Chemo-enzymatic synthesis and analysis of isomeric glycans

Gaël Vos

Doctoral thesis, Utrecht University

Chemo-enzymatic synthesis and analysis of isomeric glycans

Chemo-enzymatische synthese en analyse van isomere glycanen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

woensdag 19 april 2023 des middags te 12.15 uur

door

Gaël Martijn Vos

geboren op 4 oktober 1991
te Amersfoort

Promotor:

Prof. dr. G.J.P.H. Boons

Copromotor

Dr. R.P. de Vries

Beoordelingscommissie:

Prof. dr. S. Flitsch

Prof. dr. A.J.R. Heck

Prof. dr. D. Lefeber

Prof. dr. J. Garssen

Prof. dr. R.J. Pieters

Table of contents

List of abbreviations

AAL	Aleuria Aurantia Lectin
Ac	Acetyl
AcOSu	Succinimidyl acetate
Alloc	Allylloxycarbonyl
Asn	Asparagine
ATD	Arrival time distribution
BCoV	Bovine corona virus
Bn	Benzyl
Boc	tert-Butyloxycarbonyl
Boc_2O	tert-Butyloxycarbonylanhydride
BSM	Bovine submaxillary mucin
Cbz	Carboxybenzyl
CCS	Collision cross section
CHO	Chinese hamster ovary
CID	Collision induced dissociation
cIM	Cyclic ion mobility
CMP	Cytidine-5'-monophosphate
COSY	Correlated spectroscopy
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCM	Dichloromethane
DMSO	Dimethyl sulfoxide
Dol	Dolichol
DTIMS	Drift tube ion mobility spectrometry
ECD	Electron-capture dissociation
ECL	Erythrina cristagalli Lectin
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
TFA	Trifluoroacetic acid
ESI	Electronspray ionization
ETD	Electron-transfer dissociation
ExD	Electron activated dissociation
Fmoc	Fluorenylmethyloxycarbonyl
Fuc	Fucose
GAG	Glucosaminoglycan
Gal	Galactose
GalNAc	*N*-Acetylgalactosamine
GDP	Guanosine-5'-diphosphate
GlcN	Glucosamine
GlcNAc	*N*-Acetylglucosamine
GlcNBoc	*N*-tert-Butyloxycarbonylglucosamine
GlcNTFA	*N*-trifluoroacetylglucosamine
GPI	Glycosylphosphatidylinositol
HDFN	Hemolytic disease of the fetus and newborn
HE	Hemagglutinin esterase
HEK	Human embryonic kidney
HexNAc	*N*-Acetylhexose
HILIC	Hydrophilic interaction chromatography
HMO	Human milk oligosaccharide

HNK	Human natural killer
HPLC	High performance liquid chromatography
HRdm	High resolution demultiplexing
HSQC	Heteronuclear single quantum coherence
IM	Ion mobility
IRMPD	Infrared multiphoton dissociation
KS	Keratan sulfate
LacNAc	*N*-Acetyllactosamine
LacNTFA	*N*-Trifluoroacetyllactosamine
LC	Liquid chromatography
Lev	Levulinate
LnNT	Lacto-*N*-neotetraose
LNT	Lacto-*N*-tetraose
LOS	Lipooligosaccharide
MAH	Maackia Amurensis Lectin II
MALDI	Matrix-assisted laser desorption/ionization
Man	Mannose
MHV-S	Mouse Hepatitis Virus Strain S
MS	Mass spectrometry
Nap	2-Naphthylmethyl
Neu5Ac	*N*-Acetylneuraminic acid
Neu5Gc	*N*-Glycolylneuraminic acid
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser Effect spectrometry
PAPS	3'-Phosphoadenosine-5'-phosphosulfate
Pd/C	Palladium on carbon
PGC	Porous graphitic carbon
PMP	1-Phenyl-3-methyl-5-pyrazolone
PrA	Procainamide
QTOF	Quadrupole Time-of-Flight
Ser	Serine
SNA	Sambucus Nigra Lectin
SPE	Solid phase extraction
STD	Saturation transfer difference
TEA	Triethylamine
TEM	Transmission electron microscopy
Glc	Glucose
Thr	Threonine
TIMS	Trapped ion mobility spectrometry
TOCSY	Total correlation spectrometry
TRIS	Tris(hydroxylmethyl)aminomethane
Troc	2,2,2-Trichloroethoxycarbonyl
TWIMS	Travelling wave ion mobility mass spectrometry
UDP	Uridine-5'-diphosphate
UEA	Ulex Europaeus Agglutinin I Lectin
UMP	Uridine-5'-monophosphate
WGA	Wheat germ agglutinin Lectin

CHAPTER 1

General Introduction

The glycocalyx

Carbohydrates, also named glycans, are essential components of all living organisms. All cellular membranes of eukaryotic cells are coated with a dense gel-like network called the glycocalyx (Fig. 1). As the name implies, this viscous matrix consists mainly of carbohydrates attached to various support structures such as protein and lipids. The glycocalyx forms the first point of contact for all bodies in the extracellular space with the cell. It is especially pronounced on epitheleal and endothelial cells, where is can reach a size of several microns. [1–3] The long-established perception of the glycocalyx has been that of a physical barrier, which shields entry receptors on epithelial cells thus protecting from bacteria and virusses.[4,5] In the blood stream, the endothelial glycocalyx influences vascular permeability and prevents leukocyte adhesion.[6,7] Eventually, a more nuanced view on the glycocalyx emerged which recognised the role of carbohydrates in many more processes. These includes cell-cell and cell-ECM signalling interactions as well as fertilization, embryogenesis and differentation. [8–13] Detailed characterization of the glycosylation of the glycocalyx can give important diagnostic information regarding the health status of the tissue since glycosylation can often be disrupted or altered in disease states such as inflammatory and metabolic disorders as well as malignacies.[14–16]

Carbohydrate conjugates including those in the glycocalyx are classified by their underlying support structure: Lipid-linked structures include glycolipids, composed of a glycan that is attached to a ceramide lipid, and GPI-anchored glycoproteins which are proteins linked to a lipid through a carbohydrate linker. Protein-linked glycans can be classefied as *N*-glycans, attached to the side-chain amine of an asparagine, and *O*-glycans, attached to the side-chain oxygen of either threonine or serine. Proteoglycans are highly charged large glucosaminoglycans (GAG) that are linked to specific GAG carrier proteins through a xylosylated serine or threonine. The highly heterogenious architecture of these carbohydrates combined with the diversity in support structures and the structural organization of glycocalyx components results in an incredibly complex matrix. The complexity of these carbohydrates has greatly hindered the in-depth study of exact functional roles of glycocalyx components.

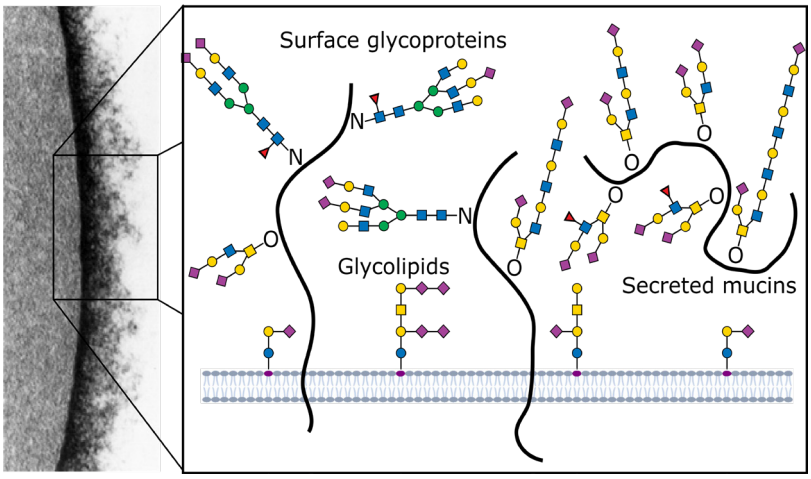


Figure 1: Left) TEM image of the erythrocyte glycocalyx, adapted from Biochemistry (4th edition). Right) Simplified depiction of the diversity of glycoconjugates found on eukaryotic cells.

The sugar code

One of the main challenges in the determination of the exact function of carbohydrates originates from the high heterogeniety in carbohydrate composition, which directly arises from the biosynthetic pathway itself. Complex carbohydrates biosynthesis occurs predominantly within the membrane compartement of the secretory pathway. The transfer of carbohydrates to scaffold structures or previous attached carbohydrates is controlled by glycosyltransferase enzymes that are able to transfer monosaccharides from activated sugar nucleotides to the acceptor structure. The action of glycosyltransferases is not template-driven but controlled by the transcription, localization and turnover rates of the glycosyltransferases, and the availability of sugar nucleotides.[17] Genomic studies revealed that humans have about 230 glycosyltransferase genes, often including multiple enzymes capable of preforming the same glycosylation reaction with different specificities and activities.[18] The action of competing glycosyltransferases for the same substrate and dependencies on other glycosyltransferases to generate the acceptor structures results in high structural heterogeneity among glycans. This high heterogeneity combined with structural similarities between oligosaccharides, makes it very challenging to accurately map the glycome.

The first part of the introduction will adress the structure of glycans found in humans. The second part will discuss the use of glycosyltransferases *in vitro* to generate isomerically pure glycans and techniques to direct glycosyltransferase activity towards single product formation. The third part will describe analytical techniques for the structural characterization of glycans and the need for pure glycan standards to study the glycome in greater detail.

Structure of carbohydrates

Complex carbohydrates are oligomers composed of monosaccharide building blocks that are linked through different sequences, positions and linkage types. Most common monosaccharides in humans are six carbon hexoses or nine carbon sialic acids (Fig. 2). The monosaccharides are linked through the 1-Carbon (C1) position of hexoses and C2 position of sialic acids, also referred to as the anomeric centre. The anomeric center can be linked in either an axial or equatorial configuration forming either an α- or β-linkage. An anomeric centre can be linked by glycosyltransferases to the scaffold protein/lipid or to oxygen atoms of the growing carbohydrate chain. A saccharide that is not anomerically linked to other monosaccharides can convert to an open configuration containing an aldehyde that acts as a reducing agent. Even when this reducing saccharide is attached to a scaffold structure and thus not able to act as a reducing agent, it is still referred to as the reducing end of the carbohydrate. The other extremity of the carbohydrate chain is called the non-reducing end. The linkages between monosaccharides are described from the non-reducing to the reducing end of the glycan. First the anomeric configuration is noted (α/β), followed by the position of the anomeric centre, and the position of the acceptor hydroxyl linked to the saccharide (Galβ1,4GlcNAcβ1-). Some monosaccharides, such as non-terminal galactoses, can be glycosylated multiple times resulting in branched structures which are denoted using brackets (e.g. GlcNAcβ1,3[GlcNAcβ1,6]Galβ1-).

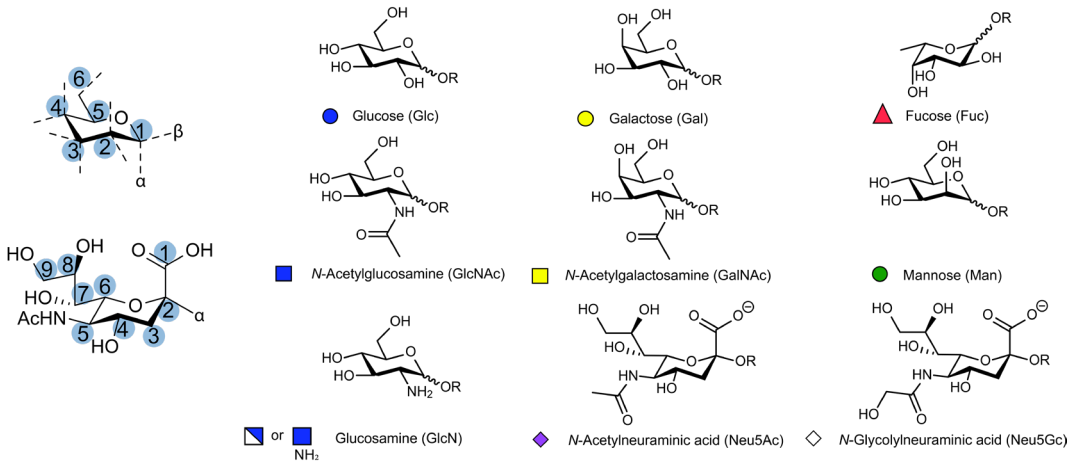


Figure 2: Left) Carbon numbering of hexose (upper) and sialic acid (lower). Right) Chemical structure of common mammalian monosaccharides and their associated graphical representation.

The oligosaccharides, built up by glycosyltransferases, can be positioned on various scaffold such as lipids, proteins, or can exist as unconjugated/free reducing end saccharides. The support structure of the carbohydrate greatly influences the selectivity of glycosyltransferase, resulting in distinct core structures. *O*-linked glycoproteins commonly contain an α-linked GalNAc core although the much rarer *O*-mannoside, *O*-GlcNAc and *O*-fucoside *O*-glycan also exist (Fig. 3). This GalNAc can be extended by β1,3 Gal (C1GalT1) or β1,3 GlcNAc (C3GnT1) to form the core 1 and 3 which can be further branched with a β1,6 GlcNAc by C2GnT2 to core 2 and 4. These four cores can be extended by glycosyltransferases to form complex *O*-glycans. The much rarer core 5, 6, 7 and 8 *O*-glycans are not a substrate for other glycosyltransferases and are only detected as disaccharides.

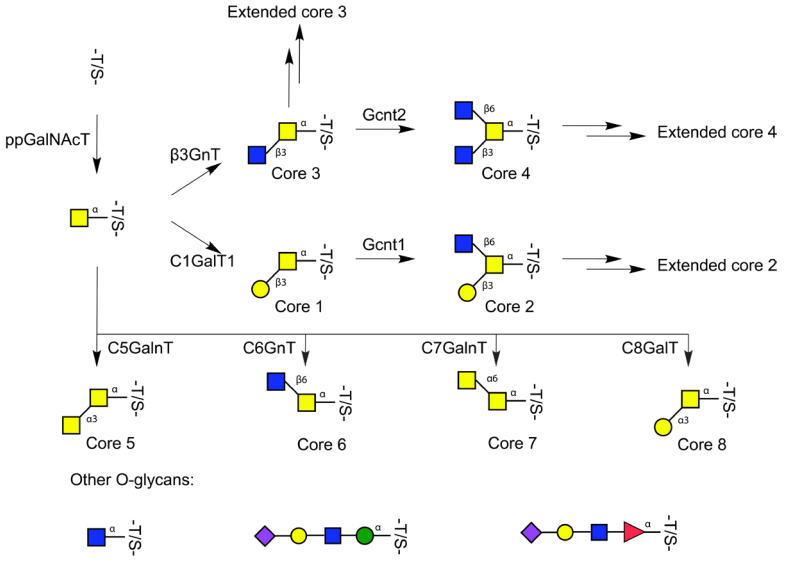


Figure 3: Upper) Biosynthetic pathways of the synthesis of GalNAc *O*-glycan core 1-8 in humans. Lower) Examples of *O*-GlcNAc, *O*-Mannose and *O*-Fucose *O*-glycans.

N-linked glycoproteins are based on a pentasaccharide $Man_3GlcNAc_2$ core which consists of a mannose that is glycosylated with two other mannoses on the 3 and 6 positions, these three mannoses are linked to two β-linked GlcNAcs (chitobiose). This pentasaccharide *N*-glycan core can be α1,6 fucosylated on the reducing end GlcNAc forming a core fucosylated *N*-glycan. *N*-glycans can be classified under one of the following three classes: high mannose stuctures with up to nine mannoses linked to chitobiose; Complex stuctures have one or two GlcNAcs on both the α3 and α6 mannose arms, which can be extended to form complex-type di-, tri- and tetraantennery *N*-glycans; Hybrid *N*-glycans contain a combination of the two previous types (Fig. 4). Glycolipids contain a lactose (Galβ1,4Glc) disaccharide core attached to a lipid. Human milk oligosaccharides share the same lactose core but are not attached to a platform and instead are in solution as a free reducing end sugar.

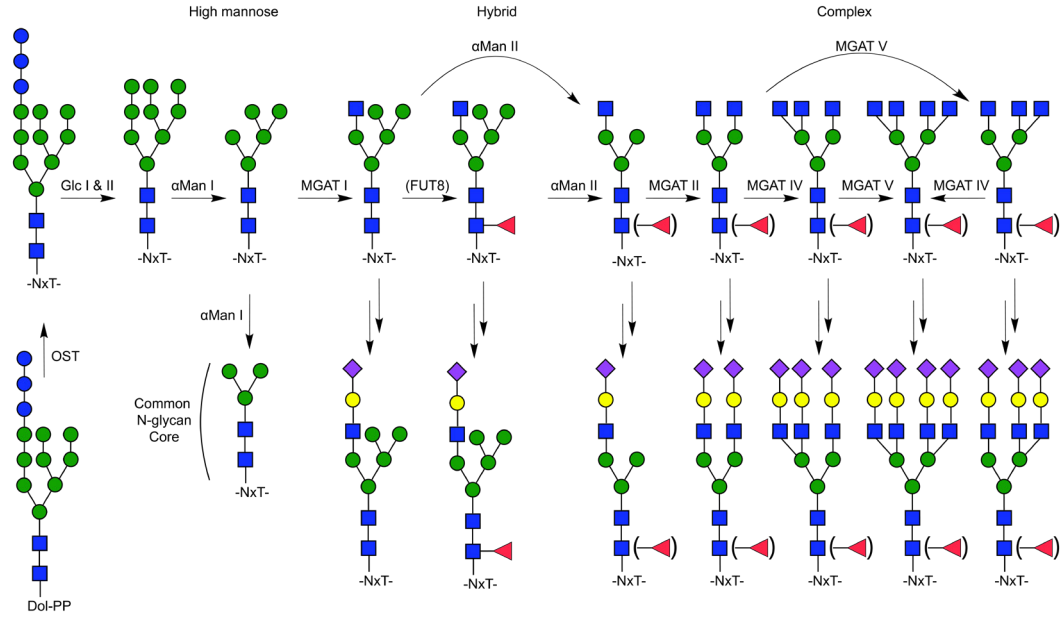


Figure 4: Biosynthetic pathway of *N*-glycan synthesis. A *N*-glycan precursor attached to dolichol is transferred to asparagine on a protein carrying the *N*-glycan glycosylation motif NxT. Glucosidases remove glucose and generate high mannose type *N*-glycan which can be trimmed by manosidases. MGAT I GlcNAc transfer on a trimmed arm of the *N*-glycan core results in the formation of hybrid type *N*-glycan. The GlcNAc can be enzymatically extended to form more complex, hybrid type *N*-glycans. MGAT II can install a second GlcNAc on the other mannose of the *N*-glycan core, resulting in the formation of bi-antennary, complex type *N*-glycans. Additional branching points can be installed by MGAT IV and MGAT V resulting in the formation of tri- and tetra-antennary *N*-glycans. Core fucose can be installed by FUT8 before enzymatic elongation of the GlcNAcs installed on the bi-antennary complex type *N*-glycan or the trimmed hybrid type *N*-glycan.

Glycan cores be enzymatically extended to form more complex glycans. Two types of glycan epitopes can be distinguished on these complex glycans: terminal epitopes that can only be present on the non-reducing end of glycans forming the end of the carbohydrate, and internal epitopes that can be installed on non-terminal positions. The glycan epitopes found on complex type glycans consist of modifications on a LacNAc repeating disaccharide backbone and can be classified under two types based on their configuration. Type 1 structures are based on a Galβ1,3GlcNAc LacNAc disaccharide and are always terminal whereas type

2 structures have a β1,4 linkage which can be further extended to form larger polyLacNAc glycans. The LacNAc backbone can be branched by I-branching enzymes that are able to transfer a GlcNAcβ1,6 to a non-terminal galactose. This GlcNAc can form a new chain that can be enzymatically extended in a similar fashion as the main backbone, including the installation of additional I-branching points. The length and architecture of the LacNAc backbone is controlled by β1,3GlcNAc transferases B3GNT2, 3, 4, 7, 8 and 9; β1,6 I-branching enzymes GCNT2 and GCNT7, and substrate competition of the β1,4Gal transferases B4GalT1-7 with the β1,3Gal transferases B3GalT1,2,4,5 and 6. Apart from LacNAc/I-branches, the only common internal glycosylation is the Lewis^{x} epitope which is generated by the action of FUT3,4,5,6,7,9,10 or 11 on type 2 LacNAc. The large number of competing enzymes involved in the biosynthesis of these internal structural features indicates that the internal glycan structure is important for cellular function. Knockout experiments of some key enzymes involved in the synthesis of this LacNAc backbone have confirmed this, however exact structure-activity relationships have not been established for most involved glycosyltransferases.[19–21]

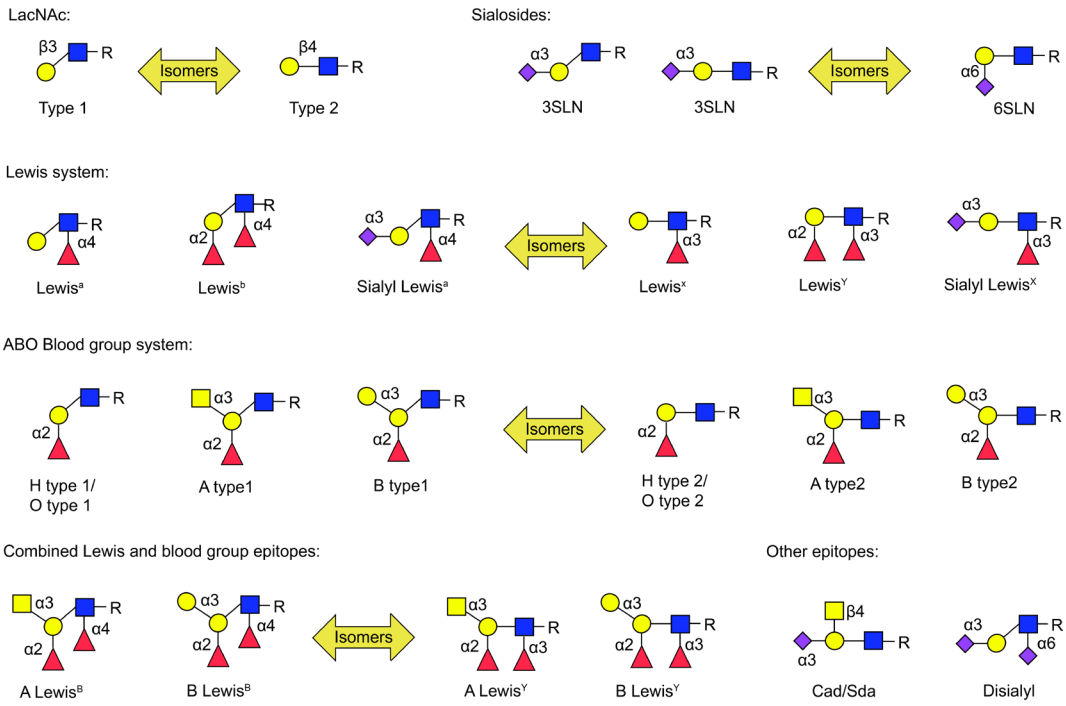


Figure 5: Structure of common terminal glycan epitopes. To the left are structures containing type 1 LacNAc (Galβ1,3GlcNAc). To the right are structures containing type 2 LacNAc (Galβ1,4GlcNAc). From top to bottom are draws: sialosides, Lewis antigens, ABO blood group antigens, combinatory ABO/Lewis antigens and rare terminal glycan epitopes.

