New Methods for Characterization of *N*-type Glycosylation of Proteins by Integration of LC-MS/MS and NMR

Thesis

Submitted in fulfillment of the requirements
of the degree Doctor rer. nat.
at the Department of Chemistry,
Faculty of Sciences, University of Hamburg

by

Alena Wiegandt

from Hamburg

Hamburg, 2016

This thesis was prepared at the Institute for Organic Chemistry from October 2012 to July 2016 (managing director: Prof. Dr. C.B.W. Stark).

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

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

1st Reviewer:	Prof. Dr. Bernd Meyer
2nd Reviewer:	Prof. Dr. Dr. h.c. mult. Wittko Francke
Date of defense:	07.04.2017

Outline

ABBREVIATIONS

1D	one-dimensional
3DCC	three-dimensional cross correlation
aa	amino acid
ACN	acetonitrile
ADCC	antibody-dependent cellular cytotoxicity
AFP	α-Fetoprotein
AIEX	anion exchange chromatography
BPC	base peak chromatogram
CDG	congenital disorder of glycosylation
CCSD	complex carbohydrate structure database
CFG	Consortium for Functional Glycomics
CHO	chinese hamster ovary
CID	collision-induced dissociation
COSY	correlation spectroscopy
CRC	colorectal cancer
CRD	carbohydrate-recognition domain
CV	column volume
DAC	diammonium citrate
DHAP	2,5-dihydroxyacetophenone
DHB	2,5-dihydroxybenzoic acid
DLS	dynamic light scattering
DTT	1,4-D/L-dithiothreitol
EDC	extracted delta chromatogram
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
EIC	extracted ion chromatogram
ER	endoplasmic reticulum
ESI	electrospray ionization
FA	formic acid
GBP	glycan-binding protein

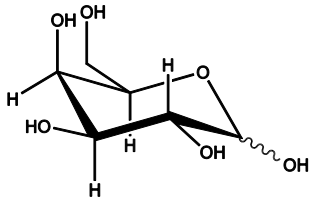
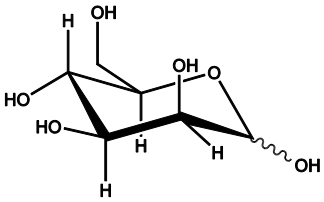
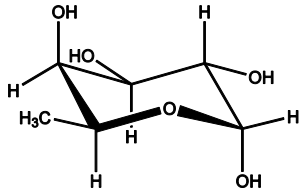
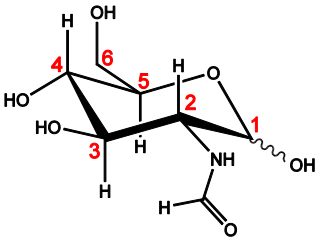
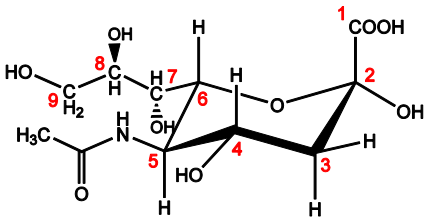
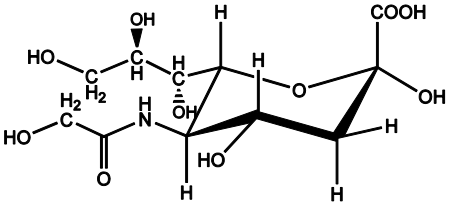
GCDFP-15	gross cystic disease fluid protein 15
gp17	glycoprotein 17
hc	heavy chain
HCC	hepatocellular cancer
HPLC	high performance liquid chromatography
HR	high resolution
HRG	histidine-rich glycoprotein
IAA	iodoacetamide
IEX	ion exchange chromatography
IgG	immunoglobulin G
IMAC	immobilized metal affinity chromatography
ISCID	in-source collision-induced dissociation
IU	international unit
LB	line broadening
LC	liquid chromatography
lc	light chain
LID	laser-induced dissociation
mAb	monoclonal antibody
MAF	minor allele frequency
MALDI	matrix-assisted laser desorption/ionization
MEM	maximum entropy method
MS	mass spectrometry
MW	molecular weight
MWCO	molecular weight cut-off
m/z	mass-to-charge ratio
NIST	National Institute for Standards and Technology
NMR	nuclear magnetic resonance
NS	number of scans
NTA	nitrilotriacetic acid
PGC	porous graphitized carbon
PIP	prolactin-inducible protein
PNGase F	protein *N*-glycosidase F
PRL	prolactin

PSA	prostate specific antigen
PTM	post-translational modification
rcf	relative centrifugal force
RNA	ribonucleic acid
RP	reversed phase
SABP	secretory actin-binding protein
SEC	size exclusion chromatography
S/N	signal-to-noise ratio
SNP	single nucleotide polymorphism
SPE	solid phase extraction
SVD	singular value decomposition
TA30	30:70 [v/v] acetonitrile:water/0.1% TFA
TFA	trifluoroacetic acid
TIC	total ion chromatogram
TOCSY	total correlated spectroscopy
TOF	time of flight
t_r	retention time
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
UHPLC	ultra high performance liquid chromatography

Amino acids

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

Monosaccharide units

Residue	Abbreviation	CFG notation	Structure
α-/β-D-galactose	Gal		
α-/β-D-mannose	Man		
α-L-fucose	Fuc		
α-/β-D-*N*-acetyl glucosamine	GlcNAc		
α-D-*N*-acetyl neuraminic acid	NeuAc		
α-D-*N*-glycolyl neuraminic acid	NeuGc		

1 SUMMARY

As one of the most important post-translational modifications of biomolecules, glycosylation significantly affects a variety of biological functions. Thus, the correlation between altered glycan structures and pathogenic or immunogenic processes is not surprising, often resulting in a specific marker for such processes. The analysis of the glycome is among others especially relevant within two biomedical areas that were therefore addressed in this thesis: A) quality control of biopharmaceuticals and B) development of biomarkers for improved diagnostics.

Recent studies have revealed, that Cetuximab, a monoclonal antibody used for treatment of cancer, causes severe, even deadly, allergic immune reactions due to anaphylactic shocks. In this work, non-human glycosylation epitopes that cause these fatal reactions were unambiguously identified for the first time. From 37 *N*-glycan compositions that were identified by a bottom up approach, the ten most abundant ones were structurally characterized by developing an optimized version of the three-dimensional cross correlation (3DCC) that mathematically combines LC-MS with NMR data. 24% of all detected glycans possess the immunogenic α-(1→3)-Gal epitope and/or *N*-glycolyl neuraminic acid that can cause severe immune reactions.

Complementary to this detailed but demanding analysis, a bioinformatic tool for a fast glycosylation analysis on the intact glycoprotein level was developed. It is based on a least-square-optimized fit of the isotope-resolved experimental MS spectrum by theoretical glycoform spectra. The most challenging aspect was the implementation of multiple glycosylation sites per molecule. The determination of the attached *N*-glycans allows conclusions about safety as well as efficiency of recombinant mAbs. A middle up approach, analyzing mAb sub-domains, was established in order to enable a site-specific analysis of glycosylation and localization of potentially inappropriate glycan epitopes. By use of the developed tool, different batches of commercially available Bevacizumab, a mAb used in therapy of various cancer diseases, were analyzed regarding their glycosylation profile. It was possible to identify distinct variations of glycosylation between batches that are associated with biopharmaceutical efficiency. In an

interlaboratory study organized by the National Institute of Standards and Technology (NIST), the multi-dimensional approach developed here, combining bottom up as well as intact glycoprotein analyses, was applied. Statistical analysis of data from all submitted reports (n = 93) by NIST confirmed that this approach yields excellent results.

Prolactin-Inducible Protein (PIP), a glycoprotein found in human secretions, has been demonstrated to be involved in important biological functions, such as tumor progression. Glycan structures attached to PIP have not been studied in detail so far. Therefore, PIP isolated from saliva was analyzed by LC-MS/MS and NMR in this thesis. It was found that PIP holds a very unusual *N*-glycosylation characterized by an extremely high degree of fucosylation. NMR analyses revealed highly fucosylated *N*-glycan epitopes as Lewis^y (Le^y) structures, that are typically up-regulated in tumors and important for prognosis. These results represent the first unambiguous characterization of *N*-type Le^y antigens on a specific glycoprotein. Variations of PIP´s glycosylation was analyzed for multiple individuals. An extremely high degree of fucosylation was observed without exception. An inverse correlation between the amounts of fucoses as well as sialic acid residues was detected: The more sialic aids are linked terminally to the *N*-glycans, the less fucose residues are attached. 40% of all individuals are heterozygous for a nonsynonymous single nucleotide polymorphism (SNP) which causes duplicate signals for the glycoprotein. The amino acid exchange influences the glycosylation pattern: The overall fucosylation is reduced while the overall sialylation is enhanced compared to PIP without the exchange. A pilot study was performed for comparing salivary PIP from healthy individuals and breast cancer patients. A distinct difference was observed for biantennary *N*-glycan structures. The amount of mono- and bisialylated biantennae is drastically increased and shows that PIP has potential as a non-invasive glyco-based biomarker for breast cancer.

Lectin-studies have revealed that Histidine-Rich Glycoprotein (HRG) shows an altered glycosylation during colorectal cancer (CRC). CRC is associated with poor survival rates, typically due to late diagnosis. Diagnostics could be improved, if a simple blood test was available. Here, an isolation strategy for HRG from blood plasma was

successfully established by introducing a two-dimensional chromatographic protocol. First analyses show that the glycosylation of HRG possesses a high uniformity between individuals. By plasmin-cleavage of HRG, it was possible to obtain a more site-specific analysis of glycosylation. On that basis, a comparison of HRG glycosylation during non-pathogenic and diseased conditions can be established in order to develop a glyco-based biomarker for CRC.

2 ZUSAMMENFASSUNG

Als eine der wichtigsten post-translationalen Modifikationen von Biomolekülen beeinflusst die Glycosylierung eine Vielzahl regulatorischer Funktionen. Daher kann der Zusammenhang zwischen veränderter Glycosylierung und pathogenen sowie immunogenen Prozessen einer Entwicklung eines spezifischeren Biomarkers dienen, als es die Proteinkonzentration selbst könnte. Die Analyse des Glycoms ist insbesondere für zwei biomedizinische Bereiche relevant, welche daher im Rahmen dieser Arbeit betrachtet wurden: A) Qualitätskontrolle von Biopharmazeutika sowie B) Entwicklung von Biomarkern für eine verbesserte Diagnostik.

Studien haben kürzlich gezeigt, dass Cetuximab, ein in der Krebstherapie eingesetzter mono-klonaler Antikörper, aufgrund anaphylaktischer Schocks schwere allergische Reaktionen bis hin zum Tod hervorrufen kann. In dieser Arbeit konnten zum ersten Mal nicht-humane Glycosylierungsepitope eindeutig identifiziert werden, welche für diese Reaktionen verantwortlich sind. Von den 37 *N*-Glycan-Zusammensetzungen, die in einem *bottom up*-Ansatz identifiziert wurden, konnten die zehn häufigsten durch Anwendung einer hier optimierten Variante der sogenannten 3DCC-Methode (*three-dimensional cross correlation*), welche LC-MS- und NMR-Daten kombiniert, strukturell charakterisiert werden. So besitzen 24% aller detektierten Glycane α-(1→3)-Gal-Epitope und/oder *N*-Glycolylneuraminsäuren, die die genannten schweren Immunreaktionen auslösen.

Komplementär zu dieser detaillierten, aber aufwendigen Analyse wurde eine schnelle bio-informatische Glycosylierungsanalyse auf der Ebene intakter Glycoproteine entwickelt. Sie basiert auf einem *Least-Square*-optimierten Fit des isotopenaufgelösten experimentellen MS-Spektrums durch theoretische Glycoform-Spektren. Die größte Herausforderung stellte dabei die Implementierung mehrerer Glycosylierungsstellen pro Protein dar. Die Bestimmung der *N*-Glycane erlaubt Rückschlüsse auf die Sicherheit sowie die Effizienz rekombinanter Antikörper. Ein *middle up*-Ansatz, der auf der Analyse von mAb-Subdomänen basiert, wurde etabliert, um eine positionsspezifische Glycosylierungsanalyse zu ermöglichen und potentiell unerwünschte Glycanepitope zu

lokalisieren. Anhand dieser Methode wurden verschiedene kommerziell erwerbliche Chargen von Bevacizumab, einem in der Krebstherapie verwendeten mAb, hinsichtlich ihrer Glycosylierung untersucht. Hiermit war es möglich, Unterschiede der Glycosylierung zu identifizieren und die Chargen hinsichtlich ihrer therapeutischen Effizienz zu vergleichen. In einer laborübergreifenden Studie, die durch das *National Institute of Standards and Technology* (NIST) organisiert wurde, wurden sowohl die oben beschriebenen *bottom up*-Analysen als auch solche von intakten Glycoproteinen durchgeführt. Eine erste statistische Analyse der eingereichten Datensätze (n = 93) durch NIST bestätigte anschließend, dass dieser multi-dimensionale Ansatz exzellente Ergebnisse liefert.

Prolactin-Inducible Protein (PIP), ein Glycoprotein, das in humanen Sekreten auftritt, konnte mit diversen biologischen Funktionen in Verbindung gebracht werden, beispielsweise der Tumorentstehung. Seine Glycanstrukturen wurden bisher nicht detailliert beschrieben. Daher wurde PIP in dieser Arbeit aus Speichel isoliert und mittels LC-MS/MS und NMR analysiert. Es zeigte sich, dass PIP eine ungewöhnliche *N*-Glycosylierung aufweist, die durch einen extrem hohen Fucosylierungsgrad gekennzeichnet ist. NMR-Analysen identifizierten diese hoch-fucosylierten *N*-Glycanepitope als Lewis^{y} (Le^{y})-Strukturen, deren Expression typischerweise während der Tumorentstehung hochreguliert wird. Somit wurden zum ersten Mal *N*-Typ-Le^{y}-Antigene auf einem spezifischen Glycoprotein eindeutig charakterisiert. Die Analyse der PIP-Glycosylierung verschiedener gesunder Individuen zeigte, dass bei diesen ohne Ausnahme ein hoher Fucosylierungsgrad herrscht. Es besteht eine inverse Korrelation zwischen der Menge von Fucosen sowie Sialinsäuren: Je mehr Sialinsäuren an die Glycane gebunden sind, desto geringer wird der Fucosylierungsgrad. 40% aller Individuen sind heterozygot für einen *single nucleotide polymorphism* (SNP), der die Aufspaltung der Proteinsignale bewirkt. Dieser Aminosäuren-austausch beeinflusst das Glycosylierungsmuster: Die Gesamtfucosylierung ist reduziert, während die Gesamtsialylierung erhöht ist. In einer Pilotstudie wurde die Glycosylierung von Gesunden und an Brustkrebs erkrankten Patientinnen verglichen. Ein Unterschied konnte

bei den biantennären *N*-Glycanen beobachtet werden. Die Menge an mono- und bisialylierten Biantennen ist signifikant erhöht und deutet auf eine Eignung PIPs als nicht-invasiver Glyco-basierter Biomarker hin.

Lectinstudien haben gezeigt, dass *Histidine-Rich Glycoprotein* (HRG) eine veränderte Glycosylierung bei Darmkrebserkankungen aufweist. Darmkrebs ist durch geringe Überlebensraten gekennzeichnet, typischerweise begründet durch eine späte Diagnose. Eine frühere Diagnose könnte durch die Entwicklung eines Bluttestes ermöglicht werden. In dieser Arbeit konnte eine Isolierungsstrategie für HRG aus Blutplasma erfolgreich etabliert werden. Erste Glycosylierungsanalysen zeigten eine hohe Einheitlichkeit zwischen verschiedenen Individuen. Ein Plasminverdau von HRG stellt eine positionsspezifischere Analyse dar. Auf dieser Grundlage kann ein Vergleich zwischen der HRG-Glycosylierung im nicht-pathogenen sowie erkrankten Zustand erfolgen.

3 INTRODUCTION

Within the last two decades, the analysis of the glycome gained significant prominence in biomedical research. Every cell is covered by a layer of glycans and glycoconjugates (the so-called glycocalyx), and the majority of all secreted proteins within multicellular organisms is glycosylated. Thus, the enormous relevance of glycosylation is not surprising.[1-5] Glycosylation of the cell surfaces becomes especially interesting, since several pathogenic organisms mimic the host glycosylation to avoid immunogenic reactions.[6] As one of the most important post-translational modifications of biomolecules, glycosylation controls manifold biological processes. For example, intra- and inter-cellular contacts are regulated by glycoconjugates and covalently attached glycans have a significant effect on physico-chemical properties such as the solubility, the folding or the stability of the corresponding proteins.[7-9] Specific alteration of glycan expression is associated with pathogenic processes such as inflammation or tumor progression.[10-11] Furthermore, recent studies have shown that biopharmaceuticals such as recombinant antibodies can cause severe immune reactions in patients that were induced by non-human glycosylation.[12] These and other glyco-related discoveries caused the emerging interest in glycomics, most importantly in medical and pharmaceutical research areas.

Especially the progress of analytical techniques provided new opportunities for more detailed insights into structural properties of glycosylation. These developments comprise new and improved chromatographic techniques and material for isolation and separation of biomolecules and in particular the enhanced sensitivity and accuracy of analytical instruments. Enormous instrumental progress of techniques such as mass spectrometry and NMR spectroscopy as well as development of bioinformatic tools for data interpretation led to improved insights and a deeper understanding of glycomics within structural biology and medical research.[13-17] These improvements led to biomarker detection, which in turn enables early diagnosis and disease monitoring, and to assessment of human compatibility within quality control of biopharmaceuticals.

3.1 Glycosylation of Proteins - Biosynthesis, Structures, Functions

Glycosylation is not directly controlled by the genome, in contrast to proteins that are synthesized by means of translation of RNA. However, glycosyltransferases that catalyze the transfer of monosaccharides to build up oligosaccharides are encoded in the genome.[18]

The majority of glycans is synthesized in the endoplasmic reticulum (ER) and the Golgi apparatus.[19] However, biosynthesis of glycans may also occur within the cytosol and the nucleus. In eukaryotic cells, the *N*-glycan biosynthesis starts by building a fourteen sugar oligosaccharide linked to a polyisoprenol lipid.[20] The first steps are carried out on the cytoplasmic side of the ER. Translocation into the lumen of the ER happens at the heptasaccharide stage. After completion of the tetradecasaccharide, it is transferred to the protein. Then, the carbohydrate is partly degraded and finally re-built by several enzymes within the ER and the Golgi apparatus. These glycosyltransferases are located within the mentioned organelles and catalyze the transfer of nucleotide donors (monosaccharides that were activated in the cytoplasm). Degradation of glycoproteins commonly occurs in the lysosome by glycosidases.[21]

More than a hundred native monosaccharides exist. However, only a small subset is regularly present in humans, namely hexoses, deoxyhexoses, hexosamines, uronic acids and sialic acids.[22] In animals and plants pentoses also occur frequently.[23] In comparison to nucleotides or amino acids that are linked linearly to form polymeric DNA/RNA molecules and proteins, respectively, monosaccharides can build up several glycosidic bonds resulting in branching points.[24] Monosaccharides occur in α- and β-configuration of the anomeric carbon and can be linked to theoretically every hydroxyl group from other mono-, oligo- or polysaccharides. Thus, the combination of various monosaccharides leads to isomers with an enormous structural diversity, demanding highly complex structure analyses (cf. Figure 1, bottom). Furthermore, modifications, such as phosphorylation and acetylation, increase complexity even more.[25-26]

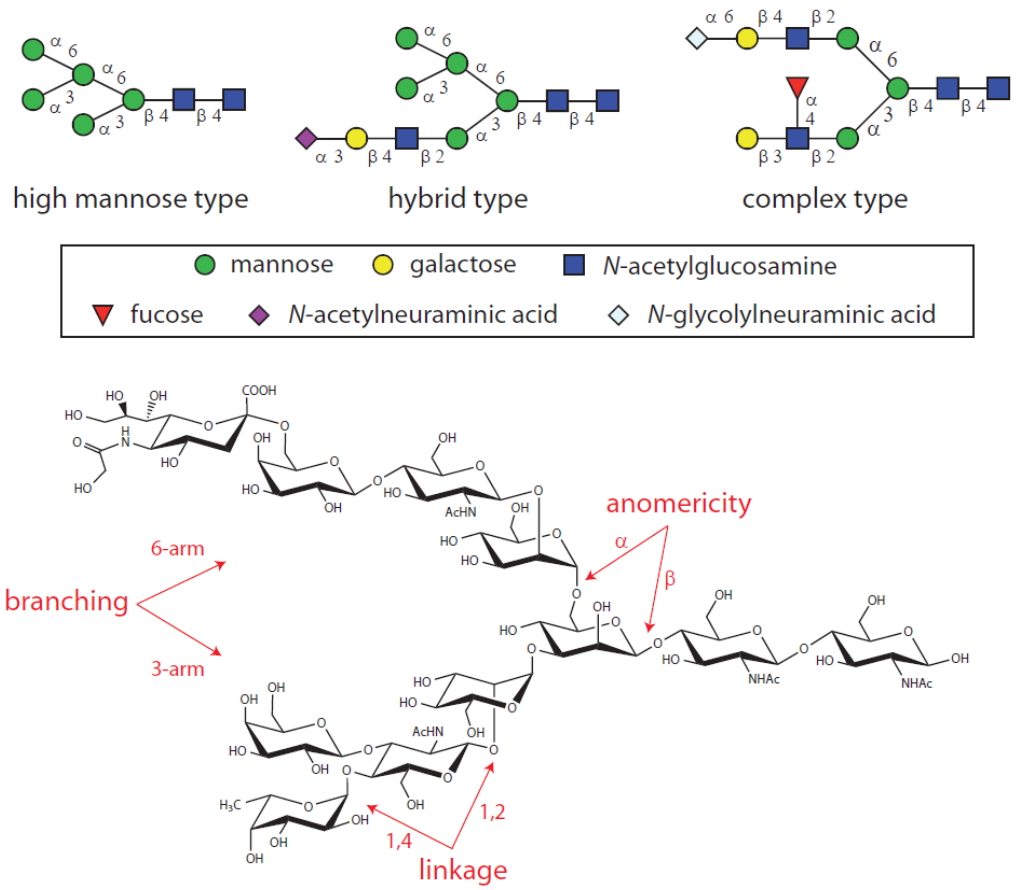


Figure 1: **Types and structural precision of *N*-glycans. Top: All *N*-glycans share the common pentasaccharide core $Man_3GlcNAc_2$. *N*-glycans are covalently attached to proteins by linkage of their proximal GlcNAc residue to asparagines within a consensus sequence (Asn-X-Ser/Thr, X: any amino acid except proline) and can be sub-divided into three groups. In high mannose type glycans, only mannoses are attached to the pentasaccharide core. The hybrid type glycan carries only mannoses on the 3-linked antenna, the other antenna is like those found in complex type *N*-glycans. In complex type *N*-glycans, both antennae are linked with *N*-acetyllactosamine that can be is further modified. Further branching at the core-mannose residues results in structures with multiple antennae, most commonly in tri- or tetraantennary-glycans. Bottom: Detailed chemical structure of the complex type *N*-glycan shown on the right in the top part. Linkage, branching and anomericity determine multiple physico-chemical properties and functions of the *N*-glycan.**

Two types of protein glycosylation are known to be predominant: *O*-glycans are covalently attached to serine or threonine residues, *N*-glycans are linked to asparagines (within a consensus sequence Asn-X-Thr/Ser, X: any amino acid except proline) in polypeptide chains.[26] While *O*-glycans can have several core structures, all *N*-glycans exhibit a pentasaccharide core ($Man_3GlcNAc_2$) and are modified by further attachment of monosaccharides (cf. Figure 1, top).

The diversity and structural precision of glycan structures enable glycoconjugates to fulfill a great number of physical and biological purposes, e.g. through their stabilizing or protecting properties and their action as specific recognition elements.[27-28] For example, glycans serve as a quality control for proper protein folding.[29-31] Regardless of their biological function, the *N*-glycans confirm correct protein folding within the ER and initiate degradation within the cytoplasm or lysosome if the protein is misfolded.

After correct biosynthesis, the glycoprotein accomplishes diverse biological objectives. One of the most important functions is the mediation as recognition unit for glycan-binding proteins (GBPs) within diverse biological processes. A fundamental group of GBPs are lectins, which possess carbohydrate-recognition domains (CRDs) that typically bind to distinct glycosylation structures.[32] Such lectins can be highly specific, e.g. the human influenza virus only binds to the terminal disaccharide sialic acid linked $\alpha(2\rightarrow6)$ to galactose.[33] Often, terminal residues, uncommon sequences or derivatizations of glycans provide specific functions in physiological processes or serve, presumably for evolutionary reasons, as targets for pathogens.

3.2 Glycans in Human Diseases

The manifold functional properties of glycans determined by their structure explain the diversity of diseases that are associated with glycosylation. Some examples for glyco-based diseases are given in this section. Especially the influence of glycans in biomarker detection as well as in quality control of biopharmaceuticals is discussed.

Glycans are involved in the pathogenesis of (bacterial and viral) infections and can therefore act as a starting point for developing therapeutics.[34-35] As mentioned, almost all cells carry a glycocalyx, and in the same way, bacteria and viruses exhibit saccharide structures on their surfaces. Therefore, it is not surprising, that interactions between host cells and bacteria or viruses are primarily determined by carbohydrate interactions.[36] For example, viruses present receptors on their surfaces that are specific for host-cell glycans.[37] Binding to the host-cell surface is required for a viral entry and therefore initiates the infection. Infection by the influenza virus is induced by specific recognition

of epithelial sialic acids in the respiratory tract of the host by surface lectins.[38] Therefore, crystal structures of the purified surface proteins were used for designing drugs that can bind and block the binding pocket for sialic acids.[39] In these studies, inhibitor drugs with high affinities to neuraminidases were developed, e.g. Zanamivir.[40]

Glycan biosynthesis can also be affected by inherited diseases (CDG: congenital disorders of glycosylation).[41] Thereby, glycans in diverse organs are synthesized structurally incorrect due to genetic defects. The extent can be variable up to a complete absence of major glycan types. The latter often causes embryonic death, whereas less distinct disorders, e.g. the absence of discrete residues, result in variable phenotypes, that are determined by severity and affected system causing several dysfunctions, e.g. impairment of cognitive skills.

More examples for diseases that are associated with abnormal glycosylation structures and/or functions include, among others, inflammation, cardiovascular diseases such as arteriosclerosis, pulmonary diseases such as bronchitis and neurological disorders.[27, 42-43]

3.2.1 Potential of Glycans as Specific Biomarkers for Human Diseases

Many disorders and diseases originate from incorrect glycan biosynthesis, or the disease causes structurally changed glycans as discussed above. Therefore, identification of altered glycosylation profiles can act as a specific biomarker for diagnosis or for monitoring of disease progression. Body fluids or tissues can be used for developing specific diagnostic assays.[44] In particular, easily accessible fluids such as blood, urine or saliva are favored as biomarkers since they are non- or minimally-invasively obtained sample materials. By such assays based on altered glycosylation profiles, a specific diagnosis in an early stage of disease could improve cure and/or survival rates significantly.

Likewise, biomarkers are useful for diseases that have no distinct phenotype. Patients that suffer from CDG can show diverse disorders that are difficult to correlate with the physiological origin.[41] Therefore, the detection of a physiological change observed from

a blood test aids in diagnosis and treatment. In CDG, many proteins possess an altered glycosylation. Especially the serum protein transferrin has a glycosylation that can be used as reliable biomarker for CDG diagnostics by

proving a significantly reduced sialylation.[45] Such a detection enables the treatment of CDG patients.

Behind cardiovascular diseases, cancer is the second most common cause of death worldwide (*World Health Organization*). A reason for severe or lethal disease progression is often its late diagnosis. Several diagnostic assays for cancer are based on the determination of the concentration of specific blood components. Unfortunately, several of these biomarkers often lead to false positive or false negative diagnoses, e.g. the determination of the concentration of prostate specific antigen glycoprotein (PSA).[46] A promising approach for developing more specific and reliable biomarkers is the analysis of the glycosylation profile of proteins. For example, α-Fetoprotein (AFP) was found to be specifically altered in patients suffering from hepatocellular carcinoma (HCC).[47] I.e., the heptasaccharide *N*-glycan ($Man_5GlcNAc_2$) detected on AFP of healthy donors as well as liver cirrhosis patients is modified by core-fucosylation in case of HCC.

It is well known, that proteins with a significantly altered glycosylation can serve as a function of tumor development and progression.[48] As in rapidly growing cells such as cells during embryonic development that are characterized by glycosylation changes, cancer cells also show altered glycan profiles. Some specific changes were observed frequently and can therefore be exploited for the development of specific biomarkers for early diagnosis or monitoring of cancer progression.[49] Some of these specific alterations are e.g. increased *N*-glycan branching, higher degree of sialylation, increased observation of (sialyl) Tn antigens on *O*-glycans and detection of significantly elevated concentrations of Lewis structures (cf. Figure 2).[50-53]

Many of these alterations have been studied in detail. To give one example, elevated amounts of $\beta(1\rightarrow6)$ branching are caused by the increased expression of the enzyme GlcNAc transferase V. In biological models, the enhanced expression of this

glycosyltransferase leads to an elevated degree of metastasis, while mutants lacking the gene for GlcNAc transferase V possess a reduced potential metastasis.[54]

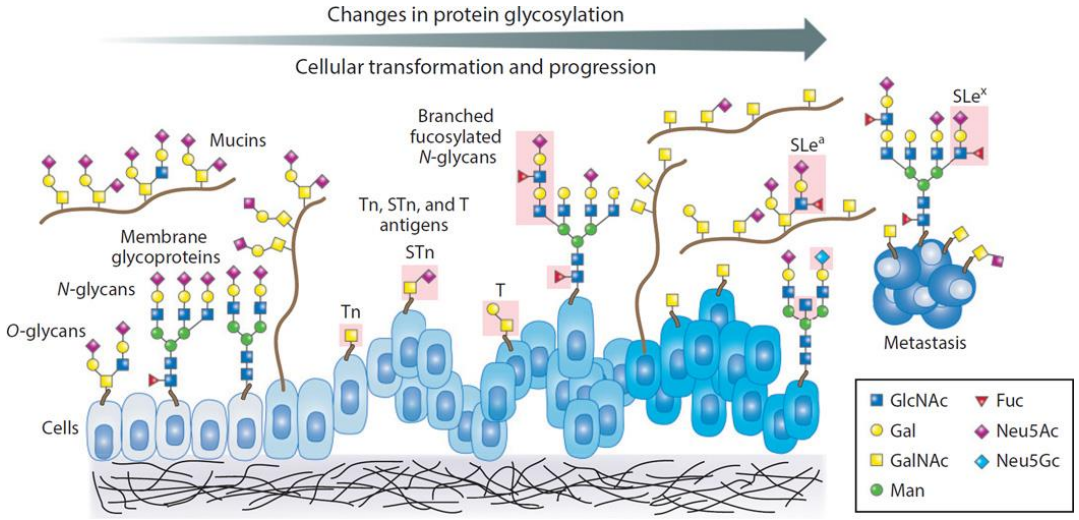


Figure 2: **Changes in protein glycosylation are common during carcinogenesis. Cancer-specific alterations are e.g. an elevated degree of branching of *N*-glycans, increased amounts of T, Tn, and STn antigens and Lewis antigens (pink-framed residues). GlcNAc: *N*-acetylglucosamine, Gal: galactose, GalNAc: *N*-acetylgalactosamine, Man: mannose; Fuc: fucose; Neu5Ac: *N*-acetyl neuraminic acid, Neu5Gc: *N*-glycolyl neuraminic acid, T: Gal-GalNAc-Ser/Thr, Tn: GalNAc-Ser/Thr, STn: sialyl Tn, SLe: sialyl Lewis.[55]**

3.2.2 Limits of Glyco-Based Biopharmaceuticals - Quality Control

Many commercially available biopharmaceuticals such as hormones or monoclonal antibodies possess carbohydrate residues.[56] This includes drugs against Alzheimer´s disease, against influenza viruses, as anticoagulants or during *in vitro* fertilization. Especially recombinant glycoproteins serve as profitable biopharmaceuticals.[57] Their glycosylation residues can affect their stability, activity or safety, respectively.[58-59]

Monoclonal antibodies are most rapidly emerging as therapeutic drugs for the treatment of cancer and of various other diseases like autoimmunity or inflammation.[60] Recently, it was found that non-human glycosylation of recombinant antibodies can cause tremendous problems for some patients.[12] α-(1→3)-Gal-epitopes and the glycan moiety *N*-glycolyl neuraminic acid (Neu5Gc) can both cause anaphylaxis.[61-62] These terminal residues can be found in glycans produced by non-human cell lines such as CHO or mouse cell lines.[63] Humans possess antibodies against α-(1→3)-Gal and Neu5Gc units

that can cause severe immunogenic reactions.[62] As a positive effect of an altered glycosylation, it was found that non-fucosylated glycan core structures as well as bisecting *N*-acetylglucosamines in monoclonal antibodies can enhance antibody-dependent cellular cytotoxicity (ADCC).[64-65] Therefore, unambiguous assignment of the glycosylation pattern of therapeutic antibodies is of high importance for assessment of human compatibility. The proportion of patients with a hypersensitive reaction against Cetuximab, a monoclonal antibody for the treatment of head, neck and colorectal cancer, illustrates the urgent need for reliable analysis of the glycosylation. Recombinant expression of glycoproteins without heterogeneity in glycosylation is so far impossible. Structures of glycans on antibodies vary greatly with the expression system.[56] Therefore, glycoprotein-based pharmaceuticals are typically allowed to show a certain glycoform variance.[66] Astonishingly, the glycosylation patterns in already licensed biopharmaceuticals have to be reproduced in generic versions, even if the glyco-profile caused severe side-effects in patients. As a result, this variance has to be ensured within quality control as well as for the approval of biosimilar mAb versions.

3.3 Analytics of Protein Glycosylation

The analysis of glycans is challenging due to their enormous structural diversity. Diversity results from variation in linkage, branching, monosaccharides present and from the anomericity of glycosidic bonds.[26] Every glycosylation site within a protein carries a number of glycan structures, resulting in a microheterogeneity.[67] For a complete structural characterization, this complex microheterogeneity has to be elucidated. Manifold analytical techniques were established so far and the choice of method has to be made regarding the purpose: Which structural features have to be identified for the corresponding sample? Is an unambiguous characterization including linkage as well as anomericity of the glycosidic bonds necessary? Is the variance of attached glycans to a specific glycosylation site required? Is the overall composition of all glycoforms necessary? Thus, the more structural details are required, the more time for analyses is required.

Since no standardized method for all these purposes exists, multifaceted approaches have to be applied and combined. Classical approaches include the use of glycan-binding

lectins, endo- and exoglycosidases, chemical treatments as well as comparing LC profiles of glycans with databases.[68] More recently applied methods for glycan characterization focus on LC-MS(/MS) and NMR techniques as well as various stationary phases for (HP)LC separations.[69-75]

Many lectins, also called glycoreceptors, were discovered that specifically recognize distinct monosaccharides or carbohydrate epitopes, most commonly terminal, but in some cases also internal residues.[32, 76] Also, endo- and exoglycosidases are used.[68] These glycosidases remove single monosaccharides or larger units leading to known reference glycans and therefore the carbohydrate can be sequenced. Several exoglycosidases are highly specific not only for the monosaccharide residue, but also for the anomericity of the glycosidic bond, which allows detailed characterization. Drawbacks are the limited number of endo- and exoglycosidases (not every residue has a convenient glycosidase counterpart) as well as the resistance of enzymatic cleavage in case of glycan derivatization. In such cases, chemical release of glycan fragments, e.g. β-elimination, can also be performed.

Much effort was made to develop improved stationary phases for chromatographic separation of biomolecules, such as (glyco)proteins and glycans. Incorporation of such chromatography materials, e.g. reversed phase-, ion exchange-, affinity- and size exclusion-phases into columns convenient for HPLC, has greatly improved proteomic as well as glycomic analyses.[69-70] Another development that has benefited glycan analysis is the development of the porous graphitized carbon (PGC) phase for chromatographic separation and often even mass isobaric glycans that differ only in one branching position can be separated.[71]

Mass spectrometry (MS) has evolved to the most commonly used technique for glycome analysis.[72] The development of soft ionization sources (MALDI and ESI) enabled intact ionization processes of biomolecules. Furthermore, by coupling of LC online with ESI-MS, analyte mixtures can be separated by diverse chromatographic techniques and rapidly analyzed by MS. Much effort was made to develop several analyzers for

separating molecules according to their size by high mass accuracy.[73, 77-79] Furthermore, several analyzers that are characterized by high analyzation speed, e.g. quadrupole magnets, can be used in combination with highly mass accurate devices, such as orbitraps. The possibility to fragment analytes is of particular importance.[80] Therewith, it is possible to sequence glycans by fragmentation. Typically, the glycans are fragmented by collision with inert gases which produce characteristic fragment ions, most likely (but not exclusively) resulting from dissociation of the glycosidic bonds. This typically allows identification of parts or even full sequences of the glycan composition. Furthermore, the analysis of glycopeptides allows a determination of the glycosylation site.[14] It is possible to vary the MS/MS parameters to obtain either fragments of the glycan, the peptide backbone or the linkage glycan-peptide. Without labelling and database comparison, linkages can typically not be determined unambiguously by MS/MS. Furthermore, the anomericity, i.e. α- or β-configuration, cannot be identified.

NMR is a non-destructive method and produces information of the linkage pattern, chirality and anomericity. Furthermore, in recent years it was also shown that NMR is highly sensitive: 15 pmol of material are sufficient for ^{1}H-NMR spectra so that the technique gains importance for glycome analysis.[74] In addition, a large amount of data is present for the analysis of isolated glycans by NMR spectroscopy.[75] By means of the structural reporter group concept it is possible to assign anomeric protons, H-2 protons from mannose, H-3, H-4 and H-5 protons from galactose, H-3 protons from neuraminic acid, H-5 and methyl protons from fucose, *N*-glycolyl neuraminic acid as well as *N*-acetate signals from *N*-acetyl neuraminic acid and *N*-acetylhexosamines from one-dimensional ^{1}H-NMR spectra.[17] By two-dimensional NMR analyses such as COSY or TOCSY, the analysis of the glycan can be fully accomplished without database comparison.

The use of MS and NMR analyses is a common strategy to obtain reliable data for structure interpretation. 3DCC (three-dimensional cross correlation) is a procedure that does no longer treat these data as independent, but integrates them with LC information to allow a consistent characterization with standard laboratory equipment and data

analysis software.[81] 3DCC enables the unambiguous assignment of glycans that are strongly overlapping in LC. By mathematical correlation of NMR signals with information from the LC-run (retention time and accurate *m/z* values), it is possible to extract NMR spectra of pure compounds. Therewith, simultaneously eluting *N*-glycans with different monosaccharide compositions can be structurally characterized by the integrated use of mass spectrometry and NMR spectroscopy.

As mentioned before, it is important to determine the level of structural details required. Summarized, there are three distinct analytical categories within glycoproteomics. The glycosylation can be analyzed on the level of released glycans, of glycopeptides or of intact glycoproteins: It has to be differentiated between so-called *bottom up* and *top down* approaches, respectively.[82] In the past, more effort had been made in the field of bottom up glycomics.[14] Therefore, the glycans are enzymatically or chemically cleaved from the glycoconjugates. The characterization of released glycans allows to draw conclusions for the overall glycan composition of the sample material, e.g. an isolated glycoprotein, a cell culture or a biological sample material such as blood plasma or tissue. Bottom up analyses are typically more detailed in terms of structural items. On the other hand, bottom up approaches show several disadvantages, e.g. the analyses typically require a lot of time and an intense sample preparation such as enzymatic/chemical cleavage and purification. This increases the risk of sample modification or selective enrichment which hampers, among other effects, reliable quantification. To overcome such problems, so-called top down approaches emerged recently.[83-84] Therein, the sample preparation is reduced to a minimum, since the analyte is not modified and typically injected intact into the MS. This tremendously accelerates the workflow. Furthermore, the risk of sample modification is drastically decreased. Also, quantification is more reliable for glycoproteins, as the ionization capability should be rather determined by the protein than from the smaller glycans. The glycoprotein can be analyzed intactly, but additional information about structural features such as amino acid sequences can be achieved by fragmentation. A draw-back of top down analyses is the missing detailed information about glycan compositions such as branching and linkage.

However, NMR and MS analyses produce enormous amounts of data. Furthermore, distinct identifications, e.g. determination of anomericity or linkage, require a high level of expertise and manual data interpretation. Development of supporting bioinformatic tools is essential to process the acquired data. Furthermore, reliable and complete databases to facilitate correct data interpretation are essential. Moreover, effort has to be made to develop algorithms for easier and automatic spectrum interpretation.

The recently developed methods and techniques immensely facilitated the structural analyses within glycomics. Ongoing progress and efforts within bioanalytics are essential to gain deeper understanding of basic principles of glycosylation. Therewith it will be possible to strongly contribute to biomedical research especially for diagnosis and treatment of glyco-associated disorders and for reliable development as well as quality control of glyco-based biopharmaceuticals.

4 OBJECTIVE

Glycosylation of proteins has long been known to significantly affect various of their properties as well as regulatory functions such as mediation of intra- and inter-cellular events.[7, 85] Thus, alteration of glycosylation associated with pathogenic or immunogenic processes is not surprising.[86] Carbohydrate structures have to be elucidated in detail to obtain a deeper understanding of such biologically relevant processes.

The quality control of biopharmaceuticals in terms of their glycan structures is very important for drug safety. Cetuximab, a biopharmaceutically used monoclonal antibody (mAb), is known to cause severe immunogenic side effects to the point of death. Epitopes that cause these fatal reactions should be unambiguously identified by a new optimized combination of LC-MS and NMR. Furthermore, new methods for a fast and reliable glycosylation analysis of mAbs should be developed. The critical aspect involves the attachment of multiple glycans per molecule. Focus should be laid on an application within quality control and development of new biosimilars.

The development of biomarkers offers tremendous potential for improving early diagnosis of various diseases. Altered glycosylation is often a specific marker for pathogenic processes. For developing glyco-based biomarkers for diseases, firstly a detailed characterization of healthy individuals is required. Two proteins were chosen for this: Prolactin-Inducible Protein (PIP) and Histidine-Rich Glycoprotein (HRG). PIP from seminal fluid was characterized to hold a very unusual *N*-glycosylation characterized by an extremely high degree of fucosylation. This aspect should be analyzed in detail from salivary PIP. Especially the nature of the high fucosylation should be elucidated by means of possibly occurring antigenic structures, i.e. Lewis^y epitopes, that are typically expressed during pathogenesis. Gender-specific as well as disease-related analyses should be performed to elucidate the suitability of PIP as a non-invasive biomarker. Lectin-studies have revealed that HRG shows an altered glycosylation during colorectal cancer, that is associated with poor survival rates, typically due to late diagnosis. A blood test could improve early diagnosis. HRG should be purified and analyzed from blood to achieve a first overview of HRG glycosylation.

5 RESULTS AND DISCUSSION

5.1 Glycosylation of Biopharmaceuticals

With more than 30 clinically approved monoclonal antibodies (mAbs) and mAb derivatives currently used in therapy within the US and the EU and 30 additional mAbs being in investigation for late clinical trial, IgGs represent the fastest growing class of therapeutics.[87-89] The antibodies are developed as treatment for various diseases like autoimmunity, cancer, inflammation, infection and cardiovascular diseases.[60] In 2009, mAbs used in the medical sector generated a revenue of 36 billion dollars in the US alone.[90] Since several therapeutic mAbs will go off patent in 2016, biosimilar products are in development and production.[91] These generic versions have to be verified in terms of structural similarity to ensure the biosimilar's quality and safety. Due to the enormous variety of effects of glycans on protein properties, the influence of glycosylation on potency and safety of biopharmaceuticals is not surprising. For example, in monoclonal antibodies an enhanced antibody-dependent cellular cytotoxicity (ADCC) was observed for non-fucosylated glycan core structures as well as bisecting *N*-acetylglucosamines.[92] Structures of glycans located on antibodies vary greatly with the expression system. Several of these non-human glycan motifs affect immunogenicity of mAbs, such as α-(1→3)-Gal-epitopes and *N*-glycolyl neuraminic acids that can cause anaphylaxis in humans.

Therefore, analysis of the glycome is important for development and quality assurance of biopharmaceuticals, like Cetuximab, as well as for the approval of biosimilar mAb versions by the authorities.

Cetuximab is used for the treatment of head, neck and colorectal cancer. It blocks the EGFR activation and is often combined with chemotherapy, especially in case of colorectal cancer.[93-94] Studies revealed that many different *N*-glycan structures can be linked to Cetuximab at four glycosylation sites - one at each ^{88}Asn in the Fab domain and one at each ^{299}Asn in the Fc domain (cf. Figure 3).[95]

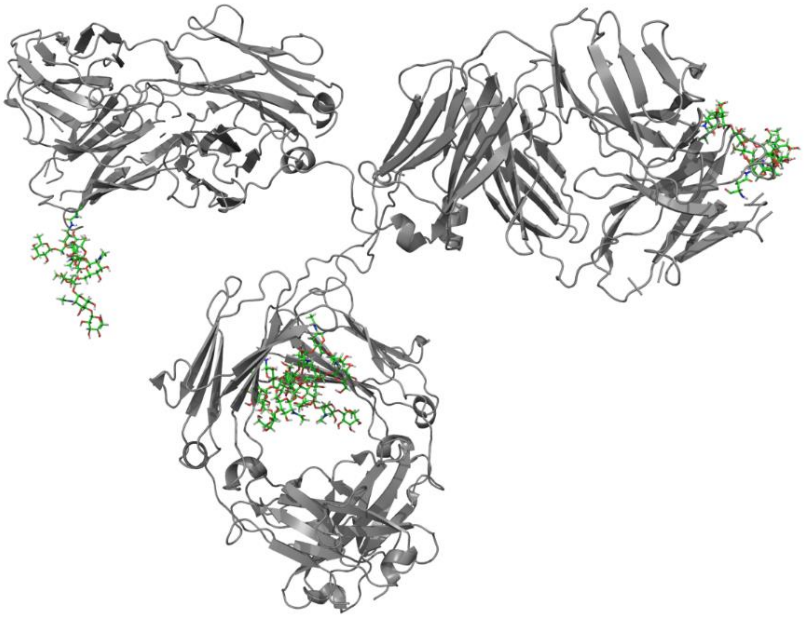


Figure 3: **Exemplified IgG_1 with four *N*-glycosylation sites. IgG_1 type antibodies (MW ≈ 160 kDa) are composed of two identical heavy and two identical light chains (MW_H ≈ 55 kDa, MW_L ≈ 25 kDa). The light and heavy chains are linked to a Y-shaped structure via disulfide bridges and strong hydrophobic interactions. All chains possess an N-terminal variable domain (Fab: antigen-binding fragment) and the C-terminal constant domain (Fc: fragment, crystallizable). Recombinant as well as native IgG_1 antibodies contain a highly conserved *N*-glycosylation site in the Fc domain. 15-20% of all human IgGs contain additional *N*-glycans in the variable light and/or heavy chain (here: additional oligosaccharides are linked to the variable heavy chain as present on the chimeric mAb Cetuximab). IgG structure was taken from PDB ID 1IGT (IgG from *mus musculus*) and extended with two *N*-glycans in the Fab domain (glycan structure taken from the Fc domain).**

30% of patients showed hypersensitive reaction against Cetuximab and therefore illustrate the crucial need for reliable glycosylation analysis.[12] Furthermore, besides undesirable immune reactions, distributers warn patients of other severe side effects. For example, the distributing US company *Lilly* informs that 2% of all patients that receive Cetuximab in combination with radiation therapy suffer heart attack and/or sudden death. Even worse are the statistics for patients that receive both Cetuximab and platinum-based chemotherapy with 5-fluorouracil: 3% of treatments cause A) heart problems resulting in death or B) sudden death. Besides necessary glycoform analyses, recent studies revealed an incorrect primary structure of Cetuximab in the databases and therefore demonstrate the requirement of extensive characterization of amino acid sequences to ensure the safety of first generation and biosimilar versions.[91]

In this work, versatile methods for the characterization of *N*-glycans are developed and applied to monoclonal antibodies. Detailed characterization of released *N*-glycans correlating LC, MS and ^{1}H-NMR data (cf. section 5.1.1) is shown and a procedure for a broad identification of multiple glycosylated mAbs is demonstrated to obtain a fast and robust analytical method for intact glycoproteins (cf. section 5.1.2) and sub-domains of these IgGs (cf. section 5.1.2.1), respectively.

5.1.1 Three-Dimensional Cross Correlation of Cetuximab

Unambiguous assignment of the glycosylation pattern of therapeutic antibodies is of high importance for assessment of human compatibility. Therefore, a detailed *N*-glycan analysis of the therapeutic antibody Cetuximab was targeted by LC-MS/MS analyses tightly integrated with ^{1}H-NMR spectroscopy was targeted to obtain unambiguous structural characterization.

Different schemes exist in literature for the analysis of *N*-glycans linked to proteins. Most of the time, LC-MS(/MS) techniques are employed. Historically, bottom up approaches starting from the glycoprotein followed by enzymatic cleavage either into glycopeptides or glycans and subsequent analysis of the compositions dominate the field.[14, 96] Combination of separation of *N*-glycans with high precision MS analysis as well as characteristic fragmentation patterns has also tremendously benefited structural elucidation.[97-101] Furthermore, NMR is a non-destructive method and produces information on the linkage pattern, chirality and anomericity. By means of the structural reporter group concept it is possible to unambiguously assign anomeric protons as well as several other protons of diverse monosaccharides from one-dimensional ^{1}H-NMR spectra.[17]

The use of MS and NMR analyses is a common strategy to obtain reliable data for interpretation of the structure. 3DCC (*three-dimensional cross correlation*) is a procedure that integrates MS, NMR and LC information to allow a consistent characterization with standard laboratory equipment and data analysis software,

developed in preceding work in this laboratory.[81, 102] Analytes similar in structure often overlap in chromatographic separations. 3DCC allows the assignment of glycans that are strongly overlapping in LC. By mathematical correlation of NMR signals with information from the LC-MS run (retention time and accurate *m/z* values) it is possible to extract NMR spectra of pure compounds.[81] Previous work applied the 3DCC procedure to analyze well-known *N*-glycan structures of the model system bovine fibrinogen.

In this procedure, EDCs (extracted delta chromatograms) are defined as the intensity at each data point in the NMR spectra as a function of the fraction number, i.e. retention time, that is, a plot of the intensity at each chemical shift as a function of retention time. All NMR signals belonging to one analyte rise and fall at the same time the analyte generates a signal in the MS spectrum (this latter one is referred to as EIC). At a given chemical shift in the NMR often more than one analyte exhibits a signal. This results in an EDC that is composed of several signals rising and decaying differently with the retention time. Thus, if an EDC contains signals from several analytes, the EICs of all structures that cause this experimental EDC are summed up in a weighted proportion (according to their abundance) to generate a calculated EDC (cf. Figure 8). In the precedent work, this was realized in two mathematical ways - $3DCC_L$ that uses least square optimization and $3DCC_C$ that applies correlation coefficients (Pearson). Factors thus obtained have to be multiplied with each EIC for the calculation of the theoretical EDC and are proportional to the abundance of the corresponding analyte in the mixture. An unambiguous clarification of glycan epitopes that cause the reported severe immunogenic reactions in patients after receiving Cetuximab was intended.

In this work, the *N*-glycans of Cetuximab were analyzed and quantified using different analytical methods. The glycans were prepared and analyzed by tight integration of LC-MS(/MS) and ^{1}H-NMR, respectively (cf. Figure 4).

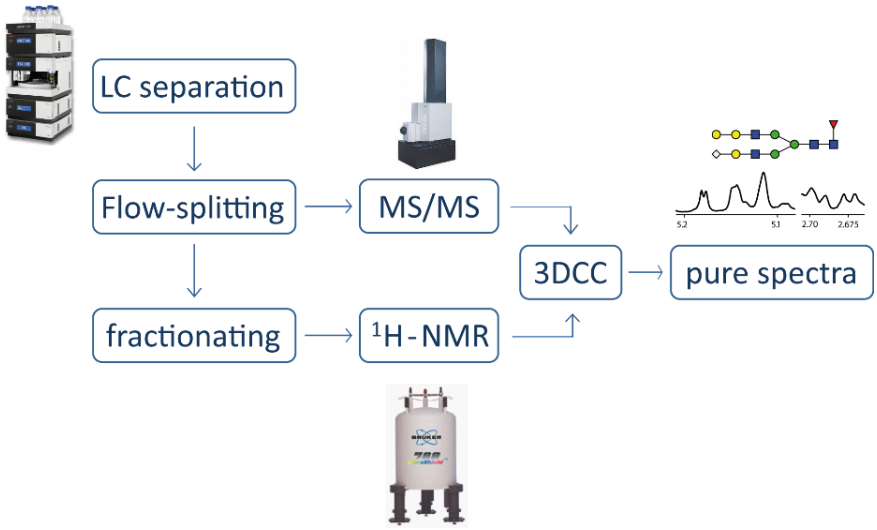
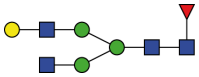
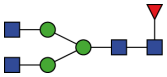
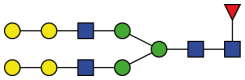
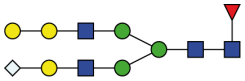
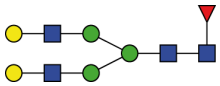
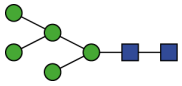
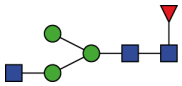
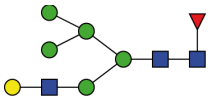
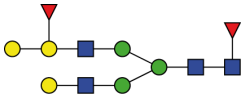
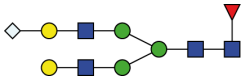
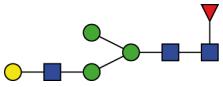
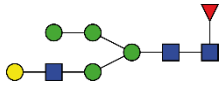
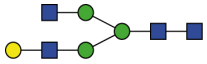


Figure 4: **Schematic setting of experimental workflow. For correlation of LC, MS(/MS) and NMR data, in the first instance analytes are separated by LC. The flow is split post-column. A defined volume is directly injected online into the mass spectrometer (here: ESI-q/TOF) to continuously generate (tandem-)mass spectrometric data. Simultaneously, the remaining flow is fractionated into a well plate. For NMR analyses, fractions are lyophilized and re-suspended in D_2O. Subsequently, LC-MS and ^{1}H-NMR data are mathematically correlated by 3DCC to obtain pure spectra of every analyte.**

Cetuximab was treated with 6 M urea and subsequently reduced with DTT and alkylated with iodoacetamide to denature the protein. Enzymatic cleavage of the glycans from the protein was achieved by addition of PNGase F. After purification of the glycan/glycopeptide/peptide mixture by C18-SPE (solid phase extraction), the released oligosaccharides were separated by PGC-HPLC. Ammonia was added to the eluent to avoid mutarotation of the monosaccharides at the reducing end during chromatography. Thus, glycans elute as one LC peak instead of two peaks for the α- and β-isomers of each oligosaccharide, respectively. LC separation was coupled online to an ESI-q/TOF to acquire MS(/MS) spectra. Within a 10-min period, 37 underivatized N-glycans eluted (cf. Figure 89 in appendix and Table 1). The compositions of the sugars were identified by ESI-q/TOF mass spectrometry yielding a precise mass with only one exceeding 4 ppm accuracy. Monosaccharide arrangement was suggested by comparing mass accuracy of parent ions with glycan databases, partly by inclusion of fragment spectra. The amount of the N-glycan with the lowest abundance was approximately 38 pmol (50 ng). Among the more abundant structures, critical motifs like α-(1→3)-Gal-epitopes and *N*-glycolyl neuraminic acids were detected.

Table 1: **Compositions of N-glycans of Cetuximab characterized by LC-MS based on high mass accuracy and several of them by CID-ESI fragment spectra (cf. Figure 6).**

Glycan composition[a]	Identified by		Quant.
	MS [Da]	Acc. [ppm]	MS [%]
	**1624.604***	**3.7**	**31.4**
	1462.552*	3.7	19.6
	2110.762*	2.7	13.9
	2255.800*	2.3	7.7
	1786.656*	2.7	7.5
	1234.436	2.2	2.9
	1259.465	1.0	2.0
	1745.626*	1.7	1.9
	2094.763*	1.3	1.5
	**2093.746***	**2.2**	**1.5**
	1421.519	1.1	0.4
	1583.574*	1.3	1.0
	1478.542*	1.4	1.0

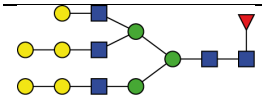
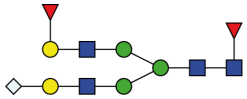
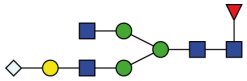
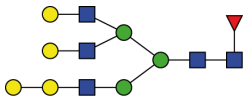
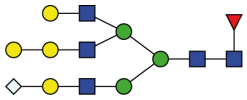
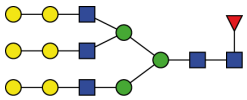
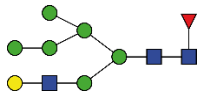
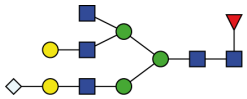
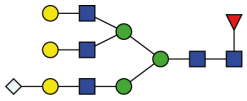
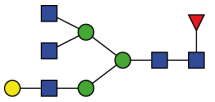
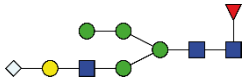
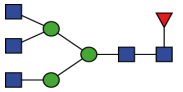
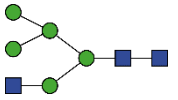
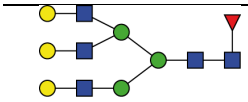
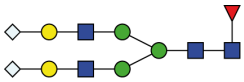
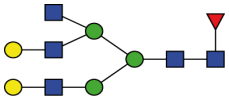
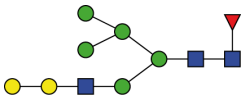
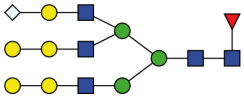
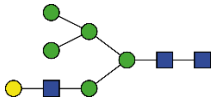
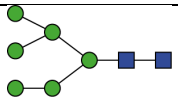
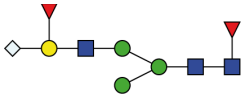
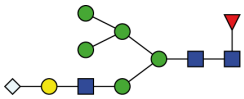
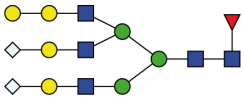
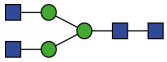
Glycan composition[a]	Identified by		Quant.
	MS [Da]	Acc. [ppm]	MS [%]
	2475.888*	1.9	0.4
	2239.802	1.5	0.4
	1931.692	1.9	0.3
	2313.838	0.6	0.3
	2620.931*	1.5	0.3
	2637.945*	1.6	0.3
	1761.621	1.6	0.3
	2296.8198	0.3	0.2
	2458.868	0.6	0.1
	1827.678*	0.2	0.3
	1890.758	2.2	0.2
	1665.625	0.6	0.2
	1437.514	1.8	0.2

Table 1: Continued

Glycan composition[a]	Identified by MS [Da]	Identified by Acc. [ppm]	Quant. MS [%]
	2151.782	1.3	0.8
	2400.832*	1.9	0.6
	1989.734*	0.8	0.6
	1907.680*	1.5	0.6
	2782.984*	1.4	0.5
	1599.566*	1.4	0.5
	1396.486	0.7	0.2
	2077.751	4.5	0.1
	2052.718	1.9	0.1
	2928.016*	1.9	0.1
	1316.486*	2.3	<0.1

[a] : *N*-acetylhexosamine, : fucose, : *N*-glycolylneuraminic acid, : mannose, : galactose. * Analyzed by CID-ESI fragment spectra. The two structures highlighted in boldface each contain two isomers (cf. Table 2) identified by 3DCC.

Figure 5 shows the summed up full scan MS spectra within the 10-min period of oligosaccharide elution. Besides the protonated analyte that is most abundant, for each glycan precursor several ammonium adducts exist. Biantennary complex type *N*-glycans are typically doubly charged, larger glycans partly carry three units of charge. High mannose type *N*-glycans usually carry one charge, e.g. glycan 7 (cf. Table 2) at m/z = 1235.4401.

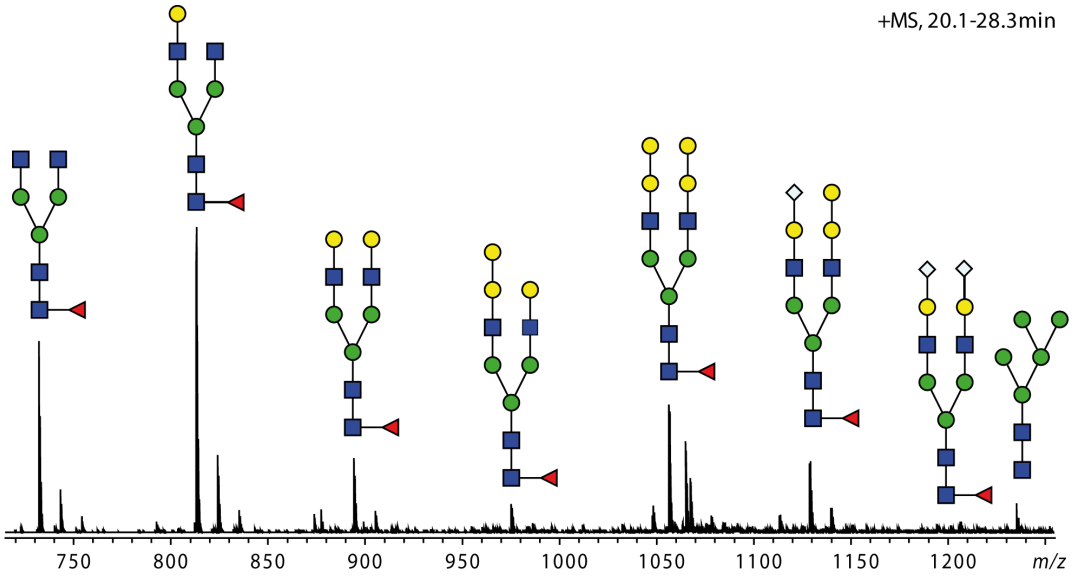


Figure 5: **Summed up full scan spectrum (MS^1, minute 20 to 28) of *N*-glycans released from Cetuximab. Accurate masses allow determination of overall monosaccharide combinations. Glycan compositions (see annotations) can be predicted by comparison to databases containing *N*-glycan structures. Only glycan structures that are identified by 3DCC (cf. Table 2) are annotated. Glycan $Hex_5HexNAc_4dHex_1NeuGc_2$ can be identified if the mass range is zoomed.**

Fragment spectra were recorded and analyzed for half of them (cf. Table 1 for details, marked by an asterisk). Figure 6 shows the fragment spectrum of glycan 3 (cf. Table 2, compound number). Fragmentation of glycans generates glycosidic and cross ring fragments. Glycosidic fragments allow determination of monosaccharide sequences, while cross ring fragments hint linkage positions. However, only monosaccharide sequences can be determined with a high probability, while linkages of glycosidic bonds are typically not resolved by MS of native glycans. Modification such as methylation and spectra comparison with model glycans have to be included for information about linkages.[103] Nevertheless, structural information about glycosidic bonds by MS is by far not as precise as NMR analyses or digestion with highly specific glycosidases, although several fragments may provide an indication.

