Universität für Bodenkultur Wien

Department of Chemistry

Division of Biochemistry
Head of Institute: Ao.Univ.Prof. Dipl.-Ing. Dr.nat.techn. Erika Staudacher

Advisor: Ao.Univ.Prof. Dipl.-Ing. Dr.nat.techn. Friedrich Altmann

ISOMER SPECIFIC ANALYSIS OF N-GLYCANS AS A PREREQUISITE FOR UNDERSTANDING BIOLOGICAL FUNCTIONS AT THE EXAMPLES OF INFLUENZA A VIRUS RECEPTORS AND MOUSE BRAIN GLYCOPROTEINS

Dissertation
for obtaining a doctorate degree
at the University of Natural Resources and Applied Life Sciences Vienna, Division of Biochemistry

Submitted by
Mag. Laura Neumann

Vienna, Oktober 2014

So eine Arbeit wird eigentlich nie fertig,

man muss sie für fertig erklären,

wenn man nach der Zeit und den Umständen das Möglichste getan hat.

- *Johann Wolfgang von Goethe*

Acknowledgements

Herewith, I want to thank my supervisor Fritz Altmann for the great opportunity to perform my PhD thesis in his lab where I had the possibility to gain profound knowledge in mass spectrometric and biochemical methods.

Furthermore I want to thank my co-supervisor Reingard Grabherr for her support and help with all upcoming problems.

A special thanks goes to Martin Pabst for his constant support, excellent knowledge and endowing supervision even after he left the lab. His patience and modesty with me and all my questions on a professional and also personal level helped me immensely to accomplish this work.

I'm grateful for all my colleagues in the lab for the great time we had during coffee breaks work discussions and all the activities we made together. Without them I would have missed a number of great experiences.

I also want to thank Anna, Johannes, Krista, Martin, Melanie and Simone for their companionship during this challenging period of life, the coffee breaks, after work's and finally the friendship that arose. I'm happy that I was allowed to meet such great people.

Last but definitely not least I want to thank my family, my parents and my sister. Not only financially but more important mentally they were always there for me and supported me in any possible way.

Abstract

Protein glycosylation is a pivotal post translational protein modification essential for a number of different cellular functions such as brain development or memory formation. Furthermore, pathogens are known to take advantage of the presence of unique carbohydrates. The complexity of the underlying glyco-code is based on the possibility of generating a large number of different glycan structures with just a few different monosaccharides. Many of these naturally occurring glycans appear to be isomeric, differing in the linkage or position of residues, or the position of further modifications. Selective, sensitive and quantitative methods allowing a deeper insight and understanding are therefore of great importance for the understanding of the protein function.

Influenza A virus strains are known to attack their hosts in a species-specific manner by an initial binding of their hemagglutinins to sialic acid-decorated receptors. In the first part of this study, the glycan binding specificities of recombinantly expressed H1 and H3 hemagglutinin were closely investigated. For this investigation, N-glycans isolated from porcine naso-pharyngeal tissues were either analysed by affinity chromatography or by glycan arrays using MALDI-TOF-MS for identification glycan structures. Surprisingly, the affinity chromatography as well as the glycan array experiments showed binding to both oligomannosidic as well as sialic acid-containing glycan structures. However, the overall retention in the affinity chromatography experiments was rather poor due to the low affinity of the hemagglutinins to free glycans and especially to monovalently presented sialic acids.

The second part of this work focused on the analysis of isomeric fucosylated N-glycans. Different recombinantly expressed fucosyltransferases were used to generate glycan standards with defined fucose linkages (Fuc α1,3- or α1,4-linked to GlcNAc and Fuc α1,2-Gal on the distal side and a Fuc α1,6- core linkage). The glycans were characterised using porous graphitic carbon (PGC) capillary chromatography coupled online to an ESI-QTOF-MS instrument, which allows high shape-selectivity even for closely related isomeric compounds. Stable isotopically labelled internal standards helped to overcome retention time shifts. Furthermore, fragmentation experiments were performed on the separated standards using CID with an ion trap mass spectrometer aiming to identify diagnostic fragments characteristic for different types of fucosylation. In fact, the majority of the singly and doubly fucosylated reference glycans gave peaks with unambiguous retention behaviour. This knowledge was then applied to the secretory component of IgA, where Lewis x-antigens were confirmed. Finally, the LC-ESI-MS/MS method was used to investigate fucosylated glycans from mouse brain. The latter surprisingly did not contain complex-type di-antennary fucosylated N-glycans, but instead presumably contained substantial amounts of hybrid-type glycans with bisecting GlcNAc. This approach is particularly attractive, as – in principle – all structures within a biological sample can be assigned and quantitated within a short analysis time, showing highest selectivity, sensitivity and unambiguous identification using relative retention times and mass spectrometric detection.
Not least, glycans are linked at well-defined positions to the protein backbone. Therefore in addition to the type of glycan, the position within a protein is also of great importance. Within the broader scope of this work, we also applied glycoproteomic and proteomic analysis on different model proteins to perform and further develop a site-specific glycosylation analysis approach based on RP-ESI-MS or MS/MS.

Key words: Glycan-isomers, PGC-chromatography, mass spectrometry, glycan arrays, affinity chromatography

Zusammenfassung (German)

Eine der wichtigsten post-translationalen Modifikationen stellt die Glykosylierung von Proteinen dar. Glykane beeinflussen eine Vielzahl zellulärer Vorgängen und sind unter anderem an der Gehirn- und Gedächtnisentwicklung beteiligt, sowie essenzieller Bindungspartner für unterschiedliche Pathogene. Die Identifizierung der Glykan-zusammensetzung eines Proteins oder Gewebes stellt eine erhebliche analytische Aufgabe dar, da die einzelnen Zuckermoleküle in den unterschiedlichsten Zusammensetzungen und Bindungen vorkommen können, wobei jede Zusammensetzung unterschiedliche Effekte hervor rufen kann. Daher ist es wichtig eine hoch empfindliche, selektive und quantitative Methode für die Glykananalyse zu entwickeln, um selbst die äußerst geringen Unterschiede der Glykanisomerie zweifelsfrei unterscheiden zu können. In dieser Arbeit wurden unterschiedliche biochemische und analytische Methoden entwickelt, mit der an Hand von Referenzproteinen, isomerspezifische Bindungspräferenzen untersucht wurden, sowie Glykanisomere in unterschiedlichen Glykoproteinen identifiziert werden konnten.

Ein Teil der Arbeit beschäftigte sich mit dem Influenza A virus-Membranprotein Hämagglutinin. Es ist bewiesen, dass das Bindungsverhalten unterschiedlicher Virusstämme mit der unterschiedlichen Bindung von Sialinsäure an den Glykoproteinrezeptoren der jeweiligen Spezies einhergeht. Aus diesem Grund wurden affinitätschromatographische Experimente sowie Glykan-Arrays mit unterschiedlichen Hämagglutininen durchgeführt. Das Bindungsverhalten der Hämagglutinine konnte einerseits hinsichtlich der Präferenz der jeweiligen Sialinsäurebindung bestätigt werden, allerdings zeigten die rekombinant hergestellten Proteine zusätzlich auch ein sehr breites generelles Bindungsspektrum an neutrale hochmannosidische Strukturen. Dieses Ergebnis kann möglicherweise durch die Wahl des Expressionssystems, den Insektenzellen, erklärt werden.

Ein weiterer Teil der Arbeit beschäftigte sich mit der Herstellung so genannter Referenzglykane welche, sämtliche Glykanisomere einer komplexen di-antennären einfach fukosylierten Struktur darstellen. Diese Glykanisomere zeigen ein hochspezifisches Elutions-profil auf einer *Porous Graphitized Carbon* (PGC)-Säule, welche die chromatographische Trennung solcher Isomere erlaubt. Durch Zugabe isotopenmarkierter interner Glykan-standards, ließen sich relative Retentionszeiten berechnen, mit Hilfe derer Glykanisomere von unbekannten Proteinproben, wie der sekretorischen Kette des IgA Moleküls oder von Mäusegehirnproteinen, identifiziert werden konnten. Da das Retentionsverhalten der PGC-Säulen nicht völlig stabil ist und immer wieder leichte Verschiebungen im Elutionsmuster der Glykane vorkommen können, wurden zusätzlich Fragementierungsanalysen der Referenz-glykane sowie der unbekannten Glykanisomere durchgeführt. Die Fragmentierungen wurden mit Hilfe einer Ionenfalle im negativen Ionisierungsmodus durchgeführt, da dies zur Herstellung sogenannter „diagnostischer Ionen“ führt, meist aus der A- und Z-Ionenserie, und dadurch die Identifizierung der unterschiedlichen Fukosylierungsarten ermöglicht wird. Zusammen mit dem eindeutigen Elutionsverhalten unterschiedlicher Glykanisomere, ermöglicht diese Methode eine hochspezifische, empfindliche und eindeutige Identifizierung unbekannter Isomere.
Zu guter Letzt beschäftigte sich auch ein maßgeblicher Teil der Arbeit mit der Identifizierung von Glykopeptiden, Glykosylierungsstellen und Proteinen anhand unterschiedlicher Modellproteine.

Schlüsselwörter: Glykanisomerie, PGC-Chromatographie, Massenspektrometrie, Glykan Arrays, Affinitäts-chromatographie

1 Introduction

1.1 Protein-glycosylation

Glycosylation is one of the most common post translational modifications of proteins. More than 50 % of all proteins are glycosylated implying its high functional importance [1].

In contrast to carbohydrate intake for nutritional purposes to supply the energy metabolism, which is nowadays mostly served by starch (polymer of glucose) or saccharose intake (disaccharide of glucose and fructose), the discussed glycoconjugates in this work are rather complex, mostly branched hetero-polymers consisting of a large variety of different glycosidically linked monosaccharides (mainly hexoses and pentoses) [2].

In mammalian systems glycoconjugates can be divided into three major groups as shown in Figure 1. Two of the three major classes are protein related. For the groups of glycoproteins and proteoglycans, carbohydrates are bound to proteins, whereas within the third class, the glycosphingolipids, the glycan moiety is linked to lipids [2].

Proteoglycans, in contrast to glycoproteins, are built up from one single core protein with one or more glycosaminoglycan (GAG) chains attached. These unbranched polysaccharides consist of repeating disaccharide units, which can vary in the disaccharide composition, linkage and sulfatation [3].

Glycolipids within mammals comprise of two major subclasses. The first subclass of glycosphingolipids consists of a hydrophilic oligosaccharide chain and a hydrophobic lipid diacylglycerol or ceramide, which anchors the glycolipid to cell membranes. The second subclass of glycophospholipids (GPI anchors) is made of a phosphatidylglycerol core instead of the ceramide [4].

Generally speaking, glycans are assemblies of highly structural diverse carbohydrate moieties (mono-, oligo- and polysaccharides) linked to their conjugates (glycoproteins, glycolipids and proteoglycans). The structural diversity lies in the different monosaccharide compositions, glycosidic linkages, anomeric states, branching and carbohydrate modifications such as phosphorylation, sulfatation or methylation. Each of these structural diversifications is determined during glycan biosynthesis by the “glycosylation - machinery” consisting of a big set of enzymes, cofactors, transporters, activated sugar donors and many

more [5]. Approximately 1 % of the genome is dedicated to glycosyltransferases and considering all genes of the “glycosylation machinery” the number would rise to 3-4 %. This again, reflects the high structural diversity within glycosylation patterns [6].

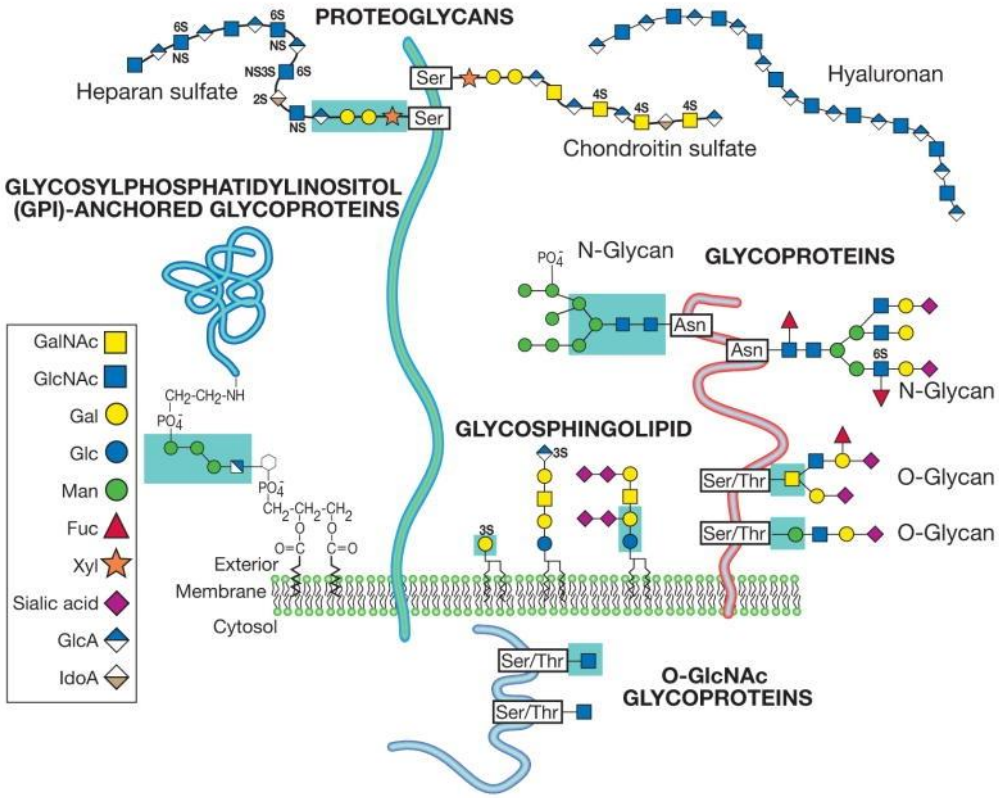


Figure 1 Different classes of mammalian glycosylation [2]

About 50 % mammalian cellular and membrane bound proteins are glycosylated implicating the crucial role of glycosylation for protein and cell functions.

Carbohydrates are known to participate in protein folding and maturation by providing structural properties, charge and hydrophility. Furthermore they mediate cellular interactions, host-pathogen interactions and many more which will be described in detail later. Essential carbohydrate mediated interactions are found in the immune system. Several studies revealed their importance, as they participate in immune cell differentiation [7] and trafficking [8], T-cell receptor signalling and apoptosis [9], antibody function [10], pathogen recognition [11] and immune homeostasis [12].

1.2 Biosynthesis - glycosylation pathway

This work focuses on glycoproteins, which are proteins carrying covalently linked oligosaccharides to the peptide backbone. These oligosaccharides consist of different monomers and are mostly branched [13]. Protein bound glycans are either linked via an amide-bond to an asparagine (N-linked glycosylation) or via the oxygen of a serine, threonine, tyrosin or hydroxyproline (O-linked glycosylation) [14] [15].

Glycosylation is restricted mainly to proteins that are synthesized and sorted in the secretory pathway, which includes ER, Golgi, lysosomal- and plasma-membrane as well as secretory proteins. One exception is the modification of nuclear and cytosolic proteins with a single O-linked GlcNAc [16].

The probably best studied type of glycosylation is the N-linked glycosylation which is conserved within all eukaryotic cells [2].

Glycan precursors are synthesized (Figure 2) in the endoplasmatic reticulum (ER) partly on the cytoplasmic side and partly on the luminal side of the ER membrane. The initiation takes place on the cytosolic side by the transfer of a GlcNAc-P from UDP-GlcNAc to a membrane linked dolichol-P lipid which is catalysed by the GlcNAc-1-phosphotransferase. Subsequent stepwise addition of a second GlcNAc and five mannoses takes place by processing UDP-GlcNAc and GDP-Man, respectively [17]. The resulting precursor $Man_5GlcNAc_2$-P-P-Dol is translocalized through the membrane to the luminal side of the ER by a “flippase” where further processing takes place [18]. The glycan processing on the luminal side is accomplished by dolichol phosphate activated sugars, in contrast to the direct processing and subsequent incorporation of nucleotid activated sugars on the cytosolic side. The Dol-P-Man and Dol-P-Glc donors are generated by processing GDP-Man and UDP-Glc on the cytosolic side by attaching the sugar to dolichol phosphate which is subsequently flipped into the lumen, where the incorporation of these sugars takes place. Different mannosyltransferases and glucosyltransferases generate the final glycan precursor, $Glc_3Man_9GlcNAc_2$-P-P-Dol, which is finally transferred by an oligosaccharyltransferase complex (OST) to the nascent translated protein [17] [19]. The linkage to the protein occurs via an amide bond to the asparagine with in the N-glycan consensus sequence Asn-X-Ser/Thr (in unusual cases Cys) where the amino acid X can be anything but a proline residue [2].

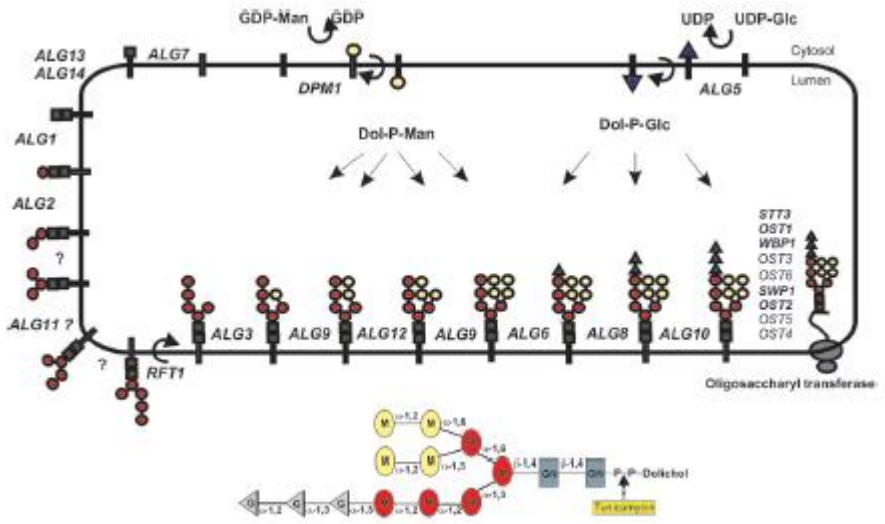


Figure 2 Synthesis of dolichol-P-P-GlcNAc$_2$Man$_9$Glc$_3$ [13]

In contrast to N-linked glycosylation O-linked glycosylation can be much more diverse as 7 different types of O-glycosylation can occur (mucin-type, glycosaminoglycans (GAG), O-linked GlcNAc, O-linked Gal, O-linked Man, O-linked Glc and O-linked Fuc) which are α or β linked to Ser, Thr, Tyr, Hyp [20].

Within O-glycosylation no “en bloc” transfer of the glycan moiety to the protein backbone occurs, like it is found in N-glycosylation, but a sequential enzymatic transfer of individual monosaccharides starting with the glycosylation of hydroxyl-groups of side chains of various amino acids [2]. The biosynthesis of O-glycans is initiated in the late ER or the Golgi compartments after proper protein folding and eventual oligomerization of the protein [20].

In vertebrates O-glycosylation is mainly initialized by the addition of a GalNAc to the hydroxyl-group of a serinee or threonine by a polypeptide GalNAc transferase (ppGalNAcT) [21]. In contrast to N-glycosylation there is no known consensus sequence which leads to the GalNAc transfer though the frequency of adjacent proline, serine and threonine residues seem to enhance whereas, adjacent negatively charged amino acids seem to hinder the occurrence of O-glycosylation [22]

1.2.1 N-glycosylation

After the transfer of the Glc$_3$Man$_9$GlcNAc$_2$ N-glycan precursor to the protein backbone glucosidases trim the two terminal glucoses. The resulting monoglucosylated high mannose oligosaccharide is required for the interaction with calnexin and calreticulin, two chaperones

which participate in glycoprotein folding. After the accurate folding, deglucosylation takes place and further processing of the $Man_9GlcNAc_2$ – intermediate happens in the ER and the Golgi compartments. The first trimming with α1,2-mannosidases IA and IB starts in the Golgi and results in $Man_5GlcNAc_2$ which is the substrate for further diversification [23].

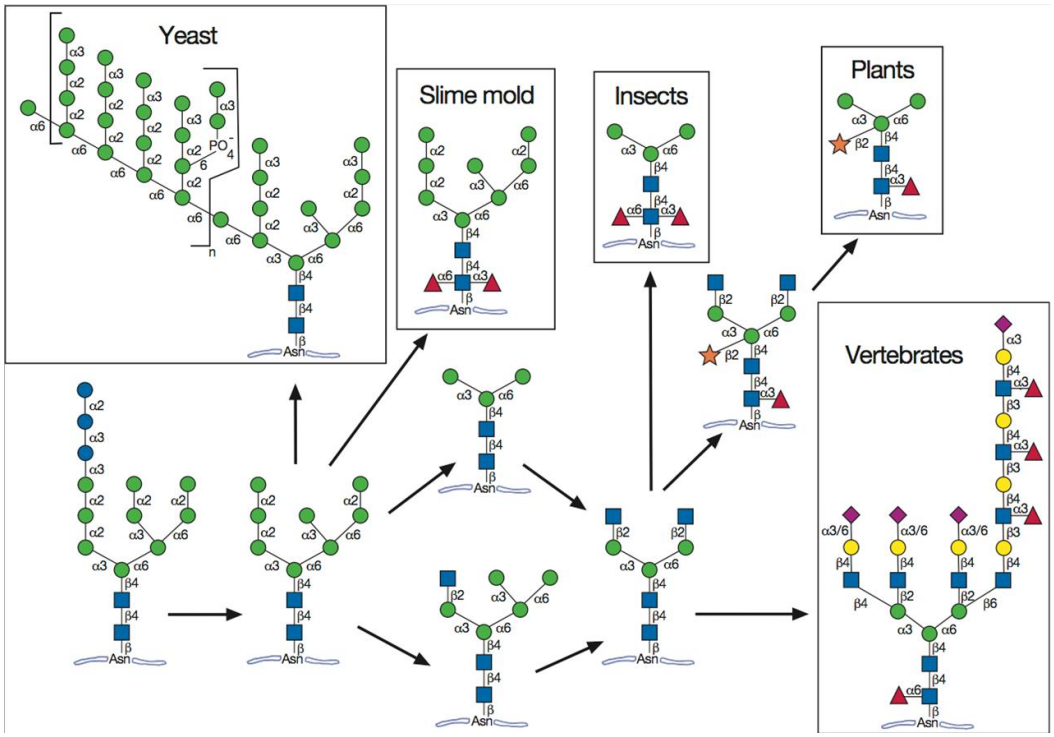


Figure 3 N-glycan processing and maturation [2]

Within different species different types of N-glycosylation can occur, like highly branched paucimannosidic N-glycans (yeast) in contrast to rather small complex N-glycans with different decorations like α1,6 and α1,3 linked core fucoses (insects) or a combination of an α1,3 linked core fucoses and a β1,2 linked xylose (plants). Mammalian N-glycosylation reaches from oligomannosidic and hybrid type to highly complex and branched type glycans (vertebrates) as indicated in Figure 3.

In most cases processing of the N-glycans starts after a the removal of the α1,2 Man from the central arm of the $Man_9GlcNAc_2$ in the ER by the ER α-mannosidase I producing $Man_8GlcNAc_2$ for further maturation in the distinct Golgi compartments. Here the glycoproteins can be catalysed by the N-acetyl glucosaminyl phosphotransferase and the N-acetylglucosamin-1-phosphodiester-α-acetylglucosaminidase resulting in mannose 6-phosphorylated residues on mannose-rich glycoproteins. This mannose-6-phosphorylated glycans are directly targeted to the lysosomes, the degrading mammalian organelle [19].

To process oligomannosidic ($Man_{5-9}GlcNAc_2$), hybrid and complex N-glycans, non lysosomal glycoproteins are further trimmed by Golgi α-mannosidase I (three isoforms are present in humans) removing the remaining three mannose residues to generate $Man_5GlcNAc_2$ [24].

In the medial cisternae of the Golgi the GlcNAc-transferase I (GnT-I) is adding a GlcNAc to the "3-arm" resulting in a hybrid-type N-glycan. To obtain complex-type N-glycans, the remaining α-mannose on the "6-arm" needs to be removed by Golgi α-mannosidase II prior the addition of a second GlcNAc on the "6-arm" catalysed by the GlcNAc transferase II (GnT-II). In total, five different GlcNAc-transferases (GnT-I-V) can act on the glycan, leading to a large variety of branched (GnT-IV and -V) and bisecting (GnT-III) glycan structures [19]. Furthermore the addition of an α1,6 linked fucose to the inner most core GlcNAc requires the prior action of the GnT-I.

In plant-N-glycans the asparagine adjacent GlcNAc can be decorated with an α1,3 linked fucose and an additional β1,2 xylose to the innermost mannose residue of the N-glycan core, which are both immunogenic glycan epitopes for humans [25]. In invertebrates the core GlcNAc residue can be decorated with both (α1,3 and α1,6) linkage type fucoses [26]. In vertebrates a large variety of glycosyltransferases can produce complex-type N-glycans which can be highly decorated with up to four branches, LacNAc repeats terminating sialic acids, sulfatation, phosphorylation and Lewis as well as core fucosylation [19] [2].

1.2.2 O-glycosylation

Figure 4 shows the most commonly found eight different O-glycosylation core subtypes in vertebrates. This type of glycosylation is called mucin-type and is initiated by the addition of a GalNAc in α1-linkage to the hydroxyl group of a serine or threonine residue. This type of glycosylation is mostly found on secreted and membrane bound mucins as well as on other glycoproteins [27] [28].

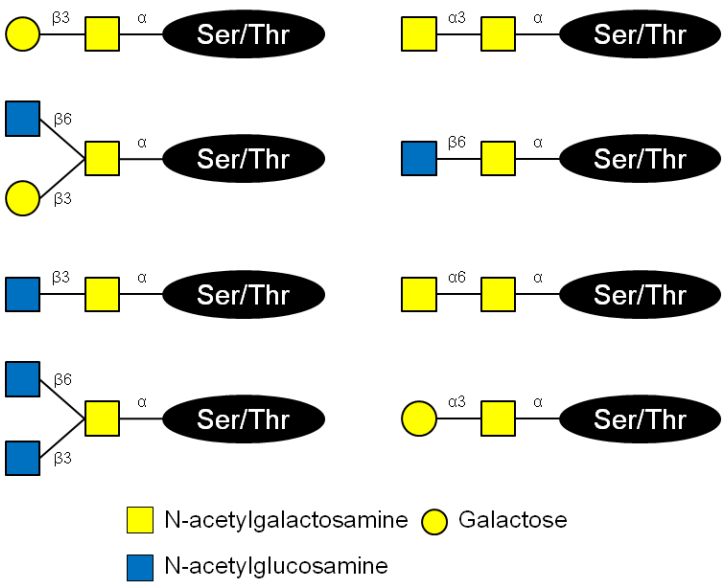


Figure 4 Different O-glycan core structures found in mammals [28]

The simplest mucin type O-glycan is also termed Tn antigen and represents a singly GalNAc residue linked to Ser/Thr. The further decoration of the Tn antigen with an additional β1,3 linked Gal is called T-antigen or core 1 O-glycan. Both glycan structures are often antigenic and may be further modified by sialic acid forming sialylated-Tn or -T antigens. Core 2 O-glycans bear an addition β1,6 linked GlcNAc residue and are often found on glycoproteins and mucins. Core 3 and core 4 O-glycans (GlcNAc β1,3 GalNAc and GlcNAcβ1,6(GlcNAcβ1,3)GalNAc) have been found in secreted mucins from colon- or bronchi tissues and salivary glands. The remaining 4 other mucin type O-glycan core structures have only very restricted appearance, like intestinal adenomacarcinoma (core 5), human intestinal mucin (core 6), human respiratory mucin (core 8) and bovine submaxillary mucine (core 7) [2] [29] [30].

Heavy O-glycosylation can also occur on proteins when, unlike in mucin type O-glycosylation, a xylose is bound to Ser/Thr building the core for the glycosaminoglycans (GAGs) on proteoglycans. GAGs are long unbranched polysaccharides bearing a high negative charge due to the presence of acidic sugars and/or modification by sulphate groups. Repeating disaccharide units consisting of alternating acidic- and amino-sugars adopt an extended conformation on the protein backbone, leading to attraction of cations binding water. These hydrated GAGs enable the proteoglycans in joints and tissues to buffer large pressure changes [31].

1.3 Glycosylation in mammals

The most common core modification within mammals is the α1,6 linked fucose to the GlcNAc adjacent to the asparagine residue of the protein backbone. This type fucosylation commonly found in mammals is often linked to cancer [32].

The N-glycan core in mammals is typically elongated with LacNAc repeats (Galβ1,4GlcNAc) Figure 5. Up to four antennas can be substituted due to the initial catalytic activity of different GlcNAc-transferases (GnT-I, -II, -IV and -V), whereas GnT-III produces bisecting N-glycans, where the GlcNAc is β1,4 linked to the β-mannose of the core [33]. The activity of GnT-V (addition of a GlcNAc in β1,6 linkage to the "6-arm" of the core) has been mentioned on combination with cancer metastasis [34].

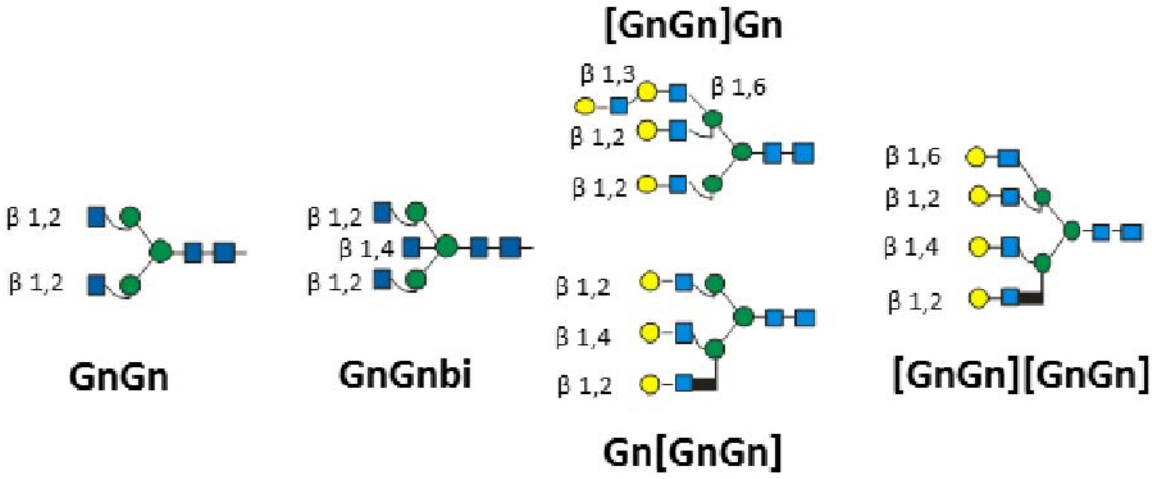


Figure 5 Different linkages of N-acetyl-glucosamine (Gn)

The most prominent and species specific mammalian glycan modifications are capping or decorating the glycans by adding galactoses usually β14 linked but also possibly β1,3 linked in humans. Furthermore the so called Galili epitope (Figure 6), an additional galactose α1,3 linked to the β1,4 linked galactose, can be found in mammals. This epitope can be found in mouse or pigs, but not in humans, as it is known to be antigenic to humans [35] [36].

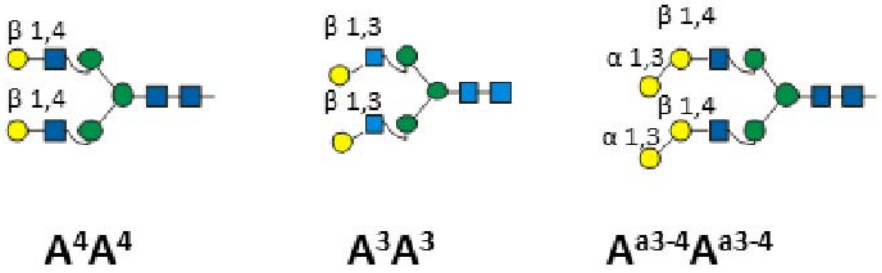


Figure 6 Different linkages of galactose (A)

Terminal galactoses can be modified by sulfatation on the 3-O position, which is found on the Tamm Horsfall protein, but also further modification by adding a terminal GalNAc residue can occur, which can be sulphated on the 4-O position [37] [38]. Another type of sulfatation is found on N-glycans isolated from recombinant erythropoietin in Chinese hamster ovary cells where the GlcNAc residue exhibits a sulfo group on the 6-O position [38]. Furthermore the type of sialic acid can differ among different species as in humans the N-acetylneuraminic acid is most abundant, whereas in many other mammals N-glycolylneuraminic acid is found [39] as depicted in Figure 7.

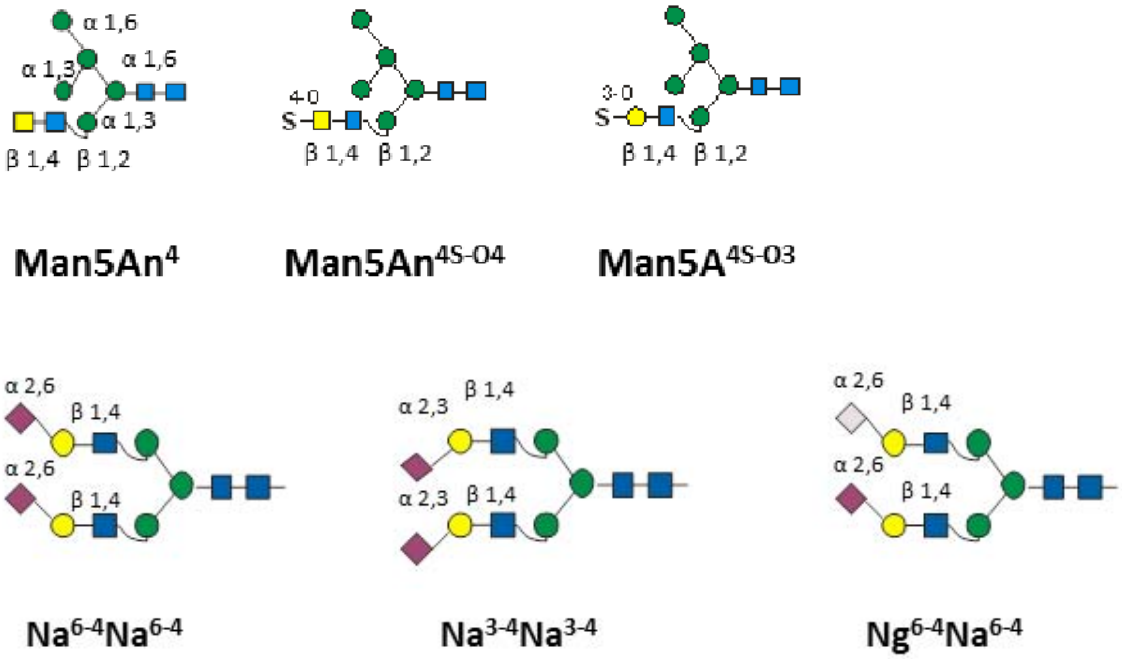


Figure 7 Different types and linkages of sialic acid (Na, Ng) and sulfatation (S) on galactose (A) or N-acetyl galacosamine (An)

Additionally to the α1,6 core fucosylation, mammalian glycans can be decorated with terminal fucoses. Up to date 11 different fucosyltransferase genes have been identified in

the human genome [40]. Different types of fucose linkage have distinct biological importance. Besides α1,2 linked fucose to terminal galactose residues (H-antigen) the fucose can also be attached α1,3 to the terminal GlcNAc (Lewis-type x) or α1,4 to the terminal GlcNAc if the galactose is linked β13 instead (Lewis-type a) [40] [41]. The Lewis-type x glycan epitope with an additional α1,2 linked fucose to the galactose residue is called Lewis-type y whereas the Lewis-type a epitope with the additional α1,2 linked fucose is called Lewis-type b [42] [43].

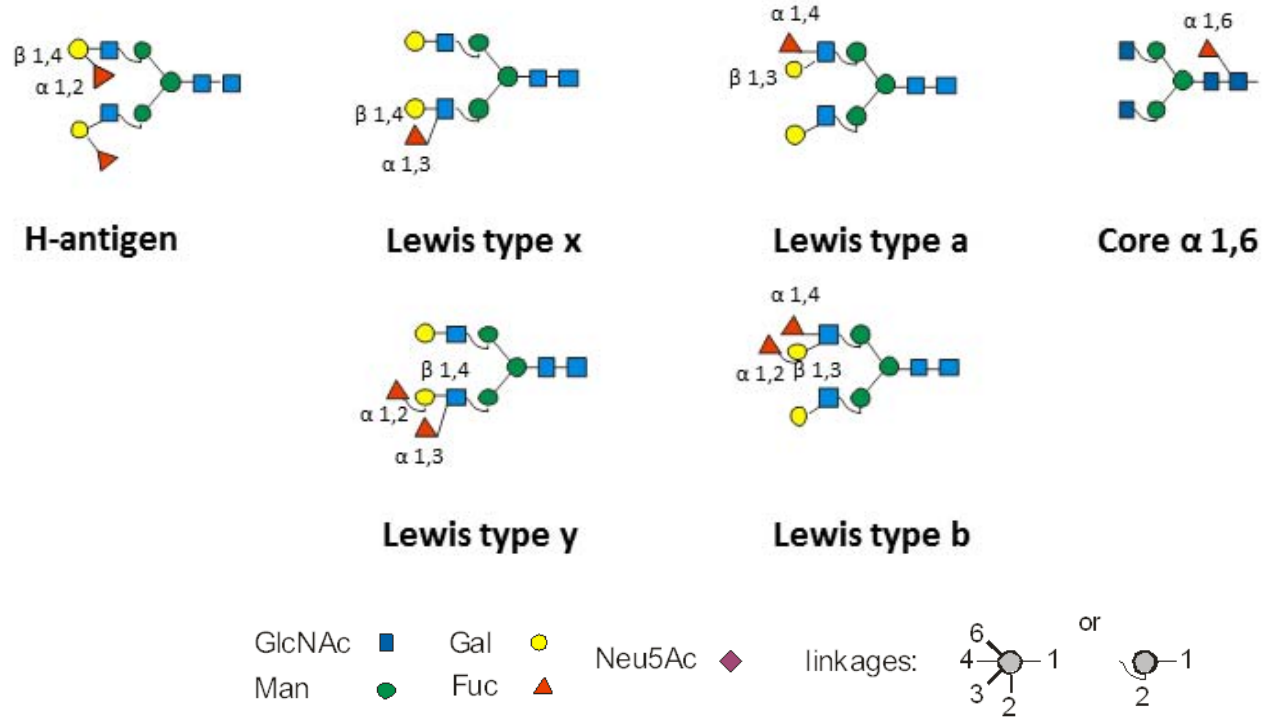


Figure 8 Different types of fucosylation (F)

Beside N-glycans also complex O-GalNAc glycans can occur within mammals of the different core structures. The complex extended core 1, 2 and 4 structures, shown in Figure 9, are found in human respiratory mucins, whereas the core 3 structure was found in human colon mucins. All four core structures can be extensively branched, and extended and furthermore decorated with terminating fucose in various linkages, sialic acid in α2,3 linkage or blood group antigenic determinants. Core 1 and 3 structures additionally can also carry sialic acid α2,6-linked to the core N-acetylgalactosamine [2].

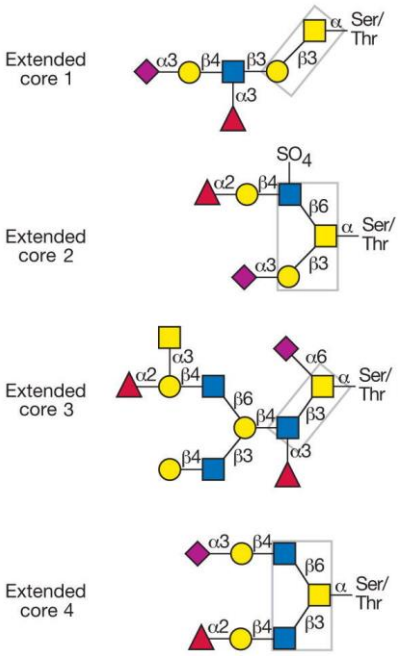


Figure 9 Complex O-GalNAc glycans with different core structures taken from [2]

1.4 Lectins - Glycan-binding proteins

Glycan binding proteins, also called lectins, can be classified on the basis of the nature of their glycan ligand, their dependence on the availability of divalent cations, their subcellular localization or the biological process they are participating in. As the genomic and amino acid sequences of hundreds of lectins are available classification can be done by shared sequence alignments. Though this sequence alignments do not function for all existing lectin classes, this classification can be used for most mammalian lectin classes (mostly C-type lectins)(Figure 10), which share conserved sequences in the active centre, the so called carbohydrate-recognition domain (CRD) [44]. C-type lectins require the presence of C^{++} cations for activity and as they are extracellular proteins they bind a high diversity of carbohydrate molecules. Another lectin class are galectins (formally S-type lectins), which can be found intra- and extracellularly, are mostly depend on reducing agents (thiols) and exclusively bind β-galactosides. P-type lectins, a third lectin class, bind mannose – phosphate [45]. In the 1990 it was found that members of the immunoglobulin superfamily can also recognize carbohydrates and thereby the class of I-type lectins was introduced with a subgroup of so called siglecs, which specifically recognize sialic acids [46]. This classification according to sequence homologies include the majority of known animal lectins but many

other glycan binding proteins are known, which do not fit into this genomic sequence based grouping but show structural similarities or also no similarity at all [2].

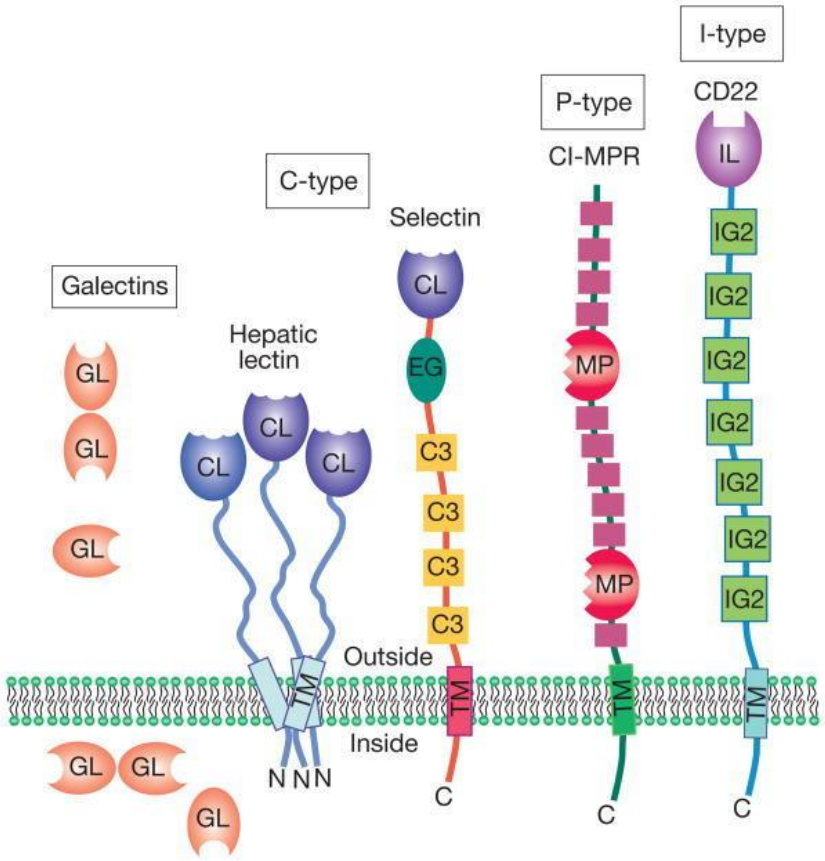


Figure 10 Classes of animal lectins according to CRD homologies [2]

The biochemical usage of lectins was greatly facilitated as in 1965 the affinity chromatography with immobilized carbohydrates was introduced to isolate plant lectins [47]. Later, the application of lectins to isolate glycoproteins as well as to separate glycoproteins into their glycoforms was discovered. This technique is for example used to fractionate varying glycoforms of IgG with different degrees of sialylation (using the sialic acid specific *Sambucus nigra* agglutinin), which can be correlated to different degrees of anti-inflammatory activity [48]. The finding of the ability of PNA (peanut agglutinin) to agglutinate immature thymocytes can be used as a marker for thymocyte maturation [49] which subsequently paved the way for the usage of SBA (soya bean agglutinin) and PNA to fractionate mouse bone marrow and spleen cells for bone marrow transplantation across histocompatibility barriers [50]. Furthermore SBA was used for purifying human bone marrow of haploidentical donors for subsequent transplantation into children born with SCID (severe combined immune deficiency or “bubble children”) which led to the cure of 75% of bubble children [51].

As it was found that lectins do not only occur in the plant kingdom, but also in bacteria and mammals, great effort was done to investigate lectins function and abilities of usage. *E-coli* was found to bind to buccal epithelial cells via mannoses, as the binding was inhibited by adding mannose and methyl α-mannoside as well as by the pre-treatment of the cells with the mannose binding lectin concanavalin A (ConA). This gave arise to the conclusion that lectins can function in cell-cell recognition [52].

The first mammalian lectin was a galactose specific liver lectin, which binds serum asialo-glycoproteins and mediates the uptake and eventual degradation in the liver [53]. The discovery of the selectins furthermore proved the function of lectins in cell recognition as they control lymphocyte migration to specific lymphoid organs and sites of inflammation [54]. Since then it was found that glycan-dependent lectin-mediated interactions are crucial for the adaptive and innate immune system as they regulate immune cell trafficking and play a role in synapse formation, activation and survival [55].

1.5 Role of specific glycan linkages/structures (isomeric glycans) in biological systems

Isomeric glycans are highly common within the complex field of glycosylation. As the focus on glycan analysis gained more and more importance within the last decades the importance on the identification of isomeric glycosylation emerged simultaneously. Not only the appearance of distinct isomers can be crucial for many biological systems (e.g. differently linked sialic acids mandatory for Influenza A virus receptor binding), also the relative quantity of specific isomers can be linked to certain diseases and therefore can be used as a biomarker for cancer diagnosis or progress [56].

This study focuses on the investigation of glycan isomers by different biochemical and analytical methods. The biological samples used are the lectin hemagglutinin from Influenza A virus and Lewis type fucoses in the S-chain of IgA and in the mouse brain.

1.5.1 Influenza A virus

The key player in the influenza A virus infection process is hemagglutinin (HA), one of the three integral membrane proteins found on the viral surface. In 1981 the three dimensional structure of this meanwhile well studied membrane protein was resolved by [57], rendering the large amount of already existing data for this protein newly interpretable.

Hemagglutinin has three major functions that make it highly important and crucial for the entire viral infection process. The first and for this study probably most interesting function is the binding of hemagglutinin to sialic acid-containing receptors on the host cell surface that initiates the viral infection cycle. The second function is the hemagglutinin-mediated entry of the virus into the cytoplasm of the affected cell by membrane fusion achieved by the HA itself. Third, HA functions as the major viral antigen against which most vaccinations are designed by producing neutralizing antibodies [58]. Glycosylation plays a crucial role in the infection process as the hemagglutinin binds to glycoproteins (receptors) that are decorated with N-linked glycans bearing terminal sialic acids. These sialic acids can be either α2,6 or α2,3 linked to the underlying galactose residue. This small difference in the linkage (isomeric glycans) has a remarkably huge effect considering the species specificities of influenza strains: It is known that human viruses prefer to bind α2,6-linked sialic acid whereas avian and equine viruses prefer an α2,3 linkage. Viruses infecting swine appear to have the ability to recognize both sialic acid linkages [59] [60]. These specificities reflect the relative abundances of differently linked sialic acids on the tissue sites of infection. Lectin labelling of the respiratory tract from different species with specific lectins, binding either α2,6-linked (*Sambucus nigra* lectin - SNA) or α2,3-linked sialic acids (*Maackia amurensis* lectin – MAL), support this theory [61]. However, the binding preferences can be broader than just the differently linked sialic acid. For example, avian viruses share the preference of α2,3-linked sialic acid to Gal-2 but viruses from different avian species can alternate between linkages of Gal-2 to GlcNAc-3. The highly pathogenic H5N1 from Hong Kong 1997 prefers Galβ1,4GlcNAc and sulphated-GlcNAc whereas H5N1 from ducks prefers Galβ1,3GlcNAc [62].

The binding specificity of hemagglutinin is highly conserved among the different existing subtypes. Hemagglutinin is a homotrimer and each monomer contains a receptor binding site at the distal tip. At the base of each monomer Tyr-98, Trp-153, His-183 and on the distal

edge Tyr-195 are conserved amino acids. Similarly, three secondary structure elements are conserved within the protein family (the 130- and 220-loops and the 190-α-Helix) [63]. Crystal structure experiments of a recombinant H1 hemagglutinin from H1N1 swine-origin influenza A virus Cal04/09 co-crystallized together with LSTc and LSTa (pentameric glycoconjugates NeuAcα2,6/3Galβ1,4GlcNAcβ1,3Galβ1,4Glc with either α2,6 or α2,3-linked sialic acid respectively) were performed and compared to binding data obtained in a previous study where H3 was described by [63]. Electron density for LSTa (the α 2,3 glycoform) in the catalytic centre of H1 was very weak and only Sia-1 and Gal-2 could be modelled for glycan ligand interactions. The α2,6 glycoform in contrast showed a high electron density and the catalytical centre with LSTc could be modelled for Sia-1, Gal-2, GlcNA-3 and Gal-4. The crucial amino acids in H1 (H1N1) that allow human receptor binding specificity and viral transmission in humans are Asp190 and Asp225 as they allow for a reliable recognition of α2,6-linked sialic acid [64]. When it comes to the hemagglutinin binding specificity of other viruses such as H3N2 (A/Hong Kong/1/68), other amino acids in the catalytic centre are known to be responsible for the distinct sialic acid binding prevalence. Therefore the substitutions of a glutamine at positions 226 by a leucine together with a serine instead of a glycine at position 228 (L226Q and S228G) result in the shift of a human virus like preference (α2,6 linked sialic acid) towards an avian virus like receptor specificity (α2,3 linked sialic acid) [65]. The mutations necessary to change the receptor binding specificity of Influenza A virus subtypes usually occur in genetic events called antigenic drift, which is a genetical re-assortment of two different viruses infecting the same cell, though interestingly most epidemic viruses are of avian origin [66].

1.5.2 Lewis – type fucose epitopes

The second focus of this work is the elucidation of the impact of differently liked fucoses to the glycan moiety on the retention times and the fragmentation pattern.

A Fucose is a deoxy-hexose existing in a wide variety of organisms. In mammals fucose-containing carbohydrates play an important role in many biological contexts. In the *ABO* blood group antigens, fucosylated glycans on erythrocyte surfaces differ in their structural appearance. An α1,2 fucosyltransferase expressed in erythroid precursors synthesizes the so-called H-antigen (Fucα2Galβ4GlcNAc-R). Within the different blood groups the *ABO* gene

locus encodes an N-acetyl galactosamintransferase (adding a terminal GalNAc residue) on the A allele, a galactosyltransferase (adding a terminal Gal residue) on the B allele and an inactive truncated protein leading to no further modification of the glycan structure on the O allele [67]. Another very well-investigated effect of fucose is the essential role as the carbohydrate ligand for the selectin-mediated cell adhesion receptors. Selectins are C-type lectins that are expressed by plathels (P-selectin), endothelial cells (E- and P-selectins) and leukocytes (L-selectins) and they bind to carbohydrate ligands carrying sialyl Lewis x structures. These structures are catalysed by two α1,3 fucosyltransferase resulting in Siaα3Galβ4[Fucα3]GlcNAc-R glycan epitope [68] [69]. Additionally, fucosylated blood group antigens can be important in host-microbe interactions as e.g. *heliobacter pylori*, a gastric pathogen, recognizes the Lewis b antigen, which is a modified H-antigen carrying an α1,2 and an additional α1,4-linked fucose to the GlcNAc residue in the gastric epithelial [70].

More recently, investigations have suggested the importance of fucosylated glycoproteins to brain function and development. In a study presented by Bleckmann et al 2009 the cell-adhesion glycoprotein CD24 was investigated according to its glycosylation profile, as it has been proposed that CD24s functions are mediated by its N- and O-glycans. CD24 plays an important role in the central nervous system (CNS). It is highly expressed on immature neurons and other neuronal cells during the CNS development. Furthermore CD24 is expressed by hematopoietic cells like T- and B-lymphocytes influencing the immune system. The glycomic profile revealed in addition to α2,3-linked sialic acid, which modulates the interaction of CD24 and the cell adhesion molecule L1 [71], the occurrence of highly complex multi Lewis x fucosylated glycan structures. Moreover, to a lesser extent, additional H antigens were found [72]. Other studies investigated Fucα1,2Gal glycan epitopes, which are known to occur on proteins important in the molecular mechanisms underlying neuronal development, learning and memory. This was shown in α1,2 fucosyltransferase deficient mice, which exhibit developmental defects like smaller and fewer glomeruli and a thinner olfactory nerve layer [73]. The same glycan epitope is also found on synapsin Ia and Ib, synaps-specific proteins involved in synaptogenesis and neurotransmitter release. Defucosylation leads to delayed synaps-formation and stunted neurite outgrowth, implying the importance of Fucα1,2Gal glycans in the regulation of neuronal proteins and morphological changes maybe underlying synaptic plasticity [74].

In addition to the influence of fucosylated glycan-epitopes on brain development, correlations of fucosylated glycoproteins to cancer metastasis have been found. The up-regulated activity of certain fucosyltransferase in cancer cells leads to a higher occurrence of sialyl-Lewis x and sialyl-Lewis a glycan epitopes, which can be used as a tumour marker [75].

All in all, many examples show the necessity to be able to distinguish between glycan isomers. On the one hand this knowledge can help in order to design vaccines against different influenza A viruses but on the other hand analytical tools for selective, sensitive and quantitative high-throughput screenings of biological glycan samples are sought.

1.6 One experimental mass but thousand possible structures (Isomeric / isobaric glycans)

Isobaric structures are defined as compounds of the same nominal mass but of different elemental composition whereas isomeric structures consist of the same nominal mass as well as the same elemental composition differing in the linkage of the substituents [76] [35].

The importance to distinguish between isomers can be elucidated in the example of Influenza A virus binding specificities. The binding and initiation of the infection process is achieved by one of the three integral membrane proteins of the viral surface, the hemagglutinin. Hemagglutinin acts as a viral lectin as it specifically binds sialic acid-terminating glycan structures on the host cell receptors. The linkage of the sialic acid is the crucial factor that protects different species from cross species infections as human Influenza A viruses specifically bind α2,6-linked sialic acid whereas avian Influenza A virus only recognize an α2,3 linkage [60].

The ability to distinguish between glycan isomers is not only of biological importance but also an analytical challenge. One possible way to meet this challenge is to chromatographically separate fluorescently-labelled glycans in a two or three dimensional HPLC (high performance liquid chromatography) approach [77] [78]. However, this approach is quite laborious and often requires for additional mass spectrometric detection and verification of the glycan structures. Especially when considering the analysis of complex biological systems the retention time alone might not always be conclusive enough.

A second classical approach for structural elucidation of glycans is the monitoring of products from exoglycosidase digestions by gel filtration chromatography [79] HPLC [80] or MALDI-MS [81]. Exoglycosidases give structural information about the number, species and linkage of the removed monosaccharides [82]. However, the drawback of using exoglycosidases is their limited availability in sufficient amount and purity, which can often result in misleading structural conclusions.

The approach of MS/MS fragmentation gained more and more importance within the last years as even small amounts of glycan can be fragmented. The most frequently used techniques are MALDI TOF-MS and ESI-TOF-MS, two fairly sensitive ionisation methods coupled to a time of flight (TOF) mass analyser. MALDI-TOF-MS is a relatively simple and fast technique that allows a high mass range but only small concentrations of contaminants. However, a limitation of this method is the analysis of native sialylated glycans as the sialic acids are fairly labile and often get lost during the ionisation. This can be circumvented by esterification of the acidic groups [83]. Moreover, MALDI-TOF-MS only provides an overview about possible glycan structures while giving poor structural information in terms of sequence and linkage of the single sugar residues, and thus does not provide information about isomeric glycans [84]. An ESI-source MS coupled to a liquid chromatography (LC) can be used to overcome this problem as the addition of a high-resolving separation technique can give additional retention time related structural insight. The preferred method, which was also used in this work, is the combination of a quadrupol time of flight, or ion trap mass spectrometer with an UPLC using a porous graphitized carbon (PGC) column for chromatographic separation, which has the ability to distinguish between structural isomers as shown in Figure 11 by [35].

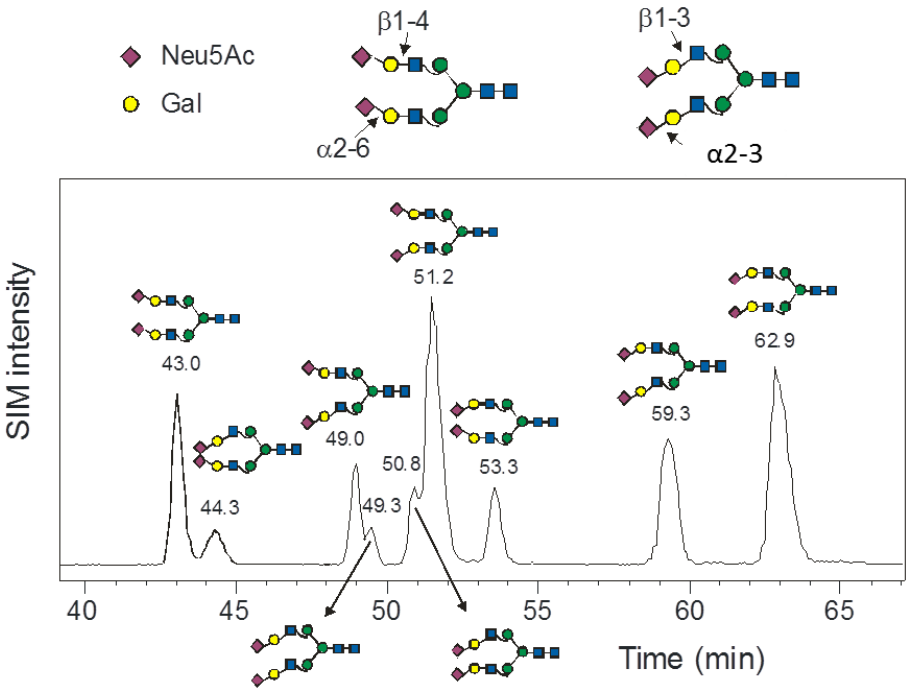


Figure 11 Isomeric distribution of all possible 9 isomers of NaNa using a PGC-column, taken and modified from [35]

The additional fragmentation of isomeric glycans in negative ionisation mode reveals highly structural-specific "diagnostic fragments or feature ions", which are specific ions resulting from glycosidic- and cross ring cleavages that are only found in the fragmentation pattern of glycan structures carrying specific epitopes [85] [86] [87].

This combination of reduced glycans separated on a PGC-column with mass detection and subsequent fragmentation in negative ionisation mode using an ion trap mass spectrometer leads to a rather accurate possibility to structurally identify glycans, even within a rather complex biological sample [88].

1.7 Detailed investigation of released glycan structures

Glycomic and glycol-proteomic investigations are of considerable interest and its importance is widely accepted since the nature of the ABO blood group glycans was recognized [89]. It has been reported that alteration in the glycan profiles can have sever functional influences as for example the slight modification of IgG N-glycans by the addition of one single sialic acid converts the formally pro-inflammatory IgG molecule into an anti-inflammatory acting antibody [48]. Furthermore the binding of the IgG molecule to the FcγRIIIa, which initiates antibody dependent cellular cytotoxicity (ADCC), can be disabled by an additional core fucose on IgG Fc N-glycans [90].

As the structural variability of glycans leads to considerable challenges it is necessary to use the appropriate biochemical and analytical techniques to gain enough insight into the glycan species, monosaccharide composition and linkage of the different sugar residues.

The starting point in glycomic analysis workflows commonly starts with the enzymatic release of N-glycans using N-glycosidase F (PNGaseF) or peptide-N4-(N-acetyl-β-glucosaminyl)-asparagine amidase F from *Flavobacterium meningosepticum* to cleave N-glycans from the intact protein backbone if no core α1,3 fucose is present Figure 12. Glycoproteins carrying glycans with an α1,3 core-fucose, which is known to exist in plants and insects, can be release after enzymatic protein digestion from the peptide backbone with N-glycosidase A (PNGase A) isolated from almonds [91] .

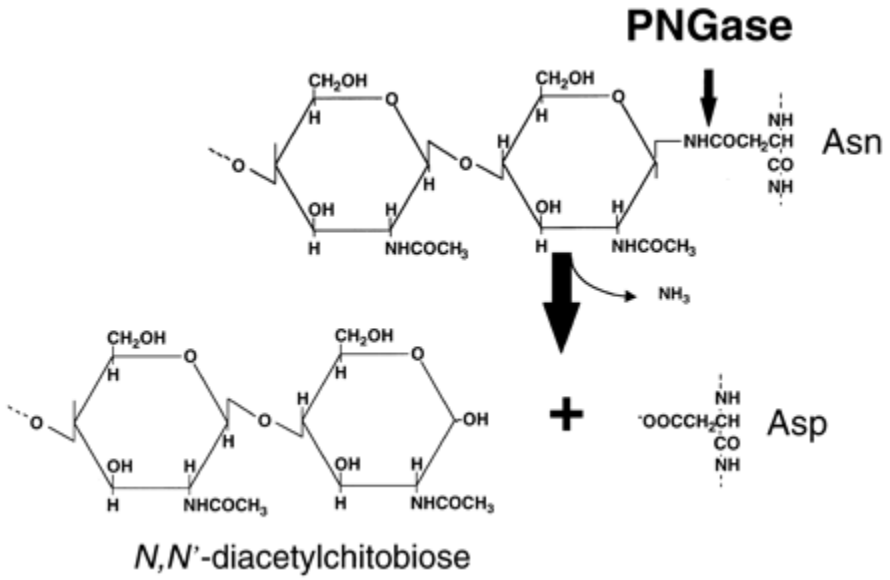


Figure 12 Schematic cleavage of N-glycosidase taken from [92]

The N-glycosidase cleaves the bond between the N and C atom from the N-acetylglucosamine and the asparagine (Asn) of the protein backbone converting the Asn to aspartic acid and the released sugars to a N-glycosylamine. The latter is subsequently hydrolysed at slightly acidic pH to an intact reducing end [92].

Beside the enzymatic release, which only functions for N-glycans as there is no O-glycanase available, oligosaccharide structures can be released from the protein backbone by chemical treatment. Both glycan species (N- and O-glycans) can be released by anhydrous hydrazinolysis though with this technique all N-acetylgroups are cleaved and subsequent "re-N-acetylation" is necessary [93]. The ability of hydrazine to cleave N-acetyl groups

though can be used for isotopic labelling of glycans by re-N-acetylation with deuteroacetyl as described in 3.3.3.

For O-glycan-release alkaline as well as hydrazine degradation can be used. The alkaline release can by performed by the reductive alkaline β-elimination where glycoproteins are incubated in a sodium hydroxide solution containing 1-2 M sodium borohydride ($NaBH_4$) at 45 °D for 16 h [94]. If the released sugars are needed for further reductive derivatisation the glycans can also be eliminated by 28% ammonia in saturated ammonium bicarbonate buffer at mild temperatures over 40 h [95] or by ethylamine [96]. The mostly used type of non-reductive glycan release though is mild hydrazinolysis [97].

1.7.1 NMR spectroscopy

The most conclusive, detailed and definite answer about structural properties of glycans is given by NMR (Nucular Magnetic Resonance) spectroscopy. With NMR-spectroscopy, all structural features including monosaccharide composition, ring size and anomericity, the conformational preference and linkage of the monosaccharide residues as well as type and position of possible substituents can be assigned. Complete structural assignment of a glycan structure requires both the ^{1}H and ^{13}C-NMR spectra of the glycan. The drawbacks of NMR spectroscopy lie in the requirement of a large amount of highly pure glycan sample which is hardly ever the case when working with biological samples and glycomic mixtures [98].

1.7.2 Glycan arrays

Beside well established biochemical and analytical tools like MS, HPLC, NMR etc. a lot of data about glycan expression in biological systems has been obtained by anti-glycan antibodies or lectins. It is crucial to completely understand the glycan binding specificities of these molecules for the analysis of glycans in biological systems. In the last years glycan microarrays have emerged more and more as very suitable tool to investigate these glycan binding specificities [99]. The traditional methods to evaluate the interaction of glycans with glycan binding proteins are isothermal calorimetry [100], surface plasmon resonance [101] or enzyme linked lectin assays [102]. Glycan arrays are used as a novel high throughput screening method for glycomic studies evaluating the glycan binding of proteins, antibodies, viruses or whole cells. Another advantage of glycan arrays is the mimicry of natural cell

surfaces due to the multivalent presentation of glycan epitopes in a narrow space. The basis for large scale simplified glycan synthesis was reported by Wong [103] [104] and Seeberger [105]. These developmental advances could be used to generate a high diversity of synthesized biomolecules. The consortium of functional glycomics [106] is an online resource providing synthetic glycan arrays, synthetic glycans as well as results of tested arrays with various biological counterparts (virus, antibodies, recombinant proteins etc.). An alternative approach is to generate natural-glycan arrays, which was also performed in this work. Therefore glycans were released from natural sources, labelled and separated prior printing the glycan moieties onto array slides [107]. Many different glycan array fabricates are available reaching from covalent or non-covalent bound glycans, numerous different possibilities of surface linkers (e.g. NHS esters, epoxide, amine..), differently tagged glycans (e.g. Thiol, Amine, Neo-glycolipids…) to highly diverse substrates (e.g. Glass, Nitrocellulose, Microtiter…), though in the most common approach differently coated glass slides are used for glycan immobilization. A detailed list of the different glycan arrays is available in a review by Liang and Wu, 2009 [108]. As new technologies emerge all the time the task to find the appropriate glycan array matrix to gain optimal results can be a challenging task, as the quality of the substrate, surface linker and tagged carbohydrates influences the results and lead to misinterpretation [99].

Nevertheless glycan arrays are a valuable technological novelty and can be used for many biological applications. Van Diepen et al. 2012 generated glycan arrays of glycans isolated from the different life stages of the human parasite *schistosoma mansoni* and applied IgG and IgM antibodies in sera from children and adults living in endemic areas to the array, as the immune response triggered by the parasite, is directed against glycan epitopes [109]. Within the last years glycan arrays also gained high importance in assessing the binding specificities of viruses. A recent study for example showed specificity changes from pandemic H1N1 influenza A virus towards an elevated binding to α2,6 as well as α2,3 linked sialic acid due to a D222N mutation in the catalytic centre of the hemagglutinin [110]. Other studies compared different H1N1 viruses according to their binding specificity, again showing decreased over all binding and specificity of the A/California/04/09 strain, which was used as a reference strain [64]. The immense amount of binding data obtained in a single glycan array experiments logically calls for bioinformatical software for analysis but also comparison of published findings. Therefore a recently published study used a data

mining technique to analyse already available data of influenza virus glycan arrays and found a sulphated glycan motif that also increased viral infection in cells overexpressing the responsible sulfotransferase [111].

1.7.3 Lectin-Affinity chromatography

Lectin affinity enrichment of glycoproteins from proteomic mixtures is a widely used and well established method. Affinity columns, where the lectins are immobilized, can either be used to enrich a specific glycoform according to the used lectin, or to enrich the entire glycoprotein pool if different lectin sepharoses are used [112]. A detailed structural glycan assignment though does not seem to be achievable with the application of lectins due to their generally broader binding specificity compared to antibodies. The well characterized leguminose concanavalin A (ConA) lectin for example is known to bind oligomannosidic structures but furthermore also shows recognizable binding of the trimannosylcore of complex N-glycans [113]. The binding specificity of lectins though can be engineered by genetic mutations of the carbohydrate –recognition domains (CDR'S) and therefor can be used for detection or isolation of glycoproteins bearing specific glycan epitopes [114].

1.7.4 Separation techniques coupled to MS

1.7.4.1 Overview of common separation techniques

Released glycans can be detected by refractive index, pulsed amperometric detection [115] and UV detection at 210 nm after separation by HPLC. These detection methods though are often not very sensitive and also detect coeluting impurities [116].

The complexity of glycan structures occurring in biological systems, especially the high abundance of isomeric and close related structures, calls for a high performance separation prior to detection. The used tool in this work is liquid chromatography which was invented over hundred years ago by Mikhail Tswett (Tsvet) [117]. This first demonstration of a chromatographic separation built the technological base for the very powerful separation approaches used nowadays. The method of high performance liquid chromatography (HPLC) was finally introduced by Howard and Martin 1950 [118]. Since then a broad facility of different stationary phases for all kinds of analyses were developed [119]. Two types of

columns are nowadays commonly used as stationary phases for routine HPLC. One type is a monolithic column, where porous silica and cross linked polymers are used as stationary phase and thereby give fast separation at low pressure, which is fairly suitable for large biomolecules. The drawbacks of this stationary are the poor reproducibility and weak stability of the material, which makes their use in glycan analysis rather obsolete. The second type of column is the packed, mostly silica based column with porous, non-porous, fused core, or recently also developed core shell particles with particle sizes smaller than 2 µm. These columns require high performing LC (UPLC) instruments but also lead to highly efficient separation and shorter columns with higher flow rates [120].

1.7.4.1.1 Reversed Phase Chromatography

Reversed phase columns are widely used for many biological applications as they are cheap to produce, robust and the packed silica as the stationary phase can by versatile chemically modified. However hydrophilic compounds such as released unmodified glycans hardly interfere with the stationary phase due to their hydrophilic character. Using ion pairing of acidic glycans with amine compounds in the mobile phase, retention of the glycans can be obtained, though the ion pairing reagent can lead to signal suppression in subsequent MS analysis [121]. Alternatively released glycans can be subjected to reductive amination via a Schiff base to increase hydrophobicity and thereby interfere with the column [122]. Commonly glycans are derivatised with 2-aminopyridin (PA), 2-aminobenzamide (2-AB), 2-aminobenzoic acid (2-AA), 2-aminonaphthalene trisulfonic acid (ANTS) or 1-aminopyrene-3,6,8-trisulfonic acid (APTS) [123]. Labelling procedures, however, are a time consuming task and need subsequent purification to remove the excess label reagent. Furthermore partial sugar degradation of the more labile sugars like fucose or sialic acids can occur [124] [125]. Structural glycan variability is not only constituted due to different glycan species but also because of isomeric differences. The combined approach of 2-AA labelled glycans separated by reversed phase chromatography leads to isomeric separation of oligomannosidic glycans as well as AGnF and GnAF which could not be shown with 2-AB labelled glycans [126].

1.7.4.1.2 Normal Phase Chromatography / HILIC

The term hydrophilic interaction chromatography (HILIC), a variant of normal phase chromatography was first defined by Alpert 1990 [127] and is widely used for the analysis of carbohydrates. The principle of separation lies in the interaction of analyses with highly

polar (hydrophilic) stationary phases by applying a polar mobile phase with high concentrations of organic solvents. Typically a binary gradient is applied using an aqueous part and a less polar organic solvent (usually acetonitrile). Increasing the water concentration leads to the elution of hydrophilic molecules such as glycans [128]. The retention of the analyte on the stationary hydrophilic phase is achieved by hydrogen bonding (separation according to size), ionic interactions according to charge and dipole-dipole interaction [129]. In contrast to traditional adsorption mechanisms on normal phase materials, HILIC stationary phases, such as silica particles, amine-, hydroxy-, amide-bonded or zwitterionic (ZIC-HILIC) particles involve a partitioning mechanism [127] [129]. Silica based columns can exhibit charges as basic pH due to deprotonation of the silanol groups. This leads to a higher retention of basic compounds when the pH is below the pKa of the molecules. In contrast to that amine and aminopropyl phases act as weak anion exchange columns when performed at pH below 9 due to the positive charged amine groups [130]. ZIC-HILIC phases, however, reveal constant positive and negative charges which act as weak cation-exchange columns [129].

HILIC columns can be used to separate fluorescently labelled or native glycans according to their size and hydrophility are often applied in LC-MS approaches or multidimensional HPLC separations, fractionations and subsequent MALDI-MS mass detections [82]. It is often stated that HILIC columns are not suitable for isomeric separation [131]. But various research groups have shown that, taking the right HILIC material isomeric separation can be achieved to some extent. The widely used TSK gel Amide-80 columns 3 µm particle size can separate 2-AB labelled complex sialylated N-glycans from bovine fetuin and separation is further improved using a BEH Glycan column with 1,7 µm particle size and a UPLC system [132]. Furthermore the application of ZIC-HILIC columns has also been shown to be suitable for at least partial isomeric separation of PA-labelled sialylated N-glycans [133]. Additionally, as the separation according to size and hydrophility is a very conclusive one, the combination with an Ion mobility mass spectrometer can lead to a confident structural identification of N-glycans [134]. An ion mobility chamber separates isomers due to three dimensional shape differences in a high gas pressure chamber prior MS or tandem MS [131].

1.7.4.1.3 Porous graphitized carbon (PGC) column

If the isomeric separation and structural elucidation of glycans is the main purpose of a chromatographic separation porous graphitized carbon columns are a remarkable good choice. About 30 years ago Knox and Gilbert invented the preparation of porous graphitized carbon particles of 7-5 µm size [135]. The available Hypercarb columns (Hypercarb^{TM}, Thermo, Electron Corporation) are 100 % porous graphitic carbon with spherical and fully porous particles. Flat sheets of hexagonally arranged carbon atoms compose the PGC surface as a very large polynuclear aromatic molecule [136]. The PGC material can be considers as strong or modified RP material as one mechanism of retention is hydrophobic interaction. This effect is very strong as non-polar molecules like proteins or peptides are very difficult to desorb again, once bound to the material [137]. The second type of interactions are electronic forces by increased pi-electron interactions. These are charge-induced interactions of the polar analyte with the polarized surface of the graphite, which leads to increased retention of highly sialylated (charged) glycans. Furthermore the strength of the retaining interactions also depends on the molecular area that is in contact with the graphite surface also depending on type and positioning of the attached functional groups and glycan shapes leading to retention differences of isomeric and even isobaric glycan structures [136] [138] [137].

Taken together graphitized carbon is known to strongly bind glycans under aqueous conditions and the retention increases with the glycan size and number of charged (acidic) groups including sialylation, phosphorylation and sulfatation of N-glycans, O-glycans and glycosaminoglycans. To obtain a stable chromatographic surrounding, constant pH and ionic strength is necessary for peak shapes and elution times. Additionally a grounding of the column is necessary if a coupling of the PGC column to an ESI source, if an LC-MS approach is performed. This is necessary as the PGC material is a conductive material and electrosorption of acidic and basic molecules can occur originating from the high voltage ESI source [139].

Graphitized carbon is an extremely stable material tolerating high temperatures and acidic conditions. Therefor a clocked column with high back pressure can be regenerated by treatment with 4 M TFA under boiling conditions [94].

All these mechanisms together, lead to a high chromatographic resolution and additionally to a good separation of structural glycan isomers. In combination with mass spectrometric detection retention time libraries can be developed, against which unknown biological samples can be referenced and structural elucidation can be performed [140]. For the generation of retention time libraries released N-glycans should be reduced to prevent splitting of the peaks due to the anomeric reducing end, which is typically done directly after the enzymatic cleavage of the N-glycan from the protein backbone. The released glycans are incubated at basic conditions with the reducing agent $NaBH_4$ for several hours at room temperature and subsequent purification is performed to remove excess salt by using solid phase extraction (SPE) PGC cartridges [35].

1.7.4.1.4 Capillary electrophoresis

The separation principle of analytes in capillary electrophoresis (CE) lies in the migration velocity of charged analytes in an electric field. This electric field is applied to the ends of a capillary column and allows high speed of analysis and high separation efficacy of charged analytes like sulphated or sialylated glycans. Neutral glycans require complexion with borate or labelling with a charged fluorophor, like ANTS or ATPS, to obtain a charge. The usage of borate buffers (non-volatile buffer) is especially suitable for isomeric separation with subsequent laser induced fluorescence detection [141]. Due to the fast and efficient separation properties coupling to MS instruments is attractive and has been realized by various groups [142] [143].

1.7.4.1.5 Affinity chromatography

Affinity chromatography is a method widely used for glycan analysis and analysis of glycan binding proteins. Beside the purification purpose, which is one of the major biochemical applications, it can also be used to investigate binding specificities. This is realized either in the commonly used batch-approach, which is elucidated in Figure 13 (A), where the basic principle is loading a sample, incubating, washing away unbound molecules and elution of the bound molecules. Another very conclusive affinity approach is called frontal affinity chromatography (FAC) (Figure 13 B), which is used for molecules that are retained but not necessarily completely bound to the stationary phase. With this method a mixture of e.g. glycans is constantly loaded on a column and due to different binding affinities the molecules are differently retained and reach a detector unit at different time points.

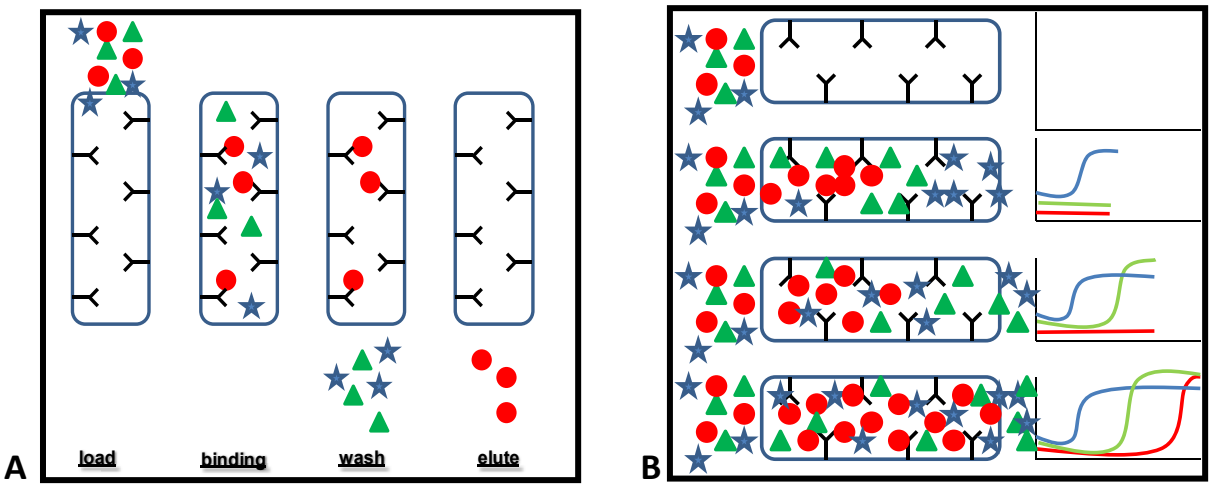


Figure 13 Affinity chromatography methods
Schematic overview of two experimental procedure of affinity based chromatography. (A) Commonly used batch approach with four experimental steps of loading, binding, washing and eluting. (B) frontal affinity chromatography (FAC) with continuous loading and online monitoring of retained molecules.

1.7.4.2 Overview of mass spectrometric techniques

Nowadays the majority of published work about structural analysis of carbohydrates is performed with mass spectrometry. This analytical tool offers accurate results high sensitivity and high versatility. Commonly used techniques in glycomics are matrix assisted laser desorption ionisation (MALDI) and electron spray ionisation (ESI) [121]. LC systems can either be combined with subsequent offline MALDI-MS detection where glycan fractions are spotted and can be re-analysed or with a direct online approach the LC-ESI-MS. One advantage of the online approach lies in the possibility of peak quantification [144].

1.7.4.2.1 Ionization techniques

Matrix assisted laser desorption ionization (MALDI) is a soft ionization method used for large, labile and non-volatile molecules like proteins, DNA fragments or carbohydrates. The sample is spotted on a target plate dispersed in matrix and crystallized upon solvent evaporation. MALDI is a very suitable method for the detection of neutral oligosaccharides due to the tendency of the technique to produce $[M+Na]^+$ ions or to a smaller extend $[M+K]^+$ ions in contrast to electron spray ionisation where differently charged ions and with different adducts can occur. On the other hand MALDI suffers from problems with charged groups found on sialylated, sulphated or phosphorylated glycans. There glycans are rather unstable in MALDI and need to be analysed under special conditions or preceding

derivatisation [121] [145] [144]. To overcome the problem permethylation of the glycans can be performed to prevent fragmentation during the MALDI process, which additionally gives advantages when performing MS/MS analysis [144] [131]. The matrix usually consists of low molecular weight molecules with optical absorbance in the UV or IR range containing acidic groups, which serve as proton donors during the ionization process. During the crystallization process the matrix integrates the analyte into their crystals and by adsorbing the laser energy the matrix gets ionized subsequently transferring the charge to the analyte. This occurs in a molecular plum created by the irradiation with the pulsed laser light [146]. Commonly used matrices for the investigation of glycan structures are DHB (2,5-dihydroxybencoic acid) or THAP (2,4,6,-trihydroxyacetophenone) which can be used specially for acidic glycans [146].

The second most predominantly used ionization technique is electron spray ionisation (ESI), which belongs to the atmospheric pressure ionization techniques. A solution containing the analyte is infused into the source through a conductive needle, where an elevated electric field in positive or negative polarity is applied. The high voltage together with an inert nebulizer gas produces a fine spray of charged droplets that evaporate to form analyte ions. During the ionization process almost no vibrational energy confers on the analyte preventing in source fragmentation. Furthermore sialic acids are generally not lost with this technique and all types of glycans can be analysed in positive as well as negative ionisation mode, thus differing in the relative intensities [144]. The generated ions are mostly doubly and triply charged proton adducts in contrast to the singly charged sodium adduct ions occurring in MALDI. This is not only influencing MS/MS experiments, as the fragmentation pathway considerably differs, but also produces less conclusive spectra [121] [144]. ESI can be used coupled to a LC system or by direct infusion, though the separation using a LC-system prior the ionisation and subsequent detection facilitates latter quantification of glycan structures and minimizes sample impurities that would lead to signal suppression [121] [35].

1.7.4.2.2 Mass analysers

After the ions are generated the detection of the analytes in a mass analyser is the next crucial step in glycomics to gain the most information. The important criteria are sensitivity, resolution, mass accuracy and speed of data acquisition [144]. Furthermore the possibility to perform MS and additional MS/MS or even MS^n experiments can lead to detailed and fast

data acquisition which simultaneously results in a load of data generated in a single measurement and calls for subsequent computational tools for interpretation [88]. The ESI-source is often coupled to a classical ion trap (IT), which shows on the one hand limited resolution, but on the other hand the capability to perform MS^n experiments, which is crucial for structural elucidation of glycan structures down to the isomers [88]. A QTOF (Quadrupol-Time of Flight) instrument, another commonly used mass analyser coupled to an ESI-source, provides a higher dynamic range and mass accuracy as well as a far better resolution. The breakthrough in terms of mass accuracy and resolution is represented by the Fourier transform-ion cyclotron resonance (FT-ICR) mass spectrometers [147]. The Orbitrap-mass spectrometer separates ions in an oscillating electric field and thereby performs analysis with a similar mass accuracy and resolution like FT-ICR [144] [147]. Collision induces fragmentation (CID) is a well-established fragmentation method for glycan analysis. Depending on the ion polarity applied to the MS instrument different fragment-species are obtained. The fragmentation usually takes place in the collision cell where ionized molecules collide with gas atoms rupturing the weakest bond of the native, reduced or reductive aminated glycans. In positive ionization mode the glycan fragments typically are b- and y ions whereas in negative ionization mode additional cross ring cleavages (a-ions and x-ions) and D as well as D-18 ions are produced, resulting in the appearance of so called diagnostic ions, which give detailed structural information about antennae compositions and for example isomeric fucosylation [131] [88]. Wuhrer et al 2011 described the phenomenon of rearrangements of protonated ions during tandem MS analysis which can be circumvented by analysing e.g. sodiated (cationized) ions [148]. Another limitation occurs when analysing sialylated native glycans by MALDI-TOF-MS. Sialic acids tend to get lost during the ionization process which can be circumvented by esterification of the acidic groups [83]. Furthermore permethylation of glycans in combination with CID fragmentation using MALDI-TOF/TOF-MS instruments leads to kinetically controlled bone rupture and cross ring cleavages which are particularly useful for determining linkages [131].

Ion mobility spectrometry (IMS) has become a very suitable tool for investigating isomeric glycan structures as the glycans separate according to the molecular shapes. Drift times can be measured when the molecules drift through an electric field colliding with a counter flow of neutral gas [134] [131].

Taken all the different possibilities together the most suitable approach for the elucidation of glycan structures, that was used in one part of the subsequently presented work, is a combination of UPLC separation of reduced glycans using a PGC-column coupled to an ESI-QTOF-MS or ESI-IT-MS/MS operated in positive or in case of the ESI-IT-MS/MS approach also negative ion mode. This method allows for both the retention time correlated structural assignment of glycan isomers and the additional MS/MS fragmentation which helps to distinguish between coeluting molecules and to confirm structural assignments according to the retention time.

2 Purpose of this work

Protein glycosylation is a crucial post translational modification essential for a number of different cellular functions and is even known to play a role in brain development as well as in memory formation. Furthermore, pathogens are known to take advantage of the presence of the unique carbohydrates. Influenza A Virus for example relies on recognition of and binding to distinct glycan epitopes in order to successfully invade the hosts.

The complexity of the underlying glyco-code is based on the possibility to generate a huge number of different glycan structures with just a few different monosaccharides. Many of the natural glycans occur as isomers, either due to linkages or positions of single or multiple residues or simply due to positions of further modifications on a saccharide

Sialylation and fucosylation are two prominent structural features that introduce a plethora of isomeric diversity.

Two methods for monitoring binding of sialo-oligosaccharides to carbohydrate binding proteins in an isomer specific manner should be applied at the example of H1 hemagglutinin of the H1N1 Influenza A virus. The one approach was affinity chromatography with structural elucidation by mass spectrometry and the alternative route was via glycan arrays generated by extensive fractionation of glycans present on the influenza virus host tissues.

For the much more complex isomer diversity of fucosylated glycans, porous graphitized carbon chromatography with mass spectrometric detection and additional MS/MS fragmentation should be developed as a tool for complete structural elucidation.

The unambiguous retention behaviour of the fucosylated reference structures should be used for retention time based identification of fucosylated glycans from secretory chain N-glycans from IgA and mouse brain proteins.

In addition to the structure of the glycans themselves their distinct position of N- and O-glycosylation on the protein backbone can give important quantitative and protein structure related information. Therefore the location of glycans on several model proteins, mostly expressed in green plants, should be analysed.

.

3 Materials and Methods

3.1 Biochemical methods for glycan arrays

3.1.1 Isolation of N-glycans

Approximately 165 mg dry tissue was grinded using a mortar, solubilized with approximately 2 ml PBS and subsequently homogenization using a potter. Solubilized tissue was brought to a volume of 5 ml with PBS and 1,3 % SDS and 0,1 % b-mercaptoethanole were added.

The solution was boiled at 95 °C in a water bath and incubated for 10 min to denaturate proteins and solubilize membrane proteins. To allow for subsequent tryptic digestions the SDS was neutralized with the addition of 1,3 % NP40 after cooling down the sample on ice.

The tryptic digestion was performed with the addition of 150 µl trypsin-coupled agarose beads to the protein solution followed by incubation at 37°C over night while shaking. The next day a centrifugation step was performed at 400 rpm for 3 min to spin down the beads. The supernatant was transferred to a new tube and 0,1 % acid and 22 µl N-glycosidase F (PNGaseF, Merck) were added followed by another 24 h incubation at 37 °C.

3.1.1.1 C18 – SPE (solid phase extraction)

To obtain remaining glycopeptides a C18 solid phase extraction (SPE) purification was performed. Therefore a 250 mg C18 SPE-cartridge was washed with 5 ml ACN (100%) followed by 3 ml 60/40 % ACN/0,1 % TFA, to remove possible impurities from the manufacturing process. Subsequent equilibration of the cartridge was performed with 5 ml dH_2O prior loading the sample on the column. The sample was centrifuged (10 min, 13.000 g) and the supernatant was applied on the cartridge followed by two washing steps with 2ml 10% ACN and 4 ml dH_2O. The flow through fractions were pooled and kept as they contain the released glycans. Peptides and remaining glycopeptides, which bind to the C18 column, were eluted in two steps with 4 ml 30 % ACN / 0,1 % TFA and a second elution step with 4 ml 60 5 ACN / 0,1 % TFA. The combined eluates were dried in the speed vac.

3.1.1.2 PGC – SPE (solid phase extraction)

The combined flow-throughs were used for PGC (porous graphitized carbon)-SPE purification to desalt the sample. In a first step the column was washed with 5 ml 100% ACN followed by another washing step with 3 ml 50 % ACN / 0,1 % TFA to clean the columns. The column was equilibrated with 5 ml dH_2O to obtain an hydrophilic environment for appropriate glycan binding. The flow through and wash-fractions of the preceding C18-purification were applied to the column and desalting was performed with a 5 ml dH_2O washing step. The glycans were eluted from the column in two steps with 3 ml 25 % ACN / 0,1 % TFA and additional 3 ml 50 % ACN / 0,1 % TFA. The combined eluates were dried in the speed vac.

3.1.2 AA – labelling of reducing Glycans

For fluorescent detection and for immobilizing glycans on a glycan array glass slide the purified reducing glycans were labelled with anthranillic acid (AA) which gives a fluorescent signal at excitation 330 nm and emission 420 nm. Furthermore the addition of AA to the reducing end of a glycan structure presents a primary amino group which is necessary for the subsequent binding to the epoxy silane coated glass slides which are used for array printing.

The labelling procedure starts with dissolving the dried glycans in 50 µl dH_2O. The labelling reagent is anthranillic acid dissolved at a concentration of 48 mg/ml in a solution containing 500 µl DMSO mixed with 150 µl acetic acid. The resulting AA solution is used to dissolve the 2-picoline-borane complex to a concentration of 107,33 mg/ml.

Equal amounts of glycan solution and labelling reagent were mixed properly and incubated at 65°C for 2-3 h. After the incubation the sample is cooled down to room temperature (RT) prior to purification to get rid of excess labelling reagent.

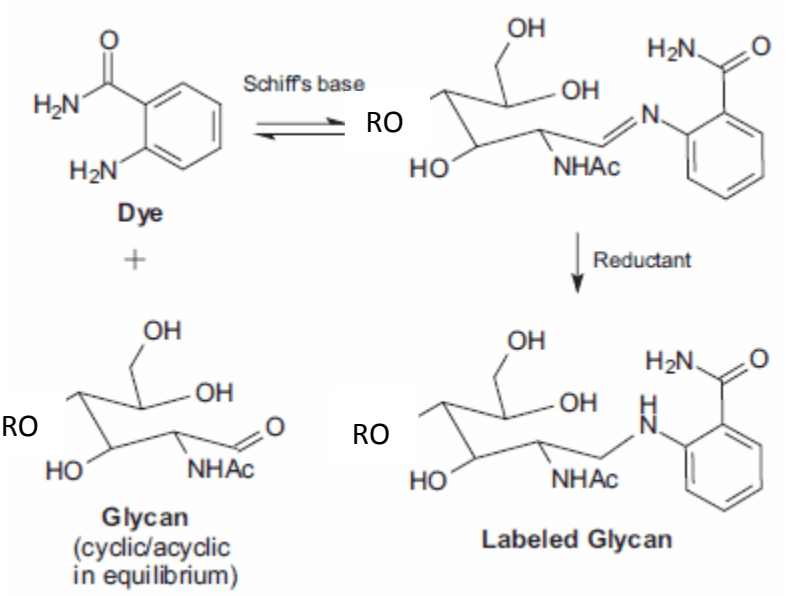


Figure 14 Reductive amination of released glycans [149]

3.1.2.1 Purification with biogel-SPE (Hydrophilic Interaction Liquid Chromatography - HILIC)

This purification method can be used to remove access labelling reagent as well as to desalt glycans to obtain high purity. For this method glycans can be native, reduced or fluorescent labelled as the interaction with the stationary phase is accomplished by the glycan moiety and the terminal modification of the glycan doesn't interfere with the binding to the stationary phase.

The stationary phase consists of polyacrylamide beads with a wet bead size of 45 – 90 µm (Biogel P10 – Bio-Rad Laboratories GmbH, Munich, Germany) and is dissolved in 10 % ACN (100 mg/ml). Prior usage the solution is stirred for 10 min to allow swelling. 200 µl of the biogel solution are added to each well of a acropep 96 well filter plate (0,45 µm pore size) which is mounted on a vacuum – manifold with a waste container for purification.

In a first step the excess liquid (10 % ACN) is pulled through by applying low pressure using a vacuum pump and each well is washed twice with 200 µl dH_2O each. Equilibration of the Biogel-beads is performed by rinsing the plate three times with 200 µl 80 % ACN per well to obtain the hydrophobic environment. The sample is brought to 75 % ACN and aliquots of 200 µl per well are applied to the filter plate. Subsequent washing of the wells is performed by adding four times 200 µl 80% ACN to each well. The glycans are eluted in a new 96 deep

well plate with the addition of 200 µl dH_2O twice. Eluates of each aliquot are pooled and dried in the speedvac.

3.1.3 Glycan analysis with MALDI-TOF-MS

An Ultraflex II MALDI-TOF/TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) was used for glycan analysis. In order to be able to measure sialylated glycans the instrument was only used in linear negative mode to prevent potential loss of sialic acids as the operation of the MALDI-TOF-MS in reflectron positive mode often leads to loss of weakly bound monosaccharides, especially sialic acid.

For appropriate measurements the matrix DHB (2,5-dihydroxybenzoic acid - 20 mg per ml 50 % ACN) was used which allows AA-labelled glycan measurement in positive as well as negative ion mode. The addition of the AA-label is crucial for MALDI-MS measurements in negative ionisation mode as the fluorescent label adds an additional charge to the glycans allowing simultaneous detection of neutral as well as charged glycans [116].

3.1.4 Three dimensional Glycan separation by HPLC (High performance liquid chromatography)

For glycan analysis and subsequent separations an Ultimate 3000 HPLC system (Dionex Corporation, Sunnyvale, CA) was used. Glycans were separated and fractionated in three dimensions to achieve an optimal separation of the different glycan structures. The principal workflow of the three dimensional separation can be seen in Figure 15 completed by a final validation step of each fraction by MALDI-TOF-MS.

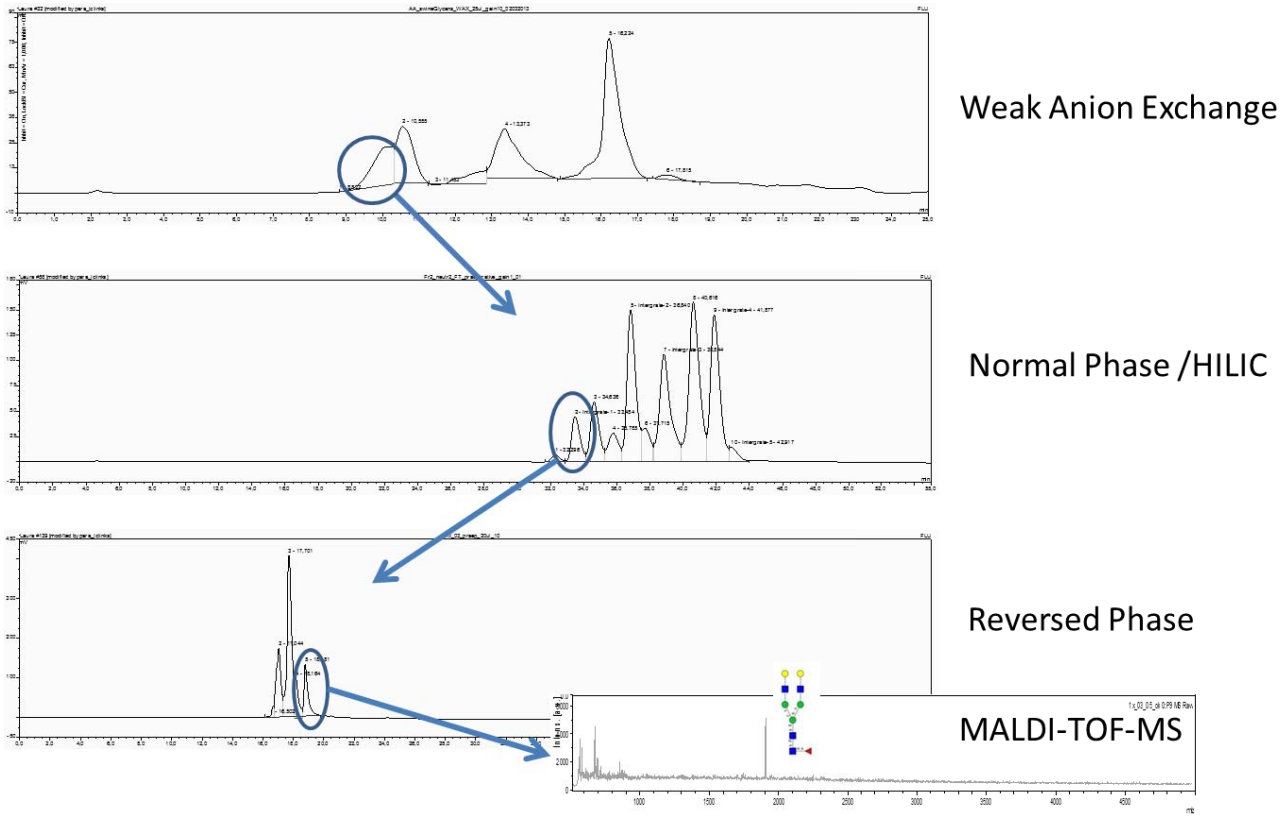


Figure 15 Three dimensional glycan separation using different types of HPLC columns for subsequent glycan array printing - workflow
AA-labelled swine glycans were chromatographically separated in three dimensions to obtain a maximum of separated glycan fractions for subsequent glycan array production. The first dimension was according to charge (weak anion exchange) followed by hydrophilic and size related separation (NP/HILIC) and a final hydrophobic related separation by (reversed phase). Each of the obtained reversed phase fractions was checked for glycan content using MALDI-TOF-MS.

3.1.4.1 Weak anion exchange chromatography (WAX)

A separation of glycans according to their charge is achieved when using an anion exchange chromatography. For the separation a DEAE TSK-Gel-5PW column, 7,9 mm x 7,5 cm (TOSOHAAS - Stuttgart, Deutschland) at a flow rate of 1 ml/min was used. Buffer A consisted of 10 % ACN and a two-step gradient with buffer B (0,3 M NH_4Acetate pH 9) ranging from 0 to 10 % B within minute 5 to 10 and from 10 to 65% B from 10 to 20 min was performed.

Peaks were collected according to their fluorescent signal at wavelengths of excitation – emission, 360 -425 nm.

After this separation a total amount of 7 fractions was obtained.

3.1.4.2 Normal Phase chromatography (NP) / Hydrophilic interaction liquid chromatography (HILIC)

In a second step of HPLC-separation a HILIC-column (Amide-80 column, 4.6 mm x 25 cm, particle size 5 µm, Tosoh Bioscience, Stuttgart, Germany) was used to separate glycans according to their hydrophility and size. Eluent A consisted of 50 mmol/L formic acid (pH 4.4) and a gradient with eluent B (eluent A/acetonitrile, 20/80) was performed from 100 to 40 % B within 40 min followed by a short washing step with 20 % B for two min and subsequent equilibration of the column. The flow was set to 0,5 ml/min and peaks were detected by coupling the column to fluorescent detection (excitation – emission, 360 -425 nm).

3.1.4.3 Reversed phase chromatography (RP)

Fractions obtained after HILIC separation were further fractionated by reversed phase, a chromatographic method used to separate glycans according to hydrophobicity. Therefor a Hypersil ODS column (2 mm x 25 cm; particle size 3 µm; Thermo Electron, Waltham, MA) with a flow rate of 0,4 ml/min was used. The used eluent A was 0,4 % ACN with 0,1 % formic acid and eluent B consisted of 95% ACN with 0,1 % formic acid added. The used gradient for appropriate separation had a 6,5 min equilibration phase at 5 % eluent B followed by a linear gradient over 25 min till a maximum of 50 % eluent B. Peaks were again fractionated according to their fluorescent signal followed by subsequent MALDI-TOF-MS analysis for precise content determination of each peak fraction.

Glycan amounts were calculated by analysing 5 pmole of a biantennary bisialylated AA labelled reference glycan structure and the peak area (mV*min) of the detected peak was used for relative quantification.

3.1.5 Construction of the Glycan array

Glycan fractions as well as synthetic glycol-conjugates were added in a 384-wells V-bottom plate (Genetix, New Milton, UK) at different concentrations. The glycan samples were dried

in the speedvac and for subsequent printing of the array each glycan fraction as well as synthetic glycol-conjugates were dissolved in 20 µl printing buffer consisting of 1:1 spotting buffer (Nexterion Spot, Schott Nexterion) and 20 % DMSO. A total number of 164 glycan samples (74 neutral -, 28 singly charged -, 14 mixture of singly and double charged -, 28 doubly charged - and 6 triply charged –N-glycans and 13 synthetic glycol-conjugates) were printed in five replicas on epoxasilane – coated glass slides (Slide E, Schott, Nexterion). The printing procedure was performed by contact printing using the Omnigrid 100 microarrayer (Genomic Solutions, Ann Arbor, MI). The application of the glycan solution was achieved with SMP3 pins with uptake channels that deposit 0.7 nl at each contact. On each epoxy-silane coated glass slide four identical arrays were printed. Dot spacing between fraction triplicates and different fractions was 290 mm (X) and 245 mm (Y), and array spacing was 6000 mm. To allow the covalent binding of the primary amines of the 2-AA labelled glycan fractions and synthetic glycol-conjugates and to prevent drying of the spots the printed slides were incubated overnight in a humidity chamber at room temperature [107].

3.1.6 Glycan array the assay

The principle and workflow of the assay can be seen in Figure 16

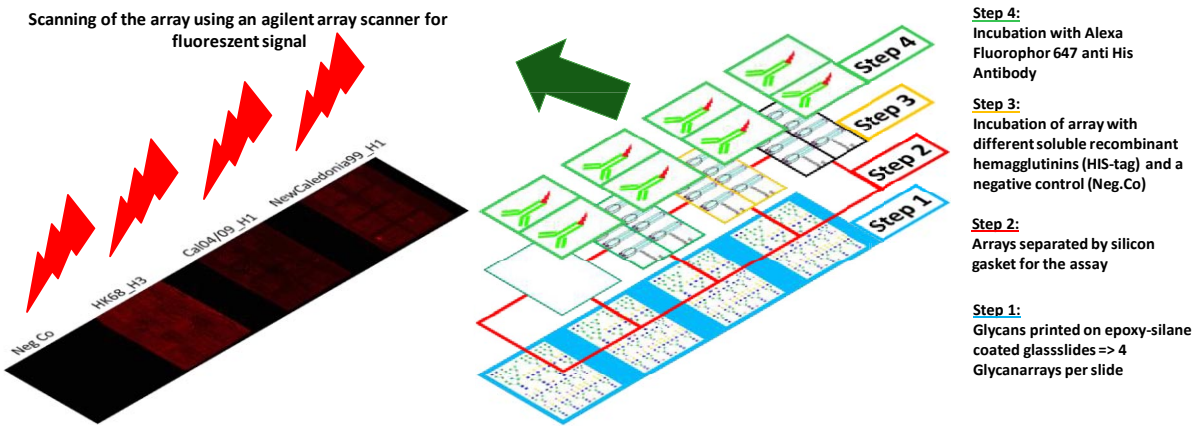


Figure 16 Glycan array - the assay workflow

The glycan array assay follows the principles of a sandwich immune sorbent assay. In the assay recombinant expressed hemagglutinins as soluble homotrimeres carrying a poly-Histidine-Tag for subsequent detection were applied onto the array to investigate hemagglutinin specific binding to different glycan fractions. Each glass slide carries four separate arrays and to allow the simultaneous performance of four different assays at a time the assays are separated by applying a silicon gasket onto the glass slide to obtain separate

incubation chambers. In a first step the arrays were blocked to prevent unspecific binding to the glass surface. This blocking step was performed at RT for 1 hour using 0,5 ml per array of a solution containing 2 % BSA in 1xPBS with 50 mM ethanolamine. After the blocking step two washing steps were performed with 1xPBS 0,05 % Tween20 and 1x PBS by excessive rinsing of the arrays. Different hemagglutinins (0,5 ml each) were applied to separate arrays at a concentration ranging from 2 to 10 µg/ml 1xPBS, with 0,01 % Tween20 and 1 % BSA added, according to their binding optimum. Incubation was again performed at RT for 1 hour followed by the same washing steps as described previously. To detect the bound hemagglutinin a fluorescently labelled Anti-His-Tag-Alexa Fluor 647 monoclonal antibody (MBL, International Corporation, Woburn, MA 01801, USA) was used at a concentration of 1 µg/ml 1xPBS, with 0,01 % Tween20 and 1 % BSA added. The incubation was performed at RT for 30 min under light protection. As a final step the array was washed as described above with an additional excessive washing step with dH_2O after the removal of the silicon gasket followed by a drying step using centrifugation. Dried and light protected arrays were scanned for specific fluorescence by the G265BA scanner (Agilent Technologies, Santa Clara, CA) using one laser at 532 nm a wavelength were the AA label does not fluoresce.

3.1.7 Data analysis - Software

Data and image analysis was performed using the GenePix Pro 6.0 software (Molecular Devices, Sunnyvale, CA). Spots were integrated and re-sized using round feature alignment with CPI threshold set to 50. Total background and negative controls subtracted median intensities of the replicates were averaged. For subsequent cluster analysis the datasets were $\log^{10}$ transformed and plotted against the sample numbers as described by Oyelaran et al 2009 [150].

Hierarchical Clustering on ferret and swine glycan arrays was performed using R (Version 3.0) [151]. The $\log^{10}$ transformed data was first mean-centred and then range-scaled between values of -1 and 1 according to van den Berg et al. 2006 [152]. For the construction of the dendrogram shown on each side of the heatmap euclidean distance and ward's linkage was used.

3.2 Biochemical methods for affinity chromatography

3.2.1 Affinity chromatography

3.2.1.1 Protein coupling to NHS activated sepharose

NHS-activated Sepharose 4 Fast Flow (GE Healthcare) was used to immobilize purified recombinant hemagglutinin to investigate glycan binding.

Binding of the recombinant proteins was performed as follows. Approximately 1 ml of the NHS-activated sepharose slurry was transferred into a 2 ml tube and allowed for sedimentation by gravity. The supernatant was discarded and the beads were washed with 10-15 times medium volume with 1 mM HCl. The supernatant was removed carefully and 1 mg of the protein dissolved in coupling buffer (0,2 M $NaHCO_3$, 0,5 M NaCl, pH 7-8) was added to the sepharose beads at a volume ratio of 2 x beads : 1 x protein. A 10 µl aliquot of the protein solution was kept for subsequent monitoring of binding efficacy to the sepharose. The slurry was incubated at RT for 4 h under gentle agitation. After the coupling was completed the supernatant was removed and again a 10 µl aliquot was taken and loaded on a SDS-PAGE (3.4.1) together with the aliquot taken before the incubation. The non-reacted groups were blocked by incubating the beads for 4 more h under gentle agitation with 0,5 M Ethanolamine in 0,5 M NaCl pH 8,3.

Before the affinity matrix is ready for use, several washing steps with two buffers differing in the pH is mandatory. First three times the medium volume of a high pH buffer (0,1 M Tris.HCl pH 8-9) were added and after gentle mixing the beads sedimented by gravity and the supernatant was removed. A second washing step with three times the volume of a low pH buffer (0,5 M NaCl pH 4-5) with gentle mixing and gravity sedimentation followed. These cycles were repeated 3 to 6 times.

3.2.1.2 Batch affinity experiment and self-packed affinity column

The previously prepared protein-coupled NHS-activated sepharose was used for affinity chromatography studies. This was either done by preparing a self-packed column with a 1 mm inner diameter tubing (Fisher Scientific UK, Loughborough, Leicestershire) to be able to use this column in further HPLC-like experiments or by loading the sepharose in batch

columns (Bio-Rad Laboratories GmbH, Munich, Germany) with 4 mm inner diameter used for manually washing, loading the AA-labelled or unlabelled glycan samples, washing away unbound glycans and eluting bound glycans. The flow was applied by gravity and fractions of flow through washing and elution were collected in eppendorf tubes (SARSTEDT, Wr. Neudorf, Austria) and further subjected to LC-ESI-MS (3.3.5) or MALDI-TOF-MS (3.1.3) to check their content.

For the self-packed column approach a separate switchboard and external pumps were used coupled to a fluorescent detector to allow chromatographic detection of eluted glycans using 1 mM HCl or 1 M NaCl. Fractions were collected and analysed by LC-ESI-MS (3.3.5) or MALDI-TOF-MS (3.1.3).

3.2.2 Amicon centrifugation tubes

To avoid the binding procedure to the NHS-activated-sepharose, amicon tubes (Amicon Ultra-0.5 Centrifugal Filter Unit with Ultracel-10 membrane, Merck Millipore, Darmstadt, Germany) were also used to investigate bound and unbound glycans to recombinantly expressed protein. Therefore the tubes were wetted with 1 x PBS before loading a previously incubated solution of protein and labelled glycans.

The mixture was applied to the amicon tube, spinned at 14,000 x g for 10 min and the flow through was collected. Three to five washing steps were performed to remove the unbound glycans and subsequent elution was performed by adding 500 µl of 1 mM HCl or 1 M NaCl. All the fractions were either checked by LC-ESI-MS (3.3.5) or MALDI-TOF-MS (3.1.3).

3.3 Combined Glycomics: biochemical methods for retention time studies

3.3.1 Glycan isolation

Glycan preparation from swine tissue

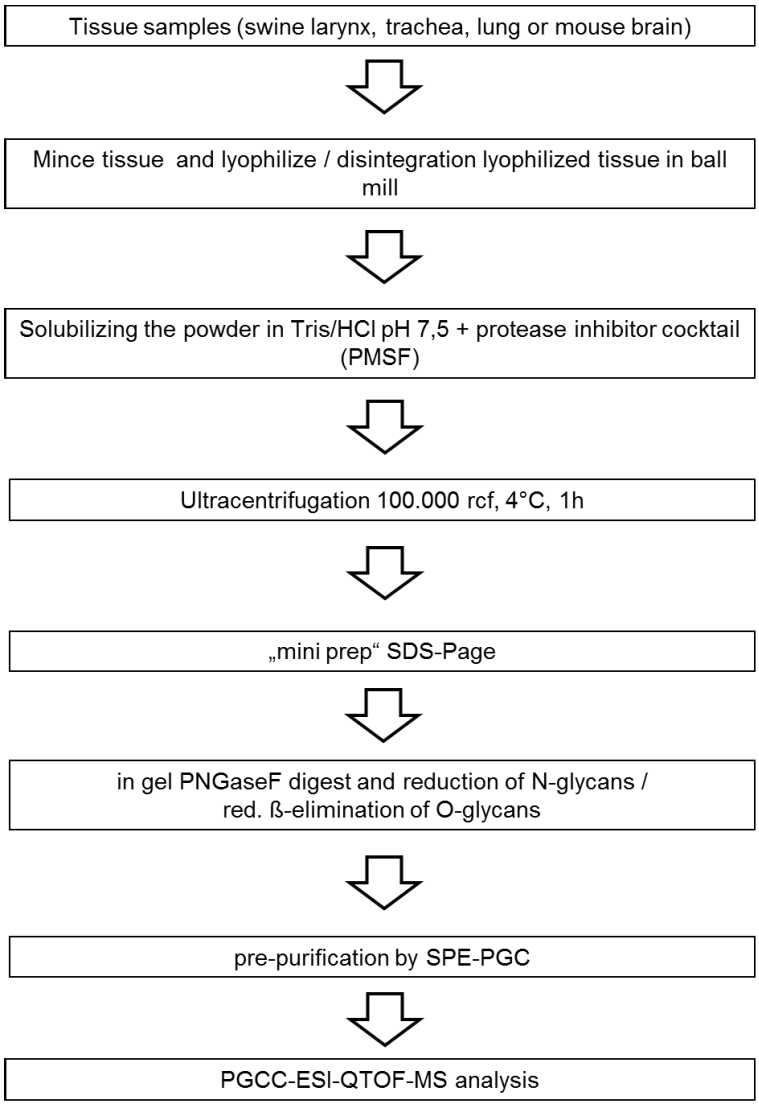


Figure 17 Glycan preparation from swine tissue - workflow

3.3.1.1 N-glycan-isolation of swine glycans/ mouse brain glycans

Tissue samples (swine: trachea, larynx and lung; mouse brain) were prepared, either by isolating the mucosal layer from the surrounding cartilage tissue followed by cutting it into small pieces (larynx and trachea) or directly by cutting the lung or brain tissue in small pieces. Each sample was transferred into a 15 ml falcon tube and washed with 1 x PBS. After removal of excess liquid a small volume of PBS was used to overlay the samples and homogenize it with an Ultra Turrax T25 homogenizer (Janke & Kunkel, IKA Laboratories, Staufen, Germany). The resulting homogenate is dried using a lyophilizer and stored at -20 °C until further usage.

3.3.1.2 Mini-prep gel

The miniprep gel is a modified SDS-PAGE (3.4.1) with a 17 % acrylamide gel instead of the commonly used 12,5 %. Furthermore the stacking gel is about 1 cm long to allow a rather simultaneous diffusion of all proteins into the denser separation gel.

Table: Mini-prep - gel formulation

	Separation gel (µl)	**Stacking gel (µl)**
Acrylamide (AA) 30 %	4 x 842	285
Bisacrylamide (Bis) 1 %	2 x 525	390
Separation gel buffer 1,5 M Tris.HCL pH 8,8	2 x 750	/
Stacking gel buffer 0,5 M Tris.HCL pH 6,8	/	750
H_2O	/	2 x 640
SDS (sodium dodecyl sulphate)	60	30
APS (Ammonium per sulfate)	36	24
TEMED (BioRAD)	3,6	2,4

Table 1 Mini-prep - gel formulation

The lyophilized and homogenized tissue samples were solubilized in SDS-PAGE sample buffer with DDT added and boiled for 10 min at 96 °C to solubilize membrane proteins. Approximately 2 mg of dried tissue sample were dissolved in 60 µl sample buffer of which ~40 µl were used per lane of a 5 lane gel. After the protein denaturation the samples were carefully spinned down at full speed (14.000 g) for 10 min and 40 µl were transferred into a new tube. To each sample 3 µl of the prestained protein ladder were added and the entire amount was loaded in one slot of the 5 slot gel. Electrophoretic separation (3.4.1) at 80 V was performed for about one hour until only minor separation of the proteins was achieved as shown in Figure 17. The purpose of this experiment is not to separate the entire membrane proteins, but to obtain a rather pure and protein dens gel piece for subsequent

glycan isolation. Electrophoresis was finished when the prestained ladder, prior added to each sample, has separated for about 1 cm.

The inner area between the highest and lowest band of the prestained ladder was excised, washed twice with dH_2O and cut into small pieces.

3.3.1.3 N-glycan isolation

The gel pieces were directly used to isolate N-glycans from the glycoproteins by a direct PNGase F digest.

Therefore the gel pieces were equilibrated twice with 500 µl PNGase F-buffer (50 mM NH_4Ac, pH 8,4) and finally covered with buffer. To each sample 1-2 µl PNGase F were added (N-Glycosidase F, Roche Diagnostics GmbH, Mannheim, Germany) and enzymatic digestion took place at 37 °C over night. The glycans are released in the liquid phase and can be further processed by reduction (3.3.2) or fluorescent labelling (3.1.2).

3.3.1.4 O-glycan isolation – reductive β-elimination

The gel pieces can also be subjected to O-glycan isolation by reductive b elimination. Therefore the gel pieces were rinsed with water and after careful removal of excess liquid the pieces were generously covered with approximately 500 µl of 1 M $NaBrH_4$ (Merck Millipore, Darmstadt, Germany) in 0,5 M NaOH (VWR Chemicals, Vienna, Austria).

Incubation was performed over night at 50 °C with a firmly and tight closed tube-lid to prevent drying out of the sample.

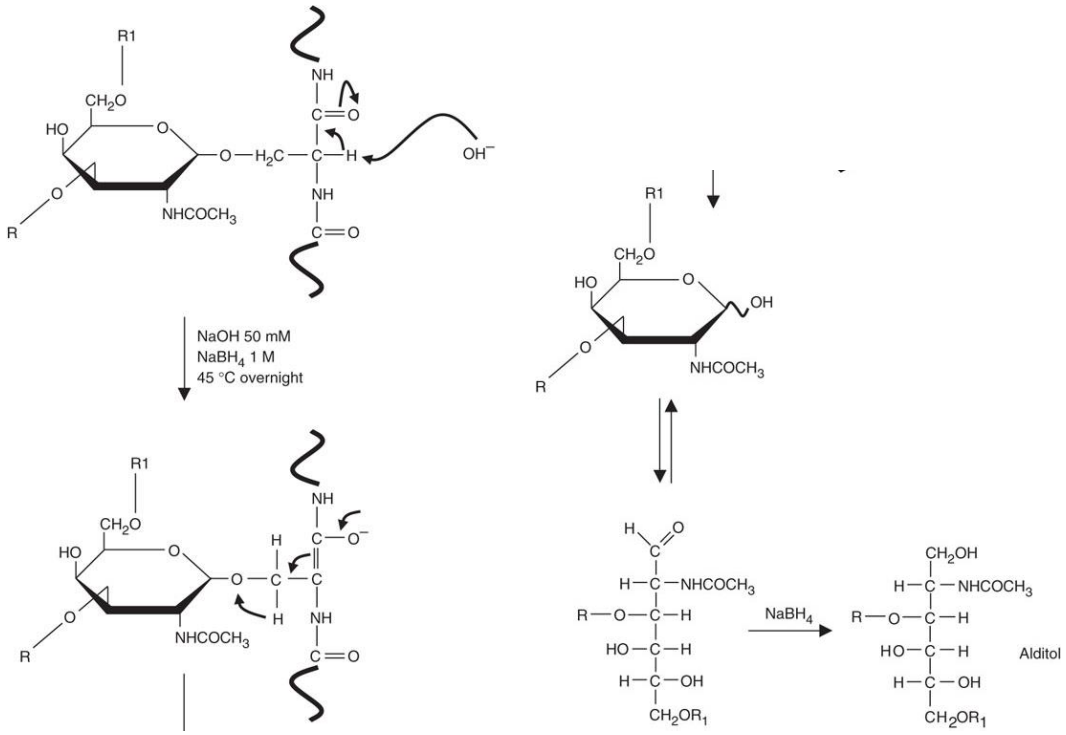


Figure 18 Reductive β-elimination of O-glycans [153]

3.3.2 Reduction

Reduction of glycans is a method that allows isomeric separation of glycans by PGC (porous graphitized carbon) chromatography resulting in a single peak per molecule in contrast to native glycans which are reviled directly after PNGase F digestion. Native glycans result in a double peak due to the reducing end-standing hydroxyl group, which constantly changes the configuration from α to β.

Reduction of the reducing glycan was achieved by a first incubation of the released glycans with 5 % formic acid (Carl Roth GmbH + Co. KG, Karlsruhe, Germany) at a pH around 3 for 30 min to convert amine groups on the reducing end, which are a by-product of the enzymatic release, back into hydroxyl groups. After the incubation, the pH was adjusted to pH 8-9 with ammonia (VWR Chemicals, Vienna, Austria) followed by the addition of 10 % $NaBrH_4$ in dH_2O to a final concentration of 1 %. The samples were incubated at RT for 4 h followed by a final SPE-PGC cartridges purification (3.4.4.1) to remove excess salt and buffers.

3.3.3 Stable isotopic labelling with hydrazine monohydrate

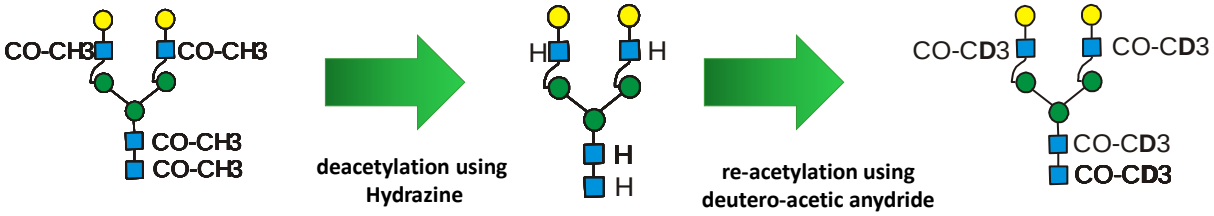


Figure 19 Stable isotopic labelling with hydrazine monohydrate and deutero-acetic anhydrid - workflow

To generate stable isotopically labelled glycan standards reduced (3.3.2) and SPE-PGC (3.4.4.1) purified glycans were solubilized in 50 µl Hydrazin monohydrate 99 % purity (Alfa Aesar, Karlsruhe, Germany) and incubated at 100 °C in a tightly closed tube for 24 – 48 h. The sample was dried in the speed vac and fractionated by HPLC (3.3.4) with a large scale PGC column to separate mono- double- triply- and four times - deacetylated glycoforms. This can be monitored by a mass shift of 42 Da per lost acetyl group. The fraction containing the doubly and triply de N-acetylated glycans was used for re N-acetylation with deuteron acetic anhydrid leading to a mass shift of + 3 Da per acetyl group. Re N-acetylation was performed by adding 125 µl 0,1 M $NaHCO_3$ to the dried fraction as well as 12.5 µl of deuterated acetic anhydride. After the incubation at 4 °C for 1 h the sample was purified by a SPE-PGC cartridge and dried again. Labelled glycans can be used as internal standards for structural investigations and were used in the retention time studies.

3.3.4 Fractionation large scale

Large scale fractionation experiments for preparing glycan standards were performed with a Hypercarb column operated with a flow of 0,6 ml/min (Thermo Hypersil-Keystone, Hypercarm, 150 x 3 mm, 5 µ particle size). The used buffers and gradients were the same as described in 3.3.5.1. Fractionation was performed using an HPLC system (Shimadzu) and 300 µl fractions were collected using the fraction collector. After each run the collected fractions were dried and checked by LC-ESI-QTOF-MS (3.3.5) for the glycan content and positive fractions containing the correct glycan moiety were used for further enzymatic processing (3.4.6) to generate the isomeric and isobaric glycan standards (4.4.1).

3.3.5 LC –ESI-QTOF-MS

The major analytical tool used in this part of the work was liquid chromatography coupled to mass spectrometric detection (LC-MS). Different instruments were available in the Lab. Most retention time related PGC-MS measurements were performed with an ESI-QTOF-MS, Electron Spray Ionization - Quadrupol - Time Of Flight - Mass Spectrometer (Ultima Global Waters Micromass) in positive ionization mode (ES+). Standard values were set as described in Table 2. Data analysis was performed with the MassLynx V4.0 software (Waters, Micromass).

Ionization mode	**ES+ (positive ion mode)**
Collision energy (collision gas – Argon)	10.0
m/z range	200 – 1900
Capillary voltage (kV)	3.20
Cone voltage (V)	80
Source temperature	100 °C
Desolvation temperature	200 °C
Cone gas flow (L/h)	75
Desolvation gas flow (L/h)	300

Table 2 MS parameters of positive ion mode QTOF

The LC system used was a capillary liquid chromatography system Ultimate 3000 (Dionex) with flow rates up to 50 µl without solvent splitting. The software to operate the LC system was Chromeleon (Dionex, Thermo Fisher).

3.3.5.1 PGCC (Porous Graphitized Carbon Chromatography)

For retention time studies the SPE-PGC purified samples were subjected directly to Hypercarb columns (0.32 or 0.18 mm _ 100 mm_3 or 5 µm particle size; Thermo Electron). Different gradients were used during this work. For separation of singly fucosylated complex

and hybrid/bisecting glycans to investigate the relative retention times a 0.32 mm x 100 mm column with 3 µm particle size and a gradient as follows was used with a flow of 4 µl / min. Total glycan pool analysis were performed with a long gradient to achieve optimal isomeric separation with a 0.32 x 100 mm column with 5 µm particle size and a flow rate of 6 µl / min. The solvents used in this method were solvent A 65 mM ammonium formeat buffer pH 3 and as solvent B 100 % ACN with 0,1% formic acid. Samples were solubilized in dH_2O and a partial loop (20 µl loop) injection was performed. The PGC columns used during this work have experienced several rounds of acidic treatments with 4 M trifluoroacetic acid (TFA) at 100 °C over night followed by 24 h flushing with solvent A as a regeneration process for the clotted columns. The regenerated columns were mounted using Viper™ fittings (Thermo Fisher Scientific, Waltham, MA USA) and endowed with grounding clamps to prevent electric charging of the column urging from the ESI source. LC gradients used can be seen in Table 3.

Time 35 min	% Solvent B	Time 95 min	% Solvent B
0	0	0	0
2	0	7	0
2.1	8	8	5.1
25	14	63	17.4
30	40	80	60
30.10	0	81	0
35	0	95	0

Table 3 PGC-LC gradients used for RT studies

3.3.5.2 Retention time assignments, using RT reference standards for PGC

For calculation of relative retention times to each reference standard chromatographic run, internal isotopically labelled standards were added for adaption of the gradient and subsequent calculation of the reference standard elution time relative to the elution of deutero acetylated A^4A^4. The used isotopically labelled internal standards were deuterated

Man8 (M+2H+ = 871.3) deutero acetylated $GnGnF^6$ (M+2H+ = 736.8), A^4A^4 (M+2H+ = 825.3) as well as $A^4F^2A^4F^2$ (M+2H+ = 971.4) and $A^4F^2A^4$ (M+2H+ = 898.4).

To calculate the relative retention times (RRT) the RT of $GnGnF^6$ and $A^4F^2A^4F^2$ was used to calculate the deviation from a reference run. All elution times were correlated to the Retention Time of A^4A^4.