The terminal LacNAc of both type 1 and type 2 are often glycosylated with chain-ending glycan features (Fig. 5). Most epitopes exist both in a type 1 and type 2 form resulting in isomeric glycan structures. Neu5Ac can be present on terminal galactose of both LacNAc types in an α2,3 linkage. ST3Gal1 and ST3Gal2 can install α2,3 sialic acid on type 1 LacNAc while ST3Gal3-6 are responsible for type 2 α2,3 sialylation. α2,6 linked sialic acid is generated by ST6Gal1 and ST6Gal2 and can only be installed on type 2 structures resulting

in a total of three possible isomeric mono-sialylated LacNAc structures. The Lewis antigen system consists of the previous discussed Lewis^{x} structure as well as the type 1 isomer Lewis^{a}. These modifications can also be present on α2,3 sialylated structures, resulting in sialyl Lewis^{x} and sialyl Lewis^{a}. FUT1 and FUT2 can transfer an α1,2 fucose to terminal galactose, resulting in synthesis of the H-antigen identical to the O-antigen in the ABO blood group system. Both H/O type 1 and 2 can be fucosylated by the Lewis$^{a/x}$ generating enzymes to form Lewis^{b} and Lewis^{y} difucosides. The ABO blood group system is used to denote the presence of α1,3 GalNAc (A) α1,3 Gal (B), both (AB) or none (O) on the H/O antigen of erythrocytes. Blood group A type 1 and 2 are synthesized by GTA and blood group B type 1 and 2 are made by GTB; both enzymes share a high homology and have only four different amino acids. Lower abundant epitopes include glycans that feature the combination of Lewis$^{a/x}$ and blood group A/B which are A Lewis$^{b/y}$ and B Lewis$^{b/y}$; disialylated type 1 LacNAc which is generated on type 1 sialyl LacNAc by St6GalNAc6 and does not have a type 2 isomer; and the Cad/Sda antigen which is made by the action of B4GalNT2 installing a β1,4 GalNAc on α2,3 sialyl LacNAc type 2.

Another layer of complexity in carbohydrate structure, in addition to the previously described glycan epitopes, is generated by non-carbohydrate modifications of glycans with the most common ones being *O*-sulfation and *O*-acetylation. Sulfation is an important modification in GAGs but can also exist on LacNAcs of complex glycans. This complex glycan sulfation is mostly present on the 6-position of GlcNAc or on both the 6-positions of GlcNAc and Gal. GlcNAc-6-sulfate can also be present on more complex α2,3-sialyl or sialyl Lewis^{x} epitopes. Keratan sulfate is part of the GAG family and consists of ordered domains of these sulfated LacNAcs. Keratan sulfate does not share the xylose core but exists on *N*- and *O*-glycans. Other sulfates can be found on the 3-position of glucuronic acid attached to type 2 LacNAc, forming the HNK-1 epitope.

O-acetylation in humans, and to a lesser degree *O*-lactylation, can be present on the 7, 8 and 9 position of sialic acids.[22] CasD1 is the only a *O*-acetyl transferase described in humans and can only transfer to the 9 position of the sialic acid sugar nucleotide CMP-Neu5Ac, with the others being generated upon acetyl migration. Multiple *O*-acetylation modifications can be present on a single sialic acid by repeated CasD1 action after migration of the initially introduced acetyl ester.[23] The multiple possible positions of these non-carbohydrate modifications, together with the isomeric nature of carbohydrate epitopes and the often added complexity of branched structures, make complete characterization of glycan structure as well as the synthesis of complex carbohydrates highly challenging.

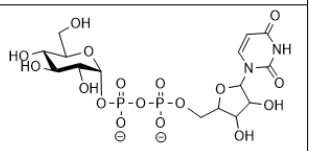
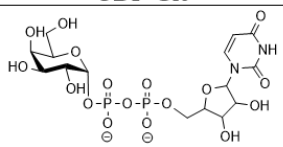
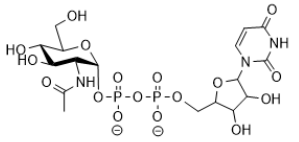
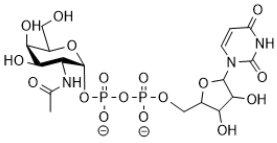
Strategies towards the synthesis of complex carbohydrates

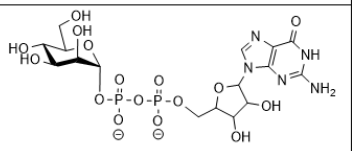
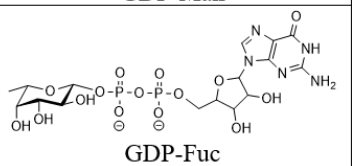
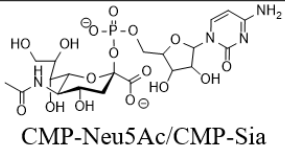
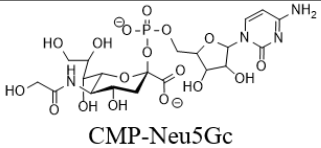
The synthetic preparation of homogeneous complex oligosaccharides provides important reference compounds for the detailed structural characterization of glycans. Moreover, it offers access to biochemical tools such as probes, well-defined substrates for biological research and lead compounds for the development of vaccines and drugs.[24] The high stereochemical complexity of complex carbohydrates has significantly hindered their preparation. The chemical preparation of complex carbohydrates has suffered from elaborate procedures and synthetic planning due to the need for protecting groups on almost all hydroxyl groups to direct both regio- and stereoselectivity. This results in large step counts, low overall yields, and low atom efficiencies. The *in vitro* utilization of carbohydrate active enzymes can greatly reduce regio- and stereoselectivity issues by exploiting the often-strict natural selectivity of the enzymes. Carbohydrate active enzymes can broadly be divided in glycosyltransferases that catalyze the formation of a glycosidic linkage, and glycosidases that can remove saccharides through the hydrolysis of the glycosidic linkages.

The mechanisms of enzymatic glycosylation by glycosyltransferases

Glycosyltransferases can catalyze the formation of a glycosidic linkage and require activated sugar phosphate donors to transfer a saccharide unto a nucleophile. The transferases either utilize sugar mono- or di-phosphonucleotides coined Leloir type transferases, or non-nucleotide sugar phosphates and phospholipids termed non-Leloir type transferases.[25] Most human glycosyltransferases are of the Leloir type and thus utilize sugar nucleotides. These sugar nucleotides are synthesized in mammals by the conversion of exogenous monosaccharides obtained out of diet to sugar monophosphates. After conversion to the sugar monophosphate, they are enzymatically conjugated to a phospho-nucleotide by a pyrophosphorylase. The human metabolism is able to interconvert all sugar nucleotides into each other to adapt to dietary changes and ensure adequate availability of sugar nucleotides.[26] From the nine common human sugar nucleotides, there are seven that are relevant for the synthesis of complex *N*- and *O*-glycans (Table I). These sugar nucleotides have either a uridine-diphosphate (UDP), guanosine-diphosphate (GDP) or cytidine-monophosphate (CMP) leaving group.

Table I: Chemical structure of common sugar nucleotide donors in mammals

Monosaccharide	Nucleotide donor
Glucose (Glc)	 UDP-Glc
Galactose (Gal)	 UDP-Gal
N-Acetyl glucosamine (GlcNAc)	 UDP-GlcNAc
N-Acetyl galactosamine (GalNAc)	 UDP-GalNAc

Monosaccharide	Nucleotide donor
Mannose (Man)	 GDP-Man
Fucose (Fuc)	 GDP-Fuc
N-Acetyl neuraminic acid (Neu5Ac)	 CMP-Neu5Ac/CMP-Sia
N-Glycolyl neuraminic acid (Neu5Gc) non-human	 CMP-Neu5Gc

Leloir-type glycosyltransferase can operate through two main mechanisms: a mechanism that retains the stereochemistry of the sugar nucleotide or a mechanism that results in the opposing stereochemistry (Fig. 6). The inverting mechanism deprotonates the acceptor hydroxyl, activating it as a nucleophile. The di- or mono-phosphonucleotide is then displaced by the nucleophile, which acts through an oxocarbenium ion-like transition state in a S_N2-like fashion. The formed glycosidic linkage by this mechanism contains inverted stereochemistry relative to the sugar nucleotide.

The mechanism for glycosyltransferases that retains stereochemistry has not been fully established, though it has been suggested that it operates through a Koshland-type intermediate. This intermediate is characterized by a covalent glycosyl-enzyme transition state with inverted stereochemistry. This transition state is then attacked by the nucleophile acceptor, which results in the formation of a glycosidic linkage with retained stereochemistry. However, for most retaining glycosyltransferases no putative nucleophile can be detected. An alternative S_NI mechanism without Koshland-type intermediate is suggested to be the more dominant mechanism, which also results in retained stereochemistry.[27]

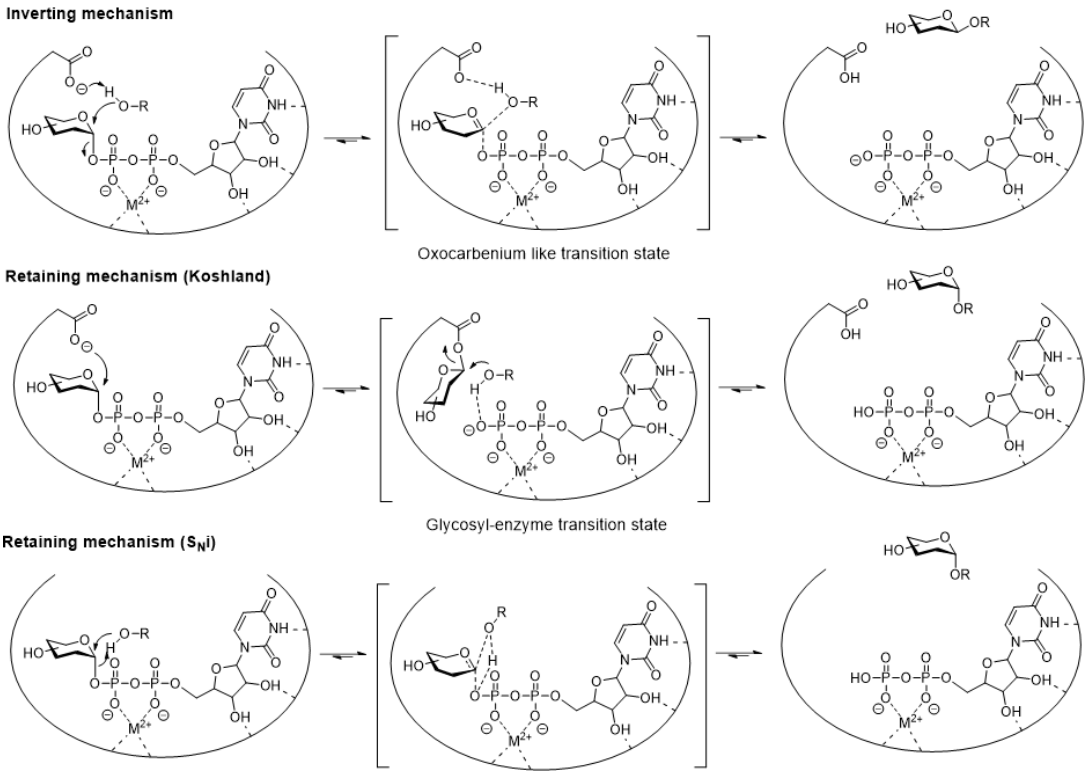


Figure 6: Proposed general mechanisms of cation dependent glycosyltransferase. The sugar nucleotide donor is bound to the active site mediated through a 2+ metal ion. Top) A side chain of the active site carboxylate containing amino acid deprotonated the acceptor nucleophile acting as a base catalyst. The diphosponucleotide is displaced through a S_N2-like transition state and a glycosidic bond is formed with stereochemistry that is inverted from the sugar nucleotide donor. Middle) The Koshland mechanism proposes that a side chain of the active site carboxylate containing amino acid displaces the diphosphonucleotide, which results in the formation of a covalent enzyme-glycan intermediate. The intermediate is then attacked by an incoming nucleophile. This leads to the formation of a glycosidic linkage with retained stereochemistry. Most retaining glycosyltransferase classes lack a potential nucleophilic side chain. Bottom) The suggested S_Ni-type mechanism can generate retained stereochemistry through an oxocarbenium like transition state.

Glycosyltransferases and their application in the synthesis of complex oligosaccharides

Enzymatic synthesis approaches that make use of glycosyltransferases can convert starting material with high step efficiency, making these strategies highly suitable for the synthesis of complex glycans which require multi-step procedures. The application of non-Leloir glycosyltransferases for the synthesis of complex carbohydrates is limited due to the scarcity of reported enzymes. Whilst only limited useful bonds can be generated through these enzymes, the use of non-Leloir enzymes is scalable.[28] Thus, non-Leloir enzymes are applied in the large-scale production of disaccharides such as the kilogram preparation of lacto-N-biose I but have limited use in the synthesis of more complex glycans.

The application of Leloir-type glycosyltransferases was pioneered by the Whitesides group who also introduced in-situ sugar nucleotide regeneration to reduce the need for expensive sugar nucleotide.[30] However, for lower scale enzymatic synthesis it is more convenient to prepare and isolate sugar nucleotides in advance, as this increases control over the concentration of sugar nucleotide in the reaction mixture. In recent years, most natural sugar nucleotides have become commercially available at acceptable prices and advances in the synthesis and isolation of sugar nucleotides have made both natural and derivatized sugar nucleotides more accessible.[31,32]

Traditionally, bacterial glycosyltransferases have been used that can be recombinantly expressed, and steadily more enzymes of this type have been characterized.[33] Some bacteria have even evolved to produce mammalian like glycans on their surface; these glycosyltransferases are especially interesting as they can provide useful tools for the synthesis of mammalian complex glycans.[38] Additionally, bacterial glycosyltransferases often exhibit broad substrate and acceptor selectivity that can be used to make a wide variety of oligosaccharides.[34–37]

Recently, a large investment has been made in the development of expression systems for functional mammalian and plant glycosyltransferases, including all known human glycosyltransferases.[42,43] Mammalian expressed glycosyltransferases often need to be glycosylated themselves in order to be active. The expression of mammalian glycosyltransferases requires more complex systems such as mammalian (HEK293) cells or baculovirus-infected insect cells since recombinant expression systems lack the correct glycosylation pathways. The complexity and high cost of these expression systems results in relatively low expression yields, but the use of human enzymes in the synthesis of complex carbohydrates can provide valuable information on subtle differences in substrate selectivity of the often structurally very similar human glycosyltransferases. For example, humans express eight fucosyltransferases that are able to glycosylate a GlcNAc with an α1,3 fucose. These glycosyltransferases can be expressed in different cells or tissues and can be important factors that influence the expressed local glycome. Characterization of subtle difference between human glycosyltransferases can also help identify enzymes involved in disease related distortions in glycosylation and further our understanding of structural differences in the glycome.[44] Although the general activity of all glycosyltransferases involved in the biosynthesis of complex oligosaccharide is known (Fig. 7), the specific acceptor selectivity is often not yet characterized in great detail. Therefore it can be useful to screen a selection of glycosyltransferases in order to find the most active transferase for a certain transformation.[45]

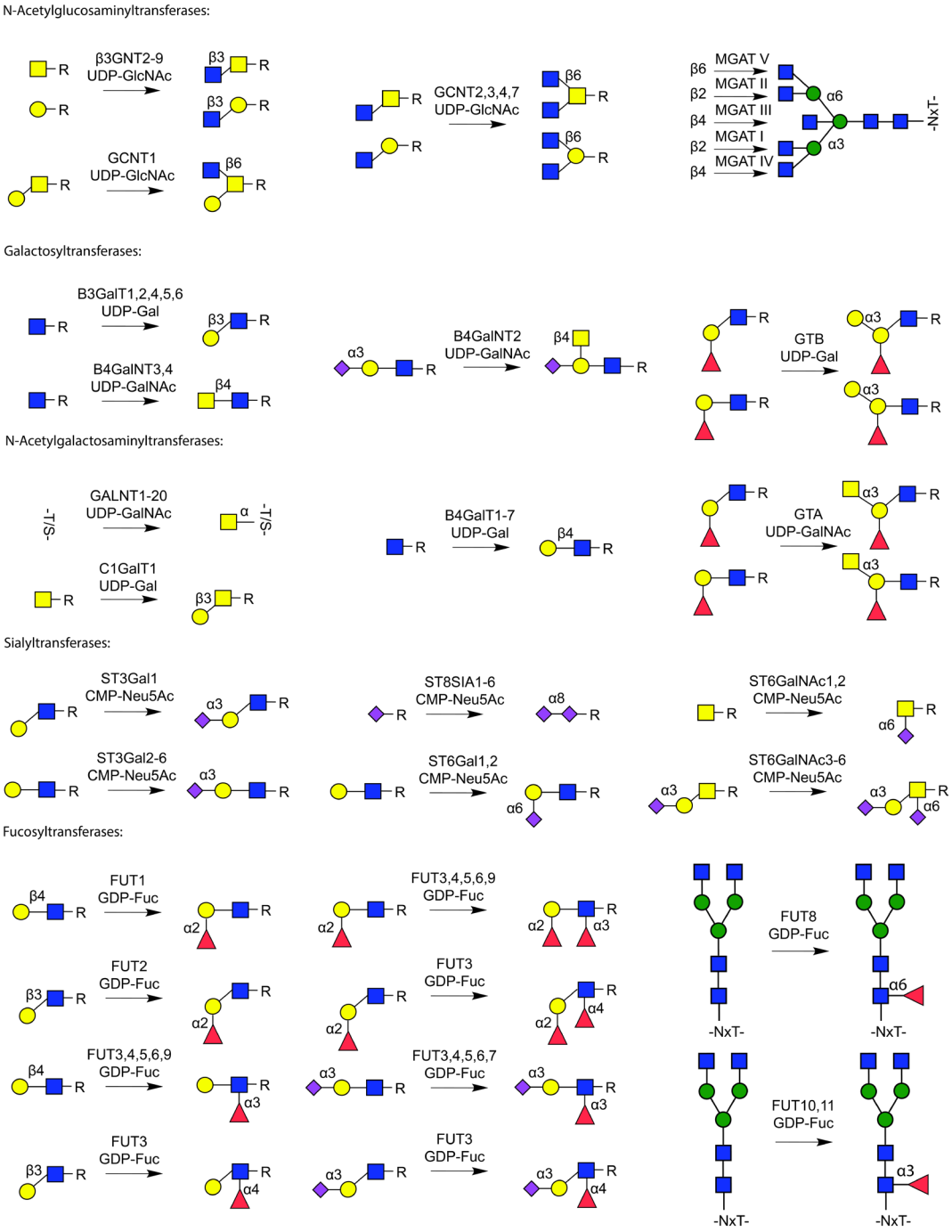


Figure 7: Enzymatic transformations in the biosynthesis of complex oligosaccharides catalyzed by human glycosyltransferases.