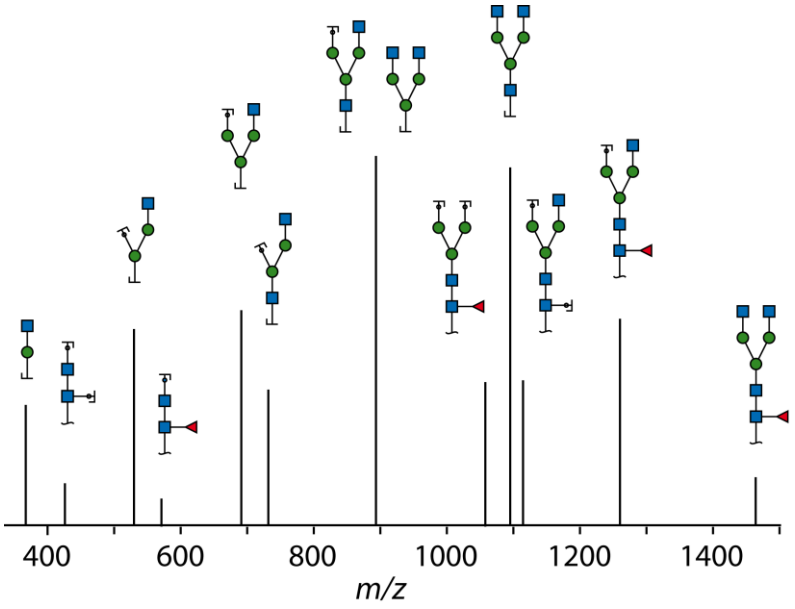


Figure 6: **CID MS/MS spectrum of *N*-glycan 3 (cf. Table 2) with the composition $Hex_3HexNAc_4dHex_1$ at *m/z* = 1463.540. B- and Y-ions occur most frequently in fragmentation of *N*-glycans in positive mode mass spectra. Fragment spectra give information about the monosaccharide sequence. However, details about linkages are hard or even impossible to obtain by MS. For symbol explanation see legend to Table 1. Signal at *m/z* = 893.325 can originate from the two annotated fragment ions.**

MS alone is not sufficient to elucidate the full structure of glycans; hexoses, e.g. mannose, galactose and glucose, cannot be differentiated by MS because of their identical molecular formula. However, these hexoses exhibit altered biological activities and therefore have to be distinguished clearly. Since mass isobaric structures that differ only in linkage positions and branching are hard or even impossible to differentiate by MS/MS, integration of MS and ^{1}H-NMR was used for unambiguous structure characterization. Analysis of glycan mixtures solely by ^{1}H-NMR is also hardly appropriate for unambiguous characterization, since signals for all present glycans overlay and cannot be distinguished. The LC flow was fractionated (15 sec/fraction) in parallel to the MS acquisition into a deep well plate. Each fraction measured by 1D-^{1}H-NMR contains multiple *N*-glycans and as a result signals of various analytes (cf. Figure 7). 3DCC was applied to mathematically separate NMR spectra using information from the LC-MS run, i.e. the EICs containing retention time and accurate *m/z* values. Therewith, ten *N*-glycans were unambiguously identified by the 3DCC

procedure (Figure 7, Table 2). *N*-Glycans with an abundance as low as 0.1% (relative quantification by integration of EIC signals for all detected 37 structures) were detected.

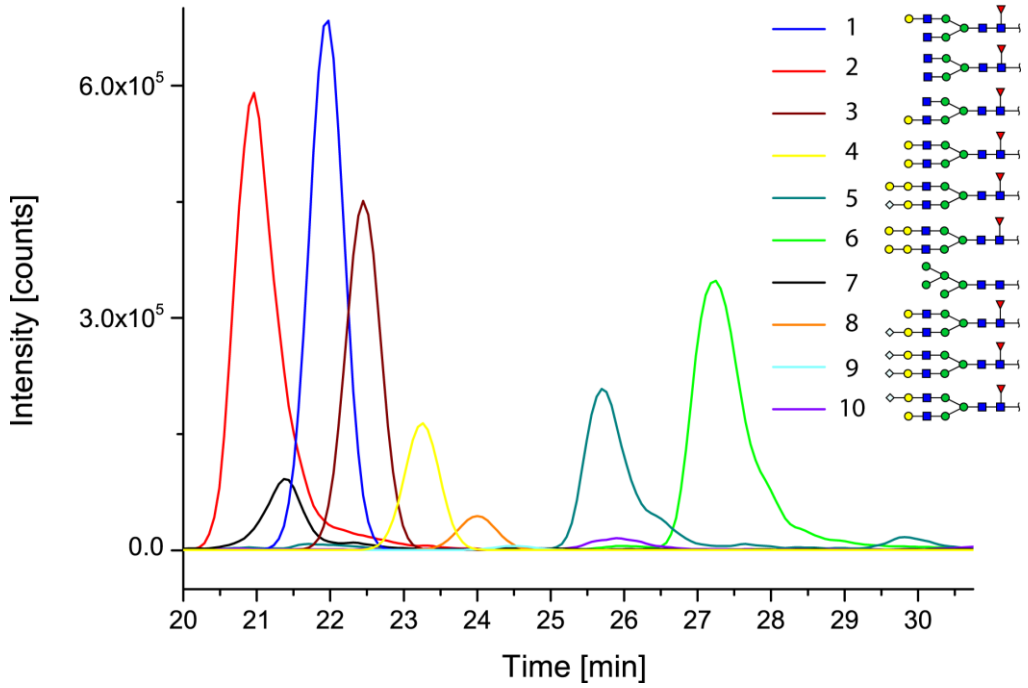
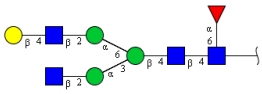
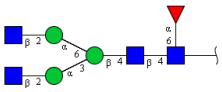
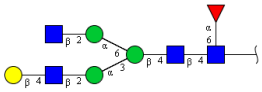
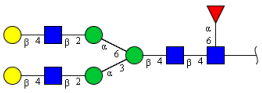
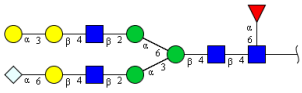
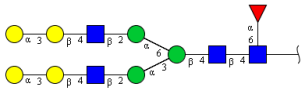
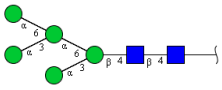
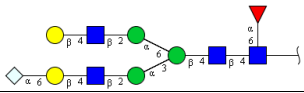
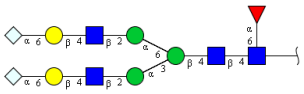
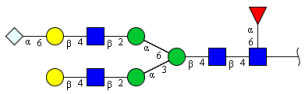


Figure 7: **LC-MS chromatogram of underivatized *N*-glycans enzymatically cleaved from Cetuximab separated on a PGC column. Ammonia was used as co-eluent to avoid separation of α- and β-anomers. A detail of ten extracted ion chromatograms of *N*-glycans that yielded unambiguous structure characterization by ^{1}H-NMR analysis is shown. Within the time depicted, 29 further *N*-glycans co-elute that were identified by MS/(MS) spectra. See Table 2 for configuration of glycosidic bonds.**

Compounds 1 and 3 as well as 8 and 10 are pairwise mass isobaric structures that could only be characterized by 3DCC methods. These structures cannot be distinguished by LC-MS alone which demonstrates the analytical power of combining LC-MS and NMR in the 3DCC procedure (cf. Figure 11).

Table 2: **Results obtained by applying the $3DCC_L$, $3DCC_C$ and $3DCC_S$ methods to elucidate the structure of the prominent glycans. About 67% of the structures are covered by the 3DCC technique. The EICs obtained from the separation of the LC-MS run were used for the generation of the ^1H NMR spectra (cf. Fig. 1 and Fig. 3). Compounds 1 and 3 as well as 8 and 10 are pairwise mass isobaric structures that could only be characterized by 3DCC methods.**

Glycan structure	Identified by			Quantification		Compound Number
	MS [Da]	Accuracy [ppm]	NMR (L,C,S)[a]	MS [%][b]	NMR [%/nmol][c]	
	1624.604	3.9	L,C,S	20.4	13.9/5.2	1
	1462.552	4.9	L,C,S	19.6	22.9/8.6	2
	1624.604	3.9	L,C,S	11.4	14.6/5.5	3
	1786.656	3.1	L,S	7.5	3.0/1.1	4
	2255.800	2.8	L,C,S	7.7	5.8/2.2	5
	2110.762	2.8	L,C,S	13.9	15.9/5.9	6
	1234.436	1.9	L,C,S	2.9	7.9/3.0	7
	2093.746	2.5	L,S	1.4	4.8/1.8	8
	2400.832	0.4	L,S	0.6	10.0/3.7	9
	2093.746	2.5	L,S	0.1	1.2/0.4	10

[a]: L: identification by applying $3DCC_L$ (*least square optimization*), C: identification by applying $3DCC_C$ (*Pearson coefficients*), S: identification by applying $3DCC_S$ (*singular value decomposition*) cf. text; [b]: MS quantification by integration of EIC signals for all 37 detected *N*-glycans by MS. [c]: Relative NMR quantification by integration of H-1 (β-D-GlcNAc) signal. Absolute quantification by calibration relative to an external sucrose standard.

Values of quantification by MS and NMR differ (cf. Table 3). It is known, that the ionization capability of small molecules, such as glycans, varies in dependency on their

structure, so that the amount of glycans is not directly reflected by the MS intensity. For example, phosphorylated peptides show a significantly decreased ionization capability when compared to the same peptides without modification. Thus, MS intensity is not appropriate for exact quantification. Therefore, quantification should be based on NMR analyses that are not dependent on structural features. Integrals of ^{1}H-NMR peaks are proportional to the corresponding amount of the analyte in the sample. Absolute quantification can be performed by calibration to an internal or an external standard.

Table 3: Comparison of quantity determined by MS and NMR. Values from Table 2 were normed and the absolute deviation was determined (right column).

MS normalized [%]	NMR normalized [%]	Deviation, absolute [%]
24	14	10
23	23	0
13	15	2
9	3	6
9	6	3
16	16	0
3	8	5
2	5	3
1	10	9
0	1	1

Originally, 3DCC was realized in two mathematical ways - $3DCC_L$ that uses least square optimization and $3DCC_C$ that applies correlation coefficients (Pearson). Here, 3DCC was extended with a new variant to solve the set of equations by using singular value decomposition (SVD). The new variant is called $3DCC_S$. SVD is considered to be the best method for dealing with sets of equations or matrices that are singular or very close to singular.

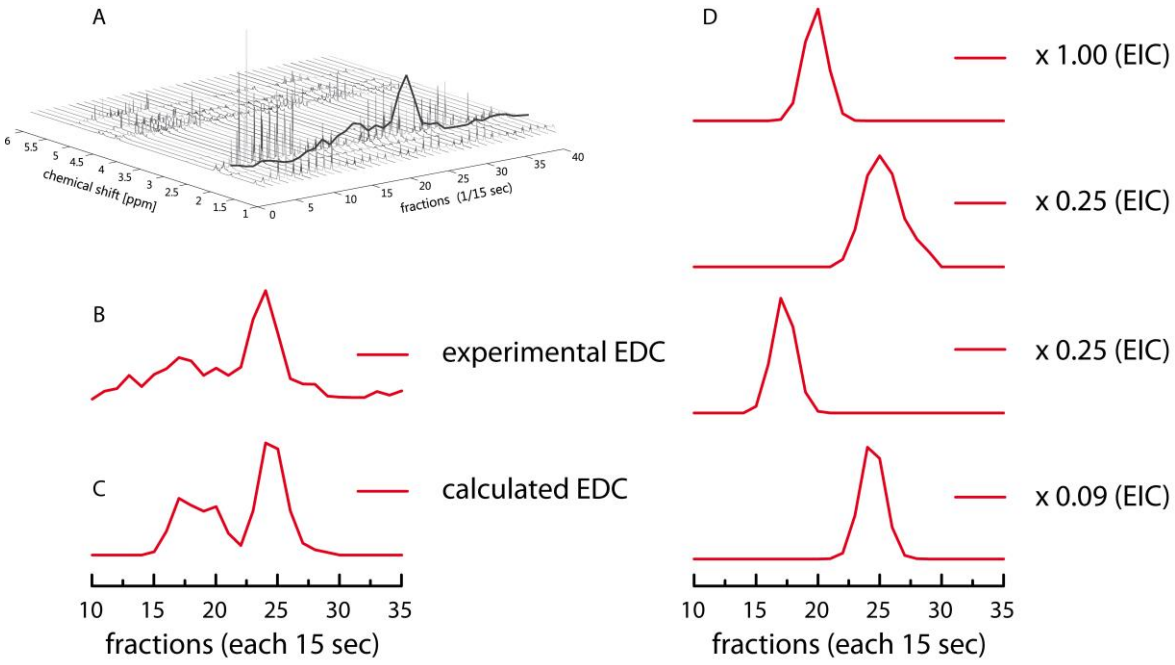


Figure 8: **Exemplified 3DCC results by applying singular value decomposition using LC-MS and ^{1}H-NMR information. 40 LC-fractions were analyzed by ^{1}H-NMR (A). An EDC represents a distinct chemical shift over all fractions (A+B). The 3DCC method calculates a theoretical EDC (C) by adding EICs (D1-D4) with weighting factors that fit an arbitrary experimental EDC (B). The factor for the weighting of an EIC is proportional to the portion of the glycan in the mixture that generates the considered EDC signal.**

Implementation of experimental data is identical for all variants. NMR data are imported to MATLAB by using the "rbnmr" script of Nils Nyberg. As a result, two objects are generated: a vector holding the chemical shift for each data point and a 2-dimensional matrix (matrix_NMR) containing the signal intensities of all acquired NMR spectra as a function of chemical shifts and of fraction number. EICs were generated (DataAnalysis, Bruker) and the number of data points was reduced to the number of LC fractions (Origin, OriginLab Corporation), which is equal to the number of acquired NMR spectra. In case an EIC for a given mass has more than one peak, the EIC is separated into as many EICs as peaks occur. This is achieved by a line shape fitting procedure. A two-dimensional matrix containing the intensities for all EIC data points as a function of LC retention time and the number of analytes that can be extracted from the EICs is imported to MATLAB (matrix_MS). EDCs are vectors (i.e. intensities at a given NMR data point as a function of the retention time, which is proportional to the fraction number) and are contained as

columns in matrix_NMR. The following subsections describe briefly the two recently introduced solutions to the cross correlation problem of correlating the EICs with the EDCs by least square minimization ($3DCC_L$) and by correlation coefficients ($3DCC_C$) followed by a detailed section on the new variant $3DCC_S$ utilizing SVD.

$3DCC_L$

Theoretical EDCs (the number of EDCs is equal to the number of NMR data points) are calculated to yield the fit to each experimental EDC using a linear combination of all EICs. EICs are summed up with individual weights to generate the theoretical EDC (cf. Figure 8). This variant uses least square optimization (*lsqlin* command of MATLAB) to fit the EDCs. Each EIC receives a coefficient that is required for the calculation of each theoretical EDC. The coefficients for one particular EIC are then plotted for all NMR data points resulting in the $3DCC_L$ NMR spectra for pure components for each EIC, i.e. for each analyte.

$3DCC_C$

The second previously described variant uses cross correlation (Pearson coefficients) of every EIC with every EDC. Therefore, the correlation coefficients are built pairwise (*corr* command in MATLAB). The 3DCC NMR spectra are then calculated by pairwise multiplying these correlation coefficients with the experimental NMR sum spectrum at a given chemical shift and *m/z* value. Negative intensity values were set to zero. This procedure leads to 3DCC NMR spectra that contain unique signals. In combination with MS spectra the characterization is achieved.

$3DCC_S$

The new variant $3DCC_S$ uses matrix_MS to generate individual EDCs with singular value decomposition (SVD) using the corresponding MATLAB routine (*svd* command). SVD is known to be the method of choice to solve most linear squares problems and can be seen to enhance effectiveness and accuracy of $3DCC_L$. The system of equations

$$A \cdot x = b \qquad (1)$$

has to be solved for every EDC, so that A is represented by the (known) matrix_MS, b is represented by one (known) experimental EDC (individual columns in matrix_NMR) and x denotes a vector of the coefficients (unknown) that is necessary to fit the weighted sum of EICs to the experimental EDC (compare $3DCC_L$) and has the same size as the number of fractions. Therefore, the SVD routine decomposes matrix A to prepare the inverse matrix A^{-1}, since

$$x = A^{-1} \cdot b \tag{2}$$

Summarized, any $j \times k$ matrix A can be expressed as the product of three unique matrices: a $j \times j$ matrix U_A, a $k \times k$ matrix $(V_A)^+$ ($^+$: conjugate transpose) and a $j \times k$ diagonal matrix D_A:

$$A = U_A \cdot D_A \cdot (V_A^{\cdot})^+ \tag{3}$$

Since U_A and V_A are unitary matrices $(U_A)^+ = (U_A)^{-1}$ and $(V_A)^+ = (V_A)^{-1}$, it is

$$A^{-1} = ((V_A)^+)^{-1} \cdot (D_A)^{-1} \cdot (U_A)^{-1} \tag{4}$$

$$= V_A \cdot (D_A)^{-1} \cdot (U_A)^+ \tag{5}$$

With (2), coefficient vector x can be calculated for every EDC, e.g. b:

$$x = V_A \cdot (D_A)^{-1} \cdot (U_A)^+ \cdot b \tag{6}$$

$3DCC_S$ spectra are then calculated in the same way as for $3DCC_L$ spectra.

$3DCC_L$ and $3DCC_S$ generate complete extracted spectra of the compounds. In comparison, $3DCC_C$ results in unique signals, but shows significantly higher signal-to-noise (S/N) ratios. Therefore, $3DCC_C$ should be used in case of low S/N ratios. However, quantification is possible only with $3DCC_L$ and $3DCC_S$. $3DCC_S$ spectra show a more continuous progress of curves than $3DCC_L$ spectra (cf. Figure 9). In addition, effectivity in terms of computing time is enhanced drastically. In case of $3DCC_S$ less than 1 second is required for the whole calculation on a modern workstation PC compared to at least 20 minutes for the $3DCC_L$ variant.

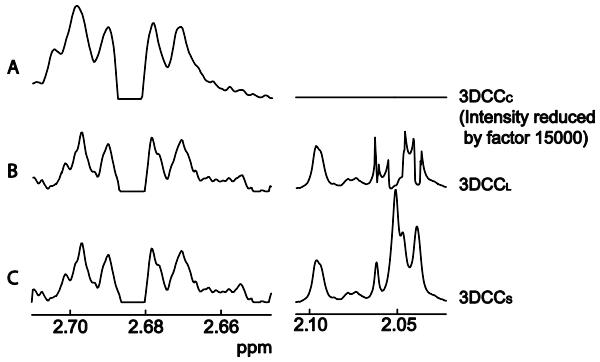


Figure 9: **Comparison of spectra obtained by the $3DCC_C$, $3DCC_L$ and $3DCC_S$ methods, respectively. $3DCC_L$ and $3DCC_S$ generate complete spectra containing all signals of the compounds. In comparison, $3DCC_C$ results in unique signals (see NAc signals at about 2.0 ppm, A: right hand side), but shows significantly higher signal-to-noise (S/N) ratios (see *N*-glycolyl neuraminic acid signals at about 2.68 ppm, A, left hand side). The depicted detail of the $3DCC_S$ spectrum (C, right hand side) shows more continuous curves than $3DCC_L$ (B, right hand side).**

Correlation of increasing and decreasing NMR signals of one analyte with the rising and falling of the corresponding *m/z* values detected by MS, led to ten extracted NMR spectra of pure compounds (Table 2, Figure 10). By means of 3DCC, it is possible to identify the presence of the immunogenic α-(1→3)-Gal epitope utilizing highly specific signals at chemical shifts of 5.145 (H-1), 4.019 (H-4) or 4.192 (H-5) ppm and of the immunogenic *N*-glycolyl neuraminic acid moieties at chemical shifts of 4.118 (NeuGc), 2.684 (H-3e) and 1.739 (H-3a) ppm, respectively. Figure 10 shows four 1D-^{1}H-NMR spectra of these critical glycosylation patterns. This represents the first unambiguous identification by MS/MS and NMR of these structures. At the top of Figure 10 the sum of all NMR spectra is shown. The spectrum expectedly contains all characteristic chemical shifts. Only by means of the mathematical separation into spectra of pure analytes, the unambiguous characterization of each structure was possible. Accordingly, spectra of analytes that do not carry a specific motif, e.g. an *N*-glycolyl neuraminic acid, do not show signals at the chemical shifts of 2.684 ppm (cf. traces for compounds 1, 3 and 6) characteristic for these residues.

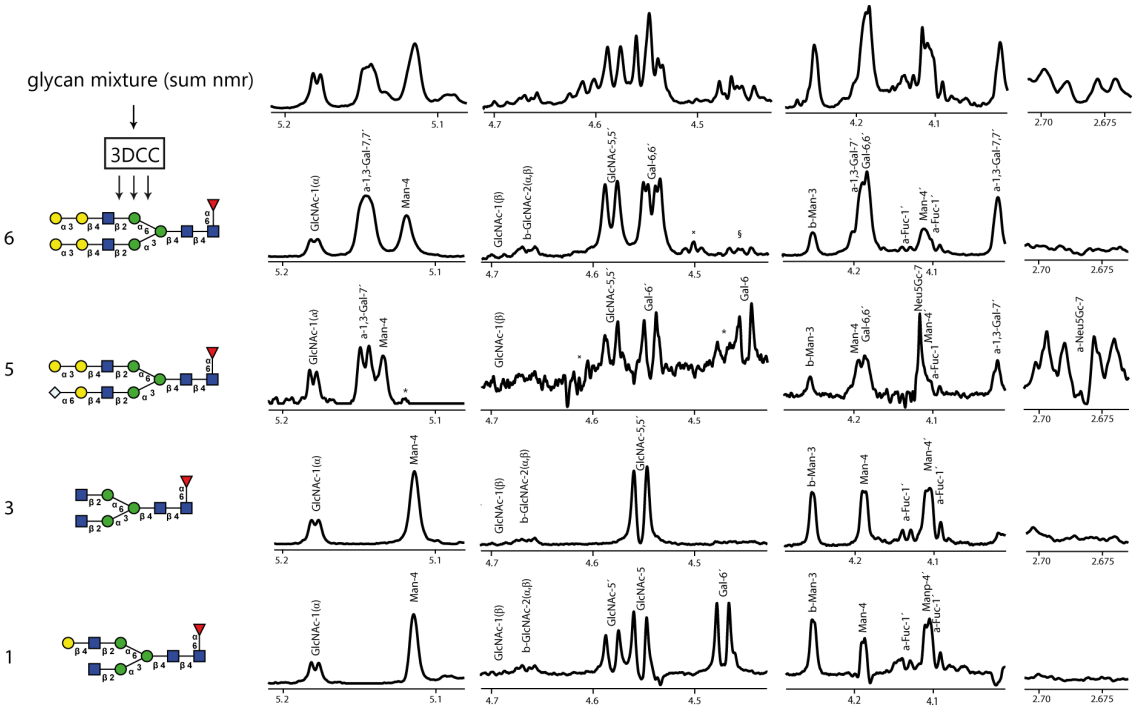


Figure 10: **Extracted NMR spectra as a result of applying 3DCC$_S$. The uppermost spectrum shows the sum of the NMR spectra of all fractions of the glycan mixture. 3DCC mathematically correlates LC-MS information with rising and falling NMR signals to extract NMR spectra of pure compounds. In total, ten structures were unambiguously identified; four of these are shown. Critical moieties were identified by characteristic chemical shifts, e.g. *N*-glycolyl neuraminic acids at 2.684 (H-3e) and 4.119 (NeuGc) or α-(1→3)-Gal epitopes at 5.147 (H-1). *: Minor signals of simultaneously eluting structure in the LC run with the composition Hex$_6$HexNAc$_4$dHex$_1$ (glycan structure with very low abundance). §: Minor signal of a glycan (Hex$_5$HexNAc$_4$dHex$_1$NeuGc$_2$) that is overlapping in the elution profile. ^{x}: Impurity probably resulting from a short peptide that is simultaneously eluting.**

The data presented here demonstrate clearly the benefit of combining NMR and LC-MS for an unambiguous identification and characterization of glycan structures. Using this technology on a routine basis could avoid problems that arose in the case of Cetuximab. 3DCC yields a reliable determination of critical epitopes and provides an improved quantification. Using the combination of NMR and LC-MS, it is easy to analyze and quantify the immunogenic α-(1→3)-Gal and the *N*-glycolyl neuraminic acid epitopes. 37 *N*-glycan compositions were identified by LC-MS(/MS). Subsequently, 10 *N*-Glycans strongly overlapping in chromatographic separation and even mass isobaric structures with equal monosaccharide composition were unambiguously identified (cf. Figure 11). It was possible to extract NMR spectra of pure *N*-glycans that were heavily overlapping

in a chromatographic separation by mathematically dissecting the NMR spectra obtained from chromatographic fractions. NMR is certainly less sensitive than LC-MS, but it allows identification of compounds down to about 15 pmol (approximately 30 ng of a biantennary complex type oligosaccharide). Given the quantities available from therapeutic antibodies, it is easy to perform this type of analysis on a routine basis for quality assurance of the products. The 3DCC procedure was improved in terms of robustness and effectivity by implementing SVD (*singular value decomposition*) as an optimized scheme compared to the least square technique. The quality of the new method in comparison to the previously published version is being demonstrated for the glycans of Cetuximab to provide new insights into biological relevant glycan structures. The smallest amount of the *N*-glycan characterized here by 3DCC was approximately 400 pmol (836 ng). Among the ten unambiguously identified glycans, six *N*-glycans possess the immunogenic α-(1→3)-Gal epitope and/or *N*-glycolyl neuraminic acid. Furthermore, 19 from the 37 detected *N*-glycans by MS(/MS) analyses - representing above 50% of all detected glycans - possess immunogenic units. These results illustrate the importance of integrated use of LC-MS(/MS) and ^{1}H-NMR for the glycome analysis of biopharmaceuticals in research, development and quality control.

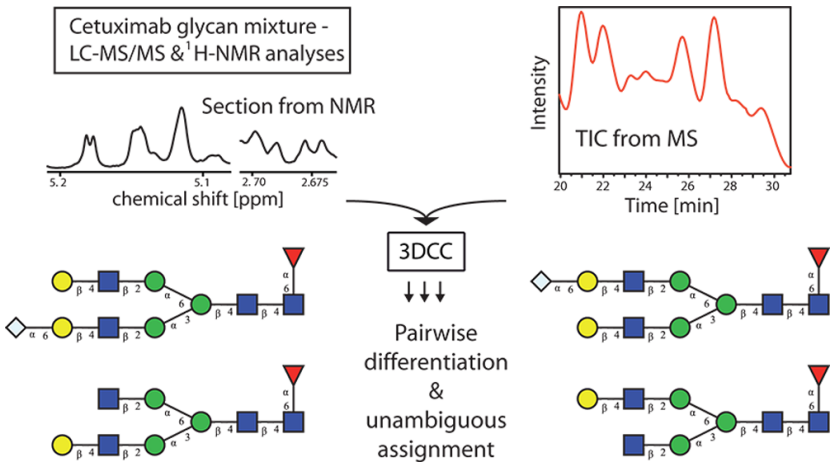


Figure 11: **37 *N*-glycan compositions were identified by LC-MS(/MS). Subsequently, ten abundant structures were structurally characterized by applying the recently introduced method called three-dimensional cross correlation (3DCC). It was possible to extract NMR spectra of pure *N*-glycans that were heavily overlapping in a chromatographic separation by mathematically dissecting the NMR spectra obtained from chromatographic fractions. Even mass isobaric structures that differ only in the branching position of one monosaccharide unit were distinguished and characterized.**

5.1.2 Glycosylation Analysis on the Level of Intact Proteins

More recently, top down approaches that directly analyze the intact glycoprotein by (LC-)MS/MS techniques have gained ground.[104] The analysis of intact proteins offers several advantages compared to bottom up approaches. Commonly used methods for bottom up proteomics/glycomics involve e.g. enzymatic digestion, chemical treatment, sample enrichment and purification procedures. Analyzing the glycosylation from the intact glycoprotein itself decreases time, resources and cost by reducing sample preparation to a minimum. At the same time the risk of sample modification is reduced and data from intact (glyco-)protein analyses obtain more reliability in reflecting the native structure. Another advantage of analyzing the glycosylation of proteins on the intact level is a more consistent quantification of the various attached glycans. The structural change between different glycans has an enormous impact on the ionization ability and may lead to false quantification, when analyzing free *N*-glycans. As the ionization capability of large intact glycoproteins is mainly dependent on the polypeptide part, quantification of glycans and other PTMs is far more reliable.

Further verification of the 37 detected *N*-glycans in the bottom up approach (cf. section 5.1.1) was realized by analyzing intact Cetuximab. Therefore, Cetuximab was purified by C_8-RP-HPLC, and MS data were acquired on an ESI-q/TOF. Dehydration of proteins in the MS is achieved by applying an in-source collision induced decay (ISCID). ISCID voltages from 0-150 eV were tested, and 120 eV was found to be best suited for sufficient MS signals of intact Cetuximab. Higher ISCID voltages were avoided to minimize the risk of fragmentation. The full scan mass spectrum contains multiple charge states of the analytes. Since the intensity is distributed over all charge states ($[M]^{62+}$ to $[M]^{35+}$), the intensity for each peak decreases causing a reduced sensitivity (cf. Figure 12). To gain intensity and improve the signal-to-noise ratio, spectra were summed up over time and MS data were deconvolved by an algorithm using the maximum entropy method (MEM, implemented in Bruker Daltonics Software, version 4.2). The deconvolved spectrum of Cetuximab clearly indicates a comprehensive presence of glycans due to multiple

characteristic mass distances, e.g. 162 Da for hexoses and 147 Da for deoxy hexoses (cf. Figure 12).

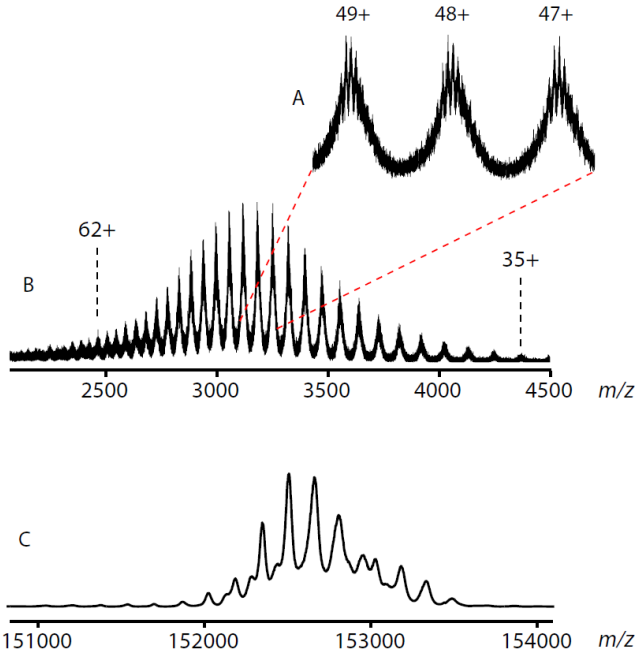


Figure 12: **MS data from intact Cetuximab. Non-treated Cetuximab was purified by a C_8-reversed phase column and directly injected to an ESI-q/TOF MS. MS raw data (B) show multiple charge states of the antibody (unambiguous assignment from charge state $[M]^{62+}$ to $[M]^{35+}$). Every charge state displays the same pattern of signals (A). The deconvolution process (maximum entropy method) generates the singly-charged MS spectrum of intact Cetuximab (C). Multiple mass distances of 162 Da and 147 Da (hexose and deoxy hexoses, respectively) between signal maxima indicate the diversity of attached glycans.**

Due to the presence of four *N*-glycosylation sites on Cetuximab, that can each be occupied by multiple *N*-glycans, signals are heavily overlapping and an unambiguous identification of the *N*-glycans is impossible. However, in biopharmaceutical quality control as well as for the development of biosimilars, it is necessary to perform a fast and reliable examination of different glycoforms of mAbs.

In a precedent work (Behnken, Dissertation, 2013) a MATLAB script was developed for automatic data interpretation of a protein containing one *N*-glycosylation site.[105-106] The script mainly used the following input information: on the one hand the protein sequence depicted as its atomic composition, the number of disulfide bridges and a table containing all possible glycans in terms of their composition of monosaccharides (hexoses, hexosamines, deoxy hexoses, acetyl neuraminic acids, glycolyl neuraminic acids),

glycan phosphorylations as well as glycan sulfations and on the other hand the measured (experimental) spectrum of the full glycoprotein.

For each glycoform the atomic composition was calculated, and a MATLAB-intrinsic function (*isotopicdist*) served to compute the masses and their relative frequencies of all possible isotopes of each glycoform. From this, a realistic theoretical spectrum for each glycoform was calculated by broadening every delta-function peak (related to a single isotope) to a Gaussian peak with a predefined linewidth greater than zero. This procedure as a whole comprised the first step.

In a second step, all these realistic single-glycoform spectra were fitted at the same time to the experimental spectrum by using least square optimization (MATLAB function *lsqlin*). By this, an optimal combination of weighting coefficients for all the possible glycoforms was calculated. Since the iterative least square algorithm will always retain an error, the single-glycoform spectra together with their weighting coefficients were used to calculate one single combined spectrum (joining theoretical and experimental information) which in turn could be used for direct comparison with the fully experimental spectrum. The whole described mathematical analysis was developed for the isotope pattern of prostate specific antigen glycoprotein (PSA, molecular weight of the peptide backbone approx. 26.1 kDa) that exhibits only one *N*-glycosylation site.

For multiply glycosylated proteins, e.g. IgGs, the complexity of the glycosylation pattern is significantly enhanced. Due to numerous glycans that can be covalently attached to two (or even four, in case of a Fab-glycosylation) *N*-glycosylation sites, glycoforms are heavily overlapping. Thus, a robust method for a fast analysis of glycosylation pattern of multiply glycosylated proteins was developed. For that reason, the MATLAB script was extended and modified in multiple directions which are briefly described below. The basic solution structure, comprising single-glycoform spectra with isotopic Gaussian peaks and the least square optimization, however, is still used. The case of the precedent work with only one glycosylation site can still be retrieved.

Firstly, a number S of glycosylation sites greater than one (e.g. S = 2 or S = 4) leads to the fact that combinations of different glycans have to be considered. This greatly enhances the number of possible glycoforms to, in principle, G^S where G is the number of single possible glycans. This number, however, is effectively reduced, but in a non-trivial way:

A) Symmetry properties of the glycoproteins have to be considered: As an example, a symmetric version of a glycoprotein with S = 2 in fact only allows for G*(G+1)/2 different glycoforms. The other glycoforms can be retrieved by pure rotation in space. Symmetry information can be provided to the MATLAB script by stating if the protein in question is a reduced one or not.

B) Since the mass spectrum is insensitive to the glycosylation site on the protein, some glycoforms have to be considered as being equivalent for our analysis method, although they are in fact different molecules and cannot be transferred into each other by pure rotation. An example would be a non-symmetric glycoprotein with S = 2 where different glycans are attached to the two glycosylation sites. The two existing permutations of the two specific glycans are equivalent here.

C) Not on the level of glycans, but on the level of atomic composition, it is possible that completely different molecules become equivalent.

In order to permit a successful least square optimization without arbitrariness, these effects are taken into account.

Secondly, the number of possible different glycoproteins can further be modified by distinguishing between the Fab and the Fc domain: Some glycans might be too large for the Fc domain. A switch variable is provided to either take this into account or not. To the same end, two lists of possible glycans have to be provided as input for the MATLAB script.

Thirdly, besides oxidation states, adduct ions (in the following: "states") can now be included in the analysis, further raising the number of different glycoproteins. Different

options are included how the least squares problem can be solved under these conditions. Either the different "states" are allowed to get their individual sets of weighting coefficients: in this case, they are all fitted to the experimental spectrum at the same time. Or it is assumed that the weighting coefficients of each glycan combination are the same for all the "states": in this case only the single "state" which is assumed to be the most abundant one (e.g. the non-oxidized state without an adduct ion) is used for the least square procedure in a first step by which the relative weights of the glycan combinations are calculated; in a second step, another least square procedure is conducted to get the relative weights of the different "states". Multiplying these two types of relative weighting coefficients yields the desired absolute weights.

Fourthly, due to the observed occurrence of commonly prevalent lysine clipping (cf. Figure 19), the option of multiple protein sequences was implemented. This feature will also be of use when analyzing other glycoforms that occur in several splicing variants. The different proteins, for each of which the atomic composition has to be specified, again raise the number of possible different glycoproteins. They can be treated the same way as the oxidization states and adduct ions above. But another solution strategy is also possible: The ratio of the different protein species can be specified in advance. Before the first least square fitting procedure, the completely theoretical spectra of one specific glycan combination (but for all the different protein species) are combined to one spectrum with a third set of weighting coefficients supplied as input.

Fifthly, the least square optimization itself now is conducted under the constraint that all the weighting coefficients be ≥ 0 (lower bound), since negative factors would be meaningless.

An output file comprising the parameters and settings used, an extensive list of all the different glycans used for the fitting procedure, the calculated weighting coefficients and the different combined spectra is written.

This script was applied to the analysis of Cetuximab, as well as Bevacizumab and is described in the next sections.

The MS spectrum of intact Cetuximab was analyzed in the way described. Figure 13 shows the deconvolved experimental MS spectrum, the theoretical (calculated) spectrum and the difference between both. As glycan input, the glycan structures obtained by bottom up analyses were used plus one additional structure ($Hex_6HexNAc_4dHex_1$) that was identified during middle up analysis (cf. section 5.1.2.2). The small values of the difference spectrum indicate the correctness of the *N*-glycans identified by the bottom up approach (cf. section 5.1.1). 3419 different combinations are possible based on this input. The manual comparison of these combinations with the experimental spectrum is hardly possible. The highest coefficients were obtained for the following combinations of glycans at the four binding sites:

1) $2*Hex_7HexNAc_4dHex_1+2*Hex_3HexNAc_4dHex_1$

2) $2*Hex_7HexNAc_4dHex_1+1*Hex_3HexNAc_4dHex_1+1*Hex_4HexNAc_4dHex_1$,

3) $2*Hex_7HexNAc_4dHex_1+2*Hex_4HexNAc_4dHex_1$.

These structures match with the most abundant structures obtained by the bottom up analysis (cf. 5.1.1) as well as the middle up approach described in section 5.1.2.2. Note that some of the compositions can also be explained by other mass isobaric combinations, e.g. $2*Hex_4HexNAc_4dHex_1$ can also be provided by $1*Hex_3HexNAc_4dHex_1$ and $1*Hex_5HexNAc_4dHex_1$. For simplicity, these were not separately listed above (cf. 2).

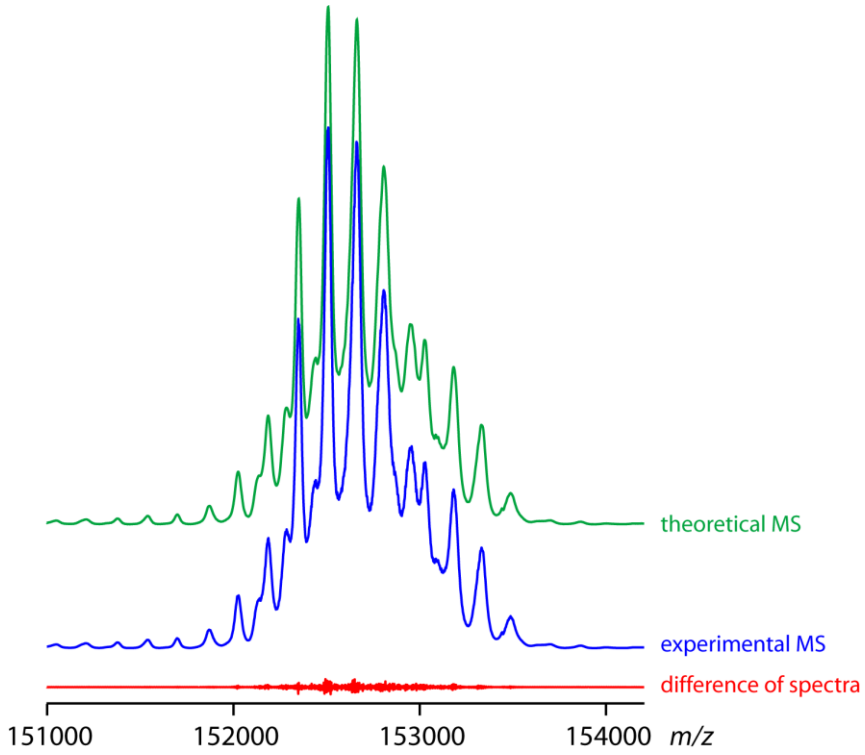


Figure 13: **Deconvolved MS spectrum of intact Cetuximab. For automated data interpretation, a MATLAB script was written, that calculates theoretical spectra by inclusion of isotopic patterns of antibodies with four glycosylation sites (arbitrary *N*-glycans can be attached) and of oxidized states as well as adduct ions. The figure displays the calculated, the experimental and for easier comparison the difference spectrum of the former two. For the theoretically calculated spectrum 38 *N*-glycan structures were considered. The flat difference spectrum containing only small values displays the high agreement between theoretical and experimental spectra and verifies therewith the presence of the previously detected glycan structures.**

Nonetheless, it should be possible to quickly and reliably identify changes of the glycosylation pattern, e.g. within quality control of biopharmaceuticals or biosimilar development, where glycan homogeneity among all batches is required. If a difference in the profile is detected, coefficients in the output data file give information about the glycan combinations that are affected.

To evaluate the efficiency of the data analysis script for quality control of biopharmaceuticals, several batches of Bevacizumab (humanized mAb, type IgG_1), that is used for treatment of cancer, were analyzed.[107-111] Bevacizumab holds two *N*-glycosylation sites, one at each half of the Fc domain (glycosylation site: ^{303}**N**ST^{305}).

Figure 14 shows the MS spectra of intact Bevacizumab. Peaks constitute different glycoforms. The overall glycosylation pattern looks similar for all three batches. However, especially batch 1 differs already visually in relative intensities of some peaks. Analysis of sub-domains of mAbs improves identification of all attached glycan structures. Therefore, a middle up approach was developed, which is described in the next section.

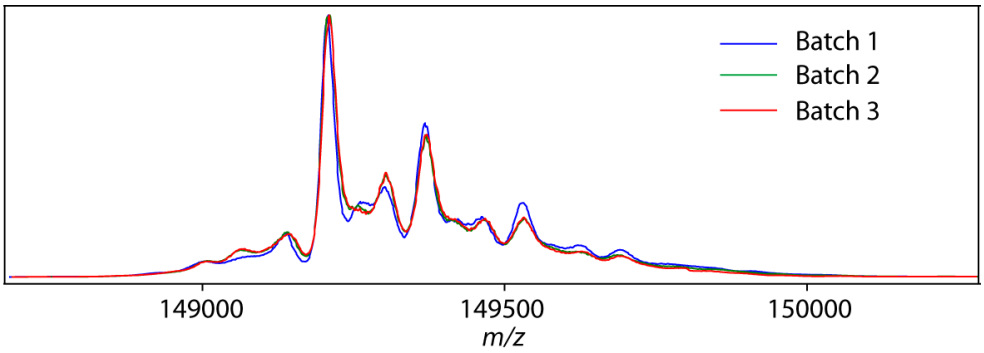


Figure 14: **MS spectra of intact Bevacizumab. Three different batches of the mAb were analyzed by RP-LC-MS. Peaks constitute different glycoforms. Since Bevacizumab holds two *N*-glycosylation sites, every peak is composed of two *N*-glycans. The overall glycosylation patterns look similar for all three batches. However, especially batch 1 differs in height.**

5.1.2.1 Middle Up Approach for Analysis of Monoclonal Antibodies

For a more detailed and site-specific analysis of the *N*-glycosylation of mAbs a middle up approach was established. By splitting the molecule into large sub-fragments an accurate mass determination as well as a better assignment of the *N*-glycans should be possible. At the same time, the sample preparation is still significantly faster than the complete protein digestion and release of *N*-glycans in a bottom up approach. Figure 15 depicts the techniques developed for middle up analysis. In one approach, the mAb is digested with papain, that cleaves N-terminal to the hinge region resulting in the intact Fc domain as well as two Fab/2 domains (cf. Figure 15 A2). These sub-domains were either directly analyzed, i.e. 2*Fab/2+1*Fc, or after reducing the disulfide bridges with DTT. This procedure separates the light chain from the N-terminus of the heavy chain (Fd) as well as the Fc domain (cf. Figure 15 A1). In another approach, the intact IgG was directly treated with DTT (or other reagents such as tris(2-carboxyethyl)phosphine) to

reduce the disulfide bridges resulting in the separation of the light and the heavy chains (cf. Figure 15 B).

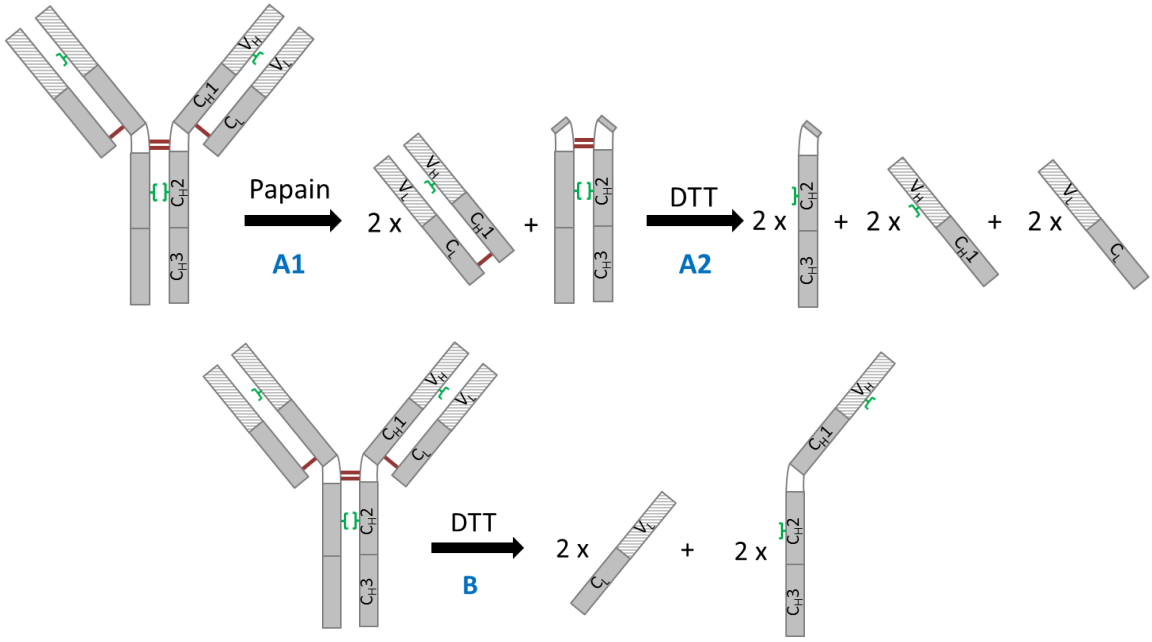


Figure 15: **Middle up approach for analysis of the *N*-glycosylation of mAbs. Various workflows were established to cleave mAbs into sub-domains for a more detailed and site-specific glycan characterization. Papain cleaves N-terminal to the hinge region resulting in the intact Fc domain as well as two Fab/2 domains (A1), thus separating the *N*-glycosylation sites localized in the Fc- and the Fab-domain. To further reduce size and number of glycosylation sites in the Fc domain, inter-domain disulfide bridges can be cleaved by reducing agents, leaving ~25 kDa big sub-domains with a maximum of one *N*-glycosylation site (A2). Another approach includes the reduction of the intact IgG resulting in a separation of the light and the heavy chain (B). This reduces the MW of the sub-domains and the number of glycosylation sites per protein domain. Green: *N*-glycan, red: inter-domain disulfide bridge. Note that only inter-domain disulfide bridges are indicated. Cetuximab contains 16 disulfide bridges in total.**

5.1.2.2 Middle Up Analysis of Cetuximab for Glycosylation Analysis

Within quality control of biopharmaceuticals, the *N*-glycosylation profile has to be analyzed by reproducible analytical methods. Therefore, the reliable generation of sub-domains has to be ensured. Thus, the digestion of mAbs with papain was optimized.

The cysteine-protease papain is known to cleave unspecifically within the Fc domain. Therefore, conditions had to be established for optimal mAb digestion. Several parameters for digestion were varied: 1) enzyme-to-protein ratio, 2) concentration of

mAb during digestion, 3) buffer system and 4) time of incubation. Figure 16 summarizes the results from these variations. All approaches were analyzed by RP-LC-MS. The Fc domain elutes at retention time (t_r) 21 min, thus prior to the two Fab/2 domains at t_r = 25 min. Cetuximab was cleaved by papain between the Fc and the two Fab domains C-terminal to ^{226}H. Integrity of mAb digestion as well as other possibly occurring side products were also analyzed.

The enzyme-to-protein ratio was evaluated, to ensure a fast cleavage while retaining a highly specific cleavage. Comparison of different ratios clearly reveals that an enzyme-to-protein ratio of 1:100 (D) is too low to ensure a quantitative cleavage. No Fc domain is observable after one hour of digestion. Initial formation of the Fc domain could only be observed after six hours of incubation. After 25 hours, a fragment of ~100 kDa is still observable originating from an incompletely digested mAb (t_r = 24 min). Contrary, a ratio of 1:10 resulted in a complete digestion after six hours. No further degradation is observed after 25 hours (A). In Figure 17 the cleavage efficiency is monitored as a function of time (for parameters see Figure 16A).

The effect of the overall protein concentration (enzyme and mAb) was evaluated at a fixed enzyme-to-protein ratio of 1:10. As expected, an increased concentration resulted in a faster cleavage. For a mAb concentration of 2.5 µM (C) a quantitative cleavage was observed after 25 hours, for a mAb concentration of 5 µM (A) after six hours and for a mAb concentration of 10 µM (B) after three hours. The chromatogram of the highest concentration (B) shows an increase of a signal (t_r = 19.5 min) that consists of undesired cleavage products. The chromatogram at the lowest concentrations (C) shows undesired peaks that overlap with the signal of the Fc domain.

To evaluate the effect of the cleavage buffer Tris-HCl (A) was compared to sodium phosphate (E). Both buffers contain EDTA and L-cysteine for papain activation. Digestion in Tris-HCl buffer resulted in a quantitative cleavage after six hours. In contrast, digestion in sodium phosphate buffer was completed after two hours. No other

signals were observed, thus the reaction is significantly enhanced and more appropriate for a fast analysis.

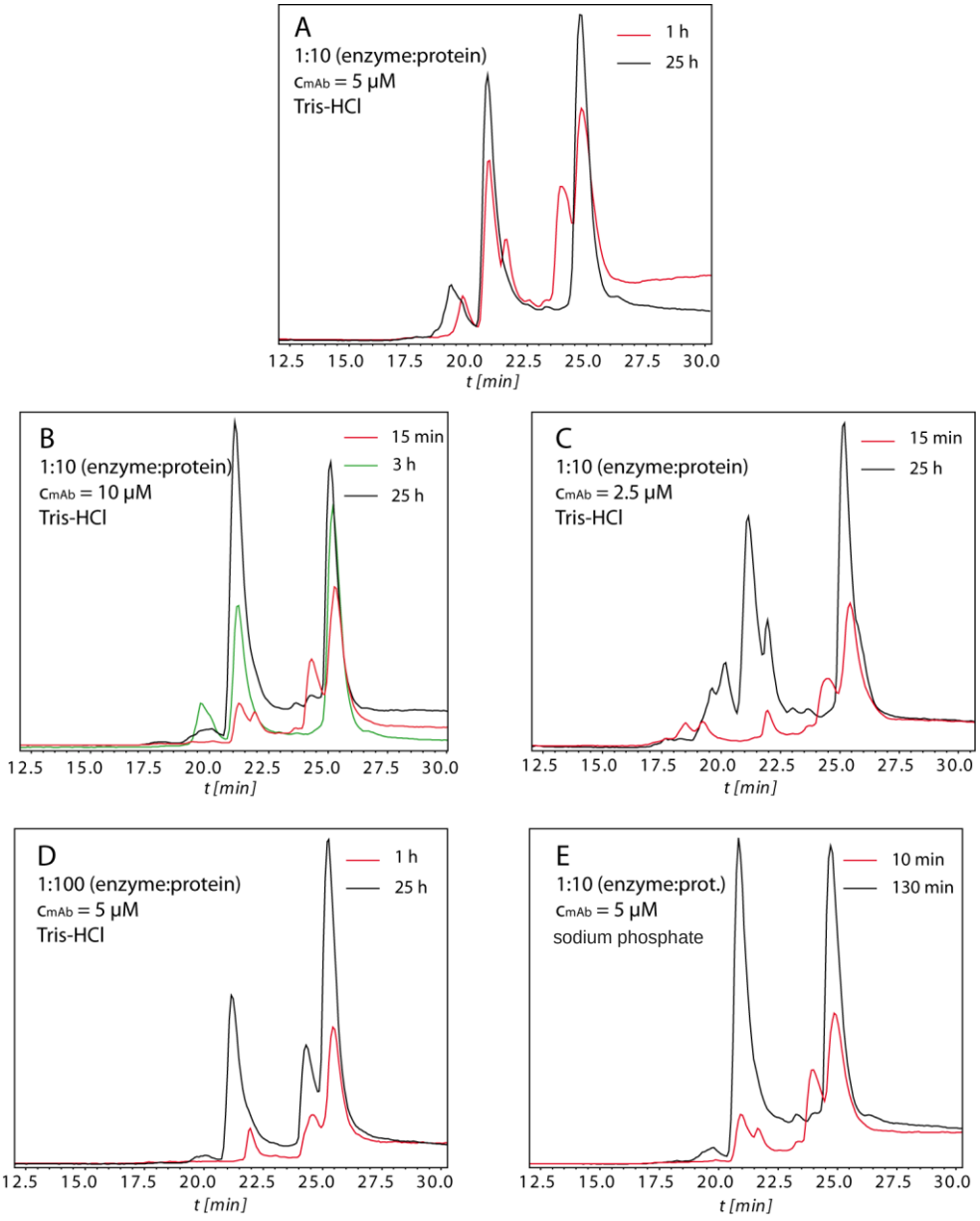


Figure 16: **Total ion chromatograms from LC-MS analyses of papain-digested Cetuximab. For reliable interpretation of MS spectra of mAb sub-domains, a consistent cleavage with papain is required. Parameters such as incubation time, reaction buffer, enzyme-to-protein ratio as well as protein concentration during digestion were varied (details given in the individual chromatograms). For a detailed comparison see text.**

To ensure a complete digestion, the following parameters were used for further middle up experiments: enzyme-to-protein ratio of 1:10, mAb concentration of 5 µM, Tris-HCl as digestion buffer. Tris-HCl was chosen as a buffer system to prohibit unspecific

cleavages at longer incubation times during the following steps (e.g. further processing such as reduction).

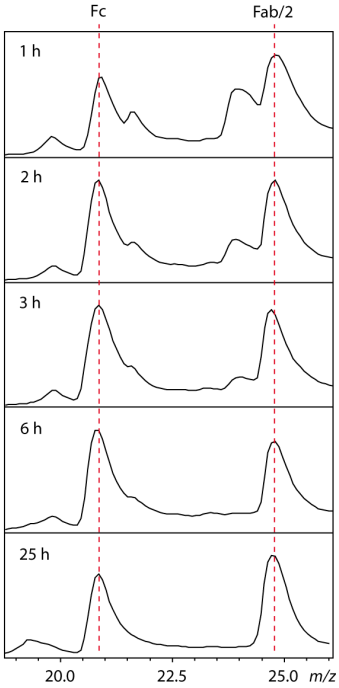


Figure 17: **Total ion chromatogram of LC-MS analyses of papain-digested Cetuximab. Amount of cleavage was monitored as a function of time. After six hours of incubation, the main part is cleaved into the Fab and the Fc domain. In the first time points several side products are observable (e.g. partly uncleaved mAb, t_r = 24 min).**

Following these optimizations, the domain-specific analysis of glycans (attached to the Fc or the Fab domain) can be obtained. While only one glycan is attached to the Fab/2 domain, the Fc domain still carries two oligosaccharides. To ensure a more detailed Fc characterization, a cleavage of the disulfide-linked heavy chains is required. Separation was evaluated with varying concentrations of DTT as reduction reagent. Figure 18 shows the obtained sub-domains analyzed by LC-MS (reduction of papain-digested Cetuximab in 450 mM DTT for 30 min at 37 °C). All sub-domains were separated by RP-LC. The Fc/2 domain elutes in a broad peak caused by incomplete C-terminal lysine-clipping and partial disulfide rearrangement (cf. Figure 19).

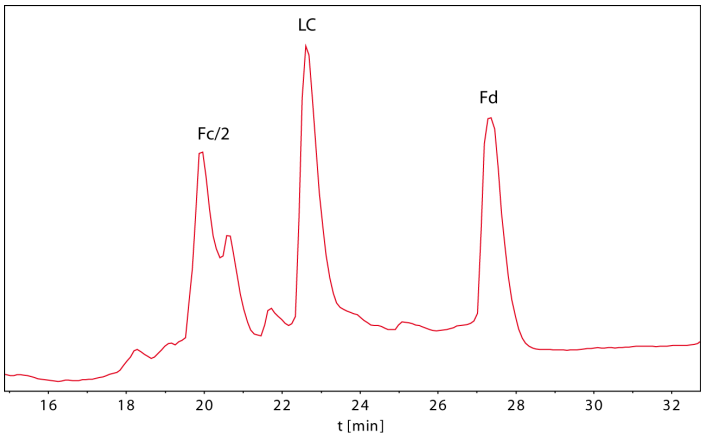


Figure 18: **Total ion chromatogram of LC-MS analysis of sub-domains obtained by a papain digestion of Cetuximab followed by a reduction of disulfides. Fc/2: half Fc domain (reduced Fc domain), LC: light chain, Fd: heavy chain part of the Fab domain. The Fc/2 sub-domain elutes in a broad peak caused by partial lysine clipping and disulfide rearrangement.**

The separation of the sub-domains now allows a glycosylation site specific analysis. Furthermore, a maximum of one glycan structure is attached to each protein fragment. Data analysis was done by least square optimized fitting of the glycosylation pattern. Only major signals were annotated in the MS spectra (cf. Figure 19 and Table 17 in appendix). It is observed that the large glycans, i.e. triantennae as well as complex type glycans with α-(1→3)-Gal epitopes and/or *N*-glycolyl neuraminic acids, are attached only to the Fab domain. Also, the microheterogeneity of the glycans on the Fab domain is enhanced compared to that of the Fc domain. In the Fc spectra for each glycoform two signals occur in the MS spectrum. This is due to C-terminal clipping of lysine, a common modification of mAbs. However, this clipping is incomplete (about one part native and three parts lysine-clipped) and will be considered during data interpretation. Therefore, possibility of using several protein species for fitting was included into the MATLAB script. If the ratio is equal for all glycoforms, the ratios can be kept fixed. Coefficients are listed in the appendix (cf. Table 17).

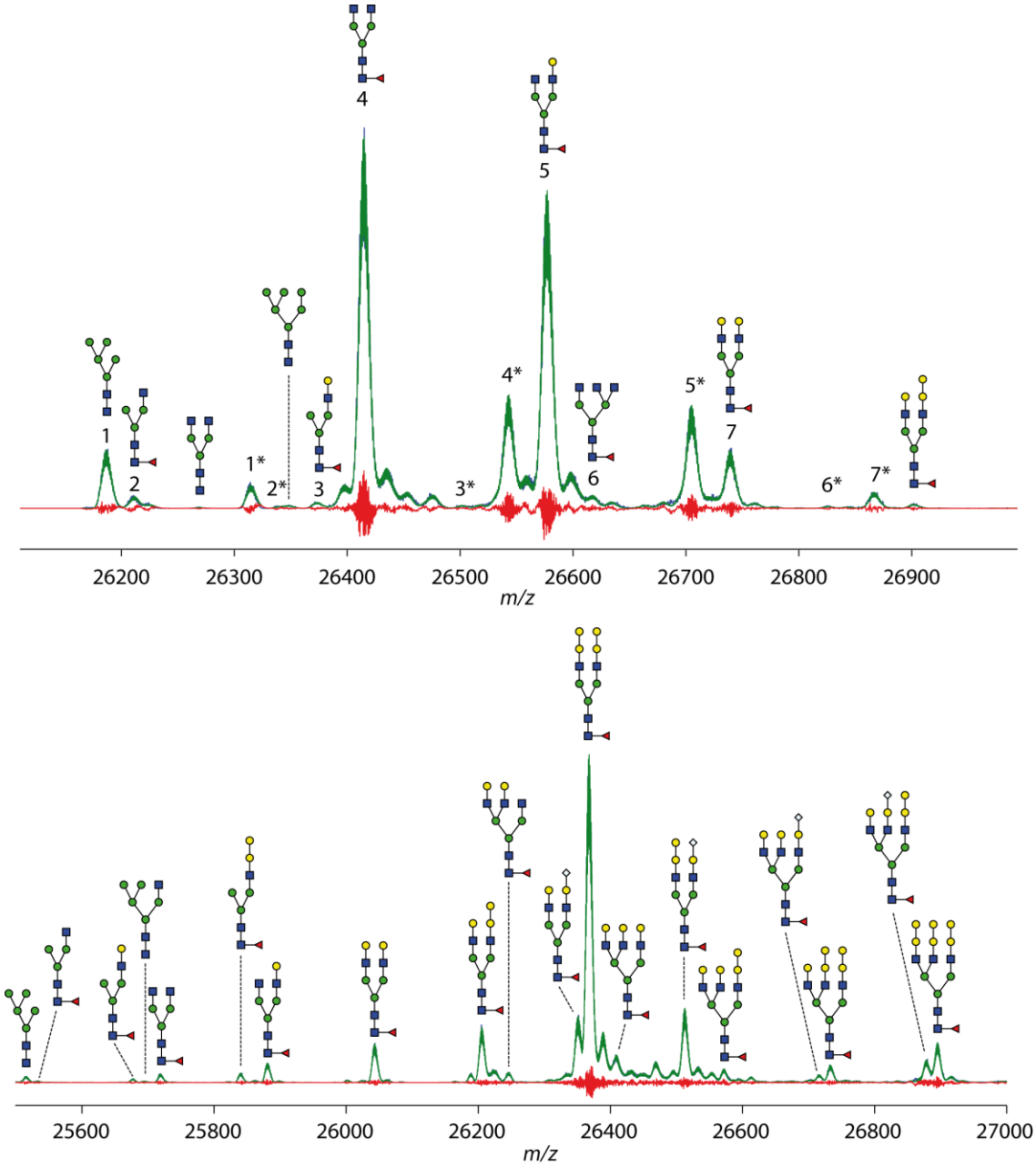


Figure 19: **Deconvolved MS spectra of the pure Fc/2 domain (top) and the Fd domain (bottom) of Cetuximab. Domains were obtained by papain digestion followed by reduction of disulfide bridges. Thus, only one glycosylation site per protein fragment exists. Analysis was performed by RP-LC-MS. Data was interpreted by fitting the isotope pattern of the individual glycoforms using a least square optimization. The dominant glycan structures are annotated. *: For analysis of the Fc/2 domain partial lysine clipping was observed. Thus, every glycoform occurs twice (Fc/2 with and without lysine). Blue: experimental spectrum (hidden behind green), green: theoretical spectrum, red: difference spectrum.**

The non-reduced domains can be analyzed in the same manner. The same glycosylation profile is observed with the discrepancy, that it is shifted to higher masses. In fact, only the mass of the (non-fucosylated) light chain has to be added. Another glycosylation

pattern is observable for the Fc domain (cf. Figure 20). This is due to the fact, that two *N*-glycosylation sites are present. Thus, every signal in the MS spectrum is constituted by a combination of two *N*-glycans. Coefficients are listed in Table 18 in appendix.

