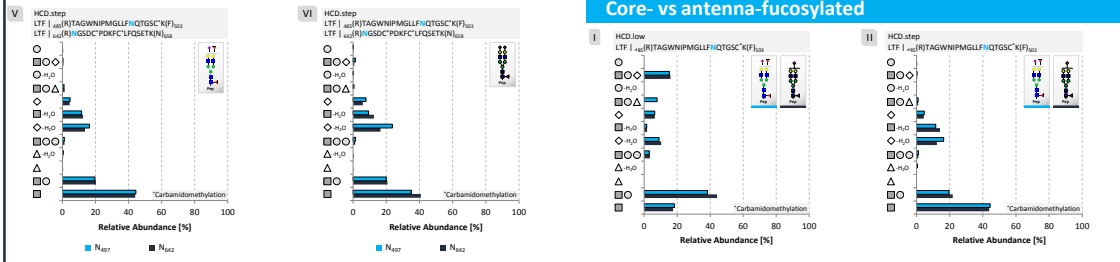
Lactotransferrin: complex-type *N*-glycopeptides — Part E

N-glycosylation sites



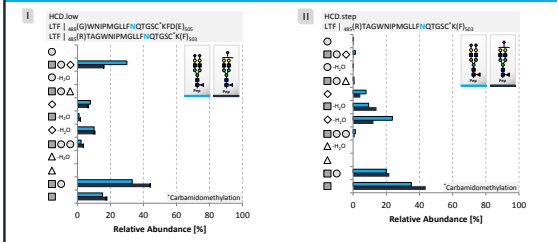
Lactotransferrin: complex-type *N*-glycopeptides — Part F

Core- vs antenna-fucosylated

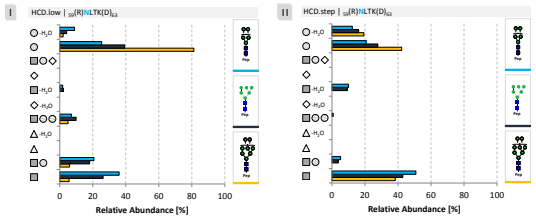


Lactotransferrin: complex-type *N*-glycopeptides — Part G

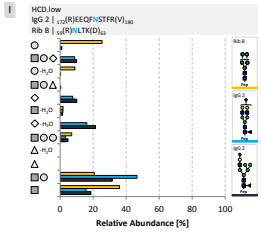
Mono- vs disialylated



Ribonuclease B: high-mannose-type *N*-glycopeptides — Part H

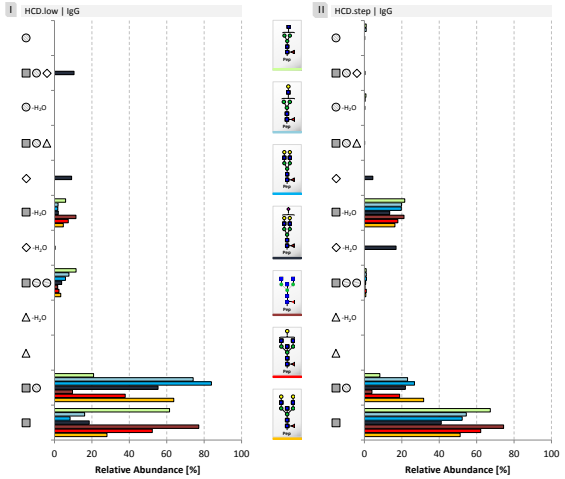


Rib B / IgG: high-mannose-type vs complex-type vs hybrid-type N-glycopeptides — Part I



Immunoglobulin gamma : complex-type *N*-glycopeptides — Part J

Overview



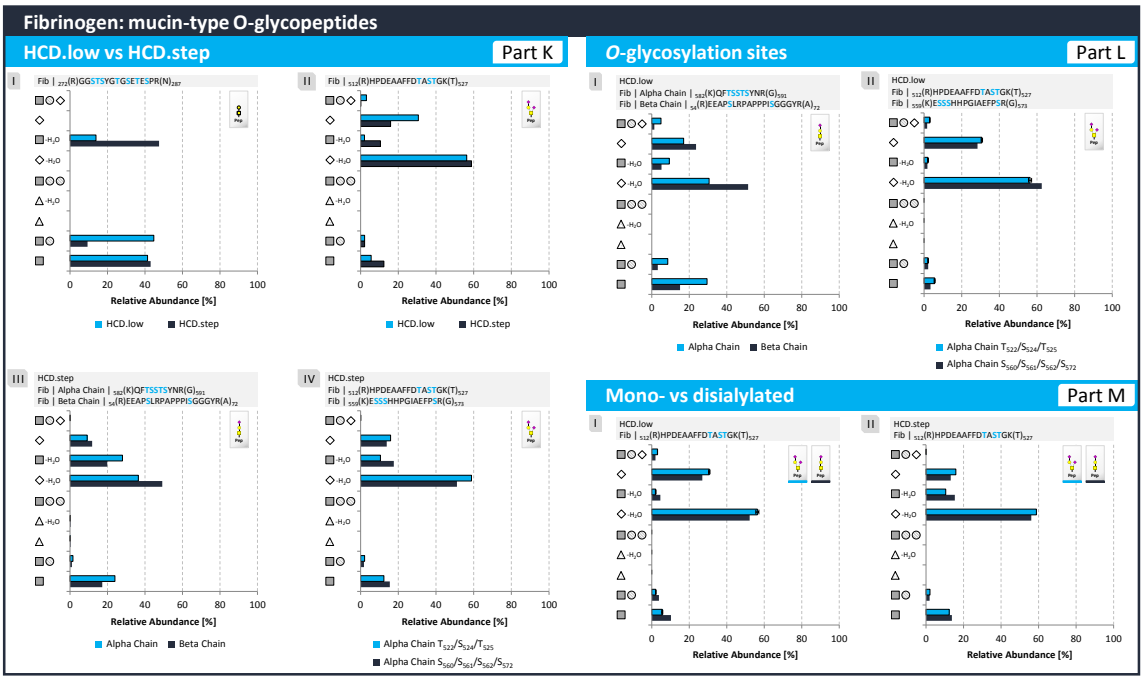


Figure 40: Relative quantitation of *N*- and *O*-glycopeptide oxonium ions acquired by HCD fragmentation using varying collisional energies (*HCD.low*, *HCD.step*): intensities of detected oxonium ions are represented relative to the summed intensity of all detected oxonium ions within a fragment spectrum (relative abundance). For glycopeptides with more than one corresponding fragment ion spectrum (scans) the average relative abundance is given. Error bars indicate the standard deviation of the relative oxonium ion abundances for glycopeptides with more than two scans. Relative oxonium ion abundances are compared with respect to the different collisional energies and between: a) different *N*- and *O*-glycoforms present on the same peptide/glycosylation site, b) the same glycoform present on different peptides/proteins, and c) different precursor ion charge states of the same glycopeptide. For comparison *N*- and *O*-glycopeptides derived from human immunoglobulin gamma 1 and 2 (IgG), human fibrinogen (Fib), human lactotransferrin (LTF), and bovine ribonuclease B (RNase B) were considered.

The comparison of the relative oxonium ion abundance detected for seven IgG 1 and 2 *N*-glycopeptides exemplifies this effect (Figure 40 Part J (I-II)): while the relative abundance of the $HexNAc_1Hex_1NeuAc_1$ oxonium ion is about 10% with *HCD.low*, the relative abundance drops to about 1% with *HCD.step*. The same holds true for the $HexNAc_1Hex_2$ oxonium ion (average relative abundance: *HCD.low* 6%, *HCD.step* 0.9%). A second example is given in Figure 40 Part E (II, V): while *HCD.low* fragmentation of the lactotransferrin *N*-glycopeptide $HexNAc_4Hex_5Fuc_2NeuAc_1$ (G2F2S1) produces a strong $HexNAc_1Hex_1Fuc_1$ oxonium ion signal of about 8%, the relative abundance with *HCD.step* is only about 1%. Another striking characteristic that differentiates *HCD.step* from *HCD.low* is that *HCD.step* fragmentation of complex-type *N*-glycopeptides always results in the HexNAc oxonium ion peak being the dominant peak among the oxonium ions – independent of the present complex-type *N*-glycoform (relative abundance always >38%, Figure 40 Parts A (VI-XIII), D, E (IV-VI), J (II)). In addition, at higher collisional energy, the relative abundance of corresponding water-loss species is also increased compared to lower energy fragmentation (e.g. average relative abundance for HexNAc -H_2O: *HCD.low* 6%, *HCD.step* 19%, Figure 40 Part J (I-II)). At lower collisional energy, in contrast, either the HexNAc or $HexNAc_1Hex_1$ oxonium ion dominates, depending on the present glycoforms (Figure 40 Part J (I)). The $HexNAc_1Hex_1$ oxonium ion

dominates whenever at least one galactose is present in the *N*-glycopeptide glycoform (Figure 40 Part D (II-IV)). In case there is no galactose the HexNAc oxonium ion dominates (Figure 40 Parts A (III), D (I)). An exception seems to be the presence of a single galactose as part of a di- or more-antennary *N*-glycan (with/without bisecting GlcNAc): in this case either the $HexNAc_1Hex_1$ or the HexNAc oxonium ion dominates depending on the precursor ion charge state (Figure 40 Part A (I, II)). Another exception is the presence of a single GlcNAc attached to the trimannosyl-chitobiose core without an additional galactose and without being a bisecting GlcNAc (single antenna GlcNAc without additional galactose): in this case, the $HexNAc_1Hex_1$ oxonium ion dominates (Figure 40 Part D(I)).

Complex-Type N-Glycopeptides: Antenna GlcNAc vs Bisecting GlcNAc

Comparing the relative oxonium ion abundance of individual fragment ion scans acquired for a particular *N*-glycopeptide with each other, for both *HCD.low* and *HCD.step*, only marginal differences can be detected in most cases, as indicted by the standard deviation (<0.5; Figure 40 Parts A, D). Supplemental Figure S(II)5-1 (Part A) exemplarily shows the individual *HCD.low* fragment ion scans acquired for the IgG 1 *N*-glycopeptide $HexNAc_5Hex_3Fuc_1$ (G0FN) over time. This *N*-glycopeptide features a non-galactosylated *N*-glycan with a bisecting GlcNAc. In agreement with our general observations, the low-energy fragmentation of a non-galactosylated complex-type *N*-glycopeptide results in the HexNAc oxonium ion being the dominant peak among the oxonium ions. For the IgG 1 *N*-glycopeptide $HexNAc_5Hex_3Fuc_1$ this was consistently observed across all acquired scans. The presence of a bisecting GlcNAc gives rise to two characteristic fragment ions when using low-energy fragmentation: (I) peptide+$HexNAc_3Hex_1$ (Y-ion, $[M+H]^+$, $[M+2H]^{2+}$) and (II) $HexNAc_2$ (oxonium ion, B-ion, $[M+H]^+$). These fragment ions were consistently found in all *HCD.low* scans of the IgG 1 *N*-glycopeptide $HexNAc_5Hex_3Fuc_1$ (Supplemental Figure Figure S(II)5-1 Part A). With *HCD.step* these fragment ions are not present or only at very low intensity (not shown).

The same fragmentation behavior was also observed for all acquired *HCD.low* scans for the IgG 2 *N*-glycopeptide $HexNAc_5Hex_3Fuc_1$ (G0FN) (Supplemental Figure Figure S(II)5-1 Part B) and for the IgG 1 *N*-glycopeptide $HexNAc_3Hex_2Fuc_1$ (Supplemental Figure Figure S(II)5-1 Part C). Again, the relative abundance of the oxonium ions and the presence of the peptide+$HexNAc_3Hex_1$ and the $HexNAc_2$ fragment ions suggest a non-galactosylated *N*-glycan with a bisecting GlcNAc being attached to the peptide. The two diagnostic fragment ions (I) peptide+$HexNAc_3Hex_1$ and (II) $HexNAc_2$ were absent in *N*-glycopeptides derived from glycoproteins largely lacking *N*-glycans with a bisecting GlcNAc, such as Fib and LTF (both have almost exclusively complex-type *N*-glycans without bisecting GlcNAc), or RNase B (only high-mannose-type *N*-glycans). Interestingly, *HCD.low* fragmentation of the IgG 1 *N*-glycopeptide $HexNAc_3Hex_3Fuc_1$ (Supplemental Figure Figure S(II)5-2 Part D) showed an inconsistent oxonium ion pattern across the acquired scans (which also explains the high standard deviation in Supplemental Figure S(II)5-2 Part D(I)). In the first four scans the $HexNAc_1Hex_1$ oxonium ion dominates, which along with the absence of a

peptide+$HexNAc_3Hex_1$ fragment ion suggests a non-galactosylated *N*-glycan with an antenna GlcNAc attached to the peptide (Supplemental Figure S(II)5-2 Part D, scan #1-4). In scans #5 and #6, however, the $HexNAc_1$ oxonium ion dominates, which along with the presence of the peptide+$HexNAc_3Hex_1$ and the $HexNAc_2$ fragment ion in scan #6 suggests a non-galactosylated *N*-glycan with a bisecting GlcNAc attached (Figure 41, Supplemental Figure S(II)5-2 Part D, scan #5-6). The same fragmentation behavior could also be observed for the IgG 2 *N*-glycopeptide $HexNAc_3Hex_3Fuc_1$ (Supplemental S(II)5-2 Part F). Again, the earlier eluting non-galactosylated *N*-glycopeptide with an antenna GlcNAc shows a different oxonium ion pattern (Supplemental Figure S(II)5-2 Part F, scan #1) and no peptide+$HexNAc_3Hex_1$ fragment ion, compared to the later eluting non-galactosylated *N*-glycopeptide with a bisecting GlcNAc (Supplemental Figure S(II)5-2 Part F, scan #2-3). Thus, this finding might enable discrimination between isobaric *N*-glycopeptides featuring either an antenna GlcNAc or a bisecting GlcNAc, based on differences in the retention time, characteristic *HCD.low* oxonium ion pattern, and diagnostic fragment ions. With *HCD.step* the observed change in the oxonium ion pattern between antenna GlcNAc and bisecting GlcNAc could not be observed (Supplemental Figure S(II)5-2 Parts E and G).

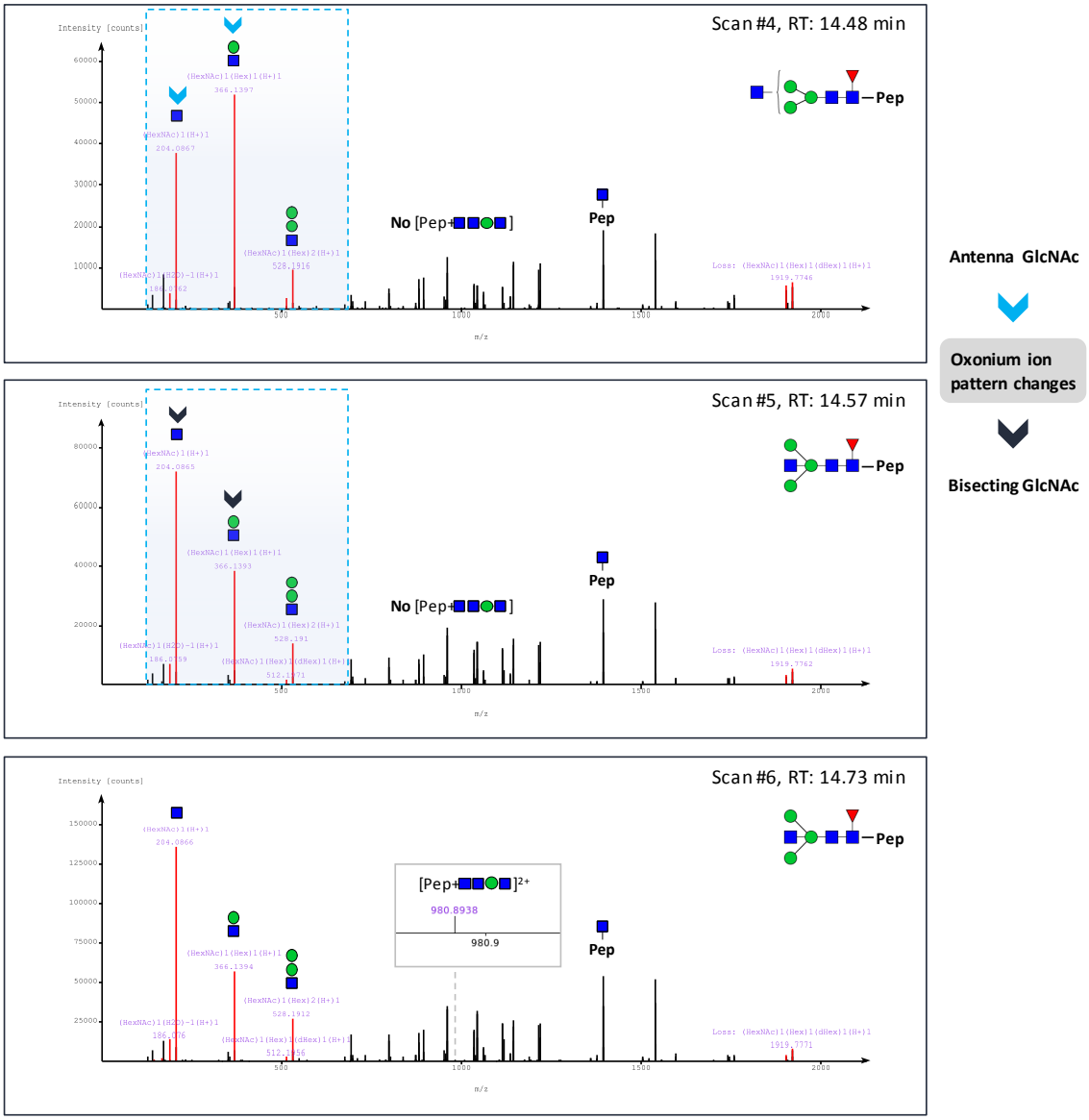


Figure 41: Antenna GlcNAc vs bisecting GlcNAc: Change of the *N*-glycopeptide oxonium ion pattern between associated fragment ion scans allows discrimination of antenna GlcNAc and bisecting GlcNAc *N*-glycoforms. The example shows consecutive *HCD.low* fragment ion scans derived from the IgG1 *N*-glycopeptide $_{76}$EEQY**N**STYR$_{184}$ + $HexNAc_3Hex_3Fuc_1$.