Regioselective enzymatic synthesis

While the installation of terminal glycan epitopes on linear oligosaccharides can be readily achieved through enzymatic synthesis, synthetic challenges arise when internal residues are acceptor substrates or when multiple non-reducing end glycans are present due to branching. Desymmeterization of multi-antennary oligosaccharides can be achieved by appropriate enzyme selection and route design which subsequently allows regioselective enzymatic activity. For example, the bottom (α3) arm of biantennary *N*-glycans is more accessible to ST6Gal1, which can be selectively sialylated acid when the donor is limited to <1 equivalent.[46] Now, the top (α6) arm can be selectively extended because the sialic acid prevents enzymatic extension of the bottom arm (Fig. 8A). A similar selectivity can be observed with *E.coli* galactosidase, which can selectively cleave the galactose from the bottom arm.[47] The resulting product can then be sialylated on the top arm because the bottom, more reactive galactose is not present anymore. Restoration of the galactose reactivates the bottom arm which can then selectively be extended (Fig. 8B). Both these strategies together can be applied in the selective enzymatic synthesis of asymmetric biantennary *N*-glycans.[48] Tri- and tetra-antennary *N*-glycans have an larger regioselectivity challenge because of the additional antenna, these prohibit the regioselective enzymatic synthesis of most asymmetric glycans.

Another form of multi-antennary oligosaccharides results from β1,6 I-branching of non-terminal galactoses by I-branching enzymes GCNT1,2,3,4 or 7. While the minimal epitope of the acceptor is GlcNAcβ1,3Gal, the larger Galβ1,4GlcNAcβ1,3Gal or Galβ1,3GlcNAcβ1,3Gal epitopes are also acceptors. The use of these extended acceptors results in the formation of an asymmetric branching point which can be exploited by the installation of a terminal glycan epitope on the bottom (LacNAc) arm (Fig. 8C). The branched arm can be enzymatically extended after blocking the bottom arm acceptor site.[49] Methods for regioselective enzymatic synthesis of asymmetric I-branched oligosaccharides on longer polyLacNAc's, featuring multiple I-branching sites, have not been developed yet.

Similar regioselectivity challenges arise with other internally active glycosyltransferases such as those that install the Lewis^x epitope. On a diLacNAc scaffold, all possible combinations of Lewis^x fucosylation can be acquired by an appropriate choice of glycosyltransferase (Fig. 8D). Installation of a α2,3 sialic acid activates the acceptor to FUT7 which can selectively synthesize sialyl Lewis^x but not Lewis^x. FUT9 on the other hand cannot generate sialyl Lewis^x and is therefore only able to fucosylated the penultimate LacNAc.[50–53] This selective fucosylation is no longer able to generate all possible patterns on larger polyLacNAc's. Tools for the regioselective enzymatic synthesis of glycans exist but are not sufficient to govern selectivity on highly complex scaffolds such as polyLacNAc's. Thus additional (chemical) factors are needed to increase the regioselectivity of glycosyltransferases.

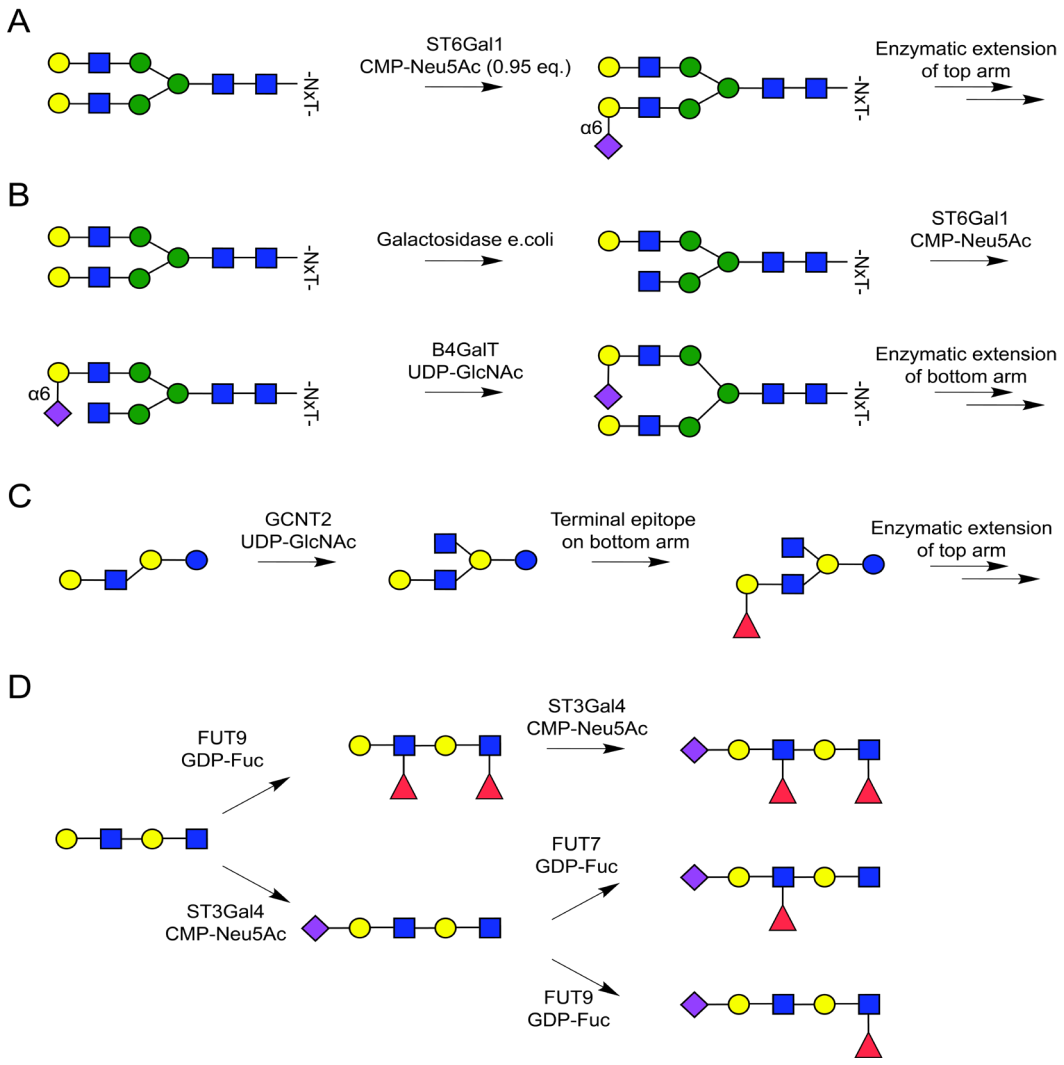


Figure 8: Methods for regioselective enzymatic synthesis. A) Desymmeterization of bi-antennary *N*-glycan by selective α2,6 sialylation of the bottom arm. B) Inverted glycosylation can be achieved in a three step procedure using e.coli galactosidase, which can selectively cleave the bottom arm galactose. ST6Gal1 can then transfer sialic acid to the top arm. In the final step the galactose of the bottom arm is reintroduced. C) Selective desymmeterization of I-branched glycans can be achieved by blocking enzymatic activity towards LacNAc generating enzymes on the bottom arm. D) Selective fucosylation on diLacNAc can be achieved by exploiting innate fucosyltransferase selectivity. α2,3 Sialic acid can direct selective fucosylation by blocking FUT9 activity on the terminal LacNAc. FUT9 will thus only act on the proximal position. FUT7 on the other hand is only able to fucosylate sialylated LacNAc and is therefore selective for the terminal LacNAc.

Chemo-enzymatic synthesis of complex carbohydrates

Enzymatic synthesis is a efficient method for preforming glycosylation reactions, often only requiring a single synthetic step per glycosidic linkage. Regioselectivity of glycosyltransferases, however, can only provide enough control to selectively synthesize a small fraction of all

possible (natural) complex oligosaccharides. In contrast, chemical synthesis provides high regioselective control by employing protecting groups, but every glycosidic linkage significantly increases the synthetic effort required. Chemo-enzymatic synthesis attempts to use the best features of both methods by combining the high-yielding glycosylation of glycosyltransferases with regioselective control introduced by chemical synthesis or manipulation of non-natural glycans.

One strategy that has been applied extensively is a chemical synthesis of the core substrate, which is subsequently enzymatically extended. The chemical synthesis of the core oligosaccharide is especially attractive when the core saccharide is difficult to generate through enzymatic synthesis or when asymmetry needs to be introduced on multi-antennary oligosaccharides. Challenging cores that have been made by this approach include glycolipids and bacterial LOS that mimics human glycans.[54,55] The lipidic portions of these molecules induces micelle formation which interferes with efficient enzymatic synthesis of the core structure. In addition, the core of bacterial LOS also contains non-human monosaccharides. The core structures of *O*-glycans are particularly difficult to access using purely enzymatic means due to the low enzymatic activity when no peptidic portion is present. Multiple groups have recently contributed to the chemo-enzymatic synthesis of *O*-glycans by applying a chemical core synthesis strategy, but methods to access asymmetrically extended bi-antennary core 4 *O*-glycans are notably lacking.[56–59] Desymmetrized, chemically prepared core structures have also successfully found use in the synthesis of other multi-antennary glycan classes such as branched human milk oligosaccharides and *O*-mannosyl glycans.[60,61] Boons et al. managed to synthesize a orthogonally protected tetra-antennary *N*-glycan core where each of the four arms could be orthogonally deprotected and functionalized (Fig. 9A).[62] Despite the heavy synthetic effort required for the synthesis of the *N*-glycan core, this approach has provided access to incredibly complex oligosaccharides.[63–68]

An alternative chemo-enzymatic approach does not rely on the initial chemical synthesis of the core structure but uses an enzymatic introduction of chemical functionalities. These functionalities can then be manipulated to assert regioselectivity in the context of a complex, unprotected oligosaccharide. These chemical functionalities can be introduced in a complex oligosaccharide by enzymatic reactions. An example is the introduction of an aldehyde on the 6-position of a terminal galactose with galactose oxidase. The generated aldehyde can be used as a substrate for reductive amination reactions with anomerically aminated oligosaccharides for a facile synthesis of I-branch like polyLacNAc glycomimetics.[69] Cao et al. extended this chemo-enzymatic methodology by using 6-aldehyde galactose as a blocking group for internal α2,6 sialylation of galactoses (Fig. 9B).[70] The α2,6 sialic acids were exploited in later work to control Lewis^x and Lewis^a fucosylation on polyLacNAc scaffolds.[71]

Instead of relying on specific enzymatic activities of a carbohydrate modifying enzyme it is also possible to exploit donor promiscuity of glycosyltransferases. For example, UDP-*N*-trifluoroacetyl glucosamine (GlcNTFA) can be accepted by the heparan sulfate polymerase enzyme and integrated in heparan sulfate polymers. The GlcNTFA can then be used to control *N*-sulfation as the protected glucosamine is not recognized as a substrate.[72] UDP-GlcNTFA can also be selectively introduced in arms of multi-antennary *N*-glycans using MGAT enzymes. Deprotection of the trifluoroacetamide and subsequent introduction of other

orthogonal protecting groups stops enzymatic activity on multiple arms while allowing for selective re-activation (Fig. 9C).[73] Enzymatic introduction of GlcNTFA enables the use of natural extracted bi-antennary *N*-glycan as a starting material, thereby evading the large synthetic effort previously required for the synthesis of the *N*-glycan core. Apart from controlling arm selectivity, *N*-protected glucosamine can activate/deactivate Lewis^x fucosylation on polyLacNAc scaffolds, although enzymatic methods to introduce GlcNTFA in a polyLacNAc chain have not yet been reported.[74] In recent years, the development of new chemo-enzymatic strategies has increased the synthetic accessibility of high complexity and asymmetric glycans while greatly reducing step count. Further developments in chemo-enzymatic synthesis might allow the synthesis of all naturally occurring glycans in a selective manner by using glycosyltransferases.

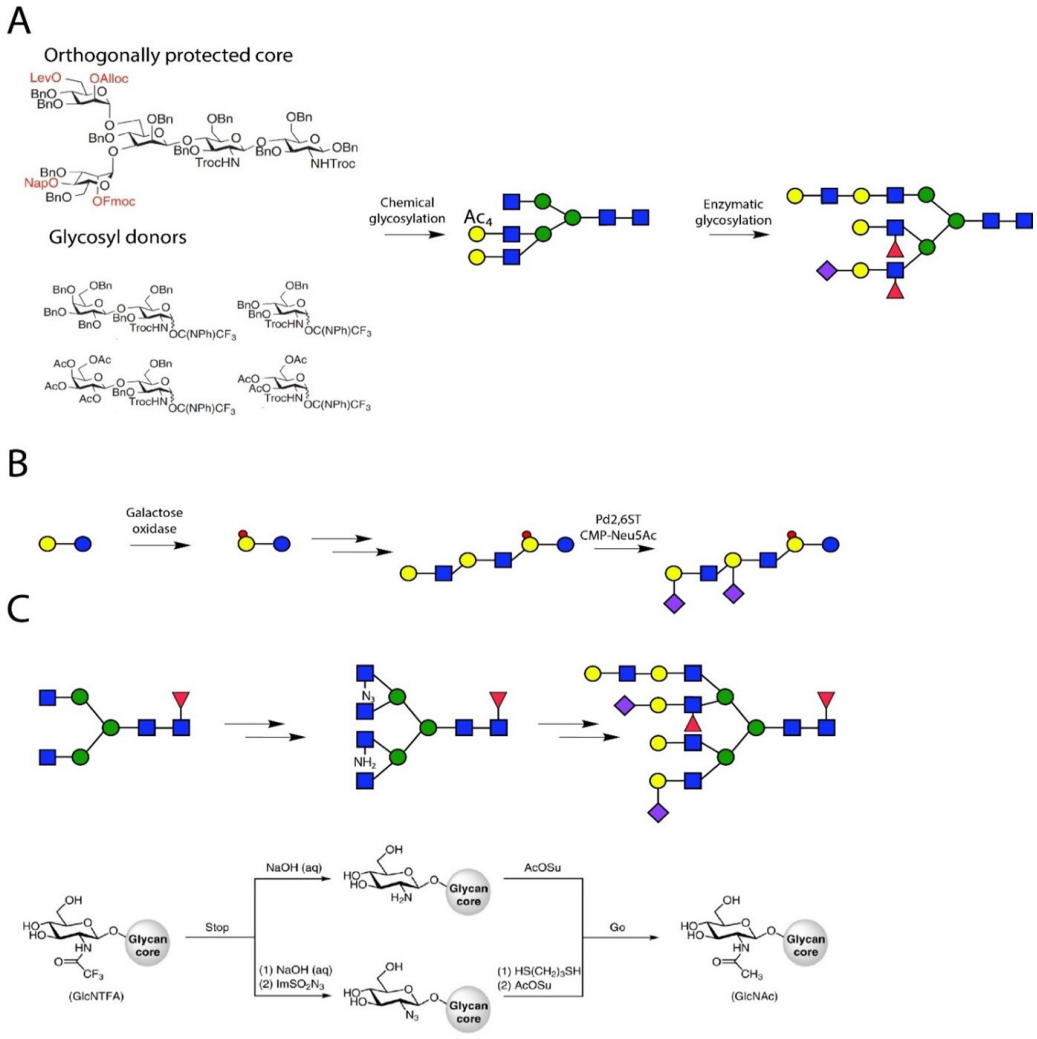


Figure 9: Approaches for the chemo-enzymatic synthesis of complex glycans. A) Chemical synthesis of *N*-glycan core followed by enzymatic diversification. B) Enzymatic oxidation introduces unnatural 6-aldehyde (red dot) galactose, which can be used to direct glycosyltransferase activity. C) Enzymatic incorporation of unnatural sugar nucleotide UDP-GlcNTFA which can be chemically deactivated towards enzymatic activity by forming glucosamine. Glucosamine can then be protected by amino protecting groups to generate orthogonality between installed unnatural glucosamines.

Methods for the structural characterization of complex carbohydrates

The choice of methodology for the characterization of complex carbohydrates is highly dependent on the source of the material, purity of the sample, and level of information of interest. Full identification of all monosaccharides, linkages, positions, and location of non-charbohydrate modifications is increadibly challenging even for homogeneous samples. Furthermore, labile groups on the glycan can be lost during preparation of complex carbohydrate samples, resulting in misassignment of the glycome. In order to improve our understanding of the glycome, methodology is required to completely resolve glycan structure in complex samples without losing information on labile groups. The development of technologies that can adress these issues is currently a very active area of investigation.[71]

Nuclear magnetic resonance (NMR)

A high degree of information on the glycan structure can be obtained through NMR (Fig. 10). NMR tyically requires millgram quantities of high purity carbohydrates. ^{1}H-NMR can indicate how many glycan residues are present through analysis of the anomeric (H-1) resonance peaks that appear in a well-resolved region of the spectrum. A characteristic larger doublet splitting for β-anomers over α-anomers in this region facilitates facile linkage assignment.

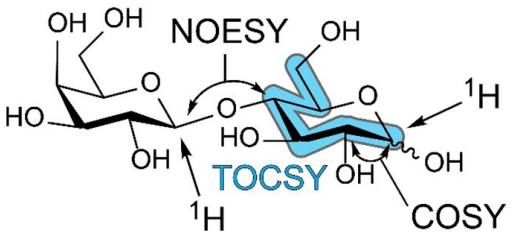


Figure 10: Chemical structure of lactose denoted with NMR experiments used to determine correct stereo- and regiochemistry.

Further characterization of the glycan structure requires more advanced two-dimensional NMR techniques. Correlated spectroscopy (COSY) and Total correlation spectroscopy (TOCSY) are useful techniques to determine the signals of neighbouring protons, which help identify the H2-H6/H9 protons depending on the spin system. Nuclear Overhauser effect spectroscopy (NOESY) can be used to identify spins that undergo cross-relaxtion. ^{1}H dipole -dipole spin coupling generates this cross-relaxation in a spatial manner, allowing the identification of protons that are near each other in space. Combining NOESY information with that of TOCSY/COSY allows the identification of cross-relaxation of protons that are not part of the monosaccharide itself, instead belonging to the attached carbohydrate. The combination of these 2D techniques can be used to identify linkage positions of the glycan. Heteronuclear sigle quantum coherence (HSQC) can be used to determine the ^{13}C spin by ^{1}H-^{13}C magnetization transfer after the ^{1}H shift has been assigned. A significant drawback of NMR characterization of carbohydrates is that signals of similar monosaccharides are often partially overlapping, especially in larger glycan structures. Completely resolving all signals is a labourious process and requires a high degree of expertise. The usability of NMR on mixtures of glycans is very low because peak overlap makes it near inpossible to determine which signal belongs to which glycan.

Histochemistry

Lectin histochemistry is a microscopy based technique used for the evaluation of tissues or micro-array glycans. Carbohydrate binding proteins (lectins) are fluorescently tagged and used to stain 2D surfaces mediated by the carbohydrate binding selectivity of the lectin. A large collection of lectins have been described with widely varying glycan binding preferences, but detailed characterization of the lectin binding preferences is still lacking. Lectin histochemistry has widespread use in the descriptive assesment of glycosylation on a tissue level, altough limited information is obtained on the actual glycan structure. Sequential exoglycosidase treatment together with assesment of the treated product with lectins or other techniques such as mass spectrometry has also been used to gain additional structural information based on the activity profile of the glycosidases.[72–74]

Analysis of glycosylation on glycoproteins

The analysis of glycoproteins or glycopeptide has traditionally relied on the separation of the glycan from the peptide or protein, which incurs a loss of information on the site of glycosylation. Glycoproteomics aims to combine information on both protein and glycan identity.[75] Top-down glycoproteomic approaches attempt to gather detailed information on the complete amino-acid sequence, including post-translational modifications such as glycosylation. The downside of this approach is that is involves extremely complex data sets. [76] More common bottom-up approaches rely on digestion of the glycoprotein to smaller glycopeptides, which loses potential meta-heterogeneity information.[77] Bottom-up approaches are developed to efficiently characterize the site of-glycosylation and both the occupation density and composition of the glycan site.[78] In order to determine the complete glycan structure of naturally occurring glycopeptides requires more laborious techniques such as MS/MS fragmentation which are not very efficient due to the reduced signal intensity because of peptide heterogeneity. Moreover, sufficient information needs to be generated by both glycan and peptide fragmentation to fully resolve the structure, a data and sample intensive process of which the complexity rapidly increases in conjunction with sample complexity.[79,80] Because of these still standing challenges, it is currently more convenient to separate the glycan from the protein/peptide when interested in the glycan structure.

Chemical as well as enzymatic methods have been described to release glycans from glycoproteins and glycopeptides. The gold standard to release *N*-glycans is by the action of the broadly acting *N*-glycosidase PNGaseF (Fig. 11A). PNGaseF can cleave between the innermost GlcNAc and asparagine resulting in a free reducing end glycan. The minimal epitope required by PNGase F is a chitobiose attached to a asparagine. Provided the core is not α1-3 fucosylated, PNGaseF can cleave all complex, hybrid, and high mannose oligosaccharides. [81] PNGase F activity is not always able to cleave all native *N*-glycans, therefore sometimes glycoproteins denaturization is required to facilitate complete cleavage by PNGaseF.

Endoglycosidases (Endo) are a class of *N*-glycan cleaving enzymes that are less dependent on protein structure and suitable for deglycosylation of native proteins.[82] Endo enzymes can cleave between the GlcNAcs of the chitobiose which does result in a loss of information on core fucosylation. Since Endo enzymes have more stringent substrate requirements a cocktail

of Endo enzymes can be used to broadly cover the *N*-glycome.[83] The use of a single Endo enzyme can be beneficial to study a subset of the *N*-glycome because subsets of *N*-glycans can be released, such as the release of biantennary complex *N*-glycans with EndoF2. Chemical release methods for *N*-glycans have been described that are similar to those described below for *O*-glycans. However, these methods suffer from low recovery and glycan stability compared to enzymatic release of *N*-glycans due to the harsh conditions required.[84]

N-glycans can be removed by broadly acting *N*-glycosidases, but no similar efficient enzymes exist for *O*-glycans (Fig. 11B). The analysis of mucin type *O*-glycosylation is severely hindered by the harsh conditions required to release the glycan structure from the peptide backbone. Chemical release methods are complicated by the presence of labile sulfates and *O*-acetylation modifications on *O*-glycans, which are unstable to mild acid and mild base respectively. Recent advances in enzymatic and chemical *O*-glycan release have only had limited success. The development of an efficient, clean, and non-destructive *O*-glycan release method is highly desirable.