Retention times of extracted selected ion monitoring (SIM) chromatograms of the given m/z value of 895,3, which comprises the doubly charged mass of Hex5HexNAc4dHex1, were used for calculation after a smoothing the chromatogram with a window size (scans) 1 and two rounds of smoothing.

rRT_{GnGnF6} ….reference Retention Time of $GnGnF^6$

$rRT_{A4F2A4F2}$ ….reference Retention Time of $A^4F^2A^4F^2$

$RT_{A4F2A4F2}$ Retention Time of $A^4F^2A^4F^2$

RT_{GnGnF6} Retention Time of $GnGnF^6$

RT_{A4A4} Retention Time of reference standard all the elution times are correlated to

$$\Delta RT_{GnGnF6} = rRT_{GnGnF6} - RT_{GnGnF6}$$

$$\Delta RT_{A4F2A4F2} = rRT_{A4F2A4F2} - RT_{A4F2A4F2}$$

$$X = (\Delta RT_{GnGnF6} - \Delta RT_{A4F2A4F2}) / (RT_{GnGnF6} - RT_{A4F2A4F2})$$

$$\text{Correlation Factor (cF)} = \Delta RT_{A4F2A4F2} - (X \times RT_{A4F2A4F2})$$

$$\text{Corrected RT (cRT}_{sample}\text{)} = (RT_{sample} \times X) + cF + RT_{sample}$$

$$\text{Relative Retention Time (RRT)} = cRT_{sample} / RT_{A4A4}$$

Table 4 Equations for Relative Retention Time (RRT) calculations

3.3.6 LC - Ion trap MS

The ion trap MS was used in positive as well as negative ion mode as well as for fragmentation analysis to investigate reduced and fluorescent labelled glycans. The standard settings of the mass spectrometer were:

negativ	Ionization mode	positive
Enhanced resolution	**Scan mode**	Enhanced resolution
7	**Dry gas flow (L/min)**	7
250	**Dry gas temperature (°C)**	160
20,00	**Nebulizer gas (p.s.i.)**	20,00
-4500	**HV capillary (V)**	-4500
140	**Capillary exit (V)**	140
-500	**HV End Plate Offset (V)**	-500
66,3	**Tap drive**	62,4
On	**ICC**	On
80.000	**Target**	100.000
200	**Max. accumulation time (ms)**	200
200 - 2200	**Scan range (m/z)**	500 – 1900

Table 5 MS parameters of positive/negative ion mode Ion trap

3.3.6.1 MS/MS fragmentation

For fragmentation in negative ion mode the additional values were:

Isolation window	3 m/z
MS/MS fragmentation amplitude	1.00 V
Smart fragmentation options	on
MS(n) averages (CID)	3 Spectra
Abs. Treshhold AutoMS(n)	500

Rel. Treshhold AutoMS(n)	1 %
Preferred charge state	double

3.3.6.2 Neg mode MS/MS diagnostic ions

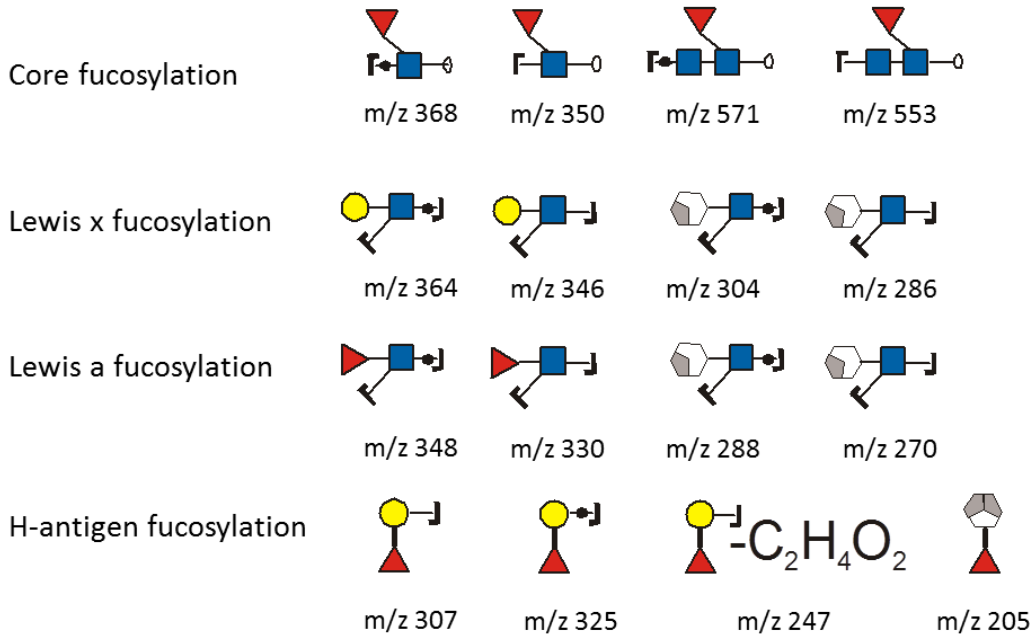


Figure 20 Diagnostic Ions for negative ion mode MS/MS fragmentation [88]

In negative ionisation mode fragmentation experiments using the ion trap result in the generation of so called diagnostic ions (Figure 20). These diagnostic ions can give additional structural information as well as confirmation of putative identified structures. For better fragmentation results solvent A was changed to 10 mM NH_4HCO_3 buffer pH 8 to avoid formeate adducts obtained with using the usual ammonium formeat buffer pH 3.

3.4 General biochemical techniques

3.4.1 Gel electrophoresis (SDS –PAGE)

For a reducing SDS-PAGE protein samples were diluted 1:1 with sample buffer and boiled at 95 °C for 10 min to denaturate the protein structure. DTT (dithiothreitol, Carl Roth GmbH + Co. KG, Karlsruhe, Germany) was added to the sample buffer to obtain reducing conditions and thereby break disulfide bonds stabilizing the tertiary structure of a protein.

After the denaturation step the samples were centrifuged at full speed (21.000 g) for 3 min at room temperature. This step is mandatory when working with whole cells or tissue preparations to pellet undissolved cell membrane- or tissue-particles and thereby avoid a smear on the gel.

20 µl of the supernatant were loaded in a single slot of the 10 slot gel. 3 µl of unheated protein ladder (Page Ruler prestained protein ladder, Fermentas SM067, Thermo scientific) were additionally loaded on the gel for latter size determination. Gels were prepared by mixing the components of the separation gel pipetting the solution between two glass slides and overlying it with 20 % EtOH while polymerization to achieve an even interface between the separation gel and the stacking gel. After 30-45 min of polymerization the EtOH was removed and the mixed components for the stacking gel were pipetted onto the separation gel followed by direct application of the slot comb. After further 30 min of polymerization the gel can be used for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE).

The gel electrophoresis was performed in a special chamber filled with running buffer (25 mM Tris, 200 mM Glycine, 3,5 mM SDS). After loading the sample and closing the lid the electrophoresis ran at a constant voltage of 200 V for approximately 50 min.

Subsequent fixing (50 % MetOH, 7 % acetic acid) and staining of the gel with Coomassie brilliant Blue solution (3,5 % perchloricacid, 0,04 % Coomassie Blue G250) for protein visualization or semi dry blotting of the separated proteins on a nitrocellulose membrane was performed.

Sample Buffer 2x

200 mg SDS (final conc. 2%)

154 mg DTT (dithiothreitol) (100 mM, Sigma D0632) freshly added

5 ml stacking gel buffer pH 6.8 (250mM)

3.6 ml glycerol 87% (30%)

Water was finally added to a total volume of 10 ml as well as some crystals bromphenol blue.

Separation gel & Stacking gel (for 2 gels)

	Separation gel (µl)	Stacking gel (µl)
Acrylamide (AA) 30 %	4 x 938	760
Bisacrylamide (Bis) 1 %	2 x 585	520
Separation gel buffer 1,5 M Tris.HCL pH 8,8	3 x 750	/
Stacking gel buffer 0,5 M Tris.HCL pH 6,8	/	1000
H_2O	2 x 640	2 x 855
SDS (sodium dodecyl sulphate)	90	40
APS (Ammonium per sulfate)	54	32
TEMED (BioRAD)	5,4	3,2

Table 6 SDS-PAGE gel formulation

Running buffer 5 x

15 g / L Tris

72 g / L Glycine

5 g / L SDS

Fixing buffer

500 ml Methanole

70 ml Acetic acid (100 M)

Filled up with dH_2O to 1000 ml

Coomassie brilliant blue

3,5 % Perchloracid

0,04 % Coomassie Brilliant Blue G-25

3.4.2 Western blott

The transfer of the separated proteins onto a nitrocellulose membrane was performed by a semi dry blotting (25 min 15 V) procedure with filters and membrane soaked with blotting buffer.

After blotting the membrane was blocked for 1 h at RT (shaking) or at 4 °C over night with blocking solution to prevent unspecific binding of antibodies. Subsequent incubation with the first antibody was carried out without any intermediate washing steps. As all proteins used in this work carried a poly-Histidine-Tag used for purification as well as immune detection. The primary antibody used for protein detection was a Monoclonal Anti-poly-Histidine antibody produced in mouse (Sigma-Aldrich) diluted 1:3000 in blocking buffer and incubated for 1 h at RT. After three TBS-T-washing steps for 5 min each, the detection antibody was applied, usually an alkaline phosphatase labelled anti mouse IgG antibody produced in rabbit (Sigma-Aldrich). Colorimetric detection of the bound antibodies was achieved by incubation with SIGMAFAST™BCIP®/NBT (Sigma-Aldrich) tablets dissolved under light protection in 10 ml dH_2O. Appearing bands were visible after a few min of incubation with the solution followed by washing with dH_20, drying and scanning of the blott.

Blotting buffer

25 mM Tris

200 mM Glycine

20 % methanol in dH_2O

TBS 10x

1 M Tris

1 M NaCl

pH 7,5

TBS-T

1 x TBS

0,05 % Tween 20

Blocking Buffer

1 x TBS-T

0,5 % BSA

3.4.3 Ni-NTA-Affinity purification (using ÄKTA purifier)

For purification of recombinantly expressed protein a 2 ml bead volume Ni-NTA Agarose (Qiagen, Venlo, The Netherlands) column was used together with the ÄKTA-purifier. Prior to the purification the column was washed with 100% solvent B (250 mM imidazole) to wash away possible impurities bound to the column. After equilibrating the column with 4 % solvent B (10 mM Imidazole) the cell culture supernatant containing the recombinant protein was directly applied onto the column via the sample pump. Simultaneously the UV adsorption at 210 nm and the fraction collector was started. After the sample was loaded onto the column the system was flushed with loading pump-buffer to make sure that the entire supernatant is applied. As soon as the supernatant passes the detector the signal rises due to the proteins in the cell culture supernatant, therefore careful washing of the column prior elution is mandatory. When the UV chromatogram has reached the baseline again, a linear gradient ranging from 4 % up to 100 % solvent B within 30 min was applied to elute the bound protein.

Typically the bound proteins elute between 40 and 120 mM Imidazole concentration which was visualized in the UV spectrum. Corresponding fractions were subsequently checked by western blott (3.4.2) and purity was checked by SDS-PAGE and Coomassie staining (3.4.1).

Solvent A	25 mM Tris.HCL, 300 mM NaCl, pH 7,2
Solvent B	25 mM Tris.HCL, 300 mM NaCl, pH 7,2, 250 mM Imidazole
Loadingpump Buffer	25 mM Tris.HCL, 300 mM NaCl, pH 7,2, 10 mM Imidazole

3.4.4 Solid Phase Extraction (SPE)

Solid phase extraction was performed to purify and desalt glycans- as well as glycopeptide – samples. As the samples were further analysed by liquid chromatography coupled to a ion trap or quadrupol time of flight mass spectrometer (LC-IT / QTOF-MS) or by a matrix assisted

laser desorption ionization time of flight mass spectrometer (MALDI TOF MS) the desalting is crucial as a high salt concentration would interfere with the measurements and would hinder appropriate ionization, matrix crystallization and subsequent detection.

3.4.4.1 PGC

HyperSep™ Hypercarb™ SPE Cartridges (Thermo Scientific) were used to purify and desalt native and reduced glycans. According to the glycan amount cartridges could be chosen between 10, 25 and 250 mg graphite material. For about 3 nmole the 25 mg cartridges were used as follows.

To wash away impurities remaining on the column from the manufacturing process the cartridges were washed with 500 µl of 80 % ACN followed by 500 µl 50 % ACN. Liquid was spinned through the cartridge at 1400 rpm for 30 sec. The flow through was discarded and equilibration of the cartridge was performed with three times 500 µl dH_2O. This equilibration allows the loaded glycan samples to bind to the graphite material. Each washing, equilibration, loading and elution step was performed by gentle centrifugation at 1400 rpm for 30 sec to 1 min. After the sample was loaded again three washing steps with 500 µl H_2O each were performed to desalt the sample. The flow through of each step was discarded. Finally the glycans were eluted with 50 % ACN in 65 mM formic acid buffered to pH 3 with ammonia [35]. Eluted samples were dried in the speedvac and ready for subsequent analysis.

3.4.4.2 C18

HyperSep™ C18 SPE Cartridges (Thermo Scientific) were used to purify and desalt peptides and glycopeptides for subsequent analysis by mass spectrometry. Different columns sizes were available and volume amounts of buffers are adjusted accordingly. The protocol is performed as follows with 25 mg columns. To remove impurities from the manufacturing process columns were first washed with 500 µl 100 % methanol followed by a second washing step with 500 µl 80 % ACN and 0,1 % FA (formic acid). Equilibration of the column was performed with three times 500 µl 5 % ACN and 0,1 % FA and the flow through was discarded. The sample is loaded onto the column followed by three more washing steps with 500 µl 5 % ACN and 0,1 % FA each, to achieve appropriate desalting. Elution of the bound

peptides and glycopeptides was performed in a new tube with 500 µl 80 % ACN and 0,1 % FA. Samples were dried in the speedvac for subsequent MS analysis.

3.4.4.3 Zip-C18 Tip

ZipTip® Pipette Tips with C18 resin (MERCK, MILLIPORE) were used to purify fluorescently labelled glycans and simultaneously spot them on a MADLI plate. The pipette tips were washed twice with 10 µl 50 % ACN with 0,1 % TFA (Trifluoracetic acid) added followed by equilibration with 0,1 % TFA. After that the sample, dissolved in dH_2O, was loaded onto the C18 material by repeatedly pipetting the solution up and down. The glycans are bound to the C18 material via the fluorescent label and therefore the tip was washed by flushing it with 40 µl of 0,1 % TFA buffer. Elution and spotting of the bound purified glycans was performed by repeatedly pipetting 2 µl of DHB matrix (10 mg / ml DHB in 50 % ACN + 0,1 % TFA) up and down directly on the MALDI plate. After the sample has dried at RT it was directly analysed by MALDI-MS.

3.4.4.4 HILIC TIP

This is a method to purify and simultaneously spot glycan samples on a MADLI plate. Therefore 10 µl pipette tips were prepared by stuffing small amounts of cotton wool into the very rear part of the tip. The cotton wool serves equilibrated with organic solvent as a HILIC phase and can bind glycans. To wash away impurities in the cotton the tips were washed three times with 20 µl dH_2O. Equilibration was achieved by flushing the tips three times with 10 µl 83 % ACN with 0,1 % TFA. The sample was brought to 83 % ACN and loaded onto the stationary phase by pipetting 20 to 50 times up and down. After a washing procedure with three times 20 µl 83 % ACN with 0,1 % TFA the sample was eluted with 3 µl water directly on the MALDI plate and after the spot has dried at RT 1 µl DHB matrix (3.1.3) was added and dried again at RT.

3.4.5 Glycopeptide and proteomic analysis

For glycopeptide as well as proteomic analysis, proteins needed to be enzymatically digested for subsequent investigation using RP-LC-ESI-MS. Therefore at least 5 µg of protein samples were loaded on a SDS-PAGE and stained with Coomassie brilliant blue (3.4.1). The stained

bands corresponding to the glycoprotein of interest were excised from the gel, cut into small pieces transferred to a tube and rinsed with water for subsequent destaining procedure.

In a first step the water was discarded and 50 µl of 50 % ACN were added for 15 min followed by a second incubation with 50 µl of 50 % ACN for 10 min. After the liquid was discarded again a third incubation with 100 % ACN was performed till the pieces turn white. Removal of the liquid was followed by adding 30 µl of freshly prepared 0,1 M NH_4HCO_3 for 5 min and a final destaining step again with 30 µl of 100 % ACN for 10 min. Finally the destained gel pieces were dried in the speedvac after discarding the liquid.

For the tryptic digest disulfide bonds in the protein were broken by an incubation of the destained gel pieces with 50 µl 0,1 M NH_4HCO_3 with 10 mM DTT (Carl Roth GmbH + Co. KG, Karlsruhe, Germany) added for 30 min at 56 °C in the heating block, followed by a second incubation with 50 µl of 55 mM Iodacetamide (FLUKA, Sigma-Aldrich) in the same buffer. The incubation was at RT in the dark. After discarding the liquid washing steps were performed exactly according to the destaining protocol.

When the gel pieces were dried again the tryptic digest was started by adding one part trypsin (Promega) stock (50 ng / µl sequence grading trypsin in 1 mM HCl) and 2 parts 25 mM NH_4HCO_3 – buffer to allow the dried pieces to swell for 10 minutes. The gel pieces were covered by 25 mM NH_4HCO_3 –buffer and incubated at 37 °C over night.

In a final step the tryptic peptides/glycopeptides were extracted from the gel by adding 30 µl of 25 mM NH_4HCO_3 –Buffer and incubating the samples for 10 min on a shaker followed by adding 30 µl of 100 % ACN and further shaking. The supernatant was collected in a fresh tube and 30 µl of 5 % formic acid (Fisher scientific) were added to the gel pieces. After another 10 min shaking the supernatant was combined with the other one and the last step was repeated again.

Finally the combined supernatants were dried in the speedvac and were ready for subsequent RP-ESI-QTOF-MS or RP-ESI-IT-MS/MS experiments.

3.4.5.1 RP-LC-QTOF-MS for glycopeptides

For the investigation of glycopeptides a BioBasic C18 analytical column (150 mm60.18 mm, 5 mm, Thermo Scientific) was used for the chromatographic separation . The used method was described by [154]. In brief the solvents used in this method were the same as described in

3.3.5. Solvent A consisted of 65 mM ammonium formeat buffer pH 3 and 100 % ACN with 0,1% formic acid was used as solvent B. Samples were solubilised in dH_2O and a partial loop (20 µl loop) injection was performed. The used gradient for appropriate separation depended on the amount of glycosylation sites and differed in the length. In most cases an equilibration of the column was performed for 5 min at 1 % B followed by a first raise of solvent B within 15 min to 15 % and a second steeper raise within 16 min to 55 % B. A final cleaning step up to 80 % within three min followed and the column equilibration was performed for 11 min to 1 % B. The mass spectrometer used was a QTOF-MS (Ultima Global Waters Micromass) with a standard ESI source in positive ionization mode (ES+). Standard values were set as described in Table 2. Data analysis was performed with the MassLynx V4.0 software (Waters, Micromass) and herein, notably, the MaxEnt3 deconvolution/deisotoping feature [155].

3.4.5.2 RP-LC-QTOF/IT-MS for proteomics

After in-gel tryptic digest of the separated proteins on SDS-PAGE gels, generated peptides were extracted as described in 3.4.5. Subsequently peptides were subjected to reverse phase ESI-MS/MS peptide mapping using a Bruker Ion trap AmaZon speed ETD and a Bruker Maxis4G. For protein identifications protein database search of tandem MS data was done using the ProteinScape 3 software tool (Bruker-Daltonik, Germany) with Mascot 2.3.02 algorithm for peptide identification or the free available software GPM (The Global Proteome Machine Organization) using the X!tandem algorithm.

3.4.6 Enzymatic assays

For the RRT experiments the standard tool used were enzymatic assays with recombinantly expressed enzymes as well as purchased ones. The used enzymes are listed in 3.6. Glycosyltransferases as well as glycosidases were used to generate standards and elucidate the glycan structure of isomers in biological samples. The for the glycosyltransferase assay a standard protocol described by [35] was used. Briefly, the assay consisted of 10 µl enzyme solution, 5 µl of 10 µM nucleotide sugar (GPD-Fucose, UDP-N-acetyl Galactosamine, UDP-Galactose, KYOWA Hakko Kogyo CO), 2,5 µl $MnCl_2$ (0,4 M), 10 µl glycan substrate (~ 300 pmole) and to a final volume of 50 µl 50 mM MES-buffer, pH 7,0 was added. Incubation over

night at 37 °C was performed. The glycosidase assays were also performed at 37 °C between 3 h and overnight incubations. The used buffer was 100 mM sodium citrate/phosphate buffer at pH 5.0 with the addition of 5 µl enzyme and 10 µ glycan substrate with an approximate amount of 300 pmole to a total volume of 50 µl [156]. Prior to analysis by LC-MS, to monitor substrate conversion, a SPE-PGC cartridge purification was performed as described in 3.4.4.1.

3.4.7 Hemagglutination assay

The Hemagglutination assay is a widely used tool to investigate erythrocyte agglutinating activity of proteins. Therefore serial dilutions of recombinant expressed hemagglutinins and as a positive control the lectin ConA (Concanavalin A, *Canavalia ensiformis*) were prepared in V-Bottom 96-well plates.

The serial dilutions were prepared with either 0,9 % NaCl solution or 1 x PBS pH 7,4 (phosphated buffered saline) with or without the addition of divalent cations (1 mM $CaCl_2$ and 1 mM $MgCl_2$). Positive control protein concentration ranged from 50 µg to 0,4 µg per well. Besides that, a 1 % erythrocyte solution was prepared by diluting pelleted erythrocytes, which can be considered as 75 % concentrated solution after soft centrifugation at 500 g for 10 min at 4 °C. Remaining pelleted erythrocytes were washed twice with 1 x PBS by gentle centrifugation as described above. Erythrocytes can be stored at 4 °C for up to 2 weeks when overlaid with 1x PBS. Different hemagglutinins as well as different erythrocytes were tested. Beside human blood group 0, A and B also chicken erythrocytes were tested.

In a first step protein solutions, positive control and negative control were applied to each v-bottom well at a volume of 50 µl per well in duplicates. Subsequently 50 µl of the 1 % erythrocyte solution were applied to each well. After gentle shaking of the plate to achieve a proper dilution, incubation at RT for 30 min was performed. The result could be read out directly as agglutinated erythrocytes appear as a turbid solution whereas not agglutinated cells sediment to the bottom of the plate and due to the V-shape a clear red spot in the centre of each well can be seen.

3.5 Molecular biological methods

3.5.1 Recombinant expression of Hemagglutinins in *Sf9* cells

HA hemagglutinin from influenza A virus A/California/07/2009(H1N1)

ATGAAGGCAATACTAGTAGTTCTGCTATATACATTTGCAACCGCAAATGCAGACACATTATGTATAGGTTATCAT
GCGAACAATTCAACAGACACTGTAGACACAGTACTAGAAAAGAATGTAACAGTAACACACTCTGTTAACCTTCTA
GAAGACAAGCATAACGGGAAACTATGCAAACTAAGAGGGGTAGCCCCATTGCATTTGGGTAAATGTAACATTGCT
GGCTGGATCCTGGGAAATCCAGAGTGTGAATCACTCTCCACAGCAAGCTCATGGTCCTACATTGTGGAAACACCT
AGTTCAGACAATGGAACGTGTTACCCAGGAGATTTCATCGATTATGAGGAGCTAAGAGAGCAATTGAGCTCAGTG
TCATCATTTGAAAGGTTTGAGATATTCCCCAAGACAAGTTCATGGCCCAATCATGACTCGAACAAAGGTGTAACG
GCAGCATGTCCTCATGCTGGAGCAAAAAGCTTCTACAAAAATTTAATATGGCTAGTTAAAAAAGGAAATTCATAC
CCAAAGCTCAGCAAATCCTACATTAATGATAAAGGGAAAGAAGTCCTCGTGCTATGGGGCATTCACCATCCATCT
ACTAGTGCTGACCAACAAAGTCTCTATCAGAATGCAGATGCATATGTTTTTGTGGGGTCATCAAGATACAGCAAG
AAGTTCAAGCCGGAAATAGCAATAAGACCCAAAGTGAGGGrTCrAGAAGGGAGAATGAACTATTACTGGACACTA
GTAGAGCCGGGAGACAAAATAACATTCGAAGCAACTGGAAATCTAGTGGTACCGAGATATGCATTCGCAATGGAA
AGAAATGCTGGATCTGGTATTATCATTTCAGATACACCAGTCCACGATTGCAATACAACTTGTCAAACACCCAAG
GGTGCTATAAACACCAGCCTCCCATTTCAGAATATACATCCGATCACAATTGGAAAATGTCCAAAATATGTAAAA
AGCACAAAATTGAGACTGGCCACAGGATTGAGGAATATCCCGTCTATTCAATCTAGAGGCCTATTTGGGGCCATT
GCCGGTTTCATTGAAGGGGGGTGGACAGGGATGGTAGATGGATGGTACGGTTATCACCATCAAAATGAGCAGGGG
TCAGGATATGCAGCCGACCTGAAGAGCACACAGAATGCCATTGACGAGATTACTAACAAAGTAAATTCTGTTATT
GAAAAGATGAATACACAGTTCACAGCAGTAGGTAAAGAGTTCAACCACCTGGAAAAAAGAATAGAGAATTTAAAT
AAAAAAGTTGATGATGGTTTCCTGGACATTTGGACTTACAATGCCGAACTGTTGGTTCTATTGGAAAATGAAAGA
ACTTTGGACTACCACGATTCAAATGTGAAGAACTTATATGAAAAGGTAAGAAGCCAGCTAAAAAACAATGCCAAG
GAAATTGGAAACGGCTGCTTTGAATTTTACCACAAATGCGATAACACGTGCATGGAAAGTGTCAAAAATGGGACT
TATGACTACCCAAAATACTCAGAGGAAGCAAAATTAAACAGAGAAGAAATAGATGGGGTAAAGCTGGAATCAACA
AGGATTTACCAGATTTTGGCGATCTATTCAACTGTCGCCAGTTCATTGGTACTGGTAGTCTCCCTGGGGGCAATC
AGTTTCTGGATGTGCTCTAATGGGTCTCTACAGTGTAGAATATGTATTTAA

HA hemagglutinin from influenza A virus A/Hiroshima/52/2005(H3N2)

ATGAAGACTATCATTGCTTTGAGCTACATTCTATGTCTGGCTTTCGCTCAAAAACTTCCCGGAAATGACAACAGC
ACGGCAACGCTGTGCCTTGGGCACCATGCAGTACCAAACGGAACGATAGTGAAAACAATCACGAATGACCAAATT
GAAGTTACTAATGCTACTGAGCTGGTTCAGAGTTCCTCAACAGGTGGAATATGCGACAGTCCTCATCAGATCCTT
GATGGAGAAAACTGCACACTAATAGATGCTCTATTGGGAGACCCTCAGTGTGATGGCTTCCAAAATAAGAAATGG
GACCTTTTTGTTGAACGCAGCAAAGCCTACAGCAACTGTTACCCTTATGATGTGCCGGATTATGCCTCCCTTAGG
TCACTAGTTGCCTCATCCGGCACACTGGAGTTTAACAATGAAAGCTTCAATTGGACTGGAGTCACTCAAAATGGA
ACAAGCTCTGCTTGCAAAAGGAGATCTAATAACAGTTTCTTTAGTAGATTGAATTGGTTGACCCAATTAAAATTC
AAATACCCAGCATTGAAAGTGACTATGCCAAACAATGAAAAATTTGACAAATTGTACATTTGGGGGGTTCACCAC
CCGGTTACGGACAATGACCAAATCTTCCTGTATGCTCAAGCATCAGGAAGAATCACAGTCTCTACCAAAAGAAGC
CAACAAACTGTAATCCCGAATATCGGATCTAGACCCAGAGTAAGGAATATCCCCAGCAGAATAAGCATCTATTGG
ACAATAGTAAAACCGGGAGACATACTTTTGATTAACAGCACAGGGAATCTAATTGCTCCTAGGGGTTACTTCAAA
ATACGAAGTGGGAAAAGCTCAATAATGAGATCAGATGCACCCATTGGCAAATGCAATTCTGAATGCATCACTCCA
AATGGAAGCATCCCCAATGACAAACCATTTCAAAATGTAAACAGGATCACATATGGGGCCTGTCCCAGATATGTT
AAGCAAAACACTCTGAAATTGGCAACAGGGATGCGAAATGTACCAGAGAAACAAACTAGAGGCATATTTGGCGCA
ATCGCGGGTTTCATAGAAAATGGTTGGGAGGGAATGGTGGATGGTTGGTACGGTTTCAGGCATCAAAATTCTGAG
GGAATAGGACAAGCAGCAGATCTCAAAAGCACTCAAGCAGCAATCAATCAAATCAATGGGAAGCTGAATAGGTTG
ATCGGGAAAACCAACGAGAAATTCCATCAGATTGAAAAAGAATTCTCAGAAGTAGAAGGGAGAATTCAGGACCTC
GAGAAATATGTTGAGGACATTAAAATAGATCTCTGGTCATACAACGCGGAGCTTCTTGTTGCCCTGGAGAACCAA
CATACAATTGATCTAACTGACTCAGAAATGAACAAACTGTTTGAAAGAACAAAGAAGCAACTGAGGGAAAATGCT
GAGGATATGGGCAATGGTTGTTTCAAAATATACCACAAATGTGACAATGCCTGCATAGGATCAATCAGAAATGGA
ACTTATGACCATGATGTATACAGAGATGAAGCATTAAACAACCGGTTCCAGATCAAAGGCGTTGAGCTGAAGTCA
GGATACAAAGATTGGATCCTATGGATTTCCTTTGCCATATCATGTTTTTTGCTTTGTGTTGCTTTGTTGGGGTTC
ATCATGTGGGCCTGCCAAAAAGGCAACATTAGGTGCAACATTTGCATTTGA

Figure 21 Nucleotide sequence of H1 (A/California/07/2009(H1N1) and H3 (A/Hiroshima/52/2005(H3N2) hemagglutinin

Different recombinantly expressed hemagglutinins provided by Dr. Florian Krammer (Ass. Prof. Dr, Icahn School of Medicine at Mount Sinai, New York, 10029, USA) were used in this study in order to investigate their binding specificity on glycan arrays.

Different hemagglutinins are depicted in Table 7.

H1 Hemagglutinin (H1N1)	
A/PR/8/34 (Uniprot: **P03452** (HEMA_I34A1)	**900 µg**
A/California/04/09 (Uniprot: **C3W5X2** (C3W5X2_9INFA)	**600 µg**
A/New Caledonia/20/99 (Uniprot: **Q6WG00** (Q6WG00_9INFA)	**225 µg**
H3 Hemagglutinin (H3N2)	
A/Hong Kong/1/68 (Uniprot: **Q91MA7** (HEMA_I68A4)	**210 µg**
A/Panama/2007/99 (Uniprot: **Q1K9M3** (Q1K9M3_9INFA)	**220 µg**
H2 Hemagglutinin (H2N2)	
A/Japan/305/57 (Uniprot: **P03451** (HEMA_I57A0)	**400 µg**
H5 Hemagglutinin (H5N1)	
A/Vietnam/1203/04 (Uniprot: **D1LP57** (D1LP57_9INFA)	**1260 µg**

Table 7 Recombinant hemagglutinins provided by Florian Krammer

Sf9 cells were seeded in 6 cm diameter cell culture dishes at a density on 1 x 10^6 cells per dish followed by incubation for 30 min at 27 °C, which allowed the cell adhesion to the dish surface. The medium was removed and cells were washed three times with IPL-41 to remove fetal calf serum for the transfection process.

While the cells were allowed to adhere to the surface the two co-transfection solution A and B were prepared. Solution A was a 2 % lipofectin solution (Invitrogen™ Life technologies, Thermo Fisher Scientific) in 1,5 ml IPL-41 medium without FCS. Solution B comprised 50 ng BaculoGold™ DNA (BD Bioscience, San Diego, CA, USA), 1 µg transfer vector (either pVT-Bac-

His- or pBacPAK8-constructs see Figure 23 and Figure 24) and 1,5 ml IPL-41 medium without FCS. Both solutions were mixed and incubated for 30 min at RT prior to the addition of the solution onto the cells. Co-transfection took place at 27 °C for 5 h followed by incubation of the cells for 5 days with medium containing 50 µg/ml Penicillium/Streptomycin and 5 % FCS under standard conditions. On a daily base cells were checked by fluorescent-microscopy as the pVT-BAC_His transfected cells co-expressed the green fluorescent protein (GFP) encoded on the vector backbone after successful co-transfection.

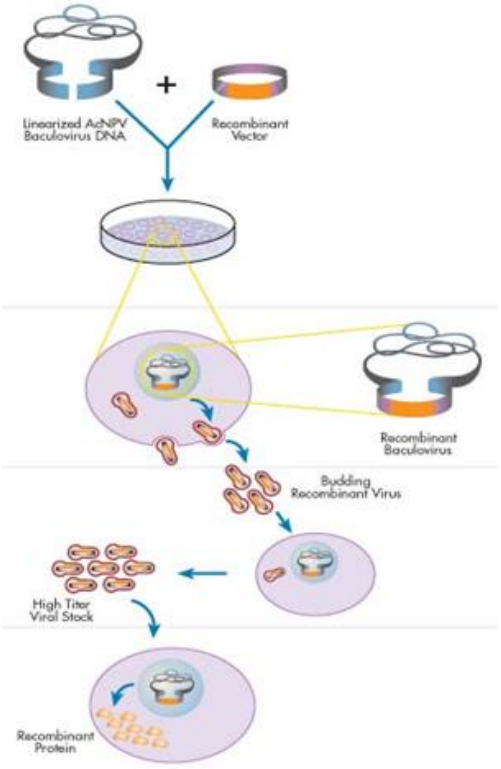


Figure 22 Baculovirus co-transfection of *SF9* insect cells [157]

A schematic scheme of the co-transfection protocol of *Sf9* insect cells with linearized BaculoGold™ DNA and the pVT-Bac_His vector (Figure 23) or the pBacPAK8 vector (Figure 24) containing the coding sequence for the desired hemagglutinin can be seen in Figure 22.

3.5.1.1 Baculovirus

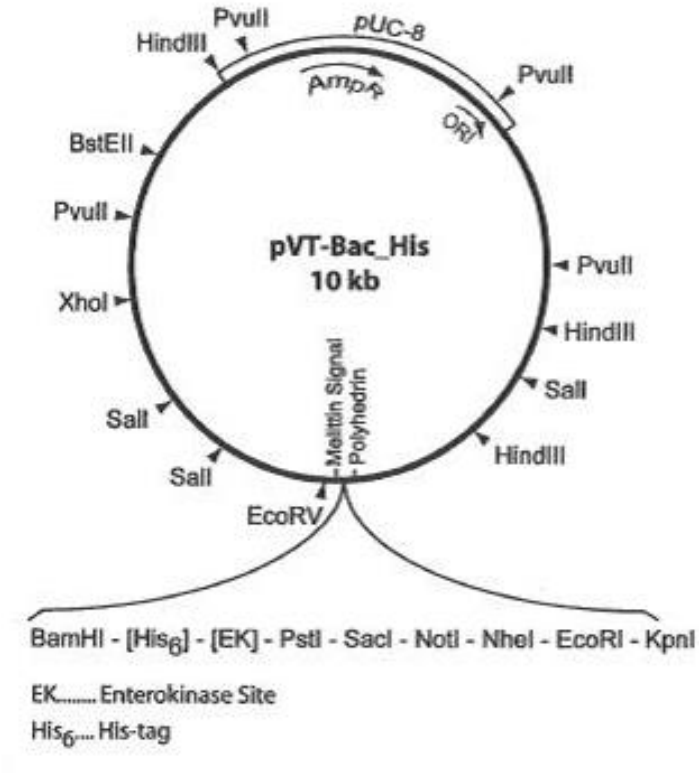


Figure 23 Vectormap of pVT-Bac_His

H1 and H3 hemagglutinins were cloned into the pVT-Bac_His vector and used for co-transfection of *Sf9* insect cells. Co-transfection and protein expression was performed as described in 3.5.1.

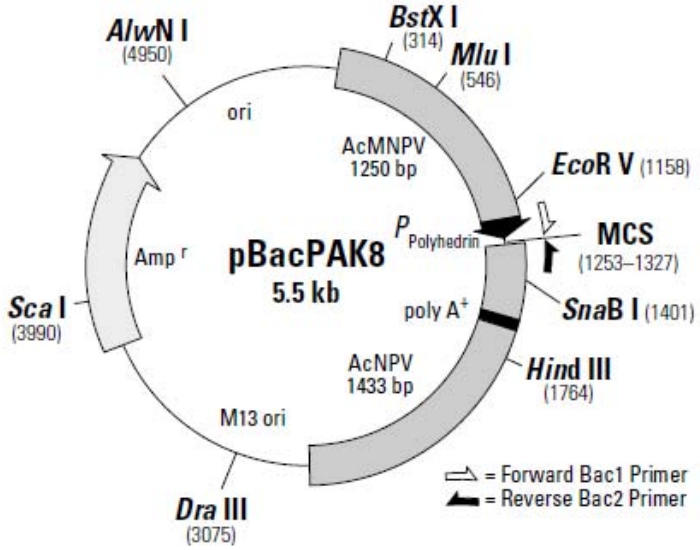


Figure 24 Vectormap of pBacPAK8

H3 containing the C-terminal trimerization sequence was cloned into the pBacPAK8 vector by Miriam Klausberger (Dipl.Ing, Boku, Department of Biotechnology). Subsequent co-transfection of *Sf9* insect cells and protein expression was performed as described in 3.5.1

3.5.1.2 Insect cells

Sporodopta frugiperda Sf9 cells [158] were cultivated in IPL-41 medium (PAN Biotech, Aidenbach, Germany) complemented with 5 % fetal calf serum (FCS) and 50 µg/ml Penicillium/Streptomycin. Cells were splitted (diluted 1:3 or 1:5 with new cell medium) two to three times a week to retain appropriate growth.

3.6 Glycosyltransferases for enzymatic assays

- core α1,6-Fuc-T, a crude microsome preparation from chicken heart was used prepared by Katharina Wozny (Msc, former BOKU, Department of Chemistry)

Cloned and expressed in *Sf9* insect cells (**Error! Reference source not found.**) by Andreas Thader (Mag., Boku, Department of Chemistry)

- Murine α1,2-Fuc-T Enzyme name: Fucosyltransferase 2; Gene name: FUT2
- Human α1,3/4-Fuc-T (which turned out to be only an a1,4-Fuc-T) Enzyme name: FucT-III; Gene name: FUT4)
- Human α1,3-Fuc-T IV (for Lewis X) Enzyme name: FucT-IV; Gene name: FUT3
- Human β1,3-Gal-T Enzyme name: β3GalT5; Gene name: B3GALT5
- Human β1,4-GlcNAc-T III (for bisected structures) Enzyme name: GnT-III; Gene name: MGAT3
- Rabbit β1,2-GlcNAc-T I (for hybrid type structures) Enzyme name: GnT-I; Gene name: MGAT1

Commercially available transferases were β1,4-Gal-T and α1,3-Gal-T (Sigma-Aldrich)

4 Results

4.1 Investigation of glycans isolated from swine tissues using different analytical methods

In this part of the study two different analytical tools were chosen to investigate isomeric glycans isolated from swine tissues. One combined a PGC-LC separation of reduced glycans, coupled to ESI-QTOF-MS for detection. In the other analytical approach a multidimensional HPLC separation and fractionation of fluorescently AA-labelled glycans with subsequent MALDI-TOF-MS detection was performed. The fractionated AA-labelled glycans were further used for the preparation of a glycan array.

4.1.1 Strategy for the Investigation of isomeric N-glycans using LC-ESI-QTOF-MS

Isomer specific analysis of glycans can be a crucial necessity as different isomers can provoke different immunogenic functions [48]. The isolated N-linked glycans from the different swine tissues were investigated by LC-ESI-QTOF-MS with a PGC column to obtain isomeric separation.

Figure 25 shows the base peak chromatogram (PBC) of a PGC chromatographic run. From this overview chromatogram it is clearly visible how the different glycan species interfere with the PGC material. Oligomannosidic glycans elute from the column already with a rather low percentage of organic solvent, whereas the more complex the structures get the later the elution. The highest complexity of multiple antennas in combination with sialic acid decoration and an α1,6 linked core fucose leads to the highest interference with the stationary phase. The RT influences of different glycan species are shown later in more detail.

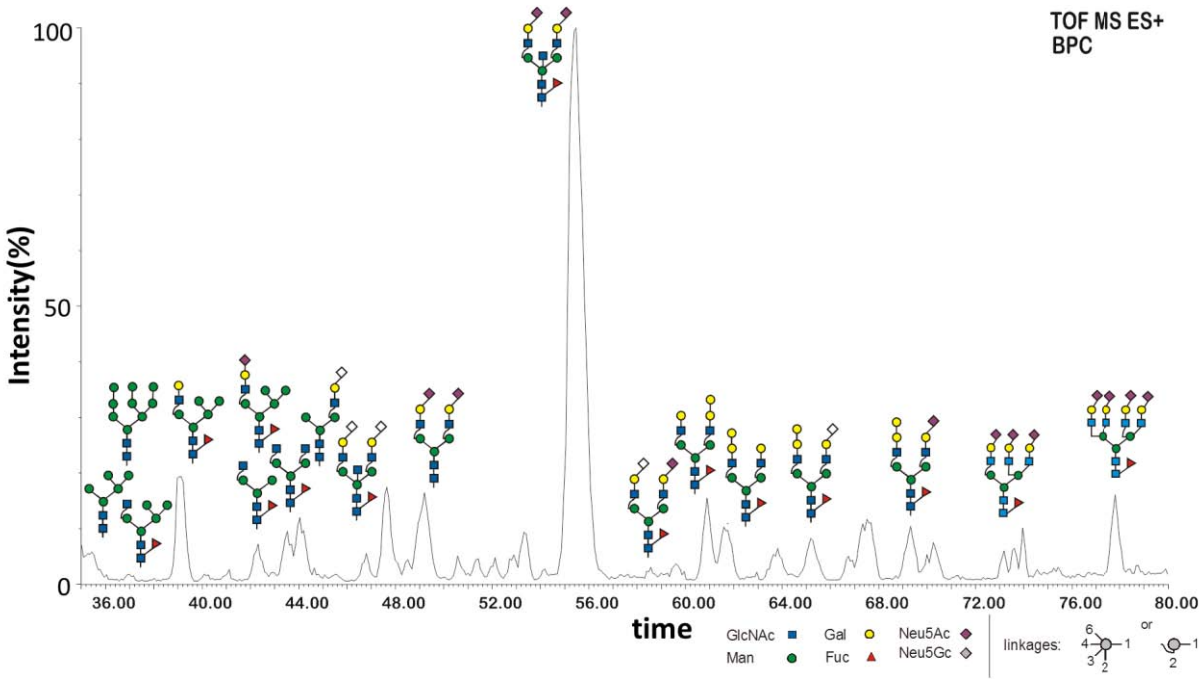


Figure 25PGCC-Chromatogram of swine N-glycans on the example of larynx using LC-ESI-MS analysis
Interaction of complex glycan mixtures isolated from biological samples with a PGC-column. Oligomannosidic and hybrid structures elute rather early whereas the more complex the structures get, the later is the elution.

When investigating the elution profiles of different glycan structures in more detail the efficacy of the PGC material to separate isomeric and isobaric structures gets clearly visible.

In Figure 26 panel A,B,D and E show different oligomannosidic glycan structures eluting as single peaks whereas panel C shows two separate peaks indicating the occurrence of two stereoisomers of Man7. The elution profiles show a rather narrow elution window of Man9-Man6 whereas the Man5 isomer interferes with the PGC material remarkably more. The assigned glycan structures are putative but in accordance with the calculated RRT of oligomannosidic structures previously published by Pabst et al. 2012 [140].

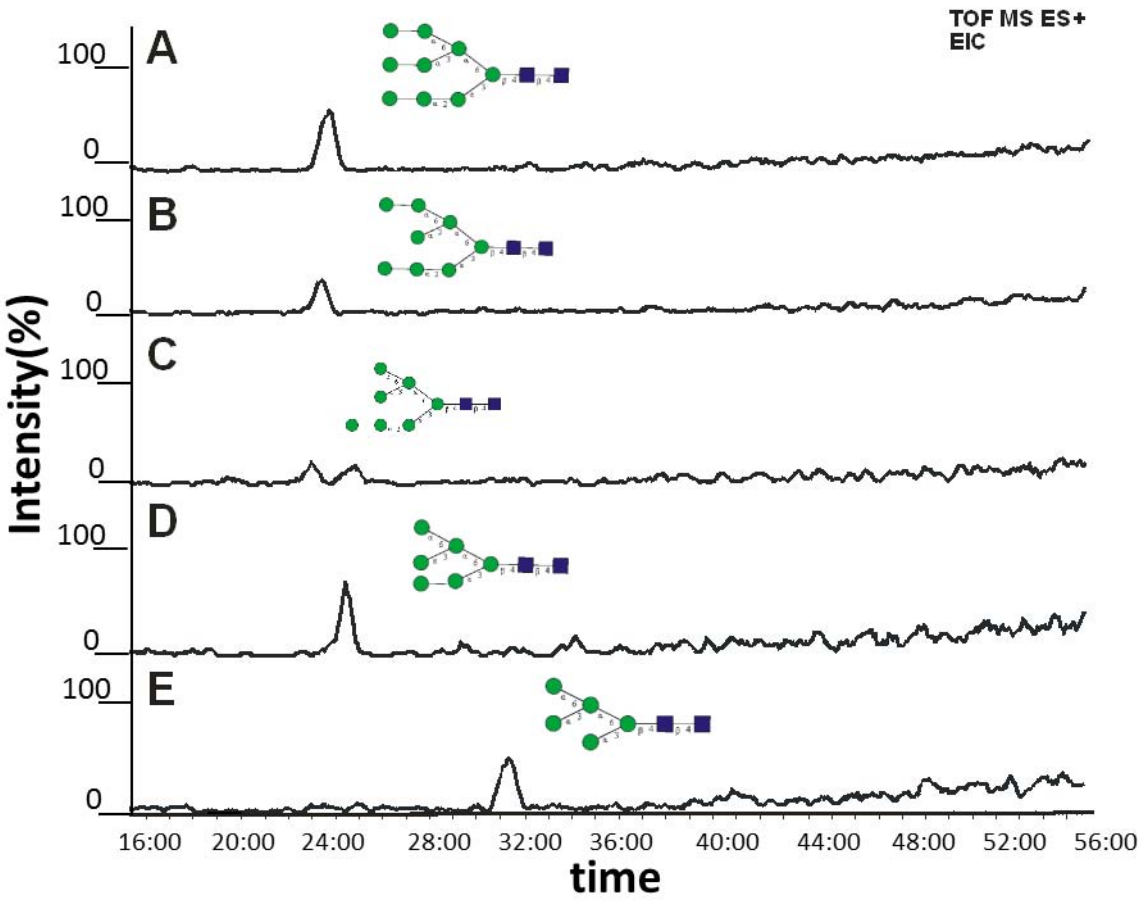


Figure 26 Elutionprofiles of oligomannosidic glycans in swine tissue using PGCC-QTOF-MS
Extracted ion chromatograms (EIC) of different oligomannosidic glycan structures interfering differently, in a time dependant manner, with the PGC-column. The elution profiles of (A-E) Man9-Man5 are shown respectively.

In mammals the occurrence of oligomannosidic glycans is of minor importance as the major glycan species found are complex type N-linked glycans and the more complex the glycan structures get the more possibilities of isomers come up. In Figure 27 the elution profiles of the different glycan species step wise building up a typical biantennary bisialylated complex glycan structure is depicted to elucidate the isomeric separation potential of a PGC column. In panel A a single peak for GnGnF^6 can be seen while in panel B the addition of a single galactose residue results in two peaks, as the galactose can be linked to the β1,3 or β1,6 antennae of the N-glycan core structure. The further addition of a N-acetyl neuraminic acid (sialic acid) leads to two additional peaks as besides the difference in the antennae the sialic acid can be either β2,3 or β2,6 linked to the galactose (panel C). Panel D shows the isomeric separation of a doubly galactosylated biantennary complex N-glycan structure. The occurrence of three peaks proves that beside the depicted structure also α1,3 linked galactose residues can be found in swine tissue as previously described by [159] which is

elucidated in more detail in Figure 28. Panel E and F show further isomeric separation due to linkage possibilities of the sialic acids.

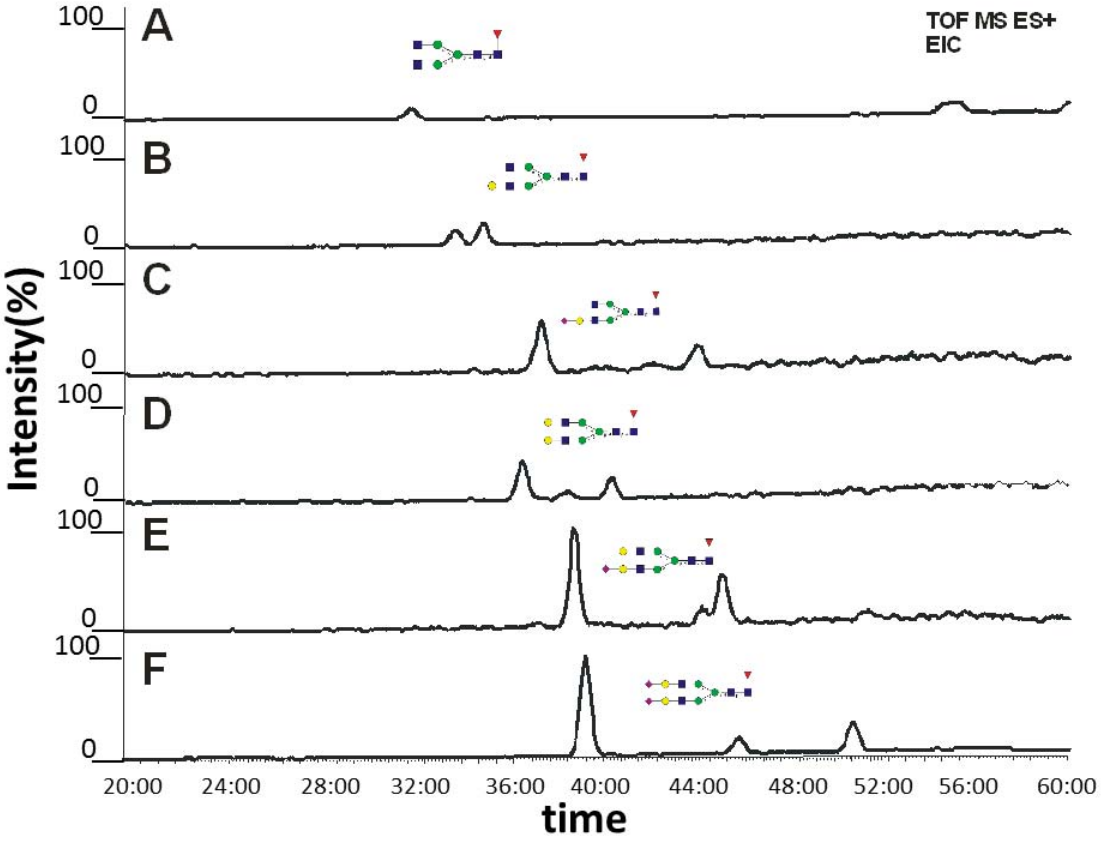


Figure 27 Elutionprofiles of step by step built up NaNaF^6 in swine tissue using PGCC-QTOF-MS
EIC of (A) GnGnF^6 till (F) NaNaF^6 showing isomeric resolution and retention time shifts the more complex the glycan structure gets.

To monitor the occurrence of the immunogenic α-galactose epitopes in the glycome of swine tissues, extracted ion chromatograms (EIC) of non, singly, doubly, triply and four times galactosylated structures (Figure 28 panel E-A) were isolated. Panel C shows three peaks for the three possible doubly galactosylated isomers. Two with an α-galactose epitope on each antennae and the first peak depicts the isomer with two β1,4 galactosylated antennas. The possibility of isomers reduces with the amount of attached galactoses as in panel B only two possible triply galactosylated peaks occur and the four times galactosylated glycan structure results in a single peak. Due to the addition of β1,4 linked galactoses as well as α1,3 linked galactoses to each antennae a clear retention time shift to the back can be observed.

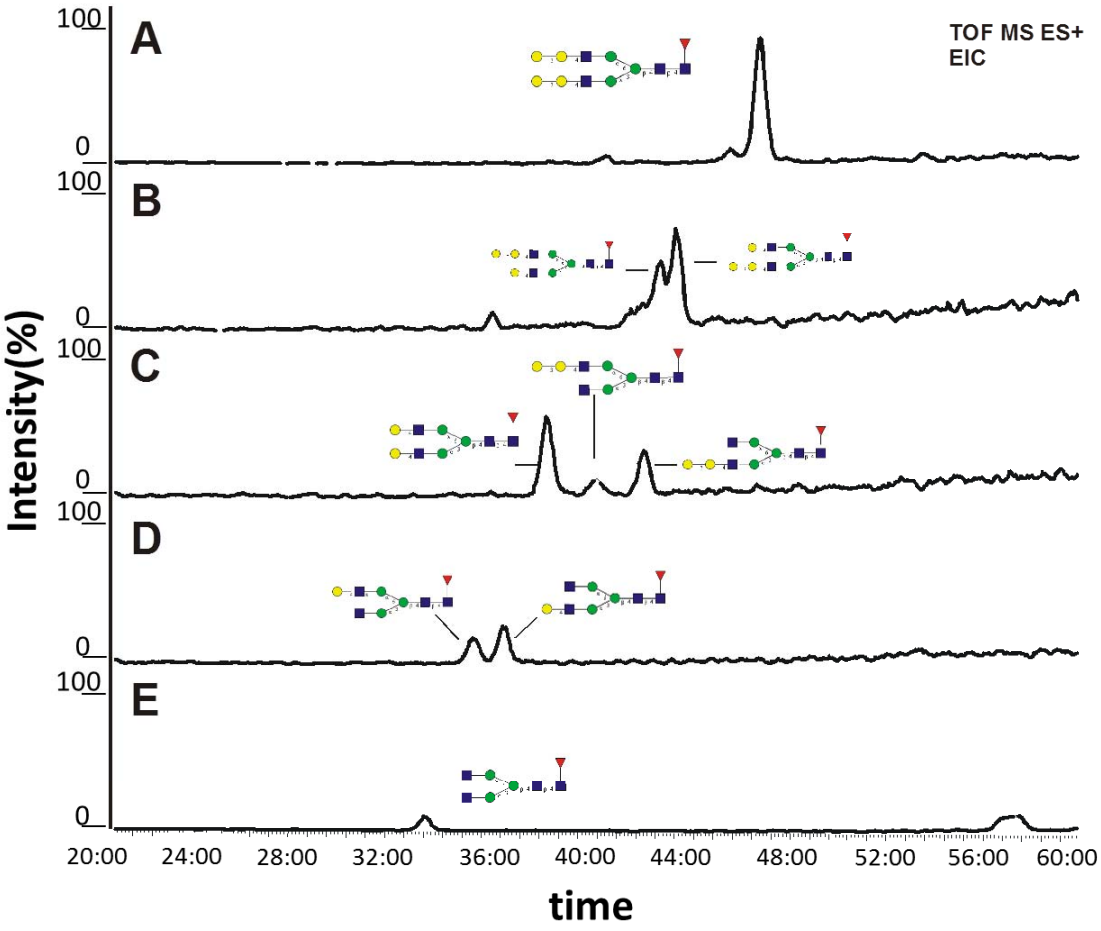


Figure 28 Elutionprofiles of α-Galactose epitopes in swine tissue by PGCC-QTOF-MS
EIC of (E) $GnGnF^6$ till (A) $(A^4A^3)(A^4A^3)F^6$ showing isomeric resolution and retention time shifts the more complex the glycan structure gets.

The PGC column can also be used to distinguish between isobaric structures. These structures, similarly to isomeric structures, consist of the same mass but differ in the molecular composition [160]. The example used here is the differentiation between $NgGnF^6$ and NaA^4 (Figure 29). Both structures have the monoisotopic mass of 1931,68 Da. To distinguish between as well as to prove the existence of both isobaric structures the EIC of the mass of 1931,68 Da m/z (panel C) can be compared to EIC of a core fucosylated singly sialylated structure carrying a N-acetylneuraminic acid instead of a N-glycolylneuraminic acid. These two structures are known to elute similarly due to their chemical appearance but they differ in mass (N-acetylneuraminic acid = 291 Da, N-glycolylneuraminic acid = 307 Da). The same can be seen in panel A where the EIC of a singly sialylated (with N-glycolylneuraminic acid) structure carrying a galactose on the other antenna is shown. The elution profile of this structures helps to elucidate the different peaks in panel C whereas the

picture gets more complicated when taking differently linked sialic acids into account. Therefore the addition of stable isotopically labelled standards or treatment with specific exoglycosidases can help to elucidate the structures in more detail.

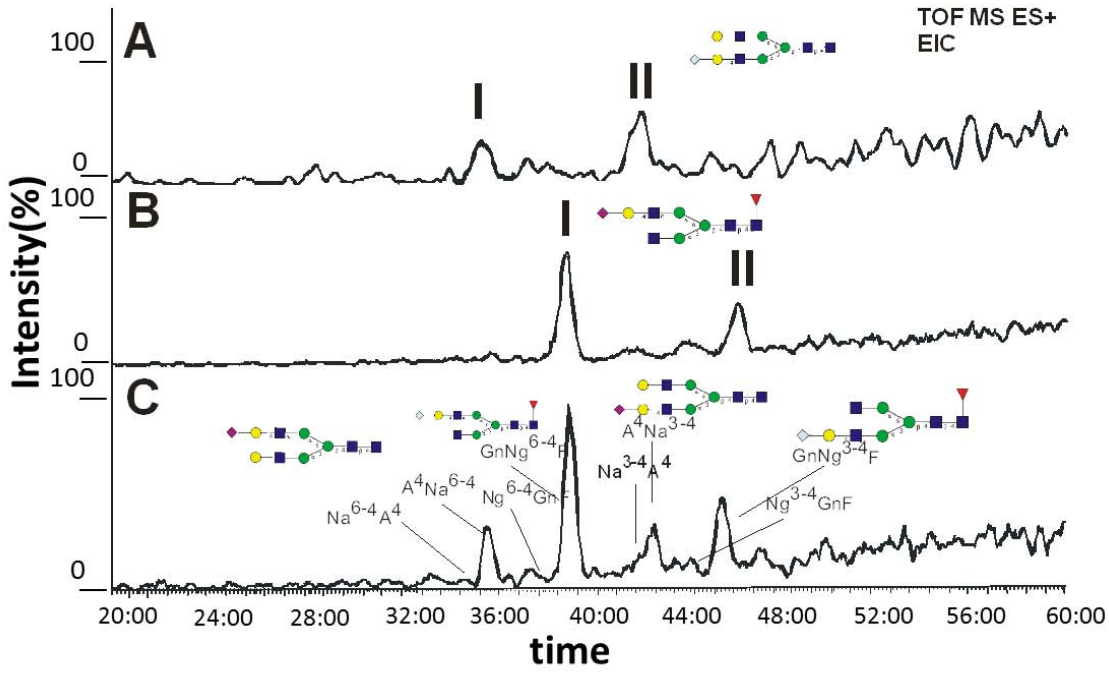


Figure 29 Elutionprofiles of how to distinguish between isobaric NaA^4 and GnNgF^6 glycans in swine tissue by PGCC-QTOF-MS
EIC of (A) NgA^4 isomers (B) NaGnF^6 isomers and (C) NaA^4 and GnNgF^6 showing isomeric resolution and how the elution profiles of (A) and (B) can be used to resolve isobaric glycans.

4.1.2 Investigation of swine O-glycans using LC-ESI-QTOF-MS

Beside N-glycan analysis also O-linked glycans from swine tissues were investigated by PGCC-ESI-QTOF-MS. The O-glycans were isolated by an in-gel reductive β-elimination (3.3.1.4).

As depicted in Figure 30 the O-glycans are fairly low abundant and mostly consist of core 1 type and some core 2 type glycan structures [2] decorated with one or two sialic acids either N-acetylneuraminic acid or to a lower extent N-glycolylneuraminic acid as well as additional galactoses.

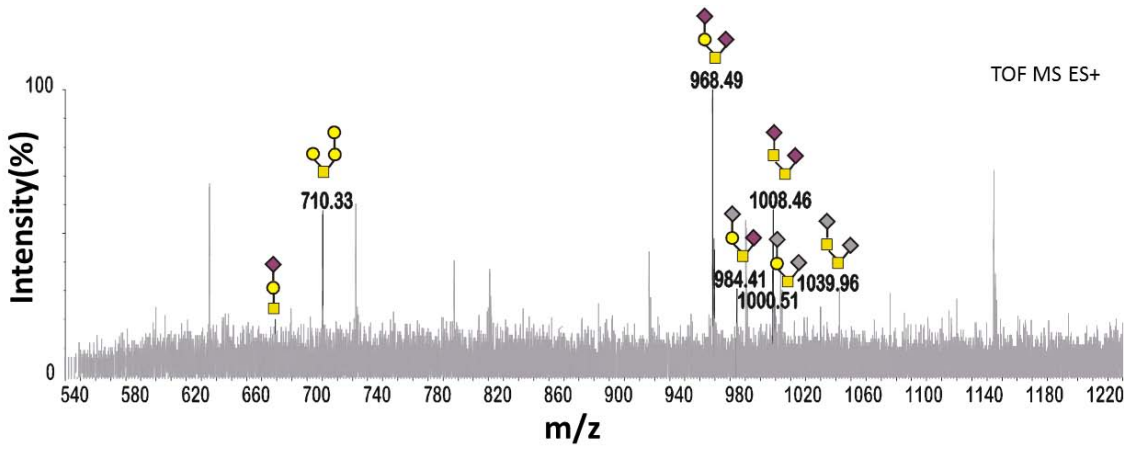


Figure 30 O-glycans in swine tissue by PGCC-QTOF-MS
$[M-H+]^+$ mass spectrum of O-glycans isolated from swine larynx. Core 1 O-glycans with one or two sialic acid decorations (m/z :677,3, 968,5, 984,4, and 1000,5) or additional galactoses (m/z 710,3) as well as core 2 type O-glycans with N-acetyl neuraminic acid or to a lower extent N-glycolylneuraminic (m/z 1008,46 and 1039,96) can be seen.

4.1.3 Investigation of swine N-glycans using MALDI-TOF-MS

Beside analysis of reduced N-glycans with LC-MS also another way of investigation glycan structures was chosen in this study. In this part of the work the main goal was to separate glycans by different dimensions of HPLC to produce fractions with best possible separation of structures. The goal to end up with fractions containing one single structure for subsequent glycan array preparation could be partly accomplished. The more complex the structures are the more difficult it is to separate them by the used stationary phases.

As for glycan array printing a primary or secondary amide group is crucial to allow covalent binding to the epoxysilane coated glass slide [161] glycans were labelled with the fluorescent label 2-AA (anthranilic acid) via reductive amination. This label also allows for the possibility of fractionation after HPLC separation as the aromatic ring of the 2-AA label has a signal optimum at 330 nm excitation and 420 nm emission which can be visualized by a fluorescent detector. Another positive side effect of the label is that it carries an intrinsic negative charge leading to the possibility of analysing the labelled glycans in negative ion mode MALDI-TOF-MS [162]. This allows for simultaneous analysis of neutral and charged N-glycans [163]. In the chromatographic behaviour the AA-label can be compared to other fluorescent labels like 2-Amino pyridine (PA) or 2- amino bencoic acid (AB), but the negative charge makes it versatile for HPLC separation, immobilization on epoxy-silane glass slides and simultaneous monitoring of neutral and charged glycans by MADLI-TOF-MS operated in linear negative mode.

Figure 31 shows total glycan spectra of N-glycans isolated from swine – lung, trachea upper (I) and lower (II) part as well as larynx tissue. The glycan content is fairly similar within the three tissue types and therefore it was decided to pool all glycans together prior the chromatographic separation steps. The major differences between the tissues lie in the higher abundance of oligomannosidic glycans (m/z 1516,6, 1678,6, 1840,6 and 2002,6; Man6-Man9) in lung and larynx tissues (panel A, D) in contrast to tracheal N-glycans (panel B, D). Furthermore (panel A) the abundance of N-glycolylneuraminic acid (Ng) (m/z 2375,7 = NgNg and m/z 2520,9 = $NgNgF^6$)in contrast to N-acetylneuraminic acid (Na) is much more abundant in lung tissue compared to the other tissue sources. The combining of the different glycan pools is also more feasible considering the subsequent fraction amount printed on the glycan array.

The overall glycan content is very well comparable to the investigations performed with PGCC-ESI-MS analysis though the isomeric diversity cannot be visualized by MALDI-TOF-MS which demands for additional exo- and endo glycosidic treatment.

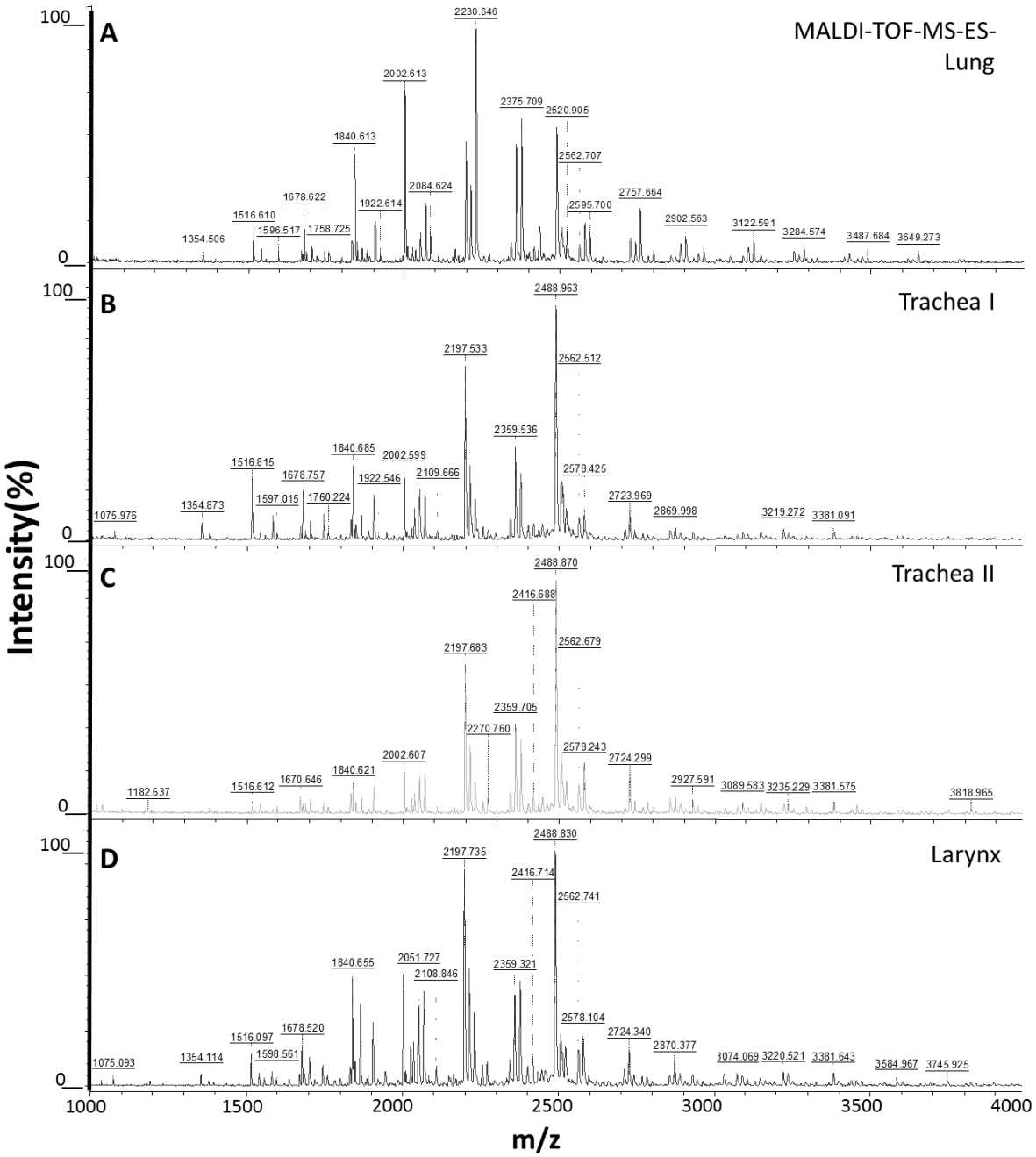


Figure 31 AA-labelled N-glycans from swine tissues using MALDI-TOF-MS
The N-glycans from different porcine sources of the respiratory tract (A) lung, (B) trachea I – upper part, (C) trachea II – lower part and (D) larynx were analysed by MALDI-TOF-MS in linear negative mode after PNGase F release and reductive amination with AA.

Figure 32 shows an annotated MALDI-TOF-MS spectrum of N-glycans isolated from swine larynx, as a reference for the other tissues, operated in linear negative ion mode. The upper panel shows an overall spectrum of the total glycan content including relative abundances of the different glycan species. To facilitate the annotation panel A-C are zoomed-in spectrum parts as indicated in the total glycan spectrum.

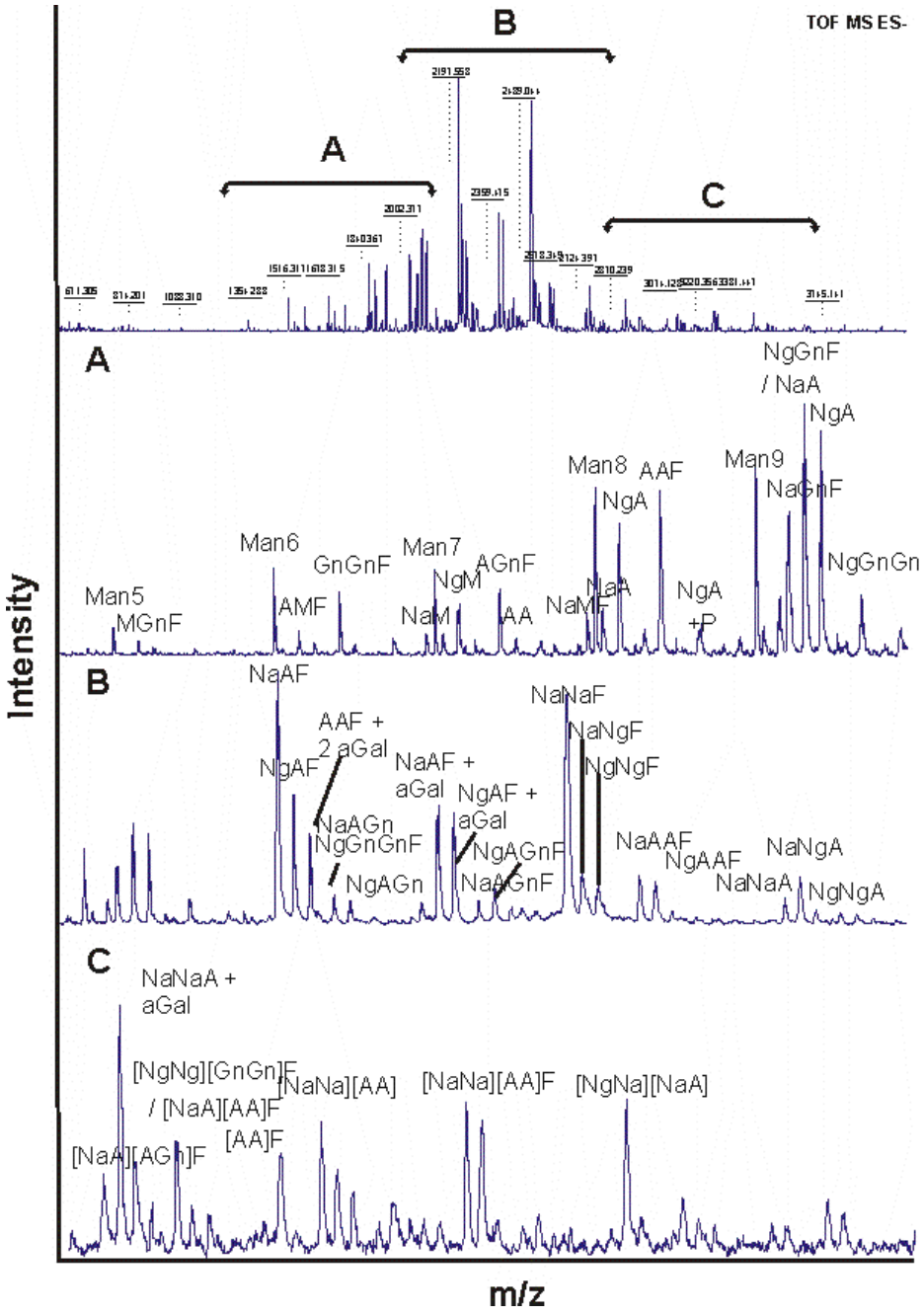


Figure 32 Detailed glycome of swine N-glycans on the example of larynx using MALDI-TOF-MS
Detailed glycan assignment of N-glycans isolated from porcine larynx tissue. Upper panel shows the spectrum overview and (A-C) show magnified spectrum parts with glycan assignments according to the proglycan nomenclature (http://www.proglycan.com/) [164]

For the preparation of a glycan array with glycans originating from a natural source it is crucial to choose a separation strategy resulting in best possible separation of the different glycan species naturally occurring in biological systems. The so far described glycome of swine glycans [159] consists of 43 different glycan structures occurring to various extents within the different analysed tissue types.

The separation strategy used in this study consisted of three different chromatographic dimensions. The strategies described so far used separation techniques with only two dimensions of a first HILIC separation and a second RP separation [109]. The addition of the weak anion exchange column as a first dimension followed by a HILIC separation (second dimension) and a RP separation (third dimension) should help to obtain rather pure fractions in the end as the charge state of the glycans should be clear due to the addition of this first dimension.

4.1.4 Separation of swine N-glycans by weak anion exchange chromatography (WAX)

In a first chromatographic run AA-labelled N-glycans isolated from swine tissue were separated using a weak anion exchange column. With this method glycans different charges could be separated [165]. Peaks were collected as shown in Figure 33 and dried for further separation. N1 and N2 indicate both neutral N-glycans. Additionally the intermediate between neutral and singly charged glycans (1x) was collected, as well as the intermediate between singly and doubly charged glycans (2x). The last peak collected were triply charged glycans (3x) though the content was very low compared to the other fractions.

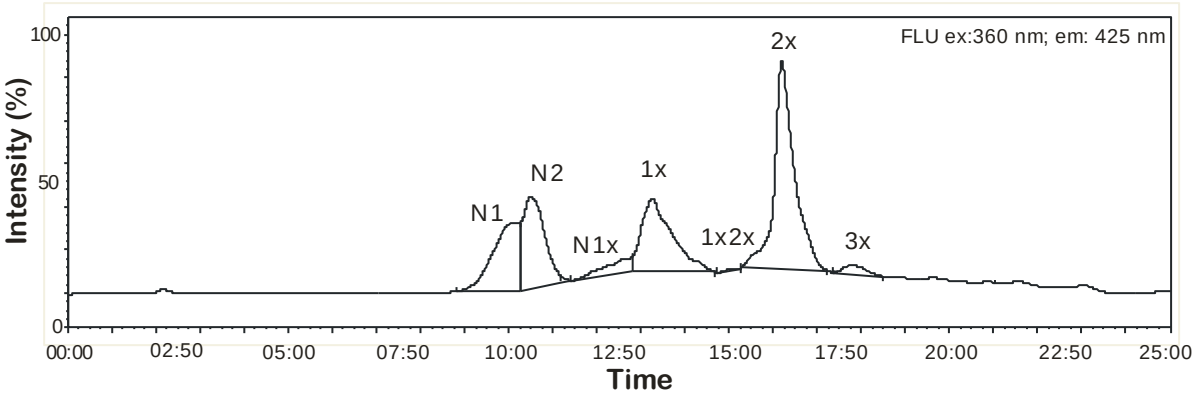


Figure 33 Swine glycans separated by Weak Anion Excange chromatography (WAX) as first separation dimension
HPLC chromatogram of WAX separated AA labelled swine N-glycans. 7 fractions (N1, N2, N1x, 1x, 1x2x, 2x, 3x) were collected differing in the charge state. N = neutral, 1x = singly charged, 2x = doubly charge and 3x = triply charged

4.1.5 Separation of swine N-glycans using hydrophilic interaction chromatography (HILIC)

Each of the seven fractions obtained by the first WAX separation was further subjected to normal phase (NP) or hydrophilic interaction liquid chromatography (HILIC) to separate the glycans according to their polarity and hydrophility. With this separation technique the aqueous content of the mobile phase creates a water rich layer on the surface of the

stationary phase [133]. Glycans can interact with this aqueous layer via dipole-dipole interactions and hydrogen bonding and thereby a separation according to size and hydrophility is achieved. In Figure 34 a representative chromatogram of the N1 fraction is shown. After the HILIC – separation step a total number of 82 fractions was obtained, which were collected and dried in the speed vac prior further separation by reversed phase chromatography.

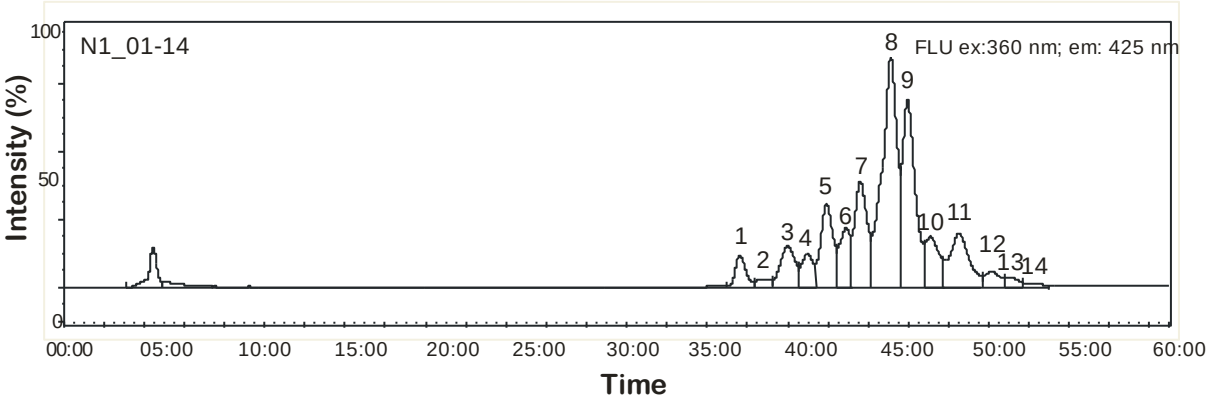


Figure 34 Normal phase (NP) separation of swine glycan fractions obtained by WAX chromatography on the example of fraction N1 as second separation dimension
NP-HPLC chromatogram of previously WAX-separated fraction N1. The neutral glycans from fraction N1 are separated by NP chromatography resulting in 14 fractions (N1_01 – N1_14).

4.1.6 Separation of swine N-glycans using reversed phase chromatography (RP)

As a final separation step a reversed phase column was used. The separation principle of the C18 column lies in the effects of hydrophobicity and polarity [126]. Each of the 82 HILIC fractions was run on the RP column leading to further separation and simultaneous purification and desalting of the glycans. After this last separation a total number of 372 fractions were collected. Quantification of each fraction was performed after peak integration by correlating the peak area to the area of an AA-labelled Man5 – standard. Each fraction was checked by MALDI-TOF-MS for the glycan content. All MALDI-positive fractions were used for subsequent printing on the glycan array in different concentrations (<0,5 µmole – 30 µmole) if available.

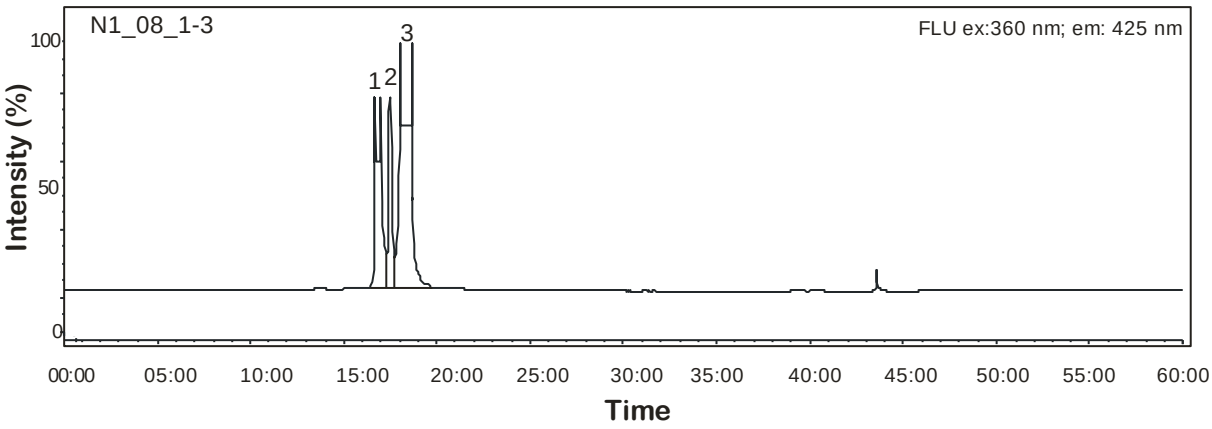


Figure 35 Reversed phase (RP) separation of swine glycan fractions obtained by NP-chromatography on the example of fraction N1_08 as third separation dimension
RP-HPLC chromatogram of previously NP-separated fraction N1_08. The normal phase separated fraction N1_08 is further separated by reversed phase chromatography resulting in three fractions (N1_08_01 – N1_08_03).

Figure 35 shows a representative RP - chromatogram of fraction N1_08 resulting in three major peaks (N1_08_01- N1_08_03). Each peak was analysed after drying by MALDI-TOF-MS as depicted representative in Figure 36.

4.1.7 MALDI-TOF-MS

Even after three separation dimensions it is clearly visible that not all of the fractions are 100 % pure and contain only one singly structure. Though in most cases one glycan structure is the most prominent one as shown in panel C of Figure 36.

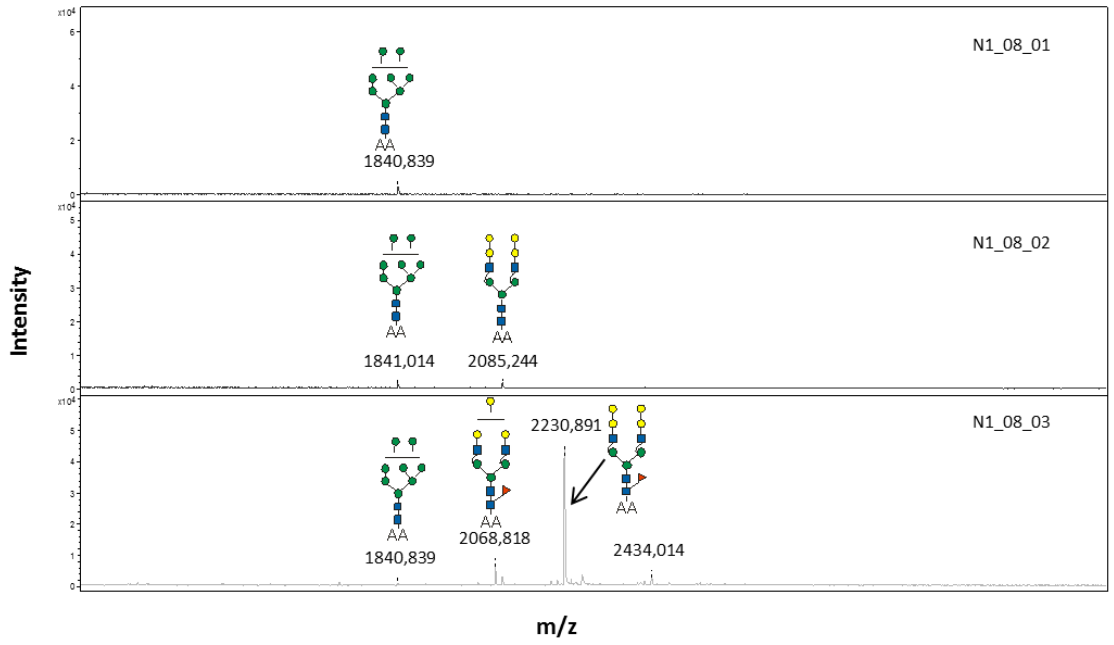


Figure 36 MALDI-TOF-MS analysis of three dimensional separated swine glycans on the example of fractions N1_08_01-N1_08_03
MADLI-TOF-MS spectra in linear negative mode were taken of all 372 reversed phase fractions (data not shown) to check for glycan content. (A) shows the MALDI spectra of fraction N1_08_01 - a single Man8 peak. (B) shows fraction N2_08_02 containing Man8 and $(A^4A^3)(A^4A^3)F^6$ in equal intensities whereas (C) fraction N1_08_03 most prominently contains $(A^4A^3)(A^4A^3)F^6$ but also other neutral glycans.

4.2 Glycan arrays

The principle of the glycan array assay can be seen in Figure 16. It follows the outlines of an immuno-adsorbent assay and uses immobilized glycans as a first binding partner for recombinant expressed hemagglutinin. The hemagglutinins were recombinantly expressed in insect cells by Florian Krammer. The 7 different constructs all contained a 6x His-Tag for purification as well as for detection in the glycan array experiment. Furthermore the constructs were expressed without transmembrane region to obtain a soluble expressed protein in the cells supernatant. The third necessity to gain soluble and active recombinant hemagglutinin is the addition of a trimerization signal [166] in the N-terminus of the protein as this trimerization is crucial for protein activity. The total number of different glycan fractions printed on the array after verification by MALDI-MS was 152 and together with positive controls and blanks (only printing buffer) and different concentrations of the single fractions 348 spots were printed in quintuplicates (5x) on each array. This small number of glycan fractions allowed a total number of four single arrays printed on one glass slide and the possibility to perform four assays simultaneously by separating the arrays with a silicon gasket. The first blocking step of the array was performed with ethanol amine to prevent

unwanted binding of the hemagglutinin as well as antibodies to the glass surface. In a next step the recombinant hemagglutinin was added in tested concentrations of 0,5; 1; 2; 4 and 8 µg/ml where the optimal concentration was at 2 µg/ml. The detection of the bound hemagglutinins was performed with the anti-His- Alexa fluorophor 647 labelled antibody which gave a fluorescent signal after scanning with a laser at 633 nm a wavelength where the fluorescent label (AA) of the immobilized glycans does not give a signal.

The analysis of the data was performed with the genepix 6.0 software to obtain the signal intensity per spot and further calculations were performed with Microsoft excel, like averaging spot replicates, setting thresholds to sort out positive and negative results, blank subtractions of median values as well as log 10 value calculations for subsequent cluster analysis by the multi experiment viewer (MeV).

4.2.1 Ferret arrays

In this work two different glycan arrays were used. One was produced by Miranda de Graaf (Erasmus MC, Rotterdam, The Netherlands) with glycans isolated from the respiratory tissues from ferrets. The array production was the same as described for the swine array though only a two dimensional separation of the glycans was performed with HILIC and subsequent RP separation. All different recombinant expressed hemagglutinins obtained by Florian Krammer were tested on the ferret array.

Figure 37 shows the binding profile of H3 (HK68) and two H1 hemagglutinins (New Caledonia 99 and Cal0409) to the synthetic glycol-conjugates printed on each array as a positive control. All three hemagglutinins bind α2,6 linked sialic acid (6-sialyllactose, a1 acid) as well as α2,3 linked sialic acid (3-sialyllactose) in a concentration dependant manner. The results of the most prominent binding specificity to α2,6 linked sialic acid is not surprising and reported in the literature, though the comparable strong binding of the α2,3 linked sialic acid was only reported for H3 (HK68) [65] [106]. The H1 hemagglutinins show a clear preference of the α2,6 linked sialic acids.

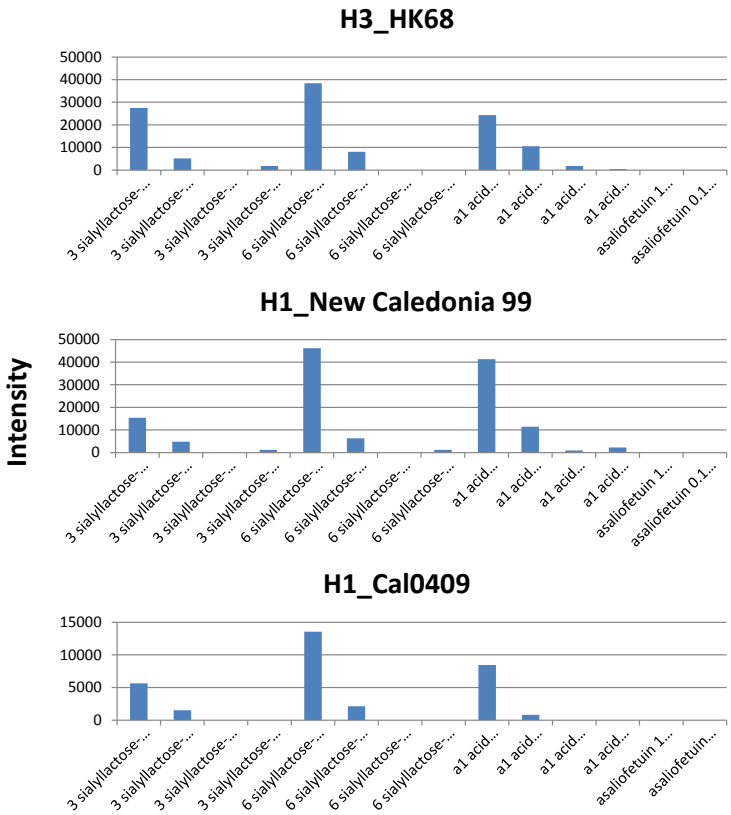


Figure 37 Binding profile of H3-HK68, H1_New Caledonia 99 and H1_Cal0409 to synthetic glycol-conjugates on ferret glycan array

Synthetic substrates were printed on each glycan array in different concentrations. 3-sialyllactose (α2,3 linked sialic acid), 6-silyllactose (α2,6 linked sialic acid), a1 acid (α2,6 linked sialic acid) and asialyfetuin (no sialic acid). Recombinant hemagglutinins (H3_HK68, N1_NewCaledonia_99, H1_Cal0409) carrying a HIS-Tag were used for binding experiments. Signals show laser induced fluorescence intensities at 532 nm. For detection of binding the Anti-His-Tag-Alexa Fluor 647 antibody was used. All three hemagglutinins bind α2,3 linked as well as α2,6 linked sialic acid, but to higher extends the α2,6 linked isomer.

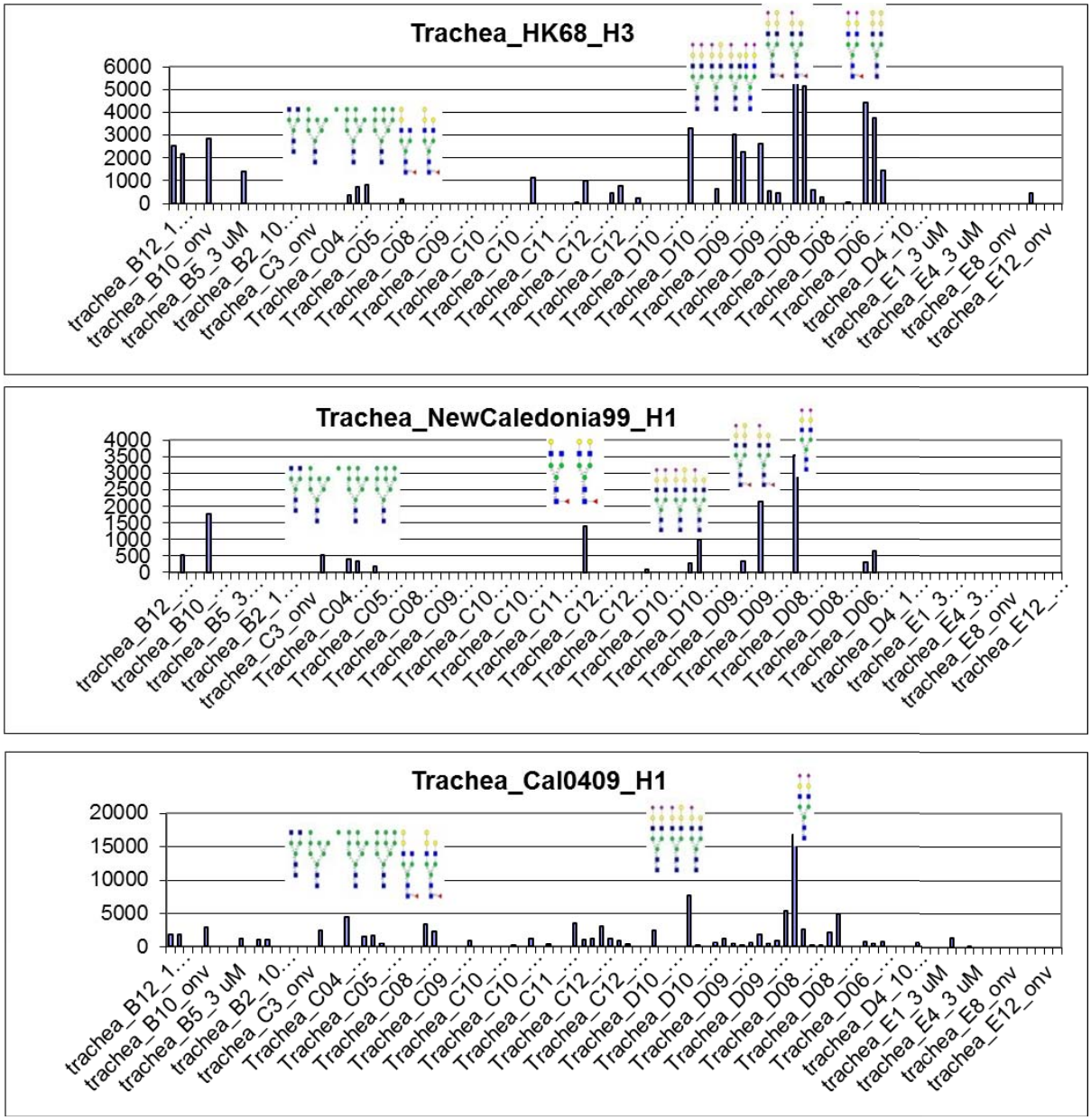


Figure 38 Hemagglutinin binding to ferret glycan arrays
AA-labelled glycan fractions from N-Glycans isolated from ferret trachea tissue were printed on epoxysilane coated glass slides. Recombinant hemagglutinins (H3_HK68, N1_NewCaledonia_99, H1_Cal0409) carrying a HIS-Tag were used for binding experiments. Signals show laser induced fluorescence intensities at 532 nm . For detection of binding the Anti-His-Tag-Alexa Fluor 647 antibody was used.

Figure 38 shows the binding specificity of H3 (HK68) and the two H1 hemagglutinins (New Caledonia 99 and Cal0409) to N-glycans isolated from ferret tracheal tissue. All recombinantly expressed proteins show rather broad binding specificities. Beside the expected binding to sialic acid (α2,3 and α2,6 linked sialic acids occur in ferret tissues) containing spots also binding to fractions containing neutral glycans was observed. Oligomannosidic as well as galactosylated N-glycans were bound by the recombinant hemagglutinins, though the broadest or highest unspecific binding was observed by the H1 hemagglutinin from Cal9409. The combined binding specificity to α2,3 and α2,6 linked sialic

acids can be clearly seen for the H3 hemagglutinin whereas the H1 hemagglutinins more specifically bind α2,6 linked sialic acids containing spots.

4.2.2 Swine arrays

Comparable to the ferret arrays prepared by Miranda de Graaf on each of the swine glycan arrays, synthetic glycan structures in different concentrations were printed to evaluate the quality of the array and give additional information of sialic acid preferences of the tested hemagglutinins. Results are shown in Figure 39 again showing a broad binding specificity to both sialic acid linkages of H3_HK68 and H1_New Caledonia 99 though the concentration dependence was not as accurate as it should be compared to the ferret arrays. H1_Cal0409 showed a rather inconclusive binding pattern to α2,3 (YDS-_biotin) and α2,6 (6-sialyllactose) linked sialic acids.

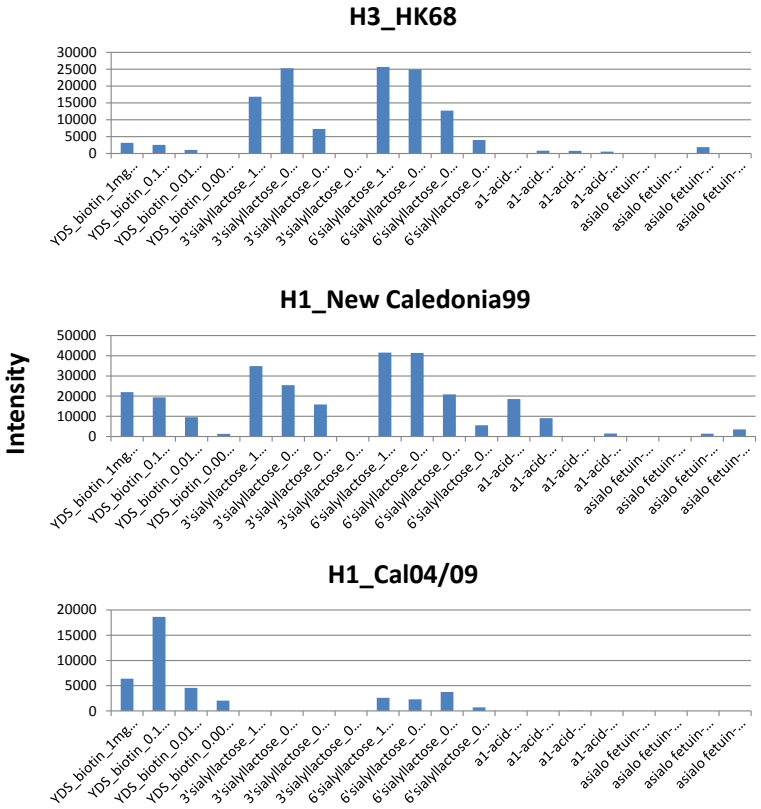


Figure 39 Binding profile of H3-HK68, H1_New Caledonia 99 and H1_Cal0409 to synthetic glycol-conjugates on swine glycan array
Synthetic substrates were printed on each glycan array in different concentrations. YDS-biotin (α2,3 linked sialic acid) 3-sialyllactose (α2,3 linked sialic acid), 6-silyllactose (α2,6 linked sialic acid), a1 acid (α2,6 linked sialic acid) and asialyfetuin (no sialic acid). Recombinant hemagglutinins (H3_HK68, N1_NewCaledonia_99, H1_Cal0409) carrying a HIS-Tag were used for binding experiments. Signals show laser induced fluorescence intensities at 532 nm. For detection of binding the Anti-His-Tag-Alexa Fluor 647 antibody was used. All three hemagglutinins bind α2,3 linked as well as α2,6 linked sialic acid, but to higher extends the α2,6 linked isomer.

The binding results to the swine glycan array of the three different hemagglutinins are shown in Figure 40. Compared to the results of the ferret array, the binding profiles showed an even broader binding affinity to sialic acid containing fractions as well as fractions containing only neutral glycans. Glycan fractions containing glycans with only one sialic acid (yellow) were comparably weak bound whereas fractions containing oligomannosidic glycans as well as additional α-galactose epitopes (green) were comparably strong bound by H3 (HK68) and H1 (New Caledonia 99). The strongest binding though was observed to the fractions carrying two sialic acids. H1 (Cal94/09) behaved slightly different. The most prominent glycan fractions bound contained two sialic acids but comparably weak binding was also observed for fractions containing one sialic acid as well as neutral fractions.

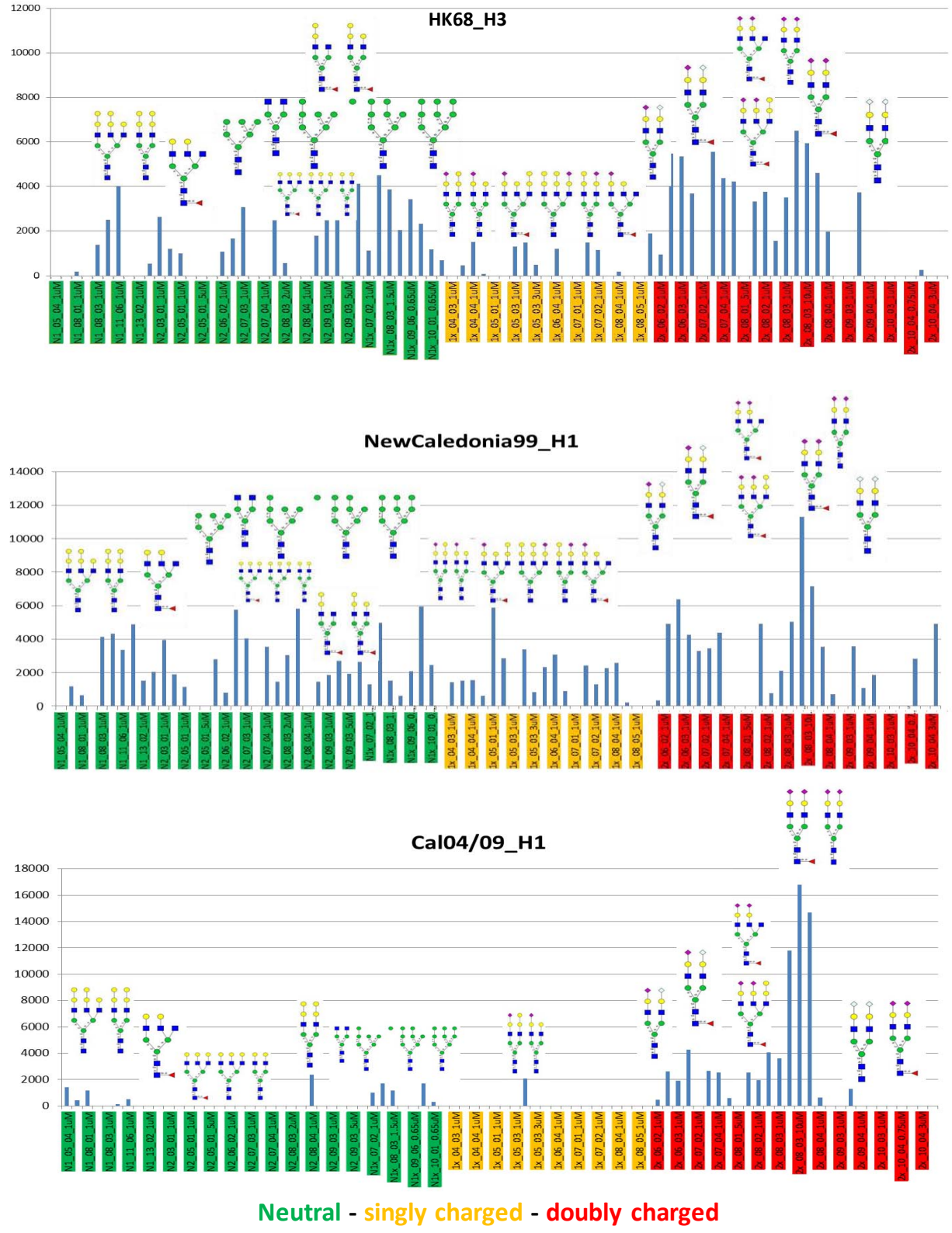


Figure 40 Hemagglutinin binding to swine glycan arrays

AA-labelled glycan fractions from N-Glycans isolated from swine tissue were printed on epoxysilane coated glass slides. Recombinant hemagglutinins (H3_HK68, N1_NewCaledonia_99, H1_Cal04/09) carrying a HIS-Tag were used for binding experiments. Signals show laser induced fluorescence intensities at 532 nm . For detection of binding the Anti-His-Tag-Alexa Fluor 647 antibody was used.

4.2.3 Cluster analysis of ferret and swine arrays

To get more insights into the rather broad binding profiles of the recombinant proteins to the ferret as well as swine arrays cluster analysis using R software were performed. Data was log^{10} transformed followed by range-scaling according to van den Berg et al. 2006 [152]. This scaling method was used because the focus was lying on the difference in the dataset according to their biological relevance instead of the technical variance inherent to immunosorbent assays itself. As expected the assay replica of the glycan array (swine and ferret arrays) experiments cluster together. To get even more conclusive results glycan fractions of only one single concentration were used (10 µm – ferret array and 1 µM – swine array) to prevent biased results in the cluster analysis.

Red coloured fractions show high glycan binding whereas blue fractions indicate a negative binding result. In the ferret array results the highest binding was observed to glycan fractions mainly isolated from bronchus and nasal turbinate (NT) tissue whereas trachea and lung glycans were weaker bound. These results are constant within all compared assays and can be seen by the clear separation of glycan fractions into two major clusters (cluster one: NT and Bronchus; cluster two lung and trachea). N-glycans from ferret tissues are known to contain more α2,3 linked than α2,6 linked sialic acid. Therefore the higher binding of H3 (HK68) and H1 (New Caledonia) to the bronchus and NT fractions (all mainly containing sialic acid decorated N-glycans) in contrast to the less intense binding of H1 (Cal04/09) is in accordance with the binding results to the synthetic glycol-conjugates (Figure 37).

A similar picture can be seen when comparing the cluster analysis results of the swine glycan array. Again array replica of H3 (HK68) together with H1 (NewCaledonia99) and replica of H1 (Cal04/09) tend to cluster together. Furthermore the highest binding can be detected to doubly charged glycan fractions, which was expected. As the H1 hemagglutinin (Cal04/09) shows more intense binding, indicated by the deep red colour, in contrast to H3 and the second investigated H1, which show a lighter red colour, the conclusion of the more intense binding to α2,6 linked sialic acid can be drawn. This result is in accordance with the binding profile to the synthetic glycol-conjugates as well as the general binding profiles to the swine N-glycans shown in Figure 39 and Figure 40, respectively.

Furthermore intense binding of all hemagglutinins can be seen to several fractions which are putatively annotated as containing neutral and singly charged N-glycans.

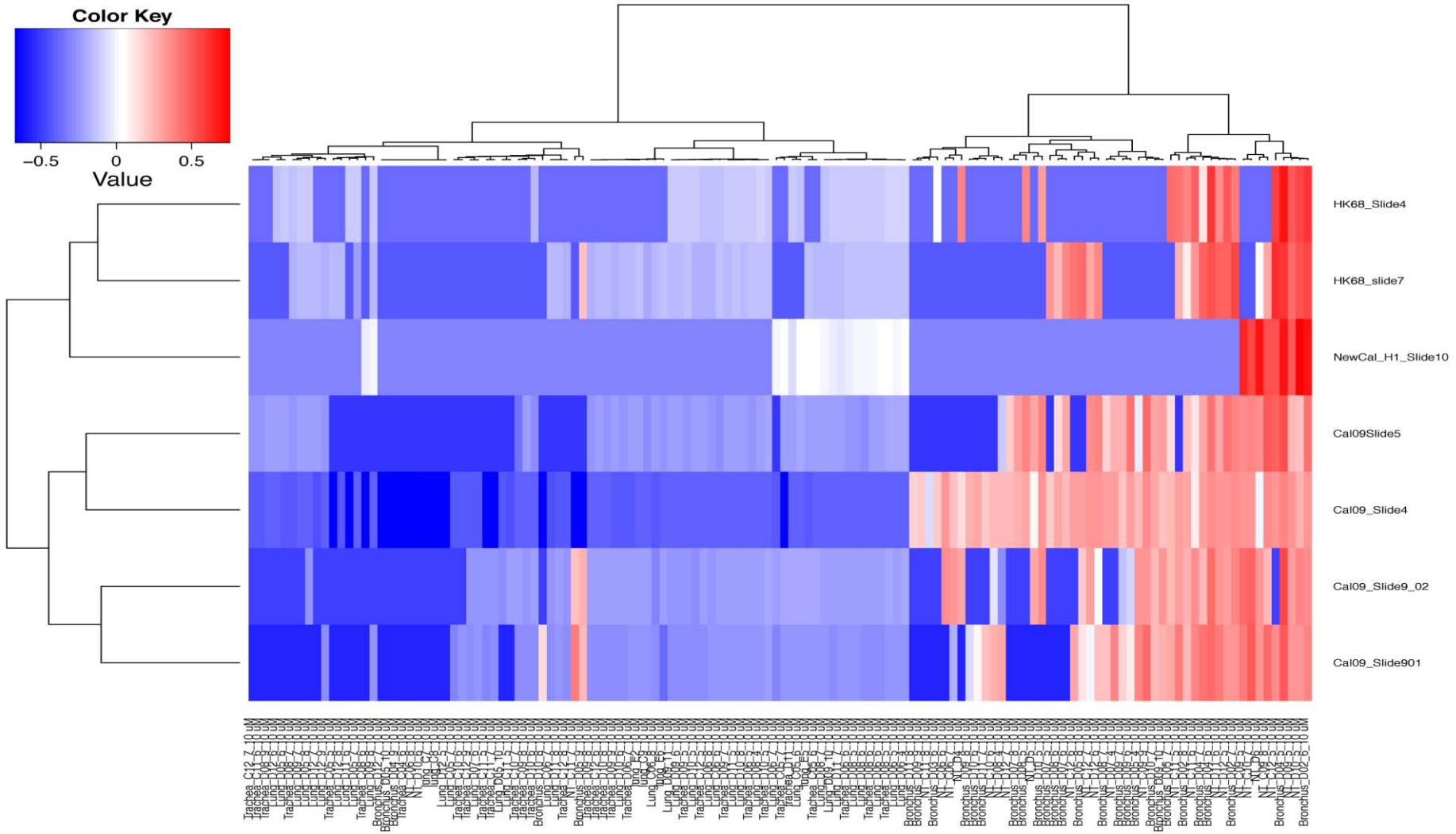


Figure 41 Cluster ananlysis of H3 (HK68) and H1 (New Caledonia 99 and Cal04/09) hemagglutinins ferret glycan arrays

R was used for creating the heatmap and for the hierarchical cluster analysis (HCA using Euclidean distance and Ward's linkage clustering) on both sides of the heatmap. The dataset was $\log^{10}$ transformed and range-scaled according to van den Berg et al.

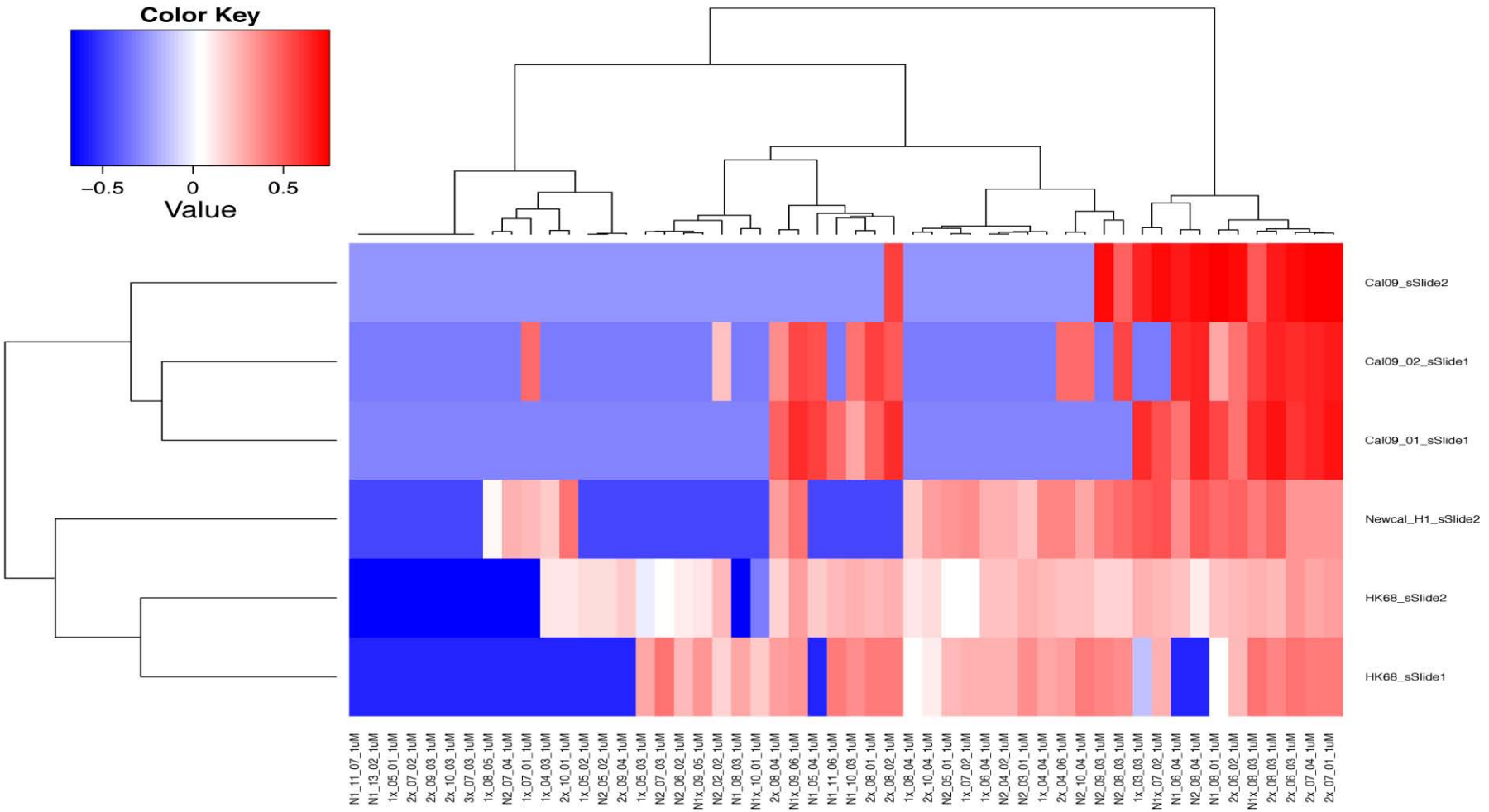


Figure 42 Cluster ananlysis of H3 (HK68) and H1 (New Caledonia 99 and Cal04/09) hemagglutinins swine glycan arrays
R was used for creating the heatmap and for the hierarchical cluster analysis (HCA using Euclidean distance and Ward's linkage clustering) on both sides of the heatmap. The dataset was $\log^{10}$ transformed and range-scaled according to van den Berg et al.2006.

4.3 Affinity chromatography

In this study several glycan binding proteins like the LCA lectin, ConA lectin, galectin or recombinantly expressed hemagglutinins were coupled to NHS-activated sepharose as described in 3.2.1.1. LCA lectin, ConA and galectin were model proteins used for the method set up for subsequent experiments with the H1 hemagglutinin from swine-origin H1N1 influenza A virus. To compare binding results two different hemagglutinins were used (H1 – H1N1 California/04/09 and H3 - H3N2 Victoria/361/11) to investigate binding affinities of different N-glycans isolated from swine tissue.

4.3.1 LCA-lectin (*lens culinaris* agglutinin) coupling to NHS activated sepharose

One pre-experiment was the isolation of the LCA lectin isolated from the *lens culinaris*. This lectin is known to bind α1,6 core fucosylated glycan structures. Therefore the isolated protein was successfully immobilized on NHS-activated sepharose as shown in Figure 43. Subsequent glycan binding experiments were performed in batch columns with an inner diameter of 4 mm.

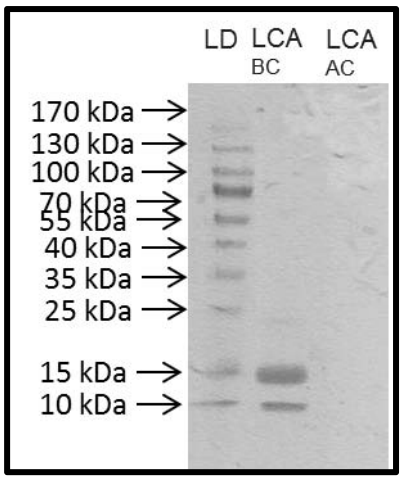


Figure 43 Coupling of LCA-lectin to NHS-activated sepahrose for affinity chromatography experiments
SDS-PAGE of 10µl aliquots of LCA-protein solution before (LCA BC) and after (LCA AC) coupling to NHS-activated sepharose.17 kDa Band diminishes after 4 h incubation indicating binding to the sepharose.

4.3.2 LCA sepharose glycan binding

The used glycans, to show binding to the immobilized LCA lectin, were GnGn, which served as a negative control and $GnGnF^6$ as a positive control. Both glycans were labelled with the fluorescent label 2-aminopyridine (PA). Flow through, wash and elution (15 mM

methylmannosid) fractions were collected and subsequently analysed by LC-ESI-IT-MS. The chromatographic separation was performed with a C18-Reversed Phase column which allows retention of the glycans via the PA label. Table 8 shows the results of the relative intensities indicating that no retention or binding of the glycans was detectable. As the lectin is binding to the core fucose the question arose about the possibility of a steric hindrance due to the PA – label which is in close proximity to the core fucose.

Wash

Fraction No	GnGn-PA	GnGnF6-PA
1	+++	+++
2	+++	+++
3	++	++
4	++	++

Elution

Fraction No	GnGn-PA	GnGnF6-PA
1	++	++
2	+	+
3	~	~
4	-	-

Table 8 Binding of PA-labelled glycans to LCA-lectin coupled sepharose using NP – HPLC for detection

Fluorescently labelled GnGn-PA and GnGnF^6-PA were loaded onto the LCA-lectin column and wash fractions (1-4) as well as elution fractions (1-4) with 15 mM methylmannosid were collected and checked for glycan content by NP-HPLC with fluorescence detection. Both glycan species showed no binding and were washed off the column. +++ = very strong positive, ++ = strong positive, + = positive, ~ = detectable, - = negative

Therefore another set of experiments was performed using soluble LCA lectin and native instead of fluorescently labelled glycans to see if binding can be achieved when the glycan core is not derivatized. The experiment was performed with amicon centrifugation tubes (Merck Millipore, Darmstadt, Germany) with a cut off of 10 kDa, which allows the unbound glycans to be washed away whereas the protein, with a molecular weight of ~ 17 kDa, as well as the protein-N-glycan complex are kept inside the membrane. After two rounds of washing the elution was again performed with 15 mM methylmannosid followed by subsequent analysis by PGCC-ESI-QTOF-MS. Figure 44 shows the normalized intensities of the EIC of two glycan structures (A^4A^4 – negative control; $GnGnF^6$ – positive control)

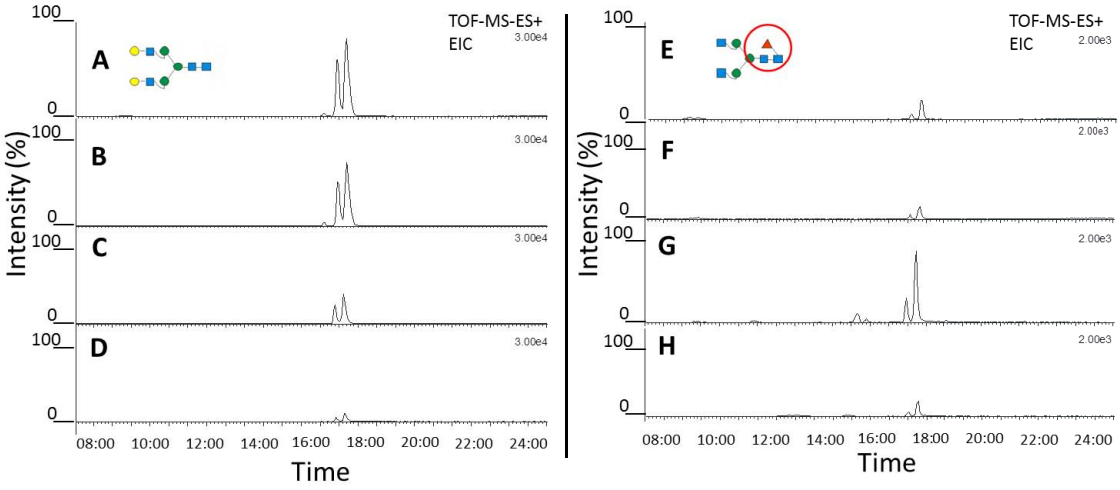


Figure 44 Binding of native gycans by LCA-lectin using amicon centrifugation tubes and PGCC-ESI-QTOF-MS for detection
EICs collected fractions of native glycans incubated with LCA-protein and separated using amicon centrifugation tubes with a 10 kDa cut off. A-D show chromatograms of A^4A^4 as a negative control continuously washed off. E-H show chromatograms of $GnGnF^6$ as a positive control bound to the LCA-lectin. (A,B, E and F) represent the washing fractions and (C,D,E and F) the elution fractions with 15 mM methylmannosid. Fractions were analysed using PGCC-ESI-QTOF-MS.

Panel A and B represent the glycan amount of A^4A^4 (negative control) in the wash fractions whereas C and D are both elution fractions indicating the continuous washing – off of the glycan without any binding effect. In contrast to that panel E and F show the glycan abundance of $GnGnF^6$ (positive control) indicating a minor overload of the protein binding capacity. The elution fractions (panel G and H) prove binding of the core fucosylated native glycan as the elution buffer released bound glycans shown by the higher abundance of the peaks in the EIC.

4.3.3 ConA coupling to NHS sepharose

Another set of preliminary experiments were performed with the mannose binding lectin ConA (concanavalin A). As described previously the lectin was covalently coupled to NHS-activated sepharose and different fluorescently labelled glycans were loaded onto the self-prepared batch column. The loading buffer was 1x PBS and after collecting the flow through a washing and elution of weak binding glycans was performed with 15 mM methylmannosid followed by a high concentration elution with 500 mM methylmannosid in 1x PBS to elute the tightly bound glycans. The fractions were investigated by normal phase HPLC and checked for the abundance of the different glycan molecules (Table 9). As expected the complex biantennary AA-PA was not binding to the column and was washed off the column immediately, whereas GnGn-PA was surprisingly well bound due to the easier access of the

lectin to the core mannoses which is reported in the literature [113]. Oligomannosidic structures (Man6-PA and Man9-PA) were both bound as expected though the binding to Man6-PA was weaker than to the Man9-PA. The same set of experiments was also performed with a ready-prepared purchased ConA coupled sepharose (GE Healthcare Life Science) which gave the same results as seen in Table 9 (data not shown).

15 mM methylmannosid

Fraction No	Man6-PA	Man9-PA	GnGn-PA	AA-PA
1	--	--	--	+++
2	--	--	--	+++
3	+	~	--	+++
4	++	~	--	+++
5	+++	~	--	+++

500 mM methylmannosid

Fraction No	Man6-PA	Man9-PA	GnGn-PA	AA-PA
1	+++	+	+	+
2	+++	++	+	~
3	+++	+++	++	~
4	+	+++	++	--
5	+	+++	++	--

Table 9 Binding of PA-labelled glycans to a Concanavalin A(ConA) - affinity column using NP – HPLC for detection
Fluorescently labelled Man6-PA, Man9-PA, GnGn-PA and A^4A^4-PA were loaded onto a ConA-lectin column and wash fractions (1-5, 15 mM methylmannosid) as well as elution fractions (1-5) with 500 mM methylmannosid were collected and checked for glycan content by NP-HPLC with fluorescence detection. A^4A^4-PA showed no binding and was immediately washed off the column, whereas Man6-PA, Man9-PA and GnGn-PA showed retention to different extends. +++ = very strong positive, ++ = strong positive, + = positive, ~ = detectable, - = negative

4.3.4 Galectin affinity chromatography

Another affinity experiments was performed with the galactose binding protein galectin. 250 µg of protein were bound to 100 µl of NHS activated sepharose as shown in Figure 45. After successful coupling, the sepharose was packed in a batch column with an inner diameter of 4 mm as described previously in 3.2.1.1. Loading of PA-labelled A^4A^4, incubation, washing and elution with buffer containing 100 mM lactose was performed. Fractions were collected and analysed by NP-HPLC. Panel A and B of Figure 45 show the wash fraction II and V respectively indicating a lot of unbound glycans. After elution with lactose which competes with the bound glycans due to its higher affinity to the galectin panel C and D prove the presence of bound A^4A^4. The drawback in these experiments is the necessity of rather big amounts of protein in the mg range. The maximum binding capacity of NHS-activated sepharose is, depending on the protein size, around 15-25 mg/ml.

Therefore the method of investigating lectin affinities by coupling the protein to sepharose is not always feasible as obtaining sufficient protein amounts can be very costly and challenging.

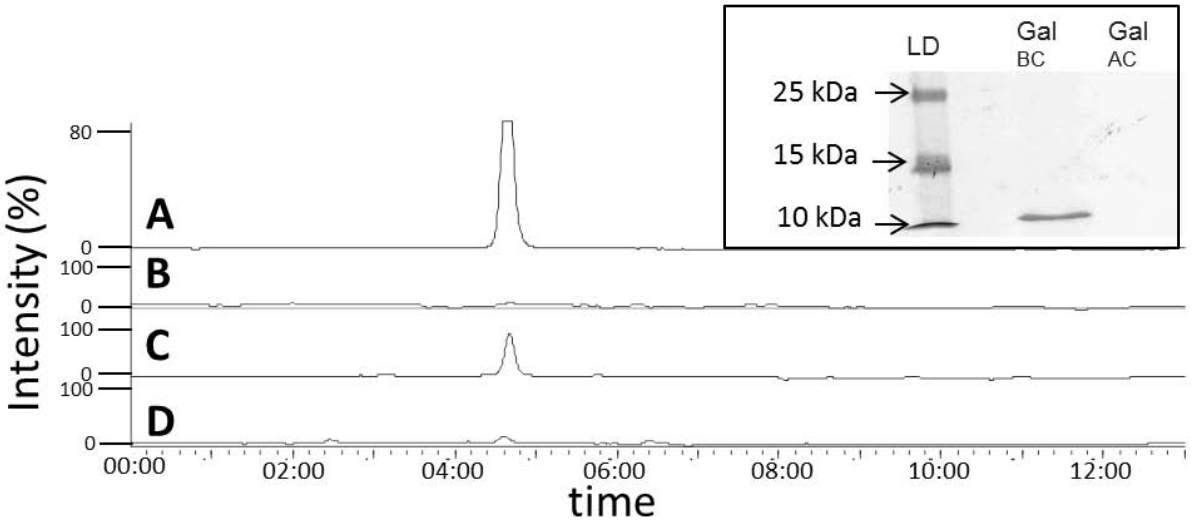


Figure 45 Binding of PA-labelled A^4A^4 by a galectin (Gal) - affinity column using NP – HPLC for detection
Upper right corner: SDS-PAGE of 10µl aliquots of Gal-protein solution before (Gal BC) and after (Gal AC) coupling to NHS-activated sepharose.~10 kDa Band diminishes after 4 h incubation indicating binding to the sepharose. NP-HPLC chromatograms (A-D): Fluorescently labelled A^4A^4-PA was loaded onto a galectin coupled affinity column and wash fractions (A,B) showed an overload of the affinity column whereas C and D show the elution of A^4A^4-PA with 100 mM lactose solution.