Complex-Type N-Glycopeptides: Core- vs Antenna-fucosylation

The reliable detection and discrimination of core- and antenna-fucosylation is another important objective in many *N*-glycoproteomic studies. In this work, the presence of a core-fucose did not produce any unique and characteristic oxonium ion, neither for *HCD.low* nor for *HCD.step* (Figure 40 Part J (I-II)). However, in many cases a $HexNAc_1Hex_1Fuc_1$ oxonium ion was detected for both core- and antenna-fucosylated *N*-glycopeptides. For core-fucosylated *N*-glycopeptides this oxonium ion is present only at very low relative abundance (<1%) – independent of the applied collisional energy – and seems to be an artifact generated by fucose rearrangement [325]. The presence of an antenna-fucose, instead, produced a much higher signal of the $HexNAc_1Hex_1Fuc_1$ oxonium ion – but only at lower collisional energy (*HCD.low*: 10%; *HCD.step*: 2%) (Figure 40 Part F (I-II)). Thus, at lower collisional energy discrimination

between core- and antenna-fucosylation based on the relative abundance of the $HexNAc_1Hex_1Fuc_1$ oxonium ion seems to be possible. Interestingly, for *N*-glycopeptides presumed to carry an antenna-fucose, we also detected a $HexNAc_1Hex_1Fuc_1NeuAc_1$ oxonium ion when using *HCD.low*, thus giving further evidence for this annotation. This finding was very recently confirmed in a publication by Acs *et al.* [326].

Complex-type N-Glycopeptides: Type of Sialic acid, Degree of Sialylation

NeuAc and NeuGc are the two most commonly found types of sialic acids in vertebrates [327]. High- as well as low-energy fragmentation of NeuAc or NeuGc containing *N*-glycopeptides produces intense and distinct oxonium ions allowing a clear distinction of these two types of sialic acid: NeuAc oxonium ions (NeuAc; NeuAc -H_2O; $HexNAc_1Hex_1NeuAc_1$), NeuGc oxonium ions (NeuGc; NeuGc -H_2O; $HexNAc_1Hex_1NeuGc_1$) (NeuGc not shown in this study). To assess the degree of NeuAc sialylation the relative abundance of oxonium ions derived from mono- and disialylated glycoforms of LTF glycopeptides were compared (Figure 40 Part G). With *HCD.low* relative abundance of the NeuAc and NeuAc -H_2O oxonium ions did not differ significantly between mono- and disialylated LTF *N*-glycopeptides. The relative abundance of the $HexNAc_1Hex_1NeuAc_1$ oxonium ion, though, was found to be higher with disialylated LTF *N*-glycopeptides in most, yet not all, cases (relative abundance $HexNAc_1Hex_1NeuAc_1$: monosialylated, between 10-18%; disialylated, between 22-28%). With *HCD.step* relative abundance of the NeuAc and NeuAc -H_2O oxonium ions were slightly higher with disialylated LTF *N*-glycopeptides (NeuAc: monosialylated, about 5%; disialylated, about 10%. NeuAc -H_2O: monosialylated, about 12%; disialylated, about 20%). The relative abundance of $HexNAc_1Hex_1NeuAc_1$ oxonium ion, in contrast, was very low (<2%), and did not differ significantly between mono- and disialylated LTF *N*-glycopeptides when applying higher collisional energy ($HexNAc_1Hex_1NeuAc_1$: monosialylated, about 1%; disialylated, about 2%). Overall, predicting the degree of NeuAc sialylation based on the relative abundance of NeuAc, NeuAc -H_2O, and $HexNAc_1Hex_1NeuAc_1$ oxonium ions seems promising, but needs to be further investigated, as for instance the influence of different sialic acid linkages needs to be evaluated, and might have caused the observed inconsistencies. Also of note, and in agreement with reports by Halim *et al.* [328], along with NeuAc and NeuAc -H_2O six lower abundant oxonium ions corresponding to $[NeuAc -2xH_2O]^+$ *m/z* 256.0821, $[NeuAc -3xH_2O]^+$ *m/z* 238.0715, $[NeuAc -2\times H_2O -NH_2C(O)CH_3]^+$ *m/z* 197.043, $[NeuAc -2\times H_2O -COH -NH_2C(O)CH_3]^+$ *m/z* 167.0375, $[NeuAc -4xH_2O]^+$ *m/z* 220.061, and $[NeuAc -2\times H_2O -CHO -NH_2C(O)CH_3 -COOH]^+$ *m/z* 121.032 were consistently detected; the last two were only detected with *HCD.step*, though.

High-mannose-type N-glycopeptides

The relative oxonium ion abundance detected for high-mannose-type *N*-glycopeptides are remarkably different from those previously described for complex-type *N*-glycopeptides. *HCD.low* and *HCD.step* generated fragment ion spectra of complex-type *N*-glycopeptide are usually dominated by HexNAc and/or $HexNAc_1Hex_1$ oxonium ions as well as

$HexNAc_1Hex_1NeuAc_1$ oxonium ions in some cases. In contrast, high-mannose-type *N*-glycopeptide fragment ion spectra, by nature feature only Hex- or HexNAc-related oxonium ions (HexNAc, HexNAc -H_2O, Hex, Hex -H_2O, $HexNAc_1Hex_1$, $HexNAc_1Hex_2$). The most striking feature of high-mannose-type *N*-glycopeptides, however, is the high relative abundance of Hex and Hex -H_2O oxonium ions, which directly reflects the high content of mannoses in the *N*-glycan structures.

While for complex-type *N*-glycopeptides, if present at all, the relative abundance of Hex and Hex -H_2O oxonium ions did not exceed 0.5%, when low collisional energy was applied (Figure 40 Part J (I-II)) for high-mannose-type *N*-glycopeptides, the relative abundance of the Hex oxonium ion, for instance, ranged from 25-80% when using *HCD.low* fragmentation, as exemplified for RNase B Man-5/6/8 *N*-glycopeptides (Figure 40 Part H (I)). A similar effect was also observed for *HCD.step*: here the relative abundance of the Hex oxonium ion ranged from 20-42% for the RNase B Man-5/6/8 *N*-glycopeptides (for complex-type *N*-glycopeptides the relative abundance of Hex and Hex -H_2O oxonium ions was normally below 1%). Compared to *HCD.low*, the increased collisional energy applied with *HCD.step* resulted in an increase of the Hex -H_2O oxonium ion abundance (*HCD.low*: 2-10%; *HCD.step*: 12-20%. Figure 40 Part H (I-II)). Overall, these findings suggest that both fragmentation regimes enable the discrimination of high-mannose-type *N*-glycopeptides from complex-type *N*-glycopeptides based on the relative abundance of Hex and Hex -H_2O oxonium ions. Moreover, *HCD.low* results indicate a direct correlation between the degree of mannosylation and the relative abundance of the Hex oxonium ion.

Hybrid-Type N-Glycopeptides

Hybrid-type *N*-glycopeptides combine features of complex- and high-mannose-type *N*-glycopeptides alike. Comparing the relative oxonium ion abundance detected for an *HCD.low* measurement of an IgG 2 complex-type *N*-glycopeptide ($HexNAc_4Hex_5Fuc_1NeuAc_1$) with those detected for an IgG 2 hybrid-type *N*-glycopeptide ($HexNAc_3Hex_6Fuc_1NeuAc_1$), major differences relate to the relative abundance of the $HexNAc_1Hex_1$ oxonium ion, which reflects the two fully galactosylated antenna of the complex-type *N*-glycan (Figure 40 Part I). A minor, though significant, difference relates to the relative abundance of the Hex oxonium ion: for complex-type *N*-glycopeptides this oxonium is either not present or only present at a relative abundance of less than 0.5% at *HCD.low*; for hybrid-type *N*-glycopeptides the relative abundance of this oxonium ion was consistently found to be higher than 1.1%. This finding is in agreement with the observation previously made for high-mannose-type *N*-glycopeptides; also there an increase of the relative abundance of the Hex_1 oxonium ion was observed, correlating with the degree of mannosylation (Figure 40 Part I). Overall, the relative abundance of the Hex oxonium ion and the presence of non-high-mannose-type oxonium ions, such as $HexNAc_1Hex_1NeuAc_1$, may allow discriminating hybrid-type *N*-glycopeptides from complex-type or high-mannose-type *N*-glycopeptides. Another interesting finding in this context is the presence of a Hex_2 oxonium ion exclusively in *N*-glycopeptides presumed to be

of the hybrid-type, high-mannose-type, or not fully galactosylated complex-type (Hex_2 oxonium ion exclusively found in IgG and RNase B *N*-glycopeptide fragment ion spectra; not present in Fib and LTF) (Supplemental Figure S(II)6-1). Produced by *HCD.low* fragmentation, this oxonium ion can give additional confidence in the *N*-glycopeptide annotation.

Mucin-Type O-Glycopeptides

Apart from various *N*-glycopeptide also *O*-glycopeptide fragment ion spectra have been investigated with respect to relative oxonium ion abundance. To this end, core-1 mucin-type *O*-glycopeptides derived from human fibrinogen (Fib) were fragmented by *HCD.low* or *HCD.step* and the resulting oxonium ion abundance were analyzed. As for *N*-glycopeptides, also for core-1 mucin-type *O*-glycopeptides differences in the relative abundance of the oxonium ions depending on the applied collisional energy were observed. Again, the increased collisional energy accompanied by *HCD.step* leads to a decrease or absence of signals corresponding to di- and trisaccharide oxonium ions, as can be seen for the relative abundance of $HexNAc_1Hex_1$ and $HexNAc_1Hex_1NeuAc_1$ oxonium ions (Figure 40 Part K (I, II)). While there were striking differences between *HCD.low* and *HCD.step* for the oxonium ion abundance of non-sialylated *O*-glycopeptides (Figure 40 Part K (I)), differences were less prominent for sialylated *O*-glycopeptides (Figure 40 Part K (II)). The latter showed differences primarily in the relative abundance of NeuAc (NeuAc, NeuAc $-H_2O$, $HexNAc_1Hex_1NeuAc_1$) and HexNAc related oxonium ions (HexNAc, HexNAc $-H_2O$); the overall distribution, with NeuAc $-H_2O$ being the dominant oxonium ion, remained unaffected, though. With non-sialylated *O*-glycopeptides, in contrast, HexNAc and $HexNAc_1Hex_1$ oxonium ions dominated at *HCD.low*, while HexNAc and HexNAc $-H_2O$ oxonium ions dominated at *HCD.step*. Comparing the relative oxonium ion abundance of the very same *O*-glycoform between different Fib *O*-glycosylation sites/regions (α and β chain), differences between mono- and disialylated *O*-glycoforms become apparent (Figure 40 Part L). For disialylated core-1 mucin-type *O*-glycopeptides only slight differences between different *O*-glycosylation sites/regions were observed, independent of the applied collisional energy (Figure 40 Part L (II, IV)). Thus, it seems that the relative oxonium ion abundance of disialylated *O*-glycopeptides are conserved, and independent of the peptide backbone. On the contrary, for monosialylated core-1 mucin-type *O*-glycopeptides, and particularly for *HCD.low*, significant differences between the relative oxonium ion abundance of the fibrinogen α and β chain were detected (Figure 40 Part L (I, III)). This suggests that the relative oxonium ion abundance of monosialylated *O*-glycopeptides are not conserved, and not independent of the peptide backbone. Surprisingly, a direct comparison, between mono- and disialylated *O*-glycoforms present on the very same peptide, revealed no significant differences in the relative oxonium ion abundance – neither for *HCD.low* nor for *HCD.step*. Hence, differentiating mono- and disialylated core-1 mucin-type *O*-glycoforms based on their relative oxonium ion abundance seems to be not possible (Figure 40 Part M (I, II)). However, a general comparison of the relative oxonium ion abundance acquired for core-1 mucin-type *O*-glycopeptides with those acquired for *N*-glycopeptides, revealed significant differences that enable differentiation of these two forms of protein glycosylation. The most striking difference

is the lack of the $HexNAc_1Hex_2$ oxonium ion for core-1 mucin-type *O*-glycopeptides. This oxonium ion was found consistently across all analyzed *N*-glycopeptides, independent of the applied collision energy (Figure 40 Parts A-J). It appears to represent a characteristic oxonium ion that occurs upon fragmentation of (I) the chitobiose core – a main fragmentation event during *N*-glycopeptide fragmentation – and (II), optionally, further antenna-directed fragmentation steps, to ultimately generate a fragment ion corresponding to 2nd chitobiose GlcNAc with two attached mannoses ($HexNAc_1Hex_2$). Additionally, also fragment ions derived from *N*-glycan antennae LacNAc (+mannose) residues can give rise to a HexNAc1Hex2 oxonium ion. Another important observation relates to the HexNAc and HexNAc -H2O oxonium ion ratio. In agreement with Halim *et al.* [328], using HCD.step this ratio allows to discriminate *N*-glycopeptides from *O*-glycopeptides (HexNAc -H2O/ HexNAc ratio: for *N*-glycopeptides 0.1-0.3; for *O*-glycopeptides 0.85-1.2, Figure 42).

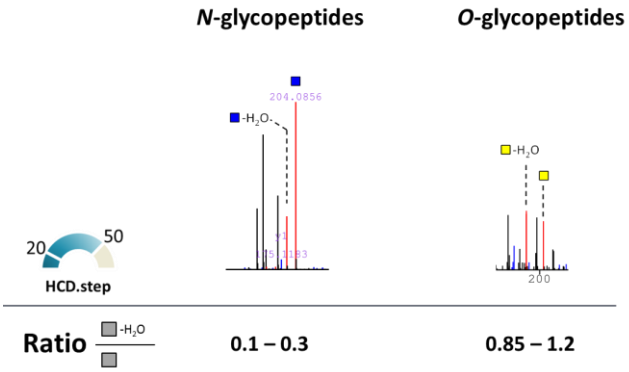


Figure 42: Discrimination of *N*- and *O*-glycopeptides. Using *HCD.step N*- and *O*-glycopeptide fragment ion spectra can be discriminated from each other with the help of characteristic HexNAc-H_2O:HexNAc: oxonium ion ratios.

A summary of all findings related to the relative oxonium ion quantitation is given in Table 3.

Table 3: Diagnostic potential of oxonium ions for *N*- and *O*-glycoproteomics – overview.

		Figure
Complex-type *N*-glycopeptides		
General Observations	▪ Oxonium ion abundance are characteristic for specific *N*-glycan features	Figure 40: A and B
	▪ Oxonium ion abundance of a specific *N*-glycoform are independent of the peptide backbone	Figure 40: A and B
	▪ Oxonium ion abundance of a specific *N*-glycoform are mostly independent of the precursor ion charge state	Figure 40: C
	▪ Oxonium ion abundance depend on applied collisional energy (*HCD.low*, *HCD.step*)	Figure 40: D
	▪ *HCD.step*: decrease/absence of di- and trisaccharide oxonium ions compared to *HCD.low*	Figure 40: E (II, V), J (I-II)
	▪ Differentiation of complex-, hybrid- and high-mannose-type *N*-glycopeptides based on oxonium ion abundance possible (*HCD.low*)	Figure 40: I
	▪ Diagnostic oxonium ion (*HCD.low*): Hex Relative abundance: high-mannose-type >> hybrid-type > complex-type	
Antenna GlcNAc vs Bisecting GlcNAc	▪ Differentiation of antenna GlcNAc and bisecting GlcNAc based on oxonium ion abundance and retention time possible (*HCD.low*)	Figure 41, S(II)5-1 - 2: A-D, and F
	▪ Diagnostic fragment ions for bisecting GlcNAc (*HCD.low*): ▪ Peptide+$HexNAc_3Hex_1$ (Y-ion, $[M+H]^+$, $[M+2H]^{2+}$) ▪ $HexNAc_2$ (oxonium ion, B-ion, $[M+H]^+$)	
	▪ Retention time: bisecting GlcNAc > antenna GlcNAc	
Core- vs Antenna-Fucosylation	▪ Differentiation of core- and antenna-fucosylation based on oxonium ion abundance possible (*HCD.low*)	Figure 40: F (I-II)
	▪ Diagnostic oxonium ion (*HCD.low*): $HexNAc_1Hex_1Fuc_1$ Relative abundance: antenna fucose > core fucose	
Type of Sialic Acid, Degree of Sialylation	▪ Differentiation of the type of sialic acid (NANA, NGNA) based on diagnostic oxonium ions possible (*HCD.low*, *HCD.step*)	S(II)1-98 - 100
	▪ Predicting the degree of *N*-acetylneuraminic acid sialylation based on diagnostic oxonium ion abundance seems promising, though further analyses are necessary (*HCD.low*, *HCD.step*)	Figure 40: G
High-Mannose-Type *N*-Glycopeptides	▪ Direct correlation between the degree of mannosylation and the relative abundance of the diagnostic Hex oxonium ion (*HCD.low*) Relative abundance of Hex oxonium ion: e.g. $Man_1<Man_5<Man_9$	Figure 40: H (I-II)
Hybrid-Type *N*-Glycopeptides	▪ Combination of complex- and high-mannose-type related diagnostic oxonium ions	Figure 40: I
Mucin-Type *O*-Glycopeptides	▪ Differentiation of core-1 mucin-type *O*-glycopeptides and *N*-glycopeptides based on diagnostic oxonium ion abundance possible (*HCD.low*, *HCD.step*)	Figure 40: K (I-II)
	▪ Diagnostic oxonium ion (*HCD.low*, *HCD.step*): $HexNAc_1Hex_2$ $HexNAc_1Hex_2$ absent in *O*-glycopeptide fragment ion spectra	Figure 40: A-J
	▪ HexNAc -H_2O:HexNAc ratio (*HCD.step*): for *N*-glycopeptides: 0.1-0.3; for *O*-glycopeptides: 0.85-1.2	Figure 42
	▪ Differentiation of mono- and disialylated *O*-glycopeptides (NANA) based on diagnostic oxonium ion abundance not possible (*HCD.low*, *HCD.step*)	Figure 40: L and M

5.3. Summary

Over the last years site-specific mass spectrometry-based *N*- and *O*-glycoproteomic analyses have become more and more advanced, more reliable, and more popular. By introducing novel techniques and software solutions, the detailed characterization of single glycoproteins or mixtures of glycoproteins with different types of glycosylation and multiple glycosylation sites has become feasible. However, the analysis of intact glycosylated peptides remains challenging and still has potential for improvements, for instance with respect to quantitation and structural analysis of the glycan moiety. Facing multiple layers of complexity by unambiguously identifying two structurally diverse entities – the glycan moiety and the peptide backbone – dedicated analysis workflows along with a basic understanding of the glycopeptide fragmentation process and *a priori* knowledge about the sample to be analyzed are crucial. Here, we present a universal glycoproteomic workflow that allows the in-depth analysis of *N*- and *O*-glycopeptides. To this end, glycopeptides generated by tryptic or proteinase K digest, were enriched by spin-cotton-HILIC-SPE and subjected to reversed-phase liquid chromatography coupled to tandem mass spectrometry analysis. The use of two HCD fragmentation regimes, *HCD.low* and *HCD.step*, enabled the unambiguous identification of the peptide backbone and allowed the detailed analysis of the glycan moiety. Using a set of four representative *N*- and *O*-glycosylated proteins (IgG, Fib, LTF, and RNase B) the versatility, robustness and comprehensiveness of the developed workflow was demonstrated. In addition, diagnostic oxonium ion patterns that are characteristic for certain glycoforms and glycan features are discussed in detail, and we report on a conserved *N*-glycopeptide fragmentation pattern, which enables the reliable determination of the peptide mass. Our findings are applicable for the analysis of a broad range of glycoproteins, can significantly increase the confidence in manual and software-assisted analysis of *N*- and *O*-glycopeptides, and can thus improve MS-based glycoanalytics.