Enzymatic cleavage of *O*-glycan is hindered by the limited substrate scope of natural occurring Endo-α-N-acetylgalactosaminidases (O-glycosidases). *O*-glycosidases can only cleave core 1 and core 3 disaccharides as well as T antigen monosaccharides from the peptide backbone but have no activity on more complex *O*-glycans.[85] Enzymatic digestion of complex *O*-glycans in nature is often preceded by exoglycosidases until the substrate is trimmed back enough to be an active substrate for *O*-glycosidases. The stringent substrate scope of *O*-glycosidases is unable to capture structural diversity in the *O*-glycome. Recently, it was attempted to increase the substrate scope of *O*-glycosidases by the Withers group. Through functional metagenomic screening, a slightly more promiscuous *O*-glycosidase was discovered which can cleave a sialylated trisaccharide.[86] Although current success in developing broad *O*-glycosidases is limited, such an enzyme without substrate bias would revolutionize enzymatic *O*-glycan release and *O*-glycomics in general.

Chemical *O*-glycan release mostly relies on the acidic α-proton next to the glycosidic linkage (Fig. 11C). Under alkaline conditions a β-elimination can occur, forming a dehydroalanine (from serine) or 3-methyldehydroalanine (from threonine) and a free reducing end carbohydrate. Different basic release agents have been described, e.g. hydrazine, LiOH, NaOH, ammonia and hydroxylamine.[87,88] Significant drawbacks of these base mediated methods are a loss or migration of *O*-acetylation and a degradation reaction of the formed free reducing end sugar, called the peeling reaction. The peeling reaction is favored on the common β1-3 linked glycan. It is initiated by base promoted deprotonation of the C-2 axial proton, followed by elimination of the linked glycan to form an alkene and a new free reducing end glycan. The newly formed free reducing end can undergo additional peeling, thus leading to further misassignment of the *O*-glycan structure. Chemical additives have been explored which exploit the fact that the peeling reaction requires the reducing end glycan to be in its acyclic configuration. For example, the addition of a mild reductant in the reaction mixture results in the formation of the alditol glycan, which is chemically resistant to peeling.[89,90] Reductive amination using an aromatic amine in combination with a mild reductant, such as sodium cyanoborohydride, can achieve a similar result while incorporating useful chemical entities such as purification

or ionization labels in a single step or two step procedure.[91,92] A role for calcium in the peeling reaction is suggested by reports of significant peeling reduction by the addition of 100 mM EDTA or 0.1% TFA.[93,94] Other non-reductive approaches are the addition of DBU to hydroxylamine release, one-pot PMP labelling, or the incorporation of ammonium salts that form the more stable glycosylamine derivative.[88,95–97]

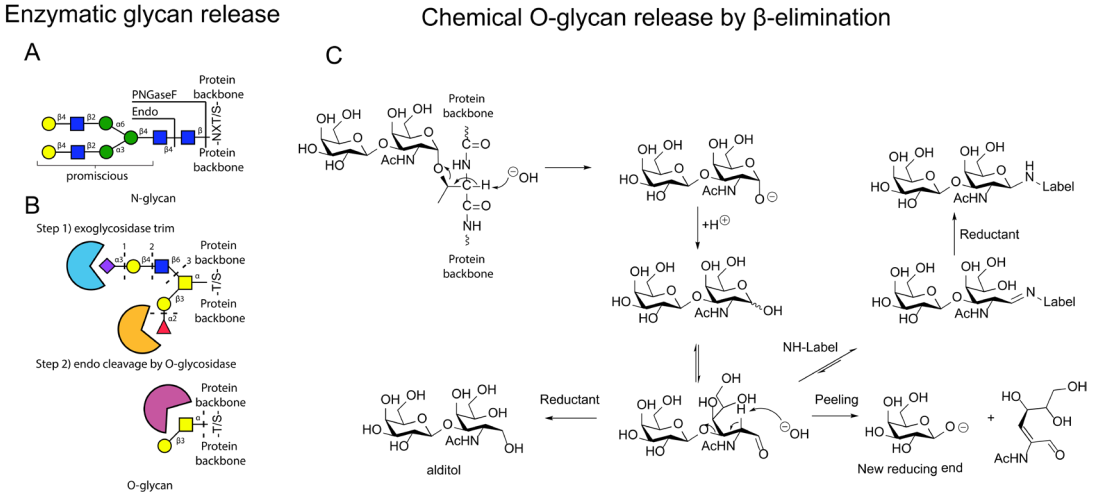


Figure 11: Common glycan release methods. A) *N*-glycans can be released by promiscuous endo glycosidases. Endo enzymes can cleave between the two GlcNAc of the chitibiose core and have divers *N*-glycan selectivities. PNGase F is often used due to its high substrate tolerance. It can cleave the complete *N*-glycan but sometimes requires denaturation of the protein. B) Enzymatic *O*-glycan digestion occurs through an initial trim of the *O*-glycan by exoglycosidases followed by endo acting glycosidases that cleave the *O*-glycan from the protein. The endo acting glycosidases have a highly limited substrate scope. C) Mechanism of base mediated β-elimination of *O*-glycans. The resulting free anomeric end is prone to base catalyzed peeling side reactions leading to potential misassignment of the *O*-glycome. Peeling can be prevented by in-situ reduction towards the alditol derivative or it can be minimized using amino based reagents.

Alkaline hypochlorite release of *O*-glycans is a promising new chemical release method for the analysis of *O*-glycans.[98–100] At present, the mechanism of action of hypochlorite release is unknown. The reaction produces intact *O*-glycan on a glycolic or lactic acid aglycone for serine and threonine, respectively. This oxidative release gives unique information of the parent amino acid (serine/threonine) that is lost when using β-elimination. Another major benefit of this reaction is the direct installation of a charged carboxylate that can be exploited as a negative mode ionization tag in mass spectroscopy, resulting in higher sensitivity. Furthermore, the anomeric aglycone locks the glycan in a closed configuration and thereby prevents peeling. While oxidative release seems promising, the currently developed method requires a large amount of material and has a low recovery: only around 1 µmol glycan was obtained out of 25g porcine stomach mucin.[84] Moreover, oxidative release is conducted at high pH which causes migration of *O*-acetyl groups and can allow partial β-elimination. A mixture of multiple products is obtained with a significant fraction of *O*-glycan released as a free reducing end carbohydrate, but this still leads to the formation of unwanted peeling products. Optimization of the reaction conditions is hindered by the lack of *O*-glycan standards, making it difficult to differentiate side-products/intermediates from impurities in the natural samples that were studied.[98–100] Further developments in the mild and efficient release of *O*-glycans are necessary to bring our understanding of the *O*-glycome to a similar level of that of the *N*-glycome.

Glycan detection by mass spectrometry

Currently, the go-to method for the analysis of complex glycans are either by liquid chromatography coupled with electrospray ionization mass spectrometry (LC-ESI-MS) or by matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) due to their high sensitivity, high throughput and widespread applicability.[101] MS with either ionization techniques can provide the mass-to-charge ratios (m/z) of compounds to obtain a comprehensive glycan composition but often cannot distinguish isomeric glycans.[102] The presence of multiple isomeric glycans can be detected by MS after LC separation using HILIC or PGC, which are often able to separate glycan isomers, and are analyzed as multiple peaks with identical molecular weight. Identification of the exact glycan structure by MS requires an additional dimension which can be achieved with a variety of techniques.

Multistage mass spectrometry

Multiple mass analyzers can be coupled together which can provide additional structural information on glycan structure. Usually, ions of a particular m/z are separated and subsequently split into smaller fragments (Fig. 12). Several ion activation techniques are applied to gain additional structural information on complex glycans based on vibrational heating/excitation such as collision-induced dissosciation (CID) and infrared multiple photon dissociation (IRMPD). Other methods based on activated electron dissociation (ExD) such as electron-capture dissociation (ECD) and electron-transfer dissociation (ETD) are also used.[103] CID or IRMPD generally fragment glycosidic bonds (B, C, Y, Z type ions), while ExD methods especially yield fragmentation across the glycan ring (A and X type ions).[104] Complementary sequence information on both glycan and peptide can be collected by ECD/IRMPD analysis, which shows promise for the analysis of glycopeptides and branched or highly substituted glycans.[105] While multistage MS can provide useful information, especially in determining glycosylation of branched glycans and the position of internal modifications, these techniques are much less suited to assign linkage types.[106] Derivatization methods have been developed that selectively react with only one linkage type of sialic acids e.g. linkage specific esterification, resulting in a mass diverence that can be picked up by MS.[107,108] Another level of ambiguity in multistage MS analysis of complex glycans stems from the ability of fucose residues, and to a less extent hexoses, to rearange.[109–111] Furthermore, differentation of type 1 and 2 isomeric structures and the positional assesment of labile glycan modifications remain highly challenging with multistage MS. There thus exists a high demand for the development of techniques to distinguish isomeric glycan fragments in line with (multistage) MS.

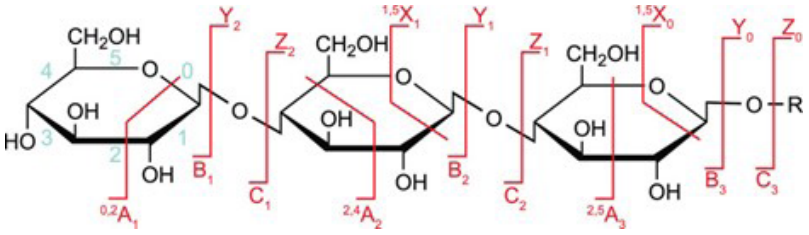


Figure 12: Domon–Costello nomenclature of carbohydrate fragmentation. A- and X-fragments result from cross-ring cleavages, whereas B- and C-fragments and their Y and Z counterparts originate from glycosidic cleavage between two monosaccharide units. The numbering of bonds in the sugar ring, which is indicated as superscript numerals in cross-ring fragments, is exemplarily shown for one monosaccharide.[109]

Ion mobility spectrometry

Despite the capabilities of multistage MS, the full structural elucidation of glycans with MS is still exceedingly difficult. The complex fragmentation spectra obtained are often difficult to interpret on homogeneous glycans and this difficulty increases dramatically when dealing with natural samples. Moreover, the presence of numerous isomers, which often have identical MS/MS spectra, makes exact structures identification using these techniques challenging. Ion mobility spectrometry (IMS) is a fast developing separation technique that can be combined with MS and has shown promise in the analysis of glycans.[112] In IMS, ions are separated based on their mobility, which is determined by the time it takes for ions to travel through a buffer gas under influence of an electric field. This arrival time depends on the charge and size, but also on the shape of an ion due to the interaction with the buffer gas.

Multiple types of IMS mobility cells have been developed, such as drift tube (DTIMS), trapped (TIMS) and travelling wave (TWIMS) (Fig.13A-D). DTIMS passes ions through a static, inert buffer gas mediated by the action of a homogeneous electric field. This process separates ions based on ion mobility due to their interaction with the buffer gas. TIMS uses the same physical principle as DTIMS but inverts the experimental setup where the electric field is used to hold the ions stationary against a moving gas. TWIMS uses periodically spaced ring electrodes that move ions through a static buffer gas by variable potential forming travelling electrostatic waves. Radio frequency potential is applied to confine ions to the wave. Ions with a lower interaction with the buffer gas can "surf" the wave. An improved form of TWIMS, cycling ion mobility (cIM) has been developed by placing a looped TWIMS cell orthogonal to the main axis. The cyclic geometry enables the ions to pass multiple times through the TWIMS cell, which provides additional resolution.

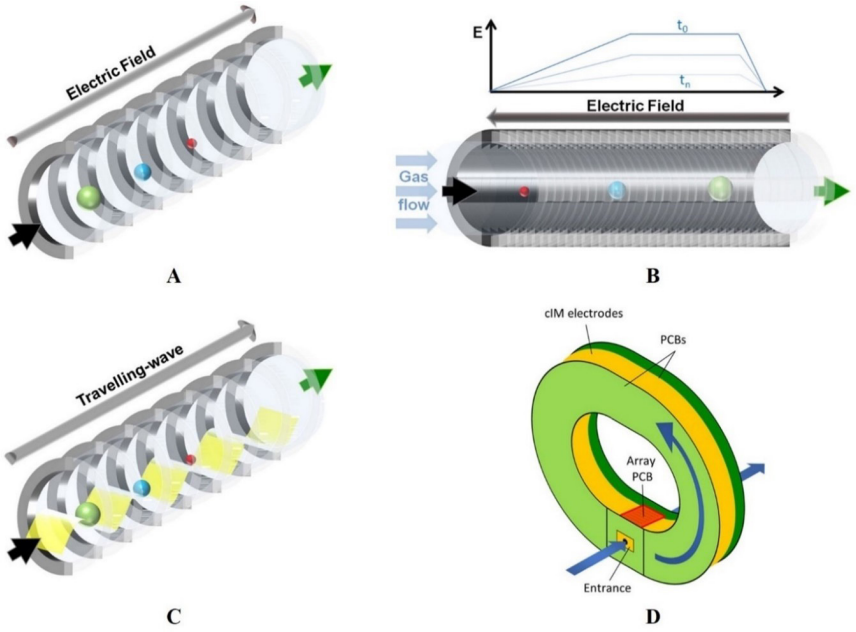


Figure 13: Mobility cells for drift tube (A), trapped (B), travelling wave (C) and cyclic (D) ion mobility analyzers. Ions with increasing CCS are depicted as red, blue and green spheres. Black arrows (A-C) show where the ions enter the cell and green arrows show where they exit the cell. Blue arrows (D) show the direction of ions towards the MS. Figure copied with permission.[113]

The ability of IMS to separate based on the shape of the ion makes it suitable to analyze isomeric glycans. Glycans can adopt unique conformations in the gas phase, depending on branching and linkage-type. These conformers can be resolved in their IMS arrival time distribution (ATD) and the unique ATD fingerprints can be used for facile and confident identification of isomeric glycans. Furthermore, IMS can be combined with ion fragmentation techniques to allow the separation of isomeric fragment ions before MS.[112,114,115] IMS has been shown to separate challenging isomeric glycans: branching architecture of isomeric N-glycans of glycopeptides could be identified.[116] Multiple IMS studies were able to discriminate α2-3 and α2-6 linked sialic acids, even on complex *N*-glycans (Fig 14.).[117–119] Furthermore, type 1 and 2 isomers of human milk oligosaccharides could be separated by ion mobility.[120] Fucosylation positions were also determined on fragmentation products of simple human milk oligosaccharides, which were subsequently applied to assign fucosylation on a higher complexity glycan.[121] Where migration of fucosylation can easily lead to misassignment on multistage MS approaches, it was demonstrated that the IMS distribution of migrated fucosides does not resemble known natural fucosides.[122] Previous mentioned methods rely on pure reference glycans to distinguish isomeric ionization fragments but ATDs are instrument-dependent measurements which requires calibration with suitable standards.

Collisional cross-section (CCS) is used as a reproducible, instrument-independent variable describing the surface area of the ion. The determination of CCS values for complete glycans or key ionization fragments allows the differentiation of isomers without the need for further standards. The elimination of specialized glycan standards is essential for high-throughput analysis using IMS since calibration with reference compounds for each instrument and each glycan epitope is not feasible.[112] In order to determine CCS values it is required to determine the reduced mobility (K_0) of an ion through a gas which can be calculated using:

$$K_0 = \frac{L}{t_d E} \cdot \frac{p}{760} \cdot \frac{273.2}{T}$$

Formula 1: L is the length of the drift tube, t_d is drift time, E is the electric field strength, p is the pressure of the buffer gas and T is the experimental temperature.

CCS can be subsequently calculated from the reduced mobility by the Mason-Schamp equation:[123]

$$\Omega(\text{Å}^2) = \frac{3}{16} \cdot \left(\frac{2\pi}{\mu k T}\right)^{\frac{1}{2}} \cdot \frac{ze}{K_0 N_0}$$

Formula 2: Ω is the integrated CCS, μ is the reduced mass of analyte ion and drift gas, k is the Boltzmann constant, T is temperature, z is the charge of the ion, e is the elementary charge and N_0 is the Loschmidt constant.

While it is possible to determine CCS values with most IMS analyzers, DTIMS has contributed to the vast majority of CCS value determinations. DTIMS is able to establish high precision CCS measurements and has a direct relationship between drift time and cross section which is more complex for other types of analyzers.[124] When the reduced field strength is experimentally considered low (<15 Td), an expanded form of the Mason-Schamp equation can be derived, using experimentally measured p, E, T and t_d for DTIMS in a specific buffer gas:

$$^{DT}CCS_{gas}(\text{Å}^2) = \frac{1}{K_0} \cdot \frac{(18\pi)^{\frac{1}{2}}}{16} \cdot \frac{ze}{(kT)^{\frac{1}{2}}} \cdot \left[\frac{1}{m_{ion}} + \frac{1}{m_{gas}}\right]^{\frac{1}{2}} \cdot \frac{1}{N}$$

Formula 3: m_{ion} is the mass of the ion, m_{gas} is the mass of the buffer gas and N is the density of buffer gas.

To increase ion mobility resolution in DTIMS it is possible to increase the tube length or gas pressure which has practical limitations. Recently, a post-processing technique has been developed that improves the resolving power of their DTIMS systems.[125] This technique improves sensitivity by employing multiple ion pulses per IMS cycle, leading to an increased number of data points per cycle in multiplexed measurements. The pulses are demultiplexed into a single pulse data file, which can be processed with high resolution demultiplexing software (HRdm). HRdm uses the multiplexed data to further improve peak shape and resolution of the single pulse data file. This improved resolution allows separation of previously unresolvable isomers and provides CCS values for these isomers with high certainty. Increased resolving power allows for the separation of smaller isomeric structures, which are more difficult to resolve by IMS than larger structures.[126,127] The ability to distinguishing these smaller glycans is needed for the analysis of glycan fragments in IMS and can open the door to fragmentation-based glycan sequencing approaches of complex glycans/glycan mixtures.[128]

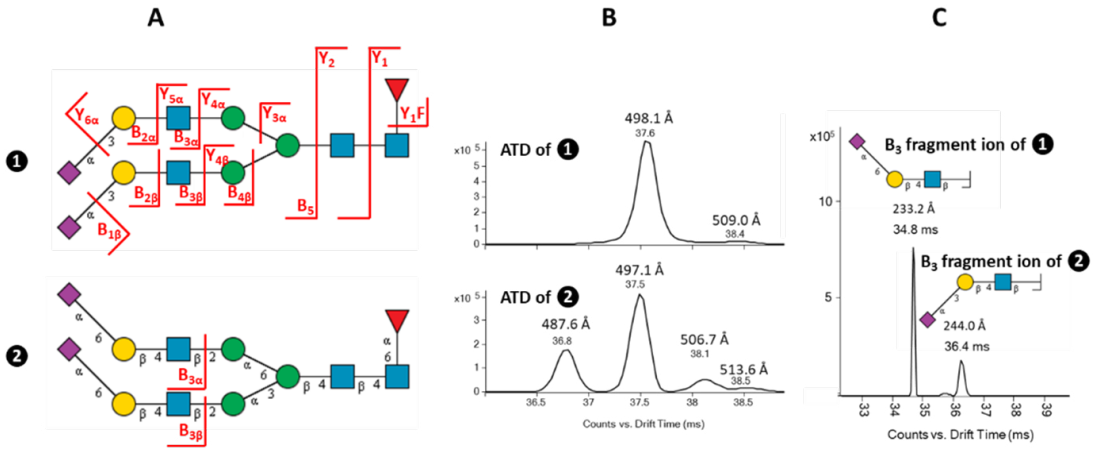


Figure 14: A) Structures of isomeric glycans **1** and **2**, differing only in sialic acid linkage, and fragmentation (in red). B) Top-down differentiation of intact glycans 1 and 2 by HR-ATDs of $[M-2H]^{2-}$ ions. C) Bottom-up differentiation of glyco-epitopes in **1** and **2** by the HR-ATDs and CCS values of B_3 fragments as $[M+H]^+$ ions. Figure copied with permission.

Challenges that isomeric glycans bring in glycoscience

It is clear that researchers have only scratched the surface of understanding the detailed biological function of the glycome. The complexity of the glycome and lack of suitable analytical techniques to study this complexity have long held back developments in glycoscience. In recent years, advancements in instrumentation and synthetic methodology have revitalized efforts to disect the glycome. A key feature of the recent developments in glycoscience stems from the interplay of synthesis, analysis and biology. Developments in efficient synthetic strategies such as enzymatic and chemo-enzymatic synthesis have enabled the synthesis of well-defined glycan standards. The increased availability of glycan standards that match the complexity of naturally occurring glycans has begun to facilitate the development of analytical methodology that can distinguish isomeric complex glycans. Once the ability to determine exact glycan structure enters mainstream glycoscience, it will enable the determination of structure-activity relationships of glycans in much greater detail.