4.3.5 HA recombinant expression

After the preliminary experiments showed that the principle of immobilizing a protein on the sepharose worked nicely and was achieved with different types of glycan binding proteins, the next experiments were planned to continue with recombinant expressed hemagglutinins to elucidate the binding profile and affinity to glycans isolated from swine tissues.

In a first set of experiments truncated (without transmembrane region) versions of H1 (Cal07/09) as well as H3 (Hiroshima52/05) were recombinantly expressed in insect cells using the baculo virus expression system. The constructs contained a His-Tag in the C-terminus for subsequent purification and detection. As indicated in Figure 46 the purification using a Ni-NTA-agarose column worked nicely and clear bands occurred in the 40 and 80 mM Imidazole fractions. As well as in the positive control. The sizes of the proteins H1 (panel A) and H3 (panel B) are at around 70 kDa which is slightly more than stated in the literature and can be explained by the high extent of N-glycosylation on the protein.

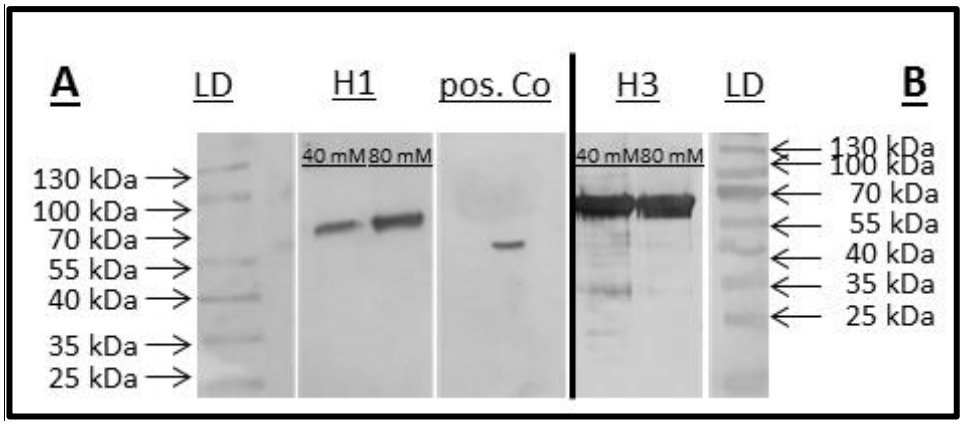


Figure 46 Hemagglutinin (H1 and H3) expression and purification using Ni-NTA agarose
Western blott experiment using of collected fractions from HIS-Tag affinity purification using Ni-NTA-agarose of recombinantly expresses H1 and H3 hemagglutinin in *Sf9* insect cells. A stepwise elution of HIS-Tag bound recombinant proteins was performed by the addition of up to 250 mM imidazole. Fractions were collected and checked for protein content using SDS-PAGE and western blott with an Anti-HIS antibody to detect positive fractions. Fractions containing recombinant protein are shown for H1 hemagglutinin (A) lanes H1 40 and 80 mM and for H3 hemagglutinin (B) lanes H3 40 and 80 mM.

4.3.6 Hemagglutination assay

Both purified proteins were used to check activity by a hemagglutination assay. Several assays were performed with the recombinant H1 and H3 with different blood groups. The first assays were performed with chicken erythrocytes but no activity could be detected. According to the literature [167] the usage of the right erythrocytes can be crucial to assess protein activity and thereby beside chicken also human erythrocytes from blood group A, B and 0 were tested without success. Furthermore different buffer conditions like 1x PBS with and without the addition of divalent cations ($CaCl_2$ and $MgCl_2$) as well as 0,9 % NaCl solution were tested. As none of the tested assay conditions worked for the proof of the HA's activity, different insect cell systems for the expression were used as well as VLP's (virus like particles) displaying the H3 hemagglutinin on the surface. In Figure 47 the results of an assay with H3 VLP's (lane 1) or the full length H3 expressed in *Sf9* cells (lane 2), resulting in insect cells displaying the protein on the outer membrane, furthermore truncated H1 (lane 3 & 4) and H3 (lane 5 & 6) (without transmembrane region) expressed in Hi5 and Ao38 insect cells, respectively, are shown. As a positive control (lane 7 & 8) the lectin ConA was used and the negative control (lane 9 & 10) consisted of buffer and chicken erythrocytes without any recombinant protein.

The only slightly positive signals can be seen in lane 1 and 2 of the H3 presenting VLP's or *Sf9* cells, respectively. Positive signals as depicted in lane 7 & 8 (pos Co, ConA) appear as a fuzzy

solutions in contrast to the negative control (lane 9 & 10) or the soluble recombinant hemagglutinins from different cell systems (lanes 3-6) which appear to be sharp spots at the bottom of the V-shaped 96 well plate. The fuzziness is a result of agglutinating protein activity where erythrocytes clot together in cell-bundles of different sizes and thereby float around instead of sedimenting as single cells to the bottom as in seen in the negative results.

The negative result of the recombinant hemagglutinins from the different cell types (lane 3-6) can be explained due to the lackage of a trimerization signal in the carboxy-terminus of the protein. This signal seems to be crucial for trimerization and accordingly activity of hemagglutinin, when expressed soluble and without the transmembrane domain [166]. VLPs displaying H3 on the outer surface show agglutinating activity but only to a very low content, which is rather surprising and can be probably explained due the low concentration of the VLP stock. The same explanation is valid for the rather weak positive signal for full length H3 expressed on the surface of *Sf9* cells.

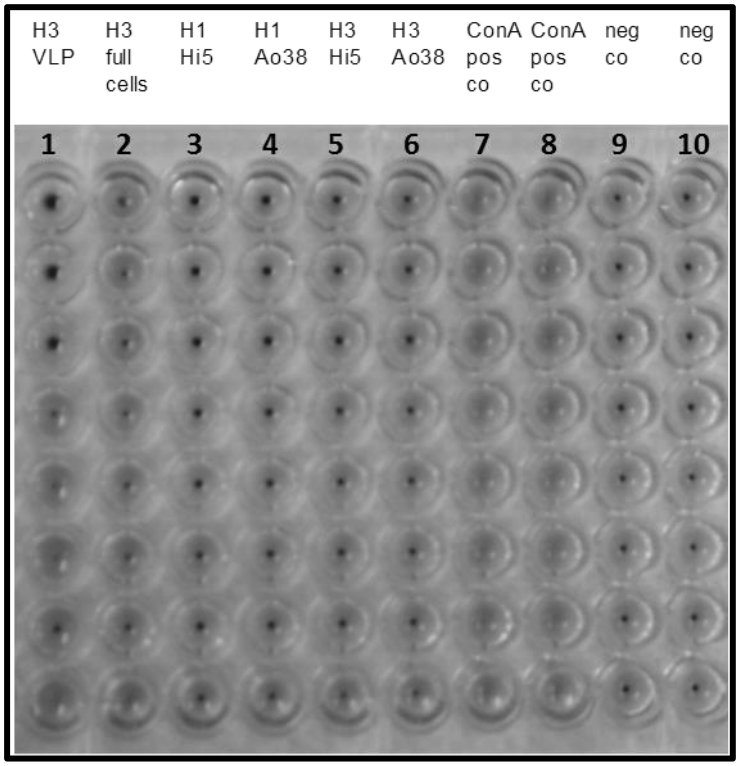


Figure 47 Heamagglutination assay of different recombinant heamgglutinins
Agglutininating activity was assessed for (lane 1) VLP's carrying H3 on the surface, (lane 2) insect cells expressing the full length protein on the cell surface and (lane 3-6) recombinant truncated H1 and H3 expressed in Hi5 and Ao38 insect cells, respectively. ConA (lane 7 and 8) served as positive control and (lane 9 and 10) were negative controls without protein addition. Serial dilutions of the samples were performed from the first to the last row. Recombinant proteins as well as VLP's and full cells displaying the entire H3 were obtained from the group of Reingard Grabherr from Dieter Palmberger and Miriam Klausberger.

For subsequent analysis by affinity chromatography and glycan array experiments only constructs containing the trimerization signal in the C-terminus, obtained by Florian Krammer, were used.

4.3.7 HA coupling to NHS activated sepharose

For affinity chromatography measurements two different hemagglutinins (H1 - H1N1 California/04/09 and H3 - H3N2 Victoria/361/11) were used to additionally monitor differences between HA's from distinct viral origin.

Figure 48 shows the successful binding of the recombinantly expressed proteins to the NHS-activated sepharose. Both proteins were coupled at a concentration of 1,5 mg protein / 0,5 ml NHS-activated sepharose. 10 µl aliquots of the coupling solution were taken prior and after the coupling incubation and loaded on a SDS-PAGE gel to monitor the diminishing of the signal due to binding to the sepharose.

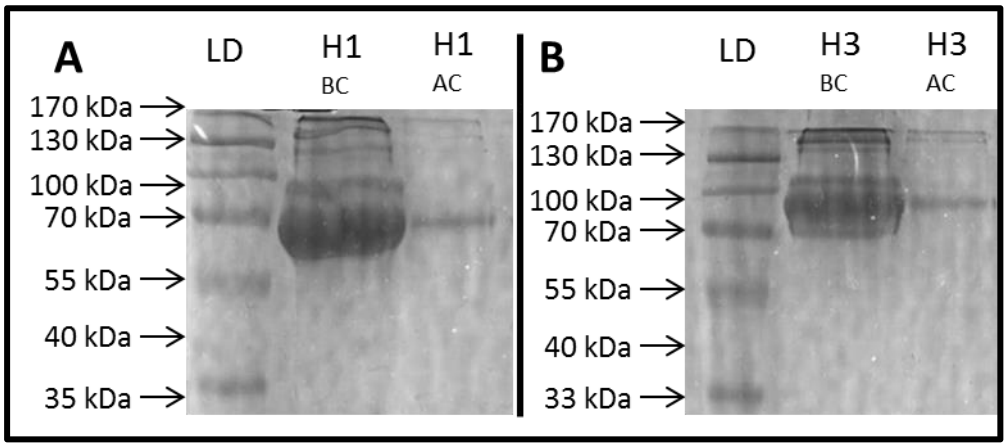


Figure 48 Coupling of H1 and H3 heamgglutinin to NHS-activated sepahrose for affinity chromatography experiments
SDS-PAGE of 10µl aliquots of (A) H1-protein solution before (H1 BC) and after (H1 AC) coupling as well as (B) H3-protein solution before (H3 BC) and after (H3 AC) coupling to NHS-activated sepharose was performed. Bands both around 72 kDa Band diminish after 4 h incubation indicating binding to the sepharose.

After the binding was completed 300 µl of the protein coupled sepharose were loaded in small scale batch columns with an inner diameter of 4 mm to perform batch experiments with AA labelled glycans in 1x PBS. For the affinity experiments the sepharose was equilibrated with 1x PBS to keep the protein in its pH optimum and thereby keep it active. The glycans were loaded onto the sepharose and incubation at RT for 30 min was performed followed by collecting the flow through and washing fractions. The chosen elution procedure was lowering the pH with 1 mM HCl to release possibly bound glycans, rather than using high salt concentrations, which would be another elution possibility, but would have needed

excessive purification steps as high salt concentrations impair ionization in subsequent MALDI-TOF-MS experiments.

4.3.8 HA Batch experiments

Figure 49 shows the ZIP-C18-TIP purified MALDI-TOF-MS spectra of the different fractions collected in the big scale batch-experiments. Panel A shows the mass spectrum of the elution fraction clearly showing that no binding of the AA-labelled glycans could be achieved. The flow through (panel B) and three wash fractions (panel C-E) indicate that the AA-glycans are continuously washed off, on first sight without any interference with the immobilized hemagglutinin. Interestingly though, when comparing the relative intensities of the different wash-spectra (C-E) panel D clearly shows a comparably higher appearance of the oligomannosidic structures man7, man8 and man9 indicating a slightly higher retaining of oligomannosidic structures. The same effect was also observed in the parallel performed experiment with H3 hemagglutinin immobilized on the NHS-activated sepharose. But taken together there is a very weak over all retaining effect of the recombinant proteins and as this set of experiments is designed to investigate only bound vs unbound glycans a possible retaining or low affinity binding effect of the recombinant proteins can't be visualised.

The batch experiments performed with H3 bound to the column also did not show any binding or retaining effect of most glycans except the oligomannosidic structures.

As hemagglutinins are known to bind sialic acid containing glycan structures it was rather surprising to see the possible retaining effect on oligomannosidic structures rather than on sialylated glycans. A possible explanation for the negative binding results can be that the binding or retaining effect probably could not be visualized due to the experiment design in combination with a generally low affinity of the recombinant hemagglutinin. The binding activity of the hemagglutinins was proven by the glycan array experiments but as the sialic acid presentation in an affinity chromatography experiment does not mimic cell surfaces like in glycan array experiments, the affinity of the HAs might be too low to firmly bind only monovalently presented sialic acid containing glycans [168]. Another theory is that the protein integrity was probably partly destroyed due to the binding procedure to the sepharose as the final washing steps of the coupling procedure are performed with different

buffers (pH 4-5 and pH 8-9). Therefore another idea was to test the intact protein without modifications in a “shot gun” approach using amicon centrifugation tubes.

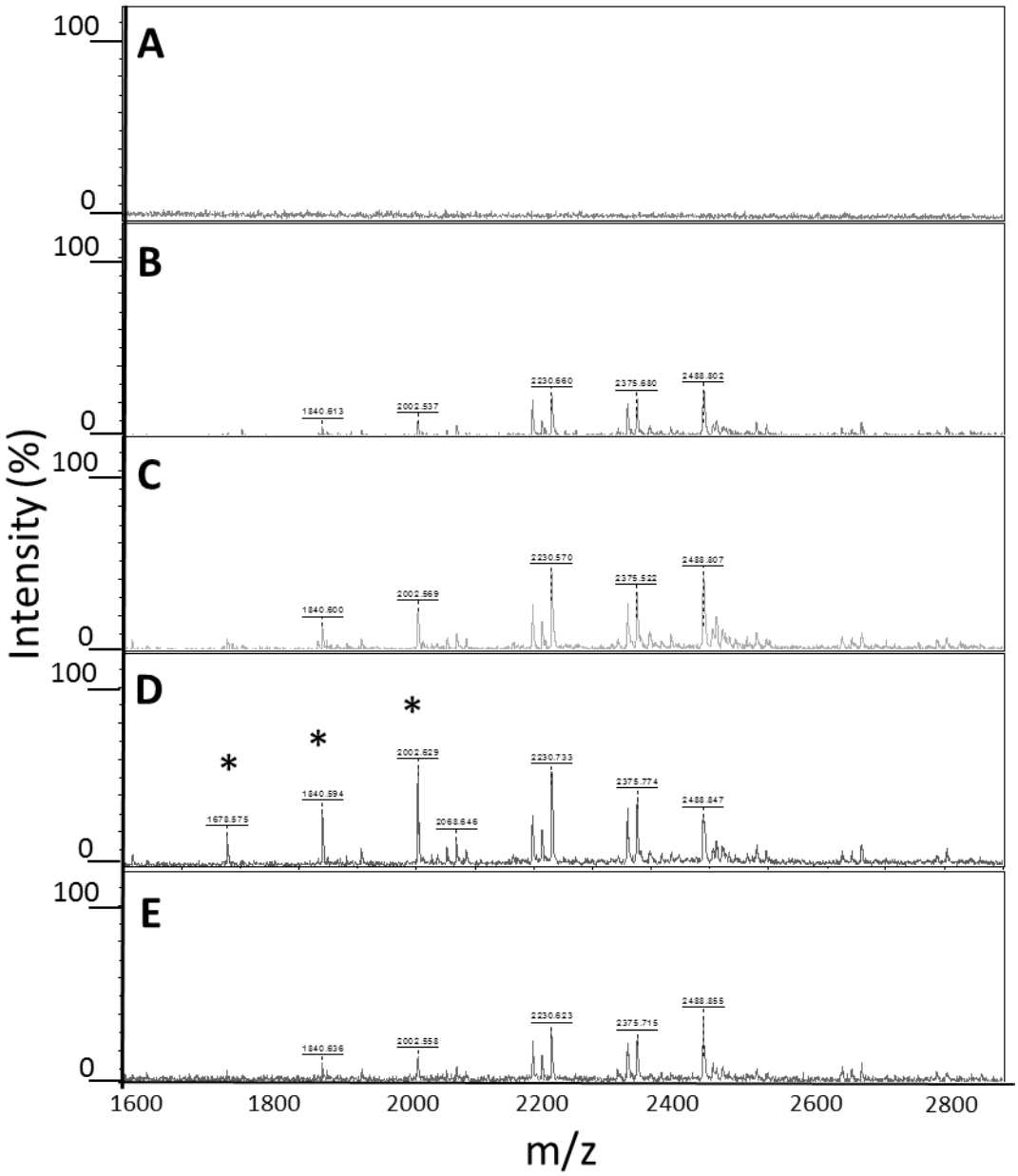


Figure 49 Binding of AB-glycans by H1-affinity–batch column using MALDI-TOF-MS
Mass spectra were taken of different fractions from a H1 affinity batch column experiment with AB-labelled swine glycans. (A) shows the elution fraction (lowering the pH with 1 mM HCl), where no glycans could be detected. (B-E) represent mass spectra of (B) flowtrough and (C-E) subsequent wash fractions where fraction D shows slightly higher abundance of oligomannosidic structures (*).

4.3.9 Shotgun with amicon tubes

To complete the experiments a shot gun approach using amicon tubes with a cut off of 10 kDa was also tested. H1 and H3 protein solution was incubated at RT for 30 min with AA-labelled glycans from swine tissues. Figure 50 shows the MADLI-TOF-MS spectra of the

collected fractions. Panel C and D show the elution fraction whereas A and B depict the flow through and a washing fraction, respectively. Again in the second elution fraction (panel D) the oligomannosidic structures appear in high intensities similar as in the flow through fraction (panel A). These results indicate that the oligomannosidic structures (* Man7 – Man9) seem to be retained by the hemagglutinins as the wash fraction (panel B) didn't contain the high amount of oligomannosidic glycans and they start to re-appear in the first elution fraction (panel C). Elution was again performed by lowering the pH by the addition of 1 mM HCl.

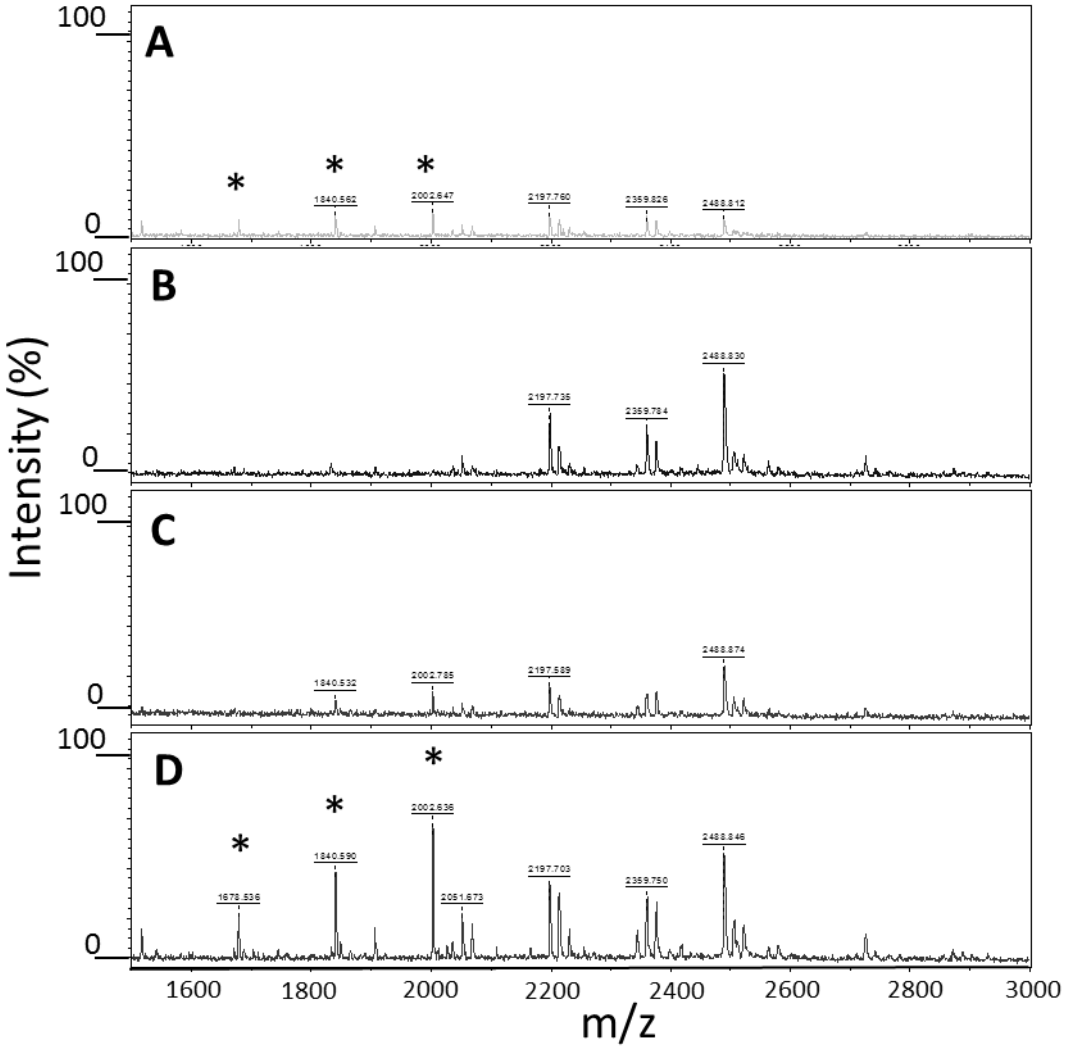


Figure 50 Binding of AB-labelled gycans by H1 protein using amicon centrifugation tubes and MALDI-TOF-MS for detection

Mass spectra of collected fractions of AB-labelled swine glycans incubated with H1 hemagglutinin protein, separated using amicon centrifugation tubes with a 10 kDa cut off. (A) and (B) show mass spectra of the flow through and wash fraction, respectively, whereas (B) and (D) depict the spectra of subsequent elution fractions with 1 mM HCl. Peaks labelled with (*) are oligomannosidic glcans appearing relatively high in the (A) flow through and (D) last elution fraction. Mass spectra were obtained using MALDO-TOF-MS operated in linear negative mode.

4.4 Fucose linkage and positional isomers for retention time studies (RRT)

4.4.1 Glycan standard sets

The purpose of this work is to elucidate specific retention times of isomeric glycan standard sets in correlation to the elution of a doubly galactosylated, α1,2 fucosylated, complex, biantennary N-glycan structure ($A^4(F^2)A^4$). As differently linked fucoses, found on glycoproteins in varying biological systems, are often correlated to high differences of the biological context it is our goal to facilitate the isomeric investigation of fucoses in complex biological systems.

Four different fucosyltransferases were used in this study. The FucT-IV catalyses the addition of an α1,3 linked fucose to the GlcNAc residue on the antennas (Gal(β1,4)[Fuc(α1,3)]GlcNAc) resulting in Lewis x type fucosylation. The second Lewis type fucosylation (Lewis type a) is produced by the FucT-III where the fucose is α1,4 linked to the GlcNAc residue on the antennas (Gal(β1,3)[Fuc(α1,4)]GlcNAc). The third fucosyltransferase (fucosyltransferase-2) attaches an α1,2 linked fucose to the terminal galactose residues of both antennas either Fuc(α1,2)Gal(β1,3)GlcNAc or Fuc(α1,2)Gal(β1,4)GlcNAc. Additionally core α1,6 fucosylated standards were used to complete the set of complex-type isomeric glycans.

4.4.1.1 Classical AA isomers

The first set of isomeric reference glycans is depicted in Figure 52. This glycan set consists only of complex glycan structures differing in the linkage of the terminal galactoses as well as in the differently attached fucoses.

The generation of the different fucosylated isomeric standards longed for the foremost generation of the differently linked galactoses. According to [154] the first and higher abundant peak of a singly galactosylated core fucosylated structure from hIgG glycans corresponds to A^4GnF^6 carrying the galactose on the α1,6 linked antennae whereas the second and lower abundant peak represents the isomer with the galactose attached to the α1,3 linked antennae. This knowledge was used to prepare the differently linked galactosylated isomers for subsequent fucosylation. A flat gradient (3.3.4) was used to separate and fractionate $A^4A^4F^6$, A^4GnF^6, GnA^4F^6 and $GnGnF^6$ from reduced hIgG glycans using HPLC together with a large scale PGC-column. After the fractions were checked for the

glycan content by PGCC-ESI-MS, as reduced unlabelled glycans neither can be visualized by UV nor fluorescence, $A^4A^3F^6$ / $A^3A^4F^6$ as well as $A^3A^3F^6$ were generated using the β1,3 galactosyltransferase (β3GalT5). The core fucosylated glycans were also treated with core fucosidase (α-L-Fucosidase from bovine kidney) to obtain the same glycan set without core fucosylation. Figure 51 shows the elution profiles of the differently linked AA standards in a single run (panel A) or each structure in a separate chromatographic run (A^4A^4, A^4A^3 A^3A^4 and A^3A^3; panel B-E) respectively.

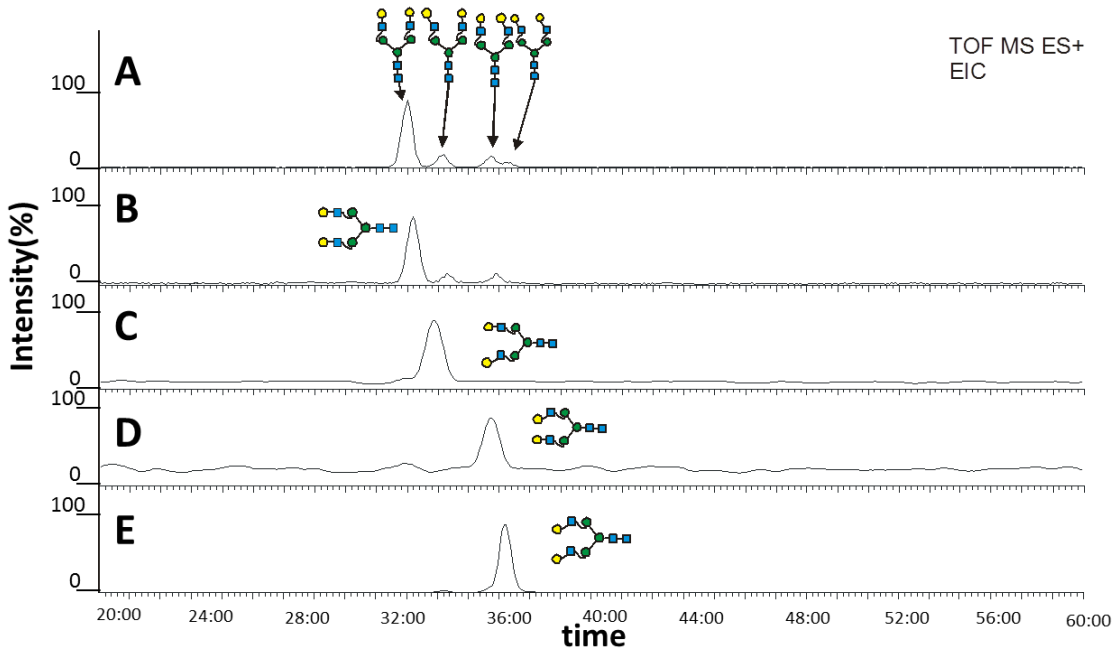


Figure 51 Elution profiles of AA isomers prepared from hIgG N-glycans using PGCC-ESI-QTOF-MS
Extracted ion chromatograms of m/z 822,5 [M+2H$^+$] of differently linked galactosylated glycan standards showing isomeric influence on retention time in a PGCC-ESI-QTOF-MS experiment. (A) shows a mixture of all 4 possible isomers whereas (B-E) show the single elution profile of each reference standard (A^4A^4, A^4A^3 A^3A^4 and A^3A^3)respectively.

Figure 52 shows the first glycan standard set that was prepared to investigate relative retention times (RRT) on a PGC chromatography. Panel A and B show the differently galactosylated isomers, with and without core α1,6 fucosylation. Those 8 reference glycans standards built the basis for subsequent fucosyltransferase treatments. Panel C shows the galactose isomers treated with an α1,2 fucosyltransferase (fucosyltransferase 2) resulting in 8 different singly fucosylated reference standards differing in the linkage of the terminal galactose either on the α1,6 linked antennae or the α1,3 linked antennae. Furthermore they differ in the antennae position of the α1,2 linked fucose. Panel D represents the 4 possibilities that result after the decoration of the galactose isomers with a Lewis x fucose which is an α1,3 linked fucose (FucT-IV) to the GlcNAc residues of the two antennas. The

same experiments were performed with the FucT-III enzyme resulting in Lewis a fucosylation which is an α1,4 linked fucose to the GlcNAc residues of both antennas.

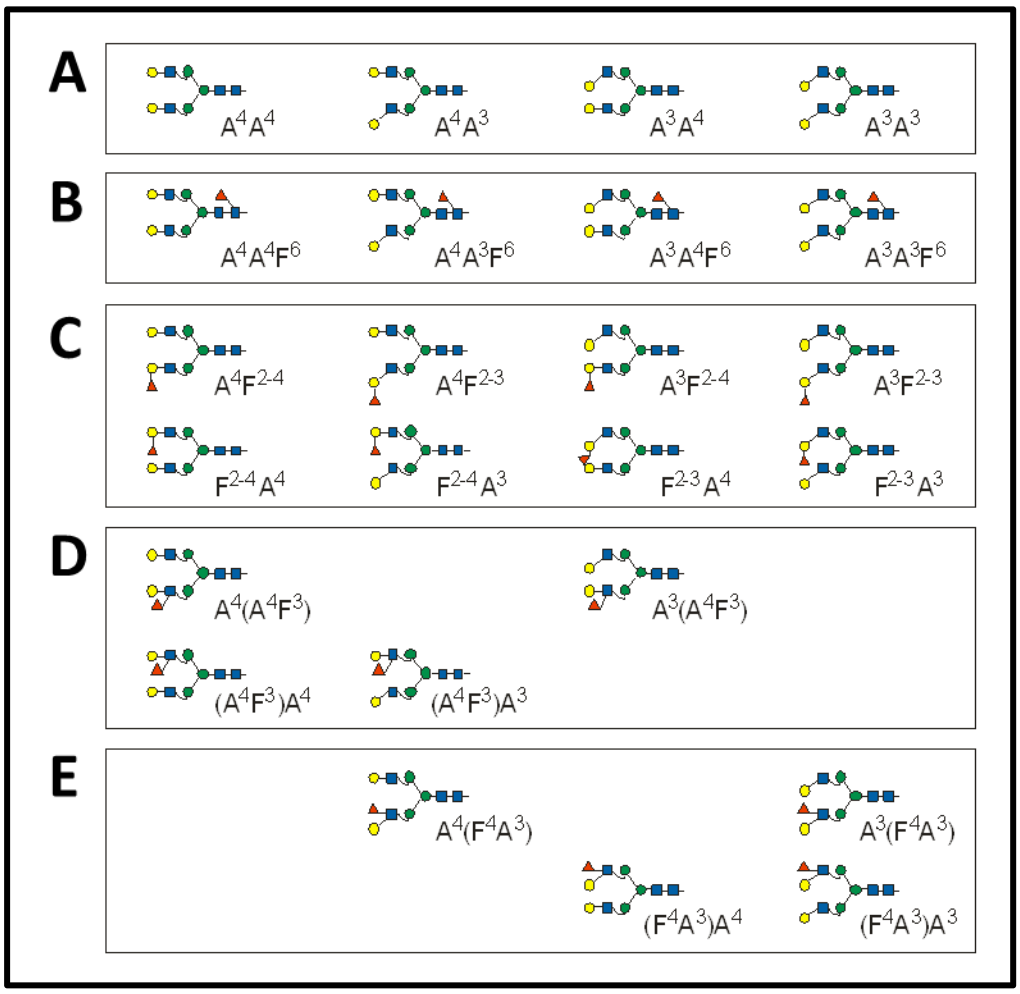


Figure 52 Classical complex AA isomeres

Producing this set of glycan reference isomers results in a high variety of different elution profiles of the distinct structures on a PGC column. Even on a fairly short gradient of 35 min the separation of structural isomers works well. By simply looking on relative retention times a lot of structural information is gained which is very helpful in elucidating structures of unknown biological samples. The number of possibilities of an unknown glycan structure can be narrowed down to a feasible amount of isomers simply by comparing relative retention times. Figure 53 shows the retention time shifts resulting from a glycan digest with the fucosyltransferase-2. The decoration of the terminal galactoses (β1,3) with an α1,2 linked fucose (panel B) leads to a RT shift to the back which is even higher when both antennas bear a terminal fucose (panel C).

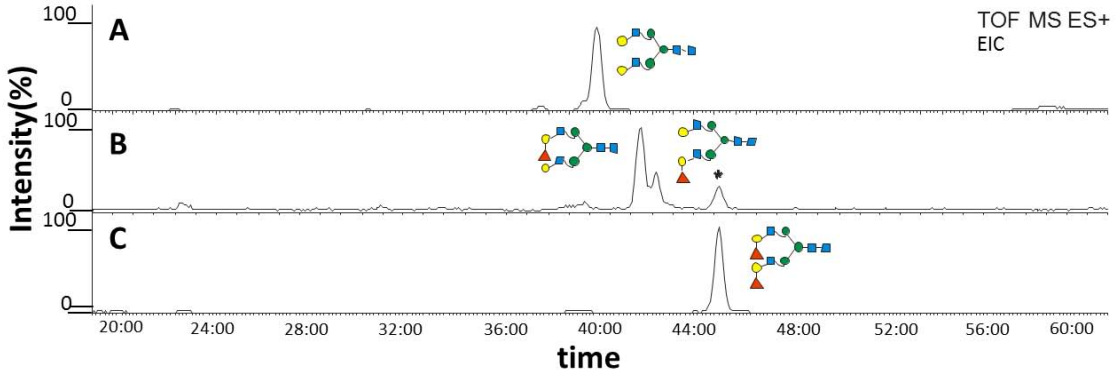


Figure 53 RT shift of FucT I on the example of A^3A^3
Extracted ion chromatograms [M+2H$^+$] of (A) m/z 822,5, (B) m/z 895,5 and (C) 968,5 m/z of A^3A^3 with the addition of one or two α-1,2 linked fucoses showing isomeric influence on retention time in a PGCC-ESI-QTOF-MS experiment. The peak assigned with a dot (B) indicates an in source fragment resulting from the doubly fucosylated glycan species.

The addition of a Lewis x fucose (α1,3 linked fucose) to the GlcNAc residue with a β1,4 linked galactose is achieved by the FucT-IV enzyme. In this experiment A^4A^4 was incubated with FucT-IV resulting in a mixture of unfucosylated, singly and double fucosylated glycan moieties. The Lewis x fucosylation (Figure 54) leads to an early elution from the PGC column (panel B) and is even less retained when two Lewis fucoses (panel A) are present on both antennas.

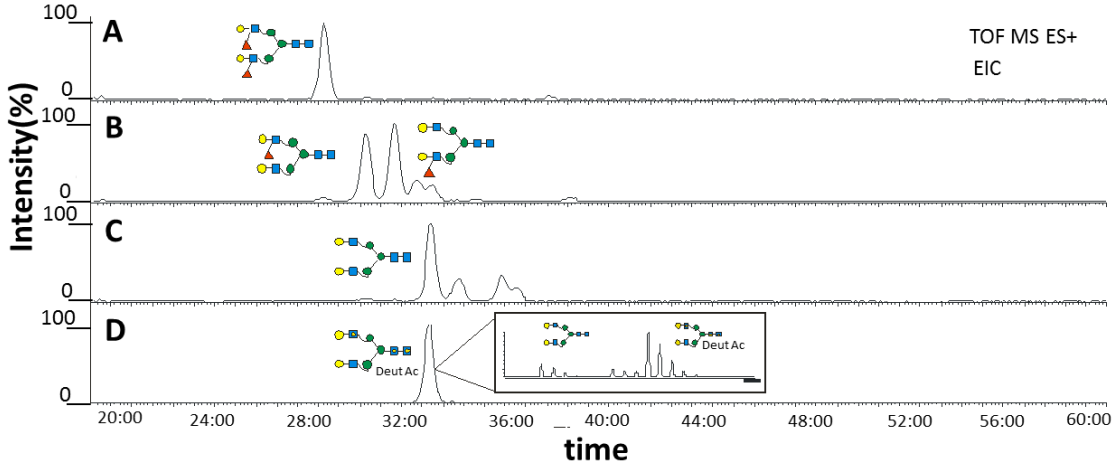


Figure 54 RT shift of FucT IV on the example A^4A^4 and the usage of deutero - acetylated internal standards
Extracted ion chromatograms of [M+2H$^+$] (A) doubly Lewis x fucosylated m/z 968,5, (B) singly Lewis x fucosylated m/z 895,5 and (C) unfucosylated m/z 822,5 A^4A^4 treated with FucT-IV. (D) shows the elution profile of the internally added isotopically labelled reference standard A^4A^4 (Deut Ac) m/z 825,5 exactly coeluting with the unlabelled A^4A^4 (C). The corresponding mass spectrum in (D) shows the existence of both structures as they differ in mass.

For latter relative retention time calculations to each reference standard run internal isotopically labelled standards were added. The lowest panel (D) shows the co-elution of the unfucosylated substrate A^4A^4 and the isotopically labelled internal standard A^4A^4 (Deut Ac).

Both glycan species exactly coelute but can be clearly distinguished by their mass (mass spectra panel D) as the isotopic label leads to a monoisotopic mass shift of + 3Da per isotopically substituted n-acetyl group of the four possible GlcNAcs.

The third possible fucosylation, an α1,6 core fucose, is shown in Figure 55. The 4 chromatograms represent the relative retention effect of the core fucosylation which clearly shifts the retention backwards when comparing the elution of A^4Gn and GnA^4 (panel D and C) to the elution A^4GnF^6 and GnA^4F^6 (panel B and A).

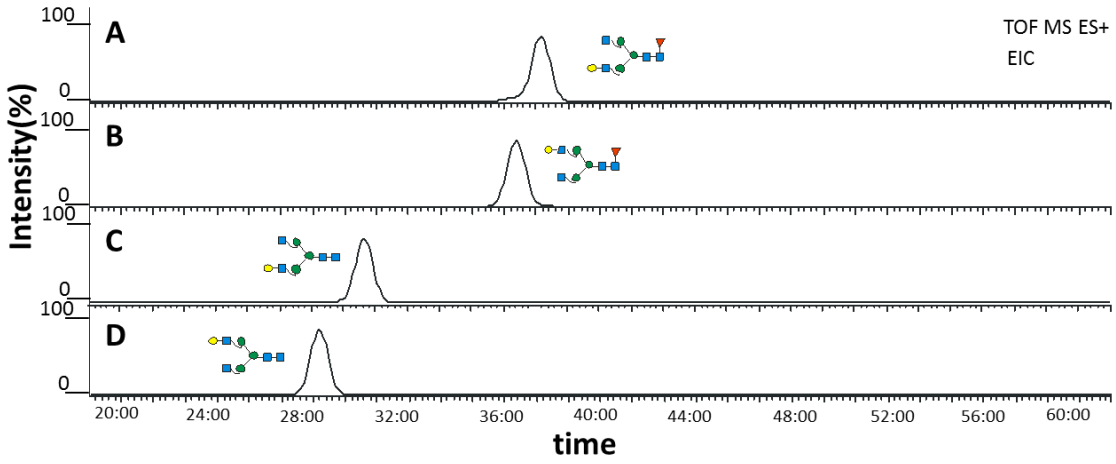


Figure 55 RT shifts of A^4Gn GnA^4 with or without core α1,6 fucosylation
Extracted ion chromatograms of [M+2H+] (A) GnA4F6, (B) A^4GnF^6 both m/z 814,5 in comparison to the unfucosylated isomers (C) GnA^4 and (D) A^4Gn both m/z 741,5. Isomeric separation was performed in a PGCC-ESI-QTOF-MS experiment operated in positive ionisation mode.

Similarly the addition of a β1,3 linked galactose to A^4Gn and GnA^4 results in a later retention of the glycan moieties (Figure 56). Furthermore the antennae, which is singly (panels (D) A^4Gn and (C) GnA^4) or doubly (panels (B) A^4A^3 and (A) A^3A^4) galactosylated has an impact on the glycan retention. Generalized it can be stated that the galactosylation of the α1,6 linked antennae leads to an earlier elution from the column whereas a glycan moiety where the decoration is found on the α1,3 linked antennae leads to a higher interference with the stationary phase of the PGC column.

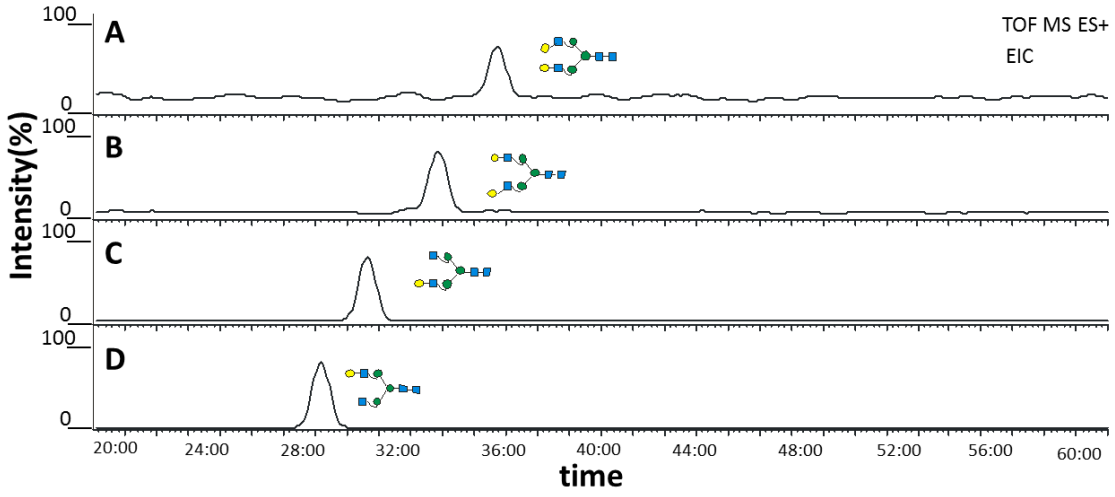


Figure 56 RT shift of $A^4Gn\ GnA^4$ treated with β1,3 galactosyltransferase (β3GalT5)
Extracted ion chromatograms of [M+2H+] (A) A^3A^4, (B) A^4A^3 both m/z 822,5 in comparison to the glycan isomers lacking the β1,3 linked galactose (C) GnA^4 and (D) A^4Gn both m/z 741,5. Isomeric separation was performed in a PGCC-ESI-QTOF-MS experiment operated in positive ionisation mode.

4.4.2 PGC elution profiles of singly, doubly and triply fucosylated reference standards on the example of hIgA

To show that the elution profiles can be used to elucidate glycan structures of complex biological samples N-glycans from the secretory chain of human IgA molecules were isolated and reduced. The s-chain of hIgA is known to carry Lewis x fucosylation on the N-glycome [169] that's why the protein was chosen to prove the principle.

Figure 57 depicts the elution profiles of the extracted ion chromatograms of singly (panel A), doubly (panel B) and triply fucosylated (panel C) glycan moieties. The red chromatograms indicate the EIC of the unknown s-chain IgA N-glycan pool in comparison to the black chromatograms which are the EIC elution profiles of some of the prepared reference standards. In panel A the red chromatogram from the s-chain IgA N-glycan pool represents the singly fucosylated possibly complex N-glycan structures. Three peaks can be detected, all of which the first two peaks are eluting in close proximity rather early and the third singly peak elutes rather late. By comparing the elution profiles to the reference standards strong evidence can be seen, that the first two peaks represent two β1,4 linked galactoses on each antennae decorated with an α1,3 linked Lewis x fucose, respectively. The third peak coelutes with $A^4A^4F^6$. When comparing the doubly fucosylated EIC (panel B – red chromatogram) of the s-chain glycans again three peaks can be detected. In contrast to the singly fucosylated chromatogram there is only one single peak eluting early from the column and two peaks in

close proximity which are retained longer on the column. As the first peak is eluting even earlier as the singly Lewis x fucosylated isomers it is very likely that this structure is doubly Lewis x fucosylated as this type of fucosylation shifts the retention time to an early time point. The same can be seen for the two narrow doubly fucosylated peaks eluting later from the PGC column as compared to the singly core fucosylated structure from panel A a retention time shift to the front can be observed. By comparing the elution profiles to the reference standards it is clearly visible that the first peak indicates a doubly Lewis x fucosylated isomer whereas the two later eluting glycans are both core fucosylated with an additional single Lewis x fucose on each antennas respectively.

The picture gets even clearer when comparing the triply fucosylated N-glycans (panel C). Instead of more peaks only a single peak is left supporting the theory of a combination of core and Lewis x fucoses in combination with two β1,4 linked galactoses. Those results are in accordance to the literature [169] and again a coelution with the reference standard (panel C - black chromatogram) is observed.

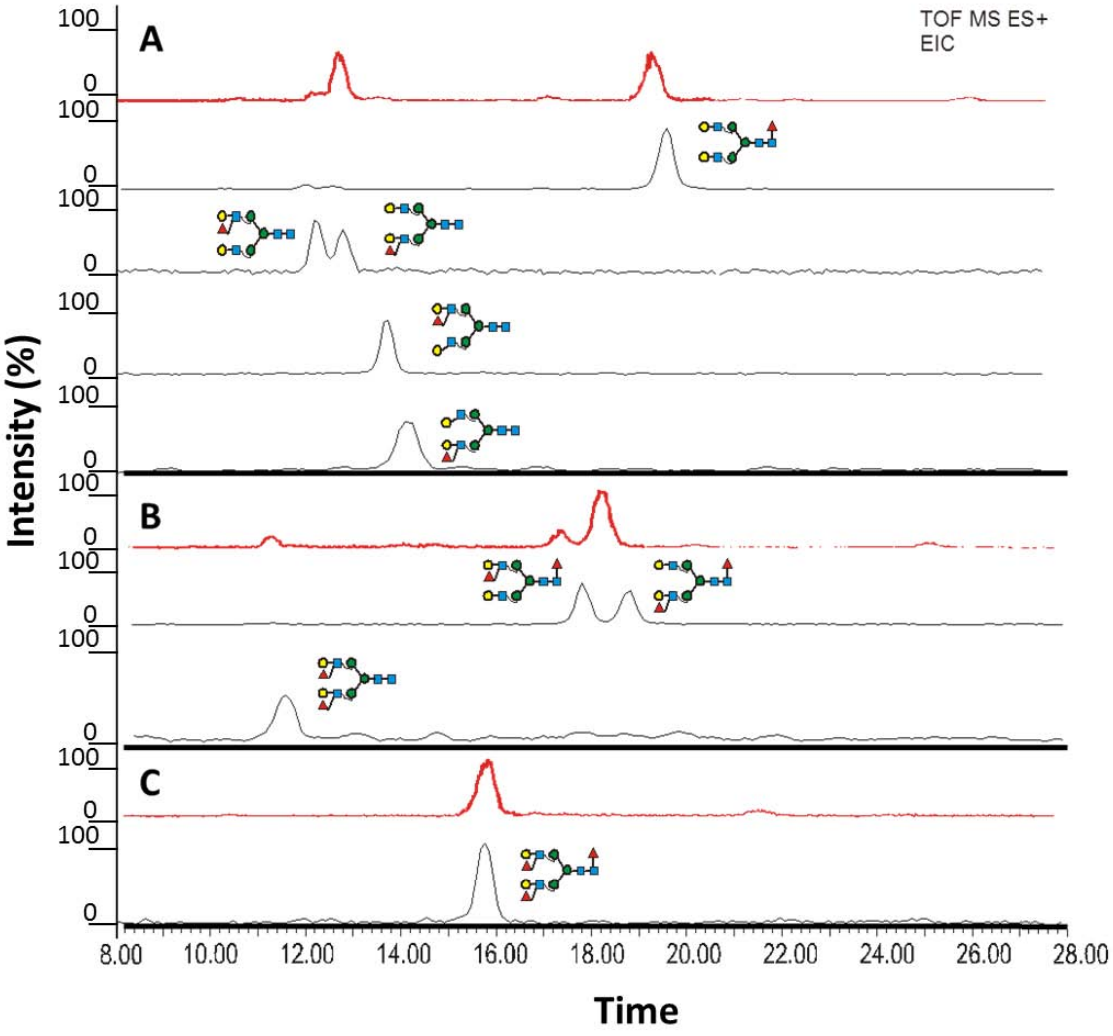


Figure 57 Elution profiles of singly, doubly and triply fucosylated standards on the example of hIgA for structural identification using PGCC-QTOF-MS
EIC of singly (A), doubly (B), and triply (C) fucosylated glycan isomers isolated from s-chain of hIgA are shown in red. RT profiles of the corresponding reference standards are shown in black. Analysis was performed using PGCC-ES-QTOF-MS in positive ionisation mode.

4.4.3 PGC elution order of fucosylated reference glycans and fucosylated samples (RRTs)

As the retention times of glycans eluting from a PGC column are not entirely stable and can vary slightly from run to run as well as after changing columns, it is necessary to add internal standards to each of the reference runs as well as to N-glycan runs of biological samples to be able to calculate relative retention times and thereby overcome the problem of slight retention time shifts or the instability of the PGC material.

The added internal standards were isotopically labelled glycan standards that can be distinguished from the naturally occurring glycan counterpart by a mass shift due to the isotopic labelling. Therefore deuterated Man8, deutero acetylated GnGnF^{6}, deutero-

acetylated A^4A^4 and deutero-acetylated $A^4(F^{2-3})A^4(F^{2-3})$ were added to each chromatographic run. By using the elution time of these glycan standards the used ACN concentration of the gradient was adjusted regularly by using the regression line of a reference run to facilitate a stable and comparable elution system which is mandatory for the subsequent RRT calculations.

To calculate the Relative Retention Times (RRT) the RT of $GnGnF^6$ and $A^4F^2A^4F^2$ was used to calculate the deviation from a reference run. All elution times were correlated to the retention time of the isotopically labelled A^4A^4.

Table 10 shows the calculated relative retention times (RRTs) of all the reference standards as well as the RRTs of the corresponding masses found in different biological samples (for the equations of the calculations see 3.3.5.2).

The RRTs of the glycans isolated from s-chain IgA nicely correlate to the reference standard A^4A^4 treated with α1,3-FucT. All possibilities occur, starting from the single fucosylation on both A^4A^4 antennas respectively (Lewis x) as well as the doubly fucosylated combinations of α1,6 core fucosylation with an additional Lewis x fucosylation or two Lewis x fucoses on both antennas and all three possible fucosylation (Lewis x and core fucoses) together.

References

References			
α1.2-FucT	1F	2F	3F
A4A4	1,34	1,99	
	1,42		
A4A3	1,58	1,74	
A3A4	1,46	1,84	
A3A3	1,49	1,85	
A4A4F6	1,58	1,96	2,39
α1.3-FucT	1F	2F	3F
A4A4	0,82	0,72	
	0,88		
A4A3	0,96		
A3A4	1,01		
A3A3			
A4A4F6	1,57	1,31	1,15
		1,40	
α1.4-FucT	1F	2F	3F
A4A4			
A4A3	1,13		
A3A4	1,34		
A3A3	1,25		
	1,46		

samples

S-chain IgA	1F	2F	3F
	1,57	1,33	1,18
	0,85	1,41	
	0,88	0,74	
mouse brain	1F	2F	3F
	0,48	0,57	1,15
	0,62		
H-chain IgA	1F	2F	3F
	1,58		

Table 10 RRT of fucosylated reference glycans and different fucosylated samples

Column (1F) shows the RRTs for singly fucosylated glycan structures, (2F) doubly fucosylated and (3F) triply fucosylated glycan structures. The reference standards (References) are generated using the glycan isomers A^4A^4, A^4A^3, A^3A^4, A^3A^3 and $A^4A^4F^6$ treated with fucosidase-2 (α1,2 FucT), FucT-IV (α1,3 FucT) and FucT-III (α1,4 FucT). (samples) shows the RRTs of glycan isomers of biological samples. RRTs are calculated relative to $A^4F^2A^4$ from chromatographically separations using PGCC-ESI-QTOF-MS.

4.4.4 Deutero acetylated standards

To simplify the calculation of the RRT's internal standards were added to each of the reference runs. These internal standards were stably isotopically labelled with deuterated acetyl groups and could be used to identify structural isomers due to the retention time shifts. If naturally occurring structures in the biological samples are identical and not structural isomers, they are exactly coeluting and no retention time shift can be observed. But never the less those glycans still can be identified as they can be distinguished due to the isotopic labelling, which results in a monisotopical mass shift of +3 Da per deuterated acetyl group. Figure 58 indicates the elution profile of the isotopically labelled $A^4F^2A^4F^2$ (Deut Ac) reference glycan (panel A) and the elution profile of three glycan isomers $A^4F^2A^3F^2$, $A^3F^2A^4F^2$ and $A^4F^2A^4F^2$ (panel B). The isotopically labelled reference glycan elutes as a single peak (A') and the corresponding mass spectrum (panel A' = $[M+2H^+]^{2+}$) shows a mixture of doubly [m/z 971,4] and triply [m/z 972,9] deuterated acetyl groups of the 4 possible substituted N-acetyl glucosamines. Each substitution results in a mass addition of 1,5 Da in the doubly charged mass spectrum $[M+2H^+]^{2+}$. Panel B shows the elution profile of the isomeric mixture of $A^4F^2A^3F^2$, $A^3F^2A^4F^2$ and $A^4F^2A^4F^2$ [m/z 968,4] with the addition of the internal standard $A^4F^2A^4F^2$ (Deut Ac). The three peaks can be identified as glycan isomers whereas the corresponding mass spectra of the peak B' clearly shows the coelution of the unlabelled isomeric glycan structure [m/z 968,4] and the structurally identical isotopically labelled reference glycan [m/z 971,4 & m/z 972,9].

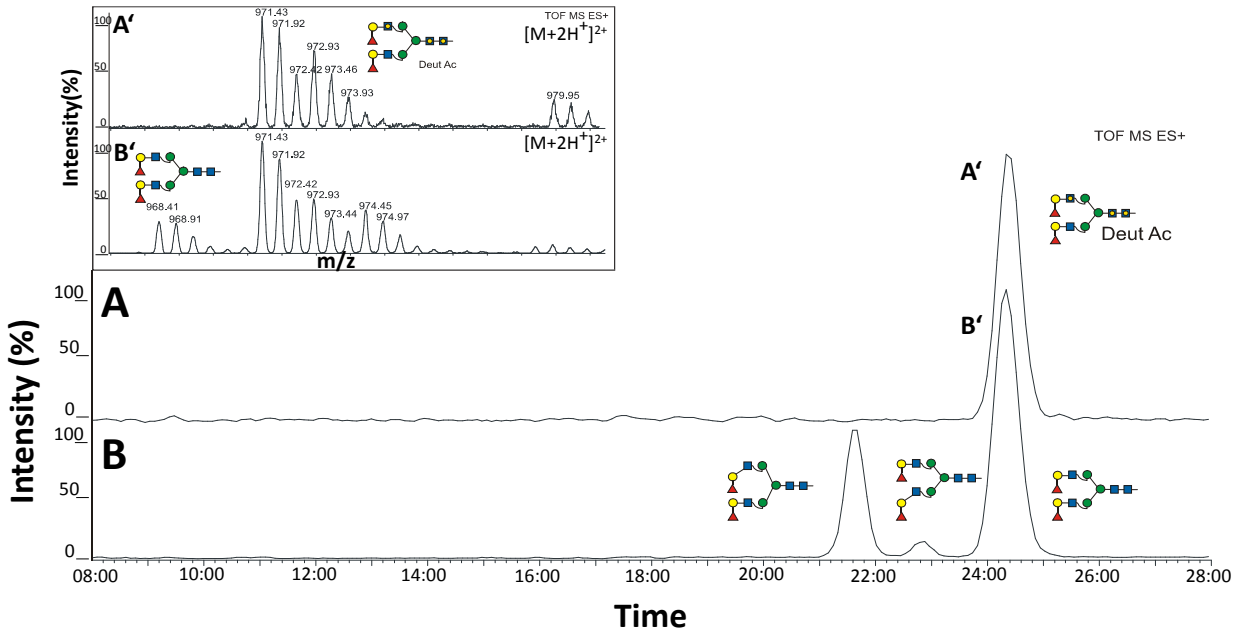


Figure 58 Addition of isotopically labelled internal standards for structural glycan identification using PGCC-QTOF-MS
EIC of $A^4F^2A^4F^2$ (Deut Ac) (A) and of the glycan isomers $A^4F^2A^3F^2$, $A^3F^2A^4F^2$ and $A^4F^2A^4F^2$ (B) show exact coelution of A' and B'. The corresponding mass spectra show the mass shift of the isotopically labelled $A^4F^2A^4F^2$ (Deut Ac) [m/z 971,4 & m/z 972,9] (A') in contrast to the unlabelled isomer [m/z 968,4]where the labelled standard was added (B'). Analysis was performed in positive ionisation mode using PGCC-ESI-QTOF-MS.

The benefit of using isotopically labelled reference standards lies on the one hand in the identification of exactly identical glycan structures due to the coelution of both glycan species, on the other side the resulting mass shift and unique isotopic pattern, due to the isotopic label, allows for the addition of those standards to any biological glycan sample without facing the problem of possible mass identical glycan structures. Additionally the elution times of those reference standards can be used for relative retention time calculations and to monitor the separation stability of the PGC material.

Taken together the PGC-column is a quite robust system but the RT's can be easily influenced by solvent systems, sample impurities, and reverse polarity voltage from the ESI-source. Therefore it is difficult to obtain 100 % stable retention times which can be seen when comparing the RRTs of the S-Chain IgA glycans and the reference glycans. They don't match a 100 % e.g. RRT of the singly Lewis x fucosylated reference standard is 0,82 and 0,88, respectively, whereas the RRTs of the biological sample are 0,85 and 0,88.

To overcome the problem of slight RT-shifts as well as possible coeluting glycan isomers additional fragmentation analysis of the differently eluting peaks can be performed to prove the structures.

4.4.5 Negative ionisation mode versus positive ionisation mode for MS/MS fragmentation experiments

In Figure 59 MS/MS fragmentation of experiments of $A^3A^4F^2$ in comparison to $A^4F^3A^3$ in positive as well as negative ion mode can be seen. Panel A and C show the fragmentation pattern in negative ion mode whereas panel B and D show the positive ion mode fragmentation.

It is clearly visible by comparing the positive ion mode experiments that the structures hardly can be distinguished as both structures give the same b and y ion series typically obtained by this fragmentation mode.

In contrast to that panel A and C can be easily distinguished as in negative ion mode so called diagnostic ions can be detected, which facilitated the structural identification and indicate the occurrence of different types of Lewis fucosylation. The diagnostic ions are mostly fragments of the “a and z” ion-series, which are cross ring cleavages, and so called D ions.

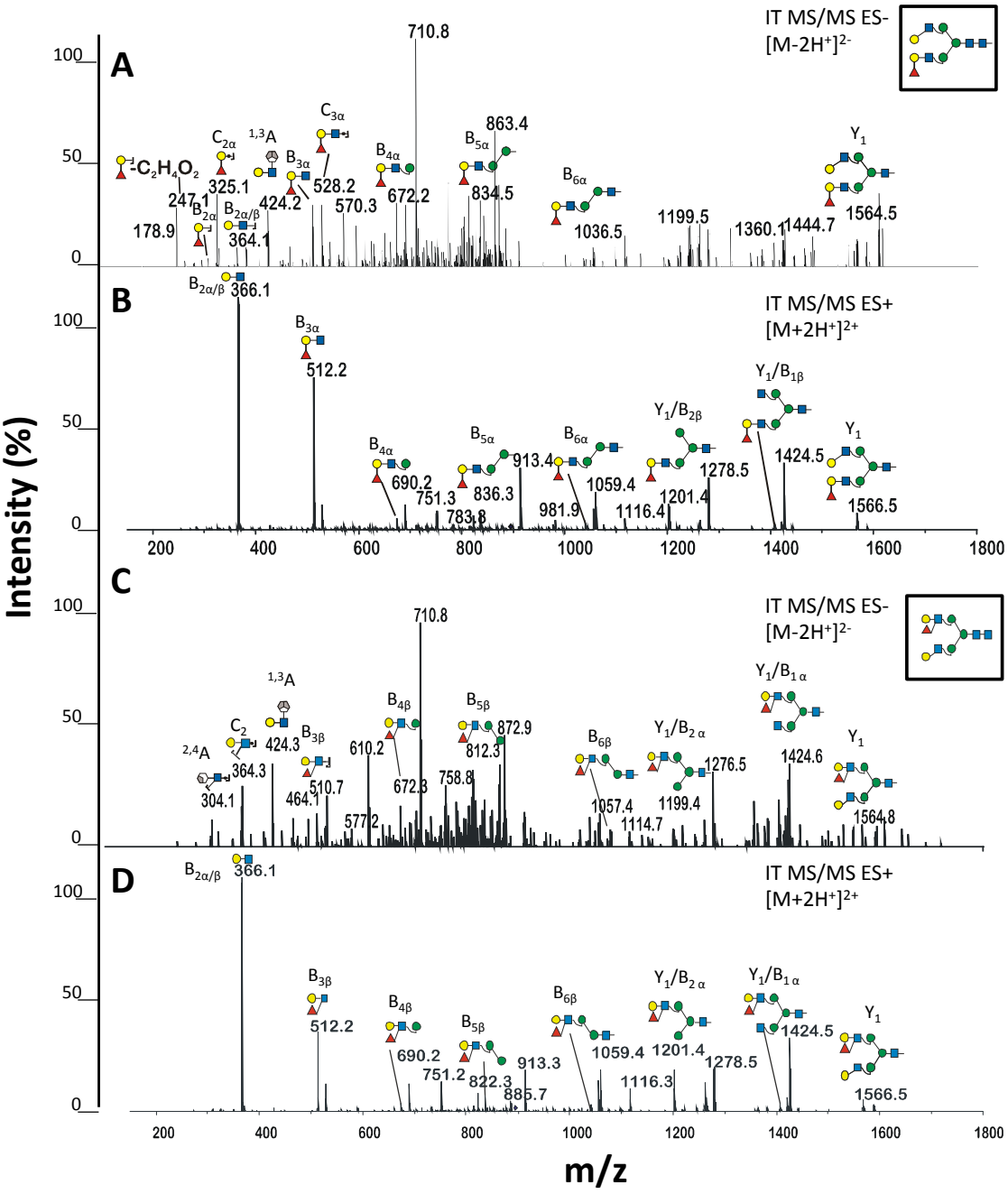


Figure 59 Negative ionisation mode versus positive ionisation mode for MS/MS fragmentation experiments using PGCC-ESI-IT-MS/MS

Fragmentation of two isomeric glycans ($A^3A^4F^2$ & $A^4F^3A^3$) was performed with PGCC-ESI-IT-MS/MS in negative ionisation mode (A and C) or in positive ionisation mode (B and D). Peaks are labelled with fragment symbols and names.

4.4.6 Positive and negative mode MS and MS/MS experiments of s-chain N-glycans from hIgA

Figure 60 shows negative mode MS/MS fragmentation of the different singly and doubly fucosylated isoforms found in the s-chain of hIgA. In panel A the elution profile of the singly fucosylated glycan isomers is depicted and the corresponding mass spectra of peaks labelled

with A 1 and A 2 can be seen underneath, respectively. The mass spectra were obtained in negative ionisation mode and the fragmentation pattern of A 1 clearly shows the occurrence of the diagnostic ions indicative for Lewis x structures. The cross ring cleavages 264m/z (^{1,3}A) and 304 m/z (^{2,4}A) as well as C-ions 364 m/z (C_2) are typical results of negative mode fragmentation. Furthermore typical B ions occur, which are found in negative as well as positive ionisation mode resulting from cleavage from the terminal glycan side. Peak A 2 is the latest eluting peak among the singly fucosylated isomers which was assigned as α1,6 core fucosylated in the elution profiles and RRT calculations. The fragmentation pattern confirms the occurrence of core fucosylation by the diagnostic ions of 368 m/z (Y_1) and 350 m/z (Z_1) additionally cross ring fragments like ^{1,3}A ions and Z ions were found. Panel B 1 is the corresponding mass spectra of the peak B 1 labelled in panel B, the doubly fucosylated glycan species. Diagnostic ions for both fucosylation species were found, confirming the presence of both types of fucosylation. The characteristic ions are 304 m/z (^{2,4}A) and 364 m/z (C_2) for the presence of a Lewis x fucose as well as 350 m/z (Z_1) and 368 m/z (Y_1) for core fucosylation. Furthermore several structural fragments of the B as well as Z ions series were found.

Taken together, the additional use of negative ion mode fragmentation, using an ammonium carbonate buffer at high pH, leads to highly informative MS/MS fragmentation patterns, which give additional proof to assign a distinct structure. However due to the used ionisation charge the fragmentation spectra get a lot more complex as beside the major B and Y ion series, which are usually obtained in positive ionisation mode, A-,X-, C-and Z- ions as well as D-Ions are occurring. The resulting diagnostic ions though can be very helpful but the complexity of data calls for bioinformatics tools to analyse, when considering this analytical approach for the high throughput screening of biological samples.

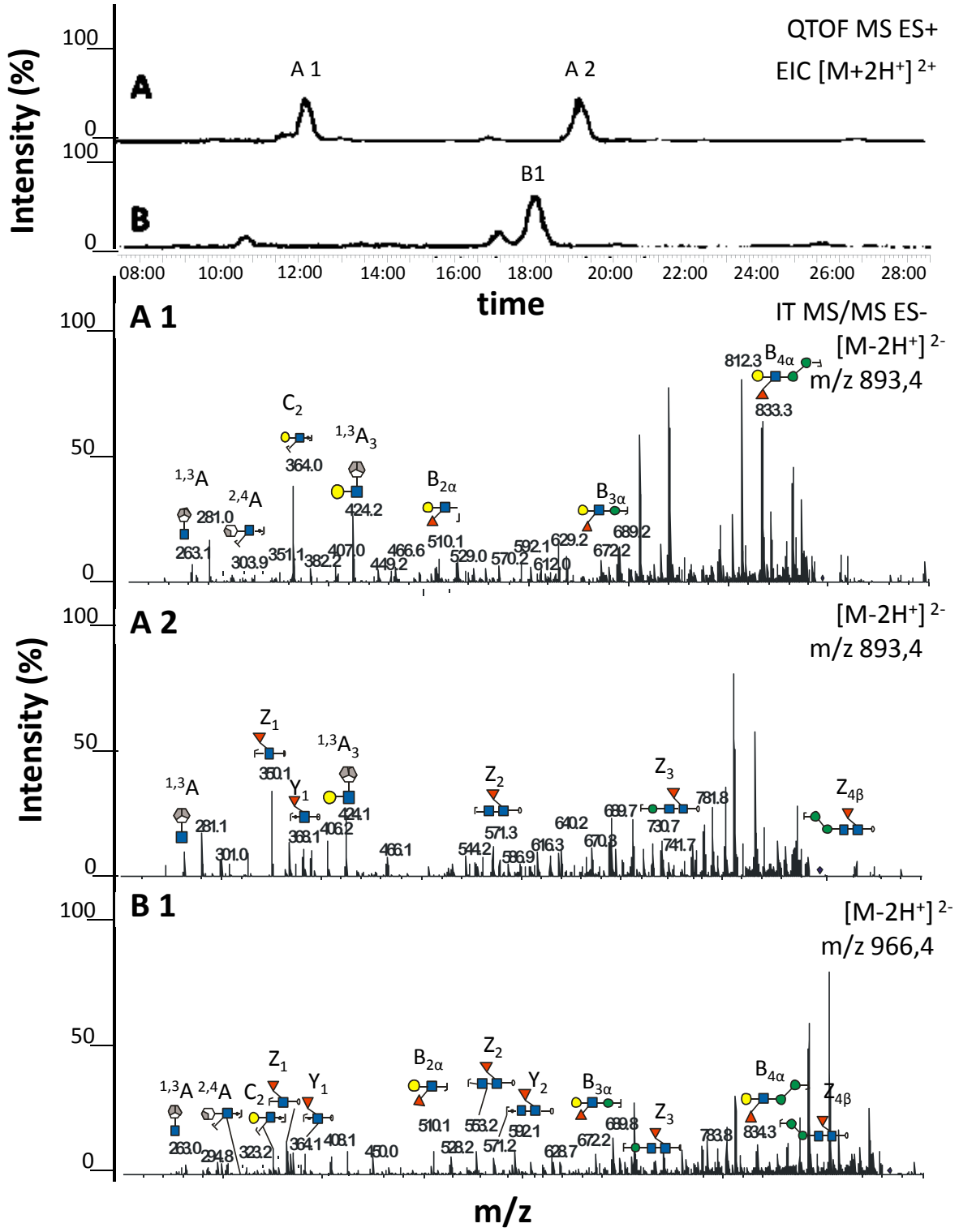


Figure 60 Positive and negative mode MS and MS/MS experiments of isolated glycans from s-chain hIgA
RRT experiments of N-glycan isomers isolated from the s-chain of hIgA (A,B) were performed with PGCC-ESI-QTOF-MS in positive ionisation mode. Fragmentation of isomeric glycans A1-B1 was performed with PGCC-ESI-IT-MS/MS in negative ionisation mode. Peaks are labelled with fragment symbols and names.

4.4.7 Elution order of fucosylated mouse brain glycans on PGC

When comparing the elution profiles of the singly and doubly fucosylated elution profiles from N-glycans isolated from mouse brain, the chromatographic peaks show no overlay with the elution profiles of the reference glycans as indicated in Figure 61. Furthermore comparing the RRT (Table 10) of the singly and doubly fucosylated glycan structures found in the N-glycans isolated from mouse brain to the RRT's of the reference standards no coeluting structures could be found. The RRTs of the mouse brain structures are comparably small (0,48, 0,62 and 0,57) compared to the RRTs of the prepared reference standards which range between 0,72 – 1,99. This led to the conclusion that the fucosylated glycan isomers found in mouse brain samples are not complex glycan species but probably hybrid and or bisecting glycans.

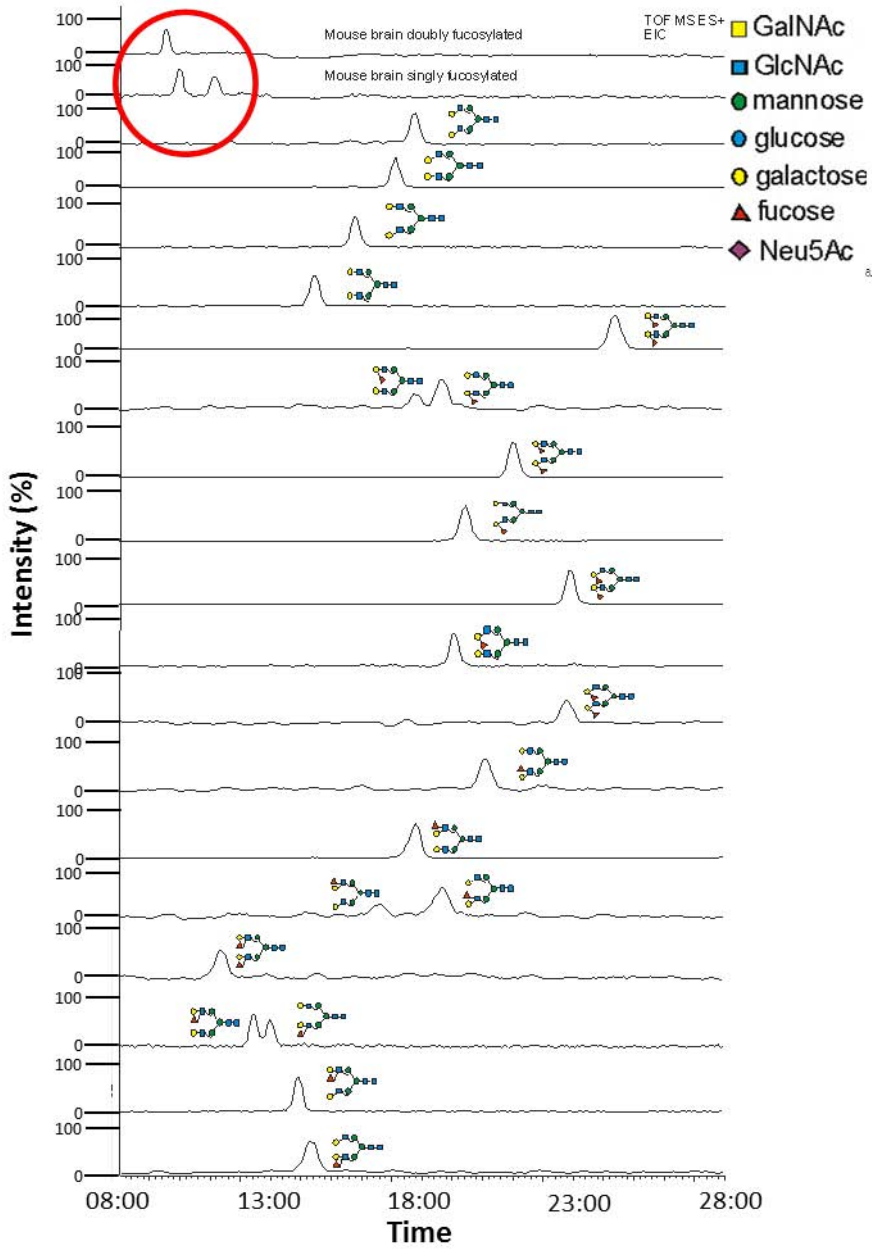


Figure 61 Elution profile of single and double fucosylated mouse brain glycan isomers compared to the elution profiles of all complex reference standards using PGCC-ESI-QTOF-MS
EICs of singly and doubly fucosylated mouse brain glycan isomers (highlighted in red) in comparison to the elution profiles of the complex reference standard set from Figure 52. No coelution can be detected.

4.4.7.1 Hybrid and bisecting isomers

As it is known from the literature mouse brain contains a lot of hybrid and bisecting glycan structures with different types of fucosylation [170]. According to that knowledge a new set of reference glycans was composed consisting of structural isomers with the same mass as the glycan structures depicted in Figure 52. Panel A shows different isomers carrying a bisecting GlcNAc and differing in the attachment of the galactose (either β1,3 or β1,4) as

well as the addition of a fucose to the terminal antennae (either α1,2, α1,3 or α1,4). All glycans depicted in panel A are hybrid and bisecting glycans isomers.

In panel B possible glycan isomers carrying a core α1,6 linked fucose are shown. The composition is again mainly hybrid and bisecting glycans though also an β1,4 linked GalNAc could be a possible terminal modification of the complex antennae instead of the addition of a bisecting GlcNAc. This is also shown in panel C where only hybrid structures with additional β1,4 or β1,3 linked GalNAc residues in combination with α1,2, α1,3 or α1,4 fucosylation is depicted.

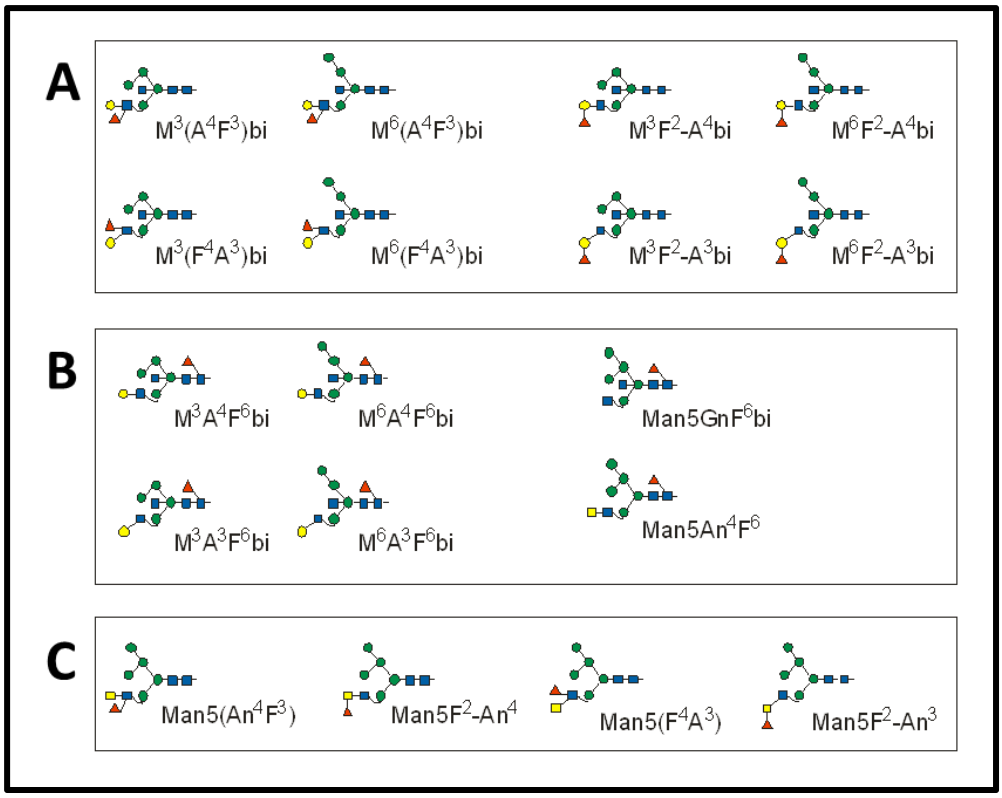


Figure 62 Hybrid and bisecting glycan isomers

4.4.7.2 α-Galactose and LacNAc isomers

To complete the set of possible glycan isomers also α1,3 linked galactoses could be present on already attached β1,3 or β1,4 linked terminal galactoses. This glycan feature is known to occur on monoclonal antibodies produced in mouse cells and is further known to be immunogenic for humans [171].

Figure 63 panel A shows different α1,6 core-fucosylated glycans in combination with alpha galactose epitopes either on β1,3 or β1,4 linked galactose differing further in the decoration of the antennae, in case of the complex gycan species, or the presence of a bisecting GlcNAc

on a hybrid glycan species. Panel B shows different Lewis x fucosylated isomers and in panel C different types of fucosylation in combination with the occurrence of GalNAc residues or LacNAc repeats are shown. These standards were not generated in this work due to time issues.

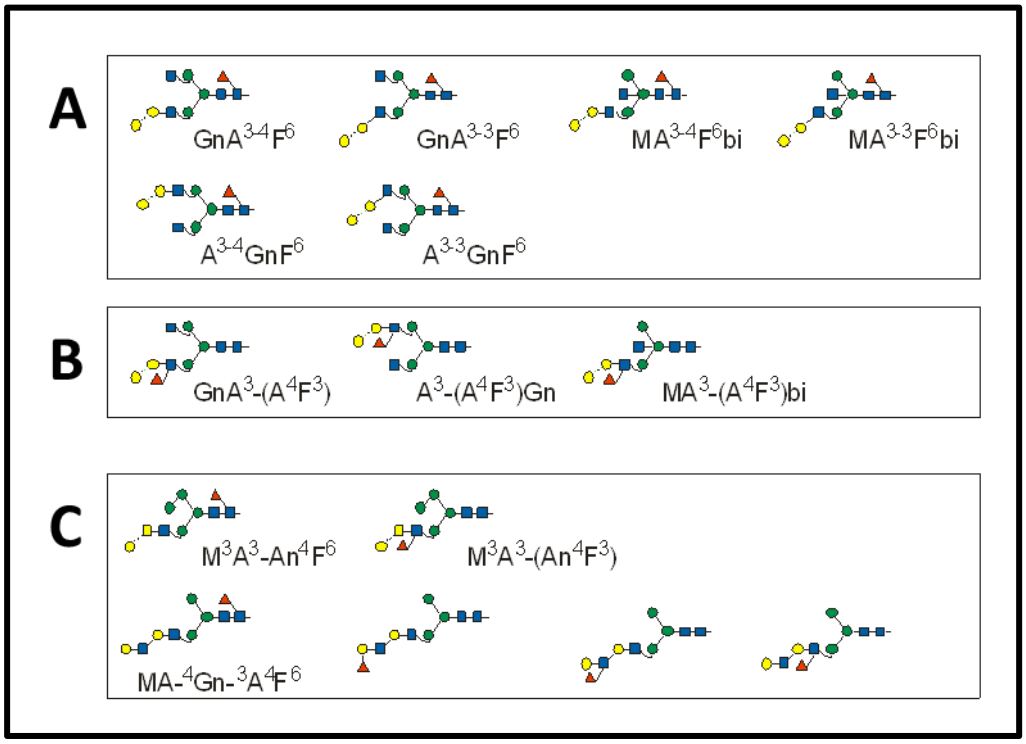


Figure 63 α-Galactose and LacNAc glycan isomers

4.4.8 Preparation of hybrid standards - workflow

The workflow of the preparation of the different reference standards is depicted in Figure 64 on the example of Man4A^{4}F^{6}. Several enzymatic steps are needed to produce one single reference structure in combination with purification steps and monitoring of product conversions after each single step. The preparation starts with a Man5 glycopeptide isolated from the protein Taka-amylase A (*Aspergillus oryzae*) which carried mainly the oligomannosidic structures Man5 and Man6. This glycopeptid was treated with jack bean mannosidase removing a mannose and additionally with the GnT-I which transfers a GlcNAc residue to the "3-arm" of the trimannosyl core. The prior action of GnT-I is mandatory for the subsequent activity of the FucT-VIII, which transfers a fucose residue in α1,6 linkage to the innermost core-GlcNAc. Beside the prior activity of GnT-I, the FucT-VIII also needs a remaining peptide backbone or a reducing-end glycan for successful fucosylation. The

enzyme is found in microsomes isolated from chicken heart. After each of these enzymatic steps a C-18 SPE purification was performed and product conversion was monitored by MS. After the successful generation of the Man4GnF^{6}-glycopeptide the peptide backbone was removed by enzymatic cleavage with PNGase F and subsequent reduction of the glycan was performed. Subsequent incubation with GnT-III was performed, which transfers the GlcNAc to the β-mannose of the core structure resulting in a bisecting N-glycan (Man4GnF^{6}bi). This glycan structure was further modified with a terminating β1,4 linked galactoses. Each of the resulting reference glycans was analysed using PGCC-ESI-QTOF-MS together with the addition of internal isotopically labelled standards for subsequent calculations of the relative retention times. Furthermore the standards were analysed using PGCC-ESI-IT-MS operated in negative ionisation mode to perform MS/MS fragmentation experiments. The used buffer for the PGC separation was NH_4formeat pH 3 as solvent A and 100 % acetonitrile with 0,1% TFA as solvent B. Due to the used buffer the resulting ions were mostly ammonia adducts which were only poorly accessible to the CID fragmentation. This led to a change of the solvent system for the fragmentation analysis to 10 mM NH_4HCO_3, pH 8 which gave far better fragmentation pattern.

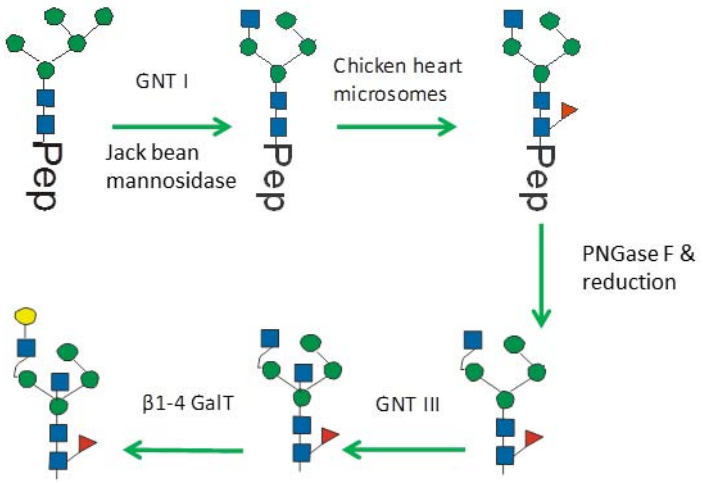


Figure 64 Schematic workflow of enzymatic reations for the generation of hybrid and bisecting reference glycans

The generation of this glycan isomers beard unexpected challenges due to highly inefficient substrate turnover and a high occurrence of side activities of the used glycosyltransferases. As the singly and doubly fucosylated N-glycan isomers from the mouse brain eluted very early from the column it is likely that the glycans are bisecting and hybrid. As shown in Figure 65 and Figure 66 the addition of a bisecting GlcNAc with the GnT-III as well as the treatment

of a Man5Gn with α mannosidase leads to retention time shifts forward which can be a possible explanation for the early elution of the isomers in the mouse brain. The poor turnover of Man5Gn with the GnT-III can be explained due to the very low expression levels of the enzyme. Furthermore the crude cell supernatant was used for the enzymatic assays without any preceding purifications e.g. via a HIS-Tag affinity column as the genetic construct didn't contain any amino acid sequences which could be used as purification-Tag.

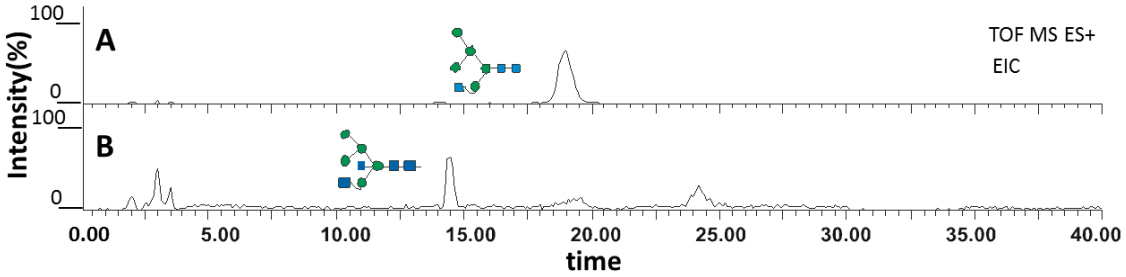


Figure 65 RT shift of Man5Gn treated with GnT-III
Extracted ion chromatograms of [M+2H+] (A) Man5Gn m/z 721,0 treated with GnT-III enzyme producing a bisecting glycan structure (B) Man5Gn(bi) m/z 822,5.Analysis was performed in a PGCC-ESI-QTOF-MS experiment operated in positive ionisation mode.

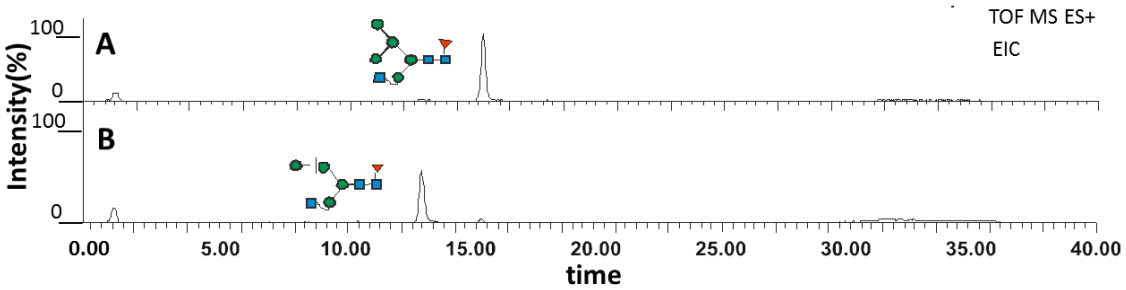


Figure 66 RT shift of Man5GnF^6 treated with Jackbean mannosidase
Extracted ion chromatograms of [M+2H+] (A) Man5GnF^6 m/z 794,0 treated with jackbean mannosidase enzyme producing (B) Man4GnF^6 m/z 713,0 .Analysis was performed in a PGCC-ESI-QTOF-MS experiment operated in positive ionisation mode

Figure 67 shows the substrate conversion of the GnT-III enzyme, which led to misleading product conversions at first. Panel A shows the partly conversion of GnGn (1417,9 Da) to the bisected counterpart GnGn(bi) (1621 Da). The addition of the GlcNAc is working, though the substrate conversion is only very low even after 24 h of incubation. Panel B shows the conversion of Man5Gn (1538.5 Da) to Man5Gn(bi) (1741,5 Da) where almost nothing is turned over. In contrast to that a relatively high conversion of Man5 to Man5Gn could be shown. This substrate conversion led to the conclusion of relatively high side activities of insect cell derived GnT-I due to the impurity of the enzyme.

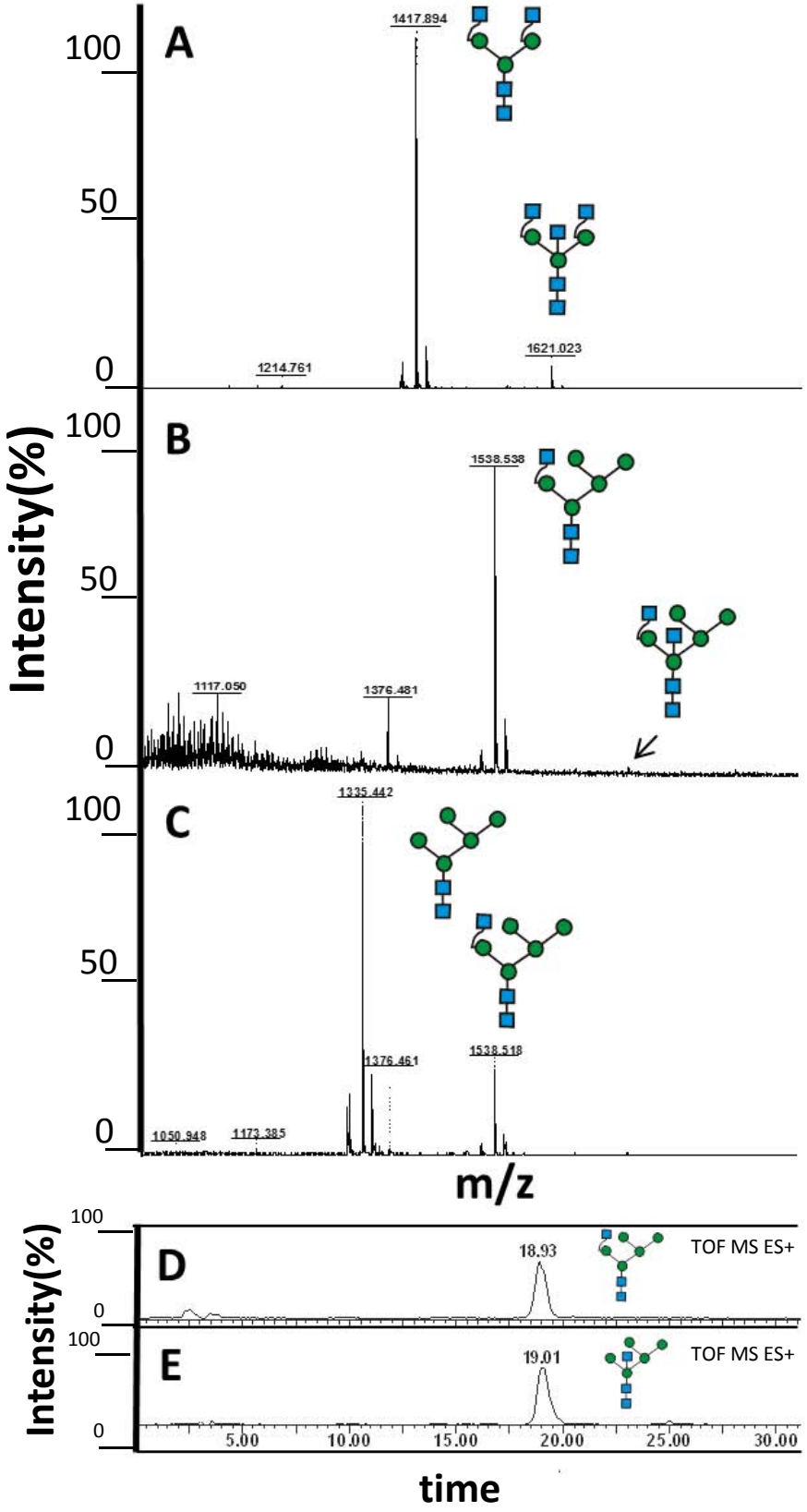


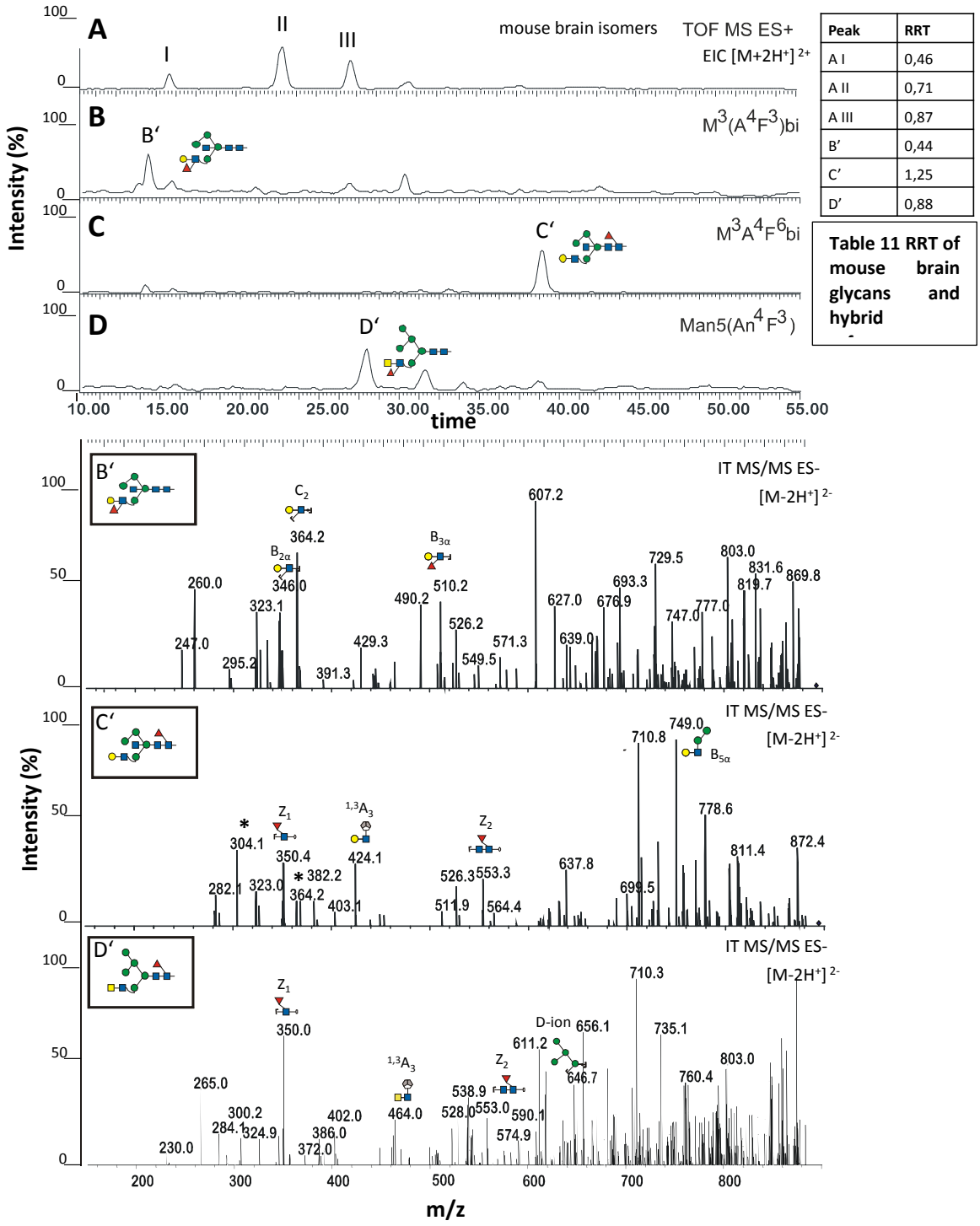
Figure 67 GnT-III substrate conversion using MALDI-TOF-MS and product verification using PGCC-ESI-QTOF-MS
(A-C) MALDI-TOF-MS mass spectra monitoring product conversion of GnT-III on the example of (A) GnGn as a substrate and GnGn(bi) as a product, (B) Man5Gn as a substrate and Man5Gn(bi) as a product and as a negative control for GnT-III activity (C) Man5 as a substrate and Man5Gn as a product of GnT-I side activity. (D) EIC of Man5 treated with GnT-I as a positive control resulting in Man5Gn (RT: 18,93) and (E) Man5 treated with GnT-III resulting in a peak at (RT: 19,03). As the peaks (D and E) elute at the same time (E) is a product of GnT-I side activity showing no GnT-III activity.

The presence of the side activities was proven by comparing the elution profiles of Man5 treated with either a recombinant GnT-I (Figure 67 Panel D) or in parallel with the impure GnT-III (Panel E). Both enzymatic assays give a single peak with the correct mass eluting at exactly at the same time. GnT-III shouldn't be able to convert Man5 as a substrate, as it is known to work poorly if the substrate does not bear β1,2 linked GlcNAc residues to the α1,3 core – mannose [172]. The resulting product conversion seen in Figure 67, panel C can thereby only by the activity of GnT-I originating from lysed insect cells. Although the unwanted by-product Man5Gn would be the appropriate substrate for the GnT-III to produce the hybrid bisecting glycan structure Man5Gnbi, which is necessary for the subsequent generation of some of the glycan standards shown in Figure 62. But unfortunately the activity of the recombinant enzyme was far too weak to produce sufficient amounts of hybrid bisecting glycans.

4.4.9 Positive and negative mode MS and MS/MS experiments of hybrid reference glycans

Nevertheless, some glycan standards could be produced in rather low amounts (Figure 68), and by calculating the RRT in comparison to the three peaks occurring in the mouse brain a possible identification could be detected. The RRT (Table 11) of the very first peak of the mouse brain glycans A I is 0,46 which elutes in close proximity to B', a Lewis x fucosylated bisecting glycan structure with an β1,4 linked galactose. Peak D', which is a also Lewis x fucosylated but carries a β1,4 linked GalNAc instead of the Gal and lacks a bisecting GlcNAc, elutes similar to peak A III. The likelihood of the occurrence of such a structure though is rather low. As the retention times were not entirely conclusive additional negative mode MS/MS fragmentation of the reference standards as well as the mouse brain glycans was performed. Figure 68 shows the fragmentation patterns (panel B'-D') of the prepared hybrid reference glycans (panel B-D). Some diagnostic ions (panel B': 346 m/z, 364 m/z and 510 m/z for Lewis x ; panel C': 350 m/z and 553 m/z for core α1,6 and D': 350 m/z and 553 m/z for core α1,6) were found, which confirm the composition of the reference glycans, though in this experiments Nh_4FA-Buffer was used instead of NH_4CO_3-Buffer which lead to rather poor fragmentation results as well as inconclusive diagnostic ions (Panel C'). Fragments labelled

with * are masses corresponding to diagnostic fragments indicative for Lewis x fucosylation, whereas the prepared reference standard is supposed to carry only a core α1,6 fucosylation.



Peak	RRT
A I	0,46
A II	0,71
A III	0,87
B'	0,44
C'	1,25
D'	0,88

Table 11 RRT of mouse brain glycans and hybrid

Figure 68 Positive and negative mode MS and MS/MS experiments of hybrid reference glycans
RRT experiments of hybrid reference glycans and isolated mouse brain glycans (A-D) were performed with PGCC-ESI-QTOF-MS in positive ionisation mode. Fragmentation of reference glycans B'-D' was performed was performed with PGCC-ESI-IT-MS/MS in negative ionisation mode. Peaks are labelled with fragment names.

4.4.10 Positive and negative mode MS and MS/MS experiments of mouse brain glycans

As stated previously the amount and purity of the enzymes was not sufficient enough to produce pure reference glycans in high amounts which led to the last experiment. The three singly as well as the doubly fucosylated glycan isomers found in the mouse brain were fractionated and separately analysed by PGCC-ESI-IT-MS/MS in negative ionisation mode. The used buffer was a NH_4CO_3-Buffer pH: 8 for better fragmentation results, to identify the nature of fucosylation according the detection of diagnostic ions. The results are shown in Figure 69. Panel A shows the chromatogram of the doubly fucosylated structure $[M\text{-}2H^+]^{2+}$ 966,4 m/z. The corresponding MS/MS fragmentation spectrum (panel A 1) clearly confirms the presence of diagnostic fragments for α1,6 core fucosylation 350 m/z (Z_1), 368 m/z (Y_1) and 553 m/z (Z_2) as well as fragments for Lewis x fucosylation (α1,3 linked fucose) 262 m/z ($^{1,3}A$), 304 ($^{2,4}A$), 346 m/z (C_2), 510 m/z (B_2) and 528 m/z (C_3). Investigating the three singly fucosylated peaks (panel B, B1 – B3) lead to the conclusion that B 1 is Lewis x fucosylated (diagnostic fragments: 262 m/z ($^{1,3}A$), 304 m/z ($^{2,4}A$), 346 m/z (C_2), 510 m/z (B_2) and 528 m/z (C_3)). This confirms the possibility of the RRT results shown in Figure 68, though the exact glycan composition could not be identified in the fragmentation results. Panel B2 and B3 in contrast only showed diagnostic fragments for an α1,6 core fucosylation: 350 m/z (Z_1), 368 m/z (Y_1) and 553 m/z (Z_2).

Taking into account that the RRT for the three peaks are rather low indicating a weaker interaction with the PGC material, together with the knowledge of RT influences of Lewis x fucosylation (RT shift forward), bisecting GlcNAc's (RT shift forward) and α1,6 core fucosylation (RT shift backwards) allows to draw some structural conclusions.

Peak B 1 might be a the hybrid bisecting Lewis x fucosylated glycan structure also shown in Figure 68, whereas peak B2 and B3 are probably structural isomers differing in the decoration of the antennae. Both isomers are core fucosylated and due to the early elution they are probably bisecting, complex glycan isomers carrying an additional α1,3 linked galactose, which is a glycan epitope often found in murine tissues on each antennae, respectively.

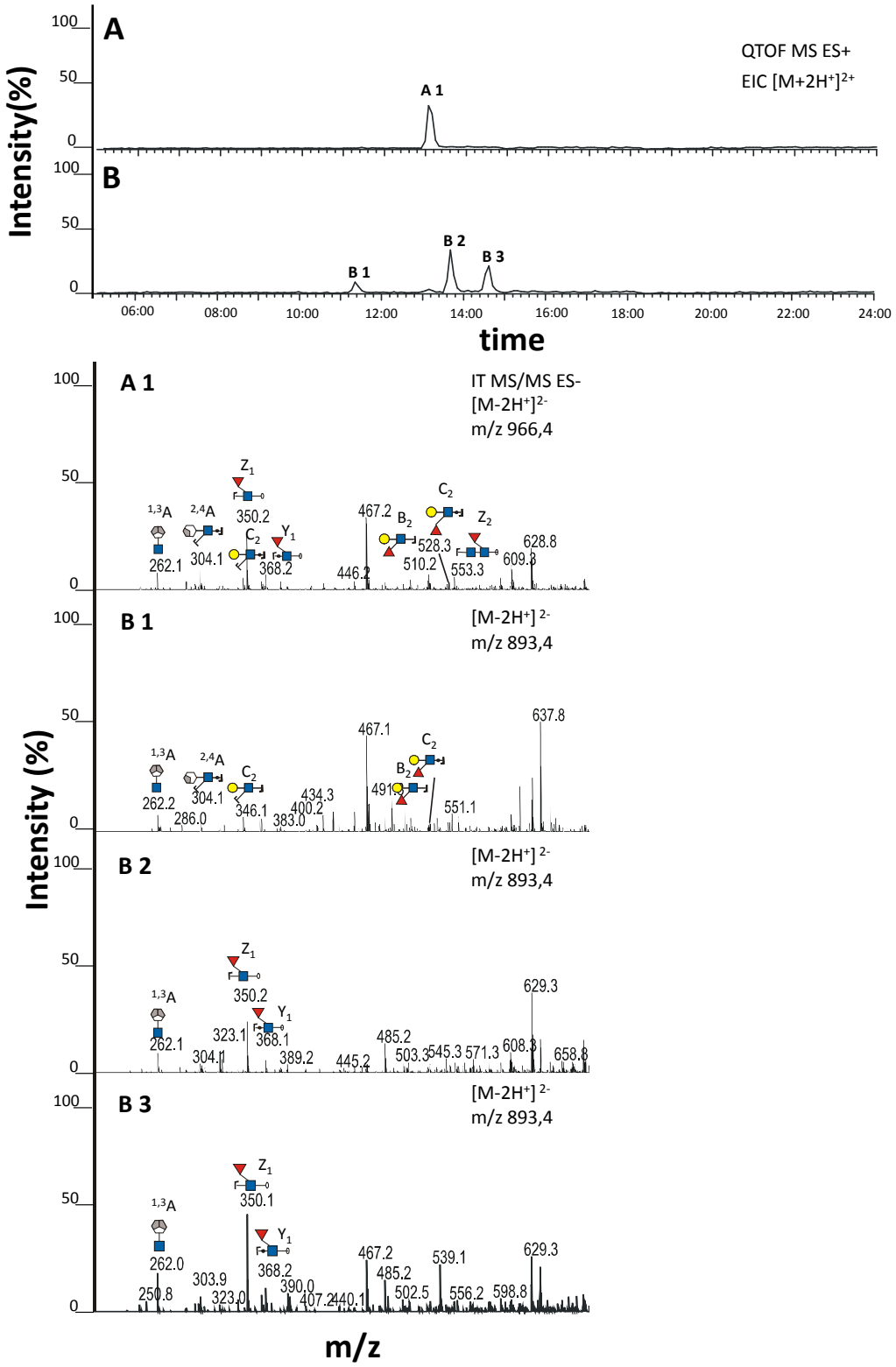
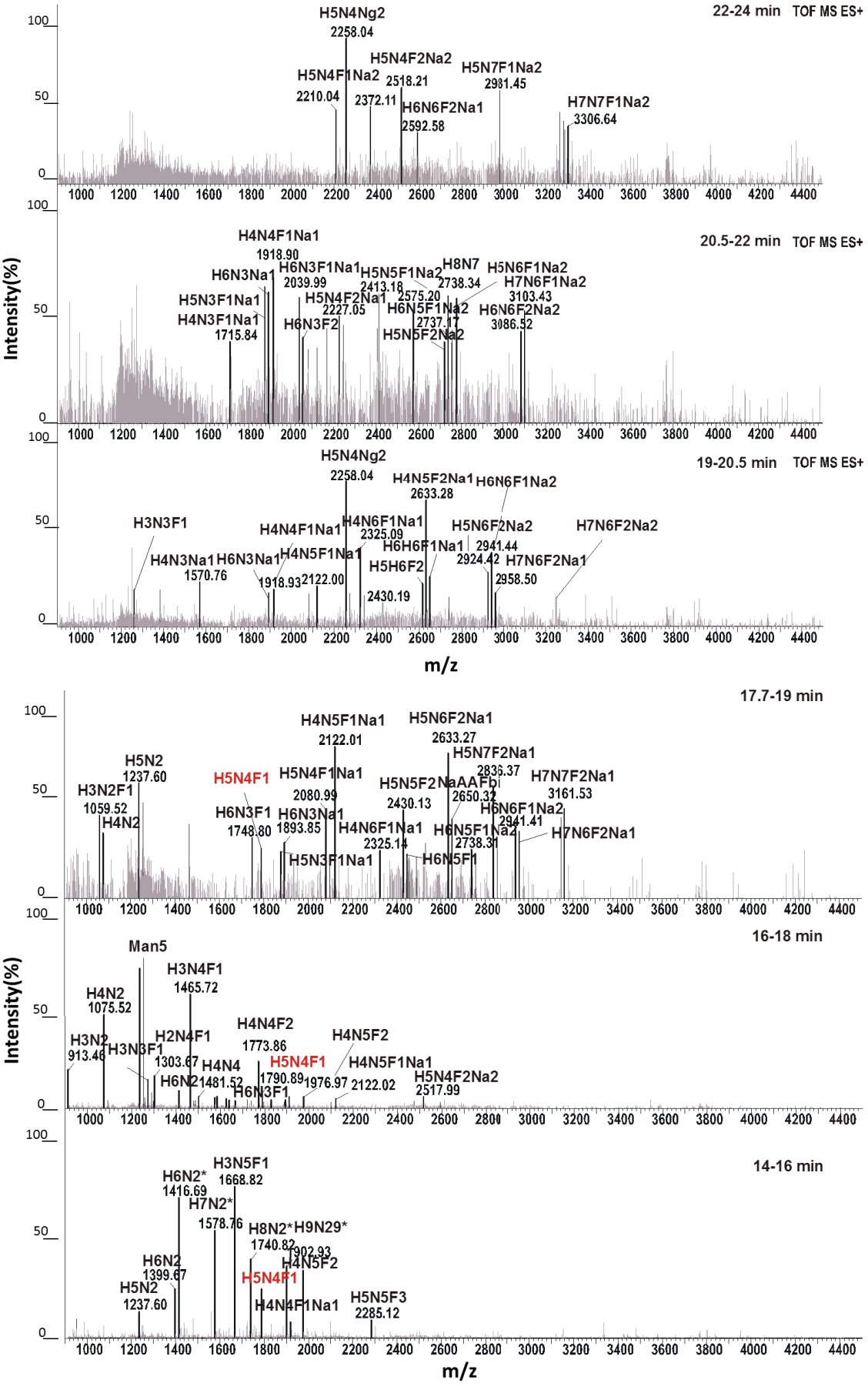


Figure 69 Positive and negative mode MS and MS/MS experiments of singly and doubly fucosylated mouse brain glycans
RRT experiments of glycan isomers isolated from mouse brain (A,B) were performed with PGCC-ESI-QTOF-MS in positive ionisation mode. Fragmentation of isomeric glycans A1-B3 was performed with PGCC-ESI-IT-MS/MS in negative ionisation mode. Peaks are labelled with fragment symbols and names.

Complex biological glycan mixtures are high analytical challenges. The singly fucosylated glycan species shown in Figure 69 are only a few glycan species found in the complex glycome of mouse brain glycans. A putative annotation of the mouse brain glycome is shown in Figure 70, where $[M+H+]^+$ mass spectra of ~ 2 min elution windows (10-12 min, 12-14 min, 14-16 min, 16-18 min, 17,7-19 min, 19-20,5 min, 20,5-22 min and 22-24 min) from a chromatographic separation performed with a PGC column are shown. Putative glycan species ranging from oligomannosidic structures, complex and possibly hybrid structures up to multi antennary highly sialylated structures can be found.

The annotated structure AAF/Man5GnFbi labelled in red is found in the elution windows 10-12 min, 12-14 min, 14-16 min, 16-18 min and 17,7-19 min, indicating the different singly fucosylated glycan isomers present in the biological sample of mouse brain N-glycans, which were investigated in more detail previously. The later the elution-window, the more complex are the found glycan structures. Peaks are labelled with putative structure compositions according to the proglycan nomenclature [164]. For exact structural identification, large scale fractionation, additional exoglycosidase treatments, RT correlations and fragmentation experiments in positive and negative ionisation mode would be necessary.


22-24 min TOF MS ES+
H5N4Ng2
2258.04
H5N4F2Na2
H5N4F1Na2
2210.04
2518.21
H5N7F1Na2
2981.45
2372.11
H6N6F2Na1
2592.58
H7N7F1Na2
3306.64
20.5-22 min TOF MS ES+
H4N4F1Na1
1918.90
H6N3F1Na1
H6N3Na1
2039.99
H5N5F1Na2
2413.18
H8N7
2738.34
H5N6F1Na2
H7N6F1Na2
3103.43
H5N3F1Na1
H5N4F2Na1
2227.05
2575.20
H6N5F1Na2
2737.17
H6N6F2Na2
3086.52
H4N3F1Na1
1715.84
H6N3F2
H5N5F2Na2
19-20.5 min TOF MS ES+
H5N4Ng2
2258.04
H4N5F2Na1
2633.28
H6N6F1Na2
H3N3F1
H4N4F1Na1
H4N6F1Na1
2325.09
H5N6F2Na2
H7N6F2Na2
H6H6F1Na1
2941.44
H4N3Na1
1570.76
H6N3Na1
H4N5F1Na1
2924.42
H7N6F2Na1
H5H6F2
1918.93
2122.00
2958.50
2430.19
m/z
17.7-19 min
H4N5F1Na1
2122.01
H5N6F2Na1
2633.27
H5N7F2Na1
H5N2
1237.60
H5N4F1
H5N4F1Na1
2836.37
H7N7F2Na1
H5N5F2NaAAFbi
2650.32
3161.53
H3N2F1
1059.52
2080.99
2430.13
H6N6F1Na2
H6N3F1
H6N3Na1
2941.41
H4N2
1748.80
1893.85
H4N6F1Na1
H6N5F1Na2
H7N6F2Na1
2325.14
2738.31
H5N3F1Na1
H6N5F1
16-18 min
Man5
H3N4F1
1465.72
H4N2
1075.52
H4N4F2
H4N5F2
1773.86
H3N2
913.46
H3N3F1
H2N4F1
1303.67
H4N4
H5N4F1
H4N5F1Na1
H6N2
1481.52
1790.89
1976.97
2122.02
H5N4F2Na2
2517.99
H6N3F1
14-16 min
H3N5F1
1668.82
H6N2*
1416.69
H7N2*
1578.76
H8N2*H9N29*
1740.82
1902.93
H6N2
1399.67
H4N5F2
H5N4F1
H5N2
1237.60
H4N4F1Na1
H5N5F3
2285.12
Intensity(%)
m/z

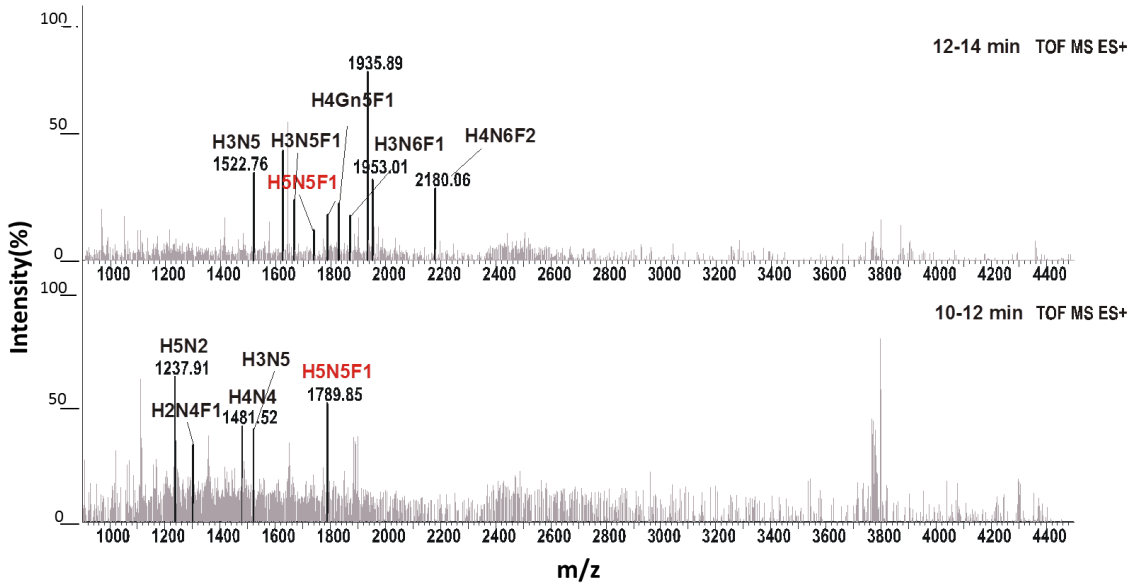


Figure 70 ESI-QTOF-MS mass spectra of mouse brain N-glycans from different PGC elution windows
Deconvoluted TOF MS ES+ $[M+H+]^+$ mass spectra of ~ 2 min elution windows from a PGC chromatographic separation are shown. Peaks are putatively labelled according to their possible structural composition using the proglycan nomenclature **[164]**. Glycan structure labelled in red (AAF/Man5GnFbi) represents the occurrence of the singly fucosylated glycan isomers present in the mouse brain tissue.

4.5 Method development for site specific glycopeptide analysis using the model protein rhEPO-Fc

Beside the isomeric glycan analysis a lot of analytical work during the theses was performed on glycopeptide analysis of proteins recombinantly expressed in plants. Recombinant protein expression and its use in the pharmaceutical industry is very well established. The example used here is the work that was done on the well investigated recombinant human erythropoietin (rhEPO) see also **Appendix I** and **Appendix II**. This protein was one of the first hematopoetic growth factors, which was approved to treat anaemia correlated to cancer, kidney failure and other pathological conditions [173]. The mature glycoprotein EPO consists of 166 amino acids and carries 3 N-glycosylation sites (Asp-24, -38, -83) as well as one O-glycosylation site (Ser-126) with an overall weight of ~ 30 kDa all of which the glycosylation accounts for 40 % of the total weight [174]. The glycosylation of recombinant hEPO plays a crucial role in terms of overall stability and activity as several studies report that terminal sialic acids increase the circulatory half-life and furthermore show that the ration of tetra- to bi-antennary glycans positively affects the *in vivo* biological activity of the glycoprotein [175] [176]. Due to the importance of the complex glycosylation pattern rhEPO is usually produced in mammalian cell cultures mainly in Chinese hamster ovary cells (CHO) which is a very costly way. Additionally to the glycosylation and sialylation the fusion of rhEPO to an Fc-fusion peptide leads to a stabilizing effect and extended serum half-life [177]. A highly attractive alternative in terms of large-scale and low-cost production of biopharmaceuticals is the use of plants due to their ability to carry out complex N-glycosylation with remarkably homogeneity which makes them suitable for N-glyco-engineering in the direction of complex tetra-antennary multi sialylated glycosylation [178]. Figure 71 shows glycopetide analysis using RP-ESI-QTOF-MS of the three different glycopeptides of rhEPO expressed in *N. benthamiana* ΔXTFT plants co-expressing mammalian genes for synthesis of tri-antennary sialylated *N*-glycans. Double digest of the purified rhEPO-Fc using two proteolytic enzymes (trypsin and endoproteinase Glu-C) were performed, generating four separate glyco-peptides, three originating from rhEPO and the fourth from the Fc fusion-peptide (data not shown). Results show the highest abundance of a tri-antennary tri-sialylated glycan structures present on all three glycopeptides which could

further been processed to tetra-antennary and tetra-sialylated N-glycans by co-expression of the respective mammalian genes.

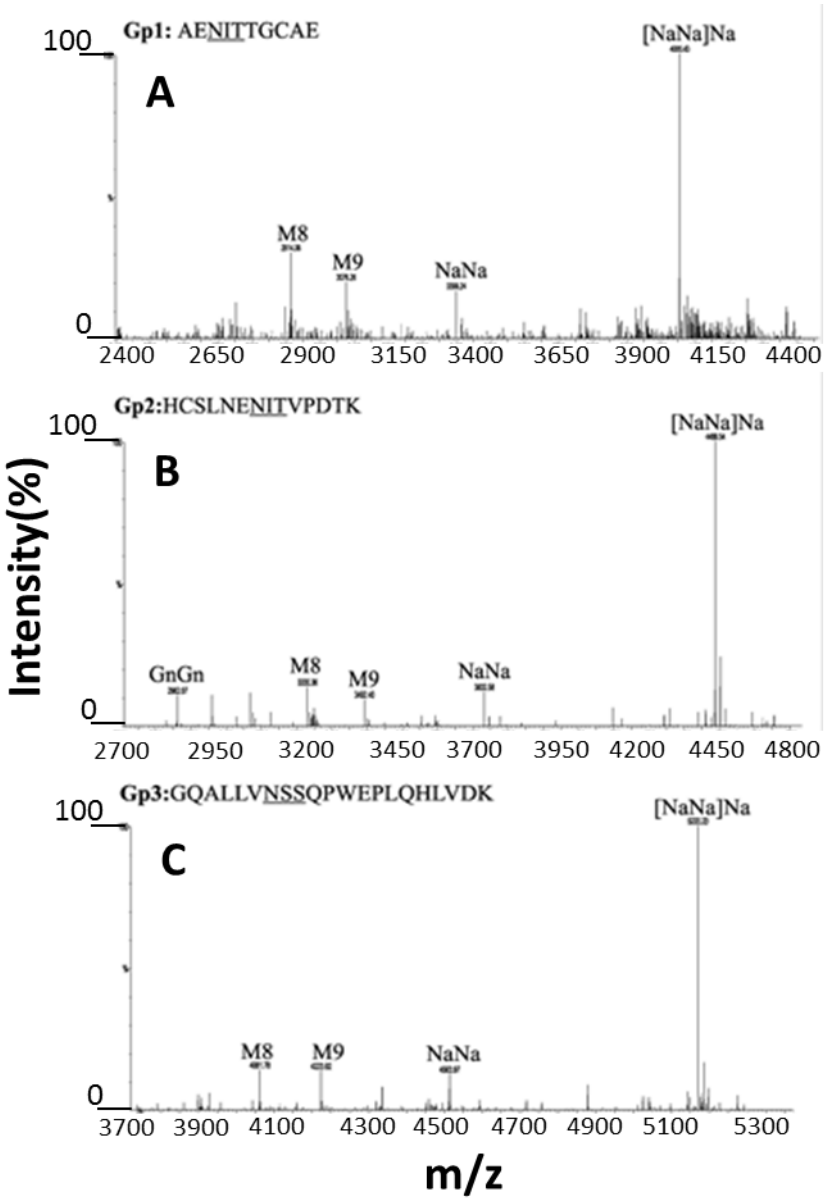


Figure 71 N-glycopeptide analysis of rhEPO-Fc expressed in plants generating tri-sialylated N-glycans (Figure taken and modified from [178]).
RP-ESI-QTOF-MS mass spectra of trypsin and endoproteinase Glu-C double-digested rhEPOFc co-expressed in *N. benthamiana* ΔXTFT with mammalian genes for synthesis of tri-antennary sialylated *N*-glycans ($rhEPO_{TriSia}$). Glycosylation patterns of rhEPO (A) Gp1: E^{21}AENITGCAE^{31}; (B) Gp2: H^{32}CSLNENITVPDTK^{45} and (C) Gp3: G^{77}QALLVNSSQPWEPLQHLVDK^{97} are shown. Peak labels were made according to the ProGlycan system (www.proglycan.com) [164].

It has been suggested that the clearance of EPO from the blood circulation, which is influenced by the sialic acid content of the glycosylation, might be even enhanced by a sialic acid containing core 1 mucin-type O-glycosylation [179]. Therefore, besides the N-glycan engineering of rhEPO in plants also successful O-glycan engineering of the O-glycan site Ser-126 was performed in *N. benthamiana* ΔXTFT. Transient co-expression of genes necessary

for the generation of mucin-type core 1 O-glycans decorated with two sialic acids, together with the machinery for generation sialylated N-glycans was performed by Alexandra Castilho (Figure 72) [180]. Glycopeptide analysis using RP-ESI-QTOF-MS was performed showing (A) successful generation of a doubly sialylated core 1- O-glycan structure on the Ser-126 which is naturally not glycosylated when expressed in plants due to the lacking of an O-glycosylation machinery in this expression host. Panel B-D represents glycopeptide analysis of the three simultaneously analysed sialylated N-glycopeptides from rhEPO-Fc.

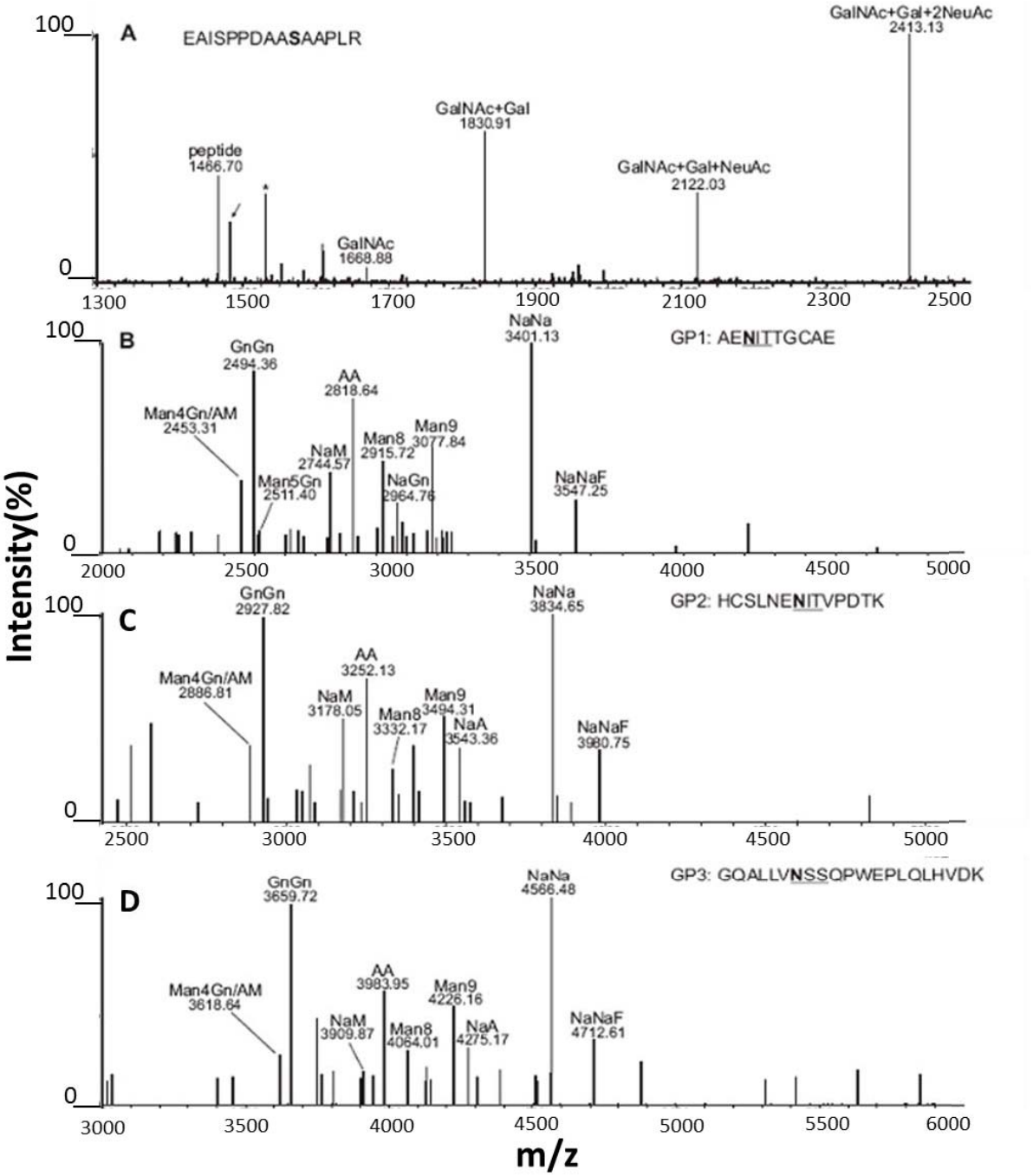


Figure 72 O- and N-glycopeptide analysis of rhEPO-Fc expressed in plants generating disyalylated O-glycans (Figure taken and modified from [180])

RP-ESI-QTOF-MS mass spectra of trypsin and endoproteinase Glu-C double-digested rhEPOFc co-expressed in *N. benthamiana* ΔXTFT (B-D) together with mammalian genes for synthesis of bi-antennary sialylated N-glycans ($rhEPO_{DiSia}$) as well as the genes of the glycosylation machinery to produce all possible glycosylation variants up to a doubly sialylated core-1 O-glycan structure. The presence of a peptide corresponding to the hydroxylation of one proline residue is indicated by an arrow. The Asterisk denotes the presence of a contamination. Peak labels were made according to the ProGlycan system (www.proglycan.com) [164].