6 Chapter Six

Conclusion and Outlook

The overarching objective of this thesis was to establish an analytical platform that enables the reliable identification and characterization of glycoproteins. This glycoproteomic analysis platform should extend and complement already existing xCGE-LIF-based *N*-glycomics and the MS-based proteomics platforms. By providing information on the peptide moiety and the glycan moiety alike, the glycoproteomics approach can bridge the gap between the solely glycan-centered glycomics approach and the solely protein-centered proteomics approach. It enables to derive protein- and site-specific glycosylation information in a qualitative and quantitative manner and therefore to address questions related to glycan micro- and macroheterogeneity. Moreover, it supports the identification and characterization of yet unknown glycosylation sites, thus offering the opportunity to apply this approach for explorative glycoproteomic studies. Establishing such a method allows to tackle new scientific questions and to unravel yet unknown biological implications of *N*- and *O*-glycans and the glycoproteins carrying these glycans.

6.1. Human Blood Plasma *O*-Glycoproteomics

With this project we have successfully developed an *O*-glycoproteomic workflow capable of analyzing intact mucin-type core-1 and 2 *O*-glycopeptides derived from human blood plasma glycoproteins. The developed workflow addresses many of the initially stated requirements: it proved to be robust and reliable and able to provide detailed site-specific *O*-glycoproteomic information for glycoproteins present in a complex matrix such as blood plasma (unambiguous

identification of the peptide moiety, compositional description of the intact glycan moiety, localization of the occupied glycosylation sites). The use of proteinase K for the glycoproteomic analysis of such a complex sample sets this study apart from comparable studies where conventionally trypsin is the protease of choice. We could demonstrate that, by using proteinase K, a number of *O*-glycosylation sites could be discovered that have not been detected before by trypsin-based approaches. For *N*- and *O*-glycoproteomic studies we therefore advocate the use of proteolytic enzymes with a broad cleavage specificity, such as proteinase K, as an addition to trypsin. Explorative site-specific *N*- and *O*-glycoproteomic studies of biofluids, like human blood plasma, human milk, urine or cerebrospinal fluid hold an enormous potential to better understand the implications of protein glycosylation under normal physiological conditions, but also under pathophysiological conditions. By serving as a diagnostic tool, the detection/discovery of relevant glycopeptides (biomarker candidates) can be the basis for targeted quantitative glycoproteomic analyses, which allow for a site-specific monitoring of glycosylation alterations, e.g. during disease progression. Site-specific glycosylation analyses are, moreover, important to produce biopharmaceuticals according to quality by design requirements, in particular if these biopharmaceuticals are produced in heterologous expression systems. In this regard site-specific glycosylation analyses might also enable understanding/controlling important glycan-related features of the final product including its efficacy, half-life or antigenicity.

Although the *O*-glycoproteomic workflow described in Chapter four represents a solid foundation to tackle future glycoproteomic studies there are still technical challenges that need to be addressed. In the following a number of options to overcome caveats identified throughout this project are specified.

With regard to the sample preparation, implementation of protein fractionation steps, such as mixed-mode ion exchange chromatography with subsequent fractionation, or immuno-affinity-based enrichment/depletion of specific glycoproteins, may enable to further increase the depth of the glycoproteomic analysis in future studies.

Proteinase K has a broad cleavage specificity. This prevents prediction of the peptide moiety of glycopeptides generated by this protease, which consequently complicates the peptide identification (need to do CID MS^3 to confirm the peptide identity, low peptide identification scores, etc.). To alleviate this problem, at least partly, implementation of a two-stage glycopeptide identification procedure should be considered. This would require to measure each sample twice by nano-RP-LC-ESI-IT-MS^2 (CID): in the first stage the glycosylation of each *O*-glycopeptide is disregarded and the focus is to reliably identify the peptide moiety of all potential *O*-glycopeptides present in the sample. The basis for this approach is to either enzymatically truncate the *O*-glycan moiety and to only keep the core GalNAc residue attached to Ser or Thr; or to completely remove all *O*-glycans either chemically or enzymatically. Either way, this step has to be done after the glycopeptide enrichment step. By truncating or removing the *O*-glycan moiety, the inherent heterogeneity of each *O*-glycopeptide is reduced, leading to

more intense MS^1 signals and to less complex fragment ion spectra. This simplifies and improves the peptide identification, and allows to replace the time-consuming CID-MS^3 step by a much faster CID MS^2-based peptide identification. Since the proteinase K digest is reproducible, the identified deglycosylated/truncated glycopeptides from this first stage measurement can be translated to subsequent measurements (peptide data base, i.e. spectral library). In the second stage, this data base of unambiguously identified peptides represents the basis for the identification of the corresponding peptide moieties of intact *O*-glycopeptides by CID MS^2 (detection and comparison of the respective peptide mass in the CID MS^2 fragment ion spectrum is in this case sufficient for the identification, i.e. peptide mass fingerprint). This allows to consider the peptide moiety as quasi-constant and to focus on the identification and characterization of the glycan moiety via CID MS^2.

Identification and characterization of intact *N*- and *O*-glycopeptides would also profit from improvements on the ETD fragmentation. ETD is a powerful technique that allows to pinpoint occupied glycosylation sites. The ETD fragmentation efficiency, and therefore the number and the quality of the acquired fragment ion spectra, however, is still inferior to CID. One option to increase the ETD fragmentation efficiency is to artificially increase the precursor ion charge state of glycopeptides during LC-MS ionization. This can be achieved by supplementation of the LC mobile phase with supercharging reagents such as dimethyl sulphoxide or *m*-nitrobenzyl alcohol. Another option is to introduce additional charges to the glycopeptides by appropriate labeling methods, such as tandem mass tag (TMT) labeling.

The mass spectrometric identification and characterization of *N*- and *O*-glycopeptides is very much driven by the capabilities of the mass spectrometer that is used. Crucial parameters in this regard are the maximal resolution, scan rate, mass accuracy, sensitivity, as well as the performance of the available fragmentation modes. In many of these aspects, the amazon ETD ion-trap mass spectrometer, employed in this study, is inferior compared to modern QTOF or Orbitrap mass spectrometers. It is therefore advisable to perform future glycoproteomics studies with one of those instruments.

6.2. In-Depth N- and O-Glycoproteomics

Chapter five describes the consistent further development of the glycoproteomic workflow introduced in chapter four. Thereby, limitations identified within the previous workflow were tackled, along with initial requirements that have not yet been met. A central element of the newly developed workflow are the HCD fragmentation capabilities of the Orbitrap mass spectrometer that was employed. Two HCD fragmentation regimes were established: *HCD.low* operates at fixed low collisional energy and provides detailed information on the glycan moiety; *HCD.step* operates at low and high collisional energy in a stepped manner and allows to unambiguously identify the corresponding peptide moiety. The new workflow also features an improved cotton-HILIC SPE procedure for the enrichment of *N*- and *O*-glycopeptides, which

offers a higher loading capacity, higher throughput, and better ease of use. We have demonstrated the efficiency, flexibility, merits, and limitations of the developed workflow by analyzing four representative and biopharmaceutically-relevant glycoproteins (IgG, Fib, LTF, and RNase B). Aside from confirming already known *N*- and *O*-glycosylation sites on these glycoproteins, we could also detect and characterize a number of new glycosylation sites and regions. Using *HCD.step* a conserved fragmentation signature $[M_{peptide}+H+^{0,2}X\ GlcNAc]^+$ was consistently detected for all analyzed *N*-glycopeptides. This signature represents a valuable, yet rarely used, tool during Orbitrap MS-based glycoproteomics, as it increases the confidence in the identification of the peptide moiety of *N*-glycopeptides. In this thesis, for the first time, the occurrence of this pattern on Orbitrap instruments has been systematically evaluated. The obtained findings thereby allow for a more frequent use of this pattern in future glycoproteomic studies. Another objective of the developed workflow was to provide more in-depth information on the glycan moiety of *N*- and *O*-glycopeptides. In this thesis relative oxonium ion abundances obtained by *HCD.low* and *HCD.step* fragmentation of *N*- and *O*-glycopeptides were comprehensively investigated for their potential to provide structural glycan information. Within this thesis, we could demonstrate the diagnostic potential of oxonium ion abundances for *N*- and *O*-glycoproteomics. We were able to discriminate *N*-glycopeptides from mucin-type core 1 and 2 *O*-glycopeptides based on the HexNAc-H_2O:HexNAc ratio. Moreover, *N*-glycopeptide fragment ion spectra could be classified as high-mannose-type, hybrid-type, or complex-type *N*-glycopeptide spectra solely based on characteristic oxonium ion abundances. Furthermore, characteristic oxonium ion abundances were detected allowing for a discrimination of antenna GlcNAc from bisecting GlcNAc as well as discrimination of a core fucose from an antenna fucose. Many of these oxonium ion-related information have not been adopted by glycopeptide analysis software programs yet. Therefore, this thesis may contribute to the development of new glycopeptide scoring and identification algorithms, which in turn may allow for a more detailed description of the glycan moiety of glycopeptides.

In the future, the presented glycoproteomic workflow may also feature the use of characteristic oxonium ion abundances enabling the discrimination of α2,3- and α2,3-linked NeuAc, as recently described by Pett *et al.* [329]. Also desirable would be the discovery of characteristic oxonium ion abundance for the identification of α-Gal residues in the future. Throughout the glycoproteomic workflows described in Chapter four and five, the annotation and interpretation of glycopeptide fragment ion spectra was identified as the most critical and most time-consuming step. To improve the software-assisted analysis of glycopeptide fragment ion spectra the in-house development of glyXtool^{MS} was initiated in the course of this thesis (development of glyXtool^{MS} in collaboration with M. Pioch [216]). The software glyXtool^{MS} allows for semi-automated in-depth analysis of *N*- and *O*-glycopeptide fragment ion spectra. It thereby considers findings and requirements derived from the present thesis, such as the detection of the conserved *N*-glycopeptide fragmentation pattern, the ability to perform manual *de novo* glycan and peptide annotations, as well as the ability to display graphs for the

relative oxonium ion abundances. This thesis therefore contributed to – but also benefitted from – the in-house development of glyXtool^{MS}.

Overall, in this thesis the in-depth and site-specific glycoproteomic analysis of *N*- and *O*-glycoproteins was demonstrated. Thereby, the developed workflows and obtained findings offer reliable and detailed information on glycopeptide level, along with applicability to a broad range of glycoproteins. With this thesis an LC-MS-based glycoproteomic analysis platform was developed that enables to tackle different *N*- and *O*-glycoproteomic projects in the future. This can include the analysis of biopharmaceutical proteins, such as mAbs, the characterization of viral glycoproteins for vaccine production, or the explorative analysis of human blood plasma glycoproteins derived from a patient cohort. Moreover, the developed workflow may also be applied to other types of protein glycosylation in the future. This may include for instance *O*-mannosylation or *C*-mannosylation.

For the future, glycoproteomic analyses, in general, still require significant efforts and innovations in several regards. This concerns for instance: I) standardization of measurement and analysis procedures to improve inter-laboratory comparability, II) increase of the analysis throughput, e.g. by development of faster and more reliable glycopeptide search engines, III) integration of glycomic and proteomic analyses, IV) quantification of glycopeptides, e.g. by MRM V) intact glycoprotein analysis, VI) establishment of new MS-based analysis techniques for glycoproteomics such as ion mobility MS, VII) providing more structural glycan information on glycopeptide level and integration of such information in software tools.

Bibliography

1. Morell, A. G., Gregoriadis, G., Scheinberg, I. H., Hickman, J., and Ashwell, G. (1971) The role of sialic acid in determining the survival of glycoproteins in the circulation. *J. Biol. Chem.* 246, 1461-1467
2. Eichwald, E. (1865) Beiträge zur Chemie der gewebbildenden Substanzen und ihrer Abkömmlinge. Über das Mucin, besonders der Weinbergschnecke. *Ann. Chem. Pharm.* 134
3. Burger, M. M. (1969) A difference in the architecture of the surface membrane of normal and virally transformed cells. *Proc. Natl. Acad. Sci. U. S. A.* 62, 994-1001
4. Inbar, M., and Sachs, L. (1969) Interaction of the carbohydrate-binding protein concanavalin A with normal and transformed cells. *Proc. Natl. Acad. Sci. U. S. A.* 63, 1418-1425
5. Sharon, N., and Lis, H. (1982) Glycoproteins: research booming on long-ignored ubiquitous compounds. *Mol. Cell. Biochem.* 42, 167-187
6. Varki, A. (2017) Biological roles of glycans. *Glycobiology* 27, 3-49
7. Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (2017) *Essentials of Glycobiology, Third Edition*, Cold Spring Harbor Laboratory Press
8. Moremen, K. W., Tiemeyer, M., and Nairn, A. V. (2012) Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* 13, 448-462
9. Bielik, A. M., and Zaia, J. (2010) Historical Overview of Glycoanalysis. In: Li, J., ed. *Functional Glycomics: Methods and Protocols*, pp. 9-30, Humana Press, Totowa, NJ
10. Alley, W. R., Jr., Mann, B. F., and Novotny, M. V. (2013) High-sensitivity analytical approaches for the structural characterization of glycoproteins. *Chem. Rev.* 113, 2668-2732
11. Montreuil, J., Vliegenthart, J. F. G., and Schachter, H. (1995) *Glycoproteins I*, Elsevier, Amsterdam, The Netehrlands
12. Montreuil, J., Vliegenthart, J. F. G., and Schachter, H. (1997) *Glycoproteins II*, Elsevier, Amsterdam, The Netherlands
13. Hofsteenge, J., Mueller, D. R., de Beer, T., Loeffler, A., Richter, W. J., and Vliegenthart, J. F. G. (1994) New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us. *Biochemistry (Mosc.)* 33, 13524-13530
14. Oman, T. J., Boettcher, J. M., Wang, H., Okalibe, X. N., and Van Der Donk, W. A. (2011) Sublancin is not a lantibiotic but an S-linked glycopeptide. *Nat. Chem. Biol.* 7, 78-80
15. Stepper, J., Shastri, S., Loo, T. S., Preston, J. C., Novak, P., Man, P., Moore, C. H., Havlíč ek, V., Patchett, M. L., and Norris, G. E. (2011) Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Lett.* 585, 645-650
16. Abeln, M., Borst, K. M., Cajic, S., Thiesler, H., Kats, E., Albers, I., Kuhn, M., Kaever, V., Rapp, E., Munster-Kuhnel, A., and Weinhold, B. (2017) Sialylation Is Dispensable for Early Murine Embryonic Development in Vitro. *Chembiochem* 18, 1305-1316
17. Zielinska, D. F., Gnad, F., Wisniewski, J. R., and Mann, M. (2010) Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* 141, 897-907
18. Landsteiner, K. (1901) Ueber Agglutinationserscheinungen normalen menschlichen Blutes. *Wien. Klin. Wochenschr.* 14, 1132-1134
19. Couceiro, J. N. S. S., Paulson, J. C., and Baum, L. G. (1993) Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity. *Virus Res.* 29, 155-165
20. Rhodes, J., Campbell, B. J., and Yu, L.-G. (2010) Glycosylation and Disease. *eLS*, pp. 1-15, John Wiley & Sons, Ltd
21. Lauc, G., Pezer, M., Rudan, I., and Campbell, H. (2015) Mechanisms of disease: The human N-glycome. *Biochim. Biophys. Acta*
22. Ohtsubo, K., and Marth, J. D. (2006) Glycosylation in cellular mechanisms of health and disease. *Cell* 126, 855-867
23. Walsh, G. (2014) Biopharmaceutical benchmarks 2014. *Nat. Biotechnol.* 32, 992-1000
24. Ecker, D. M., Jones, S. D., and Levine, H. L. (2015) The therapeutic monoclonal antibody market. *mAbs* 7, 9-14