The synthesis of complex glycans has been revolutionized by the maturation of enzymatic and chemo-enzymatic synthesis approaches. Altough most glycan structures can now be accessed, there are some notable exceptions. For example, the enzymatic synthesis of *O*-glycans has been greatly hindered by low enzymatic activity on unnatural substrates. It also remains challenging to integrate labile modifications such as *O*-acetylation and sulfation in enzymatic synthesis due to their instability. A notable glycan class where progress has been especially difficult is the synthesis of keratan sulfates which has so far been unsuccesful. Other, much more rare modifications such as methylation or lactylations have neither been integrated into enzymatic nor in chemo-enzymatic syntheses. Even for glycans only containing common glycan epitopes, the synthetic effort required for their preparation is still significant. The integration of enzymatic and chemo-enzymatic methodology in robust automated synthesis platforms could greatly release the synthetic burden of complex glycan synthesis. The amount of potential structures given a certain composition increases exponentially when considering larger structures. There thus is a growing need for new analytical techniques that can determine exact glycan structure to guide glycan synthesis.

There are multiple factors that currently thwart exact determination of glycan structure: the ability to discriminate type 1 and 2 isomers of terminal glycan epitopes, the exact assignment of the arm glycosylation in multiantennary glycans, and the determination of internal Lewis^x fucosylation position for example. Just solving a single one of these factors currently requires specialized expertise, hence facile techniques need to be developed that can unambiguously assign exact glycan structure. Furthermore, the loss of labile modifications is often ignored in current analytical methods but leads to a false interpretation of the glycome. Conservation of labile groups, or the ability to assess these losses, should be enforced in the design of glycomic strategies. To enable this, glycan release strategies need to be developed that reflect the intact glycan structure.

The above described opportunities to advance glycoscience only consider the complexity of the glycan, whereas the true complexity of the glycome also encompasses the protein or lipid conjugate. The synthesis of all glycan standards is impossible; the added complexity when taking peptide/protein into account makes synthesis of even the most common glycopeptide/protein standards challenging. To determine exact glycosylation in complex glycoconjugates, techniques need to be developed that allows the assignment of key glycan features in complex glycans from information collected from limited glycan standards.

Thesis outline

The research described in this dissertation covers the development of chemo-enzymatic methodology for the regioselective synthesis of isomeric complex oligosaccharides. Synthetic standards are used to develop and assess mild *O*-glycan release methodology compatible with highly labile glycan epitopes. Furthermore, synthetic complex oligosaccharides are applied as reference compounds for the development of analytical methodology that can structurally elucidate challenging carbohydrate epitopes.

The first two research chapters describe the regioselective synthesis of complex carbohydrates with a focus on controlling the synthesis of internal glycan epitopes. In **Chapter 2**, synthetic methodology is developed that allows for the efficient chemo-enzymatic synthesis of orthogonally *N*-protected polyLacNAc. A single hexassacharide precursor made by this method is chemically manipulated to selectively (de)activate positions on the oligosaccharide towards enzymatic fucosylation. This strategy is applied in the synthesis of a glycan library consisting of polyLacNAc glycans functionalized by blood group antigens. These polyLacNAc glycans contain well-defined fucosylation patterns in addition to blood group antigens. A glycanmicroarray is contructed with the synthesized glycans and used to gain insights in the binding glycan binding proteins to blood group epitopes and the effect that α1,3 fucosylation can have the interactions.

Chapter 3 further builds on the chemo-enzymatic methodology developed in chapter 2. In this chapter, chemo-enzymatically introduced protecting groups are used to control activity of the I-branching enzyme GCNT2. Complex polyLacNAc architectures are generated for the first time in a regioselective manner by controlling GCNT2 activity on polyLacNAc featuring multiple acceptor sites. Efficient invertion of the activated and deactivated branching positions is achieved by exploiting glycosyltransferase selectivity and glycan hydrolysis kinetics.

The next two chapters demonstrate how high purity glycan standards are tools for the development and optimization of analytical methodology. In **Chapter 4**, glycopeptide standards containing labile glycan epitopes are used as tools to develop mild *O*-glycan release conditions. The effect of pH and time on the oxidative release of synthetic glycopeptide standards with hypochlorite is evaluated. Neutralization of hypochlorite is identified as an oxidative release condition that is able to efficiently release *O*-glycans with only single product formation. This method is the first time *O*-glycans are chemically released at neutral pH. The chemical stability of labile sialic acid, fucose and sulfate groups on *O*-glycans are evaluated by subjecting synthetic glycopeptide standards to the developed release conditions. The release kinetics of isomeric core 2 and 3 glycopeptide standards are evaluated to determine if *O*-glycan core structure influenced the release. Pure synthetic *O*-glycopeptides are used as starting material to identify intermediate products of the release reaction. These products allow the proposition of a putative reaction mechanism. The *O*-glycan content of the model mucin bovine submaxilary mucin was evaluated after neutralized hypochlorite release to determine the compability with natural mucin type proteins which can be sterically more crowded than the model *O*-glycopeptides. LC-MS shows that this purified mucin is efficiently cleaved under neutralized oxidative release conditions and observed *O*-glycans match previous reports. A higher concentration of the base labile *O*-acetyl modification on sialic acid is observed, as well as di- and tri-*O*-acetyled sialid acids and sulfated glycans that were not previously reported.

In **Chapter 5**, ion mobility is used in combination with mass spectrometry to distinguish isomeric glycan fragments. A library of synthetic, well-defined glycan standards featuring *O*-acetylated sialic acids is used to develop an approach to identify *O*-acetylation patterns in complex samples. Under in-source fragmentation conditions, key ionization fragments were characterized that are informative of the *O*-acetylation of the sialic acid. Distinctive collision cross-section values were determined of mono- di- and tri-O-acetylated Neu5Ac and Neu5Gc fragment ions by seperation of the isomeric fragments through a drift tube containing inert gas. These cross sections areas are molecular properties and we show that *O*-acetylated sialic acid fragments can be used to determine *O*-acetylation position. For the common sialyl LacNAc epitope, an additional ionization fragment could be identified that can be used to determine the sialic acid linkage type of the *O*-acetylated sialic acids in conjunction with *O*-acetyl position. The strength of this ion mobility method is demonstrated by detailed characterization of *O*-acetylation positions and sialic acid linkage of the *N*-glycans in two biotherapeutics. In the upper-airway tissues of a horse, this method could detect relationships between *O*-acetylation, sialic acid linkage and core-fucosylation that are not detectable with conventional techniques. This method was also applied to analyze the *O*-glycan content of bovine submaxillary mucin released by neutralized hypochlorite in chapter 4. By applying ion mobility analysis to this sample, we determined that the 7-*O*-acetylation is the most abbundant *O*-acetylation position in this sample. The observation of acetyl esters on this very labile 7 position strongly suggests that neutralized hypochlorite limits *O*-acetyl migration.

Chapter 6 summarizes the research described herein. This section concludes with future prospects inspired by this work which may influence future research to address stereoselective synthesis challenges and develop methods for the exact characterization of glycans.

References

[1] R. T. A. Megens, S. Reitsma, P. H. M. Schiffers, R. H. P. Hilgers, J. G. R. De Mey, D. W. Slaaf, M. G. A. oude Egbrink, M. a. M. J. van Zandvoort, J. Vasc. Res. 2007, 44, 87–98.

[2] S. Reitsma, M. G. A. oude Egbrink, H. Vink, B. M. van den Berg, V. L. Passos, W. Engels, D. W. Slaaf, M. A. M. J. van Zandvoort, J. Vasc. Res. 2011, 48, 297–306.

[3] B. M. van den Berg, J. A. E. Spaan, H. Vink, Pflüg. Arch. - Eur. J. Physiol. 2009, 457, 1199–1206.

[4] M. A. McGuckin, S. K. Lindén, P. Sutton, T. H. Florin, Nat. Rev. Microbiol. 2011, 9, 265–278.

[5] M. Bomsel, A. Alfsen, Nat. Rev. Mol. Cell Biol. 2003, 4, 57–68.

[6] C. B. Henry, B. R. Duling, Am. J. Physiol. 1999, 277, H508-514.

[7] H. H. Lipowsky, Adv. Exp. Med. Biol. 2018, 1097, 51–68.

[8] E. Tecle, P. Gagneux, Mol. Reprod. Dev. 2015, 82, 635–650.

[9] H. H. Lipowsky, Ann. Biomed. Eng. 2012, 40, 840–848.

[10] P. E. Constantinou, M. Morgado, D. D. Carson, in Regul. Implant. Establ. Pregnancy Mamm. Tribute 45 Year Anniv. Roger V Shorts Matern. Recognit. Pregnancy, 2015, pp. 51–68.

[11] Y. Zeng, X. F. Zhang, B. M. Fu, J. M. Tarbell, Adv. Exp. Med. Biol. 2018, 1097, 1–27.

[12] M. M. Thi, J. M. Tarbell, S. Weinbaum, D. C. Spray, Proc. Natl. Acad. Sci. 2004, 101, 16483–16488.

[13] M. J. Paszek, C. C. DuFort, O. Rossier, R. Bainer, J. K. Mouw, K. Godula, J. E. Hudak, J. N. Lakins, A. C. Wijekoon, L. Cassereau, M. G. Rubashkin, M. J. Magbanua, K. S. Thorn, M. W. Davidson, H. S. Rugo, J. W. Park, D. A. Hammer, G. Giannone, C. R. Bertozzi, V. M. Weaver, Nature 2014, 511, 319–325.

[14] S. S. Pinho, C. A. Reis, Nat. Rev. Cancer 2015, 15, 540–555.

[15] J. Qu, Y. Cheng, W. Wu, L. Yuan, X. Liu, Front. Cell Dev. Biol. 2021, 9.

[16] A. M. Dias, A. Correia, M. S. Pereira, C. R. Almeida, I. Alves, V. Pinto, T. A. Catarino, N. Mendes, M. Leander, M. T. Oliva-Teles, L. Maia, C. Delerue-Matos, N. Taniguchi, M. Lima, I. Pedroto, R. Marcos-Pinto, P. Lago, C. A. Reis, M. Vilanova, S. S. Pinho, PNAS. 2018, 115, E4651–E4660.

[17] S. Neelamegham, L. K. Mahal, Curr. Opin. Struct. Biol. 2016, 40, 145–152.

[18] L. L. Lairson, B. Henrissat, G. J. Davies, S. G. Withers, Annu. Rev. Biochem. 2008, 77, 521–555.

[19] Y. Liu, P. Huang, B. Jiang, M. Tan, A. L. Morrow, X. Jiang, PLOS ONE 2013, 8, e78113.

[20] A. Togayachi, Y. Kozono, H. Ishida, S. Abe, N. Suzuki, Y. Tsunoda, K. Hagiwara, A. Kuno, T. Ohkura, N. Sato, T. Sato, J. Hirabayashi, Y. Ikehara, K. Tachibana, H. Narimatsu, PNAS 2007, 104, 15829–15834.

[21] J. G. Sweeney, J. Liang, A. Antonopoulos, N. Giovannone, S. Kang, T. S. Mondala, S. R. Head, S. L. King, Y. Tani, D. Brackett, A. Dell, G. F. Murphy, S. M. Haslam, H. R. Widlund, C. J. Dimitroff, Nat. Commun. 2018, 9, 3368.
[22] A. P. Corfield, S. A. Wagner, A. Safe, R. A. Mountford, J. R. Clamp, J. P. Kamerling, J. F. Vliegenthart, R. Schauer, Clin. Sci. Lond. Engl. 1979 1993, 84, 573–579.
[23] E. A. Visser, S. J. Moons, S. B. P. E. Timmermans, H. de Jong, T. J. Boltje, C. Büll, J. Biol. Chem. 2021, 297, 100906.
[24] H. Geyer, R. Geyer, Biochim. Biophys. Acta BBA - Proteins Proteomics 2006, 1764, 1853–1869.
[25] C. E. Cardini, A. C. Paladini, R. Caputto, L. F. Leloir, Nature 1950, 165, 191–192.
[26] S. Goon, C. R. Bertozzi, J. Carbohydr. Chem. 2002, 21, 943–977.
[27] A. Ardèvol, J. Iglesias-Fernández, V. Rojas-Cervellera, C. Rovira, Biochem. Soc. Trans. 2016, 44, 51–60.
[28] E. C. O'Neill, R. A. Field, Carbohydr. Res. 2015, 403, 23–37.
[29] C. H. Wong, S. L. Haynie, G. M. Whitesides, J. Org. Chem. 1982, 47, 5416–5418.
[30] Y. Chen, V. Thon, Y. Li, H. Yu, L. Ding, K. Lau, J. Qu, L. Hie, X. Chen, Chem. Commun. 2011, 47, 10815–10817.
[31] H. Yu, H. Yu, R. Karpel, X. Chen, Bioorg. Med. Chem. 2004, 12, 6427–6435.
[32] R. M. Schmaltz, S. R. Hanson, C.-H. Wong, Chem. Rev. 2011, 111, 4259–4307.
[33] H. de Jong, M. M. S. M. Wösten, T. Wennekes, Glycobiology 2021, 32, 11–22.
[34] H. Yu, S. Huang, H. Chokhawala, M. Sun, H. Zheng, X. Chen, Angew. Chem. Int. Ed Engl. 2006, 45, 3938–3944.
[35] J. E. Park, K.-Y. Lee, S.-I. Do, S. S. Lee, J. Biochem. Mol. Biol. 2002, 35, 330–336.
[36] Y. Li, M. Xue, X. Sheng, H. Yu, J. Zeng, V. Thon, Y. Chena, M. M. Muthanaa, P. G. Wangb, X. Chena, Bioorg. Med. Chem. 2016, 24, 1696–1705.
[37] T. Yamamoto, M. Nakashizuka, I. Terada, J. Biochem. (Tokyo) 1998, 123, 94–100.
[38] K. W. Moremen, A. Ramiah, M. Stuart, J. Steel, L. Meng, F. Forouhar, H. A. Moniz, G. Gahlay, Z. Gao, D. Chapla, S. Wang, J.-Y. Yang, P. K. Prabhakar, R. Johnson, M. D. Rosa, C. Geisler, A. V. Nairn, J. Seetharaman, S.-C. Wu, L. Tong, H. J. Gilbert, J. LaBaer, D. L. Jarvis, Nat. Chem. Biol. 2018, 14, 156–162.
[39] P. K. Prabhakar, H.-T. Wang, P. J. Smith, J.-Y. Yang, W. J. Barnes, M. J. Peña, K. W. Moremen, B. R. Urbanowicz, Methods Cell Biol. 2020, 160, 145–165.
[40] J. M. Pierce, R. D. Cummings, Handbook of Glycomics - 1st Edition.
[41] N. Taniguchi, K. Honke, M. Fukuda, Handbook of Glycosyltransferases and Related Genes, Springer Science & Business Media, 2011.
[42] D. H. Joziasse, W. E. Schiphorst, D. H. van den Eijnden, J. A. van Kuik, H. van Halbeek, J. F. Vliegenthart, J. Biol. Chem. 1985, 260, 714–719.
[43] D. H. van den Eijnden, W. M. Blanken, A. van Vliet, Carbohydr. Res. 1986, 151, 329–335.
[44] F. Broszeit, "N-Glycans as Receptors for Influenza A Viruses," 2021.
[45] A. R. Prudden, L. Liu, C. J. Capicciotti, M. A. Wolfert, S. Wang, Z. Gao, L. Meng, K. W. Moremen, G.-J. Boons, Proc. Natl. Acad. Sci. 2017, 114, 6954–6959.

[46] H. Narimatsu, in Handb. Glycosyltransferases Relat. Genes. Springer Japan, Tokyo, 2002, pp. 246–251.
[47] S. Nishihara, H. Iwasaki, M. Kaneko, A. Tawada, M. Ito, H. Narimatsu, FEBS Lett. 1999, 462, 289–294.
[48] S. Nishihara, H. Iwasaki, K. Nakajima, A. Togayachi, Y. Ikehara, T. Kudo, Y. Kushi, A. Furuya, K. Shitara, H. Narimatsu, Glycobiology 2003, 13, 445–455.
[49] K. Handa, D. A. Withers, S. Hakomori, Biochem. Biophys. Res. Commun. 1998, 243, 199–204.
[50] T. Li, M. A. Wolfert, N. Wei, R. Huizinga, B. C. Jacobs, G.-J. Boons, J. Am. Chem. Soc. 2020, 142, 19611–19621.
[51] A. Santra, Y. Li, H. Yu, T. J. Slack, P. G. Wang, X. Chen, Chem. Commun. 2017, 53, 8280–8283.
[52] W. Ma, Y. Deng, Z. Xu, X. Liu, D. G. Chapla, K. W. Moremen, L. Wen, T. Li, J. Am. Chem. Soc. 2022, 144, 9057–9065.
[53] Z. Xu, Y. Deng, Z. Zhang, W. Ma, W. Li, L. Wen, T. Li, J. Org. Chem. 2021, 86, 10819–10828.
[54] S. Wang, C. Chen, M. R. Gadi, V. Saikam, D. Liu, H. Zhu, R. Bollag, K. Liu, X. Chen, F. Wang, P. G. Wang, P. Ling, W. Guan, L. Li, Nat. Commun. 2021, 12, 3573.
[55] A. Santra, Y. Li, T. Ghosh, R. Li, H. Yu, X. Chen, ACS Catal. 2019, 9, 211–215.
[56] Z. Xiao, Y. Guo, Y. Liu, L. Li, Q. Zhang, L. Wen, X. Wang, S. M. Kondengaden, Z. Wu, J. Zhou, X. Cao, X. Li, C. Ma, P. G. Wang, J. Org. Chem. 2016, 81, 5851–5865.
[57] S. Wang, Q. Zhang, C. Chen, Y. Guo, M. R. Gadi, J. Yu, U. Westerlind, Y. Liu, X. Cao, P. G. Wang, L. Li, Angew. Chem. 2018, 130, 9412–9417.
[58] Z. Wang, Z. S. Chinoy, S. G. Ambre, W. Peng, R. McBride, R. P. de Vries, J. Glush ka, J. C. Paulson, G.-J. Boons, Science 2013, 341.
[59] L. Li, Y. Liu, C. Ma, J. Qu, A. D. Calderon, B. Wu, N. Wei, X. Wang, Y. Guo, Z. Xiao, J. Song, G. Sugiarto, Y. Li, H. Yu, X. Chen, P. George Wang, Chem. Sci. 2015, 6, 5652–5661.
[60] T. Li, M. Huang, L. Liu, S. Wang, K. W. Moremen, G.-J. Boons, Chem. – Eur. J. 2016, 22, 18742–18746.
[61] S. Pawar, L. Hsu, T. Narendar Reddy, M. Ravinder, C.-T. Ren, Y.-W. Lin, Y.-Y. Cheng, T.-W. Lin, T.-L. Hsu, S.-K. Wang, C.-H. Wong, C.-Y. Wu, ACS Chem. Biol. 2020, 15, 2382–2394.
[62] I. A. Gagarinov, T. Li, J. S. Toraño, T. Caval, A. D. Srivastava, J. A. W. Kruijtzer, A. J. R. Heck, G.-J. Boons, J. Am. Chem. Soc. 2017, 139, 1011–1018.
[63] B. Echeverria, S. Serna, S. Achilli, C. Vivès, J. Pham, M. Thépaut, C. H. Hokke, F. Fieschi, N.-C. Reichardt, ACS Chem. Biol. 2018.
[64] A. D. Srivastava, L. Unione, M. Bunyatov, I. A. Gagarinov, S. Delgado, N.G.A. Abrescia, A. Ardá, G.-J. Boons, Angew. Chem. 2021, 60, 19287–19296.
[65] C. E. Kupper, R. R. Rosencrantz, B. Henßen, H. Pelantová, S. Thönes, A. Drozdová, V. Křen, L. Elling, Beilstein J. Org. Chem. 2012, 8, 712–725.
[66] N. Lu, J. Ye, J. Cheng, A. Sasmal, C.-C. Liu, W. Yao, J. Yan, N. Khan, W. Yi, A. Varki, H. Cao, J. Am. Chem. Soc. 2019, 141, 4547–4552.