4.6 Glycomic-, glycoproteomic-, proteomic- and metabolomic - work performed in collaborations

Additionally to the work performed on rhEPO-Fc several other projects analyzing glycopeptides from proteins expressed in plants (**Appendix III** and **Appendix IV**) or glycan analysis as well as glycopeptide- and proteomic-analysis of bacteria derived glycoproteins (**Appendix V**, **Appendix VI** and **Appendix VII**) was performed. Furthermore some experiments on the analysis of nucleotide sugars for metablomics (**Appendix VIII**) and glycan analysis for the characterization of proteins (**Appendix IX**) was performed.

5 Discussion

5.1 Influenza A Virus - Hemagglutinin

5.1.1 Affinity chromatography

When the project about binding specificities of the swine-origin Influenza A virus was started in 2009 the recently emerged pandemic Influenza A virus A/California/07/2009(H1N1) was already a very well-investigated influenza strain due to the pandemic appearance of H1N1 in 1918, the so called Spanish Flu. Nevertheless the task was to identify glycan epitopes beyond the already known isomeric sialic acid linkage prevalence to gain a more detailed insight into influenza binding specificities and the isomeric glycan epitope structures.

After successful establishment of the insect cell culture in our laboratory, together with my PhD colleague Andreas Thader (Mag. Boku. Department of Chemistry), the recombinant expression of two hemagglutinin constructs (H1, A/California/07/2009(H1N1) and H3 A/Hiroshima/52/2005(H3N2)) was started with the aim to perform affinity chromatography measurements of immobilized protein. Several pre-experiments with different isolated (LCA-lectin) and purchased (ConA, Galectin) lectins were performed using NHS-activated sepharose for protein immobilisation and fluorescently labelled (ConA and Galectin) or native (LCA-lectin) glycans to perform binding experiments. Glycan binding according to the respective binding specificities could be shown either by using fluorescence detection after a prior NP-HPLC experiment (Galectin) or by using RP/PGC-ESI-IT/QTOF-MS experiments to analyse obtained fractions from batch or centrifugation tube affinity experiments with ConA and LCA lectin. Both hemagglutinins were successfully recombinantly expressed and purified. Subsequent activity testing was performed with hemagglutination experiments using chicken erythrocytes as well as human erythrocytes from the different blood groups. However, none of the tested assay conditions led to a positive result, which can be explained due to the cDNA sequence of the used constructs. For soluble protein expression into the cell culture supernatant, a truncated protein version lacking the transmembrane region was initially used. This transmembrane region was later discovered to be crucial for protein trimerization and thus its activity. To overcome this problem a trimerization signal can be cloned into the C-terminus of the protein, which mediates the trimerization and thereby

leads to an active product [166]. Subsequent affinity measurements and glycan array experiments were performed with several of the new and active constructs obtained from Florian Krammer (Ass. Prof. Dr., Mount Sinai School of Medicine, New York, Microbiology). The recombinant hemagglutinins were immobilized on NHS-activated sepharose and used in different affinity chromatography experiments. All affinity-based experiments performed with the newly obtained proteins showed only weak glycan retention, probably due to the low affinity of hemagglutinin. Surprisingly a clear retention of oligomannosidic structures could be seen in affinity chromatography experiments of H1 as well as H3. As the retention of sialic acid-containing glycans was very weak the question arose if the coupling procedure or the elution with a low pH buffer probably harmed the protein structure and thereby hindered or impaired sufficient binding. As it was shown in previous experiments with LCA lectin a centrifugation based affinity assay using amicon centrifugation tubes with a cuf off of 10 kDa worked nicely to prove glycan binding to a soluble protein. Therefore the same experiment was performed with soluble H1 and H3 hemagglutinin together with AA-labelled glycans. Oligomannosidic glycans were again found in relatively high intensity in the elution fractions but the retention of the sialic acid containing glycans still did not give the expected results. It is known that hemagglutinin affinities are relatively low when investigated by surface plasmon resonance [181] and most probably need a tetravalent glycan presentation and simultaneously up to 1 µM hemagglutinin concentrations to perform sufficient binding experiments [168]. As the glycans isolated from swine tissue mainly contained one and two sialic acids the necessary multivalence could not be provided with free glycans.

5.1.2 Glycan arrays

Further binding-specificity analyses were performed with two different glycan arrays. Glycan arrays are a very well established method of investigating binding specificities of pathogens resulting in infectious disease such as influenza A viruses or recombinant proteins such as hemagglutinins [182]. Many different glycan array fabricates are nowadays available for the assessment of binding specificities as well as for the identification of possible glycan epitopes usable for subsequent vaccine designs. However, acquiring a large collection of glycans in a suitable format for immobilization on an array-surface is still a challenge. Therefore, the glycan arrays are designed using glycans from two very different origins, either isolated from natural sources or obtained by chemical synthesis. H1 hemagglutinin has already been extensively studied using glycan arrays with chemically synthesized glycans, which have the benefit of having the control of the synthesized target structure and the ability to be produced in large, materially homogenous quantities.

In this study the approach of a glycan array produced with glycans isolated from the natural source of different swine or ferret tissues was chosen. This approach offers several advantages as the glycans are synthesized by the organism itself. They can be isolated, fluorescently derivatised and chromatographically separated into groups of structures or if possible into single structures. Obtained glycan fractions can then be immobilized on glass slides through the reaction of the secondary amide group from the fluorescent labels with an epoxy silane linker on the glass surface. Furthermore the usage of natural glycans offers the possibility to immobilize the entire glycome of a biological source. [183]. In this study one array used N-glycans isolated from different ferret tissues (lung, trachea, nasal turbinate (NT) and bronchus) and was prepared by Miranda de Graaf (PhD, Erasmus MC, Utrecht, NL). The other array was prepared with AA-labelled N-glycans isolated from different swine tissue types (trachea, larynx and lung). Due to the rather similar glycan content within the different tissues, glycans were pooled and three dimensionally separated adapting a previously published two dimensional separation technique [109]. Three different HPLC columns were used to separate and fractionate the swine N-glycan pool by charge (weak anion exchange), hydrophility and size (HILIC) and hydrophobicity (RP). A resulting number of 164 different glycan fractions together with synthetic glycol-conjugates were printed in different concentrations on NHS-activated glass slides to produce a “shotgun” glycan array similar to

the one produced by van Diepen et al. [109]. All available 7 different recombinant hemagglutinins were tested on the ferret- as well as swine-arrays at a minimum in duplicates. Results shown in this work concentrate on the comparison of H1 (A/California/04/09) to H3 (A/Hong Kong/1/68) and another H1 (A/New Caledonia/20/99). The two reference hemagglutinins were chosen for the comparison, as the H1 hemagglutinin from Cal04/09 is known to bind specifically α2,6-linked sialic acid according to glycan array data available online from the Consortium of Functional Glycomics (CFG) [106] whereas H1 from New Caledonia 99 was found to exhibit a broader overall binding and a preference for α2,6-linked sialic acid but also high intensity binding to α2,3-linked sialic acid. H3 form HK68 is known to efficiently replicate in swine trachea and bronchus tissue bearing mostly α2,6-linked sialic acid but also in lung tissue from swine which is known to additionally carry α2,3 linked sialic acid [65]. These published results are in accordance with the findings in this study indicated by the simultaneous binding of synthetic glycol-conjugates carrying both types of sialic acid linkages. However, the clear preference of H1 (Cal04/09) towards the α2,6 linkage can be seen in contrast to the two other hemagglutinins, which show sufficient binding to glycoconjugates carrying both linked sugar moieties. These preferences are further confirmed by comparing the binding profiles of the three proteins to both glycan arrays. H1 (New Caledonia 99) and H3 (HK68) show a higher intensity binding to sialic acid-containing fractions on the ferret array, which are known to be mostly α2,3-linked, whereas the binding of H1 (Cal04/09) is more prominent on the swine arrays, where the overall occurrence of α2,6-linked sialic acid is higher. Nevertheless the initial aim of this study, besides the successful generation of a shotgun glycan array, was to possibly identify glycan features beyond the differently attached sialic acid. When comparing the overall glycan binding profile to both glycan arrays, and furthermore taking the affinity results into account, a rather broad binding of the recombinant hemagglutinins was detected, which surprisingly showed additional binding to neutral glycans such as oligosaccharidic structures or glycans carrying α-galactose epitopes or several fucoses. This fairly unspecific binding might be a result of the used expression system. A study by de Vries et al, 2010 investigated the binding specificities of hemagglutinins carrying different types of glycosylation as a result of the usage of different mammalian- or insect cell-cultures for expression. The hemagglutinins derived from insect cells carrying paucimannosidic N-glycans bound less specifically and far more broadly on glycan arrays compared to hemagglutinins carrying

complex galactosylated N-glycosylation derived from a mammalian expression system HEK293T [184]. Therefore the idea was to express the hemagglutinin in Chinese hamster ovary (CHO) cells which are known to produce human-like glycosylation, but the expression of the protein did not work in sufficient amounts and due to time issues the project was not further pursued and could not be finished. Besides the usage of a different cell expression system the broad binding could also be a result of the not optimized assay conditions itself. Though several concentrations (0,5 – 8,0 µM) of the recombinant expressed proteins were tested and the concentration of 2,0 µM gave the optimum in the ratio between binding and background signals, the conditions should also be adapted in terms of blocking or antibody incubation time and washing procedure. Due to the limited availability of produced arrays, which is a cost-intensive procedure and the very time consuming analysis of each performed array, the optimization was only performed towards the protein concentration. The other incubation times were adapted from assay conditions performed with whole viruses, which gave similar results in terms of bound fractions, but also showed a far more specific binding profile without comparable high interference with neutral glycan fractions. The work using whole viruses was performed by Miranda de Graaf using her ferret glycan array.

Optimized assay conditions in combination with hemagglutinins carrying different types of glycosylation would be necessary to verify the obtained results for the binding affinities presented in the current study.

5.2 Isomer specific analysis of fucosylated N-glycans

The second part of the thesis focused on the investigation of released glycans using porous graphitized carbon phase chromatography (PGCC) online coupled to electrospray ionisation mass spectrometry (ESI-MS). Here, the focus was on the separation of released glycans, which appear isomeric with regards to their fucose linkage or to their fucose position within the glycan. Distinct isomers can be crucial for many biological systems as shown in the example of differently linked sialic acids responsible for species specific Influenza A virus receptor binding [59]. Furthermore, also the relative quantity of existing isomers in certain tissues can be correlated to several diseases [56] as for example up regulated occurrence of sialyl-Lewis x and sialyl-Lewis a glycan epitopes is found in gastrointestinal cancer [75]. Additionally, some specific linkages of isomeric fucose epitopes were found to be necessary in mediating cell-cell adhesion reactions like the selectin mediated leucocyte adhesion [68]. Moreover, specific fucose linkages or known to be an essential part in the generation of the different blood group antigens [67]. Furthermore, CD24, a cell adhesion molecule which is known to play an important role in the central nervous system, was discovered to be highly Lewis x fucosylated [72]. Another study revealed the importance of glycoproteins carrying Fuc α1,2 Gal epitopes in neuronal development and memory [73]. The importance of the function of different isomeric N-glycans within complex biological samples calls for a straightforward analytical tool to selectively identify specific glycan epitopes. Therefore, in this study we aimed to generate a retention time library by generating reference glycans carrying core α1,6- and/or different Lewis – type fucose residues. The ordinary complex type asialo diantennary N-glycan with one single fucose can already form up to 30 isomeric combinations. Additionally, taking in addition the possibility of hybrid type structures, bisecting GlcNAcs, α-galactose epitopes, GalNAc residues and LacNAc repeats into account, this number increases drastically. Based on previous studies, which demonstrated the separation of isomers showing differently linked sialic acids as well as oligomannosidic structures, here the aim was to systematically analyse fucosylated N-glycans [35] [140].

The chosen approach aims to separate the isomeric glycans before detection using a porous graphitized carbon (PGC) phase coupled via electron spray ionisation (ESI) to a quadrupol time of flight (QTOF) or to an ion trap (IT) mass spectrometer (MS). In recent years, PGC-ESI-MS has evolved into the most powerful method for the straightforward analysis of complex

glycan mixtures [144]. Starting from the two isomeric glycan forms A^4Gn and GnA^4, singly galactosylated glycans isolated from hIgG, the four different galactosylated glycan isomers were enzymatically produced (β3GalT5) with or without core fucosylation. The generation of these differently galactosylated isomers provided the basis for further enzymatic treatments with fucosyltransferase-2, FucT-III and FucT-IV. A reference standard set of different galactosylated and Lewis- as well as core-fucosylated N-glycans with one, two or even three fucoses was produced. Each of the reference glycans was separated by PGC-chromatography in order to generate a retention time library. To overcome run-to-run retention time fluctuations, simultaneously stable isotopically labelled internal standards were added. For each glycan standard, relative retention times were calculated (Table 4). This allowed a good compensation of the retention time shifts caused by not-yet fully understood red-ox reactions of the graphite phase and by the clogging of the column pores [139] [185]. The proof of principle for the generated retention time library was tested with singly, doubly and triply fucosylated N-glycans isolated from the S-chain of hIgA, which are known to be Lewis x fucosylated [169]. A successful identification of the respective glycan isomers could be achieved by comparing the elution profiles of the reference standards carrying Lewis x fucosylation on β1,4-galactoslyated complex glycans. Additionally a mixture of α1,6-linked core fucosylation in combination with Lewis x fucosylation could be structurally defined down to the decoration of the respective antenna. The calculated relative retention times (RRTs) further supported an unambiguous identification. Three singly-, three doubly- and one triply- fucosylated N-glycan structure from the complex glycome of the S-chain IgA glycans were identified in one single chromatographic run. Though PGC columns have an amazingly highly reproducible and shape selective power, slight retention time shifts are a constant problem during analysis. The separation performance can be further optimized by using smaller particles (3µm diameter), shallow gradients and columns with increased length [138]. Besides the structural elucidation of S-chain IgA glycans the RT-library was also used to elucidate singly and doubly fucosylated N-glycan species from the biological sample of mouse brain. Reduced N-glycans originating from mouse brain were separated using PGCC with the addition of internal standards for RRT calculation. Interestingly none of the prepared reference standards showed coelution with the three singly- and one doubly-fucosylated isomers found in the mouse brain N-glycome. Due to the rather early elution of the unknown isomers and the knowledge of the elution range of hybrid and bisecting type

glycans, we were able to presume that the detected fucose isomers were of hybrid or bisecting type nature. This led to the generation of a second and third set of reference-glycans.

5.2.1 Reference-standard generation

For the generation of the first set of reference glycans consisting of complex biantennary glycan isomers, only two or a maximum of three enzymatic steps were needed. This was found to be reasonable, since several more steps would account for a too large sample loss during the multiple purification steps. Each enzymatic step calls for careful purification via PGC-SPE cartridges prior the monitoring of substrate conversion using PGCC-ESI-QTOF-MS. Without purification the strong hydrophobic character of the stationary PGC phase would greatly interfere with the added enzymes and lead to the previously mentioned clogging of the column. Furthermore, highly active and pure enzymes are required for fully converted and pure substrates. When the generation of the hybrid and bisecting reference standards was considered, the number of enzymatic and biochemical steps to generate one reference glycan rose to seven. This is a relatively large number of follow-up steps for the generation of a single glycan standard. The generation started here with the catalysis of a Man5-glycopeptide originating from a mixture of mainly Man5- and Man6-glycopetides (GP) isolated from the TAKA-amylase. Using an α-mannosidase from jackbean the major glycoform of a Man4-glycopeptide was generated, which was used for further treatment with GnT-I to obtain Man4Gn-GP. For the addition of a core α1,6 fucose the glycopeptides were incubated with a microsome [186] preparation from chicken hearts which are known to contain the FucT-VIII enzyme [26], but also mannosidase and several other glycan catalysing enzymes, which were acting to some extent on the Man4Gn-GP substrate. At this point the glycan could be released from the peptide-backbone using PNGaseF for the enzymatic cleavage. Subsequent borohydride reduction of the released sugar alditol could be performed to allow for conclusive PGCC analysis obtaining sharp peaks for each glycan moiety [144]. Several glycosyltransferases were used for remodelling the distinct glycan isomers adding galactoses, GalNAcs and the different fucoses. Most of those enzymes were recombinantly expressed by my PhD colleague, Andreas Thader (Mag., Boku, Department of Chemistry), using the baculo virus insect cell expression system. Enzyme activity was tested using the appropriate fluorescently labelled glycan-substrates and NP-HPLC for monitoring of product conversion by peak shifts. Unfortunately several problems arose with the usage of some of these enzymes. The partly very low substrate conversion of all enzymes constantly needed for newly produced proteins, which is a time consuming procedure.

Furthermore the new expression of enzymes would have asked for further protein purification, as the enzymes were usually expressed in the insect cells' supernatant. However, a further purification step (usually HIS-Tag purification) was not possible for all of the enzymes as e.g. the GnT-III enzyme did not contain the necessary polyhistidine tag in the protein sequence and thus could only be used directly from the crude cell culture supernatant, which was ultra-filtrated for protein up-concentration. At first it was believed that the poor substrate conversion of Man4Gn using the GnT-III, which was tested for activity using GnGn as a substrate, was a result of activity loss. However, after several rounds of newly prepared enzymes a more detailed enzyme characterization was performed by testing several substrates. It could be shown that the enzyme contained high GnT-I side activities as well as other side activities originating from the insect cell culture. Therefore the sub-optimal substrate Man4Gn was hardly catalysed. Since partly new cloning, expression and extensive purification are all necessary, this part of the generation of reference standards will be continued in further ongoing projects. Nevertheless, some hybrid and bisecting reference glycan standards could be generated albeit in a very low concentration. Nevertheless these reference standards were used for assessing relative retention times and a putative glycan identification of one single-fucosylated isomer from the mouse brain glycome could be accomplished. This glycan structure was assigned to be a Lewis x fucosylated hybrid type isomer with a bisecting GlcNAc.

5.2.2 Collision induced dissociation (CID) experiments on reference glycans and glycans from model proteins

State of the art characterization and analysis of isomeric glycans usually involves fragmentation analysis using mass spectrometry. Therefore, in this study we also emphasized to evaluate the usefulness of this strategy. Several fragmentation experiments were performed on the generated reference glycans as well as the fucosylated glycans isolated from the different natural sources used in this study. In a first set of experiments the fragmentations were performed using PGCC-ESI-IT-MS/MS operated in positive ionisation mode. This fragmentation experiment resulted in rather a simple series of fragments typically consisting of B- and Y-ions as well as a few B/Y ions [140]. Though the fragmentation patterns were fairly clear, a comprehensive structural assignment of the different glycan isomers could not be accomplished, as the resulting fragments hardly differed within the different types of Lewis fucose containing structures. Therefore, a relatively new approach was chosen: We analysed the glycans in negative ionization mode and performed the CID experiments. Here we additionally obtained A- and X-ion fragments, which result from cross ring cleavages. Furthermore we noticed more intense C- and Z-ions as well as characteristic D and D-18 fragments. [87] [85]. The excellent work of Harvey and Rudd [87] [85] enabled the generation of a list of so called “diagnostic ions” indicative for specific glycan epitopes, such as the different Lewis type fucosylation, core fucosylation and composition of the different antennas. A summarized list of those diagnostic fragments can be found in a study published by Everest-Dass et al [88]. This list was used in this study as a base for structural assignments of the reference glycans and glycans isolated from natural sources according to their fragmentation patterns. This highly informative fragmentation pattern of a negative ion mode MS/MS fragmentation analysis can be used to clearly distinguish between possible coeluting glycan isomers due to the occurrence of diagnostic fragments and additionally can help to confirm structures when the retention time based results are not 100% certain due to the steadily occurring retention time shifts. Therefore negative mode MS/MS fragmentation analysis was also used in this study to confirm the structural composition of the reference glycan standards. However, the first fragmentation experiments were performed with the standard solvent A (65 mM ammonium formate buffer pH 3) leading to the high abundance of formate adducts of the parent ions, which in

turn led to poor fragmentation accessibility. To obtain better fragmentation results with higher abundance of the diagnostic fragments, the buffer system was changed according to the ones used in [88] — an ammonium carbonate buffer at pH 8. The change of the buffer system led to the generation of conclusive and informative fragmentation spectra. This method was further used for the analysis of singly and doubly fucosylated N-glycan species found in the S-chain of IgA, clearly proofing the retention time based assignment of Lewis x fucosylation in combination with α1,6 core fucosylation.

As the generation of the entire set of proposed hybrid and bisecting reference glycans was above the scope of this work and thereby the structural assignment according to their relative retention time was out of reach, this fragmentation strategy was additionally used for the three singly fucosylated and one doubly fucosylated glycan structures found in N-glycans isolated from mouse brain. For this purpose the fucosylated isomers were fractionated and separately analysed by PGCC-ESI-IT-MS/MS in negative ionisation mode. These experiments gave the expected diagnostic fragments for the one already retention time-based assigned isomer carrying a Lewis x fucosylation. Furthermore, the fragmentation pattern for both later eluting, not yet identified isomers, showed indicative fragments for α1,6 core fucosylation. The relatively early elution time in the PGCC experiments along with the indication for a core fucosylation allowed for a structural conclusion to be drawn: As the first of the three isomers was assigned as a hybrid and bisecting N-glycan isomer, the two later-eluting peaks are most likely also hybrid and/or bisecting since the elution profile compared to the complex reference glycans still shows a comparably weak retention. The later elution of the two mouse brain isomers can be explained by the core α1,6 fucosylation. The difference between the two remaining isomers probably can be found in the occurrence of an αGal epitope or an β1,3-linked galactose on the latest eluting peak as it has been shown that both galactose linkages lead to a similar stronger retention compared to a commonly found β1,4-linked galactose [35]. Therefore the isomers were putatively assigned as two narrow-eluting possibly hybrid isomers, differing in the decoration of the antenna both carrying a core fucose, a bisecting GlcNAc and possibly an α galactose epitope. This hypothesis of course needs additional proof via exoglycosidase treatments and the generation of the described reference standards for retention time-based structural assignments. The doubly fucosylated glycan structure found in mouse brain N-glycans carries two different fucose epitopes according to the found diagnostic fragments. Lewis x

fucosylation and additional α1,6 core fucosylation is proposed. The Lewis x fucosylation was additionally confirmed by using an exoglycosidase sensitive for α1,3 fucosylation. In this experiment performed on the total glycome of mouse brain N-glycans the putatively assigned Lewis x fucosylated isomer diminished, whereas the other two isomers remained. Furthermore the doubly fucosylated isomer diminished after the fucosidase treatment and additionally lead to a signal increase of the last singly fucosylated isomer. This unparalleled signal increase of the two unidentified isomers after the Lewis x fucosidase treatment indicates a more substantial difference between the two structures than the linkage of galactoses decorating the antennas.

6 Bibliography

[1] C. H. Wong, “Protein Glycosylation: New Challenges and Opportunities,” *J Org Chem,* vol. 70, no. 11, pp. 4219-25, 2005.

[2] A. Varki, R. Cummings, J. Esko, H. Freeze, G. Hart and J. Math, Essentials of Glycobiology, 2nd edition, 2nd Edition ed., Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2009.

[3] S. J. Clark, P. N. Bishop and A. J. Day, “The proteoglycan glycomatrix: a sugar microenvironment essential for complement regulation,” *Front Immuno,* vol. 4, p. 412, 2013.

[4] P. H. H. Lopez and R. L. Schnaar, “Determination of Glycolipid–Protein Interaction Specificity,” *Methods Enzymol,* vol. 417, pp. 205-220, 2006.

[5] D. B. Werz, R. Ranzinger, S. Herget, A. Adibekian, C.-W. von der Lieth and P. H. Seeberger, “Exploring the structural diversity of mammalian carbohydrates (“glycospace”) by statistical databank analysis.,” *ACS Chem Biol,* vol. 2, pp. 685-91, 2007.

[6] Y. Van Kooyk, H. Kalay and J. J. Garcia-Vallejo, “Analytical tools for the study of cellular glycosylation in the immune system,” *Front Immunol,* vol. 4, p. 451, 2013.

[7] J. D. Marth and P. K. Grewal, “Mammalian glycosylation in immunity,” *Nat Rev Immunol,* vol. 8, pp. 874-87, 2008.

[8] M. Sperandio, C. A. Gleissner and K. Ley, “Glycosylation in immune cell trafficking,” *Immunol Rev,* vol. 230, pp. 97-113, 2009.

[9] G. A. Rabinovich and D. O. Croci, “Regulatory circuits mediated by lectin-glycan interactions in autoimmunity and cancer,” *Immunity,* vol. 36, pp. 322-35, 2012.

[10] J. N. Arnold, M. R. Wormald, R. B. Sim, P. M. Rudd and R. A. Dwek, “The impact of glycosylation on the biological function and structure of human immunoglobulins,” *Annu Rev Immunol,* vol. 25, pp. 21-50, 2007.

[11] F. Osorio and C. Reis e Sousa, “ Myeloid C-type lectin receptors in pathogen recognition and host defense,” *Immunity,* vol. 34, pp. 22-37, 2011.

[12] J. J. García-Vallejo and Y. van Kooyk, “Endogenous ligands for C-type lectin receptors: the true regulators of immune homeostasis,” *Immunol Rev,* vol. 230, pp. 22-37, 2009.

[13] L. Lehle, S. Strahl and W. Tanner, “Protein Glycosylation, Conserved from Yeast to Man: A,” *Angew Chem Int Ed Engl,* vol. 45, no. 41, pp. 6802-18, 2006.

[14] P. Messner, K. Steiner, K. Zarschler and C. Schäffer, “S-layer nanoglycobiology of bacteria.,” *Carbohydr Res.,* vol. 343, no. 12, pp. 1934-51, 2008.

[15] R. Leonard, B. O. Petersen, M. Himly, W. Kaar, N. Wopfner, D. Kolarich, R. van Ree, C. Ebner, J. O. Duus, F. Ferreira and F. Altmann, “Two novel types of O-glycans on the mugwort pollen allergen Art v 1 and their role in antibody binding,” *J Biol Chem.,* vol. 280, no. 9, pp. 7932-40, 2005.

[16] L. Wells and G. W. Hart, “O-GlcNAc turns twenty: functional implications for post-translational modification of nuclear and cytosolic proteins with sugars,” *FEBS Lett,* vol. 546, pp. 154-8, 2003.

[17] P. Burda and M. Aebi, “The dolichol pathway of N-linked glycosylation,” *Biochim Biophys Acta,* vol. 1426, no. 2, pp. 239-57, 1999.

[18] M. D. Snider, L. A. Sultzman and P. W. Robbins, “Transmembrane localisation of oligosaccharide-lipid synthesis in microsomal vesicles,” *Cell,* vol. 21, no. 2, pp. 385-92, 1980.

[19] S. Kornfeld, “Trafficking of lysosomal enzymes,” *Faseb J,* vol. 1, no. 6, pp. 462-8, 1987.

[20] S. Wopereis, D. J. Lefeber, E. Morava and R. A. Wevers, “Mechanisms in Protein O-Glycan Biosynthesis and Clinical and Molecular Aspects of Protein O-Glycan Biosynthesis Defects: A Review,” *Clinical Chemistry,* vol. 52, no. 4, pp. 574-600, 2006.

[21] K. G. Ten Hagen, D. T. Tran, T. A. Gerken, D. S. Stein and Z. Zhang, “Functional Characterization and Expression Analysis of Members of the UDP-GalNAc:Polypeptide N-Acetylgalactosaminyltransferase Family from Drosophila melanogaster,” *J Biol Chem,* vol. 278, no. 37, pp. 35039-35048, 2003.

[22] J. D. Marth, “Complexity in O-linked oligosaccharide biosynthesis engendered by multiple polypeptide N-acetylgalactosaminyltransferases,” *Glycobiology,* vol. 6, no. 7, pp. 701-705, 1996.

[23] A. Herscovics, “Importance of glycosidases in mammalian glycoprotein biosynthesis,” *Biochim Biophys Acta,* vol. 1473, pp. 96-107, 1999.

[24] E. Liebminger, S. Huttner, U. Vavra, R. Fischl, J. Schoberer, J. Grass, C. Blaukopf, G. J. Seifert, F. Altmann, L. Mach and R. Strasser , “Class I alpha-mannosidases are required for N-glycan processing and root development in Arabidopsis thaliana,” *Plant Cell,* vol. 21, no. 12, pp. 3850-67, 2009.

[25] V. Gomord, A.-C. Fitchette, L. Menu-Bouaouiche, C. Saint-Jore-Dupas, C. Plasson, D. Michaud and L. Faye, “Plant-specific glycosylation patterns in the context of therapeutic protein production,” *Plant Biotechnology Journal,* vol. 8, no. 5, p. 564–587, 2010.

[26] E. Staudacher, F. Altmann, I. B. Wilson and L. Maerz, “Fucose in N-glycans: from plant to man,” *Biochim Biophys Acta,* vol. 1473, no. 1, pp. 216-36, 1999.

[27] K. Julenius, A. Molgaard, R. Gupta and S. Brunak, “Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites,” *Glycobiology,* vol. 15, no. 2, pp. 153-164, 2004.

[28] J. E. Nettleship, Structural Biology of Glycoproteins, Glycosylation, D. S. Petrescu, Ed., InTech, 2012.

[29] T. Ju, Q. Zheng and R. D. Cummings, “Identification of core 1 O-glycan T-synthase from Caenorhabditis elegans,” *Glycobiology,* vol. 16, no. 10, pp. 947-958, 2006.

[30] T. Ju, V. I. Otto and R. D. Cummings, “The Tn Antigen—Structural Simplicity and Biological Complexity,” *Angew. Chem. Int. Ed.,* vol. 50, p. 1770 – 1791, 2011.

[31] K. Prydz and K. T. Dalen, “Synthesis and sorting of proteoglycans,” *Journal of Cell Science,* vol. 113, pp. 193-205, 2000.

[32] Q. Cao, Q. Zhao, X. Lv, X. Li, Y. Zhao, B. Peng, W. Ying and X. Qian, “A strategy integrating stepped fragmentation and glycan diagnostic ion-based spectrum refinement for the identification of core fucosylated glycoproteome using mass spectrometry,” *Anal Chem,* vol. Just accepted manuscript, 2014.

[33] M. Takahashi, Y. Kuroki, K. Ohtsubo and N. Taniguchi, “Core fucose and bisecting GlcNAc, the direct modifiers of the N-glycan core: their functions and target proteins,” *Carbohydrate Research,* vol. 344, no. 12, pp. 1387-1390, 2009.

[34] N. Taniguchi, S. Ihara, T. Saito, E. Miyoshi, Y. Ikeda and K. Honke, “Implication of GnT-V in cancer metastasis: a glycomic approach for identification of a target protein and its unique function as an angiogenic cofactor.,” *Glycoconj J,* vol. 18, no. 11-12, pp. 859-65, 2001.

[35] M. Pabst, J. S. Bondii, J. Stadlmann, L. March and F. Altmann, “Mass + Retention Time = Structure: A Strategy forthe Analysis of N-Glycans by Carbon LC-ESI-MS and Its Application to Fibrin N-Glycans,” *Anal. Chem.,* vol. 79, no. 13, pp. 5051-7, 2007.

[36] D. M. Sheeley, B. M. Merrill and L. C. Taylor, “Characterization of monoclonal antibody glycosylation: comparison of expression systems and identification of terminal alpha-linked galactose,” *Anal Biochem,* vol. 247, no. 1, pp. 102-10, 1997.

[37] J. J. van Roojien, J. P. Kamerling and J. F. Vliegenthart, “Sulfated di-, tri- and tetraantennary N-glycans in human Tamm-Horsfall glycoprotein,” *Eur. J. Biochem,* vol. 256, pp. 471-487, 1998.

[38] N. Kawasaki, Y. Haishima, M. Ohta, S. Itoh, M. Hyuga, S. Hyuga and T. Hyakawa, “Structural analysis of sulfated N-linked oligosaccharides in erythropoietin.,”

Glycobiology, vol. 11, no. 12, pp. 1043-9, 2001.

[39] A. Varki, "N-glycolylneuraminic acid deficiency in humans," *Biochimie,* vol. 83, pp. 615-622, 2001.

[40] D. J. Becker and J. B. Lowe, "Fucose: biosynthesis and biological function in mammals," *Glycobiology,* vol. 13, no. 7, pp. 41R-53R, 2003.

[41] D. Kolarich, A. Weber, M. Pabst, J. Stadlmann, W. Teschner, Ehrlich H, H. P. Schwarz and F. Altmann, "Glycoproteomic characterization of butyrylcholinesterase from human plasma," *Proteomics,* vol. 8, pp. 254-263, 2008.

[42] P. C. Pang, B. Tissot, E. Z. Drobnis, P. Sutovsky, H. R. Morris, G. F. Clark and A. J. Dell, "Expression of Bisecting Type and Lewis x / Lweis y Terminated N-Glycans on Human Sperm," *J. Biol. Chem.,* vol. 282, pp. 36593-36602, 2007.

[43] M. Nonaka, H. Imaeda, S. Matsumoto, M. B. Yong, N. Kawasaki, E. Mekata, A. Andoh, Y. Saito, T. Tani, Y. Fujiyama and T. Kawasaki, "Mannan-binding protein, a C-type serum lectin, recognizes primary colorectal carcinomas through tumor- associated Lewis glycans," *J. Immunol,* vol. 192, no. 3, pp. 1294-301, 2014.

[44] K. Drickamer, "Two Distinct Classes of Carbohydrate-recognition Domains in Animal Lectin," *J Biol Chem,* vol. 263, no. 20, pp. 9557-9560, 1988.

[45] K. Drickamer and M. E. Taylor, "Biology of Animal Lectins," *Annu Rev Cell Biol,* vol. 9, pp. 237-64, 1993.

[46] T. Angata and E. Brinkman-Van der Linden, "I-type lectins," *Biochim Biophys Acta,* vol. 1572, no. 2-3, pp. 294-316, 2002.

[47] B. B. Agrawal and I. J. Goldstein, "Specific binding of concanavalin A to cross-linked dextran gels," *Biochem J,* vol. 96, pp. 23-25, 1965.

[48] Y. Kaneko, F. Nimmerjahn and J. V. Ravetch, "Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation," *Science,* vol. 313, pp. 670-674, 2006.

[49] M. A. Daniels, K. A. Hogquist and C. S. Jameson, "Sweet 'n' sour: the impact of differential glycosylation on T cell differentiation," *Nat Immunol,* vol. 3, pp. 903-910, 2002.

[50] Y. Reisner , L. Itzicovitch, A. Meshorer and N. Sharon, "Hemopoietic stem cell transplantation using mouse bone marrow and spleen cells fractionated by lectins," *Proc. Natl. Acad. Sci. U.S.A.,* vol. 75, pp. 2933-2936, 1978.

[51] Y. Reisner, N. Kapoor, D. Kirkpatrick, M. S. Pollack, S. Cunningham-Rundles, B. Dupont, M. Z. Hodes, R. A. Good and R. J. O´Reilly, "Transplantation for severe combined immunodeficiency with HLA-A,B,D,DR-incompatible parental marrow cells fractionated

by soybean agglutinin and sheep red blood cells.," *Blood,* vol. 61, pp. 341-348, 1983.

[52] N. Sharon , "Lectins: past, present and future.," *Biochem Soc Trans,* vol. 36, no. 6, pp. 1457-60, 2008.

[53] R. L. Hudgin, W. E. Pricer Jr, G. Ashwell, R. J. Stockert and A. G. Morell, "The isolation and properties of a rabbit liver binding protein specific for asialoglycoproteins," *J. Biol. Chem,* vol. 249, pp. 5536-5543, 1974.

[54] K. Ley, "The role of selectins in inflammation and disease," *Trends Mol. Med,* vol. 9, pp. 263-268, 2003.

[55] Y. van Kooyk and G. A. Rabinovich, "Protein-glycan interactions in the control of innate and adaptive immune responses.," *Nat Immunol,* vol. 9, no. 6, pp. 593-601, 2008.

[56] L. R. Ruhaak, S. Miyamoto and C. B. Lebrilla, "Developments in the identification of glycan biomarkers for the detection of cancer.," *Mol Cell Proteomics,* vol. 12, no. 4, pp. 846-855, 2013.

[57] I. A. Wilson, J. J. Skehel and D. C. Wiley, "Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution.," *Nature,* vol. 289, no. 5796, pp. 366-73, 1981.

[58] D. C. Wiley and J. J. Skehel, "The Structure and Function of the Hemaglutinin Membran Glycoprotein of Influenza Virus," *Annu. Rev. Biochem.,* vol. 56, pp. 365-394, 1987.

[59] M. N. Matrosovich, A. S. Gambaryan, S. Teneberg, V. E. Piskarev, S. S. Yamnikova, D. K. Lvov, J. S. Robertson and K. A. Karlsson, "Avian Influenza A Viruses Differ from Human Viruses by Recognition of Sialyloligosaccharides and Gangliosides and by a Higher Conservation of the HA Receptor-Binding Site," *Virology,* vol. 233, pp. 224-234, 1997.

[60] A. S. Gambarayan, A. B. Tuzikov, V. E. Piskarev, S. S. Yamnikova, D. K. Lvov, J. S. Robertson, N. V. Bovin and M. N. Matrosovich, "Specification of receptor-binding phenotypes of influenza virus isolates from different hosts using synthetic sialylglycopolymers: non-egg-adapted human H1 and H3 influenza A and influenza B viruses share a common high binding affinity for 6'-sialyl(N-ace," *Virology,* vol. 232, no. 2, pp. 345-50, 1997.

[61] K. Shinya, M. Ebina, S. Yamada, M. Ono, N. Kasai and Y. Kawaoka, "Avian flu: Influenza virus receptors in the human airway," *Nature,* vol. 440, pp. 435-436, 2006.

[62] A. Gambaryan, S. Yamnikova, S. Lvov, A. Tuzikov, A. Chinarev, G. Pazynina, R. Webster, M. Matrosovich and N. Bovin, "Polymer-bound 6' sialyl-N-acetyllactosamine protects mice infected by influenza virus.," *Virology,* vol. 334, pp. 476-283, 2005.

[63] J. J. Skehel and D. C. Wiley, "Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin.," *Annu Rev Biochem,* vol. 69, pp. 531-569, 2000.

[64] R. Xu, R. McBride, C. M. Nycholat, J. C. Paulson and I. A. Wilson, “Structural Characterization of the Hemagglutinin Rezeptor Specificity from the 2009 H1N1 Influenza Pandemic,” *J Virol,* vol. 86, no. 2, pp. 982-990, 2012.

[65] S. Van Poucke, J. Uhlendorff, Z. Wang, V. Billiau, J. Nicholls, M. Matrosovich and K. Van Reeth, “Effect of receptor specificity of A/Hong Kong/1/68 (H3N2) influenza virus variants on replication and transmission in pigs,” *Influenza Other Respir Viruses. ,* vol. 7, no. 2, pp. 151-9, 2012.

[66] R. J. Gartner, C. T. Davis, C. A. Russel, B. Shu, S. Lindstrom, A. Balish, W. M. Session, X. Xu, E. Skepner, V. Deyde, M. Okomo-Adhiambo, L. Gubareva, J. Barnes, C. B. Smith, S. L. Emery, M. J. Hillman, P. Rivailler, J. Smagala, M. de Graaf , D. F. Burke, R. A. Fouchier, C. Pappas, C. M. Alpuche-Aranda, H. Lopez-Gatell, H. Olivera, I. Lopez, C. A. Myers, D. Faix, P. J. Blair, C. Yu, K. M. Keen, P. D. Dotson Jr, D. Boxrud, A. R. Sambol, S. H. Abid, K. S. Georg, T. Bannerman, A. L. Moore, D. J. Stringer, P. Blevins, G. J. Demmler-Harrison, M. Ginsberg, P. Kriner, S. Waterman, S. Smole, H. F. Guevara, E. A. Belongia, P. A. Clark, S. T. Beatrice, R. Donis, J. Katz, L. Finelli, C. B. Bridges, M. Shaw, D. B. Jernigan, T. M. Uyeki, D. J. Smith, A. I. Klimov and N. J. Cox, “Antigenic and Genetic Characteristics of Swine-Origine 2009 A(H1N1) Influenza Viruses Circulating in Humans,” *Science,* vol. 325, pp. 197-201, 2009.

[67] J. B. Lowe, “The blood group specific human glycosylatransferases,” *Baillieres Clin Haematol,* vol. 6, pp. 465-492, 1993.

[68] J. B. Lowe, “Selectin ligands, leukocyte trafficking, and fucosyltransferase genes.,” *Kidney Int,* vol. 51, pp. 1418-1826, 1997.

[69] T. A. Springer, “Traffic signals for lymphocyte recirculation and leukocyte emigration: the multistep paradigm,” *Cell,* vol. 76, pp. 301-314, 1994.

[70] L. V. Hooper and J. I. Gordon, “Glycans as legislators of host-microbial interactions: spanning the spectrum from symbiosis to pathogenicity,” *Glycobiology,* vol. 11, pp. 1R-10R, 2001.

[71] R. Kleene, H. Yang, M. Kutsche and M. Schachner, “The neural recognition molecule L1 is a sialic acid binding lectin for CD24, which induces promotion and inhibition of neurite outgrowth,” *J Biol Chem,* vol. 276, no. 24, pp. 21656-63, 2001.

[72] C. Bleckmann, H. Geyer, V. Reinold, A. Lieberoth, M. Schachner, R. Kleene and R. Geyer, “Glycomic Analysis of N-Linked Carbohydrate Epitopes from CD24 of Mouse Brain,” *Journal of Proteome Research,* vol. 8, pp. 567-582, 2009.

[73] H. E. Murrey, S. B. Ficarro, C. Krishnamurthy, S. E. Domino, E. C. Peters and L. C. Hsieh-Wilson, “Identification of the Plasticity-Relevan Fucose-α(1-2)-Galactose Proteom from the Mouse Olfactory Bulb,” *Biochemistry,* vol. 48, pp. 7261-7270, 2009.

[74] H. E. Murrey, C. I. Gama, S. A. Kalovidouris, W. I. Luo, E. M. Driggers, B. Porton and L. C. Hsieh-Wilson, “Protein fucosylation regulates synapsin Ia/Ib expression and neuronal morphology in primary hippocampal neurons,” *PNAS,* vol. 103, no. 1, pp. 21-26, 2006.

[75] K. Moriwaki and E. Miyoshi, “Fucosylation in gastrointestinal cancer,” *World J Hepatol,* vol. 2, no. 4, pp. 151-161, 2010.

[76] L. Aalberg, C. Randall Clark and J. DeRuiter, “Chromatographic and Mass Spectral Studies on Isobaric and Isomeric Substances Related to 3,4-Methylenedioxymethamphetamine,” *Journal of Chromatographic Science,* vol. 42, pp. 464-469, 2004.

[77] S. Natsuka and S. Hase, “Analysis of N- and O-glycans by pyridylamination,” *Methods Mol Bio,* vol. 78, pp. 101-13, 1996.

[78] L. Royle , C. M. Radcliffe, R. A. Dwek and P. M. Rudd, “Detailed structural analysis of N-glycans released from glycoproteins in SDS-PAGE gel bands using HPLC combined with exoglycosidase array digestions,” *Methods Mol Biol,* vol. 347, pp. 125-43, 2006.

[79] R. A. Dwek, C. J. Edge, D. J. Harvey, M. R. Wormald and R. B. Parekh, “Analysis of Glycoprotein-Associated Oligosaccharides,” *Annu Rev Biochem,* vol. 62, pp. 65-100, 1993.

[80] P. M. Rudd, C. Colominas, L. Royle, N. Murphy, E. Hart, A. H. Merry, H. F. Hebestreit and R. A. Dwek, “A high-performance liquid chromatography based strategy for rapid, sensitive sequencing of N-linked oligosaccharide modifications to proteins in sodium dodecyl sulphate polyacrylamide electrophoresis gel bands.,” *Proteomics,* vol. 1, no. 2, pp. 285-94, 2001.

[81] C. W. Scutton , J. A. O'Neill and J. S. Cottrell, “Site-specific characterization of glycoprotein carbohydrates by exoglycosidase digestion and laser desorption mass spectrometry.,” *Anal Biochem,* vol. 218, no. 1, pp. 34-46, 1994.

[82] L. Royle, T. S. Mattu, E. Hart, J. I. Langridge, A. H. Merry, N. Murphy, D. J. Harvey, R. A. Dwek and P. M. Rudd, “An Analytical and Structural Database Provides a Strategy for Sequencing O-Glycans from Microgram Quantities of Glycoproteins,” *Anal Biochem,* vol. 304, no. 1, pp. 70-90, 2002.

[83] A. M. Bielik and J. Zaia, “Historical overview of glycoanalysis,” *Mathods Mol Biol,* vol. 600, pp. 9-30, 2010.

[84] J. M. Prien, B. D. Prater and S. L. Cockrill, “A multi-method approach toward de novo glycan characterization: a Man-5 case study.,” *Glycobiology,* vol. 20, no. 5, pp. 629-47, 2010.

[85] D. J. Harvey, “Fragmentation of negative ions from carbohydrates. Part 2. Fragmentation of high-mannose N-linked glycans,” *J Am Soc Mass Spectrom,* vol. 16,

pp. 631-646, 2005.

[86] D. J. Harvey, R. A. Dwek and P. M. Rudd, “Determining the structure of glycan moieties by mass spectrometry.,” *Curr Protoc Protein Sci,* vol. 43, pp. 12.17.1-12.17.18, 2006.

[87] D. J. Harvey and P. M. Rudd, “Fragmentation of negative ions from N-linked carbohydrates. Part 5. Anionic N-linked glycans,” *Int J Mass Spectrom,* vol. 305, pp. 120-130, 2011.

[88] A. V. Everest-Dass, J. L. Abrahams, Kolarich, D. Kolarich, N. H. Packer and M. P. Campbell, “Structural Feature Ions for Distinguishing N- and O-Linked Glycan Isomers by LC-ESI-IT MS/MS,” *J. Am. Soc. Mass Spectrom.,* vol. 24, pp. 895-906, 2013.

[89] W. T. Morgan and M. Watkins W, “Genetic and bichemical aspects of human blood-group A-, B-, H-, Le-a- and Le-b-specificity,” *Br Med Bull,* vol. 25, pp. 30-34, 1969.

[90] S. Iida, H. Misaka, M. Inoue, M. Shibata, R. Nakano, N. Yamane-Ohnuki, M. Wakitani, K. Yano, K. Shitara and M. Satoh, “Nonfucosylated therapeutic IgG1 antibody can evade the inhibitory effect of serum immunoglobulin G on antibody dependent cellular cytotoxicity through its high binding to FcgammaRIIIa,” *Clin Cancer Res,* vol. 12, pp. 2879-2887, 2006.

[91] N. Takahashi, “Glycoamidases,” in *CRC Handbook of Endoglycosidases and Glycoamidases*, CRC Press Boca Raton, FL, 1992, pp. 183-198.

[92] T. Suzuki, H. Park and W. J. Lennarz, “Cytoplasmic peptide:N-glycanase (PNGase) in eukaryotic cells: occurrence, primary structure, and potential functions,” *FASEB,* vol. 16, no. 7, pp. 635-641, 2002.

[93] T. P. Patel and R. B. Parekh, “Release of oligosaccharides from glycoproteins by hydrazinolysis.,” *Methods Enzymol,* vol. 230, pp. 57-66, 1994.

[94] M. Pabst, J. Grass, R. Fischl, R. Leonard, C. Jin, G. Hinterkorner, N. Borth and F. Altmann, “Nucleotide and nucleotide sugar analysis by liquid chromatography - electronspray ionization - mass spectrometry on surface conditioned porous graphitic carbon,” *Anal Chem,* vol. 82, no. 23, pp. 9782-8, 2010.

[95] Y. Huang, Y. Mechref and M. V. Novotny, “Microscale nonreductive release of O-linked glycans for subsequent analysis through MALDI mass spectrometry and capillary electrophoresis.,” *Anal Chem,* vol. 73, no. 24, pp. 6063-9, 2001.

[96] W. Chai, T. Feizi, C. T. Yuen and A. M. Lawson, “Nonreductive release of O-linked oligosaccharides frim mucin glycoproteins for structure/funktion assignments as neoglycolipids: application in the detection of novel ligands for E-selectin,” *Glycobiology,* vol. 7, no. 6, pp. 861-72, 1997.

[97] A. H. Merry, D. C. Neville, L. Royle, B. Matthews, D. J. Harvey , R. A. Dwek and P. M.

Rudd, “Recovery of intact 2-aminobenzamide-labeled O-glycans released from glycoproteins by hydrazinolysis.,” *Anal Biochem,* vol. 304, no. 1, pp. 91-9, 2002.

[98] J.-R. Brisson, E. Vinogradov, D. J. McNally, N. H. Khieu, I. C. Schoenhofen, S. M. Logan and H. Jarrell, “The Application of NMR Spectroscopy to Functional Glycomics,” *Methods Mol Biol,* vol. 600, pp. 155-73, 2010.

[99] L. Wang, R. D. Cummings, D. F. Smith, M. Huflejt, C. T. Campbell, J. C. Gildersleeve, J. Q. Gerlach, M. Kilcoyne, L. Joshi, S. Serna, N.-C. Reichardt, N. P. Pera, R. J. Pieters, W. Eng and L. K. Mahal, “Cross-platform comparison of glycan microarray formats,” *Glycobilogy,* vol. 24, no. 6, pp. 507-17, 2014.

[100] T. K. Dam and C. F. Brewer, “Multivalent protein-carbohydrate interactions: isothermal titration microcalorimetry studies,” *Methods Enzymol,* vol. 379, pp. 107-128, 2004.

[101] Y. Shinohara, F. Kim , M. Shimizu, M. Goto, M. Tosu and Y. Hasegawa, “Kinetic measurement of the interaction between an oligosaccharide and lectins by a biosensor based on surface plasmon resonance,” *Eur J Biochem,* vol. 223, no. 1, pp. 189-194, 1994.

[102] J. Hirabayashi, T. Hashidate and Y. Arata, “Oligosaccharide specificity of galectins: a search by frontal affinity chromatography,” *Biochim Biophys Acta,* vol. 1572, no. 2-3, pp. 232-254, 2002.

[103] Z. Y. Zhang, I. R. Ollmann, X. S. Ye, R. Wischnat, T. Baasiv and C. H. Wong, “Programmable one-pot oligosaccharide synthesis,” *J Am Chem Soc,* vol. 121, no. 4, pp. 734-753, 1999.

[104] P. Sears and C. H. Wong, “Toward automated synthesis of oligosaccharides and glycoproteins,” *Science,* vol. 291, no. 5512, pp. 2344-2350, 2001.

[105] O. J. Plante, E. R. Palmacci and P. H. Seeberger, “Automated solid-phase synthesis of oligosaccharides,” *Science,* vol. 291, no. 5508, pp. 1523-1527, 2001.

[106] “CFG funtional glycomics gateway,” Consortium for Functional Glycomics, 2001. [Online]. Available: http://www.functionalglycomics.org/. [Accessed 24 06 2014].

[107] A. R. de Boer, C. H. Hokke, A. M. Deelder and M. Wuhrer, “General microarray technique for immobilization and screening of natural glycans,” *Anal Chem,* vol. 79, no. 21, pp. 8107-8113, 2007.

[108] C.-H. Liang and C.-Y. Wu, “Glycan array: a powerful tool for glycomic studies,” *Expert Rev Proteomics,* vol. 6, no. 6, pp. 631-645, 2009.

[109] A. van Diepen, C. H. Smit, L. van Egmond, N. B. Kabatereine, A. Pinot de Moria, D. W. Dunne and C. H. Hokke, “Differential Anti-Glycan Antibody Responses in Schistosoma mansoni-Infected Children and Adults Studied by Shotgun Glycan Microarray,” *PLoS*

Negl Trop Dis, vol. 6, no. 11, p. e1922, 2012.

[110] W. Kong, L. Liu, Y. Wang, H. Gao, K. Wei, Y. Sun, J. Liu, G. Ma and J. Pu, "Hemagglutinin mutation D222N of the 2009 pandemic H1N1 influenza virus alters receptor specificity without affecting virulence in mice," *Virus Res,* vol. 189C, pp. 79-86, 2014.

[111] T. Ichimiya, S. Nishihara, S. Takase-Yoden, H. Kida and K. Aoki-Kinoshita, "Frequent glycan structure mining of influenza virus data revealed a sulfated glycan motif that increased viral infection," *Bioinformatics,* vol. 30, no. 5, pp. 706-11, 2014.

[112] Y. H. Ahn, J. Y. Kim and J. S. Yoo, "Quantitative mass spectrometric analysis of glycoproteins combined with enrichment methods," *Mass Spectrom Rev,* vol. Epub ahead of print, 2014.

[113] H. Tateno, A. Kuno and J. Hirabayashi, "How to Determin Specificity: From Lectin Profiling to Glycan Mapping and Arrays," in *The Sugar Code: Fundamentals in Glycoscience*, Weinheim, Wiley-VCH Verlag GmbH & Co. KGaA, 2009, pp. 247-253.

[114] A. S. Powlesland, A. Quintero-Martinez, P. G. Lim, Z. Pipirou, M. E. Taylor and K. Drickamer, "Engineered carbohydrate-recognition domains for glycoproteomic analysis of cell surface glycosylation and ligands for glycan-binding receptors," *Methods Enzymol,* vol. 480, pp. 165-79, 2010.

[115] J. M. Prien , J. Ashline D, A. J. Lapadula, H. Zhang and V. N. Reinhold, "The high mannose glycans from bovine bibonuclease B isomer characterization by ion trap MS," *J Am Soc Mass Spectrom,* vol. 20, no. 4, pp. 539-56, 2009.

[116] L. R. Ruhaak, G. Zauner, C. Huhn, C. Bruggink, A. M. Deelder and Wuhrer M, "Glycan labeling strategies and their use in identification and quantification," *Anal Bioanal Chem,* vol. 397, pp. 3457-3481, 2010.

[117] M. Tswett, "Physikalisch-Chemische Studien über das Chlorophyll. Die Adsorption," *Ber Deutsch Bot Ges,* vol. 24, pp. 316-326, 1906.

[118] G. A. Howard and A. J. Martin, "The Separation of the C12-C18 Fatty Acids by Reversed-phase Partition Chromatography," *Biochem J,* vol. 46, no. 5, pp. 532-8, 1950.

[119] H. Engelhardt, "One century of liquid chromatography. From Tswett's columns to modern high speed and high performance separations," *J Chromatogr B Analyt Technol Biomed Life Sci,* vol. 800, no. 1-2, pp. 3-6, 2004.

[120] R. Hayes, A. Ahmed, T. Edge and H. Zhang, "Core–shell particles: Preparation, fundamentals and applications in high performance liquid chromatography," *J Chromatogr A,* vol. Epub ahead of print, 2014.

[121] J. Zaia, "Mass Spectrometry and Glyomics," *OMICS,* vol. 14, no. 4, pp. 401-418, 2010.

[122] S. Hase, S. Hara and Y. Matsushima, “Tagging of sugars with a fluorescent compound, 2-aminopyridine,” *J Biochem,* vol. 85, no. 1, pp. 217-20, 1979.

[123] M. Pabst , D. Kolarich, G. Pöltl , T. Dalik, G. Lubec, A. Hofinger and F. Altmann, “Comparison of fluorescent labels for oligosaccharides and introduction of a new postlabeling purification method,” *Anal Biochem,* vol. 384, no. 2, pp. 263-73, 2009.

[124] K. R. Anumula, “Advances in fluorescence derivatization methods for high-performance liquid chromatographic analysis of glycoprotein carbohydrates,” *Anal Biochem,* vol. 350, no. 1, pp. 1-23, 2006.

[125] A. Kondo, J. Suzuki, N. Kuraya, S. Hase, I. Kato and T. Ikenaka, “Improved method for fluorescence labeling of sugar chains with sialic acid residues,” *Agric Biol Chem,* vol. 54, no. 8, pp. 2169-70, 1990.

[126] F. Higel, U. Demelbauer, A. Seidl, W. Friess and F. Sörgel, “Reversed-phase liquid-chromatographic mass spectrometric N-glycan analysis of biopharmaceuticals,” *Anal Bioanal Chem,* vol. 405, no. 8, pp. 2481-93, 2013.

[127] A. J. Alpert, “Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compound,” *J Chromatogr,* vol. 499, pp. 177-96, 1990.

[128] G. Zauner, A. M. Deelder and M. Wuhrer, “Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics,” *Electrophoresis,* vol. 32, no. 24, pp. 3456-66, 2011.

[129] M. Wuhrer, A. R. de Boer and A. M. Deelder, “Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry,” *Mass Spectrom Rev,* vol. 28, pp. 192-206, 2009.

[130] R. R. Townsend, P. H. Lipniunas, C. Bigge, A. Ventom and R. Parekh, “Multimode high-performance liquid chromatography of fluorescently labeled oligosaccharides from glycoproteins,” *Anal Biochem,* vol. 239, no. 2, pp. 200-7, 1996.

[131] J. Zaia, “Mass spectrometry and the emerging field of glycomics,” *Chem Biol,* vol. 15, no. 9, pp. 881-92, 2008.

[132] J. Ahn, J. Bones, Y. Q. Yu, P. M. Rudd and M. Gilar, “Separation of 2-aminobenzamide labeled glycans using hydrophilic interaction chromatography columns packed with 1.7 µm sorbent,” *J Chromatogr,* vol. 1216, pp. 3252-3259, 2009.

[133] Y. Takegawa, K. Deguchi, T. Keira, H. Ito, H. Nakagawa and S.-I. Nichimura, “Separation of isomeric 2-aminopyridine derivatized N-glycans and N-glycopeptides of human serum immunoglobulin G by using a zwitterionic type of hydrophilic-interaction chromatography,” *J Chromatogr A,* vol. 29, no. 16, pp. 2533-40, 2006.

[134] Y. Yamaguchi, W. Nishima, S. Re and Y. Sugita, “Confident identification of isomeric N-glycan structures by combined ion mobility mass spectrometry and hydrophilic interaction liquid chromatography,” *Rapid Commun Mass Spectrom,* vol. 51, no. 41, pp. 8125-31, 2012.

[135] J. H. Knox and M. T. Gilbert, “Preparation of Porous Carbon”. United Kingdome Patent UK Patent 7939449, US Patent 4263268, FRG Patent P2946688-4, 20 Nov 1979.

[136] [Online]. Available: http://www.thermo.com/eThermo/CMA/PDFs/Product/productPDF_28871.pdf.

[137] T. Hanai, “Separation of polar compounds using carbon columns,” *J Chromatogr A,* vol. 989, no. 2, pp. 183-96, 2003.

[138] M. Melmer, T. Stangler, A. Premstaller and W. Lindner, “Comparison of hydrophilic-interaction, reversed-phase and porous graphitic carbon chromatography for glycan analysis,” *J Chromatogr A,* vol. 1218, pp. 118-123, 2011.

[139] M. Pabst and F. Altmann, “Influence of electrosorption, solvent, temperature and ion polarity on the performance of LC-ESI-MS using graphitic carbon for acidic oligosaccharides,” *Anal Chem,* vol. 80, no. 19, pp. 7534-42, 2008.

[140] M. Pabst, J. Grass, S. Toegel, E. Liebminger, R. Strasser and F. Altmann, “Isomeric analysis of oligomannosidic N-glycans and their dolichol-linked precursors,” *Glycobiology,* vol. 22, no. 3, pp. 389-399, 2012.

[141] S. Kamoda, R. Ishikawa and K. Kakehi, “Capillary electrophoresis with laser induced fluorescence detection for detailed studies on N-linked oligosaccharide profile of therapeutic recombinant monoclonal antibodies,” *J Chromatogr A,* vol. 1133, no. 1-2, pp. 332-9, 2006.

[142] A. F. Zamfir, C. Flangea, A. Serb, A. M. Zagrean, A. M. Rizzi and E. Sisu, “Separation and identification of glycoforms by capillary electrophoresis with electrospray ionization mass spectrometric detection,” *Methods Mol Biol,* vol. 951, pp. 145-69, 2013.

[143] Z. Szabo, A. Guttman, T. Rejtar and B. L. Karger, “Improved sample preparation method for glycan analysis of glycoproteins by CE-LIF and CE-MS,” *Electrophoresis,* vol. 31, no. 8, pp. 1389-95, 2010.

[144] M. Pabst and F. Altmann, “Glycan analysis by modern instrumental methods,” *Proteomics,* vol. 11, no. 4, pp. 631-43, 2011.

[145] D. J. Harvey, “Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: an update for 2007-2008,” *Mass Spectrom Rev,* vol. 31, no. 2, pp. 183-311, 2012.

[146] F. Hillenkamp and J. P. Katalinic, “Chapter 2. MALDI Mass Spectrometry Instrumentation,” in *MALDI MS: A Practical Guide to Instrumentation, Methods and Applications*, Weinheim, Germany, Wiley-VCH Verlag GmbH & Co KGaA, 2007.

[147] B. Domon and R. Aebersold, “Mass spectrometry and protein analysis,” *Science,* vol. 312, no. 5771, pp. 212-7, 2006.

[148] M. Wuhrer, A. M. Deelder and Y. E. van der Burgt, “Mass spectrometric glycan rearrangements,” *Mass Spectrom Rev,* vol. 30, no. 4, pp. 664-80, 2011.

[149] s. aldrich, “sigma aldrich,” [Online]. Available: http://www.sigmaaldrich.com/technical-documents/articles/biofiles/glycoprofi-le-labelling.html.

[150] O. Oyelaran, L. M. McShane, L. Dodd and J. C. Gildersleeve, “Profiling human serum antibodies with a carbohydrate antigen microarray,” *J Proteome Res,* vol. 8, no. 9, pp. 4301-4310, 2009.

[151] R. C. Team, *R. A language and environement for statistical computing,* Vienna: R Foundation for Statistical Computing, 2013.

[152] R. A. van den Berg, H. C. Hoefsloot, J. A. Westerhuis, A. K. Smilde and M. J. van der Werf, “Centering, scaling, and transformations: improving the biological information content of metabolomics data.,” *BMC Genomics,* vol. 7, no. 1, p. 142, 2006.

[153] W. Morell and J.-C. Michalski, “Analysis of protein glycosylation by mass spectrometry,” *Nature Protocols,* vol. 2, pp. 1585-1602, 2007.

[154] J. Stadlmann, M. Pabst, D. Kolarich , R. Kunert and F. Altmann, “Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides,” *Proteomics,* vol. 8, no. 14, pp. 2858-2871, 2008.

[155] X. Zhang, J. M. Asara, J. Adamec and M. Ouzzani, “Data preprocessing in liquid chromatography-mass spectrometry based proteomics,” *Bioinformatics,* vol. 21, pp. 4054-4059, 2005.

[156] R. Zeleny, R. Leonard, G. Dorfner, T. Dalik, D. Kolarich and F. Altmann, “Molecular cloning and characterization of a plant a1,3/4-fucosidase based on sequence tags from almond fucosidase I,” *Phytochemistry,* vol. 67, pp. 641-648, 2006.

[157] B. Bioscience. [Online]. Available: https://www.bdj.co.jp/pdf/55-02_03-7900030-23-A1.pdf.

[158] Summers, M. D., Smith, G. E.,, “A manual of methods for baculavirus vectors and insect cell culture procedures,” *TX Ag. Expt. Stn. Bull.,* p. 1555, 1987.

[159] N. Sriwilaijaroen, S. Kondo, H. Yagi, N. Takemae, T. Saito, H. Hiramatsu, K. Kato and Y. Suzuki, “N-Glycans from Porcine Trachea and Lung: Predominant NeuAca2-6Gal Could

Be a Selective Pressure for Influenza Variants in Favor of Human-Type Receptor," *Plos one,* vol. 6, no. 2, p. e16302, 2011.

[160] L. Aalberg, C. R. Clark and J. deRiuter, "Chromatographic and Mass Spectral Studies on Isobaric and Isomeric Substances Related to 3,4-Methylenedioxymethamphetamine," *Journal of Chromatographic Science,* vol. 42, pp. 465-469, 2004.

[161] A. R. de Boer, C. H. Hokke, A. M. Deelder and M. Wuhrer, "General Microarray Technique for Immobilization and Screening of Natural Glycans," *Anal. Chem.,* vol. 79, no. 21, pp. 8107-8113, 2007.

[162] J. C. Bigge, T. P. Patel, J. A. Bruce, P. N. Goulding, S. M. Charles and R. B. Parekh, "Nonselective and Efficient Fluorescent Labeling of Glycans Using 2-Amino Benzamide and Anthranilic Acid," *Analytical Biochemistry,* vol. 230, pp. 229-238, 1995.

[163] L. R. Ruhaak, C. huhn, W.-J. Waerreus, A. R. de Boer, C. H. Hokke, A. M. Deelder and M. Wuhrer, "Hydrophilic Interaction Chromatography-Based High-Throughput Sample Preparation Method for N-Glycan Analysis from Total Human Plasma Glycoproteins," *Anal. Chem.,* vol. 20, no. 15, pp. 6119-6126, 2008.

[164] "ProGlycan," [Online]. Available: http://www.proglycan.com/index.php?page=pga_nomenclature.

[165] Thermo Scientific, [Online]. Available: http://www.dionex.com/en-us/webdocs/115143-AN-LC-GlycanPac-AXR1-Glycans-Glycoproteins-AN20910_E.pdf.

[166] F. Krammer, I. Margine, G. S. Tan, N. Pica, J. C. Krause and P. Palese , "A carboxy-terminal trimerization domain stabilizes conformational epitopes on the stalk domain of soluble recombinant hemagglutinin substrates.," *Plos One,* vol. 7, no. 8, p. e43603, 2012.

[167] S. D. Pawar, S. S. Parkhi, S. S. Koratkas and A. C. Mishra, "Receptor specificity and erythrocyte binding preferences of avian influenza viruses isolated from India," *Virol J,* vol. 9, p. 251, 2012.

[168] E. Suenaga, H. Mizuno and P. K. Kumar, "Influenza virus surveillance using surface plasmon resonance.," *Virulance,* vol. 3, no. 5, pp. 464-70, 2012.

[169] N. Deshpande, P. H. Jensen, N. H. Packer and D. Kolarich, "GlycoSpectrumScan: Fishing Glycopeptides from MS Spectra of Protease Digests of Human Colostrum sIgA," *Journal of Proteome Research,* vol. 9, no. 2, pp. 1063-1075, 2010.

[170] Y. J. Chen, D. R. Wing, G. R. Guile , R. A. Dwek, D. J. Harvey and S. Zamze, "Neutral N-glycans in adult rat brain tissue Complete characterisation reveals fucosylated hybrid and complex structures," *Eur. J. Biochem,* vol. 251, pp. 691 - 703, 1998.

[171] M. H. Chung, B. Mirakhur, E. Chan, Q. T. Le, J. Berlin, M. Morse, B. A. Murphy, S. M.

Santinover, J. Hosen, D. Mauro, R. J. Slebos, Q. Zhou, D. Gold, T. Hatley, D. J. Hicklin and T. A. Platts-Mills, “Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose.,” *N Engl J Med.,* vol. 358, no. 11, pp. 1109-1117, 2008.

[172] S. Narasimhan, “UDP-GlcNAc:Glycopeptide b4-N-acetylglucosaminyltransferase III, an enzyme in hen oviduct which adds GlcNAc in b1-4 linkage to the b-linked Mannose of the trimannosyl core of N-Glycosyl Oligosaccharides,” *J.Biol. Chem.,* vol. 257, no. 17, pp. 102350-10242, 1982.

[173] K. Maiese, Z. Z. Chong, F. Li and Y. C. Shang, “Erythropoietin: elucidating new cellular targets that broaden therapeutic strategies,” *Prog Neurobiol,* vol. 85, pp. 194-213, 2008.

[174] S. Dube, J. W. Fisher and J. S. Powell, “Glycosylation at specific sites of erythropoietin is essential for biosynthesis, secretion, and biological function,” *J Biol Chem,* vol. 263, pp. 17516-17521, 1988.

[175] C. H. Hokke, A. A. Bergwerff, G. W. Van Dedem, J. P. Kamerling and J. F. Vliegenthart, “Structural analysis of the sialylated N- and O-linked carbohydrate chains of recombinant human erythropoietin expressed in Chinese hamster ovary cells. Sialylation patterns and branch location of dimeric N-acetyllactosamine units.,” *Eur J Biochem,* vol. 228, p. 981–1008, 1995.

[176] C. T. Yuen, P. L. Storring, R. J. Tiplady, M. Izquierdo and R. Wait, “Relationships between the N-glycan structures and biological activities of recombinant human erythropoietins produced using different culture conditions and purification procedures,” *Br J Haematol,* vol. 121, pp. 511-526, 2003.

[177] A. J. Bitonti, J. A. Dumont, S. C. Low, R. T. Peters and K. E. Kropp, “Pulmonary delivery of an erythropoietin Fc fusion protein in non-human primates through an immunoglobulin transport pathway,” *Proc Natl Acad Sci U S A,* vol. 101, pp. 9763-9768, 2004.

[178] A. Castilho, L. Neumann, P. Gattinger, R. Strasser , K. Vorauer-Uhl, T. Sterovsky, F. Altmann and H. Steinkellner, “Generation of Biologically Active Multi-Sialylated Recombinant Human EPOFc in Plants,” *PLoS ONE,* vol. 8, no. 1, p. e54836, 2013.

[179] S. Elliott, J. Egrie, J. Browne, T. Lorenzini, L. Busse, N. Rogers and I. Ponting, “Control of rHuEPO biological activity. The role of carbohydrate,” *Exp. Hematol.,* vol. 32, pp. 1146-1155, 2004.

[180] A. Castilho, L. Neumann, S. Daskalova, H. S. Mason, H. Steinkellner, F. Altmann and R. Strasser, “Engineering of Sialylated Mucin-type O-Glycosylation in Plants,” *J. Biol. Chem.,* vol. 287, pp. 36518-36526, 2012.

[181] C.-F. Mandenius, R. Wang, A. Alden, G. Bergström, S. Thebault, C. Lutsch and S. Ohlson, “Monitoring of influenza virus hemagglutinin in process samples using weak affinity

ligands and surface plasmon resonance," *Acta Anal Chim,* vol. 623, pp. 66-75, 2008.

[182] A. Geissner, C. Anish and P. H. Seeberger, "Glycan arrays as tools for infectious disease research.," *Curr Opin Chem Biol,* vol. 18, pp. 38-45, 2014.

[183] O. Oyelaran and J. C. Gildersleeve, "Glycan arrays: recent advances and future challenges," *Curr Opin Chem Biol.,* vol. 13, no. 4, pp. 406-413, 2009.

[184] R. P. de Vries, E. de Vries, B. J. Bosch, R. J. de Groot, P. J. Rottier and C. A. de Haan, "The influenza A virus hemagglutinin glycosylation state affects receptor-binding specificity.," *Virology,* vol. 403, no. 1, pp. 17-25, 2010.

[185] M. Melmer, T. Stangler, A. Premstaller and W. Lindner, "Effects of the redox state of porous graphitic carbon on the retention of oligosaccharides.," *J Chromatogr A,* vol. 1217, pp. 6097-6101, 2010.

[186] R. M. Bill, L. Revers and I. B. Wilson, "Branching out: Constructing The Antennae Of N-Linked Sugars," in *Protein Glycosyaltion*, Dordrecht, Kluwer Academic Publishers, 1998, pp. 236-237.

[187] J. Stadlmann, A. Weber, M. Pabst, H. Anderle, R. Kunert, H. J. Ehrlich, H. P. Schwarz and F. Altmann, "A close look at human IgG sialylation and subclass distribution after lectin fractionation," *Proteomics,* vol. 9, pp. 4143-4153, 2009.

[188] C. M. Chu, "Inactivation of haemagglutinin and infectivity of influenza and Newcastle disease viruses by heat and by formalin.," *J Hyg,* vol. 46, pp. 247-251, 1948.

[189] J. C. Paulson , J. E. Sadler and R. L. Hill, "Restoration of specific myxovirus receptors to asialoerythrocytes by incorporation of sialic acid with pure sialyltransferases.," *J Biol Chem,* vol. 254, p. 2120–2124, 1979.

[190] Y. Kaneko, F. Nimmerjahn and J. V. Ravetch, "Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation," *Science,* vol. 313, pp. 670-673, 2006.

7 List of Figures

8 List of Tables

9 Selected publications and posters

Appendix I

Castilho, A., **Neumann, L.**, Gattinger, P., Strasser, R., Vorauer-Uhl, K., Sterovsky, T., Altmann, F., Steinkellner, H. (2013) Generation of Biologically Active Multi-Sialylated Recombinant Human EPOFc in Plants. PLOS ONE 2013;8(1):e54836. doi: 10.1371/journal.pone.0054836.

Appendix II

Castilho, A., **Neumann, L**., Daskalova, S., Mason, H. S., Steinkellner, H., Altmann, F., Strasser, R. (2012) Engineering of Sialylated Mucin-type O-glycosylation in Plants. *J. Biol. Chem*. 287(43):36518-26. doi:10.1074/jbc.M112.402685

Appendix III

Schneider, J.D., Castilho, A., **Neumann, L.**, Altmann, F., Loos, A., Kannan, L., Mor, T.S., Steinkellner, H. (2013) Expression of human butyrylcholinesterase with a glycosylation profile resembling the plasma derived orthologue. Biotechnol. J. 9, 501-510. doi: 10.1002/biot.201300229.

Appendix IV

Liebminger, E., Grass, J., Jez, J., **Neumann, L.**, Altmann, F., Strasser, R. (2012) Myrosinases TGG1 and TGG2 from Arabidopsis thaliana contain exclusively oligomannosidic N-glycans. *Phytochemistry* 84:24-30. doi: 10.1016/j.phytochem.2012.08.023

Appendix V

Tomek, M., **Neumann, L.,** Nimeth, I., Koerdt, A, Andesner, P., Mach, L., Messner, P**.,** Potempa, J., Schaeffer, C. (2014) S-layer glycoproteins of Tannerella forsythia are secreted via a type IX secretion system that is decoupled from protein glycosylation. Mol. Oral Microbiol., doi: 10.1111/omi.12062

Appendix VI

Posch, G., Pabst, M., **Neumann, L.**, Coyne, M. J., Altmann, F., Messner, P., Comstock, L. E., Schäffer, C. (2012) "Cross-glycosylation" of proteins in Bacteriodales species. *Glycobiology*, 23(5):568-77. doi: 10.1093/glycob/cws172

Appendix VII

Anzengruber, J., Pabst, M., **Neumann, L.**, Sekot, G., Heinl, S., Grabherr, R., Altmann, F.,Messner, P., Schaeffer, C. (2014) Protein *O*-glucosylation in *Lactobacillus buchneri*. *Glycoconj. J.* 31, 117-131. doi: 10.1007/s10719-013-9505-7

Appendix VIII

Hernandez Bort, J. A., Shanmukam, V., Pabst, M., Windwarder, M., **Neumann, L.**, Alchalabi, A., Krebiehl, G, Koellensperger, G., Hann, S., Sonntag, D., Altmann, F., Heel, C., Borth, N. (2014) Reduced quenching and extraction time for mammalian cells using filtration and syringe extraction. J. Biotechnol., doi: 10.1016/j.jbiotec.2014.04.014.

Appendix IX

Jung, G., Pabst, M., **Neumann, L.**, Berger, A., Lubec, G. (2012) Characterization of α-l-Iduronidase (Aldurazyme®) and its complexes. *J Proteomics*, doi:10.1016/j.jprot.2012.09.022

Appendix X

Neumann, L., Thader, A., Wozny, K., Pabst, M., Altmann, F., (2013): Isomer specific analysis of mouse brain N-glycans by LC-MS (Talk) [APRS-Meeting 2013, Innsbruck, Austria, September 23-25, 2013]

Appendix XI

Neumann, L., Krammer, F., de Graaf, M., Hokke, C.H., Altmann, F., (2013): Binding Specificity of the H1N1 Swine-Origin Influenza A Virus (S-OIV) (Poster), [Glyco 22, Dalian, China, June 22-29, 2013]

Appendix XI

Neumann, L., Thader, A., Altmann, F., Pabst, M. (2012): Comparison of N-Glycans of different Influenza Virus-susceptible tissues using PGCLC-ESI-MS and stable isotope labelled internal standards (Poster)[Summer Course Glycoscience, 12th European Training Course on Carbohydrates, Groningen, The Netherlands, June 3-7, 2012]

Appendix XII

Neumann, L., Thader, A., Altmann, F., Pabst, M., (2012): Structural investigation of fucose epitopes on N-Glycans by PGC-ESI-MS using the example of secretory IgA and mouse brain glycans, (Poster) [Carbohydrate analysis and glycomics: where next?, London, UK, March 21, 2012]

Appendix I

Castilho, A., **Neumann, L.**, Gattinger, P., Strasser, R., Vorauer-Uhl, K., Sterovsky, T., Altmann, F., Steinkellner, H. (2013) Generation of Biologically Active Multi-Sialylated Recombinant Human EPOFc in Plants. PLOS ONE 2013;8(1):e54836. doi: 10.1371/journal.pone.0054836.

OPEN ACCESS Freely available online

PLOS ONE

Generation of Biologically Active Multi-Sialylated Recombinant Human EPOFc in Plants

Alexandra Castilho[1], Laura Neumann[2], Pia Gattinger[1], Richard Strasser[1], Karola Vorauer-Uhl[3], Thomas Sterovsky[4], Friedrich Altmann[2], Herta Steinkellner[1]*

1 Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Vienna, Austria, **2** Department of Chemistry, University of Natural Resources and Life Sciences, Vienna, Austria, **3** Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria, **4** Polymun Scientific GmbH, Klosterneuburg, Austria

Abstract

Hyperglycosylated proteins are more stable, show increased serum half-life and less sensitivity to proteolysis compared to non-sialylated forms. This applies particularly to recombinant human erythropoietin (rhEPO). Recent progress in *N*-glycoengineering of non-mammalian expression hosts resulted in *in vivo* protein sialylation at great homogeneity. However the synthesis of multi-sialylated *N*-glycans is so far restricted to mammalian cells. Here we used a plant based expression system to accomplish multi-antennary protein sialylation. A human erythropoietin fusion protein (EPOFc) was transiently expressed in *Nicotiana benthamiana ΔXTFT*, a glycosylation mutant that lacks plant specific N-glycan residues. cDNA of the hormone was co-delivered into plants with the necessary genes for (i) branching (ii) β1,4-galactosylation as well as for the (iii) synthesis, transport and transfer of sialic acid. This resulted in the production of recombinant EPOFc carrying bi- tri- and tetra-sialylated complex *N*-glycans. The formation of this highly complex oligosaccharide structure required the coordinated expression of 11 human proteins acting in different subcellular compartments at different stages of the glycosylation pathway. *In vitro* receptor binding assays demonstrate the generation of biologically active molecules. We demonstrate the *in planta* synthesis of one of the most complex mammalian glycoforms pointing to an outstanding high degree of tolerance to changes in the glycosylation pathway in plants.

Citation: Castilho A, Neumann L, Gattinger P, Strasser R, Vorauer-Uhl K, et al. (2013) Generation of Biologically Active Multi-Sialylated Recombinant Human EPOFc in Plants. PLoS ONE 8(1): e54836. doi:10.1371/journal.pone.0054836

Editor: M. Lucrecia Alvarez, TGen, United States of America

Received September 10, 2012; **Accepted** December 19, 2012; **Published** January 25, 2013

Funding: This work was supported by the Laura Bassi Centres of Expertise (Grant Number 822757) and the Wiener Wissenschafts-, Forschungs- und Technologie fonds (WWTF) (Grant Number L575-B13). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Thomas Sterovsky is employed by Polymun Scientific GmbH. There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

* E-mail: herta.steinkellner@boku.ac.at

Introduction

Recombinant human erythropoietin (rhEPO) was the first hematopoietic growth factor approved to treat anemia associated with kidney failure, cancer and other pathological conditions [1]. Mature EPO is a 30 kDa glycoprotein with 166 amino acids carrying three *N*-linked (Asn-24, -38 and -83) and one O-linked (Ser-126) carbohydrate chains which account for 40% of the total molecular weight [2,3]. Glycosylation has a profound effect in maintaining the overall stability and *in vivo* hematopoietic activity of hEPO [4–6]. Several studies report that terminal sialic acid increases the circulatory half- life of rhEPO, moreover a positive correlation between the *in vivo* biological activity and the ratio of tetra- to bi-antennary sialylated oligosaccharides was shown [7,8]. Due to the complexity of the glycosylation pattern, therapeutic rhEPO is exclusively produced in mammalian cell cultures, mainly in Chinese hamster ovary (CHO) [9–11]. Many efforts have been made to improve the sialylation content of the hormone [12–14]. Indeed, hyper-sialylated rhEPOs with prolonged half-life and subsequent enhanced drug efficacy were produced [6]. Another strategy to improve drug efficacy of rhEPO is its fusion to stabilizing peptides/proteins. The application of immunoglobulin Fc-fusions to therapeutic proteins has become very popular since the Fc fragment can extend the conjugated protein serum half-life by being recycled via the neonatal Fc receptor (FcRn). EPOFc fusions have been successfully explored in this direction [15].

The limited production capacity and expensive mammalian cell based production facilities make the recombinant hormone very costly. A viable alternative for the large-scale and low cost production of biopharmaceuticals is the use of plants [16,17]. Recent progress in expression levels, production speed and up-scaling, have placed this expression system into an encouraging position. Another important feature of using plants as production platform is their ability to carry out human-like complex N-glycosylation. Due to their comparable small repertoire of glycosylation reactions, plants carry out complex *N*-glycosylation with remarkable homogeneity, which makes them especially amenable for *N*-glycoengineering. Indeed, over the past years many research groups have concentrated their efforts on modulating plant *N*-glycosylation to enable the production of recombinant proteins with human-like structures (Review [18,19] One of the most impressive results is the introduction of the mammalian biosynthetic pathway for *in planta* protein sialylation [20].

Previous attempts to produce rhEPO in plants resulted in the generation of a recombinant hormone that shows *in vitro* activity

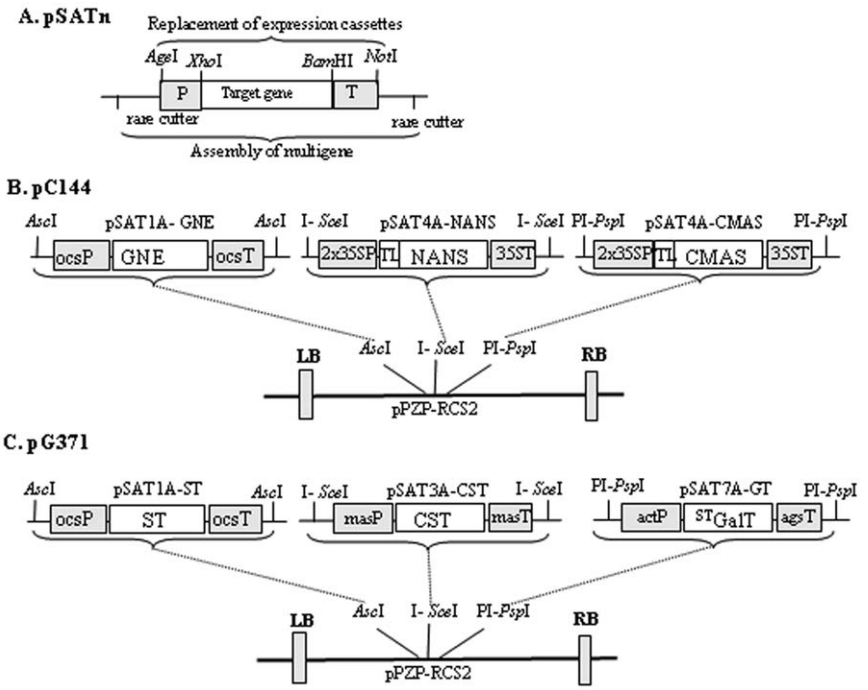


Figure 1. Schematic representation of the multi-gene vectors used in this investigation. A. Structural features of the pSAT series of vectors (pSATn) suitable for the expression of target genes under the control of various constitutive promoters and terminators (pSAT1A, pSAT3A, pSAT4A and pSAT7A). Expression cassettes are interchangeable within pSATn as *Age*I-*Not*I fragments. Rare-cutting enzymes flanking each pSAT vector are used to transfer the expression cassettes into the expression vector pPZP-RCS2. B. Outline of the cloning strategy to assemble the mammalian genes necessary for the synthesis of sialic acid, pC144. The GNE, NANS and CMAS [20] open reading frames were subcloned into pSAT auxiliary vectors and were then sequentially assembled in pPZP-RCS2 using specific rare-cutting enzymes. C. Outline of the cloning strategy to assemble the mammalian genes acting in the Golgi apparatus for *in planta* protein sialylation, pG371. STGalT, CST and ST [20] genes were put under control of different promoters and terminators in pSAT vectors. These were then sequentially assembled into pPZP-RCS2 vector using appropriate rare-cutting enzymes. 35SP: cauliflower mosaic virus (CaMV) 35S promoter; TL: translational enhancer 5′-UTR from tobacco etch; 35ST: CaMV 35S terminator; OcsP: octopine synthase promoter; OcsT: octopine synthase terminator; actP: actin promoter; agsT: agropin synthase terminator; masP: manopine synthase promoter masT: manopine synthase terminator; GNE: mouse UDP-*N*-acetylglucosamine-2-epimerase/*N*-acetylmannosamine kinase; NANS: Homo sapiens *N*-acetylneuraminic acid phosphate synthase; CMAS: Homo sapiens CMP-*N*-acetylneuraminic acid synthase; STGalT: β1,4-galactosyltransfease fused to the cytoplasmic tail, transmembrane domain and stem region of the rat α2,6-sialyltransferase; CST: Mouse CMP-sialic acid transporter; ST: rat α2,6-sialyltransferase; LB: left border; RB: right border.
doi:10.1371/journal.pone.0054836.g001

[21–23]. However, plant-derived rEPO was not active *in vivo* most probably due the lack of sialylation [24]. Regrettably, most of these studies did not consider the glycosylation status of the recombinant hormone. rhEPO and rhEPOFc produced in glycoengineered moss and *N. benthamiana* carried mainly human type complex $GlcNAc_2Man_3GlcNAc_2$ (GnGn) structures, lacking plant specific xylose and fucose [25,26]. Moreover the production of rhEPOFc with tetra-antennary bisected complex *N*-glycans was achieved upon overexpression of mammalian *N*-acetylglucosaminyltransferases (GnTIII, GnTIV and GnTV) [26,27]. Overall, the results demonstrate the feasibility of plants to generate active rhEPOFc with a targeted N-glycosylation profile, however plant derived (multi-) sialylated rhEPOFc remains elusive.

In this investigation we set out to express in plants rhEPOFc carrying tri- and tetra-sialylated N-glycans. Agrobacterium containing rhEPOFc cDNA was delivered to *N. benthamiana* ΔXTFT mutants (lacking the plant specific *N*-glycan residues β1,2-xylose and core α1,3-fucose) together with the mammalian genes required for *in planta* protein sialylation (i.e. 6 genes, [20]). *N*-glycosylation profiling of the recombinant protein using LC-ESI-MS exhibited the synthesis of mainly complex bi-antennary sialylated *N*-glycans, i.e. $NeuAc_2Gal_2GlcNAc_2Man_3GlcNAc_2$ (NaNa). Transient co-expression of rhEPOFc with mammalian genes necessary for the branching and sialylation of *N*-glycans (in total 11 genes) resulted in the synthesis of rhEPOFc decorated with tri- and tetra-sialylated oligosaccharides. All glycoforms exhibit biological activities comparable to the CHO derived rhEPOFc, as determined by cell-based receptor binding assays.

Materials and Methods

Vectors for single gene expression

MagnICON pro-vector system was used for the expression of rhEPOFc chimeric proteins as described before [26]. For modulation of rhEPOFc N-glycosylation profile we used the previously described binary vectors each carrying a single gene necessary to produced multi-antennary *N*-glycans (^{FUT11}GnTIV and ^{FUT11}GnTV, [26]) and to assemble *in planta* the metabolic pathway for protein sialylation (GNE, NANS, CMAS, CST, STGalT and ST, [20]).

Binary vectors for multiple gene expression

We combined six expression cassettes in two different binary plasmids: one for the expression of the genes necessary for the synthesis of sugar-activated sialic acid, CMP-Neu5Ac (GNE, NANS and CMAS) and another for the expression of genes

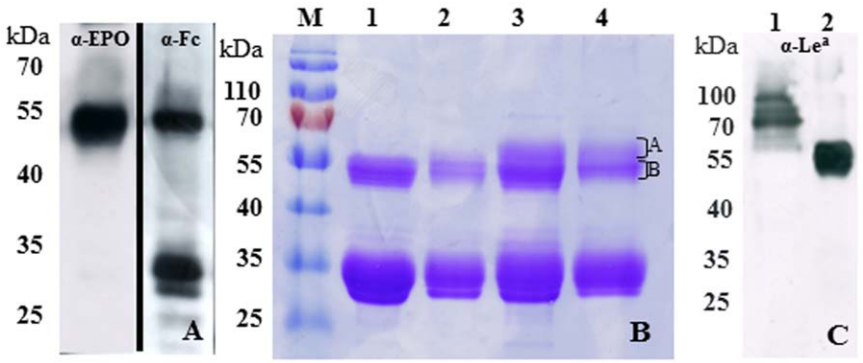


Figure 2. Expression of rhEPOFc in *N. benthamiana*. A. Western blot analysis of total soluble proteins extracted from *N. benthamiana* expressing rhEPOFc. 55 kDa protein reacts to both anti-EPO (α-EPO) and anti-Fc (α-Fc) antibodies while the ~30 kDa band reacts only with α-Fc antibodies. B. Protein A purified rhEPOFc fractionated by SDS PAGE and stained with Coomassie-brilliant blue R-250. lane 1: rhEPOFc expressed in *N. benthamiana* mutants lacking plant specific β1,2-xylose and α1,3-fucose ($rhEPOFc_{\Delta XTFT}$); lane 2: rhEPOFc co-expressed with mammalian genes for protein sialylation (GNE, NANS, CMAS, CST, STGalT and ST) ($rhEPOFc_{Sia}$); lane 3: rhEPOFc co-expressed with mammalian genes necessary for sialylation and synthesis of tri-antennary *N*-glycans GnTIV or GnTV, ($rhEPO_{TriSia}$); lane 4: rhEPOFc co-expressed with mammalian genes for sialylation and synthesis of tetra-antennary *N*-glycans, GnTIV and GnTV ($rhEPO_{TetraSia}$). A and B represent distinct protein fractions from the 55 kDa band of $rhEpoFc_{TriSia}$ and $rhEPO_{TetraSia}$, used for N-glycan analysis; the ~30 kDa band represent free Fc. C. Western blot analysis of total soluble proteins (5 μg TSP) extracted from *N. benthamiana* ΔXTFT mutants (control; lane 1) and of purified $rhEPOFc_{\Delta XTFT}$ (lane 2) using antibodies against Lewis-A epitopes (JIM 84). Several proteins in TSP and the 55 kDa protein band corresponding to intact rhEPOFc reacted to JIM 84 revealing the presence of *N*-glycans with Lewis-a epitopes. (M) protein marker.
doi:10.1371/journal.pone.0054836.g002

necessary for synthesis of the acceptor substrate (β1,4-galactosylation), Golgi transport and transfer of sialic acid (CST, STGalT and ST). To this intent we used the versatile pSAT family that allows target genes to be cloned under a large choice of promoters and terminators that are easily interchangeable (Fig 1A, [17]). cDNA from each gene were amplified from the correspondent binary vector described in Castilho et al [20]. Appropriate rare-cutting enzymes flanking the expression cassettes in each pSAT vector were used to assemble several cassettes into plant transformation RCS2-based vectors carrying the same rare-cutting enzymes [17]. The pSAT auxiliary vectors and the pPZP-RCS2 binary vectors were purchased from University of Michigan, USA.

Construction of vector for the expression of GNE, NANS and CMAS

The cDNA of GNE was amplified with primers GNE R1/F1 digested with *Xho*I/*BgL*II and cloned into pSAT1A digested *Xho*I/*Bam*HI (pSAT1A-GNE). NANS and CMAS cDNAs were amplified with primers NANS F1/R1 and CMAS F1/R1 respectively, digested with *Xho*I/*Bam*HI and cloned into pSAT4A (pSAT4A-NANS and pSAT4A-CMAS). The expression cassette of pSAT4A-CMAS was transferred into the *Age*I-*Not*I sites of the pSAT6A. To obtain the construct for simultaneous expression of the three proteins the expression cassette of pSAT1A-GNE was removed by *Asc*I digestion and cloned into the *Asc*I site of pPZP-RCS2, the expression cassette from pSAT4A-NANS was removed by I-*Sce*I digestion and cloned into the I-*Sce*I site of pPZP-RCS2 and finally the CMAS expression cassette was inserted into the site PI-*Psp*I of pPZP-RCS2 (pC144, Figure 1B)

Table 1. Expression of rhEPOFc in *N. benthamiana*.

rhEPOFc	mg/kg	% TSP
ΔXTFT	9.13	0.18
Sia	6.14	0.12
TriSia	9.12	0.18
TetraSia	9.11	0.18

Concentration of transiently expressed rhEPOFc was determined using a commercially available immunoassay. For each sample the concentration is given in mg/kg of fresh leaf. The percentage of the total soluble protein (TSP) was also calculated. rhEPOFc was expressed in *N. benthamiana* ΔXTFT mutants (ΔXTFT); co-expressed in ΔXTFT with mammalian genes for protein sialylation (Sia); co-expressed in ΔXTFT with mammalian genes for synthesis of tri-antennary sialylated *N*-glycans (TriaSia) and co-expressed in ΔXTFT with mammalian genes for synthesis of tetra-sialylated *N*-glycans (TetraSia)
doi:10.1371/journal.pone.0054836.t001

Construction of vector for the expression of CST, STGalT and ST

The cDNA from CMP-Neu5Ac transporter was amplified from the correspondent binary vector with the primer pair CST F1/R1, digested with *Xho*I/*Bam*HI and cloned into pSAT3A digested the same way (pSAT3A-CST). cDNA from the modified β1,4-galactosyltransferase (STGalT [28]) was amplified with ST F1/STGalT R1 primers digested and cloned into the *Xho*I-*Bam*HI sites of pSAT7A (SAT7A-GT). Lastly, the cDNA from the rat α2,6-sialyltranferase was amplified with the primer pair STF1/R1 digested with *Xho*I/*Bam*HI and cloned into pSAT1A (pSAT1A-ST). Upon restriction with appropriate rare-cutting enzymes, the pSAT1A-ST, pSAT3A-CST and pSAT7A-GT were assembled into pPZP-RCS2 vector as *Asc*I, I-*Sce*I and PI-*Psp*I fragments, respectively (pG371, Figure 1C).

All binary vectors were transformed into the *Agrobacterium tumefaciens* strain UIA 143 and magnICON constructs were transformed into strain GV3101 pMP90. All primers used in this investigation are listed in Table S1.

Plant material and transient protein expression

Nicotiana benthamiana ΔXTFT plants [29] were grown in a growth chamber at 22°C with a 16 h light/8 h dark photoperiod.

Transient expression of rhEPOFc was done in four-to-five-week old plants by agroinfiltration. The magnICON 3′- vector containing cDNA was co-infiltrated with the corresponding 5- vector carrying the signal peptide for secretion in combination

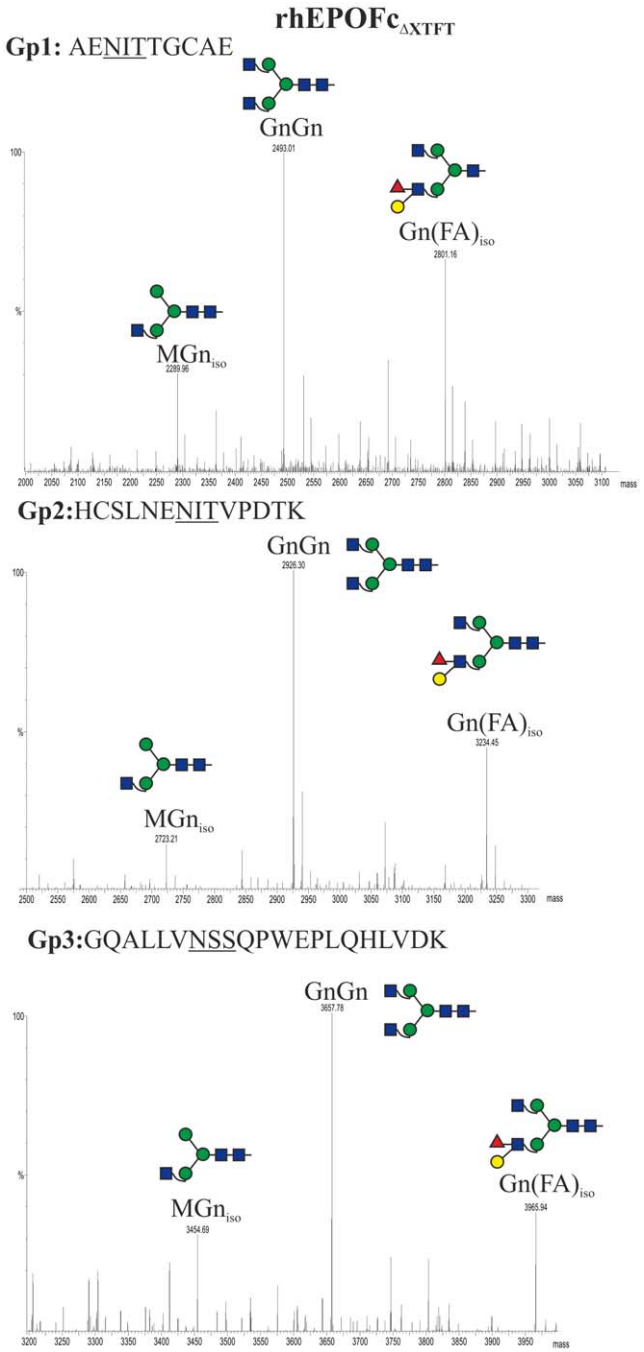


Figure 3. Generation of GnGn structures in rhEPOFc. Mass spectra of trypsin and endoproteinase Glu-C double-digested rhEPOFc expressed in *N. benthamiana* ΔXTFT (rhEPOFc$_{\Delta XTFT}$; Figure 2B, lane 1). Glycosylation patterns of rhEPO glycopeptide 1 (Gp1): E/A^{22}ENITTGCAE^{31}; glycopeptide 2 (Gp2): E/H^{32}CSLNENITVPDTK^{45} and glycopeptide 3 (Gp3): R/G^{77}QALLVNSSQPWEPLQHLVDK^{97} are shown. The corresponding *N*-glycosylation profile of the Fc glycopeptide (R/EEQYNSTYR) is shown in Figure S1. Peak labels were made according to the ProGlycAn system (www.proglycan.com). Illustrations display *N*-glycans on assigned peaks, for interpretation of other assigned glycoforms see Figure S5.
doi:10.1371/journal.pone.0054836.g003

with the binary vector for the expression of the recombinase [30]. For modulation of the *N*-glycosylation profiles, binary vectors containing the cDNA of the different mammalian genes were co-infiltrated with the magnICON viral-based vectors. Agrobacteria carrying the magnICON constructs were infiltrated using optical density (OD_{600}) 0.1 and 0.05 for agrobacteria carrying binary constructs.

Protein purification, N-glycan analysis and peptide mapping

rhEPOFc was purified from agroinfiltrated leaves (200–300 mg) with rProteinA Sepharose^{TM}Fast Flow (GE Healthcare) as described previously [26]. For glycopeptide analysis, purified rhEPOFc were resolved by SDS-PAGE and bands corresponding to 55 kDa were cut out, S-alkylated and double-digested with trypsin and endoproteinase Glu-C. This double digestion allows site-specific analysis of all four *N*-glycosylation sites (GPs): EPO GP1: E/A^{22}ENITTGCAE^{31}; EPO GP2: E/H^{32}CSLNENITVPDTK^{45}, EPO GP3: R/G^{77}QALLVNSSQPWEPLQHLVDK^{97} and Fc glycopeptide: R/EEQYNSTYR. Subsequently samples were analysed by liquid-chromatography electrospray ionization-mass spectrometry, LC-ESI-MS [31,32]. Briefly, a BioBasic C18 column (150×0.32 mm, 5 µm; Thermo Scientific) was eluted with 0.3% formic acid buffered to pH 3.0 with ammonia as the aqueous solvent and a gradient from 10% to 55% acetonitrile developed over 40 min of 1.5 µL/min. The glycoforms of a given peptide co-eluted due to the use of buffered eluent [32]. The elution zone of each peak was summed and the spectra were deconvoluted using MaxEnt3 (Waters Micromass). Peak heights were taken as indicators of the molar ratios of glycoforms, which was recently shown to give meaningful results for Fc-glycopeptides [31].

The 30 kDa protein band corresponding to free Fc was analysed by LC ESI MS/MS for peptide mapping in order to identify the N-terminus. The data was analyzed using the X! Tandem open source software to match tandem mass spectra with the EPO-Fc protein sequence. The N-terminal peptide was identified by the GPM (Global Protein Machine) search engine.

Immunoblot Analysis

Five micrograms of total soluble protein and Protein A purified rhEPOFc were subjected to 12% SDS-PAGE under reducing conditions and blotted onto Hybond Enhanced Chemiluminescence nitrocellulose membranes (GE Healthcare). The blots were blocked in 1xPBS containing 0.1% (v/v) Tween 20 and 3% (w/v) BSA for 1 h and the protein bands were analysed by immunoblotting using either anti-hEPO (1:3000 dilution MAB2871, R&D Systems, Minneapolis, MN), anti-human IgG (1:5000 dilution anti-Fc, W4031 Promega, Mannheim, Germany) or anti-Lewis-A (1:40 dilution JIM84, kindly provided by Paul Knox, University of Leeds, UK) antibodies.

rhEPOFc quantification and in vitro assay

The expression level of plant-derived rhEPOFc was measured in total soluble proteins using the Quantikine IVD ELISA for human EPO (DEPOO, R&D Systems) according to manufacturer's instructions. The biological activity of protein A purified rhEPOFc was measured in a UT-7 cell based proliferation assay. Briefly, the UT-7 cell line [33] was maintained in RPMI 1640 (Biochrome AG) supplemented with 10% fetal calf serum (PAN Biotech.), 4 mM L-glutamine and 5 ng/mL EPO. The cells were washed with EPO free culture medium and incubated for 4 h at 37°C and 7% CO_2. In a 96-well culture plate increasing amounts of CHO-derived rhEPOFc (0.009–60 ng/ml) and plant-derived rhEPOFc (ranging from 0.003–20 ng/mL) were added to 100 µL of medium containing about 10^4 cells. After 4 days at 37°C and 7% CO_2, 10 µL of a MTT (Thiazolyl Blue Tetrazolium Bromide;

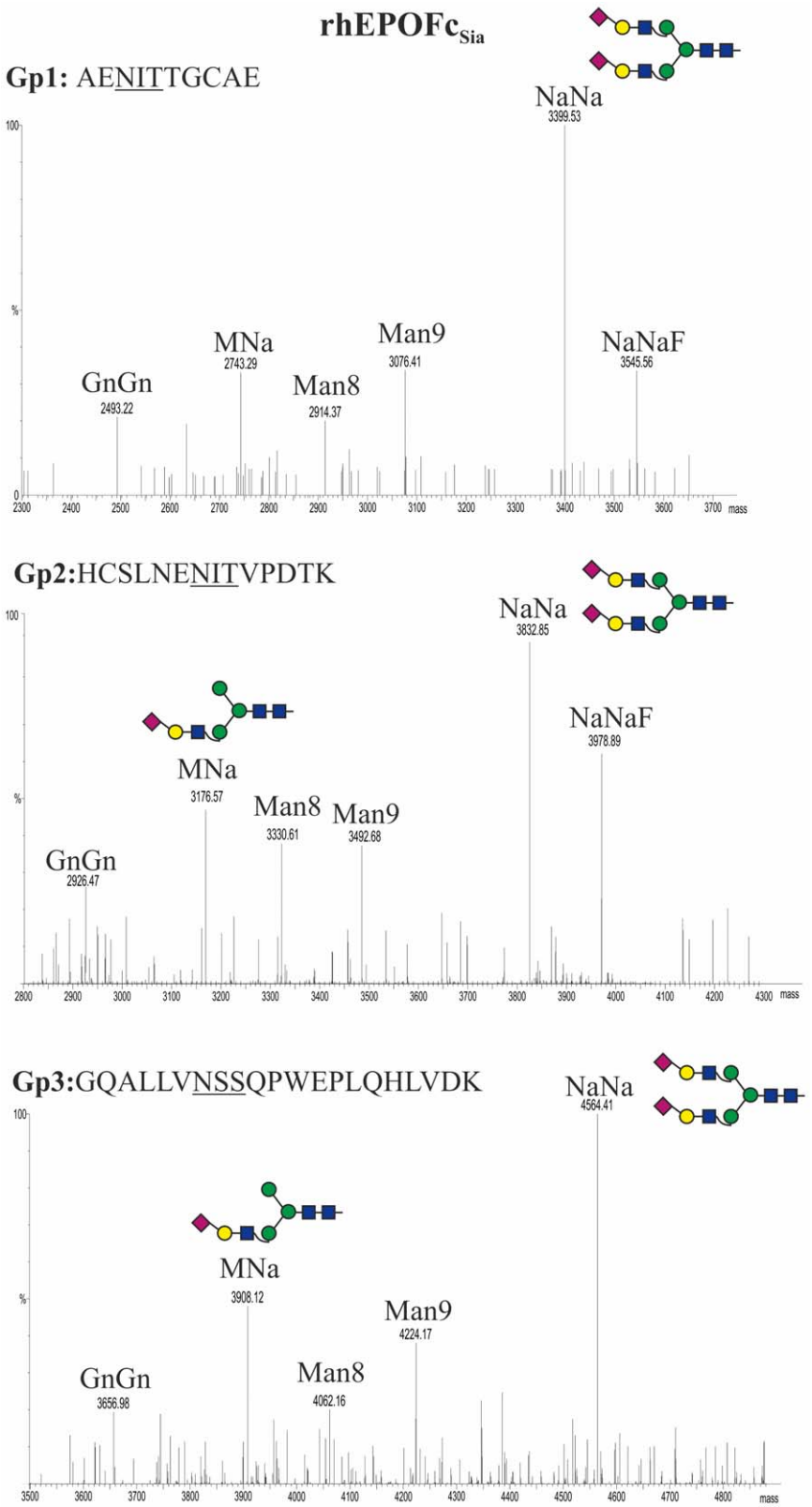


Figure 4. Generation of bi-sialylated structures in rhEPOFc. Mass spectra of trypsin and endoproteinase Glu-C double-digested rhEPOFc co-expressed in *N. benthamiana* ΔXTFT with mammalian genes for protein sialylation (GNE, NANS, CMAS, CST, STGalT and ST) (rhEPOFc$_{Sia}$; Figure 2B, lane 2). Glycosylation patterns of rhEPO Gp1: E/A^{22}ENITTGCAE^{31}; Gp2: E/H^{32}CSLNENITVPDTK^{45} and Gp3: R/G^{77}QALLVNSSQPWEPLQHLVDK^{97} are shown. *N*-glycosylation profile of the Fc glycopeptide is shown in Figure S1. Peak labels were made according to the ProGlycAn system (www.proglycan.com). Illustrations display *N*-glycans on assigned peaks, for interpretation of other assigned glycoforms see Figure S5.
doi:10.1371/journal.pone.0054836.g004

Table 2. Relative abundance of different complex glycoforms detected in rhEPOFc. (oligomannosidic structures that are present in all samples are not included).

	$EpoFc_{Sia}$			$EpoFc_{TriSia}$			$EpoFc_{TetraSia}$		
Glycoform (%)	**Gp1**	**Gp2**	**Gp3**	**Gp1**	**Gp2**	**Gp3**	**Gp1**	**Gp2**	**Gp3**
GnGn	9.1	10.1	10.9	4.8	3.7	6.6	5.6	4.4	3.4
MNa_{iso}	**12.1**	**17.6**	**22.5**	-	-	-	12.3	-	11.2
NaNa	**62**	**49.7**	**66.6**	**4**	**12.4**	**15.3**	**13.5**	**29.4**	**26.9**
NaNaF	**16.8**	**22.6**	-	-	-	-		-	-
[GnGn]Gn				18.8	10.1	10.4	10.7	2	2.1
[GnGn]GnF				-	2.2	-	-	-	-
[AGn]Gn				-	2.1	-	-	-	-
[AGn]GnF				-	1.4	-	-	-	-
[NaNa]Na				**72.4**	**68.1**	**67.7**	**56.3**	**39.6**	**21.5**
[GnGn][GnGn]							1.6	7.7	10.4
$[GnGn][GnA]_{iso}$							-	5.3	9.7
$[GnA][GnA]_{iso}$							-	1.2	1.3
[NaNa][NaNa]							-	**10.4**	**13.5**
$\sum$ Sialylation	**90.9**	**89.9**	**89.1**	**76.4**	**81.5**	**83**	**82.1**	**79.4**	**73.1**

Relative abundance of complex *N*-glycans determined by LC-ESI-MS. **$rhEPOFc_{Sia}$**: rhEPOFc co-expressed with mammalian genes for protein sialylation; **$rhEPOFc_{TriaSia}$**: rhEPOFc co-expressed with mammalian genes for synthesis of tri-antennary sialylated *N*-glycans; **$rhEPOFc_{TetraSia}$:** rhEPOFc co-expressed with mammalian genes for synthesis of tetra-sialylated *N*-glycans. ΔXTFT was used as expression host. Values are in percentages. Quantifications were done for complex *N*-glycans (oligomannosidic structures were not included in calculations). Gp1: glycopeptide 1; Gp2: Glycopeptide 2; Gp3: Glycopeptide 3.
doi:10.1371/journal.pone.0054836.t002

Sigma) solution (5 mg/mL) were supplied to each well and the plate was incubated for 4 h as before. Finally, 100 µL of 10% SDS (in 0.01 M HCl) were added to each well and mixed thoroughly at 37°C before reading absorbance at 570 nm (reference wavelength 690 nm). The experiments were performed in 5 replicates and the results were evaluated using MS Excel Solver. The half maximal effective UT-7 cell proliferation dose (ED_{50}) was used to compare the activities of plant- and CHO-derived rhEPOFc.

Results

Transient expression of EPOFc in N. benthamiana ΔXTFT

We used *N. benthamiana ΔXTFT*, a glycosylation mutant that synthesizes complex *N*-glycans devoid of plant specific β1,2-xylose and core α1,3-fucose, as expression platform [29]. In previous studies we have shown the versatility of these plants for the modulation of plant *N*-glycosylation towards mammalian-like structures (recently reviewed [19]). Using the potent viral-based expression system magnICON, [30] appropriate agrobacteria carrying hEPOFc cDNA were delivered to *ΔXTFT* leaves. 4–5 days post infiltration (dpi) expression was monitored by Western blotting. Antibodies against EPO and Fc enabled the detection of a 55 kDa band which corresponds to the expected size of the fusion protein and an additional 30 kDa band that reacted only with anti Fc antibodies (Figure 2A). The expression level of the intact protein was up to 9 mg/kg leaves, corresponding to 0.2% of total soluble protein (Table 1). rhEPO was purified via protein A-based chromatography and separated by SDS PAGE. Coomassie staining revealed the presence of two bands as already detected by immunoblotting (Figure 2B). Peptide mapping and MS analyses demonstrated that the 55 kDa band corresponds to the intact rhEPOFc, while the 30 kDa band refers to free Fc (data not shown). Similar observations of rhEPOFc fragmentation have been reported in earlier studies in transgenic chickens [34] and in plants [26,27]. In our attempts to enhance the expression of full-length rhEPOFc, different fusion constructs were generated. These included plant codon-optimization of the hEPO fragment using the GeneArt® Gene Synthesis and GeneOptimizer® process (www.lifetechnologies.com, GenBank accession No. KC329647), amino acid variations in the hinge region of Fc, and exchange of the hinge-Fc fragment from IgG1 for the IgGD hinge and the IgG4-Fc regions [35]. Another concern is the post translational elimination of the arginyl (Arg^{166}) amino acid residue. Analysis of the C-terminus of CHO-rhEPO and human EPO purified from the urine demonstrates that the Arg^{166} predicted to be at the C-terminus is missing. This is presumably due to the enzymatic activity of endogenous carboxypeptidases [36]. Since plants contain several types of carboxypeptidases the trimming of Arg^{166} and the consequent loss of tags fused to the C-terminus cannot be excluded. To possibly prevent this eventual cleavage we generated a hEPOFc fusion lacking this amino acid. Unfortunately none of the strategies led to improved expression of full length EPOFc (data not shown). Moreover, the identification of the N-terminus on the free Fc fraction by LC ESI MS/MS was not clear and the results showed that the ~30 kDa band consist of a mixture of Fc fragment fused to varying sizes of EPO sequence. It was therefore not possible to identify an exact cleavage site between the hEPO and the Fc.

LC-ESI-MS analysis was performed to determine the *N*-glycosylation profile of purified rhEPOFc expressed in *ΔXTFT* ($rhEPOFc_{\Delta XTFT}$, Figure 2B, lane 1). MS data revealed that all three *N*-glycosylation sites of rhEPO carry a similar glycosylation pattern (Figure 3), with a major glycoform, GnGn. In addition significant amounts of structures compatible to $Gn(FA)_{iso}$ were present, a carbohydrate formation already detected previously on plant derived rhEPO and rhEPOFc [25,26]. The presence of the terminal trisaccharide consisting of α1,4-fucose and β1,3-galactose linked to *N*-acetylglucosamine also known as Lewis-a epitope can

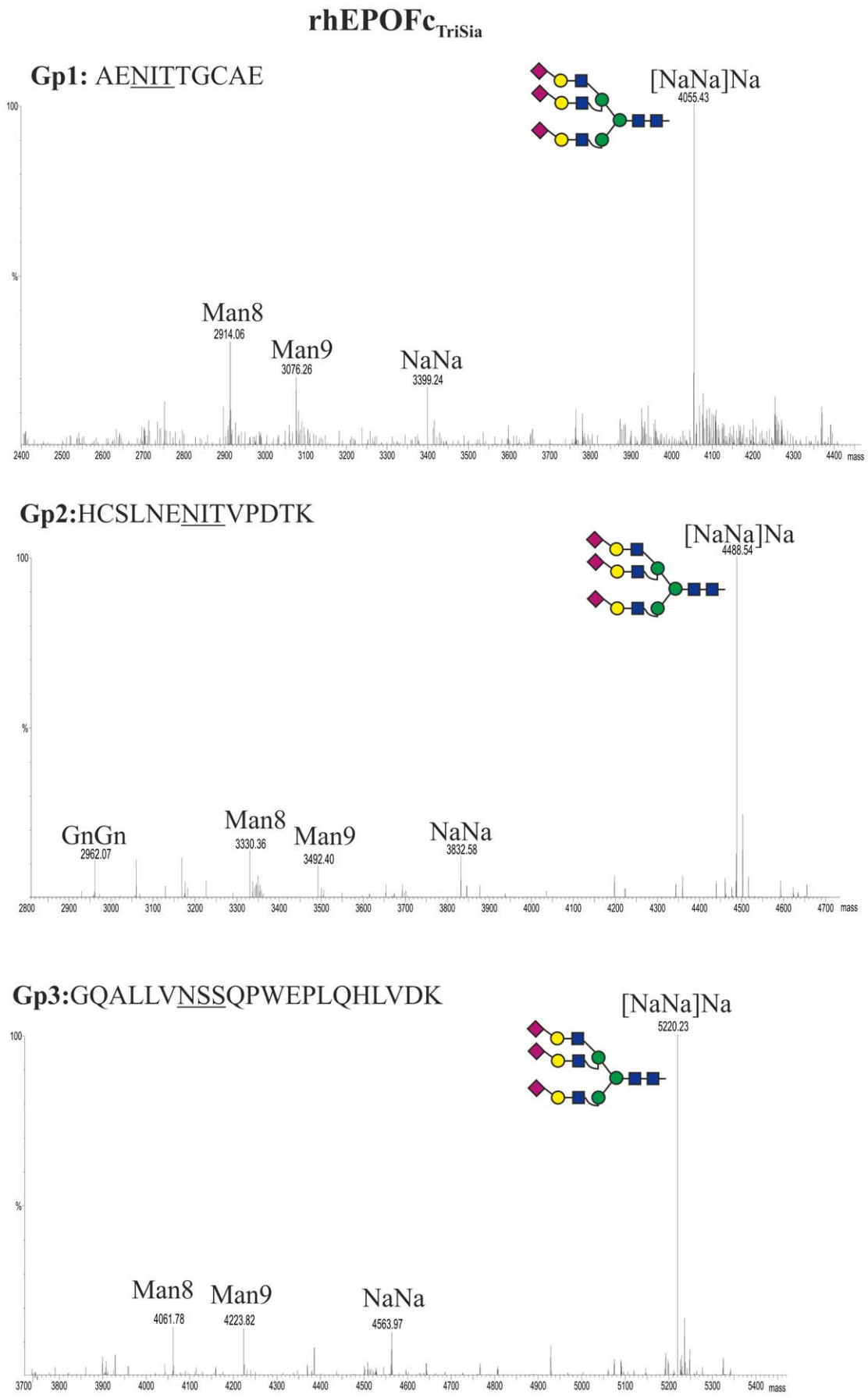

rhEPOFc TriSia
Gp1: AENITTGCAE
[NaNa]Na
4055.43
Man8
2914.06
Man9
3076.26
NaNa
3399.24
Gp2:HCSLNENITVPDTK
[NaNa]Na
4488.54
GnGn
2962.07
Man8
3330.36
Man9
3492.40
NaNa
3832.58
Gp3:GQALLVNSSQPWEPLQHLVDK
[NaNa]Na
5220.23
Man8
4061.78
Man9
4223.82
NaNa
4563.97
mass

Figure 5. Generation of tri-sialylated structures in rhEPOFc. Mass spectra of trypsin and endoproteinase Glu-C double-digested rhEPOFc co-expressed in *N. benthamiana* ΔXTFT with mammalian genes for synthesis of tri-antennary sialylated *N*-glycans (rhEPO$_{TriSia}$). The analysis was performed on rhEPOFc$_{TriSia}$ present on fraction A of the 55kDa band (Figure 2B, lane 3). Glycosylation patterns of rhEPO Gp1: E/A^{22}ENITTGCAE^{31}; Gp2: E/H^{32}CSLNENITVPDTK^{45} and Gp3: R/G^{77}QALLVNSSQPWEPLQHLVDK^{97} are shown. *N*-glycosylation profile of the Fc glycopeptide is shown in Figure S1. Glycosylation profile of rhEPOFc present on fraction B of the 55 kDa band is shown in Figure S2. Peak labels were made according to the ProGlycAn system (www.proglycan.com). Illustrations display *N*-glycans on assigned peaks, for interpretation of other assigned glycoforms see Figure S5.
doi:10.1371/journal.pone.0054836.g005

be detected by immunoreaction to the monoclonal antibody, JIM84 [23,37]. Total soluble proteins (TSP) and protein A purified rhEPOFc$_{\Delta XTFT}$ analysed by Western blotting showed that the 55 kDa band corresponding to the intact rhEPOFc reacts with anti-Lewis-a antibodies, while the free Fc 30 kDa band does not (Figure 2C). In fact, the glycosylation profile of Fc exhibits exclusively GnGn structures (Figure S1).

Multiple gene expression vectors

In previous studies we have shown that *in planta* sialylation can be accomplished by co-infiltration of 6 agrobacteria cultures into a plant leaf (each carries a binary vector with a mammalian glycosylation gene) [20]. To achieve this, all recombinant proteins (including the target protein, which is also co-delivered) must work in the same cell in a highly coordinated fashion. However, the infection of a single cell via agro-infiltration is a random procedure, thus the delivery of single constructs might lead to inefficiencies. To facilitate the simultaneous delivery of all cDNAs to the same cell, two multi gene vectors were generated, each carrying three mammalian glycosylation genes. The pSAT-family vectors allow target genes to be cloned under a large choice of promoters and terminators and the expression cassettes are easily interchangeable (Figure 1A [17]). The six different cDNAs were initially cloned into pSAT vector and subsequently groups of three expression cassettes were assembled in two binary vectors: (i) pC144, carries the genes necessary for the synthesis of nucleotide sugar activated sialic acid, CMP-Neu5Ac (GNE, NANS and CMAS, Figure 1B); (ii) pG371, carries the genes necessary for the synthesis of the β1,4-galactosylated acceptor substrate, Golgi transport and transfer of sialic acid (CST, STGal and ST, Figure 1C). For detailed description of the vectors see Experimental Procedures.

Generation of bi-sialylated N-glycans on rhEPOFc

In order to elongate the GnGn glycoforms present on rhEPOFc$_{\Delta XTFT}$ with β1,4-galactose and α2,6-linked sialic acid, the hormone was co-expressed with the multi gene vectors, pC144 and pG371, allowing a total of 9 genes to be simultaneously delivered to *Δ*XTFT. Site specific *N*-glycosylation-profiling of the purified recombinant hormone (rhEPOFc$_{Sia}$, Figure 2B, lane 2) showed that all *N*-glycosylation sites on the rhEPO are similarly occupied and were efficiently modulated (Figure 4). MS analysis of the 55 kDa band revealed that about 90% of complex *N*-glycans was sialylated (Table 2). Notably, we observed a dominant *N*-linked glycoform, i.e. bi-antennary sialylated structures (NaNa), which accounts for more than 60% of all complex structures. In addition fucosylated (NaNaF) and incompletely sialylated (MNa) glycoforms were detected and about 10–15% of rhEPOFc$_{Sia}$ carried oligomannosidic structures (not included in Table 2). Surprisingly, no Lewis-a structures were detected. In contrast, the *N*-glycan profile of Fc exhibited a largely heterogeneous glycosylation profile, including GnGn, mono and bi-galactosylated structures (GnA, AA), incompletely processed structures (MNa) and oligomannosidic glycoforms (Figure S1). Notably, the procedure worked in a similar way when single binary vectors were used [20].

Generation of multi-sialylated N-glycans on rhEPOFc

The generation of plant derived rhEPOFc carrying branched (tri- and tetra-antennary) *N*-glycans has been reported previously [26,27]. This was achieved by the co-expression of rhEPOFc with mammalian *N*-acetylglucosaminyltransferases IV and V targeted to medial Golgi compartment (^{FUT11}GnTIV or ^{FUT11}GnTV; [26]). Here we set out to generate multi-antennary sialylated rhEPOFc. To approach this issue, we first co-expressed rhEPOFc with pC144 and pG371 in combination with either ^{FUT11}GnTIV or ^{FUT11}GnTV. SDS-PAGE analysis of purified rhEPOFc (rhEPOFc$_{TriSia}$, Figure 2B lane 3) showed that the 55 kDa band corresponding to the fusion protein appears as a "smeary" band compared to rhEPOFc$_{\Delta XTFT}$ or rhEPOFc$_{Sia}$ (Fig 2B, lanes 1 and 2, respectively). The relative occurrence of the different complex glycoforms present in rhEPOFc$_{TriSia}$ is displayed in Table 2. In total about 80% of all glycans were sialylated, with the dominant *N*-glycan, being tri-sialylated oligosaccharide. Low amounts of tri-antennary non-sialylated structures are also detected and oligomannosidic structures account for ca. 10–12% of the total *N*-glycans. In addition the "smeary" 55 kDa band was separated into two fractions (A and B, Figure 2B) and they were individually analysed. The *N*-glycosylation profile of fraction A (which corresponds to a size slightly larger than 55 kDa) exhibits almost exclusively tri-antennary sialylated carbohydrates in all three glycosites ([NaNa]Na) (Figure 5). In contrast, fraction B (which corresponds to the lower part of the 55 kDa band) was decorated mainly with tri-antennary non-sialylated *N*-glycans with or without galactosylation ([GnGn]Gn, [AGn]Gn), accompanied by oligomannosidic *N*-glycans (Figure S2).

Finally rhEPOFc was co-expressed with pC144 and pG371 in combination with both ^{FUT11}GnTIV and ^{FUT11}GnTV. This procedure encompasses a coordinated action of eleven heterologous proteins. The purified product (rhEPOFc$_{TetraSia}$) exhibited on Coommassie stained SDS-PAGE a "smeary" 55 kDa band as observed for rhEPOFc$_{TriSia}$.