25. (2017) Monoclonal Antibodies (mAbs) Market Analysis By Source (Chimeric, Murine, Humanized, Human), By Type of Production, By Indication (Cancer, Autoimmune, Inflammatory, Infectious, Microbial, Viral Diseases), By End-use (Hospitals, Research, Academic Institutes, Clinics, Diagnostic Laboratories) And Segment Forecasts, 2013 - 2024. *Market Research Reports*
26. Reusch, D., and Tejada, M. L. (2015) Fc glycans of therapeutic antibodies as critical quality attributes. *Glycobiology*
27. Hudak, J. E., and Bertozzi, C. R. (2014) Glycotherapy: new advances inspire a reemergence of glycans in medicine. *Chem. Biol.* 21, 16-37
28. Butler, M., and Spearman, M. (2014) The choice of mammalian cell host and possibilities for glycosylation engineering. *Curr. Opin. Biotechnol.* 30, 107-112
29. Dicker, M., and Strasser, R. (2015) Using glyco-engineering to produce therapeutic proteins. *Expert Opin. Biol. Ther.* 15, 1501-1516
30. Kunert, R., and Reinhart, D. (2016) Advances in recombinant antibody manufacturing. *Appl. Microbiol. Biotechnol.* 100, 3451-3461
31. Dinnis, D. M., and James, D. C. (2005) Engineering mammalian cell factories for improved recombinant monoclonal antibody production: lessons from nature? *Biotechnol. Bioeng.* 91, 180-189
32. Frenzel, A., Hust, M., and Schirrmann, T. (2013) Expression of recombinant antibodies. *Frontiers in immunology* 4, 217
33. Stech, M., Nikolaeva, O., Thoring, L., Stocklein, W. F. M., Wustenhagen, D. A., Hust, M., Dubel, S., and Kubick, S. (2017) Cell-free synthesis of functional antibodies using a coupled in vitro transcription-translation system based on CHO cell lysates. *Sci Rep* 7, 12030
34. *U.S. Food and Drug Administration* (2015) Quality Considerations in Demonstrating Biosimilarity of a Therapeutic Protein Product to a Reference Product. Retrieved 24.10.2017 https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM291134.pdf
35. *U.S. Food and Drug Administration* (2016) Clinical Pharmacology Data to Support a Demonstration of Biosimilarity to a Reference Product - Guidance for Industry. Retrieved 24.10.2017 https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM397017.pdf
36. *U.S. Food and Drug Administration* (2015) Scientific Considerations in Demonstrating Biosimilarity to a Reference Product. Retrieved 24.10.2017 https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM291128.pdf
37. *European Medicines Agency* (2016) Guideline on development, production, characterisation and specification for monoclonal antibodies and related products. Retrieved 24.10.2017 http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2016/08/WC500211640.pdf
38. Hossler, P., Khattak, S. F., and Li, Z. J. (2009) Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology* 19, 936-949
39. Brühlmann, D., Jordan, M., Hemberger, J., Sauer, M., Stettler, M., and Broly, H. (2015) Tailoring recombinant protein quality by rational media design. *Biotechnol. Prog.* 31, 615-629
40. Gaunitz, S., Nagy, G., Pohl, N. L. B., and Novotny, M. V. (2017) Recent Advances in the Analysis of Complex Glycoproteins. *Anal. Chem.* 89, 389-413
41. Morelle, W., and Michalski, J.-C. (2007) Analysis of protein glycosylation by mass spectrometry. *Nat. Protocols* 2, 1585-1602
42. Thaysen-Andersen, M., and Packer, N. H. (2014) Advances in LC-MS/MS-based glycoproteomics: Getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. *Biochim. Biophys. Acta* 1844, 1437-1452
43. Wuhrer, M. (2013) Glycomics using mass spectrometry. *Glycoconj. J.* 30, 11-22
44. Wuhrer, M., Catalina, M. I., Deelder, A. M., and Hokke, C. H. (2007) Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J Chromatogr B Analyt Technol Biomed Life Sci* 849, 115-128
45. Walsh, I., Zhao, S., Campbell, M., Taron, C. H., and Rudd, P. M. (2016) Quantitative profiling of glycans and glycopeptides: an informatics' perspective. *Curr. Opin. Struct. Biol.* 40, 70-80
46. Woodin, C. L., Maxon, M., and Desaire, H. (2013) Software for automated interpretation of mass spectrometry data from glycans and glycopeptides. *Analyst* 138, 2793-2803

47. Behne, A., Muth, T., Borowiak, M., Reichl, U., and Rapp, E. (2013) glyXalign: High-throughput migration time alignment preprocessing of electrophoretic data retrieved via multiplexed capillary gel electrophoresis with laser-induced fluorescence detection-based glycoprofiling. *Electrophoresis* 34, 2311-2315
48. Hennig, R., Cajic, S., Borowiak, M., Hoffmann, M., Kottler, R., Reichl, U., and Rapp, E. (2016) Towards personalized diagnostics via longitudinal study of the human plasma N-glycome. *Biochim. Biophys. Acta* 1860, 1728-1738
49. Hennig, R., Rapp, E., Kottler, R., Cajic, S., Borowiak, M., and Reichl, U. (2015) N-Glycosylation Fingerprinting of Viral Glycoproteins by xCGE-LIF. *Methods Mol Biol* 1331, 123-143
50. Huffman, J. E., Pučić-Baković, M., Klarić, L., Hennig, R., Selman, M. H. J., Vučković, F., Novokmet, M., Krištić, J., Borowiak, M., Muth, T., Polašek, O., Razdorov, G., Gornik, O., Plomp, R., Theodoratou, E., Wright, A. F., Rudan, I., Hayward, C., Campbell, H., Deelder, A. M., Reichl, U., Aulchenko, Y. S., Rapp, E., Wuhrer, M., and Lauc, G. (2014) Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol. Cell. Proteomics* 13, 1598-1610
51. Hütter, J., Rödig, J. V., Höper, D., Seeberger, P. H., Reichl, U., Rapp, E., and Lepenies, B. (2013) Toward animal cell culture-based influenza vaccine design: Viral hemagglutinin N-glycosylation markedly impacts immunogenicity. *J. Immunol.* 190, 220-230
52. Ruhaak, L. R., Hennig, R., Huhn, C., Borowiak, M., Dolhain, R. J. E. M., Deelder, A. M., Rapp, E., and Wuhrer, M. (2010) Optimized workflow for preparation of APTS-labeled N-glycans allowing high-throughput analysis of human plasma glycomes using 48-channel multiplexed CGE-LIF. *J. Proteome Res.* 9, 6655-6664
53. Schwarzer, J., Rapp, E., and Reichl, U. (2008) N-glycan analysis by CGE-LIF: profiling influenza A virus hemagglutinin N-glycosylation during vaccine production. *Electrophoresis* 29, 4203-4214
54. Kolarich, D., Jensen, P. H., Altmann, F., and Packer, N. H. (2012) Determination of site-specific glycan heterogeneity on glycoproteins. *Nat. Protoc.* 7, 1285-1298
55. Apweiler, R., Hermjakob, H., and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* 1473, 4-8
56. Wratil, P. R., Horstkorte, R., and Reutter, W. (2016) Metabolic Glycoengineering with N-Acyl Side Chain Modified Mannosamines. *Angewandte Chemie International Edition* 55, 9482-9512
57. Sugisawa, H., and Edo, H. (2006) *The Thermal Degradation of Sugars I. Thermal Polymerization of Glucose*
58. Clamp, J. R. (1974) Analysis of glycoproteins. *Biochem. Soc. Symp.*, 3-16
59. Kornfeld, S., Li, E., and Tabas, I. (1978) The synthesis of complex-type oligosaccharides. II. Characterization of the processing intermediates in the synthesis of the complex oligosaccharide units of the vesicular stomatitis virus G protein. *J. Biol. Chem.* 253, 7771-7778
60. Harvey, D. J., Merry, A. H., Royle, L., Campbell, M. P., Dwek, R. A., and Rudd, P. M. (2009) Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. *Proteomics* 9, 3796-3801
61. Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Marth, J. D., Bertozzi, C. R., Hart, G. W., and Etzler, M. E. (2009) Symbol nomenclature for glycan representation. *Proteomics* 9, 5398-5399
62. Cherepanova, N., Shrimal, S., and Gilmore, R. (2016) N-linked glycosylation and homeostasis of the endoplasmic reticulum. *Curr. Opin. Cell Biol.* 41, 57-65
63. Shrimal, S., and Gilmore, R. (2013) Glycosylation of closely spaced acceptor sites in human glycoproteins. *J. Cell Sci.* 126, 5513-5523
64. (1984) IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Nomenclature and symbolism for amino acids and peptides. Recommendations 1983. *Biochem. J.* 219, 345-373
65. Elliott, S., Chang, D., Delorme, E., Eris, T., and Lorenzini, T. (2004) Structural requirements for additional N-linked carbohydrate on recombinant human erythropoietin. *J. Biol. Chem.* 279, 16854-16862
66. Ellgaard, L., McCaul, N., Chatsisvili, A., and Braakman, I. (2016) Co- and Post-Translational Protein Folding in the ER. *Traffic* 17, 615-638
67. Schauer, R. (2009) Sialic acids as regulators of molecular and cellular interactions. *Curr. Opin. Struct. Biol.* 19, 507-514

68. Wuhrer, M., Koeleman, C. A., Deelder, A. M., and Hokke, C. H. (2006) Repeats of LacdiNAc and fucosylated LacdiNAc on N-glycans of the human parasite Schistosoma mansoni. *FEBS J.* 273, 347-361
69. Dotz, V., and Wuhrer, M. (2016) Histo-blood group glycans in the context of personalized medicine. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1860, 1596-1607
70. Jones, A. (2017) N-glycan analysis of biotherapeutic proteins. *BioPharm Int.* 30, 20-25
71. van Beers, M. M., and Bardor, M. (2012) Minimizing immunogenicity of biopharmaceuticals by controlling critical quality attributes of proteins. *Biotechnol. J.* 7, 1473-1484
72. Macher, B. A., and Galili, U. (2008) The Galalpha1,3Galbeta1,4GlcNAc-R (alpha-Gal) epitope: a carbohydrate of unique evolution and clinical relevance. *Biochim. Biophys. Acta* 1780, 75-88
73. Kuo, C. W., Guu, S. Y., and Khoo, K. H. (2018) Distinctive and Complementary MS(2) Fragmentation Characteristics for Identification of Sulfated Sialylated N-Glycopeptides by nanoLC-MS/MS Workflow. *J. Am. Soc. Mass Spectrom.* 29, 1166-1178
74. Khoo, K. H., and Yu, S. Y. (2010) Mass spectrometric analysis of sulfated N- and O-glycans. *Methods Enzymol.* 478, 3-26
75. Joshi, H. J., Narimatsu, Y., Schjoldager, K. T., Tytgat, H. L. P., Aebi, M., Clausen, H., and Halim, A. (2018) SnapShot: O-Glycosylation Pathways across Kingdoms. *Cell* 172, 632-632 e632
76. Brockhausen, I. (1999) Pathways of O-glycan biosynthesis in cancer cells. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1473, 67-95
77. Darula, Z., Sherman, J., and Medzihradszky, K. F. (2012) How to dig deeper? Improved enrichment methods for mucin core-1 type glycopeptides. *Mol. Cell. Proteomics* 11, O111.016774
78. Varki, A. (2011) Evolutionary forces shaping the Golgi glycosylation machinery: why cell surface glycans are universal to living cells. *Cold Spring Harb Perspect Biol* 3
79. Ozcan, S., Kim, B. J., Ro, G., Kim, J.-H., Bereuter, T. L., Reiter, C., Dimapasoc, L., Garrido, D., Mills, D. A., Grimm, R., Lebrilla, C. B., and An, H. J. (2014) Glycosylated proteins preserved over millennia: N-glycan analysis of Tyrolean Iceman, Scythian Princess and Warrior. *Scientific Reports* 4, 4963
80. Dell, A., Galadari, A., Sastre, F., and Hitchen, P. (2010) Similarities and Differences in the Glycosylation Mechanisms in Prokaryotes and Eukaryotes. *International Journal of Microbiology* 2010, 14
81. Nothaft, H., and Szymanski, C. M. (2013) Bacterial Protein N-Glycosylation: New Perspectives and Applications. *The Journal of Biological Chemistry* 288, 6912-6920
82. (2012) *Transforming Glycoscience: A Roadmap for the Future*, National Academy of Sciences., Washington DC
83. Defaus, S., Gupta, P., Andreu, D., and Gutierrez-Gallego, R. (2014) Mammalian protein glycosylation - structure versus function. *Analyst* 139, 2944-2967
84. Ellgaard, L., and Helenius, A. (2003) Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* 4, 181-191
85. Lopez-Otin, C., and Bond, J. S. (2008) Proteases: multifunctional enzymes in life and disease. *J. Biol. Chem.* 283, 30433-30437
86. Srinivasan, S., Romagnoli, M., Bohm, A., and Sonenshein, G. E. (2014) N-glycosylation regulates ADAM8 processing and activation. *J. Biol. Chem.* 289, 33676-33688
87. Parekh, R. B., Dwek, R. A., Sutton, B. J., Fernandes, D. L., Leung, A., Stanworth, D., Rademacher, T. W., Mizuochi, T., Taniguchi, T., Matsuta, K., and et al. (1985) Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature* 316, 452-457
88. Anthony, R. M., Nimmerjahn, F., Ashline, D. J., Reinhold, V. N., Paulson, J. C., and Ravetch, J. V. (2008) Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. *Science* 320, 373-376
89. Ferrara, C., Grau, S., Jager, C., Sondermann, P., Brunker, P., Waldhauer, I., Hennig, M., Ruf, A., Rufer, A. C., Stihle, M., Umana, P., and Benz, J. (2011) Unique carbohydrate-carbohydrate interactions are required for high affinity binding between FcgammaRIII and antibodies lacking core fucose. *Proc. Natl. Acad. Sci. U. S. A.* 108, 12669-12674
90. Vogel, J., Sperandio, M., Pries, A. R., Linderkamp, O., Gaehtgens, P., and Kuschinsky, W. (2000) Influence of the endothelial glycocalyx on cerebral blood flow in mice. *J. Cereb. Blood Flow Metab.* 20, 1571-1578

91. Zhu, F., and Wu, H. (2016) Insights into bacterial protein glycosylation in human microbiota. *Sci China Life Sci* 59, 11-18
92. Cohen, M., Zhang, X.-Q., Senaati, H. P., Chen, H.-W., Varki, N. M., Schooley, R. T., and Gagneux, P. (2013) Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase. *Virol. J.* 10, 321
93. Yang, J., Liu, S., Du, L., and Jiang, S. (2016) A new role of neuraminidase (NA) in the influenza virus life cycle: implication for developing NA inhibitors with novel mechanism of action. *Rev. Med. Virol.* 26, 242-250
94. McAuley, J. L., Corcilius, L., Tan, H. X., Payne, R. J., McGuckin, M. A., and Brown, L. E. (2017) The cell surface mucin MUC1 limits the severity of influenza A virus infection. *Mucosal Immunol*
95. Moscona, A. (2005) Neuraminidase inhibitors for influenza. *N. Engl. J. Med.* 353, 1363-1373
96. Chandrasekaran, A., Srinivasan, A., Raman, R., Viswanathan, K., Raguram, S., Tumpey, T. M., Sasisekharan, V., and Sasisekharan, R. (2008) Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nat. Biotechnol.* 26, 107-113
97. Marquardt, T., and Denecke, J. (2003) Congenital disorders of glycosylation: review of their molecular bases, clinical presentations and specific therapies. *Eur. J. Pediatr.* 162, 359-379
98. Scott, K., Gadomski, T., Kozicz, T., and Morava, E. (2014) Congenital disorders of glycosylation: new defects and still counting. *J. Inherit. Metab. Dis.* 37, 609-617
99. Thiel, C., and Körner, C. (2013) Therapies and therapeutic approaches in Congenital Disorders of Glycosylation. *Glycoconj. J.* 30, 77-84
100. Lehle, L., Strahl, S., and Tanner, W. (2006) Protein glycosylation, conserved from yeast to man: a model organism helps elucidate congenital human diseases. *Angew. Chem. Int. Ed. Engl.* 45, 6802-6818
101. Nakayama, Y., Nakamura, N., Tsuji, D., Itoh, K., and Kurosak, A. (2013) Genetic Diseases Associated with Protein Glycosylation Disorders in Mammals.
102. Altmann, F. (2007) The role of protein glycosylation in allergy. *Int. Arch. Allergy Immunol.* 142, 99-115
103. Thaysen-Andersen, M., Packer, N. H., and Schulz, B. L. (2016) Maturing glycoproteomics technologies provide unique structural insights into the N-glycoproteome and its regulation in health and disease. *Mol. Cell. Proteomics* 15, 1773-1790
104. Varki, A. Sialic acids in human health and disease. *Trends Mol. Med.* 14, 351-360
105. Imberty, A., and Varrot, A. (2008) Microbial recognition of human cell surface glycoconjugates. *Curr. Opin. Struct. Biol.* 18, 567-576
106. Marth, J. D., and Grewal, P. K. (2008) Mammalian glycosylation in immunity. *Nat. Rev. Immunol.* 8, 874-887
107. Connelly, M. A., Gruppen, E. G., Otvos, J. D., and Dullaart, R. P. F. (2016) Inflammatory glycoproteins in cardiometabolic disorders, autoimmune diseases and cancer. *Clin. Chim. Acta* 459, 177-186
108. Boren, T., Falk, P., Roth, K. A., Larson, G., and Normark, S. (1993) Attachment of Helicobacter pylori to human gastric epithelium mediated by blood group antigens. *Science* 262, 1892-1895
109. Baumler, A. J., and Sperandio, V. (2016) Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* 535, 85-93
110. Marcobal, A., Southwick, A. M., Earle, K. A., and Sonnenburg, J. L. (2013) A refined palate: bacterial consumption of host glycans in the gut. *Glycobiology* 23, 1038-1046
111. Trbojevic Akmacic, I., Ventham, N. T., Theodoratou, E., Vuckovic, F., Kennedy, N. A., Kristic, J., Nimmo, E. R., Kalla, R., Drummond, H., Stambuk, J., Dunlop, M. G., Novokmet, M., Aulchenko, Y., Gornik, O., Campbell, H., Pucic Bakovic, M., Satsangi, J., Lauc, G., and Consortium, I.-B. (2015) Inflammatory bowel disease associates with proinflammatory potential of the immunoglobulin G glycome. *Inflamm. Bowel Dis.* 21, 1237-1247
112. Wuhrer, M., Stam, J. C., van de Geijn, F. E., Koeleman, C. A., Verrips, C. T., Dolhain, R. J., Hokke, C. H., and Deelder, A. M. (2007) Glycosylation profiling of immunoglobulin G (IgG) subclasses from human serum. *Proteomics* 7, 4070-4081
113. Plomp, R., Bondt, A., de Haan, N., Rombouts, Y., and Wuhrer, M. (2016) Recent Advances in Clinical Glycoproteomics of Immunoglobulins (Igs). *Mol. Cell. Proteomics* 15, 2217-2228
114. Bondt, A., Selman, M. H., Deelder, A. M., Hazes, J. M., Willemsen, S. P., Wuhrer, M., and Dolhain, R. J. (2013) Association between galactosylation of immunoglobulin G and improvement of rheumatoid arthritis during pregnancy is independent of sialylation. *J Proteome Res* 12, 4522-4531