[67] J. Ye, H. Xia, N. Sun, C.-C. Liu, A. Sheng, L. Chi, X.-W. Liu, G. Gu, S.-Q. Wang, J. Zhao, P. Wang, M. Xiao, F. Wang, H. Cao, Nat. Catal. 2019, 2, 514–522.
[68] R. Liu, Y. Xu, M. Chen, M. Weïwer, X. Zhou, A. S. Bridges, P. L. DeAngelis, Q. Zhang, R. J. Linhardt, J. Liu, J. Biol. Chem. 2010, 285, 34240–34249.
[69] L. Liu, A. R. Prudden, C. J. Capicciotti, G. P. Bosman, J.-Y. Yang, D. G. Chapla, K. W. Moremen, G.-J. Boons, Nat. Chem. 2019, 11, 161–169.
[70] I. A. Gagarinov, T. Li, N. Wei, J. Sastre Toraño, R. P. de Vries, M. A. Wolfert, G.-J. Boons, Angew. Chem. Int. Ed. 2019, 58, 10547–10552.
[71] C. J. Gray, L. G. Migas, P. E. Barran, K. Pagel, P. H. Seeberger, C. E. Eyers, G.-J. Boons, N. L. B. Pohl, I. Compagnon, G. Widmalm, S. L. Flitsch, J. Am. Chem. Soc. 2019, 141, 14463–14479.
[72] G. R. Guile, P. M. Rudd, D. R. Wing, S. B. Prime, R. A. Dwek, Anal. Biochem. 1996, 240, 210–226.
[73] G. Menghi, D. Accili, A. M. Bondi, M. G. Gabrielli, Histochemistry 1989, 90, 331–338.
[74] L. Ali, D. T. Kenny, C. A. Hayes, N. G. Karlsson, Metabolites 2012, 2, 648–666.
[75] H. J. An, J. W. Froehlich, C. B. Lebrilla, Curr. Opin. Chem. Biol. 2009, 13, 421–426.
[76] W. B. Struwe, A. Stuckmann, A.-J. Behrens, K. Pagel, M. Crispin, ACS Chem. Biol. 2017, 12, 357–361.
[77] T. Čaval, A. J. R. Heck, K. R. Reiding, Mol. Cell. Proteomics MCP 2021, 20, 100010.
[78] L. Turiák, S. Sugár, A. Ács, G. Tóth, Á. Gömöry, A. Telekes, K. Vékey, L. Drahos, Sci. Rep. 2019, 9, 14822.
[79] M. I. Kelly, E. D. Dodds, J. Am. Soc. Mass Spectrom. 2020, 31, 624–632.
[80] N. M. Riley, S. A. Malaker, M. D. Driessen, C. R. Bertozzi, J. Proteome Res. 2020, 19, 3286–3301.
[81] A. L. Tarentino, R. B. Trimble, T. H. Plummer, in Methods Cell Biol. Academic Press, 1989, pp. 111–139.
[82] T. Muramatsu, N. Koide, K. Maeyama, J. Biochem. (Tokyo) 1978, 83, 363–370.
[83] R. B. Trimble, A. L. Tarentino, J. Biol. Chem. 1991, 266, 1646–1651.
[84] D. A. Fischler, R. Orlando, J. Biomol. Tech. JBT 2019, 30, 58–63.
[85] P. Rahfeld, J. F. Wardman, K. Mehr, D. Huff, C. Morgan-Lang, H.-M. Chen, S. J. Hallam, S. G. Withers, J. Biol. Chem. 2019, 294, 16400–16415.
[86] J. F. Wardman, P. Rahfeld, F. Liu, C. Morgan-Lang, L. Sim, S. J. Hallam, S. G. Withers, ACS Chem. Biol. 2021, 16, 2004–2015.
[87] Y. Goso, T. Sugaya, K. Ishihara, M. Kurihara, Anal. Chem. 2017, 89, 8870–8876.
[88] A. Kameyama, W. W. Thet Tin, M. Toyoda, M. Sakaguchi, Biochem. Biophys. Res. Commun. 2019, 513, 186–192.
[89] D. M. Carlson, J. Biol. Chem. 1968, 243, 616–626.
[90] Y. Huang, T. Konse, Y. Mechref, M. V. Novotny, Rapid Commun. Mass Spectrom. RCM 2002, 16, 1199–1204.
[91] L. R. Ruhaak, E. Steenvoorden, C. A. M. Koeleman, A. M. Deelder, M. Wuhrer, PROTEOMICS 2010, 10, 2330–2336.

[92] J. C. Bigge, T. P. Patel, J. A. Bruce, P. N. Goulding, S. M. Charles, R. B. Parekh, Anal. Biochem. 1995, 230, 229–238.
[93] R. P. Kozak, L. Royle, R. A. Gardner, D. L. Fernandes, M. Wuhrer, Anal. Biochem. 2012, 423, 119–128.
[94] R. P. Kozak, L. Royle, R. A. Gardner, A. Bondt, D. L. Fernandes, M. Wuhrer, Anal. Biochem. 2014, 453, 29–37.
[95] Y. Miura, K. Kato, Y. Takegawa, M. Kurogochi, J.-I. Furukawa, Y. Shinohara, N. Nagahori, M. Amano, H. Hinou, S.-I. Nishimura, Anal. Chem. 2010, 82, 10021–10029.
[96] G. Yu, Y. Zhang, Z. Zhang, L. Song, P. Wang, W. Chai, Anal. Chem. 2010, 82, 9534–9542.
[97] C. Wang, W. Fan, P. Zhang, Z. Wang, L. Huang, PROTEOMICS 2011, 11, 4229–4242.
[98] X. Song, H. Ju, Y. Lasanajak, M. R. Kudelka, D. F. Smith, R. D. Cummings, Nat. Methods 2016, 13, 528–534.
[99] Q. Zhang, Z. Li, X. Song, Front. Chem. 2020, 8, 508.
[100] Y. Zhu, M. Yan, Y. Lasanajak, D. F. Smith, X. Song, Carbohydr. Res. 2018, 464, 19–27.
[101] L. R. Ruhaak, G. Xu, Q. Li, E. Goonatilleke, C. B. Lebrilla, Chem. Rev. 2018, 118, 7886–7930.
[102] P. H. Jensen, N. G. Karlsson, D. Kolarich, N. H. Packer, Nat. Protoc. 2012, 7, 1299–1310.
[103] N. Leymarie, J. Zaia, Anal. Chem. 2012, 84, 3040–3048.
[104] B. Domon, C. Costello, Glycoconj. J. 2005, DOI 10.1007/BF01049915.
[105] K. Håkansson, H. J. Cooper, M. R. Emmett, C. E. Costello, A. G. Marshall, C. L. Nilsson, Anal. Chem. 2001, 73, 4530–4536.
[106] J. Shen, L. Jia, L. Dang, Y. Su, J. Zhang, Y. Xu, B. Zhu, Z. Chen, J. Wu, R. Lan, Z. Hao, C. Ma, T. Zhao, N. Gao, J. Bai, Y. Zhi, J. Li, J. Zhang, S. Sun, Nat. Methods 2021, 18, 921–929.
[107] K. R. Reiding, D. Blank, D. M. Kuijper, A. M. Deelder, M. Wuhrer, Anal. Chem. 2014, 86, 5784–5793.
[108] T. Pongracz, A. Verhoeven, M. Wuhrer, N. de Haan, Glycoconj. J. 2021, 38, 157–166.
[109] E. Mucha, M. Lettow, M. Marianski, D. A. Thomas, W. B. Struwe, D. J. Harvey, G. Meijer, P. H. Seeberger, G. von Helden, K. Pagel, Angew. Chem. Int. Ed. 2018, 57, 7440–7443.
[110] M. Wuhrer, C. A. M. Koeleman, C. H. Hokke, A. M. Deelder, Rapid Commun. Mass Spectrom. RCM 2006, 20, 1747–1754.
[111] M. Wuhrer, C. A. M. Koeleman, A. M. Deelder, Anal. Chem. 2009, 81, 4422–4432.
[112] C. J. Gray, B. Thomas, R. Upton, L. G. Migas, C. E. Eyers, P. E. Barran, S. L. Flitsch, Biochim. Biophys. Acta BBA - Gen. Subj. 2016, 1860, 1688–1709.

[113] J. Sastre Torano, "Analysis of Glycans by Electrophoretic Approaches: Highly Sensitive Capillary Electrophoretic Separation and Ion Mobility-Mass Spectrometric Structure Assignment,"
[114] H. Li, B. Bendiak, W. F. Siems, D. R. Gang, H. H. Hill, Rapid Commun. Mass Spectrom. RCM 2013, 27, 2699–2709.
[115] J. Hofmann, K. Pagel, Angew. Chem. Int. Ed Engl. 2017, 56, 8342–8349.
[116] E. G. Pallister, M. S. F. Choo, I. Walsh, J. N. Tai, S. J. Tay, Y. S. Yang, S. K. Ng, P. M. Rudd, S. L. Flitsch, T. Nguyen-Khuong, Anal. Chem. 2020, 92, 15323–15335.
[117] H. Hinneburg, J. Hofmann, W. B. Struwe, A. Thader, F. Altmann, D. V. Silva, P. H. Seeberger, K. Pagel, D. Kolarich, Chem. Commun. 2016, 52, 4381–4384.
[118] M. Guttman, K. K. Lee, Anal. Chem. 2016, 88, 5212–5217.
[119] C. S. Lane, K. McManus, P. Widdowson, S. A. Flowers, G. Powell, I. Anderson, J. L. Campbell, Anal. Chem. 2019, 91, 9916–9924.
[120] Y. Pu, M. E. Ridgeway, R. S. Glaskin, M. A. Park, C. E. Costello, C. Lin, Anal. Chem. 2016, 88, 3440–3443.
[121] J. Hofmann, A. Stuckmann, M. Crispin, D. J. Harvey, K. Pagel, W. B. Struwe, Anal. Chem. 2017, 89, 2318–2325.
[122] J. Sastre Toraño, I. A. Gagarinov, G. M. Vos, F. Broszeit, A. D. Srivastava, M. Palmer, J. I. Langridge, O. Aizpurua-Olaizola, V. J. Somovilla, G.-J. Boons, Angew. Chem. Int. Ed Engl. 2019, 58, 17616–17620.
[123] E. A. Mason, H. W. Schamp, Ann. Phys. 1958, 4, 233–270.
[124] A. Marchand, S. Livet, F. Rosu, V. Gabelica, Anal. Chem. 2017, 89, 12674–12681.
[125] J. C. May, R. Knochenmuss, J. C. Fjeldsted, J. A. McLean, Anal. Chem. 2020, 92, 9482–9492.
[126] T. L. Peterson, G. Nagy, Anal. Chem. 2021, 93, 9397–9407.
[127] G. Nagy, I. K. Attah, S. V. B. Garimella, K. Tang, Y. M. Ibrahim, E. S. Baker, R. D. Smith, Chem. Commun. 2018, 54, 11701–11704.
[128] P. Both, A. P. Green, C. J. Gray, R. Šardzík, J. Voglmeir, C. Fontana, M. Austeri, M. Rejzek, D. Richardson, R. A. Field, G. Widmalm, S. L. Flitsch, C. E. Eyers, Nat. Chem. 2014, 6, 65–74.

CHAPTER 2

A Stop-and-Go Strategy for the Chemo-Enzymatic Synthesis of selectively Fucosylated Polylactosamine Derivatives

Gaël M. Vos, Gerlof Bosman, Yunfei Wu, Lin Liu, Margreet Wolfert, Geert-Jan Boons

Introduction

Poly-*N*-acetyllactosamines (polyLacNAc's), which are also known as i-antigens, consist of repeating units of type 2 *N*-acetyllactosamine (LacNAc) [Galβ1,4GlcNAcβ1,3]. These structures are found on a variety of mammalian glycan classes, including *N*- and *O*-glycans, glycosphingolipids, and human milk oligosaccharides.[1–4] PolyLacNAc's can be substituted at the terminus with a variety of epitopes such as sialosides, HNK-1, and Lewis- and blood group antigens.[5] PolyLacNAc chains can also undergo internal glycosylation such as β1,6-branching by *N*-acetyl-glucosamine at galactose to give I-antigens or by α1,3-fucosylation at GlcNAc to form Lewis^x epitopes. These internal epitopes can modulate interactions with glycan binding proteins such as galectins and E-selectin that have regulatory functions and are involved in inflammation, immunity and cancer.[6,7] A number of pathogens, such as *H. pylori* and *S. mansoni*, also express polyLacNAc chains modified by Lewis^x epitopes, thereby mimicking human glycans resulting in host immune evasion.[8–12]

Although it is appreciated that the complex architecture of polyLacNAc chains is important for many biological and disease processes, exact ligand requirements of relevant glycan binding proteins are not well understood. This is in part is due to a lack of convenient methods to synthesize panels of polyLacNAc derivatives having different patterns of internal fucosylation and modified by a variety of terminal epitopes.[13] Lewis^x containing polyLacNAc chains have been chemically synthesized, however, such approaches only provided one-compound-at-a-time that require a relatively large number of steps for oligosaccharide assembly, and are associated with difficulties to stereoselectively install fucosides.[14–16] On the other hand, enzymatic approaches can provide polyLacNAc chains much faster but are hindered by lack of regioselectivity of fucosyl transferases to install Lewis^x moieties.[17–22] To address these limitations, we reported a chemo-enzymatic strategy to access all possible fucosylation patterns of an oligo-LacNAc chain from a single advanced precursor (Fig. 1).[23] Key to the approach was the chemical synthesis of a tri-LacNAc derivative in which the amines are differently modified including a free amine, a trifluoroacetamide (TFA) and a tert-butyloxycarbamate (Boc) (Fig. 1B). This precursor could easily be converted into a panel of compounds having various patterns of natural GlcNAc, GlcNH_2 and GlcNBoc. We exploited the find that the NHBoc and NH_2 block fucosylation by the fucosyl transferases FUT5 and Hp39-FT to give a range of oligo-LacNAc chains having different patterns of Lewis^x epitopes. Next, the Boc groups could be removed by treatment with trifluoroacetic acid (TFA) and acetylation of the free amines gave a panel of Lewis^x containing compounds. In a subsequent study, it was shown that 2,6-sialylation of galactosides can also block 1,3-fucosylation to provide differently fucosylated compounds.[24]

Here, we report a stop-and-go-strategy[25] for the facile assembly of polyLacNAc chains having various types of terminal epitopes and different patterns of fucosylation. Integral to this strategy is the finding that the unnatural donor UDP-GlcNTFA is readily accepted by the microbial glycosyl transferase Hpβ3GlcNAcT to install a GlcNTFA moiety (Fig. 1C), which can be extended by B4GalT1 in the presence of UDP-Gal to provide LacNTFA.[26,27] Base mediated cleave of the TFA moiety will give LacNH_2 that can be protected by for example Boc to afford LacNBoc. An important observation was that the resulting compound is an appropriate substrate to undergo a similar cycle of enzymatic and chemical modifications to

give oligosaccharides in which the amines are orthogonally modified. The methodology was employed to prepare hexasaccharide **24**, in which the amines are modifed in such a way that it can control fucosylation. Furthermore, by employing appropriate glycosyl transferases, it was possible to install a variety of terminal epitopes. The compounds are equipped with an anomeric aminopentyl linker that is protected a benzyloxycarbamate (Cbz) which is stable under conditions to remove TFA and Boc but at the end of a synthetic sequence can readily be deprotected by hydrogenation to give a free amine that, for example, can be employed for microarray printing. The attraction of the new chemo-enzymatic strategy is that far fewer steps are required to prepare a common precursor that can provide a wide variety of derivatives. To demonstrate the versatility, the methodology was applied for the preparation of glycans **1**-**18** which have various patterns of fucosylation and terminate in ABO blood group epitopes (Fig. 1D). The compounds were printed as a glycan microarray to investigate binding selectivities of several plant lectins, galectins and ABO blood group binding antibodies.

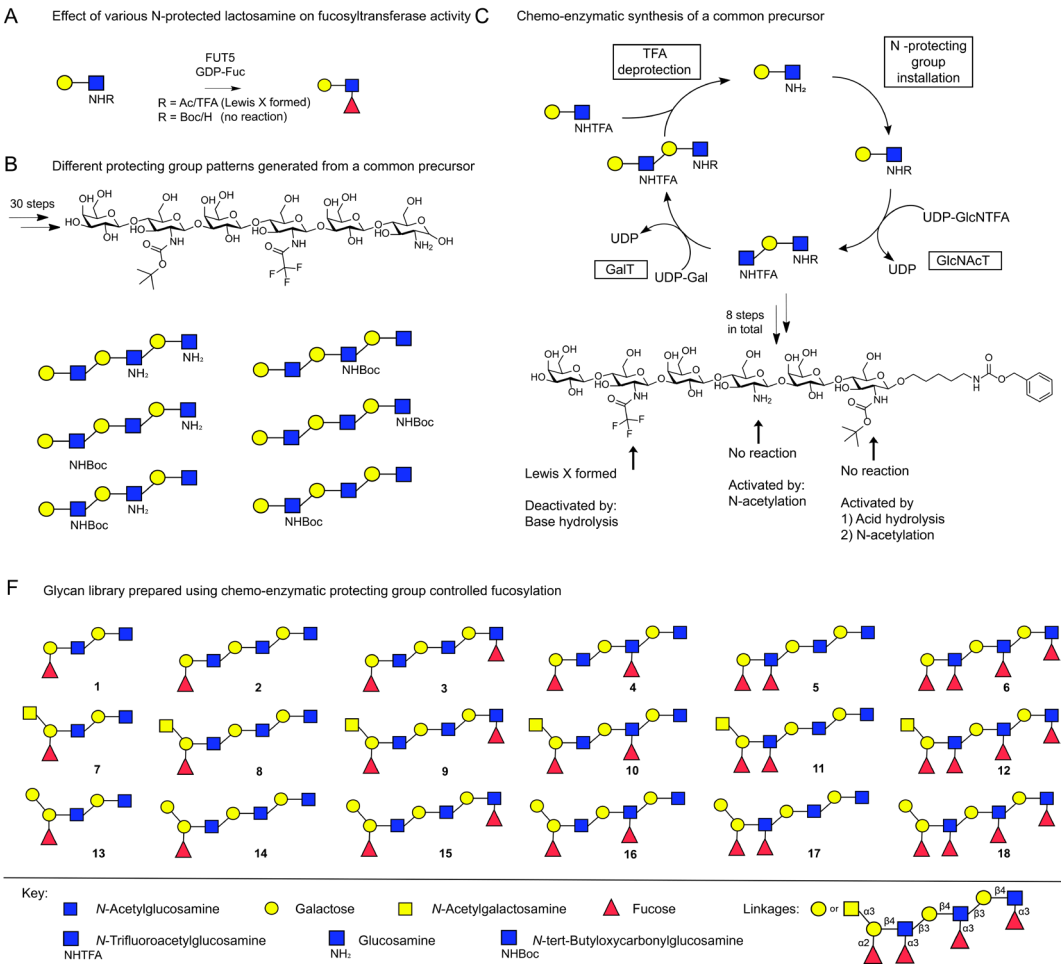


Figure 1: A) The fucosyl transferase FUT5 tolerates Ac and TFA groups but is blocked by NH_2 and NHBoc. B) A chemical synthesized common precursor was used to generate six protecting group patterns that allow for selective fucosylation. C) (top) Scheme depicting the chemo-enzymatic strategy for the synthesis of orthogonally *N*-protected polyLacNAc. R = $(CH_2)_5NHCbz$. (bottom) Chemical structure of the common precursor obtained by chemo-enzymatic synthesis. D) Overview of printed glycans on glycan microarray.

Results and discussion

First, attention was focused on the chemoenzymatic assembly of common precursor **24**. Thus, GlcNTFA derivative **19** modified by a Cbz protected aminopentyl linker was subjected to enzymatic extention by recombinant B4GalT1 in the presence of UDP-Gal to give a disaccharide that was treated with aqueous ammonia to provide, after purification by size exclusion chromatography, lactosamine derivative **20** in a yield of 89% (Fig. 2). Importantly, disaccharide **20** could readily be converted into a trisaccharide by the microbial glycosyl transferase Hpβ3GlcNAcT[26,27] in the presence of UDP-GlcTFA and further treatment by B4GalT1 and UDP-Gal gave tetrasaccharide **21**. To retain the ability to differentiate the amines of compounds **21**, its free amine was protected as Boc by treatment with Boc_2O in the presence of sodium bicarbonate to give LacNTFA-LacNBoc **22**. The Boc protecting was selected because it is orthogonal to TFA and Cbz, allowing the three different amino protecting groups to be selectively removed by moderate acid, mild base or hydrogenation, respectively. Subsequent cleavage of the TFA moiety of **22** gave tetrasaccharide **23**, which was extended by a LacNTFA moiety using the afore employed reaction cycle to generate common precursor **24** that has NHBoc, NHTFA and NH_2 as chemical differentiating groups. This procedure greatly reduced synthetic efforts to prepare a common precursor compared to our previously reported chemical approach.[23]

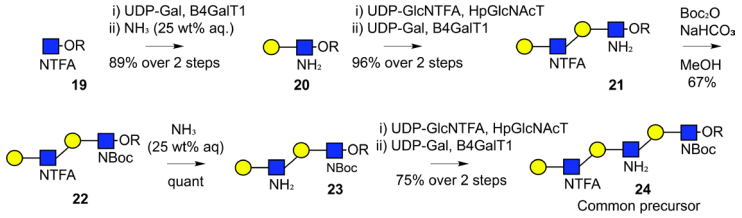


Figure 2: Chemo-enzymatic synthesis of compound **24**. R = $(CH_2)_5NHCbz$

Hexasaccharide **24** was employed to prepare the three isomeric α1,3-fucosylated glycans **3**-**5** (Fig. 3). Thus, treatment of **24** with FUT1 and GDP-Fuc installed an H-type antigen (blood group O), which was used without purification to install a Lewis^y motif (→**25**) by treatment with FUT9 in the presence of GDP-Fuc. In the latter transformation, it was exploited that the NHTFA of the terminal LacNAc moiety is tolerated by FUT9, whereas the free amine and Boc of the central and reducing end LacNAc moiety, respectively block the activity of this fucosyltransferase. Furthermore, FUT9 has higher distal activity of on neutral polyLacNAc's which may also contribute to the highly selective fucosylation of this unnatural acceptor.[21] Next, the Boc protecting group of **25** was removed by treatment with 20% TFA, which was followed by aminolysis of the TFA group to give a di-amine that was acetylated with acetic anhydride in aqueous sodium bicarbonate to give Lewis^y containing octasaccharide **26**. The latter compound was subject to hydrogenation to remove the benzyloxycarbamate of the anomeric linker to give target compound **5**.