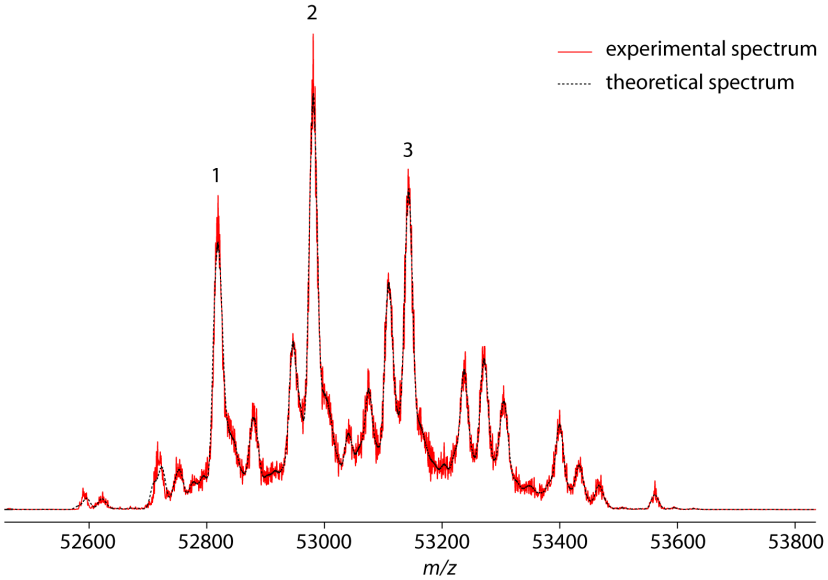


Figure 20: **Deconvolved MS spectra of the pure Fc domain of Cetuximab. The domain was obtained by papain digestion. Two glycosylation sites are present. Data was interpreted by least square optimized fitting of the isotope pattern of the individual peaks. For data interpretation, partial lysine clipping had to be considered. Thus, every glycoform combination occurs three times (both Fc/2 chains with lysine, both without lysine and one chain with and one without lysine). Major peaks can be explained by the Fc domain without lysines with the following attached glycan combinations: 1) two $Hex_3HexNAc_4dHex_1$, 2) one $Hex_3HexNAc_4dHex_1$ combined with one $Hex_4HexNAc_4dHex_1$, 3) two $Hex_4HexNAc_4dHex_1$. Red: Experimental deconvolved MS spectrum, black dotted: theoretical spectrum obtained by a mathematical fit.**

Another possibility of generating sub-domains is the reduction of disulfide bridges without papain digestion. Reduction of the disulfide bridges leads to a decreased molecular mass resulting in an easier mass determination due to an increased resolution as well as a less complex glycosylation patterns, since each heavy chain carries two glycans instead of four for the intact mAb (cf. Figure 15 B). After treatment with 1,4-D/L-dithiothreitol as reducing agent, the heavy and light chains were separated in a reversed phase column (C_8-LC) and directly injected into the ESI-q/TOF (cf. Figure 21).

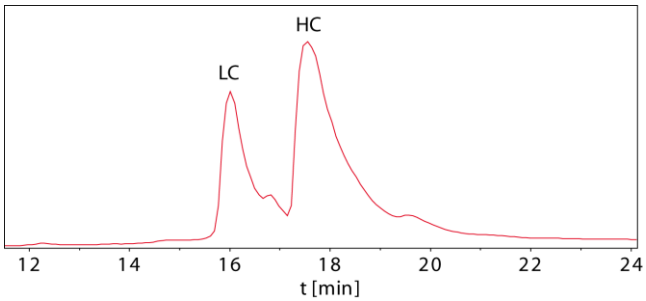


Figure 21: **Total ion chromatogram of LC-MS analysis of reduced Cetuximab. The chains were separated by reduction of disulfide bridges by addition of DTT followed by by RP chromatography. LC: light chain, HC: heavy chain.**

Figure 22 shows the deconvolved spectrum of the heavy chain (blue) and the calculated spectrum (red) obtained by the MATLAB script for automatic fitting. Using the information from above one obtains 305 different masses that result in peaks. Differences can easily be detected and located to several glycan compositions. However, this approach is more appropriate for mAbs that contain only glycosylation within the Fc domain, such as Bevacizumab whose analysis is described in the next section.

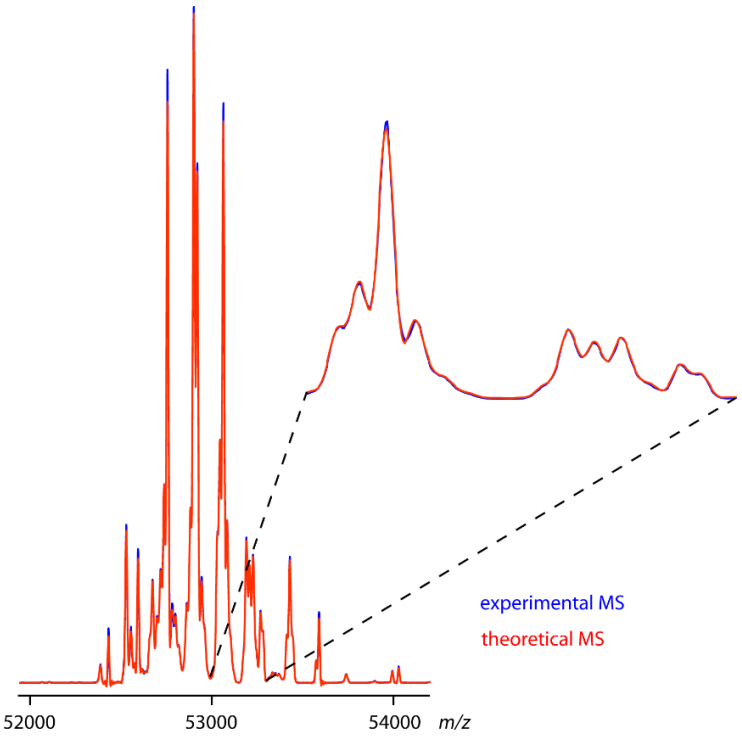


Figure 22: **Deconvolved MS spectrum of the heavy chain (HC) of Cetuximab. Cetuximab was reduced to separate light and heavy chains. The protein domains were isolated and analyzed by online C_8-RP-LC-MS. For automatic data interpretation, the experimental MS spectrum of the HC (blue) is used for fitting a theoretical MS spectrum. This multivariate fit yields coefficients for glycan combinations that contribute to each peak. Thus, differences of various batches of recombinant mAbs (or other glycoproteins) can easily be localized and characterized.**

5.1.2.3 Middle Up Analysis of Bevacizumab for Glycosylation Analysis

The same procedure as for Cetuximab was performed for samples of Bevacizumab (different batch numbers from the same producer "Roche"). First, the optimal incubation time was determined. After incubation for 21 hours cleavage is complete. Cleavage occurs specifically at the C-terminal side of ^{232}H.

Differences in glycosylation can easily be detected by manual inspection of the deconvolved mass spectrum (cf. Figure 24). Besides variances in signal heights (that is the relative quantity), several signals are present or absent exclusively for select batches (e.g. between 52.5 and 57 kDa).

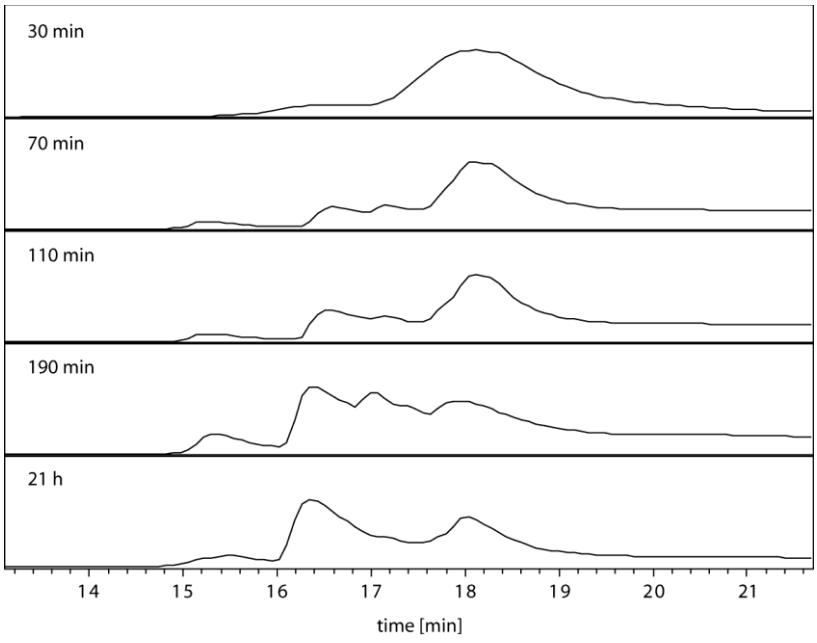


Figure 23: **Total ion chromatogram of LC-MS analyses of papain-digested Bevacizumab. Quantitative cleavage was monitored as a function of time. After incubation for 21 hours, cleavage into the Fab (t_r = ~ 18 min) and the Fc (t_r = ~ 16.5 min) domain is completed. In the chromatograms of previous time points several side products are observable, e.g. partly uncleaved mAb at t_r = minute 24 (not shown).**

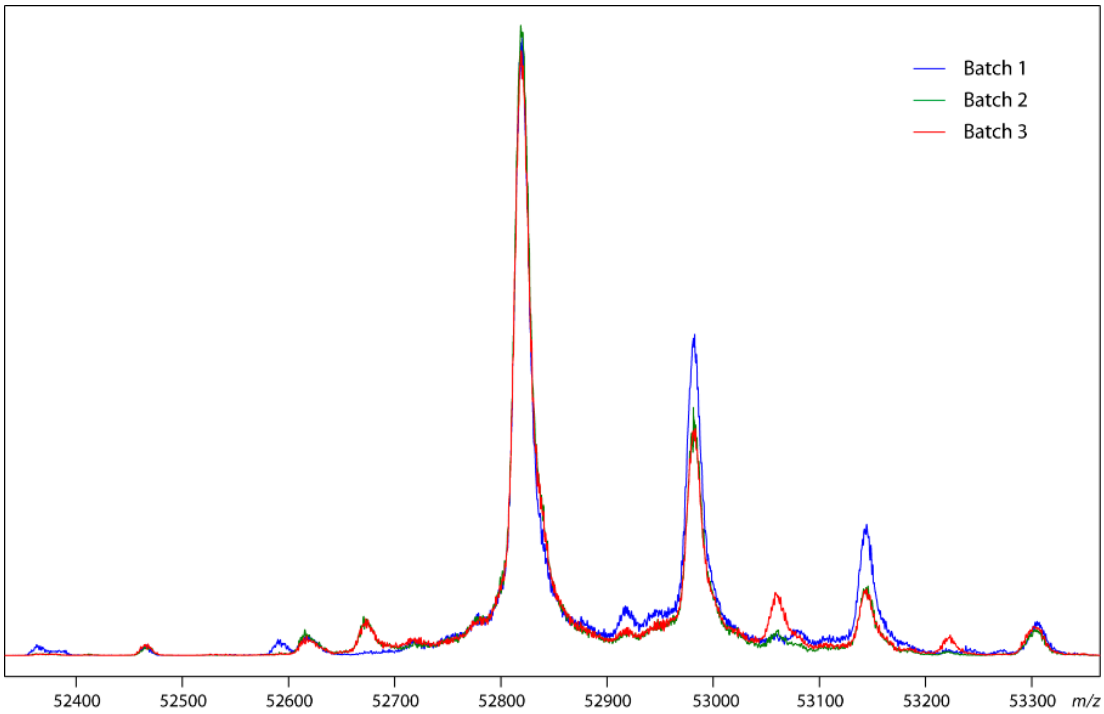


Figure 24: **MS spectrum of Fc domains of three different batches of Bevacizumab. The Fc domain was obtained by cleavage with papain and analyzed by RP-LC-MS. Every peak is composed of two monosaccharides, since each Fc chain holds one *N*-glycosylation site. An overall similarity of glycan structures is observable. However, several signals significantly differ between batches. Furthermore, the relative quantity (= height) of glycoforms differs between batches, especially for batch 1 compared to the others.**

After digestion with papain the deconvolved spectra of the Fc domain were analyzed by means of the MATLAB script (no glycosylation on the Fab domain). C-terminal lysines are not present. With six disulfide bridges and an atomic composition of $C_{2236}H_{3460}N_{588}O_{674}S_{18}$ the spectra were fitted with two occupied glycosylation sites (one per HC). Table 4 summarizes the eight *N*-glycans used for automatic fitting yielding in 30 possible different monosaccharide compositions. The MATLAB output revealed an overall similar glycosylation pattern of the different batches. However, several discrepancies were observed (cf. Figure 25).

Table 4: **Monosaccharide composition of *N*-glycans that were used for automatic data interpretation.**

Glycan Composition	n(Hex)	n(HexNAc)	n(dHex)
1	3	4	1
2	3	3	1
3	3	4	0
4	5	2	0
5	6	2	0
6	4	3	1
7	4	4	1
8	5	4	1

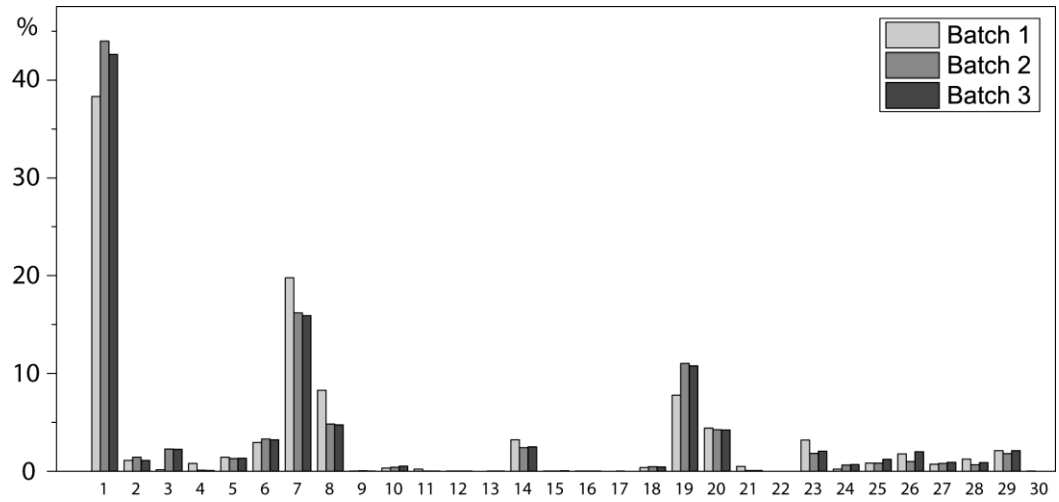


Figure 25: **Distribution of glycan combinations attached to the Fc domain of three different batches of Bevacizumab. Fc domain was obtained by papain digestion. Fc and Fab domain were separated and analyzed by RP-LC-MS. Eight *N*-glycan structures were automatically combined to all possible pairs which effectively yielded 30 possible different monosaccharide compositions. Every number (x-axis) constitutes one combination (cf. Table 4). While an overall similarity is observable, distinct differences can be identified. See Figure 26 for a more detailed view and discussion of preferable and undesired glycan structure attached to Bevacizumab.**

Two structural properties were considered in more detail. On the one hand glycan structures that are not core-fucosylated are preferable glycans because of their antibody-dependent cell-mediated cytotoxicity (ADCC)-enhancing character. On the other hand, high mannose structures are undesired structures, because they enhance mAb clearance. This results in a higher antibody dose required for efficient treatment, which can also

enhance the risk of undesired side effects for patients. Figure 26 summarizes (A+B) and contrasts (C) the amount of such preferable and undesired structures in the measured batches (cf. Table 19 in appendix for combination number). The results indicate that batch 1 contains a suboptimal glycosylation pattern compared to batch 2 and 3. Batches 2 and 3 seem to be more appropriate biopharmaceuticals in terms of efficiency.

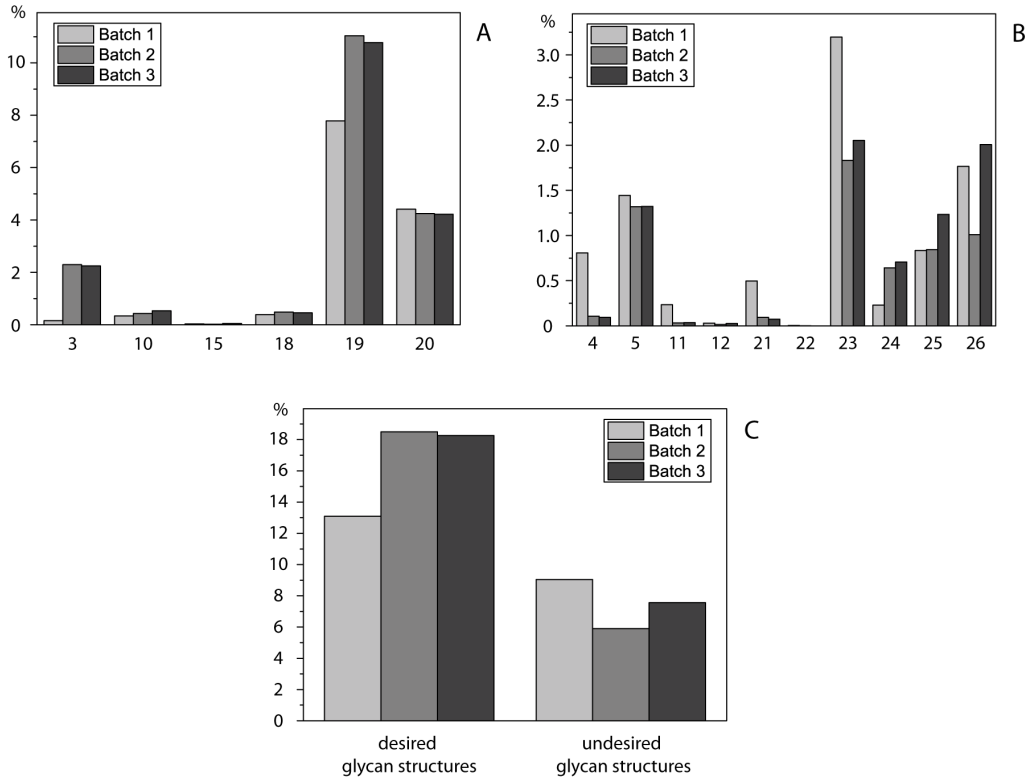


Figure 26: **Detailed overview of desired and undesired glycan combinations with respect to their monosaccharide composition attached to the intact Fc domain, i.e. two glycans per domain. A: Relative amounts of preferable glycan structure combinations, B: relative amounts of undesired glycan structure combinations, C: summary of undesired and undesired structures attached to different batches. See Table 19 in appendix for combination numbers. Combinations that contain non-core fucosylated complex type *N*-glycans enhance ADCC and are therefore preferable combinations. Combinations that contain high mannose type *N*-glycans are undesired due to their clearance enhancing character. Batches 2 and 3 hold preferable structures for an efficient medical treatment compared to batch 1.**

By analyzing the intact Fc domain, one obtains the monosaccharide composition for each peak in the MS spectrum. Each peak represents two glycans, thus, the distinct amount of a single glycan cannot be determined. Therefore, the Fc domain was separated by reducing the disulfide bridges by addition of DTT. The spectra were interpreted by means of fitting the experimental isotope pattern using a least square optimization. For

coefficients see Table 5. Major glycans are $Hex_3NAc_4dHex_1$ followed by $Hex_4HexNAc_4dHex_1$ that together constitute over 90% of all glycans. Structures with preferable/undesired monosaccharide units are minor, but the same tendency as detected for the intact Fc domain can be verified here: while batch 1 contains undesired structures in the ninefold amount compared to batches 2 and 3, the preferable non-fucosylated structure is fivefold more abundant in batches 2 and 3.

Table 5: MATLAB coefficients for Fc/2 domains of reduced, papain digested Bevacizumab.

Glycan	n(Hex)	n(HexNAc)	n(dHex)	Batch 1	Batch 2	Batch 3
1	3	4	1	67.01	74.79	74.24
2	3	3	1	1.97	1.84	1.99
3	3	4	0	0.40	2.05	2.03
4	5	2	0	1.82	0.23	0.35
5	6	2	0	0.77	0.39	0.45
6	4	3	1	2.35	2.25	2.43
7	4	4	1	23.29	17.29	17.25
8	5	4	1	2.38	1.16	1.26

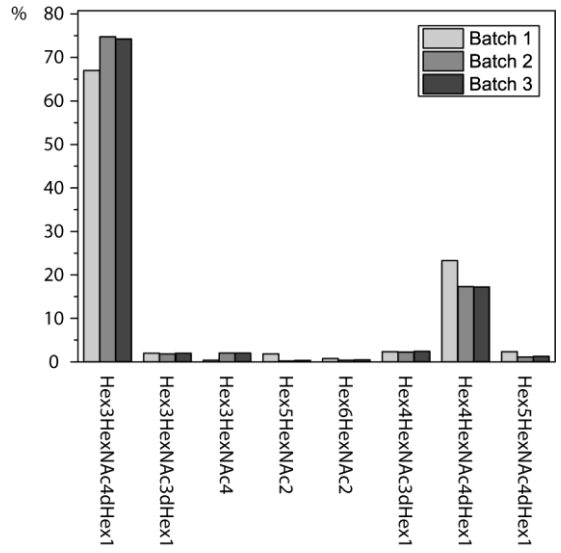


Figure 27: Detailed overview of preferable and undesired single glycan structures with respect to their monosaccharide composition attached to the Fc/2 domain (thus one glycan per protein). Structures that are non-core fucosylated complex type *N*-glycans enhance the ADCC and are therefore preferable. Structures that are high mannose type *N*-glycans are undesired structures due to their clearance-enhancing character. Batches 2 and 3 hold preferable structures for an efficient medical treatment compared to batch 1. Preferable structures: $Hex_3HexNAc_4$, undesired structures: $Hex_5HexNAc_2$, $Hex_6HexNAc_2$.

5.1.3 Interlaboratory Study on Glycan Analysis

The National Institute for Standards and Technology (NIST) invited selected laboratories associated to glycan analysis to participate in an interlaboratory study. NIST distributed two samples to all participating laboratories: the original mAb that is developed for a later use as a well-characterized mAb reference material as well as a modified version of this monoclonal antibody. The goal was to characterize and to compare the *N*-glycosylation of the mAbs for determining the variability of both, the identification and the quantification of *N*-glycans across different institutions and applied techniques.

Here, the methods described above were applied, to perform a multi-dimensional approach for a broad and detailed characterization. Several techniques such as MS of the intact, the reduced and the papain-digested protein were combined with MS and MS/MS (ESI and MALDI MS) analyses of the released *N*-glycans. About 50 pmol of each NIST sample were analyzed by C_8-RP-LC-MS, to obtain information of the sample´s properties, e.g. purity, concentration, overall protein and/or PTM pattern and differences between the two samples NIST_A and NIST_B. The TICs reveal that the overall composition is the same for both samples (cf. Figure 90). The MS spectra of the intact mAbs are shown in Figure 28. Charge states of +69 to +43 for NIST_A and +68 to +42 for NIST_B appear in the *m/z* range of about 2200 to 3400 (Figure 28 A). Each charge state represents a PTM pattern of the IgG (Figure 28 B) which is dominantly composed of glycosylation due to characteristic mass distances. Deconvolution of all charge states results in a calculated MS spectrum of the neutral (non-charged) analytes with a better S/N ratio compared to individual charge states (Figure 28 C). Both NIST samples show a similar overall glycosylation pattern with different amounts of the oligosaccharides.

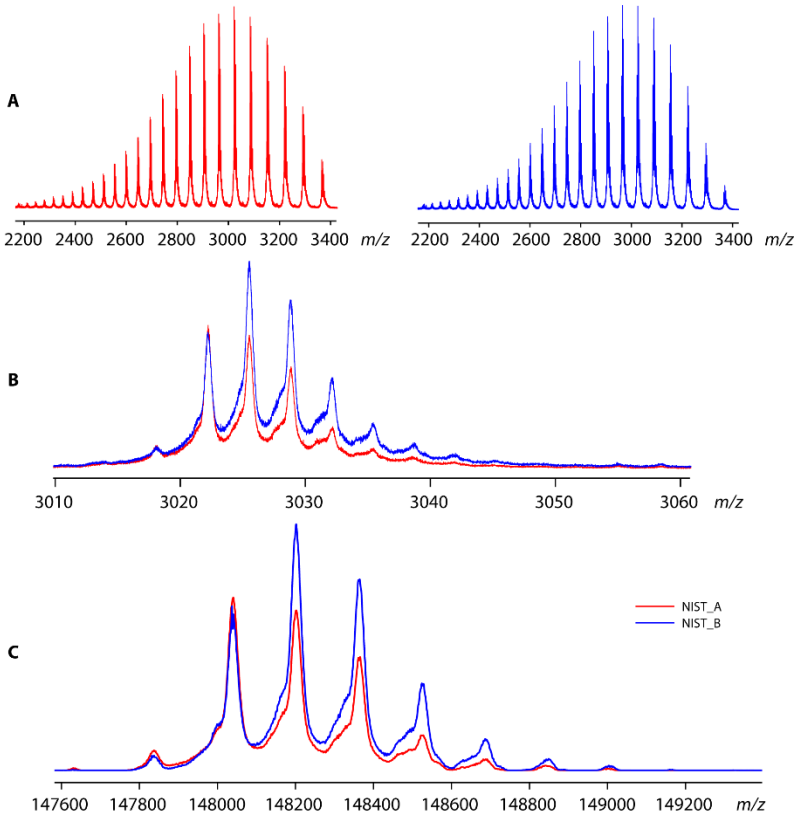


Figure 28: **Raw and deconvolved MS spectra obtained from intact NIST mAbs. A: The top panel shows the full scan MS spectra of the mAbs summed up over a time period of two minutes. The number of signal peaks arises from the multiple charge states of the proteins. Charge states from +69 to +43 for sample A and +68 to +42 for sample B are observable. B: Each charge state represents the overall peak pattern of the protein, e.g. the distribution of PTMs such as glycosylation. C: Deconvolving these multiple charge states improves the S/N ratio. The distances between maxima agree with masses of monosaccharides Thus, spectra represent the glycosylation pattern of the mAbs. Despite the high similarity between the two samples, variances in terms of signal intensity are obvious. Red: NIST_A; blue: NIST_B.**

For determination of the glycosylation site (Fab- or Fc-domain) as well as for obtaining more detailed information about other (post-translational) modifications, structural analyses of smaller sub-domains were performed. Analyses of the intact mAbs as well as middle up analyses were performed. The middle up approach comprises the digestion with papain. Papain cleaves *N*-terminal to the hinge region forming an intact Fc domain as well as two Fab/2 domains. The domains were analyzed as intact protein fragments by online-C_8-RP-MS. The domains can easily be differentiated by integrating the corresponding signals (calculating areas below signal curves) of the TIC, since Fab/2 occurs in twofold amount compared to the Fc domain. The deconvolved spectrum of the

Fab/2 domain (Figure 29, top) shows that the two NIST samples exhibit nearly identical patterns. A peak with a mass that is 162 Da higher than the mass of the most intense peak can be observed with a frequency of 3.9% (calculated by integration), that most likely represents a glycosylation, a frequently occurring modification of recombinant mAbs. Considering the deconvolved MS spectra of the Fab and Fc domains, it is obvious that the glycosylation pattern seen in the MS of the intact mAbs (cf. Figure 28) solely originates from glycosylation of the Fc domain. The peak pattern of the most intense signals of the Fc domain alone is consistent with the most intense signals from the MS of the intact mAb. Smaller signals emerge in a more prominent way due to less overlapping and better S/N ratios and resolution, respectively.

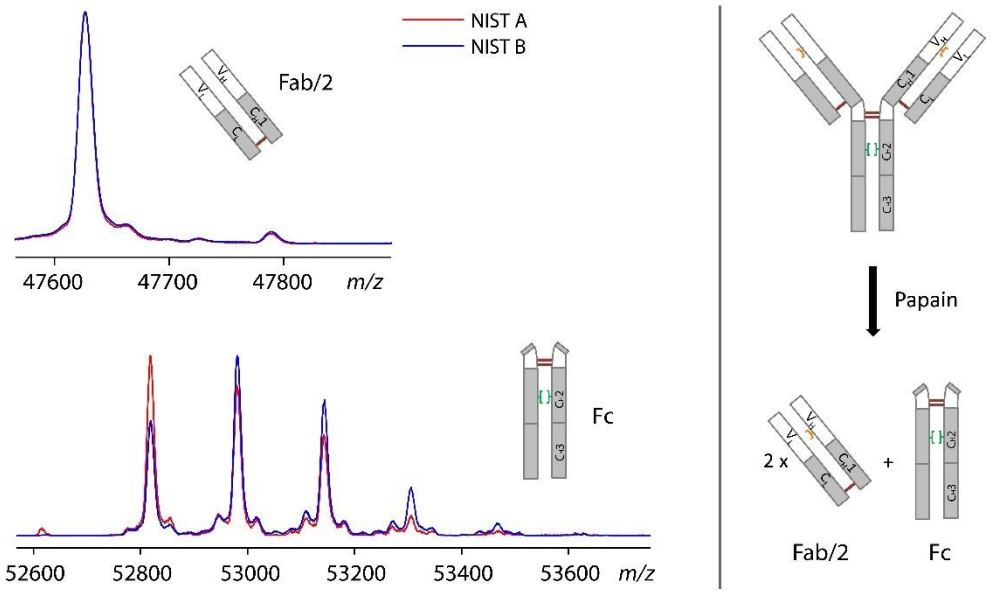


Figure 29: **MS spectra of the products of the digestion of the mAb with papain. Papain cleaves IgGs N-terminal to the hinge region leaving an intact Fc domain and two Fab/2 domains (right: scheme for a two (green bracket) or four (green + orange brackets) times glycosylated IgG). The resulting domains were separated on a C_8-RP-column and analyzed by MS. The deconvolved spectrum of the Fab/2 domain (top) shows a nearly identical pattern between the two samples. Besides alkali adducts of the most intense peak, one more peak occurs with a mass distance of 162 Da and a frequency of 3.9% (calculated by integration), representing a glycation. Bottom: The deconvolved MS spectra of the Fab/2 and Fc domains show a glycosylation pattern also seen in the MS of the intact mAbs solely originates from glycosylation of the Fc domain. The most intense peaks are consistent with the most intense peaks of the Fc domain from the MS of the intact mAb, the minor signals appear stronger. Red: NIST_A, blue: NIST_B.**

More details were obtained by an alternative middle up approach. Denaturation and reduction of disulfide bridges with DTT followed by LC-MS result in separated light and heavy chains. Figure 30 shows the MS spectrum of the isotope-resolved light chain. The high resolution as well as the accurate mass allow the identification of minor modifications and adducts. Both NIST samples show similar peaks at *m/z* = -17 Da (formation of pyroglutamate of an *N*-terminal pyroglutamic acid), +22 Da (sodium adduct), +38 Da (potassium adduct), +56 Da (acetonation by acetoacetate)[112], +100 Da (lysine succinylation), +162 Da (hexose). The high-resolution MS from this isolated light chain points out that the *N*-glycosylation pattern seen in the MS of the intact mAbs is located solely within the heavy chain.

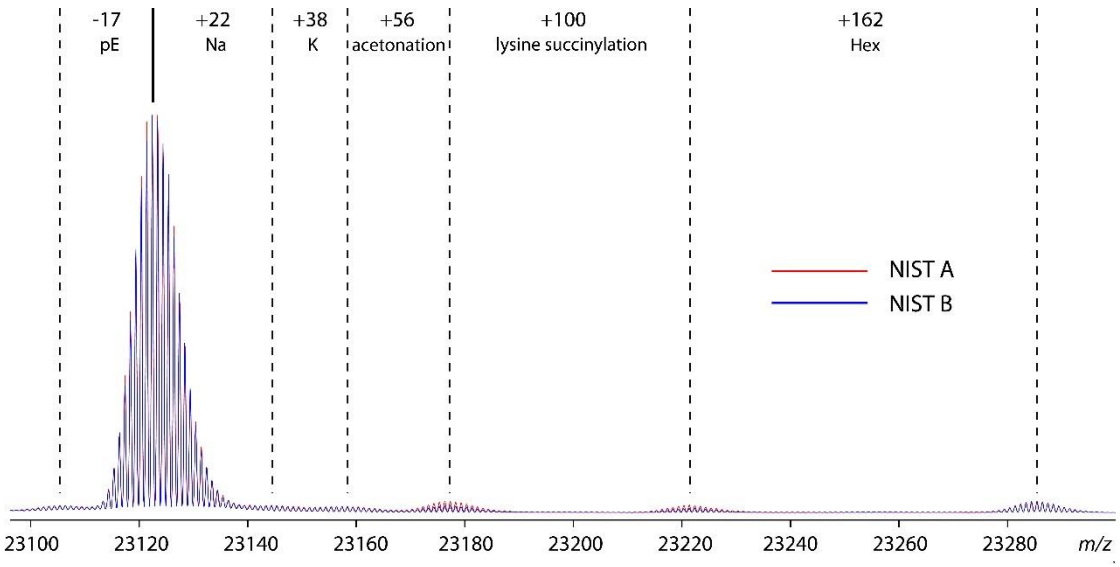


Figure 30: **HR-MS spectra of light chains of NIST mAbs. After denaturation and reduction, the heavy and light chains were separated and analyzed by C_8-RP-LC-MS by applying an ISCID of 20 eV. The spectrum of the light chain shows several modifications and adducts (denoted mass distances refer to the most intense signal): -17 Da (formation of pyroglutamate of an *N*-terminal pyroglutamic acid), +22 Da (sodium adduct), +38 Da (potassium adduct), +56 Da (acetonation), +100 Da (lysine succinylation), +162 Da (hexose). Red: NIST_A, blue: NIST_B.**

The deconvolved MS signals of the heavy chain appear as expected above ~ 50000 g/mol. This high mass as well as the glycosylation preclude isotope-resolved peaks of the HC (cf. Figure 31). The glycosylation patterns of the two heavy chains show a consensus in structures, but wide differences in the quantities of the glycoforms.

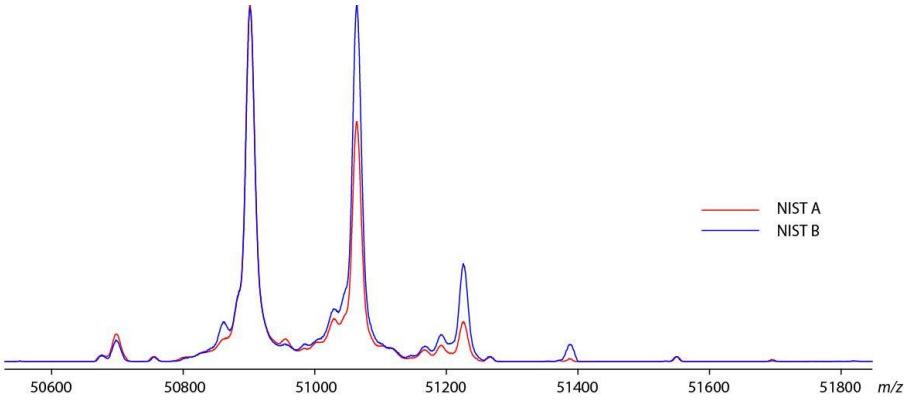


Figure 31: **Deconvolved MS spectra of the heavy chains of the NIST mAbs. After denaturing and reduction, the heavy and light chains were separated and analyzed by C_8-RP-LC-MS by applying an ISCID of 20 eV. The spectra of the HC show large consensus, although differences in glycoform quantity can be observed.**

Due to the fact that the protein sequence was not specified by the organizers of this study the composition of the glycans cannot be determined from the MS of the intact heavy chain alone. Therefore, *N*-glycans were released from the tryptically digested mAbs by PNGase F, separated on a porous graphitized carbon column and analyzed online by MS/MS on an ESI-q/TOF MS. The most intense glycans that could be identified in these LC-MS/MS runs correspond to the fucosylated biantennary structures with the monosaccharide compositions $Hex_3HexNAc_4dHex_1$ at m/z = 732.28 and $Hex_4HexNAc_4dHex_1$ at m/z = 813.31 (for detected released *N*-glycans see Table 8). The integrated areas of the extracted ion chromatograms agree with the corresponding signals in the spectra of the intact heavy chains. With the determination of the most intense glycoforms in the MS spectra of the intact sub-domains the other peaks of the heavy chain could be assigned to further glycan compositions (cf. Table 6). This was possible by the summation of possible *N*-glycan compositions combined with a slightly changed amino acid sequence of a known murine monoclonal antibody.

Table 6: ***N*-type glycosylation of NIST mAbs determined by LC-MS analysis of the reduced heavy chain (HC). By glycan release with PNGase F the most abundant glycans were determined and assigned to the signals of the deconvolved MS spectra of the HC. Based on this interpretation, further signals could be assigned to specific *N*-glycan compositions.**

Mass intact [Da]*	Monosaccharide composition[#]
50552.8	H3N3
50698.4	H3N3F1
50756.1	H3N4
50861.8	H4N3F1
50901.8	H3N4F1
51064.0	H4N4F1
51167.8	H4N3F1G1 - H5N3A1 - H5N3F2
51226.2	H5N4F1
51266.0	H4N5F1
51329.9	H4N4F1A1
51355.5	H5N4F2 - H4N4F1G1 - H5N4A1
51388.1	H6N4F1 - H5N4G1
51533.8	H6N4F2 - H5N4F1G1 - H6N4A1
51550.4	H7N4F1 - H6N4G1
51695.7	H6N4F1G1 - H5N4G2
51775.1	H7N5F1

*: denoted is the center of the corresponding deconvolved peak; [#]: H: hexose, N: hexosamine, F: deoxy hexose, G: glycolyl neuraminic acid, A: acetyl neuraminic acid;

As a result of the absence of an isotopic resolution in the MS spectra of the heavy chains, glycoforms exhibiting relatively similar molecular weights (e.g. heavy chain glycoforms H6N4F1 vs. H5N4G1: $\Delta_{m/z} = 1.0204$ Da ≈ 21 ppm) could not be differentiated. Such differences occur due to similar masses of several monosaccharide compositions (cf. Table 7).

Table 7: **Selection of monosaccharide/disaccharide combinations resulting in similar and/or equal molecular weights (MW). In the absence of isotopic resolution of glycoproteins, minor differences in MW of oligosaccharides cannot be distinguished (variant A/B). Other glycan combinations exhibit exactly the same MW and cannot be distinguished even for isotopically resolved glycoconjugates (variant C).**

Monosaccharides	MW [g/mol]	Δ_{MW} [g/mol]	Variant
2 fucoses	310.1264	1.0204	A
1 acetyl neuraminic acid	309.1060		
1 hexose + 1 fucose	326.1213	1.0204	B
1 glycolic neuraminic acid	325.1009		
1 hexose + 1 acetyl neuraminic acid	471.1588	0.0000	C
1 fucose + 1 glycolic neuraminic acid	471.1588		

To overcome ambiguity resulting from similar mass variants of glycan combinations that contain neuraminic acids, these terminal monosaccharides were removed by digestion with neuraminidase. Peaks that disappear upon digestion correspond to glycans that exhibit neuraminic acids. The corresponding desialylated glycans should arise in the spectra of the neuraminidase-treated mAbs as shown in Figure 32.

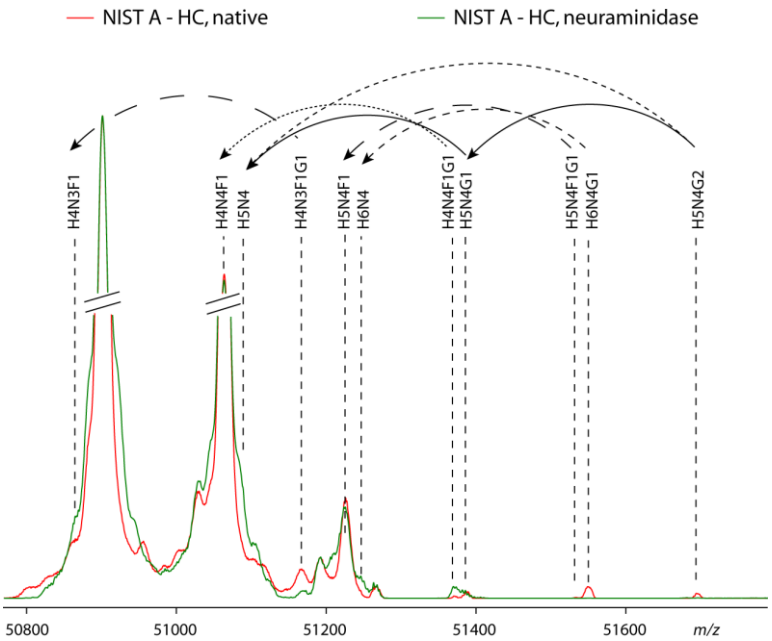


Figure 32: **Native (red) and neuraminidase-treated (green) heavy chain (HC) of NIST sample A. For a more certain assignment of glycan compositions, mAbs were reduced and desialylated. Signals decreasing or disappearing in this process are therefore composed of sialylated glycans. H: hexose, N: hexosamine, F: deoxy hexose, G: glycolyl neuraminic acid, A: acetyl neuraminic acid.**

Verification as well as further determination of the precise carbohydrate composition (e.g. differentiation of variant C, cf. Table 7) was performed by the analyses of free *N*-glycans. After tryptic digestion, the *N*-glycans were released by PNGase F digestion and analyzed by MS/MS (ESI and MALDI) without further derivatization.

Structural assignment of free *N*-glycans by MS(/MS) relies on mass accuracy of the intact oligosaccharide as well as the interpretation of the corresponding fragment spectra. Since several monosaccharide combinations yield the same precursor masses, fragmentation is an essential tool for the correct determination of the glycan composition. Fragmentation was realized by collision-induced dissociation (CID) within the collision cell of the ESI-q/TOF MS and laser-induced dissociation (LID) using a MALDI-TOF/TOF MS. An example for such two glycan compositions that would stay ambiguous if characterized solely by the precursor mass are the two undecasaccharides shown in Figure 33.

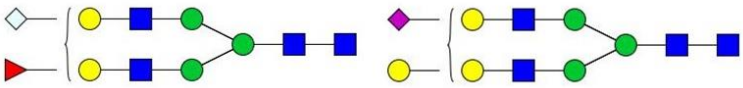


Figure 33: **Mass isobaric complex type *N*-glycan structures (both *m/z* = 1047.8775, monoisotopic, $2H^+$). LC-MS analysis of *N*-glycans released from NIST samples exhibited signals for glycan precursors that could not be distinguished solely by accurate molecular weight (MW), due to their isobaric nature. This is based on the fact that the saccharide combinations *N*-glycolyl neuraminic acid/deoxyhexose and *N*-acetyl neuraminic acid/hexose exhibit the same molecular composition and therefore the same MW.**

To overcome this ambiguity, the doubly charged precursor at *m/z* = 1047.878 was isolated and fragmented by CID. The resulting MS/MS spectrum contains several fragments, that unambiguously reveals the fucosylated *N*-glycan containing one glycolyl neuraminic acid (cf. Figure 33, left) due to characteristic signals at *m/z* = 290.086 (B-ion, loss of H_2O), *m/z* = 625.150 ($X_{Man}Z$-ion) and *m/z* = 673.230 (B-ion). The presence of *N*-glycolyl neuraminic acids instead of *N*-acetyl neuraminic acids revealed by ESI-q/TOF verifies the results from the neuraminic acid digestions, since only NeuGc monosaccharides can be cleaved by neuraminidase.

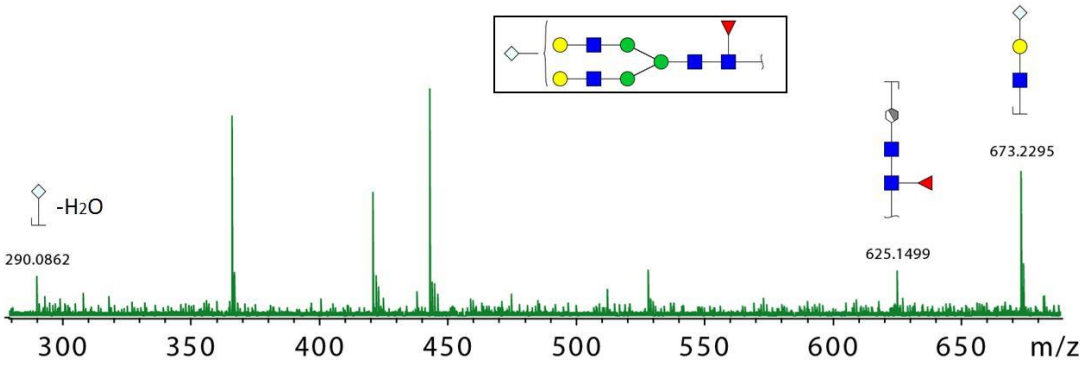


Figure 34: **MS/MS spectrum of doubly charged precursor at *m/z* = 1047.878. The parent mass was isolated and fragmented within the collision cell (CID: collision-induced dissociation). Solely by the precursor mass, it is not possible to distinguish between the two undecasaccharides shown in Figure 33. Characteristic fragments assigned in the MS/MS spectrum (*m/z* = 290.0862, 625.1499 and 673.2295) provide the information necessary to determine the correct *N*-glycan (shown in the inset).**

A more reliable assignment was performed by further verification with MALDI-MS analyses, of both sets of released *N*-glycans. The free *N*-glycans were separated on a PGC column and the LC flow was split post-column in a ratio of 1:20. 5% were directly injected into the ESI-q/TOF (online MS), 95% were collected in fractions in a deep well plate. These fractions were analyzed by MALDI-MS.

Neutral glycans were also verified by MALDI-MS by accurate masses of the sodium adduct. Glycans that contain sialic acids could not be identified by MALDI-MS due to typical low resolution of such glycans (instability of sialic acids in the gas phase).

Table 8: **Glycan structures characterized by a multi-dimensional approach. Data was achieved by analyzing the intact mAbs, papain-digested sub-domains (partly reduced) and on the free *N*-glycan level. ESI-q/TOF and MALDI-TOF/TOF mass spectra were combined with online and offline chromatographic techniques (RP for glycoprotein, PGC for released *N*-glycans). H: hexose, N: hexosamine, F: deoxy hexose, G: *N*-glycolyl neuraminic acid.**

Glycan Composition	Quantity* NIST A	Quantity* NIST B	Identified by		
			Reduced mAb - HC	Released *N*-Glycan	
			ESI	ESI	MALDI
H3N4F1	50.30	39.10	+	+	+
H4N4F1	33.80	38.60	+	+	+
H5N4F1	5.30	10.30	+	+	+
H3N3F1	3.70	2.00	+	+	+
H4N3F1	3.40	5.70	+	+	-
H4N5F1	0.42	0.55	+	+	+
H7N5F1	<0.01	<0.01	+	-	-
H3N4	0.47	0.10	+	+	-
H3N3	0.02	0.01	+	+	-
H4N3F1G1	1.90	1.30	+	+	-
H4N4F1G1	0.02	0.04	+	+	-
H5N4F1G1	0.01	0.06	+	+	-
H5N4G1	0.14	1.70	+	-	-
H5N4G2	0.16	0.07	+	-	-
H6N4G1	0.36	0.46	+	-	-

*: Quantity was obtained by middle up analyses (sub-domains of NIST samples).

After data submission, the organizers of the study revealed the nature of the mAbs. NIST B was the original "NISTmAb", NIST A was the "modified" mAb. Modification occurred by taking off a portion of NIST B that was subjected to beta-1,4-galactosidase digestion. Afterwards the sample was added back to a portion of NISTmAb (thus, NIST A contains untreated as well as enzymatically digested NISTmAb).

The results obtained here correspond very well to the described modifications, i.e. the treatment of the samples by NIST. For example, the amounts of the structures H4N4F1 and H5N4F1 are reduced by ~ 10%, while the amount of H3N4F1 is enhanced by ~ 10%, which is a glycan without terminal galactoses (substrate for galactosidases). The structure H3N3F1 is increased by the same amount (1.7%) by which that of H4N3F1 is reduced.

Detailed statistical data analysis has not been published by NIST at the time of submission of this thesis, however, preliminary results were communicated. For example, the Youden plot shown in Figure 35 displays the distribution of values for a specific quantity, that is in this case the proportion of structure H3N4F1, that have been found by all the participating laboratories (taken from a NIST presentation at the conference ASMS 2016). The Youden plot is designed to provide an instrument for inter-laboratory comparison. By plotting the same quantity on both axes, only for different experimental runs and in this case slightly different samples, it distinguishes between random and systematic errors: If a certain laboratory's entry in this scatter plot, denoted by a dot, lies along the diagonal line but far off the central "consensus value", this hints to a systematic measurement error in this laboratory. If the laboratory's dot does not lie on the diagonal line at all, this hints to (additional) random errors. Since the value obtained by us (denoted by the only grey dot, named "3 analytes") lies very close to the consensus value, this first feedback by NIST indicates that the multi-dimensional approach performed here, is very well suited to this kind of analysis problem.

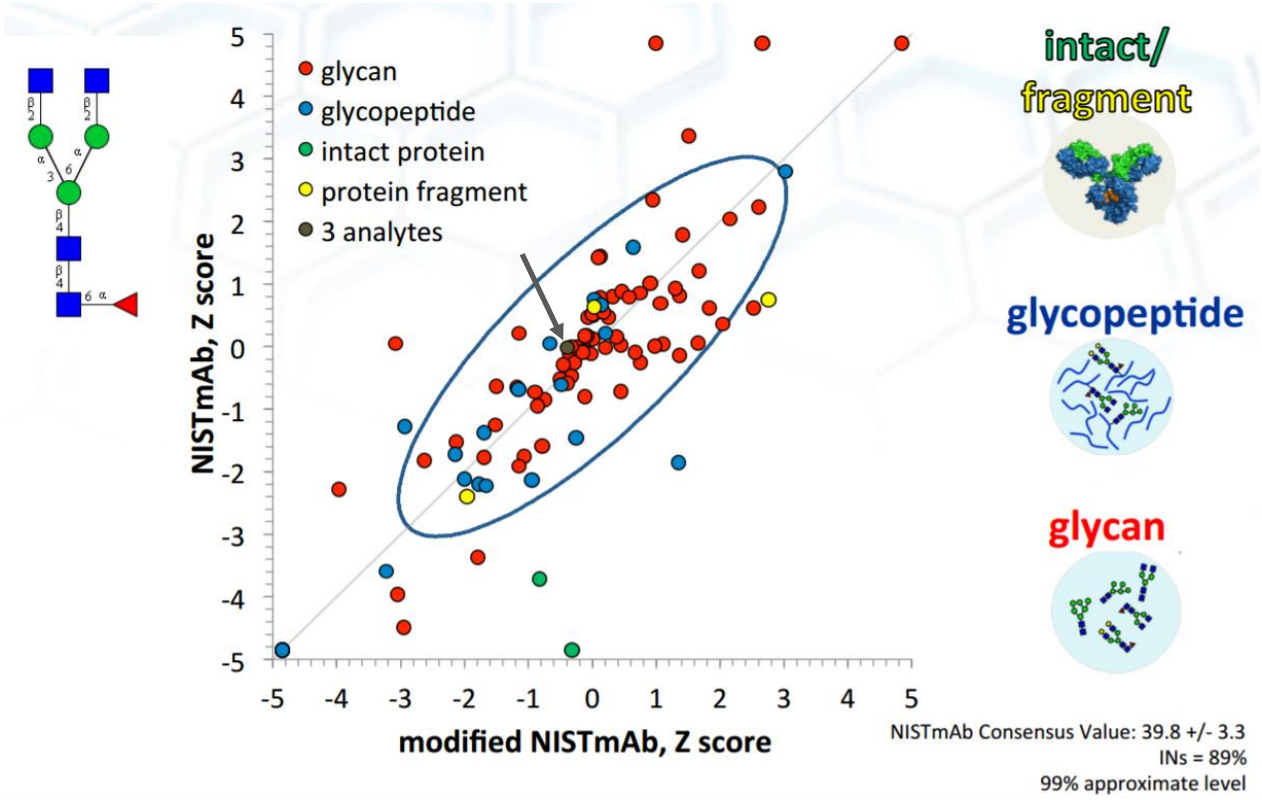


Figure 35: **Youden plot showing the distribution of values for a specific quantity, that is in this case the proportion of structure H3N4F1, that have been found by all the participating laboratories. Random and systematic errors can be distinguished. The grey dot (marked by a grey arrow) is the value that was obtained by the multidimensional approach described here. The figure was taken from a NIST presentation at the conference ASMS 2016.**

5.2 Histidine-Rich Glycoprotein

Histidine-rich glycoprotein (HRG), also known as histidine-proline-rich glycoprotein (HPRG), is a single-chain glycoprotein of multidomain structure identified in the plasma of diverse vertebrates.[113] HRG synthesis occurs in parenchymal liver cells and the propeptide exhibits 525 amino acids (aa). Cleavage of the signal peptide results in the mature protein (507 aa) with a molecular weight of ~ 65 kDa including posttranslational modifications. With a concentration of 100-150 µg/mL in human plasma HRG is an abundant plasma glycoprotein. The multidomain structure of HRG is composed of two N-terminal domains that are homologous to cystatin, two proline-rich domains surrounding a histidine-rich region and the C-terminal region. The domains are cross-linked by six disulfide bridges.[114-115] Hulett *et al.* summarize multiple physiological functions based on the multidomain structure often caused by interactions between HRG and various molecules such as metal ions[116-117], proteins[114] and DNA[118]. Several polymorphisms (n = 11) of HRG have been identified and partly associated to physiological disorders, e.g. the success of *in vitro* fertilization is significantly determined by such a SNP.[119]

Several studies revealed correlation of HRG concentration and cancer. For example, HRG is downregulated in ovarian and endometrial cancer as well as in hepatocellular cancer (HCC) and may therefore act as a biomarker.[120-122] Interestingly, HRG concentration was found to be a specific indicator for HCC patients that are α-fetoprotein (AFP) negative. Concentration of HRG might have an effect on tumor progression, since HRG was shown to exhibit regulatory functions on cell proliferation and adhesion as well as angiogenesis. In fact, mouse studies have shown that HRG inhibits vascularization as well as growth of fibrosarcoma tumors and affects the development of glioma.[123-124] Contrary results were obtained regarding the influence of HRG on tumor angiogenesis. While Klenotic *et al.* showed that HRG expressed in glioma cells (brain tumor models) causes increased tumor angiogenesis and size, Thulin *et al.* revealed that HRG knock-out mice (pancreatic tumor models) exhibit increased angiogenesis.[125-126] These contrary influences of HRG on tumor development and progression (angiogenetic

and anti-angiogenetic effects) indicate that the level of HRG alone is not an appropriate biomarker for physiological disorders. However, HRG was shown to affect and regulate multiple physiological processes, such that HRG could act as a control or marker in some way.

As mentioned before glycosylation is often a better and more specific marker than the protein level itself. Lubman *et al.* revealed by lectin studies that the glycosylation of several human plasma glycoproteins is altered in patients suffering from colorectal cancer (CRC) compared to healthy controls as well as adenoma patients.[127] Glycan patterns were determined by lectin arrays and revealed altered levels of sialylation and fucosylation in HRG, complement C3 and kininogen-1. These results indicate a potential value of these proteins as biomarkers for CRC. Behind prostate cancer in male patients, breast cancer in female patients and lung cancer for both sexes, CRC is the third-most common cancer worldwide with 694000 deaths in 2012 (source: World Health Organization, February 2015). Mortality of CRC is very high in particular due to late diagnosis. Current diagnose procedures such as colonoscopy are inconvenient and often avoided causing diagnosis only at a late disease stage with decreased rates for survival. Therefore, improved diagnosis of CRC, e.g. by blood tests, could lead to earlier diagnosis and better cure as well as survival rates.

Since Lubman *et al.* revealed altered glycosylation patterns in blood plasma of CRC patients by lectin studies, glycosylation of HRG from healthy test persons should be analyzed in detail. So far, no detailed structural analysis of the *N*-glycosylation of HRG has been performed.[128] Four consensus sequences for *N*-type glycosylation are present in HRG. Two are in direct vicinity (^{345}NNSS^{348}) and thus an occupation of both glycosylation sites is questionable.

5.2.1 Development of a Purification Strategy for HRG from Human Blood Plasma

Enrichment of HRG was supposed to be based on immobilized metal affinity chromatography (IMAC). Human blood plasma contains protein at a level of 60-80 mg/mL, distributed over an enormous dynamic range. For example, the main

protein component of blood plasma is albumin with 45±10 mg/mL. HRG has a concentration of 125±25 µg/mL, which is lower by a factor of 350. Other proteins like prolactin are even less abundant with 13±12 ng/mL, a factor of 10000 lower than HRG. Thus, isolation of a single protein from blood plasma is extremely complex. The HRG protein sequence consists of 13% histidines that contain imidazole side chains and are known to interact with bivalent cations such as Ni^{2+}, Cu^{2+}, Zn^{2+} or Co^{2+}. IMAC set-ups contain such bivalent cations in order to build complexes with free valencies available for binding to imidazoles.

Co^{2+}- as well as Ni^{2+}-attached resins were tested for efficiency in HRG isolation. Additives such as chaotrops as well as imidazole can be added to decrease unspecific binding of proteins to the matrices and of HRG to other plasma proteins to ensure an optimal purification of HRG on the other hand. Since enrichment of proteins from complex matrices such as blood plasma is unlikely to result in pure protein, multiple chromatographic procedures have to be combined. Besides anion exchange chromatography (AIEX) and C_8-reversed phase chromatography (RP), size exclusion chromatography (SEC) was tested. Here, best performance was reached by combination of nickel-based IMAC with size exclusion chromatography. Alternatively, affinity chromatography could be used for purification.

With Ni-NTA-IMAC all components with an affinity to Ni^{2+} ions are retained. To reduce unspecific interactions, imidazole was added to the plasma samples to a final concentration of 20 mM prior to IMAC. The bulk of human plasma components does not bind to the resin as shown in Figure 36. Wash fractions (W1: 20 mM, W2: 50 mM) contain several proteins that do not have a strong affinity to the matrix. Elution of proteins that are strongly interacting with Ni-NTA resin was performed by great excess of imidazole (E1-E3, 400 mM imidazole, 3 CV). However, despite depletion of major human blood plasma components, several other proteins including HRG at a MW ~ 66 kDa are present after IMAC.

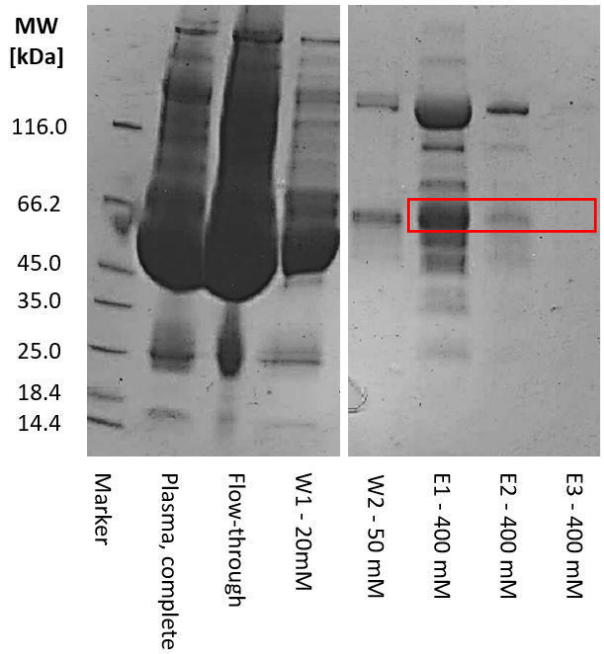


Figure 36: **IMAC with human blood plasma analyzed by reducing SDS-PAGE. Ni^{2+}-NTA resin was used for enrichment of histidine-rich glycoprotein (HRG) since imidazole side-chains of histidines show a strong affinity to bivalent cations. After sample loading (blood plasma), the resin is washed with increasing concentrations of imidazole (W1, W2). Strongly bound blood components are eluted with 400 mM imidazole (E1-E3). HRG (red box) together with several impurities is found in elution fractions.**

Therefore, SEC on a Superdex75 was introduced as a second dimension (flow-rate: 0.3 mL/min). The UV chromatogram is shown in Figure 37. The peak at t_r = 27 min contains components above 75 kDa that could not be resolved by the SEC column (flow-through) and overlaps with another peak at t_r = 30 min. The third peak at t_r = 35 min contains HRG, other proteins in that fraction include fibrinogen and albumin (as determined by MALDI MS, cf. Figure 38). Both are major plasma components. Whereas albumin does not contain *N*- or *O*-glycans, fibrinogen possesses both of them and has to be eliminated before analyzing the glycans of HRG in detail to prevent false interpretation of HRG glycosylation.

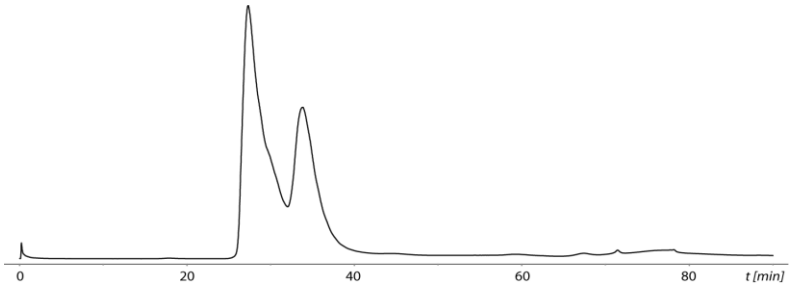


Figure 37: **Size exclusion chromatography of Ni^{2+}-NTA-enriched components (cf. Figure 36) from human blood plasma. IMAC eluate was subjected to a Superdex75 column (300×10 mm, GE Healthcare) and separated with Tris-HCl buffer (pH: 8.0) with a flow-rate of 300 µL/min.**

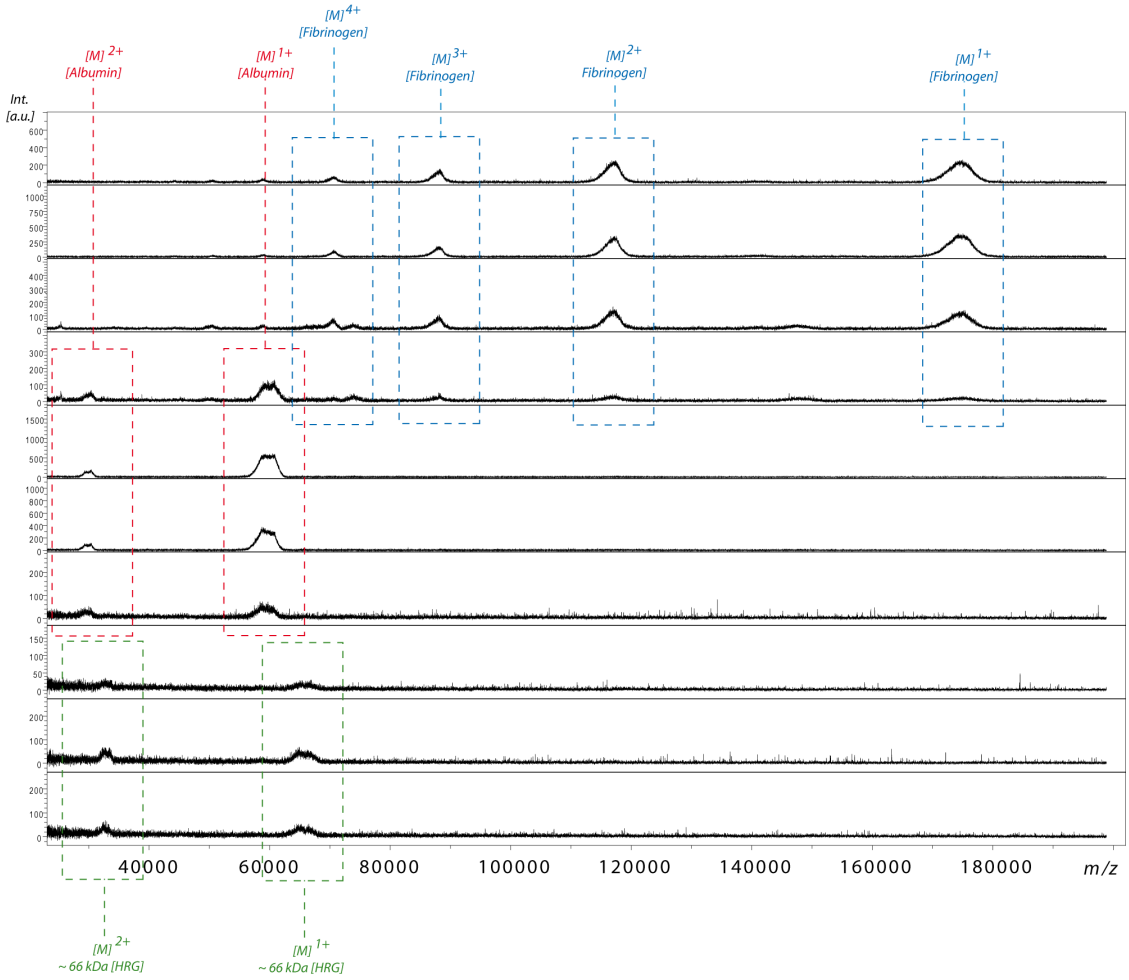


Figure 38: **MALDI spectra of SEC fractions (positive ions linear mode). HRG was isolated by IMAC and further purified by SEC. Masses indicate that the peak at t_r = 35 min in the UV chromatogram contains HRG as well as fibrinogen and albumin as impurities.**

The reducing SDS-PAGE gel from the eluates after IMAC (cf. Figure 36) indicated that several components smaller than HRG were present. However, no masses smaller than ~ 60 kDa eluted from the SEC column. HRG is known to be composed of a single chain

protein of multidomain structure whose domains are linked by intra-protein disulfide bridges. Although these domains are part of the single chain protein, a human enzyme called plasmin is known to cleave HRG *in vivo*.[129-130] Due to intra-protein disulfide bridges, domains cleaved by plasmin are still linked. By adding DTT as reducing reagent for SDS-PAGE gels the cleaved domains are separated. Without addition of a reducing reagent, the domains are still covalently linked and therefore appear as one protein in SDS-PAGE. This is shown in SDS-PAGE gels with and without addition of DTT (cf. Figure 39).