115. Pinho, S. S., and Reis, C. A. (2015) Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* 15, 540-555
116. Cazet, A., Julien, S., Bobowski, M., Burchell, J., and Delannoy, P. (2010) Tumour-associated carbohydrate antigens in breast cancer. *Breast Cancer Res.* 12, 204
117. Lommel, M., and Strahl, S. (2009) Protein O-mannosylation: conserved from bacteria to humans. *Glycobiology* 19, 816-828
118. Loibl, M., and Strahl, S. (2013) Protein O-mannosylation: what we have learned from baker's yeast. *Biochim. Biophys. Acta* 1833, 2438-2446
119. Sentandreu, R., and Northcote, D. H. (1968) The structure of a glycopeptide isolated from the yeast cell wall. *Biochem. J.* 109, 419-432
120. Jurado, L. A., Coloma, A., and Cruces, J. (1999) Identification of a human homolog of the Drosophila rotated abdomen gene (POMT1) encoding a putative protein O-mannosyl-transferase, and assignment to human chromosome 9q34.1. *Genomics* 58, 171-180
121. Steentoft, C., Bennett, E. P., Schjoldager, K. T., Vakhrushev, S. Y., Wandall, H. H., and Clausen, H. (2014) Precision genome editing: a small revolution for glycobiology. *Glycobiology* 24, 663-680
122. Griffin, M. E., and Hsieh-Wilson, L. C. (2016) Glycan Engineering for Cell and Developmental Biology. *Cell Chem Biol* 23, 108-121
123. Steentoft, C., Vakhrushev, S. Y., Vester-Christensen, M. B., Schjoldager, K. T. B. G., Kong, Y., Bennett, E. P., Mandel, U., Wandall, H., Levery, S. B., and Clausen, H. (2011) Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat. Methods* 8, 977-982
124. Meuris, L., Santens, F., Elson, G., Festjens, N., Boone, M., Dos Santos, A., Devos, S., Rousseau, F., Plets, E., Houthuys, E., Malinge, P., Magistrelli, G., Cons, L., Chatel, L., Devreese, B., and Callewaert, N. (2014) GlycoDelete engineering of mammalian cells simplifies N-glycosylation of recombinant proteins. *Nat Biotech* 32, 485-489
125. Piron, R., Santens, F., De Paepe, A., Depicker, A., and Callewaert, N. (2015) Using GlycoDelete to produce proteins lacking plant-specific N-glycan modification in seeds. *Nat. Biotechnol.* 33, 1135-1137
126. Lannoo, N., and Van Damme, E. J. (2015) Review/N-glycans: The making of a varied toolbox. *Plant Sci* 239, 67-83
127. Beck, A. (2011) Biosimilar, biobetter and next generation therapeutic antibodies. *mAbs* 3, 107-110
128. Val, I. J. d., Kyriakopoulos, S., Kontoravdi, C., Jedrzejewski, P. M., Exley, K., Sou, S. N., and Polizzi, M. (2012) *Application of Quality by Design Paradigm to the Manufacture of Protein Therapeutics*
129. Eon-Duval, A., Broly, H., and Gleixner, R. (2012) Quality attributes of recombinant therapeutic proteins: an assessment of impact on safety and efficacy as part of a quality by design development approach. *Biotechnol. Prog.* 28, 608-622
130. del Val, I. J., Kontoravdi, C., and Nagy, J. M. (2010) Towards the implementation of quality by design to the production of therapeutic monoclonal antibodies with desired glycosylation patterns. *Biotechnol. Prog.* 26, 1505-1527
131. Beck, A., and Reichert, J. M. (2013) Approval of the first biosimilar antibodies in Europe: a major landmark for the biopharmaceutical industry. *mAbs* 5, 621-623
132. Lingg, N., Zhang, P., Song, Z., and Bardor, M. (2012) The sweet tooth of biopharmaceuticals: importance of recombinant protein glycosylation analysis. *Biotechnol. J.* 7, 1462-1472
133. Yang, Z., Wang, S., Halim, A., Schulz, M. A., Frodin, M., Rahman, S. H., Vester-Christensen, M. B., Behrens, C., Kristensen, C., Vakhrushev, S. Y., Bennett, E. P., Wandall, H. H., and Clausen, H. (2015) Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat. Biotechnol.* 33, 842-844
134. DeFrees, S., Wang, Z. G., Xing, R., Scott, A. E., Wang, J., Zopf, D., Gouty, D. L., Sjoberg, E. R., Panneerselvam, K., Brinkman-Van der Linden, E. C., Bayer, R. J., Tarp, M. A., and Clausen, H. (2006) GlycoPEGylation of recombinant therapeutic proteins produced in Escherichia coli. *Glycobiology* 16, 833-843
135. Dalziel, M., Crispin, M., Scanlan, C. N., Zitzmann, N., and Dwek, R. A. (2014) Emerging principles for the therapeutic exploitation of glycosylation. *Science* 343, 1235681
136. Elgundi, Z., Reslan, M., Cruz, E., Sifniotis, V., and Kayser, V. (2017) The state-of-play and future of antibody therapeutics. *Adv Drug Deliv Rev* 122, 2-19
137. Solá, R. J., and Griebenow, K. (2010) Glycosylation of Therapeutic Proteins: An Effective Strategy to Optimize Efficacy. *BioDrugs : clinical immunotherapeutics, biopharmaceuticals and gene therapy* 24, 9-21

138. Mimura, Y., Katoh, T., Saldova, R., O'Flaherty, R., Izumi, T., Mimura-Kimura, Y., Utsunomiya, T., Mizukami, Y., Yamamoto, K., Matsumoto, T., and Rudd, P. M. (2017) Glycosylation engineering of therapeutic IgG antibodies: challenges for the safety, functionality and efficacy. *Protein & Cell*
139. Shields, R. L., Lai, J., Keck, R., O'Connell, L. Y., Hong, K., Meng, Y. G., Weikert, S. H., and Presta, L. G. (2002) Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcgamma RIII and antibody-dependent cellular toxicity. *J. Biol. Chem.* 277, 26733-26740
140. Shinkawa, T., Nakamura, K., Yamane, N., Shoji-Hosaka, E., Kanda, Y., Sakurada, M., Uchida, K., Anazawa, H., Satoh, M., Yamasaki, M., Hanai, N., and Shitara, K. (2003) The absence of fucose but not the presence of galactose or bisecting N-acetylglucosamine of human IgG1 complex-type oligosaccharides shows the critical role of enhancing antibody-dependent cellular cytotoxicity. *J. Biol. Chem.* 278, 3466-3473
141. Ferrara, C., Brunker, P., Suter, T., Moser, S., Puntener, U., and Umana, P. (2006) Modulation of therapeutic antibody effector functions by glycosylation engineering: influence of Golgi enzyme localization domain and co-expression of heterologous beta1, 4-N-acetylglucosaminyltransferase III and Golgi alpha-mannosidase II. *Biotechnol. Bioeng.* 93, 851-861
142. Beck, A., and Reichert, J. M. (2012) Marketing approval of mogamulizumab: a triumph for glyco-engineering. *mAbs* 4, 419-425
143. Salles, G., Morschhauser, F., Lamy, T., Milpied, N., Thieblemont, C., Tilly, H., Bieska, G., Asikanius, E., Carlile, D., Birkett, J., Pisa, P., and Cartron, G. (2012) Phase 1 study results of the type II glycoengineered humanized anti-CD20 monoclonal antibody obinutuzumab (GA101) in B-cell lymphoma patients. *Blood* 119, 5126-5132
144. Sehn, L. H., Assouline, S. E., Stewart, D. A., Mangel, J., Gascoyne, R. D., Fine, G., Frances-Lasserre, S., Carlile, D. J., and Crump, M. (2012) A phase 1 study of obinutuzumab induction followed by 2 years of maintenance in patients with relapsed CD20-positive B-cell malignancies. *Blood* 119, 5118-5125
145. Elbein, A. D., Tropea, J. E., Mitchell, M., and Kaushal, G. P. (1990) Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. *J. Biol. Chem.* 265, 15599-15605
146. Elbein, A. D., Solf, R., Dorling, P. R., and Vosbeck, K. (1981) Swainsonine: an inhibitor of glycoprotein processing. *Proc. Natl. Acad. Sci. U. S. A.* 78, 7393-7397
147. Zhou, Q., Shankara, S., Roy, A., Qiu, H., Estes, S., McVie-Wylie, A., Culm-Merdek, K., Park, A., Pan, C., and Edmunds, T. (2008) Development of a simple and rapid method for producing non-fucosylated oligomannose containing antibodies with increased effector function. *Biotechnol. Bioeng.* 99, 652-665
148. Surve, T., and Gadgil, M. (2015) Manganese increases high mannose glycoform on monoclonal antibody expressed in CHO when glucose is absent or limiting: Implications for use of alternate sugars. *Biotechnol. Prog.* 31, 460-467
149. Dumont, J., Euwart, D., Mei, B., Estes, S., and Kshirsagar, R. (2016) Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Crit. Rev. Biotechnol.* 36, 1110-1122
150. North, S. J., Huang, H. H., Sundaram, S., Jang-Lee, J., Etienne, A. T., Trollope, A., Chalabi, S., Dell, A., Stanley, P., and Haslam, S. M. (2010) Glycomics profiling of Chinese hamster ovary cell glycosylation mutants reveals N-glycans of a novel size and complexity. *J. Biol. Chem.* 285, 5759-5775
151. Kesik-Brodacka, M. (2018) Progress in biopharmaceutical development. *Biotechnol. Appl. Biochem.* 65, 306-322
152. Genzel, Y. (2015) Designing cell lines for viral vaccine production: Where do we stand? *Biotechnol. J.* 10, 728-740
153. Baeshen, M. N., Al-Hejin, A. M., Bora, R. S., Ahmed, M. M., Ramadan, H. A., Saini, K. S., Baeshen, N. A., and Redwan, E. M. (2015) Production of Biopharmaceuticals in E. coli: Current Scenario and Future Perspectives. *Journal of microbiology and biotechnology* 25, 953-962
154. Ward, M., Lin, C., Victoria, D. C., Fox, B. P., Fox, J. A., Wong, D. L., Meerman, H. J., Pucci, J. P., Fong, R. B., Heng, M. H., Tsurushita, N., Gieswein, C., Park, M., and Wang, H. (2004) Characterization of humanized antibodies secreted by Aspergillus niger. *Appl. Environ. Microbiol.* 70, 2567-2576
155. Gerngross, T. U. (2004) Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. *Nat. Biotechnol.* 22, 1409-1414

156. Hamilton, S. R., Davidson, R. C., Sethuraman, N., Nett, J. H., Jiang, Y., Rios, S., Bobrowicz, P., Stadheim, T. A., Li, H., Choi, B. K., Hopkins, D., Wischnewski, H., Roser, J., Mitchell, T., Strawbridge, R. R., Hoopes, J., Wildt, S., and Gerngross, T. U. (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* 313, 1441-1443
157. Jacobs, P. P., Geysens, S., Vervecken, W., Contreras, R., and Callewaert, N. (2009) Engineering complex-type N-glycosylation in Pichia pastoris using GlycoSwitch technology. *Nat. Protoc.* 4, 58-70
158. Aumiller, J. J., Mabashi-Asazuma, H., Hillar, A., Shi, X., and Jarvis, D. L. (2012) A new glycoengineered insect cell line with an inducibly mammalianized protein N-glycosylation pathway. *Glycobiology* 22, 417-428
159. Cox, M. M. (2012) Recombinant protein vaccines produced in insect cells. *Vaccine* 30, 1759-1766
160. Steinkellner, H., and Castilho, A. (2015) N-Glyco-Engineering in Plants: Update on Strategies and Major Achievements. *Methods Mol Biol* 1321, 195-212
161. Zhu, L., van de Lavoir, M. C., Albanese, J., Beenhouwer, D. O., Cardarelli, P. M., Cuison, S., Deng, D. F., Deshpande, S., Diamond, J. H., Green, L., Halk, E. L., Heyer, B. S., Kay, R. M., Kerchner, A., Leighton, P. A., Mather, C. M., Morrison, S. L., Nikolov, Z. L., Passmore, D. B., Pradas-Monne, A., Preston, B. T., Rangan, V. S., Shi, M., Srinivasan, M., White, S. G., Winters-Digiacinto, P., Wong, S., Zhou, W., and Etches, R. J. (2005) Production of human monoclonal antibody in eggs of chimeric chickens. *Nat. Biotechnol.* 23, 1159-1169
162. Bertolini, L. R., Meade, H., Lazzarotto, C. R., Martins, L. T., Tavares, K. C., Bertolini, M., and Murray, J. D. (2016) The transgenic animal platform for biopharmaceutical production. *Transgenic Res.* 25, 329-343
163. Thoring, L., Wustenhagen, D. A., Borowiak, M., Stech, M., Sonnabend, A., and Kubick, S. (2016) Cell-Free Systems Based on CHO Cell Lysates: Optimization Strategies, Synthesis of "Difficult-to-Express" Proteins and Future Perspectives. *PLoS ONE* 11, e0163670
164. Beck, A., Wagner-Rousset, E., Bussat, M. C., Lokteff, M., Klinguer-Hamour, C., Haeuw, J. F., Goetsch, L., Wurch, T., Van Dorsselaer, A., and Corvaia, N. (2008) Trends in glycosylation, glycoanalysis and glycoengineering of therapeutic antibodies and Fc-fusion proteins. *Curr. Pharm. Biotechnol.* 9, 482-501
165. Keser, T., Pavic, T., Lauc, G., and Gornik, O. (2018) Comparison of 2-Aminobenzamide, Procainamide and RapiFluor-MS as Derivatizing Agents for High-Throughput HILIC-UPLC-FLR-MS N-glycan Analysis. *Front Chem* 6, 324
166. Zhou, S., Veillon, L., Dong, X., Huang, Y., and Mechref, Y. (2017) Direct comparison of derivatization strategies for LC-MS/MS analysis of N-glycans. *Analyst* 142, 4446-4455
167. Zhou, S., Wooding, K. M., and Mechref, Y. (2017) Analysis of Permethylated Glycan by Liquid Chromatography (LC) and Mass Spectrometry (MS). *Methods Mol Biol* 1503, 83-96
168. Marino, K., Bones, J., Kattla, J. J., and Rudd, P. M. (2010) A systematic approach to protein glycosylation analysis: a path through the maze. *Nat. Chem. Biol.* 6, 713-723
169. Mantovani, V., Galeotti, F., Maccari, F., and Volpi, N. (2018) Recent advances in capillary electrophoresis separation of monosaccharides, oligosaccharides, and polysaccharides. *Electrophoresis* 39, 179-189
170. Hardy, M. R., Townsend, R. R., and Lee, Y. C. (1988) Monosaccharide analysis of glycoconjugates by anion exchange chromatography with pulsed amperometric detection. *Anal Biochem* 170, 54-62
171. Galermo, A. G., Nandita, E., Barboza, M., Amicucci, M. J., Vo, T. T., and Lebrilla, C. B. (2018) Liquid Chromatography-Tandem Mass Spectrometry Approach for Determining Glycosidic Linkages. *Anal Chem*
172. Xu, G., Amicucci, M. J., Cheng, Z., Galermo, A. G., and Lebrilla, C. B. (2017) Revisiting monosaccharide analysis - quantitation of a comprehensive set of monosaccharides using dynamic multiple reaction monitoring. *Analyst* 143, 200-207
173. Medeiros, P. M., and Simoneit, B. R. (2007) Analysis of sugars in environmental samples by gas chromatography-mass spectrometry. *J Chromatogr A* 1141, 271-278
174. Vliegenthart, J. F. G. (2017) The complexity of glycoprotein-derived glycans. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 93, 64-86
175. Vanderschaeghe, D., Festjens, N., Delanghe, J., and Callewaert, N. (2010) Glycome profiling using modern glycomics technology: technical aspects and applications. *Biol. Chem.* 391, 149-161

176. Pucic-Bakovic, M. (2017) High-Throughput Analysis of the IgG N-Glycome by UPLC-FLR. *Methods Mol Biol* 1503, 21-29
177. Zauner, G., Deelder, A. M., and Wuhrer, M. (2011) Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics. *Electrophoresis* 32, 3456-3466
178. Jensen, P. H., Karlsson, N. G., Kolarich, D., and Packer, N. H. (2012) Structural analysis of N- and O-glycans released from glycoproteins. *Nat. Protoc.* 7, 1299-1310
179. Hofmann, J., and Pagel, K. (2017) Glycan Analysis by Ion Mobility–Mass Spectrometry. *Angew. Chem. Int. Ed.* 56, 8342-8349
180. Reiding, K. R., Blank, D., Kuijper, D. M., Deelder, A. M., and Wuhrer, M. (2014) High-Throughput Profiling of Protein N-Glycosylation by MALDI-TOF-MS Employing Linkage-Specific Sialic Acid Esterification. *Anal. Chem.* 86, 5784-5793
181. Bladergroen, M. R., Reiding, K. R., Hipgrave-Ederveen, A. L., Vreeker, G. C. M., Clerc, F., Holst, S., Bondt, A., Wuhrer, M., and van der Burgt, Y. E. M. (2015) Automation of high-throughput mass spectrometry-based plasma N-glycome analysis with linkage-specific sialic acid esterification. *J. Proteome Res.*
182. Reiding, K. R., Bondt, A., Hennig, R., Gardner, R. A., O'Flaherty, R., Trbojevic-Akmacic, I., Shubhakar, A., Hazes, J., Reichl, U., Fernandes, D. L., Pucic-Bakovic, M., Rapp, E., Spencer, D. I. R., Dolhain, R., Rudd, P., Lauc, G., and Wuhrer, M. (2018) High-throughput serum N-glycomics: method comparison and application to study rheumatoid arthritis and pregnancy-associated changes. *Mol. Cell. Proteomics*
183. Barroso, A., Gimenez, E., Konijnenberg, A., Sancho, J., Sanz-Nebot, V., and Sobott, F. (2018) Evaluation of ion mobility for the separation of glycoconjugate isomers due to different types of sialic acid linkage, at the intact glycoprotein, glycopeptide and glycan level. *Journal of Proteomics* 173, 22-31
184. Sanz-Nebot, V., Balaguer, E., Benavente, F., Neususs, C., and Barbosa, J. (2007) Characterization of transferrin glycoforms in human serum by CE-UV and CE-ESI-MS. *Electrophoresis* 28, 1949-1957
185. Balaguer, E., and Neususs, C. (2006) Glycoprotein characterization combining intact protein and glycan analysis by capillary electrophoresis-electrospray ionization-mass spectrometry. *Anal Chem* 78, 5384-5393
186. Hernandez-Alba, O., Wagner-Rousset, E., Beck, A., and Cianferani, S. (2018) Native Mass Spectrometry, Ion Mobility, and Collision-Induced Unfolding for Conformational Characterization of IgG4 Monoclonal Antibodies. *Anal Chem* 90, 8865-8872
187. Bern, M., Caval, T., Kil, Y. J., Tang, W., Becker, C., Carlson, E., Kletter, D., Sen, K. I., Galy, N., Hagemans, D., Franc, V., and Heck, A. J. R. (2018) Parsimonious Charge Deconvolution for Native Mass Spectrometry. *J Proteome Res* 17, 1216-1226
188. Baerenfaenger, M., and Meyer, B. (2018) Intact Human Alpha-Acid Glycoprotein Analyzed by ESI-qTOF-MS: Simultaneous Determination of the Glycan Composition of Multiple Glycosylation Sites. *J Proteome Res*
189. Belov, A. M., Zang, L., Sebastiano, R., Santos, M. R., Bush, D. R., Karger, B. L., and Ivanov, A. R. (2018) Complementary middle-down and intact monoclonal antibody proteoform characterization by capillary zone electrophoresis - mass spectrometry. *Electrophoresis*
190. Levery, S. B., Steentoft, C., Halim, A., Narimatsu, Y., Clausen, H., and Vakhrushev, S. Y. (2014) Advances in mass spectrometry driven O-glycoproteomics. *Biochim. Biophys. Acta* 1850, 33-42
191. Yang, Y., Franc, V., and Heck, A. J. R. (2017) Glycoproteomics: A Balance between High-Throughput and In-Depth Analysis. *Trends Biotechnol.* 35, 598-609
192. Nishikaze, T., Kawabata, S. I., and Tanaka, K. (2014) Fragmentation characteristics of deprotonated N-linked glycopeptides: Influences of amino acid composition and sequence. *J. Am. Soc. Mass Spectrom.* 25, 988-998
193. de Haan, N., Reiding, K. R., Haberger, M., Reusch, D., Falck, D., and Wuhrer, M. (2015) Linkage-specific sialic acid derivatization for MALDI-TOF-MS profiling of IgG glycopeptides. *Anal Chem*
194. Selman, M. H., Hoffmann, M., Zauner, G., McDonnell, L. A., Balog, C. I., Rapp, E., Deelder, A. M., and Wuhrer, M. (2012) MALDI-TOF-MS analysis of sialylated glycans and glycopeptides using 4-chloro-alpha-cyanocinnamic acid matrix. *Proteomics* 12, 1337-1348
195. Selman, M. H., McDonnell, L. A., Palmblad, M., Ruhaak, L. R., Deelder, A. M., and Wuhrer, M. (2010) Immunoglobulin G glycopeptide profiling by matrix-assisted laser desorption ionization Fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* 82, 1073-1081