Next, attention was focused on the preparation of glycan **4** which has a terminal H-type antigen and an internal Lewis^x moiety. For the synthesis of this compound, common precursor **24** was transformed into derivative **27** by aminolysis of the TFA moiety. We exploited that the free amine of the distal and Boc on the proximal LacNAc moiety of **27** prevent fucosylation by FUT5, which is a fucosyl transferase that installs Lewis^x epitopes. As anticipated exposure

of **27** to FUT5 and GDP-Fuc resulted in mono-fucosylation of the central LacNAc moiety. Next, we exploited that FUT1 only modifies terminal galactosides to introduce an H-type antigen.[28] Interestingly, it was found that that the latter fucosyl transferase can modify a $LacNH_2$ moiety readily providing octasaccharide **28**. Cleavage of the Boc group of **28** followed by acetylation of the free amines gave **29**, which was subjected to hydrogenation to provide targeted compound **4**.

Finally, derivative **3** was prepared by conversion of the common intermediate **24** into **30** by protection of the free amine as TFA by treatment with TFAOMe in the presence of Et_3N in methanol to give an intermediate that was purified by benchtop porous graphatized carbon solid phase extraction. The Boc group was removed under standard conditions which was followed by acetylation of the resulting free amine followed by aminolyis of the TFA protecting groups to give **30**. The free amines of the central and distal LacNAc residue of **30** prevented fucosylation by FUT5 to give readily a mono-fucosylated derivative, which was subjected to FUT1 to afford compound **31**. Next, free amine of **31** were acetylated and the Cbz protecting

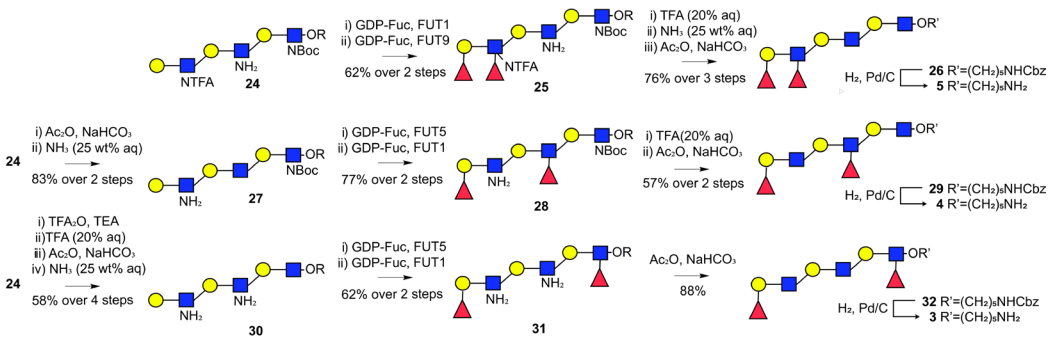


Figure 3: Selective diversification of compound **1** into blood group O glycans **3**-**5**. R = $(CH_2)_5NHCbz$.

Having successfully prepared isomeric glycans **3**-**5**, similar derivatives were prepared modified by a blood group A or B antigen (**9**-**11** and **15**-**17**). It exploited the microbial glycosyltransferases BoGT6a[29] and the human glycosyltransferase hGTB[30] that can modify a blood group O antigen (H-type) by β(1,3)GalNAc or β(1,3)Gal, respectively to form blood A and B antigens. Thus, exposure of **29** and **32** to BoGT6a and UDP-GalNAc or hGTB and UDP-Gal generated blood group A containing compounds **33** and **35** (Fig. 4A), and blood group B containing compounds **34** and **36**, which were readily converted into amino-pentyl derivatives **9**, **10**, **15** and **16** by standard manipulations. Both enzymes did not accept Lewis^y containing derivative **26** as acceptor and thus, we first installed the blood group epitope on common precursor **24** (Fig. 4B) to form compounds **37** and **40**, respectively, which were subjected to FUT9, which resulted in the formation of **38** and **41**, having an A-Lewis^y or B-Lewis^y epitope, respectively. After removal of the protecting groups and *N*-acetylation, glycans **39** (A-Lewis^y) and **42** (B-Lewis^y) were obtained, which were subjected to hydrogenation to remove the Cbz protecting group to give **11** and **17**, respectively and complete the set of mono-α1,3-fucosylated glycans.

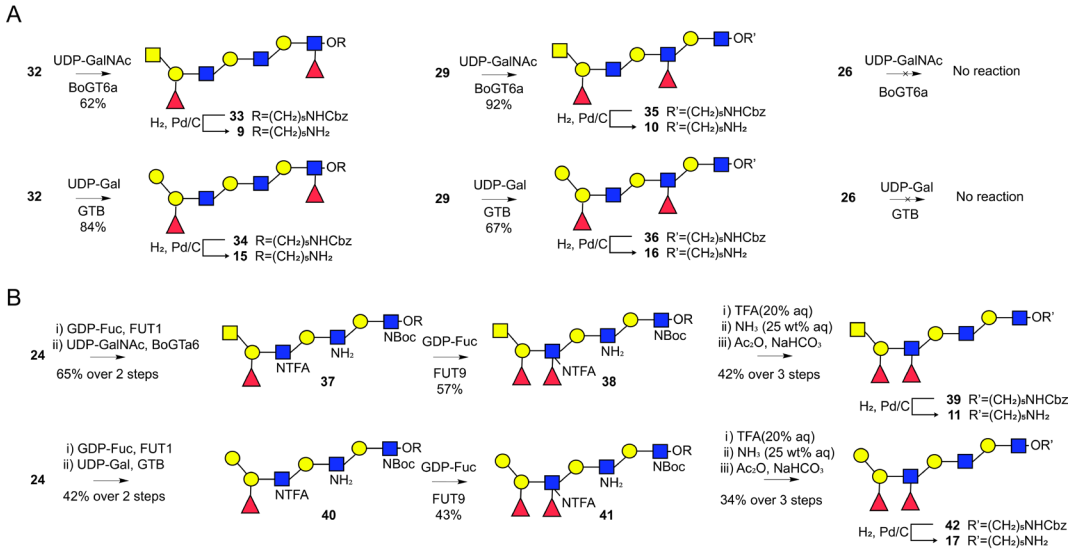


Figure 4: A) Installation of blood group A and B on **29** and **32**. B) Chemo-enzymatic synthesis of A-Lewis^y and B-Lewis^y glycan epitopes. R = $(CH_2)_5NHCbz$.

Glycans **1**, **2**, **6**, **7**, **8**, **12**, **13**, **14** and **18** were selected as additional compounds to uncover the effects of α1,3-fucosylation on binding selectivities of glycan binding proteins. These glycans have either none or all GlcNAc positions modified by α1,3-fucosides. Glycans **2**, **6**, **8**, **12**, **14** and **18** could in principle be prepared from common precursor **24** by protecting group removal followed by *N*-acetylation, however, such a strategy would have poor atom efficiency due to unnecessary protecting group manipulations. Therefore, oligo-LacNAc backbones were assembled using natural sugar nucleotide donors. Linker modified LacNAc acceptor **43** was conveniently accessed by *N*-acetylation of **20**. This derivative was extended by additional LacNAc units by the alternate action of B3GNT2 and B4GalT1 in combination with natural donor sugar-nucleotides UDP-GlcNAc and UDP-Gal, respectively (Fig. 5) and by employing one or two extension cycles tetrasaccharide **44** and hexasaccharide **48** were readily obtained. Compounds **44** and **48** could be modified by an H-type epitope by treatment with FUT1, and subsequently to blood group A and B epitopes by treatment with BoGT6a and hGTB, respectively in one pot reactions to form compounds **45**, **46**, **47**, **50**, **51** and **53**. Structures having Lewis^y-di-Lewis^x epitopes in which every GlcNAc is α1,3-fucosylated, could be generated by treatment of **50**, **51** and **53** with FUT9 in the presence of an excess of GDP-Fuc forming **49**, **52** and **54**, respectively, completing the targeted compound library.

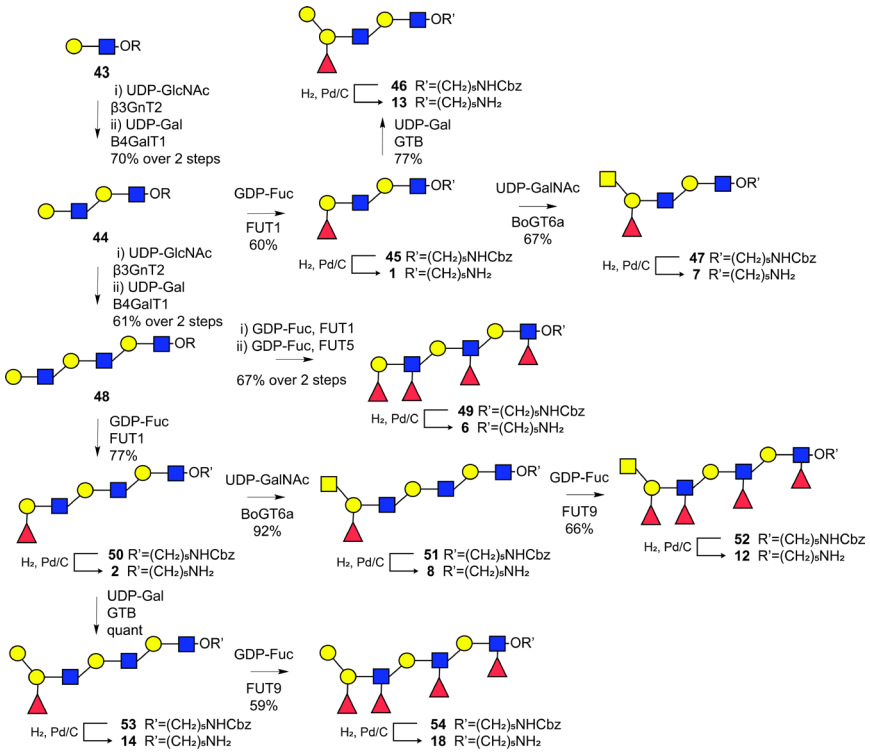


Figure 5: Enzymatic synthesis of di-LacNAc ABO antigens, tri-LacNAc ABO antigens and, tri-LacNAc ABO antigens with Lewis^y di-Lewis^x motif. R = $(CH_2)_5NHCbz$.

Unraveling the galectin/ABO blood group interactome by glycan microarray analysis

Galectins are a class of glycan binding proteins that play important roles in diverse physiological processes such as in immunity, inflammation and cancer progression and have therapeutic potential.[7,31–35] Evidence is accumulating that ABO blood group epitopes, commonly expressed in the cardiovascular system, can bind galectins.[7,36–41] When these ABO epitopes contain an α1,3-fucoside, forming di-fucosylated Lewis^y, A-Lewis^y and B-Lewis^y antigens, the binding to galectins is largely impaired.[42] This indicates that α1,3-fucosylation can modulate galectin/ABO antigen interactions. However, these studies have focused on mono-LacNAc epitopes even though most galectins have a preference for larger polyLacNAc's that are present on the *N*-glycans, *O*-glycans and glycolipids.[43–45]

Galectins 1, 3, 8 and 9 are the main galectins expressed in vascular endothelial cells, and endothelial activation by tumor cells leads to alterations in the quantities and localization of these galectins.[46] Similar modulations are reported in response to injury, interferon γ or viral RNA which underscores the role of galectins in vascular immune and inflammation responses.[47–49] The multimeric binding mode of galectins allows galectins to modulate the interaction of T-cells, B-cells and leukocytes with the vascular endothelium.[50,51] For these galectins, interactions with ABO antigens have been reported, which is interesting because of the presence of type 2 ABO antigens in vascular endothelium.[52,53] Binding to ABO antigens has also been reported for galectin 4, which is expressed by vascular endothelial

cells albeit at lower mRNA levels.[46] Furthermore, it was demonstrated that knockout of the polyLacNAc generating enzyme B3GNT2 results in a decreased binding of galectins 1, 3, 8 and 9 to CHO cells.[54] This polyLacNAc component of a glycan can be important factor in galectin interactions as most galectins such as galectin 3, 7, 8 and 9 are reported to bind both internal and terminal epitopes.[34,55] The interaction of mono-LacNAc ABO epitopes with galectins can only substitute for the terminal binding component of galectins. The effects that internal α1,3-fucosylation can have on the binding of polyLacNAc ABO antigens to galectins remains elusive. To better understand the internal binding component of galectins and how these are modulated by α1,3-fucosylation, we investigated the binding of galectins expressed in vascular endothelium with the synthesized library of complex glycans.

An ABO glycan microarray was generated from the synthetic glycans **1-18** by piezoelectric printing onto NHS activated glass slides. The slides were examined with *Aleuria aurantia* (AAL: α1-2, α1-3, or α1-6 fucose), *Erythrina cristagalli lectin* (ECL: Galβ1-4GlcNAc), *Wheat germ agglutinin* (WGA: β-GlcNAc) and *Ulex europaeus agglutinin I* (UEA: α1-2 fucose) which showed consistent binding with proper substrate disposition and spot integrity (see supplementary information, Fig. S1). The presence of blood group antigens was evaluated with anti-Blood Group H antibody 87-N, anti-Blood Group B antibody HEB-29 and, anti-Blood Group A antibody HE-193 and most blood groups were detected as expected, except 87-N did not show the expected binding to compound **4** (see supplementary information, Fig. S2). After confirming proper printing of the glycans, the binding of galectins 1, 3, 4, 7, 8 and 9 was investigated (Fig. 6A-F).

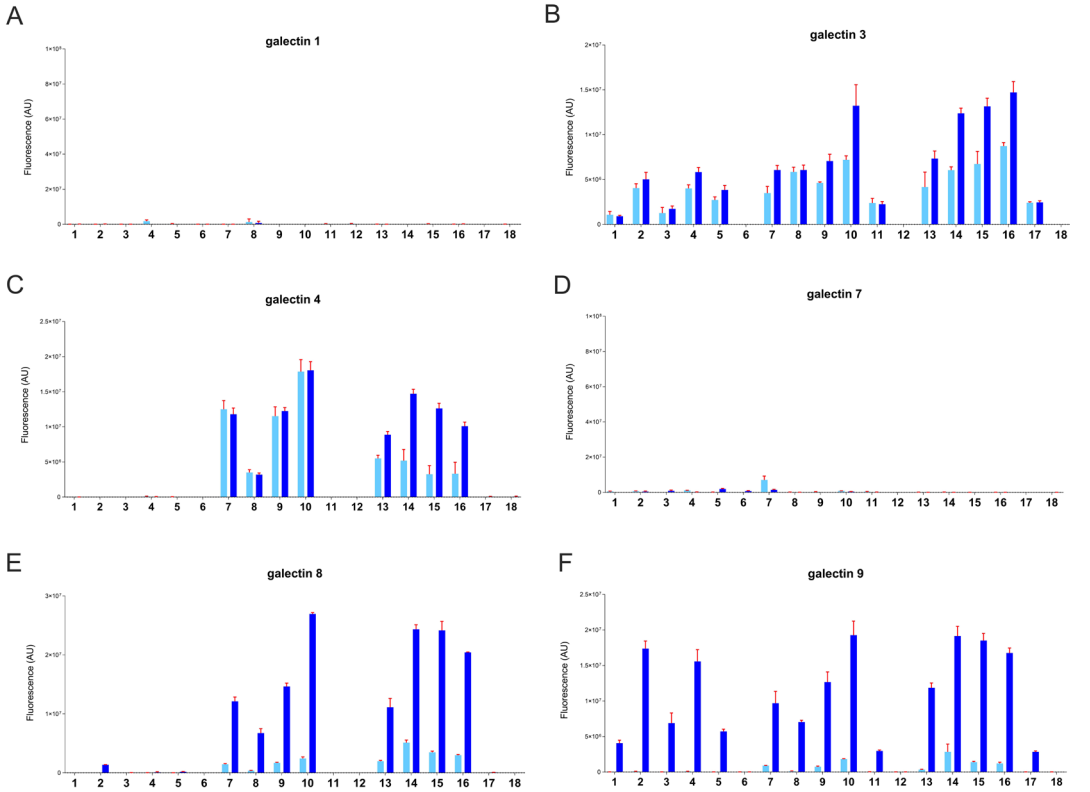


Figure 6: Microarray results of the synthetic ABO glycan library (**1-18**) at 100 µM with (A) galectin 1, (B) 3, (C) 4, (D) 7, (E) 8, and (F) 9. Galectin concentrations are (A-E) 1 and 10 µg/mL and (F) 0.1 and 1 µg/mL (light and dark blue bars, respectively). Bars represent the mean ± SD (n=4). Results of each galectin are depicted at a scale whereby the background is approximately 0.02% of the maximum value.

Galectin 3 and 9 showed a similar binding profile and recognized compounds **1-5**, **7-11** and **13-17** whereas no responsiveness was observed for derivatives **6**, **12** and **18**. Earlier studies, using mono-LacNAc ABO antigens, showed that galectin 3 binds blood group A and B epitopes with higher affinity than blood group O.[40,56] This trend was also observed in the binding profile obtained with the new polyLacNAc microarray, where A and B epitopes showed indeed higher fluorescent readouts. Galectin 9 is a tandem repeat galectin, composed of two distinct carbohydrate binding domains with different but overlapping binding to ABO blood group antigens.[42,57] Full length galectin 9 showed binding to blood group A, B and O epitopes, which was consistent with previous reports. For both galectins, it appears that mono-Lewis^x fucosylation on the middle (**4**, **10** and **16**) and proximal LacNAc (**3**, **9** and **15**) does not influence binding. When the terminal LacNAc is fucosylated, forming (A/B) Lewis^y epitopes (**5**, **11** and **17**), a substantial reduction in binding was observed. This binding to Lewis^y epitope containing oligosaccharides was not reflected in reported data that focused on smaller mono-LacNAc oligosaccharides.[42] The introduction of α1,3-fucoside on all GlcNAc positions (**6**, **12** and **18**) abolished binding to both galectin 3 and 9. This observation suggests that the binding interaction of **5**, **11** and **17** is not mediated through the terminal epitope but through the unsubstituted, internal LacNAc motifs. Previous STD-HSQC NMR experiments of unsubstituted polyLacNAc showed that both galectin 3 and 9 can indeed interact with internal LacNAc moieties.[34] Using the new glycan microarray, binding patterns of galectins 3 and 9 to ABO antigens are observed that are consistent with a flexible binding mode that can be regulated by α1,3-fucosylation.

Next, attention was focused on galectins 4 and 8, which are both tandem repeat galectins that showed a similar binding profile on the blood group glycan microarray. Both galectins showed interactions to the A and B antigens (**7-10** and **13-16**) but not to O antigens. Proximal fucosylated compounds **9** and **15** and compounds **10** and **16** that are fucosylated on the central LacNAc showed similar binding as A/B blood groups that do not contain Lewis^x (**7**, **8**, **13** and **14**). This binding behavior suggests that internal fucosylation has limited or no effect on the binding of galectins 4 and 8. In contrast, compounds **11**, **12**, **17** and **18** that contain Lewis^y on the terminal LacNAc abolished binding. Interestingly, substantial binding to internal LacNAc of unsubstituted polyLacNAc has been reported for galectin 8, which we did not observe on the glycan microarray. It is possible that this internal binding is weak relative to the stronger interaction of galectin 8 with blood group A and B epitopes, and is therefore not detected on glycan microarray.[34] This result demonstrates that the ability of galectins to bind to internal LacNAc epitopes does not always contribute substantially to binding.

Lastly, we examined the binding of galectins 1 and 7, which are reported to interact mainly with terminal epitopes[34] and therefore no influence of the internal fucosylation was expected. No binding was observed for galectin 7 which is in agreement with previous data that it has very low affinity for ABO antigens and prefers type 1 LacNAc and multi-antennary *N*-glycans.[42] Literature reports for galectin 1 have shown binding to unsubstituted and blood group O containing $(LacNAc)_3$ but not to shorter blood group A/B and O antigens.[7] Our new glycan microarray did detect any binding although compound **2** contains the same glycan structure. Our microarray data supports the notion that galectin 1 does not interact with type 2 blood group antigens.

In short, three distinct binding patterns were observed for the examined galectins: 1) Galectins 1 and 7 showed no binding to type 2 ABO antigens. 2) Galectins 3 and 9 bind to type 2 blood group A, B and O epitopes and can accept one α1,3-fucoside on a HexNAc scaffold. Both galectins lose binding when every GlcNAc is fucosylated on a polyLacNAc chain indicting that they interact with glycan regions that do not contain such residues. 3) Galectins 4 and 8 bind only to blood group A and B and not to blood group O epitopes of type 2 and cannot accept A/B Lewis^y as ligand. The binding of these galectins to blood group A and B appear to be mediated only through terminal epitopes.

Conclusions

An efficient chemo-enzymatic methodology has been developed that can control enzymatic fucosylation of polyLacNAc oligosaccharides. The methodology was used to synthesize a library of ABO blood group oligosaccharides carrying α1,3-fucoside on selected residues. This library was used to fabricate a glycan microarray platform that revealed that the effect of α1,3-fucosylation on galectin binding is dependent on the position and differs between class members. This platform can be used to gain insights in binding specificities of other blood group antigen binding proteins such as anti-ABO antibodies or viral/bacterial ABO attachment factors. Because of the modular nature of the synthetic methodology, we anticipate that it is compatible with more complex glycan topologies and diverse terminal epitopes. It is also expected that this method can be used to generate well-defined fucosylation patterns on polyLacNAc's conjugated to complex glycan scaffolds such as *N*-glycan, *O*-glycans and glycolipids.

References

[1] K. Watanabe, S. I. Hakomori, R. A. Childs, T. Feizi, J. Biol. Chem. 1979, 254, 3221–3228.