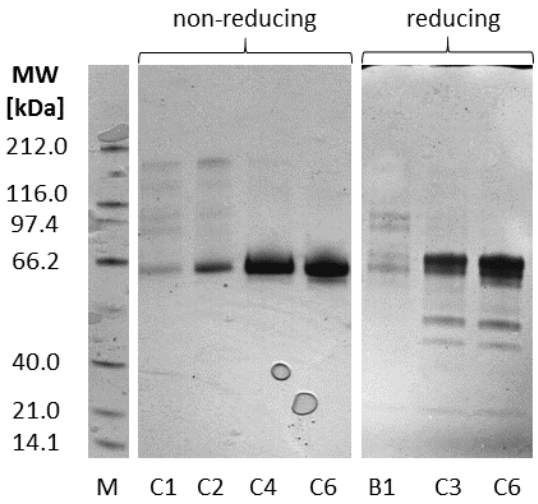


Figure 39: **SDS-PAGE of histidine-rich glycoprotein containing fractions under non-reducing (left) and reducing (right) conditions. Fractions arise from SEC-separated IMAC enriched blood components. Without addition of DTT HRG appears at one gel band (left, C4/C6, ~ 66 kDa), while the amount of *in vivo* digested HRG is separated (< 66 kDa) after addition of DTT (right, C3/C6). HRG is known to be digested by the human enzyme plasmin. However, the domains are still linked by intra-protein disulfide bridges. Fraction order from SEC column is B1-C1-C2-C3-C4-C5-C6.**

Purification of HRG still has to be improved by reducing contaminants such as fibrinogen. Therefore, addition of chaotrops (here: urea) as well as varying concentration of imidazole during IMAC (prior to sample loading and during washing procedures) were tested for efficient HRG isolation. Furthermore, separation by SEC was improved by decreased flow-rates. Efficiency was controlled by SDS-PAGE prior and after SEC (cf. Figure 40). The protein bands reveal that washing with a high concentration of imidazole has a great effect on removing impurities during IMAC, while HRG is

remaining on the column due to strong interactions (washing procedure: A1/C1: 20 mM imidazole, B1/D1: 80 mM imidazole). The impact of denaturation by addition of urea prior to IMAC (A1/B1) is smaller, but the number of proteins visible without previous urea addition seems smaller (B1: with urea, D1: without urea - both washed with 80 mM imidazole). Further reduction of impurities was achieved by increasing imidazole concentration to 80 mM prior to IMAC. Thereby, blood components with a lower affinity to the Ni^{2+}-resin such as fibrinogen do not bind during sample loading (E1). However, the yield of HRG is also reduced such that a compromise in imidazole concentration in blood plasma prior to IMAC must be found.

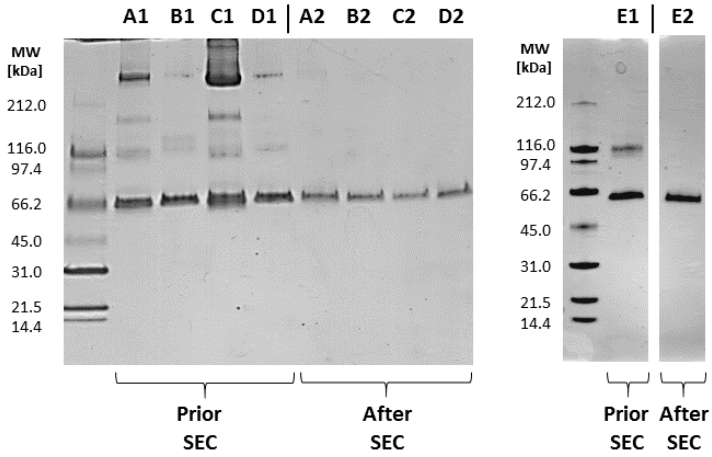


Figure 40: **Optimizing IMAC parameters to reduce enrichment of other proteins than HRG from human blood plasma. Best results were achieved for samples washed with 80 mM imidazole after sample loading to IMAC resin (B1, D1, E1). Addition of urea prior to IMAC (B1) led to more impurities in eluates than were present without urea (D1). Purest HRG was obtained for samples containing 80 mM imidazole while loading (E1 compared to A1-D1). After SEC (A2-E2) HRG was pure in all samples. A1-E1: IMAC eluates before SEC. A2-E2: IMAC eluates after SEC. A+B: with addition of urea prior to IMAC, C+D: without addition of urea prior to IMAC, A+C: washing step with 20 mM imidazole, B+D+E: washing step with 80 mM imidazole, A-D: 20 mM imidazole in sample prior to IMAC, E: 80 mM imidazole in sample prior to IMAC.**

Optimal results for isolation of pure HRG were obtained by adjusting plasma samples to a final imidazole concentration of 35 mM prior to IMAC experiments. After binding to the Ni-NTA resin, unspecific binding proteins were eluted with 20 CV 80 mM imidazole containing Tris-HCl buffer. Elution was performed twice with 2 CV 400 mM imidazole. After concentrating the eluates to 500 µL, ~ 5 mg urea were added and rigorously mixed to inhibit interaction of HRG with other proteins. Separation of remaining impurities was

achieved by SEC with a Superdex200 column (10×300 mm, GE Healthcare) at a reduced flow-rate of 150 µL/min. Baseline separation was achieved to obtain pure HRG (cf. Figure 37 for comparison). Corresponding HRG protein bands in SDS-PAGE gels for control of purification appear as in E2, Figure 40 (only difference in sample preparation: 20 mM instead of 35 mM imidazole in the sample prior to IMAC).

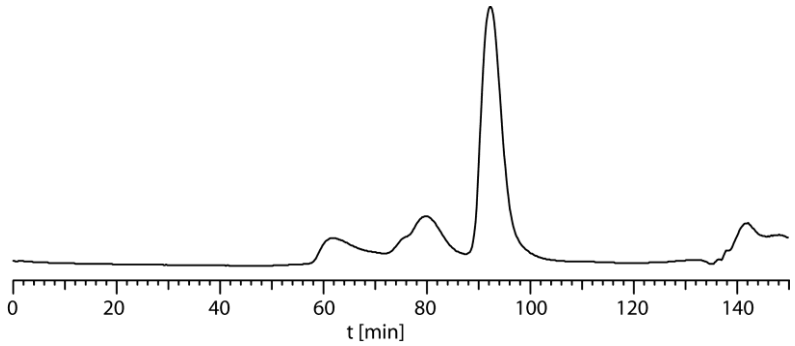


Figure 41: **UV chromatogram of SEC of optimized IMAC and SEC conditions. For enrichment of HRG, blood plasma was treated with 35 mM imidazole prior to IMAC experiments. After sample loading Ni-NTA resin was washed with 80 mM imidazole to remove unspecifically bound proteins. After eluting with 400 mM imidazole, the sample was concentrated and ~ 5 mg urea were added to inhibit interaction of HRG with other proteins. Purification was achieved by SEC with a Superdex200 column (10×300 mm, GE Healthcare) at a flow-rate of 150 µL/min. Baseline separation was achieved to obtain pure HRG (t_r = 93 min).**

Purity of HRG was further evaluated by dynamic light scattering (DLS) experiments. DLS is used to determine the size distribution of analytes in solution, e.g. the hydrodynamic radius of proteins. DLS was performed with native HRG as well as deglycosylated HRG. DLS spectra show presence of a pure component by a uniform size distribution. Treatment with PNGase F results in a smaller size distribution of the unglycosylated protein and shows indirectly the presence of *N*-glycosylation of the original protein.

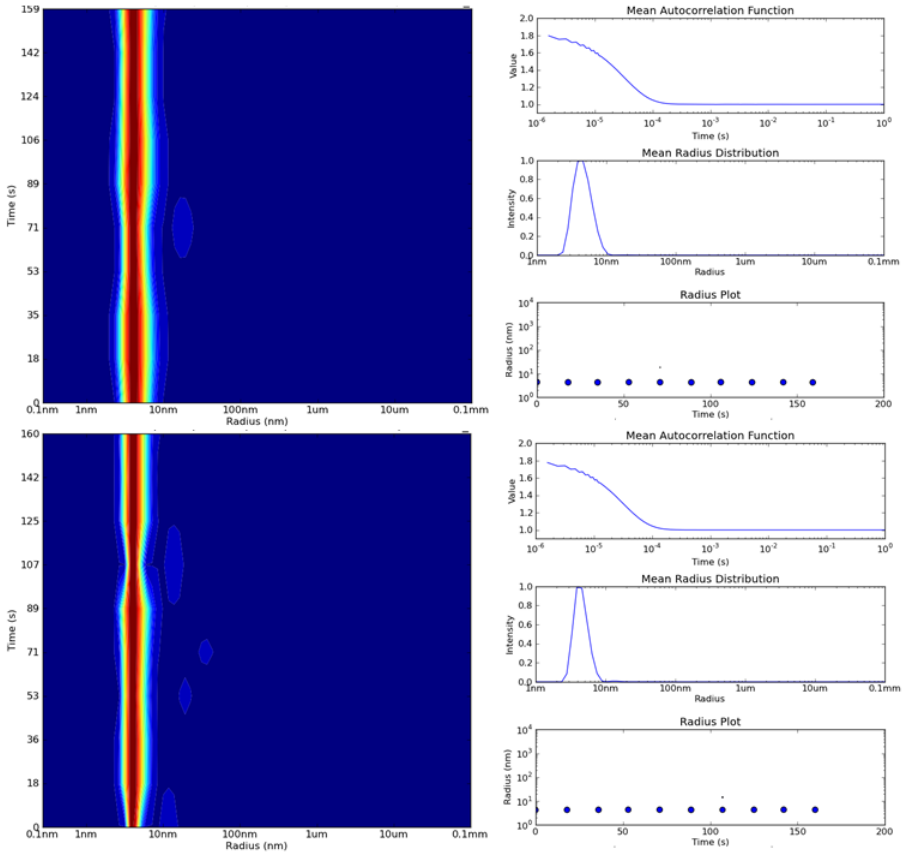


Figure 42: **Dynamic light scattering (DLS) experiments of purified HRG. The DLS spectra show a steady distribution of the hydrodynamic radius, verifying a suspension containing uniform/isobaric analytes. Deglycosylation (bottom) with PNGase F results in a smaller hydrodynamic radius confirming the presence of *N*-glycosylation in the original protein.**

5.2.2 Mass Spectrometric Analysis of Histidine-Rich Glycoprotein

An in-gel digestion of purified HRG from blood plasma was performed. The extracted tryptic peptides were subjected to RP-LC-MS/MS. Data analysis was performed automatically by comparison with the human proteome. The sequence coverage is 67%. Peptides containing the *N*-glycosylation sites at ^{63}N and $^{344/345}$N were not detected by MS analysis. Kluft *et al.* performed sequence analyses of exons and demonstrated that the amino acid exchange ^{204}P to ^{204}S frequently occurs and thus generates a new *N*-glycosylation site. Here, no corresponding peptide was identified by LC-MS/MS indicating that probably an amino acid exchange was present, leading to a non-detectable glycopeptide. The peptide containing the *N*-glycosylation site at ^{125}N was detected. However, only one spectrum was acquired for this peptide, indicating that the abundancy

of this peptide is extremely low or constitutes an incorrect assignment. Thus, an occupation of this glycosylation site is possible.

One phosphorylation site was identified at ^{356}S, however, the same peptide without a phosphorylation was detected with a 5- to 10-fold intensity/area. Thus, only partial phosphorylation of HRG seems to be present. However, a distinct value of phosphorylation amount cannot be obtained by a comparison of signal intensity in MS spectra, since ionization of a phosphopeptides is known to be significantly lower compared to their non-phosphorylated analog.

1	11	21	31	41
MKALIAALLL	ITLQYSCAVS	PTDCSAVEPE	AEKALDLINK	RRRDGYLFQL
51	61	71	81	91
LRIADAHLDR	VENTTVYYLV	LDVQESDCSV	LSRKYWNDcE	PPDSRRPSEI
101	111	121	131	141
VIGQcKVIAT	RHSHESQDLR	VIDFNcTTSS	VSSALANTKD	SPVLIDFFED
151	161	171	181	191
TERYRKQANK	ALEKYKEEND	DFASFRVDRI	ERVARVRGGE	GIGYFVDFSV
201	211	221	231	241
RNCPRHHFPR	HPNVFGFcRA	DLFYDVEALD	LESPKNLVIN	cEVFDPQEHE
251	261	271	281	291
NINGVPPHLG	HPFHWGGHER	SSTTKPPFKP	HGSRDHHHPH	KPHEHGPPPP
301	311	321	331	341
PDERDHSHGP	PLPQGPPPLL	PmScSScQHA	TFGTNGAQRH	SHNNNSSDLH
351	361	371	381	391
PHKHHsHEQH	PHGHHPHAHH	PHEHDTHRQH	PHGHHPHGHH	PHGHHPHGHH
401	411	421	431	441
PHGHHPHCHD	FQDYGPCDPP	PHNQGHCCHG	HGPPPGHLRR	RGPGKGPRPF
451	461	471	481	491
HcRQIGSVYR	LPPLRKGEVL	PLPEANFPSF	PLPHHKHPLK	PDNQPFPQSV
501	511	521		
SEScPGKFKS	GFPQVSmFFT	HTFPK		

Figure 43: **Sequence coverage of HRG achieved by in-gel digestion followed by LC-MS/MS. Data analysis was performed by comparison with the human proteome. Green: Identified peptide sequences. Grey: signal peptide, not existing in mature protein; bold and underlined: *N*-glycosylation sequon by a SNP causing amino acid exchange from ^{202}NCP^{204} to ^{202}NCS^{204}; underlined: *N*-glycosylation sequences; small letter c: modified as carbamidomethylated cysteine; small letter m: identified as oxidized methionine; small letter s: identified as phosphorylated serine. Sequence coverage was 67%.**

Based on this purification procedure, the glycosylation pattern of HRG can be analyzed. Initially, the global glycosylation of HRG from 13 test persons was analyzed (cf. Figure 44).

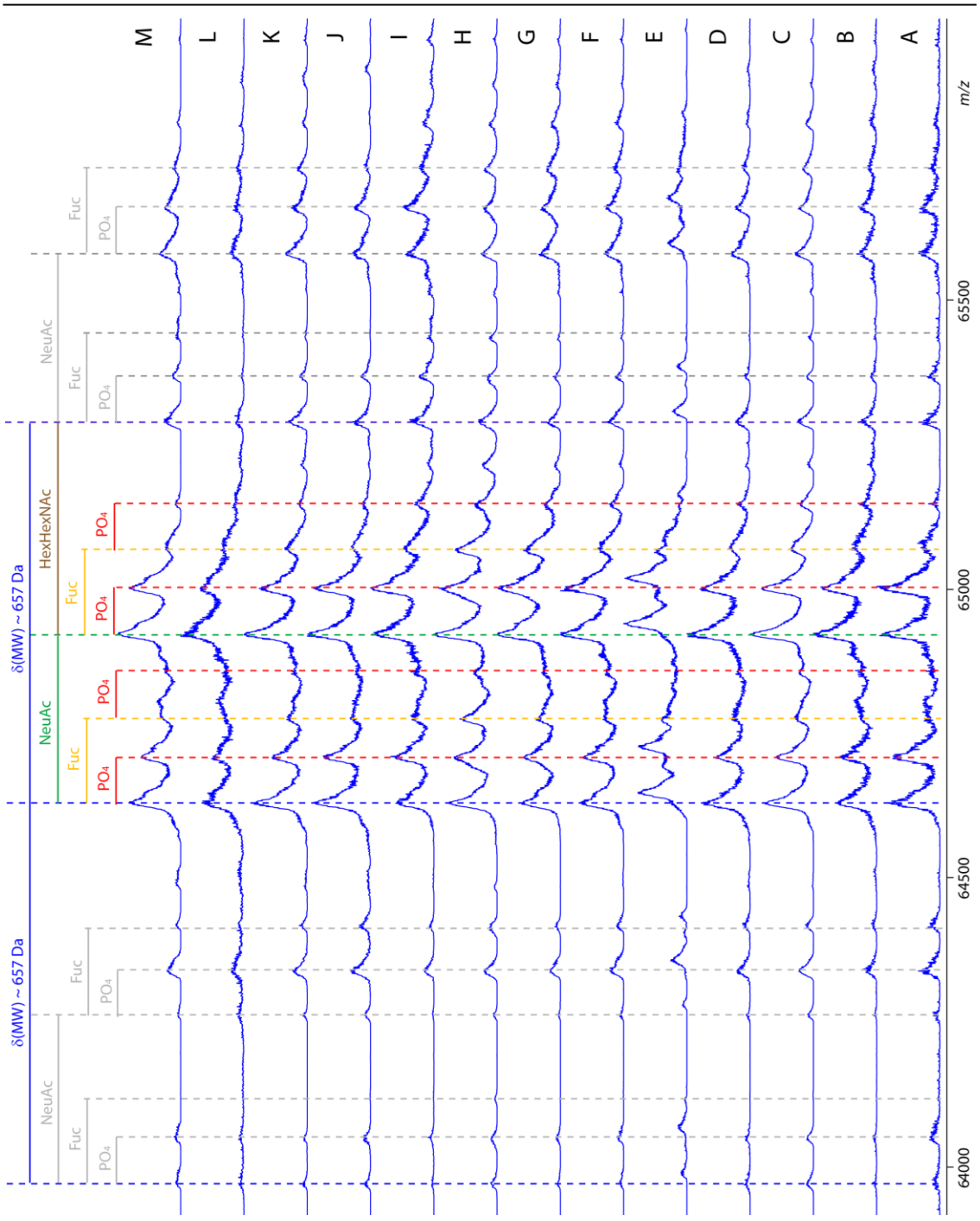


Figure 44: **MS spectra of intact HRG isolated from blood plasma from 13 individuals. Major mass distances can be explained by monosaccharide units and phosphorylation (colored annotation). The glycosylation profile between *m/z* 64600-65100 repeats twice with mass distance of ~ 657 Da at *m/z* 63900-64500 and *m/z* 65300-65800 between main signals (657 Da: $Hex_1HexNAc_1dHex_2$ or $Hex_1HexNAc_1NeuAc_1$). Cf. Table 20 in appendix for sample ID (A-M).**

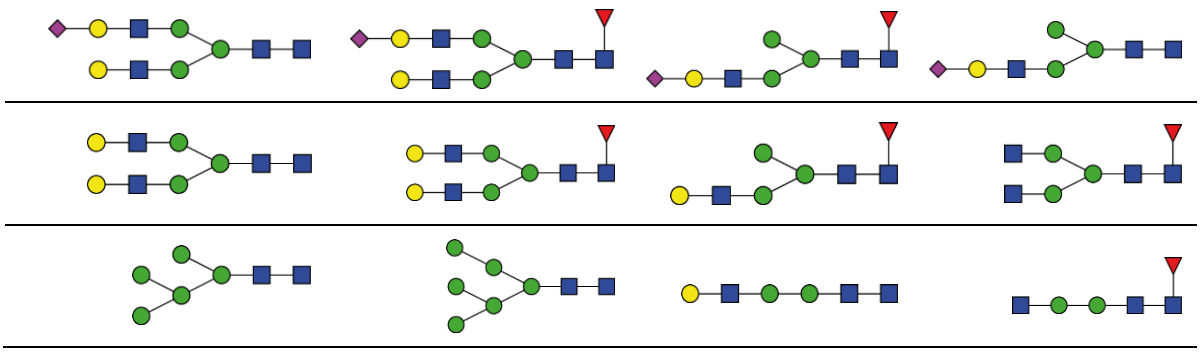
The glycosylation pattern ranges from ~ 64000 to ~ 66000 Da and is extremely similar for all individuals. Characteristic distances for saccharide structures are observable and

annotated. Furthermore, phosphorylation seems to be present in approximately 40-50%. Every peak is accompanied by an oxidation peak (+16 Da). One individual (cf. Figure 44, E) shows a mass shift of approximately 19 Da to higher masses. This could be due to a single nucleotide polymorphism that causes an amino acid exchange, H→R (ss1687332202, MAF: 80%). However, the amount of individuals holding no such SNP is surprising, thus, potentially another SNP is responsible. However, the peak pattern is not split, neither for the twelve individuals nor for the one individual, thus, all seems to be homozygots.

The overall mass is approximately 7000 Da higher than that for the pure amino acid sequence. To explain this mass difference with three *N*-glycosylation sites (assuming that the two vicinal ^{344}NNSS^{347} glycosylation sites are not both occupied), relatively large glycans have to be linked to HRG. One of the most prominent peaks at 64630 Da corresponds to the amino acid sequence taken from literature plus 18 hexoses, 15 hexosamines, 1 *N*-acetyl neuraminic acid as well as 3 fucoses. However, due to the high mass of the protein as well as the heavy glycosylation, no isotope pattern is observed, so that the composition of monosacchrides could also differ.

To elucidate the structure of the *N*-glycans, the tryptic peptides were digested with PNGase F that cleaves at the proximal GlcNAc from the peptide backbone. Table 9 summarizes twelve hereby detected *N*-glycans. However, the *N*-glycan with the composition $Hex_5HexNAc_4NeuAc_1$ is the most prominent glycan. EIC integrals for all further glycan structures were < 1% compared to the major *N*-glycan. With respect to this result, the mass difference of ~ 7 kDa could be explained by three times the major glycan structure (1931.688 Da) plus an additional small glycan, e.g. resulting from biosynthesis rupture (such structures were detected after PNGase F digestion, cf. Table 9) or caused by *O*-glycosylation. The mass distance of ~ 657 Da (visualized in Figure 44) could be explained by addition of one of the following two compositions: $Hex_1HexNAc_1dHex_2$ (MW = 657.244) or $Hex_1HexNAc_1NeuAc_1$ (MW = 656.223). These structures could either be explained by an additional antenna on an *N*-glycan or constitute for an *O*-glycan.

Table 9: **Identified oligosaccharide compositions attached to HRG by LC-MS analysis of free *N*-glycans (released with PNGase F from tryptic peptides). First row/first column: Major *N*-glycan $Hex_5HexNAc_4NeuAc_1$. The other depictured *N*-glycans constitute less than 1% to the total glycan amount.**



A glycosylation site-specific analysis could be of interest, since HRG is known to be cleaved *in vivo* by plasmin. Studies revealed that plasmin cleavage of HRG can activate and disrupt HRG binding to other ligands.[129] Therefore, HRG was cleaved with human plasmin. Without reduction of disulfide bridges, only two domains were separated (~ 15 and 49 kDa, cf. Figure 45) that together have the mass of intact HRG. The mass differences between signals found on intact HRG split here: Whereas the small domain carries distances indicating dominantly attached fucose residues as well as the phosphorylation site, the larger domain shows mass distances corresponding to neuraminic acids. However, this domain possesses a poor resolution and signal intensity, so that other distances cannot be excluded.

A plasmin digestion followed by a reduction of the disulfide bridges separates the domains. Besides smaller domains lacking glycosylation patterns, the domain at ~ 15 kDa above described as well as a domain at ~ 33 kDa (cf. Figure 46) possess glycosylation patterns. In-gel digests of the separated domains were performed and indicated that the 33 kDa domain contains the N-terminus of HRG.

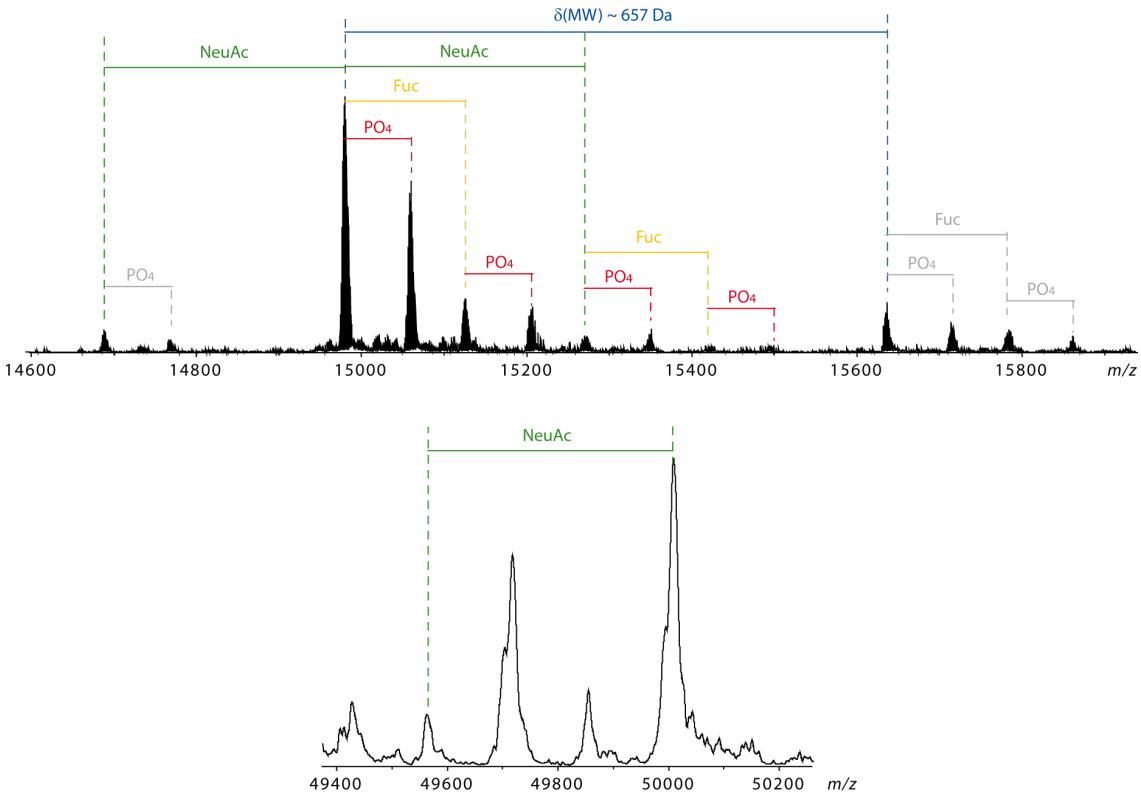


Figure 45: **HRG domains resulting from a plasmin digestion without subsequent disulfide bridge reduction. The glycosylation profile seen for intact HRG is split here: Only the small ~ 15 kDa (top) domain carries the phosphorylation as well as the fucosylation. Distances for NeuAc residues can be seen in both domains. More monosaccharide distances may be found hidden under signals of the larger domain (poor resolution as well as signal intensity).**

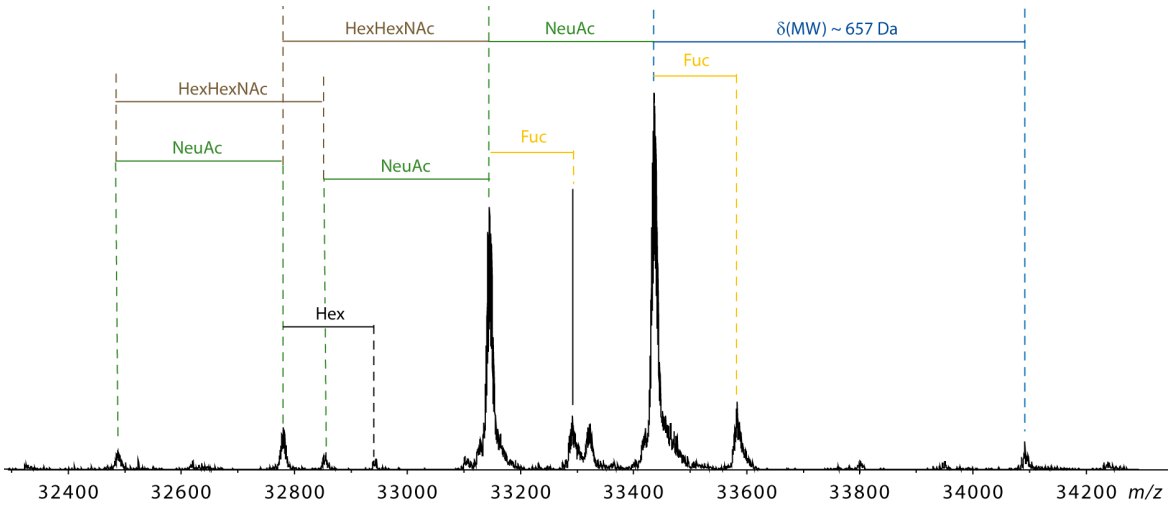


Figure 46: **HRG domain resulting from disulfide bridge cleavage following a plasmin digestion. Tryptic digestion suggests that this domain contains the N-terminus of HRG.**

However, none of these domains could be successfully fitted by the protein sequence described in the literature combined with a large set of arbitrary glycans. Since plasmin is known to cleave specifically, the amino acid sequence in literature may be partly incorrect. This consequence also originated from data analysis of tryptic HRG glycopeptides as well as HRG digestion with pronase (mixture of endo- and exoproteases resulting in extremely short peptides). Diverse fragments in MS/MS spectra indicated comprehensive glycosylation. For some glycopeptides, large areas of the corresponding peptide backbone were found in MS/MS spectra, but none of these precursor masses could be explained by the peptides plus glycosylation. The *N*-glycosylation sites at ^{62}NTT^{64} and ^{125}NCT^{127} were detected in many fragment ions. However, precursors could not be explained by the amino acid sequence combined with *N*- and/or *O*-glycans. Simultaneously, the glycosylation site-containing peptides were not detected without glycosylation, either. After *N*-glycan release of tryptic peptides, the resulting sample mixture was analyzed. Diverse tryptic glycopeptides are still observable. Thus, *O*-glycosylation is likely, since these structures are not released by PNGase F.

Therefore, future studies should address complete protein sequencing. Special focus should be laid on the domains containing the glycosylation patterns. For this purpose, an isolation strategy was successfully established. First analyses of the *N*-glycosylation revealed a high uniformity between individuals. Such an absence of glycosylation variance is appropriate for an application as glyco-based biomarker. The major *N*-glycan attached to HRG was identified as $Hex_5HexNAc_4NeuAc_1$. A phosphorylation site present with high amounts in intact HRG was localized by proteolytic digestion. Plasmin cleavage results in separation of glycosylation sites which enables more unambiguous characterization, which can be applied to identify glycosylation variance between individuals.

5.3 Prolactin-Inducible Protein

Prolactin-inducible protein (PIP), also known as gross cystic disease fluid protein 15 (GCDFP-15), secretory actin-binding protein (SABP) and glycoprotein 17 (gp17), is a 13.5 kDa single-chain protein showing one *N*-glycosylation site at ^{105}Asn.[131] PIP is secreted from several exocrine glands and thus found e.g. in saliva, tears, seminal fluid and during pathogenic states of mammary glands.[132-134]

Expression of PIP is induced by the peptide hormone prolactin (PRL) and by androgens and down-regulated by estrogen.[132, 135] PRL is a peptide hormone secreted mainly by pituitary glands.[136] Functions of PRL could be identified in reproduction, immunoregulation, growth, development, neurotransmission or metabolism.[137-138] PRL was identified as an important factor during breast oncogenesis.[139] Breast cancer cells secrete high levels of PRL that stimulates the growth of these cells. Besides, most breast cancer cells possess PRL receptors. As a pituitary hormone PRL is also involved in the neural system, e.g. serum concentration of PRL is correlated with psychiatric conditions.[140-143] Since the expression of PIP is induced by PRL, it is natural to suppose a function of PIP in the medical areas described.

Many important biological functions have been identified for PIP, such as inhibition of bacterial growth, involvement in fertilization, immune regulation or tumor progression. PIP also interacts with fibrinogen, fibronectin, the Fc domain of IgGs, zinc-α2-glycoprotein and the CD4-T cell receptor.[133, 144] Recently, PIP was shown to be significantly increased in patients suffering from Autism Spectrum Disorder (ASD).[145]

Multiple functions of PIP during breast cancer progression have been demonstrated.[146] The level of PIP is regularly determined to control the success of chemotherapy in breast cancer. It is well known that the structure of the glycans is often a better marker for cancer than the concentration of the protein in plasma.[147] The appearance of several plasma glycoproteins such as PIP within saliva could simplify standard clinical procedures by introducing non-invasive saliva analyses.[148-149] PIP occurs at a relatively high concentration in saliva. Glycan analysis of PIP can thus be performed using a non-

invasive procedure.[148-149] Caputo *et al.* revealed that the *N*-glycosylation of PIP from different sources differs between diseased (isolated from breast cyst fluid of women with gross cystic disease) and healthy conditions (isolated from seminal fluids from healthy donors).[150]

5.3.1 Structural Analysis of the *N*-Glycosylation of Human Salivary PIP

In a previous work in this laboratory, it was found that seminal PIP carries an uncommon *N*-glycosylation pattern (Behnken, Dissertation, 2013). It was shown, that only biantennary *N*-glycan structures ($Hex_5HexNAc_4$) are present that hold a varying degree of fucosylation (n = 1-5). Here, the *N*-glycosylation of PIP from saliva was analyzed by various methods such as LC-MS(/MS) and ^{1}H-NMR and is shown in detail in the following sections.

5.3.1.1 Analysis of Highly Fucosylated *N*-Glycans by HR-LC-MS of Intact Salivary PIP

PIP isolated from saliva carries only highly fucosylated *N*-linked glycans. Therefore, salivary PIP can easily be detected during enrichment steps by MALDI-MS by a characteristic signal pattern at a MW of ~ 16.5 kDa (cf. Figure 47). Signal maxima show a distance of ~ 147 Da indicating varying amounts of fucosyl residues. Compared to seminal PIP (~ 15200-15900 Da), signals of salivary PIP also appear at higher masses (up to ~ 17300 Da) in MALDI spectra. Thus, salivary PIP carries structurally larger *N*-glycans than found for seminal PIP. Besides multiple mass differences indicating the presence of extensive fucosylation, a higher-ranked glycosylation pattern was detected. Three fucose ladders were found with a mass shift of approximately 657 Da between their maxima. These mass distances indicate the attachment of up to two additional antennae to the biantennary core structure found in seminal PIP, resulting in tri- and tetraantennary *N*-glycans.

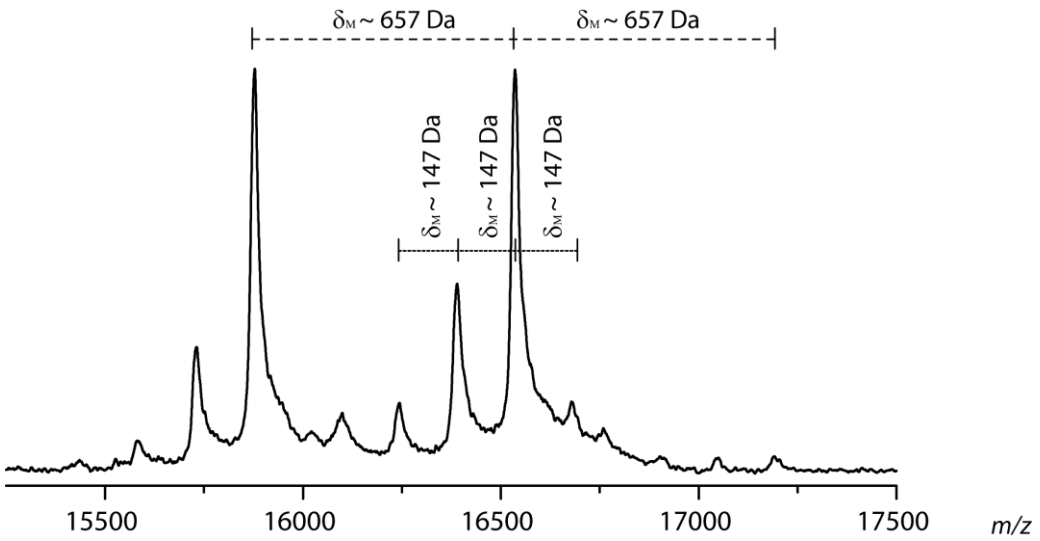


Figure 47: **MALDI-MS spectrum of salivary prolactin-inducible protein (PIP). Human PIP was enriched from saliva by miscellaneous chromatographic procedures. To monitor PIP enrichment, samples were analyzed by MALDI-MS in linear positive mode (matrix: dihydroxyacetophenone). PIP was detected at *m/z* ≈ 15500 - 17500 by the occurrence of characteristic mass distances of ~ 147 Da between signal maxima. The mass distance of 657 Da is indicative for an additional antenna.**

Due to the low resolution in MALDI-MS of larger proteins, high-resolution ESI-MS was used for detailed characterization of intact salivary PIP (cf. Figure 48). Native PIP was separated on a C_8-RP column and directly injected into the ESI-q/TOF. PIP occurs in several charge states (z = 8 to z = 13). A theoretical MS spectrum of neutral PIP is calculated by a deconvolution process of these charge states using a maximum entropy algorithm (cf. Figure 48 D).

These highly resolved isotope patterns of salivary PIP can be used for exact determination of the masses of individual glycoforms. A calculated isotope pattern of the glycoprotein is fit by a least square optimization to the experimental high-resolution MS data (see section 5.1.2 for details). In the ESI spectra of the intact glycoprotein it was possible to identify bi-, tri- and tetraantennary *N*-glycans. All of these antennae carry exceedingly high amounts of fucosyl residues. One can observe ladders of peaks that differ only by the mass of one fucosyl residue. For example, the highest intensity in the mass spectrum of intact PIP is representing a triantennary structure with a mass equivalent to seven fucosyl residues. PIP glycan species include bi-, tri- and

tetraantennary structures from two to five, seven and nine fucosyl residues, respectively (cf. Figure 50).

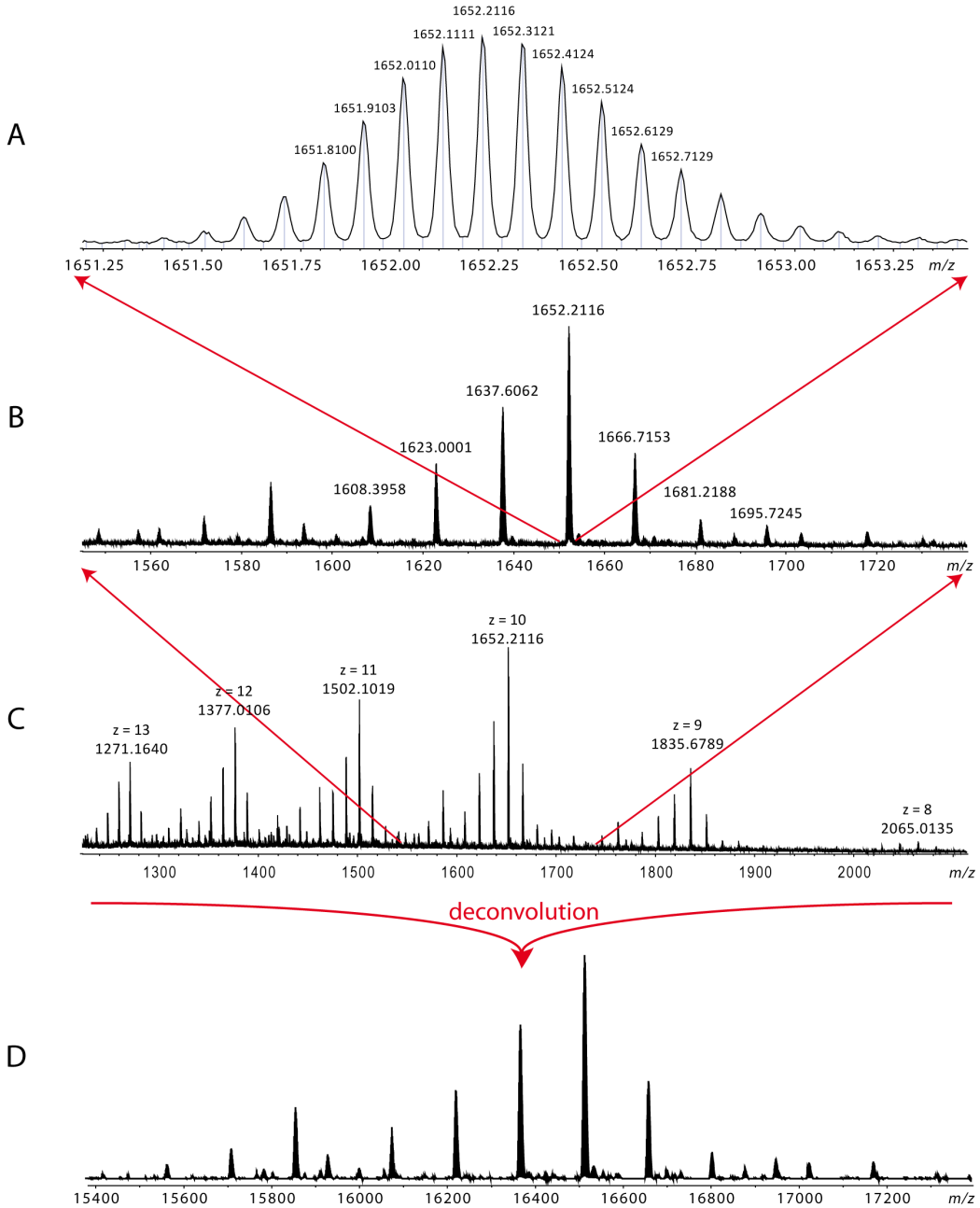


Figure 48: **ESI-MS spectra of salivary prolactin-inducible protein (PIP). Human PIP was enriched from saliva by miscellaneous chromatographic procedures. To monitor PIP enrichment, samples were analyzed by HR-LC-MS using an ESI-q/TOF MS. A: Baseline-separated peaks provide isotopically resolved components. B: Every charge state exhibits the glycoforms present on PIP (here: z = 10). C: PIP occurs in several charge states (z = 8 to z = 13). D: Deconvolution process of these charge states by a maximum entropy algorithm generates a calculated MS spectrum of neutral PIP.**

Data interpretation revealed three basic glycan structures, the bi-, the tri- and the tetra-antennary complex type *N*-glycans with an intense occupation by fucosyl residues. However, the signals cannot be explained by the isotope patterns of a single glycoform. A multivariate fit of the isotope patterns revealed that also *N*-acetyl neuraminic acid occurs, that has a mass of 1.0204 Da lower than two fucosyl residues combined (cf. Figure 49).

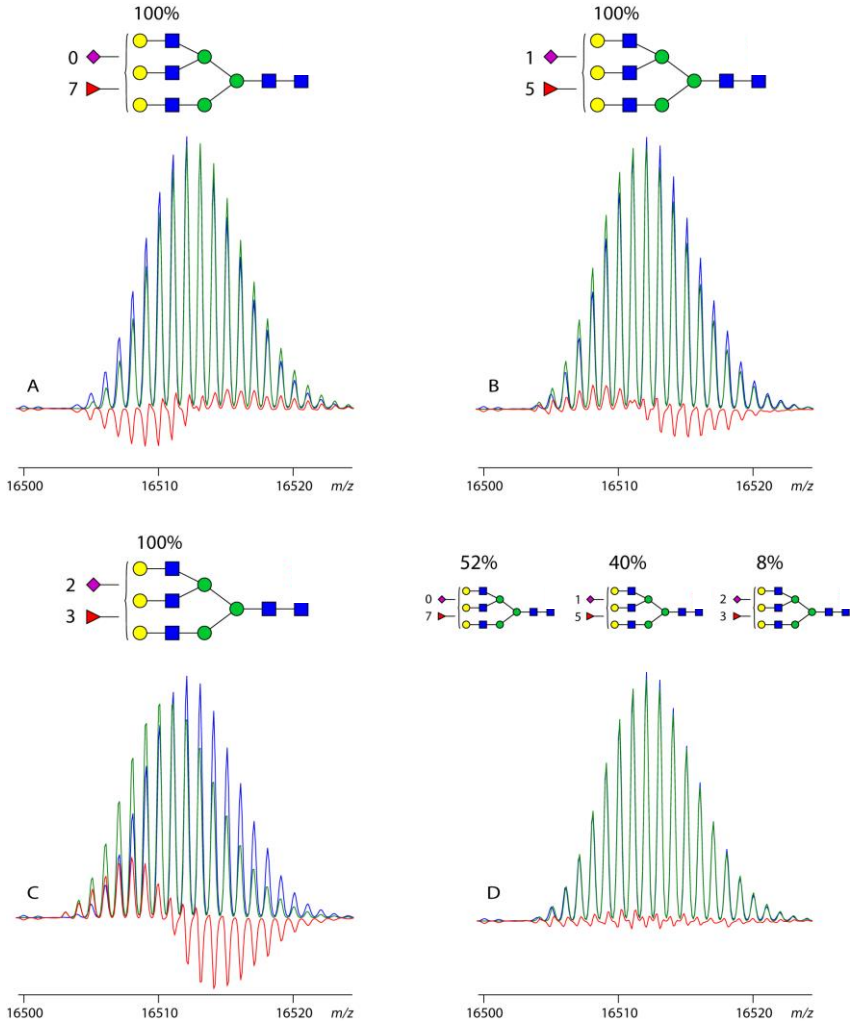


Figure 49: **Multivariate fit of native salivary PIP. The baseline-separated peaks allow the back-calculation of the isotope pattern of each glycoform. By summation of the isotope patterns of the protein backbone and variable glycan compositions, theoretical spectra are calculated and can be fitted to the experimental spectrum. By combination of multiple glycan combinations, the best fit was identified (the difference between experimental and theoretical spectrum reveals the best fit/glycan combination by the smallest integral). Blue: experimental spectrum, green: theoretical spectrum, red: difference spectrum. The experimental spectrum was fitted by a variation of different glycans or glycan combinations (A-D). The best fit was achieved for considering all three glycan variants (D).**

This procedure can be performed for the whole glycosylation profile. Figure 50 shows a spectrum of native PIP (female, age: 22) which was interpreted in this way. This interpretation is based on the fit of a calculated isotope pattern of the glycoprotein to the experimental high-resolution MS data. Therefore, it was possible to identify and quantify the composition of 33 *N*-glycans of PIP (cf. Table 21 in appendix).

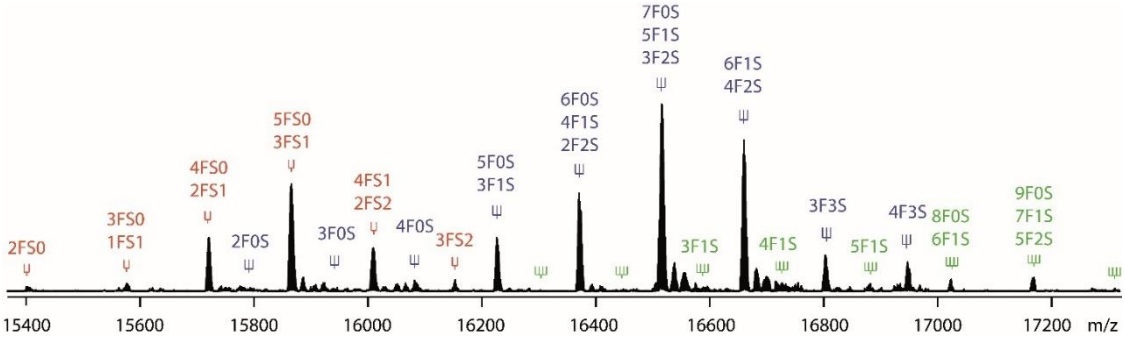


Figure 50: **High-resolution MS spectrum of intact PIP was interpreted by considering the isotope pattern of the glycoprotein. Bi-, tri- and tetra-antennary structures were detected with a full occupation of the antennae by fucoses (biantennary: ⊏, triantennary: ⊏, tetraantennary: ⊟). F: deoxyhexose (fucose), S: sialic acid (*N*-acetyl neuraminic acid). Only signals above an intensity threshold were interpreted, that allow unambiguous assignment of the isotope patterns.**

The presence of glycoforms that have an *m/z* that is one or two units lower than the isotope pattern of the exclusively fucosylated species could already be seen in several peaks when compared to the theoretical pattern. The mass of one *N*-acetyl neuraminic acid is only lower by 1.0204 Da than that of two fucosyl residues. Using a multivariate fit procedure to calculate the percentages of the individual components, it was revealed that e.g. the most intense peak in the region of ~ 16505-16525 Da (cf. Figure 51 A) arises from three triantennary glycans $H_6N_5F_7$, $H_6N_5F_5S_1$ and $H_6N_5F_3S_2$ (H = hexose, N = *N*-acetylhexosamine, F = fucose, S = neuraminic acid).

In order to test the multivariate fit, the molecules were also treated with neuraminidase. After neuraminidase treatment, the peak at 16515 Da splits into three molecular species with the following monosaccharide compositions: $H_6N_5F_7$ (unchanged due to a lack of neuraminic acids), $H_6N_5F_5$ (loss of one neuraminic acid) and $H_6N_5F_3$ (loss of two neuraminic acids). The peak at ~ 16501 Da (after neuraminidase treatment) shows only the isotope pattern of $H_6N_5F_7$ (cf. Figure 51 B).

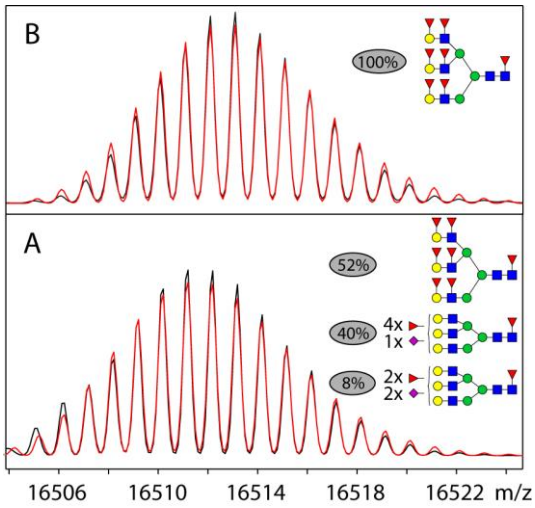


Figure 51: **Analysis of the baseline-separated isotope pattern of PIP shows a high number of fucosyl residues bound to *N*-glycan structures. Using a multivariate fit the glycan composition prior to (A) and after (B) neuraminidase digest was determined (black: experimental, red: calculated). (A) Native PIP shows an MS with peaks that are the result of the overlay of three species, two of them containing *N*-acetyl neuraminic acids. The intensities of the peaks are in agreement with the composition of the three compounds shown in the spectrum. (B) Neuraminidase treatment moves the two compounds with neuraminic acids from the spectrum window to lower masses and leaves only the isotope pattern of the heptafucosylated structure. Therefore, the isotope pattern is slightly shifted to the right (top): The heptafucosylated structure is expected to have a mass which is one to two units higher than that of the structure containing neuraminic acids. The theoretical spectrum of the heptafucosylated triantennary glycan is in excellent agreement with the experimental data. The braces indicate that the exact attachment points of the *N*-neuraminic acids and the fucoses are not known. Red triangle: fucose, pink diamond: acetyl neuraminic acid, yellow circle: galactose, green circle: mannose, blue square: *N*-acetyl glucosamine.**

The same procedure was executed for the other peaks (cf. Figure 52 B). After digestion, no glycans with neuraminic acids were chosen by the automatic MATLAB least square optimized fit (cf. Table 24), combination 1-50 for bi- and triantennae, all coefficients for glycans containing neuraminic acids are < 0.3%, although two exceptions with 0.8 and 0.7% exist, that did not appear in native PIP and therefore most likely stem from signals with low S/N). Amount of bi- and triantennae were equal prior and after neuraminidase digestion. Prior to neuraminidase digestion only 48% of all glycans could be explained by structures without neuraminic acids. Figure 53 shows the distribution of bi- and triantennary glycans after neuraminidase digestion, differing only by the amount of fucosyl residues. The complexity is reduced from 33 to 18 different glycans.

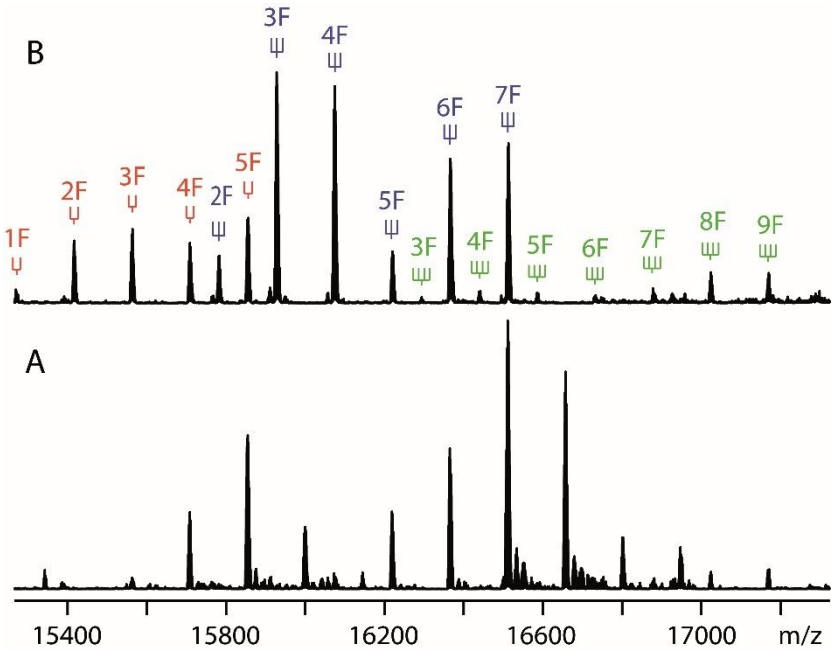


Figure 52: **ESI-q/TOF MS spectra of intact prolactin-inducible protein (PIP) are shown prior to (A) and after (B) neuraminidase digest. The peak pattern in (B) is shifted to lower masses validating the presence of *N*-acetyl neuraminic acids in native PIP. However, highly fucosylated structures with up to nine fucosyl residues are present. Ladders of peaks can be found that differ in the mass of one fucosyl residue each. Bi-, tri- and tetraantennary structures were detected with a full occupation of the antennae by fucoses (biantennary: ⊏, triantennary: ⊏, tetraantennary: ⊏).**

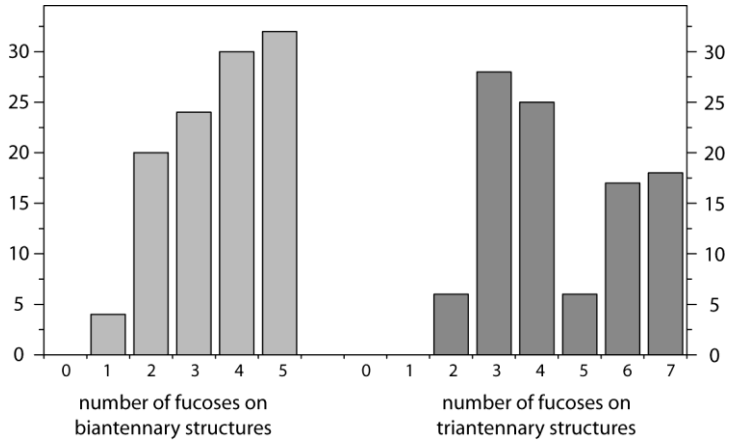


Figure 53: **Distribution of fucose residues on sialidase-treated salivary PIP. Bi- and triantennae carry predominantly a high degree of fucosylation. The major biantennary glycoform is fully occupied with Fuc residues on antennae and core. Some triantennary structures also carry a full fucosylation, however, the most abundant structure is a triantenna with three fucoses (most likely core-fucosylated with two fucoses in the antennae).**

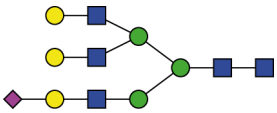
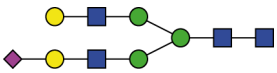
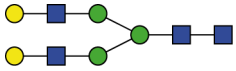
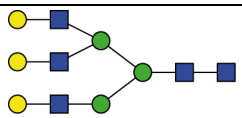
Still, we find triantennary structures with two to up to seven fucosyl (Fuc) residues. Likewise, the biantennae carry one to five Fuc residues and the tetraantennae two to nine

Fuc residues. This represents an unusually pattern of *N*-glycan structures from an individual glycoprotein with an extremely high amount of fucosylation.

5.3.1.2 Analysis of Free *N*-Glycans from Salivary PIP by LC-MS/MS

For further structural confirmation of the presence of highly fucosylated oligosaccharides, the *N*-glycans were released with Protein-*N*-glycosidase F (PNGase F). PNGase F cleaves *N*-glycans at the proximal GlcNAc from the protein backbone. The *N*-glycans released were separated on a PGC (porous graphitized carbon) column and analyzed by MS(/MS). Table 10 summarizes the *N*-glycans identified. Seven core structures with varying amounts of fucose residues were detected. Several highly-fucosylated triantennae were also identified, although structurally large *N*-glycans often show a reduced release because of steric hindrance of PNGase F. Correspondingly, the absence of free (mainly highly fucosylated) tetrasaccharides is not surprising.

Table 10: **Released *N*-glycans from human salivary PIP. Oligosaccharides were cleaved with protein-*N*-glycosidase F from PIP and analyzed by LC-MS(/MS). Separation occurred on a porous graphitized carbon column. *: Low intensity and S/N ratio.**

Core Structure	n (Fucoses)
	0, 1, 2, 3, 4, 5, 6, 7
	0, 1, 2, 3, 4
	1, 2, 3, 4, 5
	1, 2, 3, 4, 5, 6*, 7

Thus, the high degree of fucosylation is also observed for *N*-glycans released from PIP. Up to two fucosyl residues per Gal-GlcNAc disaccharide were observed per antenna, which leaves two options for the antigenic nature of the terminal tetrasaccharides: either Lewis^{b} (Fucα[1-2]Galß[1-**3**](Fucα[1-**4**])GlcNAcß) and/or the isobaric structure Lewis^{y}

(Fucα[1-2]Galß[1-**4**](Fucα[1-**3**])GlcNAcß). Because MS analyses of the released oligosaccharides and the intact glycoprotein (cf. section 5.3.1.1) do not give information about the linkage, MS/MS was performed. MS/MS data show that even in cases with a low number of fucosyl residues, such that also Le^x type structures could explain the precursor masses, significant amounts of Le^y type structures are present indicating a preferred clustering of fucoses on the same branch (cf. Figure 55). Figure 55 shows one example of the summed up MS/MS spectrum for the precursor ion at *m/z* = 1040.390. This mass corresponds to the biantennary structure $H_5N_4F_3$ (H: hexose, N: hexosamine, F: deoxyhexose). Several fragments indicate the linkage of the fucose residues to various monosaccharides in the antennae, thus distinguishing between two Le^x motifs (e.g. fragments at *m/z* = 512.2 and 674.2) and one Le^y motif (e.g. fragments at *m/z* = 658.3 and 893.3). Consequently, the structure contains already a significant amount of the Le^y motif. However, because of rearrangement of fucosyl residues in the gas phase, the antigenic nature cannot be unambiguously identified by MS/MS.[151]

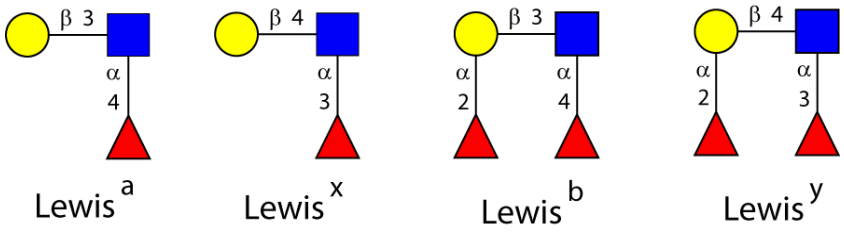


Figure 54: **Possible attachments of fucoses (n = 1-2) to the terminal disaccharides, derived from the Lewis antigen system (human blood group system). Yellow circle: galactose, blue square: *N*-acetylhexosamine, red triangle: fucose.**

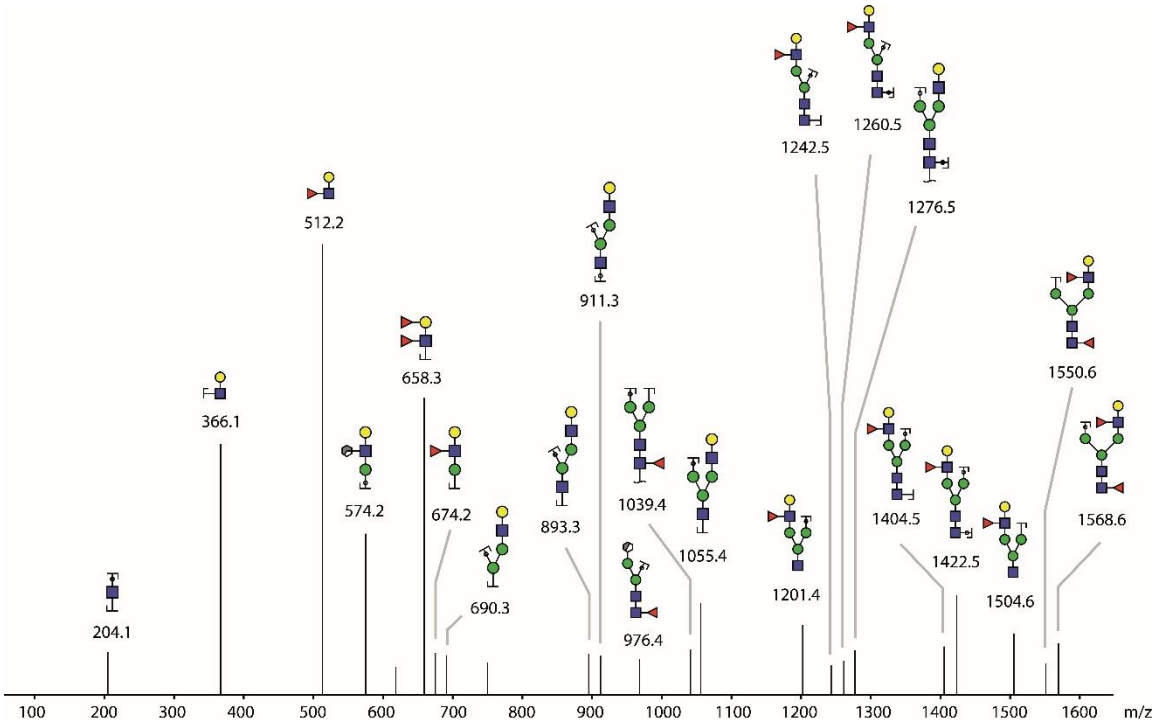


Figure 55: **Summed up CID MS/MS spectrum of a biantennary *N*-glycan with three fucoses ($Hex_5HexNAc_4dHex_3$) Fragments indicate that two isobaric structures occur, namely Lewis^x (one fucose at each antenna) and Lewis^y (two fucoses in one antenna). Several fragments can be explained by others than the depicted structures. Therefore, for annotation only the most likely fragments were used (e.g. B/Y fragments).**

5.3.1.3 Analysis of Free *N*-Glycans from Salivary PIP by ^{1}H-NMR and TOCSY-NMR

High-resolution MS data led to the conclusion that PIP contains a high amount of Lewis epitopes. Many glycan structures hold antennae with two fucoses each. MS alone cannot be used to unambiguously characterize linkages of the *N*-glycan monosaccharides. In order to assign the structures of the terminal highly fucosylated tetrasaccharides, it is important to elucidate the linkage configuration as well as the attachment points of the fucosyl residues. Therefore, the *N*-glycans were enzymatically released from PIP with PNGase F and separated on a PGC column (porous graphitized carbon). The LC flow was fractionated, lyophilized and re-dissolved in D_2O for NMR analysis. To elucidate the fucosylated antigenic structures, ^{1}H-NMR and TOCSY spectra were acquired. Signals characteristic for Le^y were found at δ = 5.278 (α-1,2-Fuc, H1) and 5.268 ppm (α-1,3-Fuc, H-1) as well as δ = 1.274 ppm (α-1,2-Fuc, CH_3) (cf. Figure 56). These chemical shifts are not present in other antigenic epitopes.