LC-ESI-MS analysis revealed that rhEPOFc$_{TetraSia}$ glycopeptides carried about 80% sialylated structures including tri- and tetra-sialylation (Table 2). Tri-antennary sialylated structures were the major glycoform in all three rhEPO glycosites (up to 56%). While GP 2 and 3 carried about 10-13% tetra-sialylated structures, surprisingly, this complex carbohydrate was not present on GP 1. Moreover ~15% of rhEPOFc$_{TetraSia}$ are decorated with oligomannosidic structures (not included in Table 2). As before, *N*-glycosylation analysis was individually performed on the two fractions A and B. (Figure 2B, lane 4). The main glycoform of rhEPOFc in fraction A is tri-sialylated with significant amounts of bi- and tetra-sialylated *N*-glycans (NaNa and [NaNa][NaNa]) on glycopeptide 2 (Gp2, Asn-38) and Gp3 (Asn-83) (Figure 6). Interestingly, on Gp1 (Asn-24) a single dominant peak corresponding to tri-sialylated structures is detected as well as smaller fractions of bi-sialylated glycans (NaNa) but no tetra-sialylated N-glycans were detected (Figure 6). Fraction B exhibited a variety of non-sialylated branched *N*-glycans some carrying one or two galactose residues ([GnGn][GnGn], Gn[GnGn]$_{iso}$, GnGn, [GnGn][GnA] and [GnA][GnA]$_{iso}$. Consistently with fraction A, Gp1 carries only GnGn and tri-antennary *N*-glycans (Figure S3).

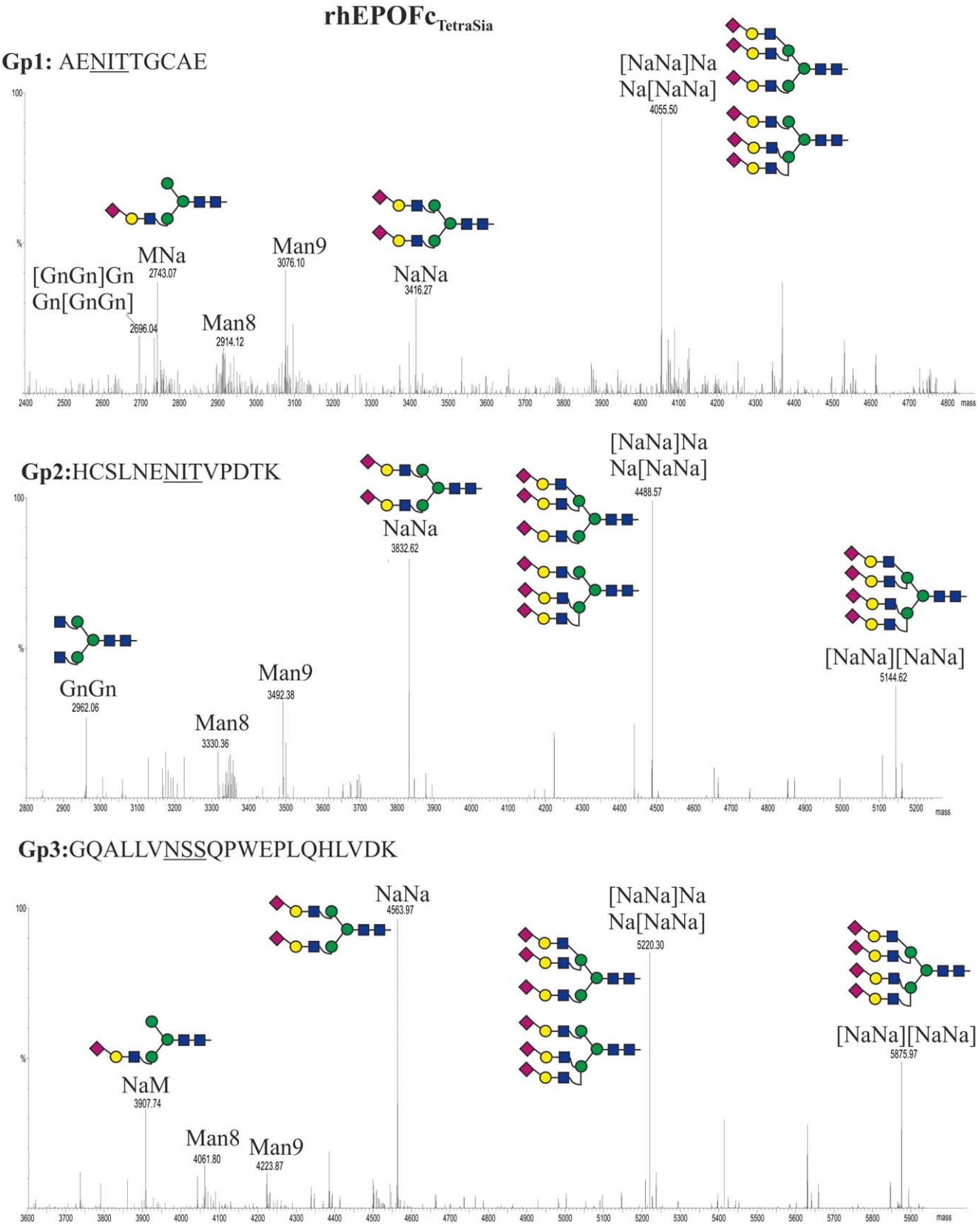


Figure 6. Generation of tetra-sialylated structures in rhEPOFc. Mass spectra of trypsin and endoproteinase Glu-C double-digested rhEPOFc co-expressed in *N. benthamiana* ΔXTFT with mammalian genes for synthesis of tetra-sialylated *N*-glycans (rhEPO$_{TetraSia}$). The analysis was performed on rhEPOFc$_{TetraSia}$ present on fraction A of the 55kDa band (Figure 2B, lane 4). Glycosylation patterns of rhEPO Gp1: E/A^{22}ENITTGCAE^{31}; Gp2: E/H^{32}CSLNENITVPDTK^{45} and Gp3: R/G^{77}QALLVNSSQPWEPLQHLVDK^{97} are shown. *N*-glycosylation profile of the Fc glycopeptide is shown in Figure S1.

Glycosylation profile of rhEPOFc present on fraction B of the 55kDa band is shown in Figure S3. Peak labels were made according to the ProGlycAn system (www.proglycan.com Illustrations display *N*-glycans on assigned peaks, for interpretation of other assigned glycoforms see Figure S5.
doi:10.1371/journal.pone.0054836.g006

Analysis of Fc glycosylation in rhEPoFc$_{TriSia}$ and rhEPOFc$_{TetraSia}$ shows a largely heterogenous *N*-glycosylation profile with a mixture of GnGn and oligomannosidic glycoforms, but also minor amounts of tri-antennary, galactosylated and sialylated structures (Figure S1). Notably expression levels of all glycoforms were in the same range (Table 1) indicating that co-infiltration of human glycosylation enzymes did not alter expression level of the recombinant fusion protein.

In vitro activity of different EPOFc glycoforms

Finally the biological activity of the plant-derived rhEPOFc variants was analysed using an erythropoietin-dependent human leukemia cell line, UT-7. Proliferation of the UT-7 cells is induced by the presence of EPO. The proliferation of UT-7 cells was measured and half maximal effective dose (ED_{50}) values were compared. All plant-derived rhEPOFc glycoforms had similar ED_{50} values ranging 0.26–0.54 ng/mL. Comparably a slightly reduced receptor binding was obtained for the CHO derived counterpart (ED_{50} 1.7 ng/mL) (Table 3). This might be due to different downstream procedures of plant and CHO derived recombinant hormones, e. g. CHO derived rhEPOFc but not the plant derived counterparts was subjected to a virus inactivation test.

Discussion

With the recognition of the N-glycan nature of the AB0 blood group types, glycoconjugates were accepted to elicit specific reactions [38]. Since then numerous studies have highlighted the impact of this important posttranslational modification on the function of proteins. A well-known example is rhEPO, one of the leading biopharmaceutical products. The human EPO is a highly glycosylated molecule with three *N*- and one *O*-linked glycans. The relevance of *O*-glycosylation for the biological activity of EPO is unclear, and implications for a role in secretion are not conclusive [39]. On the contrary the biological implications of *N*-glycosylation are well characterized [4–6]. In the course of enhancing drug efficacy in anemia treatment, increased *in vivo* half-life *via* enhanced terminal sialylation was achieved. Moreover, fusing the hormone to an IgG-Fc domain resulted in a significant extension of the serum half-life of the recombinant hormone [40].

Table 3. *in vitro* activity of CHO- and plant-derived rhEPOFC.

rhEPOFc	ED_{50} (ng/mL)
CHO	1.7
ΔXTFT	0.45
Sia	0.54
TriSia	0.25
TetraSia	0.26

In vitro activity assay of plant- and CHO- derived rhEPOFc. Half maximal effective doses (ED_{50}) are displayed. rhEPOFc was expressed in CHO cells (CHO); in *N. benthamiana* ΔXTFT mutants (ΔXTFT); co-expressed in ΔXTFT with mammalian genes for protein sialylation (Sia); co-expressed in ΔXTFT with mammalian genes for synthesis of tri-antennary sialylated *N*-glycans (TriaSia) and co-expressed in ΔXTFT with mammalian genes for synthesis of tetra-sialylated *N*-glycans (TetraSia).
doi:10.1371/journal.pone.0054836.t003

Here we report the transient expression of rhEPOFc in plants. Using the *N. benthamiana* ΔXTFT in combination with the magnICON based expression system, we achieved expression levels of rhEPOFc of up to 9 mg/kg leaves, which accounts for 0.2% of TSP. This is a relatively modest expression level in comparison to amounts reported for other recombinant proteins with the magnICON systems [30,41], however they are in agreement with rhEPOFc expressed previously in plants [27]. Low expression could result from the fact that a large portion of the recombinantly expressed protein (about 30–50 times) refers to Fc lacking the hEPO fragment. We designed different rhEPOFc chimeras to address this issue. However neither manipulation on the Fc-hinge region nor the presence/absence of the EPO Arginine^{166} residue had a significant influence on expression of the full length fusion protein. The generation of free Fc has been already reported previously upon expression of hEPOFc in chicken [34], however it is not present when produced in mammalian cells [9,42]. The reason for this phenomenon has not been investigated in detail, although degradation of the fusion protein by plant proteases is a plausible explanation. Several studies refer to the proteolytic degradation of heterologous proteins in plants [43] and the outcome indicated that this occurs preferentially in the apoplast [44]. Importantly human proteins like EPOFc have not evolved in the context of plant proteases and thereby they represent novel targets. The apoplastic fluid of *N. benthamiana* is enriched of acidic proteases, e.g. the presence of papain-like cystein family was reported [45]. Papain is a non-specific protease that cleaves monoclonal antibodies preferentially in the *N*-terminal side of the hinge region and was effectively used to separate the rhEPO from the Fc fragment during *N*-glycan profiling of rhEPOFc produced in CHO cells [9]. Proteolysis is a major issue of recombinant proteins affecting the product yield, not only in plants but also in other expression systems. Different strategies are being considered to avoid or minimize proteolysis of heterologous proteins expressed in plants [46]. And the outcome hopefully will allow enhanced expression of full length rhEPOFc in plants.

Here we report the generation of rhEPOFc glyco-variants which largely resembles that of the CHO derived counterparts (Figure S4). Expression of rhEPOFc in ΔXTFT mutants results in the formation of almost exclusively GnGn structures on all glycosylation sites. Interestingly, although present in total soluble proteins extracted from *Δ*XTFT mutants and in some recombinantly expressed proteins [47], no truncated paucimannosidic structures, i.e. MM, were detected. Co-expression of hEPOFc with the mammalian genes involved in protein sialylation permitted the production of a hormone largely decorated with bi-antennary sialylated complex *N*-glycans. These structures are one of the major glycoforms of EPO present human serum [48], however accounts only for about 15% on the CHO derived counterpart [49]. Moreover, we report the synthesis of plant-derived rhEPOFc carrying multi-antennary sialylated *N*-glycans, the major structures of mammalian cell derived therapeutic rhEPO. Co-expression of rhEPOFc with the genes necessary for *N*-glycan branching and sialylation resulted in a mixture of neutral and charged oligosaccharides. In total approximately 10–16 glycoforms, with different relative amounts, are distributed by the four glycopeptides, similar to the observation for CHO-derived rhEPOFc [9]. In summary, rhEPOFc carried 80–90% sialylated structures upon co-expression with the sialylation pathway. Bi- and tri-sialylation were the major

glycoform in recombinant products rhEPOFc$_{Sia}$ and rhEPOFc-$_{TriSia}$, respectively. In contrast to mammalian cell derived rhEPO, tetra-sialylation is inefficiently synthesized in plants and rhEPOFc$_{TetraSia}$ carries only about 10–14% of this highly complex carbohydrate formation. Beside complex *N*-glycans, plant derived rhEPOFc carries about 10–15% oligomannosidic structures. Notably, and as for mammalian cells, there are significant differences on the *N*-glycan profile of rhEPO and of Fc domain. While efficient modulation towards mammalian-like structures was observed on all rhEPO glycosites, Fc glycosylation exhibited unusual structures. A homogeneous glycosylation profile (namely GnGn oligosaccharides) was obtained for Fc produced in *Δ*XTFT, all further modification steps (branching, sialylation) led to the synthesis of a largely heterogeneous glycosylation profile with unusual incompletely processed structures. In particular, sialylation was modest. The unusual Fc glycan-modulation was already observed previously for the synthesis of tetra-antennary *N*-glycans in rhEPOFc [26]. The reason for this different performance in glyco-modulation of rhEPO and Fc is currently not understood. One explanation could be different accessibility of the *N*-glycosylation sites, EPO glycosites are considered very exposed while the Fc-glycosites are buried within the protein backbone [50] and as a consequence they have restricted accessibility to *N*-glycan processing enzymes.

Multi-sialylation of rhEPOFc requires the coordinated expression of 11 exogenous genes in a single cell. To reduce the number of agrobacteria cultures and to facilitate the simultaneous delivery of glycosylation genes into the same cell, two multi gene vectors carrying the six genes for *in planta* sialylation were constructed (pC144 and pG371). With a future intention of using these vectors to stably introduce the sialic acid pathway into plants, different plant selection markers have been placed to the vector backbones. Also a combination of several promoter and terminator sequences were used to reduce the risk of transgene silencing when attempting to express a series of genes stably from a single plasmid. Here, by transient expression, we demonstrate that these multi-gene vectors efficiently sialylate their target protein making them valuable tools for plant transformation.

Transgenic *N. benthamiana* plants stable expressing mammalian glycosyltransferases can be extremely useful for the production of recombinant proteins with a highly homogenous human-like glycosylation profile as recently demonstrated [27,28].

Importantly all plant-derived rhEPOFc glycoforms are biologically active as seen in receptor binding assays. These results are a good starting point for follow up advanced structure-function studies, with the aim to determine the most suitable glycoforms. These will be the focus of future experiments.

With the generation of multi-sialylated glycans we display *in vivo* engineering of one of the most complex human *N*-glycan structures *in planta* and thereby demonstrate the enormous plasticity of plants to tolerate modifications on their protein *N*-glycosylation. The results presented here together with other achievements in plant *N*-glycoengineering (reviewed by [19]) provide the know-how for the generation of recombinant proteins with targeted *N*-glycosylation profiles. This allows advanced protein-carbohydrate structure-function studies to better understand the impact of *N*-glycans and to develop next generation drugs, where patients would benefit from optimally glycosylated drugs.

Supporting Information

Figure S1 *N*-glycosylation profile observed in the Fc glycopeptide (R/EEQYNSTYR) of rhEPOFc$_{ΔXTFT}$: rhEPOFc expressed in *N. benthamiana* *Δ*XTFT mutants; rhEPOFc$_{Sia}$: rhEPOFc co-expressed in *Δ*XTFT with mammalian genes for protein sialylation; rhEPO$_{TriSia}$: rhEPOFc co-expressed in *Δ*XTFT with mammalian genes for synthesis of tri-antennary sialylated *N*-glycans; rhEPO$_{TetraSia}$: rhEPOFc co-expressed in *Δ*XTFT with mammalian genes for synthesis of tetra-sialylated *N*-glycans. For interpretation of glycoforms present in assigned peaks see Figure S5.
(TIF)

Figure S2 *N*-glycosylation profile of rhEPOFc$_{TriSia}$ present in fraction B of the 55kDa band (Figure 2B, lane 3). Glycosylation patterns of rhEPO Gp1: E/A^{22}ENITTGCAE^{31}; Gp2: E/H^{32}CSLNENITVPDTK^{45} and Gp3: R/G^{77}QALLVNSSQPWEPLQHLVDK^{97} are shown. Peak labels were made according to the ProGlycAn system (www.proglycan.com). For interpretation of glycoforms present in assigned peaks see Figure S5.
(TIF)

Figure S3 *N*-glycosylation profile of rhEPOFc$_{TetraSia}$ present in fraction B of the 55kDa band (Figure 2B, lane 4). Glycosylation patterns of rhEPO Gp1: E/A^{22}ENITTGCAE^{31}; Gp2: E/H^{32}CSLNENITVPDTK^{45} and Gp3: R/G^{77}QALLVNSSQPWEPLQHLVDK^{97} are shown. Peak labels were made according to the ProGlycAn system (www.proglycan.com). For interpretation of glycoforms present in assigned peaks see Figure S5.
(TIF)

Figure S4 *N*-glycosylation profile of rhEPOFc expressed in CHO cells. Glycosylation patterns of rhEPO Gp1: E/A^{22}ENITTGCAE^{31}; Gp2: E/H^{32}CSLNENITVPDTK^{45}; Gp3: R/G^{77}QALLVNSSQPWEPLQHLVDK^{97}and the Fc glycopeptide (R/EEQYNSTYR) are shown. Peak labels were made according to the ProGlycAn system (www.proglycan.com). For interpretation of glycoforms present in assigned peaks see Figure S5.
(TIF)

Figure S5 Illustration of *N*-glycan structures on transiently expressed rhEPOFc. Oligomannosidic structures (Man5–Man9), complex *N*-glycans typical of plant-derived proteins (GnGnXF) and glycans carrying Lewis-a epitopes (Gn(FA)$_{iso}$) are also illustrated. Schematic representations are based on the nomenclature proposed by the consortium for Functional Glycomics.
(TIF)

Table S1 List of primers as cited in Material and Methods.
(DOCX)

Acknowledgments

We thank Thomas Hackl and Hanna Weindorfer, Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Vienna, Austria for excellent technical support, also Dr. Eva Decker, Plant Biotechnology University of Freiburg, Freiburg, Germany for providing the hEPO quantification kit.

Footnote: Nucleotide sequence data for EPOFc codon optimized for plants is available in the Genbank under the accession number KC329647.

Author Contributions

Conceived and designed the experiments: AC HS. Performed the experiments: AC LN PG. Analyzed the data: AC LN FA TS KVU RS.

Contributed reagents/materials/analysis tools: TS KVU. Wrote the paper: AC HS.

References

1. Maiese K, Chong ZZ, Li F, Shang YC (2008) Erythropoietin: elucidating new cellular targets that broaden therapeutic strategies. Prog Neurobiol 85: 194–213.
2. Sasaki H, Ochi N, Dell A, Fukuda M (1988) Site-specific glycosylation of human recombinant erythropoietin: analysis of glycopeptides or peptides at each glycosylation site by fast atom bombardment mass spectrometry. Biochemistry 27: 8618–8626.
3. Takeuchi M, Takasaki S, Miyazaki H, Kato T, Hoshi S, et al. (1988) Comparative study of the asparagine-linked sugar chains of human erythropoietins purified from urine and the culture medium of recombinant Chinese hamster ovary cells. J Biol Chem 263: 3657–3663.
4. Dubé S, Fisher JW, Powell JS (1988) Glycosylation at specific sites of erythropoietin is essential for biosynthesis, secretion, and biological function. J Biol Chem 263: 17516–17521.
5. Higuchi M, Oh-eda M, Kuboniwa H, Tomonoh K, Shimonaka Y, et al. (1992) Role of sugar chains in the expression of the biological activity of human erythropoietin. J Biol Chem 267: 7703–7709.
6. Egrie JC, Browne JK (2001) Development and characterization of novel erythropoiesis stimulating protein (NESP). Nephrol Dial Transplant 16 Suppl 3: 3–13.
7. Hokke CH, Bergwerff AA, Van Dedem GW, Kamerling JP, Vliegenthart JF (1995) Structural analysis of the sialylated N- and O-linked carbohydrate chains of recombinant human erythropoietin expressed in Chinese hamster ovary cells. Sialylation patterns and branch location of dimeric N-acetyllactosamine units. Eur J Biochem 228: 981–1008.
8. Yuen CT, Storring PL, Tiplady RJ, Izquierdo M, Wait R, et al. (2003) Relationships between the N-glycan structures and biological activities of recombinant human erythropoietins produced using different culture conditions and purification procedures. Br J Haematol 121: 511–526.
9. Schriebl K, Trummer E, Lattenmayer C, Weik R, Kunert R, et al. (2006) Biochemical characterization of rhEpo-Fc fusion protein expressed in CHO cells. Protein Expr Purif 49: 265–275.
10. Wang MD, Yang M, Huzel N, Butler M (2002) Erythropoietin production from CHO cells grown by continuous culture in a fluidized-bed bioreactor. Biotechnol Bioeng 77: 194–203.
11. Elliott S, Lorenzini T, Asher S, Aoki K, Brankow D, et al. (2003) Enhancement of therapeutic protein in vivo activities through glycoengineering. Nat Biotechnol 21: 414–421.
12. Bragonzi A, Distefano G, Buckberry LD, Acerbis G, Foglieni C, et al. (2000) A new Chinese hamster ovary cell line expressing alpha2,6-sialyltransferase used as universal host for the production of human-like sialylated recombinant glycoproteins. Biochim Biophys Acta 1474: 273–282.
13. Wang Z, Park JH, Park HH, Tan W, Park TH (2011) Enhancement of recombinant human EPO production and sialylation in chinese hamster ovary cells through Bombyx mori 30Kc19 gene expression. Biotechnol Bioeng 108: 1634–1642.
14. Bork K, Reutter W, Weidemann W, Horstkorte R (2007) Enhanced sialylation of EPO by overexpression of UDP-GlcNAc 2-epimerase/ManAc kinase containing a sialuria mutation in CHO cells. FEBS Lett 581: 4195–4198.
15. Bitonti AJ, Dumont JA, Low SC, Peters RT, Kropp KE, et al. (2004) Pulmonary delivery of an erythropoietin Fc fusion protein in non-human primates through an immunoglobulin transport pathway. Proc Natl Acad Sci U S A 101: 9763–9768.
16. Davies HM (2010) Review article: commercialization of whole-plant systems for biomanufacturing of protein products: evolution and prospects. Plant Biotechnol J 8: 845–861.
17. Chung SM, Frankman EL, Tzfira T (2005) A versatile vector system for multiple gene expression in plants. Trends Plant Sci 10: 357–361.
18. Loos A, Steinkellner H (2012) IgG-Fc glycoengineering in non-mammalian expression hosts. Arch Biochem Biophys.
19. Castilho A, Steinkellner H (2012) Glyco-engineering in plants to produce human-like N-glycan structures. Biotechnology Journal. doi: 10.1002/biot.201200032
20. Castilho A, Strasser R, Stadlmann J, Grass J, Jez J, et al. (2010) In planta protein sialylation through overexpression of the respective mammalian pathway. J Biol Chem 285: 15923–15930.
21. Conley AJ, Mohib K, Jevnikar AM, Brandle JE (2009) Plant recombinant erythropoietin attenuates inflammatory kidney cell injury. Plant Biotechnol J 7: 183–199.
22. Kittur FS, Hung CY, Darlington DE, Sane DC, Xie J (2012) N-Glycosylation engineering of tobacco plants to produce asialoerythropoietin. Plant Cell Rep.31: 1233–43.
23. Parsons J, Altmann F, Arrenberg CK, Koprivova A, Beike AK, et al. (2012) Moss-based production of asialo-erythropoietin devoid of Lewis A and other plant-typical carbohydrate determinants. Plant Biotechnol J.10: 851–61.
24. Matsumoto S, Ikura K, Ueda M, Sasaki R (1995) Characterization of a human glycoprotein (erythropoietin) produced in cultured tobacco cells. Plant Mol Biol 27: 1163–1172.
25. Weise A, Altmann F, Rodriguez-Franco M, Sjoberg ER, BÃ¤umer W, et al. (2007) High-level expression of secreted complex glycosylated recombinant human erythropoietin in the Physcomitrella Delta-fuc-t Delta-xyl-t mutant. Plant Biotechnol J 5: 389–401.
26. Castilho A, Gattinger P, Grass J, Jez J, Pabst M, et al. (2011) N-glycosylation engineering of plants for the biosynthesis of glycoproteins with bisected and branched complex N-glycans. Glycobiology 21: 813–823.
27. Nagels B, Van Damme EJ, Callewaert N, Zabeau L, Tavernier J, et al. (2012) Biologically active, magnICON(®)-expressed EPO-Fc from stably transformed Nicotiana benthamiana plants presenting tetra-antennary N-glycan structures. J Biotechnol.160: 242–50.
28. Strasser R, Castilho A, Stadlmann J, Kunert R, Quendler H, et al. (2009) Improved virus neutralization by plant-produced anti-HIV antibodies with a homogeneous beta1,4-galactosylated N-glycan profile. J Biol Chem 284: 20479–20485.
29. Strasser R, Stadlmann J, Schähs M, Stiegler G, Quendler H, et al. (2008) Generation of glyco-engineered Nicotiana benthamiana for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure. Plant Biotechnol J 6: 392–402.
30. Marillonnet S, Thoeringer C, Kandzia R, Klimyuk V, Gleba Y (2005) Systemic Agrobacterium tumefaciens-mediated transfection of viral replicons for efficient transient expression in plants. Nat Biotechnol 23: 718–723.
31. Stadlmann J, Pabst M, Kolarich D, Kunert R, Altmann F (2008) Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. Proteomics 8: 2858–2871.
32. Pabst M, Chang M, Stadlmann J, Altmann F (2012) Glycan profiles of the 27 N-glycosylation sites of the HIV envelope protein CN54gp140. Biol Chem 393: 719–730.
33. Komatsu N, Nakauchi H, Miwa A, Ishihara T, Eguchi M, et al. (1991) Establishment and characterization of a human leukemic cell line with megakaryocytic features: dependency on granulocyte-macrophage colony-stimulating factor, interleukin 3, or erythropoietin for growth and survival. Cancer Res 51: 341–348.
34. Penno CA, Kawabe Y, Ito A, Kamihira M (2010) Production of recombinant human erythropoietin/Fc fusion protein by genetically manipulated chickens. Transgenic Res 19: 187–195.
35. Im SJ, Yang SI, Yang SH, Choi DH, Choi SY, et al. (2011) Natural form of noncytolytic flexible human Fc as a long-acting carrier of agonistic ligand, erythropoietin. PloS one 6: e24574.
36. Recny MA, Scoble HA, Kim Y (1987) Structural characterization of natural human urinary and recombinant DNA-derived erythropoietin. Identification of des-arginine 166 erythropoietin. J Biol Chem 262: 17156–17163.
37. Strasser R, Bondili JS, Vavra U, Schoberer J, Svoboda B, et al. (2007) A unique beta1,3-galactosyltransferase is indispensable for the biosynthesis of N-glycans containing Lewis a structures in Arabidopsis thaliana. Plant Cell 19: 2278–2292.
38. Morgan WT, Watkins WM (1969) Genetic and biochemical aspects of human blood-group A-, B-, H-, Le-a- and Le-b-specificity. Br Med Bull 25: 30–34.
39. Wasley LC, Timony G, Murtha P, Stoudemire J, Dorner AJ, et al. (1991) The importance of N- and O-linked oligosaccharides for the biosynthesis and in vitro and in vivo biologic activities of erythropoietin. Blood 77: 2624–2632.
40. Way JC, Lauder S, Brunkhorst B, Kong SM, Qi A, et al. (2005) Improvement of Fc-erythropoietin structure and pharmacokinetics by modification at a disulfide bond. Protein Eng Des Sel 18: 111–118.
41. Giritch A, Marillonnet S, Engler C, van Eldik G, Botterman J, et al. (2006) Rapid high-yield expression of full-size IgG antibodies in plants coinfected with noncompeting viral vectors. Proc Natl Acad Sci U S A 103: 14701–14706.
42. Taschwer M, Hackl M, Hernández Bort JA, Leitner C, Kumar N, et al. (2012) Growth, productivity and protein glycosylation in a CHO EpoFc producer cell line adapted to glutamine-free growth. J Biotechnol 157: 295–303.
43. Lombardi R, Donini M, Villani ME, Brunetti P, Fujiyama K, et al. (2012) Production of different glycosylation variants of the tumour-targeting mAb H10 in Nicotiana benthamiana: influence on expression yield and antibody degradation. Transgenic ResDOI 10.1007/s11248-012-9587-1.
44. Hehle VK, Paul MJ, Drake PM, Ma JK, van Dolleweerd CJ (2011) Antibody degradation in tobacco plants: a predominantly apoplastic process. BMC Biotechnol 11: 128.
45. van der Hoorn RA (2008) Plant proteases: from phenotypes to molecular mechanisms. Annu Rev Plant Biol 59: 191–223.
46. Adamczyk M, Gebler JC, Wu J (2000) Papain digestion of different mouse IgG subclasses as studied by electrospray mass spectrometry. J Immunol Methods 237: 95–104.
47. Dirnberger D, Steinkellner H, Abdennebi L, Remy JJ, van de Wiel D (2001) Secretion of biologically active glycoforms of bovine follicle stimulating hormone in plants. Eur J Biochem.268:4570–9.
48. Skibeli V, Nissen-Lie G, Torjesen P (2001) Sugar profiling proves that human serum erythropoietin differs from recombinant human erythropoietin. Blood 98: 3626–3634.

49. Shahrokh Z, Royle L, Saldova R, Bones J, Abrahams JL, et al. (2011) Erythropoietin produced in a human cell line (Dynepo) has significant differences in glycosylation compared with erythropoietins produced in CHO cell lines. Mol Pharm 8: 286–296.

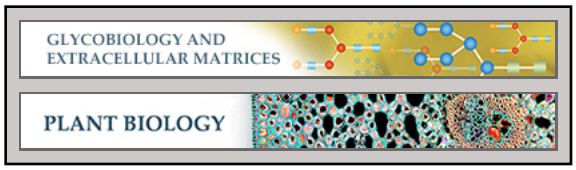
50. Sondermann P, Huber R, Oosthuizen V, Jacob U (2000) The 3.2-[angst] crystal structure of the human IgG1 Fc fragment-Fc[gamma]RIII complex. Nature 406: 267–273.

Appendix II

Castilho, A., **Neumann, L**., Daskalova, S., Mason, H. S., Steinkellner, H., Altmann, F., Strasser, R. (2012) Engineering of Sialylated Mucin-type O-glycosylation in Plants. *J. Biol. Chem*. 287(43):36518-26. doi:10.1074/jbc.M112.402685

Glycobiology and Extracellular Matrices:

Engineering of Sialylated Mucin-type *O*-Glycosylation in Plants



Alexandra Castilho, Laura Neumann, Sasha Daskalova, Hugh S. Mason, Herta Steinkellner, Friedrich Altmann and Richard Strasser

J. Biol. Chem. 2012, 287:36518-36526.
doi: 10.1074/jbc.M112.402685 originally published online September 4, 2012

⌘ *Author's Choice*

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 287, NO. 43, pp. 36518–36526, October 19, 2012

Engineering of Sialylated Mucin-type *O*-Glycosylation in Plants*[S]

Received for publication, July 26, 2012, and in revised form, August 24, 2012 Published, JBC Papers in Press, September 4, 2012, DOI 10.1074/jbc.M112.402685

Alexandra Castilho[‡], Laura Neumann[§], Sasha Daskalova[¶], Hugh S. Mason[¶], Herta Steinkellner[‡], Friedrich Altmann[§], and Richard Strasser[‡1]

From the [‡]Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna Austria, the [§]Department of Chemistry, University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria, and [¶]The Biodesign Institute and School of Life Sciences, Arizona State University, Tempe, Arizona 85287

Background: Plants lack the machinery for mucin-type *O*-glycosylation.
Results: Transient expression of the mammalian *O*-glycosylation pathway in *Nicotiana benthamiana* resulted in the formation of sialylated mucin-type *O*-glycans on recombinant erythropoietin.
Conclusion: Therapeutic proteins with engineered *N*- and *O*-glycosylation can be produced in plants.
Significance: Plants are attractive hosts for the production of glycosylated recombinant proteins with defined glycan structures.

Proper *N*- and *O*-glycosylation of recombinant proteins is important for their biological function. Although the *N*-glycan processing pathway of different expression hosts has been successfully modified in the past, comparatively little attention has been paid to the generation of customized *O*-linked glycans. Plants are attractive hosts for engineering of *O*-glycosylation steps, as they contain no endogenous glycosyltransferases that perform mammalian-type Ser/Thr glycosylation and could interfere with the production of defined *O*-glycans. Here, we produced mucin-type *O*-GalNAc and core 1 *O*-linked glycan structures on recombinant human erythropoietin fused to an IgG heavy chain fragment (EPO-Fc) by transient expression in *Nicotiana benthamiana* plants. Furthermore, for the generation of sialylated core 1 structures constructs encoding human polypeptide:*N*-acetylgalactosaminyltransferase 2, *Drosophila melanogaster* core 1 β1,3-galactosyltransferase, human α2,3-sialyltransferase, and *Mus musculus* α2,6-sialyltransferase were transiently co-expressed in *N. benthamiana* together with EPO-Fc and the machinery for sialylation of *N*-glycans. The formation of significant amounts of mono- and disialylated *O*-linked glycans was confirmed by liquid chromatography-electrospray ionization-mass spectrometry. Analysis of the three EPO glycopeptides carrying *N*-glycans revealed the presence of biantennary structures with terminal sialic acid residues. Our data demonstrate that *N. benthamiana* plants are amenable to engineering of the *O*-glycosylation pathway and can produce well defined human-type *O*- and *N*-linked glycans on recombinant therapeutics.

Recombinant protein-based drugs are one of the fastest growing branches of the pharmaceutical industry. Consequently, there is a demand for increasing current manufacturing capacities and exploring novel expression systems that are faster, more flexible, and possibly cheaper than established hosts. During the last 10 years there have been remarkable advances in expression technology and manipulation of post-translational modifications in plants, which makes them promising hosts for the production of proteins of interest. In fact, a carrot cell-produced recombinant human β-glucocerebrosidase for treatment of Gaucher disease is now approved for enzyme replacement therapy in humans (1, 2). Plants are versatile production platforms as they are essentially free of human pathogens and naturally lack certain glycan epitopes that are present on recombinant glycoproteins produced in non-human mammalian cell lines (3).

Many therapeutic proteins are glycosylated, and the presence or absence of distinct sugar residues can have a significant impact on protein function *in vitro* and *in vivo*. One of the best examples demonstrating the importance of distinct glycoforms for its function is the Fc region of IgG molecules (4, 5). For the glycoprotein hormone erythropoietin (EPO),[2] which regulates red blood cell formation and is the standard drug for the treatment of anemia, proper glycosylation is crucial for its *in vivo* function (6). In particular there is a direct, positive correlation between the sialic acid content, increased half-life in the blood, and its bioactivity. Consequently, the potency of EPO can be increased by engineering of glycosylation, such as the formation of highly branched sialylated *N*-glycans, introduction of additional glycosylation sites, or metabolic engineering to provide more CMP-sialic acid for sialylation (7, 8). The majority of these glycosylation engineering approaches have focused on modification of *N*-glycan structures on EPO, whereas little attention has so far been paid to *O*-glycan engineering. Both

* This work was supported by a grant (to R. S.) from the Federal Ministry of Transport, Innovation, and Technology (bmvit) and Austrian Science Fund (FWF) TRP 242-B20 (to R. S.), by the PhD program "BioToP-Biomolecular Technology of Proteins" from the Austrian Science Fund (FWF) Project W1224-B09 (to L. N.), and by a grant from the Austrian Research Promotion Agency (FFG) Laura Bassi Centres of Expertise project 822757 (to H. S.).
⌘ *Author's Choice*—Final version full access.
[S] This article contains supplemental Figs. S1–S3.
[1] To whom the correspondence should be addressed. Tel.: 43-1-47654-6705; Fax: 43-1-47654-6392; E-mail: richard.strasser@boku.ac.at.

[2] The abbreviations used are: EPO, erythropoietin; Fc, heavy chain fragment; ESI, electrospray ionization; ΔXTFT, *N. benthamiana* glycosylation mutant deficient in β1,2-xylosyltransferase and core α1,3-fucosyltransferase; CST, CMP-sialic acid transporter; Glu-C, endoglucosaminidase C; GalNAc-T, *N*-acetylgalactosaminyltransferase.

urinary and recombinant human EPO contain a single *O*-glycosylation site at Ser-126 that is mostly decorated with a disialylated mucin-type *O*-glycan (9–11). The importance of *O*-glycan structures for EPO stability or biological function is not entirely clear (12, 13). However, it has been suggested that the sialic acid residues from the mucin-type *O*-glycans could also contribute to the total sialic acid content and subsequently might influence the clearance of EPO from circulation (14).

In comparison to *N*-glycans, the role of different glycoforms and individual sugar residues on *O*-linked glycans of recombinant proteins is much less understood. Dependent on the size and composition of the *O*-glycan, the conformation of the protein or its activity might be affected (15). The presence of *O*-linked structures can mask recognition sites for receptors and other interacting proteins or protect them from degradation by proteases (16). Moreover, changes in *O*-glycosylation are often seen in diseases (*e.g.* in IgA nephropathy), and specific *O*-glycan structures are aberrant in tumors and are, therefore, potential targets for the development of glycopeptide-based anticancer vaccines (17, 18). An expression system that enables the production of recombinant glycoproteins with well defined homogenous *O*-linked glycans would be highly desirable not only for the biopharmaceutical industry but also for structure function studies involving glycoproteins with *O*-glycans.

Here we describe the production of tumor-associated antigens (Tn and T antigen) and disialylated mucin-type core 1 *O*-glycans on recombinant EPO transiently expressed in leaves of *Nicotiana benthamiana* plants. In addition, we combine *N*- and *O*-glycan modification strategies to generate a production platform equipped with a fully functional human-type *N*- and *O*-glycosylation machinery. Our data demonstrate that plants are amenable to extensive *O*-glycan engineering, which greatly expands the potential of plants as novel expression hosts for the production of recombinant glycoproteins with customized glycans.

EXPERIMENTAL PROCEDURES

Construction of Plant Binary Expression Vector—The binary vector for expression of human GalNAc-T2 (pH7WG2:GNT2) was described previously (19). Binary vectors for expression of *Caenorhabditis elegans* UDP-GlcNAc/UDP-GalNAc transporter (pH7WG2:GT) and *Yersinia enterocolitica* UDP-GlcNAc 4-epimerase (pH7WG2:GE) were generated by Gateway LR-reaction transfer (Invitrogen) of the appropriate gene from the pENTR/D-Topo-based shuttle vectors (19) to pH7WG2 plant expression vector (20). EPO-Fc was expressed using the magnICON viral-based expression system (21) as described previously (22). A codon-optimized open reading frame coding for human core 1 β1,3-galactosyltransferase (C1GALT1) was obtained from GeneArt Gene Synthesis (Invitrogen). XhoI and BamHI restriction enzyme recognition sites were incorporated at the 5′- and 3′-ends, respectively, to facilitate subsequent cloning into the XhoI/BamHI sites of the auxiliary vector pSAT1A (pSAT1A-C1GALT1) (23). The rare-cutting enzyme AscI was used to clone the expression cassette of pSAT1A-C1GALT1 into pPZP-RCS2 binary expression vector.

A clone (IMAGE ID: 5724507) coding for human COSMC was purchased from Source BioScience (Cambridge, UK). The open reading frame was amplified by PCR using oligos Chaperon-F1 (5′-TATACTCGAGATGCTTTCTGAAAGCAGC-3′) and Chaperon-R1 (5′-TATAAGATCTTCAGTCATTGTCAGAACC-3′), digested with XhoI/BglII, and ligated into XhoI/BamHI digested pSAT1A vector (pSAT1A-Cosmc). The rare-cutting enzyme AscI was used to transfer the expression cassette from pSAT1A-Cosmc to pPZP-RCS2. A codon-optimized clone of *Drosophila melanogaster* C1GALT1 was synthesized by GeneArt Gene Synthesis with flanking XbaI and BamHI restriction sites. The XbaI/BamHI fragment was cloned into the binary expression vector pPT2M (pPT2M-C1GALT1) (24). In this vector, expression is under control of the cauliflower mosaic virus 35S promoter. A clone (IMAGE ID: 3925036) coding for human α2,3-sialyltransferase (ST3Gal-I) was purchased from Source BioScience, amplified with oligos S3GAL1-F1 (5′-TATACTCGAGATGGTGACCCTGCGGAAG-3′)/S3GAL1-R1 (5′-TATAGGATCCTCATCTCCCCTTGAAGATC-3′), XhoI/BamHI-digested, and cloned into pSAT6A to generate vector pSAT6A-ST3Gal-I. A clone (IMAGE ID: 6844232) coding for *Mus musculus* α2,6-sialyltransferase (ST6GalNAc-III/IV) was purchased from Source BioScience. The corresponding open reading frame was amplified by PCR using oligos ST6GAL-F1 (5′-TATACTCGAGATGAAGGCCCCGGGCCGC-3′)/ST6GAL-R1 (5′-TATAGGATCCCTACTTGGCCCTCCAGGAC-3′), XhoI/BamHI-digested, and cloned into pSAT1A vector (pSAT1A-ST6GalNAc). To reduce the number of constructs during the agroinfiltration procedure, ST3Gal-I and ST6GalNAc-III/IV were expressed from one construct together with the Golgi CMP-sialic acid transporter (CST) (25). CST was amplified from the *M. musculus* cDNA clone using oligos CST-F1 (5′-TATACTCGAGATGGCTCCGGCGAGAGAAAATG-3′) and CST-R1 (5′-TATAGGATCCTCACACACCAATGATTCTCTC-3′) and cloned into XhoI/BamHI-digested pSAT3A vector (pSAT3A-CST). To obtain the construct for simultaneous expression of the three proteins, the expression cassette of pSAT1A-ST6GalNAc was removed by AscI digestion and cloned into the AscI site of pPZP-RCS2, the expression cassette from pSAT6A-ST3Gal-I was removed by digestion with the homing endonuclease PI-PspI and cloned into the PI-PspI site of pPZP-RCS2, and the CST expression cassette was inserted into the I-SceI site of pPZP-RCS2.

All binary vectors except the magnICON constructs were transformed into the *Agrobacterium tumefaciens* strain UIA 143. All magnICON constructs were transformed into strain GV3101 pMP90. Bacterial suspensions were infiltrated at the following optical densities (OD_{600}): magnICON constructs, 0.1; binary vectors, 0.05. In all co-expression experiments the respective bacterial suspensions were mixed 1:1 before infiltration.

Plant Material—N. benthamiana wild-type and glycoengineered ΔXTFT line (26) were grown in a growth chamber at 22 °C with a 16-h light/8-h dark photoperiod. All constructs were expressed by agroinfiltration of leaves as described in detail previously (26).

Analysis of N- and O-Linked Glycans—EPO-Fc was purified from infiltrated leaves by affinity chromatography using rProteinA-Sepharose ™ Fast Flow (GE Healthcare) as described in detail previously (22). Purified EPO-Fc was separated by

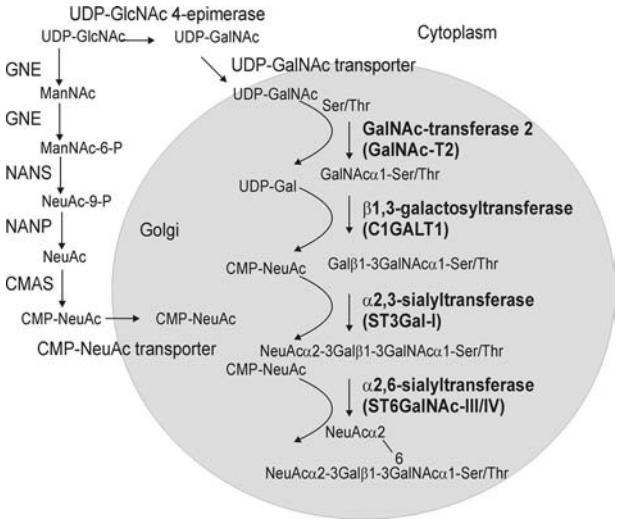


FIGURE 1. **Schematic representation of the pathway for the formation of disialylated *O*-glycans in plants.** UDP-GalNAc formation and CMP-NeuAc biosynthesis start from the nucleotide sugar UDP-GlcNAc. The steps for conversion of UDP-GlcNAc to CMP-NeuAc and transport of CMP-NeuAc to the Golgi have been engineered in plants previously (25, 34). The specific steps that seem absolutely required for sialylated mucin-type *O*-glycan biosynthesis are depicted in *bold*. *GNE*, UDP-*N*-acetylglucosamine 2-epimerase/*N*-acetylmannosamine kinase; *NANS*, *N*-acetylneuraminic acid phosphate synthase; *CMAS*, CMP-*N*-acetylneuraminic acid synthetase; *NANP*, *N*-acetylneuraminate-9-phosphate phosphatase. Conversion of NeuAc-9-P to NeuAc is very likely carried out by an endogenous plant enzyme.

SDS-PAGE, and protein bands were stained with Coomassie Brilliant Blue or analyzed by immunoblotting using anti-EPO (MAB2871, R&D Systems, Minneapolis, MN) or anti-human IgG (anti-Fc) (Promega, Mannheim, Germany) antibodies. The corresponding band was excised from the gel and double-digested with trypsin and endoglucosaminidase C (Glu-C) (sequencing grade, Roche Applied Science). Glycopeptide analysis was carried out by liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) as described in detail previously (27, 28).

RESULTS

Strategy for Sialylated Mucin-type O-Glycan Engineering in Plants—Biosynthesis of sialylated mucin-type core 1 structures in *N. benthamiana* requires enzymatic reactions as well as transport steps from the cytosol to the Golgi lumen. For the transfer of GalNAc residues to Ser/Thr, which is the initiation step of mucin-type *O*-glycosylation and presumably takes place in the Golgi apparatus, UDP-GlcNAc must be converted to UDP-GalNAc in the cytosol, and UDP-GalNAc must be transported into the Golgi lumen where it is used by a polypeptide: *N*-acetylgalactosaminyltransferase (GalNAc-T) to attach GalNAc residues to specific *O*-glycosylation sites within a protein (Fig. 1). Constructs for expression of the corresponding proteins are available from a previous study (19), and similar constructs have been successfully used by another group to initiate *O*-glycan formation in plants (29).

One of the most common *O*-glycan extensions is the formation of the core 1 structure (T-antigen). Core 1 β1,3-galactosyltransferase (C1GALT1, also called core 1 or T-synthase) catalyzes the transfer of a single galactose residue from UDP-Gal to GalNAcα1-*O*-Ser/Thr to generate Galβ1–3GalNAcα1-*O*-Ser/Thr. In humans, this particular step requires the co-expression of a specific chaperone termed COSMC (30). COSMC binds to human C1GALT1 in the endoplasmic reticulum and is required for correct folding of the glycosyltransferase and subsequently for localization and activity in the Golgi. Invertebrates and plants lack COSMC homologs, and consequently, we hypothesized that human C1GALT1 expression depends on the co-expression of COSMC. Alternatively C1GALT1 from *Drosophila* or another invertebrate could be functional in plants without any additional chaperone.

Core 1 structures are frequently capped with sialic acid residues. This terminal modification step requires the co-expression of the respective sialyltransferases, *e.g.* transfer of *N*-acetylneuraminic acid (NeuAc) in α2,3-linkage to galactose by α2,3-sialyltransferase (ST3Gal-I) and transfer of NeuAc in α2,6-linkage to the GalNAc residue, which is catalyzed by α2,6-sialyltransferase (ST6GalNAc-III/IV) (Fig. 1). All used constructs are listed in Fig. 2.

Expression of Recombinant EPO-Fc in Plants—We previously showed production of human EPO fused to the Fc domain of an IgG molecule (EPO-Fc) by transient expression in the glycoengineered *N. benthamiana* line ΔXTFT (22). All three *N*-glycosylation sites of ΔXTFT-produced recombinant EPO are occupied by complex *N*-glycans almost completely lacking the plant-specific β1,2-xylose and core α1,3-fucose residues. In another study it was reported that this EPO-Fc fusion protein expressed in *N. benthamiana* using the viral-based magnICON expression system accumulated to ~1–5 mg per kg fresh weight and was biologically active (31). Here we used the same magnICON construct for expression of EPO-Fc in ΔXTFT (Fig. 3) and purified the recombinant protein by protein A affinity chromatography. As previously reported (22, 31), a 55-kDa band corresponding to the full-length EPO-Fc as well as a smaller 30-kDa band corresponding to the molecular mass of free Fc was obtained (Fig. 3). The 55-kDa EPO-Fc was excised from the SDS-PAGE gel, and trypsin/Glu-C double-digested peptides were analyzed by LC-ESI-MS, which gave the expected peptide AISPPDAASAAPLR and the cleavage variant EAISPPDAASAAPLR (Fig. 4*a*). Tandem mass spectrometry of peptides confirmed the identity of the non-glycosylated peptide comprising Ser-126 in recombinant human EPO (data not shown). A smaller peak corresponding to a non-glycosylated peptide with a hydroxylated proline residue (mass shift +16 Da) was observed, indicating that one of the three proline residues that are adjacent to Ser-126 is modified by an endogenous prolyl 4-hydroxylase. Furthermore a peak indicative of a double-hydroxylated peptide occurred. Each substitution of Pro by Hyp brought about a lowering of the retention time by about 2 min. As a result, the various peptides are spread over a considerable elution time range. Because hydroxyproline residues are sites for plant-specific *O*-glycosylation, we looked for the presence of arabinose chains, but no signals indicating the presence of one to four arabinose residues were found.

Initiation of Mucin-type O-Glycosylation on Recombinant Plant-produced EPO-Fc—Previously it has been shown by lectin blotting that initiation of mucin-type *O*-glycan formation on a recombinant reporter protein expressed in *N. benthamiana* can be achieved by co-expression of human GalNAc-T2, a

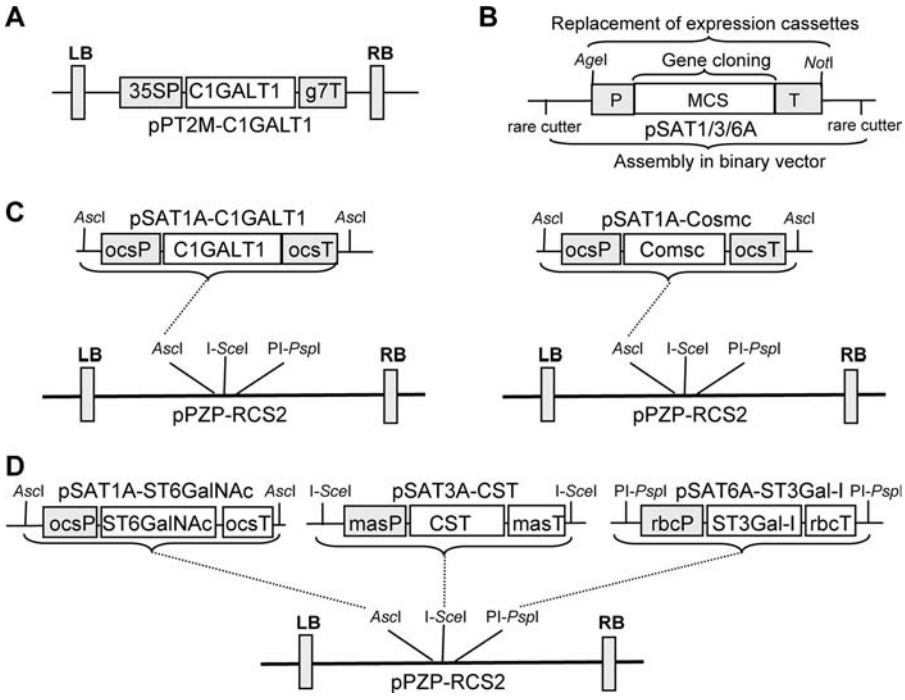


FIGURE 2. **Schematic representation of newly generated vectors.** *A*, binary vector for the expression of the *Drosophila* C1GALT1 is shown. *B*, shown are structural features of the pSAT series of auxiliary vectors (pSAT1A, pSAT3A, and pSAT6A) for the assembly of promoter-gene-terminator cassettes. Rare-cutting enzymes flanking each pSAT vector are used to transfer the expression cassettes into the expression vector pPZP-RCS2. *C*, shown is the outline of the cloning strategy for expression of human C1GALT1 and its chaperone COSMC. *D*, shown is a schematic representation of the cloning strategy for the multiple gene expression vector. The CST, ST3Gal-I, and ST6GALNAc-III/IV open reading frames were cloned into different pSAT auxiliary vectors and were then sequentially assembled in pPZP-RCS2 using specific rare-cutting enzymes. In the final constructs all three proteins are expressed under different promoter and terminator sequences. *35SP*, cauliflower mosaic virus 35S promoter; *g7T*, *Agrobacterium* gene 7 terminator; *ocsP*, octopine synthase promoter; *ocsT*, octopine synthase terminator; *rbcP*, rubisco small subunit promoter; *rbcT*, rubisco small subunit terminator; *masP*, manopine synthase promoter; *masT*, manopine synthase terminator; *LB*, left border sequence; *RB*, right border sequence.

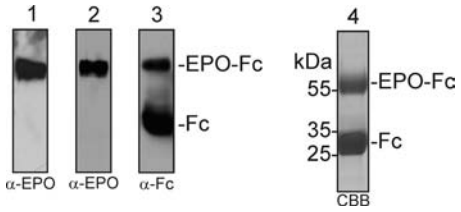


FIGURE 3. **SDS-PAGE and immunoblot analysis of EPO-Fc expressed in *N. benthamiana*.** Total soluble protein extracts were subjected to SDS-PAGE followed by immunoblotting with anti-EPO antibodies (α-*EPO*) (1). Eluates from the protein A purification were separated by SDS-PAGE and analyzed by immunoblotting with α-EPO (2) with anti human IgG antibodies (α-*Fc*) (3) or by Coomassie Brilliant Blue staining (4). Representative images are shown.

Y. enterocolitica UDP-GlcNAc 4-epimerase, and a *C. elegans* UDP-GlcNAc/UDP-GalNAc transporter (19). To investigate whether the co-expression of this *O*-glycosylation machinery leads to the initiation of *O*-GalNAc formation on recombinant EPO-Fc, we expressed all four proteins transiently in ΔXTFT. LC-ESI-MS analysis of trypsin/Glu-C digested EPO-Fc (Fig. 4*b*) showed a peak that corresponds to the mass of an *O*-glycosylated peptide. Tandem mass spectrometry of peptides confirmed the identity of the glycosylated peptide comprising Ser-126 in recombinant human EPO (Fig. 4*c*). These data show the successful initiation of mucin-type *O*-glycosylation on recombinant EPO-Fc.

Production of Core 1 Structures on Recombinant Plant-produced EPO-Fc—The next *O*-glycan elongation step is the transfer of a galactose residue in β1,3-linkage to produce the core 1 structure on EPO-Fc. In our first attempt we transiently expressed human C1GALT1 with human COSMC and the three constructs for *O*-GalNAc formation together with EPO-Fc in *N. benthamiana*. Analysis of trypsin/Glu-C double-digested peptides by mass spectrometry revealed the presence of a peak corresponding to a galactosylated *O*-GalNAc structure (Fig. 5, *a* and *b*). However, the major peak corresponded to the *O*-GalNAc-modified peptide, indicating that the transfer of galactose was not very efficient. Next, we chose to transiently express *Drosophila* C1GALT1 together with GalNAc-T2 and the UDP-GlcNAc/UDP-GalNAc transporter as well as the UDP-GlcNAc 4-epimerase and analyzed the glycopeptide containing Ser-126 from EPO-Fc (Fig. 5*c*). In contrast to human C1GalT1, expression of *Drosophila* C1GALT1 resulted in an almost complete conversion of *O*-GalNAc to Galβ1–3GalNAc. These data suggest that human C1GALT1 is not properly expressed in plants or is rapidly degraded, presumably because the C1GALT1-COSMC interaction is less efficient when expressed in a heterologous system. *Drosophila* C1GALT1 on the other hand was very effective in synthesis of core 1 structures on recombinant EPO-Fc.

Decoration of Core 1 Structures on Recombinant Plant-produced EPO-Fc with Sialic Acids—The final steps in our engineering approach were the capping of the core 1 structure by sialic acid. Co-expression of the *O*-glycosylation machinery for the core 1 structure formation with human ST3Gal-I, *M. musculus* ST6GalNAc-III/IV, and the sialylation machinery resulted in the formation of peaks corresponding to the incorporation of one and two sialic acid residues into the core 1

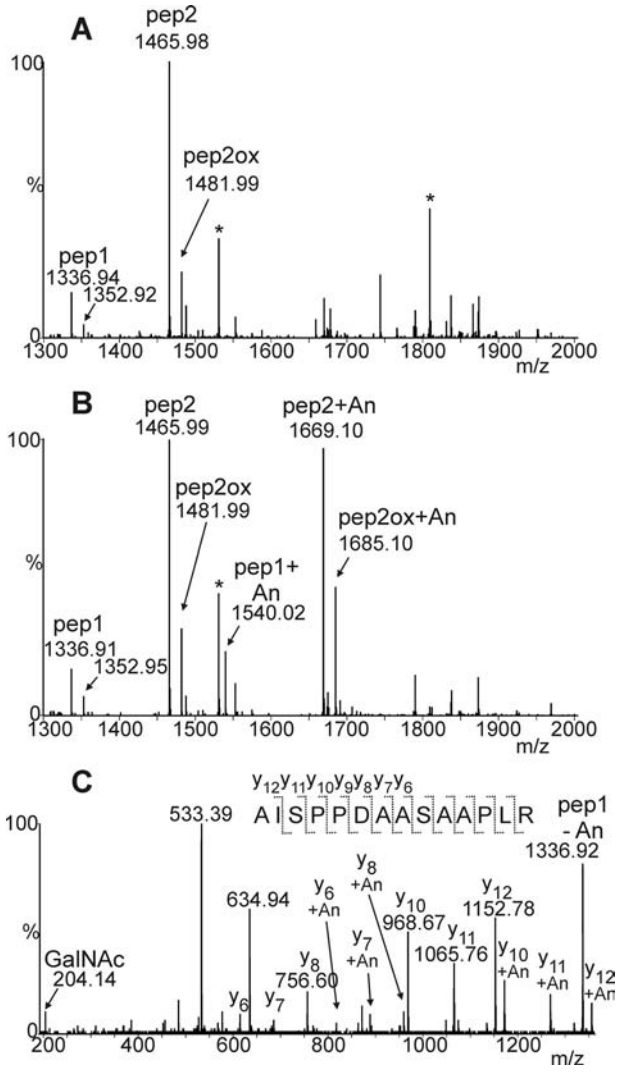


FIGURE 4. **Initiation of *O*-GalNAc formation on Ser-126 of recombinant plant-produced EPO-Fc.** Mass spectra of trypsin and endoproteinase Glu-C double-digested EPO-Fc expressed in *N. benthamiana* ΔXTFT line are shown. *A*, shown is a spectrum of the EPO-Fc Ser-126-containing peptide(s) in the absence of any *O*-glycan machinery; due to partial miscleavage, two peptides containing Ser-126 are generated, pep1 (AISPPDAASAAPLR) and pep2 (EAISPPDAASAAPLR). *B*, shown is a spectrum of the EPO-Fc Ser-126 peptides from plants co-expressing GalNAc-T2, UPD-GlcNAc 4-epimerase, and UDP-GlcNAc/UDP-GalNAc transporter with EPO-Fc. The presence of glycosylated peptides is indicated (+*An* indicates the presence of GalNAc residues). The presence of peptides with hydroxyproline residues (Pro to Hyp conversion: +16 Da, *e.g.* pep2ox) is indicated by *arrows*. The glycosylated versions of this peptide as well as the double- hydroxylated peptide eluted outside of the displayed time window. *Asterisks* denote the presence of co-eluting peptides or contaminations. *C*, shown is the Y-ion series of LC-ESI-MS/MS fragmentation experiment of the *O*-glycosylated EPO-Fc peptide (+*An* indicates the presence of a single GalNAc residue).

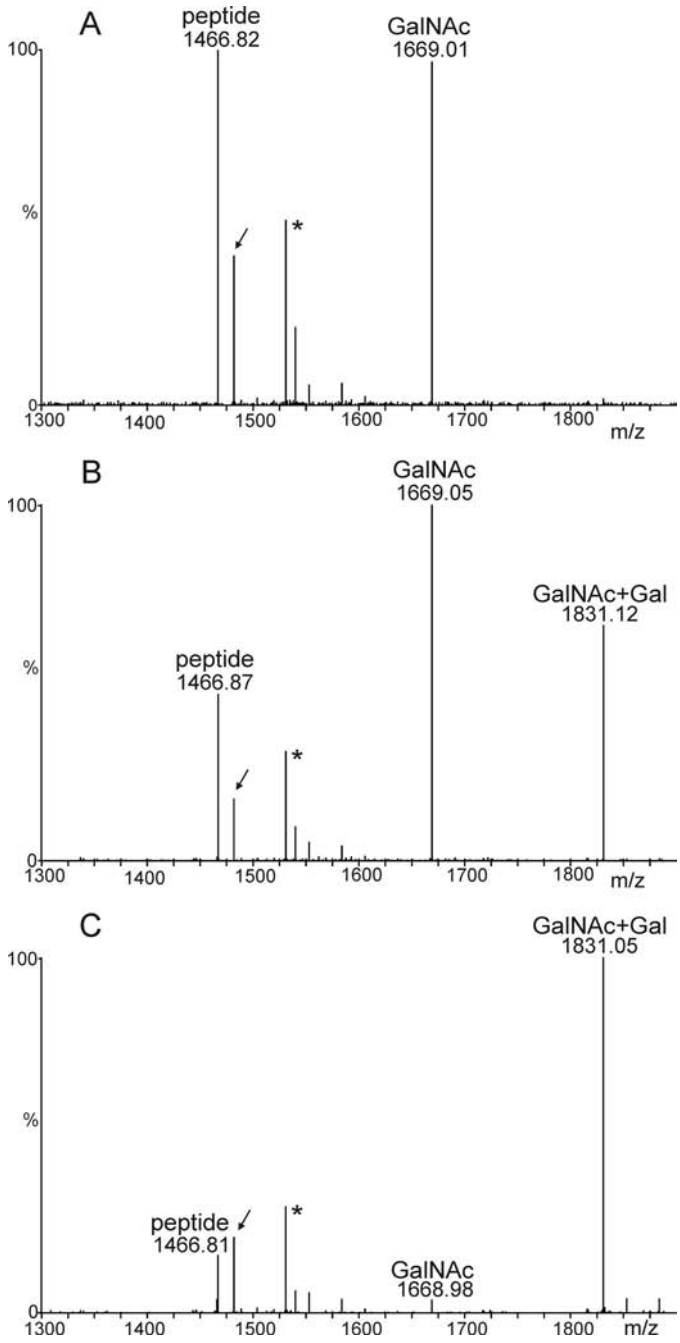


FIGURE 5. **Generation of T-antigen (Galβ1–3GalNAc) by co-expression of C1GALT1.** *A*, co-expression of EPO-Fc with the machinery for *O*-GalNAc formation is shown. *B*, co-expression of EPO-Fc with the machinery for *O*-GalNAc formation, human C1GALT1, and its specific chaperone COSMC is shown. *C*, co-expression of EPO-Fc with the machinery for *O*-GalNAc formation and *Drosophila* C1GALT1 is shown. The spectra show the *O*-glycosylated EPO-Fc Ser-126-containing peptide (^{117}EAISPPDAASAAPLR^{131}). The *arrow* points at the peptide with one hydroxyproline residue. Unrelated peaks are denoted by an *asterisk*.

structure (Fig. 6*a*), thus showing that sialylated core 1 structures can be produced on recombinant EPO-Fc in our glycoengineered plants.

The requirement of heterologous expression of UDP-GlcNAc 4-epimerase and a UDP-GlcNAc/UDP-GalNAc transporter has been controversial (19, 29). To investigate whether these two proteins are necessary for *O*-glycan initiation and modification on EPO-Fc, we compared the engineered *O*-glycan structures in the presence and absence of *Y. enterocolitica* UDP-GlcNAc 4-epimerase and a *C. elegans* UDP-GlcNAc/UDP-GalNAc transporter. We found that expression of human GalNAc-T2 is sufficient for the generation of *O*-GalNAc on EPO-Fc (supplemental Fig. S1). Moreover, core 1 and sialylated core 1 structures were also generated in the absence of an additionally co-expressed epimerase or transporter (supplemental Fig. S1), suggesting that endogenous plant proteins can produce sufficient amounts of UDP-GalNAc and perform UDP-

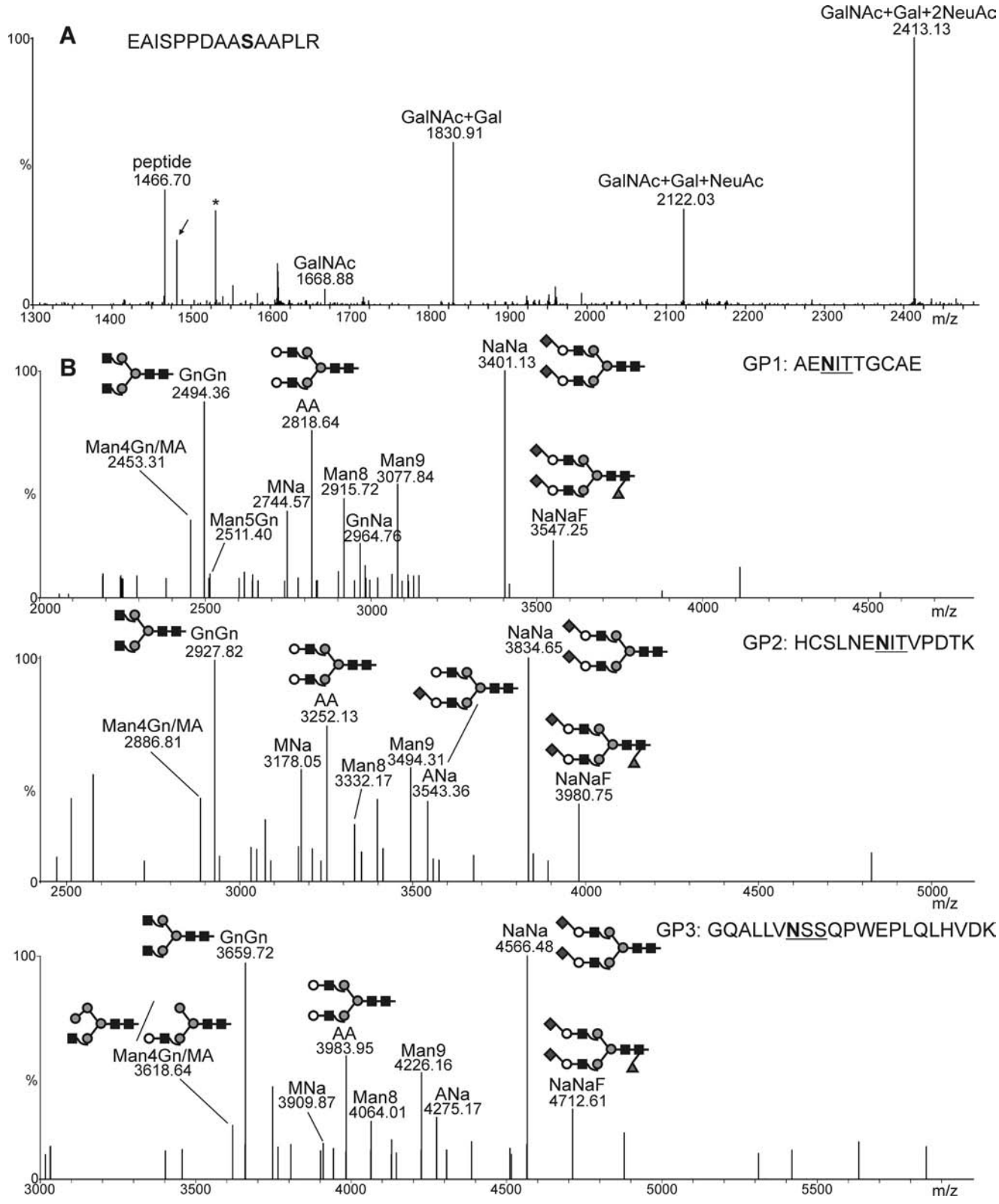


FIGURE 6. **Generation of disialylated *O*-glycans on EPO-Fc.** *A*, shown are co-expression of EPO-Fc with the mammalian pathway for CMP-sialic acid synthesis, its transport to the Golgi, and the *O*-glycosylation machinery including UDP-GlcNAc 4-epimerase, UDP-GlcNAc/UDP-GalNAc transporter, GalNAc-T2, *Drosophila* C1GALT1, ST3Gal-I, and ST6GalNAc-III/IV. The LC-ESI-MS spectrum depicts the presence of all possible glycosylation variants up to a doubly sialylated core-1 *O*-glycan structure. The presence of a peptide corresponding to the hydroxylation of a single proline residue is indicated by an *arrow*. The *asterisk* denotes the presence of a contamination. *B*, analysis of the three EPO-Fc peptides containing *N*-linked glycans is shown. *N*-Glycan analysis was carried out by LC-ESI-MS of tryptic/Glu-C double-digested EPO-Fc.

GalNAc transport into the Golgi where it is used as donor substrate by GalNAc-T2.

Structural Analysis of N-Linked Glycans on Recombinant Plant-produced EPO-Fc—Efficient *N*- and *O*-glycosylation requires sufficient pools of the nucleotide sugar UDP-GlcNAc, which serves as the precursor for conversion into UDP-GalNAc or CMP-sialic acid (Fig. 1). Furthermore, UDP-GlcNAc is also used as the donor substrate by endogenous *N*-acetylglucosaminyltransferases during synthesis and processing of *N*-glycans. In addition, introduction of the Golgi-resident mammalian glycosyltransferases and the additionally expressed Golgi transporter could disturb the organization of endogenous *N*-glycosylation enzymes in the Golgi and subsequently affect sialylation of EPO-Fc *N*-glycans. To investigate whether co-expression of the six proteins for *O*-glycosylation (UDP-GlcNAc 4-epimerase, UDP-GlcNAc/UDP-GalNAc transporter, GalNAc-T2, *Drosophila* C1GALT1, ST6GalNAc-III/IV and ST3Gal-I) affect the *N*-glycan processing on EPO-Fc, we co-expressed them together with the pathway for CMP-sialic acid biosynthesis, transport, and transfer of sialic acid to β1,4-galactosylated *N*-glycans (additional six proteins: UDP-*N*-acetylglucosamine 2-epimerase/*N*-acetylmannosamine kinase, *N*-acetylneuraminic acid phosphate synthase, CMP-*N*-acetylneuraminic acid synthetase, CMP-NeuAc transporter, β1,4-galactosyltransferase, and α2,6-sialyltransferase) (25). In total, 12 proteins were co-expressed transiently in ΔXTFT plants together with EPO-Fc. All three glycopeptides from EPO-Fc displayed as a major peak a disialylated biantennary *N*-glycan structure that corresponds to $NeuAc_2Gal_2GlcNAc_2Man_3GlcNAc_2$ (*NaNa*) (Fig. 6*b*), showing that the generation of disialylated mucin-type *O*-glycans does not interfere with *N*-glycan engineering.

To rule out the possibility of cross-talk between the two different glycosylation pathways, we analyzed whether sialylated *N*-glycans can be generated by the two sialyltransferases (ST3Gal-I and ST6GalNAc-III/IV) used for *O*-glycan modification. In the absence of α2,6-sialyltransferase, no sialylated *N*-glycans were generated, showing that the two mammalian *O*-glycan-specific sialyltransferases retained their specificity when expressed in plants (supplemental Fig. S2). In addition, the *O*-GalNAc structure was not modified by β1,4-galactosyltransferase and α2,6-sialyltransferase (supplemental Fig. S3).

DISCUSSION

Here we report for the first time the generation of sialylated mucin-type *O*-glycans on a plant-produced recombinant glycoprotein intended for therapeutic use in humans. In addition to the disialyl core 1 structure, we could also very efficiently produce the T-antigen on EPO by expression of the *Drosophila* C1GALT1, and our data show that *O*-glycan modifications do not interfere with the *N*-glycosylation capacity of glycoengineered plants. Notably, the leaves of infiltrated plants expressing the whole *O*-glycosylation machinery did not display any obvious phenotype. Biomass production and expression of recombinant EPO-Fc was also unchanged, which highlights the flexibility of our *N. benthamiana*-based transient expression platform. However, although *N. benthamiana* plants are amenable to rapid scale-up and high biomass yield can be obtained within a short time period, the expression levels of EPO-Fc and protocols for downstream processing have to be optimized to make the production system commercially more attractive.

The first attempts toward formation of mammalian-like mucin-type *O*-glycosylation in plants were performed by expression of the machinery to transfer GalNAc residues to *O*-glycosylation sites of recombinant proteins expressed in *N. benthamiana* (19, 29). In the latter study the incorporation of GalNAc residues was confirmed by structural analysis of the corresponding glycopeptides. Several Ser/Thr residues of a reporter protein containing a MUC1 tandem repeat motif were modified by the transient co-expression of a GlcNAc 4-epimerase together with human GalNAc-T2 and GalNAc-T4 to achieve high density *O*-glycosylation. A similar approach was used to generate *O*-GalNAc modifications by stable expression of the genes in *Arabidopsis thaliana* (32). Here, we demonstrate that *O*-glycosylation initiation on EPO-Fc requires only GalNAc-T2, and we extend the mucin-type *O*-GalNAc structure further by transfer of galactose and decoration of the terminal positions with sialic acid. Especially the generation of sialic acid containing *O*-glycans is challenging because plants lack the whole machinery to produce CMP-NeuAc and related sialic acid (33, 34). Consequently, the nucleotide sugar CMP-NeuAc must be generated in the cytosol/nucleus (25) and transported to the proper Golgi compartment where it is transferred to *N*- and *O*-linked glycans.

Furthermore, because our studies were carried out in our glycoengineered ΔXTFT line (26), we could combine *N*- and *O*-glycan engineering approaches to generate a platform capable for human-type glycoform production. Similar extensive modifications of two interacting glycosylation pathways have not been reported so far in other engineered expression hosts (35–38).

The production of defined homogenous carbohydrate structures on recombinant proteins is a prerequisite for in-depth structure function analysis of glycoforms. The role of *O*-linked glycans is typically investigated by mutagenesis of the respective *O*-glycosylation sites (39) but not by analysis of the contribution of individual glycan structures. Here, in this proof-of-concept study the *O*-glycosylation site of EPO served as a reporter protein to show that customized biosynthesis of *O*-glycan structures is possible. The role of the *O*-linked glycan for EPO is not very well understood. In one study Ser-126 was changed to Gly and the mutated EPO protein was not properly secreted, which led to the conclusion that the *O*-glycan might play a role in efficient secretion of EPO (40). However, a block in *O*-glycan formation in a mutant CHO line did not affect the secretion of EPO or its *in vitro* and *in vivo* biological activities (12). In comparison, our plant-based production system offers greater flexibility with respect to production of defined glycoforms and may provide a useful tool to discover novel functions of *O*-linked glycans on recombinant EPO or other *O*-glycosylated proteins in the future.

In contrast to CHO cells, *N. benthamiana* plants contain a limited protein glycosylation repertoire and are, therefore, better amenable to glycoengineering by expression of only those mammalian proteins (*e.g.* α2,6-sialyltransferase) that generate the desired glycan modification. Due to the presence of the

endogenous *O*-glycosylation machinery heterogeneity is frequently observed on recombinant proteins produced in mammalian cells. Analysis of *O*-glycosylation in an EPO-Fc CHO cell-based producer line (41) showed that under different growth conditions, only ~50% of all *O*-glycosylation sites on EPO-Fc are occupied by glycans, and similar to our findings, mono- as well as disialylated *O*-glycans were detected. Consequently strategies are employed to reduce the background glycosylation and the observed glycan heterogeneity in mammalian cells (42). Apart from differences in glycosyltransferase expression and substrate specificity, the number of homogenous glycan structures might also be increased by strategies that increase the availability of nucleotide sugars like CMP-NeuAc, *e.g.* by avoiding feedback regulation (43) or elimination of unwanted nucleotide sugar hydrolyzing enzyme activities (44).

One potential hurdle for the broad use of plants for the manufacturing of *O*-glycosylated therapeutic proteins or peptides could be the presence of hydroxyproline residues close to *O*-glycosylation sites. These hydroxyproline residues can be further modified by the endogenous plant *O*-glycosylation machinery, leading to the attachment of arabinose residues or arabinogalactans. Arabinose residues close to *O*-glycosylation sites have been described for maize-produced IgA and for a mucin-tandem repeat containing peptide expressed in *N. benthamiana* (45, 46). These plant-specific glycans could cause immunogenic or allergenic reactions in humans (47). Although small amounts of hydroxyproline residues were found on recombinant EPO-Fc, we could not detect any incorporation of arabinose to adjacent hydroxyproline residues or other plant-specific *O*-glycosylation. These data are consistent with reports on expression of recombinant EPO in *Physcomitrella patens* and a mucin-tandem repeat containing protein in *N. benthamiana* (29, 48). Future research directions will focus on strategies to prevent hydroxyproline formation by using inhibitors for prolyl 4-hydroxylases or specific knockdown/knock-out of the corresponding enzymes (32, 49, 50).

The approach presented here is the first step toward the generation of elongated and branched *O*-glycans in plants and is not restricted to recombinant EPO-Fc but can very likely also be used for other glycoproteins. In the future we will extend the glycosylation repertoire of *N. benthamiana* further by engineering of additional *O*-glycosylation steps like the formation of other core structures (17) and the production of glycopeptide vaccines containing well defined *O*-glycan structures like T, Tn, and sialyl-Tn that are associated with cancer cells and might be used to break self-antigen tolerance when applied as a vaccine. Moreover, optimization of the expression of the different non-plant proteins and their subcellular localization will help to achieve even more homogenous *N*- and *O*-glycan structures on recombinant proteins and thus will make glycoengineered *N. benthamiana* an exciting new host for the production of glycoprotein pharmaceuticals.

Acknowledgments—We thank Pia Gattinger, Eva Liebminger, and Christiane Veit, Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Vienna, Austria for excellent technical support.

REFERENCES

1. Zimran, A., Brill-Almon, E., Chertkoff, R., Petakov, M., Blanco-Favela, F., Muñoz, E. T., Solorio-Meza, S. E., Amato, D., Duran, G., Giona, F., Heitner, R., Rosenbaum, H., Giraldo, P., Mehta, A., Park, G., Phillips, M., Elstein, D., Altarescu, G., Szleifer, M., Hashmueli, S., and Aviezer, D. (2011) Pivotal trial with plant cell-expressed recombinant glucocerebrosidase, taliglucerase alfa, a novel enzyme replacement therapy for Gaucher disease. *Blood* **118,** 5767–5773
2. Maxmen, A. (2012) Drug-making plant blooms. *Nature* **485,** 160
3. Ghaderi, D., Zhang, M., Hurtado-Ziola, N., and Varki, A. (2012) Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. *Biotechnol. Genet. Eng. Rev.* **28,** 147–175
4. Jefferis, R. (2012) Isotype and glycoform selection for antibody therapeutics. *Arch. Biochem. Biophys.* **526,** 159–166
5. Loos, A., and Steinkellner, H. (2012) IgG-Fc glycoengineering in non-mammalian expression hosts. *Arch. Biochem. Biophys.* **526,** 167–173
6. Fukuda, M. N., Sasaki, H., Lopez, L., and Fukuda, M. (1989) Survival of recombinant erythropoietin in the circulation: the role of carbohydrates. *Blood* **73,** 84–89
7. Egrie, J. C., Dwyer, E., Browne, J. K., Hitz, A., and Lykos, M. A. (2003) Darbepoetin alfa has a longer circulating half-life and greater *in vivo* potency than recombinant human erythropoietin. *Exp. Hematol.* **31,** 290–299
8. Son, Y. D., Jeong, Y. T., Park, S. Y., and Kim, J. H. (2011) Enhanced sialylation of recombinant human erythropoietin in Chinese hamster ovary cells by combinatorial engineering of selected genes. *Glycobiology* **21,** 1019–1028
9. Sasaki, H., Bothner, B., Dell, A., and Fukuda, M. (1987) Carbohydrate structure of erythropoietin expressed in Chinese hamster ovary cells by a human erythropoietin cDNA. *J. Biol. Chem.* **262,** 12059–12076
10. Tsuda, E., Kawanishi, G., Ueda, M., Masuda, S., and Sasaki, R. (1990) The role of carbohydrate in recombinant human erythropoietin. *Eur. J. Biochem.* **188,** 405–411
11. Hokke, C. H., Bergwerff, A. A., Van Dedem, G. W., Kamerling, J. P., and Vliegenthart, J. F. (1995) Structural analysis of the sialylated *N*- and *O*-linked carbohydrate chains of recombinant human erythropoietin expressed in Chinese hamster ovary cells. Sialylation patterns and branch location of dimeric *N*-acetyllactosamine units. *Eur. J. Biochem.* **228,** 981–1008
12. Wasley, L. C., Timony, G., Murtha, P., Stoudemire, J., Dorner, A. J., Caro, J., Krieger, M., and Kaufman, R. J. (1991) The importance of *N*- and *O*-linked oligosaccharides for the biosynthesis and *in vitro* and *in vivo* biologic activities of erythropoietin. *Blood* **77,** 2624–2632
13. Delorme, E., Lorenzini, T., Giffin, J., Martin, F., Jacobsen, F., Boone, T., and Elliott, S. (1992) Role of glycosylation on the secretion and biological activity of erythropoietin. *Biochemistry* **31,** 9871–9876
14. Elliott, S., Egrie, J., Browne, J., Lorenzini, T., Busse, L., Rogers, N., and Ponting, I. (2004) Control of rHuEPO biological activity. The role of carbohydrate. *Exp. Hematol.* **32,** 1146–1155
15. Kodama, S., Tsujimoto, M., Tsuruoka, N., Sugo, T., Endo, T., and Kobata, A. (1993) Role of sugar chains in the *in vitro* activity of recombinant human interleukin 5. *Eur. J. Biochem.* **211,** 903–908
16. Mattu, T. S., Pleass, R. J., Willis, A. C., Kilian, M., Wormald, M. R., Lellouch, A. C., Rudd, P. M., Woof, J. M., and Dwek, R. A. (1998) The glycosylation and structure of human serum IgA1, Fab, and Fc regions and the role of *N*-glycosylation on Fc alpha receptor interactions. *J. Biol. Chem.* **273,** 2260–2272
17. Tarp, M. A., and Clausen, H. (2008) Mucin-type *O*-glycosylation and its potential use in drug and vaccine development. *Biochim. Biophys. Acta.* **1780,** 546–563
18. Ju, T., Otto, V. I., and Cummings, R. D. (2011) The Tn antigen-structural simplicity and biological complexity. *Angew. Chem. Int. Ed. Engl.* **50,** 1770–1791
19. Daskalova, S. M., Radder, J. E., Cichacz, Z. A., Olsen, S. H., Tsaprailis, G., Mason, H., and Lopez, L. C. (2010) Engineering of *N. benthamiana* L. plants for production of *N*-acetylgalactosamine-glycosylated proteins to-

ward development of a plant-based platform for production of protein therapeutics with mucin-type *O*-glycosylation. *BMC Biotechnol.* **10,** 62
20. Karimi, M., Inzé, D., and Depicker, A. (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* **7,** 193–195
21. Marillonnet, S., Thoeringer, C., Kandzia, R., Klimyuk, V., and Gleba, Y. (2005) Systemic *Agrobacterium tumefaciens*-mediated transfection of viral replicons for efficient transient expression in plants. *Nat. Biotechnol.* **23,** 718–723
22. Castilho, A., Gattinger, P., Grass, J., Jez, J., Pabst, M., Altmann, F., Gorfer, M., Strasser, R., and Steinkellner, H. (2011) *N*-Glycosylation engineering of plants for the biosynthesis of glycoproteins with bisected and branched complex *N*-glycans. *Glycobiology* **21,** 813–823
23. Chung, S. M., Frankman, E. L., and Tzfira, T. (2005) A versatile vector system for multiple gene expression in plants. *Trends Plant Sci.* **10,** 357–361
24. Strasser, R., Bondili, J. S., Schoberer, J., Svoboda, B., Liebminger, E., Glössl, J., Altmann, F., Steinkellner, H., and Mach, L. (2007) Enzymatic properties and subcellular localization of *Arabidopsis* β-*N*-acetylhexosaminidases. *Plant Physiol.* **145,** 5–16
25. Castilho, A., Strasser, R., Stadlmann, J., Grass, J., Jez, J., Gattinger, P., Kunert, R., Quendler, H., Pabst, M., Leonard, R., Altmann, F., and Steinkellner, H. (2010) *In planta* protein sialylation through overexpression of the respective mammalian pathway. *J. Biol. Chem.* **285,** 15923–15930
26. Strasser, R., Stadlmann, J., Schähs, M., Stiegler, G., Quendler, H., Mach, L., Glössl, J., Weterings, K., Pabst, M., and Steinkellner, H. (2008) Generation of glyco-engineered *Nicotiana benthamiana* for the production of monoclonal antibodies with a homogeneous human-like *N*-glycan structure. *Plant Biotechnol. J.* **6,** 392–402
27. Stadlmann, J., Pabst, M., Kolarich, D., Kunert, R., and Altmann, F. (2008) Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. *Proteomics* **8,** 2858–2871
28. Kolarich, D., Jensen, P. H., Altmann, F., and Packer, N. H. (2012) Determination of site-specific glycan heterogeneity on glycoproteins. *Nat. Protoc.* **7,** 1285–1298
29. Yang, Z., Drew, D. P., Jørgensen, B., Mandel, U., Bach, S. S., Ulvskov, P., Levery, S. B., Bennett, E. P., Clausen, H., and Petersen, B. L. (2012) Engineering mammalian mucin-type *O*-glycosylation in plants. *J. Biol. Chem.* **287,** 11911–11923
30. Ju, T., and Cummings, R. D. (2002) A unique molecular chaperone Cosmc required for activity of the mammalian core 1 β3-galactosyltransferase. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 16613–16618
31. Nagels, B., Van Damme, E. J., Callewaert, N., Zabeau, L., Tavernier, J., Delanghe, J. R., Boets, A., Castilho, A., and Weterings, K. (2012) Biologically active, magnICON®-expressed EPO-Fc from stably transformed *Nicotiana benthamiana* plants presenting tetra-antennary *N*-glycan structures. *J. Biotechnol.* **160,** 242–250
32. Yang, Z., Bennett, E. P., Jørgensen, B., Drew, D. P., Arigi, E., Mandel, U., Ulvskov, P., Levery, S. B., Clausen, H., and Petersen, B. L. (2012) Toward stable genetic engineering of human *O*-glycosylation in plants. *Plant Physiol.* **160,** 450–463
33. Zeleny, R., Kolarich, D., Strasser, R., and Altmann, F. (2006) Sialic acid concentrations in plants are in the range of inadvertent contamination. *Planta* **224,** 222–227
34. Castilho, A., Pabst, M., Leonard, R., Veit, C., Altmann, F., Mach, L., Glössl, J., Strasser, R., and Steinkellner, H. (2008) Construction of a functional CMP-sialic acid biosynthesis pathway in *Arabidopsis*. *Plant Physiol.* **147,** 331–339
35. Amano, K., Chiba, Y., Kasahara, Y., Kato, Y., Kaneko, M. K., Kuno, A., Ito, H., Kobayashi, K., Hirabayashi, J., Jigami, Y., and Narimatsu, H. (2008) Engineering of mucin-type human glycoproteins in yeast cells. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 3232–3237
36. Hamilton, S. R., Davidson, R. C., Sethuraman, N., Nett, J. H., Jiang, Y., Rios, S., Bobrowicz, P., Stadheim, T. A., Li, H., Choi, B. K., Hopkins, D., Wischnewski, H., Roser, J., Mitchell, T., Strawbridge, R. R., Hoopes, J., Wildt, S., and Gerngross, T. U. (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* **313,** 1441–1443
37. Harrison, R. L., and Jarvis, D. L. (2006) Protein *N*-glycosylation in the baculovirus-insect cell expression system and engineering of insect cells to produce "mammalianized" recombinant glycoproteins. *Adv. Virus Res.* **68,** 159–191
38. Jacobs, P. P., Geysens, S., Vervecken, W., Contreras, R., and Callewaert, N. (2009) Engineering complex-type N-glycosylation in Pichia pastoris using GlycoSwitch technology. *Nat. Protoc.* **4,** 58–70
39. Badirou, I., Kurdi, M., Legendre, P., Rayes, J., Bryckaert, M., Casari, C., Lenting, P. J., Christophe, O. D., and Denis, C. V. (2012) *In vivo* analysis of the role of *O*-glycosylations of von Willebrand factor. *PLoS One* **7,** e37508
40. Dubé, S., Fisher, J. W., and Powell, J. S. (1988) Glycosylation at specific sites of erythropoietin is essential for biosynthesis, secretion, and biological function. *J. Biol. Chem.* **263,** 17516–17521
41. Taschwer, M., Hackl, M., Hernández Bort, J. A., Leitner, C., Kumar, N., Puc, U., Grass, J., Papst, M., Kunert, R., Altmann, F., and Borth, N. (2012) Growth, productivity and protein glycosylation in a CHO EpoFc producer cell line adapted to glutamine-free growth. *J. Biotechnol.* **157,** 295–303
42. Steentoft, C., Vakhrushev, S. Y., Vester-Christensen, M. B., Schjoldager, K. T., Kong, Y., Bennett, E. P., Mandel, U., Wandall, H., Levery, S. B., and Clausen, H. (2011) Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat. Methods* **8,** 977–982
43. Bork, K., Reutter, W., Weidemann, W., and Horstkorte, R. (2007) Enhanced sialylation of EPO by overexpression of UDP-GlcNAc 2-epimerase/ManAc kinase containing a sialuria mutation in CHO cells. *FEBS Lett.* **581,** 4195–4198
44. Van Dijk, W., Lasthuis, A. M., Trippelvitz, L. A., and Muilerman, H. G. (1983) Increased glycosylation capacity in regenerating rat liver is paralleled by decreased activities of CMP-N-acetylneuraminate hydrolase and UDP-galactose pyrophosphatase. *Biochem. J.* **214,** 1003–1006
45. Karnoup, A. S., Turkelson, V., and Anderson, W. H. (2005) *O*-Linked glycosylation in maize-expressed human IgA1. *Glycobiology* **15,** 965–981
46. Pinkhasov, J., Alvarez, M. L., Rigano, M. M., Piensook, K., Larios, D., Pabst, M., Grass, J., Mukherjee, P., Gendler, S. J., Walmsley, A. M., and Mason, H. S. (2011) Recombinant plant-expressed tumor-associated MUC1 peptide is immunogenic and capable of breaking tolerance in MUC1.Tg mice. *Plant Biotechnol. J.* **9,** 991–1001
47. Leonard, R., Petersen, B. O., Himly, M., Kaar, W., Wopfner, N., Kolarich, D., van Ree, R., Ebner, C., Duus, J. Ø, Ferreira, F., and Altmann, F. (2005) Two novel types of *O*-glycans on the mugwort pollen allergen Art v 1 and their role in antibody binding. *J. Biol. Chem.* **280,** 7932–7940
48. Weise, A., Altmann, F., Rodriguez-Franco, M., Sjoberg, E. R., Bäumer, W., Launhardt, H., Kietzmann, M., and Gorr, G. (2007) High level expression of secreted complex glycosylated recombinant human erythropoietin in the *Physcomitrella* Delta-fuc-t Delta-xyl-t mutant. *Plant Biotechnol. J.* **5,** 389–401
49. Velasquez, S. M., Ricardi, M. M., Dorosz, J. G., Fernandez, P. V., Nadra, A. D., Pol-Fachin, L., Egelund, J., Gille, S., Harholt, J., Ciancia, M., Verli, H., Pauly, M., Bacic, A., Olsen, C. E., Ulvskov, P., Petersen, B. L., Somerville, C., Iusem, N. D., and Estevez, J. M. (2011) *O*-Glycosylated cell wall proteins are essential in root hair growth. *Science* **332,** 1401–1403
50. Moriguchi, R., Matsuoka, C., Suyama, A., and Matsuoka, K. (2011) Reduction of plant-specific arabinogalactan-type *O*-glycosylation by treating tobacco plants with ferrous chelator 2,2′-dipyridyl. *Biosci. Biotechnol. Biochem.* **75,** 994–996

Appendix III

Schneider, J.D., Castilho, A., **Neumann, L.**, Altmann, F., Loos, A., Kannan, L., Mor, T.S., Steinkellner, H. (2013) Expression of human butyrylcholinesterase with a glycosylation profile resembling the plasma derived orthologue. Biotechnol. J. 9, 501-510. doi: 10.1002/biot.201300229.

Biotechnology Journal
www.biotechnology-journal.com

Biotechnol. J. 2014, 9, 501–510

DOI 10.1002/biot.201300229

Research Article

Expression of human butyrylcholinesterase with an engineered glycosylation profile resembling the plasma-derived orthologue

Jeannine D. Schneider[1], Alexandra Castilho[1], Laura Neumann[2], Friedrich Altmann[2], Andreas Loos[1], Latha Kannan[3], Tsafrir S. Mor[3] and Herta Steinkellner[1]

[1]Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Vienna, Austria
[2]Department of Chemistry, University of Natural Resources and Life Sciences, Vienna, Austria
[3]The Biodesign Institute, Arizona State University, Tempe, Arizona, USA

Human butyrylcholinesterase (BChE) is considered a candidate bioscavenger of nerve agents for use in pre- and post-exposure treatment. However, the presence and functional necessity of complex *N*-glycans (i.e. sialylated structures) is a challenging issue in respect to its recombinant expression. Here we transiently co-expressed BChE cDNA in the model plant *Nicotiana benthamiana* with vectors carrying the genes necessary for *in planta* protein sialylation. Site-specific sugar profiling of secreted recombinant BChE (rBChE) collected from the intercellular fluid revealed the presence of mono- and di-sialylated *N*-glycans, which largely resembles to the plasma-derived orthologue. Attempts to increase that sialylation content of rBChE by the over-expression of an additional glycosylation enzyme that generates branched *N*-glycans (i.e. β1,4-*N*-acetylglucosaminyl-transferase IV), allowed the production of rBChE decorated with tri-sialylated structures (up to 70%). Sialylated and non-sialylated plant-derived rBChE exhibited functional in vitro activity comparable to that of its commercially available equine-derived counterpart. These results demonstrate the ability of plants to generate valuable proteins with designed sialylated glycosylation profiles optimized for therapeutic efficacy. Moreover, the efficient synthesis of carbohydrates present only in minute amounts on the native protein (tri-sialylated *N*-glycans) facilitates the generation of a product with superior efficacies and/or new therapeutic functions.

Received	12 JUN 2013
Revised	26 SEP 2013
Accepted	14 OCT 2013
Accepted article online	16 OCT 2013

Supporting information available online

Keywords: Butyrylcholinesterase · Glycoengineering · Plants · Recombinant biopharmaceuticals · Sialic acid

1 Introduction

Organophosphorous pesticides (OP) and chemical warfare nerve agents target acetylcholinesterase, a key enzyme regulating neurotransmission in cholinergic synapses. The high affinity of human cholinesterases to OPs prompted the evaluation of applying cholinesterases for prophylaxis and post-exposure treatment against chemical warfare nerve agents and has had considerable success in pre-clinical studies [1-4]. In particular, the use of the serum paralog of acetylcholinesterase, butyrylcholinesterase (BChE), has proceeded into advanced development and has undergone Phase I clinical trials (www.ClinicalTrials.gov Identifier: NCT00333528). BChE was the bioscavenger of choice in these first-in-human studies because BChE has a long half-life in serum and

Correspondence: Prof. Herta Steinkellner, Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria
E-mail: herta.steinkellner@boku.ac.at

Abbreviations: ΔXT/FT, xylosyl- and fucosyl-transferase knock-down *Nicotiana benthamiana;* **Asn**, asparagine; **BChE**, butyrylcholinesterase; **BChE$_{sia}$**, sialylated butyrylcholinesterase; **cGMP**, current good manufacturing practices; **dpi**, days post infiltration; **GnTII**, α1,6-mannosyl-β1,2-*N*-acetylglucosaminyltransferase; **GnTIV**, α1,3-mannosyl-β1,4-*N*-acetylglucosaminyltransferase; **hBChE**, human BChE; **IF**, intercellular fluid; **LC-ESI-MS**, liquid chromatography – electrospray ionization – mass spectrometry; **Na**, *N*-Acetylneuraminic acid; **OP**, organophosphate; **PAct2**, *Arabidopsis thaliana* actin 2 promoter; **rBChE**, recombinant butyrylcholinesterase; **TMV**, tobacco mosaic virus; **TSP**, total soluble protein; **WT**, wild-type

because it is readily available from outdated blood-banked plasma. However, blood-bankded plasma is an expensive source of BChE, and is limited supply and unlikely to meet the expected market demands. Recombinant DNA technology, e.g. mammalian cell-culture or plant-based production systems, could offer a solution to this problem [5, 6]. Despite the promise of such systems in providing the necessary quantities of the bioscavenger, the less-than-optimal pharmacokinetic properties of recombinant human BChE as compared to its plasma-derived counterpart remain a major hurdle [4, 6]. Human serum BChE undergoes post-translational modifications that endow it with a very long circulatory half-life of up to two weeks [7]. Factors that influence the pharmacokinetic profile are mass and glycosylation. In plasma BChE circulates as a heavily glycosylated tetramer. Each monomer carries sialylated glycans on all of its 9 *N*-glycosylation sites, increasing its molecular mass from 66 to 85 kDa [8]. Glycan sialylation is crucial for conferring the long circulatory residence time to serum cholinesterase [9, 10].

Human BChE has been recombinantly expressed in various systems [4, 5, 9, 11, 12]. However in most cases this yields a protein with reduced circulatory half-life. For example, BChE expressed in mammary glands of transgenic goats and excreted into their milk is mostly dimeric, contains a great proportion of under-sialylated glycans, and has a half-life of 6.5 h in the guinea pig model [11] as compared to 64.6 h of human plasma-derived BChE in the same model [13]. Similarly, BChE expressed in plants was enzymatically identical to the human enzyme [5], effectively protect animals against OP challenge, but also has reduced circulatory residence compared to the native counterpart [4]. This is most probably due to the fact that the plant expressed human enzyme was only 50% tetrameric and bore non-sialylated glycans. Improvement of the pharmacokinetic properties could be achieved by forced tetramerization of the protein [14], post-purification decoration of the enzyme with polyethylene-glycol [4, 11], or by chemical polysialylation [6]. Nonetheless, these in vitro modifications complicate the production and formulation process and significantly increase the cost of producing BChE. We have hypothesized that an efficient and cost-effective alternative is to adapt the production system to enable post-translational modification processes allowing the production of proteins that more closely resemble their human counterparts.

Particularly, plant production systems seem highly relevant – not only bearing lower risks of mammalian pathogens, lower cultivation costs, and allowing easier and less-costly scale-up – but also demonstrating an outstanding degree of tolerance to changes in the *N*-glycosylation pathway. Indeed, several examples of *in planta* syntheses of complex human-like glycoforms, including sialylation, have been established [15, 16].

In this study we set out to produce recombinant butyrylcholinesterase (rBChE) with a glycosylation profile that resembles that of the plasma-derived protein. We used plants as our expression system and focused on generating efficiently sialylated glycovariants. The potent magnICON plant-viral-based system was used to transiently express the enzyme in leaves of the model plant *N. benthamiana*. rBChE derived from the intercellular fluid (IF, secreted fraction) was analyzed by liquid chromatography – electrospray ionization – mass spectrometry (LC-ESI-MS). Site-specific sugar profiling showed the presence of complex *N*-glycans. Moreover, sialylated rBChE, carrying mono-, di-, and tri-sialylated glycans, was obtained by co-expression of BChE with the six mammalian genes required for *in planta* protein sialylation [17] and *N*-glycan branching [18]. In addition, differently glycosylated plant-derived rBChE exhibited normal enzymatic activity.

2 Materials and Methods

2.1 Construction of magnICON vectors

The signal peptide from the barley α-amylase gene for targeting proteins to the secretory pathway was inserted into the tobacco mosaic virus (TMV) based magnICON vector pICH26211 (kindly supplied by V. Klimyuk) resulting in pICHa26211. Dicotyledonous plant-codon optimized cDNA of the human BChE gene [5] was used as template for PCR amplification using primers with flanking *BsmB*I restriction sites (used primers BChE-F1: 5'-GCACGTCTCAAGGTGAGGATGACATCATCATTGC-3' and BChE-R1: 5'-GCACGTCTCAAAGCCTAGAGACCC ACACAGCTCTCC-3'); the resulting fragment corresponds to the full-length protein without its signal peptide. It was cloned into the Zero Blunt® TOPO® vector (Invitrogen®) and the identity was confirmed by DNA sequence analysis. After digestion with *BsmB*I, the DNA fragment was cloned into the pICHa26211 *Bsa*I sites. The resulting vector pBChE (Fig. 1A) was transformed into the *Agrobacterium tumefaciens* strain GV3101 pMP90 by electroporation.

2.2 Binary vectors for *N*-glycan engineering

For *N*-glycoengineering two binary vectors containing the six genes necessary for protein sialylation were used [17, 19]. In addition, a binary vector was generated for over-expression of *Arabidopsis thaliana* α1,6-mannosyl-β1,2-*N*-acetylglucosaminyltransferase (GnTII). cDNA was obtained by RT-PCR with the primer pair Ath-GNTII-12F (5'-TATATCTAGASATGGCAAATCTTTGGAAGAAGC-3') and Ath-GnTII-14R (5'-TATAGGATCCTCATGGAGATG CACTGCTACTG-3') as described [20]. The *Xba*I and *BamH*I restriction sites created by the primers facilitated the subsequent subcloning of the *Xba*I/*BamH*I fragment into the binary expression vector pPT2M [21]. The result-

ing binary vector (pGnTII) was transformed into agrobacterium strain UIA 143 pMP90. The plasmid containing isoform A of human α1,3-mannosyl-β1,4-*N*-acetylglucosaminyl-transferase (GnTIV) used in this investigation has been described previously [18]. Detailed information about the expression cassettes used in this investigation is displayed in Supporting information, Table S1.

2.3 Plant material and protein expression

N. benthamiana wild-type (WT) and xylosyl- and fucosyltransferase knock-down *Nicotiana benthamiana* (ΔXT/FT) mutant plants, which lack plant-specific β1,2-xylose and core α1,3-fucose residues [22], were cultivated in a growth chamber with a constant temperature of 24°C, 60% humidity, and a 16 h light/ 8 h dark photoperiod. 4- to 5-week old plants were used for agroinfiltration [17, 22]. Modifications of the *N*-glycosylation profile of rBChE was achieved by co-infiltrating binary vectors carrying mammalian genes necessary for protein sialylation [17, 19] and synthesis of tri-antennary *N*-glycans [18]. Agrobacteria containing magnICON constructs were infiltrated at an optical density at 600 nm (OD_{600}) of 0.1, while binary constructs were mixed at an OD_{600} of 0.05. The GnTII construct was added to the infiltration mixture at OD_{600} 0.8.

2.4 Isolation of intercellular fluid (IF) and total soluble protein (TSP)

Infiltrated leaves harvested 3–12 days post infiltration (dpi) were submerged in buffer (100 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 2 mM EDTA). Vacuum was applied twice for 5 min. The proteins secreted into the apoplast were collected by centrifugation at 900 × *g* for 15 min (1 leaf yielded approximately 1 mL of IF). Prior to analysis, the IF was concentrated 10-fold with 10 or 30-kDa cut-off Spin-X® UF centrifugal concentrators (Corning®). All steps were carried out at 4°C. Protein quantification was done using the Bradford-based BioRad® Protein Assay following the manufacturer's instructions. For total soluble protein extraction, infiltrated leaf material (250 mg) was submerged in liquid nitrogen. The material was ground in a swing mill (Retsch®, MM2000) for 2 min at amplitude 60 and then a double volume (v/w) of 1×PBS was added. After incubation on ice for 10 min the extracts were centrifuged (6000 rpm for 5 min at 4°C) and the supernatant was stored at -20°C for further analysis. Total soluble protein quantification was done using the Bradford-based BioRad® Protein Assay following the manufacturer's instructions.

2.5 Immunoblot analysis and quantification of rBChE

TSP- and IF-derived proteins were mixed with reducing 4× Laemmli buffer, incubated for 5 min at 95°C, and separated by 10% SDS-PAGE. The fractionated proteins were used for western blotting and Coomassie Brilliant Blue R-250 staining. Western blots were analyzed using polyclonal goat anti-BChE antibodies (N-15, Santa Cruz Biotechnology, Inc® sc-46803, diluted 1:300). Detection was performed using HRP-conjugated secondary antibodies (polyclonal anti-goat IgG-peroxidase antibody, A5420 Sigma Aldrich®, diluted 1:10000). SuperSignal West Pico Chemiluminescent was used as a substrate (Pierce). BChE expression levels were estimated by a semi-quantitative western blot analysis using plasma-derived BChE as a standard. Measurements were taken via the BioRad ChemiDoc Imaging System (BioRad).

2.6 Glycopeptide profiling

The *N*-glycan composition of rBChE was determined using LC-ESI-MS as described [23, 24]. In brief, the samples were submitted to denaturing SDS-PAGE, and the BChE-containing bands were S-alkylated, digested with trypsin, and eluted from the gel fragment with 50% acetonitrile. Samples were separated on a Biobasic C18 column (150 × 0.32 mm, Thermo Electron) with a gradient of 1-80% acetonitrile containing 65 mM ammonium formate at pH 3.0. Trypsin digestion allowed the analysis of five of the nine BChE glycopeptides (Gps): Gp2: W^{80}SDIWNATK^{88}; Gp4: N^{269}RTLNLAK^{276}; Gp5: E^{283}NETEIIK^{290}; Gp6: E^{368}NNSIITRK^{376}, and Gp7: D^{482}NYTKAEEILSR^{493}. Glycosites 8 and 9 (Y^{505}GNPNETQNNSTSWPVFK^{522}) are part of the same tryptic peptide. Therefore the discrimination of the glycosylation on the distinct glycosites is arduous, as already observed for human BChE (hBChE) in previous studies [8]. Detection of positive ions was facilitated by a quadrupole-time of flight (Q-TOF) Ultima Global mass spectrometer (Waters, Milford, MA, USA). The glycoforms were identified by summed and deconvoluted spectra of the glycopeptides' elution range. Peaks were labelled according to the ProGlycAn system (www.proglycan.com), with the ProGlycAn translator (http://www.proglycan.com/index.php?page=pga_translator) to convert abbreviations into Consortium for Functional Glycomics (CFG) type cartoons.

2.7 In vitro activity assay of rBChE

BChE activity was determined by a modified Ellman assay [25]. Total soluble protein extracts were analyzed. Hydrolysis of butyrylthiocholine (1 mM, B3253, Sigma Aldrich®) was evaluated in the presence of Ellman's reagent in sodium phosphate buffer (50 mM, pH 8.0) using 96-well microtiter plates. Substrate conversion rates were measured at room temperature with a Wallac® Victor2 plate reader at 405 nm every minute for 20 min, and thereafter at 30, 45, and 60 min. Equine serum-derived BChE (6 units, C1057, Sigma Aldrich®) served as a reference.

3 Results

3.1 Transient expression of BChE in *N. benthamiana*

The plant-expression optimized cDNA of human BChE was cloned into the TMV-based magnICON vector pICHα26211 (Fig. 1A). Agrobacteria carrying the plasmid were delivered to WT *N. benthamiana* leaves by agroinfiltration. Expression of the recombinant protein (rBChE) was analyzed in time-course experiments using TSP and IF extracted at different dpi. Western blot analysis of rBChE containing leaf extracts exhibited intensive signals at approximately 85 kDa, i.e. the expected size of the full-length monomeric protein. Signals could be detected as early as 3 dpi with increasing intensity up to 12 dpi (Fig. 1B). The expression level was approximately 10.5 µg/g leaf, which corresponds to approximately 1.3% of TSP (data not shown). Interestingly, at 10 dpi (and beyond), a 55 kDa degradation product was present in addition to the 85 kDa band (Fig. 1B). Degradation was more pronounced in rBChE collected from the IF. Here the degradation product appeared at 5 dpi (Fig. 1C), at an intensity similar to the full-length protein. LC-ESI-MS data (not shown) indicated that the bands ranging from 50 to 55 kDa were in fact truncated rBChE.

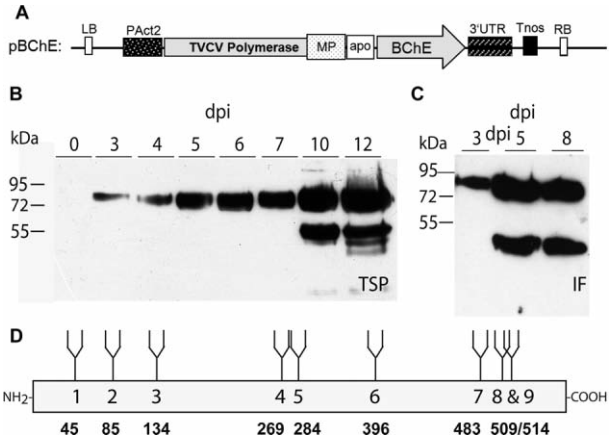


Figure 1. Expression of human BChE, (**A**) Schematic representation of the expression vector. BChE cDNA was cloned into a modified TMV-based magnICON vector (pICHα26211) resulting in pBChE. Apo: Signal peptide sequence from barley a amylase; BChE: Human butyrylcholinesterase sequence lacking the native signal peptide; LB: Left border; MP: Movement protein; PAct2: *Arabidopsis thaliana* actin 2 promoter; RB: Right border; Tnos: Nopaline synthase gene terminator; TVCV polymerase: Turnip vein clearing virus RNA-dependent RNA polymerase; 3'UTR: TVCV 3'-untranslated region; (**B** and **C**) Monitoring of rBChE expression in *N. benthamiana*. Expression of BChE was analyzed over a period of time ranging from 0-12 dpi. TSP (panel B, 4 µg protein per lane) and IF-derived proteins (**C**, 1 µg protein per lane) were fractionated using 10% SDS-PAGE and the proteins were detected using anti-BChE antibodies. Experiments were performed using four different leaves transiently expressing BChE and western blots were done in triplicates. (**D**) Protein backbone of human BChE. Numbers in the box (1-9) refer to the glycopeptides obtained upon trypsin digestion. Bold numbers below the box indicate the positions of the *N*-glycosylated asparagine residues within the amino acid sequence of the protein.

3.2 *N*-glycosylation profiling of plant-derived BChE

Human serum BChE is a heavily glycosylated 85 kDa protein with nine *N*-glycosylation sites that are decorated with sialylated *N*-glycans (Fig. 1D) [6]. The glycosylation profile of secreted rBChE expressed in WT *N. benthamiana* plants ($BChE_{WT}$) was determined by LC-ESI-MS analysis. IF-derived BChE-containing bands at 85 and 50 kDa were excised from denaturing gels and digested with trypsin. This allowed for analysis of five of the nine BChE glycosylation sites namely glycopeptides (Gps) 2, 4, 5, 6, and 7, corresponding to the *N*-glycosites Asn^{85}, Asn^{269}, Asn^{284}, Asn^{396}, and Asn^{483} (Fig. 1D). Although trypsin digestion allowed for analysis of most of the rBChE glycosites, detection of Gps 1 and 3 was not possible, and Gps 8 and 9 are part of the same tryptic peptide. Figure 2 shows the results for Gp2 and Gp7, which are representative of similar glycosylation patterns observed for other glycosites (data not shown). LC-ESI-MS profiles displayed

Table 1. Relative abundance in % of major glyco-structures detected on plant-derived BChE[a)]

Structure	BChE (WT)	BChE (ΔXT/FT)	$BChE_{sia}$ (WT)	$BChE_{sia}$ (ΔXT/FT)	$BChE_{trisia}$ (ΔXT/FT)
MGn_{iso}		7			
GnGn		88		9	
$MGnXF_{iso}$	15				
GnGnXF	81				
ANa_{iso}				7	
NaNa				71	26
NaNaF				6	
NaNaX			20		
NaNaXF			71		
Na[NaNa]					65
Σother ≤ 5 %	4	5	9	7	9

a) BChE was collected from intercellular fluid. Letters in brackets indicate the expression host, WT or ΔXT/FT plants. Σother ≤ 5%: sum of glyco-forms present at levels below 5%. The glycan structures are assigned using the ProGlycAn nomenclature (www.proglycan.com).

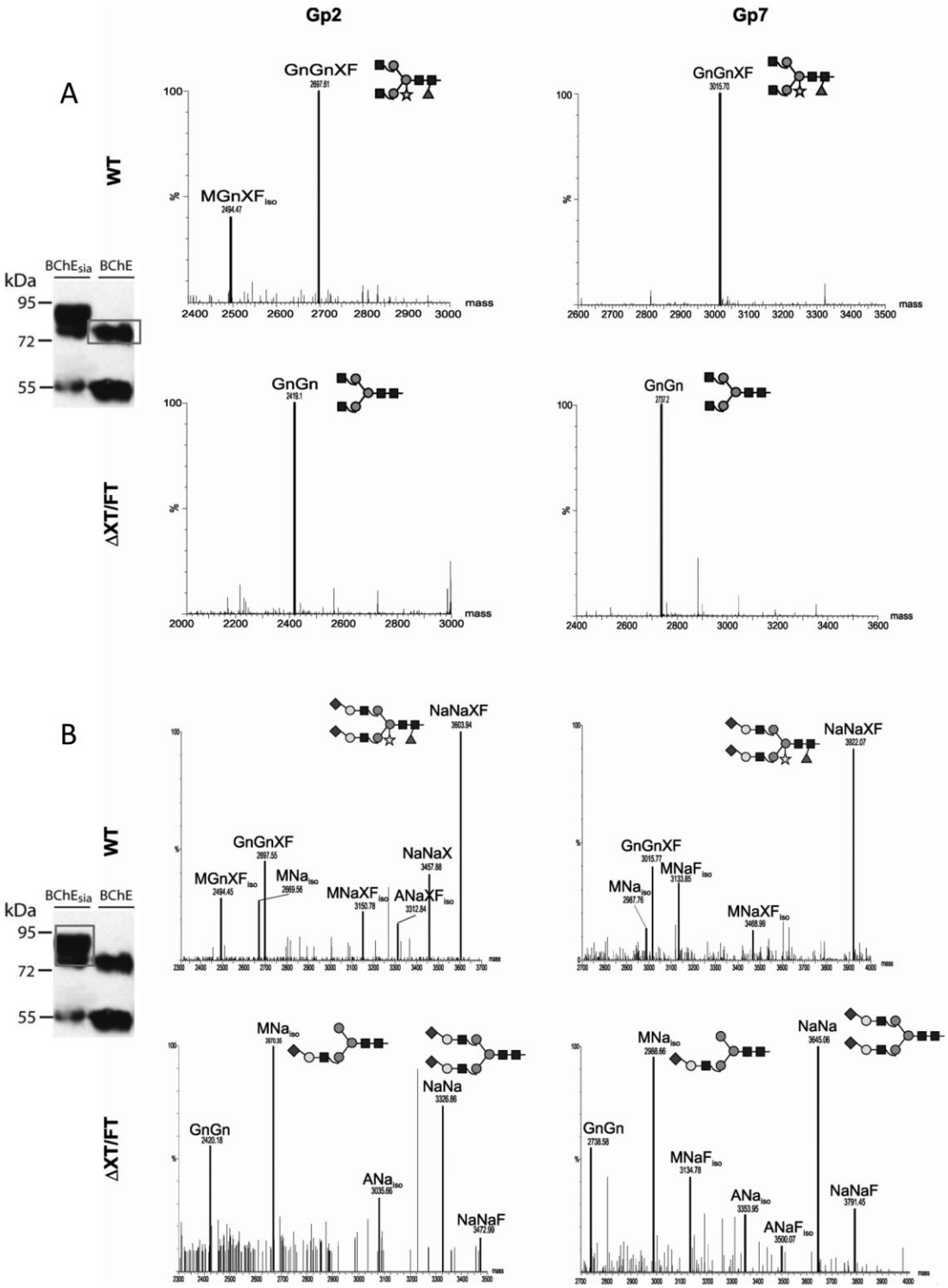


Figure 2. *N*-Glycan profiles of rBChE glycopeptides Gps 2 and 7 expressed in *N. benthamiana* WT and the glycosylation mutant ΔXT/FT. (**A**) Mass spectra of IF-derived BChE. (**B**) Mass spectra of IF-derived rBChE co-expressed with genes necessary for *in planta* sialylation (sialylated butyrylcholinesterase [$BChE_{sia}$]). Western blot analysis of $BChE_{sia}$ and BChE are shown on the left. Protein bands used for glycan analysis are boxed. Mass spectra were generated after liquid-chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) of glycopeptides (Gp) obtained upon trypsin digestion. Peaks are labeled in accordance with the ProGlycAn system (www.proglycan.com). The suffix "iso" at the end of glycan abbreviations denotes the probable presence of isomers. Major glyco-forms are shown as cartoons, for detailed description of cartoons see Supporting information, Fig. S3. Unassigned peaks are background originating from co-eluting peptides. Figure shows a representative glycoprofiling of BChE expressed in different plants out of an average of four experiments.

a single dominant *N*-glycan, namely GnGnXF, accompanied by incompletely processed MGnXF structures (Fig. 2A; Table 1). These are glyco-forms typical for secreted plant proteins. Plant-specific glycosylation, i.e. core β1,2-xylose and α1,3-fucose, was circumvented by expressing BChE in *N. benthamiana* ΔXT/FT, a mutant that lacks these two sugar residues [22] ($BChE_{\Delta XT/FT}$). $BChE_{\Delta XT/FT}$ exhibited a single human-like complex *N*-glycan, i.e. GnGn, devoid of plant-typical core β1,2-xylose and α1,3-fucose residues (Fig. 2A; Table 1). Glycosylation of $BChE_{\Delta XT/FT}$ was consistent on all glycosites analyzed (data not shown).

3.3 Sialylation of plant-derived BChE

We sought to generate terminal sialylation (sialylated butyrylcholinesterase [$BChE_{sia}$]) to further "humanize" the *N*-glycosylation of plant-derived BChE. To this end,

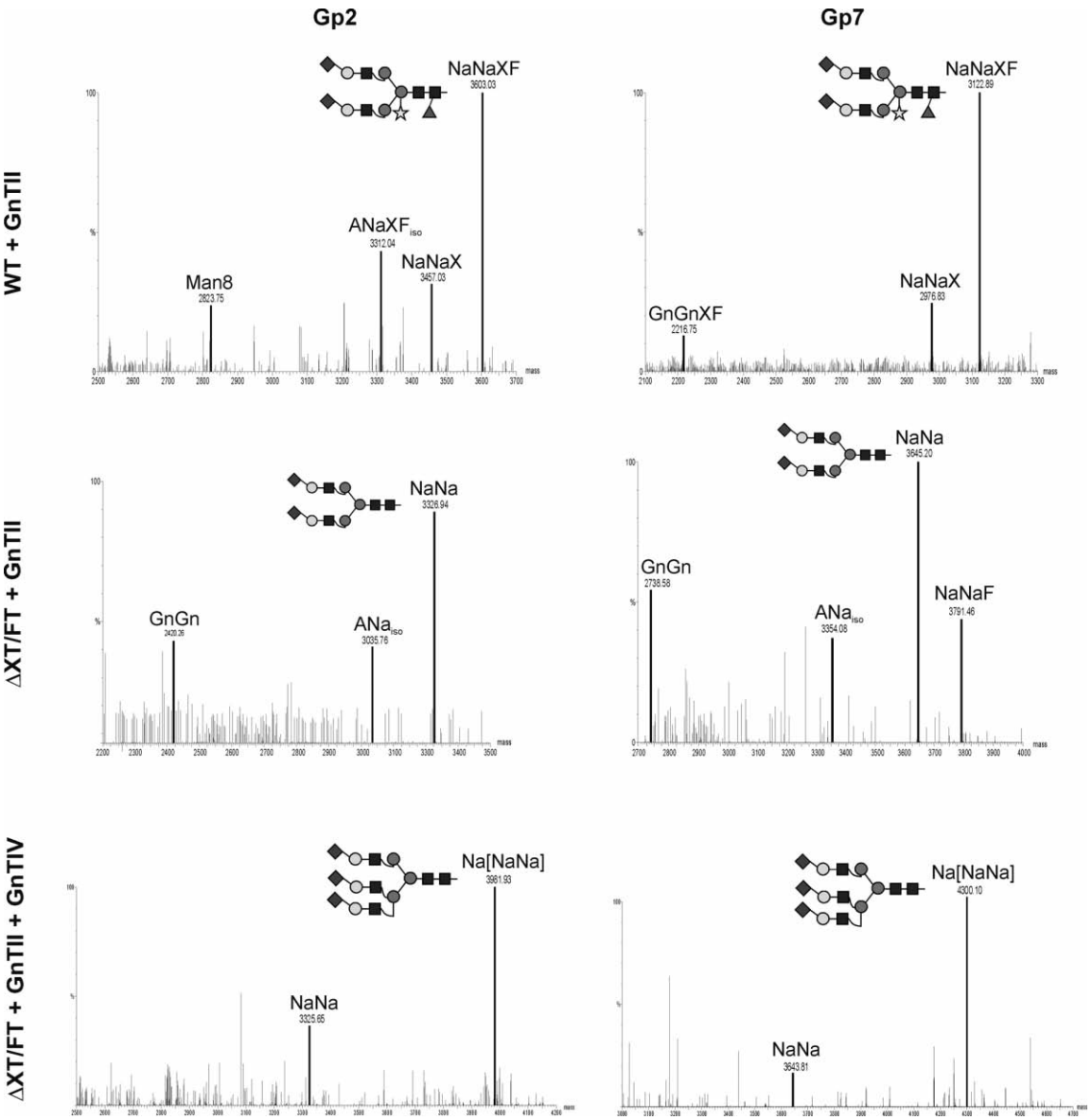


Figure 3. IF-derived *N*-Glycan profiles of IF-derived rBChE Gps 2 and 7 co-expressed with genes necessary for *in planta* sialylation ($BChE_{sia}$) and *N*-acetylglucosaminyltransferase II (GnTII). WT (top panel) and ΔXT/FT (middle and bottom panels) plants were used as expression hosts. Bottom panels show the MS profile of IF-derived $BChE_{trisia}$ obtained upon co-expression of genes for *in planta* sialyltion, GnTII, and GnTIV (glycosyltransferase responsible for branching). Mass spectra were generated after liquid-chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) of glycopeptides (Gp) obtained upon trypsin digestion. The suffix "iso" at the end of glycan abbreviations denotes the probable presence of isomers. Major glyco-forms are shown as cartoons, for detailed description of cartoons see Supporting information, Fig. S3. Unassigned peaks are background originating from co-eluting peptides. Figure shows a representative glycoprofiling of BChE expressed in different plants out of an average of four experiments.

pBChE was co-infiltrated with the genes necessary for *in planta* sialylation [17]. Two multi-gene vectors, each carrying three sialylation genes, were used [19]. IF was collected from infiltrated leaves at 5 dpi and analysed by western blot. A strong signal was obtained for $BChE_{sia}$ at a slightly higher position than $BChE_{WT}$ (Fig. 2). This size shift was most probably due to larger glyco-forms attached to $BChE_{sia}$. Interestingly, it appears that sialylation improves the stability of $BChE_{sia}$ in the apoplast, as the degradation product is less prominent than in non-sialylated $BChE_{WT}$ (Fig. 2). Glycoprofiles of IF-derived $BChE_{sia}$ expressed in WT plants displayed largely sialylated structures with the major peak representing di-sialylated glycosylation (NaNaXF; Fig. 2B). In addition, incompletely sialylated (MNa, MNaXF, ANaXF) and non-sialylated structures (GnGnXF, MGnXF) were detected (Fig. 2B). (Note, the biosynthetically more plausible isomer is given here and in the following cases). Upon co-expression of rBChE in ΔXT/FT plants together with the genes coding for the sialylation pathway, rBChE carrying efficiently sialylated oligosaccharides, lacking plant-specific xylose and core fucose, was generated. Unexpectedly, the proportion of incompletely processed sialylated MNa structures increased compared to WT plants (Fig. 2B). Incompletely processed *N*-glycans such as MNa possibly result from interference of the transiently introduced enzymes with the endogenous α1,6-mannosyl-β1,2-*N*-acetylglucosa-minyltransferase (GnTII). To address this issue, *Arabidopsis thaliana* GnTII was co-expressed with the mammalian genes needed for *in planta* sialylation. Incompletely processed carbohydrates were then indeed converted into fully processed structures, as seen by the increased appearance of di-sialylated *N*-glycans (Fig. 3, Table 1 and Supporting information, Fig. S1). Finally, we aimed to produce BChE with tri-antennary structures to increase levels of sialylation. Thus, we transiently co-expressed BChE with the genes necessary for *in planta* sialylation together with GnTII as well as GnTIV. The latter is responsible for the synthesis of tri-antennary glycans [18, 26]. In total nine foreign genes were simultaneously delivered into plant cells. Glycoprofiles of BChE expressed in this manner in ΔXT/FT plants ($BChE_{trisia}$) showed that tri-antennary sialylation could be achieved in a largely homogeneous manner (Na[NaNa], Fig. 3; Table 1) in addition to di-sialylated glycans. Again, all glycopeptides analyzed showed similar profiles (Supporting information, Fig. S2). Overall, 80–90% of BChE *N*-glycans carried sialic acid moieties upon co-expression of the sialylation pathway, irrespective of the expression host (WT and ΔXT/FT). Notably, small amounts of sialylated structures were also fucosylated (NaNaF) in some samples. Core fucose was most probably present due to incomplete silencing of the α1,3-fucosyltransferase in the ΔXT/FT plants [22]. Importantly, expression of BChE was in no case impaired by the addition of mammalian glycosylation proteins to the infiltration mixture.