196. Stavenhagen, K., Plomp, R., and Wuhrer, M. (2015) Site-Specific Protein N- and O-Glycosylation Analysis by a C18-Porous Graphitized Carbon–Liquid Chromatography-Electrospray Ionization Mass Spectrometry Approach Using Pronase Treated Glycopeptides. *Anal. Chem.* 87, 11691-11699
197. Watson, J. T., and Sparkman, O. D. (2008) *Introduction to Mass Spectrometry: Instrumentation, Applications and Strategies for Data Interpretation: Fourth Edition*
198. Domon, B., and Costello, C. E. (1988) A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* 5, 397-409
199. Biemann, K. (1990) Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol.* 193, 886-887
200. Hinneburg, H., Stavenhagen, K., Schweiger-Hufnagel, U., Pengelley, S., Jabs, W., Seeberger, P. H., Silva, D. V., Wuhrer, M., and Kolarich, D. (2016) The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics. *J. Am. Soc. Mass Spectrom.* 27, 507-519
201. Reiding, K. R., Bondt, A., Franc, V., and Heck, A. J. R. (2018) The benefits of hybrid fragmentation methods for glycoproteomics. *Trac-Trend Anal Chem* 108, 260-268
202. Zhu, J., Chen, Z., Zhang, J., An, M., Wu, J., Yu, Q., Skilton, S. J., Bern, M., Ilker Sen, K., Li, L., and Lubman, D. M. (2018) Differential Quantitative Determination of Site-Specific Intact N-Glycopeptides in Serum Haptoglobin between Hepatocellular Carcinoma and Cirrhosis Using LC-EThcD-MS/MS. *J Proteome Res*
203. Ko, B. J., and Brodbelt, J. S. (2015) Comparison of glycopeptide fragmentation by collision induced dissociation and ultraviolet photodissociation. *International Journal of Mass Spectrometry* 377, 385-392
204. Madsen, J. A., Ko, B. J., Xu, H., Iwashkiw, J. A., Robotham, S. A., Shaw, J. B., Feldman, M. F., and Brodbelt, J. S. (2013) Concurrent automated sequencing of the glycan and peptide portions of O-linked glycopeptide anions by ultraviolet photodissociation mass spectrometry. *Anal Chem* 85, 9253-9261
205. Rodriguez, J., Gupta, N., Smith, R. D., and Pevzner, P. A. (2008) Does trypsin cut before proline? *J Proteome Res* 7, 300-305
206. Burkhart, J. M., Schumbrutzki, C., Wortelkamp, S., Sickmann, A., and Zahedi, R. P. (2012) Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *J. Proteomics* 75, 1454-1462
207. Hanisch, F. G. (2001) O-glycosylation of the mucin type. *Biol. Chem.* 382, 143-149
208. Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359-362
209. Ongay, S., Boichenko, A., Govorukhina, N., and Bischoff, R. (2012) Glycopeptide enrichment and separation for protein glycosylation analysis. *J. Sep. Sci.* 35, 2341-2372
210. Khatri, K., Staples, G. O., Leymarie, N., Leon, D. R., Turiák, L., Huang, Y., Yip, S., Hu, H., Heckendorf, C. F., and Zaia, J. (2014) Confident assignment of site-specific glycosylation in complex glycoproteins in a single step. *J Proteome Res* 13, 4347-4355
211. Nilsson, J., Ruetschi, U., Halim, A., Hesse, C., Carlsohn, E., Brinkmalm, G., and Larson, G. (2009) Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat Meth* 6, 809-811
212. Rogers, J. C., and Bomgarden, R. D. (2016) Sample Preparation for Mass Spectrometry-Based Proteomics; from Proteomes to Peptides. *Adv. Exp. Med. Biol.* 919, 43-62
213. Stavenhagen, K., Hinneburg, H., Thaysen-Andersen, M., Hartmann, L., Varon Silva, D., Fuchser, J., Kaspar, S., Rapp, E., Seeberger, P. H., and Kolarich, D. (2013) Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *J. Mass Spectrom.* 48, 627-639
214. Goldman, R., and Sanda, M. (2015) Targeted methods for quantitative analysis of protein glycosylation. *Proteomics Clin Appl.* 9, 17-32
215. Miyamoto, S., Stroble, C. D., Taylor, S., Hong, Q., Lebrilla, C. B., Leiserowitz, G. S., Kim, K., and Ruhaak, L. R. (2018) Multiple Reaction Monitoring for the Quantitation of Serum Protein Glycosylation Profiles: Application to Ovarian Cancer. *J Proteome Res* 17, 222-233
216. Pioch, M., Hoffmann, M., Pralow, A., Reichl, U., and Rapp, E. (2018) glyXtool(MS): An Open-Source Pipeline for Semiautomated Analysis of Glycopeptide Mass Spectrometry Data. *Anal Chem*

217. Skala, W., Wohlschlager, T., Senn, S., Huber, G. E., and Huber, C. G. (2018) MoFi: A Software Tool for Annotating Glycoprotein Mass Spectra by Integrating Hybrid Data from the Intact Protein and Glycopeptide Level. *Anal Chem* 90, 5728-5736
218. Mariethoz, J., Alocci, D., Gastaldello, A., Horlacher, O., Gasteiger, E., Rojas-Macias, M., Karlsson, N. G., Packer, N., and Lisacek, F. (2018) Glycomics@ExPASy: Bridging the gap. *Mol. Cell. Proteomics*
219. Alocci, D., Mariethoz, J., Gastaldello, A., Gasteiger, E., Karlsson, N. G., Kolarich, D., Packer, N. H., and Lisacek, F. (2018) GlyConnect: glycoproteomics goes visual, interactive and analytical. *J Proteome Res*
220. Bern, M., Beniston, R., and Mesnage, S. (2017) Towards an automated analysis of bacterial peptidoglycan structure. *Anal. Bioanal. Chem.* 409, 551-560
221. Lee, L. Y., Moh, E. S. X., Parker, B. L., Bern, M., Packer, N. H., and Thaysen-Andersen, M. (2016) Toward Automated N-Glycopeptide Identification in Glycoproteomics. *J. Proteome Res.* 15, 3904-3915
222. Yang, Y., Liu, F., Franc, V., Halim, L. A., Schellekens, H., and Heck, A. J. (2016) Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat Commun* 7, 13397
223. Yang, Y., Wang, G., Song, T., Lebrilla, C. B., and Heck, A. J. R. (2017) Resolving the micro-heterogeneity and structural integrity of monoclonal antibodies by hybrid mass spectrometric approaches. *mAbs* 9, 638-645
224. Stavenhagen, K., Kayili, H. M., Holst, S., Koeleman, C., Engel, R., Wouters, D., Zeerleder, S., Salih, B., and Wuhrer, M. (2017) N- and O-glycosylation analysis of human C1-inhibitor reveals extensive mucin-type O-glycosylation. *Mol. Cell. Proteomics*
225. Hoffmann, M., Marx, K., Reichl, U., Wuhrer, M., and Rapp, E. (2016) Site-specific O-Glycosylation Analysis of Human Blood Plasma Proteins. *Mol. Cell. Proteomics* 15, 624-641
226. Selman, M. H., Hemayatkar, M., Deelder, A. M., and Wuhrer, M. (2011) Cotton HILIC SPE microtips for microscale purification and enrichment of glycans and glycopeptides. *Anal Chem* 83, 2492-2499
227. Hoffmann, M., Pioch, M., Pralow, A., Hennig, R., Kottler, R., Reichl, U., and Rapp, E. (2018) The Fine Art of Destruction: A Guide to In-Depth Glycoproteomic Analyses-Exploiting the Diagnostic Potential of Fragment Ions. *Proteomics* 18, e1800282
228. Pioch, M., Hoffmann, M., Pralow, A., Reichl, U., and Rapp, E. (2018) glyXtool(MS): An Open-Source Pipeline for Semiautomated Analysis of Glycopeptide Mass Spectrometry Data. *Anal Chem* 90, 11908-11916
229. Wopereis, S., Lefeber, D. J., Morava, É., and Wevers, R. A. (2006) Mechanisms in Protein O-Glycan Biosynthesis and Clinical and Molecular Aspects of Protein O-Glycan Biosynthesis Defects: A Review. *Clin. Chem.* 52, 574-600
230. Struwe, W. B., Agravat, S., Aoki-Kinoshita, K. F., Campbell, M. P., Costello, C. E., Dell, A., Ten, F., Haslam, S. M., Karlsson, N. G., Khoo, K. H., Kolarich, D., Liu, Y., McBride, R., Novotny, M. V., Packer, N. H., Paulson, J. C., Rapp, E., Ranzinger, R., Rudd, P. M., Smith, D. F., Tiemeyer, M., Wells, L., York, W. S., Zaia, J., and Kettner, C. (2016) The minimum information required for a glycomics experiment (MIRAGE) project: sample preparation guidelines for reliable reporting of glycomics datasets. *Glycobiology* 26, 907-910
231. Anderson, N. L. (2002) The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol. Cell. Proteomics* 1, 845-867
232. Schaller, J., Gerber, S., Kämpfer, U., Lejon, S., and Trachsel, C. (2008) Blood Plasma Proteins. *Human Blood Plasma Proteins*, pp. 17-20, John Wiley & Sons, Ltd
233. Putnam, F. W. (1984) *The Plasma Proteins*, 2nd Ed., Academic Press, Inc.
234. Anderson, L. (2014) Six decades searching for meaning in the proteome. *J. Proteomics* 107, 24-30
235. Ceciliani, F., and Pocacqua, V. (2007) The acute phase protein alpha1-acid glycoprotein: a model for altered glycosylation during diseases. *Curr. Protein Pept. Sci.* 8, 91-108
236. Polanski, M., and Anderson, N. L. (2007) A list of candidate cancer biomarkers for targeted proteomics. *Biomark Insights* 1, 1-48
237. Anderson, L., and Hunter, C. L. (2006) Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* 5, 573-588
238. Farrah, T., Deutsch, E. W., Omenn, G. S., Campbell, D. S., Sun, Z., Bletz, J. A., Mallick, P., Katz, J. E., Malmstrom, J., Ossola, R., Watts, J. D., Lin, B., Zhang, H., Moritz, R. L., and Aebersold, R.

(2011) A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol. Cell. Proteomics* 10, M110 006353

239. Haab, B. B., Geierstanger, B. H., Michailidis, G., Vitzthum, F., Forrester, S., Okon, R., Saviranta, P., Brinker, A., Sorette, M., Perlee, L., Suresh, S., Drwal, C., Adkins, J. N., and Omenn, G. S. (2005) Immunoassay and antibody microarray analysis of the HUPO Plasma Proteome Project reference specimens: Systematic variation between sample types and calibration of mass spectrometry data. *Proteomics* 5, 3278-3291
240. Hortin, G. L., Sviridov, D., and Anderson, N. L. (2008) High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance. *Clin. Chem.* 54, 1608-1616
241. Schenk, S., Schoenhals, G. J., de Souza, G., and Mann, M. (2008) A high confidence, manually validated human blood plasma protein reference set. *BMC Med Genomics* 1, 41
242. Zhou, M., Lucas, D. A., Chan, K. C., Issaq, H. J., Petricoin Iii, E. F., Liotta, L. A., Veenstra, T. D., and Conrads, T. P. (2004) An investigation into the human serum "interactome". *Electrophoresis* 25, 1289-1298
243. Carrascal, M., Gay, M., Ovelleiro, D., Casas, V., Gelpí, E., and Abian, J. (2010) Characterization of the human plasma phosphoproteome using linear ion trap mass spectrometry and multiple search engines. *J Proteome Res* 9, 876-884
244. Zawadzka, A. M., Schilling, B., Cusack, M. P., Sahu, A. K., Drake, P., Fisher, S. J., Benz, C. C., and Gibson, B. W. (2014) Phosphoprotein secretome of tumor cells as a source of candidates for breast cancer biomarkers in plasma. *Mol. Cell. Proteomics* 13, 1034-1049
245. Kim, Y. J., Zaidi-Ainouch, Z., Gallien, S., and Domon, B. (2012) Mass spectrometry-based detection and quantification of plasma glycoproteins using selective reaction monitoring. *Nat. Protoc.* 7, 859-871
246. Pasing, Y., Sickmann, A., and Lewandrowski, U. (2012) N-glycoproteomics: mass spectrometry-based glycosylation site annotation. *Biol. Chem.* 393, 249-258
247. Lee, H. J., Cha, H. J., Lim, J. S., Lee, S. H., Song, S. Y., Kim, H., Hancock, W. S., Yoo, J. S., and Paik, Y. K. (2014) Abundance-ratio-based semiquantitative analysis of site-specific N-linked glycopeptides present in the plasma of hepatocellular carcinoma patients. *J Proteome Res* 13, 2328-2338
248. Pacchiarotta, T., Hensbergen, P. J., Wuhrer, M., van Nieuwkoop, C., Nevedomskaya, E., Derks, R. J., Schoenmaker, B., Koeleman, C. A., van Dissel, J., Deelder, A. M., and Mayboroda, O. A. (2012) Fibrinogen alpha chain O-glycopeptides as possible markers of urinary tract infection. *J. Proteomics* 75, 1067-1073
249. Gomes, C., Almeida, A., Ferreira, J. A., Silva, L., Santos-Sousa, H., Pinto-de-Sousa, J., Santos, L. L., Amado, F., Schwientek, T., Levery, S. B., Mandel, U., Clausen, H., David, L., Reis, C. A., and Osorio, H. (2013) Glycoproteomic analysis of serum from patients with gastric precancerous lesions. *J Proteome Res* 12, 1454-1466
250. Dube, D. H., and Bertozzi, C. R. (2005) Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* 4, 477-488
251. Rhodes, J., Campbell, B. J., and Yu, L.-G. (2010) Glycosylation and Disease. *eLS*, John Wiley & Sons, Ltd
252. Ju, T., Wang, Y., Aryal, R. P., Lehoux, S. D., Ding, X., Kudelka, M. R., Cutler, C., Zeng, J., Wang, J., Sun, X., Heimburg-Molinaro, J., Smith, D. F., and Cummings, R. D. (2013) Tn and sialyl-Tn antigens, aberrant O-glycomics as human disease markers. *Proteomics Clin Appl* 7, 618-631
253. Yabu, M., Korekane, H., and Miyamoto, Y. (2014) Precise structural analysis of O-linked oligosaccharides in human serum. *Glycobiology* 24, 542-553
254. Jentoft, N. (1990) Why are proteins O-glycosylated? *Trends Biochem. Sci.* 15, 291-294
255. Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Bennett, E. P., Mandel, U., Brunak, S., Wandall, H. H., Levery, S. B., and Clausen, H. (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* 32, 1478-1488
256. Jensen, P. H., Kolarich, D., and Packer, N. H. (2010) Mucin-type O-glycosylation--putting the pieces together. *FEBS J.* 277, 81-94
257. Strous, G. J., and Dekker, J. (1992) Mucin-type glycoproteins. *Crit. Rev. Biochem. Mol. Biol.* 27, 57-92
258. Halim, A., Ruetschi, U., Larson, G., and Nilsson, J. (2013) LC-MS/MS characterization of O-glycosylation sites and glycan structures of human cerebrospinal fluid glycoproteins. *J Proteome Res* 12, 573-584