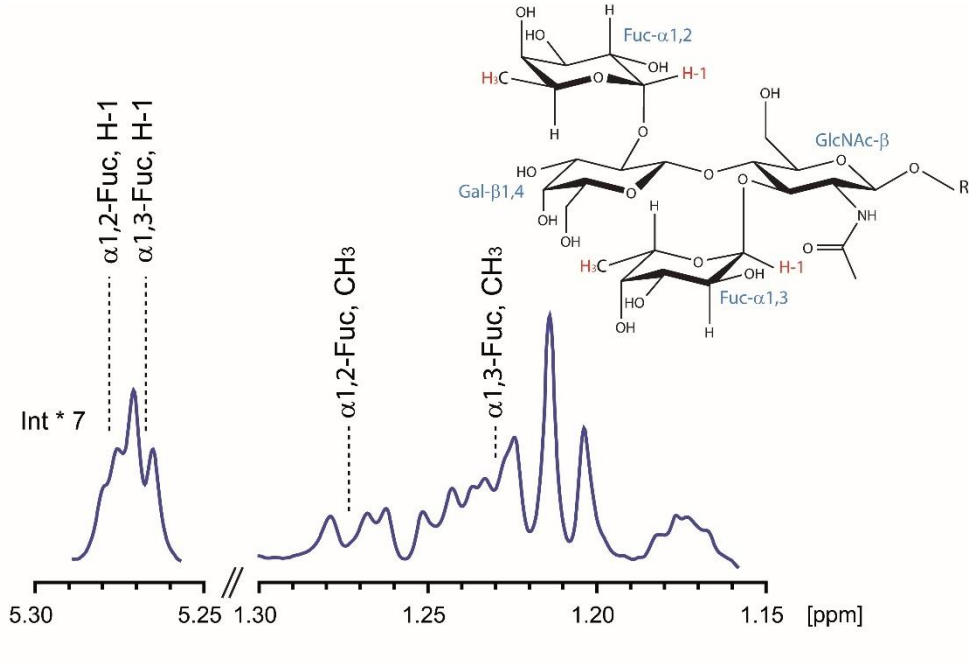


Figure 56: **^{1}H-NMR spectrum of free *N*-glycans of salivary PIP separated on a PGC column and analyzed by ^{1}H-NMR. Selected details of the summed up ^{1}H-NMR spectrum are depicted. Characteristic signals for Le^y are annotated, e.g. H-1 and CH_3 of fucosyl residues.**

In the TOCSY spectrum the chemical shifts for the H-2 and the H-5 signals could be assigned to δ_{H2} = 3.79 and 3.82 ppm and δ_{H5} = 4.26 ppm, respectively (cf. Figure 57). These values are in excellent agreement with NMR data of synthetic Le^y structures.[152] These NMR data unambiguously confirm the unusual presence of Le^y epitopes on *N*-glycans. The presence of triantennary structures was confirmed by the NMR signal at δ = 4.217 ppm representing H-2 of D-Man in triantennae. Signals at δ = 1.19-1.16 ppm represent methyl groups of α-L-Fuc of Le^x and peaks at δ = 1.220-1.208 ppm arise from the core fucose methyl groups.

Thus, PIP carries highly fucosylated *N*-linked glycans as Le^y structures. Such structural elements, e.g. attached to a triantenna, constitute an extremely bulky carbohydrate portion compared to the protein backbone of PIP (cf. Figure 58).

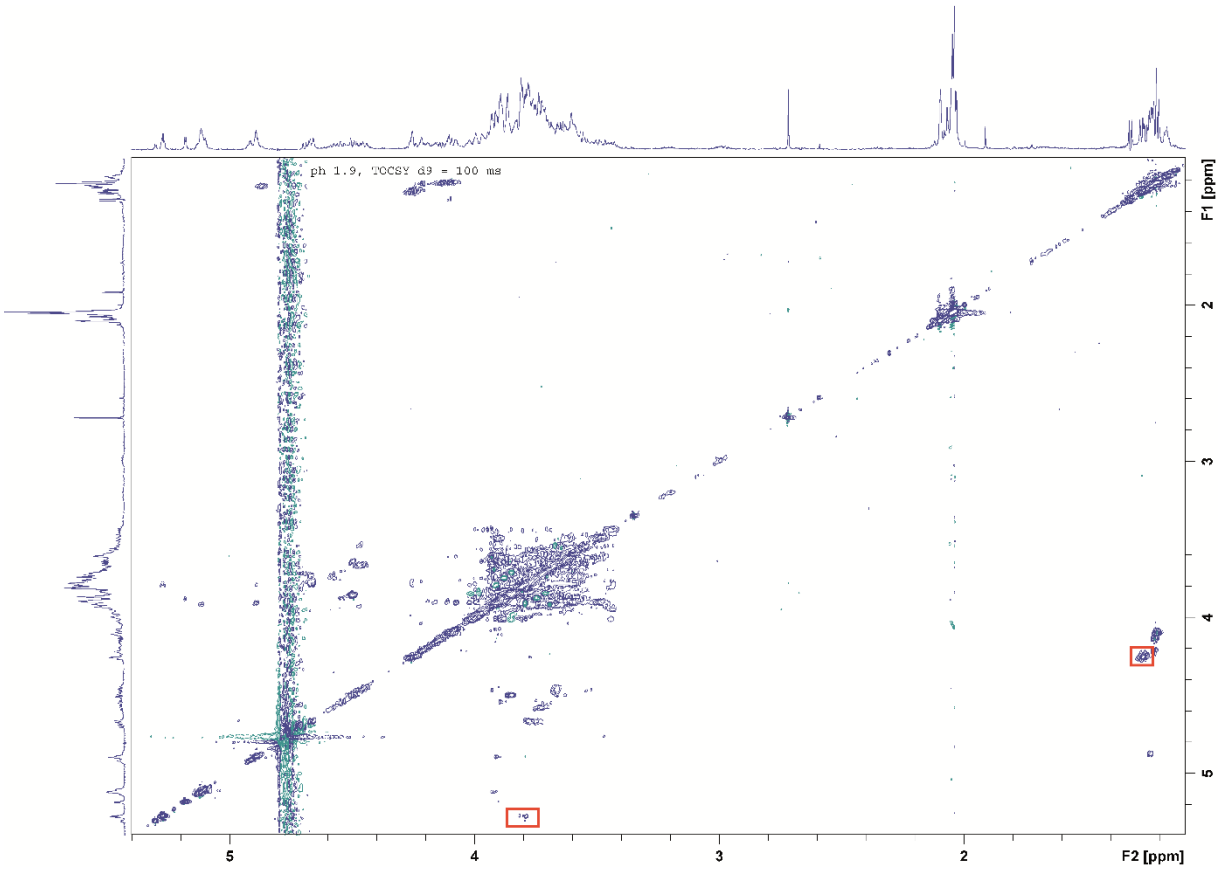


Figure 57: **TOCSY spectrum of free *N*-glycans of salivary PIP separated on a PGC column and analyzed by TOCSY-NMR. Chemical shifts for the H-2 and the H-5 signals could be assigned to δ_{H2} = 3.79 and 3.82 ppm and δ_{H5} = 4.26 ppm. Red-framed: cross signals characteristic for Le^y structures.**

In many organs, Le^y is only present and up-regulated in tumor tissue and an indication for a poor prognosis. So far only *O*-type structures with Le^y were reported for a specific glycoprotein, i.e. Mucin 1. Commonly occurring changes in tumor diseases are the increase of *N*-glycan branching, decrease of *O*-glycan chains, sialylation and the presence of Lewis antigens.[153-155] Especially Lewis^x, sialyl-Lewis^a and Lewis^y epitopes are strongly associated with a poor prognosis of disease.[156-159] These epitopes were found to be expressed in elevated levels on cell membranes in most tumor diseases.[160-163] Contrary, in non-small cell lung cancer (NSCLC) Le^y expression on tumor cells was associated with better disease progression and improved survival.[164] Le^y epitopes are naturally occurring only in the cytoplasm in epithelial cells of the gastrointestinal tract.

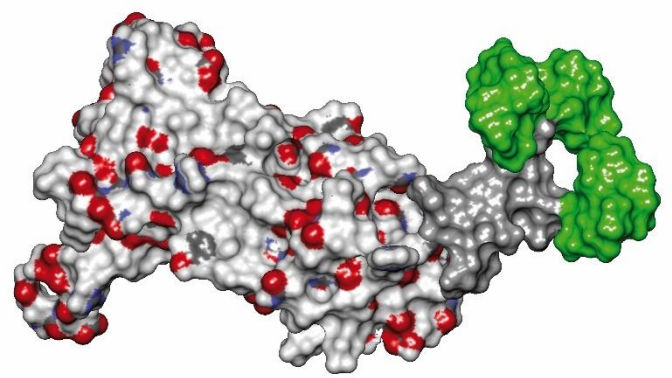


Figure 58: **Space-filling model of *N*-glycosylated salivary PIP based on results obtained by LC-MS, LC-MS/MS and NMR analyses. The most abundant *N*-glycan identified (triantenna with seven fucoses, terminal antennae constitute NMR-characterized Le^y motifs) was chosen as carbohydrate. Dark grey: fucosylated carbohydrate core structure with the composition $Hex_3HexNAc_2Fuc_1$, green: three terminal Le^y motifs.**

However, on tumor cells Le^y antigens are also found on the cell surface.[165] Le^y epitopes were up to now only detected and unambiguously characterized on *O*-linked glycans by NMR analyses.[154, 166-168] Clark *et al.* have shown oligosaccharides from pooled seminal glycoproteins containing highly fucosylated *N*-glycans carrying Le^y structures by Lectin- and MS based-studies.[169] Le^y structures on *N*-type glycoproteins have been described in the literature based on MS evidence.[169-170] These antigenic epitopes are typically expressed in many tumor tissues, here, they are found on salivary PIP of healthy individuals. This high prevalence of these Lewis^y antigens needs more research to understand its involvement as tumor marker, as tumor prognostic and also in other diseases e.g. neurodegenerative disorders.

5.3.2 Enrichment of Prolactin-Inducible Protein from Human Saliva

To evaluate the occurrence of this uncommon glycosylation pattern, salivary PIP from multiple individuals was analyzed. Therefore, a reproducible and glycosylation-independent enrichment of PIP from saliva had to be established. The following sections describe and compare the results from anion exchange-, size exclusion- and antibody affinity-chromatography.

5.3.2.1 Anion Exchange Chromatography for Enrichment of PIP

Ion exchange chromatography (IEX) is commonly applied to separate ionic analytes. In AIEX (anion exchange chromatography) experiments, anionic analytes bind to a cationic stationary phase. By stepwise increasing the ionic strength of the elution buffer, analytes can be separated according to their ionic character. Human PIP has a calculated pI of 5.4 and is therefore appropriate for AIEX.

Several stationary phases convenient for AIEX were tested for efficiency of PIP enrichment from human saliva. Besides the resin, parameters such as elution buffer composition and gradients were adjusted for optimal PIP enrichment. Three stationary phases from GE Healthcare were used: a) HiTrap Q FF - strong anion exchanger (SAX)/ligand: quaternary amine - prepacked column for FPLC, b) HiTrap DEAE FF - weak anion exchanger (WAX)/ligand: diethylaminoethyl - prepacked column for FPLC, c) DEAE Sephadex A-50 - weak anion exchanger (WAX)/ligand: diethylaminoethyl - stationary phase for batch or manual column chromatography. In Figure 59 HR-MS spectra from the eluates using these three stationary phases are contrasted with each other.

The weak anion exchanger phases (WAX, cf. Figure 59, middle/bottom) are significantly more efficient in isolating salivary PIP than the strong anion exchanger (SAX, cf. Figure 59, top). Besides enhanced S/N ratio, spectra show a significantly higher number of glycoforms. Both WAX showed reproducible results. WAX HiTrap DEAE FF shows an improved S/N ratio compared to WAX DEAE Sephadex A-50. Furthermore, less signals overlay especially in the mass range of tetraantennary glycans. Additionally, HiTrap DEAE FF is an FPLC column and thus more appropriate for reproducible chromatography on an automated LC system. Therefore, WAX HiTrap DEAE FF was chosen for the isolation of PIP from several healthy adults (buffer: 20 mM Tris-HCl, pH: 8.5, NaCl-gradient for elution up to 500 mM within 30 min, cf. section 6.6.3).

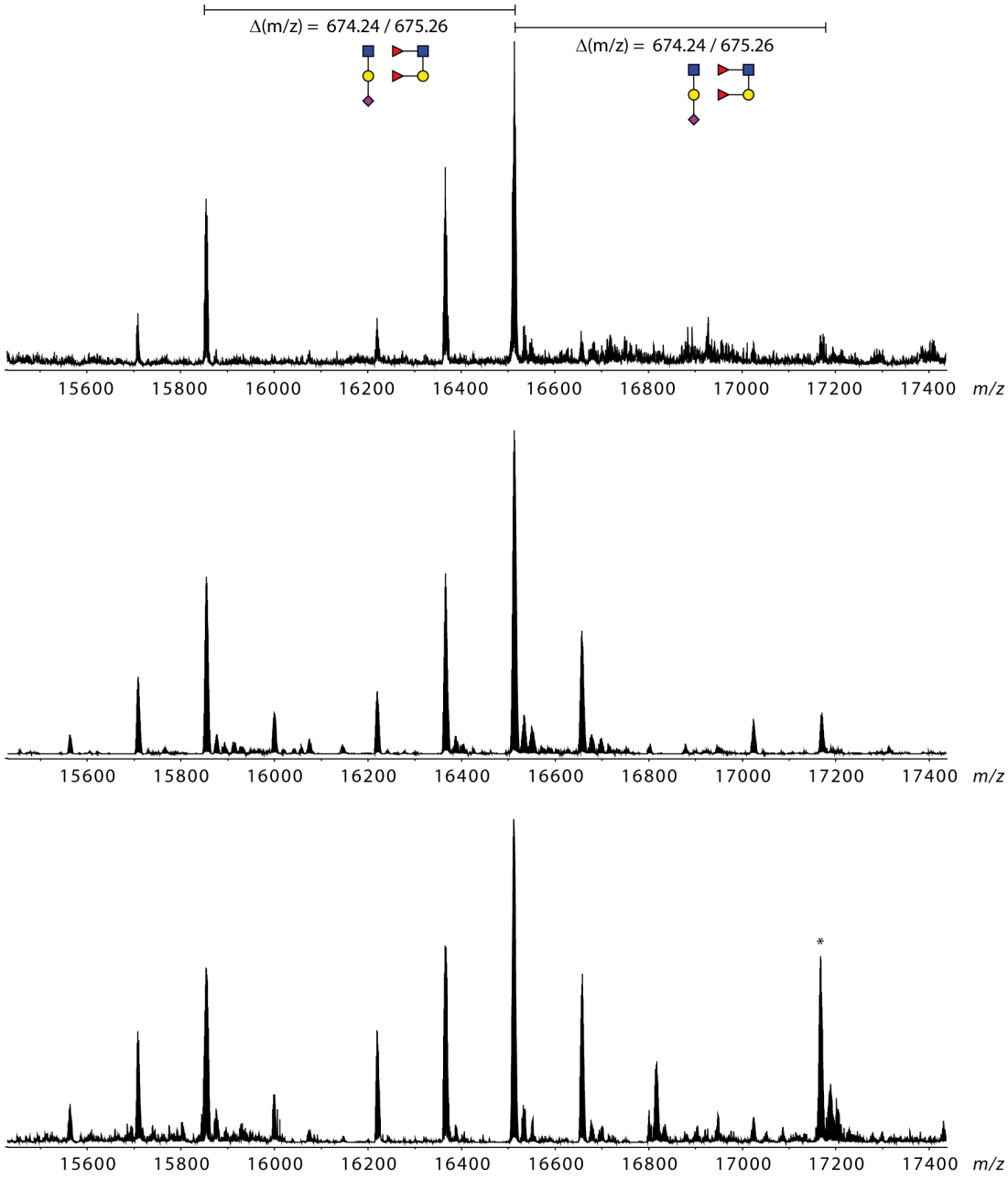


Figure 59: **Deconvolved MS spectra of native PIP enriched by three different stationary phases. Top: Stationary phase: FPLC column HiTrap Q FF, buffer A: 10 mM Tris-HCl/pH 8.5; buffer B: 10 mM Tris-HCl / 500 mM NaCl / pH 8.5; gradient A (fractions 13-16). Middle: Stationary phase: FPLC column HiTrap DEAE FF, buffer A: 20 mM Tris-HCl / pH 8.5; buffer B: 20 mM Tris-HCl / 500 mM NaCl / pH 8.5; gradient A (fractions 28-30). Bottom: Stationary phase: manual chromatography with DEAE Sephadex A-50, buffer A: 20 mM Tris-HCl / pH 6.2; buffer B: 20 mM Tris-HCl / 500 mM NaCl / pH 6.2; step gradient B, fraction 150 mM NaCl). *: Impurity, not associated with PIP.**

5.3.2.2 Enrichment of PIP by Size Exclusion Chromatography

Charged glycoforms have an influence on retention on AIEX phases. Depending on the stationary phase, different glycoforms were enriched. Therefore, size exclusion chromatography (SEC) was chosen to enrich PIP from other saliva components solely based on the hydrodynamic radius of PIP. Indeed, PIP elutes in a retention time window shortly after components with MW ~ 60 kDa. This shift to higher masses is caused by the enormous carbohydrate portion of PIP that increases its hydrodynamic radius.

Enrichment of PIP by SEC provided reproducible results. Figure 60 shows the MS spectra of intact salivary PIP from six healthy adults. All spectra confirm that presence and relative distribution of the highly-fucosylated *N*-glycans seems to be a common pattern of human salivary PIP.

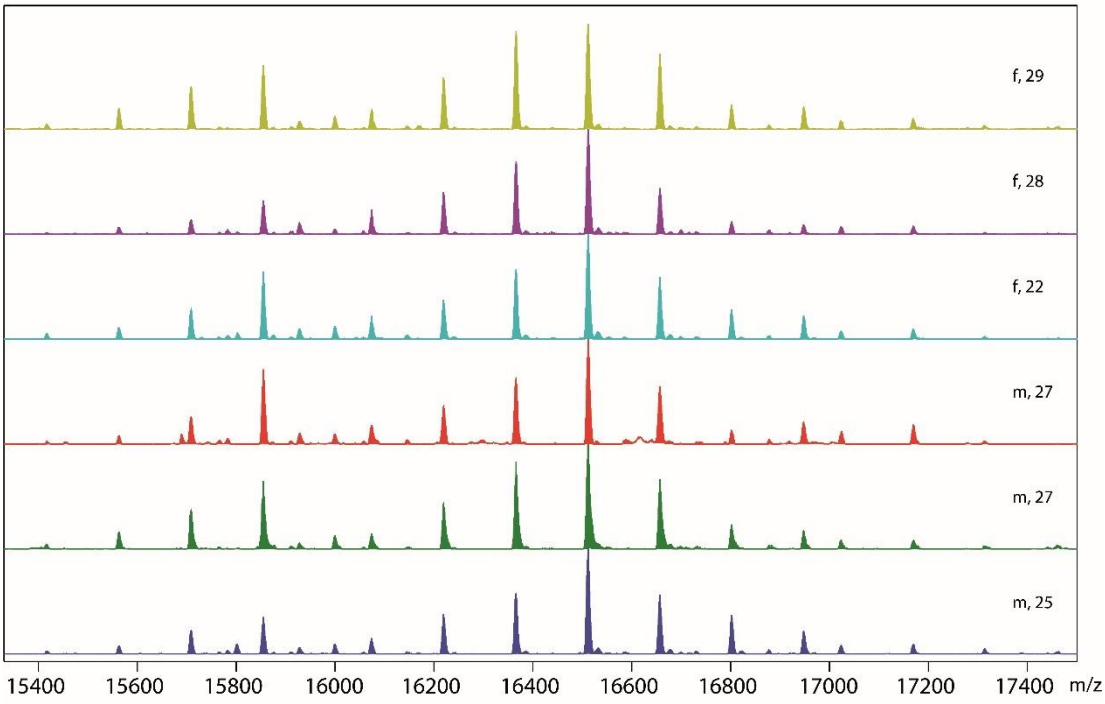


Figure 60: **Deconvolved MS spectra of intact salivary PIP from six healthy individuals (age: 22 to 29 years, f: female, m: male) enriched by SEC and further analyzed by RP-LC-MS. The unusually high degree of fucosylation seems to be a common feature of PIP in saliva.**

5.3.2.3 Enrichment of PIP by Combinatory Use of AIEX and SEC

Isolation of salivary PIP was also performed by AIEX in combination with SEC to possibly improve removal of disturbing contaminants. Saliva samples (13 individuals) were prefractionated by AIEX and subsequently combined for SEC. The full-scan MS exhibits a high purity ensuring a high degree of confidence by use of the maximum entropy method. Furthermore, a multiplicity of signals is recognizable and identifiable, although a detailed signal assignment was avoided, since the SEC was performed with pooled AIEX fractions from various test persons that are partly heterozygous for a nonsynonymous single nucleotide polymorphism (cf. section 5.3.3). The MS spectrum shows a high number of glycoforms, especially the portion of tetraantennary glycans is increased. Intensity and S/N ratio is high. However, during AIEX a complete preservation of all glycoforms has to be ensured.

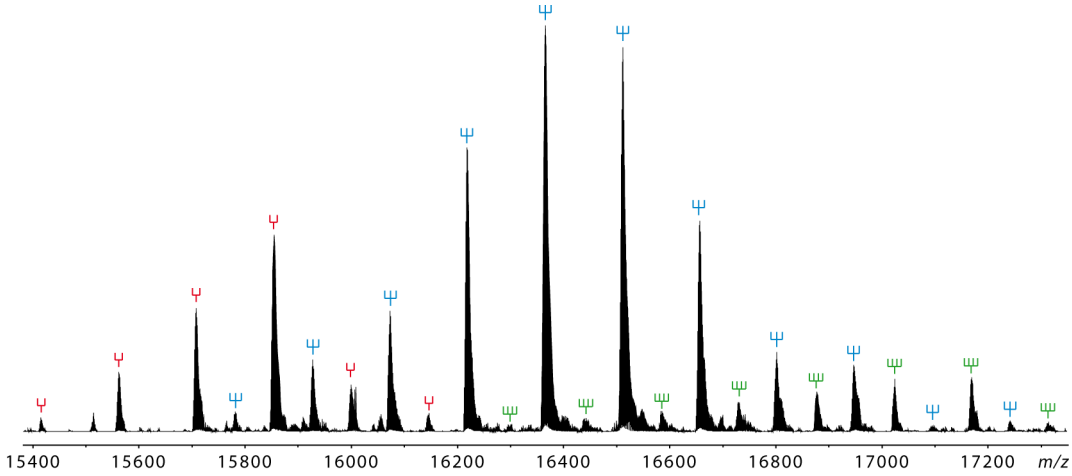


Figure 61: Deconvolved MS spectrum of intact salivary PIP (pooled from several individuals) isolated by AIEX and SEC, followed by analysis by RP-LC-MS.

5.3.3 Nonsynonymous Single Nucleotide Polymorphism in Salivary PIP

During this work a single nucleotide polymorphism (SNP, rs73170678) known in literature (characterized by genome analyses) was identified in MS spectra. It can be categorized as of the nonsynonymous, missense SNP type. The SNP causes an amino acid change from ^{79}Gln to ^{79}His and results in a mass shift of 9.01 Da to higher masses (Figure 62). Only heterozygous humans were identified showing the signal splitting. The

SNP is described with an overall frequency of 10.1% (> 60000 analyzed humans). For Europeans (exclusively saliva from Europeans was analyzed here) the MAF was determined to 14.3%.[171] In this work 40% of all individuals showed a SNP (MAF = 20%) and therefore a significantly enhanced ratio. Interestingly, the documented MAFs significantly deviate between ethnical groups (e.g. Europeans: 14.3%, Africans: 2.0%, Americans: 6.1%).

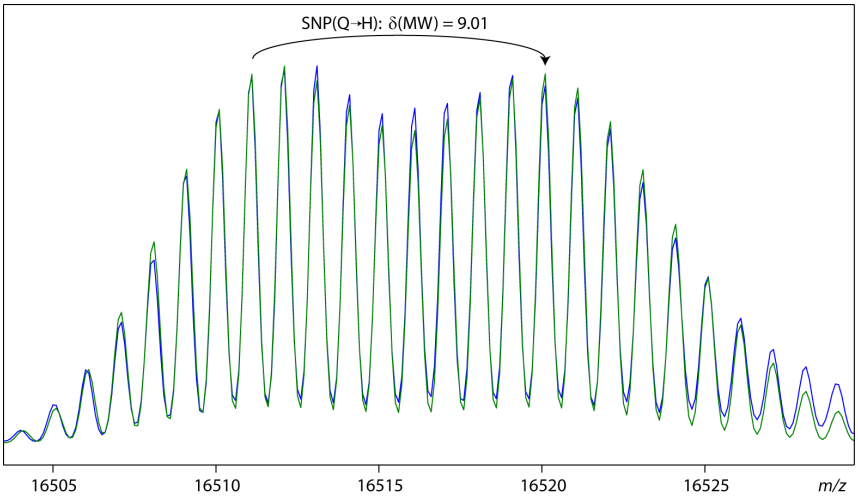


Figure 62: **Mass spectrum of a heterogeneous individual showing the SNP $^{79}Q \rightarrow {}^{79}H$ in salivary PIP resulting in a mass shift of 9.01 Da. Blue: experimental spectrum, green: theoretical spectrum of PIP with and without the amino acid exchange in combination with the underlying glycoforms ($H_6N_5F_7S_0$, $H_6N_5F_5S_1$, $H_6N_5F_3S_2$, H: hexose, N: hexosamine, F: fucose, S: sialic acid).**

Interestingly, the ratio of 1:1 of the two proteins depicted above is not universal. The ratio deviates between individuals (cf. Table 11). This observation could be explained e.g. by an RNA-interference. In this process, the expression of several proteins can be reduced by binding of siRNAs (small interfering RNA) or miRNAs (microRNA) to mRNA, that is subsequently cleaved by proteins within the so-called RNA-induced silencing complex.

Table 11: Observed SNP-ratios of expressed PIP per individuum. Only three of eleven SNPs show a distribution of 1:1. Frequency indicates the number of individuals showing the observed ratios. The first number in the column "ratio" is PIP without the amino acid exchange. See Table23 in appendix for donor information.

Sex	Ratio	Frequency	Sample ID
Female	1:0.25	I	SEC_f_2
	1:0.85	I	AIEX_f_1
	1:1.3	I	SEC_f_1
	1:1	I	AIEX_f_3
	1:1.5	II	AIEX_f_2, AIEX_f_4
Male	1:0.7	I	SEC_m_4
	1:1	I	SEC_m_3
	1:1.2	III	SEC_m_1, SEC_m_2, AIEX_m_1

However, the ratios observed for each glycoform are the same within each individual (cf. Figure 63) and thus independent of the glycoform.

An incomplete summation of one of the expressed proteins during the summation of LC-MS spectra based on a shifted retention time can be excluded, since the summation of all spectra containing PIP was ensured by manual inspection. Nonetheless, a retention time shift of different PIP species was observed. The shift in retention time caused by the amino acid exchange is less than the shift by glycoform, thus every glycoform elutes time-shifted for PIP with and without the SNP as depicted in Figure 64.

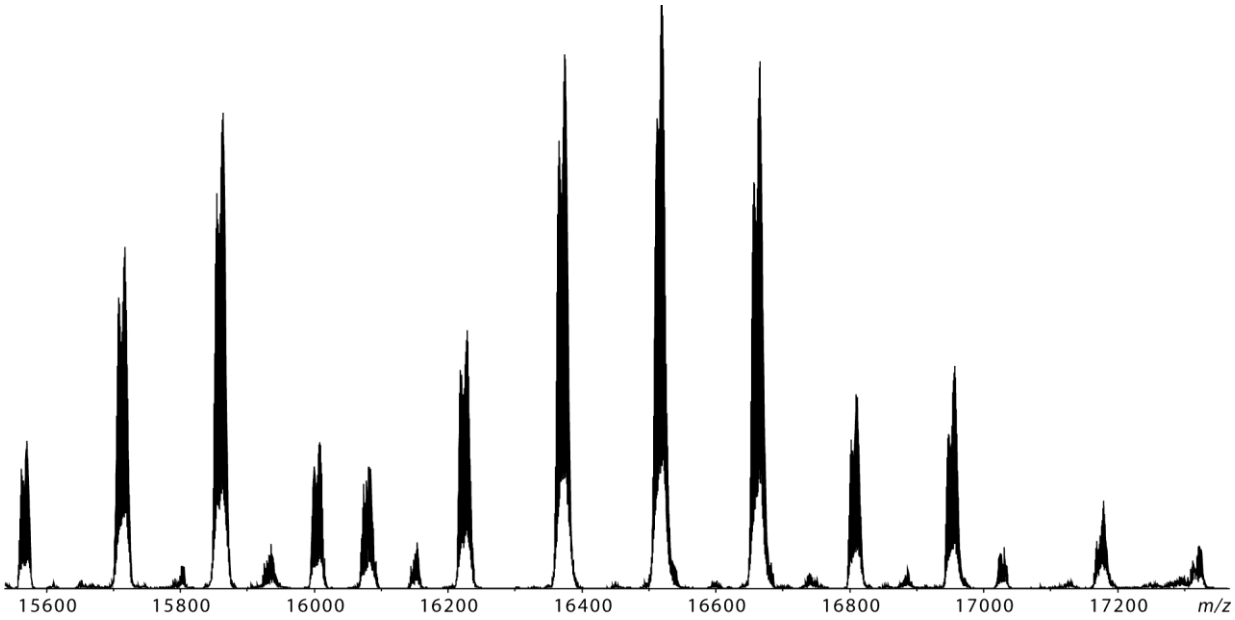


Figure 63: **Mass spectrum showing the amino acid exchange caused by a nonsynonymous SNP within PIP from a male individual (age: 34 years). The ratio of the expressed proteins is the same for all glycoforms.**

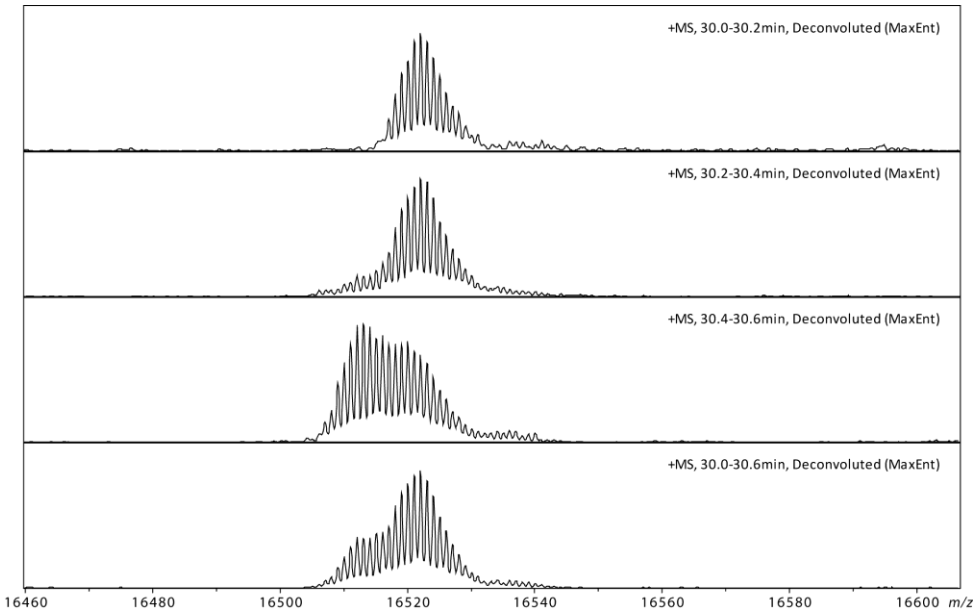


Figure 64: **Shift of PIP from a male individual (age: 34 years) based on amino acid exchange during RP-LC. MS spectra were summed up stepwise (Δt = 0.2 min, upper three spectra) and in total (Δt=0.6 min, bottom spectrum) for one glycoform. PIP with the amino acid exchange (^{79}His) elutes prior to PIP without the amino acid exchange (^{79}Gln).**

5.3.4 Comparison of the *N*-Glycosylation of Salivary PIP from Healthy Individuals

For a potential clinical testing of salivary PIP as biomarker for a pathological state, e.g. breast cancer or neurodegenerative diseases, based on an altered glycosylation profile, the glycan structures have to be analyzed from several test persons to elucidate the variance of PIP glycosylation. The analyses are based on spectra of intact PIP. Their evaluation was achieved by MATLAB using the calculation of isotope patterns described in section 5.1.2. Two experimental approaches were performed: PIP enrichment by AIEX (WAX HiTrap DEAE FF) followed by LC-ESI-HR-MS analysis using a reversed phase C_8 column and PIP enrichment by SEC (Superdex75) also followed by LC-ESI-HR-MS analysis (C_8-RP). Data of heterozygous carriers of rs73170678 and homozygotes for native PIP were analyzed separately. Tetraantennary structures were excluded in data evaluation because of their low intensity causing low quality isotope patterns. However, it was possible to identify tetraantennary structures also containing heavily fucosylated sub-structures such as $Hex_7HexNAc_6dHex_9$ or $Hex_7HexNAc_6dHex_5NeuAc_1$.

During this comparison, it is important to consider all glycoforms eluting during the LC-MS analysis. Time-separated inspection of MS spectra revealed a glycan-dependent elution of PIP on a reversed phase C_8 column. Manual inspection ($\Delta t = 0.3$ min) reveals a shift to higher masses as a function of time or rather hydrophobicity, since the content of acetonitrile is increasing during the LC run (cf. Figure 65).

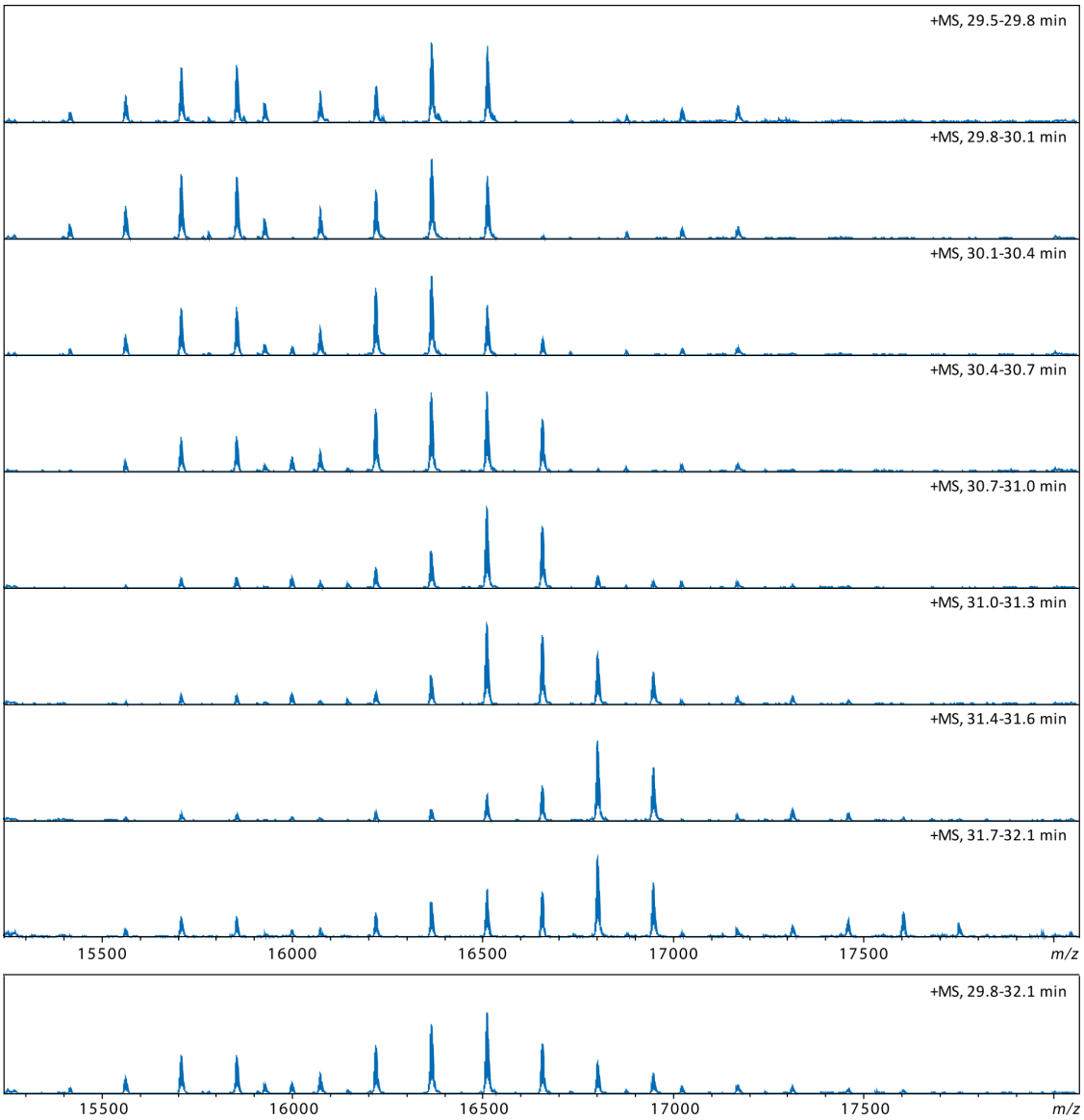


Figure 65: **Time-dependent elution of PIP glycoforms eluting on RP-C_8-LC. Saliva from one male (age: 27 years) test person was enriched by SEC and analyzed by RP-LC-MS (C_8-modified silica gel, reversed phase, elution by increasing concentration of acetonitrile). PIP elutes over a time period of approx. 2.5 minutes (bottom: sum spectrum over total elution time). Top: Separated MS spectra show that this long elution is due to the heterogeneous glycosylation pattern of PIP. Larger glycans elute time-shifted. More detailed insights in correlation of elution time and glycoform are given below.**

Coefficients (obtained using MATLAB) were interpreted with respect to the time-dependent elution behavior of glycans regarding the number of antennae, neuraminic acids and fucoses.

Elution of *N*-glycans is dependent on the number of antennae. Elution shifts to higher retention times for increasing size of the core structure (cf. Figure 66). While biantennae

show their elution maximum at time-point 2, absolute amounts of triantennary structures increase up to time-point 4 (left). Tetraantennary structures have the highest hydrophobic character and therefore elute in the end. However, triantennae represent the main component at all times (summed up over time: triantennae - 66%, biantennae - 23%, tetraantennae - 11%).

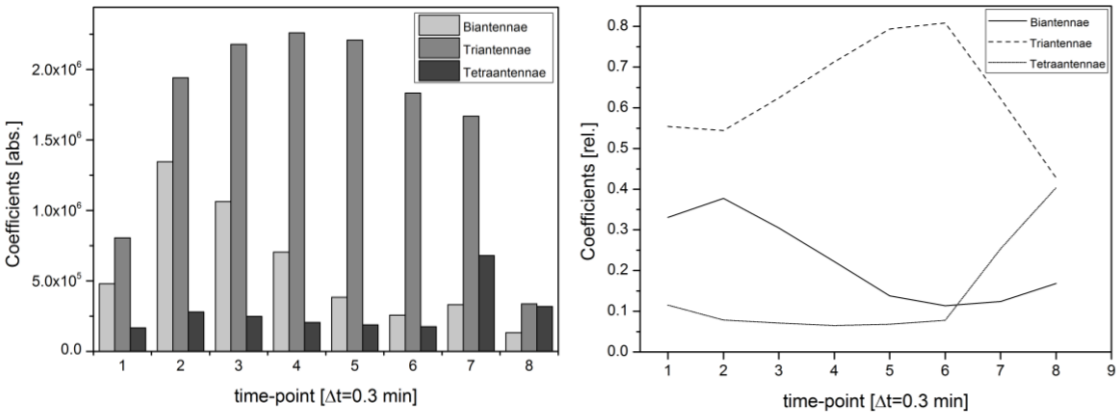


Figure 66: **Glycoforms of human salivary PIP separated by C_8-RP-LC as a function of (retention) time. PIP elutes over approx. 2.5 minutes. MS spectra of intact PIP were summed up in time periods of each 0.3 minutes. Spectra were interpreted as explained in text. Coefficients were summed up in groups to obtain the content of bi-, tri- and tetra-antennary glycans as a function of time. In total, triantennae represent the major group with 66%, followed by biantennae with an overall content of 23%. Tetraantennae represent the smallest group with 11%. The increasing number of antennae shifts elution to higher retention times (maximum of biantennae: time-point 2, maximum of triantennae (absolute/relative): time-point 4/6, maximum of tetraantennae (absolute/relative): time-point:7/8). Left: Absolute coefficients of different groups of antennae at each time-point. Right: Relative amounts of different groups of antennae at each time-point.**

The amounts of fucoses and sialic acids were separately analyzed for bi-, tri- and tetraantennary structures.

For biantennae, structures containing no neuraminic acids clearly dominate in terms of absolute glycan amount, followed by glycans that contain one neuraminic acid. Glycans with fully sialylated antennae (n = 2) represent a minor group (cf. Figure 67). However, the relative amount of structures occupied with multiple neuraminic acids considerably increases with elution time due to their enhanced hydrophobicity.

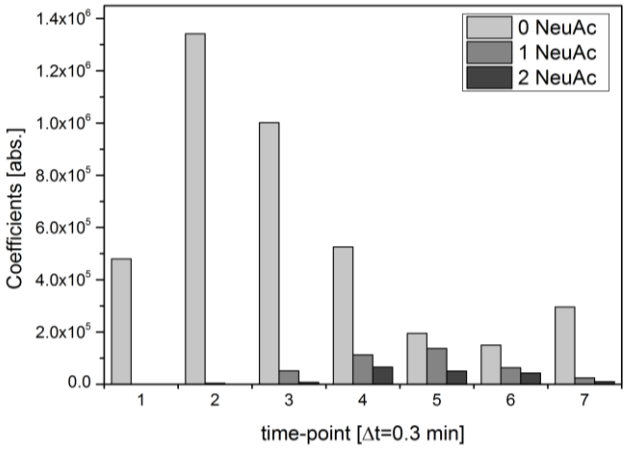


Figure 67: **Biantennary glycoforms of human salivary PIP eluting from RP column in terms of the content of terminal neuraminic acids. MS spectra were summed up over time (0.3 min per time-point). For each time-point all glycan amounts were summed up that contain no, one or two neuraminic acids, respectively. At all times, biantennary structures lacking neuraminic acids dominate. The majority of these structures elutes in the beginning. With higher retention times glycan structures containing one or two neuraminic acids arise. The shifted elution time is due to an increased hydrophobicity of glycans containing NeuAc residues.**

In the same manner, the degree of fucoses on biantennary structures attached to PIP was analyzed as a function of time in RP chromatography. The more fucoses a biantenna is holding, the more abundant is the structure (cf. Figure 68, left). Highly fucosylated structures elute first. Considering the relative amounts, the shift in elution time correlated with the degree of fucosylation becomes more obvious (cf. Figure 68, right). For example, the relative amount of glycoforms exhibiting only one or two fucose residues significantly increases as a function of time (equivalent to elevated hydrophobicity). Thus, fucose residues induce a more hydrophilic character in the individual glycoforms.

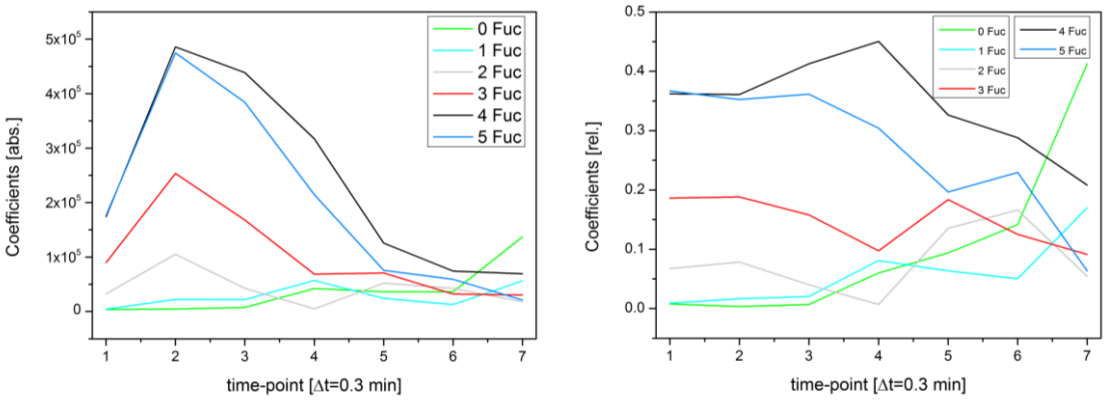


Figure 68: **Content of fucose residues attached to biantennary *N*-glycans on salivary PIP as a function of the retention time on a reversed phase LC column. MS spectra were summed up over time (each window 0.3 minutes). Highly fucosylated structures are most abundant and elute prior to less fucosylated ones. All biantennae without differentiation of their NeuAc content (n = 0-2) were summed up for this comparison. Left: absolute amounts, right: relative amounts per time-step.**

For the correlation between the number of attached fucoses and the time-shift in elution, the amount of neuraminic acids was neglected by summing up over all of their configurations. However, decreased level of fucosylation is associated with a gain in neuraminic acids. For example, a biantennary structure with five fucoses is unlikely to be occupied with neuraminic acids. Thus, the observed time-shift of elution of biantennae with a low degree of fucoses has to be individually monitored with respect to the amount of neuraminic acids.

Figure 69 shows the dependence of the degree of fucosylation on a biantennary structure without neuraminic acids and the elution time in absolute (left) and relative (right) amounts. The previously described correlation of degree of fucosylation and RP elution behavior appears again and verifies the hypothesis of dependency.

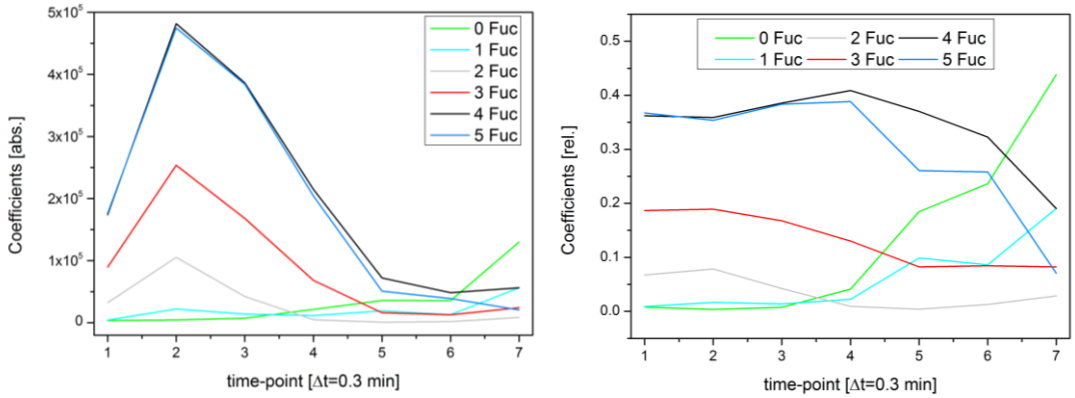


Figure 69: **Content of fucose residues attached to biantennary *N*-glycans lacking terminal neuraminic acids as a function of retention time in RP-LC. MS spectra were summed up over time (each window 0.3 minutes). Highly fucosylated structures are most abundant and elute prior to less fucosylated ones. Left: absolute amounts, right: relative amounts per time-step.**

In the same manner, the relation between degree of fucosylation and elution time was analyzed for biantennary structures with one (Figure 70) as well as two (Figure 92, appendix) neuraminic acids. All biantennae with one neuraminic acid elute at higher retention times. Together with the corresponding peaks of those structures without a neuraminic acid (see above), the broad elution profiles seen in Figure 68 (independent from NeuAc content) can be explained. Furthermore, the intensity of the structure fully occupied with one NeuAc and five Fuc residues is extremely low (dark blue, left). This may either be due to sterical hindrance of fucosyltransferases after a NeuAc residue has been attached during biosynthesis, or the attachment of a NeuAc unit acts as a stop signal for biosynthesis. Since no explicit order of glycosyltransferases is existent, this issue cannot be clarified here.

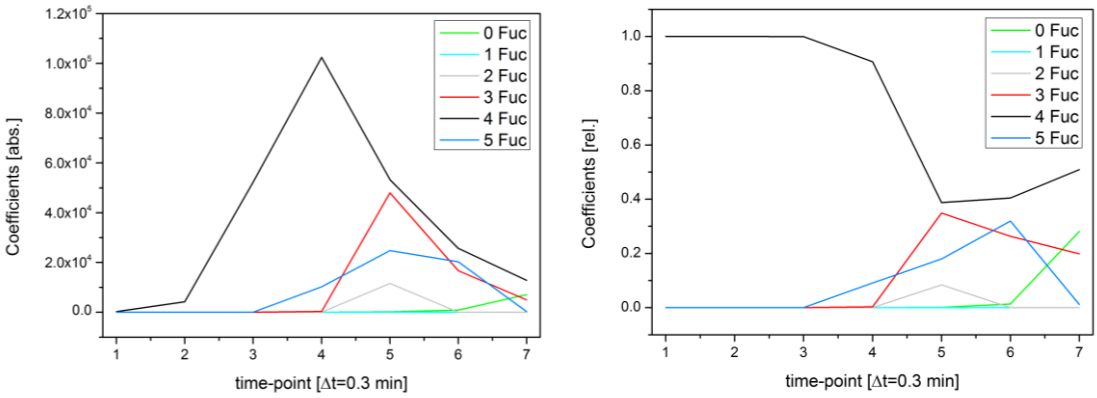


Figure 70: **Content of fucose residues attached to biantennary *N*-glycans with one terminal neuraminic acid as a function of the retention time in RP-LC. MS spectra were summed up over time (each 0.3 minutes). Highly fucosylated structures are most abundant and elute prior to less fucosylated biantennae. The level of fully fucosylated biantennae is drastically decreased compared to the same glycan without a terminal NeuAc. All glycans holding one NeuAc elute later compared to glycans lacking NeuAc due to an increased hydrophobicity. Left: absolute amounts, right: relative amounts.**

Because biantennary structures holding two NeuAc residues have reduced absolute amounts and therewith show worse signal quality, the tendencies described above are less reliable (cf. Figure 92 in appendix). However, the absence of biantennae with two NeuAc and four or five fucoses is obvious and verifies the assumption that one antenna unlikely holds two fucoses and one neuraminic acid, although this is not fully impossible (cf. Figure 70, blue: structure with five Fuc and one NeuAc residues).

The same findings were detected for tri- and tetraantennary structures. Independently from the content of fucoses, the elution profile of triantennary structures holding up to three NeuAc residues was studied (cf. Figure 71). Again, the influence of NeuAc residues leading to an increased hydrophobicity, expressed by a delayed retention time, is obvious. While the beginning of PIP elution is dominated by non-sialylated glycoforms, antennae fully occupied with NeuAc units are the major glycoforms in MS spectra at late retention times.

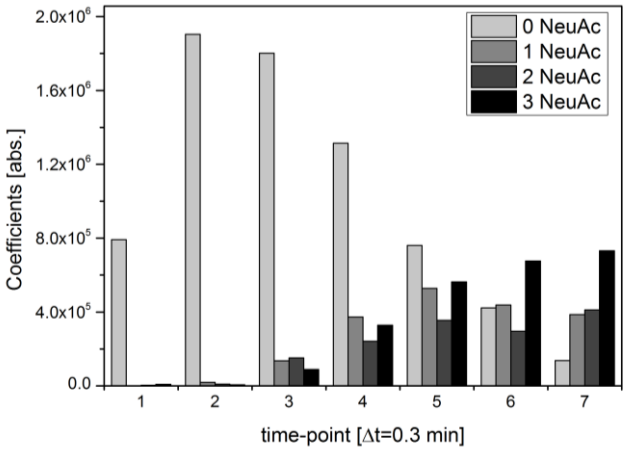


Figure 71: **Triantennary glycoforms of human salivary PIP eluting from RP column in terms of the number of terminal neuraminic acids. MS spectra were summed up over time (0.3 min per time-point). For each time-point all glycan amounts were summed up that contain a fixed number (of up to three) neuraminic acids, respectively. Triantennary structures lacking neuraminic acids dominate. The majority of these structures elutes in the beginning. With higher retention times glycan structures containing one to three neuraminic acids arise. The shifted elution time is due to an increased hydrophobicity of glycans containing NeuAc residues.**

The complementary effect, that is the fucosylation´s influence on elution time independent from the sialylation degree, was analyzed (cf. Figure 71). Highly fucosylated triantennae dominate the glycosylation pattern and elute prior to less fucosylated ones.

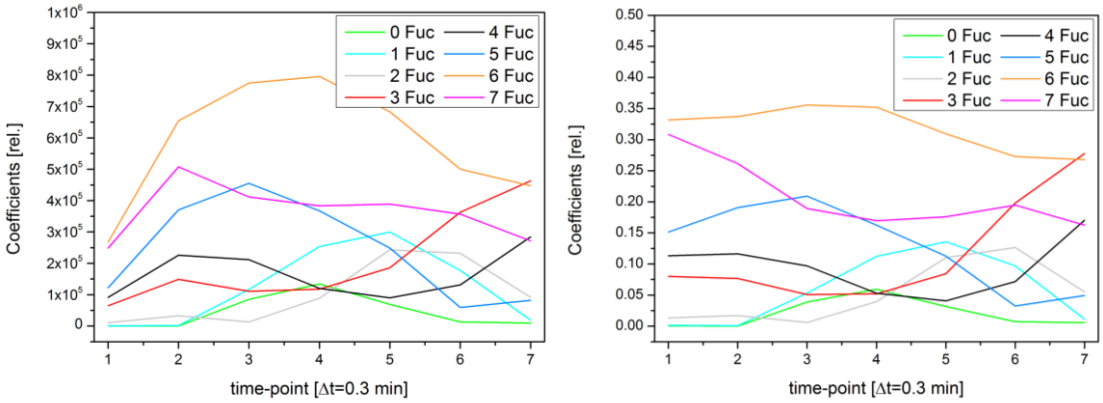


Figure 72: **Content of fucose residues attached to triantennary *N*-glycans on salivary PIP as a function of the retention time on a reversed phase-LC column. MS spectra were summed up over time (each data point: 0.3 minutes). Highly fucosylated structures are most abundant and elute prior to less fucosylated ones. All triantennae without differentiation of the NeuAc content (n = 0-3) were summed up for this comparison. Left: absolute amounts, right: relative amounts.**

The same behavior was detected for triantennae holding no *N*-acetyl neuraminic acid (Figure 73, right, relative amounts), although it is less distinct due to the high prevalence of highly fucosylated structures. Indeed, the fully fucosylated structure elutes in high amounts over all times. A long elution time could be explained by a high abundance or a separated elution of glycan isoforms.

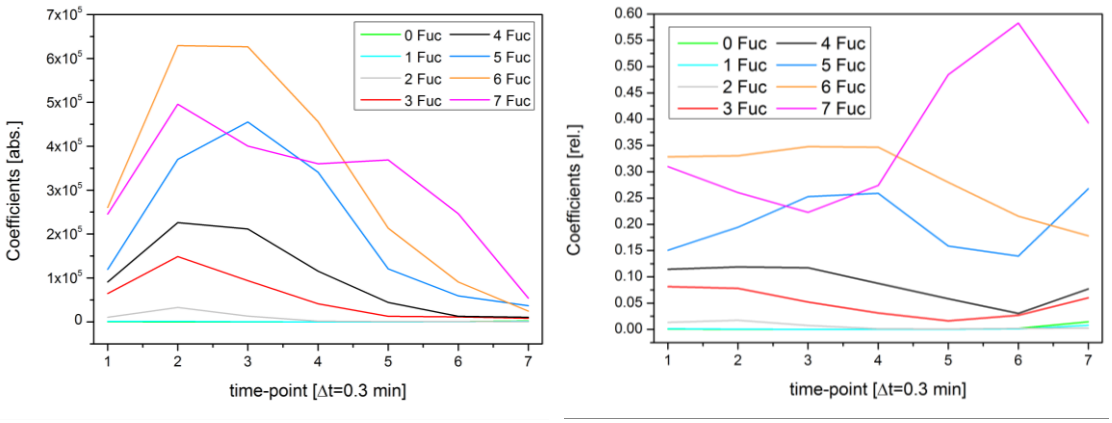


Figure 73: **Content of fucose residues attached to triantennary *N*-glycans lacking terminal neuraminic acids as a function of the retention time in RP-LC. MS spectra were summed up over time (each 0.3 minutes). Highly fucosylated structures are most abundant and elute prior to less fucosylated ones. Left: absolute amounts, right: relative amounts.**

As expected, for triantennae holding one neuraminic acid, the amount of structures with seven fucoses (full occupation) is drastically reduced (cf. Figure 74). Furthermore, the elution time is again shifted, originating from the hydrophobicity-enhancing character of the acetyl neuraminic acid.

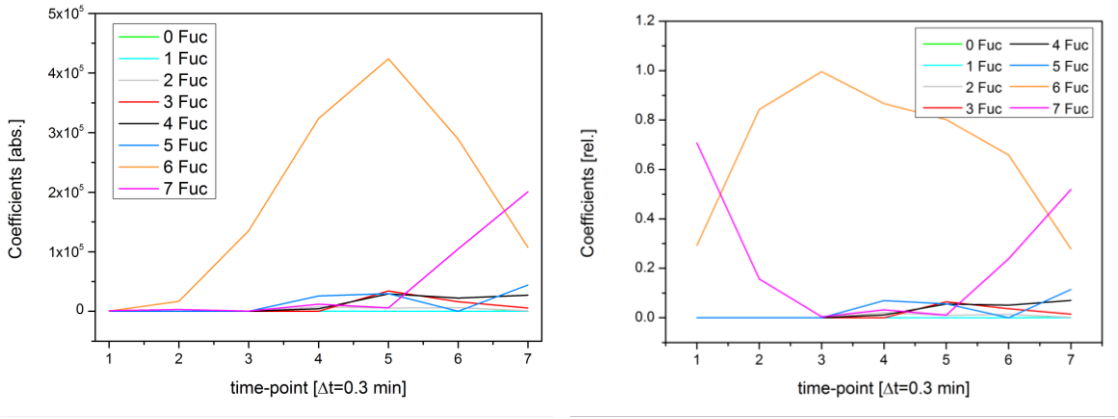


Figure 74: **Content of fucose residues attached to triantennary *N*-glycans with one terminal neuraminic acid as a function of the retention time in RP-LC. MS spectra were summed up over time (each 0.3 minutes). Highly fucosylated structures are most abundant and elute prior to less fucosylated ones. As expected, the level of fully fucosylated triantennae is drastically decreased compared to the same glycan without a terminal NeuAc. All glycans holding one NeuAc elute later compared to their counterparts lacking NeuAc due to an increased hydrophobicity. Left: absolute amounts, right: relative amounts.**

For triantennae that hold two or three *N*-acetyl neuraminic acids, the tendencies get more ambiguous (two NeuAc: Figure 93/appendix, three NeuAc: Figure 94/appendix). Potentially, this is due to the significantly enhanced occupation possibility of Fuc and NeuAc residues on each antenna, thus an increased number of possible isomers. For example, the triantennary *N*-glycan holding two NeuAc and four Fuc residues has a high number of possibilities already for the location of the Fuc residues alone. Assuming, that one Fuc residue is attached to the core, six monosaccharides (three Gal and three GlcNAc units) are available for binding of the three fucoses left. Furthermore, the NeuAc residues can be bound to different arms (6-Man, 3-Man arm). Either the 6-Man or the 3-Man arm can be occupied with one respectively two antennae, and the core fucosylation is not mandatory although likely. This high number for possibilities of Fuc and NeuAc distribution is presumably causing a variety of isobaric *N*-glycans that elute shifted in RP-LC.

Nonetheless, the decreased level of highly fucosylated triantennae in case of an elevated number of NeuAc units (n = 2-3) is clearly obvious. While structures with up to five Fuc residues dominate in case of two attached NeuAc, the fucose level is reduced to zero to

four in case of three NeuAc units. This is plausible, if considering that the occupation with two fucoses is unlikely if a NeuAc is attached to the terminal galactose.

As for bi- and triantennary *N*-glycans, tetraantennary core structures were analyzed regarding their fucose and neuraminic acid content as a function of elution time. All data have to be treated with care, since the tetraantennary structures are present only in minor quantity, which caused low quality (low S/N ratio and intensity) MS spectra. Graphical diagrams can be found in the appendix (section 8). However, the same trends described above are also detected for tetraantennae. Highly fucosylated structures elute prior to structures attached with fewer fucoses, and covalently bound neuraminic acids shift the elution time to higher retention times.

Because of these findings, it is important to sum up all MS spectra that contain PIP signals prior to deconvolution, if a comparative study between several test persons or groups is intended.

5.3.4.1 Comparison of PIP Glycoforms by 2D-LC-MS Analysis

Comparative analyses were performed by two chromatographic enrichment strategies: 1) AIEX followed by RP-LC and 2) SEC followed by RP-LC. Enrichment by AIEX or SEC resulted in very similar data with only minor variation. However, because of the glycoform-depending elution in AIEX (cf. 5.3.2.1), predominantly SEC data are shown and discussed in detail. All further data (especially AIEX data) that are not shown in this chapter can be found in the appendix.

Salivary PIP was enriched by SEC and analyzed by RP-LC-MS. Here, saliva from twelve individuals was used for data analysis. Six individuals were heterozygous for the SNP ^{79}Q→^{79}H and showed doubled signals. All saliva samples used in the data analysis were treated identically. However, samples that were analyzed slightly differently (different saliva volume, different SEC column or different enrichment strategy) lead to similar results. HR-MS spectra of PIP were analyzed with MATLAB (cf. section 5.1.2). PIP glycoforms were quantified by fitting the theoretical isotopic pattern to the deconvolved

experimental spectrum. The isotopic pattern of the following glycoforms was used for the fit: bi- and triantennary *N*-glycans with all possible numbers of fucoses (Fuc) and *N*-acetyl neuraminic acids (NeuAc), respectively (0-5 Fuc and 0-2 NeuAc for biantennary structures, 0-7 Fuc and 0-3 NeuAc for triantennary structures), thus 50 different glycoforms were used for data analysis (cf. Table 24 in appendix, combination 1-50). Data analysis was performed separately for PIP from female and male saliva and it was differentiated between individuals that are homozygous or heterozygous for the SNP resulting in the amino acid exchange ^{79}Gln→^{79}His. See Table 23 in appendix for donor information.

For homozygous individuals (without a SNP) the ratio of biantennae to triantennae is similar for both sexes (cf. Figure 75).

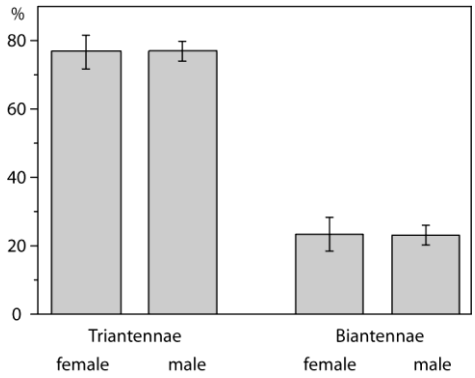


Figure 75: Ratio of bi- and triantennary *N*-glycans in human salivary PIP. PIP was enriched from six individuals (female: 3, male: 3) by SEC and analyzed by LC-MS. Female triantennae: (76.6±3.5)%, female biantennae: (23.4±3.5)%, male triantennae: (78.0±1.6)%, male biantennae: (22.0±1.6)%.