3.4 In vitro activity

The enzymatic activity of plant-derived BChE was monitored with a modified Ellman assay [25]. TSP extracts were used to assay the BChE's ability to hydrolyze the butyrylthiocholine iodide substrate. The following glyco-forms were tested: $rBChE_{WT}$ (GnGnXF), $rBChE_{\Delta XT/FT}$ (GnGn), and $rBChE_{sia}$ (NaNa) (Fig. 4). The activity was compared with that of commercially available equine BChE. All glyco-forms were active, with the sialylated BChE showing activity most similar to the equine

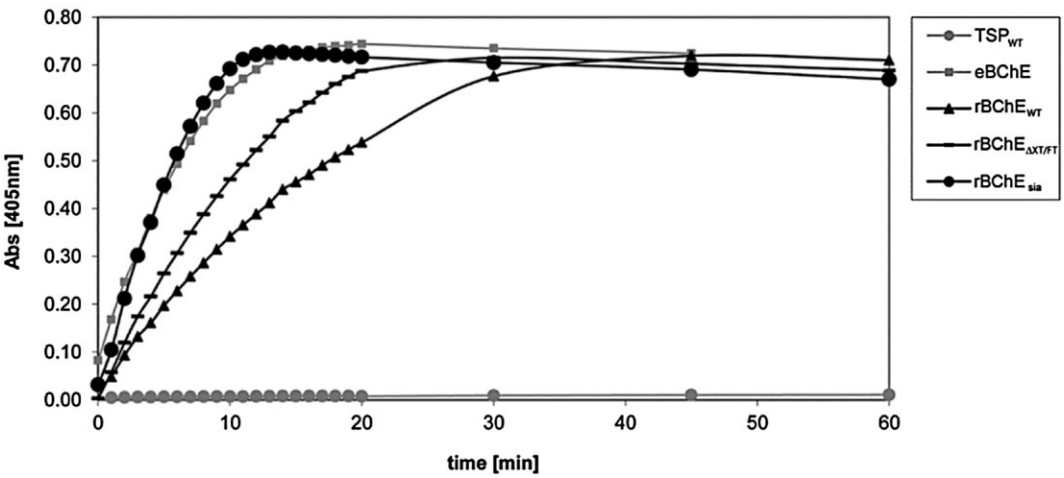


Figure 4. Evaluation of in vitro activity of plant-derived rBChE by a modified Ellman assay.Hydrolysis of butyrylthiocholine was evaluated in the presence of Ellman's reagent in sodium phosphate buffer. Substrate conversion rates were measured at different time points. Equine serum-derived BChE served as a reference. TSP_{WT}: TSP of wild-type plants infiltrated with agrobacteria that do not carry the BChE cDNA; eBChE: equine BChE; $BChE_{WT}$: BChE expressed in WT plants (main glyco-form GnGnXF); $BChE_{\Delta XT/FT}$: BChE expressed in ΔXT/FT plants (main glyco-form GnGn); $BChE_{sia}$: BChE co-expressed with the genes for *in planta* sialylation in ΔXT/FT plants (main glyco-form NaNa). All samples were analyzed in triplicates.

enzyme. That plant-proteins other than BChE hydrolyze the substrate could be excluded, because the TSP of leaves infiltrated with agrobacteria that carried "empty" plasmids did not show activity.

4 Discussion

In this study, we set out to produce rBChE in plants with a glycosylation profile that largely resembles the sialylated plasma-derived counterpart. Moreover we pursued novel in vivo glycoengineering strategies to generate glyco-forms with increased sialylation content, which potentially surpass the overall in vivo efficacy of the native counterparts. Using the magnICON-based transient expression system, we were able to generate the recombinant protein as early as 3 days after delivery of the expression vectors to the model plant *N. benthamiana*. This time-efficient procedure clearly provides advantages over the more time-consuming generation of rBChE using other (transgenic) expression platforms.

N-glycosylation analysis of secreted rBChE collected from leaf apoplast (IF-derived) exhibited a largely homogeneous glycosylation profile with a dominant glycoform, either GnGnXF or GnGn, depending on whether the expression host was WT or the glycosylation mutant ΔXT/FT. Although consistent glycosylation is typical for secreted recombinant proteins in plants [15], the presence of virtually only one oligosaccharide structure, as was the case for $BChE_{\Delta XT/FT}$, is impressive. Such uniformity was not previously demonstrated in any other expression platform and provides a perfect starting point for further elongation steps. Upon co-expression of rBChE with the six mammalian genes responsible for protein sialylation, rBChE was decorated with mono- and di-sialylated *N*-glycans. This carbohydrate profile largely resembles that of plasma-derived BChE [8]. Although tri-sialylated oligosaccharides are present only in minute amounts on plasma-derived BChE, we aimed to extend rBChE *N*-glycans toward tri-sialylated structures. Increased sialic acid content has been reported to prolong half-life of recombinant BChE and other bio-therapeutics [6, 27]. By overexpressing an additional glycosylation enzyme that carries out *N*-glycan branching (i.e. GnTIV), we were able to produce rBChE decorated with surprisingly high amounts of tri-sialylated carbohydrates (70%). Overall, sialylated structures of IF-derived rBChE accounted for up to 90% of all glycans detected in the sialylation experiments aimed at di- and tri-sialylation (Table 1). Such glyco-engineered BChE holds great promise for therapeutic applications. Here we demonstrated by the use of a modified Ellman assay that the three plant-derived glyco-forms ($BChE_{WT}$, $BChE_{\Delta XT/FT}$, and $BChE_{sia}$) are biologically active. The activity profile of $BChE_{sia}$ largely resembles that of the commercially available equine-derived protein.

During our studies we overcame unexpected obstacles, such as the appearance of significant amounts of incompletely processed structures (MNa). On IF-derived $BChE_{WT}$ and $BChE_{\Delta XT/FT}$, such hybrid carbohydrates (MGnXF, MGn) are present in only small amounts, if at all. Thus it appears that the simultaneous transient expression of "foreign" glycosyltransferases, which are required for *in planta* sialylation, impairs the functions of endogenous enzymes. From the structures synthesized (MNa) it seemed obvious that this was the case for GnTII. This enzyme generates important intermediate structures for complete processing to complex sructures, such as GnGnXF and GnGn, which are required as acceptor substrates for galactosylation and subsequent sialylation. Indeed, by transiently over-expressing GnTII we were able to convert the incomplete structures into fully processed sialylated BChE glyco-forms (NaNa, NaNaXF). It seems that apart from supplying the six genes necessary for *in planta* protein sialylation [17], some fine-tuning is needed for the generation of tailor-made structures. Such fine-tuning is not predictable, it largely depends on the *N*-glycan structure required and the protein itself. Basically, it can be accomplished by altering the mixture of glycosylation enzymes in the agro-infiltration procedures. The expression levels of certain foreign glycosylation enzymes may also interfere with the endogeneous glycosylation machinery. This has for example been observed with the expression of human β1,4-galactosyltransferase, where low expression levels are advantageous [28].

An important aspect of this study is that *in planta* sialylation can be achieved using multi-gene vectors [19], instead of single gene vectors. Two binary vectors, each carrying three mammalian genes, permitted the synthesis of sialylated rBChE. This had the advantage that fewer recombinant agrobacteria could be used in the agro-infiltration procedure, enhancing the chance of all genes being expressed and delivered simultaneously, and thus allowing batch-to-batch consistency of glycosylation. Our results point to the feasibility of creating a single multi-gene vector for the entire sialylation pathway. This further facilitates glyco-engineering by transient expression at large scale as needed, for example, for in vivo studies. Such procedures provide a viable alternative to transgenic plants, which have the disadvantage of time-intensive processes and restricted availability.

Another undesired issue we faced was partial proteolytic degradation of the recombinant enzyme, predominantly in the apoplast. Proteolytic attack on recombinant proteins in plants is well known [19, 29, 30]. Secreted proteins are particularly affected, most probably due to the abundant presence of proteases in the extracellular space [31]. Various strategies, such as co-expression of proteinase inhibitors and down-regulation/elimination of endogenous proteinases, have been reported as suitable approaches to circumvent this unwanted effect [32].

The results presented here are encouraging for generating recombinant BChE suitable for therapeutic application with a pharmacokinetic profile that is equivalent, if not superior (through enhanced sialylation content), to that of the human plasma-derived counterpart. It appears entirely feasible that this proof-of-concept study will translate into a pharmaceutically valuable product as the procedure described here allows rapid up- and down-scaling under conditions that accord with Current Good Manufacturing Practices (cGMPs). Animal in vivo studies, which are envisaged in the near future, will be the next important step toward this aim.

We thank Thomas Hackl, Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Vienna, Austria for excellent technical support and Elise Langdon-Neuner for editing the manuscript. This work was supported by a grant from the Austrian Research Promotion Agency (FFG) Laura Bassi Centres of Expertise (Number 822757, to H. S.), the Austrian Science Fund (FWF; Grant Number L575-B13, to H.S.), and the PhD program "BioToP-Biomolecular Technology of Proteins" from the Austrian Science Fund (FWF) Project W1224-B09 (to L. N.). This work was funded in part by a National Institute of Drug Abuse Avant-Garde grant (1 P1 DA031340-01) to the Mayo Clinic, subcontracted to TSM and also by the National Institutes of Health CounterACT Program through the National Institute of Neurological Disorders and Stroke under the U-54-NSO58183-01 award – a consortium grant awarded to USAMRICD and contracted to TSM under the research cooperative agreement number W81XWH-07-2-0023.

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

The authors declare no commercial or financial conflict of interest.

5 References

[1] Doctor, B. P., Saxena, A., Bioscavengers for the protection of humans against organophosphate toxicity. *Chem.-Biol. Interact.* 2005, 167–171.

[2] Mumford, H., Docx, C. J., Price, M. E., Green, A. C. et al., Human plasma-derived BuChE as a stoichiometric bioscavenger for treatment of nerve agent poisoning. *Chem.-Biol. Interact.* 2013, *203*, 160–166.

[3] Mumford, H., Troyer, J. K., Post-exposure therapy with recombinant human BuChE following percutaneous VX challenge in guinea-pigs. *Toxicol. Lett.* 2011, *206*, 29–34.

[4] Geyer, B. C., Kannan, L., Garnaud, P. E., Broomfield, C. A. et al., Plant-derived human butyrylcholinesterase, but not an organophosphorous-compound hydrolyzing variant thereof, protects rodents against nerve agents. *Proc. Natl. Acad. Sci. USA* 2010, *107*, 20251–20256.

[5] Geyer, B. C., Kannan, L., Cherni, I., Woods, R. R. et al., Transgenic plants as a source for the bioscavenging enzyme, human butyrylcholinesterase. *Plant Biotechnol. J.* 2010, *8*, 873–886.

[6] Ilyushin, D. G., Smirnov, I. V., Belogurov, A. A., Jr., Dyachenko, I. A. et al., Chemical polysialylation of human recombinant butyrylcholinesterase delivers a long-acting bioscavenger for nerve agents in vivo. *Proc. Natl. Acad. Sci. USA* 2013, *110*, 1243–1248.

[7] Li, H., Schopfer, L. M., Masson, P., Lockridge, O., Lamellipodin proline rich peptides associated with native plasma butyrylcholinesterase tetramers. *Biochem. J.* 2008, *411*, 425–432.

[8] Kolarich, D., Weber, A., Pabst, M., Stadlmann, J. et al., Glycoproteomic characterization of butyrylcholinesterase from human plasma. *Proteomics* 2008, *8*, 254–263.

[9] Saxena, A., Ashani, Y., Raveh, L., Stevenson, D. et al., Role of oligosaccharides in the pharmacokinetics of tissue-derived and genetically engineered cholinesterases. *Mol. Pharmacol.* 1998, *53*, 112–122.

[10] Kronman, C., Chitlaru, T., Elhanany, E., Velan, B., Shafferman, A., Hierarchy of post-translational modifications involved in the circulatory longevity of glycoproteins. Demonstration of concerted contributions of glycan sialylation and subunit assembly to the pharmacokinetic behavior of bovine acetylcholinesterase. *J. Biol. Chem.* 2000, *275*, 29488–29502.

[11] Huang, Y. J., Huang, Y., Baldassarre, H., Wang, B. et al., Recombinant human butyrylcholinesterase from milk of transgenic animals to protect against organophosphate poisoning. *Proc. Natl. Acad. Sci. USA* 2007, *104*, 13603–13608.

[12] Li, S., Ip, D. T., Lin, H. Q., Liu, J. M. et al., High-level expression of functional recombinant human butyrylcholinesterase in silkworm larvae by Bac-to-Bac system. *Chem.-Biol. Interact.* 2010, *187*, 101–105.

[13] Lenz, D. E., Maxwell, D. M., Koplovitz, I., Clark, C. R., et al., Protection against soman or VX poisoning by human butyrylcholinesterase in guinea pigs and cynomolgus monkeys. *Chem.-Biol. Interact.* 2005, 205–210.

[14] Duysen, E. G., Bartels, C. F., Lockridge, O., Wild-type and A328W mutant human butyrylcholinesterase tetramers expressed in Chinese hamster ovary cells have a 16-hour half-life in the circulation and protect mice from cocaine toxicity. *J. Pharmacol. Exp. Ther.* 2002, *302*, 751–758.

[15] Castilho, A., Steinkellner, H., Glyco-engineering in plants to produce human-like N-glycan structures. *Biotechnol. J.* 2012, *7*, 1088–1098.

[16] Bosch, D., Castilho, A., Loos, A., Schots, A., Steinkellner, H., N-Glycosylation of plant-produced recombinant proteins. *Curr. Pharm. Des.* 2013, *19*, 5503–5512.

[17] Castilho, A., Strasser, R., Stadlmann, J., Grass, J. et al., In planta protein sialylation through overexpression of the respective mammalian pathway. *J. Biol. Chem.* 2010, *285*, 15923–15930.

[18] Castilho, A., Gattinger, P., Grass, J., Jez, J. et al., N-glycosylation engineering of plants for the biosynthesis of glycoproteins with bisected and branched complex N-glycans. *Glycobiology* 2011, *21*, 813–823.

[19] Castilho, A., Neumann, L., Gattinger, P., Strasser, R. et al., Generation of biologically active multi-sialylated recombinant human EPOFc in plants. *PloS One* 2013, *8*, e54836.

[20] Strasser, R., Steinkellner, H., Boren, M., Altmann, F., et al., Molecular cloning of cDNA encoding N-acetylglucosaminyltransferase II from Arabidopsis thaliana. *Glycoconj. J.* 1999, *16*, 787–791.

[21] Strasser, R., Stadlmann, J., Svoboda, B., Altmann, F. et al., Molecular basis of N-acetylglucosaminyltransferase I deficiency in Arabidopsis thaliana plants lacking complex N-glycans. *Biochem. J.* 2005, *387*, 385–391.

[22] Strasser, R., Stadlmann, J., Schahs, M., Stiegler, G. et al., Generation of glyco-engineered Nicotiana benthamiana for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure. *Plant Biotechnol. J.* 2008, *6*, 392–402.
[23] Stadlmann, J., Pabst, M., Kolarich, D., Kunert, R., Altmann, F., Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. *Proteomics* 2008, *8*, 2858–2871.
[24] Pabst, M., Chang, M., Stadlmann, J., Altmann, F., Glycan profiles of the 27 N-glycosylation sites of the HIV envelope protein CN54gp140. *Biol. Chem.* 2012, *393*, 719–730.
[25] Ellman, G. L., Courtney, K. D., Andres, V., Jr., Feather-Stone, R. M., A new and rapid colorimetric determination of acetylcholinesterase activity. *Biochem. Pharmacol.* 1961, *7*, 88–95.
[26] Nagels, B., Van Damme, E. J., Pabst, M., Callewaert, N., Weterings, K., Production of complex multiantennary N-glycans in Nicotiana benthamiana plants. *Plant Physiol.* 2011, *155*, 1103–1112.
[27] Egrie, J. C., Browne, J. K., Development and characterization of novel erythropoiesis stimulating protein (NESP). *Br. J. Cancer* 2001, *84 Suppl 1*, 3–10.
[28] Castilho, A., Bohorova, N., Grass, J., Bohorov, O. et al., Rapid high yield production of different glycoforms of Ebola virus monoclonal antibody. *PloS One* 2011, *6*, e26040.
[29] Lombardi, R., Donini, M., Villani, M. E., Brunetti, P. et al., Production of different glycosylation variants of the tumour-targeting mAb H10 in Nicotiana benthamiana: influence on expression yield and antibody degradation. *Transgenic Res.* 2012, *21*, 1005–1021.
[30] Hehle, V. K., Paul, M. J., Drake, P. M., Ma, J. K., van Dolleweerd, C. J., Antibody degradation in tobacco plants: a predominantly apoplastic process. *BMC Biotechnol.* 2011, *11*, 128.
[31] van der Hoorn, R. A., Plant proteases: From phenotypes to molecular mechanisms. *Annu. Rev. Plant Biol.* 2008, *59*, 191–223.
[32] Schillberg, S., Raven, N., Fischer, R., Twyman, R. M., Schiermeyer, A., Molecular farming of pharmaceutical proteins using plant suspension cell and tissue cultures. *Curr. Pharm. Des.* 2013, *19*, 5531–5542.

This Special Issue on "Biomolecular Technology of Proteins – BioToP" compiles selected peer-reviewed publications of students from the BioToP PhD program at the Vienna Institute of BioTechnology (VIBT) of the University of Natural Resources and Life Sciences, Vienna, Austria (BOKU) and is edited by Co-Editor-in-Chief Prof. Alois Jungbauer. The cover represents the interdisciplinary and international character of BioToP. Idea: Andreas Maccani, Dagmar Brugger, Stefan Hofbauer, Vaibhav Jadhav, Gerald Klanert, Daniel Kracher, Iris Krondorfer, Irene Schaffner. Image: Dagmar Brugger.

Biotechnology Journal – list of articles published in the April 2014 issue.

Regular Articles

Appendix IV

Liebminger, E., Grass, J., Jez, J., **Neumann, L.**, Altmann, F., Strasser, R. (2012) Myrosinases TGG1 and TGG2 from Arabidopsis thaliana contain exclusively oligomannosidic N-glycans. *Phytochemistry* 84:24-30. doi: 10.1016/j.phytochem.2012.08.023

Phytochemistry 84 (2012) 24–30

Contents lists available at SciVerse ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem

Myrosinases TGG1 and TGG2 from *Arabidopsis thaliana* contain exclusively oligomannosidic N-glycans

Eva Liebminger [a], Josephine Grass [b], Jakub Jez [a], Laura Neumann [b], Friedrich Altmann [b], Richard Strasser [a,*]

[a] Department of Applied Genetics and Cell Biology, BOKU-University of Natural Resources and Life Sciences, Muthgasse 18, A-1190 Vienna, Austria
[b] Department of Chemistry, BOKU-University of Natural Resources and Life Sciences, Muthgasse 18, A-1190 Vienna, Austria

ARTICLE INFO

Article history:
Received 15 June 2012
Received in revised form 27 August 2012
Available online 23 September 2012

Keywords:
Arabidopsis thaliana
Post-translational modification
N-Glycosylation
Glycan
Glycoprotein
Mannosidase
Thioglucoside glucohydrolase

ABSTRACT

In all eukaryotes N-glycosylation is the most prevalent protein modification of secretory and membrane proteins. Although the N-glycosylation capacity and the individual steps of the N-glycan processing pathway have been well studied in the model plant *Arabidopsis thaliana*, little attention has been paid to the characterization of the glycosylation status of individual proteins. We report here the structural analysis of all N-glycans present on the endogenous thioglucoside glucohydrolases (myrosinases) TGG1 and TGG2 from *A. thaliana*. All nine glycosylation sites of TGG1 and all four glycosylation sites of TGG2 are occupied by oligomannosidic structures with $Man_5GlcNAc_2$ as the major glycoform. Analysis of the oligomannosidic isomers from wild-type plants and mannose trimming deficient mutants by liquid chromatography with porous graphitic carbon and mass spectrometry revealed that the N-glycans from both myrosinases are processed by Golgi-located α-mannosidases.

1. Introduction

Arabidopsis thaliana is the preferred organism for the investigation of the glycan structure–function relationship in plants. By analysis of mutants with defects in various maturation steps it has become clear that N-glycan processing is essential for plant development (Boisson et al., 2001; Burn et al., 2002; Gillmor et al., 2002; Liebminger et al., 2009) and for abiotic as well as biotic stress reactions (Häweker et al., 2010; Kang et al., 2008; Lu et al., 2009). N-glycosylation mutants displaying a severe developmental defect have also drastically altered cell walls and many proteins involved in cell wall synthesis or remodelling are glycoproteins (Kang et al., 2008; Liebminger et al., 2009; Zhang et al., 2011). Moreover, plasma membrane located receptor kinases essential for plant growth (e.g. BRI1) or pathogen perception (e.g. EFR, FLS2) are heavily glycosylated and contain numerous N-glycosylation sites in their extracellular ligand binding domains (Gómez-Gómez and Boller, 2000; Li and Chory, 1997; Zipfel et al., 2006). In spite of these important findings, the analysis of the glycosylation site occupancy and oligosaccharide structures present on individual *A. thaliana* glycoproteins have been largely ignored in the past. The glycosylation status of endogenous *A. thaliana* proteins has mainly been determined by shifts in mobility on SDS–PAGE upon enzymatic or chemical deglycosylation and consequently the extent and heterogeneity of the attached oligosaccharide structures are unknown. Apart from the glycan analysis of recombinant proteins expressed in different *A. thaliana* organs (e.g. Schähs et al., 2007; Van Droogenbroeck et al., 2007) structural information of N-glycans is derived mainly from studies performed by enzymatic release of the oligosaccharides from total protein extracts (Henquet et al., 2008; Kajiura et al., 2010b; Rayon et al., 1999; Rendić et al., 2007; Strasser et al., 2004). The first comprehensive analysis of N-glycosylation sites present in the *A. thaliana* proteome has been reported recently (Zielinska et al., 2012) and in one study glycopeptides from the putative cell wall proteome have been identified by MALDI-TOF MS (Zhang et al., 2011).

Here, we performed for the first time a detailed characterization of the N-glycan structures present on two endogenous glycoproteins isolated from *A. thaliana* leaves. The oligosaccharide structures present on the myrosinases THIOGLUCOSIDE GLUCOHYDROLASE1 (TGG1) and TGG2 were determined by mass spectrometry of glycopeptides and chromatographic comparison by porous graphitic carbon-liquid chromatography–electrospray ionization mass spectrometry (PGC-LC–ESI-MS). TGG1 and TGG2 belong to the glycoside hydrolase superfamily and produce toxic compounds against pathogens and insects by catalyzing the hydrolysis of glucosinolates (Barth and Jander, 2006; Husebye et al., 2002). The two myrosinases display an organ-specific expression pattern and accumulate mainly in rosette leaves, flowers and siliques, but not in roots (Ueda et al., 2006). In previous studies it has been shown that TGG1 and TGG2 have redundant functions

* Corresponding author. Tel.: +43 1 47654 6705; fax: +43 1 47654 6392.
E-mail address: richard.strasser@boku.ac.at (R. Strasser).

0031-9422/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.phytochem.2012.08.023

in terms of glucosinolate degradation and hormone signalling (Barth and Jander, 2006; Islam et al., 2009). Both enzymes are glycosylated, but the number of glycans, the type of glycans and the specific role of N-glycosylation for their biological function are unknown (Ueda et al., 2006; Zhou et al., 2012).

2. Results and discussion

2.1. TGG1 and TGG2 accumulate abnormally in mns mutants

In a previous study we have identified *mns* mutants with defects in processing of oligomannosidic N-glycans (Liebminger et al., 2009). The *mns* mutants display altered root formation and the triple mutant (*mns1 mns2 mns3*) which completely lacks class I α-mannosidase activity, exhibits reduced leaf growth and rosette diameter. To identify potential candidate glycoproteins whose functions are altered in *mns* mutants compared to wild-type plants, we analyzed total protein extracts from leaves by SDS–PAGE and Coomassie Brilliant Blue staining. Two aberrant bands of ca. 75 and 65 kDa were detectable in the mutants (Fig. 1A). The two protein bands were excised from the gel, tryptic digested and analyzed by ESI-MS. Peptides from the 75 kDa protein band could be assigned to TGG1 and peptides from the 65 kDa band to TGG2 (data not shown). Immunoblot analysis with antibodies specific for TGG1 and TGG2 confirmed that both the mobility and the abundance of the two myrosinases are altered in *mns* mutants compared to control plants (Fig. 1B and C). In particular, the *mns1 mns2 mns3* triple mutant displayed a pronounced shift and an almost 3-fold increase in signal intensity for TGG1 and TGG2. Differences in abundance of TGG1 have also been described for mutants with defects in the assembly of N-glycans (Farid et al., 2011; Koiwa et al., 2003; Zhang et al., 2008, 2009) suggesting that N-glycosylation of myrosinases could affect their stability.

2.2. TGG1 and TGG2 are glycosylated with oligomannosidic N-glycans

To assess the degree of N-glycan processing protein extracts from *A. thaliana* wild-type leaves were subjected to endoglycosidase H (Endo H) and Peptide: N-glycosidase F (PNGase F) digestion followed by immunoblotting with TGG1 and TGG2 specific antibodies. For both myrosinases shifts in mobility were visible upon enzymatic deglycosylation and the Endo H and PNGase F digested bands showed co-migration, which is indicative for the presence of predominantly oligomannosidic N-glycans (Fig. 2A and B). However, plant complex N-glycans frequently contain α1,3-fucose residues linked to the innermost GlcNAc, which make them insensitive to PNGase F digestion (Tretter et al., 1991). Therefore we performed endoglycosidase treatments of TGG1 and TGG2 in the *fut11 fut12* line, which generates complex N-glycans lacking core α1,3-fucose (Strasser et al., 2004). Deglycosylated TGG1 and TGG2 from *fut11 fut12* were indistinguishable from deglycosylated wild-type forms strongly indicating the absence of any α1,3-fucosylated complex N-glycans (Fig. 2C and D).

Next we investigated the glycan site occupancy and glycan structures of TGG1 and TGG2. Protein extracts from *A. thaliana* leaves were separated by SDS–PAGE, bands corresponding to TGG1 and TGG2 were excised from the gel, digested by proteases and analyzed by mass spectrometry. For TGG1 nine glycopeptides were identified showing that all potential N-glycosylation sites are subjected to glycosylation (Figs. 3 and S1). The TGG1 glycopeptides displayed masses matching with different oligomannosidic structures, but lack paucimannosidic or complex N-glycans. $Man_5GlcNAc_2$ was the main oligosaccharide found on the majority of TGG1 glycopeptides. In addition, $Man_6GlcNAc_2$, $Man_7GlcNAc_2$ and $Man_8GlcNAc_2$ were also present on several glycosylation sites while on Asn108 the predominant peak corresponds to $Man_9GlcNAc_2$ indicating that this glycopeptide is not accessible for further

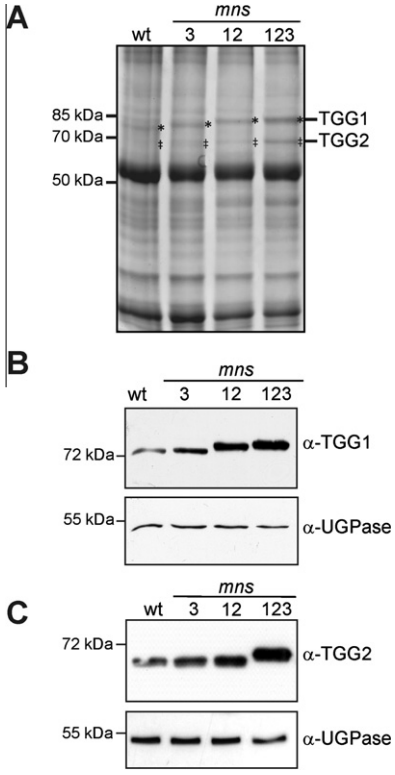


Fig. 1. TGG1 and TGG2 abundance and mobility are altered in *mns* mutants. (A) Coomassie Brilliant Blue staining of protein extracts from wild-type, *mns3* single (3), *mns1 mns2* double (12) and *mns1 mns2 mns3* triple (123) mutants. (B) Immunoblot analysis using anti-TGG1 and (C) anti-TGG2 antibodies. Antibodies against UGPase were used as a control.

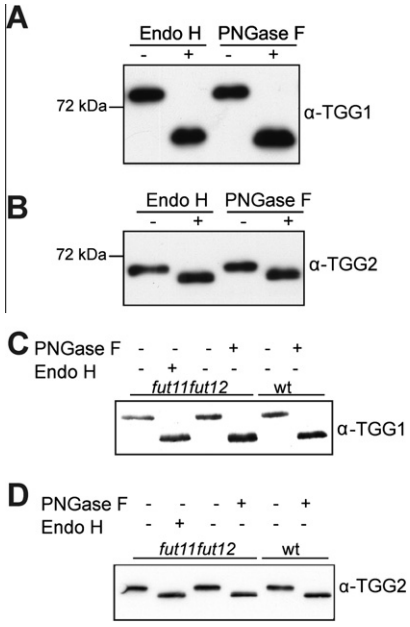


Fig. 2. TGG1 and TGG2 are sensitive to endoglycosidase digestion. Protein extracts of wild-type (A and B) and *fut11 fut12* (C and D) were digested with Endo H and PNGase F, subjected to SDS–PAGE and analyzed by immunoblotting with TGG1 (A and C) and TGG2 (B and D) specific antibodies.

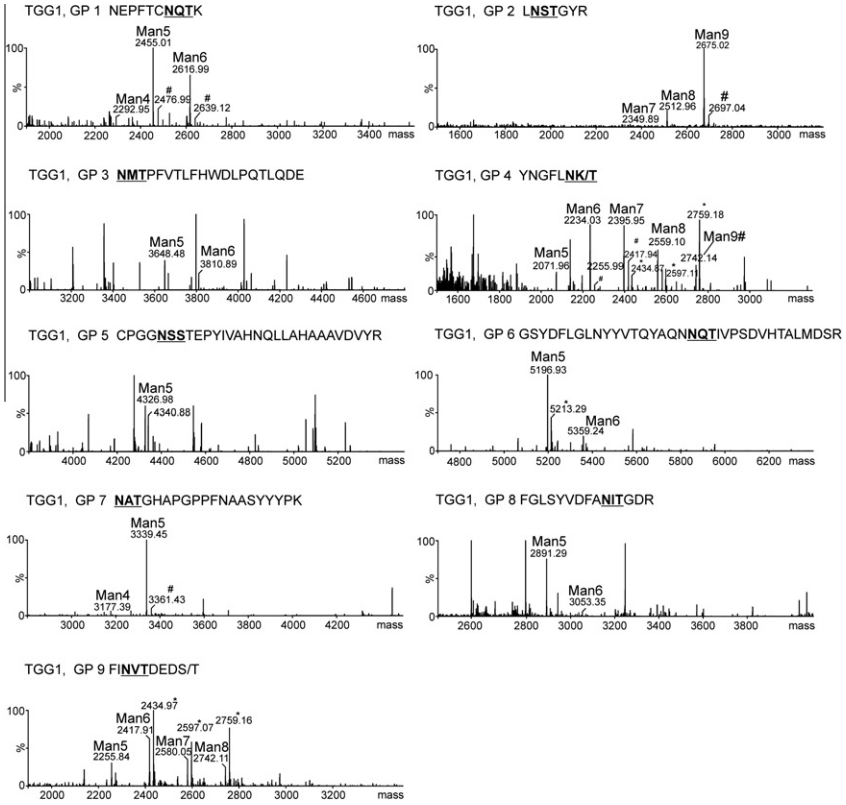


Fig. 3. TGG1 glycopeptides contain exclusively oligomannosidic N-glycans. The spectra of all nine glycopeptides (GP 1-GP 9) derived from endogenous *A. thaliana* TGG1 are shown. Peaks corresponding to N-glycan structures are labelled. Sodium adducts are indicated (#). Spectra were obtained by LC–ESI-MS analysis of trypsin/GluC double digests.

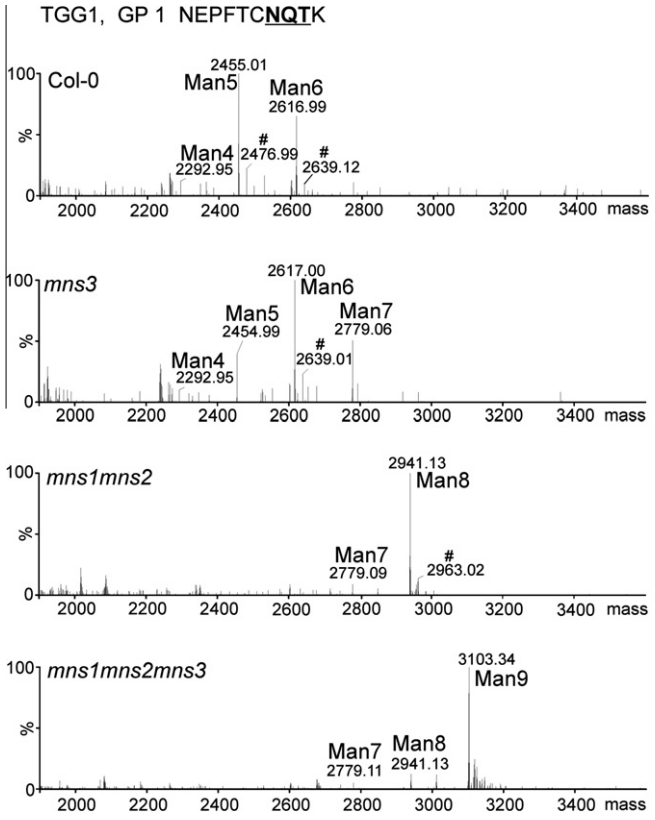


Fig. 4. TGG1 N-glycans are trimmed by class I α-mannosidases. LC–ESI-MS analysis of TGG1 glycopeptide 1 (GP 1) from wild-type, *mns3*, *mns1 mns2* and *mns1 mns2 mns3* mutants. Sodium adducts are indicated (#).

N-glycan trimming. The same N-glycan structures were observed when TGG1 was purified by Concanavalin A Sepharose from protein extracts followed by LC–ESI-MS analysis (data not shown).

For TGG2, four glycopeptides could be identified revealing that TGG2 is fully glycosylated. The assigned glycan profile of TGG2 is very similar to TGG1 showing only oligomannosidic N-glycans with $Man_5GlcNAc_2$ as predominant peak (Fig. S2).

2.3. TGG1 from mns mutants display incompletely processed oligomannosidic N-glycans

The observed shifts in mobility on protein gels suggest that both myrosinases are aberrantly glycosylated in the mannose trimming deficient *mns* mutants. Consequently, we analyzed corresponding glycopeptides of TGG1 from different *mns* mutants to figure out if the myrosinase is indeed processed by MNS proteins. Consistent with the previously described glycosylation defect (Liebminger et al., 2009) TGG1 N-glycans from *mns3* plants displayed increased levels of $Man_6GlcNAc_2$ and $Man_7GlcNAc_2$ structures. The spectra derived from *mns1 mns2* double and *mns1 mns2 mns3* triple knockouts contained almost exclusively $Man_8GlcNAc_2$ and $Man_9GlcNAc_2$ N-glycans, respectively (Figs. 4 and S3). These data show clearly that TGG1 N-glycans are processed by MNS3 and MNS1/MNS2.

Since oligomannosidic N-glycans very often represent a mixture of different isomers we subjected released N-glycans from TGG1 to PGC-LC–ESI-MS analysis, which allows the separation and identification of oligomannosidic isomers (Pabst et al., 2012). PGC-LC analysis confirmed the presence of $Man_5GlcNAc_2$ to $Man_9GlcNAc_2$ oligosaccharide structures. The $Man_5GlcNAc_2$ isomer (M^6M^3)M was the most abundant glycan on TGG1 extracted from wild-type plants (Fig. 5). Accordingly, we could assign the predominant $Man_6GlcNAc_2$ structure to (M^6M^{2-3})M in *mns3* as well as the

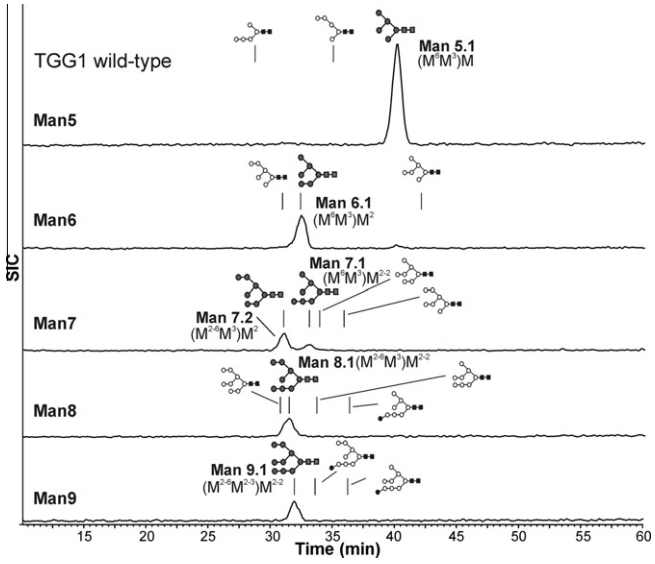


Fig. 5. Determination of the predominant N-glycan isomers on TGG1. Selected ion chromatograms (SIC) of oligomannosidic N-glycans from TGG1 are shown. TGG1 was extracted from leaves of wild-type plants and reduced N-glycans were separated by PGC-LC–ESI-MS. The assignment of structural isomers was based on reference glycans as described in detail previously (Pabst et al., 2012).

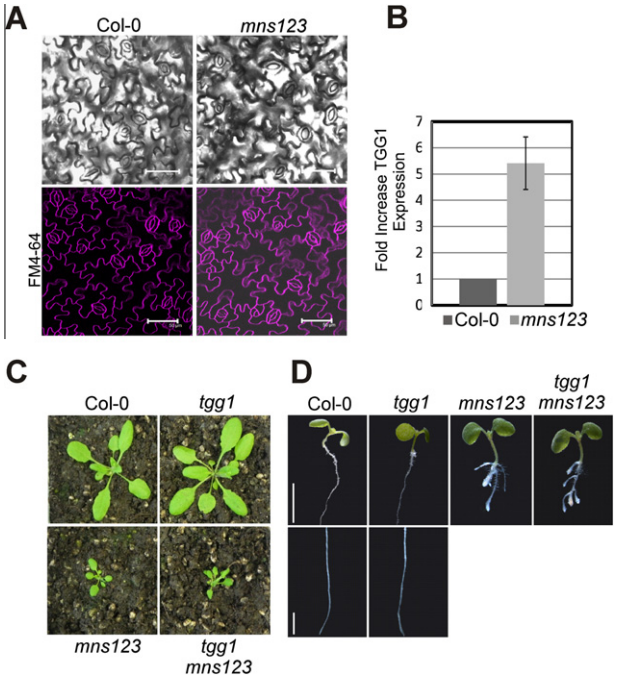


Fig. 6. Phenotypic analysis of wild-type and mutant plants. (A) Images of the epidermis of mature rosette leaves. Lower panel: staining of stomata from Col-0 and *mns1 mns2 mns3* (*mns123*) with FM4-64 dye. Bar = 50 μm. (B) Quantitative real time PCR (qPCR) analysis of TGG1 transcript levels in wild-type and mutant plants. (C) Phenotypes of 3-week-old wild type and mutant plants grown on soil under long day conditions. (D) 10-day-old seedlings grown on 1× Murashige and Skoog medium supplemented with 2% sucrose. Bar = 5 mm.

$Man_8GlcNAc_2$ to $(M^{2-6}M^3)M^{2-2}$ in *mns1 mns2* and the $Man_9GlcNAc_2$ to $(M^{2-6}M^{2-3})M^{2-2}$ in *mns1 mns2 mns3* (Table 1). These findings are in good agreement with current models of N-glycan processing in the early secretory pathway of *A. thaliana* (Liebminger et al., 2009).

2.4. TGG1 transcript levels are higher in leaves of *mns1 mns2 mns3* plants

The most prominent difference in expression levels were found for TGG1 in the *mns1 mns2 mns3* triple mutant (Fig. 1). It has been proposed that TGG1 is cell type-specific expressed in specialized myrosin cells and is also highly abundant in guard cells of leaves (Husebye et al., 2002; Ueda et al., 2006; Zhao et al., 2008). To monitor if the elevated TGG1 levels are caused by an increased number of guard cells we compared the stomata in leaves from *mns1 mns2 mns3* to wild-type plants (Fig. 6A). Neither the number nor the morphology of stomata was affected in the mutant.

Next we addressed whether TGG1 transcript expression is altered in the triple mutant. Quantitative PCR showed that TGG1 transcript levels are at least 5-fold higher in the *mns1 mns2 mns3* mutant (Fig. 6B) suggesting that under conditions of aberrant glycosylation increased amounts of TGG1 are beneficial for the plant.

To investigate the effect of TGG1 deficiency in the *mns1 mns2 mns3* line, we generated *tgg1 mns1 mns2 mns3* quadruple knockouts and compared its growth and development to *mns1 mns2 mns3*. Under our growth conditions seedlings as well as the areal parts of soil grown quadruple knockouts were indistinguishable from *mns1 mns2 mns3* (Fig. 6C and D) being consistent with the current model that TGG1 is involved in glucosinolate breakdown to deter herbivores and pathogens but plays no role for the normal development of *A. thaliana* (Barth and Jander, 2006; Husebye et al., 2002; Ueda et al., 2006).

2.5. TGG1 and TGG2 N-glycans are not altered in the mvp1 mutant

Recently, a protein trafficking mutant was identified, which displayed mislocalization of TGG1 and TGG2 to aggregates within the cell (Agee et al., 2010). Apart from mistargeting of vacuolar proteins the *modified vacuole phenotype1* (*mvp1-1* and *mvp1-2*)

Table 1
N-Glycan isomers of TGG1.

N-Glycan structure		Proglycan code	Relative abundance on individual glycosylation sites (%)			
			Wild-type	*mns3*	*mns1 mns2*	*mns1 mns2 mns3*
Man5	Man5.1	$(M^6M^3)M$	56.8	15.8	–	–
Man6	Man6.1	$(M^6M^3)M^2$	15.3	–	–	–
	Man6.10	$(M^6M^{2-3})M$	–	51.5	–	–
Man7	Man7.1	$(M^6M^3)M^{2-2}$	7.1	–	9.5	–
	Man7.2	$(M^{2-6}M^3)M^2$	3.8	–	20.8	–
	Man7.7	$(M^6M^{2-3})M^2$	–	10.5	–	–
Man8	Man8.1/8.4	$(M^{2-6}M^3)M^{2-2}/(M^{2-6}M^{2-3})M^2$	8.7	9.7	57.6	22.8
	Man8.2	$(M^6M^{2-3})M^{2-2}$	–	2.0	–	3.1
Man9	Man9.1	$(M^{2-6}M^{2-3})M^{2-2}$	8.3	10.5	12.1	74.1

The Proglycan code for oligomannosidic glycans has been described in detail recently (Pabst et al., 2012).

mutants displayed also aberrant localization of Golgi-resident proteins as observed for a fluorescent protein tagged N-acetylglucosaminyltransferase I (GnTI or NAG1), which was used as a Golgi marker (Agee et al., 2010). Since GnTI is the key enzyme for the initiation of complex N-glycan formation in plants (Strasser et al., 1999; von Schaewen et al., 1993) its aggregate formation in the *mvp1* mutants might cause aberrant glycosylation of proteins and could also affect the two myrosinases. To address whether *mvp1* plants display overall changes in N-glycan processing we first analyzed the total N-glycosylation profile of endogenous proteins. Despite the reported global defects in subcellular localization of proteins, the N-glycan profile from *mvp1-2* was essentially the same when compared with wild-type derived N-glycans (*Fig.* S4). In addition, aggregate formation of TGG1 and TGG2 in *mvp1-2* did not have an effect on their N-glycosylation profile. As shown in *Fig.* S5 the processed oligomannosidic N-glycans on TGG1 and TGG2 glycopeptides resemble the oligosaccharides found on in wild-type derived myrosinases showing that neither the aberrant localization of the TGG1 and TGG2 nor global defects in protein trafficking result in changes of N-glycosylation in MVP1 deficient plants.

3. Conclusions

The fundamental role of N-glycosylation for development and response to adverse environmental conditions is very well documented in the plant model organism *A. thaliana*. Remarkably, very little is known about the N-glycosylation of individual *A. thaliana* proteins and the specific function of their N-glycans. Here, we analyzed the glycans of two myrosinases expressed in leaves of wild-type plants and in four different mutant backgrounds. To date, this is the first comprehensive N-glycan analysis of endogenous *A. thaliana* glycoproteins with multiple N-glycosylation sites. Our structural analysis reveal that both myrosinases contain exclusively oligomannosidic N-glycans with the $(M^6M^3)M\ Man_5GlcNAc_2$ isomer as predominant structure. The data from the *mns* mutants confirm that this oligomannosidic N-glycan is processed by the α-mannosidases MNS1 to MNS3. Surprisingly, there is no further processing of TGG1 and TGG2 N-glycans by GnTI, which utilizes $(M^6M^3)M$ as substrate to initiate hybrid and complex N-glycan formation in plants (Strasser et al., 1999; von Schaewen et al., 1993). MNS1 and MNS2 are like GnTI *cis*/medial Golgi-located enzymes (Liebminger et al., 2009; Kajiura et al., 2010a; Nebenführ et al., 1999; Saint-Jore-Dupas et al., 2006; Schoberer et al., 2009; Staehelin and Kang, 2008) which raises the question how this myrosinase-specific glycosylation pattern is generated. We can provide three possible explanations: (i) the N-glycans from TGG1 and TGG2 are not accessible for GnTI and are therefore not further processed. There are no clear data from secreted plant glycoproteins available that support this explanation, but in mammals it has been shown that the occurrence of oligomannosidic N-glycans on the envelope glycoprotein from HIV-1 is at least in part caused by steric constraints that prevent further processing in the Golgi (Doores et al., 2010). (ii) TGG1 and TGG2 are only expressed in specialized cells that lack GnTI activity. While this possibility cannot be ruled out it is commonly believed that most of the Golgi-located N-glycan processing enzymes are expressed ubiquitously in plants and with the exception of the Lewis a structures (Strasser et al., 2007) no cell- or organ-specific glycosylation profile has been described for *A. thaliana* so far. (iii) In TGG1 and TGG2 expressing cells MNS proteins and GnTI are located in different sub-Golgi compartments. If the final destination for both myrosinases is the vacuole (Ueda et al., 2006), this latter hypothesis would suggest a specific sorting from the *cis*-Golgi (or MNS compartment) to the vacuole without trafficking through other Golgi cisternae. Current models for subcellular localization of glycosylation enzymes suggest the existence of some kind of assembly line across the Golgi stack (Saint-Jore-Dupas et al., 2006; Schoberer and Strasser, 2011). However, additional studies using advanced imaging and protein–protein interaction technologies (Sparkes et al., 2011) are required to reveal the organization of Golgi-located glycosylation enzymes in plants.

The biological significance of TGG1 and TGG2 N-glycosylation remains elusive and the mechanisms as well as the consequences of the increased myrosinase expression in the *mns1 mns2 mns3* mutant are unclear. Further studies are necessary to unravel a biological role for the processed oligomannosidic N-glycan structures on TGG1 and TGG2 and provide a link to myrosinase function during pathogen attack. In summary, our in depth structural analysis offers interesting new insights into N-glycan processing of individual plant proteins that are missed by mass spectrometry of N-glycans from large pools of glycoproteins.

4. Experimental

4.1. Plant material and growth conditions

A. thaliana wild-type (ecotype Columbia-0) and mutant plants were grown as described previously (Liebminger et al., 2009). Seeds of *tgg1-1* (Barth and Jander, 2006) and *mvp1-2* (Agee et al., 2010) were obtained from the Nottingham Arabidopsis Stock Centre (NASC). Line *tgg1-1* was crossed with the *mns1 mns2 mns3* triple knockout line (Liebminger et al., 2009) to generate the *tgg1 mns1 mns2 mns3* quadruple mutant. Homozygous lines were confirmed by genotyping using the following primer combinations: At1g51590-1F/2R for *mns1*, At3g21160-1F/2R for *mns2*, At1g30000-3F/2R for *mns3*; At5g26000-1F/2R for *tgg1* and MVP-1F/2R for *mvp1-2*. All primer sequences are presented in Table S1.

4.2. SDS–PAGE and LC–ESI-MS analysis of glycopeptides

Rosette leaf material from *A. thaliana* wild-type and *mns* single, double and triple knockouts was ground in liquid nitrogen and resuspended in 1× phosphate buffered saline (PBS). Samples were incubated for 10 min on ice and then centrifuged for 10 min at 9.600*g*. The supernatant was centrifuged for 5 min at 9.600*g* and mixed with SDS–PAGE loading buffer. The samples were incubated for 5 min at 95 °C, separated by SDS–PAGE (10%) under reducing conditions and polypeptides were detected by Coomassie Brilliant Blue staining. The corresponding bands were excised from the gel, destained, carbamidomethylated, in-gel digested with trypsin and GluC and analyzed by LC–ESI-MS as described previously (Stadlmann et al., 2008; Kolarich et al., 2012).

4.3. PGC-LC–ESI-MS analysis

For isomeric analysis, the above described glycopeptides were digested with peptide: N-glycosidase A (Proglycan, Vienna, Austria). The liberated glycans were reduced with sodium borohydride and subsequently purified using graphitic carbon cartridges (Pabst et al., 2012).

Reduced oligosaccharides were analyzed by PGC-LC–ESI-MS on a Hypercarb column (0.32 × 100 mm, Thermo Fisher Scientific) coupled to an Ultimate 3000 capillary HPLC (Thermo Fisher Scientific) and a Q-TOF Ultima MS (Waters) (Pabst et al., 2007) Briefly, the aqueous solvent was 0.3% formic acid buffered to pH 3.0 and a linear gradient from 8.8% to 17.2% acetonitrile was developed during 55 min. Detection was accomplished by positive mode ESI-MS. Assignment of isomers was based on reference oligosaccharides prepared from kidney beans and bovine ribonuclease B

as described in detail previously (Pabst et al., 2012). For retention time normalization, the diantennary, complex-type asialo-N-glycan A^4A^4 was added to each sample.

4.4. Deglycosylation with PNGase F and Endo H

Protein extraction from wild-type and *fuct11 fuct12* line (Strasser et al., 2004) was performed as described above. Prior to PNGase F/Endo H treatment, the protein extracts were mixed with 10x denaturation buffer (New England Biolabs, Frankfurt, Germany), boiled for 10 min at 98 °C and then cooled for 2 min on ice. For PNGase F digestion, the extract (10 μL) was mixed with 10× G7 reaction buffer containing 10% NP-40 (New England Biolabs), water and 0.5 U PNGase F (New England Biolabs) in a total volume of 15 μL. For Endo H digestion, the extract was mixed with 10× G5 reaction buffer (New England Biolabs), water and 0.5 U Endo H (New England Biolabs) in a total volume of 15 μL. Controls were treated as described above but the respective enzymes were replaced by water. After incubation of the reaction mixture for 3 h at 37 °C, SDS–PAGE loading buffer was added and the samples were heated to 95 °C for 5 min. Deglycosylated proteins and controls were subjected to SDS–PAGE (10%) followed by immunoblotting with anti-TGG1 and anti-TGG2 antibodies (kind gift of Ikuko Hara-Nishimura) (Ueda et al., 2006). A rabbit polyclonal antiserum against UDP-glucose pyrophosphorylase (anti-UGPase, Agrisera, Vännäs, Sweden) was used as a control for equal loading.

4.5. FM4-staining and imaging

A. thaliana seedlings (6 to 8-day-old) were incubated in 50 μM FM4-64 dye (Invitrogen, F34653) for an appropriate time. Seedlings were then mounted on a slide and imaging of stomata was performed using a Leica TCS SP2 confocal laser scanning microscope as described previously (Schoberer et al., 2009).

4.6. RNA Isolation, reverse transcription and qPCR

Total RNA from rosette leaves was isolated using a SV Total RNA Isolation kit (Promega). First strand cDNA was synthesized from 1.5 μg of total RNA at 42 °C using oligo dT primers and AMV reverse transcriptase (Promega) in a total volume of 20 μl. PCR reactions were performed in a Rotor-Gene RG-3000A (Corbett, Qiagen, Hilden, Germany). Reactions contained 4 μL 5× HOT FIREPol® EvaGreen® qPCR Mix Plus (no ROX) (Solis BioDyne, Tartu, Estonia), 0.1 μL cDNA and 300 nM of each gene-specific primer in a total volume of 20 μL. The following profile was used for all PCR reactions: 95 °C for 15 min, 40 cycles of 95 °C for 15 s, 60 °C for 20 s, 72 °C for 20 s. To detect TGG1 transcripts cDNA was amplified with primers At5g26000-11F/12R and normalized to the expression of the protein phosphatase 2A (PP2A) gene (primers At1g13320-1F/2R). The specificity of the PCR amplification was checked with a melting curve analysis from 65 to 99 °C and data were analyzed using Rotor-Gene software (version 6). PCR reactions were done in triplets and at least three independent biological experiments were performed.

Acknowledgments

We thank Ikuko Hara-Nishimura (Department of Botany, Graduate School of Science, Kyoto University, Kyoto, Japan) for the kind gift of anti-TGG1 and anti-TGG2 antibodies. We also thank Cornelia Konlechner (Department of Applied Genetics and Cell Biology) for help with qPCR experiments and Karin Polacsek (Department of Chemistry) for N-glycan analysis. This work was funded by the Austria Science Fund (FWF): Project P20817-B12 and by the Ph.D. programme "BioToP – Biomolecular Technology of Proteins" from the Austrian Science Funds (FWF): Project W1224-B09.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.phytochem.2012.08.023.

References

Agee, A., Surpin, M., Sohn, E., Girke, T., Rosado, A., Kram, B., Carter, C., Wentzell, A., Kliebenstein, D., Jin, H., Park, O., Jin, H., Hicks, G., Raikhel, N., 2010. MODIFIED VACUOLE PHENOTYPE1 is an Arabidopsis myrosinase-associated protein involved in endomembrane protein trafficking. Plant Physiol. 152, 120–132.

Barth, C., Jander, G., 2006. Arabidopsis myrosinases TGG1 and TGG2 have redundant function in glucosinolate breakdown and insect defense. Plant J. 46, 549–562.

Boisson, M., Gomord, V., Audran, C., Berger, N., Dubreucq, B., Granier, F., Lerouge, P., Faye, L., Caboche, M., Lepiniec, L., 2001. Arabidopsis glucosidase I mutants reveal a critical role of N-glycan trimming in seed development. EMBO J. 20, 1010–1019.

Burn, J., Hurley, U., Birch, R., Arioli, T., Cork, A., Williamson, R., 2002. The cellulose-deficient Arabidopsis mutant *rsw3* is defective in a gene encoding a putative glucosidase II, an enzyme processing N-glycans during ER quality control. Plant J. 32, 949–960.

Doores, K.J., Bonomelli, C., Harvey, D.J., Vasiljevic, S., Dwek, R.A., Burton, D.R., Crispin, M., Scanlan, C.N., 2010. Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. Proc Natl Acad Sci U S A 107, 13800–13805.

Farid, A., Pabst, M., Schoberer, J., Altmann, F., Glössl, J., Strasser, R., 2011. *Arabidopsis thaliana* alpha1,2-glucosyltransferase (ALG10) is required for efficient N-glycosylation and leaf growth. Plant J. 68, 314–325.

Gillmor, C., Poindexter, P., Lorieau, J., Palcic, M., Somerville, C., 2002. Alpha-glucosidase I is required for cellulose biosynthesis and morphogenesis in Arabidopsis. J. Cell Biol. 156, 1003–1013.

Gómez-Gómez, L., Boller, T., 2000. FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Mol. Cell 5, 1003–1011.

Henquet, M., Lehle, L., Schreuder, M., Rouwendal, G., Molthoff, J., Helsper, J., van der Krol, S., Bosch, D., 2008. Identification of the gene encoding the alpha1,3-mannosyltransferase (ALG3) in Arabidopsis and characterization of downstream N-glycan processing. Plant Cell 20, 1652–1664.

Husebye, H., Chadchawan, S., Winge, P., Thangstad, O.P., Bones, A.M., 2002. Guard cell- and phloem idioblast-specific expression of thioglucoside glucohydrolase 1 (myrosinase) in Arabidopsis. Plant Physiol. 128, 1180–1188.

Häweker, H., Rips, S., Koiwa, H., Salomon, S., Saijo, Y., Chinchilla, D., Robatzek, S., von Schaewen, A., 2010. Pattern recognition receptors require N-glycosylation to mediate plant immunity. J. Biol. Chem. 285, 4629–4636.

Islam, M., Tani, C., Watanabe-Sugimoto, M., Uraji, M., Jahan, M., Masuda, C., Nakamura, Y., Mori, I., Murata, Y., 2009. Myrosinases, TGG1 and TGG2, redundantly function in ABA and MeJA signaling in Arabidopsis guard cells. Plant Cell Physiol. 50, 1171–1175.

Kajiura, H., Koiwa, H., Nakazawa, Y., Okazawa, A., Kobayashi, A., Seki, T., Fujiyama, K., 2010a. Two *Arabidopsis thaliana* Golgi alpha-mannosidase I enzymes are responsible for plant N-glycan maturation. Glycobiology 20, 235–247.

Kajiura, H., Seki, T., Fujiyama, K., 2010b. *Arabidopsis thaliana* ALG3 mutant synthesizes immature oligosaccharides in the ER and accumulates unique N-glycans. Glycobiology 20, 736–751.

Kang, J., Frank, J., Kang, C., Kajiura, H., Vikram, M., Ueda, A., Kim, S., Bahk, J., Triplett, B., Fujiyama, K., Lee, S., von Schaewen, A., Koiwa, H., 2008. Salt tolerance of *Arabidopsis thaliana* requires maturation of N-glycosylated proteins in the Golgi apparatus. Proc. Natl. Acad. Sci. USA 105, 5933–5938.

Koiwa, H., Li, F., McCully, M., Mendoza, I., Koizumi, N., Manabe, Y., Nakagawa, Y., Zhu, J., Rus, A., Pardo, J., Bressan, R., Hasegawa, P., 2003. The STT3a subunit isoform of the Arabidopsis oligosaccharyltransferase controls adaptive responses to salt/osmotic stress. Plant Cell 15, 2273–2284.

Kolarich, D., Jensen, P.H., Altmann, F., Packer, N.H., 2012. Determination of site-specific glycan heterogeneity on glycoproteins. Nat. Protoc. 7, 1285–1298.

Li, J., Chory, J., 1997. A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. Cell 90, 929–938.

Liebminger, E., Hüttner, S., Vavra, U., Fischl, R., Schoberer, J., Grass, J., Blaukopf, C., Seifert, G., Altmann, F., Mach, L., Strasser, R., 2009. Class I alpha-mannosidases are required for N-glycan processing and root development in *Arabidopsis thaliana*. Plant Cell 21, 3850–3867.

Lu, X., Tintor, N., Mentzel, T., Kombrink, E., Boller, T., Robatzek, S., Schulze-Lefert, P., Saijo, Y., 2009. Uncoupling of sustained MAMP receptor signaling from early outputs in an Arabidopsis endoplasmic reticulum glucosidase II allele. Proc. Natl. Acad. Sci. USA 106, 22522–22527.

Nebenführ, A., Gallagher, L., Dunahay, T., Frohlick, J., Mazurkiewicz, A., Meehl, J., Staehelin, L., 1999. Stop-and-go movements of plant Golgi stacks are mediated by the acto-myosin system. Plant Physiol. 121, 1127–1142.

Pabst, M., Bondili, J.S., Stadlmann, J., Mach, L., Altmann, F., 2007. Mass + retention time = structure: a strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans. Anal. Chem. 79, 5051–5057.

Pabst, M., Grass, J., Toegel, S., Liebminger, E., Strasser, R., Altmann, F., 2012. Isomeric analysis of oligomannosidic N-glycans and their dolichol-linked precursors. Glycobiology 22, 389–399.

Rayon, C., Cabanes-Macheteau, M., Loutelier-Bourhis, C., Salliot-Maire, I., Lemoine, J., Reiter, W., Lerouge, P., Faye, L., 1999. Characterization of N-glycans from Arabidopsis. Application to a fucose-deficient mutant. Plant Physiol. 119, 725–734.

Rendić, D., Wilson, I.B., Lubec, G., Gutternigg, M., Altmann, F., Léonard, R., 2007. Adaptation of the "in-gel release method" to N-glycome analysis of low-milligram amounts of material. Electrophoresis 28, 4484–4492.

Saint-Jore-Dupas, C., Nebenführ, A., Boulaflous, A., Follet-Gueye, M., Plasson, C., Hawes, C., Driouich, A., Faye, L., Gomord, V., 2006. Plant N-glycan processing enzymes employ different targeting mechanisms for their spatial arrangement along the secretory pathway. Plant Cell 18, 3182–3200.

Schoberer, J., Strasser, R., 2011. Sub-compartmental organization of Golgi-resident N-glycan processing enzymes in plants. Mol. Plant 4, 220–228.

Schoberer, J., Vavra, U., Stadlmann, J., Hawes, C., Mach, L., Steinkellner, H., Strasser, R., 2009. Arginine/lysine residues in the cytoplasmic tail promote ER export of plant glycosylation enzymes. Traffic 10, 101–115.

Schähs, M., Strasser, R., Stadlmann, J., Kunert, R., Rademacher, T., Steinkellner, H., 2007. Production of a monoclonal antibody in plants with a humanized N-glycosylation pattern. Plant Biotechnol. J. 5, 657–663.

Sparkes, I.A., Graumann, K., Martinière, A., Schoberer, J., Wang, P., Osterrieder, A., 2011. Bleach it, switch it, bounce it, pull it: using lasers to reveal plant cell dynamics. J. Exp. Bot. 62, 1–7.

Stadlmann, J., Pabst, M., Kolarich, D., Kunert, R., Altmann, F., 2008. Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. Proteomics 8, 2858–2871.

Staehelin, L., Kang, B., 2008. Nanoscale architecture of endoplasmic reticulum export sites and of Golgi membranes as determined by electron tomography. Plant Physiol. 147, 1454–1468.

Strasser, R., Altmann, F., Mach, L., Glössl, J., Steinkellner, H., 2004. Generation of *Arabidopsis thaliana* plants with complex N-glycans lacking beta1,2-linked xylose and core alpha1,3-linked fucose. FEBS Lett. 561, 132–136.

Strasser, R., Bondili, J., Vavra, U., Schoberer, J., Svoboda, B., Glössl, J., Léonard, R., Stadlmann, J., Altmann, F., Steinkellner, H., Mach, L., 2007. A unique beta1,3-galactosyltransferase is indispensable for the biosynthesis of N-glycans containing Lewis a structures in *Arabidopsis thaliana*. Plant Cell 19, 2278–2292.

Strasser, R., Mucha, J., Schwihla, H., Altmann, F., Glössl, J., Steinkellner, H., 1999. Molecular cloning and characterization of cDNA coding for beta1, 2-N-acetylglucosaminyltransferase I (GlcNAc-TI) from *Nicotiana tabacum*. Glycobiology 9, 779–785.

Tretter, V., Altmann, F., März, L., 1991. Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha 1–3 to the asparagine-linked N-acetylglucosamine residue. Eur. J. Biochem. 199, 647–652.

Ueda, H., Nishiyama, C., Shimada, T., Koumoto, Y., Hayashi, Y., Kondo, M., Takahashi, T., Ohtomo, I., Nishimura, M., Hara-Nishimura, I., 2006. AtVAM3 is required for normal specification of idioblasts, myrosin cells. Plant Cell Physiol. 47, 164–175.

Van Droogenbroeck, B., Cao, J., Stadlmann, J., Altmann, F., Colanesi, S., Hillmer, S., Robinson, D.G., Van Lerberge, E., Terryn, N., Van Montagu, M., Liang, M., Depicker, A., De Jaeger, G., 2007. Aberrant localization and underglycosylation of highly accumulating single-chain Fv-Fc antibodies in transgenic Arabidopsis seeds. Proc. Natl. Acad. Sci. USA 104, 1430–1435.

von Schaewen, A., Sturm, A., O'Neill, J., Chrispeels, M., 1993. Isolation of a mutant Arabidopsis plant that lacks N-acetyl glucosaminyl transferase I and is unable to synthesize Golgi-modified complex N-linked glycans. Plant Physiol. 102, 1109–1118.

Zhang, H., Ohyama, K., Boudet, J., Chen, Z., Yang, J., Zhang, M., Muranaka, T., Maurel, C., Zhu, J., Gong, Z., 2008. Dolichol biosynthesis and its effects on the unfolded protein response and abiotic stress resistance in Arabidopsis. Plant Cell 20, 1879–1898.

Zhang, M., Henquet, M., Chen, Z., Zhang, H., Zhang, Y., Ren, X., van der Krol, S., Gonneau, M., Bosch, D., Gong, Z., 2009. LEW3, encoding a putative alpha-1,2-mannosyltransferase (ALG11) in N-linked glycoprotein, plays vital roles in cell-wall biosynthesis and the abiotic stress response in *Arabidopsis thaliana*. Plant J. 60, 983–999.

Zhang, Y., Giboulot, A., Zivy, M., Valot, B., Jamet, E., Albenne, C., 2011. Combining various strategies to increase the coverage of the plant cell wall glycoproteome. Phytochemistry 72, 1109–1123.

Zhao, Z., Zhang, W., Stanley, B.A., Assmann, S.M., 2008. Functional proteomics of *Arabidopsis thaliana* guard cells uncovers new stomatal signaling pathways. Plant Cell 20, 3210–3226.

Zhou, C., Tokuhisa, J.G., Bevan, D.R., Esen, A., 2012. Properties of β-thioglucoside hydrolases (TGG1 and TGG2) from leaves of *Arabidopsis thaliana*. Plant Sci. 191–192, 82–92.

Zielinska, D.F., Gnad, F., Schropp, K., Wiśniewski, J.R., Mann, M., 2012. Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. Mol. Cell 46, 542–548.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J., Boller, T., Felix, G., 2006. Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. Cell 125, 749–760.

Appendix V

Tomek, M., **Neumann, L.,** Nimeth, I., Koerdt, A, Andesner, P., Mach, L., Messner, P**.,** Potempa, J., Schaeffer, C. (2014) S-layer glycoproteins of Tannerella forsythia are secreted via a type IX secretion system that is decoupled from protein glycosylation. Mol. Oral Microbiol., doi: 10.1111/omi.12062

Received Date: 29-May-2014
Revised Date: 05-Jun-2014
Accepted Date: 07-Jun-2014
Article type: Original Article

The S-layer proteins of *Tannerella forsythia* are secreted via a type IX secretion system that is decoupled from protein *O*-glycosylation

Markus B. Tomek[1], Laura Neumann[2], Irene Nimeth[1], Andrea Koerdt[1], Philipp Andesner[1], Paul Messner[1], Lukas Mach[3], Jan S. Potempa[4,5] and Christina Schäffer[1]*

1 Department of NanoBiotechnology, *NanoGlycobiology* unit, Universität für Bodenkultur Wien, Muthgasse 11, 1190 Vienna, Austria
2 Department of Chemistry and 3 Department of Applied Genetics and Cell Biology, Universität für Bodenkultur Wien, Muthgasse 18, 1190 Vienna, Austria
4 Oral Health and Systemic Disease Group, University of Louisville, Louisville, KY 40202, USA
5 Department of Microbiology, Jagiellonian University, ul. Gronostajowa 7, 30-387 Krakow, Poland

Running head: T9SS in *Tannerella forsythia*

Keywords: C-terminal domain, protein secretion, C-terminal signal peptidase, T9SS, S-layer glycosylation

***Correspondence:** *Dr. Christina Schäffer, Department of NanoBiotechnology,* NanoGlycobiology *unit, Universität für Bodenkultur Wien, Muthgasse 11, 1190 Vienna, Austria*
Tel.: (+43)-1-47654-2203; Fax.: (+43)-1-4789112; E-Mail: christina.schaeffer@boku.ac.at

SUMMARY
Conserved C-terminal domains (CTD) have been shown to act as a signal for the translocation of certain proteins across the outer membrane of *Bacteroidetes* via a type IX secretion system (T9SS). The genome sequence of the periodontal pathogen *Tannerella forsythia* predicts the presence of the components for a T9SS in conjunction with a suite of CTD proteins. *T. forsythia* is covered with a 2-dimensional crystalline surface (S-) layer composed of the glycosylated CTD proteins TfsA and TfsB. To investigate if T9SS is functional in *T. forsythia*, T9SS-deficient mutants were generated by targeting either TF*0955* (putative C-terminal signal peptidase) or TF*2327* (PorK ortholog), and the mutants were analyzed with respect to secretion, assembly and glycosylation of the S-layer proteins as well as to proteolytic processing of the CTD and biofilm formation. In either mutant, TfsA and TfsB

were incapable of translocation, as evidenced by the absence of the S-layer in transmission electron microscopy of ultrathin-sectioned bacterial cells. Despite entrapped within the periplasm, mass spectrometry analysis revealed that the S-layer proteins were modified with the complete, mature glycan found on the secreted proteins, indicating that protein translocation and glycosylation are two independent processes. Further, the T9SS mutants showed a denser biofilm with less voids compared to the wild-type. This study demonstrates the functionality of T9SS and the requirement of CTD for the outer membrane passage of extracellular proteins in *T. forsythia*, exemplified with the two S-layer proteins. In addition, T9SS protein translocation is decoupled from *O*-glycan attachment in *T. forsythia*.

INTRODUCTION

Protein translocation across the cell envelope of Gram-negative bacteria has been intensively studied for decades and up to now, eight types of protein secretion systems (T1SS-T8SS) have been described (Desvaux *et al.*, 2009). In different species of *Proteobacteria*, these systems are employed to secrete proteins into the extracellular environment (T1SS, T2SS, T5SS), inject them into the cytoplasm of eukaryotic cells (T3SS, T4SS) or bacteria (T6SS), and build up cell surface appendages (T7SS, T8SS) (Desvaux *et al.*, 2009). Perplexingly, very little is known of how proteins are secreted by bacteria belonging to other phyla of Gram-negative prokaryotes (Pugsley, 1993, Desvaux *et al.*, 2009, Tseng *et al.*, 2009). Recently, a novel protein secretion system was identified in the crucial periodontopathogen *Porphyromonas gingivalis*, which is referred to as T9SS (Glew *et al.*, 2012).

T9SS is composed of several unique components essential for translocation of proteins with a conserved C-terminal domain (CTD), comprised of 40-70 amino acid residues, across the Gram-negative outer membrane (OM) (Nguyen *et al.*, 2007, Veith *et al.*, 2009, Sato *et al.*, 2013).In total, 31 cargo proteins of T9SS have been identified in *P. gingivalis*, several of which are important virulence factors, including gingipains (RgpA, RgpB and Kgp), peptidylarginine deiminase PPAD, and surface protein hemagglutinin HagA (Veith *et al.*, 2014). All these proteins possess a typical N-terminal signal peptide and traverse the cytoplasmic membrane into the periplasm using the general secretion (Sec) system. Subsequent translocation of these *P. gingivalis* proteins across the OM is enabled by T9SS. T9SS is associated with the cleavage of the CTD prior to carbohydrate modification of the mature protein, probably with anionic lipopolysaccharide (A-LPS) of *P. gingivalis,* which, subsequently, enables anchoring of the protein to the cell surface (Seers *et al.*, 2006, Chen *et al.*, 2011, Shoji *et al.*, 2011, Sato *et al.*, 2013). The modification of *P. gingivalis* CTD proteins with A-LPS was inferred from reactivity with a Manα1-2Manα1-phosphate A-LPS antibody, however, the mode of attachment is not yet fully understood (Saiki and Konishi, 2014), as is also the interplay of outer membrane translocation and posttranslational modification of CTD proteins in general. Recent studies revealed that CTD cleavage is catalyzed by a C-terminal signal peptidase PG0022 (PorU), an essential component of the T9SS secretion machinery, the activity of which is required for cell surface display of certain proteins or their release into the extracellular environment. Inactivation of PG0022 as well as any other T9SS component resulted in accumulation of the otherwise secreted proteins in unprocessed form in the periplasm of *P. gingivalis* (Glew *et al.*, 2012). In total, T9SS has been shown to involve 13 components. PorT is a predicted integral OM protein with a β-barrel fold possibly involved in the formation of a channel for protein translocation (Nguyen *et al.*, 2007). To the remaining OM components - PorK, PorL, PorM, PorN, PorP, PorQ, PorV

and PorW - no specific functions could be assigned, so far (Sato, 2011). LptO of the *P. gingivalis* T9SS conceivably coordinates the secretion of A-LPS and CTD proteins as well as LPS deacylation (Ishiguro *et al.*, 2009, Chen *et al.*, 2011, Saiki and Konishi, 2014). Sov was found in the OM fraction of *P. gingivalis* and shown to be involved in the secretion of gingipains (Saiki and Konishi, 2014), with PorX and PorY being regulatory proteins additionally involved in secretion (Sato *et al.*, 2010).

Bioinformatics analysis of sequenced bacterial genomes revealed the presence of genes encoding the T9SS machinery as well as CTD-proteins exclusively among members of the *Bacteroidetes* phylum, with a total of 663 such proteins predicted in 21 fully sequenced species (Nguyen *et al.*, 2007, Veith *et al.*, 2013) These also include *Tannerella forsythia* and *Prevotella intermedia*, which together with *P. gingivalis* are recognized as important pathogens implicated in development and progression of periodontal diseases (Socransky and Haffajee, 1992, Socransky *et al.*, 1998, Holt and Ebersole, 2005). Similar to *P. gingivalis,* these periodontopathogens secrete large amounts of CTD proteins, many of which have been identified as virulence factors (Veith *et al.*, 2009, Sharma, 2010, Dashper *et al.*, 2011). In the case of *T. forsythia,* the major proteins carrying CTD are BspA and the S-layer proteins TfsA and TfsB (Lee *et al.*, 2006). BspA belongs to the leucine-rich repeat (LRR) and bacterial immunoglobulin-like protein families. BspA is associated with the cell surface of *T. forsythia* and functions as an important modulator of host innate immune responses through activation of TLR2 in cooperation with TLR1 (Onishi *et al.*, 2008). The S-layer proteins TfsA (TF2661-2) and TfsB (TF2663) are intercalated on the bacterial cell surface forming a so far unique 2-dimensional crystalline monolayer (S-layer) of 22 nm thickness (Sekot *et al.*, 2012). The *T. forsythia* S-layer was shown to mediate adhesion to and invasion of carcinoma cells of the mouth (Sabet *et al.*, 2003) and to delay the bacterium's recognition by the innate immune system of the host, at least at the early stage of infection (Sekot *et al.*, 2011). The S-layer proteins are targeted by the general protein *O*-glycosylation system of *T. forsythia* for display of multiple copies of a complex oligosaccharide (Posch *et al.*, 2011). A trisaccharide branch of that oligosaccharide, composed of two *N*-acetylmannosaminuronic acid residues and one modified nonulosonic acid, was shown to act in modulating dendritic cell effector functions to suppress T-helper 17 responses, thereby ensuring the persistence of the pathogen in the host (Settem *et al.*, 2013).

The identification of proteins which are equipped with typical *Bacteroidetes* CTDs together with the presence of a complete set of genes predicted to encode orthologs of the T9SS apparatus of *P. gingivalis* strongly suggests that *T. forsythia* employs a T9SS. To verify this hypothesis, we have deleted either TF2327 or TF0955 in the *T. forsythia* genome, which are orthologs of PorK and PG0022, respectively (Glew *et al.*, 2012), essential for secretion of proteins via T9SS in *P. gingivalis*. The phenotype of isogenic mutants was analyzed with respect to secretion, glycosylation, proteolytic processing and assembly of the S-layer as well as biofilm formation. Inactivation of TF2327 and TF0955 blocked secretion of the S-layer proteins TfsA and TfsB, cleavage of CTD, and assembly of the composite S-layer on the bacterial cell surface. On the other hand, the retained S-layer proteins were modified with the mature *O*-glycan, indicative of protein glycosylation and protein secretion proceeding independently of each other.

METHODS

Bacterial strains and growth conditions

Tannerella forsythia ATCC 43037 (American Type Culture Collection, USA) and defined T9SS mutants (see below) were grown in 37 g L^{-1} of Brain-Heart-Infusion (BHI) liquid media (Oxoid, UK), containing 5 g L^{-1} yeast extract (Oxoid), 0.5 g L^{-1} L-cysteine (Sigma, Austria), 2.5 µg mL^{-1} hemin (Sigma), 2.0 µg mL^{-1} menadione (Sigma), 10 µg mL^{-1} *N*-acetylmuramic acid (Carbosynth, UK) and 5% (v/v) horse serum (Life Technologies, Austria), under anaerobic conditions at 37°C for 4-7 days. For cultivation of *T. forsythia* wild-type and mutants on BHI agar plates (0.8% w/v), the amounts of L-cysteine, hemin, and *N*-acetylmuramic acid were doubled and plates were incubated under anaerobic conditions in an anaerobic jar (AnaeroJar; Oxoid) at 37°C. Media were supplemented with gentamycin and erythromycin at a concentration of 200 µg mL^{-1} and 5 µg mL^{-1}, respectively, when appropriate.

Escherichia coli strains were grown under standard conditions in Luria-Bertani (LB) medium supplemented with 100 µg mL^{-1} ampicillin, when appropriate.

P. gingivalis W83 is used as a reference strain for comparison with predicted components of the T9SS in *T. forsythia* ATCC 43037.

DNA isolation and PCR amplification

Genomic DNA was isolated from 2 mL of bacterial suspension as published previously (Cheng and Jiang, 2006). Plasmid DNA was purified using the GeneJET Plasmid MiniPrep Kit (Thermo Scientific, Austria). PCR fragments were amplified either by Phusion High-Fidelity DNA Polymerase (Thermo Scientific) or by Herculase II Phusion Polymerase (Agilent Technologies, Germany) according to the protocols provided by manufacturers. PCR fragments were purified with the GeneJET PCR Purification Kit (Thermo Scientific). Oligonucleotide primers (Life Technologies) used in the course of this study are listed in Table 1.

Construction of T9SS deficient mutants (TFΔ*0955*, TFΔ*2327*)

Gene knock-outs were based on the homologous recombination of a gene knock-out cassette deleting or disrupting the selected gene of the *T. forsythia* T9SS. Positive clones were selected based on transferred erythromycin resistance.

The gene knock-out cassette for constructing the TFΔ*0955* mutant consisted of a 2,188-bp erythromycin resistance gene *ermF-ermAM* (primer pair 415/416) flanked by homologous up- and down-stream regions (Fig. 1A). Genomic DNA from *T. forsythia* Δ*wecC* (obtained from A. Sharma, State University of New York at Buffalo, USA) was used as a template. The 1,035-bp up-stream and 1,027-bp down-stream homology regions were amplified from genomic DNA of *T. forsythia* wild-type cells using the primer pairs 413/414 and 417/418, respectively. The three fragments were joined by overlap extension (OE-) PCR and sub-cloned into the blunt-end cloning vector pJET1.2 (Thermo Scientific) resulting in pJET1.2/TF0955ko. Approximately 5 µg of the circular gene knock-out construct were transferred by electroporation into 100 µL of *T. forsythia* culture grown to the early stationary phase. Cells were regenerated in BHI medium for 24 h before plating on BHI agar plates containing erythromycin as selection marker. Single colonies were picked and used for inoculation of liquid BHI medium. Once bacterial growth was visible, genomic DNA was isolated to confirm the integration of the knock-out cassette via PCR amplification using the primer pair 426/429.

The construction of the TFΔ*2327* mutant was done essentially the same way as described for TFΔ*0955* (Fig. 1B). Genomic integration of the gene knock-out cassette was verified using the primer pair 448/449.
Growth of the T9SS mutants was measured over a period of 5 days and the obtained growth curves were compared to that of *T. forsythia* wild-type.

SDS-PAGE and Western immunoblotting

Crude cell extracts of *T. forsythia* wild-type and mutant cells (TFΔ*0955*, TFΔ*2327*) were run on 10% SDS-PAGE gels in a Mini Protean electrophoresis apparatus (Bio-Rad, Austria) according to a standard protocol and visualized with colloidal Coomassie Brilliant Blue R-250 (Laemmli, 1970).

For Western immunoblotting, proteins were transferred onto a polyvinylidene difluoride (PVDF) membrane (Bio-Rad) using a Mini Trans-Blot Cell (Bio-Rad). Polyclonal rabbit antiserum raised against the recombinant S-layer proteins TfsA (α-TfsA) and TfsB (α-TfsB) was used in combination with goat anti-rabbit secondary antibody labeled with IRDye 800CW (LI-COR Biosciences, USA). Visualization of the S-layer proteins was done at 800 nm using the Odyssey Infrared Imaging System (LI-COR Biosciences). Staining of carbohydrates was done with the periodic acid-Schiff (PAS) reagent following a standard protocol (Doerner and White, 1990). For semi-quantification of S-layer proteins present in *T. forsythia* wild-type and TFΔ*0955* and TFΔ*2327* mutant cells, equal amounts of total cell protein were loaded on the SDS-PA gels in two different dilutions, each, followed by Western immunoblotting. The integrated intensity of the detected bands was determined using the LI-COR Odyssey Application software 3.0.21.

S-layer *O*-glycan preparation and LC ESI-MS/MS

S-layer *O*-glycans were isolated from *T. forsythia* mutant cells following a published protocol (Posch *et al.*, 2011). Briefly, S-layer glycoproteins were excised from Coomassie Brilliant Blue R-250-stained SDS-PA gels and *O*-glycans were released from the protein backbone by in-gel reductive *β*-elimination followed by removal of excess of salt using a 25-mg HyperSep Hypercarb SPE cartridge (Thermo Scientific). Borohydride-reduced *O*-glycans were analyzed by PGC-ESI-MS/MS as described recently (Hypercarb, 0.32 x 150 mm, particle size 5 µm) (Stadlmann *et al.*, 2008). Detection was done by an ESI-Q-TOF Global Ultima from Micromass (Waters, USA). Data were evaluated using MassLynx 4.0 software. MS/MS experiments were performed at 30% collision energy using CID with argon gas. Slices from the SDS-polyacrylamide gel without sample application were treated in the same way as described for the *O*-glycan samples and acquired MS spectra were used as a negative control.

Proteomics analysis

S-layer proteins as present in *T. forsythia* wild-type and mutant cells were identified as follows. After in-gel tryptic digest of the respective high molecular-mass bands on SDS-PAGE gels, generated peptides were extracted and subjected to reverse phase ESI-MS/MS peptide mapping using a Bruker IonTrap AmaZon speed ETD and a Bruker Maxis4G. Protein database search of tandem MS data was done using the ProteinScape 3 software tool (Bruker-Daltonik, Germany) with Mascot 2.3.02 algorithm for peptide identification or the X!tandem algorithm (The Global Proteome Machine Organization).

Scanning electron microscopy

T. forsythia cells with an initial optical density of OD_{600} = 0.07 were grown anaerobically on sterile glass slides at 37°C for 2-4 days. After fixation with 2.5% glutaraldehyde in phosphate buffered saline (PBS) at 4°C for 2-4 h, the cells were dehydrated with increasing concentrations of ethanol (20% to 100% ethanol in increments of 10%) for 7 min, each. Subsequently, the cells were treated with 33% and 66% bis(trimethylsilyl)amine (HMDS) in methanol and, eventually, with 100% HMDS for 30 min, each. Glass slides were air-dried and mounted onto 12.5-mm aluminum stubs covered with double-sided carbon tabs (Gröpl, Austria). At one side of the glass slide, one drop of silver solution was applied prior to sputter-coating with gold (EM SDC005, Leica). Biofilms were imagined using an Inspect S50 scanning electron microscope (FEI, The Netherlands).

Ultra-thin sectioning and transmission electron microscopy

Ultrathin-sectioning of *T. forsythia* wild-type and TFΔ*0955* and TFΔ*2327* cells was carried out as described previously (Messner *et al.*, 1984). Briefly, the room-temperature processing included fixation of samples with 2.5% (w/v) paraformaldehyde/ 2.5% (w/v) glutaraldehyde/ 0.5% (w/v) tannin in cacodylate buffer and osmium tetroxide fixation without ruthenium red. Dehydration was performed in an increasing alcohol series before samples were embedded in Epon resin (Gröpl, Austria).

Ultrathin-sectioned samples were investigated in a Tecnai G^2 20 Twin transmission electron microscope (TEM; FEI), operating at 80 keV. Pictures were taken with an FEI Eagle 4k CCD camera (4,096 x 4,096 pixels) and images were processed using software developed in house based on Fourier domain techniques (Amos *et al.*, 1982, Crowther *et al.*, 1996).

Confocal laser scanning microscopy

T. forsythia cells were grown under standard conditions as described above until an OD_{600} of 0.5 was reached (~2 days). Subsequently, the bacterial culture was inoculated into fresh growth medium. After 4 days of incubation, 3 ml of that culture were transferred into micro dishes (µ-dishes, 35 mm height; Ibidi, Germany) and biofilm cells were grown anaerobically at 37°C for 3 days. The liquid supernatant containing planktonic cells was removed and the biofilm was stained with Hoechst 33258 (Sigma) at a final concentration of 10 µg mL^{-1} in PBS (stock solution, 10 mg mL^{-1} in distilled water) for 35 min in the dark. The supernatant was discarded and the biofilm was washed with PBS. Images were taken using an inverted TCS-SP5 confocal microscope (Leica Microsystems, Germany) and images were processed with the software ImageJ (Schneider *et al.*, 2012).

RESULTS

Construction of T9SS deficient *T. forsythia*

The Gram-negative periodontopathogen *T. forsythia* encodes a full set of proteins that are orthologs of the *P. gingivalis* components constituting the recently described type IX secretory machinery (T9SS). To investigate, if T9SS is functional in *T. forsythia*, we constructed deletion mutants affecting either TF*0955* or TF*2327*, which are predicted orthologs of the C-terminal signal peptidase (sortase; PG0022) and the PorK component (PG1676), both of which have been shown to be essential for CTD-protein secretion via T9SS in *P. gingivalis* (Chen *et al.*, 2011). The *T. forsythia* T9SS orthologs were identified

using the Basic Local Alignment Search Tool BLAST (at http://blast.ncbi.nlm.nih.gov/Blast.cgi). They are scattered across the bacterial genome and include, besides the genes targeted in the course of this study, TF0188 (PorT), TF0671 (PorY), TF1742 (PorW), TF1930 (PorX), TF1959 (PorQ), TF2326 (PorP), TF2328 (PorL), TF2329 (PorM), TF2330 (PorN), TF2852 (LptO) and TF2901 (Sov) (McBride and Zhu, 2013). In the case of the ΔTF*0955* mutant, no polar effects were expected, since the downstream gene TF*0957* is separated by 106 non-coding nucleotides (Fig. 1A). In contrast, the other targeted gene, TF*2327*, is predicted to constitute an operon (BioCyc) (Caspi *et al.*, 2012) together with the immediate downstream genes TF*2328*, TF*2329* and TF*2330*, implicating that polar effects might cause the inactivation of the said genes in the mutant (Fig. 1B). However, since these latter genes code for predicted components of the T9SS, even their inactivation would not have adversely affected the usability of the ΔTF*2327* mutant for the purpose of this study. For both T9SS mutants, genomic integration of the *ermF-AM* cassette was confirmed by PCR, with the amplicons obtained for ΔTF*0955* (primer pair 426/429) and ΔTF*2327* (primer pair 448/449) conforming with the calculated size of 7,678 bp (Fig. 1C, lane 3) and 4,277 bp (Fig. 1C, lane 5), respectively. As a control, the same primer pairs were used with wild-type genomic DNA as a template (Fig. 1C, lanes 2 and 4).

The *T. forsythia* T9SS mutants had similar growth characteristics, revealing a slightly decreased growth rate in comparison to the wild-type (data not shown).

S-layer protein expression in T9SS mutants

Among several proteins of *T. forsythia* bearing a conserved CTD and, therefore, predicted to be translocated across the OM via a T9SS comparable to that established for *P. gingivalis* (Glew *et al.*, 2012), are the S-layer proteins TfsA and TfsB. Both are equipped with a typical CTD containing a possible C-terminal signal peptidase cleavage site. The S-layer proteins are modified with *O*-glycans of known structure (Posch *et al.*, 2011) and assemble in equimolar ratio into a 2-dimensional crystalline surface layer completely covering *T. forsythia* cells throughout the whole life-cycle. This translates into a high cellular protein biosynthesis effort being devoted to S-layer protein biosynthesis (Sleytr *et al.*, 1993). Thus, these abundant cell surface proteins are ideal candidates to investigate, if indeed a T9SS is employed in *T. forsythia*. SDS-PAGE analysis of crude cell extracts of *T. forsythia* wild-type shows the two prominent high molecular-mass S-layer protein bands with apparent molecular masses above 250 kDa (Fig. 2A, lane 2). These correspond to TfsA with ~230 kDa and TfsB with ~270 kDa (Posch *et al.*, 2011), the respective protein portions of which have a calculated molecular mass of 135 kDa and 152 kDa, respectively. The observed apparent molecular masses are in agreement with the migration behavior of the mature and glycosylated forms of TfsA and TfsB on SDS-PA gels as reported previously (Posch *et al.*, 2011). Conforming with the glycoprotein nature of TfsA and TfsB, both bands strongly stained for carbohydrates with the PAS reagent (Fig. 2B, lane 2). In both the ΔTF*0955* (Fig. 2B, lane 3) and the ΔTF*2327* Fig. 2B, lane 4) mutants, also two prominent protein bands that were carbohydrate-positive, were detected; however, these bands were clearly downshifted compared to the wild-type (Fig. 2AB, lanes 3 and 4). For *T. forsythia* wild-type as well as the T9SS mutants, the corresponding bands were unambiguously identified as S-layer proteins by Western-immunoblotting using polyclonal α-TfsA (Fig. 2C) and α-TfsB antibodies (Fig. 2D), respectively. Note that the α-TfsA antibody recognized a second protein band that co-migrates with the TfsA wild-type protein (Fig. C) In addition, proteomics analysis revealed tryptic peptides originating from either of the S-layer proteins (TfsA and TfsB; see Table 2 below).

Semi-quantification of the S-layer glycoprotein bands as visualized on Western-immunoblots upon separation of equal amounts of cellular protein confirmed that TfsA and TfsB are expressed in equal amounts in the *T. forsythia* wild-type as well as in the ΔTF*0955* and the ΔTF*2327* mutants. However, the integrated intensity of either S-layer protein expressed in the mutant cells was only about one third compared to that of the wild-type cells, possibly due of feed-back inhibition, since the protein secretion was blocked. Integrated intensity values were 34.5 and 35.5 for TfsB from ΔTF*0955* and ΔTF*2327*, respectively, opposed to 94.7 from wild-type, and 21.3 for TfsA and 22.0 from ΔTF*0955* and ΔTF*2327*, respectively, as opposed to 61.2 from wild-type.

C-terminal domains are not cleaved off S-layer proteins in T9SS deficient strains

The identification of glycosylated forms of the S-layer proteins TfsA and TfsB in the ΔTF*0955* and ΔTF*2327* mutants - albeit with decreased apparent molecular mass as compared to the *T. forsythia* wild-type glycoproteins (Fig. 2B) - contrasted the situation known from T9SS deficient mutants of *P. gingivalis*, where CTD proteins were shown to be retained in the periplasm in non-glycosylated form (Chen *et al.*, 2011). To investigate this finding in more detail, the S-layer proteins from the wild-type and the T9SS deficient mutants were analyzed for the presence of the CTD. For this purpose, the S-layer glycoprotein bands of *T. forsythia* wild-type and ΔTF*0955* and ΔTF*2327* mutants as revealed upon SDS-PAGE (marked with an asterisk in Fig. 2A) were digested with trypsin and subjected to MS peptide mapping (Table 2). This clearly confirmed the identity of the investigated bands as TfsA and TfsB S-layer proteins; sequence coverage for TfsA and TfsB, respectively, was 51% and 48% for *T. forsythia* wild-type, 20% and 32% for ΔTF*2327*, and 16% and 36% for ΔTF*0955*. For the mutants, several peptides mapping to the CTD were detected both for TfsA and TfsB, whereas no single peptide from this region was derived from the S-layer proteins of *T. forsythia* wild-type (Table 2). This data implicates that no cleavage of the CTD occurs from the periplasmically localized S-layer proteins in the *T. forsythia* secretion mutants, whereas the wild-type proteins are C-terminally fully processed.

T9SS is not coupled with *O*-glycan assembly and glycan transfer onto proteins

According to our data, in either T9SS deficient *T. forsythia* mutant, essentially the same TfsA and TfsB glycoproteins were produced. These were already N-terminally processed for Sec-mediated translocation across the CM as evident from the MS data obtained in the course of the peptide mapping (data not shown). Inactivation of either the signal peptidase (TF0955) or PorK (TF2327) prevented CTD-cleavage causing entrapment of the S-layer proteins in the periplasm. While glycosylation of TfsA and TfsB from the mutants is in line with the general knowledge of protein glycosylation in Gram-negative bacteria to be elaborated in the periplasm (Wacker *et al.*, 2002), a clear downshift in molecular mass of these glycoproteins in comparison to the wild-type glycoproteins (Fig. 2B) was initially assumed to be due to incomplete protein glycosylation. To shed more light on glycosylation status of the S-layer proteins from the *T. forsythia* secretion mutants, their glycans were analyzed in detail. To this end, in-gel reductive *β*-elimination was applied to release *O*-glycans from the protein backbones of TfsA and TfsB from ΔTF*2327*, ΔTF*0955* as well as from the wild-type that served as a reference. Mass analyses presented the wild-type *O*-glycan with the highest mass peak at 1898, which is in agreement with the known structure of the *T. forsythia* dekasaccharide (Fig. 3) (Posch *et al.*, 2011). Additional glycan fragments at mass to charge ratios of 1752 and 1622, corresponding to the *O*-glycan devoid of a branching fucose and a

digitoxose residue, respectively, were also resolved; this structure are due to previously described *O*-glycan microheterogeneity (Coyne *et al.*, 2013). Significantly, S-layer glycans from the T9SS deficient *T. forsythia* mutants showed the same pattern and masses on deconvoluted ESI-TOF-MS spectra as the wild-type *O*-glycan, proving the complete glycan to be attached to the S-layer proteins. In negative controls, performed with SDS-PA gel slices without sample application and Coomassie Brilliant Blue R-250 staining, no glycans were detected. Concluding, full assembly of the *T. forsythia O*-glycan and its transfer onto the proteins occur prior to T9SS-mediated translocation across the OM and, therefore CTD protein glycosylation and CTD protein secretion are independent processes. With respect to the still missing explanation of the downshifted S-layer glycoprotein bands from the ΔTF*2327* and ΔTF*0955* mutants on SDS-PA gels (Fig. 2), it is tempting to speculate that either less glycosylation sites might be occupied in the mutant S-layer proteins or that the mature, secreted S-layer glycoproteins are complexed with a still to be identified compound, possibly involved in anchoring of the S-layer to the cell surface (Posch *et al.*, 2013).

C-terminal domain proteins of *T. forsythia* are translocated across the T9SS

The intercalating high molecular-mass glycoproteins TfsA and TfsB self-assemble into a 2-dimensional crystalline monolayer on the cell surface of *T. forsythia* (Sekot *et al.*, 2012*)*. The square S-layer lattice has a lattice spacing of approximately 10 mm x 10 mm and a thickness of about 22 nm (Sekot *et al.*, 2012). To visualize that inactivation of T9SS blocks TfsA and TfsB secretion and, consequently, gives rise to an S-layer deficient cell envelope architecture, ultra-thin cross-sections of whole cell preparations from the ΔTF*2327* and ΔTF*0955* mutants were investigated in comparison to *T. forsythia* wild-type cells (Fig. 4). Transmission electron micrographs clearly confirmed the absence of the S-layer on either of the T9SS deficient mutant, while the wild-type cells are fully covered with a serrated S-layer typical of *T. forsythia* (Sekot *et al.*, 2012).

Phenotype of biofilm is largely affected in T9SS deficient *T. forsythia*

Following up the previous finding of the *T. forsythia* S-layer playing a role in biofilm formation (Honma *et al.*, 2007), the capacity of the T9SS deficient mutants to form a monospecies-biofilm as well as the architecture of the biofilm was investigated. Scanning electron microscopy of 3 day-old biofilms grown on a glass surface revealed that the *T. forsythia* wild-type strain formed a meshwork of well-connected individual rod-shaped cells of comparable size. The biofilm contained many voids and tunnel-like structures which are known to enable nutrition delivery to the basis of the biofilm (Fig. 5A). The biofilm of the ΔTF*0955* and ΔTF*2327* mutant, in contrast, revealed an entirely different architecture. The rods were elongated and stuck together, forming fibers and bundles of cells. The matrix was denser compared to the wild-type biofilm with less abundant voids and tunnel-like structures. Individual cells appeared to have a smoother surface indicating the lack of several proteins at the outermost layer of the bacterium (Fig. 5A).

The formation of a much denser biofilm by the mutants compared to the wild-type was confirmed by CLSM analysis (Fig. 5B). Although the overall height of the biofilm was comparable between all analyzed strains, the mutant cells were more densely packed and often stuck together forming lateral string-like structures, most clearly visible in the ΔTF*2327* mutant biofilm.

DISCUSSION

Putative CTD proteins have been found in 87 organisms so far, all belonging to the *Bacteroidetes* phylum, and their presence coincides with the occurrence of orthologous genes of essential components encoding the T9SS machinery. Of note, not all species of this phylum have the genetic information for a T9SS and for CTD-bearing proteins, such as, for instance, the human intestinal symbionts *Bacteroides fragilis* and *Bacteroides thetaiotamicron*. In *T. forsythia*, 37 proteins are predicted to possess a CTD, out of which 19 have been verified experimentally (Veith *et al.*, 2013). Here, we have shown for the first time that secretion of these proteins indeed occurs via T9SS.

To this end, we have studied the effect of the inactivation of two *T. forsythia* orthologs of the genes *porK* (TF*2327*) *and porU* (TF*0955*) that are essential for T9SS function in *P. gingivalis*, on the secretion of the two CTD-bearing S-layer proteins TfsA and TfsB and their posttranslational modification (glycosylation (Posch *et al.*, 2011)) as well as on the assembly of the 2-dimensional crystalline S-layer on the bacterial cell surface.

In analogy to the T9SS of *P. gingivalis*, TF0955 would be located at the cell surface and is supposed to function as the C-terminal signal peptidase required for cleaving off the CTD from proteins deemed for secretion (Glew *et al.*, 2012). TF2327 is a predicted integral OM protein of unknown function with a β-barrel structure. Despite differences in subcellular location and function of the targeted proteins, both mutants - ΔTF*0955* and ΔTF*2327* - showed the same phenotype. At the cellular level, the most profound difference in comparison to the *T. forsythia* wild-type was the loss of the S-layer (Fig. 2). A similar effect was seen for a *P. gingivalis lptO* (PG0027) knock-out mutant missing its electron-dense surface layer (EDSL) (Saiki and Konishi, 2014). However, this effect is most probably not attributable to the disruption of LptO alone (Glew *et al.*, 2012), but must be associated with inactivation of any other essential component of T9SS. Apart from the lack of the S-layer, T9SS mutants of *T. forsythia* show abnormalities with respect to the shape of individual cells and the architecture of monospecies biofilms as shown by SEM and CLSM analysis (Fig. 5). These abnormalities are most likely a combined effect of the blocked secretion of CTD proteins in the *T. forsythia* ΔTF*0955* and ΔTF*2327* mutants.

Inactivation of T9SS in *P. gingivalis* resulted in the retention of otherwise secreted proteins in the periplasm in non-glycosylated and unprocessed form still bearing the CTD (Chen *et al.*, 2011). Consequently, also the TfsA and TfsB S-layer proteins from the *T. forsythia* secretion mutants were found with intact CTDs (Table 2). In contrast, the CTDs from the S-layer proteins of the wild-type bacterium are cleaved at the KSPL#ANAE (TfsA) and the EDPV#ANEI (TfsB) motifs (Veith *et al.*, 2013) during or immediately after translocation across the OM. On the other hand, no difference was found in *O*-glycosylation of S-layer proteins derived from the wild-type and T9SS-deficient mutants. Clearly, fully assembled *O*-glycans were attached to both glycoproteins regardless of T9SS functionality, even with glycan microheterogeneity equaling that of the wild-type S-layer glycoproteins (Fig. 3). This finding indicates that glycan assembly and transfer of the glycan to multiple three amino acid motifs D(S/T)(A/I/L/M/T/V) within the CTD S-layer proteins is completed while proteins reside in the periplasm, which is in accordance with the general *O*-glycosylation pathway in *Bacteroidetes* (Coyne *et al.*, 2013).

The same *O*-glycosylation of the S-layer proteins together with the lack of their C-terminal processing indicates that another type of posttranslational modification might be responsible for the accounted molecular mass difference of about 50-kDa between the S-layer glycoproteins derived from the *T. forsythia* wild-type strain and the T9SS-deficient mutants (Fig. 2). In analogy to *P. gingivalis*, we hypothesize that in *T. forsythia*, proteolysis of CTD is

coupled with or succeeded by the attachment of a glycan moiety to the protein in a manner different from *O*-glycosylation. Such a glycan would have escaped the MS analysis performed in the course of the present study due to the protocol used for sample preparation and, consequently, could have remained undetected. It is conceivable that, as in *P. gingivalis*, the attached glycan moiety could be a variant of the LPS. This might be involved in anchoring of the TfsA and TfsB S-layer glycoproteins to the OM of *T. forsythia*. This is in line with our assumption of the rough-type LPS of *T. forsythia* being ideally suited for the mediation of protein cell surface attachment because of its defined length and high abundance on the cell envelope (Posch *et al.*, 2013). This hypothesis is currently under investigation in our laboratories.

Yet we have provided the first evidence of the presence of a functional T9SS in *T. forsythia* and confirm the importance of CTDs for successful secretion of two abundant *T. forsythia* proteins. The S-layer glycoproteins TfsA and TfsB were prevented from crossing the OM in T9SS deficient mutants (TFΔ*0955*, TFΔ*2327*), which, consequently, also lacked the S-layer. Further, we could show that protein translocation and *O*-glycosylation are decoupled from each other.

ACKNOWLEDGEMENTS

We thank Andrea Scheberl for excellent technical assistance with ultra-thin sectioning. Financial support came from the Austrian Science Fund FWF, project P24317-B22 (to C.S.) and the PhD programme “BioToP - Biomolecular Technology of Proteins” (Austrian Science Fund, FWF project W1224). J.P. acknowledges support by grants from US NIH (DE 09761), the European Commission (FP7-PEOPLE-2011-ITN-290246 "RAPID" and FP7-HEALTH-F3-2012-306029 "TRIGGER"), and the National Science Center (2012/04/A/NZ1/00051, NCN, Krakow, Poland). The Faculty of Biochemistry, Biophysics and Biotechnology of the Jagiellonian University is a beneficiary of structural funds from the European Union (POIG.02.01.00-12-064/08).

REFERENCES

Amos, L. A., R. Henderson & P. N. Unwin (1982) Three-dimensional structure determination by electron microscopy of two-dimensional crystals. *Prog Biophys Mol Biol* **39**: 183-231.

Caspi, R., T. Altman, K. Dreher, C. A. Fulcher, P. Subhraveti, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, A. Pujar, A. G. Shearer, M. Travers, D. Weerasinghe, P. Zhang & P. D. Karp (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **40**: D742-753.

Chen, Y. Y., B. Peng, Q. Yang, M. D. Glew, P. D. Veith, K. J. Cross, K. N. Goldie, D. Chen, N. O'Brien-Simpson, S. G. Dashper & E. C. Reynolds (2011) The outer membrane protein LptO is essential for the O-deacylation of LPS and the co-ordinated secretion and attachment of A-LPS and CTD proteins in *Porphyromonas gingivalis*. *Mol Microbiol* **79**: 1380-1401.

Cheng, H. R. & N. Jiang (2006) Extremely rapid extraction of DNA from bacteria and yeasts. *Biotechnol Lett* **28**: 55-59.

Coyne, M. J., C. M. Fletcher, M. Chatzidaki-Livanis, G. Posch, C. Schäffer & L. E. Comstock (2013) Phylum-wide general protein *O*-glycosylation system of the *Bacteroidetes*. *Mol Microbiol* **88**: 772-783.

Crowther, R. A., R. Henderson & J. M. Smith (1996) MRC image processing programs. *J Struct Biol* **116**: 9-16.

Dashper, S. G., C. A. Seers, K. H. Tan & E. C. Reynolds (2011) Virulence factors of the oral spirochete *Treponema denticola*. *J Dent Res* **90**: 691-703.

Desvaux, M., M. Hebraud, R. Talon & I. R. Henderson (2009) Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends Microbiol* **17**: 139-145.

Doerner, K. C. & B. A. White (1990) Detection of glycoproteins separated by nondenaturing polyacrylamide gel electrophoresis using the periodic acid-Schiff stain. *Anal Biochem* **187**: 147-150.

Glew, M. D., P. D. Veith, B. Peng, Y. Y. Chen, D. G. Gorasia, Q. Yang, N. Slakeski, D. Chen, C. Moore, S. Crawford & E. C. Reynolds (2012) PG0026 is the C-terminal signal peptidase of a novel secretion system of *Porphyromonas gingivalis*. *J Biol Chem* **287**: 24605-24617.

Holt, S. C. & J. L. Ebersole (2005) *Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia*: the "red complex", a prototype polybacterial pathogenic consortium in periodontitis. *Periodontol 2000* **38**: 72-122.

Honma, K., S. Inagaki, K. Okuda, H. K. Kuramitsu & A. Sharma (2007) Role of a *Tannerella forsythia* exopolysaccharide synthesis operon in biofilm development. *Microb Pathog* **42**: 156-166.

Ishiguro, I., K. Saiki & K. Konishi (2009) PG27 is a novel membrane protein essential for a *Porphyromonas gingivalis* protease secretion system. *FEMS Microbiol Lett* **292**: 261-267.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Lee, S. W., M. Sabet, H. S. Um, J. Yang, H. C. Kim & W. Zhu (2006) Identification and characterization of the genes encoding a unique surface (S-) layer of *Tannerella forsythia*. *Gene* **371**: 102-111.

McBride, M. J. & Y. Zhu (2013) Gliding motility and Por secretion system genes are widespread among members of the phylum *Bacteroidetes*. *J Bacteriol* **195**: 270-278.

Messner, P., F. Hollaus & U. B. Sleytr (1984) Paracrystalline cell wall surface layer of different *Bacillus stearothermophilus* strains. *Int J Syst Bacteriol* **34**: 202-210.

Nguyen, K. A., J. Travis & J. Potempa (2007) Does the importance of the C-terminal residues in the maturation of RgpB from *Porphyromonas gingivalis* reveal a novel mechanism for protein export in a subgroup of Gram-negative bacteria? *J Bacteriol* **189**: 833-843.

Onishi, S., K. Honma, S. Liang, P. Stathopoulou, D. Kinane, G. Hajishengallis & A. Sharma (2008) Toll-like receptor 2-mediated interleukin-8 expression in gingival epithelial cells by the *Tannerella forsythia* leucine-rich repeat protein BspA. *Infect Immun* **76**: 198-205.

Posch, G., O. Andrukhov, E. Vinogradov, B. Lindner, P. Messner, O. Holst & C. Schäffer (2013) Structure and immunogenicity of the rough-type lipopolysaccharide from the periodontal pathogen *Tannerella forsythia*. *Clin Vaccine Immunol* **20**: 945-953.

Posch, G., M. Pabst, L. Brecker, F. Altmann, P. Messner & C. Schäffer (2011) Characterization and scope of S-layer protein *O*-glycosylation in *Tannerella forsythia*. *J Biol Chem* **286**: 38714-38724.

Pugsley, A. P. (1993) The complete general secretory pathway in gram-negative bacteria. *Microbiol Rev* **57**: 50-108.

Sabet, M., S. W. Lee, R. K. Nauman, T. Sims & H. S. Um (2003) The surface (S-) layer is a virulence factor of *Bacteroides forsythus*. *Microbiology* **149**: 3617-3627.

Saiki, K. & K. Konishi (2014) *Porphyromonas gingivalis* C-terminal signal peptidase PG0026 and HagA interact with outer membrane protein PG27/LptO. *Mol Oral Microbiol* **29**: 32-44.

Sato, K. (2011) Por secretion system of *Porphyromonas gingivalis*. *J. Oral Biosci.* **53**: 187-196.

Sato, K., M. Naito, H. Yukitake, H. Hirakawa, M. Shoji, M. J. McBride, R. G. Rhodes & K. Nakayama (2010) A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc Natl Acad Sci U S A* **107**: 276-281.

Sato, K., H. Yukitake, Y. Narita, M. Shoji, M. Naito & K. Nakayama (2013) Identification of *Porphyromonas gingivalis* proteins secreted by the Por secretion system. *FEMS Microbiol Lett* **338**: 68-76.

Schneider, C. A., W. S. Rasband & K. W. Eliceiri (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**: 671-675.

Seers, C. A., N. Slakeski, P. D. Veith, T. Nikolof, Y. Y. Chen, S. G. Dashper & E. C. Reynolds (2006) The RgpB C-terminal domain has a role in attachment of RgpB to the outer membrane and belongs to a novel C-terminal-domain family found in *Porphyromonas gingivalis*. *J Bacteriol* **188**: 6376-6386.

Sekot, G., G. Posch, P. Messner, M. Matejka, X. Rausch-Fan, O. Andrukhov & C. Schäffer (2011) Potential of the *Tannerella forsythia* S-layer to delay the immune response. *J Dent Res* **90**: 109-114.

Sekot, G., G. Posch, Y. J. Oh, S. Zayni, H. F. Mayer, D. Pum, P. Messner, P. Hinterdorfer & C. Schäffer (2012) Analysis of the cell surface layer ultrastructure of the oral pathogen *Tannerella forsythia*. *Arch Microbiol* **194**: 525-539.

Settem, R. P., K. Honma, T. Nakajima, C. Phansopa, S. Roy, G. P. Stafford & A. Sharma (2013) A bacterial glycan core linked to surface (S)-layer proteins modulates host immunity through Th17 suppression. *Mucosal Immunol* **6**: 415-426.

Sharma, A. (2010) Virulence mechanisms of *Tannerella forsythia*. *Periodontol 2000* **54**: 106-116.

Shoji, M., K. Sato, H. Yukitake, Y. Kondo, Y. Narita, T. Kadowaki, M. Naito & K. Nakayama (2011) Por secretion system-dependent secretion and glycosylation of *Porphyromonas gingivalis* hemin-binding protein 35. *PLoS One* **6**: e21372.

Sleytr, U. B., P. Messner, D. Pum & M. Sára (1993) Crystalline bacterial cell surface layers. *Mol Microbiol* **10**: 911-916.

Socransky, S. S. & A. D. Haffajee (1992) The bacterial etiology of destructive periodontal disease: current concepts. *J Periodontol* **63**: 322-331.

Socransky, S. S., A. D. Haffajee, M. A. Cugini, C. Smith & R. L. Kent, Jr. (1998) Microbial complexes in subgingival plaque. *J Clin Periodontol* **25**: 134-144.

Stadlmann, J., M. Pabst, D. Kolarich, R. Kunert & F. Altmann (2008) Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. *Proteomics* **8**: 2858-2871.

Tseng, T. T., B. M. Tyler & J. C. Setubal (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol* 9 Suppl 1: S2.

Veith, P. D., Y. Y. Chen, D. G. Gorasia, D. Chen, M. D. Glew, N. M. O'Brien-Simpson, J. D. Cecil, J. A. Holden & E. C. Reynolds (2014) *Porphyromonas gingivalis* outer

membrane vesicles exclusively contain outer membrane and periplasmic proteins and carry a cargo enriched with virulence factors. *J Proteome Res*.

Veith, P. D., N. A. Nor Muhammad, S. G. Dashper, V. A. Likic, D. G. Gorasia, D. Chen, S. J. Byrne, D. V. Catmull & E. C. Reynolds (2013) Protein substrates of a novel secretion system are numerous in the *Bacteroidetes* phylum and have in common a cleavable C-terminal secretion signal, extensive post-translational modification, and cell-surface attachment. *J Proteome Res* **12**: 4449-4461.

Veith, P. D., N. M. O'Brien-Simpson, Y. Tan, D. C. Djatmiko, S. G. Dashper & E. C. Reynolds (2009) Outer membrane proteome and antigens of *Tannerella forsythia*. *J Proteome Res* **8**: 4279-4292.

Wacker, M., D. Linton, P. G. Hitchen, M. Nita-Lazar, S. M. Haslam, S. J. North, M. Panico, H. R. Morris, A. Dell, B. W. Wren & M. Aebi (2002) N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298**: 1790-1793.

[Figure legends]

Figure 1. Strategy for TF0955 gene disruption and TF2327 gene deletion at the *T. forsythia* wild-type locus and confirmation by PCR (not drawn to scale). (A) Genomic organization of the TF0955 gene showing the up- and down-stream homology regions used for homologous recombination and insertion of the selectable *ermF-AM* cassette disrupting TF0955. Primers (compare with Table 1) are indicated with arrows. (B) TF2327 gene deletion and insertion of the selectable *ermF-AM* cassette is shown in the context of its genomic region. The area of homologous recombination is indicated, the position of the primers are marked with arrows. (C) Agarose gel electrophoresis of PCR products from genomic DNA of *T. forsythia* wild-type and mutants with integrated *ermF-AM* cassette. GeneRuler 1 kb Plus DNA Ladder, Thermo Scientific (lane 1), *T. forsythia* wild-type (primer pair 426/ 429) yielding a 5,490-bp PCR-product (lane 2) and with integrated *ermF-AM* cassette, yielding a 7,678-bp PCR-product(lane 3); *T. forsythia* wild-type (primer pair 448/449) yielding a 3,529-bp PCR-product (lane 4) and with integrated *ermF-AM* cassette, yielding a 4,277-bp PCR product (lane 5).

Figure 2. (A) Coomassie Brilliant Blue R-250 stained SDS gel (10%) of crude cell extracts from *T. forsythia* wild type (lane 2), Δ*TF2327* (lane 3) and Δ*TF0955* (lane 4). The prominent S-layer proteins (identity confirmed by proteomics analyses) are significantly downshifted in both T9SS deficient mutants, which is also evident from the PAS-staining (B) and from Western-immunoblots probing for TfsA (C) and TfsB (D). Note that the α-TfsA antibody recognizes a second protein band that co-migrates with the TfsA wild-type protein. PageRuler Plus PreStained Protein Ladder (lane 1). Asterisks indicate S-layer glycoprotein bands that were subsequently processed for MS analysis (Table 2, Fig. 3).

Figure 3. (A) Deconvoluted ESI-TOF-MS spectrum of TfsA and TfsB *O*-glycans in *T. forsythia* wild-type and mutant strains. Identical spectra were recorded for the *O*-glycan derived from the Δ*TF0955* and Δ*TF2327* mutants and the wild-type. (B) For clarification, the complete S-layer glycan structure is shown. Gal, galactose; GlcA, glucuronic acid; Fuc, fucose; ManNAcA, N-acetyl-mannosaminuronic acid; Dig, digitoxose; Xyl, xylose; Pse5Am7Gc, 5-acetimidyl-7-glycolyl-pseudaminic acid; Ser, serine; Thr, threonine.

Figure 4. TEM micrographs showing ultra-thin cross-sectioned cells of *T. forsythia* and T9SS mutants. *T. forsythia* wild-type displays a serrated S-layer, whereas no S-layer is present on

T. forsythia Δ*TF0955* and Δ*TF2327* cells. SL, S-layer; OM, outer membrane; PP, periplasm; CM, cytoplasmic membrane; C, cytoplasm.

Figure 5. (A) SEM micrographs showing altered biofilm and cell morphology in T9SS deficient mutants in comparison to the wild-type. Biofilm was grown anaerobically on glass for two days; (B) CLSM micrographs of *T. forsythia* wild-type and mutant biofilms. Mutant biofilms revealed denser growth and a height comparable to that of the wild-type biofilm. Size bar, 20 µm.

[Tables]

Table 1. Oligonucleotides used in this study

#	Sequence (5´ → 3´)
413	TCCACGGAGTAATTCCTCAGAGCGATTG
414	*GAAGCTATGGGGGTACCTCCCCCGGG*AGTTGATTTTTCCAACTGCCGTAGGCCGGGGTC
415[a]	CCCGGGGGAGGTACCCCCGATAGCTTC
416[a]	CCCGGGGGCTAGAGGATCCCCGAAGC
417	*GCTTCGGGGATCCTCTAGCCCCCGGG*CTGCTTTATTGCCGATGATGGAAGCAGCAG
418	AGCCGTTCGCTTCAAGAGCATTCGTCTCGTC
426	GTGTCGGACGCCTTGGGGCTGCTCTG
429	CAGACATTATCAGGCCTTTC
436	TAGCTTCTGCTTGGGCGCAGGTTGAC
437	*GAAGCTATGGGGGTACCTCCCCCGGG*CTTCCTGATTGTCTATTGTCTGTTCTTATC
438	*GCTTCGGGGATCCTCTAGCCCCCGGG*ACTCCAACGACAACAATACATCAG
439	GCAGCAGACGTCCGTAAACCTGATTC
448	CTTTCGAATGACGATCAAACGATTCTG
449	GAGGCTGCTGAGGCGGATACATC

Oligonucleotides used for overlap extension PCR are written in italics

[a] (Honma *et al.*, 2007)

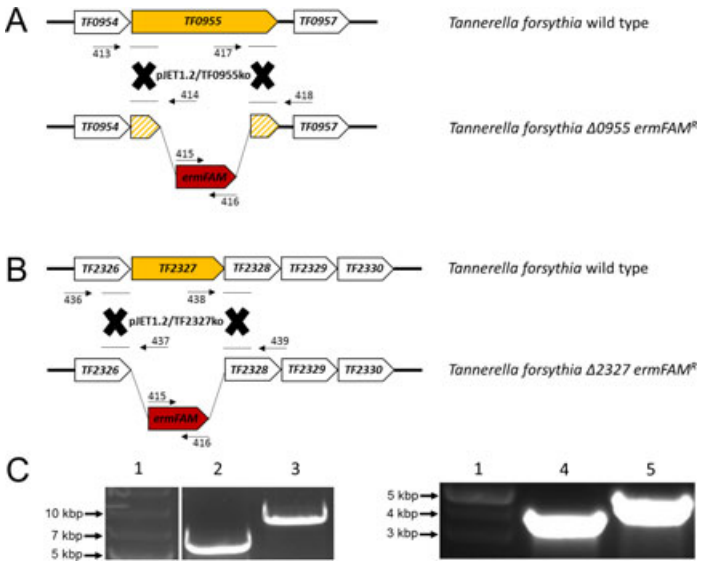
Table 2. Tryptic peptides derived from the C-terminus of TfsA and TfsB as detected in *T. forsythia* wild-type and T9SS mutants. The CTD predicted for TfsA and TfsB according to Veith and coworkers (Veith *et al.*, 2013) is typed in boldface. Peptides derived from the S-layer proteins of *T. forsythia* wild-type are written in red, those derived exclusively from the TfsA and TfsB proteins of the T9SS deficient mutants are written in green.

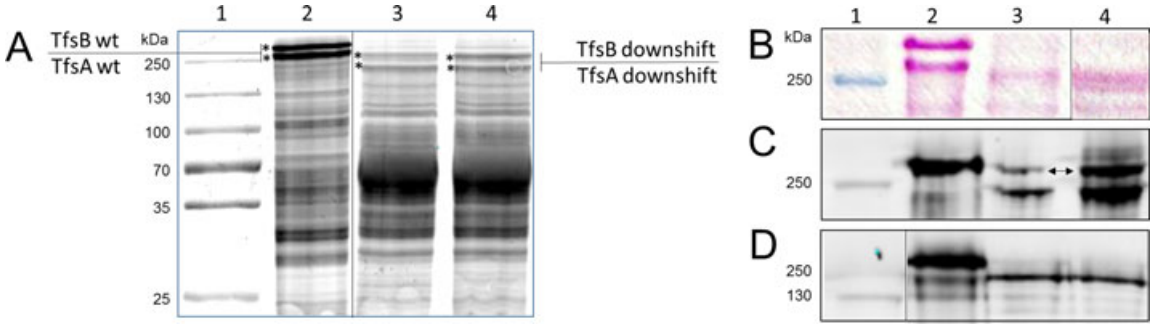
Protein	Unique peptides	log(*e*)[a]	Score[b]	Sequence[c]
TfsA 1083-SRGAYTDWIAEAETWNVKMTKKSPL**ANAEASVTDVKVVAGAGEVMIRNASGK KVTVSNILGQTIAKVVLTSDNQTIAAPQGVVVVAIEGENAVKAVVK**-1179				
TfsA wild-type	9	-2,8		visr^{1084} GAYTDWIAEAETWNVK ^{1099}mtkk
TfsA Δ*2327*	3	-3,1		kmtk^{1103} KSPLANAEASVTDVK ^{1117}vvag
	5	-2,4		mtkk^{1104} SPLANAEASVTDVK ^{1117}vvag
	2	-1,5		asgk^{1134} KVTVSNILGQTIAK ^{1147}vvlt
TfsA Δ*0955*	1	-1,7		kmtk^{1103} KSPLANAEASVTDVK ^{1117}vvag
TfsB 1268-RSDVLQAIADAEIFNVEKTKEDPV**ANEITPAITEVKVVAENGAVTILNAAGKK VVVSNVLGQTLVNTVLTSDRATVAAPQGVVVVVVEGQPAVKAMVK**-1364				
TfsB wild-type	39	-3.2		visr^{1268} SDVLQAIADAEIFNVEK ^{1284}tked
TfsB Δ*2327*	1	-5.1		nvek^{1285} TKEDPVANEITPAITEVK ^{1302}vvae
	13	-3.2		tevk^{1303} VVAENGAVTILNAAGK ^{1318}kvvv
	2		58.2	agkk^{1318} VVVSNVLGQTLVNTVLTSDR ^{1339}atva
	5		58.4	tsdr^{1339} ATVAAPQGVVVVVVEGQPAVK ^{1361}amvk
TfsB Δ*0955*	3	-3.6		nvek^{1285} TKEDPVANEITPAITEVK ^{1302}vvae
	1	-3.9		ektk^{1287} EDPVANEITPAITEVK ^{1302}vvae
	8	-2.6		tevk^{1303} VVAENGAVTILNAAGK ^{1318}kvvv
	1	-3.5		tevk^{1303} VVAENGAVTILNAAGKK ^{1319}vvvs

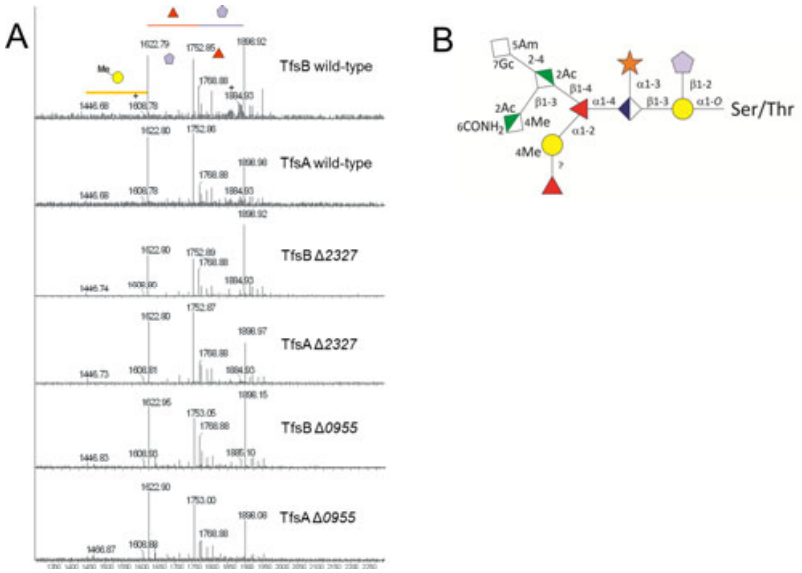
[a] log(e): the base-10 log of the expectation that any particular peptide assignment was made at random (*E*-value). If more than one peptide is found average values are shown

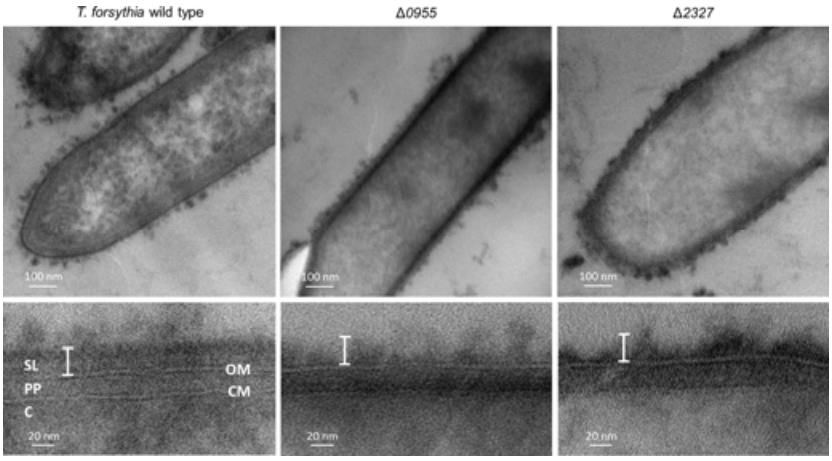
[b] Mascot Score

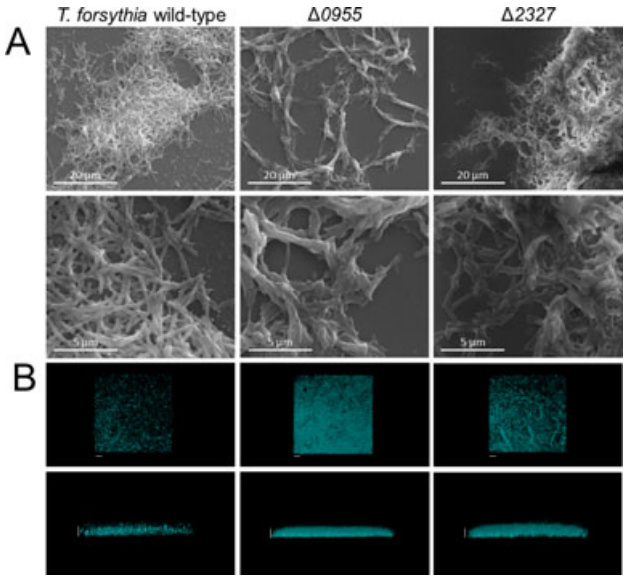
[c] Sequence: the sequence of the assigned peptide sequence. The immediate N- and C-terminal sequences of the identified peptides are also shown.