259. Julenius, K., Molgaard, A., Gupta, R., and Brunak, S. (2005) Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology* 15, 153-164
260. Zauner, G., Koeleman, C. A., Deelder, A. M., and Wuhrer, M. (2010) Protein glycosylation analysis by HILIC-LC-MS of Proteinase K-generated N- and O-glycopeptides. *J Sep Sci* 33, 903-910
261. Nwosu, C. C., Seipert, R. R., Strum, J. S., Hua, S. S., An, H. J., Zivkovic, A. M., German, B. J., and Lebrilla, C. B. (2011) Simultaneous and extensive site-specific N- and O-glycosylation analysis in protein mixtures. *J Proteome Res* 10, 2612-2624
262. Hua, S., Hu, C. Y., Kim, B. J., Totten, S. M., Oh, M. J., Yun, N., Nwosu, C. C., Yoo, J. S., Lebrilla, C. B., and An, H. J. (2013) Glyco-analytical multispecific proteolysis (Glyco-AMP): a simple method for detailed and quantitative Glycoproteomic characterization. *J Proteome Res* 12, 4414-4423
263. Alpert, A. J., Shukla, M., Shukla, A. K., Zieske, L. R., Yuen, S. W., Ferguson, M. A. J., Mehlert, A., Pauly, M., and Orlando, R. (1994) Hydrophilic-interaction chromatography of complex carbohydrates. *J. Chromatogr. A* 676, 191-202
264. Hanisch, F. G. (2012) O-glycoproteomics: Site-specific O-glycoprotein analysis by CID/ETD electrospray ionization tandem mass spectrometry and top-down glycoprotein Sequencing by In-Source Decay MALDI Mass Spectrometry. *Methods Mol. Biol.*, pp. 179-189
265. Alley, W. R., Jr., Mechref, Y., and Novotny, M. V. (2009) Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun Mass Spectrom* 23, 161-170
266. Saba, J., Dutta, S., Hemenway, E., and Viner, R. (2012) Increasing the Productivity of Glycopeptides Analysis by Using Higher-Energy Collision Dissociation-Accurate Mass-Product-Dependent Electron Transfer Dissociation. *Int. J. Proteomics* 2012, 7
267. Singh, C., Zampronio, C. G., Creese, A. J., and Cooper, H. J. (2012) Higher energy collision dissociation (HCD) product ion-triggered electron transfer dissociation (ETD) mass spectrometry for the analysis of N-linked glycoproteins. *J Proteome Res* 11, 4517-4525
268. Dallas, D. C., Martin, W. F., Hua, S., and German, J. B. (2013) Automated glycopeptide analysis--review of current state and future directions. *Brief. Bioinform.* 14, 361-374
269. Wu, S. W., Pu, T. H., Viner, R., and Khoo, K. H. (2014) Novel LC-MS2 product dependent parallel data acquisition function and data analysis workflow for sequencing and identification of intact glycopeptides. *Anal Chem* 86, 5478-5486
270. Kolarich, D., Rapp, E., Struwe, W. B., Haslam, S. M., Zaia, J., McBride, R., Agravat, S., Campbell, M. P., Kato, M., Ranzinger, R., Kettner, C., and York, W. S. (2013) The minimum information required for a glycomics experiment (MIRAGE) project: improving the standards for reporting mass-spectrometry-based glycoanalytic data. *Mol. Cell. Proteomics* 12, 991-995
271. York, W. S., Agravat, S., Aoki-Kinoshita, K. F., McBride, R., Campbell, M. P., Costello, C. E., Dell, A., Feizi, T., Haslam, S. M., Karlsson, N., Khoo, K. H., Kolarich, D., Liu, Y., Novotny, M., Packer, N. H., Paulson, J. C., Rapp, E., Ranzinger, R., Rudd, P. M., Smith, D. F., Struwe, W. B., Tiemeyer, M., Wells, L., Zaia, J., and Kettner, C. (2014) MIRAGE: the minimum information required for a glycomics experiment. *Glycobiology* 24, 402-406
272. Neue, K., Mormann, M., Peter-Katalinic, J., and Pohlentz, G. (2011) Elucidation of glycoprotein structures by unspecific proteolysis and direct nanoESI mass spectrometric analysis of ZIC-HILIC-enriched glycopeptides. *J Proteome Res* 10, 2248-2260
273. Nwosu, C. C., Huang, J., Aldredge, D. L., Strum, J. S., Hua, S., Seipert, R. R., and Lebrilla, C. B. (2013) In-gel nonspecific proteolysis for elucidating glycoproteins: A method for targeted protein-specific glycosylation analysis in complex protein mixtures. *Anal Chem* 85, 956-963
274. Zauner, G., Hoffmann, M., Rapp, E., Koeleman, C. A., Dragan, I., Deelder, A. M., Wuhrer, M., and Hensbergen, P. J. (2012) Glycoproteomic analysis of human fibrinogen reveals novel regions of O-glycosylation. *J Proteome Res* 11, 5804-5814
275. Plomp, R., Hensbergen, P. J., Rombouts, Y., Zauner, G., Dragan, I., Koeleman, C. A. M., Deelder, A. M., and Wuhrer, M. (2014) Site-specific N-glycosylation analysis of human immunoglobulin e. *J Proteome Res* 13, 536-546
276. Watzlawick, H., Walsh, M. T., Yoshioka, Y., Schmid, K., and Brossmer, R. (1992) Structure of the N- and O-glycans of the A-chain of human plasma alpha 2HS-glycoprotein as deduced from the chemical compositions of the derivatives prepared by stepwise degradation with exoglycosidases. *Biochemistry (Mosc.)* 31, 12198-12203

277. Halim, A., Nilsson, J., Ruetschi, U., Hesse, C., and Larson, G. (2012) Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. *Mol. Cell. Proteomics* 11, M111 013649
278. Nicolardi, S., van der Burgt, Y. E., Wuhrer, M., and Deelder, A. M. (2013) Mapping O-glycosylation of apolipoprotein C-III in MALDI-FT-ICR protein profiles. *Proteomics* 13, 992-1001
279. Hagglund, P., Matthiesen, R., Elortza, F., Hojrup, P., Roepstorff, P., Jensen, O. N., and Bunkenborg, J. (2007) An enzymatic deglycosylation scheme enabling identification of core fucosylated N-glycans and O-glycosylation site mapping of human plasma proteins. *J Proteome Res* 6, 3021-3031
280. Chandler, K. B., Brnakova, Z., Sanda, M., Wang, S., Stalnaker, S. H., Bridger, R., Zhao, P., Wells, L., Edwards, N. J., and Goldman, R. (2014) Site-specific glycan microheterogeneity of inter-alpha-trypsin inhibitor heavy chain h4. *J Proteome Res* 13, 3314-3329
281. Olsen, E. H., Rahbek-Nielsen, H., Thogersen, I. B., Roepstorff, P., and Enghild, J. J. (1998) Posttranslational modifications of human inter-alpha-inhibitor: identification of glycans and disulfide bridges in heavy chains 1 and 2. *Biochemistry (Mosc.)* 37, 408-416
282. Flahaut, C., Capon, C., Balduyck, M., Ricart, G., Sautiere, P., and Mizon, J. (1998) Glycosylation pattern of human inter-alpha-inhibitor heavy chains. *Biochem. J.* 333 (Pt 3), 749-756
283. Lottspeich, F., Kellermann, J., Henschen, A., and Muller-Esterl, W. (1986) Amino acid sequence of the light chain of human high molecular mass kininogen. *Adv. Exp. Med. Biol.* 198 Pt A, 85-89
284. Marti, T., Schaller, J., Rickli, E. E., Schmid, K., Kamerling, J. P., Gerwig, G. J., van Halbeek, H., and Vliegenthart, J. F. G. (1988) The N- and O-linked carbohydrate chains of human, bovine and porcine plasminogen. Species specificity in relation to sialylation and fucosylation patterns. *Eur. J. Biochem.* 173, 57-63
285. Escribano, J., Lopex-Otin, C., Hjerpe, A., Grubb, A., and Mendez, E. (1990) Location and characterization of the three carbohydrate prosthetic groups of human protein HC. *FEBS Lett.* 266, 167-170
286. Nesvizhskii, A. I., and Aebersold, R. (2005) Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* 4, 1419-1440
287. Schwartz, B. L., and Bursey, M. M. (1992) Some proline substituent effects in the tandem mass spectrum of protonated pentaalanine. *Biol. Mass Spectrom.* 21, 92-96
288. Mormann, M., Paulsen, H., and Peter-Katalinic, J. (2005) Electron capture dissociation of O-glycosylated peptides: radical site-induced fragmentation of glycosidic bonds. *Eur J Mass Spectrom (Chichester, Eng)* 11, 497-511
289. Chamot-Rooke, J., van der Rest, G., Dalleu, A., Bay, S., and Lemoine, J. (2007) The combination of electron capture dissociation and fixed charge derivatization increases sequence coverage for O-glycosylated and O-phosphorylated peptides. *J. Am. Soc. Mass Spectrom.* 18, 1405-1413
290. Kristiansen, T. Z., Bunkenborg, J., Gronborg, M., Molina, H., Thuluvath, P. J., Argani, P., Goggins, M. G., Maitra, A., and Pandey, A. (2004) A proteomic analysis of human bile. *Mol. Cell. Proteomics* 3, 715-728
291. Deshpande, N., Jensen, P. H., Packer, N. H., and Kolarich, D. (2010) GlycoSpectrumScan: fishing glycopeptides from MS spectra of protease digests of human colostrum sIgA. *J Proteome Res* 9, 1063-1075
292. Wuhrer, M., Deelder, A. M., and van der Burgt, Y. E. (2011) Mass spectrometric glycan rearrangements. *Mass Spectrom. Rev.* 30, 664-680
293. Wuhrer, M., Koeleman, C. A., and Deelder, A. M. (2009) Hexose rearrangements upon fragmentation of N-glycopeptides and reductively aminated N-glycans. *Anal Chem* 81, 4422-4432
294. Ikezu, S., and Ikezu, T. (2014) Tau-tubulin kinase. *Front Mol Neurosci* 7, 33
295. Gotz, J., Gladbach, A., Pennanen, L., van Eersel, J., Schild, A., David, D., and Ittner, L. M. (2010) Animal models reveal role for tau phosphorylation in human disease. *Biochim. Biophys. Acta* 1802, 860-871
296. Böhm, D., Keller, K., Pieter, J., Boehm, N., Wolters, D., Siggelkow, W., Lebrecht, A., Schmidt, M., Kölbl, H., Pfeiffer, N., and Grus, F.-H. (2012) Comparison of tear protein levels in breast cancer patients and healthy controls using a de novo proteomic approach. *Oncol. Rep.* 28, 429-438
297. Tagliabracci, V. S., Pinna, L. A., and Dixon, J. E. (2013) Secreted protein kinases. *Trends Biochem. Sci.* 38, 121-130

298. Goetz, Sarah C., Liem Jr, Karel F., and Anderson, Kathryn V. (2012) The Spinocerebellar Ataxia-Associated Gene Tau Tubulin Kinase 2 Controls the Initiation of Ciliogenesis. *Cell* 151, 847-858
299. Stenmark, H. (2009) Rab GTPases as coordinators of vesicle traffic. *Nat. Rev. Mol. Cell Biol.* 10, 513-525
300. Bai, X., Li, D., Zhu, J., Guan, Y., Zhang, Q., and Chi, L. (2015) From individual proteins to proteomic samples: characterization of O-glycosylation sites in human chorionic gonadotropin and human-plasma proteins. *Anal. Bioanal. Chem.* 407, 1857-1869
301. Alvarez-Manilla, G., Atwood, Guo, Y., Warren, N. L., Orlando, R., and Pierce, M. (2006) Tools for Glycoproteomic Analysis: Size Exclusion Chromatography Facilitates Identification of Tryptic Glycopeptides with N-linked Glycosylation Sites. *J. Proteome Res.* 5, 701-708
302. Canis, K., McKinnon, T. A. J., Nowak, A., Panico, M., Morris, H. R., Laffan, M., and Dell, A. (2010) The plasma von Willebrand factor O-glycome comprises a surprising variety of structures including ABH antigens and disialosyl motifs. *J. Thromb. Haemost.* 8, 137-145
303. Takahashi, K., Wall, S. B., Suzuki, H., Smith, A. D. t., Hall, S., Poulsen, K., Kilian, M., Mobley, J. A., Julian, B. A., Mestecky, J., Novak, J., and Renfrow, M. B. (2010) Clustered O-glycans of IgA1: defining macro- and microheterogeneity by use of electron capture/transfer dissociation. *Mol. Cell. Proteomics* 9, 2545-2557
304. Darula, Z., and Medzihradszky, K. F. (2009) Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum. *Mol. Cell. Proteomics* 8, 2515-2526
305. Takayasu, T., Suzuki, S., Kametani, F., Takahashi, N., Shinoda, T., Okuyama, T., and Munekata, E. (1982) Amino acid sequence of galactosamine-containing glycopeptides in the hinge region of a human immunoglobulin D. *Biochem. Biophys. Res. Commun.* 105, 1066-1071
306. Plomp, R., Dekkers, G., Rombouts, Y., Visser, R., Koeleman, C. A. M., Kammeijer, G. S. M., Jansen, B. C., Rispens, T., Hensbergen, P. J., Vidarsson, G., and Wuhrer, M. (2015) Hinge-Region O-Glycosylation of Human Immunoglobulin G3 (IgG3). *Mol. Cell. Proteomics* 14, 1373-1384
307. Zhu, Z., Su, X., Clark, D. F., Go, E. P., and Desaire, H. (2013) Characterizing O-Linked glycopeptides by electron transfer dissociation: Fragmentation rules and applications in data analysis. *Anal Chem* 85, 8403-8411
308. Good, D. M., Wirtala, M., McAlister, G. C., and Coon, J. J. (2007) Performance characteristics of electron transfer dissociation mass spectrometry. *Mol. Cell. Proteomics* 6, 1942-1951
309. Liu, H., Zhang, J., Sun, H., Xu, C., Zhu, Y., and Xie, H. (2011) The Prediction of Peptide Charge States for Electrospray Ionization in Mass Spectrometry. *Procedia Environmental Sciences* 8, 483-491
310. Falick, A. M., Hines, W. M., Medzihradszky, K. F., Baldwin, M. A., and Gibson, B. W. (1993) Low-mass ions produced from peptides by high-energy collision-induced dissociation in tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* 4, 882-893
311. Papayannopoulos, I. A. (1995) The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrom. Rev.* 14, 49-73
312. Dong, Q., Yan, X., Liang, Y., and Stein, S. E. (2016) In-Depth Characterization and Spectral Library Building of Glycopeptides in the Tryptic Digest of a Monoclonal Antibody Using 1D and 2D LC-MS/MS. *J Proteome Res* 15, 1472-1486
313. Stadlmann, J., Taubenschmid, J., Wenzel, D., Gattinger, A., Durnberger, G., Dusberger, F., Elling, U., Mach, L., Mechtler, K., and Penninger, J. M. (2017) Comparative glycoproteomics of stem cells identifies new players in ricin toxicity. *Nature* 549, 538-542
314. Breci, L. A., Tabb, D. L., Yates, J. R., and Wysocki, V. H. (2003) Cleavage N-Terminal to Proline: Analysis of a Database of Peptide Tandem Mass Spectra. *Anal. Chem.* 75, 1963-1971
315. Darula, Z., and Medzihradszky, K. F. (2015) Carbamidomethylation Side Reactions May Lead to Glycan Misassignments in Glycopeptide Analysis. *Anal. Chem.* 87, 6297-6302
316. Vidarsson, G., Dekkers, G., and Rispens, T. (2014) IgG subclasses and allotypes: from structure to effector functions. *Frontiers in immunology* 5, 520
317. Hoffmann, M., Pioch, M., Pralow, A., Hennig, R., Kottler, R., Reichl, U., and Rapp, E. (2018) The Fine Art of Destruction: A Guide to In-Depth Glycoproteomic Analyses - Exploiting the Diagnostic Potential of Fragment Ions. *Proteomics*, e1800282
318. Clerc, F., Reiding, K. R., Jansen, B. C., Kammeijer, G. S. M., Bondt, A., and Wuhrer, M. (2016) Human plasma protein N-glycosylation. *Glycoconj. J.* 33, 309-343

319. Liu, T., Qian, W. J., Gritsenko, M. A., Camp Ii, D. G., Monroe, M. E., Moore, R. J., and Smith, R. D. (2005) Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J Proteome Res* 4, 2070-2080
320. Solecka, B. A., Weise, C., Laffan, M. A., and Kannicht, C. (2016) Site-specific analysis of von Willebrand factor O-glycosylation. *J. Thromb. Haemost.* 14, 733-746
321. Karav, S., German, J. B., Rouquie, C., Le Parc, A., and Barile, D. (2017) Studying Lactoferrin N-Glycosylation. *International journal of molecular sciences* 18, 870
322. van Berkel, P. H., van Veen, H. A., Geerts, M. E., de Boer, H. A., and Nuijens, J. H. (1996) Heterogeneity in utilization of N-glycosylation sites Asn624 and Asn138 in human lactoferrin: a study with glycosylation-site mutants. *Biochem. J.* 319 (Pt 1), 117-122
323. Wuhrer, M., de Boer, A. R., and Deelder, A. M. (2009) Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry. *Mass Spectrom. Rev.* 28, 192-206
324. Jiang, K., Zhu, H., Xiao, C., Liu, D., Edmunds, G., Wen, L., Ma, C., Li, J., and Wang, P. G. (2017) Solid-phase reductive amination for glycomic analysis. *Anal. Chim. Acta* 962, 32-40
325. Wuhrer, M., Koeleman, C. A., Hokke, C. H., and Deelder, A. M. (2006) Mass spectrometry of proton adducts of fucosylated N-glycans: fucose transfer between antennae gives rise to misleading fragments. *Rapid Commun Mass Spectrom* 20, 1747-1754
326. Acs, A., Ozohanics, O., Vekey, K., Drahos, L., and Turiak, L. (2018) Distinguishing Core and Antenna Fucosylated Glycopeptides Based on Low-Energy Tandem Mass Spectra. *Anal Chem*
327. Medzihradszky, K. F., Kaasik, K., and Chalkley, R. J. (2015) Characterizing Sialic Acid Variants at the Glycopeptide Level. *Anal. Chem.* 87, 3064-3071
328. Halim, A., Westerlind, U., Pett, C., Schorlemer, M., Rüetschi, U., Brinkmalm, G., Sihlbom, C., Lengqvist, J., Larson, G., and Nilsson, J. (2014) Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. *J. Proteome Res.* 13, 6024-6032
329. Pett, C., Nasir, W., Sihlbom, C., Olsson, B. M., Caixeta, V., Schorlemer, M., Zahedi, R. P., Larson, G., Nilsson, J., and Westerlind, U. (2018) Effective Assignment of alpha2,3/alpha2,6-Sialic Acid Isomers by LC-MS/MS-Based Glycoproteomics. *Angew. Chem. Int. Ed. Engl.* 57, 9320-9324

List of Contributions

Publications

1. Hoffmann, M., Pioch, M., Pralow, A., Hennig, R., Kottler, R., Reichl, U., and Rapp, E. (2018) The Fine Art of Destruction: A Guide to In-Depth Glycoproteomic Analyses - Exploiting the Diagnostic Potential of Fragment Ions. Proteomics 18, 1800282

2. Pioch, M., Hoffmann, M., Pralow, A., Reichl, U., and Rapp, E. (2018) glyXtool(MS): An Open-Source Pipeline for Semiautomated Analysis of Glycopeptide Mass Spectrometry Data. Anal Chem 90, 11908-11916