For the following analyses, bi- and triantennary structures were evaluated separately and thus normalized within each group. Biantennary structures can hold up to two terminal neuraminic acids. Furthermore, every variant of these can be modified with up to five fucose residues. Figure 76 shows the distribution of attached Fuc residues for biantennary structures for 0, 1 and 2 NeuAc units in PIP from saliva from male individuals. For non-sialylated structures a full occupation of antennae and the proximal GlcNAc occurs most frequently. However, less fucosylated structures occur as well, in fact the number of fucoses decreases along with their abundancy. Data for biantennae holding one or two

fucoses have to be considered with care, since their intensity is drastically reduced (biantennae are only present about four times less than triantennae). Nevertheless, the tendency of increasing sialylation is inversely correlated with PIP glycan fucosylation: The more NeuAc units are present, the less fucosylation occurs. The possibility that fucosyltransferases are hindered by terminal sialic acids or terminal sialic acids work as a stop signal for biosynthesis were briefly discussed above.

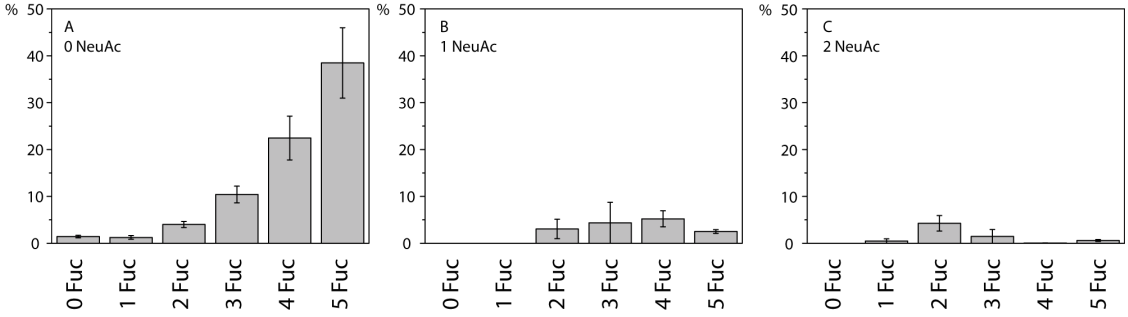


Figure 76: **Distribution of biantennary *N*-glycans from human PIP (three males). PIP was enriched by SEC from saliva and analyzed by LC-MS and a least square optimized fit of the obtained experimental isotope pattern of each glycoform signal. The obtained coefficients were sub-divided regarding their degree of terminal sialylation. A: For biantennae lacking terminal NeuAc, structures fully occupied with Fuc residues (n = 5) dominate followed by structures with decreasing numbers of Fuc residues. B: As soon as one NeuAc unit is attached to the biantenna, the major glycoform changes to the structure with four Fuc residues. Furthermore, the distribution shifts to less fucosylated structures, probably due to sterical hindrance for fucosyltransferases after sialylation. C: The tendency obtained for mono-sialylated biantennae extents to bi-sialylated structures. The fucosylation degree significantly decreases. However, due to low intensity and low S/N ratio in MS spectra, data for these structures have to be handled with care.**

The same conclusion of inverse dependency between sialylation and fucosylation can be obtained, when considering the sialylation degree as a function of fucose residues. Figure 77 summarizes the amount of 0, 1 and 2 NeuAc residues attached to biantennary glycans of PIP from saliva from male individuals with varying quantities of attached fucoses from 0 to 5. It is obvious, that the amount of NeuAc residues increases with simultaneous decrease of fucose residues. Since biantennae with no (F) or one (E) Fuc residue are rarely present, the corresponding bar diagrams should not be overinterpreted.

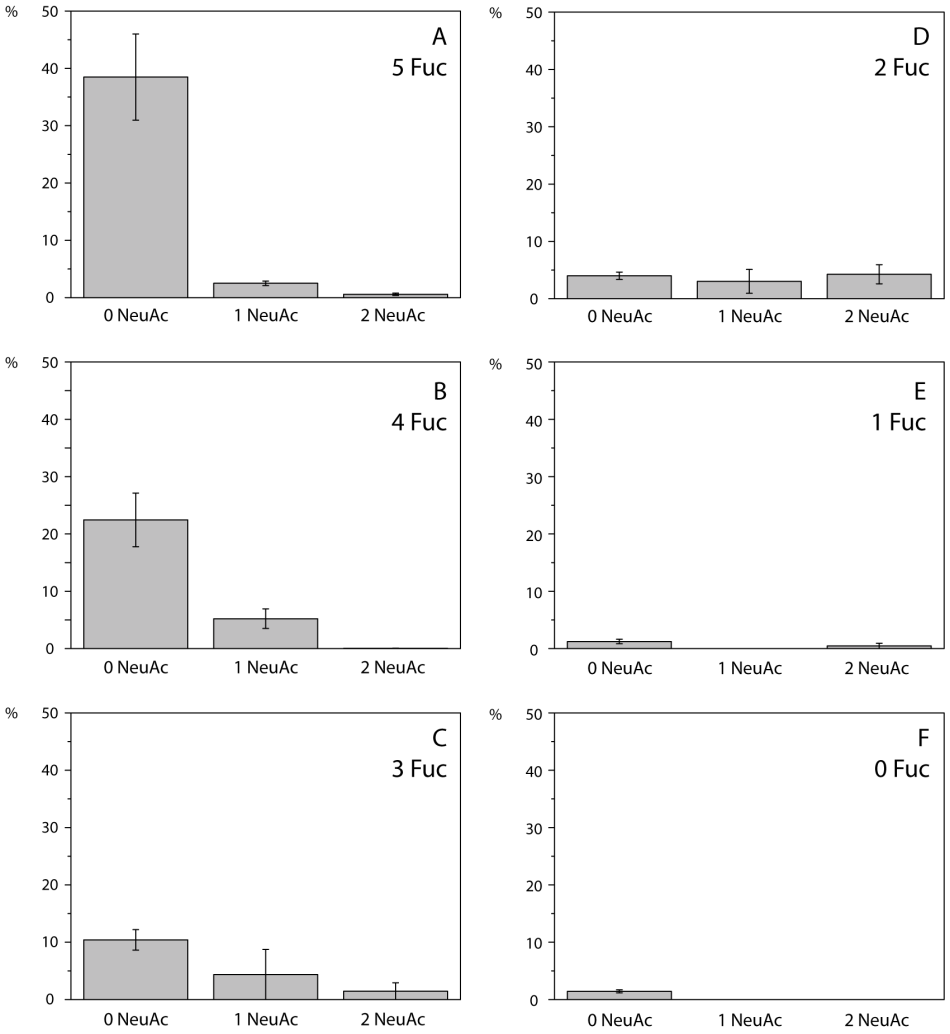


Figure 77: **Distribution of biantennary *N*-glycans from human PIP (three males). PIP was enriched by SEC from saliva and analyzed by LC-MS and mathematical back-calculation of the obtained experimental isotope pattern of each glycoform signal. The obtained coefficients were sub-divided with respect to their degree of fucosylation. A: 5 fucose (Fuc) residues, B: 4 Fuc residues, C: 3 Fuc residues, D: 2 Fuc residues, E: 1 Fuc residue, F: 0 Fuc residue. A to D: A correlation between fucosylation degree and sialylation is clearly observable. In case of a fully fucosylated *N*-glycan, sialylation is rarely detectable (A). The less fucoses are attached to the biantennae, the more NeuAc units get attached to the glycan. E+F: Due to very low intensity and low S/N ratio in MS spectra, data for these structures should be excluded from data analysis.**

The same tendencies were detected for triantennary structures. While a full occupation with fucoses is present for triantennae lacking NeuAc residues, fucosylation decreases as a function of increasing amounts of NeuAc residues (cf. Figure 78 and Figure 79). However, attachment of fucoses does not decrease one to one with increasing NeuAc units, but to a greater extent. This suggests the hypothesis that sialylation either sterically hinders further biosynthesis by fucosyltransferases or acts as a biosynthesis stop signal.

An antenna holding one NeuAc residue is likely to show only one attached Fuc residue (surprisingly, also bi-fucosylated antennae with one NeuAc unit occur).

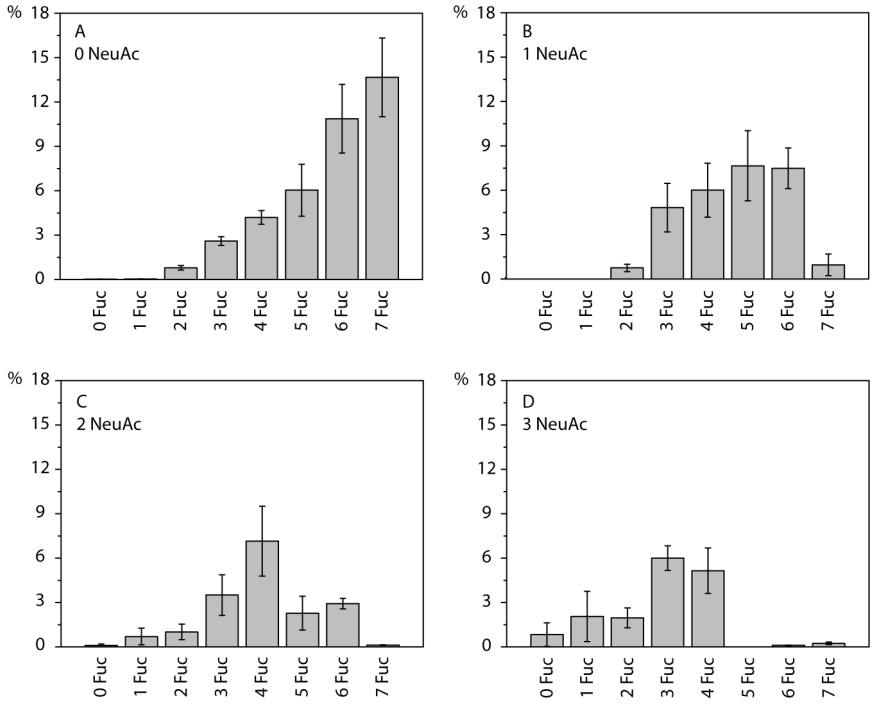


Figure 78: **Distribution of triantennary *N*-glycans from human PIP (three males). PIP was enriched by SEC from saliva and analyzed by LC-MS and mathematical back-calculation of the obtained experimental isotope pattern of each glycoform signal. The obtained coefficients were sub-divided with respect to their degree of terminal sialylation. A: For triantennae lacking terminal NeuAc structures, full occupation with Fuc residues (n = 7) dominates, followed by structures with decreasing number of Fuc residues. B: As soon as one NeuAc unit is attached to the biantenna, the major glycoform changes to structures with one to two Fuc residues less. The distribution shifts to less fucosylated structures, probably due to steric hindrance for fucosyltransferases after sialylation. C: The tendency obtained for mono-sialylated triantennae extends to bi-sialylated structures. The major glycoform is now constituted by the structure bearing one fucose residue per antenna averaged (plus core fucosylation). D: For a full occupation of antennae by terminal NeuAc residues, the fucosylation degree significantly decreases.**

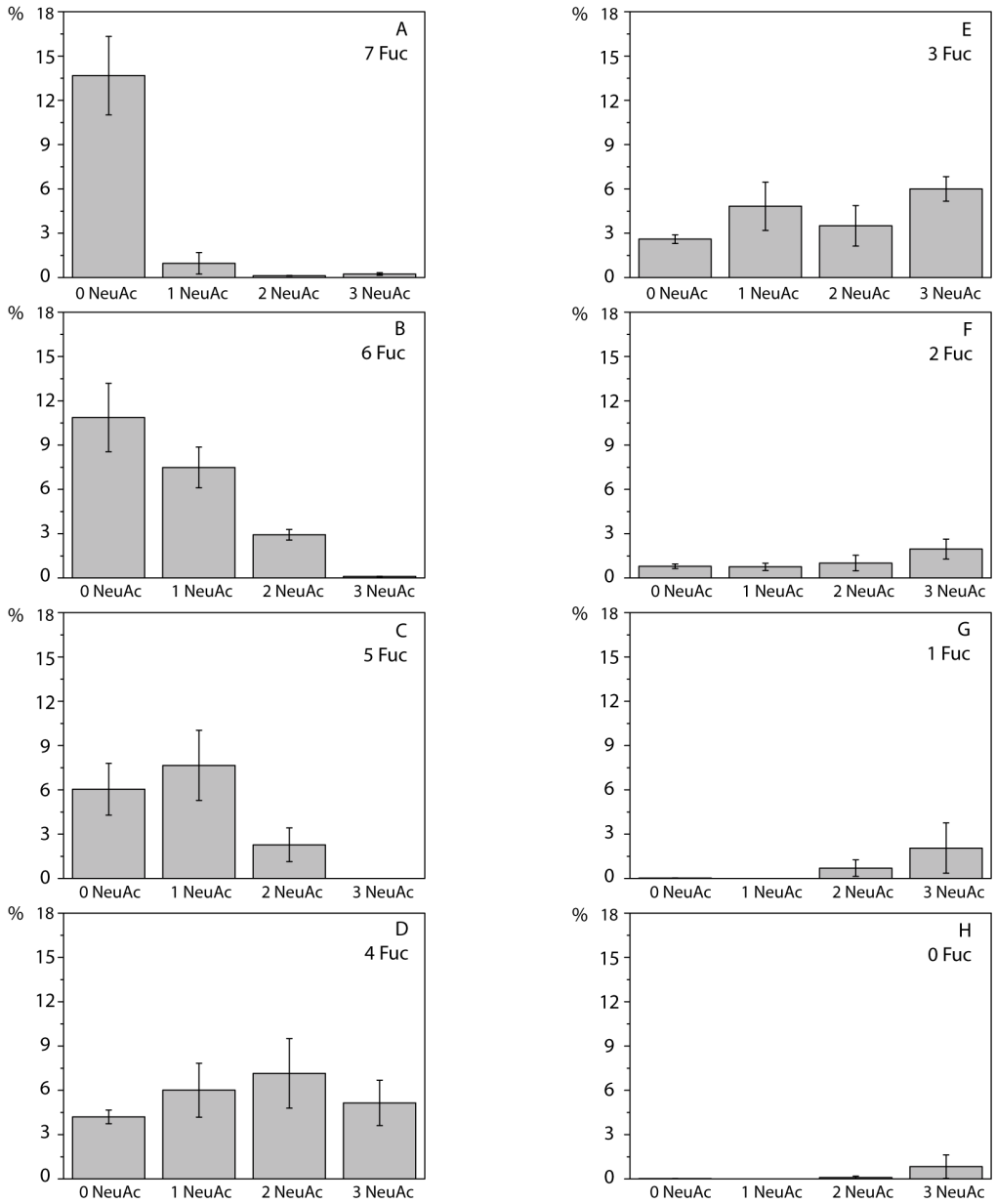


Figure 79: **Distribution of triantennary *N*-glycans from human PIP (three males). PIP was enriched by SEC from saliva and analyzed by LC-MS and mathematical back-calculation of the obtained experimental isotope pattern of each glycoform signal. The obtained coefficients were sub-divided with respect to their content of fucosylation. A: 7 fucose (Fuc) residues, B: 6 Fuc residues, C: 5 Fuc residues, D: 4 Fuc residues, E: 3 Fuc residues, F: 2 Fuc residues, G: 1 Fuc residue, H: 0 Fuc residue. A correlation between fucosylation degree and sialylation is clearly observable for all fucosylation states. In case of a fully fucosylated *N*-glycan, sialylation is rarely detectable (A). The less fucoses are attached to the biantennae, the more NeuAc units get attached to them. The less fucoses are present, the more the full sialylation dominates.**

The same data analyses were performed for test persons with the SNP causing an amino acid exchange (^{79}Gln→^{79}His). Therefore, the MATLAB script for data analysis had to be

extended regarding the number and the proportion of proteins carrying the glycosylation pattern. Instead of vectors holding the isotope pattern of one glycan composition plus one protein, vectors were calculated that are composed of one glycan composition plus the isotope pattern of the combination of two proteins (namely native PIP and PIP with one His residue instead of one Gln residue). Additionally, the possibility was included to set a ratio between the two proteins, since it was discovered that the proteins were often not expressed in equal amounts (cf. section 5.3.3).

Similar qualitative and quantitative results were obtained for PIP showing a nonsynonymous SNP. However, the amino acid exchange may have an influence on the glycosylation pattern. The comparison between the amounts of biantennae and triantennae is shown in Figure 80. Again, there is no distinct gender specificity. Also, in saliva from female donors the amounts of bi- and triantennae are similar to PIP without an amino acid exchange. The values for triantennae are slightly reduced in saliva from men and the amounts of biantennae correspondingly increased.

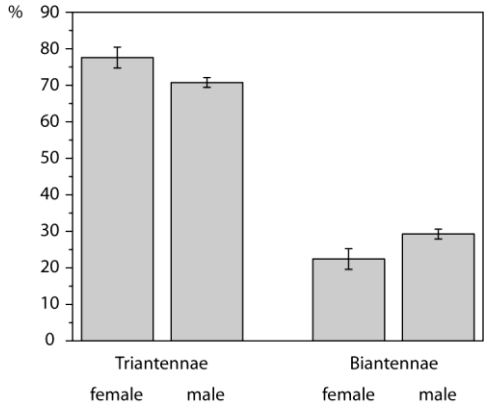


Figure 80: **Ratio of bi- and triantennary *N*-glycans in salivary PIP from test persons (female: 2, male: 4) holding a SNP (Gln→His). PIP was enriched by SEC and analyzed by LC-MS. MS spectra were analyzed by mathematical fitting of experimental spectra with theoretically calculated spectra of PIP with bi- and triantennae holding 0-5 Fuc and 0-2 NeuAc residues for biantennae and 0-7 Fuc and 0-3 NeuAc units for triantennae, respectively. Female triantennae: (77.6±2.8)%, female biantennae: (22.4±.2.8x)%, male triantennae: (70.8±1.3)%, male biantennae: (29.2±1.3)%.**

More distinct differences can be elucidated by considering the antennae separately and in more detail. The distribution of NeuAc residues on biantennae differs from PIP without an amino acid exchange, nonetheless, the tendency of an inverse ratio of

fucosylation and sialylation can still be observed. However, the portion of mono-sialylated biantennae is enhanced in case of an amino acid exchange. Furthermore, the amount of biantennae holding less than five Fuc residues (which would be the full occupation) is enhanced in non-sialylated species. Thus, in case of an amino acid exchange, overall fucosylation seems to be reduced while the overall sialylation seems to be slightly enhanced compared to PIP without the exchange.

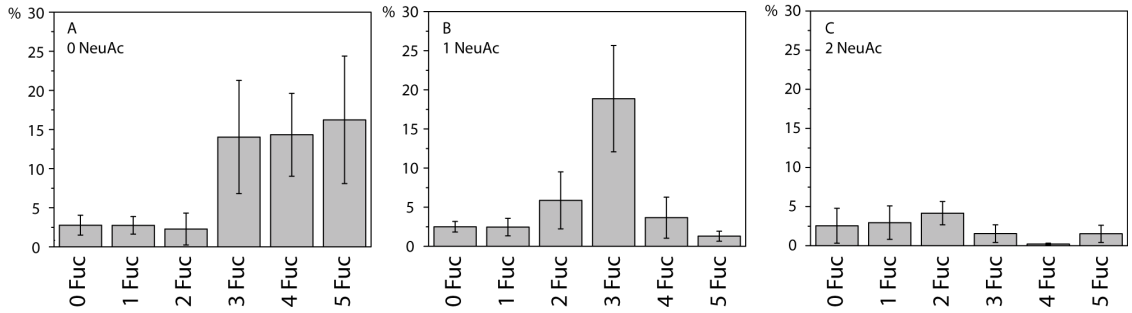


Figure 81: **Fucosylation distribution of biantennary *N*-glycans from PIP isolated from male individuals (n = 4) holding a SNP that causes an amino acid exchange (Gln→His). Sample material and data were obtained as described for individuals showing no SNP. While the tendency of decreasing amounts of fucosylation for increasing sialylation stays the same, a slightly decreased amount of fucosylation can be observed for all structures compared to PIP from individuals without a SNP. Furthermore, the overall amount of biantennae holding a NeuAc residue is significantly increased. A: 0 NeuAc, B: 1 NeuAc, C: 2 NeuAc.**

The same findings were obtained when considering the amount of NeuAc residues on biantennae as a function of fucosylation degree (cf. Figure 82). The more fucose residues are attached to the core structure, the less NeuAc units can be found. In case of an amino acid exchange, the portion of biantennae holding zero or one fucose residue is enhanced (compared to no amino acid exchange), so that the trend of an enhanced ratio of NeuAc residues can be detected also for mono- or non-fucosylated biantennae.

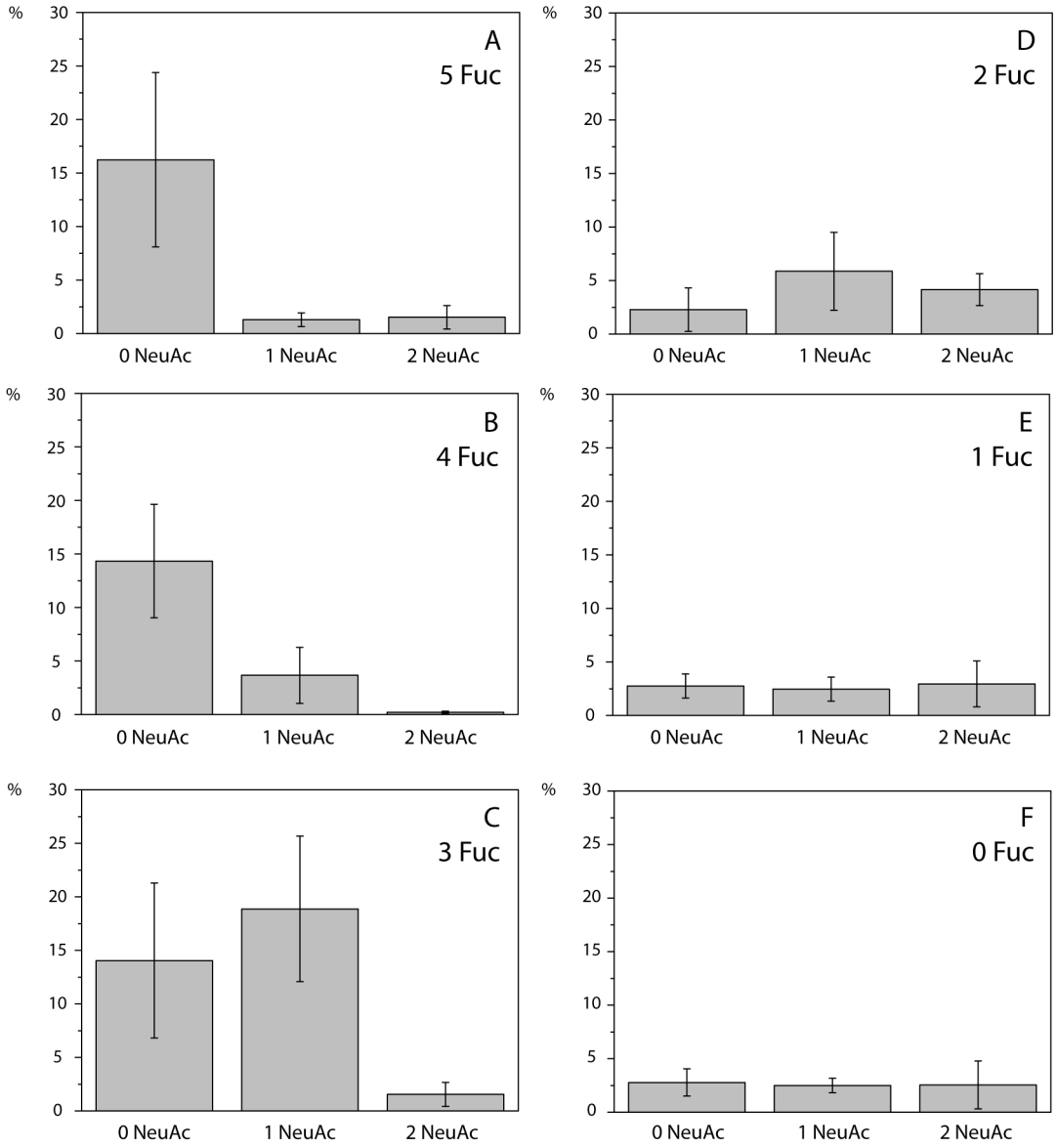


Figure 82: **Sialylation distribution of biantennary *N*-glycans from PIP isolated from male individuals (n = 4) holding a SNP that causes an amino acid exchange (Gln→His). Sample material and data were obtained as described for individuals showing no SNP. Similar to individuals holding no SNP, the sialylation degree increases for decreasing fucosylation degree. Indeed, the overall amount of fucosylation on biantennae is reduced for PIP holding the amino acid exchange. Simultaneously the amount of NeuAc residues increases. A: 5 Fuc, B: 4 Fuc, C: 3 Fuc, D: 2 Fuc, E: 1 Fuc, F: 0 Fuc.**

Also for triantennae, the number of Fuc residues is slightly reduced, while the amount of core structures with NeuAc residues is enhanced (cf. Figure 83 and Figure 84).

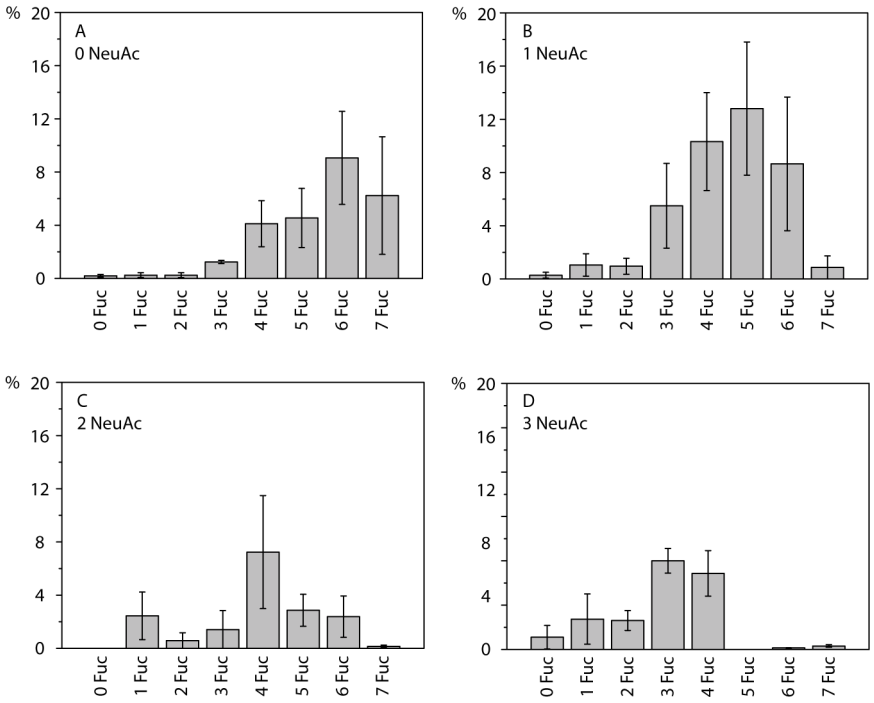


Figure 83: **Fucosylation distribution of triantennary *N*-glycans from PIP isolated from male individuals (n = 4) holding the SNP that causes the amino acid exchange (Gln→His). Sample material and data were obtained as described for individuals showing no SNP. While the tendency of decreasing amounts of fucosylation for increasing sialylation stays the same, a slightly decreased amount of fucosylation can be observed for all structures compared to PIP from individuals without a SNP. Furthermore, the overall amount of triantennae holding a NeuAc residue is significantly increased. A: 0 NeuAc, B: 1 NeuAc, C: 2 NeuAc, D: 3 NeuAc.**

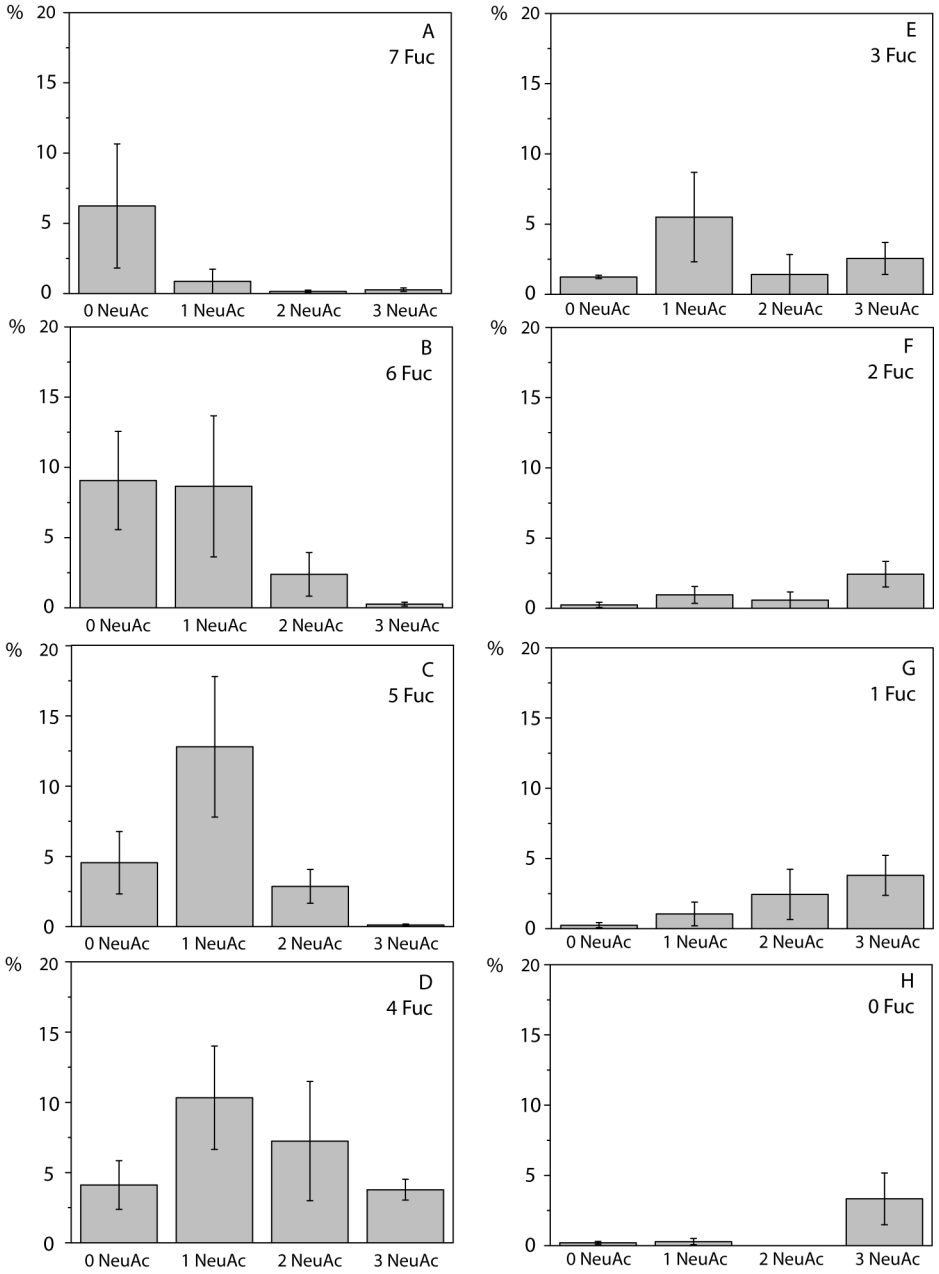


Figure 84: **Sialylation distribution of triantennary *N*-glycans from PIP isolated from male individuals (n = 4) holding the SNP that causes the amino acid exchange (Gln→His). Sample material and data were obtained as described for individuals showing no SNP. Very similar to individuals holding no such SNP, the sialylation degree increases for decreasing fucosylation degree. Indeed, the overall amount of fucosylation on biantennae is reduced for PIP holding the amino acid exchange. Simultaneously the amount of NeuAc residues increases. A: 7 Fuc, B: 6 Fuc, C: 5 Fuc, D: 4 Fuc, E: 3 Fuc, F: 2 Fuc, G: 1 Fuc, H: 0 Fuc.**

5.3.5 *N*-Glycosylation of Salivary PIP during Pregnancy

To review a possible glycosylation change of PIP during pregnancy, PIP isolated from 10 mL saliva donated by a pregnant woman (pregnancy week: 22, second trimester) was analyzed. As described in literature, the concentration of PIP is enhanced during pregnancy. As a consequence, spectra of tetraantennary *N*-glycans could successfully be obtained and quantified. Considering only bi- and triantennae for data analyses resulted in a ratio of 3:1 between them and is in agreement with saliva samples from non-pregnant women.

The glycosylation pattern is very similar to the pattern of non-pregnant women, only minor differences were observed. An overall high degree of fucosylation is present and the content of fucoses and NeuAc residues shows the same inverse behavior as was described above (cf. Figure 85).

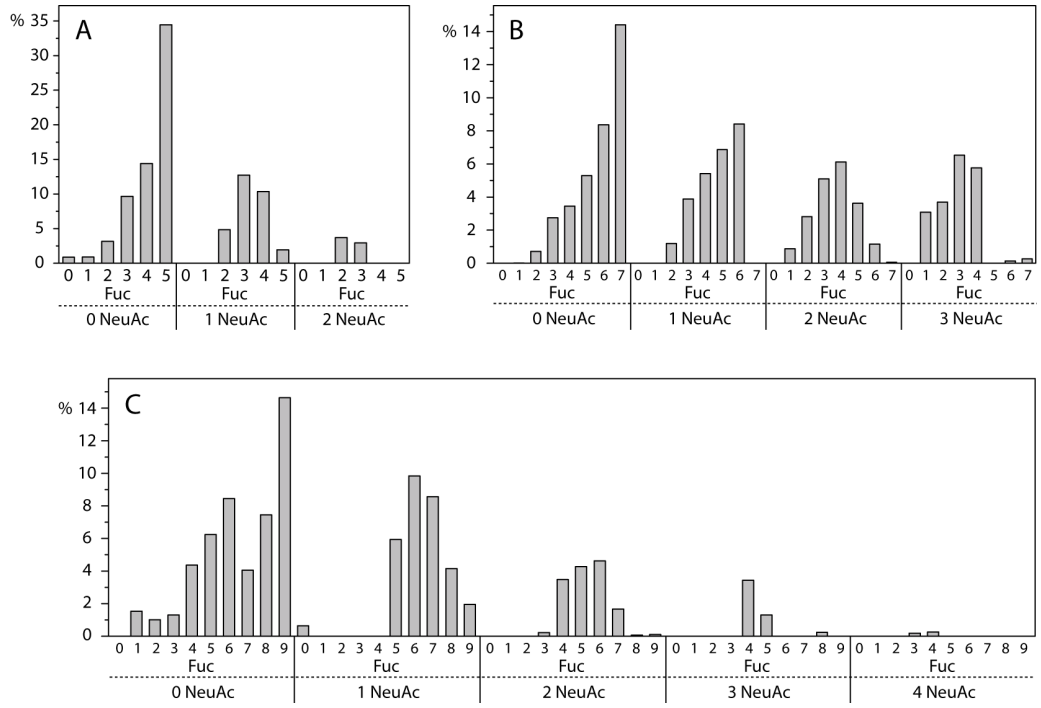


Figure 85: **Amount of fucose und NeuAc residues attached to bi-, tri- and tetraantennary structures of PIP isolated from saliva of a pregnant woman. As for nonpregnant individuals, a high fucosylation is observable, as is the fact that fucosylation is inversely correlated to the degree of sialylation. The overall content of non-sialylated bi- and triantennae is reduced, though, whereas mono-sialylated biantennae and tri-sialylated triantennae are more abundant. A: Biantennae, B: triantennae, C: tetraantennae.**

For tetraantennae highly fucosylated structures (nine fucoses per core structure) are present. The overall content of non-sialylated structures (bi- and triantennae) is reduced compared to PIP from non-pregnant women (cf. Figure 86). Instead, in biantennae the amount of structures holding one neuraminic acid is enhanced, in triantennae the amount of trifucosylated structures is enhanced (cf. Figure 87). A larger test group has to be analyzed to elucidate if the findings are specific and significant for pregnant women.

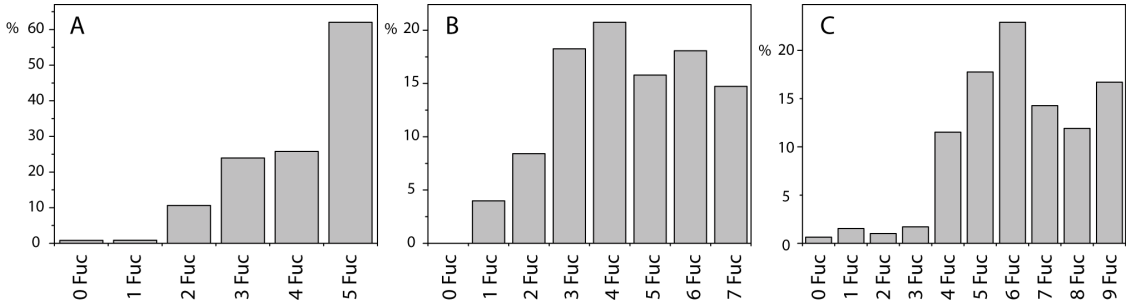


Figure 86: **Fucosylation degree of salivary PIP from a pregnant test person. A: For biantennae, mainly fully-fucosylated antennae occur (> 60%). B: The distribution of degree of fucosylation for triantennae is broader and especially spread over the range of four to seven fucoses. C: Tetraantennae hold up to nine fucose residues, that is a full occupation of the antennae and the proximal GlcNAc again, while six attached fucose residues dominate.**

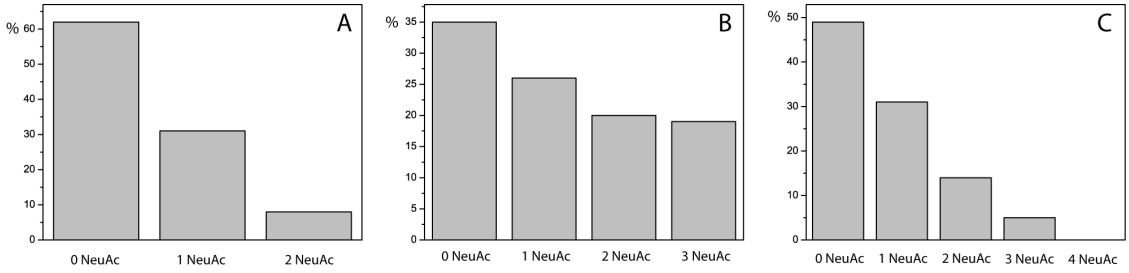


Figure 87: **Sialylation degree of salivary PIP from a pregnant test person. Although the content of non-sialylated antennae is reduced compared to non-pregnant individuals, they are still most abundant. Thus, especially for triantennae, but also for biantennae, sialylated structures are present in a higher amount. A: Biantennae, B: triantennae, C: tetraantennae.**

5.3.6 Comparative Study of *N*-Glycosylation of Salivary PIP between Breast Cancer Patients and Healthy Donors

To elucidate the use of salivary PIP as a biomarker in diagnostics of breast cancer, a pilot study was performed. Five female patients suffering from recently diagnosed breast

cancer volunteered for a donation of 10 mL saliva. Saliva was treated as described above. PIP was enriched by SEC and analyzed by RP-LC-MS. Spectra containing PIP were summed up and analyzed by automatic data interpretation (isotope pattern, MATLAB). In the sample from one test person, no PIP was detectable.

The overall degree of fucosylation is partly similar, but shows less consistent trends as have been depicted for healthy individuals. Due to the irregular fucosylation, the comparison was restricted to the numbers of antennae and the distribution of sialylation. Coefficients can be found in Table 33 in the appendix.

The distribution of bi- ((26±6)%) and triantennae ((73±6)%) in total is very similar compared to the antennae distribution of healthy test persons (cf. 5.3.4.1). Table 12 summarizes the amount of bi- and triantennae regarding their degree of sialylation. The degree of sialylation of triantennae is in agreement with data obtained from the analysis of healthy individuals. For biantennae a changed degree of sialylation is noticeable. While the amount of non-sialylated structures is significantly reduced under disease conditions, amounts of mono- and bisialylated structures are enhanced (cf. Figure 88). Further samples from breast cancer patients are needed to determine, if these changes could lead to the development of a clinically useful biomarker for breast cancer detection. Also for significant conclusions about fucosylation behavior, more samples have to be evaluated.

Table 12: Sialylation degree on bi- and triantennary *N*-glycans attached to salivary PIP from breast cancer patients (n = 4).

n (NeuAc)	Biantennae, normalized [%]	Triantennae, normalized [%]
0	46±5	41±3
1	40±7	29±2
2	14±4	18±1
3	-	11±3

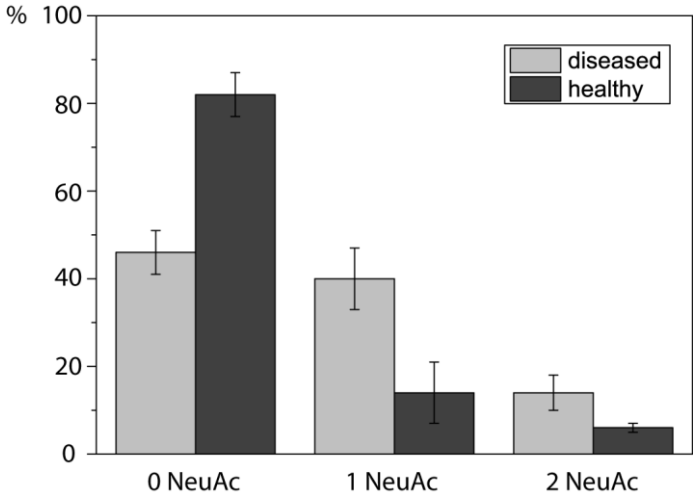


Figure 88: **Sialylation degree of biantennae of salivary PIP. Opposed to each other are values obtained for healthy individuals (n = 5, female) and breast cancer patients (n = 4, female). The amount of non-sialylated biantennae is drastically reduced within salivary PIP from patient samples. Contrary, the amount of mono- and bi-sialylated structures is enhanced.**

6 EXPERIMENTAL PROCEDURES

6.1 Instruments, Material and Software

Table 13: Instruments / Material

Software Brand Name / Version	Manufacturer
ESI-q/TOF „maXis“	Bruker Daltonics
UHPLC „Dionex 3000“	Dionex (now: Thermo scientific)
MALDI TOF-TOF „UltrafleXtreme“	Bruker Daltonics
HPLC “L7000”	Merck Hitachi (now: Merck)
HPLC “1200 Series”	Agilent Technologies
FPLC “ÄKTA Purifier P-901”	GE Healthcare
700 MHz NMR spectrometer with cryo probe	Bruker
Centrifuge “5804R” and “5417R”	eppendorf
Scale “AE21”	Mettler
Centrifugal Filters “Amicon Ultra 0.5 mL”	Merck Millipore
AIEX column “HiTrap DEAE FF”	GE Healthcare
AIEX column “HiTrap Q FF”	GE Healthcare
AIEX material “DEAE A-50”	GE Healthcare
RP column (C_8) “Aeris WIDEPORE”, 250×2.1 mm	phenomenex
RP column (C_8) “Aeris WIDEPORE”, 150×2.1 mm	phenomenex
RP column (C_{18}) “Kinetex”, 150×2.1 mm	phenomenex
SEC column “Superdex75” and “Superdex200”, 10×300 mm	GE Healthcare
SEC column “Superdex200”, 16×600 mm	GE Healthcare
Lyophilizer	CHRIST
NanoDrop	Thermo Scientific
SpeedVac “SPD121P”	Savant
Super Fraction Collector, SF-3120	Advantec
Thermomixer “comfort”	eppendorf
Vortexer “REAX top”	Heidolph
PGC column “Hypercarb”, 2.1×15 mm	Thermo Scientific
DLS instrument “spectroLITE 300”	Xtal-Concepts

Table 14: Software

Software Brand Name / Version	Distributor
Matlab R2011b, R2012a	MathWorks
OriginPro 8.5.0G SR1	OriginLab Corporation
Microsoft Excel 2010	Microsoft Corporation
GlycoWorkbench 2.0	A. Ceroni *et al.*
ProteinScape 3.0	Bruker Daltonics
Compass DataAnalysis, Version 4.0 SP 4	Bruker Daltonics
flexAnalysis, Version 3.3 (Build 65)	Bruker Daltonics
TopSpin 3.0	Bruker
Chromatography system 4.1	Merck KGaA
ChemStation B.04.03	Agilent
Biotools	Bruker Daltonics

6.2 Analysis of Released *N*-Glycans of Cetuximab

6.2.1 *N*-Glycan Release from Cetuximab and Purification

N-Glycans were enzymatically cleaved from Erbitux with trypsin and *N*-glycosidase F. Briefly, 580 µL Erbitux infusion solution (5 mg/mL) were purified from excipients by centrifugal filters (Amicon® Ultra, Millipore™), freeze-dried and dissolved in 250 µL 6 M urea. Prior to tryptic digestion, the glycoprotein was reduced by addition of 2.6 µL 100 mM DTT. After reduction for 10 min at 60 °C the protein was alkylated with 2.6 µL 300 mM iodoacetamide at room temperature in the dark (incubation time: 30 min). The solution was diluted with 2.1 mL 100 mM $NaHCO_3$, pH 8.3 and trypsin was added in a protein to enzyme ratio of 75:1. After 16 hours of incubation at 37 °C, trypsin was inactivated for 10 min at 99 °C. Glycans were enzymatically cleaved from Erbitux by incubation with 0.58 U *N*-Glycosidase F for 48 hours at 37 °C. Subsequently, the digest was freeze-dried, dissolved in 100 µL ddH_2O and subjected to solid phase extraction (SPE) to separate the glycans from the tryptic peptides. For SPE C_{18}-cartridges (100 mg) were used in reversed phase mode. The SPE-material was washed with 5 column volumes (CV) 95% ACN/0.1% FA and equilibrated with 5 CV 5% ACN/0.1% FA. After

sample loading, a three-step gradient was applied for elution of analytes. Fraction 1 was eluted with 10 CV 5% ACN/0.1% FA (mainly polar compounds such as glycans and short peptides should elute in this step), fraction 2 with 10 CV 50% ACN/0.1% FA (elution of more hydrophobic analytes such as glycopeptides) and fraction 3 with 10 CV 95% ACN/0.1% FA (elution of hydrophobic peptides). Fraction 1 was lyophilized and re-dissolved in 200 µL ddH_2O and filtered (0.22 µm cellulose acetate filter).

6.2.2 Chromatographic Separation, Mass Spectrometric Analysis and Fractionation of Released *N*-Glycans from Cetuximab by PGC-LC-ESI-q/TOF

50% of the resultant glycan mixture were analyzed by PGC-LC-MS/MS. The analysis was performed on an ESI-q/TOF (maXis, Bruker) coupled to an Ultimate 3000 UPLC (Dionex). MS data were acquired in positive ion mode with an *m/z* range of 100 to 2975 and a spectra rate of 1.0 Hz. The capillary voltage was set to 5000 V, nebulizer pressure to 3 bar, dry gas to 7 L/min and drying temperature to 190 °C. The three most intense ions (if an intensity threshold of 1000 counts was minimally reached) were automatically chosen for fragmentation. CID fragmentation occurred in the collision cell with nitrogen as collision gas and a collision energy of 40 to 55 eV depending on charge state and *m/z* range. Singly charged ions were excluded and ions that were fragmented four times were excluded for 1 min. The separation on PGC was carried out with a HyperCarb^{TM} column (4.6×150 mm, 3 µm particle size, Thermo Fisher) with ultrapure water containing 65 mM ammonium formate and 10 mM ammonia as solvent A and acetonitrile containing 0.1% FA as solvent B using the following gradient: 0 min, 5% B; 2 min, 8% B; 53 min, 35% B; 57 min, 90% B; 61 min, 90% B; 63 min, 5% B; 65 min, 5% B. Ammonia was added to avoid separation of α- and β-anomers. The column temperature was set to 40 °C and the flow-rate to 800 µL. The liquid flow was split with a post column splitter, 5% were directly injected into the mass spectrometer and 95% of the LC flow were further fractionated in well plates in 15 s (~ 190 µL) fractions. Calibration was performed by a pre-run calibration (phosphazine mix, Agilent) and a lock mass calibration.

6.2.3 ^{1}H-NMR Analysis of Released *N*-Glycans from Cetuximab

Glycan-containing fractions from the deep well plate were freeze-dried and re-dissolved in 180 µL D_2O each. NMR spectra of all fractions were acquired on an Avance 700 MHz NMR spectrometer equipped with a cryo probe. All spectra were recorded with 4 k scans utilizing an excitation sculpting technique for solvent suppression, a spectral width of 10 ppm, an acquisition time of 2.34 s per scan and a relaxation delay of 1 s.

6.2.4 Three-Dimensional Cross Correlation (3DCC) of Released *N*-Glycans from Cetuximab

NMR data were imported into MATLAB utilizing the script "rbnmr" by Nils Nyberg (SLU, 2001-05-02, nn@farma.ku.dk). Signals of impurities and residual signals of suppressed solvent were set to zero. Exported EIC data points (MS data) were reduced using Origin 8.6 to the number of collected fractions (= number of NMR spectra) and imported into MATLAB from Excel. For quantification, the integrals of the EICs were normalized to the maximum intensity of the NMR spectra. The extracted 3DCC spectra were calculated by least square optimization and by Pearson correlation coefficients as described before.[81] Furthermore, a new variant calculates 3DCC spectra by singular value decomposition using the MATLAB SVD routine. A detailed description can be found in section "Results and Discussion" (cf. section 5.1.1).

6.3 Analysis of Intact mAbs and mAb Sub-Domains

6.3.1 MS Analysis of Intact mAbs for *N*-Glycosylation Analysis

Intact mAbs (Cetuximab, Bevacizumab) from pharmaceutical vials were desalted on Amicon centrifugal filters at 10000 × g (MWCO: 10 kDa). Subsequently, samples were acidified to 0.1% FA and directly used for LC-MS analyses (n = 25-35 pmol). MAbs were purified on a RP-C_8 column (2.1×250 mm, Aeris Widepore, 300 Å, phenomenex) and injected online into the ESI-q/TOF (maXis, Bruker Daltonics). The column oven temperature was set to 60 °C, the flow-rate to 250 µL min. 0.1% FA in ddH_2O was used as buffer A, 0.1% FA in ACN as buffer B. The linear gradient was changed as follows:

0 min, 5% B; 5 min, 5% B; 33 min, 40% B; 35 min, 95% B, 37 min, 95% B, 38 min, 5% B; 40 min, 5% B. MS data were acquired in positive mode by applying an ISCID of 120 eV for Cetuximab and 80 eV for Bevacizumab. Dry heater was set to 200 °C, a dry gas of 7 L/min was used. The mass range was set to *m/z* 500 - 4498.

6.3.2 Middle Up Approach for Analyzing the *N*-Glycosylation of Cetuximab

Cetuximab sub-domains were further analyzed by three approaches: 1) Separation of heavy and light chains by reduction of disulfide bridges, 2) digestion with papain to obtain the Fc domain separated from the two Fab/2 domains, 3) reduction of disulfide bridges of papain-digested Cetuximab to separate the light chain from the part of the heavy chain that formed together the Fd domain and the Fc domain in two equal parts of the C-terminal heavy chain.

6.3.2.1 Reduction of Cetuximab for Separation of Heavy and Light Chains

Cetuximab was desalted using Amicon centrifugal filters (MWCO: 10 kDa). 2 µL mAb (= 67 pmol) were mixed with 8 µL 125 mM DTT. After incubation at 37 °C for 30 min, 10 µL 0.2% FA were added and 18 µL were injected for LC-MS analysis. The same column and the same LC-MS settings as described in section 6.3.1 were used except of the ISCID that was set to 40 eV and the linear gradient was changed as follows: 0 min, 5% B; 5 min, 5% B; 33 min, 30% B; 35 min, 95% B, 37 min, 95% B, 38 min, 5% B; 40 min, 5% B.

6.3.2.2 Papain Digestion of Cetuximab

For optimization of papain digestion of Cetuximab, several parameters were varied. Table 15 summarizes different approaches. Tris-HCl buffer contained 100 mM Tris-HCl, 2 mM EDTA and 10 µM L-cysteine (pH: 7.5). 50 mM sodium phosphate buffer contained 1 mM EDTA and 10 µM L-cysteine (pH: 7.0). Papain was dissolved in 50 mM Tris-HCl (pH 8.2). Incubation occurred for all approaches at 37 °C and 400 rpm. After incubation, the samples were frozen in liquid nitrogen to immediately stop papain

activity. Prior to LC-MS analysis, the samples were mixed with an equal volume of 0.2% FA.

Table 15: Variation of parameters for optimizing papain digestion of Cetuximab (pH of Tris-HCl buffer was adjusted to 7.0, the sodium phosphate buffer to 7.5):

Sample	Buffer System	Protein-to-Enzyme Ratio	c(mAb) [µM]	Incubation Time [min]
A	Tris-HCl	1:10	5	60, 120, 180, 360, 1500
B	Tris-HCl	1:10	10	15, 60, 185, 360, 1500
C	Tris-HCl	1:10	2.5	15, 60, 185, 360, 1500
D	Tris-HCl	1:100	5	15, 60, 185, 360, 1500
E	Sodium Phosphate	1:10	5	10, 30, 50, 78

For the detailed LC-MS analysis of the sub-domains following parameters were used: enzyme-to-protein ratio: 1 to 10, mAb concentration in reaction mixture: 10 µM, time of incubation: 16.5 h. 50 pmol of this solution were mixed 1:2 with 0.2% FA and analyzed by LC-MS with an ISCID of 50 eV. The same LC-MS settings as described in section 6.3.1 were used, but 0.1% FA in 2-propanol was used as buffer B.

6.3.2.3 Papain Digestion of Cetuximab Followed by Reduction

For disulfide bridge reduction, the Cetuximab solution digested by papain was used. 1 µL mAb (= 10 pmol) were mixed with 39 µL 250 mM DTT. LC-MS analysis was performed as described for the papain digested mAb with an ISCID set to 5 eV.

6.3.3 Middle Up Approach for Analyzing the *N*-Glycosylation of Bevacizumab

268 pmol (4 µL) Bevacizumab (Roche, batch 1: 10128296 **G**, batch 2: 10121546 **RO-BG**, batch 3: 10128294 **ELL**) were mixed with 60 µL digestion buffer (2mM L-cysteine/20mM Tris-HCl/pH: 7.0) and 1 µL papain solution (43 mM in digestion buffer) was added (enzyme-to-protein ratio: 1:6). Incubation occurred at 37 °C, incubation time was varied: 30 min, 70 min, 110 min, 190 min, 21 h. 15 µL (62 pmol) of each solution were used for LC-MS data analysis. For all approaches, LC was performed using a C_8 column (2.1×250 mm, Aeris Widepore, 300 Å, phenomenex). The column oven was set to 60 °C, the flow-rate to 250 µL min. 0.1% FA in ddH_2O was used as buffer A, 0.1% FA in ACN as buffer B. The linear gradient was changed as follows:

0 min, 5% B; 5 min, 5% B; 33 min, 60% B; 35 min, 95% B, 37 min, 95% B, 38 min, 5% B; 40 min, 5% B. MS data were acquired in positive mode by applying an ISCID of 50 eV. Dry heater was set to 200 °C, a dry gas of 7 L/min was used. The mass range was set to *m/z* 500 - 4498.

6.3.3.1 Middle Up Approach for Comparative Study of the Fc-Glycosylation of Three Batches of Bevacizumab

Three batches of Bevacizumab (c = 25 mg/mL) were compared to each other regarding their *N*-glycosylation within the Fc domain. Bevacizumab carries two *N*-glycosylation sites in total (linked to each heavy chain within the Fc domain).

For papain digestion, 5 µL of each mAb (= 835 pmol) were mixed with 5.5 µL papain solution (83.5 pmol in 2 mM EDTA / 100 mM Tris-HCl / 10 mM L-cysteine / pH:7.4, protein-to-enzyme ratio: 10:1) and 189.5 µL buffer (2 mM EDTA / 100 mM Tris-HCl / 10 mM L-cysteine / pH:7.4). Incubation occurred for 20 h at 37 °C. 10 µL (41.75 pmol mAb) were analyzed by RP-LC-MS.

For reduction of papain-digested Bevacizumab, 10 µL (41.75 pmol mAb) of above described digestion solution were mixed with 40 µL 1 M DTT and incubated at 37 °C for 30 min. The whole solution was used for RP-LC-MS.

For both approaches, LC was performed using a C_8 column (2.1×250 mm, Aeris Widepore, 300 Å, phenomenex). The column oven temperature was set to 60 °C, the flow-rate to 250 µL min. 0.1% FA in ddH_2O was used as buffer A, 0.1% FA in ACN as buffer B. The linear gradient was changed as follows: 0. min, 5% B; 5. min, 5% B; 33. min, 60% B; 35. min, 95% B, 37. min, 95% B, 38. min, 5% B; 40. min, 5% B. MS data were acquired in positive mode by applying an ISCID of 5 eV. Dry heater was set to 200 °C, a dry gas of 7 L/min was used. The mass range was set to *m/z* 500-4498.

Data processing occurred with DataAnalysis 4.2 (Bruker Daltonics). Spectra were summed up and deconvolved by a maximum entropy method. Data were interpreted by MATLAB (details described in Results and Discussions).

6.4 Interlaboratory Study on Glycan Analysis

Two frozen monoclonal antibody samples were provided from NIST (National Institute of Standards and Technology):

Sample A: 0.4 mg, 100 mg/mL mAb

Sample B: 0.4 mg, 100 mg/mL mAb

The samples were diluted with each 46 µL of a buffer containing 25 mM L-Histidine, pH 6.0 (distributed by NIST) to a final concentration of 8 mg/mL (≅ 54 pmol/µL).

6.4.1 MS Analyses of Intact NIST Monoclonal Antibodies

For mass spectrometric analyses of the intact mAbs, 1 µL of each sample (≅ 54 pmol) was diluted with 29 µL 0.1% FA in ddH_2O. The sample was purified by a C_8-RP column (1.5×300 mm, Widepore, 300 Å, Phenomenex) at 60 °C with 0.1% FA in ddH_2O as buffer A and 0.1% FA in ACN as buffer B with following gradient: 0 min, 5% B; 5 min, 5% B; 33 min, 60% B; 35 min, 95% B; 37 min, 95% B; 38 min, 5% B; 40 min, 5% B. MS data were acquired with an ESI-q/TOF in positive mode in a *m/z* range of 295 - 3845. An ISCID of 120 eV was applied.

6.4.2 Middle Up Analyses of NIST Monoclonal Antibodies

For middle up analyses the mAbs were either reduced with DTT to analyze the heavy and the light chain separately or digested with papain to obtain individual MS spectra of the Fab and the Fc domain, respectively.

For reduction 1 µL of each mAb (≅ 54 pmol) was diluted with 29 µL 250 mM DTT and incubated at 37 °C for 30 min. The samples were analyzed by LC-MS analog to the intact mAbs (cf. section 6.4.1), except for the mass range that was set to *m/z* 600 - 4790 and the ISCID that was set to 20 eV.

For digestion with papain 2 µL (106 pmol) of each mAb were diluted with 57.3 µL incubation buffer (2 mM EDTA, 100 mM Tris-HCl, 10 mM L-cysteine, pH.4) and 0.7 µL (≅ 10.6 pmol) of a papain solution (protein-to-enzyme ratio: 10:1) were added and the

solution was incubated at 37 °C for 14 hours. 30 µL of each digestion were diluted with 30 µL 0.2% FA and directly analyzed by LC-MS (cf. section 6.4.1), except for the ISCID that was set to 35 eV. The duplicate samples were lyophilized and re-dissolved in 30 µL 6 M urea/250 mM DTT and incubated at 37 °C for 40 min. The samples were analyzed by LC-MS analog to the reduced mAbs (see above).

To determine the presence of neuraminic acids, a neuraminidase (from *Clostridium perfringens*) digestion was performed. Prior to enzymatic cleavage, the mAbs were reduced to gain access to the *N*-glycosylation sites that are sterically hidden within the Fc domain. Therefore, 1 µL of each antibody (≅ 54 pmol) was added to 29 µL 250 mM DTT and incubated at 60 °C for 30 min. The samples were diluted with 970 µL 100 mM sodium acetate buffer (pH 5). The digestion was performed by adding 0.5 U neuraminidase and incubating at 37 °C for 16 hours. The sample volumes were reduced to 100 µL by freeze-drying and analyzed by LC-MS as described for the reduced mAbs (see above).

6.4.3 Bottom Up Analyses of NIST Monoclonal Antibodies

For MS/MS analysis of the free *N*-glycans, the NIST mAbs were digested with trypsin and the carbohydrates were released by cleavage with PNGase F. Therefore, 3.75 µL (≅ 200 pmol) of each mAb were lyophilized and re-dissolved in 50 µL 6 M urea. For reduction, 1.3 µL 100 mM DTT were added and the solution was incubated for 10 min at 60 °C. For alkylation, 1.3 µL 300 mM iodoacetamide were added and the reaction solution was incubated at room temperature for 30 min in the dark. The samples were diluted with 425 µL of an ammonium bicarbonate buffer (25 mM, pH 7.4), 2 µL trypsin (≅ 10 pmol) were added (protein-to-enzyme ratio: 20:1) and the digestion was performed for 16 hours at 37 °C. The digestion was terminated by inactivating trypsin at 95 °C for 15 min. For enzymatic release of *N*-glycans 2 mU PNGase F were added and incubated at 37 °C for 48 h. For separation of glycans, glycopeptides and peptides a C_{18}-reversed phase-SPE (SPE: solid phase extraction, Bond Elut-C18, 100 mg, 1 mL, Agilent Technologies) was performed. Therefore, the column was conditioned with 1 mL

95% ACN/0.1% FA and equilibrated with 1 mL 5% ACN/0.1% FA. The sample was slowly loaded onto the column, and consecutively rinsed with each 1 mL 5% ACN/0.1% FA (fraction 1), 50% ACN/0.1% FA (fraction 2) and 95% ACN/0.1% FA (fraction 3). The flow-through was pooled with fraction 1, lyophilized and re-dissolved in 100 µL 0.1% FA for MS analysis. The sample was loaded onto a PGC (porous graphitized carbon column, Hypercarb, Thermo Scientific) and separated with following gradient (buffer A: ddH_2O/0.1% FA, buffer B: ACN/0.1% FA: 0 min, 2% B; 5 min, 2% B; 60 min, 30% B; 110 min, 90% B; 135 min, 90% B; 138 min, 2% B; 140 min, 2% B. The flow-rate was set to 0.15 mL/min and the column oven temperature to 40 °C. MS data were acquired in positive mode at a *m/z* range of 100 - 2800. Dry gas was set to 6 L/min and 180 °C, the nebulizer was set to 3 bar. The three most intense peaks of each full scan spectrum were isolated as precursors for CID fragmentation to generate MS/MS spectra. Ions with a charge of 2-4 were fragmented with collisions energy values of 20-30 eV.

For complementary data, ESI- and MALDI-MS was performed for released *N*-glycans. Therefore, 35 µL of each NIST sample (1.9 nmol) were mixed with 18 mg urea and 1.3 µL 100 mM DTT solution and incubated at 60 °C for 10 min. After addition of 1.3 µL 300 mM iodoacetamide the mixture was incubated at RT for 30 min in the dark. Each sample was diluted with 425 µL ammonium bicarbonate buffer (25 mM, pH 7.4) and incubated at 37 °C for 16 hours after addition of 5 µg trypsin. Trypsin was inactivated by heating at 95 °C for 15 min. After cooling-down to RT 5 mU PNGase F were added and the solution was incubated at 37 °C for 48 hours. The separation of glycans, glycopeptides and peptides was performed by C_{18}-reversed phase-SPE (SPE: solid phase extraction, Bond Elut-C18, 100 mg, 1 mL, Agilent Technologies). SPE was performed as described before (see above). The flow-through was pooled with fraction 1, lyophilized and re-dissolved in 100 µL 0.1% FA for spectroscopic analyses. Therefore, each sample was loaded onto a PGC column (porous graphitized carbon, Hypercarb, Thermo Scientific) and separated with a flow-rate of 800 µL/min with following gradient (buffer A: ddH_2O/0.1% FA, buffer B: ACN/0.1% FA): 0 min, 2% B; 5 min, 2% B;

6 min, 8% B; 53 min, 35% B; 55 min, 90% B; 60 min, 90% B; 62 min, 2% B; 65 min, 2% B. The flow was split post-column. 5% (50 µL/min) were directly injected into the ESI-q/TOF and analyzed by MS/MS. MS data were acquired in positive mode at a *m/z* range of 100 - 2800. Dry gas was set to 5 L/min and 180 °C, the nebulizer was set to 0.8 bar. The three most intense peaks of each full scan spectrum were isolated as precursors for CID fragmentation to generate MS/MS spectra. Ions with a charge of 2-4 were fragmented with collisions energy values of 20-30 eV. 95% of the flow was further fractionated (15 sec/well) into a 96 well plate. For analysis by MALDI-MS(/MS) the fractions were freeze-dried and re-dissolved in 20 µL ddH_2O. 1 µL was spotted with DHB (2,5-dihydroxybenzoic acid, 10 mg/mL in TA30) as matrix on a polished steel target and dried. MS spectra were acquired in positive mode. MS/MS spectra were generated by LID (laser-induced dissociation) with varying laser energies.