Appendix VI

Posch, G., Pabst, M., **Neumann, L.**, Coyne, M. J., Altmann, F., Messner, P., Comstock, L. E., Schäffer, C. (2012) "Cross-glycosylation" of proteins in Bacteriodales species. *Glycobiology*, 23(5):568-77. doi: 10.1093/glycob/cws172

Glycobiology vol. 23 no. 5 pp. 568–577, 2013
doi:10.1093/glycob/cws172
Advance Access publication on December 19, 2012

"Cross-glycosylation" of proteins in *Bacteroidales* species

Gerald Posch[2], Martin Pabst[3,5], Laura Neumann[3], Michael J Coyne[4], Friedrich Altmann[3], Paul Messner[2], Laurie E Comstock[4], and Christina Schäffer[1,2]

[2]Department of NanoBiotechnology, NanoGlycobiology Unit, Universität für Bodenkultur Wien, Muthgasse 11, 1190 Vienna, Austria; [3]Department of Chemistry, Universität für Bodenkultur Wien, Muthgasse 18, 1190 Vienna, Austria; and [4]Division of Infectious Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

Received on October 18, 2012; revised on November 29, 2012; accepted on December 11, 2012

While it is now evident that the two *Bacteroidales* species *Bacteroides fragilis* and *Tannerella forsythia* both have general O-glycosylation systems and share a common glycosylation sequon, the ability of these organisms to glycosylate a protein native to the other organism has not yet been demonstrated. Here, we report on the glycosylation of heterologous proteins between these two organisms. Using genetic tools previously developed for *Bacteroides* species, two *B. fragilis* model glycoproteins were expressed in the fastidious anaerobe *T. forsythia* and the attachment of the known *T. forsythia* O-glycan to these proteins was demonstrated by liquid chromatography electrospray ionization tandem mass spectrometry. Likewise, two predominant *T. forsythia* glycoproteins were expressed in *B. fragilis* and glycosylation with the *B. fragilis* O-glycan was confirmed. Purification of these proteins from *B. fragilis* allowed the preliminary characterization of the previously uncharacterized *B. fragilis* protein O-glycan. Based on mass spectrometric data, we show that the *B. fragilis* protein O-glycan is an oligosaccharide composed of nine sugar units. Compositional and structural similarities with the *T. forsythia* O-glycan suggest commonalities in their biosynthesis. These data demonstrate the feasibility of exploiting these organisms for the design of novel glycoproteins.

Keywords: Bacteroides fragilis / glycoengineering / O-glycosylation / *Tannerella forsythia*

Introduction

Many different species of bacteria have been shown to glycosylate proteins (Nothaft and Szymanski 2010). During evolution they have developed complex systems allowing them to systematically transfer glycans onto proteins. General N- (targeting the amide nitrogen of Asn residues) and O- (targeting the hydroxyl oxygen of Ser, Thr or Tyr residues) glycosylation systems have been described in bacteria and knowledge about the molecular mechanisms underlying these processes is continuously expanding. A general *N*-glycosylation system was described for *Campylobacter jejuni* (Szymanski et al. 1999), where a heptasaccharide is transferred on to different proteins containing the conserved (D/E)X_1NX_2(S/T) glycosylation sequon (X_1 and X_2 can be any amino acid except for Pro) (Young et al. 2002; Kowarik et al. 2006). Recently, general O-glycosylation systems have been described in several bacterial species (e.g. *Neisseria gonorrhoeae* (Ku et al. 2009; Vik et al. 2009), *Bacteroides fragilis* (Fletcher et al. 2009), *Tannerella forsythia* (Posch et al. 2011, 2012) and *Acinetobacter baumanii* (Iwashkiw et al. 2012). In *B. fragilis* and *T. forsythia*, O-glycans are specifically attached to extracytoplasmic proteins containing the conserved D(S/T)(A/I/L/V/M/T) motif (Fletcher et al. 2009; Posch et al. 2011). To date, 20 glycoproteins have been experimentally confirmed in *B. fragilis*, but the actual number of proteins that are glycosylated in this organism is predicted to be much higher (Fletcher et al. 2011). Similarly, *T. forsythia* also synthesizes a wide repertoire of glycoproteins (Veith et al. 2009; Posch et al. 2011).

B. fragilis and *T. forsythia* both belong to the order *Bacteroidales*, within the *Bacteroidetes* phylum of bacteria. The *Bacteroidales* contain anaerobic species that associate with mammalian, vertebrate and invertebrate hosts, largely as commensals, symbionts or pathogens (Wexler 2007). Whereas *T. forsythia* inhabits the human oral cavity and is considered a periodontal pathogen (Holt and Ebersole 2005; Pihlstrom et al. 2005), *B. fragilis* colonizes the human intestine where it provides beneficial properties to the host (Mazmanian et al. 2008). Even though *T. forsythia* and *B. fragilis* colonize distinct ecosystems, they are phylogenetically close and thus might have evolved comparable protein O-glycosylation systems, which allow them to furnish proteins with glycan moieties using an identical glycosylation sequon (Fletcher et al. 2009; Posch et al. 2011).

In a previous study, we showed that *T. forsythia* synthesizes a complex oligosaccharide that is attached not only to the two surface (S-) layer proteins TfsA and TfsB of the organism but presumably also to numerous other proteins (Posch et al.

[1]To whom correspondence should be addressed: Tel: +43-1-47654-2203; Fax: +43-1-4789112; e-mail: christina.schaeffer@boku.ac.at
[5]Present address: Department of Chemistry and Applied Biosciences, ETH Zurich, Switzerland.

2011). The *T. forsythia* O-glycan is a complex, structurally unique oligosaccharide that is most likely involved in biofilm formation, as an isogenic mutant lacking a terminal portion of the glycan shows significantly altered biofilm formation compared with the wild-type strain (Honma et al. 2007).

The discovery of broad-spectrum bacterial glycosylation systems along with conserved glycosylation sequons prompted efforts to functionally exploit them for glycoengineering purposes. The use of bacteria for glycoengineering was successfully demonstrated when the *C. jejuni N*-glycosylation pathway was transferred into *Escherichia coli* (Wacker et al. 2002). Since then, several studies have shown that a well-directed transfer of glycans onto proteins in *E. coli* is feasible, which has also been used for the production of homogenous glycoproteins with eukaryotic *N*-glycans (Feldman et al. 2005; Schwarz et al. 2010; Lizak et al. 2011). Only recently, a bottom-up synthetic pathway in *E. coli* was engineered, allowing the production of eukaryotic trimannosyl chitobiose glycans and their transfer to specific residues in target proteins (Valderrama-Rincon et al. 2012).

The presence of a common glycosylation sequon in these *Bacteroidales* species raises the possibility of precisely targeting the addition of glycans onto proteins. In *B. fragilis*, it was previously demonstrated that introduction of a glycosylation sequon into a naturally unglycosylated *B. fragilis* protein brings about site-specific glycosylation at the engineered sites (Fletcher et al. 2011).

Here, we report on the transfer of the *B. fragilis* O-glycan onto heterologously expressed *T. forsythia* proteins and *vice versa*. Additionally, the first mass spectrometric characterization of the previously undescribed *B. fragilis* protein O-glycan is presented. We show that "cross-glycosylation" of proteins in *Bacteroidales* by utilizing the conserved D(S/T)(A/I/L/V/M/T) motif is feasible, allowing the design of novel glycoproteins.

Results

Expression of His-tagged (glycosylated) proteins in T. forsythia *and* B. fragilis

Expression of recombinant proteins in *B. fragilis* is well demonstrated (Smith et al. 1992; Bayley et al. 2000) using *E. coli-Bacteroides* shuttle plasmids, modification of which has allowed the production of C-terminally His-tagged fusion proteins (Fletcher et al. 2009).

Heterologous protein expression in *T. forsythia* has not yet been described, which may be attributed both to the inherently demanding nature of the organism's growth and to the lack of suitable genetic tools.

Considering that *T. forsythia* and *B. fragilis* are closely related organisms, we set out to determine whether the existing *B. fragilis* genetic tools could be used in *T. forsythia*. As the recipient strain, we chose to use an S-layer-deficient *T. forsythia* mutant strain, to allow better conjugation efficiency compared with the wild-type strain. Two *B. fragilis* recombinant His-tagged protein encoding genes were chosen for transfer (BF2494, GI: 60681974; BF3567, GI 60683022) as both are glycosylated in their parent strain at the D-(S/T)-(A/I/L/V/M/T) motif (Fletcher et al. 2009). Our approach thus served a double purpose: First, to determine whether expression vectors and procedures developed for *B. fragilis* would function in *T. forsythia* and, secondly, to ascertain whether *T. forsythia* would recognize the provided glycosylation sequons and produce heterologously glycosylated proteins. Likewise, we wanted to investigate whether *B. fragilis* would attach glycans to the *T. forsythia* S-layer proteins TfsA and TfsB, thus allowing "cross-glycosylation" of proteins within these two species based on the shared glycosylation target motif.

Western blot analysis of purified His-tagged proteins (Figure 1A, lanes i and ii) showed that TfsA and TfsB can be readily expressed in *B. fragilis*. Noticeably, the masses of both proteins on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS–PAGE) gels were higher than their predicted molecular masses. For TfsA-His (calculated molecular weight (Mw), 133.3 kDa with the signal peptide removed) an upshift to ~170 kDa was observed, whereas TfsB-His (calculated Mw, 150.8 kDa with the signal peptide removed) was shifted to ~200 kDa. These data indicated that both proteins are likely glycosylated in *B. fragilis*, which was further confirmed by glycostaining of the gel (Figure 1B, lanes i and ii). Additionally, comparing the migration behavior of both TfsA and TfsB recombinantly produced in *B. fragilis* with their native counterparts (Figure 1C) demonstrates that the glycosylation is different in these two species, as natively produced TfsA and TfsB glycoproteins show higher apparent molecular masses (~230 and ~270 kDa, respectively).

As shown in Figure 1A (lanes iii and iv), *T. forsythia* is also capable of expressing C-terminally His-tagged (glyco) proteins from the *B. fragilis* derived vector using the native *B. fragilis* promoter, demonstrating the feasibility of using existing *Bacteroides* genetic tools in this organism. His-tagged BF2494 and BF3567 were purified from *T. forsythia* and, interestingly, both proteins also exhibited a migration behavior different from that expected from their calculated molecular masses (BF2494-His calculated Mw, 45.9 kDa without the signal peptide, observed Mw ~55 kDa; BF3567-His calculated Mw, 66.8 kDa without the signal peptide, observed Mw >75 kDa). Again, this suggested that both proteins are posttranslationally modified in *T. forsythia*. In a western blot, probing with an anti-BF2494-His antibody, an altered migration pattern of BF2494-His expressed in *B. fragilis* and *T. forsythia* can be deduced (Figure 1D, lanes i and ii), indirectly indicating altered glycosylation. Noteworthy, BF3567-His expressed in *T. forsythia* did not show one uniform band upon probing with anti-His antibody (Figure 1A, lane iii) but several bands >75 kDa gave a positive signal, potentially indicative of a consecutive transfer of glycans onto this protein or glycoprotein degradation.

B. fragilis *O-glycan analysis*

Protein bands originating from a crude extract of a *B. fragilis* culture were excised from SDS–PAGE gels (Figure 2A, inset) and served as the initial source of *B. fragilis* glycoproteins for the characterization of its yet undescribed O-glycan. In-gel β-elimination of glycans from these protein bands followed by porous graphitized carbon electrospray ionization mass spectrometry (PGC-ESI-MS) allowed the detection of a 1571.56 (molecular ion (mi), $[M+H]^+$)–Da glycan structure

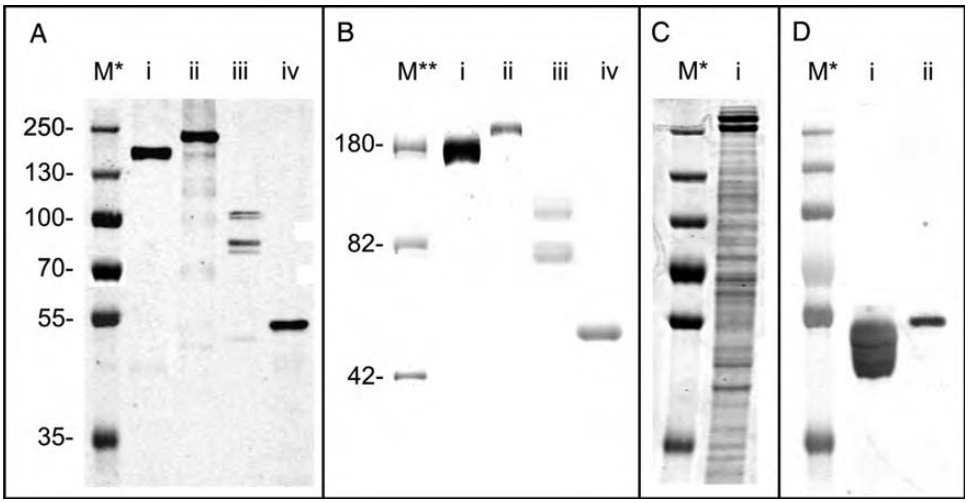


Fig. 1. (**A**) Western blot analysis of (i) TfsA-His, (ii) TfsB-His, (iii) BF3567-His and (iv) BF2494-His purified from *B. fragilis* (i and ii) and *T. forsythia* (iii, iv), respectively. The observed Mw of all four proteins is larger than the one calculated, suggesting a posttranslational modification of these proteins. (**B**) Pro-Q Emerald-stained gel showing that both (i) TfsA-His and (ii) TfsB-His purified from *B. fragilis* as well as BF3567-His and BF2494-His purified from *T. forsythia* are glycosylated. (**C**) Coomassie Blue-stained SDS–PAGE gel of a *T. forsythia* crude extract (i), showing that native (glycosylated) TfsA (lower of the two broad bands migrating around 250 kDa) and TfsB migrate at higher apparent Mw than their heterologously expressed versions shown in A (lanes i and ii), indicating a difference in posttranslational modification. (**D**) Western blot probing BF2494-His expressed in (i) *B. fragilis* wild-type (Fletcher et al. 2009) as well as in (ii) *T. forsythia* with an anti-BF2494 antibody. BF2494-His from *T. forsythia* shows altered migration behavior compared with wild-type BF2494-His (which additionally shows a banding pattern indicating a consecutive transfer of monosaccharides), indirectly indicating altered glycosylation. M*, PageRuler Plus prestained protein ladder (Thermo Fisher Scientific, Vienna, Austria); M**, CandyCane glycoprotein molecular weight ladder (Invitrogen).

(Supplementary data, Figure S1) showing in-source fragments corresponding to one methylated deoxyhexose and one potentially acetylated glycan constituent (Figure 2A). Collision-induced dissociation (CID) fragmentation analysis of the 1571.56-Da glycan revealed it to be a hetero-oligomer consisting of nine sugar residues (Figure 3). Mass increments for three hexoses (one reduced), two hexuronic acids, two *N*-acetyl hexosamines (one uncertain), one methylated deoxyhexose, as well as one deoxyhexose were identified (Supplementary data, Table S1). The terminating 203-Da residue, which is suspected to be one of the *N*-acetyl hexosamines, was found to be highly prone to deacetylation (loss of 42 Da during the strong basic reduction process). Such a propensity for deacetylation would, in our experience, indicate O-acetylation more strongly than *N*-acetylation. Thus, although the unit mass is appropriate for an *N*-acetyl hexosamine, the nature of this glycan constituent could not be fully resolved in this study.

The fragment spectra gave sufficient information to putatively assign the sequence and branching pattern of the *B. fragilis* glycan (Figure 3, Supplementary data, Figure S2). Interestingly, β-elimination and subsequent glycan analysis of TfsA-His and TfsB-His glycoprotein bands purified from *B. fragilis* gave identical mass spectra, indicating the successful O-glycan transfer onto those proteins (Figure 2B and C).

"Cross-glycosylation" of B. fragilis *and* T. forsythia *proteins*

To analyze for heterologous glycosylation events, we excised Coomassie Blue-stained bands of our selected target proteins from SDS–PAGE gels and performed in-gel tryptic digests followed by reverse phase ESI-MS/MS (glyco)peptide mapping. Modified peptides of TfsA and TfsB synthesized in *B. fragilis* were identified by the observed masses not matching those of the predicted tryptic peptide masses. We subjected one of the putative glycopeptides derived from the TfsA protein to CID MS/MS and observed a series of singly charged fragment ions within an *m/z* range between 2341.7 and 791 (Figure 4A). Detailed analysis of the fragmentation pattern revealed the presence of a peptide component corresponding to the sequence NQTDSAR, which is the predicted product of tryptic cleavage at R938 and R945 within the TfsA protein, spanning the potential glycosylation site S943 within a DSA glycosylation sequon. The modification of this peptide was determined to perfectly match all constituents of the *B. fragilis* O-glycan moiety described above.

Analysis of tryptic peptides of BF2494 and BF3567 from *T. forsythia* Δ*tfsAB* yielded a very similar result. In addition to the predicted tryptic peptides, we also found potentially modified doubly and triply charged cleavage products. Again, we selected one of those cleavage products derived from BF3567 for CID MS/MS and observed a series of fragment ions beginning at *m/z* 1496.6 $(M+2H)^{2+}$ and ending at *m/z* 1258.7 $(M+H)^{+}$ (Figure 4B). This perfectly fits peptide LGDISLDTVAVR, produced upon tryptic cleavage at K199 and R211 of BF3567, substituted with the *T. forsythia* O-glycan at T207 within the DTV glycosylation motif. By analyzing the fragment spectrum of the O-glycan, we found all known constituents including their modifications with the exception of the O-methyl group of the *N*-acetyl mannosaminuronamide residue. For yet unknown reasons, this lack of methylation is commonly observed on O-glycans of glycoproteins isolated from *T. forsythia* Δ*tfsAB* (Gerald Posch, unpublished data), and thus it may be specific to this mutant strain.

In Supplementary data, Table S2, all glycopeptides that were identified as being heterologously glycosylated are listed.

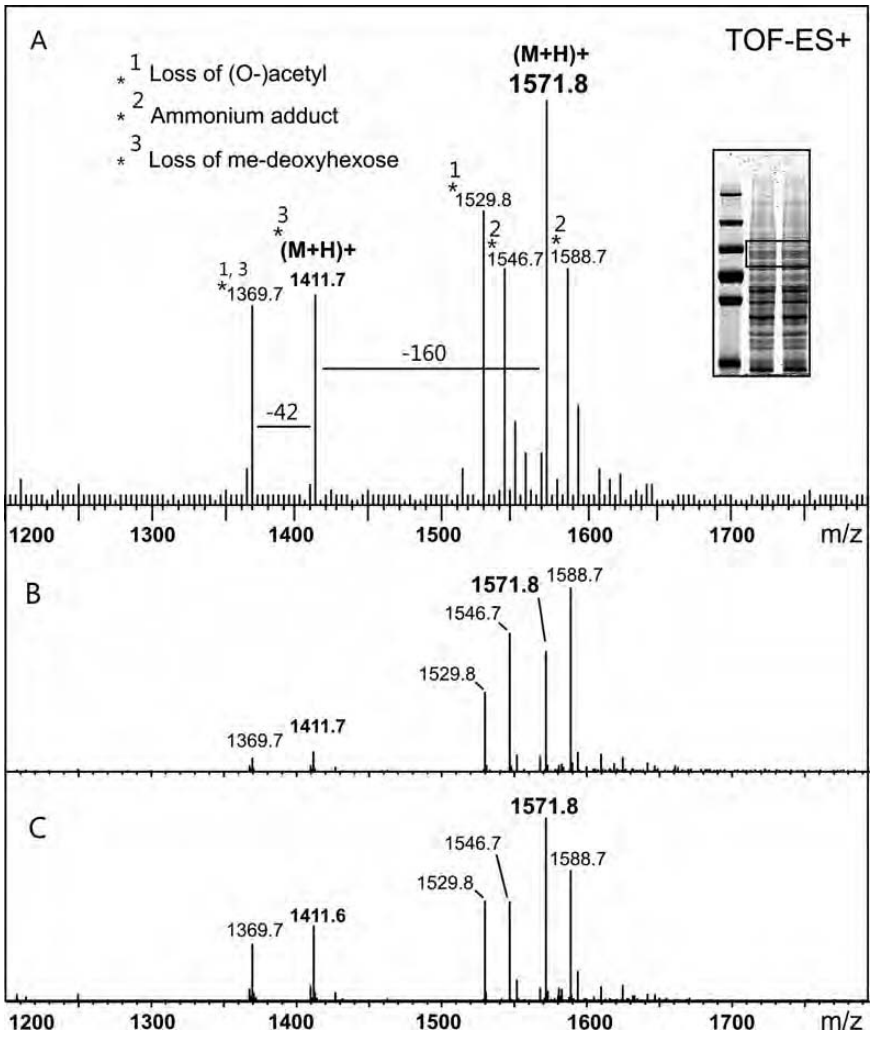


Fig. 2. (**A**) ESI-TOF-MS spectrum of the *B. fragilis* glycan as detected after in-gel reductive β-elimination of Coomassie Blue-stained SDS–PAGE bands (see inset, boxed area) originating from a *B. fragilis* crude protein extract. A dominant glycan structure of 1571.8 Da $(M+H)^+$ was observed, along with an in-source fragment corresponding to one methylated deoxyhexose (−160 Da). Also, one acetyl group was found to be partially lost during β-elimination (−42 Da). (**B**) Summarized LC-ESI-MS spectrum of the O-glycan as isolated from the *T. forsythia* S-layer protein TfsA expressed in *B. fragilis* (see Figure 1A, i). (**C**) Summarized LC-ESI-MS spectrum of the O-glycan as isolated from the *T. forsythia* S-layer protein TfsB expressed in *B. fragilis* (see Figure 1A, ii). Both O-glycan structures observed in (**B**) and (**C**) correspond to the *B. fragilis* O-glycan as isolated in (**A**), confirming the attachment of the *B. fragilis* glycan to heterologously expressed *T. forsythia* proteins. Note: Glycan screening was done with the Micromass Global Ultima instrument (Waters), showing a slight mass deviation compared with high-resolution measurements (Supplementary data, Figure S1).

In total, three glycopeptides of TfsA and TfsB were identified as being modified with the *B. fragilis* glycan. One representative "cross-glycosylated" peptide of BF2494 and two representative glycopeptides of BF3567 modified with the *T. forsythia* glycan were identified, demonstrating that the O-glycans of *T. forsythia/B. fragilis* can be added to non-native proteins. In general, the glycosylation of heterologous proteins appears to occur with high efficiency, as all of the peptides containing a glycosylation motif were either observed to be glycosylated or were not detected in their unglycosylated form.

Discussion

In recent years, bacterial protein O-glycosylation has rapidly expanded from "curiosity" to "ubiquity" (Messner 2009) with the discoveries of broad-spectrum glycosylation systems. Recent scientific efforts are directed at exploiting these systems to create novel biosynthetic pathways resulting in tailor-made glycans that can be specifically attached to target proteins (Steiner et al. 2007; Lizak et al. 2011; Valderrama-Rincon et al. 2012). Considering the manifold properties glycans provide to proteins, including stabilizing functions (Krapp et al. 2003), improved thermal stability (Mimura et al. 2000) and protection from proteases (Langsford et al. 1987), this approach seems to be promising for the future generation of glycoproteins with improved characteristics.

The present study focused on the feasibility of heterologously glycosylating proteins in the two *Bacteroidales* species *B. fragilis* and *T. forsythia*. Both organisms are known to

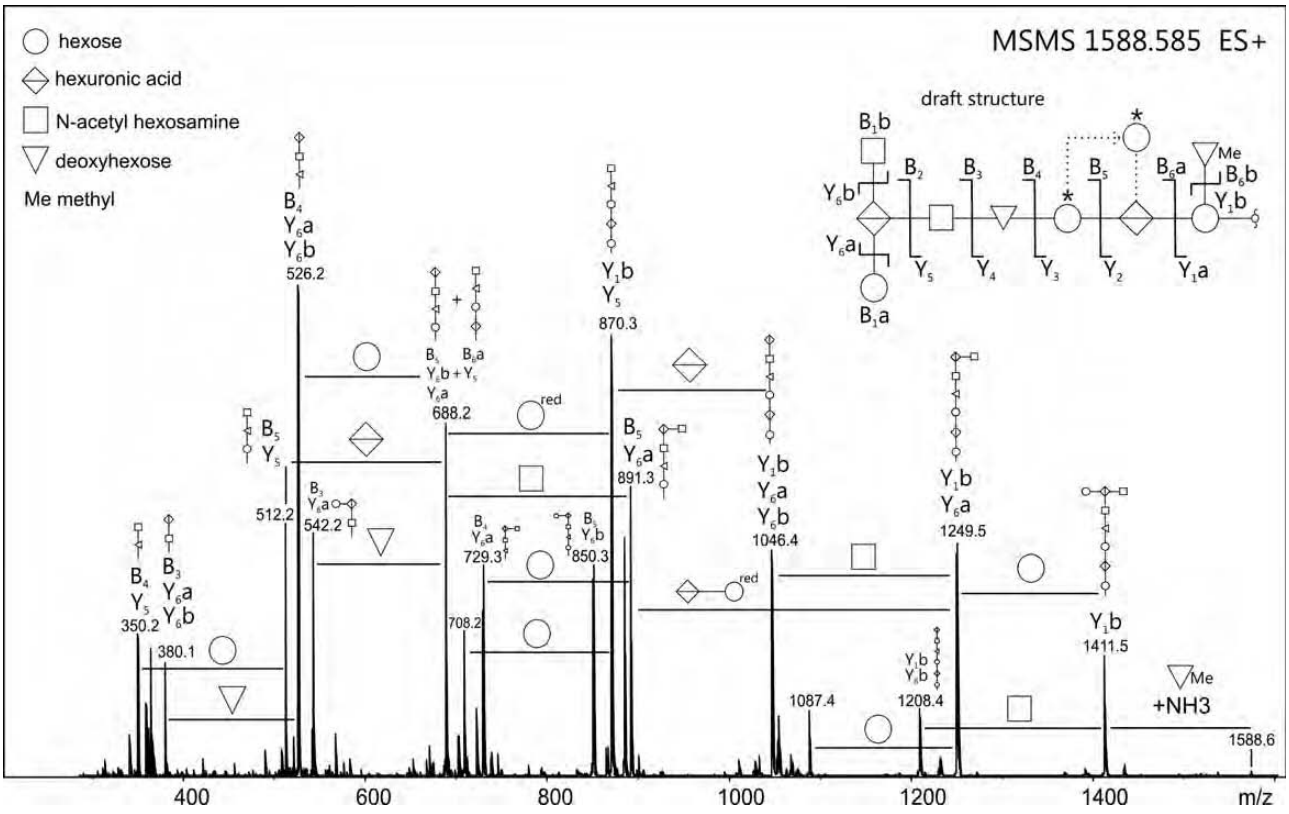


Fig. 3. ESI-TOF-MS/MS spectrum of the borohydride-reduced *B. fragilis* O-glycan, as observed from protein bands excised from SDS–PAGE gels. The fragmentation pattern revealed the presence of a nine-sugar oligosaccharide comprising three hexoses, two hexuronic acids, two *N*-acetyl hexosamines as well as two deoxyhexoses (modified or free). Based on the observed fragment ions, the *B. fragilis* O-glycan structure was drafted (see inset top right). Fragments were labeled using the nomenclature as described (Domon and Costello 1988). The terminal *N*-acetyl hexosamine residue is found with considerable uncertainty, both by its accurate mass (see Supplementary data, Table S1) and the fact that it is prone to loss of a 42 Da fragment (indicating O-acetylation rather than *N*-acetylation) during β-elimination. In addition, the existence of a glycan isoform with the middle of the three hexose residues (marked with asterisk) branching from the hexuronic acid residue cannot be completely ruled out. Potential rearrangement artifacts upon CID were excluded upon measurement of sodium adducts of the *B. fragilis* O-glycan (Wuhrer et al. 2011; data not shown). Note: Line positions between residues of the draft *B. fragilis* O-glycan structure do not represent actual linkage types.

possess general O-glycosylation systems sharing a conserved amino acid motif that is targeted for glycosylation. We first partially characterized the native *B. fragilis* O-glycan in order to determine whether the same glycan was added to *T. forsythia* proteins expressed in *B. fragilis*. Mass spectrometric analysis of native *B. fragilis* glycoproteins subjected to β-elimination revealed the glycan to be an oligosaccharide consisting of nine sugar residues. Notably, the proposed structure based on MS data resembles that of the *T. forsythia* O-glycan in several aspects. In both glycans, attachment to the protein occurs via a hexose residue which is succeeded by a hexuronic acid. Additionally, a nonpolar constituent branches from the first sugar of the *B. fragilis* O-glycan. Also, both glycans contain a deoxyhexose residue in their linear structure. As in the *T. forsythia* glycan, the deoxyhexose of the *B. fragilis* glycan is most likely a fucose (Fuc), as glycoproteins of *B. fragilis* are readily detected with *Aleuria aurantia* lectin and depend on GDP-Fuc biosynthesis genes (Fletcher et al. 2009).

Knowing the size and rough composition of the *B. fragilis* glycan, we sought to determine whether transfer of the respective glycans on to non-native proteins occurs in *B. fragilis* and *T. forsythia*. We chose to analyze the well-characterized S-layer proteins TfsA and TfsB from *T. forsythia* (Lee et al. 2006; Sekot et al. 2012) for heterologous glycosylation in *B. fragilis*. In addition, two model glycoproteins of *B. fragilis*, BF2494 and BF3567, both of unknown function (Fletcher et al. 2009), were selected for heterologous expression in *T. forsythia*. Successful transfer of the O-glycans was expected to result in novel glycoproteins with the respective O-glycans linked via S/T within the *Bacteroides* glycosylation sequon.

Our analyses of the heterologously expressed proteins revealed electrophoretic migrations different from those predicted according to their molecular masses. This observation strongly supported our concept of "cross-glycosylation" of proteins within the two species. This was confirmed by analyzing glycopeptides of the respective proteins, ultimately showing that the O-glycans had been properly transferred. Furthermore, glycan transfer appeared to be highly efficient, as none of the peptides spanning a putative glycosylation sites was observed as unmodified.

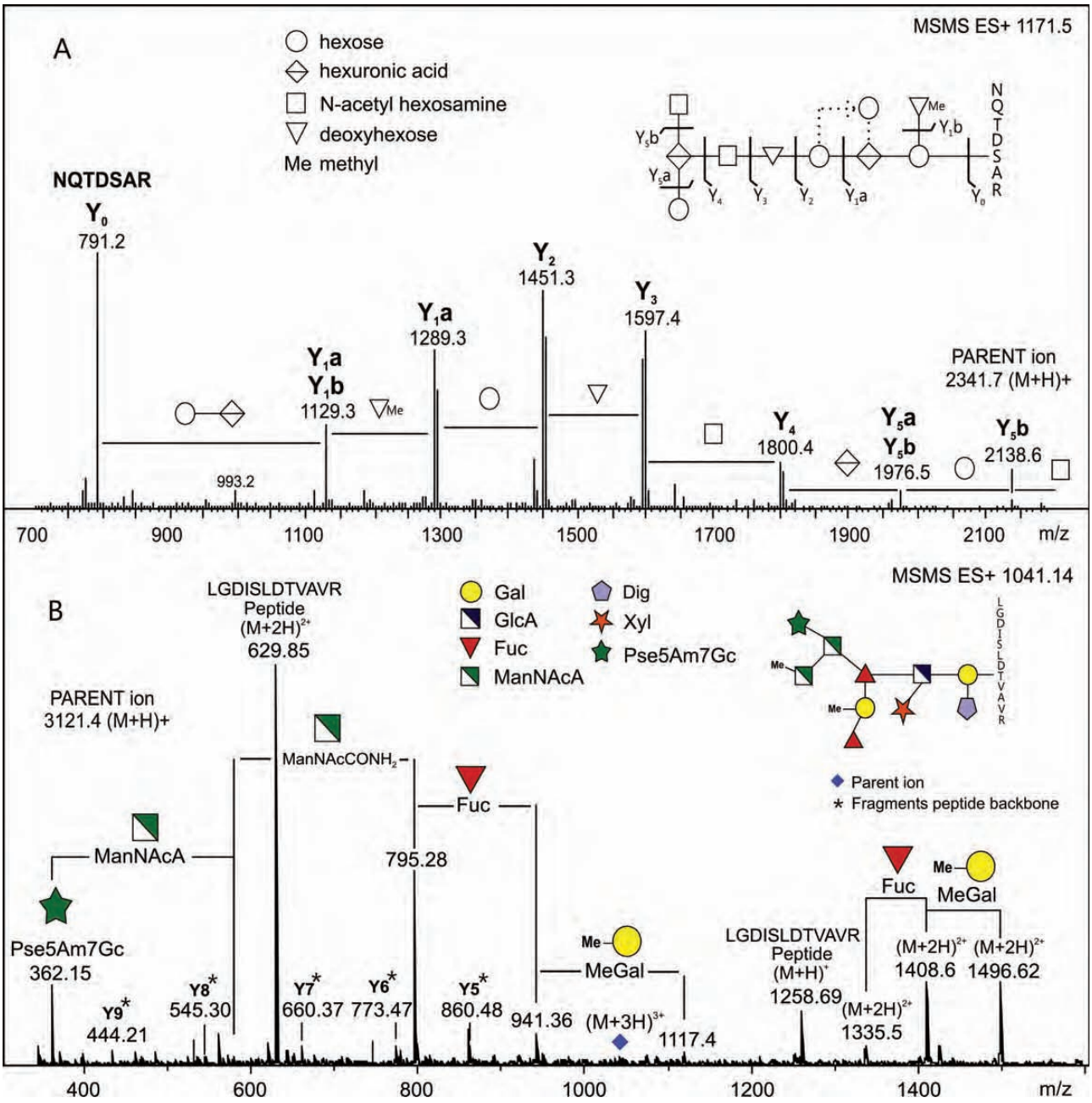


Fig. 4. (**A**) Tandem ESI-MS of a glycopeptide derived from the *T. forsythia* S-layer protein TfsA expressed in *B. fragilis* confirming the heterologous attachment of the *B. fragilis* O-glycan to the protein (attachment to S943 within the glycosylation sequon DSA is shown). The fragmentation pattern of the O-glycan and the deduced glycan structure correlate with data presented in Figure 3. Attachment to the protein backbone occurs via hexose-hexuronic acid with one methylated deoxyhexose branching from the hexose at the reducing end. (**B**) Tandem ESI-MS of a glycopeptide derived from the *B. fragilis* protein BF3567 expressed in *T. forsythia* showing the heterologous attachment of the *T. forsythia* O-glycan to the protein (binding to T207 within the glycosylation sequon DTV is shown). Fragment ions marked with an asterisk correspond to peptide backbone fragments. Note: Line positions between residues of the *B. fragilis* glycan (A, inset top right) do not represent actual linkage types.

Protein glycosylation in *T. forsythia* and *B. fragilis* is ubiquitous and seems to be an inherent feature of these organisms. Their well-established systems allow decoration of supposedly any extracytoplasmic protein bearing a glycosylation sequon (Fletcher et al. 2011). The reasons for modification of such an extensive number of proteins are still unclear. In *T. forsythia*, the protein O-glycan may be involved in the bacterium-host cross-talk, mediate cell adhesion (Sakakibara et al. 2007) and influence the biofilm formation capability of the organism (Honma et al. 2007). Also, it was shown that the glycosylated S-layer of *T. forsythia* displayed on wild-type cells delays recognition of the immune system in a macrophage cell culture model compared with the S-layer deficient *T. forsythia* Δ*tfsAB* strain (Sekot et al. 2011). Recently, the suppression of

Table I. Bacterial strains and plasmids used in this study

Strain or plasmid	Description	Reference or source
T. forsythia ATCC 43037	Wild-type strain	ATCC
T. forsythia Δ*tfsAB*	Double knock-out mutant, devoid of the S-layer proteins	Sakakibara et al. (2007)
B. fragilis NCTC9343	Wild-type strain	NCTC
E. coli DH5α	Cloning strain	Invitrogen
RK231	broad-host-range mobilizing IncP plasmid, RK2 derivative; Km®	Guiney et al. (1984)
pJET 1.2	Cloning vector, Amp®	Thermo Fisher Scientific
pCMF118	*E. coli-Bacteroides* shuttle vector, pFD340 derivative; Amp®, Em®	Coyne et al. (manuscript in preparation)
pMJC95	C-His_{10} tagged *tfsA* from *T. forsythia* cloned into pCMF118; Amp®, Em®	This study
pMJC94	C-His_{10} tagged *tfsB* from *T. forsythia* cloned into pCMF118; Amp®, Em®	This study
pGP21	C-His_{10} tagged BF2494 from *B. fragilis* cloned into pCMF118; Amp®, Em®	This study
pGP22	C-His_{10} tagged BF3567 from *B. fragilis* cloned into pCMF118; Amp®, Em®	This study

T-helper 17 responses in dendritic cells as well as increased periodontal bone loss in mice could be specifically attributed to the terminal 5-acetimidol-7-*N*-glycolylpseudaminic acid (Pse5Am7Gc)-*N*-acetylmannosamiuronamide (ManNAcCONH$_2$)-*N*-acetylmannosamiuronic acid (ManNAcA) trisaccharide branch of the *T. forsythia* O-glycan (Settem et al. 2012).

As for *B. fragilis*, protein glycosylation has been shown to be essential for normal in vitro growth and for colonization of the mammalian intestine (Fletcher et al. 2009). Additionally, surface glycoproteins isolated from an outer membrane protein preparation are proposed to confer interaction with the extracellular matrix component laminin-1 (de O. Ferreira et al. 2006). However, the identities of the glycoproteins involved in this proposed interaction are not described.

Considering the vast amount of glycoproteins synthesized by *B. fragilis*—and most likely also by *T. forsythia*—it is likely that O-glycosylation of proteins has a more general function, for instance in protein stability. As secreted proteins of *B. fragilis* do not form intramolecular disulfide bonds (Dutton et al. 2008), protein glycosylation may serve a compensatory stabilizing role. Analyzing the influence of (heterologous) glycosylation on protein stability in *T. forsythia* and *B. fragilis* will contribute to a better understanding of protein glycosylation in general, as well as trigger efforts to specifically improve protein stability through glycoengineering.

Material and methods

Bacterial strains, plasmids and growth conditions

The bacterial strains and plasmids used in this study are listed in Table I. *E. coli* strains were grown at 37°C overnight in Luria–Bertani (LB) broth supplemented with ampicillin (Amp, 100 μg mL^{-1}), kanamycin (Km, 50 μg mL^{-1}) or both.

B. fragilis and *T. forsythia* strains were grown anaerobically in brain heart infusion (BHI) broth or on agar plates (1% w/v) supplemented with yeast extract (5 g L^{-1}), cysteine (1 g L^{-1}), hemin (50 μg mL^{-1}) and menadione (10 μg mL^{-1}). The *T. forsythia* medium was additionally supplemented with *N*-acetylmuramic acid (10 μg mL^{-1}). Gentamicin (Gm, 200 μg mL^{-1}) and erythromicin (Em, 5 μg mL^{-1}) were added to the media when appropriate.

Table II. Oligonucleotide primers used in this study

Purpose	Sequence[a]
Amplification of *T. forsythia* genes for expression with His-tags in *B. fragilis*	
tfsA-for	AAATAGATCTGCGGTTTATAAGAGGAAGAAAATAAA
tfsA-rev	CTTAAGATCTTTTTACACAGCTTTCACTGCATTC
tfsB-for	CCTCGGATCCATCTCTTGCCTGCCACTCC
tfsB-rev	CGACGGATCCCTTCACCATCGCTTTTACAGC
Amplification of *B. fragilis* genes for expression with His-tags in *T. forsythia*	
BF2494-for	ATCAGGATCCACAATCATGAAAAGAGTATTATTTTC
BF2494-rev	ATCAGGATCCCATCATTTTCTCGATTTCTTCGAATTC
BF3567-for	ATCAGGATCCTACTAACTAAACGTGATTAATTTATG
BF3567-rev	ATCAGGATCCACGGGTTACTTCCAAATACTTCACC

[a]Sequences are given 5′–3′ with restriction sites underlined.

Cloning and transformation of constructs

DNA of *T. forsythia* and *B. fragilis*, respectively, was prepared as follows. 1 mL of stationary bacterial culture was harvested (6000 × *g*, 2 min) and the supernatant was discarded. The cells were resuspended in 50 μL of sterile distilled water and boiled for 5 min. Cell debris was removed by centrifugation (20,000 × *g*, 2 min) and the supernatant containing chromosomal DNA was used as template for all polymerase chain reactions (PCRs). All oligonucleotides used are listed in Table II. *TfsA* was PCR-amplified, digested with BglII and ligated into the unique BamHI site of vector pCMF118 (Coyne et al., manuscript in preparation), creating pMJC95. Similarly, PCR-amplified *tfsB* was digested with BamHI and ligated into BamHI digested pCMF118, creating pMJC94.

BF2494 and BF3567 PCR-amplified products were blunt-end cloned into pJET1.2 (Thermo Fisher Scientific, Vienna, Austria) and the inserts from positive clones were cut with BamHI and cloned into the unique BamHI site of pCMF118 creating pGP21 and pGP22, respectively. Transformants with the correct insert orientation were selected by PCR. All constructs (pMJC94, pMJC95, pGP21 and pGP22) were also confirmed by sequencing.

Plasmids were transferred from *E. coli* to *T. forsythia* Δ*tfsAB* and *B. fragilis* by conjugative transfer. First, 200 μL of *E. coli* RK231 overnight culture was mixed with 200 μL of *E. coli* DH5α cells containing the respective plasmid constructs. Cells were centrifuged (6000 × *g*, 2 min) and plated on LB agar without antibiotics. Following overnight

incubation at 37°C, the growth was struck to LB agar plates supplemented with Amp and Km to select for clones containing both RK231 and the respective expression plasmids. Subsequently, positive clones were used in a second conjugation experiment to transform *T. forsythia* Δ*tfsAB* or *B. fragilis* by combining 3 mL of an overnight culture of the respective clones with 50 mL of *T. forsythia* Δ*tfsAB* or *B. fragilis* culture ($OD_{600} \sim 0.3–0.6$) and collecting the cells by centrifugation ($6000 \times g$, 10 min). The cell pellet was resuspended in a small volume of medium and plated to BHI agar plates without antibiotics. Following aerobic incubation (37°C, overnight), the growth was struck to Gm/Em-containing BHI agar plates and incubated anaerobically at 37°C for 2 days (*B. fragilis*) or up to 14 days (*T. forsythia*). Em-resistant transconjugants were confirmed by PCR. The resulting recombinant proteins were modified by the addition of glutamine-serine-10x histidine (GSH_{10}) at the C-terminus, except TfsA, which was modified with arginine-serine-10x histidine (RSH_{10}).

Purification of His-tagged proteins from T. forsythia *and* B. fragilis

Immobilized metal affinity chromatography (IMAC) was performed to purify His-tagged proteins. Briefly, cultures (2 L) of *T. forsythia* Δ*tfsAB* and *B. fragilis* harboring plasmids encoding His-tagged proteins were grown to stationary phase, harvested ($6000 \times g$, 4°C, 15 min) and washed with buffer A (20 mM NaH_2PO_4, 20 mM imidazole; adjusted to pH 7.5 with 4 M NaOH). The bacterial pellets were resuspended at a ratio of 1:5 in buffer A (w/v; if necessary, 4 M urea was added to allow for denaturing purification). Following sonication (3 cycles of 30 pulses, each, with 1 min of cooling between the cycles), the cellular debris was removed by ultracentrifugation ($50{,}000 \times g$, 4°C, 30 min) and the supernatant was loaded on to a HiTrap HP column (GE Healthcare, Vienna, Austria V = 1 mL) connected to a BioLogic DuoFlow FPLC system (BioRad, Vienna, Austria). Bound proteins were eluted in a linear gradient of 0–0.5 M imidazole in buffer A within 10 column volumes. Fractions containing the desired proteins were pooled, concentrated via Amicon Ultra-15 centrifugal filter units (30 kDa cutoff; Millipore, Vienna, Austria) and subjected to SDS–PAGE and western blot analysis.

SDS–PAGE, Coomassie Blue staining and western blot analysis

Purified His-tagged proteins or crude cell extracts were run on 10% SDS–PAGE gels according to standard protocols (Laemmli 1970). Proteins were directly visualized with Coomassie Blue staining or transferred to polyvinylidene difluoride (PVDF) membranes for western immunoblotting using an anti-His antibody (produced in mice; Invitrogen, Vienna, Austria) in combination with an anti-mouse secondary antibody labeled with IRDye 800CW (LI-COR, Bad Homburg, Germany). Membranes were visualized using the Odyssey Infrared Imaging system (LI-COR Biosciences) at 785 nm. Glycostaining of SDS–PAGE gels was performed using the ProQ-Emerald fluorescent stain (Invitrogen).

Glycopeptide preparation

For preparation of glycopeptides, gel slices containing the protein bands were excised from Coomassie Blue-stained gels, chopped into small pieces and destained (2 cycles of 50 and 100% acetonitrile, followed by reswelling of the gel pieces in 100 mM ammonium bicarbonate, with 10 min incubation time, each). S-carbamidomethylation, trypsin digestion and extraction of (glyco-)peptides were performed by routine methods (Stadlmann et al. 2008). Briefly, cysteine bonds were reduced by treatment with 10 mM dithiothreitol (DTT) in 100 mM ammonium bicarbonate for 1 h at 56°C. Cysteine residues were S-alkylated with 55 mM iodoacetamide in ammonium bicarbonate at room temperature for 1 h. Following subsequent washing, (glyco-)proteins were digested with trypsin (Roche, Vienna, Austria, 50 ng μL^{-1} in 25 mM ammonium carbonate) overnight at 37°C and the resulting peptides were extracted with alternating washes of water, bicarbonate buffer and acetonitrile (100%) and lyophilized prior to analysis.

In-gel reductive β-elimination

O-Glycan release for further LC-ESI-MS/MS analyses of the *B. fragilis* glycoprotein glycan was performed by in-gel reductive β-elimination of Coomassie Blue-stained protein bands. Briefly, bands of purified His-tagged proteins as well as bands of miscellaneous proteins originating from a crude *B. fragilis* protein extract were excised from gels, transferred to plastic reaction tubes, covered with 1 M $NaBH_4$ in 500 mM NaOH and incubated at 50°C overnight. Excess salt was removed using a 10 mg HyperSep Hypercarb solid phase extraction (SPE) cartridge (Thermo Fisher Scientific) according to published protocols (Packer et al. 1998; Pabst and Altmann 2008).

LC ESI-MS/MS (IonTrap and Q-TOF)

Borohydride-reduced O-glycans were analyzed by PGC (Hypercarb, 0.32 × 150 mm, particle size 5 μm)-ESI-MS/MS as recently described (Pabst et al. 2007). Ammonium formate buffer (0.3% formic acid, adjusted to pH 3.0 with ammonia solution) was used as buffer A, and a 0–35% acetonitrile gradient was performed within 35 min using a Dionex Ultimate 3000 (cap flow, 8 μL min^{-1}). Glycan mass screening was performed using a Global Ultima Q-TOF from Micromass (Waters, Eschborn, Germany). Data were evaluated using the MassLynx 4.0 software. High-resolution mass spectrometry experiments with direct infusion of purified glycans were performed using a Bruker Maxis 4G Q-TOF.

Glycoproteomics analyses were performed by reversed phase LC coupled to ESI-MS/MS either on a Bruker IonTrap AmaZon speed ETD or on the high-resolution Maxis 4G Q-TOF. The X! Tandem algorithm (Craig and Beavis 2004) as implemented by the Global Proteome Machine Organization website (http://www.thegpm.org) was used for peptide identification and estimating sequence coverage. Results were further evaluated using log(*e*) values to estimate correctness of peptide assignments (Fenyö and Beavis 2003).

Supplementary data

Supplementary data for this article are available online at http://glycob.oxfordjournals.org/.

Funding

Financial support came from the Austrian Science Fund FWF, projects P20605-B12 and P24317-B22 (to C.S.) and the PhD programme "BioToP—Biomolecular Technology of Proteins" (Austrian Science Fund, FWF project W1224), and National Institute of Health/National Institute of Allergy and Infectious Diseases grant R01AI067711 (to L.C.).

Conflict of interest

None declared.

Abbreviations

BHI, brain heart infusion; CID, collision induced dissociation; Dig, digitoxose; DTT, dithiothreitol; FPLC, fast protein liquid chromatography; Fuc, fucose; GSH_{10}, glutamine-serine-10x histidine; IMAC, immobilized metal affinity chromatography; LC-ESI-MS/MS, liquid chromatography electrospray ionization tandem mass spectrometry; ManNAcA, *N*-acetylmannosamiuronic acid; $ManNAcCONH_2$, *N*-acetylmannosamiuronamide; MeGal, 4-methyl-galactose; mi, molecular ion; Mw, molecular weight; PGC-ESI-MS/MS, porous graphitized carbon electrospray ionization mass spectrometry; Pse5Am7Gc, 5-acetimidol-7-*N*-glycolylpseudaminic acid; PCR, polymerase chain reaction; PVDF, polyvinylidene difluoride; RSH_{10}, arginine-serine-10x histidine; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; SPE, solid phase extraction.

References

Bayley DP, Rocha ER, Smith CJ. 2000. Analysis of *cepA* and other *Bacteroides fragilis* genes reveals a unique promoter structure. *FEMS Microbiol Lett*. 193:149–154.

Craig R, Beavis RC. 2004. TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics*. 20:1466–1467.

de O. Ferreira E, Araújo Lobo L, Barreiros Petrópolis D, dos S, Avelar KE, Ferreira MC, e Silva Filho FC, Domingues RMCP. 2006. A *Bacteroides fragilis* surface glycoprotein mediates the interaction between the bacterium and the extracellular matrix component laminin-1. *Res Microbiol*. 157:960–966.

Domon B, Costello CE. 1988. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj J*. 5:397–409.

Dutton RJ, Boyd D, Berkmen M, Beckwith J. 2008. Bacterial species exhibit diversity in their mechanisms and capacity for protein disulfide bond formation. *Proc Natl Acad Sci USA*. 105:11933–11938.

Feldman MF, Wacker M, Hernandez M, Hitchen PG, Marolda CL, Kowarik M, Morris HR, Dell A, Valvano MA, Aebi M. 2005. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proc Natl Acad Sci USA*. 102:3016–3021.

Fenyö D, Beavis RC. 2003. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal Chem*. 75:768–774.

Fletcher CM, Coyne MJ, Comstock LE. 2011. Theoretical and experimental characterization of the scope of protein *O*-glycosylation in *Bacteroides fragilis*. *J Bio Chem*. 286:3219–3226.

Fletcher CM, Coyne MJ, Villa OF, Chatzidaki-Livanis M, Comstock LE. 2009. A general O-glycosylation system important to the physiology of a major human intestinal symbiont. *Cell*. 137:321–331.

Guiney DG, Hasegawa P, Davis CE. 1984. Plasmid transfer from *Escherichia coli* to *Bacteroides fragilis*: Differential expression of antibiotic resistance phenotypes. *Proc Natl Acad Sci USA*. 81:7203–7206.

Holt SC, Ebersole JL. 2005. *Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia*: The "red complex", a prototype polybacterial pathogenic consortium in periodontitis. *Periodontol 2000*. 38:72–122.

Honma K, Inagaki S, Okuda K, Kuramitsu HK, Sharma A. 2007. Role of a *Tannerella forsythia* exopolysaccharide synthesis operon in biofilm development. *Microb Pathog*. 42:156–166.

Iwashkiw JA, Seper A, Weber BS, Scott NE, Vinogradov E, Stratilo C, Reiz B, Cordwell SJ, Whittal R, Schild S, et al. 2012. Identification of a general *O*-linked protein glycosylation system in *Acinetobacter baumannii* and its role in virulence and biofilm formation. *PLoS Pathog*. 8:e1002758.

Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, Mills DC, Watson DC, Hernandez M, Kelly JF, et al. 2006. Definition of the bacterial *N*-glycosylation site consensus sequence. *EMBO J*. 25: 1957–1966.

Krapp S, Mimura Y, Jefferis R, Huber R, Sondermann P. 2003. Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *J Mol Biol*. 325:979–989.

Ku SC, Schulz BL, Power PM, Jennings MP. 2009. The pilin O-glycosylation pathway of pathogenic *Neisseria* is a general system that glycosylates AniA, an outer membrane nitrite reductase. *Biochem Biophys Res Commun*. 378:84–89.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 227:680–685.

Langsford ML, Gilkes NR, Singh B, Moser B, Miller RC, jr, Warren RAJ, Kilburn DG. 1987. Glycosylation of bacterial cellulases prevents proteolytic cleavage between functional domains. *FEBS Lett*. 225:163–167.

Lee SW, Sabet M, Um HS, Yang J, Kim HC, Zhu W. 2006. Identification and characterization of the genes encoding a unique surface (S-) layer of *Tannerella forsythia*. *Gene*. 371:102–111.

Lizak C, Fan Y-Y, Weber TC, Aebi M. 2011. N-linked glycosylation of antibody fragments in *Escherichia coli*. *Bioconj Chem*. 22:488–496.

Mazmanian SK, Round JL, Kasper DL. 2008. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature*. 453:620–625.

Messner P. 2009. Prokaryotic protein glycosylation is rapidly expanding from "curiosity" to "ubiquity". *ChemBioChem*. 10:2151–2154.

Mimura Y, Church S, Ghirlando R, Ashton PR, Dong S, Goodall M, Lund J, Jefferis R. 2000. 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*. 37:697–706.

Nothaft H, Szymanski CM. 2010. Protein glycosylation in bacteria: Sweeter than ever. *Nat Rev Micro*. 8:765–778.

Pabst M, Altmann F. 2008. Influence of electrosorption, solvent, temperature, and ion polarity on the performance of LC-ESI-MS using graphitic carbon for acidic oligosaccharides. *Anal Chem*. 80:7534–7542.

Pabst M, Bondili JS, Stadlmann J, Mach L, Altmann F. 2007. Mass + retention time = structure: A strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans. *Anal Chem*. 79:5051–5057.

Packer NH, Lawson MA, Jardine DR, Redmond JW. 1998. A general approach to desalting oligosaccharides released from glycoproteins. *Glycoconj J*. 15:737–747.

Pihlstrom BL, Michalowicz BS, Johnson NW. 2005. Periodontal diseases. *Lancet*. 366:1809–1820.

Posch G, Pabst M, Brecker L, Altmann F, Messner P, Schäffer C. 2011. Characterization and scope of S-layer protein *O*-glycosylation in *Tannerella forsythia*. *J Biol Chem*. 286:38714–38724.

Posch G, Sekot G, Friedrich V, Megson Z, Koerdt A, Messner P, Schäffer C. 2012. Glycobiology aspects of the of the periodontal pathogen *Tannerella forsythia*. *Biomolecules*. 2:467–482.

Sakakibara J, Nagano K, Murakami Y, Higuchi N, Shimozato K, Yoshimura F. 2007. Loss of adherence ability to human gingival epithelial cells in S-layer protein-deficient mutants of *Tannerella forsythensis*. *Microbiology*. 153:866–876.

Schwarz F, Huang W, Li C, Schulz BL, Lizak C, Palumbo A, Numao S, Neri D, Aebi M, Wang L-X. 2010. A combined method for producing homogeneous glycoproteins with eukaryotic *N*-glycosylation. *Nat Chem Biol*. 6:264–266.

Sekot G, Posch G, Messner P, Matejka M, Rausch-Fan X, Andrukhov O, Schäffer C. 2011. Potential of the *Tannerella forsythia* S-layer to delay the immune response. *J Dent Res*. 90:109–114.

Sekot G, Posch G, Oh Y, Zayni S, Mayer H, Pum D, Messner P, Hinterdorfer P, Schäffer C. 2012. Analysis of the cell surface layer ultrastructure of the oral pathogen *Tannerella forsythia*. *Arch Microbiol*. 194:525–539.

Settem RP, Honma K, Nakajima T, Phansopa C, Roy S, Stafford GP, Sharma A. 2012. A bacterial glycan core linked to surface (S)-layer proteins modulates host immunity through Th17 suppression. *Mucosal Immunol*. doi:10.1038/mi.2012.85.

Smith CJ, Rogers MB, McKee ML. 1992. Heterologous gene expression in *Bacteroides fragilis*. *Plasmid*. 27:141–154.

Stadlmann J, Pabst M, Kolarich D, Kunert R, Altmann F. 2008. Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. *Proteomics*. 8:2858–2871.

Steiner K, Novotny R, Patel K, Vinogradov E, Messner P, Schäffer C. 2007. Functional characterization of the initiation enzyme of S-Layer glycoprotein glycan biosynthesis in *Geobacillus stearothermophilus* NRS 2004/3a. *J Bacteriol*. 189:2590–2598.

Szymanski CM, Yao R, Ewing CP, Trust TJ, Guerry P. 1999. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol Microbiol*. 32:1022–1030.

Valderrama-Rincon JD, Fisher AC, Merritt JH, Fan Y-Y, Reading CA, Chhiba K, Heiss C, Azadi P, Aebi M, DeLisa MP. 2012. An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. *Nat Chem Biol*. 8:434–436.

Veith PD, O'Brien-Simpson NM, Tan Y, Djatmiko DC, Dashper SG, Reynolds EC. 2009. Outer membrane proteome and antigens of *Tannerella forsythia*. *J Proteome Res*. 8:4279–4292.

Vik Å, Aas FE, Anonsen JH, Bilsborough S, Schneider A, Egge-Jacobsen W, Koomey M. 2009. Broad spectrum O-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae*. *Proc Natl Acad Sci USA*. 106:4447–4452.

Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, et al. 2002. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science*. 298:1790–1793.

Wexler HM. 2007. *Bacteroides*: The good, the bad, and the nitty-gritty. *Clinical Microbiol Rev*. 20:593–621.

Wuhrer M, Deelder AM, van der Burgt YEM. 2011. Mass spectrometric glycan rearrangements. *Mass Spectrom Rev*. 30:664–680.

Young NM, Brisson J-R, Kelly J, Watson DC, Tessier L, Lanthier PH, Jarrell HC, Cadotte N, St. Michael F, Aberg E, et al. 2002. Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *J Biol Chem*. 277:42530–42539.

Appendix VII

Anzengruber, J., Pabst, M., **Neumann, L.**, Sekot, G., Heinl, S., Grabherr, R., Altmann, F.,Messner, P., Schaeffer, C. (2014) Protein *O*-glucosylation in *Lactobacillus buchneri*. *Glycoconj. J.* 31, 117-131. doi: 10.1007/s10719-013-9505-7

Glycoconj J (2014) 31:117–131
DOI 10.1007/s10719-013-9505-7

Protein *O*-glucosylation in *Lactobacillus buchneri*

Julia Anzengruber · Martin Pabst · Laura Neumann · Gerhard Sekot · Stefan Heinl · Reingard Grabherr · Friedrich Altmann · Paul Messner · Christina Schäffer

Received: 10 September 2013 / Revised: 4 October 2013 / Accepted: 7 October 2013 / Published online: 27 October 2013

Abstract Based on the previous demonstration of surface (S-) layer protein glycosylation in *Lactobacillus buchneri* 41021/251 and because of general advantages of lactic acid bacteria for applied research, protein glycosylation in this bacterial species was investigated in detail. The cell surface of *L. buchneri* CD034 is completely covered with an oblique 2D crystalline array (lattice parameters, $a = 5.9$ nm; $b = 6.2$ nm; $\gamma \sim 77°$) formed by self-assembly of the S-layer protein SlpB. Biochemical and mass spectrometric analyses revealed that SlpB is the most abundant protein and that it is *O*-glycosylated at four serine residues within the sequence S_{152}-A-S_{154}-S_{155}-A-S_{157} with, on average, seven Glc(α1-6) residues, each. Subcellular fractionation of strain CD034 indicated a sequential order of SlpB export and glucosylation as evidenced by lack of glucosylation of cytosolic SlpB. Protein glycosylation analysis was extended to strain *L. buchneri* NRRL B-30929 where an analogous glucosylation scenario could be detected, with the S-layer glycoprotein SlpN containing an *O*-glycosylation motif identical to that of SlpB. This corroborates previous data on S-layer protein glucosylation of strain 41021/251 and let us propose a species-wide S-layer protein *O*-glucosylation in *L. buchneri* targeted at the sequence motif S-A-S-S-A-S. Search of the *L. buchneri* genomes for the said glucosylation motif revealed one further ORF, encoding the putative glycosyl-hydrolase *Lb*GH25B and *Lb*GH25N in *L. buchneri* CD034 and NRRL B-30929, respectively, for which we have indications of a glycosylation comparable to that of the S-layer proteins. These findings demonstrate the presence of a distinct protein *O*-glucosylation system in Gram-positive and beneficial microbes.

Keywords Glycobiology · *O*-glycosylation motif · Lactic acid bacteria · Glycosyl-hydrolase · S-layer

J. Anzengruber · G. Sekot · P. Messner · C. Schäffer (✉)
Department of NanoBiotechnology, NanoGlycobiology unit, Universität für Bodenkultur Wien, Muthgasse 11, 1190 Vienna, Austria
e-mail: christina.schaeffer@boku.ac.at

J. Anzengruber
e-mail: julia.anzengruber@boku.ac.at

G. Sekot
e-mail: gerhard.sekot@acib.at

P. Messner
e-mail: paul.messner@boku.ac.at

M. Pabst · L. Neumann · F. Altmann
Department of Chemistry, Universität für Bodenkultur Wien, Muthgasse 18, 1190 Vienna, Austria

M. Pabst
e-mail: martin.pabst@org.chem.ethz.ch

L. Neumann
e-mail: laura.neumann@boku.ac.at

F. Altmann
e-mail: fierdich.altmann@boku.ac.at

S. Heinl · R. Grabherr
Christian Doppler Laboratory for Genetically Engineered Lactic Acid Bacteria, Department of Biotechnology, Universität für Bodenkultur Wien, Muthgasse 11, 1190 Vienna, Austria

S. Heinl
e-mail: setfan.heinl@boku.ac.at

R. Grabherr
e-mail: reingard.grabherr@boku.ac.at

Present Address:
M. Pabst
Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland

Present Address:
G. Sekot
Austrian Centre of Industrial Biotechnology, Muthgasse 18, 1190 Vienna, Austria

Abbreviations

SlpB	S-layer protein of *L. buchneri* CD034
SlpN	S-layer protein of *L. buchneri* NRRL B-30929
*Lb*GH25B	Glycosyl-hydrolase of *L. buchneri* CD034
*Lb*GH25N	Glycosyl-hydrolase of *L. buchneri* NRRL B-30929

Introduction

Glycosylation is one of the most common protein modifications involved in many pivotal biological processes [1]. While having long been overlooked, nowadays, glycosylation of proteins in bacteria is becoming increasingly documented, including both *O*- and *N*-linked protein glycosylation [2–8]. Most bacteria known to glycosylate proteins have specialized glycosylation systems where a few abundant polymeric surface proteins such as flagellins, pilins, or surface (S-) layer proteins [9–11] as well as virulence factors of pathogens [12–14] are glycosylated. Relatively few bacterial species have been shown to have general protein glycosylation systems where glycans are added to many proteins with diverse functions and subcellular localizations. *Campylobacter* spp. were the first bacteria demonstrated to have a general protein glycosylation system where target proteins are *N*-glycosylated [15], with *Campylobacter jejuni* being the first bacterium for which the glycosylation machinery was fully reconstituted in *E. coli* [16]. General *O*-glycosylation systems have so far been described in pathogenic *Neisseria* spp. [17], in the major gut symbiont *Bacteroides fragilis* [18] and in phylogenetically related species, such as the periodontal pathogen *Tannerella forsythia* [19].

In Gram-positive and beneficial bacteria, such as lactic acid bacteria, general glycosylation systems have not been documented so far. Lactic acid bacteria are used in various ways of applied research due to their GRAS (generally regarded as safe) status, including food and feed production [20] and production of recombinant proteins and metabolites [21–23]. Furthermore, they are promising candidates for carbohydrate engineering purposes, aimed at, for instance, the improvement of the therapeutic behavior of protein drugs [24] or the design of immunomodulatory agents [25].

Glycosylation of lactobacillar proteins was previously reported for the S-layer proteins of *Lactobacillus buchneri* 41021/251 [26], *Lactobacillus helveticus* ATCC12046 [27], *Lactobacillus acidophilus* NCFM [28], *Lactobacillus plantarum* 41021/252 [26], and various *Lactobacillus kefir* strains [29]. These S-layer glycoproteins share the common feature of self-assembling into closed, 2D crystalline arrays on the bacterial cell surface [8, 30]. While S-layer glycosylation of most of these proteins was inferred only from a positive periodic acid-Schiff staining reaction of an SDS-PAGE gel, for *L. buchneri* 41021/251, the S-layer glycans were shown to be glucose oligomers attached to serine residues of the S-layer protein [26]. Other known lactobacillar glycoproteins are extracellular, cell wall-hydrolyzing enzymes containing *O*-linked carbohydrate moieties, found in *Streptococcus faecium* ATCC 9790 [31], *L. plantarum* WCFS1 [32, 33], and *Lactobacillus rhamnosus* GG [34]. Recently, even cysteine *S*-glycosylation, a novel post-translational modification, was reported in glycopeptide bacteriocins of *L. plantarum* [35, 36]. The demonstration of glycans on several extracellular proteins of lactobacilli supports recent studies pinpointing the importance of lactobacillar cell surface glycosylation for, *e.g.* adhesion and biofilm formation [37] as wells as gastrointestinal persistence [38] and adaptation [39].

Here, S-layer protein glycosylation in the species *L. buchneri* was investigated in detail. *L. buchneri* strains are Gram-positive, obligatory hetero-fermentative, facultative anaerobes; they have been isolated from different sources, ranging from stable grass silage [40, 41], ethanol production plants [42], pickled juice [43] and acid-coagulating cheese samples [44] to the human intestine [45]. One strain was shown to have probiotic effects manifested in cholesterol reduction, acid and bile tolerance, and antimicrobial activity [43]. *L. buchneri* strains are intensively investigated because of their property of efficient preservation of animal feed silages against aerobic spoilage. The obligatory hetero-fermentative nature and acid resistance have drawn attention to this species for applications as silage starter culture [46–51]. *L. buchneri* strains used in this work include *L. buchneri* CD034, isolated from grass silage [40, 41], *L. buchneri* 41021/251, isolated from a dairy factory [26], and *L. buchneri* NRRL B-30929, isolated from an ethanol production plant [42]. The complete genome sequences of *L. buchneri* CD034 [52] and *L. buchneri* NRRL B-30929 [53] were published recently.

To assess the significance of protein glycosylation in the species *L. buchneri* this study specifically deals with the i) identification and electron-microscopic visualization of the S-layer glycoprotein SlpB of *L. buchneri* CD034, ii) elucidation of its S-layer glycan structure, iii) analysis of SlpB glycosylation sites, iv) determination of the cellular site of SlpB glycosylation, v) identification of the S-layer glycoprotein SlpN in the closely related strain *L. buchneri* NRRL B-30929, and vi) investigation of glycosylation of the putative glycosyl-hydrolases, *Lb*GH25B and *Lb*GH25N, in *L. buchneri* CD034 and NRRL B-30929, respectively.

Materials and methods

Bacterial strains and culture conditions

L. buchneri CD034 [40, 41, 52], *L. buchneri* NRRL B-30929 [42, 53, 54], and *L. buchneri* 41021/251 [26] were grown in De Man-Rogosa-Sharpe (MRS) broth (Oxoid, Basingstoke,

Hampshire, UK) [55] at 37 °C without shaking. *Escherichia coli* DH5α cells (Life Technologies, Vienna, Austria) and *E. coli* BL21 (DE3) cells (Life Technologies) were cultivated at 37 °C and 200 rpm in Luria-Bertani (LB) medium supplemented with 50 μg mL^{-1} kanamycin (Km).

Freeze-etching and transmission electron microscopy

Cell morphology of *L. buchneri* CD034 was examined by transmission electron microscopy (TEM) of freeze-etched intact bacteria [56]. Cells were harvested by centrifugation and washed three times with sterile double-distilled water (MilliQ-water). Freeze-etching was carried out in a BAF 060 unit (Leica, Wetzlar, Germany). Fracturing and etching of frozen samples was done at -96 °C for 90 s prior to platinum/carbon shadowing. Replicas were purified for 30 min in 70 % sulfuric acid, neutralized in MilliQ-water and subsequently subjected to 14 % sodium hypochloride solution (for 3-5 min), followed by three washing steps in MilliQ-water water and immobilization on 400-mesh TEM copper grids (Agar Scientific, Stansted, UK). Samples were investigated in a Tecnai G^2 20 Twin TEM (FEI, Eindhoven, The Netherlands), operating at 80 keV, and micrographs were taken with an FEI Eagle 4 k CCD camera (4,096×4,096 pixels). S-layer lattice parameters were obtained from the power spectra directly taken in the TEM as published recently [57].

Production and purification of recombinant proteins

The coding sequences for the S-layer protein SlpB of *L. buchneri* CD034 (LBUCD034_1608) devoid of the N-terminal signal peptide sequence (corresponding to amino acids 1–31 of the pre-protein) was PCR-amplified with the primers (Life Technologies) SlpB_F (5′- AATCA CCATGGGCAAATCATATGCCAAAGTTAC -3′, NcoI site is underlined) and SlpB_R (5′- AATCACTCGAG ATTAAACGGTGTAACAGTAAC -3′, XhoI site is underlined), digested with NcoI- and XhoI and ligated into pET28a+ expression vector (Novagen, Madison, WI, USA). Similarly, the coding sequences for *Lb*GH25B (LBUCD034_0240) and *Lb*GH25N (Lbuc_0200) were PCR-amplified without signal peptides (amino acids 1–29) using the primers *Lb*GH25B/N_F (5′-AATCATCTAGA AATAATTTTGTTTAACTTTAAGAAGGAGATATACCAT-GGCTTTGACGCCGTCCAGTACC-3′, XbaI site is underlined) and *Lb*GH25B/N_His_stop_R (5′- AATCAAAGCTT TTAATGATGATGATGATGATGATTCAAATAACCGCGCC-AAATCC-3′, HindIII site is underlined), digested with XbaI and HindII- and ligated into pET28a+. The resulting recombinant plasmids (pET28a-SlpB_His_6, pET28a-*Lb*GH25B_His_6 and pET28a-*Lb*GH25N_His_6) were propagated in *E. coli* BL21 Star (DE3) for production of hexahistidine-tagged proteins.

The transformed strains were grown in LB broth supplemented with kanamycin (50 μg mL^{-1}) at 37 °C and 200 rpm. At the mid-exponential growth phase (OD_{600}~0.6), protein expression was induced with 1 mM isopropyl β-D-1-thio-galactopyranoside (IPTG) and cultivation was continued for 4 h. Cells were pelleted by centrifugation (10,000 x *g*, 20 min), resuspended in lysis buffer (50 mM sodium citrate buffer, pH 6.2, 0.1 % Triton X-100) and, after addition of lysozyme (800 μg mL^{-1}; Sigma-Aldrich, Vienna, Austria) and benzonase (50 U mL^{-1}; Sigma-Aldrich), incubated for 30 min at 37 °C. Bacteria were further lysed by ultrasonication (Branson sonifier, duty cycle 50 %; output 6) applying ten cycles of 10 pulses with 30 s breaks, each, and insoluble inclusion bodies containing the recombinant proteins were pelleted. The proteins were extracted from the pellets with binding buffer (50 mM sodium citrate buffer, pH 5.5, 5 M GdHCl (guanidinium hydochloride), 20 mM imidazol, 0.5 M NaCl) for 1 h at 4 °C. The extracts were subjected to centrifugation and membrane-filtering (0.45-μm pore size). The resulting protein samples were applied to a 1-ml HisTrap HP column (GE Healthcare, Vienna, Austria) and recombinant proteins were recovered with elution buffer (50 mM sodium citrate buffer, pH 5.5, 5 M GdHCl, 1 M imidazol, 0.5 M NaCl). Recombinant proteins were dialyzed against 50 mM sodium citrate buffer, pH 5.5.

SDS-PAGE

SDS-PAGE was carried out on 15 % and 20 % slab gels in a Mini Protean electrophoresis apparatus (Bio-Rad, Vienna, Austria) according to Laemmli [58]. Laemmli buffer contained 62.5 mM Tris-HCl, pH 6.9, 2 % SDS, 1 % β-mercaptoethanol, and 10 % glycerol. Protein bands were visualized with colloidal Coomassie Brilliant Blue R-250 (CBB) staining reagent, carbohydrates with periodic acid-Schiff (PAS) staining reagent or with Pro-Q Emerald 300 fluorescent stain (Life Technologies) [59], and the gels were imaged at 700 nm using the Odyssey imaging system (LICOR, Lincoln, NE, USA) and at 300 nm using the Infinity-3000 apparatus (Vilber-Lourmat, Marne-la-Vallée, France), respectively.

Western-immunoblotting

Immunization of mice and preparation of polyclonal antiserum against purified recombinant SlpB and the predicted glycosyl-hydrolase *Lb*GH25N was done at EF-BIO s.r.o. (Bratislava, Slovakia). Due to high amino acid homology (99 %) of the predicted glycosyl-hydrolases, the *Lb*GH25N-specific antiserum recognized also the homologous protein *Lb*GH25B.

Proteins were transferred from the SDS-PAGE gel to a polyvinylidene difluoride membrane (Bio-Rad) using a Mini

Trans-Blot Cell (Bio-Rad). The S-layer and glycosyl-hydrolase specific antisera were used in combination with IR Dye 800CW goat anti-mouse antibody (LI-COR) and detection was performed at 800 nm using the Odyssey Infrared Imaging System (LI-COR).

O-glycan preparation

The S-layer proteins SlpB and SlpN as well as the predicted glycosyl-hydrolases *Lb*GH25B and *Lb*GH25N of *L. buchneri* CD034 and *L. buchneri* NRRL B-30929, respectively were excised from SDS-PAGE gels and *O*-glycans were isolated by applying in-gel reductive β-elimination [60]. Excised gel slices were transferred to a 1.5-ml Eppendorf tube, covered with 250–500 μL of 1 M $NaBH_4$ in 0.5 M NaOH, and incubated at 50 °C for 18 h. Excess of salt was removed using 10 mg HyperSepHypercarb SPE cartridges (Thermo Fisher Scientific, Vienna, Austria) according to published protocols [61–63]. For comparison, *O*-glycans of the S-layer protein *L. buchneri* 41021/251 were isolated by applying in-gel reductive β-elimination with 1 M $NaBD_4$ in 0.5 M NaOH. Borohydride- and borodeuteride-reduced glycans were vacuum-dried and further analyzed.

Monosaccharide analysis

To identify SlpB glycan constituents, borohydride-reduced *O*-glycans were hydrolyzed with 25 % TFA at 110 °C for 4 h and released monosaccharides were analyzed by high-performance anion-exchange chromatography with pulsed electrochemical detection (HPAEC-PED) (Dionex, Sunnyvale, CA, USA, [64]).

LC-ESI-MS

Borohydride- and borodeuteride-reduced *O*-glycans were analyzed by PGC-ESI-MS as described recently (Hypercarb, 0.32×150 mm, inner diameter 5 μm) [19, 63, 65]. Ammonium formate buffer (0.3 % formic acid, pH 3.0) was used as buffer A, and a gradient was performed from 0 % to 35 % acetonitrile in buffer A within 35 min using a Dionex Ultimate 3000 (cap flow, 8 μL min^{-1}). Detection was done by an ESI-Q-TOF Global Ultima from Micromass (Waters, Eschborn, Germany). Data were evaluated using MassLynx 4.0 software.

Proteomics

For the identification of peptides of the *L. buchneri* CD034 and *L. buchneri* NRRL B-30929 glycoproteins, an in-gel trypsin digest of the respective bands from SDS-PAGE gels was performed as described recently [19].

Extracted peptides from *L. buchneri* CD034 were further subjected to RP liquid chromatography coupled to ESI-MS/MS [63]. GPM and the X!tandem algorithm (http://www.thegpm.org/) were used for protein database search of the MS/MS data. Results were further evaluated using log(e) values to estimate correctness of peptide assignments.

L. buchneri NRRL B-30929 peptides were dissolved in 50 % acetonitrile and 0.1 % trifluoroacetic acid. An aliquot (0.5 μl) was directly spotted onto a ground steel sample plate and analyzed by MALDI-TOF MS with α-cyanocinnamic acid as matrix. The sample was analyzed in positive ion mode using a Bruker Autoflex Speed (Bruker Daltonics, Bremen, Germany) equipped with a Smartbeam-II solid-state 1,000 Hz laser; typically 1,000–2,000 shots were summed. Data were analyzed and predicted using the ProteinProspector MS-Fit and MS-Digest on-line software (http://prospector.ucsf.edu). SignalP 4.0 Server was used for prediction of signal peptides. Secondary structure prediction of SlpB was computed with PSIPRED protein structure prediction server.

O-glycosylation site analysis

Mass spectrometric determination of the *O*-glycosylation sites was performed according to a published protocol [66]. Briefly, an in-gel trypsin digest of the respective band on SDS-PAGE gels was performed, tryptic peptides were extracted and dried under vacuum. The whole digest was further digested with GluC in 20 mM ammonium bicarbonate overnight and the digest was further fractionated by RP chromatography (see above). Glyco-positive fractions were evaluated by ESI-MS. Glycans were released from the peptide by ammonia-based non-reductive β-elimination (25 % ammonia for 18 h at 37 °C). The peptide was further subjected to RP liquid chromatography coupled to ESI-MS/MS [63]. The deglycosylated peptide was sequenced and free serine residues could be distinguished from originally glycan-carrying serines by a mass difference of one Da.

Subcellular fractionation of *L. buchneri* CD034

L. buchneri CD034 *was* grown overnight in MRS broth at 37 °C without shaking. The bacterial cell pellet obtained after centrifugation (10,000 × *g*, 20 min) was washed with 50 mM Tris-HCl, pH 7.2 (buffer B) and, subsequently, lysed by ultrasonication in buffer B (Branson sonifier, duty cycle 50 %; output 8) by applying six cycles of 10 pulses, with 30 s breaks, each.

Cell wall fraction Cell wall fragments were pelleted (30,000 × *g*, 15 min) and resuspended in buffer B, containing 1 % Triton X-100, and pelleted again. After four washing steps with buffer

B, proteins were resuspended in PBS and mixed with Laemmli buffer prior to SDS-PAGE analysis.

Cytosol Cytosolic proteins were obtained from the supernatant of the centrifugation step directly after ultrasonication during the cell wall isolation described above. Proteins were precipitated with trichloroacetic acid (20 % final concentration) at 4 °C for 30 min. After centrifugation (9,000 × *g*, 20 min), precipitated proteins were washed twice with ice-cold 100 % acetone. The pellet was air-dried and proteins resuspended in Laemmli buffer.

Results

Description of the cell surface of *L. buchneri* CD034

L. buchneri CD034 possesses a typical Gram-positive cell wall profile with a thick peptidoglycan layer that is additionally covered by an S-layer as outermost cell envelope decoration. This S-layer is formed by self-assembly of the glycoprotein SlpB into a 2D lattice. Freeze-etching experiments of intact bacterial cells showed an oblique S-layer lattice with the base vectors $a = 5.9$ nm, $b = 6.2$ nm, and an enclosed angle of $\gamma \sim 77°$ (Fig. 1). This compares well to the S-layer lattice parameters of *L. buchneri* strains 41021/251 ($a = 5.4$ nm, $b = 6.1$ nm, and $\gamma \sim 80°$) [29] and ATCC 4005 [67]. For the latter strain, careful inspection of the published power spectrum and also of an undisturbed area of the S-layer protein lattice as revealed by a freeze-etched preparation of intact cells, unambiguously revealed the presence of an S-layer with oblique lattice symmetry (center-to-center spacing, ~ 6.0 nm) in contrast to the published hexagonal lattice symmetry.

SDS-PAGE analysis of protein glycosylation in *L. buchneri* CD034

Strong carbohydrate staining reactions with Pro-Q Emerald reagent upon separation of a crude cell extract by SDS-PAGE were observed for *L. buchneri* CD034 (Fig. 2, lane 2). The most abundant glyco-positive stained protein band with an apparent molecular mass of ~60 kDa (indicated by an arrow in Fig. 2) was excised from the SDS-PA gel, digested with trypsin, and subjected to reversed phase LC (RP)-ESI-MS/MS peptide mapping. This glycoprotein could be clearly identified as the S-layer protein of the bacterium by running the MS/MS data against the *L. buchneri* CD034 proteome database using GPM and the X!tandem algorithm (http://www.thegpm.org/). The high abundance of that protein on the CBB-stained SDS-PA gel of the whole cellular proteome (Fig. 2, lane 1) correlates with the feature of S-layer proteins being the most abundant cellular proteins, with a total protein biosynthesis effort of up to 20 % being devoted to S-layer protein biosynthesis. The high amount of S-layer proteins is required for a complete coverage of the bacterial cell with a closed S-layer lattice during all stages of the growth cycle

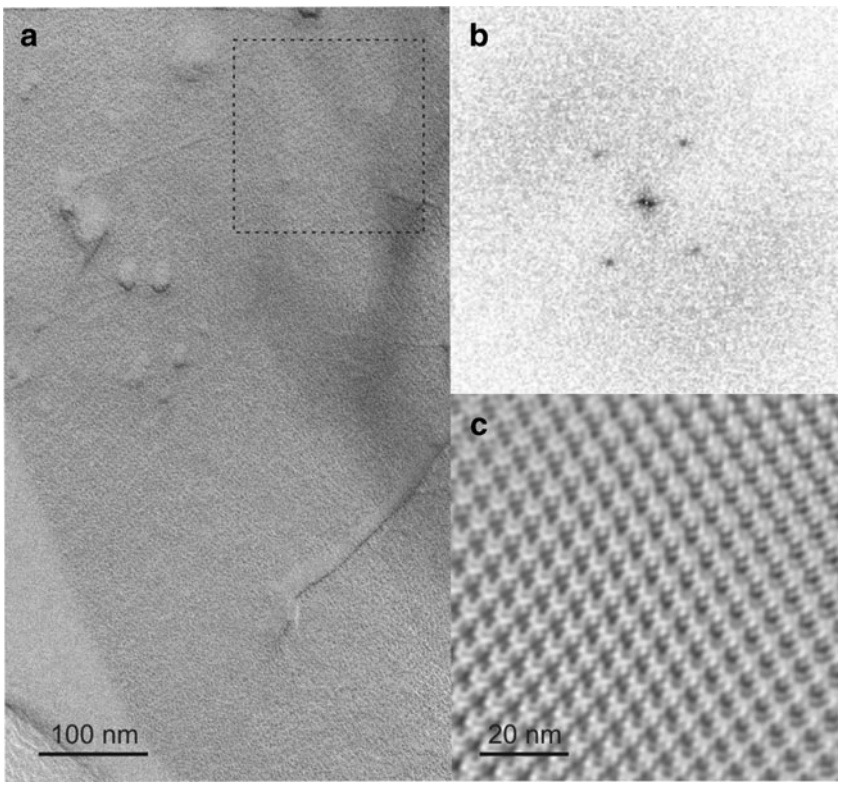


Fig. 1 TEM micrograph of a freeze-etched and metal-carbon-shadowed preparation of *L. buchneri* CD034 cells. **a** Oblique S-layer lattice; **b** power spectrum of the S-layer lattice from the boxed area in (**a**), **c** lattice reconstruction

[30]. The detected S-layer glycoprotein, SlpB (gene LBUCD034_1608), is one out of eight putative S-layer precursor proteins encoded on the genome of *L. buchneri* CD034 [52]. It is composed of 558 amino acid residues corresponding to a calculated molecular mass of 58.3 kDa and contains 11 % of α-helices, 30 % of β-sheets, and 59 % of random coils according to computed secondary structure prediction; this is in agreement with the predicted secondary structure of other lactobacillar S-layer proteins [68]. As predicted by SignalP 4.0 Server and confirmed by ESI-MS/MS peptide mapping (data not shown), the S-layer protein contains an N-terminal signal peptide of 31 amino acids, exhibiting the sequence MKKSLKKTLFAGVAALSFVAVAGVSSTNASA which includes the commonly found A-X-A motif preceding the cleavage site for type I signal peptidases [69]. Overall SlpB possesses a basic pI of 10.2 that is typical of lactobacillar S-layer proteins [30, 70]. Interestingly, the amino acid sequence of the newly identified S-layer protein SlpB includes the amino acid motif S-S-A-S-S-A-S-S-A, which is also present in the S-layer protein of *L. buchneri* strain 41021/251 [26].

Additionally, a second, ~42-kDa, slightly blurred, glyco-positive band of lower abundance was detected with Pro-Q Emerald reagent (indicated with by asterisk in Fig. 2). Details about this band are given below.

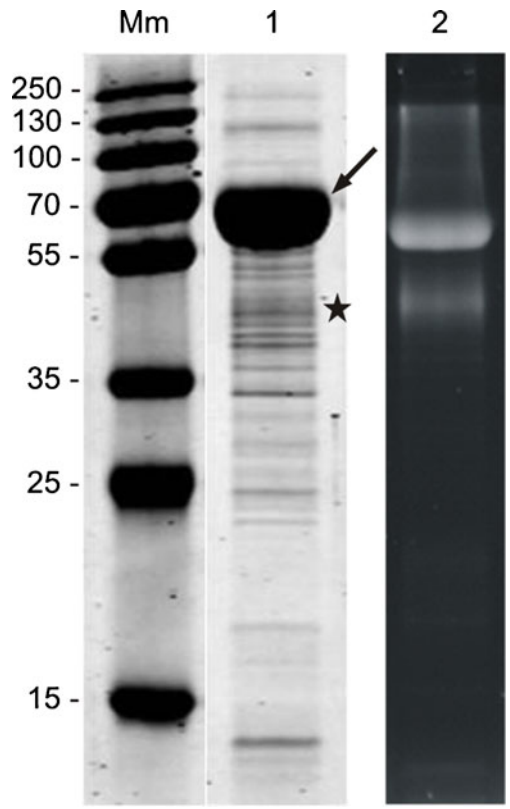


Fig. 2 SDS-PAGE analysis (15 % gel) of a crude cell extract from *L. buchneri* CD034. Mm, PageRuler Plus pre-stained protein ladder (Thermo Scientific). Gels were stained with CBB (lane 1) and Pro-Q Emerald (lane 2). Two glyco-positive protein bands were detected. The arrow indicates the S-layer glycoprotein SlpB of *L. buchneri* CD034 which makes up the main portion of the whole cellular proteome. A second ~42-kDa, glyco-positive protein was detected and is indicated by an asterisk

S-layer glycan profiles of *L. buchneri* CD034 and *L. buchneri* 41021/251

The presence of the amino acid signature motif S-S-A-S-S-A-S-S-A, ranging from amino acid 151 to 159 within SlpB of the currently investigated *L. buchneri* CD034 as well as in the *L. buchneri* 41021/251S-layer protein [26], let us hypothesize that comparable or even identical S-layer protein glycosylation might occur in these bacteria. This hypothesis was tested by first analyzing the S-layer glycan structure of SlpB of *L. buchneri* CD034 and comparing it with the known S-layer glycan structure of strain 41021/251 [26]. Extracts of cellular proteins of either bacterial strain were separated by SDS-PAGE and S-layer protein bands were excised from the 15 % SDS-PA gels. S-layer glycans from *L. buchneri* CD034 were released from the protein backbone by in-gel reductive β-elimination using borohydride, while those of strain 41021/251 were released using borodeuteride. Subsequently, glycan analysis was performed by LC-PGC (porous graphitized carbon)-ESI-MS/MS. The LC profile clearly indicated the presence of identical glycans on the S-layer proteins of *L. buchneri* CD034 and on 41021/251 (Fig. 3a), and showed both glycans to be composed of hexose units and, most probably, according to identical retention times of the S-layer glycans of either strain on the PGC column [65], of Glc(α1–6). Figure 3b shows a summarized ESI-MS spectrum representing the mass units of repeating hexose units. *L. buchneri* CD034 S-layer glycans were found with monoisotopic values of 848, 1,010, 1,172 and 1,334 Da corresponding to homooligomers composed of five, six, seven and eight units, respectively. As expected, a mass increase of 1 Da was observed for the borodeuteride-reduced S-layer glycans of strain 41021/251 compared to borohydride-reduced S-layer glycans of strain CD034 (Fig. 3b). Further, monosaccharide analysis of borohydride-reduced and TFA-hydrolyzed S-layer glycans from *L. buchneri* CD034 was performed by HPAEC-PED. The only monosaccharide constituent detected was glucose (Fig. 3c), suggesting that the SlpB glycans consist of a glucose oligomer; this is corroborated by identical PGC retention times of the S-layer glycans of either strain.

Glycosylation sites in the S-layer protein of *L. buchneri* CD034

After an in-gel trypsin digest of the glycosylated S-layer protein SlpB of *L. buchneri* CD034, the glycopeptide (amino acid sequence, $_{147}$SASASSASSASSAEQTTALTDAQK$_{170}$) was analyzed by LC-ESI-MS. A major peak in the spectrum at an m/z value of 6828.51 was found to correspond to the sum of the mass of the amino acids and 28 glucose residues (Fig. 4a). The mass accuracy of this peak compared to the

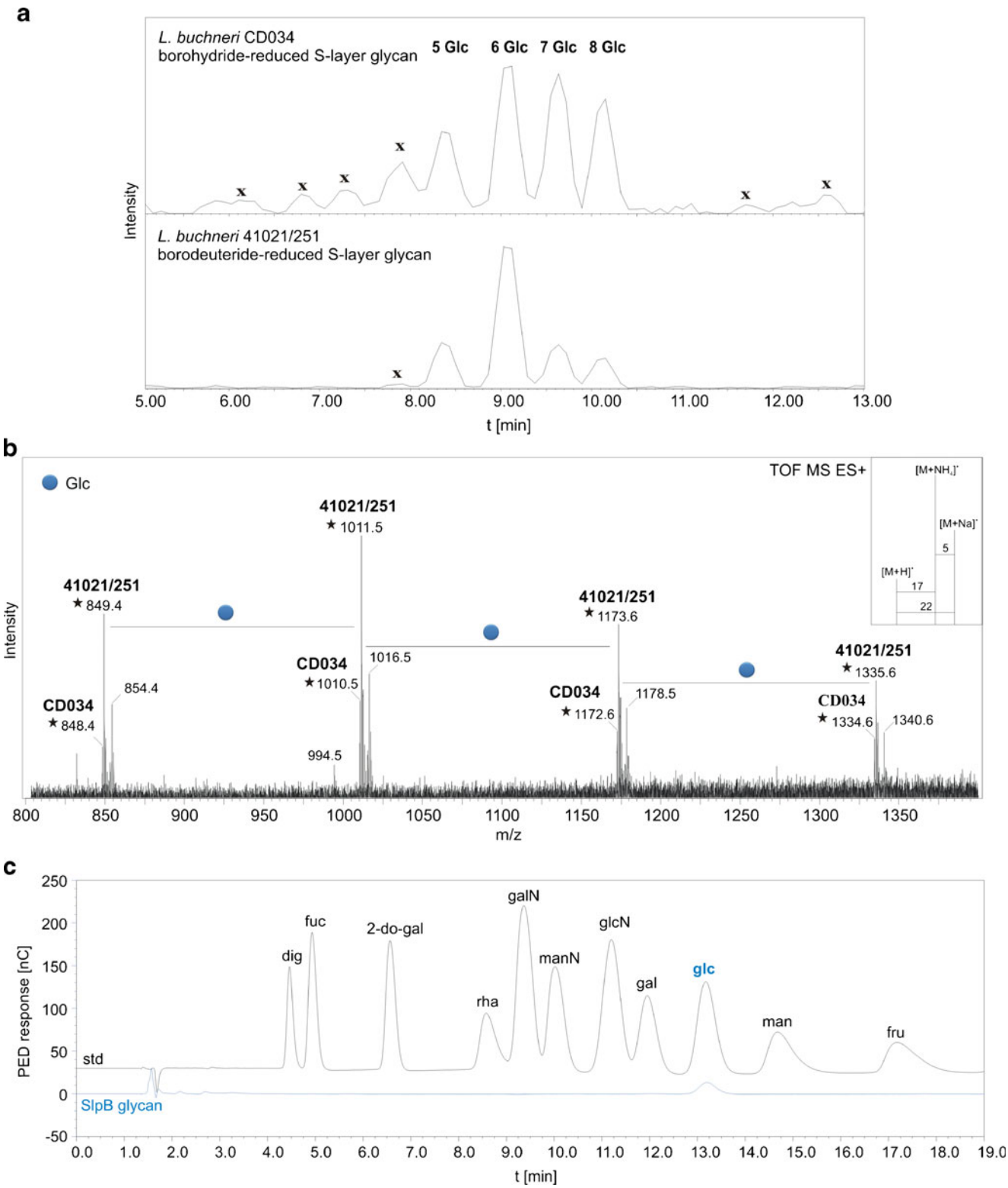


Fig. 3 Comparison of S-layer glycan profiles of *L. buchneri* CD034 and *L. buchneri* 41021/251. Glycans were released from the protein backbone *via* reductive β-elimination using $NaBH_4$ and $NaBD_4$ for *L. buchneri* strain CD034 and strain 41021/251, respectively. **a** LC profiles showing identical elution times of both glycans. Sample impurities are indicated with x. **b** Deconvoluted ESI-TOF-MS spectrum of released S-layer glycans from both strains premixed in one single run. Due to the use of deuterium for reduction of the *L. buchneri* 41021/251 S-layer glycan, a mass difference of 1 Da compared to the ^{1}H-reduced glycan of *L. buchneri* CD034 was observed. As shown in (**b**), glycans from both strains could therefore be analyzed in one single measurement. Adducts are according to ammonium and sodium as depicted in the inset. Molecular ions indicated with an asterisk are $[M+NH_4]^+$. Peaks representing repeating glucose units were observed for both strains. **c** HPAEC-PED profile of SlpB glycan monosaccharides after TFA hydrolysis

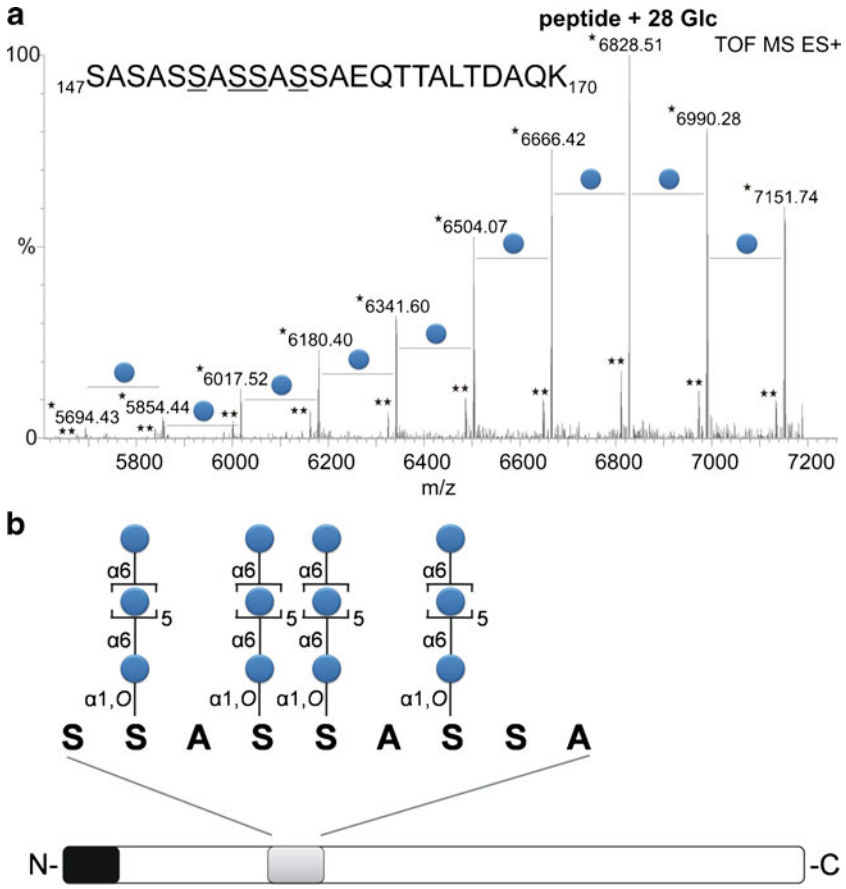


Fig. 4 **a** Mass distribution of the glycosylated peptide of the S-layer protein of *L. buchneri* CD034. After a tryptic digest the glycopeptide (amino acid sequence given in the figure) was applied to LC-ESI-MS. The m/z values correspond to the sum of the mass of the amino acids and the glucoses attached to the peptide. The major peak at an m/z value of 6828.51 refers to the mass of 28 glucoses distributed over the peptide, with a calculated m/z value of 6828.57. $[M+H+2NH_3]^+$ ions and $[M+NH_4]^+$ ions are indicated by an asterisk and a double asterisk, respectively. Derivatization of glycosylated serines *via* ammonia based non-reductive β-elimination and ESI-TOF-MS/MS analysis (Fig. 5) revealed the presence of four glycosylation sites (*underlined*). **b** Schematic drawing of the S-layer glycoprotein of *L. buchneri* CD034. Based on glycan, glycopeptide as well as protein analysis, the S-layer glycoprotein possesses a signal peptide (*black*) consisting of 31 amino acids and an *O*-glycan consisting of on average seven Glc(α1–6) units attached to four serines within the acceptor sequence motif S-S-A-S-S-A-S-S-A (*grey*), corresponding to positions Ser_{152}, Ser_{154}, Ser_{155}, and Ser_{157}

calculated m/z value of 6828.57 is −8.79 ppm. The whole tryptic sample was further digested with GluC and the smallest possible glycopeptide was fractionated by RP-HPLC. An aliquot of the vacuum-dried glycopeptide was subjected to ammonia-based non-reductive β-elimination [66]. During that procedure, glycosylated serines were transformed into ammonia-derivatives leading to a mass decrease of 1 Da compared to non-glycosylated serines. The deglycosylated peptide was further analyzed by RP-ESI-TOF-MS/MS (Fig. 5, Table 1). Four serine residues (underlined) within the sequence S-$\underline{S}_{152}$-A-$\underline{S}_{154}$-$\underline{S}_{155}$-A-$\underline{S}_{157}$-S-A could be detected with the above described mass decrease and, therefore, identified as the glycosylation sites on the S-layer protein SlpB of *L. buchneri* CD034 (compare with Fig. 4a). Distributing 28 glucoses, as found in the major glycopeptide peak, over the four glycosylation sites resulted in an *O*-glycan consisting of, on average, seven glucose residues. A schematic drawing of the S-layer glycoprotein of *L. buchneri* CD034 including the glycosylation sequence motif and the *O*-glycan structure is given in Fig. 4b.

Cellular localization of S-layer protein glycosylation in *L. buchneri* CD034

To determine the subcellular localization of the S-layer glycosylation event, (glyco)proteins were isolated from the cell wall and the cytosol, and Western-immunoblotting for S-layer detection as well as PAS staining for glycan detection was performed. While Western-immunoblotting with SlpB-specific antiserum confirmed the presence of the S-layer protein in both the cell wall and the cytosolic fractions, only the cell wall fraction exhibited strong reactivity with the PAS reagent, the cytosolic fraction, in contrast, showed no significant reactivity (Fig. 6). This observation supports our hypothesis of a sequential order of S-layer protein export to the cell surface and glycosylation.

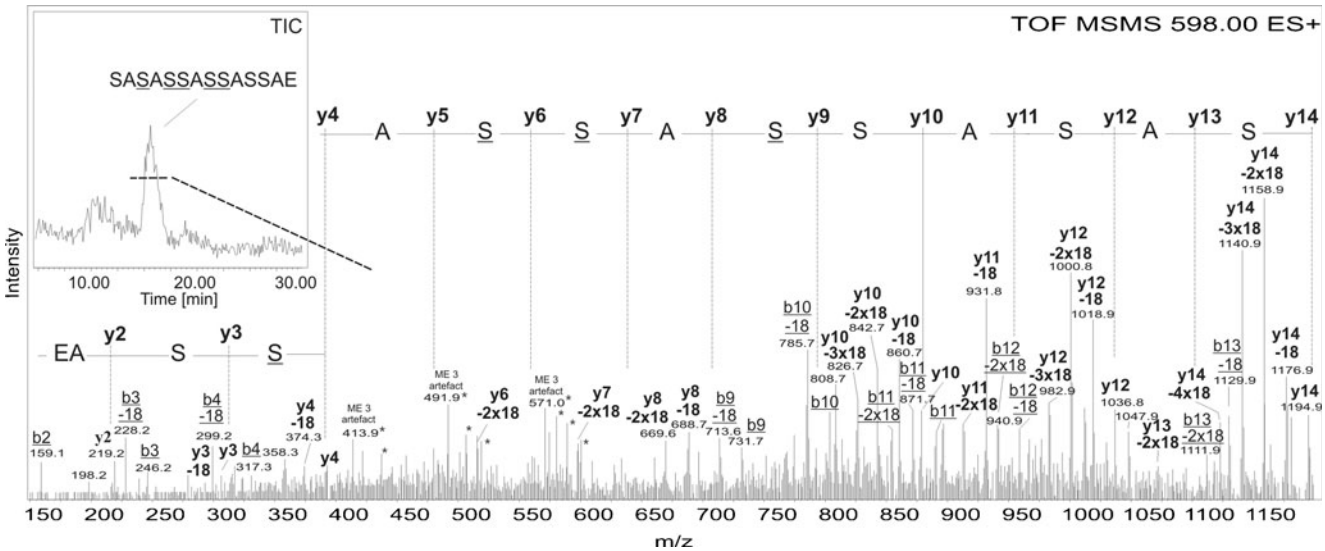


Fig. 5 Mass spectrometric determination of the *O*-glycosylation sites on the S-layer protein of *L. buchneri* CD034. The glycopeptide was subjected to RP-ESI-TOF-MS/MS after ammonia-based non-reductive β-elimination. As depicted in the inset, the deglycosylated peptide eluted at a high percentage of acetonitrile (>30 %) as a broad peak. Mainly the B- and Y- series ions were found in the MS/MS spectrum. All ions in the spectrum, if not indicated elsewise, are $[M+H]^+$. Loss of 18 Da was deteceted for each peak originating from the Ser residues, most probably, due to loss of OH-groups. Deglycoslyted Ser was found with 86 Da (−1 Da)

Table 1 MS/MS peak identification and *O*-glycosylation site determination. Red colored masses were detected after the MS/MS experiment of the deglycosylated peptide. The series of minus 18 Da originates, most probably, from the loss of the serine OH groups. Glycosylation sites, indicated in bold, were found with 86 Da. Mass deviation was found to be −0.5 Da for larger masses

mi	AA	loss of 18			B series	Y series	loss of 18		
87	S	34.0	52.0	70.0	88	1195.5	1177.5	1159.5	1141.5
71	A	105.1	123.1	141.1	159.1	1108.5	1090.5	1072.5	1054.5
87	S	192.1	210.1	228.1	246.1	1037.4	1019.4	1001.5	983.4
71	A	263.1	281.1	299.1	317.1	950.4	932.4	914.4	896.4
87	S	350.2	368.2	386.2	404.2	879.4	861.4	843.4	825.4
86	**S**	436.2	454.2	472.2	490.2	792.3	774.3	756.3	738.3
71	A	507.2	525.2	543.2	561.2	706.3	688.3	670.3	652.3
86	**S**	593.3	611.3	629.3	647.3	635.3	617.3	599.3	581.3
86	**S**	679.3	697.3	715.3	733.3	549.2	531.2	513.2	495.2
71	A	750.3	768.3	786.3	804.3	463.2	445.2	427.2	409.2
86	**S**	836.4	854.4	872.4	890.4	392.2	374.2	356.2	338.2
87	S	923.4	941.4	959.4	977.4	306.1	288.1	270.1	252.1
71	A	994.4	1012.4	1030.4	1048.4	219.1	201.1	183.1	165.1
129	E	1123.5	1141.5	1159.5	1177.5	148.1	130.1	112.1	94.1

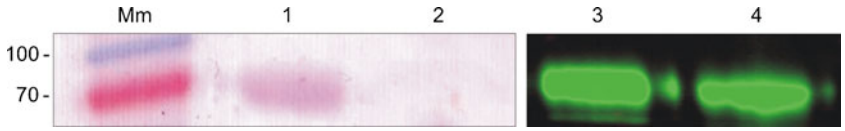


Fig. 6 Cellular localization of protein glycosylation and in *L. buchneri* CD034. *L. buchneri* proteins were isolated from the cell wall (lanes 1 and 3) and the cytosolic fraction (lanes 2 and 4) of *L. buchneri* CD034 as described in Materials and Methods. Samples were run on SDS-PAGE (15 % gel), PAS-stained for carbohydrates (lanes 1 and 2) and Western-blotted and detected with S-layer specific antiserum (lanes 3 and 4). Mm, PageRuler prestained protein ladder (Thermo Scientific)

Identical S-layer protein glycosylation among *L. buchneri* strains and determination of a common S-layer glycosylation motif

Based on the detected commonalities of *O*-glycosylation among S-layer proteins in *L. buchneri* CD034 and *L. buchneri* 41021/251, protein glycosylation analysis was extended to strain *L. buchneri* NRRL B-30929 [53] in order to potentially unravel a general S-layer protein glycosylation theme in this bacterial species. *In silico* proteome analysis of *L. buchneri* NRRL B-30929 revealed the presence of two gene products containing an S-S-A-rich region potentially representing a glycosylation motif. These are Lbuc_0578 and Lbuc_1557 containing the amino acid sequences $_{151}$S-S-A-S-S-A-S-S-A$_{159}$ and $_{149}$S-S-S-A-S-S-A-S$_{156}$, respectively. Comparable glycoprotein patterns were observed in crude cell extracts of *L. buchneri* strains NRRL B-30929 and CD034 according to SDS-PAGE analysis (Fig. 7a), revealing two glyco-positive protein bands and with apparent molecular masses of ~65 kD and 42 kDa, respectively. The prominent glycoprotein band at ~65 kDa (indicated with an arrow in Fig. 7a, lane 2) was identified as the S-layer protein SlpN (gene Lbuc_1557) of *L. buchneri* NRRL B-30929 after in-gel trypsin digest and MALDI-MS/MS analysis (data not shown). SlpN contains an N-terminal signal peptide including a type I signal peptidase cleavage site and possesses a basic pI of 9.9, which all compares well to SlpB of *L. buchneri* CD034.

For glycan analysis, the band corresponding to glycosylated SlpN was subjected to LC-ESI-MS after in-gel reductive β-elimination. *O*-Glycans, identical to the S-layer glucose homooligomer of strains CD034 and 41021/251 could be identified on the S-layer of *L. buchneri* NRRL B-30929 (Fig. 7b). Based on this observation, the S-S-A-rich sequences of all identified S-layer glycoproteins were aligned revealing the presence of the confirmed glycosylation motif S-A-S-S-A-S of the S-layer protein SlpB of *L. buchneri* CD034 (Fig. 7c).

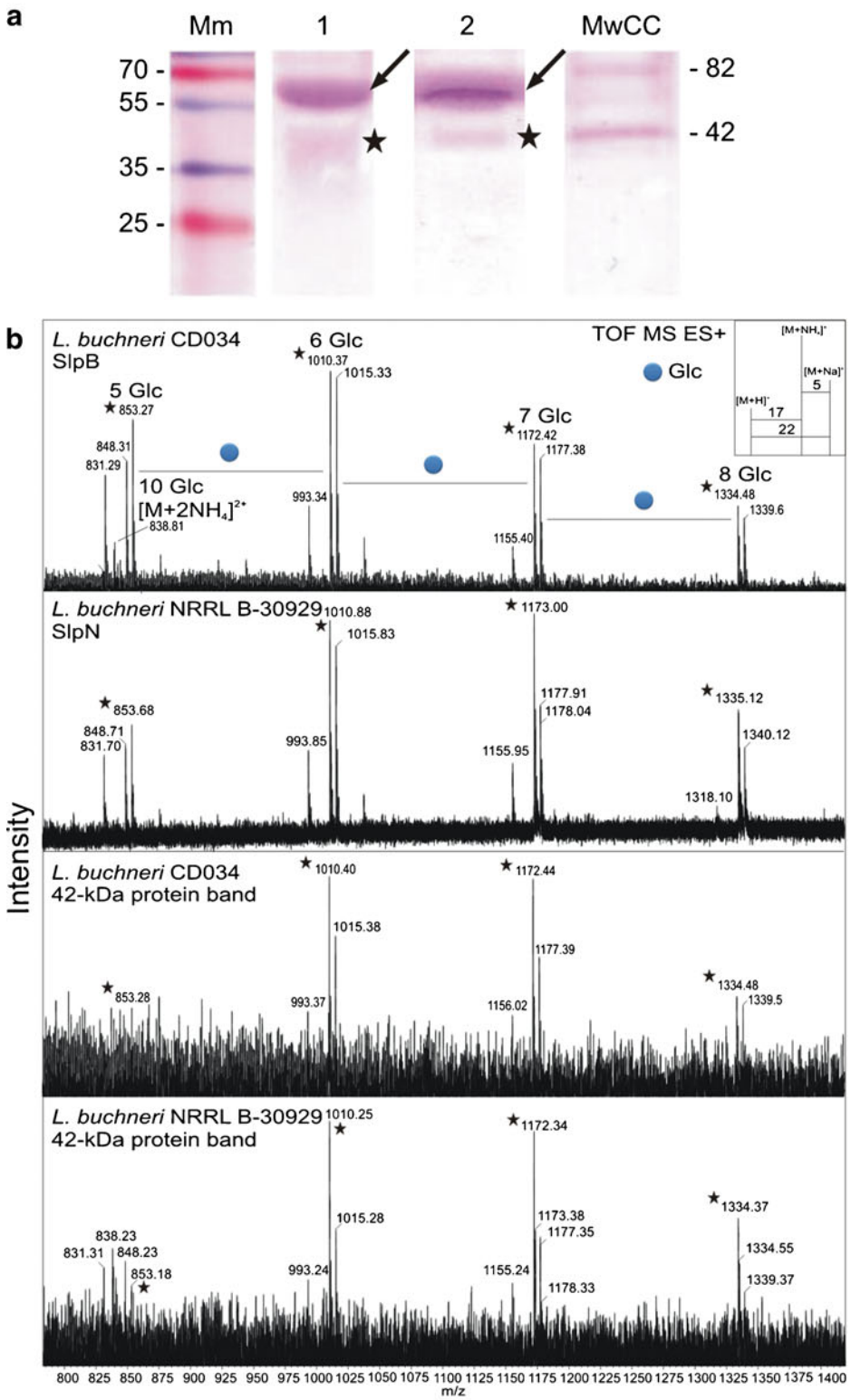
Indication of a putative glycosyl-hydrolase in *L. buchneri* possessing the S-layer glycosylation theme

Glycans from the ~42-kDa glyco-positive protein band of *L. buchneri* strains CD034 and NRRL B-30929 on SDS-PAGE (indicated with asterisks in Figs. 2 and 7a) were released from the protein backbone by in-gel borohydride-reductive β-elimination and subjected to LC-ESI-MS. Glycan profiles identical to that of the S-layer glycans were detected, showing an oligosaccharide consisting of hexose residues (Fig. 7b).

In silico proteome analysis of *L. buchneri* CD034 and NRRL B-30929 indicated the presence of two further proteins, besides the S-layer proteins, containing the characteristic S-A-S-S-A-S glycosylation motif (Fig. 7c). These are the putative extracellular glycosyl-hydrolase homologues *Lb*GH25B (gene LBUCD034_0240) and *Lb*GH25N (gene Lbuc_0200) of *L. buchneri* CD034 and NRRL B-30929, respectively, both containing a glycosyl-hydrolase family 25 domain. Similar to the S-layer proteins, the highly homologous putative glycosyl-hydrolyses *Lb*GH25B and *Lb*GH25N contain an N-terminal, 29-amino acid signal peptide including a type I signal peptidase cleavage site and possess a basic pI of 9.5.

Western-immunoblot analysis of crude cell extracts from *L. buchneri* CD034 and *L. buchneri* NRRL B-30929 using glycosyl-hydrolase specific antiserum detected both *Lb*GH25B and *Lb*GH25N at an apparent molecular mass of ~42 kDa (Fig. 8, lane 1 and 3) which is in accordance with the apparent molecular mass of the additional glyco-positive protein band as inferred from SDS-PA gels (Fig. 8a). Further, purified recombinant, non-glycosylated *Lb*GH25B and *Lb*GH25N were applied to the immunoblots, revealing a slightly downshifted apparent molecular mass when compared to the native glycosyl-hydrolases (Fig. 8, lanes 2 and 4). These data supported a posttranslational modification of the glycosyl-hydrolases and confirmed the identities of the 42-kDa bands as *Lb*GH25B and *Lb*GH25N. As evidenced by

Fig. 7 Comparative analysis of protein glycosylation in *L. buchneri* ▶ strains. **a** Crude extracts of *L. buchneri* CD034 (lane 1) and *L. buchneri* NRRL B-30929 (lane 2) were run on SDS-PAGE (20 % gels) and PAS-stained for carbohydrates. The arrow and the asterisk indicate the S-layer proteins and the 42-kDa glyco-positive band, respectively. Mm, PageRuler Plus prestained protein ladder (Thermo Scientific); MmCC, Candy Cane Glycoprotein Molecular Weight Standards (Life Technologies). **b** Deconvoluted ESI-TOF-MS spectra of glycans released from *L. buchneri* CD034 and NRRL B-30929 glycoprotein bands after in-gel β-elimination in comparison with *O*-glycans of the S-layer protein SlpB of *L. buchneri* CD034. *O*-glycans from all proteins were found to be identical. Molecular ions indicated with an asterisk are $[M+NH_4]^+$. **c** Confirmed (SlpB) and predicted glycosylation sites of *O*-glycans in *L. buchneri* strains. The common sequence motif S-A-S-S-A-S (yellow background) was found in all investigated proteins


a
Mm
1
2
MwCC
70 -
55 -
35 -
25 -
- 82
- 42
b
L. buchneri CD034
SlpB
5 Glc
853.27
6 Glc
1010.37
1015.33
7 Glc
1172.42
1177.38
8 Glc
1334.48
1339.6
10 Glc
[M+2NH4]2+
838.81
848.31
831.29
993.34
1155.40
TOF MS ES+
Glc
[M+NH4]+
[M+Na]+
[M+H]+
5
17
22
L. buchneri NRRL B-30929
SlpN
1010.88
1015.83
1173.00
853.68
848.71
831.70
993.85
1177.91
1178.04
1335.12
1340.12
1155.95
1318.10
L. buchneri CD034
42-kDa protein band
1010.40
1172.44
1015.38
1177.39
853.28
993.37
1334.48
1339.5
1156.02
L. buchneri NRRL B-30929
42-kDa protein band
1010.25
1172.34
1015.28
1173.38
1177.35
1334.37
838.23
831.31
848.23
853.18
993.24
1155.24
1178.33
1334.55
1339.37
Intensity
800 825 850 875 900 925 950 975 1000 1025 1050 1075 1100 1125 1150 1175 1200 1225 1250 1275 1300 1325 1350 1375 1400
m/z
c
S-layer proteins
SlpB
SlpN
Slp L. buchneri 41021/251 (primary structure unkown)
Glycosyl hydrolases
LbGH25B
LbGH25N
147SASASSASSASSAEQT162
146ANTSSSASSASINTQG161
SSASSASSA
248NANSSSASSASSASSS263
248NANSSSASSASSESSS263

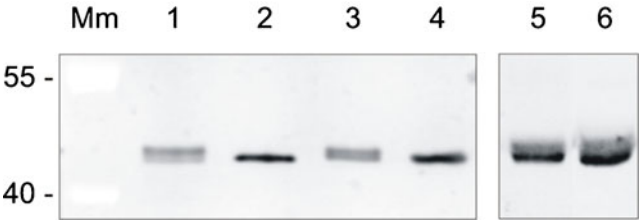


Fig. 8 Apparent mass differences between native and recombinant glycosyl-hydrolases. *L. buchneri* CD034 and NRRL B-30929 cell extracts (lanes 1 and 3), recombinant *Lb*GH25B and *Lb*GH25N (lanes 2 and 4), premixed *L. buchneri* CD034 cell extract and recombinant *Lb*GH25B (lane 5) and premixed *L. buchneri* NRRL B-30929 and recombinant *Lb*GH25N were run on SDS-PAGE (9 % gel), Western-blotted and detected with glycosyl-hydrolase specific antiserum. Mm, PageRuler prestained protein ladder (Thermo Scientific)

LC-ESI-MS the modification would be glycans identical to the *O*-glycans attached to the S-layer proteins.

Closer inspection of the SDS-PA gels revealed that the native glycosyl-hydrolases appear as double bands, with the band of lower molecular mass running at the same size as the recombinant enzyme (Fig. 8, lanes 1–4), indicative of the presence of non-glycosylated *Lb*GH25B and *Lb*GH25N in addition to the glycosylated enzymes in the native *L. buchneri* hosts. The only slight shift in molecular mass between glycosylated (upper band) and non-glycosylated (lower band) glycosyl-hydrolases can be well explained by a difference of only ~1.3 kDa resulting from the 8-unit glucose oligomer. To exclude that matrix effects would cause the different running behavior of the different *Lb*GH25B and *Lb*GH25N forms, *i.e.*, native, glycosylated *versus* recombinant, non-glycosylated, the recombinant, purified enzymes were mixed with the *L. buchneri* cell extracts prior to application onto the SDS-PA gel. On a Western blot, the double bands of both *Lb*GH25B and *Lb*GH25N could be still clearly detected by the glycosyl-hydrolase-specific antiserum, with the lower bands now appearing more intensive due to the contribution of the recombinant proteins migrating at the same size (Fig. 8, lane 5 and 6). This supported our assumption that *Lb*GH25B and *Lb*GH25N are glycoproteins with a certain fraction of the native enzyme being non-glycosylated.

Discussion

We report the presence of an S-layer with oblique lattice symmetry covering *L. buchneri* CD034 cells. We identified the S-layer protein SlpB as most abundant protein of the whole cellular proteome and as the LBUCD034_1608 translation product, one out of eight open reading frames on the genome of *L. buchneri* CD034 encoding putative S-layer proteins [52]. Furthermore, we provided evidence of S-layer protein glycosylation by using a combined biochemical and mass spectrometric approach. It was demonstrated that the S-layer protein of *L. buchneri* CD034 carries *O*-glycans composed of homooligomers of, on average, seven Glc(α1-6) residues. The glycosylation sites on the SlpB protein were identified as four serines (underlined) within the sequence S-S$_{152}$-A-S$_{154}$-S$_{155}$-A-S$_{157}$-S-A by ammonia-based non-reductive β-elimination combined with ESI-MS/MS.

These results concur with the finding that protein glycosylation in lactobacilli occurs preferentially at domains rich in alanine, serine and also threonine residues (AST domain), as it was shown for the major autolysin Acm2 and hypothesized for several other cell wall associated proteins in *L. plantarum* [32, 33, 71]. A functional role of AST domains, however, is still unexplored. Bacterial *O*-glycans containing glucose as the only carbohydrate constituent have so far only been documented for our reference strain *L. buchneri* 41021/251 [26] and have been suggested for an *N*-acetylmuramylhydrolase in the lactic acid bacterium *Streptococcus faecium* [31]. Besides that, several protein *O*-glycans were reported in lactobacilli so far, including mannose-containing structures of *L. rhamnosus* [34] and GlcNAc-containing structures of *L. plantarum* [32] as well as other glycans in *L. acidophilus* showing ConA lectin reactivity [28] and in *L. kefir* according to PAS staining [29]. However, none of these glycans has been structurally elucidated.

Concerning the cellular topology of the S-layer protein glucosylation process, our data on the cellular localization of the S-layer glycoprotein of *L. buchneri* CD034 reveal a sequential order between protein export to the cell surface and glycosylation. This corroborates previous observations indicating that (i) the S-layer glycosylation process occurs on the external face of the cytoplasmic membrane [72–74], and (ii) S-layer glycosylation is lagging behind S-layer protein biosynthesis, evidenced by the presence of non-glycosylated SlpB in the cytosol. Similar results concerning the cellular topology of protein glycosylation in lactobacilli were obtained for protein *O*-glycosylation in *L. rhamnosus* GG, where the secreted protein Msp1 was shown to be only glycosylated in the supernatant fraction [34].

We further demonstrated that the discovered S-layer glycosylation theme is also valid for the strain *L. buchneri* NRRL B-30929. The common glycosylation motif S-A-S-S-A-S which is part of the seven amino-acid stretch S-S-A-S-S-A-S was identified within the investigated S-layer protein SlpN of *L. buchneri* NRRL B-30929. Also, S-layer *O*-glucans identical to those attached to SlpB could be identified on SlpN glycoprotein band after separation of a crude cell extract on SDS-PAGE. These findings suggest the presence of species-wide S-layer protein *O*-glucosylation within the sequence motif S-A-S-S-A-S in *L. buchneri* strains.

There are strong indications that the S-layer glycosylation theme can be extended to another *L. buchneri* protein. The putative glycosyl-hydrolases *Lb*GH25B and *Lb*GH25N of *L. buchneri* CD034 and NRRL B-30929, respectively, were

found to be the only proteins in the proteome of either strain, in addition to the S-layer proteins, that contain the characteristic sequence motif S-A-S-S-A-S. Compared to the S-layer proteins, these enzymes are of markedly lower abundance according to CBB- and PAS-stained SDS-PAGE and Western-immunoblotting. The "S-layer" glucose oligomer could be unambiguously detected by mass spectrometry on the corresponding glyco-positive protein band of either *L. buchneri* strain. This, together with the observed migration difference between native and recombinant *Lb*GH25B and *Lb*GH25N detected with glycosyl-hydrolase-specific antiserum on a Western-immunoblot, supports the posttranslational modification of the glycosyl-hydrolases by glucosylation. The scenario of different proteins of a bacterium carrying the "S-layer glycan" is not new; it has been shown recently for the oral pathogen *Tannerella forsythia* [19]. We would like to note that due to the low abundance of *Lb*GH25B and *Lb*GH25N in *L. buchneri* cell extracts, a glycoproteomics approach using glycopeptides, as has been performed with the S-layer glycoprotein SlpB, has not been not successful, so far. For this purpose, homologous overexpression of the glycosyl-hydrolases is currently under way in our laboratory.

*Lb*GH25B and *Lb*GH25N, both, contain a glycosyl-hydrolase family 25 domain, which predicts them to function as bacterial cell wall hydrolases. Several reports on glycosylated extracellular cell wall-hydrolyzing enzymes in lactic acid bacteria are in line with our finding of *Lb*GH25B and *Lb*GH25N being glycosylated. It is interesting to note that the first example of a proven glycosylated enzyme in bacteria was a glycosyl-hydrolyzing enzyme, namely the endogenous, autolytic *N*-acetylmuramoyl hydrolase of *Streptococcus faecium* [31]. Glycosylation of AcmB, the peptidoglycan-hydrolyzing *N*-acetylglucosaminidase of *Lactococcus lactis*, was proposed based on the presence of potential *O*- and *N*-glycosylation sites [75]. *O*-glycans attached to the N-terminal AST domain were found on the extracellular autolysin Acm2 of *Lactobacillus plantarum* [32] and also *Lactobacillus rhamnosus* expresses *O*-glycosylated Msp1 (major secreted protein), a cell wall hydrolase [34, 76].

The glycosyltransferases involved in the protein glucosylation pathway of *L. buchneri* strains are currently unknown. Especially, *O*-oligosaccharyltransferases (*O*-OTases) as key modules of many glycosylation systems display relatively low homology and typically have rather broad substrate specificities [77, 78] which makes their identification challenging. A high number of transmembrane spanning domains of the putative glycosyl transferases and homologous gene products LBUCD034_2252 and Lbuc_2151 as well as LBUCD034_0503 and Lbuc_0464 of *L. buchneri* CD034 and NRRL B-30929, respectively, make them possible *O*-OST candidates responsible for coupling of the glucose oligomer to specific serine residues on the target proteins. More detailed studies are needed to identify involved enzymes and to obtain a deeper understanding of the protein *O*-glucosylation machinery in *L. buchneri*.

The lactobacillar glycoproteins investigated in this study, *i.e.*, the S-layer proteins SlpB and SlpN as well as the enzymes *Lb*GH25B and *Lb*GH25N, share common properties; this includes a basic pI, which is a common trait of S-layer proteins of lactobacilli [30, 70], the presence of the S-A-S-S-A-S motif, and a type I signal peptidase cleavage site required for protein export across the cytoplasmic membrane, which supports that these glycoproteins are extracellularly located. Glycosylation of extracellular proteins in *L. buchneri* might possibly facilitate adherence to solid substrates, support cell-cell interaction and contribute as cellular sugar coat to the protection against environmental stresses and proteolytic degradation.

In conclusion, we provided evidence of species wide S-layer *O*-glucosylation in *L. buchneri* strains within a common, serine rich sequence motif and provided strong indications that *L. buchneri* glycosyl-hydrolases undergo the same *O*-glucosylation process. These findings pinpoint new opportunities for glyco-engineering of customized glycans in a Gram-positive, beneficial and safe bacterial organism.

Acknowledgments We thank Siqing Liu (U.S. Department of Agriculture, Agricultural Research Service, Renewable Product Technology Research Unit, University of Illinois, USA) for kindly providing *L. buchneri* NRRL B-30929, and Andrea Scheberl and Sonja Zayni for excellent technical assistance.

Financial support came from the Austrian Science Fund FWF, projects P21954-B20 (to C.S.) and P24305-B20 (to P.M.), the PhD programme "BioToP - Biomolecular Technology of Proteins" (Austrian Science Fund, FWF project W1224), the Hochschuljubiläumsstiftung der Stadt Wien, project H-2442/2012 (to J.A.), and the Christian Doppler Laboratory for Genetically Engineered Lactic Acid Bacteria (to R.G.).