3. Pralow, A., Hoffmann, M., Nguyen-Khuong, T., Rapp, E., and Reichl, U. (2017) Improvement of the glycoproteomic toolbox with the discovery of a unique C-terminal cleavage specificity of flavastacin for N-glycosylated asparagine. Sci Rep 7, 11419

4. Hennig, R., Cajic, S., Borowiak, M., Hoffmann, M., Kottler, R., Reichl, U., and Rapp, E. (2016) Towards personalized diagnostics via longitudinal study of the human plasma N-glycome. Biochim. Biophys. Acta 1860, 1728-1738

5. Heyer, R., Benndorf, D., Kohrs, F., De Vrieze, J., Boon, N., Hoffmann, M., Rapp, E., Schluter, A., Sczyrba, A., and Reichl, U. (2016) Proteotyping of biogas plant microbiomes separates biogas plants according to process temperature and reactor type. Biotechnol Biofuels 9, 155

6. Hoffmann, M., Marx, K., Reichl, U., Wuhrer, M., and Rapp, E. (2016) Site-specific O-Glycosylation Analysis of Human Blood Plasma Proteins. Mol. Cell. Proteomics 15, 624-641

7. Kohrs, F., Wolter, S., Benndorf, D., Heyer, R., Hoffmann, M., Rapp, E., Bremges, A., Sczyrba, A., Schluter, A., and Reichl, U. (2015) Fractionation of biogas plant sludge material improves metaproteomic characterization to investigate metabolic activity of microbial communities. Proteomics 15, 3585-3589

8. Muth, T., Behne, A., Heyer, R., Kohrs, F., Benndorf, D., Hoffmann, M., Lehtevä, M., Reichl, U., Martens, L., and Rapp, E. (2015) The MetaProteomeAnalyzer: A Powerful Open-Source Software Suite for Metaproteomics Data Analysis and Interpretation. J. Proteome Res. 14, 1557-1565

9. Bakovic, M. P., Selman, M. H., Hoffmann, M., Rudan, I., Campbell, H., Deelder, A. M., Lauc, G., and Wuhrer, M. (2013) High-throughput IgG Fc N-glycosylation profiling by mass spectrometry of glycopeptides. J Proteome Res 12, 821-831

10. Selman, M. H., Hoffmann, M., Zauner, G., McDonnell, L. A., Balog, C. I., Rapp, E., Deelder, A. M., and Wuhrer, M. (2012) MALDI-TOF-MS analysis of sialylated glycans

and glycopeptides using 4-chloro-alpha-cyanocinnamic acid matrix. Proteomics 12, 1337-1348

11. Kluge, S.*, Hoffmann, M.*, Benndorf, D., Rapp, E., and Reichl, U. (2012) Proteomic tracking and analysis of a bacterial mixed culture. Proteomics 12, 1893-1901

12. Zauner, G.*, Hoffmann, M. *, Rapp, E., Koeleman, C. A., Dragan, I., Deelder, A. M., Wuhrer, M., and Hensbergen, P. J. (2012) Glycoproteomic analysis of human fibrinogen reveals novel regions of O-glycosylation. J Proteome Res 11, 5804-5814

*both authors contributed equally.

Supervision of Students

Alexander Pralow: „Glykoproteomstudie zur Charakterisierung der Glykosylierung von Lactotransferrin aus Humanmilch mittels nanoRP-LC-ESI-IT-MSn". (2015) Master Thesis, Faculty for Process and Systems Engineering, Otto von Guericke University, Magdeburg.

Talks

1. Hoffmann, M., Grote, V., Bartels, M., Hennig, R., Pioch, M., Reichl, U., Strahl, S., & Rapp, E. (2018). Towards Understanding of Congenital Disorders of Glycosylation: A Glyco(proteo)mic Approach to Unravel the Interplay between Different Types of Glycosylation. Talk presented at DFG FOR 2509 Meeting. Höchst (Odenwald), Germany.

2. Hoffmann, M., Bartels, M., Hennig, R., Pioch, M., Reichl, U., Strahl, S., & Rapp, E. (2017). Unravelling the Interplay between Protein O-Mannosylation and N-Glycosylation – An N-Glycomic and N-Glycoproteomic Study of Human N-Cadherin. Talk presented at DFG FOR 2509 Kick-off Meeting. Heidelberg, Germany.

3. Hoffmann, M., Pioch, M., Pralow, A., Rapp, E., Reichl, U. (2016). Human Blood Plasma N- and O-Glycoproteomics. Talk presented at the IVT Colloquium. Otto von Guericke University, Magdeburg, Germany.

4. Hoffmann, M., Marx, K., Wuhrer, M., Reichl, U., & Rapp, E. (2015). Human Blood Plasma O-Glycoproteomics. Talk presented at the HighGlycan Meeting. Berlin, Germany.

5. Hoffmann, M., Marx, K., Wuhrer, M., Reichl, U., & Rapp, E. (2015). Human Blood Plasma O-Glycoproteomics. Talk presented at GLYCO 23: XXIII International Symposium on Glycoconjugates. Split, Croatia.

6. Hoffmann, M. (2014). Human Blood Plasma O-Glycoproteomics. Talk presented at Bruker Anwendertreffen 2014. Berlin, Germany.

Poster

1. Hoffmann, M., Pioch, M., Pralow, A., Hennig, R., Kottler, R., Reichl, U., & Rapp, E. (2019). The Fine Art of Destruction: A Guide to In-Depth Glycoproteomic Analyses - Exploiting the Diagnostic Potential of Fragment Ions. Poster presented at GlycoBioTec 2019, Berlin, Germany.

2. Hoffmann, M., Bartels, M., Noor, S., Hennig, R., Pioch, M., Reichl, U., Strahl, S., & Rapp, E. (2018). Unravelling the Interplay between Protein O-Mannosylation and N-Glycosylation - A Glycomic and Glycoproteomic Study of Human N-Cadherin. Poster presented at DFG FOR 2509 Meeting, Höchst (Odenwald), Germany.

3. Hoffmann, M., Bartels, M., Hennig, R., Pioch, M., Reichl, U., Strahl, S., & Rapp, E. (2017). Unravelling the Interplay between Protein O-Mannosylation and N-Glycosylation-An N-Glycomic and N-Glycoproteomic Study of Human N-Cadherin. Poster presented at DFG FOR 2509 Kick-off Meeting, Heidelberg, Germany.

4. Hoffmann, M. (2017). Unravelling the Correlation between Different ER-Based Types of Glycosylation via Glyco/Proteo-Analytics. Poster presented at DFG FOR 2509 Kick-off meeting, Heidelberg, Germany.

5. Hoffmann, M., Pioch, M., Pralow, A., Nguyen-Khuong, T., Hennig, R., Kottler, R., Reichl, U., & Rapp, E. (2017). Site-Specific N-and O-Glycoproteomic Analysis Using Stepped Collisional Energy Fragmentation on a Linear Ion trap/Orbitrap Hybrid Mass Spectrometer. Poster presented at GlycoBioTec Symposium 2017, Berlin, Germany.

6. Hoffmann, M., Zauner, G., Selman, M., Wuhrer, M., Reichl, U., & Rapp, E. (2014). Site-specific O-Glycosylation Analysis of Human Blood Plasma Proteins Using Multistage Mass Spectrometry with Collision Induced Dissociation (CID) and Electron Transfer Dissociation (ETD). Poster presented at 25th Joint Glycobiology Meeting, Ghent, Belgium.

7. Hoffmann, M., Zauner, G., Rapp, E., Koeleman, C., Dragan, I., Deelder, A., Hensbergen, P., & Wuhrer, M. (2013). Glycoproteomic Analysis of Human Fibrinogen Reveals Novel Regions of O-Glycosylation. Poster presented at Proteomics Forum, Berlin, Germany.

8. Hoffmann, M. (2013). Glycoproteomic Analysis of Human Fibrinogen Reveals Novel Regions of O-Glycosylation. Poster presented at 24th Joint Glycobiology Meeting, Wittenberg, Germany.

9. Selman, M., Hoffmann, M., Zauner, G., McDonnell, L., Balog, C., Rapp, E., Deelder, A., & Wuhrer, M. (2012). MALDI-TOF-MS analysis of sialylated glycans and glycopeptides using 4-chloro-α-cyanocinnamic acid matrix. Poster presented at 6th Glycan Forum, Berlin, Germany.

10. Selman, M., Hoffmann, M., Zauner, G., McDonnell, L., Balog, C., Rapp, E., Deelder, A., & Wuhrer, M. (2012). MALDI-TOF-MS analysis of sialylated glycans and

glycopeptides using 4-chloro-α-cyanocinnamic acid matrix. Poster presented at 23rd Joint Glycobiology Meeting, Wageningen, The Netherlands.

List of Figures

Figure 1: Classes of glycoconjugates present in vertebrates 2
Figure 2: One glycoprotein may exhibit different glycoforms 3
Figure 3: Graphical representation of glycan structures 9
Figure 4: Frequency of occurrence of amino acids in the vicinity of *N*-glycosylation sites 15
Figure 5: *N*-glycan types present in mammals 16
Figure 6: Exemplary *N*-glycan structure found in humans 16
Figure 7: History of glycans in medicine 18
Figure 8: *N*-glycan epitopes causing immunogenic reactions in humans 18
Figure 9: *O*-glycan core structures 20
Figure 10: Frequency of occurrence of amino acids in the vicinity of *O*-glycosylation sites 21
Figure 11: Evolutionary conservation, informational diversity, and knowledge base of the cellular building blocks 22
Figure 12: Modeled protein structure of human IgG 24
Figure 13: Effects of *N*-glycans on therapeutic IgGs 24
Figure 14: Thickness of the endothelial glycocalyx of murine cerebral capillaries 25
Figure 15: Overview of glycoanalytical approaches 32
Figure 16: Generic workflow for the LC-ESI-MS/MS-based analysis of *N*- and *O*-glycoproteins 37
Figure 17: Nomenclature for ions generated during fragmentation 40
Figure 18: The proteomics conflict 43
Figure 19: Separation of *O*-glycopeptides via nano-RP-LC 48
Figure 20: *O*-glycoproteomic workflow for the analysis of human blood plasma glycoproteins. 51
Figure 21: Spin-Cotton-HILIC-SPE 53
Figure 22: Glycoproteomic workflow developed for the in-depth analysis of *N*- and *O*-glycosylated proteins 56
Figure 23: Relative proportion of major human blood plasma proteins 59
Figure 24: Comprehensive identification and characterization of mucin-type *O*-glycopeptides 61
Figure 25: Base peak chromatograms (BPC) of five proteinase K digests 62
Figure 26: Number of detected glycopeptides in the HILIC fractions #13-#17. 62
Figure 27: Extracted ion chromatograms (EICs) of diagnostic glycan oxonium ions 63
Figure 28: Exemplary CID fragment ion spectra of a proteinase K-generated plasminogen *O*-glycopeptide 64
Figure 29: Box-Whisker-plot of the peptide length obtained with proteinase K 71
Figure 30: Number of human blood plasma proteins identified with CID and ETD. 71
Figure 31: The use of proteinase K can facilitate identification of *O*-glycosylation sites. 72
Figure 32: Exemplary ETD fragment ion spectrum of a proteinase K-generated plasminogen *O*-glycopeptide 73
Figure 33: Preferred C-terminal cleavage sites of proteinase K 80
Figure 34: Glycopeptide fragmentation with *HCD.low* and *HCD.step* 87
Figure 35: Conserved *N*-glycopeptide fragmentation pattern 89
Figure 36: Glycopeptide fragmentation with *HCD.high* 90
Figure 37: Glycopeptide enrichment via cotton-HILIC-SPE 94
Figure 38: Relative abundance of human IgG 1 and 2 Fc glycoforms (microheterogeneity) 96
Figure 39: Relative abundance of bovine Rnase B *N*-glycoforms (microheterogeneity) 99
Figure 40: Relative quantitation of *N*- and *O*-glycopeptide oxonium ions 103
Figure 41: Antenna GlcNAc vs bisecting GlcNAc 106
Figure 42: Discrimination of *N*- and *O*-glycopeptides 110

Please note, if not stated otherwise, all supplemental figures can be found in the electronic version of this thesis.

List of Tables

Table 1: Site-specific *O*-glycan composition of identified human blood plasma glycoproteins..... 66
Table 2: Detailed overview of all identified human blood plasma *O*-glycopeptides..... 68
Table 3: Diagnostic potential of oxonium ions.....111

List of Materials and Chemicals

1

1,4-dithiothreitol · BioUltra; #43815; Sigma Aldrich, Steinheim, Germany

2

2,2,2-trifluoroethanol · #808259; Merck, Darmstadt, Germany

A

Acetonitrile · LC-MS CHROMASOLV®; # 34967; Sigma Aldrich, Steinheim, Germany
Ammonium bicarbonate · BioUltra; #09830; Sigma Aldrich, Steinheim, Germany
Ammonium formate · BioUltra; #09735; Sigma Aldrich, Steinheim, Germany

C

C18 µ-precolumn · Acclaim PepMap100, C18, 5 µm, 100 Å, 300 µm i.d. x 5 mm; Thermo Scientific, Dreieich, Germany
C18 nano pre-column · Nano Trap Column, packed with Acclaim PepMap100 C18, 5 µm, 100Å, 100 µm i.d. x 2 cm; #164564; Thermo Scientific, Waltham, MA, USA
C18 nano separation column (25 cm) · Acclaim PepMap RSLC C18, 2 µm, 100Å, 75 µm i.d. x 25 cm; #164536; Thermo Scientific, Waltham, MA, USA
C18 nano-separation column (15 cm) · Acclaim PepMap RSLC, C18, 2 µm, 100 Å, 75 µm i.d. x 15 cm; Thermo Scientific, Dreieich, Germany, Acclaim PepMap RSLC, C18, 2 µm, 100 Å, 75 µm i.d. x 15 cm; Thermo Scientific, Dreieich, Germany
Calcium chloride · #A4689; AppliChem, Darmstadt, Germany
Centrifugal filter unit · Nanosep® Omega™ with polyethersulfone membrane, molecular weight cut-off 10 kDa; # OD010c34, PALL Life Sciences, Ann Arbor, MI, USA
Cotton pads · Wattepads; ebelin, dm-drogerie markt GmbH & Co. KG, Karlsruhe, Germany

E

Erythropoietin · Glycosylated Erythropoietin, #PS-500-1; Protea Biosciences, Morgantown, WV, USA

F

Fibrinogen · #F3879; Sigma-Aldrich, Steinheim, Germany
Formic acid · Reagent grade; #F0507; Sigma Aldrich, Steinheim, Germany

H

HILIC column · ACQUITY UPLC BEH HILIC Column; 130Å, 1.7 µm, 2.1 mm X 100 mm; Waters, Manchester, UK
Human blood plasma · VisuCon-F, frozen normal control blood; #FRNCP0125; Ancaster, ON, Canada

I

Immunoglobulin gamma · #I4506-10MG; Sigma-Aldrich, Steinheim, Germany
Iodoacetamide · BioUltra; #I1149; Sigma Aldrich, Steinheim, Germany

L

Lactotransferrin · #L4894-5MG; Sigma-Aldrich, Steinheim, Germany

M

Milli-Q water · purified water, Milli-Q water purification system; Merck Millipore, Darmstadt, Germany
Milli-Q water MS · ultrapure water, Milli-Q water purification system; Merck Millipore, Darmstadt, Germany

N

Nanoelectrospray emitter · SilicaTip; FS360-20-10-D-20; New Objective, Cambridge, USA

P

Phosphate-buffered saline · Prepared according to SOP M/01
proteinase K · #A4392; AppliChem, Darmstadt, Germany
Proteinase K · BioUltra; #P2308; Sigma Aldrich, Steinheim, Germany

R

Ribonuclease B · #R7884-100MG; Sigma-Aldrich, Steinheim, Germany

T

Trifluoroacetic acid · ReagentPlus; #T6508; Sigma Aldrich, Steinheim, Germany
Tris-HCl · #A3452, AppliChem, Darmstadt, Germany
Trypsin · Sequencing Grade Modified Trypsin; # V5111; Promega, Madison, WI, USA

U

Urea · #A1049, AppliChem, Darmstadt, Germany

List of Devices and Instruments

C

Centrifuge 5424 · Eppendorf, Hamburg, Germany
Centrifuge 5804 R · Eppendorf, Hamburg, Germany

I

Ion trap mass spectrometer · AmaZon ETD; Bruker Daltonics, Bremen, Germany

L

LC-Pak Polisher · Merck Millipore, Darmstadt, Germany
linear ion trap-orbitrap hybrid mass spectrometer · LTQ Orbitrap Elite hybrid mass spectrometer; Thermo Scientific, Bremen, Germany

M

Microcentrifuge · Heraeus™ Fresco™ 17; 24 x 1.5/2.0 mL rotor, Thermo Scientific, Waltham, MA, USA
Milli-Q water purification system · Merck Millipore, Darmstadt, Germany

N

Nanoelectrospray source · Nanospray Flex™ ion source, NSI; Thermo Scientific, Waltham, MA, USA
nanoFlow ESI Sprayer · Bruker Daltonics, Bremen, Germany
NanoHPLC-System · UltiMate 3000; Thermo Scientific, Dreieich, Germany
nanoUPLC system · UltiMate 3000 RSLCnano; Thermo Scientific, Dreieich, Germany

T

Temperature-controlled incubator · Titramax 1000 + Inkubator 1000; Heidolph, Schwabach, Germany
Thermomixer · Thermomixer comfort; Eppendorf, Hamburg, Germany

V

Vacuum centrifuge · RVC 2-33 CDplus, ALPHA 2-4 LDplus; Martin Christ GmbH, Osterode am Harz, Germany

List of Software Programs

B

BioTools software 3.2 · Bruker Daltonics, Bremen, Germany

D

DataAnalysis software 4.0 · Bruker Daltonics, Bremen, Germany

F

FindPept · https://web.expasy.org/findpept/

G

glyXtool^{MS} · In-house software, developed by Pioch *et al.*

M

MASCOT version 2.2.04 · Matrix Science, London, UK
MS-Homology · http://prospector.ucsf.edu/prospector
MS-Product · http://prospector.ucsf.edu/prospector/

N

NetOGlyc 4.0 · http://www.cbs.dtu.dk/services/NetOGlyc/

P

ProteinScape 3.1 · Bruker Daltonics, Bremen, Germany
Proteome Discoverer 1.4.1.14 · Thermo Scientific, Waltham, MA, USA