6.5 Histidine-Rich Glycoprotein

Histidine-Rich Glycoprotein (HRG) was analyzed from human blood plasma. Deep-frozen plasma samples of expired blood samples of healthy individuals were obtained from a local blood bank, by courtesy of the *Zentralinstitut für Transfusionsmedizin GmbH* in Hamburg, Germany.

A method for isolation of HRG was established. Therefore, the remarkably high amount of histidine residues exhibiting imidazole side chains were utilized for enriching. Commercially available matrices developed for isolation of recombinant produced proteins showing His-tags (typically showing 6 to 10 His residues) were tested for suitability to enrich HRG by immobilized metal affinity chromatography (IMAC). “Talon super flow” is a cobalt-based IMAC medium, “Ni-NTA affinity resin” is based on nickel complexes. The Ni-NTA resin showed more specific enrichment of HRG in preliminary experiments and was chosen for further experiments.

6.5.1 Immobilized Metal Affinity Chromatography by Ni^{2+}-NTA Resin for Isolation of Human Histidine-Rich Glycoprotein

Blood plasma was thawed and centrifuged 10 min at 14000 × g at 4 °C to remove insoluble components. The supernatant was spiked with imidazole to inhibit unspecific binding of other proteins than HRG in varying concentrations (20 mM, 35 mM, 80 mM). Effect of chaotropic reagents on enrichment of HRG was tested. Urea was either added to blood plasma prior to IMAC (Immobilized Metal Affinity Chromatography) or prior to SEC. After sample loading on to the Ni-NTA column the resin was washed with Tris-HCl buffer containing imidazole in varying concentrations (20 mM, 50 mM, 80 mM) to ascertain best conditions for removal of weakly binding impurities. Elution of HRG and potential other strong-binding proteins was achieved by addition of Tris-HCl buffer containing 400 mM imidazole. Eluates were dried under vacuum for further chromatographic steps.

6.5.2 Size-Exclusion Chromatography of IMAC Enriched Blood Components for Isolation of Human Histidine-Rich Glycoprotein

Size-exclusion chromatography (SEC) was performed with gel filtration columns from GE Healthcare. Three columns were tested: Superdex75 (300×10 mm, optimal resolving power for analytes ranging between 3-70 kDa), Superdex200 (300×10 mm and 600×16 mm, optimal resolving power for analytes ranging between 10-600 kDa). 10 mM Tris-HCl containing 100 mM NaCl (pH: 8.0) was used as buffer system. Flow-rate was set to 1.3 mL/min when using Superdex200 (600×16 mm) and ranged between 150-300 µL/min when using Superdex200 (300×10 mm) or Superdex75 (300×10 mm) (cf. Results & Discussion).

6.5.3 SDS-PAGE for Control of HRG Enrichment

SDS-PAGE was performed with a one-buffer system from Bio-Rad either with 16% TRIS gels or gradients gels (4-16%). Samples were heated to 95 °C for 5 min in the presence of SDS. DTT was either added to release HRG domains that are linked only via

disulfide bridges or avoided to analyze HRG without reduction reagent to obtain an overview about purity. Gels were stained with Coomassie.

6.5.4 Dynamic Light Scattering for Determination of Size Distribution of Enriched HRG

To determine structural unity (size) of isolated HRG, DLS analyses were performed for native and deglycosylated HRG solutions (dissolved in 10 mM Tris-HCl, 20 mM NaCl, pH: 8.0). For deglycosylation, 2 µL PNGase F were added to the HRG solution and incubated at 37 °C overnight. Both, native and deglycosylated HRG were concentrated by Amicon Centrifugal Filters (MWCO: 3 kDa) and the concentrations were determined by Nanodrop analyses (MW: 60 kDa, extinction coefficient calculated based on the amino acid sequence: 26900) to c = 4.0 mg for deglycosylated HRG and to c = 8 mg/mL for native HRG. The samples were centrifuged at 16000 × g for 45 min to remove insoluble components. 15 µL of the supernatants were used for DLS analyses in a Quartz cuvette on a *spectroLITE 300* in the Institute for Biochemistry, University Hamburg. Hydrodynamic radiuses of the samples were determined.

6.5.5 Acquisition of MALDI Spectra of Intact Histidine-Rich Glycoprotein

7.6 mg dihydroxyacetophenon (DHAP) were dissolved in 375 µL and sonicated for 15 min. 125 µL diammonium hydrogen citrate (c = 18 mg/mL) were added to the solution and mixed vigorously. 1 µL of each sample was mixed with 1 µL 2% TFA and 1 µL prepared DHAP solution. Crystallization was initiated by scratching with a pipette tip within the solution. 1 µL was spotted on polished steel MALDI targets and dried. Spotted fractions were analyzed in linear positive mode on an UltrafleXtreme (Bruker Daltonics) in high mass range.

6.5.6 Plasmin Cleavage of Histidine-Rich Glycoprotein

To simulate *in vivo* occurring plasmin-digestion, purified freeze-dried HRG was dissolved in PBS buffer (pH: 7.4). Plasmin was added in a protein-to-enzyme ratio of 1:100. Incubation occurred for variable periods (from 15 min up to six hours) at 37 °C.

Digestion solutions were either analyzed directly or after incubation with 1 mg DTT for 10 min at 37 °C by LC-MS(/MS).

6.6 Prolactin-Inducible Protein

6.6.1 Saliva Preparation

Varying amounts of whole saliva were collected from voluntary test persons. Saliva was centrifuged at 5000 x g at 4 °C for 45 min to remove insoluble components. The supernatant was collected, lyophilized and re-dissolved in 10 mM Tris-HCl/100 mM NaCl (pH 8.5). Per 1 mL saliva 5 mg urea were added. Sample mixtures were filtered (0.22 µm cut-off) and applied to various chromatographic procedures.

6.6.2 Antibody Affinity Chromatography for Isolation of Salivary PIP

All experiments were performed with cyanogen bromide activated-Sepharose as matrix in a batch mode. For specific antigen-binding of PIP a polyclonal anti-PIP-antibody (*antibodies-online GmbH*) was used. 11.3 mg CNBr-activated Sepharose were mixed with 350 µL 1 mM HCl and washed at 4 °C for one hour. The supernatant was removed after centrifugation (4 °C, 1000 × g, 5 min) and the wash procedure was repeated once with equal conditions and twice with ddH_2O instead of 1 mM HCl. Subsequently, 12.4 µg anti-PIP-antibody in 136 µL 0.1 M sodium hydrogen carbonate buffer (pH: 8.3) were added to the pellet. After 20 hours at 4 °C the supernatant was removed from the pellet was washed twice with 400 µL 0.1 M sodium hydrogen carbonate buffer (+500 mM NaCl, pH: 8.3) for 5 min and with 400 µL 0.1 M Tris-HCl buffer (pH: 8.0) for three hours, both at room temperature. Afterwards, the pellet was washed with 400 µL 0.1 M Tris-HCl buffer (+500 NaCl, pH: 8.0), 400 µL sodium acetate buffer (+500 mM NaCl, pH: 4.0) and 400 µL PBS buffer (pH: 7.4). The lyophilized pellet was re-dissolved 50 µL PBS buffer (pH: 7.4), transferred to the anti-PIP-antibody matrix and incubated two hours at 4 °C. Subsequently, the supernatant was removed and the pellet was washed three times with each 300 µL PBS buffer (pH: 7.4). Elution of PIP was performed six

times with each 50 µL 0.1 M glycine (pH: 2.4). Eluates were neutralized with 5 µL 0.1 M Tris-HCl buffer (pH: 8.0).

6.6.3 Optimizing Parameters for Isolation of Salivary PIP by AIEX

Various anion exchange chromatography experiments were performed for evaluation of their suitability for enrichment of salivary PIP. For all AIEX experiments using the HPLC system the flow-rate was set to 1 mL/min. The eluate was fractionated (1 min/fraction). For manual AIEX the eluate fractions had a size of 2 mL. For each experiment 200 µL saliva were mixed with 500 µL ddH_2O (originating from 3 mL saliva, female, age: 22 years).

The experiment shown for the HiTrap Q FF column (FPLC) was performed with 10 mM Tris-HCl/pH 8.5 as buffer A and 10 mM Tris-HCl/500 mM NaCl/pH 8.5 as buffer B. The gradient was as follows: 0 min, 0% B; 10 min, 0% B; 37 min, 100% B; 40 min, 100% B; 42 min, 0% B; 45 min, 0% B. Fractions 13-16 contain PIP. After lyophilizing, samples were analyzed by RP-LC-MS.

The experiment shown for the HiTrap DEAE FF column (FPLC) was performed with 20 mM Tris-HCl/pH 8.5 as buffer A and 20 mM Tris-HCl/500 mM NaCl/pH 8.5 as buffer B. The same gradient was used as for the HiTrap Q FF column. Fractions 28-30 contained PIP. After lyophilizing, samples were analyzed by RP-LC-MS.

Manual chromatography was performed with DEAE Sephadex A-50 as stationary phase. 20 mM Tris-HCl/pH 6.2 was used as buffer A, 20 mM Tris-HCl/500 mM NaCl/pH 6.2 was used as buffer B. A stepwise gradient was used (each step: 10 mL): 10% B; 20% B; 30% B; 40% B; 50% B; 60% B; 70% B; 80% B; 90% B, 100% B. Several fractions contained PIP with a maximum at 150 mM NaCl.

6.6.4 Comparison of Human PIP from Multiple Saliva Samples by 2D-LC-MS (AIEX-RP)

All samples were pretreated by anion exchange chromatography (HiTrap DEAE FF). Eluent A was composed of 20 mM Tris-HCl buffer (pH: 8.5), eluent B additionally

contained 500 mM NaCl. 400 µL of saliva samples (originating from 8 mL saliva) were mixed with 400 µL A prior to injecting. Elution occurred by following gradient: 0 min, 0% B; 10 min, 0% B; 37 min, 100% B; 40 min, 100% B; 42 min, 0% B; 45 min, 0% B. The flow-rate was set to 1 mL/min and fractions were collected per minute. Fractions were pooled after manual evaluation of UV trace (monitoring by UV, λ = 280 nm) and lyophilized to a final volume of 150-250 µL.

6.6.5 Comparison of Human PIP from Multiple Saliva Samples by 2D-LC-MS (SEC-RP)

All samples for comparison of healthy individuals by SEC, of pregnant women and breast cancer patients were performed as described in this section.

10 mL of saliva were received from each individual. All saliva samples were immediately frozen at -32°C until further processing. Saliva samples were thawed and centrifuged at 5000 × g at 4 °C for 45 min for removal of insoluble components. The supernatant was aliquoted (2 mL/aliquot) and lyophilized. Each aliquot was re-dissolved with 5 mg urea in 100 µL 20 mM Tris-HCl/100 mM NaCl (pH: 8.0). After rigorous mixing, the sample was centrifuged for 10 min at 2 °C at 10000 × g and afterwards filtered through a 0.22 µm cellulose acetate filter at 2 °C at 11000 × g to remove remaining mucous components. Subsequently, the filtrate was subjected to size exclusion chromatography (Superdex75). 20 mM Tris-HCl/100 mM NaCl (pH: 8.0) was used as buffer and the flow-rate was set to 150 µL/min. After a wait-time of 50 min, fractions were collected (one fraction/min = 150 µL/min) in a 96 well plate. Fractions 18 to 31 were pooled and lyophilized. The samples were each re-dissolved in 100 µL 0.1% FA and subjected to C_8-LC-MS/MS (cf. 6.6.9.1 for LC-MS/MS parameters).

6.6.6 Combination of AIEX and SEC for PIP Enrichment

Advantages of both chromatographic procedures were combined in this approach. Therefore, eluates from twelve AIEX runs (HiTrap DEAE FF, procedure as described above) that contain PIP were pooled. SEC (Superdex75, 10×300 mm) was performed

with 10 mM Tris-HCl/200 mM NaCl (pH: 8.0) at a flow-rate of 250 µL/min. 1 min fractions were collected.

6.6.7 Neuraminidase Digest

For neuraminidase digest, the protein was lyophilized, re-dissolved in 100 mM sodium acetate buffer (pH: 5) and incubated with 1 U neuraminidase (from *Clostridium perfringens*) at 37 °C overnight.

6.6.8 Release of *N*-Glycans from Human Salivary PIP

For enzymatic digestion, the buffer of isolated PIP (large scale SEC) was exchanged to 25 mM ammonium bicarbonate (pH: 7.5, ABC) using centrifugal filters (MWCO: 3 kDa). The sample was reduced with 1,4-D/L-dithiothreitol for 10 min at 60 °C and alkylated with iodoacetamide for 30 min at room temperature. The sample was diluted with ABC buffer and incubated with trypsin at 37 °C overnight. Trypsin was inactivated by heating at 99 °C for 15 min and PNGase F was added to release *N*-glycans at 37 °C for 48 h. Subsequently, the digest was subjected to solid phase extraction to separate free *N*-glycans from tryptic peptides (SPE cartridges, Agilent Technologies).

6.6.9 Mass Spectrometric Analyses of PIP Samples

6.6.9.1 MALDI-MS for Intact Protein Analysis

For MALDI-MS, dihydroxyacetophenone (DHAP) was used as matrix. The matrix was prepared as follows: 7.8 mg DHAP were dissolved in 375 µL ethanol, vortexed for 1 min and homogenized in the ultrasonic bath for 15 min. 125 µL biammonium hydrogencitrate (18 mg/mL) were added to the mixture. 1 µL of the matrix was mixed with 1 µL 2% TFA and 1 µL sample. Crystallization was initialized by scratching the tube surface within the mixture. 1 µL was spotted onto the target (ground steel) and subjected to MALDI-MS in positive linear mode (UltrafleXtreme, Bruker Daltonics). The mass range was set to *m/z* 5000 - 30000.

6.6.9.2 Analysis of Intact Proteins by RP-LC-MS

The protein was injected onto a C_8 column (Aeris Widepore XB-C8, 300 Å, Phenomenex, 250×2.1 and 150×2.1 mm) using a Dionex LC system (Thermo). The column oven temperature was set to 60 °C, the flow-rate to 250 µL/min. 0.1% FA was used as buffer A, 0.1% FA in ACN as buffer B. The linear gradient was changed as follows: 0 min, 5% B; 5 min, 5% B; 34 min, 60% B; 35 min, 95% B; 37 min, 95% B; 38 min, 5% B; 40 min, 5% B. An in-source CID of 20 eV and a spectra time of 0.5 Hz was applied. The mass range was set to *m/z* 500 - 4498. Spectra were summed up over variable periods, deconvolved and interpreted by an in-house developed MATLAB script.

6.6.9.3 Analysis of Released *N*-Glycans by RP-LC-MS/MS

Released *N*-glycans were subjected to LC-MS/MS analysis. MS data were acquired in positive ion mode with an *m/z* range of 10 to 2975 and a spectra rate of 1.0 Hz. The separation was carried out with a PGC column (Hypercarb, 3 µm, Thermo Fisher) and ammonia was added to the mobile phase to avoid separation of α- and β-anomers. The liquid flow was split post-column, 5% were directly injected into the MS and 95% of the LC flow was further fractionated. Fractions were lyophilized and re-dissolved in D_2O for ^{1}H-NMR and TOCSY analyses (700 MHz NMR with cryo probe, Bruker). ^{1}H-NMR spectra were acquired with 3 k scans.

7 HAZARDS

Table 16: Hazard and precautionary statements according Globally Harmonized System of Classification and Labelling of Chemicals (GHS).

Substance	Pictograms	GHS Hazard Statements
Acetone		H225, H319, H336, P210, P240, P305+P351+P338, P403+P233
Acetonitrile		H225, H332, H302, H312, H319, P210, P240, P302+P352, P305+P351+P338, P403+P233
Ammonia (25%)		H290, H314, H335, H400, P260, P273 P280, P301+P330+P331, P303+P361+P353, P305+P351+P338
Ammonium bicarbonate		H302
Ammonium formate		H315, H319, H335, P261, 305+P351+P338
Bi-ammonium hydrogen citrate		H319
2,5-Dihydroxybenzoic acid		H302, H315, H319, H335, P261, P305+P351+P338
1,4-D/L-Dithiothreitol		H302, H315, H319, P302+P352, P305+P351+P338
ESI-TOF Tuning Mix		H225, H302+H332, H319, P210, P305+P351+P338
Ethanol		H225, H319, P210, P240, P305+P351+P338, P403+P233
Formic acid		H226, H302, H314, H331, P210, P280, P303+P361+P353, P304+P340+P310, P305+P351+P338, P403+P233
Hydrogen chloride		H290, H314, H335, P260, P280, P303+P361+P353, P304+P340+P310, P305+P351+P338
Imidazole		H302, H314, H360
Iodoacetamide		H301, H317, H334, P261, P280, P301+P310, P342+P311

Table 16: Continued.

Substances	Pictograms	GHS Hazard Statements
Methanol		H225, H331, H311, H301, H370, P210, P233, P280, P302+P352, P304+P340, P308+P310, P403+P235
Sodium dodecyl sulfate		H228, H302+H332, H315, H318, H335, H412, P210, P261, P280, P301+P312+P330, P305+P351+P338+P310, P370+P378
Sodium hydroxide		H290, H314, P280, P301+P330+P331, P305+P351+P338P308+P310
Trifluoroacetic acid		H290, H331, H314, H412, P260, P273, P280, P303+P361+P353, P305+P351+P338, P312
Tris(hydroxymethyl)amino methane		H315, H319, H335, P261, P305+P351+P338
Trypsin		H319, H335, H315, H334

8 APPENDIX

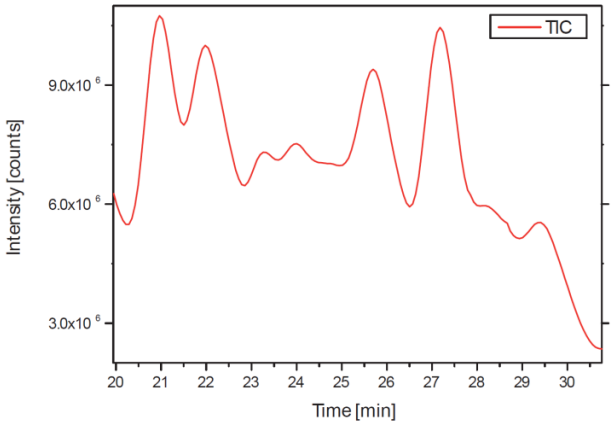


Figure 89: **LC-MS chromatogram of underivatized *N*-glycans enzymatically cleaved from Cetuximab separated on a PGC column. Ammonia was used as co-eluent to avoid separation of α- and β-anomers. A detail of the total ion chromatogram (minute 20 to 30) is shown. 37 *N*-glycans were identified by LC-MS(/MS). TIC: total ion chromatogram.**

Table 17: **MATLAB coefficients for Fd and Fc/2 domains of papain-digested, reduced Cetuximab.**

Glycan	n(Hex)	n(HexNAc)	n(dHex)	n(NeuGc)	Fd Coeff. Rel. [%]	Fc/2 Coeff. Rel. [%]
C_P_R_1	4	4	1	0	2.24	31.04
C_P_ R_2	3	4	1	0	0.96	35.33
C_P_ R_3	7	4	1	0	40.98	0.03
C_P_ R_4	6	4	1	1	9.23	0.03
C_P_ R_5	5	4	1	0	4.55	5.45
C_P_ R_6	5	2	0	0	0.02	5.95
C_P_ R_7	3	3	1	0	0.71	1.00
C_P_ R_8	6	3	1	0	0.26	1.39
C_P_ R_9	6	4	2	0	4.78	0.00
C_P_ R_10	5	4	1	1	3.89	0.02
C_P_ R_11	4	3	1	0	0.38	0.67
C_P_ R_12	5	3	1	0	1.03	1.44
C_P_ R_13	4	4	0	0	0.05	5.45
C_P_ R_14	6	5	1	0	3.97	0.04
C_P_ R_15	8	5	1	0	2.11	0.52

Table 17: Continued

Glycan	n(Hex)	n(HexNAc)	n(dHex)	n(NeuGc)	Fd Coeff. Rel. [%]	Fc/2 Coeff. Rel. [%]
C_P_ R_16	5	4	2	1	1.68	0.11
C_P_ R_17	4	4	1	1	0.93	0.05
C_P_ R_18	7	5	1	0	1.62	0.13
C_P_ R_19	7	5	1	1	2.87	0.58
C_P_ R_20	9	5	1	0	4.91	0.52
C_P_ R_21	7	3	0	0	0.14	0.65
C_P_ R_22	5	5	1	1	1.38	0.13
C_P_ R_23	6	5	1	1	0.96	0.60
C_P_ R_24	4	5	1	0	0.07	0.27
C_P_ R_25	5	3	1	1	0.01	0.11
C_P_ R_26	3	5	1	0	0.03	1.26
C_P_ R_27	5	3	0	0	0.14	0.31
C_P_ R_28	6	2	0	0	0.02	1.21
C_P_ R_29	5	4	1	2	0.17	0.22
C_P_ R_30	5	5	1	0	1.23	0.01
C_P_ R_31	7	3	1	0	0.12	0.18
C_P_ R_32	8	5	1	1	0.00	0.37
C_P_ R_33	6	3	0	0	0.13	0.70
C_P_ R_34	4	4	2	1	1.32	0.03
C_P_ R_35	6	3	1	1	0.43	0.01
C_P_ R_36	7	5	1	2	0.00	2.14
C_P_ R_37	3	4	0	0	0.02	2.04
C_P_ R_38	6	4	1	0	6.68	0.00

Table 18: MATLAB coefficients for Fab/2 domain of papain-digested Cetuximab.

Glycan	n(Hex)	n(HexNAc)	n(dHex)	n(NeuGc)	Fab/2 Coeff. Rel. [%]
C_P_1	4	4	1	0	2.46
C_P_2	3	4	1	0	0.54
C_P_3	7	4	1	0	44.68
C_P_4	6	4	1	1	1.60
C_P_5	5	4	1	0	4.46
C_P_6	5	2	0	0	0.73
C_P_7	3	3	1	0	0.14
C_P_8	6	3	1	0	0.44
C_P_9	6	4	2	0	1.09
C_P_10	5	4	1	1	6.46
C_P_11	4	3	1	0	0.03
C_P_12	5	3	1	0	0.83
C_P_13	4	4	0	0	0.00
C_P_14	6	5	1	0	6.07
C_P_15	8	5	1	0	1.78
C_P_16	5	4	2	1	0.91
C_P_17	4	4	1	1	1.09
C_P_18	7	5	1	0	1.02
C_P_19	7	5	1	1	2.56
C_P_20	9	5	1	0	6.77
C_P_21	7	3	0	0	0.63
C_P_22	5	5	1	1	0.53
C_P_23	6	5	1	1	0.09
C_P_24	4	5	1	0	0.43
C_P_25	5	3	1	1	0.11
C_P_26	3	5	1	0	0.79
C_P_27	5	3	0	0	0.02
C_P_28	6	2	0	0	0.05
C_P_29	5	4	1	2	0.00
C_P_30	5	5	1	0	2.06
C_P_31	7	3	1	0	0.63

Table 18: Continued.

Glycan	n(Hex)	n(HexNAc)	n(dHex)	n(NeuGc)	Fab/2 Coeff. Rel. [%]
C_P_31	7	3	1	0	0.63
C_P_32	8	5	1	1	0.13
C_P_33	6	3	0	0	0.19
C_P_34	4	4	2	1	3.11
C_P_35	6	3	1	1	2.05
C_P_36	7	5	1	2	0.00
C_P_37	3	4	0	0	0.00
C_P_38	6	4	1	0	5.53

Table 19: MATLAB coefficients for Fc domain obtained of papain-digested Bevacizumab.

Combination	Glyco Site	n(Hex)	n(HexNAc)	n(dHex)	Coefficients, rel. [%]		
					Batch 1	Batch 2	Batch 3
B_P_1	1	3	4	1	38.33	44.00	42.64
	2	3	4	1			
B_P_2	1	3	4	1	1.12	1.44	1.12
	2	3	3	1			
B_P_3	1	3	4	1	0.15	2.29	2.24
	2	3	4	0			
B_P_4	1	3	4	1	0.81	0.11	0.10
	2	5	2	0			
B_P_5	1	3	4	1	1.44	1.32	1.32
	2	6	2	0			
or	1	5	2	0			
	2	4	4	1			
B_P_6	1	3	4	1	2.94	3.32	3.22
	2	4	3	1			
or	1	3	3	1			
	2	4	4	1			
B_P_7	1	3	4	1	19.79	16.1	15.92
	2	4	4	1			

Table 19: Continued.

Combination	Glyco Site	n(Hex)	n(HexNAc)	n(dHex)	Coefficients, rel. [%]		
					Batch 1	Batch 2	Batch 3
B_P_8	1	3	4	1	8.28	4.84	4.76
	2	5	4	1			
or	1	4	4	1			
	2	4	4	1			
B_P_9	1	3	3	1	0.02	0.05	0.02
	2	3	3	1			
B_P_10	1	3	3	1	0.34	0.42	0.53
	2	3	4	0			
B_P_11	1	3	3	1	0.23	0.03	0.04
	2	5	2	0			
B_P_12	1	3	3	1	0.03	0.02	0.03
	2	6	2	0			
or	1	5	2	0			
	2	4	3	1			
B_P_13	1	3	3	1	0.00	0.03	0.03
	2	4	3	1			
B_P_14	1	3	3	1	3.23	2.41	2.51
	2	5	4	1			
or	1	4	3	1			
	2	4	4	1			
B_P_15	1	3	4	0	0.03	0.02	0.05
	2	3	4	0			
B_P_16	1	3	4	0	0.02	0.02	0.03
	2	5	2	0			
B_P_17	1	3	4	0	0.01	0.02	0.00
	2	6	2	0			
B_P_18	1	3	4	0	0.38	0.48	0.45
	2	4	3	1			
B_P_19	1	3	4	0	7.79	11.04	10.77
	2	4	4	1			
B_P_20	1	3	4	0	4.41	4.25	4.22
	2	5	4	1			

Table 19: Continued.

Combination	Glyco Site	n(Hex)	n(HexNAc)	n(dHex)	Coefficients, rel. [%]		
					Batch 1	Batch 2	Batch 3
B_P_21	1	5	2	0	0.50	0.09	0.07
	2	5	2	0			
B_P_22	1	5	2	0	0.00	0.00	0
	2	6	2	0			
B_P_23	1	5	2	0	3.205	1.83	2.05
	2	5	4	1			
or	1	6	2	0			
	2	4	4	1			
B_P_24	1	6	2	0	0.23	0.64	0.71
	2	6	2	0			
B_P_25	1	6	2	0	0.83	0.84	1.23
	2	4	3	1			
B_P_26	1	6	2	0	1.77	1.01	2.01
	2	5	4	1			
B_P_27	1	4	3	1	0.73	0.80	0.91
	2	4	3	1			
B_P_28	1	4	3	1	1.26	0.66	0.89
	2	5	4	1			
B_P_29	1	4	4	1	2.12	1.81	2.12
	2	5	4	1			
B_P_30	1	5	4	1	0.01	0.00	0.00
	2	5	4	1			

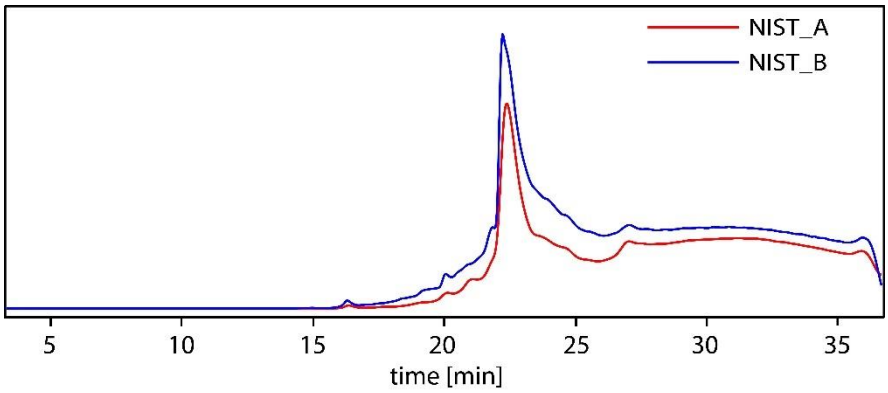


Figure 90: **Total-ion chromatograms of RP-LC-MS of the intact NIST samples. The monoclonal antibodies were purified on a C_8-RP-LC and directly injected into an ESI-q/TOF to obtain MS spectra of intact glycoproteins. An in-source decay of 120 eV was applied to obtain sufficient ionization. The total intensity counts indicate a ~ 20% larger concentration of the sample NIST_B since the ionization capability should be mainly dependent on the protein backbone which should be equal for both samples. Peaks in the slopes of the main signal were identified to be smaller proteins, likely fragments of the mAb samples or cell components formed during recombinant antibody production. Red: TIC of sample NIST_A, blue: TIC of sample NIST_B.**

Table 20: **Sample ID and information of blood plasma samples (cf. 5.2.2).**

Sample ID	age	sex	Blood group
A	19	male	?
B	58	male	A
C	60	male	A
D	21	female	0
E	48	female	A
F	45	male	B
G	21	male	0
H	54	male	A
I	28	female	A
J	24	male	B
K	55	male	B
L	46	male	A
M	47	male	B

Table 21: Summary of glycan analysis of salivary prolactin-inducible protein from MS analysis of the intact different mass spectrometric techniques.

Antennae	Hex*	HexNAc*	dHex*	NeuAc*	Rel. Int. [%]	Mass max. [Da]	Mass overlapping structures
bi	5	4	2	0	0.06	15416.7	
bi	5	4	1	1	0.30	15561.7	A
bi	5	4	3	0	0.63	15562.7	A
bi	5	4	2	1	4.56	15707.8	B
bi	5	4	4	0	1.56	15708.8	B
bi	5	4	3	1	7.74	15853.8	C
bi	5	4	5	0	5.28	15854.8	C
bi	5	4	2	2	3.41	15998.9	D
bi	5	4	4	1	2.18	15999.9	D
bi	5	4	3	2	1.03	16144.9	
tri	6	5	2	0	0.38	15781.8	
tri	6	5	3	0	0.31	15927.9	
tri	6	5	4	0	1.17	16073.9	
tri	6	5	3	1	4.57	16219.0	E
tri	6	5	5	0	1.72	16220.0	E
tri	6	5	2	2	2.40	16364.0	F
tri	6	5	4	1	4.37	16365.0	F
tri	6	5	6	0	4.82	16366.0	F
tri	6	5	3	2	10.17	16511.1	G
tri	6	5	5	1	7.90	16512.1	G
tri	6	5	7	0	5.30	16513.1	G
tri	6	5	4	2	9.35	16657.1	H
tri	6	5	6	1	8.24	16658.1	H
tri	6	5	3	3	4.29	16802.1	
tri	6	5	4	3	3.29	16948.2	
tetra	7	6	6	1	1.18	17023.3	
tetra	7	6	7	1	0.72	17169.3	I
tetra	7	6	9	0	0.84	17170.3	I
tetra	7	6	3	1	0.50	16585.1	
tetra	7	6	4	1	0.90	16731.1	
tetra	7	6	5	1	0.72	16877.2	
tetra	7	6	5	2	0.13	17168.3	

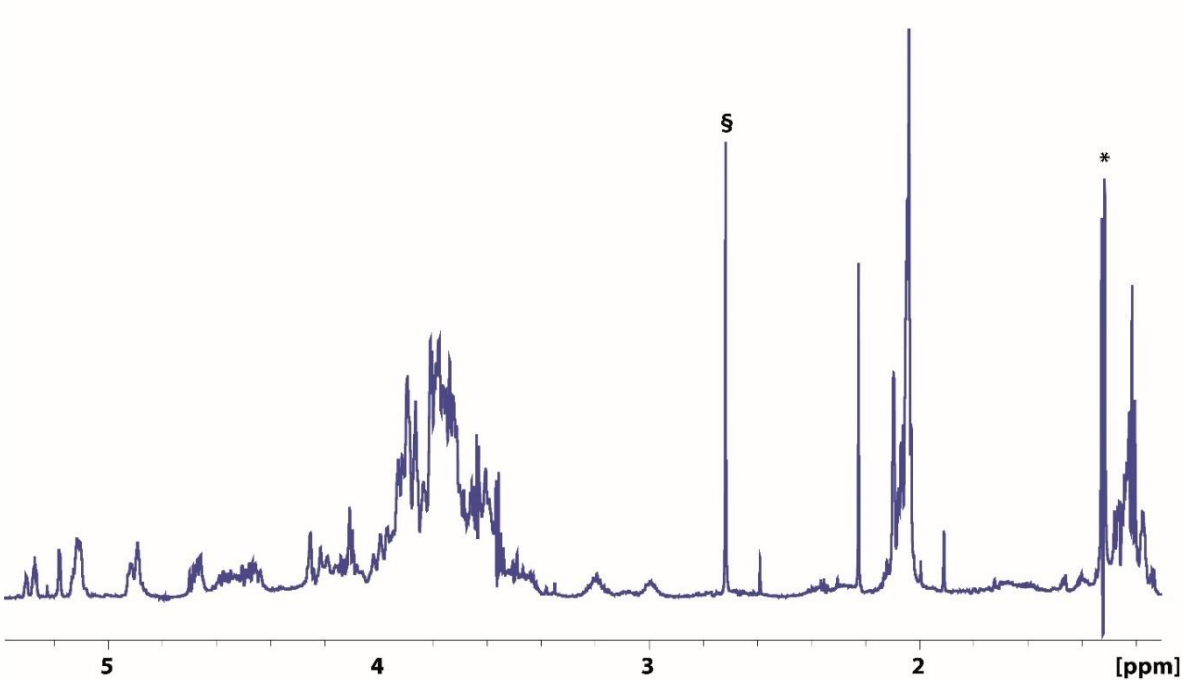


Figure 91: **Summed up ^{1}H-NMR spectrum of *N*-glycans linked to salivary PIP. §: dimethyl sulfoxide, *: lactate.**

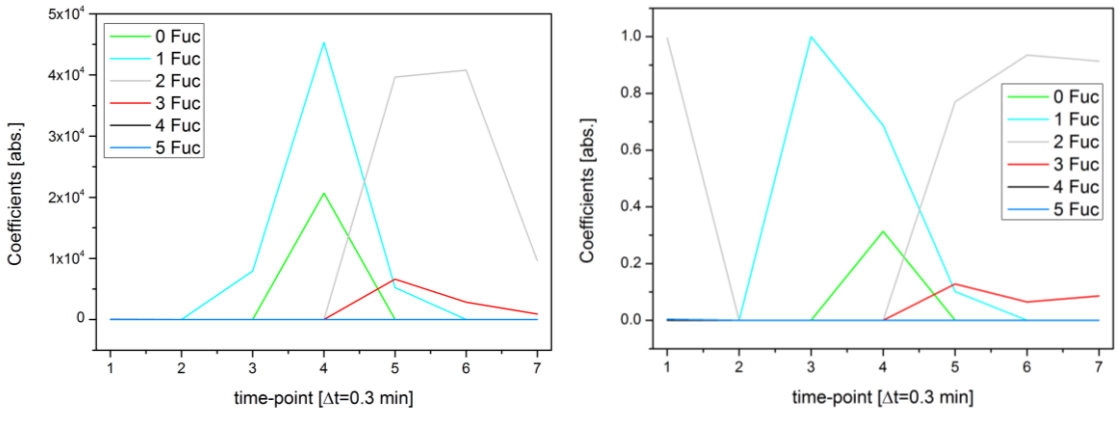


Figure 92: **Content of fucose residues attached to biantennary *N*-glycans with two terminal neuraminic acid (full occupation of antennae) as a function of the retention time in RP-LC. MS spectra were summed up over time (each 0.3 minutes). The level of fully fucosylated biantennae is drastically decreased. Because of decreased MS intensity and S/N ratio, data are less reliable than for biantennae holding no or one NeuAc. Left: absolute amounts, right: relative amounts.**

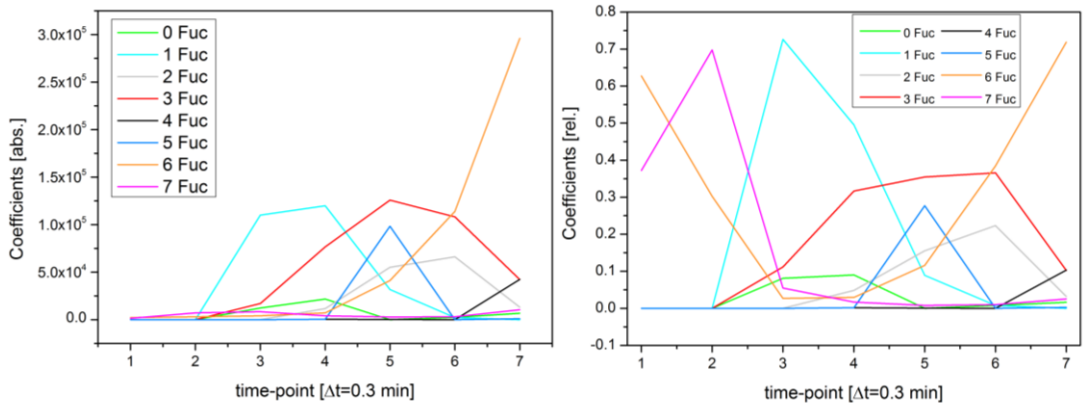


Figure 93: **Content of fucose residues attached to triantennary *N*-glycans with two terminal neuraminic acids as a function of the retention time in RP-LC. MS spectra were summed up over time (each 0.3 minutes). Highly fucosylated structures are most abundant and elute prior less fucosylated biantennae. The level of fully fucosylated triantennae is drastically decreased compared to the same glycan without a terminal NeuAc. All glycans holding two NeuAc elute later compared to glycans lacking NeuAc due to an increased hydrophobicity. Left: absolute amounts, right: relative amounts.**

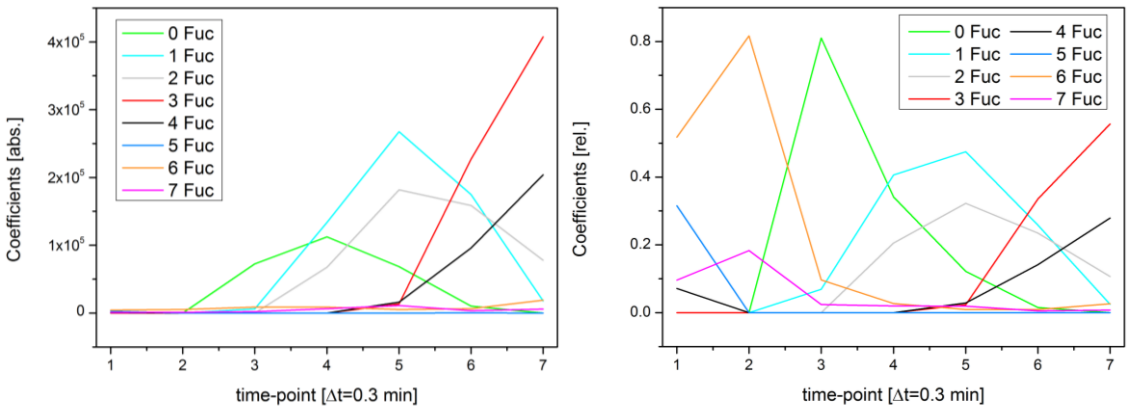


Figure 94: **Content of fucose residues attached to triantennary *N*-glycans with three terminal neuraminic acid as a function of the retention time in RP-LC. MS spectra were summed up over time (each 0.3 minutes). Highly fucosylated structures are most abundant and elute prior less fucosylated biantennae. The level of fully fucosylated triantennae is drastically decreased compared to the same glycan without a terminal NeuAc. Data is less reliable due to decreased S/N ratio and intensity in MS spectra. Left: absolute amounts, right: relative amounts.**

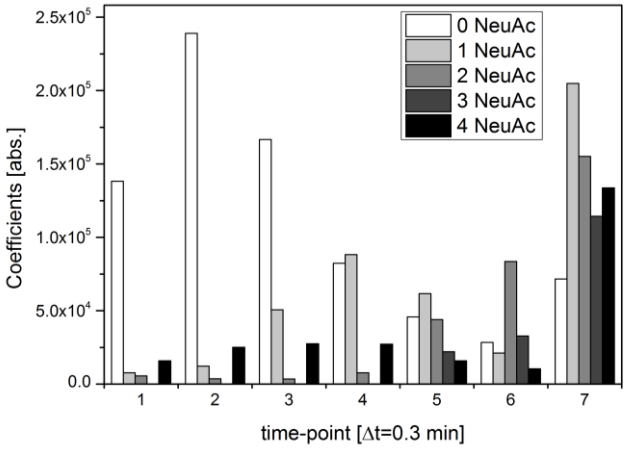


Figure 95: **Tetraantennary glycoforms of human salivary PIP eluting from RP column in terms of the content of terminal neuraminic acids. MS spectra were summed up over time (0.3 min per time-point). For each time-point all glycan amounts were summed up that contain no, one or two neuraminic acids, respectively. The majority of tetraantennary structures lacking neuraminic acids elutes in the beginning. With higher retention times glycan structures containing one to four neuraminic acids arise. The shifted elution time relies on an increased hydrophobicity of glycans containing NeuAc residues.**

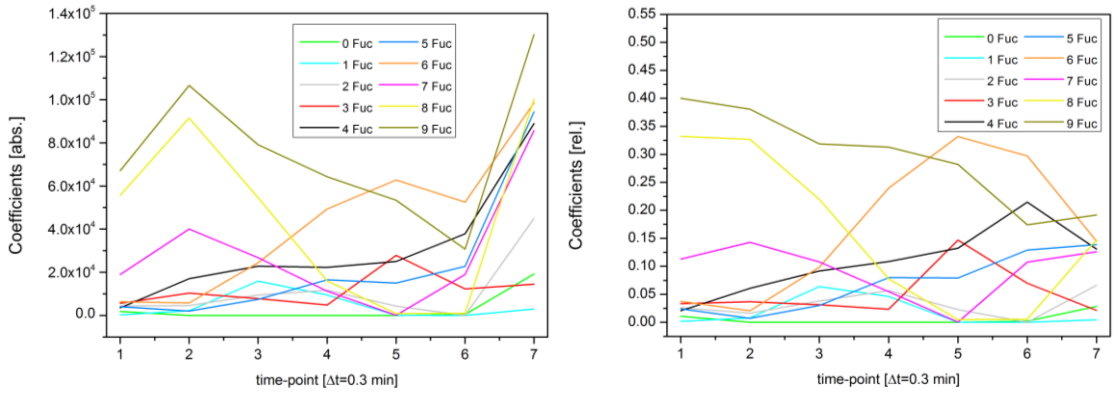


Figure 96: **Fucosylation degree as a function of NeuAc content of all tetraantennary *N*-glycans of human salivary PIP.**

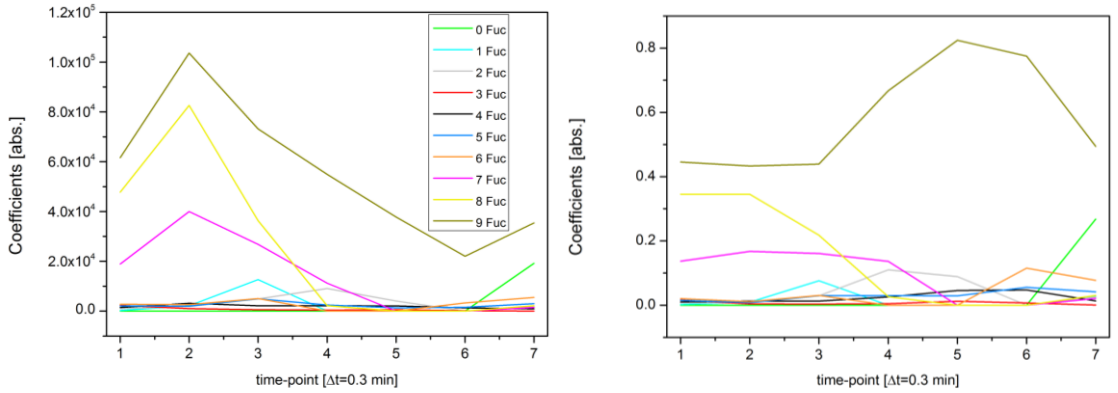


Figure 97: **Fucosylation degree as a function of NeuAc content of tetraantennary *N*-glycans lacking neuraminic acids of human salivary PIP.**

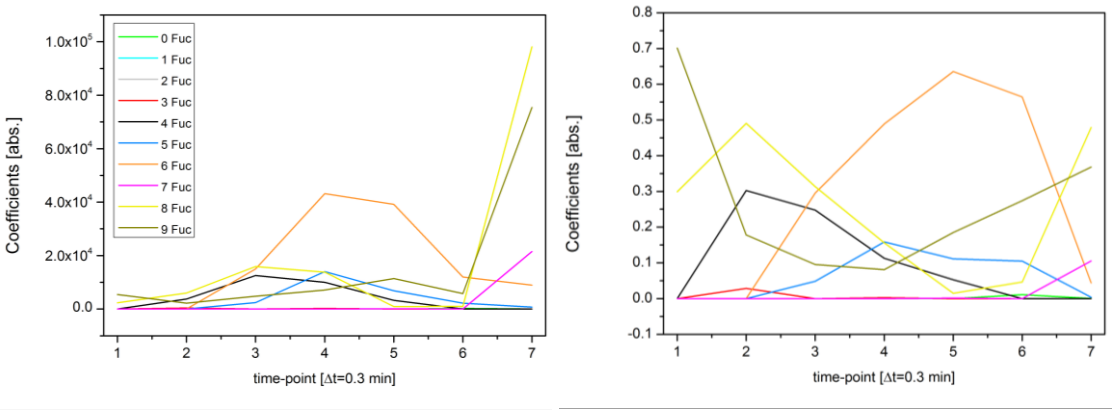


Figure 98: **Fucosylation degree as a function of NeuAc content of tetraantennary *N*-glycans with one terminal neuraminic acid of human salivary PIP.**

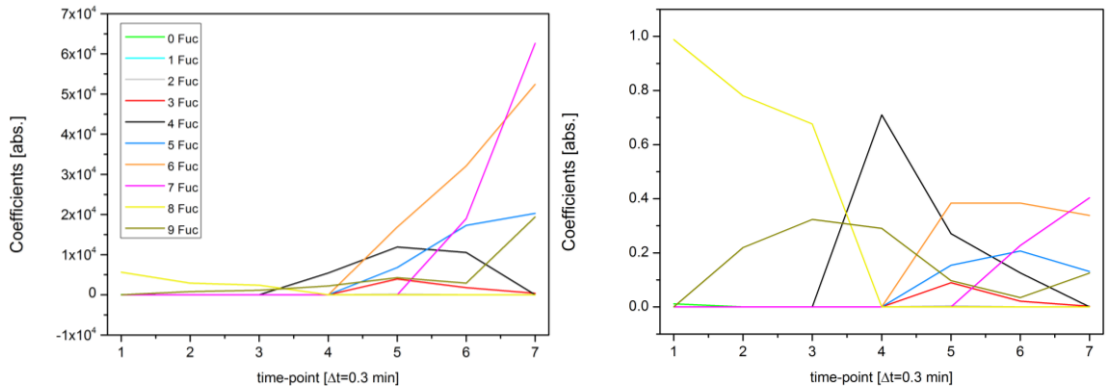


Figure 99: **Fucosylation degree as a function of NeuAc content of tetraantennary *N*-glycans with two terminal neuraminic acids of human salivary PIP.**

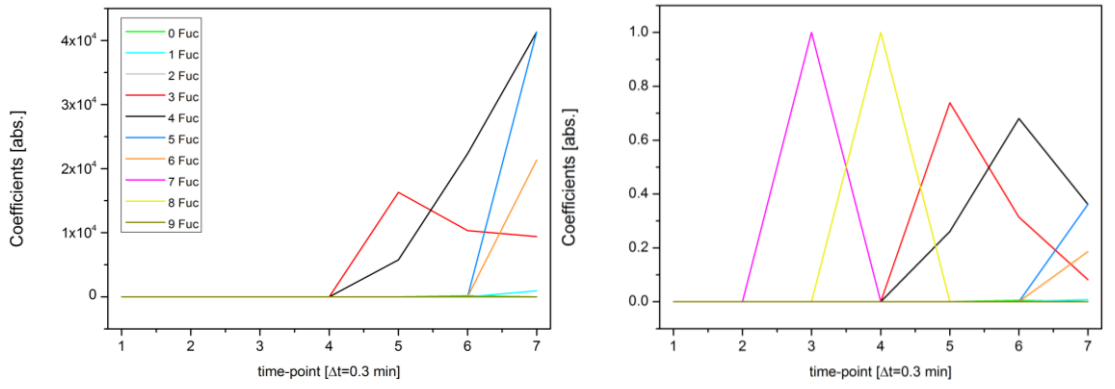


Figure 100: **Fucosylation degree as a function of NeuAc content of tetraantennary *N*-glycans with three terminal neuraminic acids of human, salivary PIP.**

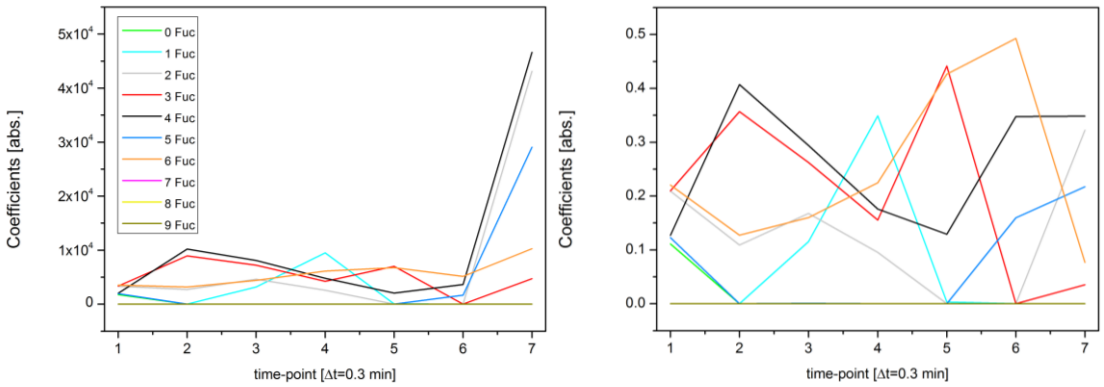


Figure 101: **Fucosylation degree as a function of NeuAc content of tetraantennary *N*-glycans with four terminal neuraminic acids of human salivary PIP.**

Table 22: **MATLAB coefficients for bi- or triantennae identified on salivary PIP.**

	SNP	**Male**		**Female**	
		bi	tri	bi	tri
SEC	n	(22.0±1.6)	(78.0±1.6)	(23.4±3.5)	(76.6±3.5)
SEC	y	(29.2±1.3)	(70.8±1.3)	(22.4±2.8)	(77.6±2.8)
AIEX	n	(25.5±2.2)	(74.5±2.2)	(29.1±2.8)	(70.9±2.8)
AIEX	y	(25.9±0.6)	(74.1±0.6)	(26.8±4.9)	(73.2±4.9)

Table 23: Sample ID for saliva donors. Saliva was enriched either by size exclusion- (SEC) or anion exchange (AIEX) chromatography. Several individuals were heterozygous for a SNP.

SampleID	SNP	Female	Male	Age [years]
SEC_f_1	y	x		21
SEC_f_2	y	x		27
SEC_m_1	y		x	34
SEC_m_2	y		x	28
SEC_m_3	y		x	27
SEC_m_4	y		x	26
SEC_f_3	n	x		28
SEC_f_4	n	x		22
SEC_f_5	n	x		29
SEC_m_5	n		x	26
SEC_m_6	n		x	28
SEC_m_7	n		x	27
AIEX_f_1	y	x		20
AIEX_f_2	y	x		22
AIEX_f_3	y	x		25
AIEX_f_4	y	x		26
AIEX_m_1	y		x	27
AIEX_f_5	n	x		22
AIEX_f_6	n	x		28
AIEX_f_7	n	x		26
AIEX_f_8	n	x		22
AIEX_f_9	n	x		53
AIEX_f_10	n	x		29
AIEX_f_11	n	x		29
AIEX_f_12	n	x		49
AIEX_m_3	n		x	27
AIEX_m_4	n		x	27
AIEX_m_5	n		x	27
AIEX_m_6	n		x	27

Table 24: Glycan compositions for automatic data interpretation of MS spectra of intact PIP. Coefficients can be found in the following tables.

Combination No.	Hex	HexNAc	dHex	NeuAc	NeuGc	Sulfate	Phosphate
PIP_1	5	4	0	0	0	0	0
PIP_2	5	4	1	0	0	0	0
PIP_3	5	4	2	0	0	0	0
PIP_4	5	4	3	0	0	0	0
PIP_5	5	4	4	0	0	0	0
PIP_6	5	4	5	0	0	0	0
PIP_7	5	4	0	1	0	0	0
PIP_8	5	4	1	1	0	0	0
PIP_9	5	4	2	1	0	0	0
PIP_10	5	4	3	1	0	0	0
PIP_11	5	4	4	1	0	0	0
PIP_12	5	4	5	1	0	0	0
PIP_13	5	4	0	2	0	0	0
PIP_14	5	4	1	2	0	0	0
PIP_15	5	4	2	2	0	0	0
PIP_16	5	4	3	2	0	0	0
PIP_17	5	4	4	2	0	0	0
PIP_18	5	4	5	2	0	0	0
PIP_19	6	5	0	0	0	0	0
PIP_20	6	5	1	0	0	0	0
PIP_21	6	5	2	0	0	0	0
PIP_22	6	5	3	0	0	0	0
PIP_23	6	5	4	0	0	0	0
PIP_24	6	5	5	0	0	0	0
PIP_25	6	5	6	0	0	0	0
PIP_26	6	5	7	0	0	0	0
PIP_27	6	5	0	1	0	0	0
PIP_28	6	5	1	1	0	0	0
PIP_29	6	5	2	1	0	0	0
PIP_30	6	5	3	1	0	0	0
PIP_31	6	5	4	1	0	0	0
PIP_32	6	5	5	1	0	0	0
PIP_33	6	5	6	1	0	0	0

Table 24: Continued.

Combination No.	Hex	HexNAc	dHex	NeuAc	NeuGc	Sulfate	Phosphate
PIP_34	6	5	7	1	0	0	0
PIP_35	6	5	0	2	0	0	0
PIP_36	6	5	1	2	0	0	0
PIP_37	6	5	2	2	0	0	0
PIP_38	6	5	3	2	0	0	0
PIP_39	6	5	4	2	0	0	0
PIP_40	6	5	5	2	0	0	0
PIP_41	6	5	6	2	0	0	0
PIP_42	6	5	7	2	0	0	0
PIP_43	6	5	0	3	0	0	0
PIP_44	6	5	1	3	0	0	0
PIP_45	6	5	2	3	0	0	0
PIP_46	6	5	3	3	0	0	0
PIP_47	6	5	4	3	0	0	0
PIP_48	6	5	5	3	0	0	0
PIP_49	6	5	6	3	0	0	0
PIP_50	6	5	7	3	0	0	0
PIP_51	7	6	0	0	0	0	0
PIP_52	7	6	1	0	0	0	0
PIP_53	7	6	2	0	0	0	0
PIP_54	7	6	3	0	0	0	0
PIP_55	7	6	4	0	0	0	0
PIP_56	7	6	5	0	0	0	0
PIP_57	7	6	6	0	0	0	0
PIP_58	7	6	7	0	0	0	0
PIP_59	7	6	8	0	0	0	0
PIP_60	7	6	9	0	0	0	0
PIP_61	7	6	0	1	0	0	0
PIP_62	7	6	1	1	0	0	0
PIP_63	7	6	2	1	0	0	0
PIP_64	7	6	3	1	0	0	0
PIP_65	7	6	4	1	0	0	0
PIP_66	7	6	5	1	0	0	0

Table 24: Continued.

Combination No.	Hex	HexNAc	dHex	NeuAc	NeuGc	Sulfate	Phosphate
PIP_67	7	6	6	1	0	0	0
PIP_68	7	6	7	1	0	0	0
PIP_69	7	6	8	1	0	0	0
PIP_70	7	6	9	1	0	0	0
PIP_71	7	6	0	2	0	0	0
PIP_72	7	6	1	2	0	0	0
PIP_73	7	6	2	2	0	0	0
PIP_74	7	6	3	2	0	0	0
PIP_75	7	6	4	2	0	0	0
PIP_76	7	6	5	2	0	0	0
PIP_77	7	6	6	2	0	0	0
PIP_78	7	6	7	2	0	0	0
PIP_79	7	6	8	2	0	0	0
PIP_80	7	6	9	2	0	0	0
PIP_81	7	6	0	3	0	0	0
PIP_82	7	6	1	3	0	0	0
PIP_83	7	6	2	3	0	0	0
PIP_84	7	6	3	3	0	0	0
PIP_85	7	6	4	3	0	0	0
PIP_86	7	6	5	3	0	0	0
PIP_87	7	6	6	3	0	0	0
PIP_88	7	6	7	3	0	0	0
PIP_89	7	6	8	3	0	0	0
PIP_90	7	6	9	3	0	0	0
PIP_91	7	6	0	4	0	0	0
PIP_92	7	6	1	4	0	0	0
PIP_93	7	6	2	4	0	0	0
PIP_94	7	6	3	4	0	0	0
PIP_95	7	6	4	4	0	0	0
PIP_96	7	6	5	4	0	0	0
PIP_97	7	6	6	4	0	0	0
PIP_98	7	6	7	4	0	0	0
PIP_99	7	6	8	4	0	0	0
PIP_100	7	6	9	4	0	0	0

Table 25: MATLAB coefficients for the overall content of neuraminic acids attached to biantennae of human PIP enriched by 2D-SEC-RP-LC followed by MS analysis.

n (NeuAc)	Male		Female	
	no SNP	SNP	no SNP	SNP
0	(76±11)	(52±10)	(79±3)	(83±6)
1	(17±7)	(35±8)	(16±2)	(11±7)
2	(8±4)	(13±5)	(6±1)	(5±1)

Table 26: MATLAB coefficients for the overall content of neuraminic acids attached to biantennae of human PIP enriched by 2D-AIEX-RP-LC followed by MS analysis.

n (NeuAc)	Male		Female	
	no SNP	SNP	no SNP	SNP
0	(76±6)	(49±7)	(82±3)	(59±9)
1	(14±3)	(35±4)	(12±3)	(18±5)
2	(10±2)	(16±11)	(6±1)	(23±12)

Table 27: MATLAB coefficients for the overall content of neuraminic acids attached to triantennae of human PIP enriched by 2D-SEC-RP-LC followed by MS analysis.

n (NeuAc)	Male		Female	
	no SNP	SNP	no SNP	SNP
0	(39±9)	(26±10)	(40±3)	(62±12)
1	(25±5)	(40±10)	(33±1)	(26±13)
2	(18±6)	(17±4)	(18±2)	(6±1)
3	(18±4)	(17±3)	(10±2)	(6±2)

Table 28: MATLAB coefficients for the overall content of neuraminic acids attached to triantennae of human PIP enriched by 2D-AIEX-RP-LC followed by MS analysis.

n (NeuAc)	Male		Female	
	no SNP	SNP	no SNP	SNP
0	(38±12)	(15±5)	(52±8)	(26±13)
1	(31±3)	(43±4)	(30±5)	(61±11)
2	(14±3)	(34±6)	(11±7)	(6±4)
3	(17±7)	(8±5)	(7±2)	(6±3)

Table 29: MATLAB coefficients for the *N*-glycans attached to biantennae of human PIP enriched by 2D-SEC-RP-LC followed by MS.

Glycan No.	Female - Biantennae		Male - Biantennae		Female - Biantennae - SNP		Male - Biantennae - SNP	
	Av. coeff, normalized	Error	Av. coeff., normalized	Error	Av. coeff, normalized	Error	Av. coeff, normalized	Error
PIP_SEC_b_1	0.011	0.003	0.016	0.002	0.033	0.009	0.028	0.013
PIP_SEC_b_2	0.012	0.003	0.015	0.004	0.021	0.009	0.027	0.011
PIP_SEC_b_3	0.041	0.004	0.045	0.005	0.026	0.017	0.023	0.020
PIP_SEC_b_4	0.106	0.013	0.115	0.020	0.133	0.007	0.140	0.072
PIP_SEC_b_5	0.213	0.022	0.242	0.061	0.290	0.034	0.143	0.053
PIP_SEC_b_6	0.402	0.041	0.322	0.058	0.329	0.056	0.162	0.082
PIP_SEC_b_7	0.000	0.000	0.000	0.000	0.017	0.006	0.025	0.007
PIP_SEC_b_8	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.011
PIP_SEC_b_9	0.034	0.006	0.029	0.029	0.000	0.000	0.059	0.036
PIP_SEC_b_10	0.061	0.031	0.058	0.058	0.052	0.052	0.189	0.068
PIP_SEC_b_11	0.048	0.012	0.059	0.022	0.036	0.008	0.037	0.026
PIP_SEC_b_12	0.016	0.005	0.022	0.003	0.009	0.008	0.013	0.006
PIP_SEC_b_13	0.000	0.000	0.000	0.000	0.005	0.005	0.026	0.022
PIP_SEC_b_14	0.000	0.000	0.006	0.006	0.019	0.019	0.030	0.021
PIP_SEC_b_15	0.037	0.005	0.043	0.024	0.014	0.004	0.042	0.015
PIP_SEC_b_16	0.005	0.003	0.020	0.020	0.001	0.001	0.015	0.011
PIP_SEC_b_17	0.001	0.001	0.001	0.000	0.001	0.000	0.002	0.001
PIP_SEC_b_18	0.012	0.005	0.006	0.003	0.013	0.010	0.015	0.011

Table 30: MATLAB coefficients for the *N*-glycans attached to biantennae of human PIP enriched by 2D-AIEX-RP-LC followed by MS.

Glycan No.	Female - Biantennae		Male - Biantennae		Female - Biantennae SNP		Male - Biantennae SNP	
	Av. coeff, normalized	Error	Av. coeff., normalized	Error	Av. coeff, normalized	Error	Av. coeff, normalized	Error
PIP_AIEX_b_1	0.015	0.003	0.037	0.009	0.039	0.001	0.035	0.005
PIP_AIEX_b_2	0.017	0.005	0.012	0.002	0.014	0.004	0.013	0.003
PIP_AIEX_b_3	0.050	0.017	0.028	0.006	0.017	0.005	0.031	0.026
PIP_AIEX_b_4	0.135	0.020	0.087	0.008	0.080	0.015	0.112	0.039
PIP_AIEX_b_5	0.231	0.023	0.257	0.014	0.159	0.021	0.072	0.070
PIP_AIEX_b_6	0.372	0.032	0.335	0.046	0.284	0.078	0.231	0.056
PIP_AIEX_b_7	0.000	0.000	0.000	0.000	0.004	0.002	0.005	0.005
PIP_AIEX_b_8	0.000	0.000	0.000	0.000	0.009	0.008	0.049	0.049
PIP_AIEX_b_9	0.038	0.013	0.000	0.000	0.063	0.034	0.125	0.010
PIP_AIEX_b_10	0.042	0.021	0.082	0.046	0.083	0.057	0.074	0.074
PIP_AIEX_b_11	0.030	0.006	0.031	0.018	0.006	0.006	0.050	0.023
PIP_AIEX_b_12	0.007	0.002	0.030	0.009	0.016	0.008	0.045	0.004
PIP_AIEX_b_13	0.004	0.004	0.000	0.000	0.065	0.030	0.015	0.014
PIP_AIEX_b_14	0.010	0.008	0.016	0.016	0.091	0.091	0.089	0.089
PIP_AIEX_b_15	0.028	0.010	0.073	0.031	0.040	0.005	0.038	0.002
PIP_AIEX_b_16	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001
PIP_AIEX_b_17	0.010	0.005	0.007	0.002	0.017	0.004	0.009	0.004
PIP_AIEX_b_18	0.012	0.006	0.004	0.002	0.017	0.002	0.007	0.001

Table 31: MATLAB coefficients for the *N*-glycans attached to triantennae of human PIP enriched by 2D-SEC-RP-LC followed by MS.