References

1. Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E.: Essentials of glycobiology, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Habor (2009). 484 pages
2. Upreti, R.K., Kumar, M., Shankar, V.: Bacterial glycoproteins: functions, biosynthesis and applications. Proteomics **3**, 363–379 (2003)
3. Schmidt, M.A., Riley, L.W., Benz, I.: Sweet new world: glycoproteins in bacterial pathogens. Trends Microbiol. **11**, 554–561 (2003)
4. Hitchen, P.G., Dell, A.: Bacterial glycoproteomics. Microbiology **152**, 1575–1580 (2006)
5. Nothaft, H., Szymanski, C.M.: Protein glycosylation in bacteria: sweeter than ever. Nat. Rev. Microbiol. **8**, 765–778 (2010)
6. Messner, P.: Bacterial glycoproteins. Glycoconj. J. **14**, 3–11 (1997)
7. Messner, P., Sleytr, U.B.: Bacterial surface layer glycoproteins. Glycobiology **1**, 545–551 (1991)
8. Ristl, R., Steiner, K., Zarschler, K., Zayni, S., Messner, P., Schäffer, C.: The S-Layer glycome—adding to the sugar coat of bacteria. Int. J. Microbiol. **2011**, 127870 (2011)

9. Messner, P., Steiner, K., Zarschler, K., Schäffer, C.: S-layer nanoglycobiology of bacteria. Carbohydr. Res. **343**, 1934–1951 (2008)
10. Hartley, M.D., Morrison, M.J., Aas, F.E., Børud, B., Koomey, M., Imperiali, B.: Biochemical characterization of the O-linked glycosylation pathway in *Neisseria gonorrhoeae* responsible for biosynthesis of protein glycans containing *N*, *N′*-diacetylbacillosamine. Biochemistry **50**, 4936–4948 (2011)
11. Logan, S.M.: Flagellar glycosylation—a new component of the motility repertoire? Microbiology **152**, 1249–1262 (2006)
12. Dobos, K.M., Swiderek, K., Khoo, K.-H., Brennann, P.J., Belisle, J.T.: Evidence for glycosylation sites on the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. Infect. Immun. **63**, 2846–2853 (1995)
13. Karlyshev, A.V., Everest, P., Linton, D., Cawthraw, S., Newell, D.G., Wren, B.W.: The *Campylobacter jejuni* general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks. Microbiology **150**, 1957–1964 (2004)
14. Szymanski, C.M., Wren, B.W.: Protein glycosylation in bacterial mucosal pathogens. Nat. Rev. Microbiol. **3**, 225–237 (2005)
15. Szymanski, C.M., Yao, R., Ewing, C.P., Trust, T.J., Guerry, P.: Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. Mol. Microbiol. **32**, 1022–1030 (1999)
16. Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, J.S., Panico, M., Morris, H.R., Dell, A., Wren, B.W., Aebi, M.: *N*-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. Science **298**, 1790–1793 (2002)
17. Ku, S.C., Schulz, B.L., Power, P.M., Jennings, M.P.: The pilin *O*-glycosylation pathway of pathogenic *Neisseria* is a general system that glycosylates AniA, an outer membrane nitrite reductase. Biochem. Biophys. Res. Comm. **378**, 84–89 (2009)
18. Fletcher, C.M., Coyne, M.J., Villa, O.F., Chatzidaki-Livanis, M., Comstock, L.E.: A general *O*-glycosylation system important to the physiology of a major human intestinal symbiont. Cell **137**, 321–331 (2009)
19. Posch, G., Pabst, M., Brecker, L., Altmann, F., Messner, P., Schäffer, C.: Characterization and scope of S-layer protein *O*-glycosylation in *Tannerella forsythia*. J. Biol. Chem. **286**, 38714–38724 (2011)
20. Giraffa, G., Chanishvili, N., Widyastuti, Y.: Importance of lactobacilli in food and feed biotechnology. Res. Microbiol. **161**, 480–487 (2010)
21. Helanto, M., Kiviharju, K., Leisola, M., Nyyssölä, A.: Metabolic engineering of *Lactobacillus plantarum* for production of L-ribulose. Appl. Environ. Microbiol. **73**, 7083–7091 (2007)
22. Peterbauer, C., Maischberger, T., Haltrich, D.: Food-grade gene expression in lactic acid bacteria. Biotechnol. J. **6**, 1147–1161 (2011)
23. Konings, W.N., Kok, J., Kuipers, O.P., Poolman, B.: Lactic acid bacteria: the bugs of the new millennium. Curr. Opin. Microbiol. **3**, 276–282 (2000)
24. Solá, R.J., Griebenow, K.: Glycosylation of therapeutic proteins. BioDrugs **24**, 9–21 (2010)
25. Wells, J.M.: Immunomodulatory mechanisms of lactobacilli. Microb. Cell Fact. **10**, S17 (2011)
26. Möschl, A., Schäffer, C., Sleytr, U.B., Messner, P., Christian, R., Schulz, G.: Characterization of the S-layer glycoproteins of two lactobacilli. In: Beveridge, T.J., Koval, S.F. (eds.) Advances in bacterial paracrystalline surface layers, vol. 252, pp. 281–284. Plenum Press, New York (1993)
27. Mozes, N., Lortal, S.: X-ray photoelectron spectroscopy and biochemical analysis of the surface of *Lactobacillus helveticus* ATCC 12046. Microbiology **141**, 11–19 (1995)
28. Konstantinov, S.R., Smidt, H., de Vos, W.M., Bruijns, S.C.M., Singh, S.K., Valence, F., Molle, D., Lortal, S., Altermann, E., Klaenhammer, T.R., van Kooyk, Y.: S layer protein A of *Lactobacillus acidophilus* NCFM regulates immature dendritic cell and T cell functions. Proc. Natl. Acad. Sci. U. S. A. **105**, 19474–19479 (2008)
29. Mobili, P., los Ángeles Serradell, M., Trejo, S.A., Avilés Puigvert, F.X., Abraham, A.G., Antoni, G.L.: Heterogeneity of S-layer proteins from aggregating and non-aggregating *Lactobacillus kefir* strains. Antonie van Leeuwenhoek **95**, 363–372 (2009)
30. Messner, P., Schäffer, C., Egelseer, E., Sleytr, U.: Occurrence, structure, chemistry, genetics, morphogenesis, and functions of S-Layers. In: König, H., Claus, H., Varma, A. (eds.) Prokaryotic cell wall compounds—structure and biochemistry, pp. 53–109. Springer, Berlin (2010)
31. Kawamura, T., Shockman, G.D.: Purification and some properties of the endogenous, autolytic *N*-acetylmuramoylhydrolase of *Streptococcus faecium*, a bacterial glycoenzyme. J. Biol. Chem. **258**, 9514–9521 (1983)
32. Fredriksen, L., Mathiesen, G., Moen, A., Bron, P.A., Kleerebezem, M., Eijsink, V.G.H., Egge-Jacobsen, W.: The major autolysin Acm2 from *Lactobacillus plantarum* undergoes cytoplasmic *O*-glycosylation. J. Bacteriol. **194**, 325–333 (2011)
33. Rolain, T., Bernard, E., Beaussart, A., Degand, H., Courtin, P., Egge-Jacobsen, W., Bron, P.A., Morsomme, P., Kleerebezem, M., Chapot-Chartier, M.P., Dufrene, Y.F., Hols, P.: *O*-glycosylation as a novel control mechanism of peptidoglycan hydrolase activity. J. Biol. Chem. **288**, 22233–22247 (2013)
34. Lebeer, S., Claes, I.J.J., Balog, C.I.A., Schoofs, G., Verhoeven, T.L.A., Nys, K., von Ossowski, I., de Vos, W.M., Tytgat, H.L.P., Agostinis, P., Palva, A., Van Damme, E.J.M., Deelder, A.M., de Keersmaecker, S.C.J., Wuhrer, M., Vanderleyden, J.: The major secreted protein Msp1/p75 is *O*-glycosylated in *Lactobacillus rhamnosus* GG. Microb. Cell Fact. **11**, 15 (2012)
35. Stepper, J., Shastri, S., Loo, T.S., Preston, J.C., Novak, P., Man, P., Moore, C.H., Havlíček, V., Patchett, M.L., Norris, G.E.: Cysteine *S*-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. FEBS Lett. **585**, 645–650 (2011)
36. Venugopal, H., Edwards, P.J.B., Schwalbe, M., Claridge, J.K., Libich, D.S., Stepper, J., Loo, T., Patchett, M.L., Norris, G.E., Pascal, S.M.: Structural, dynamic, and chemical characterization of a novel *S*-glycosylated bacteriocin. Biochemistry **50**, 2748–2755 (2011)
37. Lebeer, S., Verhoeven, T.L.A., Francius, G., Schoofs, G., Lambrichts, I., Dufrene, Y., Vanderleyden, J., De Keersmaecker, S.C.J.: Identification of a gene cluster for the biosynthesis of a long, galactose-rich exopolysaccharide in *Lactobacillus rhamnosus* GG and functional analysis of the priming glycosyltransferase. Appl. Environ. Microbiol. **75**, 3554–3563 (2009)
38. Denou, E., Pridmore, R.D., Berger, B., Panoff, J.M., Arigoni, F., Brussow, H.: Identification of genes associated with the long-gut-persistence phenotype of the probiotic *Lactobacillus johnsonii* strain NCC533 using a combination of genomics and transcriptome analysis. J. Bacteriol. **190**, 3161–3168 (2008)
39. Marco, M.L., de Vries, M.C., Wels, M., Molenaar, D., Mangell, P., Ahrne, S., de Vos, W.M., Vaughan, E.E., Kleerebezem, M.: Convergence in probiotic *Lactobacillus* gut-adaptive responses in humans and mice. ISME J. **4**, 1481–1484 (2010)
40. Heinl, S., Spath, K., Egger, E., Grabherr, R.: Sequence analysis and characterization of two cryptic plasmids derived from *Lactobacillus buchneri* CD034. Plasmid **66**, 159–168 (2011)
41. Spath, K., Heinl, S., Egger, E., Grabherr, R.: *Lactobacillus plantarum* and *Lactobacillus buchneri* as expression systems: evaluation of different origins of replication for the design of suitable shuttle vectors. Mol. Biotechnol. **52**, 40–48 (2012)
42. Liu, S., Skinner-Nemec, K.A., Leathers, T.D.: *Lactobacillus buchneri* strain NRRL B-30929 converts a concentrated mixture of xylose and glucose into ethanol and other products. J. Ind. Microbiol. Biotechnol. **35**, 75–81 (2008)
43. Zeng, X.Q., Pan, D.D., Guo, Y.X.: The probiotic properties of *Lactobacillus buchneri* P2. J. Appl. Microbiol. **108**, 2059–2066 (2010)

44. Radovanovic, R.S., Katic, V.: Influence of lactic acid bacteria isolates on *Staphylococcus aureus* growth in skimmed milk. Bulg. J. Agric. Sci. **15**, 196–203 (2009)
45. Kõll, P., Mändar, R., Smidt, I., Hütt, P., Truusalu, K., Mikelsaar, R.-H., Shchepetova, J., Krogh-Andersen, K., Marcotte, H., Hammarström, L., Mikelsaar, M.: Screening and evaluation of human intestinal lactobacilli for the development of novel gastrointestinal probiotics. Curr. Microbiol. **61**, 560–566 (2010)
46. Danner, H., Holzer, M., Mayrhuber, E., Braun, R.: Acetic acid increases stability of silage under aerobic conditions. Appl. Environ. Microbiol. **69**, 562–567 (2003)
47. Driehuis, F., Oude Elferink, S.J.W.H., Spoelstra, S.F.: Anaerobic lactic acid degradation during ensilage of whole crop maize inoculated with *Lactobacillus buchneri* inhibits yeast growth and improves aerobic stability. J. Appl. Microbiol. **87**, 583–594 (1999)
48. Holzer, M., Mayrhuber, E., Danner, H., Braun, R.: The role of *Lactobacillus buchneri* in forage preservation. Trends Biotechnol. **21**, 282–287 (2003)
49. Kung, L.J., Taylor, C.C., Lynch, M.P., Neylon, J.M.: The effect of treating alfalfa with *Lactobacillus buchneri* 40788 on silage fermentation, aerobic stability, and nutritive value for lactating dairy cows. J. Dairy Sci. **86**, 336–343 (2003)
50. Oude Elferink, S.J.W.H., Krooneman, J., Gottschal, J.C., Spoelstra, S.F., Faber, F., Driehuis, F.: Anaerobic conversion of lactic acid to acetic acid and 1,2-propanediol by *Lactobacillus buchneri*. Appl. Environ. Microbiol. **67**, 125–132 (2001)
51. Schmidt, R.J., Kung, L.J.: The effects of *Lactobacillus buchneri* with or without a homolactic bacterium on the fermentation and aerobic stability of corn silages made at different locations. J. Dairy Sci. **93**, 1616–1624 (2010)
52. Heinl, S., Wibberg, D., Eikmeyer, F., Szczepanowski, R., Blom, J., Linke, B., Goesmann, A., Grabherr, R., Schwab, H., Pühler, A., Schlüter, A.: Insights into the completely annotated genome of *Lactobacillus buchneri* CD034, a strain isolated from stable grass silage. J. Biotechnol. **161**, 153–166 (2012)
53. Liu, S., Leathers, T.D., Copeland, A., Chertkov, O., Goodwin, L., Mills, D.A.: Complete genome sequence of *Lactobacillus buchneri* NRRL B-30929, a novel strain from a commercial ethanol plant. J. Bacteriol. **193**, 4019–4020 (2011)
54. Liu, S., Bischoff, K.M., Hughes, S.R., Leathers, T.D., Price, N.P., Qureshi, N., Rich, J.O.: Conversion of biomass hydrolysates and other substrates to ethanol and other chemicals by *Lactobacillus buchneri*. Lett. Appl. Microbiol. **48**, 337–342 (2009)
55. De Man, J.C., Rogosa, M., Sharpe, M.E.: A medium for the cultivation of lactobacilli. J. Appl. Microbiol. **23**, 130–135 (1960)
56. Sleytr, U.B., Messner, P., Pum, D.: Analysis of crystalline bacterial surface layers by freeze-etching, metal shadowing, negative staining and ultrathin sectioning. Methods Microbiol. **20**, 29–60 (1988)
57. Sekot, G., Posch, G., Oh, Y., Zayni, S., Mayer, H., Pum, D., Messner, P., Hinterdorfer, P., Schäffer, C.: Analysis of the cell surface layer ultrastructure of the oral pathogen *Tannerella forsythia*. Arch. Microbiol. **196**, 525–539 (2012)
58. Laemmli, U.K.: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**, 680–685 (1970)
59. Hart, C., Schulenberg, B., Steinberg, T.H., Leung, W.-Y., Patton, W.F.: Detection of glycoproteins in polyacrylamide gels and on electroblots using Pro-Q Emerald 488 dye, a fluorescent periodate Schiff-base stain. Electrophoresis **24**, 588–598 (2003)
60. Taylor, A.M., Holst, O., Thomas-Oates, J.: Mass spectrometric profiling of *O*-linked glycans released directly from glycoproteins in gels using in-gel reductive β-elimination. Proteomics **6**, 2936–2946 (2006)
61. Packer, N.H., Lawson, M.A., Jardine, D.R., Redmond, J.W.: A general approach to desalting oligosaccharides released from glycoproteins. Glycoconj. J. **15**, 737–747 (1998)
62. Pabst, M., Altmann, F.: Influence of electrosorption, solvent, temperature, and ion polarity on the performance of LC-ESI-MS using graphitic carbon for acidic oligosaccharides. Anal. Chem. **80**, 7534–7542 (2008)
63. Stadlmann, J., Pabst, M., Kolarich, D., Kunert, R., Altmann, F.: Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. Proteomics **8**, 2858–2871 (2008)
64. Lee, Y.: High-performance anion-exchange chromatography for carbohydrate analysis. Anal. Biochem. **189**, 151–162 (1990)
65. Pabst, M., Bondili, J.S., Stadlmann, J., Mach, L., Altmann, F.: Mass + retention time = structure: a strategy for the analysis of *N*-glycans by carbon LC-ESI-MS and its application to fibrin *N*-glycans. Anal. Chem. **79**, 5051–5057 (2007)
66. Hanisch, F.-G., Müller, S.: Approaches to the *O*-glycoproteome. In: The proteomics protocols handbook. pp. 439–457. Humana Press Inc., Totowa, New York (2005)
67. Masuda, K., Kawata, T.: Characterization of a regular array in the wall of *Lactobacillus buchneri* and its reattachment to the other wall components. J. Gen. Microbiol. **124**, 81–90 (1981)
68. Åvall-Jääskeläinen, S., Palva, A.: *Lactobacillus* surface layers and their applications. FEMS Microbiol. Rev. **29**, 511–529 (2005)
69. van Roosmalen, M.L., Geukens, N., Jongbloed, J.D.H., Tjalsma, H., Dubois, J.-Y.F., Bron, S., van Dijl, J.M., Anné, J.: Type I signal peptidases of Gram-positive bacteria. Biochim. Biophys. Acta, Mol. Cell Res. **1694**, 279–297 (2004)
70. Hynönen, U., Palva, A.: *Lactobacillus* surface layer proteins: structure, function and applications. Appl. Microbiol. Biotechnol. **97**, 5225–5243 (2013)
71. Kleerebezem, M., Hols, P., Bernard, E., Rolain, T., Zhou, M., Siezen, R.J., Bron, P.A.: The extracellular biology of the lactobacilli. FEMS Microbiol. Rev. **34**, 199–230 (2010)
72. Steiner, K., Hanreich, A., Kainz, B., Hitchen, P.G., Dell, A., Messner, P., Schäffer, C.: Recombinant glycans on an S-layer self-assembly protein: a new dimension for nanopatterned biomaterials. Small **4**, 1728–1740 (2008)
73. Zarschler, K., Janesch, B., Pabst, M., Altmann, F., Messner, P., Schäffer, C.: Protein tyrosine *O*-glycosylation—a rather unexplored prokaryotic glycosylation system. Glycobiology **20**, 787–798 (2010)
74. Janesch, B., Messner, P., Schäffer, C.: Are the surface layer homology domains essential for cell surface display and glycosylation of the S-layer protein from *Paenibacillus alvei* CCM 2051^{T}? J. Bacteriol. **195**, 565–575 (2012)
75. Huard, C., Miranda, G., Wessner, F., Bolotin, A., Hansen, J., Foster, S.J., Chapot-Chartier, M.-P.: Characterization of AcmB, an *N*-acetylglucosaminidase autolysin from *Lactococcus lactis*. Microbiology **149**, 695–705 (2003)
76. Claes, I.J.J., Schoofs, G., Regulski, K., Courtin, P., Chapot-Chartier, M.-P., Rolain, T., Hols, P., von Ossowski, I., Reunanen, J., De Vos, M.D., Palva, A., Vanderleyden, J., De Keersmaecker, S.C.J., Lebeer, S.: Genetic and biochemical characterization of the cell wall hydrolase activity of the major secreted protein of *Lactobacillus rhamnosus* GG. PLoS One **7**, e31588 (2012)
77. Faridmoayer, A., Fentabil, M.A., Mills, D.C., Klassen, J.S., Feldman, M.F.: Functional characterization of bacterial oligosaccharyltransferases involved in *O*-linked protein glycosylation. J. Bacteriol. **189**, 8088–8098 (2007)
78. Hug, I., Feldman, M.F.: Analogies and homologies in lipopolysaccharide and glycoprotein biosynthesis in bacteria. Glycobiology **21**, 138–151 (2010)

Appendix VIII

Hernandez Bort, J. A., Shanmukam, V., Pabst, M., Windwarder, M., **Neumann, L.**, Alchalabi, A., Krebiehl, G, Koellensperger, G., Hann, S., Sonntag, D., Altmann, F., Heel, C., Borth, N. (2014) Reduced quenching and extraction time for mammalian cells using filtration and syringe extraction. J. Biotechnol., doi: 10.1016/j.jbiotec.2014.04.014.

Journal of Biotechnology 182–183 (2014) 97–103

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Short communication

Reduced quenching and extraction time for mammalian cells using filtration and syringe extraction

Juan A. Hernández Bort[a,1], Vinoth Shanmukam[a,b,1], Martin Pabst[c], Markus Windwarder[c], Laura Neumann[c], Ali Alchalabi[d], Guido Krebiehl[d], Gunda Koellensperger[a,c], Stephan Hann[a,c], Denise Sonntag[d], Friedrich Altmann[a,c], Christine Heel[e], Nicole Borth[a,b,*]

[a] ACIB GmbH, Austrian Centre of Industrial Biotechnology, Vienna, Austria
[b] Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria
[c] Department of Chemistry, University of Natural Resources and Life Sciences, Vienna, Austria
[d] Biocrates Life Sciences, Innsbruck, Austria
[e] Sandoz GmbH, Tirol, Austria

ARTICLE INFO

Article history:
Received 16 October 2013
Received in revised form 8 April 2014
Accepted 11 April 2014
Available online 29 April 2014

Keywords:
Metabolomics
Quenching
Metabolite extraction
Fast filtration
CHO cells

ABSTRACT

In order to preserve the *in vivo* metabolite levels of cells, a quenching protocol must be quickly executed to avoid degradation of labile metabolites either chemically or biologically. In the case of mammalian cell cultures cultivated in complex media, a wash step previous to quenching is necessary to avoid contamination of the cell pellet with extracellular metabolites, which could distort the real intracellular concentration of metabolites. This is typically achieved either by one or multiple centrifugation/wash steps which delay the time until quenching (even harsh centrifugation requires several minutes for processing until the cells are quenched) or filtration.

In this article, we describe and evaluate a two-step optimized protocol based on fast filtration by use of a vacuum pump for quenching and subsequent extraction of intracellular metabolites from CHO (Chinese hamster ovary) suspension cells, which uses commercially available components. The method allows transfer of washed cells into liquid nitrogen within 10–15 s of sampling and recovers the entire extraction solution volume. It also has the advantage to remove residual filter filaments in the final sample, thus preventing damage to separation columns during subsequent MS analysis. Relative to other methods currently used in the literature, the resulting energy charge of intracellular adenosine nucleotides was increased to 0.94 compared to 0.90 with cold PBS quenching or 0.82 with cold methanol/AMBIC quenching.

1. Introduction

In recent years, different sophisticated protocols have been developed for the quantification of intracellular metabolites (Bennett et al., 2008; Buchholz et al., 2001; Dietmair et al., 2010; Neubauer et al., 2012; Pabst et al., 2010; Sellick et al., 2011) due to the increasing interest in more rational metabolic engineering and control tools for the optimization of cell lines and processes (Dietmair et al., 2012a,b; Martínez et al., 2013; Murabito et al., 2009; Schilling et al., 2000; Sheikh et al., 2005). The use of reconstructed genome-scale models requires sufficient and reliable experimental data to predict the cellular needs of high-producer cell lines (Chung et al., 2010; Licona-Cassani et al., 2012; Selvarasu et al., 2012) and to assist with rational prediction of necessary changes in media composition, feeding strategy and process

Abbreviations: ATP, adenosine-5′-triphosphate; ADP, adenosine-5′-diphosphate; AMP, adenosine-5′-monophosphate; UTP, uridine-5′-triphosphate; UDP, uridine-5′-diphosphate; UMP, uridine-5′-monophosphate; GTP, guanosine-5′-triphosphate; GDP, guanosine-5′-diphosphate; GMP, guanosine-5′-monophosphate; UDP-Gal, uridine-5′-diphosphate galactose; UDP-Glc, uridine-5′-diphosphate glucose; UDP-HexNAc, uridine-5′-diphosphate N-acetylhexosamine.

* Corresponding author at: Department of Biotechnology, University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria. Tel.: +43 1 47654 6232; fax: +43 1 47654 6675.
E-mail address: nicole.borth@boku.ac.at (N. Borth).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jbiotec.2014.04.014

control (D'Huys et al., 2012; Martínez et al., 2013; Selvarasu et al., 2009, 2012; Zamorano et al., 2012).

To ensure that the measured data are as close to the real *in vivo* values as possible, an efficient quenching protocol should: (i) ensure a fast and complete blockage of any intracellular metabolic reaction, as most intermediates have high conversion rates (Canelas et al., 2008; de Koning and van Dam, 1992; Weibel et al., 1974), (ii) avoid contamination of the sample with extracellular metabolites present in the supernatant, (iii) lose none of the intracellular substances through leaky membranes. It has proven very difficult to achieve all of these requirements to perfection, so that most protocols try to keep a reasonable balance of preventing all three reasons of metabolite loss or contamination.

Several authors (Pabst et al., 2010; Sellick et al., 2009) designed quenching protocols in which the cell broth was directly quenched using cold solvent mixtures after sampling, as their lower freezing points allow using them below −20 °C. Even though these methods were successfully implemented for yeast and bacteria (de Jonge et al., 2012; Gonzalez et al., 1997; Moritz et al., 2000) their direct application to mammalian cells provokes some controversies due to higher leakage of intracellular metabolites, as mammalian cells lack a cell wall. Sellick (Sellick et al., 2009) and later Sengupta (Sengupta et al., 2011) used different additives to the cold methanol quenching solvent showing an increase in metabolite recovery. However, these results proved difficult to transfer between labs (Dietmair et al., 2010) and indeed, as expected, resulted in membrane damage and consequently uncontrolled leakage of intracellular metabolites.

A less aggressive quenching method uses direct dilution of the cell broth with ice-cold PBS (Pabst et al., 2010), which prevents leakage while at the same time diluting extracellular contaminants, however the final mixture does not achieve temperatures below 0 °C, which does not ensure quenching of all enzymatic reactions and it cannot be excluded that molecules from the supernatant contaminate the intracellular concentrations analyzed.

To avoid metabolite leakage, fast filtration was proposed as a washing step prior to quenching with liquid N_2 (Volmer et al., 2011a,b). Using an in-house made filter module, a determined amount of cells were passed through a filter and rapidly washed with 0.9% (w/w) NaCl. The entire procedure took between 30 and 60 s and allows obtaining cells free from any contaminating metabolites present in the cultivation media. Quenching with liquid N_2 has the advantage to rapidly freeze samples below −100 °C, thus avoiding further metabolic conversion (Buziol et al., 2002; Sellick et al., 2011). However, the system so far required individually made equipment and thus necessitated fine tuning of the individual components to identify optimal conditions of handling and use, making it difficult to transfer to other labs.

Despite the fast quenching step, the subsequent extraction of metabolites from cells captured on a filter is time consuming, as the filters have to be cut to fit into a tube, to be centrifuged together with the extraction solvent, and the final liquid extract is separated from the filter by decantation. The filter retains some fluid which makes it difficult to recover the entire volume added for the extraction. This can be partially resolved by using a suction system coupled to a column (Volmer et al., 2011a,b).

It was reported, that several methods for the intracellular metabolite extraction using solvents like hot methanol (Shryock et al., 1986), cold methanol (de Jonge et al., 2012; Volmer et al., 2011b), boiling ethanol (Gonzalez et al., 1997), chloroform/methanol (de Koning and van Dam, 1992), acid (Ryll and Wagner, 1991), alkali (Nissom et al., 2006) and hot water (Bolten et al., 2007) were successful. Though, to attain the maximum recovery of metabolites, cold methanol (Cao et al., 2011; Shin et al., 2010; Volmer et al., 2011a) and cold acetonitrile (Dietmair et al., 2010) are more preferred in mammalian cells.

All in all, the selection of a tailored protocol will depend on the cell characteristics, the metabolites required for the study or analytical equipment availability. In this work, we developed a modified protocol for fast filtration, washing and quenching using readily available parts. The entire procedure routinely takes less than 15 s to transfer washed cells into liquid nitrogen. For subsequent extraction either immediately after quenching or after storage, different cold solutions were compared and evaluated by LC–ESI-MS and FIA-MS–MS analyses of nucleotides and amino acids. The method is benchmarked against other quenching and extraction methods described in the literature and found to preserve the energy charge of AMP/ADP/ATP at the highest value (0.94 vs 0.90 and 0.82).

2. Materials and methods

2.1. Cell line and cultivation media

CHO-K1 suspension cell lines adapted to growth in glutamine free media (Bort et al., 2010) were cultivated in 500 ml spinner flasks using CD CHO media (Gibco, Invitrogen, Carlsbad, CA, USA) under 7% CO_2 and incubated at 37 °C with constant stirring at 50 rpm. Viability and cell counts were measured on a Vi-Cell analyzer (Beckman Coulter Inc., Fullerton, CA) based on the trypan-blue dye exclusion method.

2.2. Filtration and quenching

For each sample, after humidification of the filter membrane (PALL A/D Glass Fibre, 47 mm, New York, USA) with 10 ml of 0.9% (w/w) NaCl, a total of 10^7 cells harvested from a suspension culture were filtered using a standard filtration holder (Millipore, MA, USA) and immediately washed with 30 ml 0.9% (w/w) NaCl solution under controlled vacuum at 10 mbar (Controller CVC 3000 with W-B 6C, Vacuumbrand, Germany) to remove extracellular contaminating substances. The filter membrane with the captured cells was transferred into a 20 ml syringe barrel (20 ml BD™ Syringe with Luer-Lock™ connection, Becton Dickinson, NJ, USA) closed with a cap (Fig. 1). At this step a standard solution could be added (see f.i. Neubauer et al., 2012), before freezing of samples in liquid nitrogen, followed either by storage at −80 °C or immediate extraction.

Cold PBS quenching (Pabst et al., 2010) and cold methanol/AMBIC quenching (Sellick et al., 2009) were carried out exactly as published and compared to fast filtration by analysis of nucleotides and sugar nucleotides.

2.3. Metabolites extraction

Following quenching in liquid N_2, 8 ml of 50% (v/v) cold methanol were added into the syringe barrel. The plunger was carefully reinserted into the syringe barrel after removing the Luer-Lock™ cap and slightly pushing on the plunger to remove the excess air inside. The syringe barrel was relocked with a Luer-Lock™ cap and vortexed for 15 s. The cap was removed and a 0.45 μm filter membrane unit (Millex HV-Durapore PVDF filter) was attached to the syringe. The syringe plunger was squeezed to transfer the extraction solution into a fresh tube. The entire extracted volume (8 ml or as specified), was dried with a Thermo Savant SPD121P Speed Vac Concentrator equipment (Thermo Fischer Scientific, MA, USA) and stored at −80 °C for further quantification of metabolites.

2.4. LC–ESI-MS analysis for nucleotides

The sample extraction and analysis of the nucleotides and nucleotide sugars were performed as described in Pabst et al. (2010) and Taschwer et al. (2012). The data were interpreted using

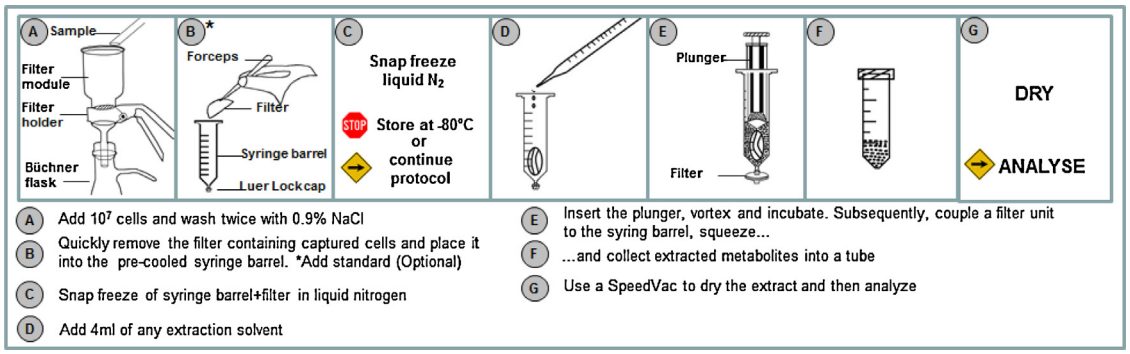


Fig. 1. Flow chart of steps for fast quenching and extraction.

Bruker's Data Analysis 4.0 software, using integration of selected ion chromatograms. All compounds were referred to the internal 13C10-ATP Standard (Sigma Aldrich, USA) for quantification. To allow correction for different ionization and detection efficiency of the compounds, equimolar mixtures of representative nucleotides were generated and measured under the same conditions (data not shown). Adenosines and Guanosines standards were purchased from Larova, Germany, Uridines and Cytidines from Sigma Aldrich, USA.

Energy charge was calculated as established by Atkinson (1968):

$$\text{Energy charge} = \frac{\text{ATP} + 0.5\text{ADP}}{\text{ATP} + \text{ADP} + \text{AMP}}$$

2.5. Lactate dehydrogenase analysis

After cell broth filtration with different vacuum pressures, the concentration of lactate dehydrogenase (LDH) in the flow-through filtrate was determined with an LDH Test Kit (Roche Diagnostics, Germany) following manufacturer instructions, in order to quantify cellular membrane damage of filtered CHO cells.

2.6. FIA-MS–MS Analysis

Biocrates' commercially available KIT plates (Innsbruck, Austria) were used for the quantification of amino acids, hexose (glucose), and biogenic amines. The fully automated assay was based on PITC (phenylisothiocyanate)-derivatization in the presence of internal standards followed by FIA-MS/MS (acylcarnitines, lipids, and hexose) and LC/MS (amino acids, biogenic amines) using an AB SCIEX 4000 QTrap™ mass spectrometer (AB SCIEX, Darmstadt, Germany) with electrospray ionization. The experimental metabolomics measurement technique is described in detail in the BIOCRATES AbsoluteIDQTM p180 KIT assay manual.

3. Results and discussion

3.1. Adaptation of protocol and optimization of vacuum for minimal leakage of cells

Fast filtration has been shown to be an efficient method for the separation of mammalian cells from complex cultivation media before quenching (Volmer et al., 2011a). We adapted the method to use only commercially available components (Fig. 1) to enable its application in any standard lab environment. A vacuum pump is used at a defined vacuum strength sufficient to increase filtration speed, but not so high as to cause damage to the cell membrane, which would result in leakage of intracellular metabolites. Therefore, a broad range of vacuum pressures (500, 80, 40, 20 and 10 mbar below the ambient air pressure) was tested and the flow-through of the wash-step analyzed for presence of lactate dehydrogenase (LDH) (as an example of a large molecule released from cells upon death, analyzed by an enzyme assay) and selected nucleotides and nucleotide sugars using LC–ESI-MS (as examples for small molecular weight metabolites with rapid turnover) (Fig. 2). As a control, three samples were filtered and washed without any vacuum at ambient air pressure (samples labeled 0-mbar). In this case the filtration and washing step took over 50 s, while the procedure took approximately 10 s when using 10 mbar of vacuum. As the filter clogged when 5×10^7 cells were applied (Murabito et al., 2009), which drastically increased the time between filtration and quenching, the applied number of cells used was 10^7 per filter for all samples. Leakage of small metabolites, especially of ATP, started to increase from 20 mbar, while larger test substances such as LDH started to significantly increase from 40 mbar onwards, similar to values published by Volmer et al. (2011a). The differences can be explained by the use of different equipment parts, however, with the protocol here described all components are readily available commercially, can be connected using standard lab equipment and require no handicraft work.

3.2. Completeness of quenching of metabolite conversion

To assess the speed and completeness of blocking all metabolic conversions in the cells, the energy charge (Ec) of cells quenched by fast filtration was compared to that of cells quenched by previously published methods, either with cold phosphate-buffered saline (pH 7.4, 0.5 °C) or with cold 60% methanol containing 0.85% (w/v) ammonium bicarbonate (AMBIC) (pH 7.4; −20 °C). For all approaches cells were isolated from the same shaker flask and each quenching procedure was performed twice by two independent operators (total $n = 4$). Depending on the nutrient state of the cells, the Ec of viable cells is expected to lie between 0.7 and 0.95 (Andersen and von Meyenburg, 1977; Atkinson, 1968; Bontemps et al., 1993; Cook et al., 1977; Sellick et al., 2011). Assuming that all samples coming from the same culture have an identical Ec, differences between quenching methods would be due to incomplete quenching and ongoing consumption of ATP by intracellular enzymes, resulting in reduced ATP and increased ADP or AMP concentrations and thus a reduced Ec (Fig. 3). The maintenance of metabolites with high conversion rates was best in the fast filtration/liquid nitrogen quenching protocol. The absolute recovery of sugar nucleotides (which have a lower conversion rate than nucleotide bi- and triphosphates) by the different methods was on average 100 for the fast filtration protocol, 150 for cold PBS and 20 for the methanol/AMBIC protocol (relative values, data not shown), indicating that with the methanol/AMBIC protocol significant loss of intracellular metabolites occurs due to leakage caused by membrane solubilisation by the methanol. The higher values for

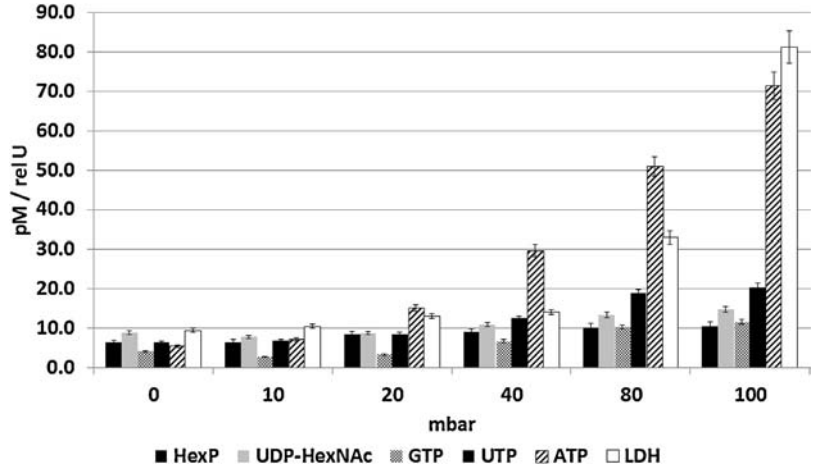


Fig. 2. Influence of vacuum on leakage of intracellular molecules into filtrate. Both low molecular weight substances and an intracellular protein (LDH) were analyzed in the filtered wash solution ($n = 3$). Results for LDH are given as relative units. Concentrations of released small molecules start to increase at 20 mbar vacuum pressure, so that 10 mbar was chosen for all subsequent experiments.

the cold PBS protocol are more difficult to explain, as they could be caused either by a low level of leakage occurring in the fast filtration protocol, or by incomplete removal of contaminating molecules derived from dead cells from the culture supernatant in the cold PBS method.

3.3. Optimization of metabolite extraction

In addition to the quenching method, the extraction of the metabolites and the subsequent sample treatment play an important role for the determination of metabolite concentrations. To provide an optimal extraction method, different volumes (4, 6, 8 ml) and ratios of the extraction solvent (methanol/ddH_2O; 100/0, 80/20, 50/50, 0/100) were analyzed by FIA-MS–MS. In general, the extraction protocols consist of disruption of cells by addition of the extraction solvent, dissolution of all cellular molecules and drying of the extract to obtain higher concentrations and better storability. Obviously, the drying requires harsh conditions that might affect the stability and recovery of metabolites. To analyze these effects, samples of different volume (1000, 500 and 250 μl) were dried in a Speed Vac and re-dissolved in 250 μl of the corresponding extraction solvent. These results were compared to samples which were directly frozen without Speed Vac treatment. This also yielded results on the effect of concentrating the metabolite amounts (concentration factor due to re-suspension: 4×, 2×, 1×).

Representative results on amino acid analyses are shown on the example of arginine and tryptophan (Fig. 4). The extraction volume used in the syringe had the least effect on the results, while the volume used for drying had a significant effect on most amino acids, especially on methionine, which was only detectable when analyzed directly and completely disappeared in the dried extracts. The larger the volume subjected to drying, the lower the observed concentration per cell. Aspartate, isoleucine, proline, phenylalanine and lysine were the least sensitive to this effect, showing only insignificant differences between the dried and directly analyzed samples. Arginine showed modest effects, while tryptophan had significant loss during drying (Fig. 4). Results were the most consistent with 100% methanol as extraction solvent, while all other

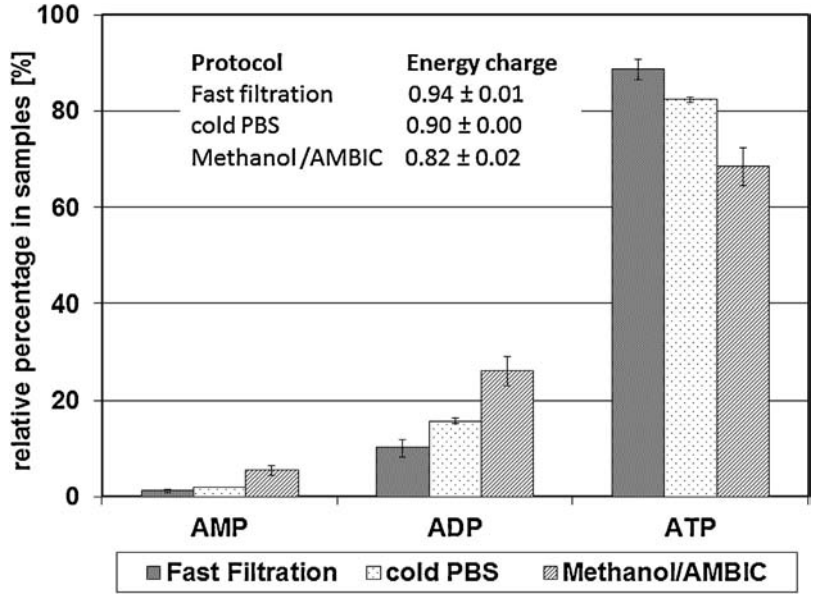


Fig. 3. Comparison of efficiency of quenching. After quenching, all samples were extracted and analyzed for nucleotides ($n = 4$). The relative distribution of AMP/ADP/ATP for each protocol is shown as well as the calculated energy charge.

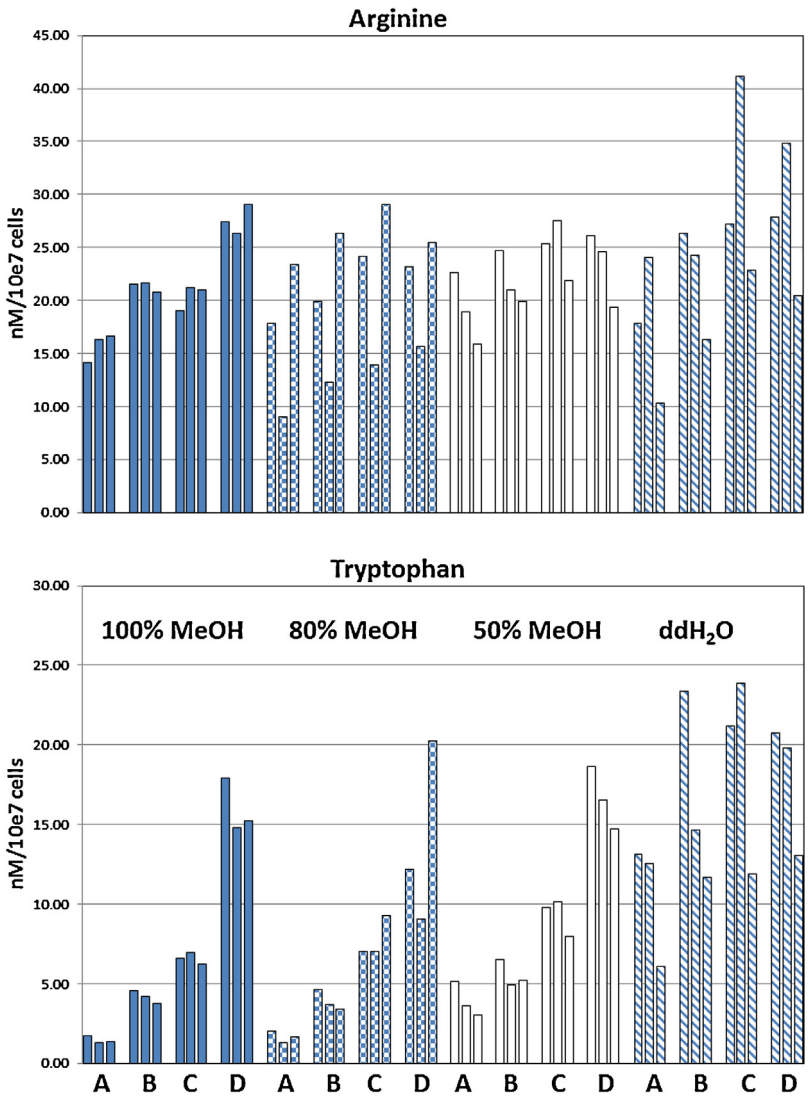


Fig. 4. Effects of extraction volume, drying and extraction solvents on intracellular arginine and tryptophan concentration: 10^7 cells were extracted using 8 ml (left bar within each group), 6 ml (middle bar) or 4 ml (right bar) of extraction solvent as specified in the graph. After extraction with the respective total volume, different amounts of extract were dried and reconstituted in 250 μl of the respective solvent. Volumes dried were 1000 μl (A), 500 μl (B) and 250 μl (C). Finally an aliquot of the extraction solution was directly analyzed without drying (D). Intracellular amino acid amounts are given as nm per 10^7 cells.

solvents showed higher variability between different handling procedures and also between replicates. This might be due to less efficient cell disruption and solution of metabolites. However, the effect of drying was also the most pronounced with 100% MeOH. Using distilled water for cell disruption worked very well when larger extraction volumes were used (8 ml or 6 ml), however, with 4 ml the recovered concentration of extracts was lower, most likely due to incomplete cell disruption in the lower volume. Detailed results for all amino acids are available as Supplement 1. Additional metabolites analyzed included biogenic amines and hexoses (Suppl. 1). Hexose gave similar results to the amino acids, with the most reliable extraction achieved by 100% methanol, while for the biogenic amines more differential results were observed: putrescine, creatinine and sarcosine behave similar to the amino acids, while acetylornithine, methionine-sulfoxide and to a lesser degree taurine were only well extracted in the methanol containing extraction solutions, with hardly any extraction possible when using water. Finally for spermidine the situation was reversed, with good extraction only possible with water and not at all with any of the methanol containing extractants. These results demonstrate the necessity to use multiple extraction solutions depending on the desired metabolites if a complete analysis of the metabolome of cells is required.

4. Conclusion

This work corroborates fast filtration as a tailored approach to be used before quenching and extraction of cells cultivated in complex media. Depending on filtration equipment and cell type, the choice of an appropriate vacuum pressure is necessary in order to reduce leakage of metabolites. Using commercially available plastic syringes which have a smooth inner surface, the filter can be quickly transferred into liquid nitrogen, without the need to cut it, which will lead to loss of time and potentially metabolites. Typical transfer times of washed cells into liquid nitrogen were below 15 s, with approximately 30 s handling time for filter change until the next replicate can be processed.

Also, the recovery of metabolites during extraction is easier and faster using the syringe. Traces or particles of the filter or cells can be removed by coupling a 0.45 μm filter membrane unit, thus increasing the life span of analytical columns. The entire protocol allows

the operator to extract metabolites in less than 2 min, thus enabling the handling of a large number of samples for bioprocess analyses.

Regarding further processing, speedvac treatment has shown a strong effect on the stability of several metabolites analyzed compared to samples which were directly frozen after extraction. The reason for this negative effect is likely to be the warm temperatures during the speedvac treatment and the time which is needed for this process. That effect was most prominent in amino acids after extraction with any methanol-solvent. In summary, the best results on the determination of metabolite concentrations in CHO cells were obtained when 4 ml of 100% methanol were used for extraction and aliquots of each sample were directly analyzed. Nevertheless, for other methods, speedvac treatment is required to remove solvents for protection of columns. In this case the use of an internal standard that contains the same metabolites as expected in the sample (Neubauer et al., 2012) is required. In addition, other classes of analytes may require different extraction solutions. Here again the speed of this method would be of advantage, as both the sampling procedure onto the filters as well as the extraction procedure are easy and fast to enable generation of several sample loaded filters and subsequent extraction with multiple extract solutions.

Acknowledgements

The authors are grateful to Dr. Tobias Thüte and Prof. Thomas Noll for kind advice during the setup of the fast filtration method. This research was performed within the Austrian Center of Industrial Biotechnology, ACIB, a COMET K2 centre of the Austrian FFG. The MS-equipment was made available by the Cellular Analysis Core Facility of Eq-VIBT. LN is supported by the FWF Doctoral Programme Biotop, W1224.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbiotec.2014.04.014.

References

Andersen, K.B., von Meyenburg, K., 1977. Charges of nicotinamide adenine nucleotides and adenylate energy charge as regulatory parameters of the metabolism in *Escherichia coli*. J. Biol. Chem. 252 (12), 4151–4156.

Atkinson, D.E., 1968. The energy charge of the adenylate pool as a regulatory parameter. Interaction with feedback modifiers. Biochemistry 7 (11), 4030–4034.

Bennett, B.D., Yuan, J., Kimball, E.H., Rabinowitz, J.D., 2008. Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. Nat. Protoc. 3 (8), 1299–1311.

Bolten, C.J., Kiefer, P., Letisse, F., Portais, J.C., Wittmann, C., 2007. Sampling for metabolome analysis of microorganisms. Anal. Chem. 79 (10), 3843–3849.

Bontemps, F., Vincent, M.F., Van den Berghe, G., 1993. Mechanisms of elevation of adenosine levels in anoxic hepatocytes. Biochem. J. 290 (Pt 3), 671–677.

Bort, J.A., Stern, B., Borth, N., 2010. CHO-K1 host cells adapted to growth in glutamine-free medium by FACS-assisted evolution. Biotechnol. J. 5 (10), 1090–1097.

Buchholz, A., Takors, R., Wandrey, C., 2001. Quantification of intracellular metabolites in *Escherichia coli* K12 using liquid chromatographic–electrospray ionization tandem mass spectrometric techniques. Anal. Biochem. 295 (2), 129–137.

Buziol, S., Bashir, I., Baumeister, A., Claassen, W., Noisommit-Rizzi, N., Mailinger, W., Reuss, M., 2002. New bioreactor-coupled rapid stopped-flow sampling technique for measurements of metabolite dynamics on a subsecond time scale. Biotechnol. Bioeng. 80 (6), 632–636.

Canelas, A.B., Ras, C., ten Perick, A., van Dam, J.C., Heijnen, J.J., van Gulik, W.M., 2008. Leakage-free rapid quenching technique for yeast metabolomics. Metabolomics 4, 226–239.

Cao, B., Aa, J., Wang, G., Wu, X., Liu, L., Li, M., Shi, J., Wang, X., Zhao, C., Zheng, T., et al., 2011. GC–TOFMS analysis of metabolites in adherent MDCK cells and a novel strategy for identifying intracellular metabolic markers for use as cell amount indicators in data normalization. Anal. Bioanal. Chem. 400 (9), 2983–2993.

Chung, B.K., Selvarasu, S., Andrea, C., Ryu, J., Lee, H., Ahn, J., Lee, D.Y., 2010. Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. Microb. Cell Fact. 9, 50.

Cook, G.A., Sullivan, A.C., Ontko, J.A., 1977. Influences of intracellular pyridine nucleotide redox states on fatty acid synthesis in isolated rat hepatocytes. Arch. Biochem. Biophys. 179 (1), 310–321.

D'Huys, P.J., Lule, I., Vercammen, D., Anné, J., Van Impe, J.F., Bernaerts, K., 2012. Genome-scale metabolic flux analysis of *Streptomyces lividans* growing on a complex medium. J. Biotechnol. 161 (1), 1–13.

de Jonge, L.P., Douma, R.D., Heijnen, J.J., van Gulik, W.M., 2012. Optimization of cold methanol quenching for quantitative metabolomics of *Penicillium chrysogenum*. Metabolomics 8 (4), 727–735.

de Koning, W., van Dam, K., 1992. A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. Anal. Biochem. 204 (1), 118–123.

Dietmair, S., Hodson, M.P., Quek, L.-E., Timmins, N.E., Chrysanthopoulos, P., Jacob, S.S., Gray, P., Nielsen, L.K., 2012a. Metabolite profiling of CHO cells with different growth characteristics. Biotechnol. Bioeng. 109 (6), 1404–1414.

Dietmair, S., Timmins, N.E., Gray, P.P., Nielsen, L.K., Krömer, J.O., 2010. Towards quantitative metabolomics of mammalian cells: development of a metabolite extraction protocol. Anal. Biochem. 404 (2), 155–164.

Dietmair, S., Nielsen, L.K., Timmins, N.E., 2012b. Mammalian cells as biopharmaceutical production hosts in the age of omics. Biotechnol. J. 7 (1), 75–89.

Gonzalez, B., François, J., Renaud, M., 1997. A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol. Yeast 13 (14), 1347–1355.

Licona-Cassani, C., Marcellin, E., Quek, L.E., Jacob, S., Nielsen, L.K., 2012. Reconstruction of the *Saccharopolyspora erythraea* genome-scale model and its use for enhancing erythromycin production. Anton. Leeuw. 102 (3), 493–502.

Martínez, V.S., Dietmair, S., Quek, L.E., Hodson, M.P., Gray, P., Nielsen, L.K., 2013. Flux balance analysis of CHO cells before and after a metabolic switch from lactate production to consumption. Biotechnol. Bioeng. 110 (2), 660–666.

Moritz, B., Striegel, K., De Graaf, A.A., Sahm, H., 2000. Kinetic properties of the glucose-6-phosphate and 6-phosphogluconate dehydrogenases from *Corynebacterium glutamicum* and their application for predicting pentose phosphate pathway flux in vivo. Eur. J. Biochem. 267 (12), 3442–3452.

Murabito, E., Simeonidis, E., Smallbone, K., Swinton, J., 2009. Capturing the essence of a metabolic network: a flux balance analysis approach. J. Theor. Biol. 260 (3), 445–452.

Neubauer, S., Haberhauer-Troyer, C., Klavins, K., Russmayer, H., Steiger, M.G., Gasser, B., Sauer, M., Mattanovich, D., Hann, S., Koellensperger, G., 2012. U(13) C cell extract of *Pichia pastoris* – a powerful tool for evaluation of sample preparation in metabolomics. J. Sep. Sci. 35 (22), 3091–3105.

Nissom, P.M., Sanny, A., Kok, Y.J., Hiang, Y.T., Chuah, S.H., Shing, T.K., Lee, Y.Y., Wong, K.T., Hu, W.S., Sim, M.Y., et al., 2006. Transcriptome and proteome profiling to understanding the biology of high productivity CHO cells. Mol. Biotechnol. 34 (2), 125–140.

Pabst, M., Grass, J., Fischl, R., Léonard, R., Jin, C., Hinterkörner, G., Borth, N., Altmann, F., 2010. Nucleotide and nucleotide sugar analysis by liquid chromatography–electrospray ionization-mass spectrometry on surface-conditioned porous graphitic carbon. Anal. Chem. 82 (23), 9782–9788.

Ryll, T., Wagner, R., 1991. Improved ion-pair high-performance liquid chromatographic method for the quantification of a wide variety of nucleotides and sugar-nucleotides in animal cells. J. Chromatogr. 570 (1), 77–88.

Schilling, C.H., Edwards, J.S., Letscher, D., Palsson, B., 2000. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. Biotechnol. Bioeng. 71 (4), 286–306.

Sellick, C.A., Hansen, R., Maqsood, A.R., Dunn, W.B., Stephens, G.M., Goodacre, R., Dickson, A.J., 2009. Effective quenching processes for physiologically valid metabolite profiling of suspension cultured Mammalian cells. Anal. Chem. 81 (1), 174–183.

Sellick, C.A., Hansen, R., Stephens, G.M., Goodacre, R., Dickson, A.J., 2011. Metabolite extraction from suspension-cultured mammalian cells for global metabolite profiling. Nat. Protoc. 6 (8), 1241–1249.

Selvarasu, S., Ho, Y.S., Chong, W.P., Wong, N.S., Yusufi, F.N., Lee, Y.Y., Yap, M.G., Lee, D.Y., 2012. Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. Biotechnol. Bioeng. 109 (6), 1415–1429.

Selvarasu, S., Wong, V.V., Karimi, I.A., Lee, D.Y., 2009. Elucidation of metabolism in hybridoma cells grown in fed-batch culture by genome-scale modeling. Biotechnol. Bioeng. 102 (5), 1494–1504.

Sengupta, N., Rose, S.T., Morgan, J.A., 2011. Metabolic flux analysis of CHO cell metabolism in the late non-growth phase. Biotechnol. Bioeng. 108 (1), 82–92.

Sheikh, K., Förster, J., Nielsen, L.K., 2005. Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of *Mus musculus*. Biotechnol. Prog. 21 (1), 112–121.

Shin, M.H., Lee, d.Y., Liu, K.H., Fiehn, O., Kim, K.H., 2010. Evaluation of sampling and extraction methodologies for the global metabolic profiling of *Saccharophagus degradans*. Anal. Chem. 82 (15), 6660–6666.

Shryock, J.C., Rubio, R., Berne, R.M., 1986. Extraction of adenine nucleotides from cultured endothelial cells. Anal. Biochem. 159 (1), 73–81.

Taschwer, M., Hackl, M., Hernández Bort, J.A., Leitner, C., Kumar, N., Puc, U., Grass, J., Papst, M., Kunert, R., Altmann, F., et al., 2012. Growth, productivity and protein glycosylation in a CHO EpoFc producer cell line adapted to glutamine-free growth. J. Biotechnol. 157 (2), 295–303.

Volmer, M., Gettmann, J., Scholz, S., Büntemeyer, H., Noll, T., 2011a. A method for metabolomic sampling of suspended animal cells using fast filtration. BMC Proc. 5 (8), P93.

Volmer, M., Northoff, S., Scholz, S., Thüte, T., Büntemeyer, H., Noll, T., 2011b. Fast filtration for metabolome sampling of suspended animal cells. Biotechnol. Lett. 33 (3), 495–502.

Weibel, K.E., Mor, J.R., Fiechter, A., 1974. Rapid sampling of yeast cells and automated assays of adenylate, citrate, pyruvate and glucose-6-phosphate pools. Anal. Biochem. 58 (1), 208–216.

Zamorano, F., Vande Wouwer, A., Jungers, R.M., Bastin, G., 2012. Dynamic metabolic models of CHO cell cultures through minimal sets of elementary flux modes. J. Biotechnol. 164 (3), 409–422.

Appendix IX

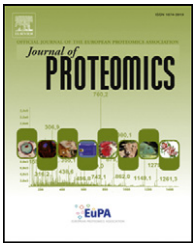
Jung, G., Pabst, M., **Neumann, L.**, Berger, A., Lubec, G. (2012) Characterization of α-l-Iduronidase (Aldurazyme®) and its complexes. *J Proteomics*, doi:10.1016/j.jprot.2012.09.022

JOURNAL OF PROTEOMICS 80 (2013) 26–33

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot



Characterization of α-L-Iduronidase (Aldurazyme®) and its complexes

Gangsoo Jung[a], Martin Pabst[b,1], Laura Neumann[b], Angelika Berger[a], Gert Lubec[a,*]

[a]Department of Pediatrics, Medical University of Vienna, Vienna, Austria
[b]Department of Chemistry, University of Natural Resource and Life Sciences, Vienna, Austria

ARTICLE INFO

Article history:
Received 19 July 2012
Accepted 22 September 2012
Available online 28 September 2012

Keywords:
α-L-Iduronidase
Mass spectrometry
Mucopolysaccharide
Mucopolysaccharidosis type I
Post-translational modification
Glycosylation

ABSTRACT

Alpha-L-Iduronidase(IDUA) was the first enzyme replacement therapy approved for mucopolysaccharidosis type I and the corresponding recombinant protein drug, Aldurazyme®, is commercially available. In the frame of gel-based mass spectrometrical characterization of protein drugs, we intended to identify protein sequence and possible protein modifications. Moreover, we were interested in which aggregation/complex form Aldurazyme® would exist, which complexes were enzymatically active and in which form the naturally occurring enzyme would be present in the brain. Aldurazyme® was run on 2DE gel electrophoresis, spots were excised, in-gel digested with several proteases and identified by nano-LC-ESI-MS/MS on an ion trap. IDUA-activity was determined by a fluorometric principle. Blue-native gel electrophoresis with subsequent immunoblotting was carried out to show the presence of protein complexes.

The protein was unambiguously identified by 100% sequence coverage; several amino acid substitutions were detected and protein modifications were novel phosphorylations on S59 and S482, histidine methylation at H572 and provide evidence for already known N-glycosylations. Four Aldurazyme® complexes that all were enzymatically active, were observed while a single complex was observed for the physiologically occurring IDUA in the brain.

The findings are relevant for understanding chemistry, physiology, pharmacology and medicine of IDUA, design of further and interpretation of previous work.

1. Introduction

Aldurazyme® is a recombinant human α-L-iduronidase (IDUA) and the first enzyme replacement therapy approved for treatment of an MPS I disorder [1,2]. Mucopolysccharidosis I is a lysosomal storage disorder that results from deficiency in the lysosomal enzyme α-L-iduronidase (IDUA; EC 3.2.1.76), which hydrolyzes terminal iduronic acid residues on glycosaminoglycans (GAGs). Deficiency of the enzyme leads to progressive accumulation of GAGs, dermatan and heparan sulfate, in all organs and tissues [3]. The clinical result of excess storage of GAGs is a diverse disorder in connective tissue, which can appear as learning difficulties, facial and skeletal abnormalities, upper airway obstruction, corneal clouding, heart disease and joint

Abbreviations: IDUA, Alpha-L-Iduronidase; GAGs, glycosaminoglycans; 2-DE, Two-dimensional gel electrophoresis; DDM, n-dodecyl β-D-maltosidase.

* *Corresponding author at:* Medical University of Vienna, Department of Pediatrics, Waehringer Guertel 18, A-1090 Vienna, Austria. Tel.: +43 1 404003215; fax: +43 1 404006065.
E-mail address: gert.lubec@meduniwien.ac.at (G. Lubec).
[1] Current affiliation: Department of Chemistry and Applied Biosystems, ETH Zurich, Switzerland.

1874-3919/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jprot.2012.09.022

stiffness [4,5]. This genetic disease is characterized by a wide spectrum of clinical severity and is classified into three syndromes: Hurler syndrome (severe, OMIM 607016), with life expectancy of less than 10 years; Hurler–Scheie syndrome (intermediate, OMIM 607015), with life expectancy of typically less than 25 years and Scheie syndrome (attenuated, OMIM 607016), with the possibility of a normal lifespan [6,7]. Weekly infusions of 0.58 mg/kg intravenous IDUA improve hepatosplenomegaly, pulmonary function, ambulation, joint mobility, cardiac function [8,9] and cognitive and neuroradiological progress has been reported [10].

Aldurazyme® (Genzyme Corp. Naarden, Netherlands) is a glycoprotein with a molecular weight of approximately 83 kDa. The recombinant protein is composed of 628 amino acids after cleavage of the N-terminus and typically, the enzyme has six N-glycosylation sites, which show a site specific glycosylation pattern: Asn-110, complex type glycans; Asn-190, complex type glycans; Asn-336, bisphosphorylated oligomannosidic glycan (P2Man7GlcNAc2); Asn-372, high mannose type glycans (mainly Man9GlcNAc2, some of which are monoglucosylated); Asn-415, mixed oligomannosidic and complex type glycans; Asn-451, bisphosphorylated oligomannosidic glycan (P2Man7GlcNAc2) [11].

To the best of our knowledge, Aldurazyme® has not been characterized by mass spectrometry. Knowledge on protein sequence, modifications and the form of complexes of the protein drug, would be important for drug safety, stability and efficiency, however. It is known that many factors influence biological and biochemical properties of protein drugs including immunogenicity and efficacy of therapeutic recombinant proteins [12]. Among them, aggregation of recombinant human protein is a key factor for the immune response that may result in adverse events and affecting protein stability: by reducing aggregation immunogenicity can be reduced and make drugs safer and more efficient [13–15]. And indeed, adverse effects of recombinant IDUA® have been reported [1,16]. They demonstrated that patients had several adverse effects including rash, arthralgia, headache, flushing, arthropathy, fever and nausea after injection of Aldurazyme® for enzyme replacement therapy.

Herein we show a series of sequence abnormalities as amino acid substitutions, protein modifications and post-translational modifications as well as that the enzyme is present as several protein complexes, for which the consequences are totally unclear. The Aldurazyme® protein was fully sequenced and modifications are shown thus forming the basis for pharmacological, immunochemical and biomedical studies.

2. Materials and methods

The design of the study is given in Supplemental Table 1.

2.1. Sample preparation for 2DE

70 μg of Aldurazyme® (Genzyme, Naarden, Netherlands) was added to 2 mL of sample buffer consisting of 8 M urea (Merck, Darmstadt, Germany), 4% CHAPS (3-[(3-cholamidopropyl) dimethylammonio]-1-propane-sulfonate) (Sigma, St. Louis, MO), 10 mM 1,4-dithioerythritol (Merck, Germany) and 0.5% carrier ampholytes "Resolyte" 3,5–10 (BDH Laboratory Supplies, Electran, UK). The suspension was transferred into Ultrafree-4 centrifugal filter units (Millipore, Bedford, MA) for desalting and concentrating proteins. The protein content of the supernatant was quantified by the Bradford protein assay system [17]. The standard curve was generated using bovine serum albumin and absorbance was measured at 595 nm.

2.2. Deglycosylation and desialylation of Aldurazyme®

70 μg of Aldurazyme® were incubated with 500 units of PNGase F (peptide N-Glycosidase F from Flavobacterium meningosepticum; New England Biolabs, Beverly, MA) in 50 mM sodium phosphate buffer (pH 7.5) at 37 °C for 2 h. Likewise, 70 μg of Aldurazyme® were incubated with 500 units of Sialidase (Neuraminidase from *Clostridium perfringens*; New England Biolabs, Beverly, MA) in 50 mM sodium citrate buffer (pH 6.0) at 37 °C for 2 h. The samples were concentrated and dialysed against sample buffer as given above.

2.3. Two-dimensional gel electrophoresis (2-DE)

Two-dimensional gel electrophoresis was performed essentially as reported [18]. 50 μg of protein were applied on immobilized pH 3–10 nonlinear gradient strips in sample cups at their basic and acidic ends. Focusing was started at 200 V and the voltage was gradually increased to 8,000 V at 4 V/min and then kept constant for a further 3 h (approximately 150,000 Vh totally). After the first dimension, strips (18 cm) were equilibrated for 15 min in the buffer containing 6 M urea, 20% glycerol, 2% SDS, 2% DTT and then for 15 min in the same buffer containing 2.5% iodoacetamide instead of DTT. After equilibration, strips were loaded on 10–16% gradient sodium dodecylsulfate polyacrylamide gels for second-dimensional separation. The gels (180×200×1.5 mm) were run at 40 mA per gel. Immediately after the second dimension run, gels were fixed for 12 h in 50% methanol, containing 10% acetic acid, the gels were stained with Colloidal Coomassie Blue (Novex, San Diego, CA) for 12 h on a rocking shaker. Molecular masses were determined by running standard protein markers (Biorad Laboratories, Hercules, CA) covering the range 10–250 kDa. pI values were used as given by the supplier of the immobilized pH gradient strips (Amersham Bioscience, Uppsala, Sweden). Excess of dye was washed out from the gels with distilled water and the gels were scanned with ImageScanner (Amersham Bioscience, Uppsala, Sweden).

2.4. In-gel digestion

Gel pieces from 2DE were put into a 1.5 mL tube and washed with 10 mM ammonium bicarbonate and 50% ACN in 10 mM ammonium bicarbonate repeatedly. Addition of ACN resulted in gel shrinking and the shrunk gel plugs were then dried in a Speedvac Concentrator 5301 (Eppendorf, Germany). Dried gel pieces were re-swollen and in-gel digested with 40 ng/μL trypsin (Promega, Madison, WI, USA) in digestion buffer (consisting of 5 mM octyl β-D-glycopyranoside (OGP) and 10 mM ammonium bicarbonate, pH 7.8) and incubated overnight at 37 °C. Digestion with chymotrypsin (Roche

Diagnostics), 25 ng/μL was done in 25 mM NH_3HCO_3 with 5 mM OGP (pH 7.8) at 30 °C for 4 h. For subtilisin digestion, the gel pieces were covered with 30 μL of 10 ng/μL subtilisin (protease from Bacillus subtilis var. biocus A., Sigma) in a digestion buffer containing of 6 M urea and 1 M Tris, pH 8.5, then rehydrated for 10 min at 4 °C. Pepsin (Sigma) digestion was performed in 0.1 M HCl (pH 1.0) and kept at 37 °C for 1 h. Peptide extraction was performed with 15 μL of 1% formic acid (FA) in 5 mM OGP for 30 min, 15 μL 0.1% FA for 30 min and 15 μL 0.1% FA in 20% ACN for 30 min. The extracted peptides were pooled for nano-LC-ESI-CID/ETD-MS/MS analysis [19].

For in-gel phosphatase treatment gel pieces were destained and dried as shown above. The dried spots were incubated in a solution of 0.5 mL of calf intestine alkaline phosphatase (New England Biolabs, Ipswich, MA, USA) in the presence of 100 mM ammonium bicarbonate, shrunk in ACN and dried in a SpeedVac. Peptides are extracted as shown above.

2.5. *Nano LC-ESI-CID/ETD-MSMS analysis*

The HPLC was used an Ultimate 3000 system (Dionex Corporation, Sunnyvale, CA) equipped with a PepMap100 C-18 trap column (300 μm×5 mm) and PepMap100 C-18 analytic column (75 μm×150 mm). The gradient was (A= 0.1% formic acid in water, B=0.08% formic acid in ACN) 4–30% B from 0 to 105 min, 80% B from 105 to 110 min, 4% B from 110 to 125 min. A HCT ultra ETD II (Bruker Daltonics, Bremen, Germany) was used to record peptide spectra over the mass range of *m*/z 350–1,500, and MS/MS spectra in information-dependent data acquisition over the mass range of *m*/z 100–2,800. Repeatedly, MS spectra were recorded followed by three data-dependent CID MS/MS spectra and three ETD MS/ MS spectra generated from the three of highest intensity precursor ions. An active exclusion of 0.4 min after two spectra was used to detect low abundant peptides. The voltage between ion spray tip and spray shield was set to 1,500 V. Drying nitrogen gas was heated to 150 °C and the flow rate was 10 L/min. The collision energy was set automatically according to the mass and charge state of the peptides chosen for fragmentation. Multiple charged peptides were chosen for MS/MS experiments due to their good fragmentation characteristics. MS/MS spectra were interpreted and peak lists were generated by Data Analysis 3.4 (Bruker Daltonics, Bremen, Germany). Searches were done by using the Mascot 2.2 (Matrix Science, London, UK) against latest NCBI and UniprotKB database for protein identification. Searching parameters were set as follows: enzyme selected as trypsin with two maximum missing cleavage sites, species limited to human, a mass tolerance of 0.2 Da for peptide tolerance, 0.2 Da for MS/MS tolerance, fixed modification of carbamidomethyl (C) and variable modification of methionine oxidation and phosphorylation (Tyr, Thr, and Ser). Positive protein identifications were based on a significant MOWSE score. After protein identification, an error-tolerant search was done to detect nonspecific cleavage and unassigned modifications. Protein identification and modification information returned from MASCOT were manually inspected and filtered to obtain confirmed protein identification and modification lists of CID MS/MS and ETD MS/MS. Modiro™ v1.1 software (Protagen, Dortmund, Germany): enzyme selected as we used with two maximum missing cleavage sites, species limited to human, a peptide mass tolerance of 0.2 Da for peptide tolerance, 0.2 Da for fragment mass tolerance, modification 1 of carbamidomethyl (C) and modification 2 of oxidation. Search for unknown mass shifts, search for amino acid substitution and calculate significance were selected on advanced PTM-explorer search strategies. Positive protein identification was first of all listed by spectra view and then each identified peptide was considered its significance based on 0.2 Da delta value, ion-charge status of peptide, b- and y-ion fragmentation quality and a significant score [20].

2.6. *Oligosaccharide analysis*

Pieces cut from the 2DE gel containing Aldurazyme® , were minced and put into a 1.5 mL Eppendorf tube. After adding 100 μL of a 50 mM ammonium acetate buffer at pH 8.4, glycans were released with 0.2U of PNGase F (Roche Diagnostics, Basel, Switzerland) at 37 °C over night. The supernatant was collected and glycans were pre-purified by the use of a 10-mg HyperSep Hypercarb SPE cartridge (Thermo Scientific, Vienna). N-glycans were then analyzed by cap flow LC-ESI-MS using a porous graphitic carbon column (100×0.32 mm, Thermo Scientific, Vienna) with 65 mM ammonium formate buffer pH 3.0 as the aqueous solvent on a Q-TOF Global Ultima (Micromass, Manchester, UK) using ESI. In contrast to a recently published work [21], glycans were not reduced and a very steep gradient from 2% to 42% ACN was developed over 20 min at a flow rate of 8 μL/min. Flow rates were maintained using a Dionex Ultimate 3000 RSLC nano. Data were acquired in ES+full scan mode, with a capillary voltage of 3.2 kV and a cone voltage of 80 V. Masses were acquired between 500 and 1800 Da, where a scan range between min 9.5 and min 15 was further summarized in order to obtain a complete glycan mass profile. Data evaluation was performed using MassLynx 4.1 software [22].

2.7. *Electroelution*

In order to allow determination of enzyme activities of individual IDUA complexes, phosphatase- and PNGase F-treated IDUA, proteins were eluted from native gels. Two longitudinal strips, i.e., one lane containing markers and one lane containing IDUA, were cut out and stained with colloidal Coomassie blue while keeping the rest of the native gel on a glass plate on ice. The stained strips were lined up along the edges of the unstained gel and used as a guide to cut out bands of interest from unstained gel. The excised bands were placed into D-Tube Midi (3.5 kDa cutoff, Novagen, Germany), 800 μL native PAGE running buffer (25 mM Tris, 192 mM glycine, p H 8.3) were added, and tubes were fixed on a horizontal electrophoresis chamber, Sub-Cell RGT cell (BioRad, Hercules, CA, USA). The chamber was filled with native PAGE running buffer and electroelution in D-Tube Midi was performed at 100 V for 5 h in the cold room. Reverse electrophoresis was done in D-Tube Midi to recover protein attached to the tube's membrane for 2 min. The electroelution tubes were dialyzed in 3 L of 10 mM phosphate buffer pH 7.0 overnight with gentle stirring in the cold room. All steps were carried out in the cold room at 4 °C [23].

2.8. Blue native (BN)-PAGE

BN-PAGE was carried out as published previously with minor modifications [19]. 40 μg of Aldurazyme®, PNGase F-treated, sialidase-treated and phosphatase-treated Aludrazyme® and a human brain sample were applied on BN-PAGE.

The human brain sample obtained from the Medical University of Leipzig, Germany, was carefully washed with ice-cold homogenization buffer containing 10 mM HEPES, pH 7.5, 300 mM sucrose, 1 mM EDTA and protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany). The human brain sample was homogenized in 10 mL of homogenization buffer using an Ultra-Turrax (IKA, Staufen, Germany). The homogenate was centrifuged for 10 min at 1,000×g and the pellet was discarded. 40 μg of the resulting supernant and 40 μg of Aldurazyme® preparations were suspended in extraction buffer containing 1.5 M 6-aminocaproic acid and 300 mM Bis-tris, pH 7.0. After resuspension, 10% n-dodecyl β-D-maltosidase (DDM) stock solution was added to achieve a 1% DDM concentration. 8 μL of BN-PAGE loading buffer [5% (w/v) Coomassie G250 in 750 mM 6-aminocaproic acid] were mixed with 50 μL of resulting supernant and loaded onto the gel. BN-PAGE was performed in a PRO-TEAN II xi Cell (BioRad, Germany) using 4% stacking and 5-18% separating gel. The BN-PAGE gel buffer contained 500 mM 6-aminocaproic acid and 50 mM Tricine, 15 mM Bis-Tris and 0.05% (w/v) Coomassie G250, pH 7.0; and the anode buffer 50 mM Bis–Tris, pH 7.0. For electrophoresis, the voltage was set to 50 V for 1 h, 75 V for 6 h, and was increased sequentially to 400 V (maximum current 15 mA/gel, maximum voltage 500 V) until the dye front reached the bottom of the gel. High-molecular mass markers were obtained from Invitrogen (Carlsbad, CA, USA).

For BN-PAGE-immunoblotting of the human brain sample the technique essentially described recently [24] was used. The antibody used was a polyclonal antibody against human IDUA (peptide 244–274; Abcam, UK, ab 103949), the secondary antibody was a goat polyclonal antibody against rabbit IgG, heavy and light chain, horseradish peroxidase conjugated (Abcam, UK, ab6721).

2.9. Enzyme activity assay

0.5 mg/mL of Aldurazyme®, Aldurazyme® treated with PNGase F, sialidase or phosphatase were diluted to 0.2 μg/mL in assay buffer (50 mM NaOAc, 150 mM NaCl, 0.02% Brij35 (w/v) pH 3.5). 200 μM of 4-methylumbelliferyl α-L-iduronide (Glycosynth, 4MU) were dissolved in assay buffer. Equal volumes of 0.2 μg/mL Aldurazyme® and 200 μM 4MU were mixed and incubated for 10 minutes at 21 °C. The mixtures were diluted to 0.005 μg/mL in developing buffer (0.1 M Tris, pH 9.0) and 100 μL of the diluted mixtures were loaded into a microtiter plate. Excitation and emission wavelengths were read at 365 nm and 445 nm in endpoint mode on Synergy 2 multi-mode microplate reader (BioTek, Germany).

2.10. Determination of IDUA activity from IDUA preparations eluted from BN-PAGE

Aldurazyme® complexes eluted from BN-PAGE were taken for the determination of the IDUA activity as given above.

3. Results

3.1. Two dimensional gel electrophoresis

Aldurazyme® was separated and moving to the known iso-electric point on 2DE. A train of spots was located near the basic region at pIs from 8 to 9 with an apparent molecular weight of 85 kDa (theoretical molecular weight is 83 kDa) (Fig. 1A). Ten punches were picked as indicated by arrowheads along the train.

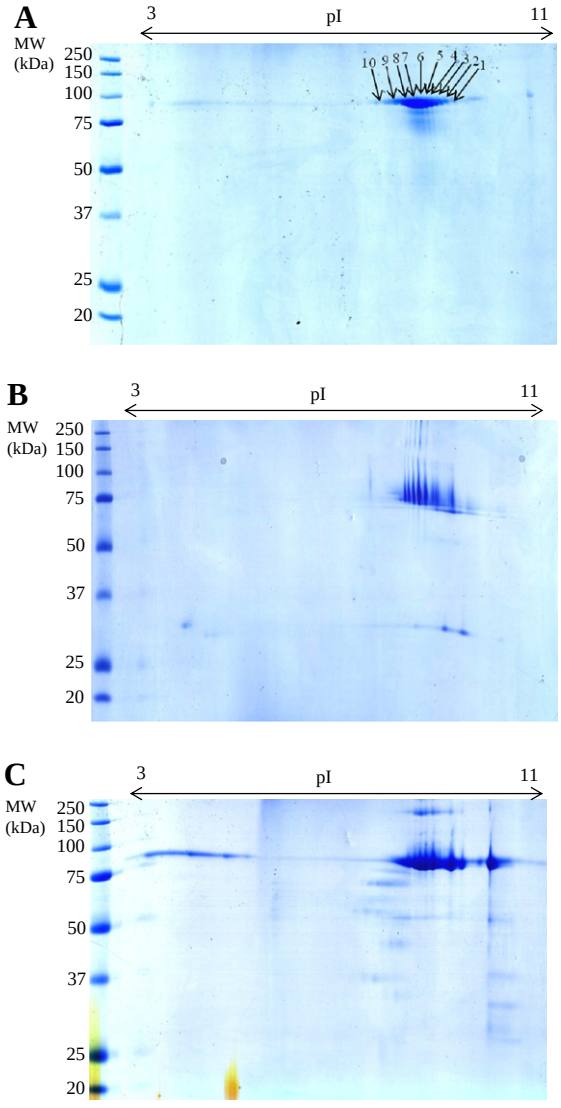


Fig. 1 – Two-dimensional electrophoresis. (A) Application of Aldurazyme® in its native form shows a series of spots as a train within pIs between 8.0 and 9.0. Arrows indicate spot picking for mass spectrometry. (B) PNGase F-treatment showed the effect of N-deglycosylation resulting into a different electrophoretic pattern. Glycosylations, i.e. the composition of the carbohydrate moiety, are revealed in the results section of the manuscript. (C) Sialidase treatment resulted into a different electrophoretic pattern with shifts to the acidic part indicating the presence of sialic acid(s).

Following PNGase-treatment of Aldurazyme®, spots were clearly separated and shifted to a lower apparent molecular weight of 75 kDa. The train of spots as shown in Fig. 1B cut out and the carbohydrate moiety was analyzed (given below).

Following sialidase-treatment no electrophoretic shifts of apparent molecular weight took place, however, a shift towards the basic part of the gel occurred as expected and individual spots were resolved (Fig. 1C).

Aldurazyme® was unambiguously identified by the gel-based mass spectrometrical method with 100% of sequence coverage (Supplemental Fig. 1). Trypsin digestion lead to 65.3%, chymotrypsin to 91.4%, pepsin to 74.4% and subtilisin to 37.4% (Supplemental Fig. 2). The list of generated peptides is shown in Table 1 along with parameters peptide sequence, observed molecular weight, expected and calculated Mr and the mass error delta.

Protein modifications are shown in Supplemental Table 2 indicating oxidations, deamidations, hydroxylations, histidine methylation on H572, that was verified by replacing methanol for gel processing by ethanol (Supplemental Fig. 3). Moreover, novel phosphorylation sites on S59 and S482 were detected and verified by phosphatase-treatment (Supplemental Figs. 4 and 5).

For each individual Aldurazyme® spot on SDS page, we explored the corresponding N-glycan pool using in gel PNGAse F release.

Overall, the N-glycan pools consisted of oligomannosidic glycans (mainly Man7, Man8, Man9 as well as Man9+Glc), complex type glycans with up to three antennae. Sialylation as well as fucosylation were observed on almost all complex type glycans. The glycan profile was found to be similar (with variations within the standard run to run variations) within all SDS page protein spots. Differences were only detected in the signal intensities, which corresponded to the different amounts of protein in the spots (Supplemental Table 3, Supplemental Fig. 6).

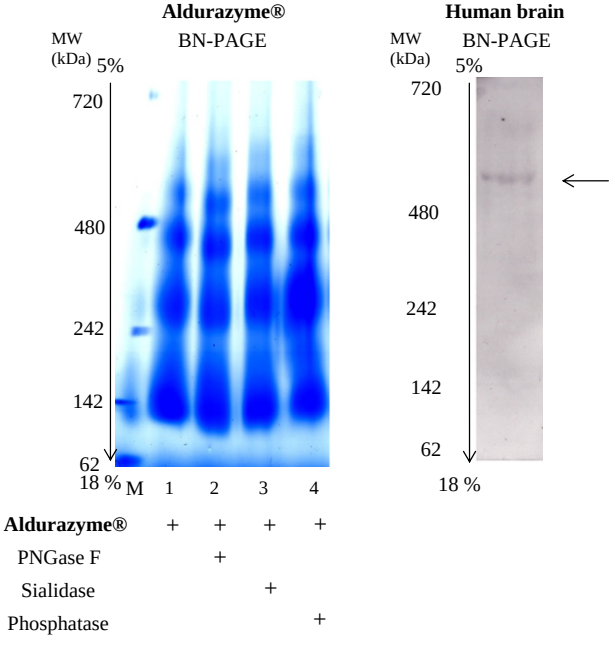


Fig. 2 – Formation of high molecular structure of Aldurazyme® on BN-PAGE. Four bands were observed for the protein drug and immunoblotting from extracted brain proteins showed a single band at an approx. apparent MW above 480 kDa. The effect on electrophoretic shift by enzymatic treatments is revealed.

3.2. *Blue native PAGE (BN-PAGE)*

BN-PAGE was performed in order to show protein complexes/aggregates in the commercially available recombinant IDUA, Aldurazyme®.

As shown in Fig. 2 four major complexes were revealed. Although the apparent molecular weight on denatured gels is approx. 85 kDa, on native gels electrophoretic mobility is changed. The monomeric band was located at about 142 kDa (marker band). Two complexes were located between 242 kDa and 480 kDa and the highest observed apparent molecular weight was above the 480 kDa marker protein. Intensity of all bands in lanes treated with PNGase F, sialidase and phosphatase (lane 2, 3 and 4) was comparable to un-treated Aldurazyme® (lane 1). The lane with PNGase F-treated Aldurazyme® showed slightly shifted bands to lower molecular weight.

A single complex for IDUA at the apparent molecular weight higher than 480 kDa was observed on BN-PAGE immunoblotting in human brain.

3.3. *Activity of IDUA*

As shown in Fig. 3 (diamonds) and Supplemental Table 4, IDUA activity was increasing with increasing substrate concentration. To determine the role of posttranslational modifications for IDUA activity, PNGase F-treatment was carried out and resulted into decreasing enzyme activity (triangles). Sialidase treatment did not change enzyme activity (cross) while phosphatase treatment reduced IDUA activity (squares).

Aldurazyme® complexes, eluted from BN-PAGE, were all enzymatically active as shown in Supplemental Fig. 7. In contrast to the native, non-eluted control, Aldurazyme® eluted from the four bands showed reduced activity, reflecting loss of activity by the elution process.

4. Discussion

The major findings of the study show that the recombinant IDUA protein drug Aldurazyme® is a protein with a series of modifications and that it consists of several enzyme complexes in contrast to the presence of a single complex in human brain.

Results from 2DE showing trains and electrophoretic shifts for the protein, already point to the presence of several modifications including glycosylations including sialylation, deamidations, oxidations, hydroxylation and methylation.

The presence of glycosylations has been reported before [25,26] but no other modifications are included in the UniProtKB database (P35475, May 16th 2012); deglycosylation but not de-sialylation led to decreased IDUA activity. In addition, herein we add mass spectrometrical informations on the carbohydrate moiety: Herein 8 kinds of carbohydrate structures of which 3 consisted of mannose and 5 of sialic acid-conjugated glycan

Table 1 – Identification and matching peptides of Aldurazyme® using multi-enzyme digestion from Mascot v2.2.06 software.

Enzyme	Sequence	Observed	Mr(expt)	Mr(calc)	Delta
Pepsin	28 P.AEAPHLVHVDA.A 36 (Ions score 43)	579.7831	1157.5516	1157.5829	−0.0313
Pepsin	37 A.ARALWPLRRF.W 46 (Ions score 34)	429.2422	1284.7048	1284.7567	−0.0519
Chymotrypsin	47 F.WRSTGFCPPLPHSQADQY.V 64 (Ions score 59)	716.3044	2145.8914	2145.9691	−0.0777
Chymotrypsin	65 Y.VLSWDQQLNL.A 74 (Ions score 32)	608.2852	1214.5558	1214.6295	−0.0737
Pepsin	75 L.AYVGAVPHRGIKQVRTHWL.L 93 (Ions score 45)	730.0725	2187.1957	2187.2178	−0.0221
Trypsin	90 R.THWLLELVTTR.G 100 (Ions score 91)	684.9587	1367.9028	1367.7561	0.1467
Chymotrypsin	93 W.LLELVTTRGSTGRGL.S 107 (Ions score 56)	786.9093	1571.804	1571.8995	−0.0955
Pepsin	121 L.LRENQLLPGFE.L 131 (Ions score 45)	658.3422	1314.6698	1314.6932	−0.0233
Trypsin	123 R.ENQLLPGFELMGSASGHFTDFEDK.Q 146 (Ions score 30)	890.4747	2668.4023	2668.2115	0.1907
Chymotrypsin	141 F.TDFEDKQQVF.E 150 (Ions score 45)	628.8526	1255.6906	1255.5721	0.1186
Trypsin	147 K.QQVFEWK.D 153 (Ions score 24)	482.8117	963.6088	963.4814	0.1274
Trypsin	154 K.DLVSSLAR.R 161 (Ions score 52)	430.778	859.5414	859.4763	0.0651
Pepsin	160 L.ARRYIGRYGLAHVSKWNF.E 177 (Ions score 20)	732.0532	2193.1378	2193.1708	−0.0330
Subtilisin	182 N.EPDHHDFDN.V 190 (Ions score 17)	413.72	1238.1382	1238.2836	−0.1454
Chymotrypsin	189 F.DNVSMTMQGF.L 198 (Ions score 21)	612.1013	1222.188	1222.3613	−0.1733
Chymotrypsin	199 F.LNYYDACSEGL.R 209 (Ions score 76)	652.8429	1303.6712	1303.539	0.1322
Pepsin	202 Y.YDACSEGLRAASPALRLGGPGDSF.H 225 (Ions score 107)	823.0213	2466.0421	2466.1598	−0.1177
Trypsin	218 R.LGGPGDSFHTPPR.S 230 (Ions score 44)	669.3475	1336.6804	1336.6524	0.0280
Trypsin	231 R.SPLSWGLLR.H 239 (Ions score 65)	514.7337	1027.4528	1027.5815	−0.1286
Trypsin	240 R.HCHDGTNFFTGEAGVR.L 255 (Ions score 29)	902.9126	1803.8106	1803.7748	0.0359
Trypsin	256 R.LDYISLHR.K 263 (Ions score 41)	508.8786	1015.7426	1015.5451	0.1976
Trypsin	268 R.SSISILEQEK.V 277 (Ions score 81)	567.3009	1132.5872	1132.5975	−0.0103
Pepsin	274 L.EQEKVVAQQIRQLF.P 287 (Ions score 30)	858.4472	1714.8798	1714.9366	−0.0568
Pepsin	288 F.PKFADTPIYNDEADPLVG.W 305 (Ions score 53)	981.3996	1960.7846	1960.9418	−0.1572
Pepsin	306 G.WSLPQPWRADVTY.A 318 (Ions score 50)	809.8801	1617.7456	1617.794	−0.0483
Trypsin	314 R.ADVTYAAMVVK.V 324 (Ions score 62)	389.8595	1166.5567	1166.6005	−0.0439
Pepsin	322 M.VVKVIAQHQNLL.L 333 (Ions score 27)	681.3942	1360.7738	1360.8191	−0.0452
Chymotrypsin	344 Y.ALLSNDNAF.L 352 (Ions score 35)	482.774	963.5334	963.4661	0.0673
Pepsin	353 F.LSYHPHPFAQRTLT.A 366 (Ions score 30)	556.6244	1666.8514	1666.858	−0.0066
Chymotrypsin	382 L.LRKPVLTAMGLL.A 393 (Ions score 22)	656.4345	1310.8544	1310.8108	0.0437
Chymotrypsin	394 L.ALLDEEQLW.A 402 (Ions score 61)	558.794	1115.5734	1115.5499	0.0236
Pepsin	402 L.WAEVSQAGTVL.D 412 (Ions score 75)	580.7842	1159.5538	1159.5873	−0.0335
Chymotrypsin	422 L.ASAHRPQGPADAW.R 434 (Ions score 26)	682.3568	1362.699	1362.6429	0.0562
Chymotrypsin	435 W.RAAVLIY.A 441 (Ions score 21)	403.2708	804.527	804.4857	0.0413
Trypsin	436 R.AAVLIYASDDTR.A 447 (Ions score 97)	647.7759	1293.5372	1293.6565	−0.1192
Trypsin	448 R.AHPNRSVAVTLR.L 459 (Ions score 21)	673.889	1345.7634	1345.7942	−0.0308
Pepsin	456 A.VTLRLRGVPPGPGLVY.V 471 (Ions score 17)	565.3292	1692.9658	1693.0039	−0.0382
Pepsin	472 Y.VTRYLDNGLCSPDGEW.R 487 (Ions score 67)	941.3592	1880.7038	1880.8363	−0.1325
Pepsin	487 E.WRRLGRPVFPTAEQ.F 500 (Ions score 20)	571.6445	1711.9117	1711.927	−0.0154
Pepsin	501 Q.FRRMRAAEDPVAAAPRPLPAGGRL.T 524 (Ions score 29)	859.1383	2574.3931	2574.4077	−0.0146
Chymotrypsin	525 L.TLRPALRLPSLL.L 536 (Ions score 25)	675.3947	1348.7748	1348.8554	−0.0806
Trypsin	532 R.LPSLLLVHVCARPEKPPGQVTR.L 553 (Ions score 78)	823.4177	2467.2313	2467.3733	−0.1421
Pepsin	536 L.LLVHVCARPEKPPGQVTRL.R 554 (Ions score 29)	724.0632	2169.1678	2169.2205	−0.0527
Pepsin	555 L.RALPLTQGQLVL.V 566 (Ions score 46)	654.8446	1307.6746	1307.7925	−0.1179
Trypsin	556 R.ALPLTQGQLVLVWSDEHVGSK.C 576 (Ions score 111)	1139.1137	2276.2128	2276.2165	−0.0036
Trypsin	577 K.CLWTYEIQFSQDGK.A 590 (Ions score 104)	887.8859	1773.7572	1773.8032	−0.0460
Pepsin	583 E.IQFSQDGKAYTPVSRKPSTFNL.F 604 (Ions score 51)	828.7536	2483.239	2483.2809	−0.0419
Pepsin	605 L.FVFSPDTGAVSGSY.R 618 (Ions score 91)	717.3006	1432.5866	1432.6511	−0.0644
Chymotrypsin	619 Y.RVRALDYW.A 626 (Ions score 27)	539.7917	1077.5688	1077.5719	−0.0031
Pepsin	626 Y.WARPGPFSDPVPYLEVPVPRGPPSPGNP.- (Ions score 47)	994.7872	2981.3398	2981.5188	−0.1791

structure were detected and all parts showed similar proportions of each carbohydrate resembling and extending previous findings [25,26].

It has to be addressed, however, that modifications including amino acid exchanges, are not observed in the potential active sites of IDUA (UniProtKB, Supplemental Fig. 1). Deamidations may be responsible for the train on the 2DE although it is not clear whether deamidation resulted from technological processing of Aldurazyme® or is a post-translational modification (Supplemental Table 2). Likewise, proline hydroxylations may be due to protein oxidation or are a result of hydroxylase activities acting on proline. Protein methylation was shown to be a post-translational modification or a result from technological processing of Aldurazyme®, but is not a result from the analytical procedure from the gel-based proteomics method as replacing methanol by ethanol did not change the electrophoretic shift (Supplemental Fig. 3).

The novel finding of two serine phosphorylation sites may play a role for protein function and indeed, dephosphorylation lead to decreased IDUA activity. As phosphorylations are not in

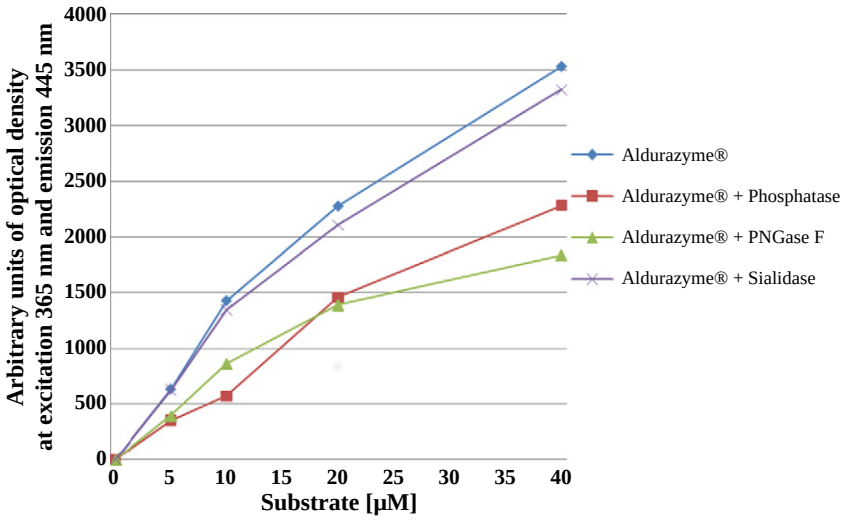


Fig. 3 – Enzyme activity of Aldurazyme® after PNGase F, sialidase and phosphatase treatment. PNGase F- and phosphatase-treatment of Aldurazyme® showed reduction of enzyme activity.

the active site of Aldurazyme®, an allosteric effect has to be considered.

As to amino acid substitutions, a large amount of natural variants were reported that were linked to mucopolysaccharidoses [27].

Given the importance of protein aggregates or complexes, it was shown that four complexes exist in the commercially available preparation. All complexes observed showed comparable IDUA activity as did the single complex at about 480 kDa observed in human brain. Clement and coworkers [28] made a major contribution to aggregate formation revealing the dependence of aggregate formation on ionic strength and pH, which has to be taken into consideration for the interpretation of the current results. The biological meaning of complex/aggregate formation in the protein drug remains elusive but it cannot be excluded that it accounts for some side effects described for IDUA therapy.

Taken together, full characterization of the Aldurazyme® protein was provided allowing further chemical and immunochemical studies including fair generation of antibodies, last but not least due to the identification of amino acid substitutions. A possible role for the presence of glycosylation and phosphorylations was proposed although consequences of other modifications remain unclear. The study is relevant for design of future and understanding of previous work on IDUA.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2012.09.022.

Disclosure statement

Herewith the authors state that there is no conflict of interests.

Acknowledgements

We acknowledge the contribution by the Verein zur Durchführung der wissenschaftlichen Forschung auf dem Gebiet der Neonatologie und Kinderintensivmedizin “Unser Kind”.

REFERENCES

[1] Clarke LA, Wraith JE, Beck M, Kolodny EH, Pastores GM, Muenzer J, et al. Long-term efficacy and safety of laronidase in the treatment of mucopolysaccharidosis I. Pediatrics 2009;123:229-40.
[2] Valayannopoulos V, Wijburg FA. Therapy for the mucopolysaccharidoses. Rheumatology (Oxford) 2011;50(Suppl. 5):v49-59.
[3] Neufeld EF, Muenzer J. The mucopolysaccharides. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Childs B, Kinzler KW, et al, editors. The metabolic and molecular bases of inherited disease. 8th ed. New York, NY, USA: McGraw-Hill, Medical Publishing Division; 2001. p. 3421-52.
[4] Coutinho MF, Lacerda L, Alves S. Glycosaminoglycan storage disorders: a review. Biochem Res Int 2012;2012:471325.
[5] Wraith JE. Enzyme replacement therapy in mucopolysaccharidosis type I: progress and emerging difficulties. J Inherit Metab Dis 2001;24:245-50.
[6] Colavita N, Orazi C, Fileni A, Leone PC, Ricci R, Segni G. A further contribution to the knowledge of mucopolysaccharidosis I H/S compound. Presentation of two cases and review of the literature. Australas Radiol 1986;30:142-9.
[7] Roubicek M, Gehler J, Spranger J. The clinical spectrum of alpha-L-iduronidase deficiency. Am J Med Genet 1985;20: 471-81.
[8] Kakkis ED, Muenzer J, Tiller GE, Waber L, Belmont J, Passage M, et al. Enzyme-replacement therapy in mucopolysaccharidosis I. N Engl J Med 2001;344:182-8.
[9] Kakkis ED, Schuchman E, He X, Wan Q, Kania S, Wiemelt S, et al. Enzyme replacement therapy in feline mucopolysaccharidosis I. Mol Genet Metab 2001;72:199-208.
[10] Valayannopoulos V, Boddaert N, Barbier V, Le Merrer M, Caillaud C, de Lonlay P. Cognitive and neuroradiological improvement in three patients with attenuated MPS I treated by laronidase. Mol Genet Metab 2010;100:20-3.
[11] Zhao KW, Faull KF, Kakkis ED, Neufeld EF. Carbohydrate structures of recombinant human alpha-L-iduronidase

secreted by Chinese hamster ovary cells. J Biol Chem 1997;272:22758-65.
[12] Schellekens H. Factors influencing the immunogenicity of therapeutic proteins. Nephrol Dial Transplant 2005;20(Suppl. 6): vi 3–9.
[13] Jefferis R. Aggregation, immune complexes and immunogenicity. MAbs 2011;3:503-4.
[14] Manning MC, Patel K, Borchardt RT. Stability of protein pharmaceuticals. Pharm Res 1989;6:903-18.
[15] Sauerborn M, Brinks V, Jiskoot W, Schellekens H. Immunological mechanism underlying the immune response to recombinant human protein therapeutics. Trends Pharmacol Sci 2010;31:53-9.
[16] Kakavanos R, Turner CT, Hopwood JJ, Kakkis ED, Brooks DA. Immune tolerance after long-term enzyme-replacement therapy among patients who have mucopolysaccharidosis I. Lancet 2003;361:1608-13.
[17] Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 1976;72:248-54.
[18] John JP, Oh JE, Pollak A, Lubec G. Identification and characterisation of arsenite (+3 oxidation state) methyltransferase (AS3MT) in mouse neuroblastoma cell line N1E-115. Amino Acids 2008;35:355-8.
[19] Kang SU, Fuchs K, Sieghart W, Pollak A, Csaszar E, Lubec G. Gel-based mass spectrometric analysis of a strongly hydrophobic GABAA-receptor subunit containing four transmembrane domains. Nat Protoc 2009;4:1093-102.
[20] Chen WQ, Priewalder H, John JP, Lubec G. Silk cocoon of *Bombyx mori*: proteins and posttranslational modifications–heavy phosphorylation and evidence for lysine-mediated cross links. Proteomics 2010;10:369-79.
[21] Pabst M, Bondili JS, Stadlmann J, Mach L, Altmann F. Mass + retention time = structure: a strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans. Anal Chem 2007;79:5051-7.
[22] Pabst M, Grass J, Fischl R, Leonard R, Jin C, Hinterkorner G, et al. Nucleotide and nucleotide sugar analysis by liquid chromatography-electrospray ionization-mass spectrometry on surface-conditioned porous graphitic carbon. Anal Chem 2010;82:9782-8.
[23] Chen WQ, Salmazo A, Myllykoski M, Sjoblom B, Bidlingmaier M, Pollak A, et al. Purification of recombinant growth hormone by clear native gels for conformational analyses: preservation of conformation and receptor binding. Amino Acids 2010;39:859-69.
[24] Ghafari M, Hoger H. Keihan Falsafi S, Russo-Schlaff N, Pollak A, Lubec G. Mass spectrometrical identification of hippocampal NMDA receptor subunits NR1, NR2A-D and five novel phosphorylation sites on NR2A and NR2B. J Proteome Res 2012;11:1891-6.
[25] Robbins AR, Myerowitz R. The mannose 6-phosphate receptor of Chinese hamster ovary cells. Compartmentalization of acid hydrolases in mutants with altered receptors. J Biol Chem 1981;256:10623-7.
[26] Tsukimura T, Tajima Y, Kawashima I, Fukushige T, Kanzaki T, Kanekura T, et al. Uptake of a recombinant human alpha-L-iduronidase (laronidase) by cultured fibroblasts and osteoblasts. Biol Pharm Bull 2008;31:1691-5.
[27] Yogalingam G, Guo XH, Muller VJ, Brooks DA, Clements PR, Kakkis ED, et al. Identification and molecular characterization of alpha-L-iduronidase mutations present in mucopolysaccharidosis type I patients undergoing enzyme replacement therapy. Hum Mutat 2004;24:199-207.
[28] Clements PR, Brooks DA, Saccone GT, Hopwood JJ. Human alpha-L-iduronidase. Purification, monoclonal antibody production, native and subunit molecular mass. Eur J Biochem 1985;152:21-8.

Appendix X

Neumann, L., Thader, A., Wozny, K., Pabst, M., Altmann, F., (2013): Isomer specific analysis of mouse brain N-glycans by LC-MS (Talk) [APRS-Meeting 2013, Innsbruck, Austria, September 23-25, 2013]

Appendix XI

Neumann, L., Krammer, F., de Graaf, M., Hokke, C.H., Altmann, F., (2013): Binding Specificity of the H1N1 Swine-Origin Influenza A Virus (S-OIV) (Poster), [Glyco 22, Dalian, China, June 22-29, 2013]

Binding Specificity of the H1N1 Swine-Origin Influenza A Virus (S-OIV)

Laura NEUMANN[1], Florian KRAMMER[2], Miranda de GRAAF[3], Cornelis H. HOKKE[4], Friedrich ALTMANN[1]

[1]University of Natural Resources and Life Sciences, Vienna, 1190, Austria; [2]Icahn School of Medicine at Mount Sinai, New York, 10029, USA; [3]ErasmusMC, Rotterdam, 3015GE, The Netherlands; [4]Leids Universitair Medisch Centrum (LUMC), Leiden, 2333ZA, The Netherlands
E-mail: friedrich.altmann@boku.ac.at

Influenza A viruses are known to attack their hosts by an initial binding to a glycoconjugate containing receptor. It is widely known that the crucial element of the glycoconjugate for Influenza A virus binding is N-acetyl neuraminic acid (sialic acid). Depending on the species origin hemagglutinins differ in their prevalence for sialic acids in different linkages (e.g. α2,6-linkage found in the human nasopharyngeal tract, α2,3-linked sialic acid found in avian tissues and α2,6- as well as α2,3-linked sialic acid in porcine tissues).

In this study we investigated the binding specificity of the recombinantly expressed H1 hemagglutinin from the H1N1 Influenza A virus (California/04/09) and other hemagglutinins (H3 of H3N2 HK68 and H1 of H1N1 NewCaledonia99) as controls to N-Glycans isolated from mammalian tissue of the nasopharyngeal tract.

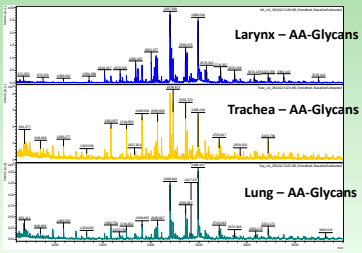
The isolated and fluorescent (AA-anthranillic acid) labelled N-Glycans were three dimensionally separated. The first dimension was according to charge (weak anion exchange), due to the number of sialic acids (neutral – no sialic acid, singly charged – one sialic acid and doubly charged – two sialic acids) followed by subsequent Normal phase separation (according to size) and in a third dimension a Reversed phase separation (according to hydrophobicity) was performed. MALDI-TOF-MS checked N-Glycan fractions were printed on a glycanarray and recombinant hemagglutinins were tested for distinct binding. The results confirm the prevalence of different hemagglutinins towards α2,3- or α2,6-linked sialic acids though the binding pattern within all hemagglutinins is rather broad as neutral and singly charged glycans are also bound. This might result from the hemagglutinin glycosylation (small complex uncharged glycosylation) itself.

Sample preparation

Membrane protein isolation
⇒ tryptic digest
⇒ PNGase F digest to release N-Glycans
⇒ AA (anthranilic acid) labeling of Glycans for subsequent separation and printing of array

Workflow

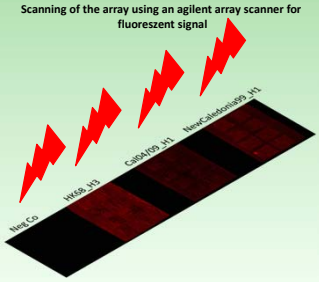
MALDI-TOF Glycan pattern of

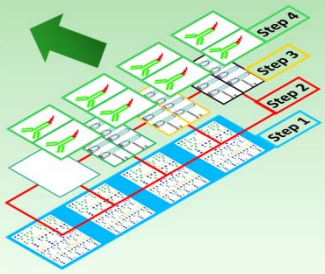


Glycans of different tissues were pooled for subsequent HPLC separation

Glycanarray – the assay

Scanning of the array using an agilent array scanner for fluoreszent signal





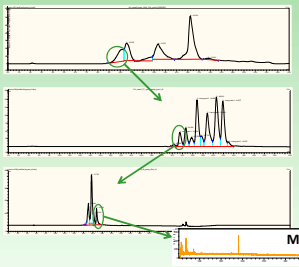
Step 4:
Incubation with Alexa Fluorophor 647 anti His Antibody

Step 3:
Incubation of array with different soluble recombinant hemagglutinins (HIS-tag) and a negative control (Neg.Co)

Step 2:
Arrays separated by silicon gasket for the assay

Step 1:
Glycans printed on epoxy-silane coated glassslides => 4 Glycanarrays per slide

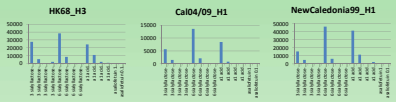
Sample separation



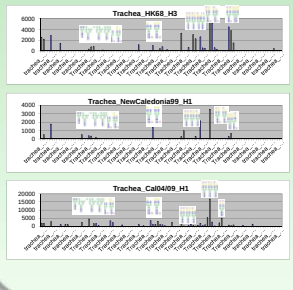
Results

Figure 1 – ferret array

a.) binding to synthetic positive controls



b.) binding N-Glycans from ferret trachea



c.) Cluster analysis

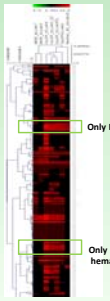
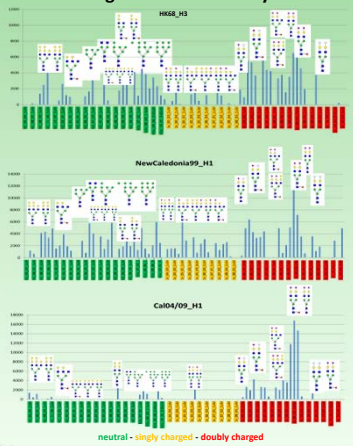


Figure 2 – swine array



Conclusion

Recombinantly expressed hemagglutinins show a fairly broad binding pattern through out both arrays. Neutral as well as singly and doubly charged glycan structures were bound, though the most intense binding could be detected towards the doubly sialylated structures.

The broad binding corroborates the finding [1] that the type of glycosylation of the hemagglutinin might play a crucial role for binding specificity.

Therefore differentially (complex galactosylated and sialylated) glycosylated hemagglutinin will be tested on both arrays.

[1] de Vries R.P. et al. Glycan-Dependent Immunogenicity of Recombinant Soluble Trimeric Hemagglutinin. J Virol. 2012 November; 86(21): 11735–11744.

Figure 1 & 2:
Figure **1a** shows the binding profile of three hemagglutinins (HK68_H3, Cal04/09_H1 and NewCaledonia99_H1) to synthetic standards printed as positive controls on the array. HK68_H3 binds slightly more to the 2-6 linked sialic acid containing control whereas the two H1 hemagglutinins clearly prefer the α2-6 linked sialic acid over α2-3 linked sialic acids. The specificity could also be shown in Figure **1b** indicating the broader binding of the HK68_H3 to the sialic acid containing fractions on the array in contrast to the two H1 hemagglutinins.
Figure **1c** depicts cluster analysis using the MeV (Multi experiment Viewer) free available software. The assay replica of the glycanarray experiments cluster as expected together whereas different glycan clusters could be detected e.g. binding solely to H1 hemagglutinins or only Cal04/09_H1 bound glycan fractions. The results of the hemagglutinin binding to glycans from swine nasopharyngeal tissue can be seen in Figure **2**. Binding pattern of HK68_H3, Cal04/09_H1 and NewCaledonia99_H1 – hemagglutinins to neutral (green), singly charged (yellow) and doubly charged (red) glycans is depicted.
Again a rather broad binding pattern of the hemagglutinins to the different glycan-fractions (charged and neutral) can be seen, though the prevalence of H1 hemagglutinins towards α2-6 linked sialic acid is clearly visible as this is the highest abundant occuring linkage in swine tissue.

BOKU VIBT Vienna Institute of BioTechnology · Mount Sinai School of Medicine · BioToP Biomolecular Technology of Proteins · LUMC Leids Universitair Medisch Centrum · FWF

Appendix XI

Neumann, L., Thader, A., Altmann, F., Pabst, M. (2012): Comparison of N-Glycans of different Influenza Virus-susceptible tissues using PGCLC-ESI-MS and stable isotope labelled internal standards (Poster)[Summer Course Glycoscience, 12th European Training Course on Carbohydrates, Groningen, The Netherlands, June 3-7, 2012]

Comparison of N-Glycans of different Influenza Virus-susceptible tissues using PGC-LC-ESI-MS and stable isotope labeled internal standards

Laura Neumann, Andreas Thader, Friedrich Altmann* and Martin Pabst
Department of Chemistry, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria
*correspondence should be addressed to friedrich.altmann@boku.ac.at

Introduction:

The relative distribution of glycoforms with epitopes such as sialic acids or different types of fucosylation determines the infectiousness of different pathogens for species or individuals.
The hemagglutinin H1 of the H1N1 swine origin influenza A virus is known to bind sialic acid-containing receptors on the host cell surface thereby initiates the infection process. This binding is highly species specific (to Neu5Ac α2,6-linked in humans, α2,3-linked in birds and α2,6 or α2,3 in swine) allowing a classification of influenza virus strains according to their host and receptor species.
In order to obtain a more detailed insight in the true receptor structure, the N-glycan pool of different swine tissues from the respiratory tract (trachea, lung and larynx) was isolated and analyzed.

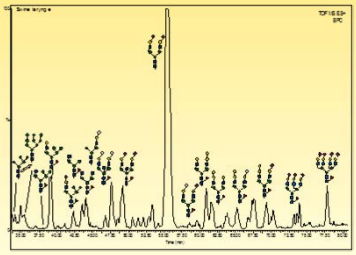
Sample preparation

Membrane protein preparation of different swine tissues

SDS-PAGE and in-gel PNGaseF digest of whole protein-extract. Reduction of released N-glycans

Workflow I

Glycan pre-purification by SPE-PGC



Glycans from swine larynx

Main structures assigned, not assigned peaks indicate all different isoforms. Glycans were analyzed using PGC-LC-ESI-MS (+ stable isotope internal standards)

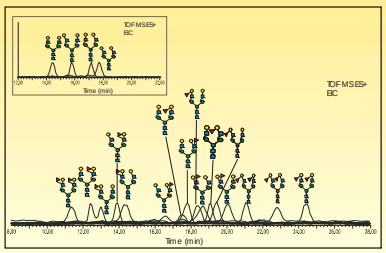
Standard / internal standard generation

were purified from bovine fetuin and used for enzyme assays

standards were generated using recombinantly expressed enzymes (FucT I, FucT III, FucT IV, ST3Gal IV, ST6Gal I)

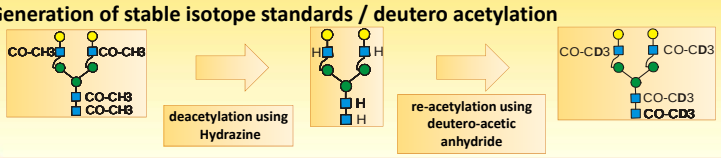
Workflow II

Glycan pre-purification by SPE-PGC



samples are analyzed using PGC-LC-ESI-MS (+ stable isotope internal standards)

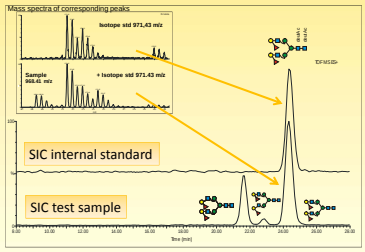
Generation of stable isotope standards / deutero acetylation



Results

Figure 1

Advantage of deutero-acetylated internal standards



Co-elution of corresponding structures

Figure 1:
For structural assignments the addition of internal standards is essential. Thereby we deuterated / deutero-acetylated glycan standards. In Figure 1 single ion chromatogramms of the enzymatically generated real sample (test sample = A^4A^4 + Fut1 - double fucosylated) and the corresponding stable isotope labeled internal standard are shown. Due to the addition of deutero-acetylated A^4A^4 (double fucosylated) as an internal standard, the resulting glycanstructures of the enzyme assay can be easily identified as the same structure as they are eluting at exactly the same time though differing in the mass.

Figure 2

Sialic acid linkage in swine larynx

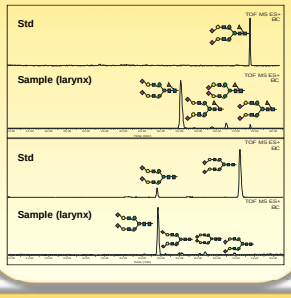


Figure 2:
According to the retention time (RT) of the standards, sialylated structures of swine tissue can be assigned to a mixture of different sialic acid linkages with the highest abundance of α2,6 linked sialic acids to a core α1,6 fucosylated or not fucosylated biantennary N-Glycan structure.

Figure 3

Structural assignment of fucosylated glycans in IgA (S-chain)

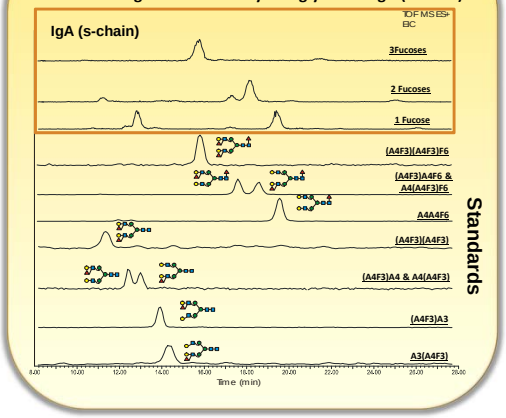


Figure 3:
Comparing the retention time (RT) of the desialylated IgA S-chain protein (Figure 3) the single fucosylated structures can be assigned to $A^4A^4F^6$ as well as $(A^4F^3)A^4$ or $A^4(A^4F^3)$. The double fucosylated A^4A^4 found in the S-chain corresponds to $(A^4F^3)(A^4F^3)$ and $(A^4F^3)A^4F^6$ & $A^4(A^4F^3)F^6$. The only trifucosylated structure found in this protein represents, according to the retention, the core-fucosylated $A^4A^4F^6$ with two additional α1,3 linked fucoses to the antennary GlcNAc's.

Conclusion

The combination of a chromatographic separation by PGC-LC with the addition of deutero-acetylated internal standards, together with the detection of eluted glycan structures by mass spectrometry provides an easy and accurate tool for structural investigation of glycans.

BOKU VIBT Vienna Institute of BioTechnology
BioToP Biomolecular Technology of Proteins
FWF

Appendix XII

Neumann, L., Thader, A., Altmann, F., Pabst, M., (2012): Structural investigation of fucose epitopes on N-Glycans by PGC-ESI-MS using the example of secretory IgA and mouse brain glycans, (Poster) [Carbohydrate analysis and glycomics: where next?, London, UK, March 21, 2012]

Structural investigation of fucose epitopes on N-Glycans, using PGC-ESI-MS and stable isotope internal standards

Laura Neumann, Andreas Thader, Friedrich Altmann* and Martin Pabst
Department of Chemistry, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria
*correspondence should be addressed to friedrich.altmann@boku.ac.at

Introduction

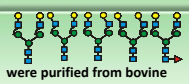
Fucosylation is a highly common feature in protein glycosylation of many species. Beside the importance of α1-2-fucosyltransferase in the ABO blood group antigen formation, fucosylated glycans can play a key role in host-pathogen interactions. A prominent example is secretory IgA which consists of a secretory component covalently linked to the dimeric IgA.
In this study we present a PGC-ESI-MS method providing the possibility for detailed structural investigation of the different possible fucosylations in N-Glycans, by using stable isotope internal standards. Recombinantly expressed fucosyl-transferases Fut1 (α1-2 to Gal), Fut3 (α1-4 to GlcNAc) and Fut4 (α1-3 to GlcNAc) were expressed in insect cells using the baculo virus expression system. Stable isotopes were introduced to the N acetyl groups by a two step chemical reaction.

Workflow

Sample preparation

SDS-PAGE of protein extract
In-gel PNGaseF digest
reduction of N-glycans

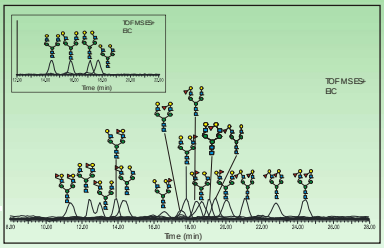
Standard / internal standard generation



were purified from bovine fetuin and used for enzyme assays

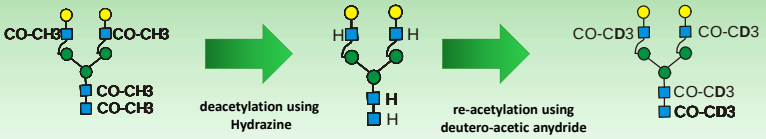
standards were generated using recombinantly expressed enzymes (Fut1, Fut3 and Fut4) in insect cells (SF9 cells)

Glycan pre-purification by SPE-PGC



samples are analyzed using PGCC-ESI-MS (+ stable isotope internal standards)

Generation of stable isotope standards / deutero acetylation



Examples, IgA S chain and H chain

Figure 1

Extracted Ion Chromatogramm (EIC) of IgA (S-chain)

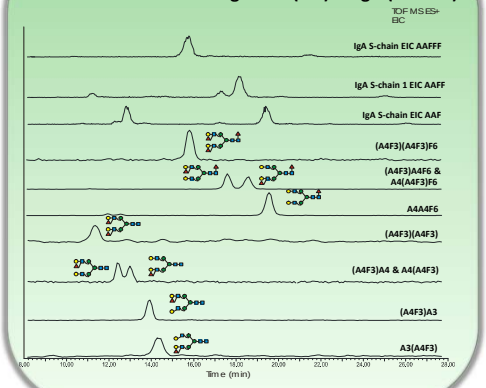


Figure 2

Extracted Ion Chromatogramm (EIC) IgA (H-chain)

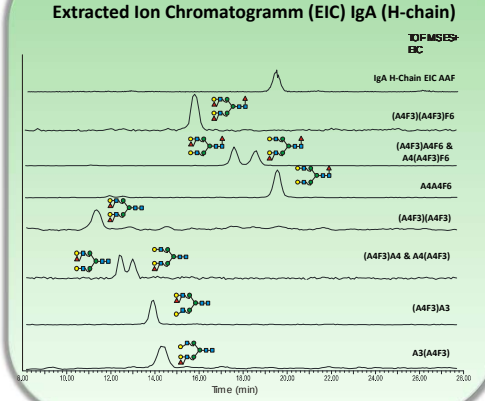
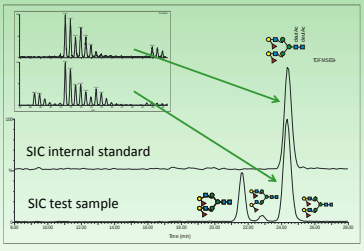


Figure 1 & 2:
Comparing the retention time (RT) of the desialylated IgA S-chain protein (Figure 1) the single fucosylated structures can be assigned to $A^4A^4F^6$ as well as $(A^4F^3)A^4$ or $A^4(A^4F^3)$. The double fucosylated A^4A^4 found in the S-chain corresponds to $(A^4F^3)(A^4F^3)$ and $(A^4F^3)A^4F^6$ & $A^4(A^4F^3)F^6$.
The only trifucosylated structure found in this protein represents, according to the retention, the core-fucosylated $A^4A^4F^6$ with two additional α1-3 linked fucoses to the antennary GlcNAc's.
As a positive control we also investigated the deialylated heavy chain of IgA. According to the RT of the single fucosylated A^4A^4, this structure can be assigned to $A^4A^4F^6$, which is widely known from the literature.

Figure 3

Advantage of deutero-acetylated internal standards



Co-elution of corresponding structures

Figure 3:
For structural assignments the addition of internal standards is essential. Thereby we deuterated / deutero-acetylated glycan standards. In Figure 3 single ion chromatogramms of the enzymatically generated real sample (test sample = A^4A^4 + Fut1 - double fucosylated) and the corresponding stable isotope labeled internal standard are shown. Due to the addition of deutero-acetylated A^4A^4 (double fucosylated) as an internal standard, the resulting glycanstructures of the enzyme assay can be easily identified as the same structure as they are eluting at exactly the same time though differing in the mass.

Conclusion

The combination of a chromatographic separation by PGC with the addition of deutero-acetylated internal standards together with the detection of eluted glycan structures by mass spectrometry provides an easy and accurate tool for structural investigation of highly fucosylated glycans.

BOKU VIBT Vienna Institute of BioTechnology
BioToP Biomolecular Technology of Proteins
FWF

10 Abbreviations

AA	2-aminobenzoic acid, anthranilic acid
A	Galactose
2-AB	2-aminobenzamide
ACN	Acetonitril
An	N-acetyl galactosamine
ANTS	2-aminonaphthalene trisulfonic acid
APTS	1-aminopyrene-3,6,8-trisulfonic acid
BSA	Bovine serum albumin
CID	Collision induced dissociation
ConA	*Concanavalin A* agglutinin
DTT	Dithiothreitol
ESI	Electron spray ionisation
EtOH	Ethanol
Fuc	Fucose
GAG	Glycosaminoglycan
Gal	Galactose
GalNAc	N-acetyl galactosamine
GlcNAc	N-acetyl glucosamine
Gn	N-acetyl glucosamine
GnT	N-acetyl glucosamine transferase
HA	Hemagglutinin
HCl	Hydrogen chloride
HILIC	Hydrophilic interaction chromatography
HPLC	High performance liquid chromatography
IT	Ion trap
LacNAc	N-acetyl lactosamine
LC	Liquid chromatography
LCA	*Lens culinaris* agglutinin
MALDI	Matrix assisted laser desorption ionisation
Man	Mannose
MetOH	Methanol
MS	Mass spectrometry
Na	N-acetyl neuraminic acid
$NaBH_4$	Sodium boro hydride
NaCl	Sodium chloride
NaOH	Sodium hydroxide
Ng	N-glycoly neuraminic acid
NHS	N-Hydroxysuccinimidyl
Ni-NTA	Nickel-nitrilotriacetic acid
NP	Normal phase
PA	2-aminopyridine
PBS	Phosphate buffered saline
PGC	Porous graphitized carbon
PNGase F/A	Peptide-N4-(N-acetyl-beta-glucosaminyl) asparagine amidase F/A
QTOF	Quadrupol time of flight
RP	Reversed phase
RRT	Relative retention time
SDS PAGE	sodium dodecyl sulphate poly acrylamide gel electrophoresis
SPE	Solid phase extraction
WAX	Weak anion exchange

11 Curriculum vitae

Name Laura Neumann, Mag.

Date of Birth 29.03.1985

Nationality Austria

Education

09.2010-10.2014 PhD thesis within the International PhD Programme in Protein Biotechnology "BioToP" at the University of Natural Resources and Life Science (BOKU), Vienna

"*Isomer specific analysis of N-glycans as a prerequisite for understanding biological functions applied to influenza A virus receptors and mouse brain glycoproteins"*

Supervisor: Ao.Univ.Prof. Dipl.-Ing. Dr.nat.techn. Friedrich Altmann

10.2003 – 11.2009 Master Degree (Mag.rer.nat) in "Nutritional Sciences" at the Department of Medical Chemistry, University of Vienna

"Influence of Hemoglobin Synthesis on the Expression of Frataxin"

Supervision: Univ. Prof. Dr. Hans Goldenberg, Assoc. Prof. PD Dr. Brigitte N. Sturm

10.1995 – 07.2003 High School, Gymnasium ORG Orth des Schulvereins der Kreuzschwestern, Gmunden

Reserach stay abroad

10.1012 – 04.2013 Research stay at the LUMC Leids Universitair Medisch Centrum, Leiden, The Netherlands
Major Methods: Glycanarrays, HPLC, MALDI-TOF-MS
Supervision: Dr. Cornelis H. Hokke.

Language skills

German (mother tongue), English (fluent), French (basics), Spanish (basics)