Glycan No.	Female - Triantennae		Male - Triantennae		Female - Triantennae SNP		Male - Triantennae SNP	
	Av. coeff, normalized	Error	Av. coeff, normalized	Error	Av. coeff, normalized	Error	Av. coeff, normalized	Error
PIP_SEC_t_1	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.001
PIP_SEC_t_2	0.000	0.000	0.000	0.000	0.001	0.000	0.003	0.002
PIP_SEC_t_3	0.008	0.003	0.009	0.002	0.004	0.001	0.002	0.002
PIP_SEC_t_4	0.029	0.005	0.026	0.004	0.019	0.008	0.012	0.001
PIP_SEC_t_5	0.040	0.001	0.043	0.007	0.055	0.011	0.041	0.017
PIP_SEC_t_6	0.055	0.002	0.070	0.024	0.107	0.032	0.045	0.022
PIP_SEC_t_7	0.122	0.010	0.119	0.032	0.199	0.057	0.091	0.035
PIP_SEC_t_8	0.143	0.021	0.128	0.037	0.233	0.031	0.062	0.044
PIP_SEC_t_9	0.000	0.000	0.000	0.000	0.008	0.000	0.003	0.002
PIP_SEC_t_10	0.000	0.000	0.000	0.000	0.011	0.011	0.011	0.008
PIP_SEC_t_11	0.013	0.004	0.007	0.003	0.012	0.012	0.010	0.006
PIP_SEC_t_12	0.059	0.004	0.040	0.022	0.036	0.036	0.055	0.032
PIP_SEC_t_13	0.077	0.006	0.046	0.024	0.044	0.044	0.103	0.037
PIP_SEC_t_14	0.111	0.010	0.063	0.032	0.063	0.063	0.128	0.050
PIP_SEC_t_15	0.066	0.003	0.080	0.019	0.081	0.013	0.087	0.050
PIP_SEC_t_16	0.000	0.000	0.013	0.010	0.007	0.001	0.009	0.009
PIP_SEC_t_17	0.000	0.000	0.001	0.001	0.006	0.006	0.000	0.000
PIP_SEC_t_18	0.007	0.003	0.009	0.008	0.011	0.011	0.024	0.018
PIP_SEC_t_19	0.018	0.001	0.011	0.007	0.000	0.000	0.006	0.006
PIP_SEC_t_20	0.046	0.012	0.037	0.019	0.000	0.000	0.014	0.014
PIP_SEC_t_21	0.083	0.008	0.059	0.032	0.032	0.032	0.072	0.042
PIP_SEC_t_22	0.014	0.007	0.027	0.016	0.007	0.003	0.029	0.012
PIP_SEC_t_23	0.010	0.005	0.032	0.005	0.006	0.006	0.024	0.016
PIP_SEC_t_24	0.001	0.000	0.001	0.000	0.001	0.000	0.002	0.001
PIP_SEC_t_25	0.000	0.000	0.011	0.011	0.020	0.020	0.033	0.018
PIP_SEC_t_26	0.000	0.000	0.027	0.024	0.016	0.016	0.038	0.014
PIP_SEC_t_27	0.013	0.002	0.022	0.009	0.004	0.004	0.024	0.009
PIP_SEC_t_28	0.043	0.007	0.065	0.011	0.000	0.000	0.026	0.011
PIP_SEC_t_29	0.040	0.007	0.049	0.022	0.014	0.014	0.038	0.007
PIP_SEC_t_30	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001
PIP_SEC_t_31	0.001	0.000	0.001	0.000	0.001	0.001	0.003	0.001
PIP_SEC_t_32	0.001	0.000	0.003	0.001	0.003	0.003	0.003	0.001

Table 32: MATLAB coefficients for the *N*-glycans attached to triantennae of PIP enriched by 2D-AIEX-RP-LC followed by MS.

Glycan No.	Female - Triantennae		Male - Triantennae		Female - Triantennae SNP		Male - Triantennae SNP	
	Av. coeff, normalized	Error	Av. coeff, normalized	Error	Av. coeff, normalized	Error	Av. coeff, normalized	Error
PIP_AIEX_t_1	0.002	0.001	0.004	0.002	0.004	0.001	0.003	0.001
PIP_AIEX_t_2	0.003	0.001	0.003	0.002	0.007	0.002	0.002	0.001
PIP_AIEX_t_3	0.016	0.005	0.035	0.017	0.006	0.003	0.011	0.007
PIP_AIEX_t_4	0.051	0.013	0.023	0.003	0.011	0.007	0.021	0.014
PIP_AIEX_t_5	0.056	0.017	0.031	0.006	0.028	0.007	0.019	0.001
PIP_AIEX_t_6	0.110	0.040	0.053	0.023	0.036	0.017	0.013	0.003
PIP_AIEX_t_7	0.153	0.031	0.112	0.033	0.076	0.041	0.050	0.030
PIP_AIEX_t_8	0.126	0.028	0.116	0.047	0.094	0.076	0.033	0.033
PIP_AIEX_t_9	0.003	0.002	0.000	0.000	0.006	0.004	0.000	0.000
PIP_AIEX_t_10	0.001	0.002	0.000	0.000	0.012	0.009	0.000	0.000
PIP_AIEX_t_11	0.005	0.004	0.002	0.002	0.013	0.005	0.018	0.018
PIP_AIEX_t_12	0.043	0.018	0.072	0.011	0.107	0.037	0.064	0.000
PIP_AIEX_t_13	0.058	0.026	0.031	0.012	0.147	0.028	0.094	0.006
PIP_AIEX_t_14	0.123	0.034	0.109	0.038	0.165	0.033	0.110	0.033
PIP_AIEX_t_15	0.045	0.006	0.073	0.023	0.086	0.035	0.061	0.023
PIP_AIEX_t_16	0.020	0.010	0.021	0.008	0.076	0.029	0.079	0.067
PIP_AIEX_t_17	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PIP_AIEX_t_18	0.001	0.003	0.000	0.000	0.006	0.006	0.028	0.028
PIP_AIEX_t_19	0.022	0.011	0.036	0.005	0.001	0.001	0.046	0.040
PIP_AIEX_t_20	0.008	0.006	0.011	0.007	0.022	0.022	0.067	0.018
PIP_AIEX_t_21	0.063	0.021	0.071	0.028	0.016	0.014	0.116	0.016
PIP_AIEX_t_22	0.001	0.001	0.001	0.001	0.002	0.002	0.037	0.037
PIP_AIEX_t_23	0.016	0.009	0.022	0.007	0.014	0.010	0.045	0.009
PIP_AIEX_t_24	0.001	0.001	0.000	0.000	0.001	0.001	0.000	0.000
PIP_AIEX_t_25	0.005	0.006	0.004	0.001	0.006	0.006	0.000	0.000
PIP_AIEX_t_26	0.001	0.001	0.008	0.008	0.009	0.009	0.000	0.000
PIP_AIEX_t_27	0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.000
PIP_AIEX_t_28	0.016	0.011	0.063	0.031	0.000	0.000	0.022	0.022
PIP_AIEX_t_29	0.027	0.018	0.067	0.034	0.016	0.009	0.048	0.026
PIP_AIEX_t_30	0.015	0.007	0.019	0.005	0.025	0.009	0.005	0.003
PIP_AIEX_t_31	0.004	0.002	0.010	0. 003	0.003	0.000	0.003	0.000
PIP_AIEX_t_32	0.002	0.001	0.003	0.001	0.003	0.001	0.002	0.000

Table 33: MATLAB coefficients for *N*-glycans attached to PIP from breast cancer patients.

Glycan No.	Pat. 1, coeff.	Pat. 1, coeff.	Pat. 1, coeff.	Pat. 1, coeff.
PIP_BC_1	0.12	0.00	0.04	0.01
PIP_BC_2	1.36	0.00	0.03	0.47
PIP_BC_3	0.00	0.05	1.48	0.10
PIP_BC_4	0.06	0.91	6.30	0.00
PIP_BC_5	9.23	1.11	7.84	1.63
PIP_BC_6	14.54	2.20	7.13	3.70
PIP_BC_7	8.78	6.56	8.93	5.52
PIP_BC_8	4.50	14.92	5.96	9.64
PIP_BC_9	3.53	0.02	0.00	0.12
PIP_BC_10	7,22	0.00	0.68	1.79
PIP_BC_11	1.76	0.38	2.61	0.59
PIP_BC_12	4.36	1.79	5.88	0.16
PIP_BC_13	4.48	4.37	7.16	3.10
PIP_BC_14	2.72	7.61	2.35	3.30
PIP_BC_15	3.73	4.22	1.13	9.55
PIP_BC_16	0.06	0.00	0.08	1.23
PIP_BC_17	2.40	0.19	0.00	0.90
PIP_BC_18	3.03	1.45	4.99	4.64
PIP_BC_19	3.49	2.55	5.23	0.65
PIP_BC_20	3.15	5.33	1.84	1.20
PIP_BC_21	0.00	4.07	2.14	0.00
PIP_BC_22	0.55	0.33	0.00	0.67
PIP_BC_23	2.02	0.47	0.54	0.68
PIP_BC_24	0.19	0.08	0.16	0.07
PIP_BC_25	0.00	0.00	0.00	3.02
PIP_BC_26	2.31	1.28	2.26	6.60
PIP_BC_27	2.55	3.03	1.29	0.00
PIP_BC_28	0.00	4.41	0.10	0.00
PIP_BC_29	0.29	2.36	0.01	0.33
PIP_BC_30	0.00	0.00	0.00	0.03
PIP_BC_31	0.28	0.13	0.37	0.05
PIP_BC_32	0.10	0.23	0.38	0.02
PIP_BC_33	0.27	0.18	0.15	0.07
PIP_BC_34	0.38	0.17	0.14	0.03
PIP_BC_35	0.00	0.00	1.19	0.07
PIP_BC_36	0.00	1.67	4.39	3.04
PIP_BC_37	3.65	4.08	3.48	6.03
PIP_BC_38	0.00	9.22	2.96	8.82
PIP_BC_39	0.83	0.05	0.17	0.06
PIP_BC_40	1.20	0.00	0.82	1.06
PIP_BC_41	1.07	2.22	2.17	2.70
PIP_BC_42	3.18	5.76	4.65	3.12
PIP_BC_43	1.41	1.21	0.47	5.25
PIP_BC_44	0.27	0.03	0.24	0.94
PIP_BC_45	0.00	0.47	0.67	2.17
PIP_BC_46	0.00	1.35	0.52	4.99
PIP_BC_47	0.00	2.78	0.53	1.49
PIP_BC_48	0.21	0.60	0.00	0.00
PIP_BC_49	0.31	0.07	0.13	0.08
PIP_BC_50	0.38	0.11	0.39	0.32

9 REFERENCES

1. Costerton, J. W.; Irvin, R. T., The bacterial glycocalyx in nature and disease. *Ann. Rev. Microbiol.* **1981,** *35*, 299-324.

2. Costerton, J. W.; Irvin, R. T.; Cheng, K. J., The role of bacterial surface structures in pathogenesis. *Crit. Rev. Microbiol.* **1981,** *8* (4), 303-338.

3. Martín, R.; Escobedo, S.; Martín, C.; Crespo, A.; Quiros, L. M.; Suarez, J. E., Surface glycosaminoglycans protect eukaryotic cells against membrane-driven peptide bacteriocins. *Antimicrob. Agents Chemother.* **2014,** *59*, 677-681.

4. Reuter, G.; Gabius, H.-J., Eukaryotic glycosylation: whim of nature or multipurpose tool? *Cell. Mol. Life Sci.* **1999,** *55*, 368-422.

5. Solera, M.; Desplat-Jegoa, S.; Vacherc, B.; Ponsonnetc, L.; Fraternob, M.; Bongranda, P.; Martinc, J.-M.; Foa, C., Adhesion-related glycocalyx study: quantitative approach with imaging-spectrum in the energy ¢ltering transmission electron microscope (EFTEM). *FEBS Lett.* **1998,** *429*, 89-94.

6. Yuki, N., Ganglioside mimicry and peripheral nerve disease. *Muscle Nerve* **2007,** *35* (6), 691-711.

7. Lis, H.; Sharon, N., Protein glycosylation - structural and functional aspects. *Eur. J. Biochem.* **1993,** *218*, 1-27.

8. Berger, E. G.; Buddecke, E.; Kamerling, J. P.; Kobata, A.; Paulson, J. C.; Vliegenthart, J. F. G., Structure, biosynthesis and functions of glycoprotein glycans *Experientia* **1982,** *38* (10), 1129-1258.

9. Spillmann, D.; Burger, M. M., Carbohydrate-carbohydrate interactions in adhesion. **1996,** *61* (J. Cell. Biochem.), 562-568.

10. Ohtsubo, K.; Marth, J. D., Glycosylation in cellular mechanisms of health and disease. *Cell* **2006,** *126* (5), 855-67.

11. Freeze, H. H., Human disorders in N-glycosylation and animal models. *Biochim. Biophys. Acta* **2002,** *1573*, 388-393.

12. Chung, C. H.; Mirakhur, B.; Chan, E.; Le, Q.-T.; Berlin, J.; Morse, M.; Murphy, B. A.; Satinover, S. M.; Hosen, J.; Mauro, D.; Slebos, R. J.; Zhou, Q.; Gold, D.; Hatley, T.; Hicklin, D. J.; Platts-Mills, T. A. E., Cetuximab-induced anaphylaxis and IgE specific for galactose-a-1,3-galactose. *N. Engl. J. Med.* **2008,** *358*, 1109-1117.

13. Dell, A.; Morris, H. R., Glycoprotein structure determination by mass spectrometry. *Science* **2001,** *291* (5512), 2351-2356.

14. Wuhrer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H., Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatogr. B* **2007,** *849*, 115-28.

15. Harvey, D. J., Glycoproteomics based on tandem mass spectrometry of glycopeptides. *Expert Rev. Proteomics* **2005,** *2* (1), 87-101.

16. Ceroni, A.; Maass, K.; Geyer, H.; Geyer, R.; Dell, A.; Haslam, S. M., GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J. Proteome Res.* **2007,** *7*, 1650-1659.

17. Vliegenthart, J. F. G.; Dorland, L.; van Halbeek, H., High-resolution, ^{1}H-nuclear magnetic resonance spectroscopy as a tool in the structural analysis of carbohydrates related to glycoproteins *Adv. Carbohydr. Chem. Biochem.* **1983,** *37*, 157-223.

18. Herscovics, A., Importance of glycosidases in mammalian glycoprotein biosynthesis. *Biochim. Biophys. Acta* **1999,** *1473*, 96-107.

19. Kornfeld, R.; Kornfeld, S., Assembly of asparagine linked oligosaccharides. *Ann. Rev. Biochem.* **1985,** *54*, 631-664.

20. Waechter, C. J.; Lennarz, W. J., The role of polyprenol-linked sugars in glycoprotein snthesis. *Ann. Rev. Biochem.* **1976,** *45*, 95-112.

21. Aronson, N. N.; Kuranda, M. J., Lysosomal degradation of Asn-linked glycoproteins. *FASEB* **1989,** *3* (14), 2615-2622.

22. Kam, R. K. T.; Poon, T. C. W., The potentials of glycomics in biomarker discovery. *Clin. Proteom.* **2008,** *4*, 67-79.

23. Rayon, C.; Lerouge, P.; Faye, L., The protein N-glycosylation in plants. *J. Exp. Bot.* **1998,** *49* (326).

24. Laine, R. A., A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 x 10^12 structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology* **1994,** *4* (6), 759-767.

25. Werz, D. B.; Ranzinger, R.; Herget, S.; Adibekian, A.; von der Lieth, C.-W.; Seeberger, P. H., Exploring the structural diversity of mammalian carbohydrates ("Glycospace") by statistical databank analysis. *ACS Chem. Biol.* **2007,** *2* (10), 685-691.

26. Cummings, R. D., The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **2009,** *5* (10), 1087-1104.

27. Arnold, J. N.; Wormald, M. R.; Sim, R. B.; Rudd, P. M.; Dwek, R. A., The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu. Rev. Immunol.* **2007,** *25*, 21-50.

28. Mimura, Y.; Church, S.; Ghirlando, R.; Ashton, P. R.; Dong, S.; Goodall, M.; Lund, J.; Jefferis, R., The influence of glycosylation on the thermal stability and effector function expression of human IgG1-Fc: properties of a series of truncated glycoforms. *Mol. Immunol.* **2000,** *37*, 697-706.

29. Kleizen, B.; Braakman, I., Protein folding and quality control in the endoplasmic reticulum. *Curr. Opin. Cell. Biol.* **2004,** *16* (4), 343-349.

30. Tifft, C. J.; Proia, R. L.; Camerini-Otero, R. D., The folding and cell surface expression of CD4 requires glycosylation. *J. Biol. Chem.* **1992,** *267*, 3268-3273.

31. Roth, J.; Zuber, C.; Park, S.; Jang, I.; Lee, Y.; Kysela, K. G.; Le Fourn, V.; Santimaria, R.; Guhl, B.; Cho, J. W., Protein N-glycosylation, protein folding, and protein quality control. *Mol. Cells* **2010,** *30* (6), 497-506.

32. Sharon, N.; Lis, H., Lectins: cell-agglutinating and sugar-specific proteins *Science* **1972,** *177* (4053), 949-959.

33. Stevens, J.; Blixt, O.; Glaser, L.; Taubenberger, J. K.; Palese, P.; Paulson, J. C.; Wilson, I. A., Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J. Mol. Biol.* **2006,** *355* (5), 1143-1155.

34. Vigerust, D. J.; Shepherd, V. L., Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol.* **2007,** *15* (5), 211-218.

35. Szymanski, C. M.; Wren, B. W., Protein glycosylation in bacterial mucosal pathogens. *Nat. Rev. Microbiol.* **2005,** *3* (3), 225-237.

36. Comstock, L. E.; Kasper, D. L., Bacterial glycans: key mediators of diverse host immune responses. *Cell* **2006,** *126* (5), 847-850.

37. Olofsson, S.; Hansen, J.-E. S., Host cell glycosylation of viral glycoproteins - a battlefield for host defence and viral resistance. *Scand. J. Infect. Dis.* **2009,** *30* (5), 435-440.

38. Nelson, J.; Couceiro, S. S.; Paulson, J. C.; Baum, L. G., Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity *Virus Res.* **1993,** *29*, 155-165

39. Russell, R. J.; Haire, L. F.; Stevens, D. J.; Collins, P. J.; Lin, Y. P.; Blackburn, G. M.; Hay, A. J.; Gamblin, S. J.; Skehel, J. J., The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* **2006,** *443* (7107), 45-49.

40. Hayden, F. G.; Osterhaus, A. D. M. E.; Treanor, J. J.; Fleming, D. M.; Aoki, F. M.; Nicholson, K. G.; Bohnen, A. M.; Hirst, H. M.; Keene, O.; Wightman, K., Efficacy and safety of the neuraminidase inhibitor zanamivir in the treatment of influenzavirus infections *N. Engl. J. Med.* **2008,** *337* (13), 874-880.

41. Eklund, E. A.; Freeze, H. H., The congenital disorders of glycosylation: a multifaceted group of syndromes. *NeuroRx* **2006,** *3* (2), 254-263.

42. Schnaar, R. L.; Gerardy-Schahn, R.; Hildebrandt, H., Sialic acids in the brain: gangliosides and polysialic acid in nervous system development, stability, disease, and regeneration. *Physiol. Rev.* **2014,** *94* (2), 461-518.

43. Wuhrer, M., Glycosylation profiling in clinical proteomics: heading for glycan biomarkers. *Expert Rev. Proteomics* **2007,** *4* (2), 135-136.

44. Durand, G.; Seta, N., Protein glycosylation and diseases: blood and urinary oligosaccharides as markers for diagnosis and therapeutic monitoring. *Clin. Chem.* **2000,** *46* (6), 795-805.

45. Yamashita, K.; Ideo, H.; Ohkura, T.; Fukushima, K.; Yuasa, I.; Ohnoll, K.; Takeshita, K., Sugar chains of serum transferrin from patients with carbohydrate deficient glycoprotein syndrome. *J. Biol. Chem.* **1993,** *268* (8), 5783-5789.

46. Brawer, M. K., Prostate-specific antigen. *Semin. Surg. Oncol.* **2000,** *18*, 3-9.

47. Miyoshi, E.; Moriwaki, K.; Nakagawa, T., Biological function of fucosylation in cancer biology. *J. Biochem.* **2008,** *143* (6), 725-729.

48. Reis, C. A.; Osorio, H.; Silva, L.; Gomes, C.; David, L., Alterations in glycosylation as biomarkers for cancer detection. *J. Clin. Pathol.* **2010,** *63* (4), 322-329.

49. Leiserowitz, G. S.; Lebrilla, C.; Miyamoto, S.; An, H. J.; Duong, H.; Kirmiz, C.; Li, B.; Liu, H.; Lam, K. S., Glycomics analysis of serum: a potential new biomarker for ovarian cancer? *Int. J. Gynecol. Cancer* **2008,** *18* (3), 470-475.

50. Yao, M.; Zhou, D.-P.; Jiang, S.-M.; Wang, Q.-H.; Zhou, X.-D.; Tang, Z.-Y.; Gu, J.-X., Elevated activity of N-acetylglucosaminyltransferase V in human hepatocellular carcinoma. *J. Cancer Res. Clin. Oncol.* **1998,** *124*, 27-30.

51. Ugorski, M.; Laskowska, A., Sialyl Lewis-a: a tumor-associated carbohydrate antigen involved in adhesion and metastatic potential of cancer cells. *Acta Biochim. Pol.* **2002,** *49* (2), 303-311.

52. Nakayama, T.; Wafanabe, M.; Kafsumafa, T.; Teramofo, T.; Kitajima, M., Expression of sialyl lewis-a as a new prognostic factor for patients with advanced colorectal carcinoma. *Cancer* **1995,** *75* (8), 2051-2056.

53. Brockhausen, I., Pathways of O-glycan biosynthesis in cancer cells. *Biochim. Biophys. Acta* **1999,** *1473*, 67-95.

54. Dennis, J. W.; Laferte, S.; Waghorne, C.; Breitman, M. L.; Kerbel, R. S., Beta1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis. *Science* **1987,** *236* (4801), 582-585.

55. Stowell, S. R.; Ju, T.; Cummings, R. D., Protein glycosylation in cancer. *Annu. Rev. Pathol.* **2015,** *10*, 473-510.

56. Kawasaki, N.; Itoh, S.; Hashi, N.; Takakura, D.; Qin, Y.; Huang, X.; Yamaguchi, T., The significance of glycosylation analysis in development of biopharmaceuticals. *Biol. Pharm. Bull.* **2009,** *32* (5), 796-800.

57. Beck, A.; Reichert, J. M., Approval of the first biosimilar antibodies in Europe: a major landmark for the biopharmaceutical industry. *MAbs* **2013,** *5* (5), 621-623.

58. Higel, F.; Seidl, A.; Sorgel, F.; Friess, W., N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *Eur. J. Pharm. Biopharm.* **2016,** *100*, 94-100.

59. Jefferis, R., Glycosylation as a strategy to improve antibody-based therapeutics. *Nat. Rev. Drug Discov.* **2009,** *8* (3), 226-234.

60. Reichert, J. M., Antibodies to watch in 2010. *mAbs* **2014,** *2* (1), 84-100.

61. Apostolovic, D.; Tran, T. A.; Sanchez-Vidaurre, S.; Cirkovic Velickovic, T.; Starkhammar, M.; Hamsten, C.; van Hage, M., Red meat allergic patients have a selective IgE response to the alpha-Gal glycan. *Allergy* **2015,** *70* (11), 1497-1500.

62. Commins, S. P.; Platts-Mills, T. A., Anaphylaxis syndromes related to a new mammalian cross-reactive carbohydrate determinant. *J. Allergy Clin. Immunol.* **2009,** *124* (4), 652-657.

63. Noguchi, A.; Mukuria, C. J.; Suzuki, E.; Naiki, M., Immunogenicity of N-glycolylneuraminic acid-containing carbohydrate chains of recombinant human erythropoietin expressed in chinese hamster ovary cells. *J. Biochem.* **1995,** *117*, 59-62.

64. Shields, R. L.; Lai, J.; Keck, R.; O'Connell, L. Y.; Hong, K.; Meng, Y. G.; Weikert, S. H.; Presta, L. G., Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcgamma RIII and antibody-dependent cellular toxicity. *J. Biol. Chem.* **2002,** *277* (30), 26733-36740.

65. Shinkawa, T.; Nakamura, K.; Yamane, N.; Shoji-Hosaka, E.; Kanda, Y.; Sakurada, M.; Uchida, K.; Anazawa, H.; Satoh, M.; Yamasaki, M.; Hanai, N.; Shitara, K., 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.* **2003,** *278* (5), 3466-3473.

66. Ucakturk, E., Analysis of glycoforms on the glycosylation site and the glycans in monoclonal antibody biopharmaceuticals. *J. Sep. Sci.* **2012,** *35* (3), 341-350.

67. van Halbeek, H.; Dorland, L.; Vliegenthart, J. F. G.; Montreuil, J.; Fournet, B.; Schmid, K., Characterization of the microheterogeneity in glycoproteins by 500-MHz 1-H-NMR spectroscopy of glycopeptide preparations *J. Biol. Chem. 256* (11), 5588-5590.

68. Rudd, P. M.; Guile, G. R.; Küster, B.; Harvey, D. J.; Opdenakker, G.; Dwek, R. A., Oligosaccharide sequencing technology. *Nature* **1997,** *388*, 205-207.

69. Zauner, G.; Deelder, A. M.; Wuhrer, M., Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics. *Electrophoresis* **2011,** *32* (24), 3456-3466.

70. Zhang, X.; Fang, A.; Riley, C. P.; Wang, M.; Regnier, F. E.; Buck, C., Multi-dimensional liquid chromatography in proteomics--a review. *Anal. Chim. Acta* **2010,** *664* (2), 101-113.

71. Davies, M. J.; Smith, K. D.; Carruthers, R. A.; Chai, W.; Lawson, A. M.; Hounsell, E. F., Use of a porous graphitised carbon column for the high-performance liquid chromatography of oligosaccharides, alditols and glycopeptides with subsequent mass spectrometry analysis *Chrom.* **1993,** *646* (2), 317-326.

72. Kailemia, M. J.; Ruhaak, L. R.; Lebrilla, C. B.; Amster, I. J., Oligosaccharide analysis by mass spectrometry: a review of recent developments. *Anal. Chem.* **2014,** *86* (1), 196-212.

73. Scigelova, M.; Makarov, A., Orbitrap mass analyzer - overview and applications in proteomics. *Proteomics* **2006,** *6 Suppl 2*, 16-21.

74. Fellenberg, M.; Coksezen, A.; Meyer, B., Characterization of picomole amounts of oligosaccharides from glycoproteins by 1H NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **2010,** *49* (14), 2630-2633.

75. Raman, R.; Raguram, S.; Venkataraman, G.; Paulson, J. C.; Sasisekharan, R., Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat Methods* **2005,** *2* (11), 817-824.

76. Pilobello, K. T.; Krishnamoorthy, L.; Slawek, D.; Mahal, L. K., Development of a lectin microarray for the rapid analysis of protein glycopatterns. *ChemBioChem* **2005,** *6* (6), 985-689.

77. Hopfgartner, G.; Varesio, E.; Tschappat, V.; Grivet, C.; Bourgogne, E.; Leuthold, L. A., Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules. *J. Mass Spectrom.* **2004,** *39* (8), 845-855.

78. Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A., An introduction to quadrupole-time-of-flight mass spectrometry. *J. Mass Spectrom.* **2001,** *36* (8), 849-865.

79. Yost, R. A.; Enke, C. G., Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Anal. Chem.* **1991,** *51* (12), 1251-1264.

80. Hunt, D. F.; Yates III, J. R.; Shabanowitz, J.; Winston, S.; Hauer, C. R., Protein sequencing by tandem mass spectrometry. *PNAS* **1986,** *83*, 6233-6237.

81. Behnken, H. N.; Fellenberg, M.; Koetzler, M. P.; Jirmann, R.; Nagel, T.; Meyer, B., Resolving the problem of chromatographic overlap by 3D cross correlation (3DCC) processing of LC, MS and NMR data for characterization of complex glycan mixtures. *Anal. Bioanal. Chem.* **2012,** *404* (5), 1427-1437.

82. Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W., Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J. Am. Chem. Soc.* **1999,** *121*, 806-812.

83. Fridriksson, E. K.; Beavil, A.; Holowka, D.; Gould, H. J.; Baird, B.; McLafferty, F. W., Heterogeneous glycosylation of Immunoglobulin E constructs characterized by top-down high-resolution 2-D mass spectrometry. *Biochemistry* **2000,** *39*, 3369-3376.

84. Kelleher, N. L., Top down proteomics. *Anal. Chem.* **2004,** *76* (11), 196–203.

85. Berger, E. G.; Buddecke, E.; Kamerling, J. P.; Kobata, A.; Paulson, J. C., Structure, biosynthesis and functions of glycoprotein glycans. *Experientia* **1982,** *38* (10), 1129-1258.

86. Rademacher, T. W.; Parekh, R. B.; Dwek, R. A., Glycobiology. *Ann. Rev. Biochem.* **1988,** *57*, 785-838.

87. Beck, A.; Wurch, T.; Bailly, C.; Corvaia, N., Strategies and challenges for the next generation of therapeutic antibodies. *Nat. Rev. Immunol.* **2010,** *10* (5), 345-352.

88. Davis, J.; Deng, R.; Boswell, C.; Zhang, Y.; Li, J.; Fielder, P.; Joshi, A.; Kenkare-Mitra, S., *Pharmaceutical Biotechnology*. 4 ed.; Springer: New York Heidelberg Dordrecht London, 2013.

89. Reichert, J. M., Marketed therapeutic antibodies compendium. *MAbs* **2012,** *4* (3), 413-415.

90. Ledford, H., 'Biosimilar' drugs poised to penetrate market. *Nature* **2010,** *468*, 18-19.

91. Ayoub, D.; Jabs, W.; Resemann, A.; Evers, W.; Evans, C.; Main, L.; Baessmann, C.; Wagner-Rousset, E.; Suckau, D.; Beck, A., Correct primary structure assessment and extensive glyco-profiling of cetuximab by a combination of intact, middle-up, middle-down and bottom-up ESI and MALDI mass spectrometry techniques. *MAbs* **2013,** *5* (5), 699-710.

92. Chan, A. C.; Carter, P. J., Therapeutic antibodies for autoimmunity and inflammation. *Nat. Rev. Immunol.* **2010,** *10* (5), 301-316.

93. Cunningham, D.; Humblet, Y.; Siena, S.; Khayat, D.; Bleiberg, H.; Santoro, A.; Bets, D.; Mueser, M.; Harstrick, A.; Verslype, C.; Chau, I.; Van Cutsem, E., Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N. Engl. J. Med.* **2004,** *351*, 337-345.

94. Humblet, Y., Cetuximab: an IgG(1) monoclonal antibody for the treatment of epidermal growth factor receptor-expressing tumours. *Expert Opin. Pharmacother.* **2004,** *5* (7), 1621-1633.

95. Qian, J.; Liu, T.; Yang, L.; Daus, A.; Crowley, R.; Zhou, Q., Structural characterization of N-linked oligosaccharides on monoclonal antibody cetuximab by the combination of orthogonal matrix-assisted laser desorption/ionization hybrid quadrupole-quadrupole time-of-flight tandem mass spectrometry and sequential enzymatic digestion. *Anal. Biochem.* **2007,** *364* (1), 8-18.

96. Fridriksson, E. K. B., A.; Holowka, D.; Gould, H. J.; Baird, B.; McLafferty, F. W., Heterogeneous Glycosylation of Immunoglobulin E Constructs Characterized by Top-Down High-Resolution 2-D Mass Spectrometry. *Biochemistry* **2000,** *39* (12), 3369-3376.

97. König, S.; Leary, J. A., Evidence for linkage position determination in cobalt coordinated pentasaccharides using ion trap mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1998,** *9*, 1125-1134.

98. Viseux, N.; de Hoffmann, E.; Domon, B., Structural assignment of permethylated oligosaccharide subunits using sequential tandem mass spectrometry. *Anal. Chem.* **1998,** *70*, 4951-4959.

99. Weiskopf, A. S.; Vouros, P.; Harvey, D. J., Electrospray ionization-ion trap mass spectrometry for structural analysis of complex N-linked glycoprotein oligosaccharides. *Anal. Chem.* **1998,** *70*, 4441-4447.

100. Domon, B.; Costello, C. E., A Systematic nomenclature for carbohydrate fragmentationsin FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate J.* **1988,** *5*, 397-409.

101. Harvey, D. J., Structural determination of N-linked glycans by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry. *Proteomics* **2005,** *5* (7), 1774-1786.

102. Fellenberg, M.; Behnken, H. N.; Nagel, T.; Wiegandt, A.; Baerenfaenger, M.; Meyer, B., Glycan analysis: scope and limitations of different techniques--a case for integrated use of LC-MS(/MS) and NMR techniques. *Anal. Bioanal. Chem.* **2013,** *405* (23), 7291-7305.

103. Sheeley, D. M.; Reinhold, V. N., Structural characterization of carbohydrate sequence, linkage, and branching in a quadrupole ion trap mass spectrometer: neutral oligosaccharides and N-linked glycans. *Anal. Chem.* **1998,** *70*, 3053-3059.

104. Han, H.; Jin, J.; Breuker, K.; McLafferty, F. W., Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science* **2006,** *314*, 109-113.

105. Behnken, H. N., New methods for the characterization of N-type glycans of glycoproteins: Integration of LC-MS, MS/MS and NMR spectroscopy. In *Dissertation*, University Hamburg, 2013.

106. Behnken, H. N.; Ruthenbeck, A.; Schulz, J. M.; Meyer, B., Glycan analysis of prostate specific antigen (PSA) directly from the intact glycoprotein by HR-ESI/TOF-MS. *J. Proteome Res.* **2014,** *13* (2), 997-1001.

107. Ferrara, N.; Hillan, K. J.; Gerber, H. P.; Novotny, W., Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nat. Rev. Drug Discov.* **2004,** *3* (5), 391-400.

108. Saltz, L. B.; Clarke, S.; Diaz-Rubio, E.; Scheithauer, W.; Figer, A.; Wong, R.; Koski, S.; Lichinitser, M.; Yang, T. S.; Rivera, F.; Couture, F.; Sirzen, F.; Cassidy, J., Bevacizumab in combination with oxaliplatin-based chemotherapy as first-line therapy in metastatic colorectal cancer: a randomized phase III study. *J. Clin. Oncol.* **2008,** *26* (12), 2013-2019.

109. Willett, C. G.; Boucher, Y.; di Tomaso, E.; Duda, D. G.; Munn, L. L.; Tong, R. T.; Chung, D. C.; Sahani, D. V.; Kalva, S. P.; Kozin, S. V.; Mino, M.; Cohen, K. S.; Scadden, D. T.; Hartford, A. C.; Fischman, A. J.; Clark, J. W.; Ryan, D. P.; Zhu, A. X.; Blaszkowsky, L. S.; Chen, H. X.; Shellito, P. C.; Lauwers, G. Y.; Jain, R. K., Direct evidence that the VEGF-specific antibody bevacizumab has antivascular effects in human rectal cancer. *Nat. Med.* **2004,** *10* (2), 145-147.

110. Giantonio, B. J.; Catalano, P. J.; Meropol, N. J.; O'Dwyer, P. J.; Mitchell, E. P.; Alberts, S. R.; Schwartz, M. A.; Benson, A. B., 3rd, Bevacizumab in combination with oxaliplatin, fluorouracil, and leucovorin (FOLFOX4) for previously treated metastatic colorectal cancer: results from the Eastern Cooperative Oncology Group Study E3200. *J. Clin. Oncol.* **2007,** *25* (12), 1539-1544.

111. Escudier, B.; Pluzanska, A.; Koralewski, P.; Ravaud, A.; Bracarda, S.; Szczylik, C.; Chevreau, C.; Filipek, M.; Melichar, B.; Bajetta, E.; Gorbunova, V.; Bay, J.-O.; Bodrogi, I.; Jagiello-Gruszfeld, A.; Moore, N., Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. *The Lancet* **2007,** *370* (9605), 2103-2111.

112. Beck, A.; Wagner-Rousset, E.; Ayoub, D.; Van Dorsselaer, A.; Sanglier-Cianferani, S., Characterization of therapeutic antibodies and related products. *Anal Chem* **2013,** *85* (2), 715-36.

113. Jones, A. L.; Hulett, M. D.; Parish, C. R., Histidine-rich glycoprotein: A novel adaptor protein in plasma that modulates the immune, vascular and coagulation systems. *Immunol Cell Biol* **2005,** *83* (2), 106-18.

114. Poon, I. K.; Patel, K. K.; Davis, D. S.; Parish, C. R.; Hulett, M. D., Histidine-rich glycoprotein: the Swiss Army knife of mammalian plasma. *Blood* **2011,** *117* (7), 2093-101.

115. Sorensen, C. B.; Krogh-Pedersen, H.; Petersen, T. E., Determination of the disulphide bridge arrangement of bovine histidine-rich glycoprotein. *FEBS Lett* **1993,** *328* (3), 285-90.

116. Morgan, W. T., Interactions of the histidine-rich glycoprotein of serum with metals. *Biochemistry* **1981,** *20* (5), 1054-61.

117. Morgan, W. T., The histidine-rich glycoprotein of serum has a domain rich in histidine, proline, and glycine that binds heme and metals. *Biochemistry* **1985,** *24* (6), 1496-501.

118. Gorgani, N. N.; Smith, B. A.; Kono, D. H.; Theofilopoulos, A. N., Histidine-rich glycoprotein binds to DNA and Fc gamma RI and potentiates the ingestion of apoptotic cells by macrophages. *J Immunol* **2002,** *169* (9), 4745-51.

119. Nordqvist, S.; Kårehed, K.; Skoog Svanberg, A.; Menezes, J.; Åkerud, H., Ovarian response is affected by a specific histidine-rich glycoprotein polymorphism: a preliminary study. *Reprod. Biomed. Online* **2015,** *30*, 74-81.

120. Wang, Y.-S.; Cao, R.; Jin, H.; Huang, Y.-P.; Zhang, X.-Y.; Cong, Q.; He, Y.-F.; Xu, C.-J., Altered protein expression in serum from endometrial hyperplasia and carcinoma patients. *J. Hematol. & Oncol.* **2011,** *4* (15), 1-8.

121. Wu, J.; Xie, X.; Liu, Y.; He, J.; Benitez, R.; Buckanovich, R. J.; Lubman, D. M., Identification and confirmation of differentially expressed fucosylated glycoproteins in the serum of ovarian cancer patients using a lectin array and LC-MS/MS. *J. Proteome Res.* **2012,** *11* (9), 4541-4552.

122. He, X.; Wang, Y.; Zhang, W.; Li, H.; Luo, R.; Zhou, Y.; Li, C.; Liao, M.; Huang, H.; Lv, X.; Xie, Z.; He, M., Screening differential expression of serum proteins in AFP-negative HBV-related hepatocellular carcinoma using iTRAQ -MALDI-MS/MS. *Neoplasma* **2013,** *61* (1), 17-26.

123. Olsson, A.-K.; Larsson, H.; Dixelius, J.; Johansson, I.; Lee, C.; Oellig, C.; Björk, I.; Claesson-Welsh, L., A fragment of histidine-rich glycoprotein is a potent inhibitor of tumor vascularization. *Cancer Res.* **2004,** *64*, 599-605.

124. Karrlander, M.; Lindberg, N.; Olofsson, T.; Kastemar, M.; Olsson, A. K.; Uhrbom, L., Histidine-rich glycoprotein can prevent development of mouse experimental glioblastoma. *PLoS One* **2009,** *4* (12), e8536.

125. Thulin, A.; Ringvall, M.; Dimberg, A.; Karehed, K.; Vaisanen, T.; Vaisanen, M. R.; Hamad, O.; Wang, J.; Bjerkvig, R.; Nilsson, B.; Pihlajaniemi, T.; Akerud, H.; Pietras, K.; Jahnen-Dechent, W.; Siegbahn, A.; Olsson, A. K., Activated platelets provide a functional microenvironment for the antiangiogenic fragment of histidine-rich glycoprotein. *Mol. Cancer Res.* **2009,** *7* (11), 1792-1802.

126. Klenotic, P. A.; Huang, P.; Palomo, J.; Kaur, B.; Van Meir, E. G.; Vogelbaum, M. A.; Febbraio, M.; Gladson, C. L.; Silverstein, R. L., Histidine-rich glycoprotein modulates the anti-angiogenic effects of vasculostatin. *Am. J. Pathol.* **2010,** *176* (4), 2039-2050.

127. Qiu, J.; Patwa, T. H.; Xu, L.; Shedden, K.; Misek, D. E.; Tuck, M.; Jin, G.; Ruffin, M. T.; Turgeon, D. K.; Synal, S.; Bresalier, R.; Marcon, N.; Brenner, D. E.; Lubman, D. M., Plasma glycoprotein profiling for colorectal cancer biomarker identification by lectin glycoarray and lectin blot. *J. Proteome Res.* **2008,** *7*, 1693-1703.

128. Clerc, F.; Reiding, K. R.; Jansen, B. C.; Kammeijer, G. S.; Bondt, A.; Wuhrer, M., Human plasma protein N-glycosylation. *Glycoconj J* **2016,** *33* (3), 309-43.

129. Poon, I. K. H.; Olsson, A.-K.; Hulett, M. D.; Parish, C. R., Regulation of histidine-rich glycoprotein (HRG) function via plasmin-mediated proteolytic cleavage. *Biochem. J.* **2009,** *424*, 27-37.

130. Smith, A.; Nuiry, T.; Morgan, W. T., Protelysis of histidine-rich glycoprotein in plasma and in patients undergoing thrombolytic therapy. *Thromb. Res.* **1985,** *40*, 653-661.

131. Schaller, J.; Akiyama, K.; Kimura, H.; Hess, D.; Affolter, M.; Rickli, E. E., Primary structure of a new actin-binding protein from human seminal plasma *Eur. J. Biochem.* **1991,** *196*, 743-750.

132. Fleissig, Y.; Reichenberg, E.; Redlich, M.; Zaks, B.; Deutsch, O.; Aframian, D. J.; Palmon, A., Comparative proteomic analysis of human oral fluids according to gender and age. *Oral Dis.* **2010,** *16* (8), 831-838.

133. Hassan, M. I.; Waheed, A.; Yadav, S.; Singh, T. P.; Ahmad, F., Prolactin inducible protein in cancer, fertility and immunoregulation: structure, function and its clinical implications. *Cell. Mol. Life Sci.* **2009,** *66* (3), 447-459.

134. Shiu, R. B. C.; Iwasiow, B. M., Prolactin-inducible protein in human breast cancer cells *J. Biol. Chem.* **1985,** *260*, 11307-11313.

135. Murphy, L. C.; Tsuyuki, D.; Myal, Y.; Shiu, R. P. C., Isolation and sequencing of cDNA clone for a prolactin-inducible protein (PIP). Regulation of PIP gene expression in the human breast cancer cell line T-47D. *J. Biol. Chem.* **1987,** *262* (15236-15241).

136. Freeman, M. E.; Kanyicska, B.; Lerant, A.; Nagy, G., Prolactin: Structure, function, and regulation of secretion. *Physiol. Rev.* **2000,** *80* (4), 1523-1631.

137. Ben-Jonathan, N.; Hugo, E. R.; Brandebourg, T. D.; LaPensee, C. R., Focus on prolactin as a metabolic hormone. *Trends Endocrinol. Metab.* **2006,** *17* (3), 110-116.

138. Bole-Feysot, C.; Goffin, V.; Edery, M.; Binart, N.; Kelly, P. A., Prolactin (PRL) and its receptor: Actions, signal transduction pathways and phenotypes observed in PRL receptor knockout mice. *Endocr. Rev.* **1998,** *19* (3), 225-268.

139. Vonderhaar, B. K., Prolactin involvement in breast cancer. *Endocr.-Rel. Cancer* **1999,** *6*, 389-404.

140. Drago, F.; D´Agata, V.; Iacona, T.; Spadaro, F.; Grassi, M.; Valerio, C.; Raffaele, R.; Astuto, C.; Lauria, N.; Vitetta, M., Prolactin as a protective factor in stress-induced biological changes. *J. Clin. Lab. Anal.* **1989,** *3*, 340-344.

141. Moore, K. E.; Demarest, K. T., Effects of prolactin on central dopaminergic neurons. *Prog. Neuro-Psychopharmacol. & Biol. Psychiat.* **1982,** *6* (4-6), 645-650.

142. Fitzgerald, P.; Dinan, T. G., Prolactin and dopamine: What is the connection? *J. Psychopharmacol.* **2008,** *22* (2), 12-19.

143. Ben-Jonathan, N.; Hnasko, R., Dopamine as a prolactin (PRL) inhibitor. *Endocr. Rev.* **2001,** *22* (6), 724-763.

144. Hassan, M. I.; Bilgrami, S.; Kumar, V.; Singh, N.; Yadav, S.; Kaur, P.; Singh, T. P., Crystal structure of the novel complex formed between zinc alpha2-glycoprotein (ZAG) and prolactin-inducible protein (PIP) from human seminal plasma. *J Mol Biol* **2008,** *384* (3), 663-72.

145. Ngounou Wetie, A. G.; Wormwood, K. L.; Russell, S.; Ryan, J. P.; Darie, C. C.; Woods, A. G., A pilot proteomic analysis of salivary biomarkers in autism spectrum disorder. *Autism Res.* **2015,** *8* (3), 338-350.

146. Naderi, A.; Vanneste, M., Prolactin-induced protein is required for cell cycle progression in breast cancer. *Neoplasia* **2014,** *16* (4), 329-342.

147. Adamczyk, B.; Tharmalingam, T.; Rudd, P. M., Glycans as cancer biomarkers. *Biochim. Biophys. Acta* **2012,** *1820* (9), 1347-1353.

148. Vitorino, R.; Lobo, M. J.; Ferrer-Correira, A. J.; Dubin, J. R.; Tomer, K. B.; Domingues, P. M.; Amado, F. M., Identification of human whole saliva protein components using proteomics. *Proteomics* **2004,** *4* (4), 1109-1115.

149. Ramachandran, P.; Boontheung, P.; Xie, Y.; Sondej, M.; Wong, D. T.; Loo, J. A., Identification of N-linked glycoproteins in human saliva by glycoprotein capture and mass spectrometry. *J. Proteome Res.* **2006,** *5*, 1493-1503.

150. Caputo, E.; Camarca, A.; Moharram, R.; Tornatore, P.; Thatcher, B.; Guardiola, J.; Martin, B. M., Structural study of GCDFP-15/gp17 in disease versus physiological conditions using a proteomic approach. *Biochemistry* **2003,** *42* (20), 6169-6178.

151. Franz, A. H.; Lebrilla, C. B., Evidence for long-range glycosyl transfer reactions in the gas phase. *J. Am. Soc. Mass Spectrom.* **2002,** *13*, 325-337.

152. Spohr, U.; Lemieux, R. U., Synthetic, conformational, and immunochemical studies of modified Lewis b and y human blood-group determinants to serve as probes for the combining site of the lectin IV of griffonia simplicifolia. *Carbohydr. Res.* **1988,** *174*, 211-237.

153. Aarnoudse, C. A.; Garcia Vallejo, J. J.; Saeland, E.; van Kooyk, Y., Recognition of tumor glycans by antigen-presenting cells. *Curr. Opin. Immunol.* **2006,** *18* (1), 105-111.

154. Isla Larrain, M.; Demichelis, S.; Crespo, M.; Lacunza, E.; Barbera, A.; Creton, A.; Terrier, F.; Segal-Eiras, A.; Croce, M. V., Breast cancer humoral immune response: involvement of Lewis y through the detection of circulating immune complexes and association with Mucin 1 (MUC1). *J. Exp. Clin. Cancer Res.* **2009,** *28*, 121-131.

155. Mortezai, N.; Behnken, H. N.; Kurze, A. K.; Ludewig, P.; Buck, F.; Meyer, B.; Wagener, C., Tumor-associated Neu5Ac-Tn and Neu5Gc-Tn antigens bind to C-type lectin CLEC10A (CD301, MGL). *Glycobiology* **2013,** *23* (7), 844-852.

156. Wu, L.-H.; Hu, P.; Wu, W.; Wu, X.-Z., Fucosylated glycans associated with development and metastasis of hepatocellular carcinoma cells. *Prog. Biochem. Biophys.* **2009,** *36* (1), 33-41.

157. Madjd, Z.; Parsons, T.; Watson, N. F.; Spendlove, I.; Ellis, I.; Durrant, L. G., High expression of Lewis y/b antigens is associated with decreased survival in lymph node negative breast carcinomas. *Breast Cancer Res.* **2005,** *7* (5), 780-787.

158. Kim, Y. S.; Yuan, M.; Itzkowitz, S. H.; Sun, Q.; Kaizu, T.; Palekar, A.; Trump, B. F.; Hakomori, S.-I., Expression of Le^y and extended Le^y blood group-related antigens in human malignant, premalignant, and nomalignant colonic tissues. *Cancer Res.* **1986,** *46*, 5885-5992.

159. Miyake, M.; Taki, T.; Hitomi, S.; Hakomori, S.-I., Correlation of expression of H/Ley/Leb antigens with survival in patients with carcinoma of the lung. *N. Engl. J. Med.* **1992,** *327*, 14-18.

160. Liu, J.-J.; Lin, B.; Hao, Y.-Y.; Li, F.-F.; Liu, D.-W.; Qi, Y.; Zhu, L.-C.; Zhang, S.-L.; Iwamori, M., Lewis (y) antigen stimulates the growth of ovarian cancer cells via regulation of the epidermal growth factor receptor pathway. *Oncol. Rep.* **2010,** *23*, 833-841.

161. van Kooyk, Y.; Rabinovich, G. A., Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat. Immunol.* **2008,** *9* (6), 593-601.

162. Saeland, E.; Belo, A. I.; Mongera, S.; van Die, I.; Meijer, G. A.; van Kooyk, Y., Differential glycosylation of MUC1 and CEACAM5 between normal mucosa and tumour tissue of colon cancer patients. *Int. J. Cancer* **2012,** *131* (1), 117-128.

163. Kitamura, K.; Stockert, E.; Garin-Chesa, P.; Welt, S.; Lloyd, K. O.; Armour, K. L.; Wallace, T. P.; Harris, W. J.; Carr, J., Specificity analysis of blood group Lewis-y (Le (y)) antibodies generated against synthetic and natural Le (y) determinants. *Immunology* **1994,** *91*, 12954-12961.

164. Kuemmel, A.; Single, K.; Bittinger, F.; Faldum, A.; Schmidt, L. H.; Sebastian, M.; Taube, C.; Buhl, R.; Wiewrodt, R., The prognostic impact of blood group-related antigen Lewis Y and the ABH blood groups in resected non-small cell lung cancer. *Tumour Biol.* **2007,** *28* (6), 340-349.

165. Schuster, M.; Umana, P.; Ferrara, C.; Brunker, P.; Gerdes, C.; Waxenecker, G.; Wiederkum, S.; Schwager, C.; Loibner, H.; Himmler, G.; Mudde, G. C., Improved effector functions of a therapeutic monoclonal Lewis Y-specific antibody by glycoform engineering. *Cancer Res.* **2005,** *65* (17), 7934-7941.

166. Klein, A.; Carnoy, C.; Lamblin, G.; Roussel, P.; van Kuik, J. A.; de Waard, P.; Vliegenthart, J. F. G., Isolation and structural characterization of novel neutral oligosaccharide-alditols from respiratory-mucus glycoproteins of a patient suffering from bronchiectasis. *Eur. J. Biochem.* **1991,** *198*, 151-168.

167. van Kuik, J. A.; de Waard, P.; Vliegenthart, J. F. G.; Klein, A.; Carnoy, C.; Lamblin, G.; Roussel, P., Isolation and structural characterization of novel neutral oligosaccharide-alditols from respiratory-mucus glycoproteins of a patient suffering from bronchiectasis. *Eur. J. Biochem.* **1991,** *198*, 169-182.

168. Klein, A.; Carnoy, C.; Lamblin, G.; Roussel, P.; van Kuik, J. A.; Vliegenthart, J. F. G., Isolation and structural charcterization of novel sialylated oligosaccharide-alditols from respiratoty-mucus glycoproteins of a patient suffering from bronchiectasis. *Eur. J. Biochem.* **1993,** *211*, 491-500.

169. Clark, G. F.; Grassi, P.; Pang, P. C.; Panico, M.; Lafrenz, D.; Drobnis, E. Z.; Baldwin, M. R.; Morris, H. R.; Haslam, S. M.; Schedin-Weiss, S.; Sun, W.; Dell, A., Tumor biomarker glycoproteins in the seminal plasma of healthy human males are endogenous ligands for DC-SIGN. *Mol. Cell Proteomics* **2012,** *11*, 1-12.

170. Pompach, P.; Brnakova, Z.; Sanda, M.; Wu, J.; Edwards, M.; Goldman, R., Site-secific glycoforms of haptoglobin in liver cirrhosis and hepatocellular carcinoma. *Mol. Cell Proteomics* **2013,** *12*, 1281-1293.

171. Lek, M.; Karczewski, K. J.; Minikel, E. V.; Samocha, K. E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A. H.; Ware, J. S.; Hill, A. J.; Cummings, B. B.; Tukiainen, T.; Birnbaum, D. P.; Kosmicki, J. A.; Duncan, L. E.; Estrada, K.; Zhao, F.; Zou, J.; Pierce-Hoffman, E.; Berghout, J.; Cooper, D. N.; Deflaux, N.; DePristo, M.; Do, R.; Flannick, J.; Fromer, M.; Gauthier, L.; Goldstein, J.; Gupta, N.; Howrigan, D.; Kiezun, A.; Kurki, M. I.; Moonshine, A. L.; Natarajan, P.; Orozco, L.; Peloso, G. M.; Poplin, R.; Rivas, M. A.; Ruano-Rubio, V.; Rose, S. A.; Ruderfer, D. M.; Shakir, K.; Stenson, P. D.; Stevens, C.; Thomas, B. P.; Tiao, G.; Tusie-Luna, M. T.; Weisburd, B.; Won, H. H.; Yu, D.; Altshuler, D. M.; Ardissino, D.; Boehnke, M.; Danesh, J.; Donnelly, S.; Elosua, R.; Florez, J. C.; Gabriel, S. B.; Getz, G.; Glatt, S. J.; Hultman, C. M.; Kathiresan, S.; Laakso, M.; McCarroll, S.; McCarthy, M. I.; McGovern, D.; McPherson, R.; Neale, B. M.; Palotie, A.; Purcell, S. M.; Saleheen, D.; Scharf, J. M.; Sklar, P.; Sullivan, P. F.; Tuomilehto, J.; Tsuang, M. T.; Watkins, H. C.; Wilson, J. G.; Daly, M. J.; MacArthur, D. G.; Exome Aggregation, C., Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **2016,** *536* (7616), 285-91.

ACKNOWLEDGEMENTS

Parts of section 5.1 were taken (partly modified) from a published paper:

Reprinted with permission from "A. Wiegandt, B. Meyer, Unambiguous characterization of *N*-glycans of monoclonal antibody cetuximab by integration of LC-MS/MS and ^{1}H-NMR spectroscopy, Analytical Chemistry, 2014, 86, 4807-4814". Copyright 2016 American Chemical Society.

Some of the experiments described in sections 5.3.2.1, 5.3.2.3, 6.6.2, 6.6.3 and 6.6.6 were performed during A. Wiegandt´s supervision of the bachelor thesis of K. Hirte in 2015.

DANKSAGUNG

Ohne eine Reihe wichtiger Leute wären diese 200 Seiten kaum möglich gewesen:

Großer Dank gilt den Analytikern - Hinnerk, Tim, Melissa und Raffi. Danke für eure schlauen Ideen und eure MS-Begeisterung. Ohren steif halten!

Meinen Mitbewohnern Hinnerk und Tim: Ich glaube, ihr wisst, wie sehr ich die Zeit im Computerraum zu schätzen weiß. Ihr (naja, wir) seid (naja, sind) echt groß in (infantiler) Ablenkung, aber ohne die würde auch die ganze Arbeit keinen so großen Spaß machen. Danke!

Nadja und Patrizia danke ich für ihre Freundschaft und die lustigen Mittagspausen. Ich vermisse euch.

Allen Bewohnern der 13a danke ich für die lustige und Musik-reiche Laborstimmung.

Meiner Bachelorandin Katha und meinen Praktis Claas, Hinnerk und René danke ich für ihre Mitarbeit. Jeder von euch hat einen Beitrag zu dieser Arbeit geleistet.

Insa und Ilona danke ich für ihre gute Laune und ihre Mithilfe bei der HRG-Aufreinigung.

Allen anderen aktuellen und ehemaligen AK-Mitgliedern danke ich für viele lustige Stunden im Labor und im Sozialraum. Insbesondere danke an Henning für die Einführung in die Glyco-MS-Welt.

Maria danke ich für ihre Hilfsbereitschaft sowie ihre stetige Ansprechbarkeit rund um die Massenspektrometrie im Allgemeinen und um das Sorgenkind „maXis“ im Speziellen. Dem MS-Team danke ich für seine immer freundliche und hilfsbereite Art.

Franky vom Spillner-AK sowie Marco und Dirk vom Redecke-AK danke ich für ihre extrem hilfsbereite und freundliche Art in den FUT8- und HRG-Projekten.

Nadja, Hinnerk, Tim, Melissa, Moritz, Raffi, Karsten und Philipp danke ich fürs Korrekturlesen von Teilen dieser Arbeit. Tim: Danke für deine NMR-Hilfe. Philipp: Danke für deine Matlab-Hilfe.

Der PIER Helmholtz Graduate School danke ich für die Unterstützung durch ein Stipendium. Besonderer Dank gilt Stefanie Tepass und Maren Arnold-Vagen, die mir einen entspannten Endspurt ermöglichten.

Hier an letzter Stelle, sonst aber an erster: Ohne meine Familie und meine außeruniversitären Freunde wären die letzten Jahre kaum möglich gewesen. Ich danke insbesondere meiner Mutter und meinen Geschwistern Katharina und Konstantin für ihre fortwährende Unterstützung, ihrem Interesse an meiner Arbeit und ihrem Verständnis, wenn ich „keine Zeit“ habe… Wird nun besser.

Der größte Dank gilt Philipp: Du weißt schon wofür!

CURRICULUM VITAE

Alena Wiegandt

born on January 16th, 1987
in Hamburg, Germany

EDUCATIONAL QUALIFICATIONS

since 08/2016	**Scientist at the University of Hamburg (Institute of Organic Chemistry)**
01/2016 - 07/2016	**Scientist at the Christian-Albrechts-University of Kiel (Institute of Experimental Medicine)** Division for "Systematic Proteome Research & Bioanalytics" - mass spectrometry based protein analytics, development of novel methods for protein- and proteome analytics
10/2012 - 12/2015	**PhD thesis at the University of Hamburg (Institute of Organic Chemistry)** "Method Development for the Characterization of *N*-type Glycosylation of Proteins by Integration of LC-MS/MS and NMR"
10/2010 - 09/2012	**Studies of Chemistry (Master of Science) at the University of Hamburg** (grade: *very good with distinction*) Master thesis (Institute of Organic Chemistry, 03/2012-09/2012): „Glycan-site specific analysis of glycoproteins by cross-correlation of chromatographic, mass spectrometric and NMR-spectroscopic analysis" (grade: *very good*)
08/2011 - 11/2011	**Research project at the University of Lund, Sweden (Centre for Molecular Protein Science, Department for Biochemistry and Structural Biology)** Project: „Mass spectrometric analysis of chemically crosslinked peptides to map interactions between small heat shock proteins"
10/2007 - 09/2010	**Studies of Chemistry (Bachelor of Science) at the Free University of Berlin** (grade: *good*) Bachelor thesis at the University Medical Centre Hamburg-Eppendorf (Institute of Clinical Chemistry, Division: Mass spectrometric proteomics, 03/2010-05/2010): „Investigation of extraction methods of human plasma peptides " (grade: *very good*)
10/2006 - 09/2007	**Studies of Industrial Engineering at the University of Hamburg**
1997 - 2006	**Secondary school: Stormarnschule in Ahrensburg** (Graduation with A-levels)

PROFESSIONAL EXPERIENCE

since 08/2016	**Scientist** at the University of Hamburg
01/2016 - 07/2016	**Scientist** at the University of Kiel
10/2012 - 12/2015	**PhD Student** at the University of Hamburg
06/2012 - 08/2012	**Tutor** in a high school student internship „Brücken in die Wissenschaft“, supported by the Joachim-Herz-Stiftung
07/2010 - 06/2011	**Work student** in the testing institute „Hansecontrol GmbH (Hermes)“, Hamburg, Division: Organic Chemistry, pollutant analytics
2009	**Internship** at the University Medical Centre Hamburg-Eppendorf, Department of Clinical Chemistry, Division: Mass spectrometric proteomics
2009	**Internship** at „Richter Helm BioLogics GmbH Co. KG”, Hamburg
2002	**Internship** at the „Federal Research Centre for ornamental plant culture and genetics“,

JOURNAL PUBLICATIONS

A. Klettner, A. Tholey, A. Wiegandt, E. Richert, B. Nölle, G. Deuschl, J. Roider, S.A. Schneider, *Reduction of GAPDH in lenses of Parkinson patients – a possible new biomarker*, *Reduction of GAPDH in lenses of Parkinson -patients – a possible new biomarker,* Movement Disorders, 2016, accepted

A. Wiegandt, B. Meyer, *Unambiguous characterization of N-glycans of monoclonal antibody cetuximab by integration of LC-MS/MS and ^{1}H-NMR spectroscopy*, Analytical Chemistry, 2014, 86, 4807-4814

M. Fellenberg, H.N. Behnken, T. Nagel, A. Wiegandt, M. Baerenfaenger, B. Meyer, *Glycan analysis: scope and limitations of different techniques - a case for integrated use of LC-MS(/MS) and NMR techniques*, Analytical and Bioanalytical Chemistry, 2013, 405, 7291-7305

C. Söderberg, W. Lambert, S. Kjellström, A. Wiegandt, R. Peterson Wulff, C. Mansson, G. Rutsdottir, C. Emanuelsson, *Detection of crosslinks within and between proteins by LC-MALDI-TOF/TOF and the software FINDX to reduce the MSMS-data to acquire for validation*, PloS one, 2012, 7 (6)

Wiegandt, B. Meyer, *Unique highly fucosylated N-glycans: Lewis Y antigens found dominantly in salivary Prolactin-Inducible Protein (PIP)*, to be submitted

SCHOLARSHIPS

Scholarship of the Helmholtz-Association (funding of doctorate, 01/2013 - 12/2015), Member of the *Helmholtz Graduate School*

Sponsorship of the German Academic Exchange Service (DAAD) for a three-month research internship in Sweden, at the University of Lund (08/2011 - 11/2011)

Travel grants for the attendance at the conference HPLC2013 in Amsterdam, NL from the “Gesellschaft Deutscher Chemiker e.V.“, the „Glycovereins e.V.“ and the „ Körperschaftsvermögens der Universität Hamburg” (06/2012)

AFFIDAVIT

ERKLÄRUNG

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

Hamburg, 27.11.2016

Alena Wiegandt