[2] M. Fukuda, S. R. Carlsson, J. C. Klock, A. Dell, J. Biol. Chem. 1986, 261, 12796–12806.

[3] M. Fukuda, B. Bothner, P. Ramsamooj, A. Dell, P. R. Tiller, A. Varki, J. C. Klock, J. Biol. Chem. 1985, 260, 12957–12967.

[4] M. Tiemeyer, S. J. Swiedler, M. Ishihara, M. Moreland, H. Schweingruber, P. Hirtzer, B. K. Brandley, Proc. Natl. Acad. Sci. U. S. A. 1991, 88, 1138–1142.

[5] P. Stanley, R. D. Cummings, in Essent. Glycobiol. (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, A.G. Darvill, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2015.

[6] K. Handa, D. A. Withers, S. Hakomori, Biochem. Biophys. Res. Commun. 1998, 243, 199–204.

[7] S. R. Stowell, C. M. Arthur, P. Mehta, K. A. Slanina, O. Blixt, H. Leffler, D. F. Smith, R. D. Cummings, J. Biol. Chem. 2008, 283, 10109–10123.

[8] M. Wuhrer, R. D. Dennis, M. J. Doenhoff, G. Lochnit, R. Geyer, Glycobiology 2000, 10, 89–101.

[9] M. L. Mickum, T. Rojsajjakul, Y. Yu, R. D. Cummings, Glycobiology 2016, 26, 270–285.

[10] M. A. Monteiro, K. H. Chan, D. A. Rasko, D. E. Taylor, P. Y. Zheng, B. J. Appelmelk, H. P. Wirth, M. Yang, M. J. Blaser, S. O. Hynes, A. P. Moran, M. B. Perry, J. Biol. Chem. 1998, 273, 11533–11543.

[11] M. Mally, C. Fontana, S. Leibundgut-Landmann, L. Laacisse, Y.-Y. Fan, G. Widmalm, M. Aebi, Mol. Microbiol. 2013, 87, 112–131.

[12] M. P. Bergman, A. Engering, H. H. Smits, S. J. van Vliet, A. A. van Bodegraven, H.-P. Wirth, M. L. Kapsenberg, C. M. J. E. Vandenbroucke-Grauls, Y. van Kooyk, B. J. Appelmelk, J. Exp. Med. 2004, 200, 979–990.

[13] J. Räbinä, J. Natunen, R. Niemelä, H. Salminen, K. Ilves, O. Aitio, H. Maaheimo, J. Helin, O. Renkonen, Carbohydr. Res. 1997, 305, 491–499.

[14] R. Phang, C.-H. Lin, Front. Immunol. 2022, 13, 858894.

[15] A. D. Srivastava, L. Unione, M. A. Wolfert, P. Valverde, A. Ardá, J. Jiménez-Barbero, G.-J. Boons, Chem. – Eur. J. 2020, 26, 15605–15612.

[16] U. Ellervik, G. Magnusson, J. Org. Chem. 1998, 63, 9314–9322.

[17] C. Chen, S. Wang, M. R. Gadi, H. Zhu, F. Liu, C.-C. Liu, L. Li, F. Wang, P. Ling, H. Cao, Chem. Commun. 2020, 56, 7549–7552.

[18] T. de Vries, C. A. Srnka, M. M. Palcic, S. J. Swiedler, D. H. van den Eijnden, B. A. Macher, J. Biol. Chem. 1995, 270, 8712–8722.

[19] T. De Vries, M. P. Palcic, P. S. Schoenmakers, D. H. Van Den Eijnden, D. H. Joziasse, Glycobiology 1997, 7, 921–927.

[20] M. Pykäri, S. Toivonen, J. Natunen, R. Niemelä, H. Salminen, O. Aitio, M. Ekström, P. Parmanne, M. Välimäki, J. Alais, C. Augé, J. B. Lowe, O. Renkonen,

R. Renkonen, J. Biol. Chem. 2000, 275, 40057–40063.
[21] S. Nishihara, H. Iwasaki, M. Kaneko, A. Tawada, M. Ito, H. Narimatsu, FEBS Lett. 1999, 462, 289–294.
[22] D. Vasiliu, N. Razi, Y. Zhang, N. Jacobsen, K. Allin, X. Liu, J. Hoffmann, O. Bohorov, O. Blixt, Carbohydr. Res. 2006, 341, 1447–1457.
[23] I. A. Gagarinov, T. Li, N. Wei, J. Sastre Toraño, R. P. de Vries, M. A. Wolfert, G.-J. Boons, Angew. Chem. Int. Ed. 2019, 58, 10547–10552.
[24] J. Ye, H. Xia, N. Sun, C.-C. Liu, A. Sheng, L. Chi, X.-W. Liu, G. Gu, S.-Q. Wang, J. Zhao, P. Wang, M. Xiao, F. Wang, H. Cao, Nat. Catal. 2019, 2, 514–522.
[25] L. Liu, A. R. Prudden, C. J. Capicciotti, G. P. Bosman, J.-Y. Yang, D. G. Chapla, K. W. Moremen, G.-J. Boons, Nat. Chem. 2019, 11, 161–169.
[26] Y. Li, M. Xue, X. Sheng, H. Yu, J. Zeng, V. Thon, Y. Chen, M. M. Muthana, P. G. Wang, X. Chen, Bioorg. Med. Chem. 2016, 24, 1696–1705.
[27] W. Peng, J. Pranskevich, C. Nycholat, M. Gilbert, W. Wakarchuk, J. C. Paulson, N. Razi, Glycobiology 2012, 22, 1453–1464.
[28] R. D. Larsen, L. K. Ernst, R. P. Nair, J. B. Lowe, Proc. Natl. Acad. Sci. 1990, 87, 6674–6678.
[29] P. Tumbale, K. Brew, J. Biol. Chem. 2009, 284, 25126–25134.
[30] F. Yamamoto, H. Clausen, T. White, J. Marken, S. Hakomori, Nature 1990, 345, 229–233.
[31] H. Leffler, S. Carlsson, M. Hedlund, Y. Qian, F. Poirier, Glycoconj. J. 2002, 19, 433–440.
[32] S. R. Stowell, M. Dias-Baruffi, L. Penttilä, O. Renkonen, A. K. Nyame, R. D. Cummings, Glycobiology 2004, 14, 157–167.
[33] Q. Zhou, R. D. Cummings, Arch. Biochem. Biophys. 1993, 300, 6–17.
[34] M. J. Moure, A. Gimeno, S. Delgado, T. Diercks, G.-J. Boons, J. Jiménez-Barbero, A. Ardá, Angew. Chem. Int. Ed. 2021, 60, 18777–18782.
[35] R.-Y. Yang, G. A. Rabinovich, F.-T. Liu, Expert Rev. Mol. Med. 2008, 10, e17.
[36] A. E. Szulman, J. Exp. Med. 1960, 111, 785–800.
[37] E. A. Gehrie, J. M. Cates, H. Nian, S. J. Olson, P. P. Young, Cardiovasc. Pathol. Off. J. Soc. Cardiovasc. Pathol. 2013, 22, 251–256.
[38] C. L. Berry, J. Amerigo, Virchows Arch. A Pathol. Anat. Histol. 1980, 388, 167–174.
[39] M. Gómez-Redondo, S. Delgado, R. Núñez-Franco, G. Jiménez-Osés, A. Ardá, J. Jiménez-Barbero, A. Gimeno, RSC Chem. Biol. n.d., 2, 932–941.
[40] A. Gimeno, S. Delgado, P. Valverde, S. Bertuzzi, M. A. Berbís, J. Echavarren, A. Lacetera, S. Martín-Santamaría, A. Surolia, F. J. Cañada, J. Jiménez-Barbero, A. Ardá, Angew. Chem. Int. Ed Engl. 2019, 58, 7268–7272.
[41] J. I. Quintana, S. Delgado, R. Núñez-Franco, F. J. Cañada, G. Jiménez-Osés, J. Jiménez-Barbero, A. Ardá, Front. Chem. 2021, 9.
[42] J. Hirabayashi, T. Hashidate, Y. Arata, N. Nishi, T. Nakamura, M. Hirashima, T. Urashima, T. Oka, M. Futai, W. E. G. Muller, F. Yagi, K. Kasai, Biochim. Biophys. Acta 2002, 1572, 232–254.
[43] R. O. Bua, A. Messina, L. Sturiale, R. Barone, D. Garozzo, A. Palmigiano, Int. J.

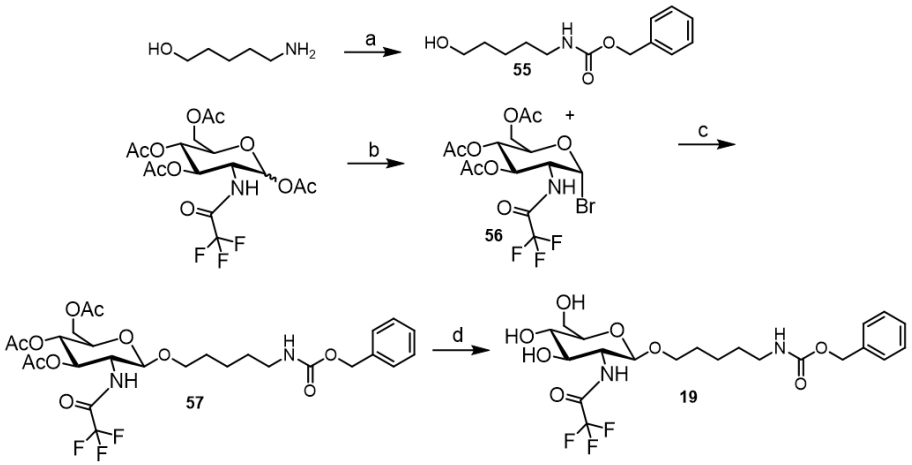
Mol. Sci. 2021, 22, 8063.
[44] Z. Wiczyńska, H. Miller-Podraza, J. Kościelak, FEBS Lett. 1980, 112, 277–279.
[45] M. Fukuda, A. Dell, J. E. Oates, M. N. Fukuda, J. Biol. Chem. 1984, 259, 8260–8273.
[46] V. L. Thijssen, S. Hulsmans, A. W. Griffioen, Am. J. Pathol. 2008, 172, 545–553.
[47] A. Ishikawa, T. Imaizumi, H. Yoshida, N. Nishi, T. Nakamura, M. Hirashima, K. Satoh, Immunol. Cell Biol. 2004, 82, 410–414.
[48] D. Ayona, P.-E. Fournier, B. Henrissat, B. Desnues, Front. Immunol. 2020, 11.
[49] T. Imaizumi, M. Kumagai, N. Sasaki, H. Kurotaki, F. Mori, M. Seki, N. Nishi, K. Fujimoto, K. Tanji, T. Shibata, W. Tamo, T. Matsumiya, H. Yoshida, X.-F. Cui, S. Takanashi, K. Hanada, K. Okumura, S. Yagihashi, K. Wakabayashi, T. Nakamura, M. Hirashima, K. Satoh, J. Leukoc. Biol. 2002, 72, 486–491.
[50] A. Chakraborty, C. Staudinger, S. L. King, F. C. Erickson, L. S. Lau, A. Bernasconi, F. W. Luscinskas, C. Perlyn, C. J. Dimitroff, J. Autoimmun. 2021, 117, 102575.
[51] M. La, T. V. Cao, G. Cerchiaro, K. Chilton, J. Hirabayashi, K. Kasai, S. M. Oliani, Y. Chernajovsky, M. Perretti, Am. J. Pathol. 2003, 163, 1505–1515.
[52] M. Jeyakanthan, K. Tao, L. Zou, P. J. Meloncelli, T. L. Lowary, K. Suzuki, D. Boland, I. Larsen, M. Burch, N. Shaw, K. Beddows, L. Addonizio, W. Zuckerman, B. Afzali, D. H. Kim, M. Mengel, A. M. J. Shapiro, L. J. West, Am. J. Transplant. 2015, 15, 2602–2615.
[53] V. Ravn, E. Dabelsteen, APMIS 2000, 108, 1–28.
[54] M. I. Nielsen, J. Stegmayr, O. C. Grant, Z. Yang, U. J. Nilsson, I. Boos, M. C. Carlsson, R. J. Woods, C. Unverzagt, H. Leffler, H. H. Wandall, J. Biol. Chem. 2018, 293, 20249–20262.
[55] M. Nagae, Y. Yamaguchi, Int. J. Mol. Sci. 2014, 15, 3768–3783.
[56] S.-C. Wu, A. D. Ho, N. A. Kamili, J. Wang, K. L. Murdock, R. D. Cummings, C. M. Arthur, S. R. Stowell, Front. Microbiol. 2021, 12.
[57] A. V. Blenda, N. A. Kamili, S.-C. Wu, W. F. Abel, D. Ayona, C. Gerner-Smidt, A. D. Ho, G. M. Benian, R. D. Cummings, C. M. Arthur, S. R. Stowell, J. Biol. Chem. 2022, 298.

Supplementary information

Materials and methods

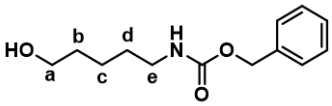
^{1}H NMR data is recorded on a 600 MHz Avance Neo NMR. Chemical shifts are reported in parts per million (ppm) relative to the residual solvent peak. NMR signals are represented as: chemical shift, multiplicity (s = singlet, d = doublet, app.t. = apparent triplet, t = triplet, dd = doublet of doublets, m = multiplet), J coupling and integration. Assignment of ^{1}H NMR signals was conducted based on ^{1}H-NMR, 2D COSY (COSYGPSW), 2D TOCSY (MLEVPHSW), 2D NOESY (NOESYPHSW) and 2D HSQC (HSQCEDETGPSISP). ^{13}C NMR signal assignment was extracted from 2D HSQC spectra. Mass spectra are recorded by ESI on a Bruker micrOTOF-QII system as m/z. Reaction progress for carbohydrates was monitored using a SeQuant Zic HILIC guard column (20x2.1 mm) coupled to a Bruker micrOTOF-QII MS system running a gradient of 90% Acetonitrile:H_2O to 50% Acetonitrile:$H_2$0 over 5 minutes followed by isocratic 50% Acetonitrile:H_2O until product was detected. MilliQ water (MQ) was purified by a Millipore Synergy water purification system. Size exclusion was conducted using Bio-Gel P2, P4 or P6 gel polyacrylamide beads from BioRad, eluding with MilliQ water or 20 mM ammonium bicarbonate for compounds containing a primary amine only when no TFA protecting group was present. Column chromatography was performed on silica gel G60. Thin Layer Chromatography (TLC) was conducted on Silicagel 60 F254 (EMD Chemicals Inc.) on prefabricated glass slides with visualization by UV and, spraying with 10% H_2SO_4 in ethanol and subsequent charring of the TLC plate. UDP-Galactose, UDP-*N*-Acetylglucosamine and UDP-*N*-Acetylgalactosamine were bought from Roche Diagnostics. Calf intestine alkaline phosphatase (CIAP) was obtained from Invitrogen. GDP-Fucose was prepared using L-fucokinase/GDP-fucose pyrophosphorylase.[1] Glycosyltransferases B3GNT2, Hpβ3GlcNAcT, B4GalT1, FUT1, FUT5 and FUT5 were expressed and purified according to published protocols.[2–4] Other chemical reagents were bought from Sigma-Aldrich or Carbosynth. Reagents purchased from commercial sources were used without further purification.

Synthesis of *N*-Benzyloxycarbonyl-5-amino-pentyl 2-trifluoroacetamido-2-deoxy-β-D-glucopyranoside



a) Benzylchloroformate(0.9 eq.), Et_3N(1 eq.), DCM, 0 °C; b) 33% HBr in AcOH, DCM; c) **55**(1.4 eq.), **56**(1 eq.) AgOTf(1.2 eq.), tetrazole(2 eq.), Tol, -78 °C to r.t.; d) NaOH pH 8, MeOH;

Benzyl *N*-(5-hydroxypentyl)carbamate (55)

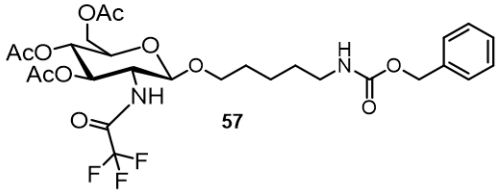


55

5-Aminopentinol (11.5 mmol, 1.18 g) was dissolved in DCM (117 mL) and Et_3N (11.5 mmol, 850 µL) was added under N_2 atmosphere. The solution was cooled to 0 ºC and benzyl chloroformate (10.4 mmol, 1.45 mL) was added dropwise. The reaction mixture was stirred for 16 h at room temperature. The reaction was cooled (0 ºC) and subsequently quenched with sat. aq. NH_4Cl solution (100 mL) when TLC analysis indicated complete consumption of benzyl chloroformate. DCM was removed under reduced atmosphere and the resulting aqueous phase was extracted with EtOAc (2x100 mL). Combined organic phase was washed with brine (100 mL), dried ($MgSO_4$) and filtrated. The filtrate was concentrated *in vacuo* to yield a clear oil. The residue was chromatographed on silica gel (EtOAc:DCM = 1:1) and fractions containing desired product were pooled and concentrated *in vacuo* (EtOAc:DCM = 1:1, R_f = 0.5, white crystalline wax, 1.9 g, 79% yield). ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.39–7.24 (m, 5H, Ar, Bn), 5.11 (broad s, 1H, NH) 5.07 (s, 2H, CH_2, Bn, 3.59 (app. t, *J* = 6.5 Hz, 2H, (CH_2 a)), 3.17 (dd, *J* = 6.7 Hz, 2H, (CH_2 e)), 1.58– 1.43 (m, 4H, (CH_2 b,d)), 1.40–1.31 (m, 2H, (CH_2 c)). ESI TOF-MS m/z calculated for $C_{26}H_{39}N_2O_6Na$ $[2M + H]^+$ = 475.2808, found 475.2825. In agreement with literature.[5]

2

***N*-Benzyloxycarbonyl-5-amino-pentyl 2-deoxy-2-trifluoroacetamido-3,4,6-tri-*O*-acetyl-β-D-glucopyranoside (57)**

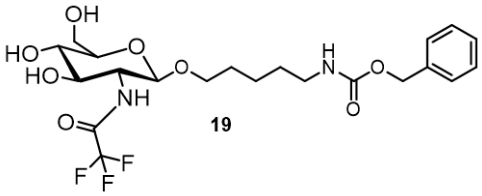


57

To a stirred mixture of (1:1 α:β) 1,3,4,6-Tetra-*O*-acetyl-2-trifluoroacetamido-2-deoxy-D-glucopyranose (5.73 mmol, 2.94 g) dissolved in DCM (20 mL) was added dropwise a solution of 33% HBr in AcOH (7 mL). The reaction mixture was stirred at room temperature for 4 h. The solution was diluted with DCM (40 mL) after TLC analysis indicated complete conversion of starting material. The resulting organic phase was washed with water (40 mL), sat.aq. bicarbonate (2x40 mL) and brine (40 mL). The organic phase was dried (MgSO4), filtrated and the filtrate was concentrated *in vacuo* (EtOAc:DCM = 1:1, R_f = 0.9, white foam, 2.66 g). This white foam was dissolved in dry toluene (30 mL) and Benzyl *N*-(5-hydroxypentyl) carbamate (8.0 mmol, 1.9 g) was added. Under argon atmosphere the mixture was cooled (-78 ºC) and AgOTf (6.9 mmol, 1.77 g) was added. The reaction was stirred and allowed to warm to room temperature. TLC analysis indicated complete consumption of starting material. The mixture was diluted with DCM (200 mL) and filtered. The filtrate was washed with water (200 mL), brine (100 mL) and dried ($MgSO_4$). After filtration, the filtrate was concentrated *in vacuo* (EtOAc:DCM = 1:2, R_f = 0.5, white foam, 3.07 g, 74% yield). ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.49 (d, *J* = 8.9 Hz, 1H, NHTFA), 7.38 – 7.23 (m, 5H, Ar, Bn), 5.37 – 5.28 (m, 1H, H-3),

5.14 – 5.02 (m, 3H, (CH_2, Bn), H-4), 4.94 (t, *J* = 6.1 Hz, 1H, NHC(=O)O), 4.66 (d, *J* = 8.5 Hz, 1H, H-1), 4.25 (dd, *J* = 12.4, 4.8 Hz, 1H, H-6a), 4.12 (dd, *J* = 12.2, 2.4 Hz, 1H, H-6b), 3.95 (dd, *J* = 9.2 Hz, 1H, H-2), 3.87 (dt, *J* = 9.7, 5.9 Hz, 2H, (CHH a, CHH d), 3.70 – 3.63 (m, 1H, H-5), 3.44 (dt, *J* = 9.9, 6.5 Hz, 2H, (CHH a, CHH b), 3.24 – 3.08 (m, 2H, (CH_2 e)), 2.11 – 1.99 (m, 9H, Ac), 1.61 – 1.53 (m, 1H, (CHH d)), 1.52 – 1.44 (m, 1H, (CHH b)), 1.42 – 1.28 (m, 2H, (CH_2 c)). ^{13}C NMR (150 MHz, $CDCl_3$) derived from HSQC δ 128.30, 100.38, 71.83, 71.64, 69.68, 68.54, 66.76, 61.94, 54.77, 41.00, 29.14, 28.57, 23.11, 20.52. ESI TOF-MS m/z calculated for $C_{27}H_{35}F_3N_2O_{11}Na$ $[M + Na]^+$ = 643.2091, found 643.1959.

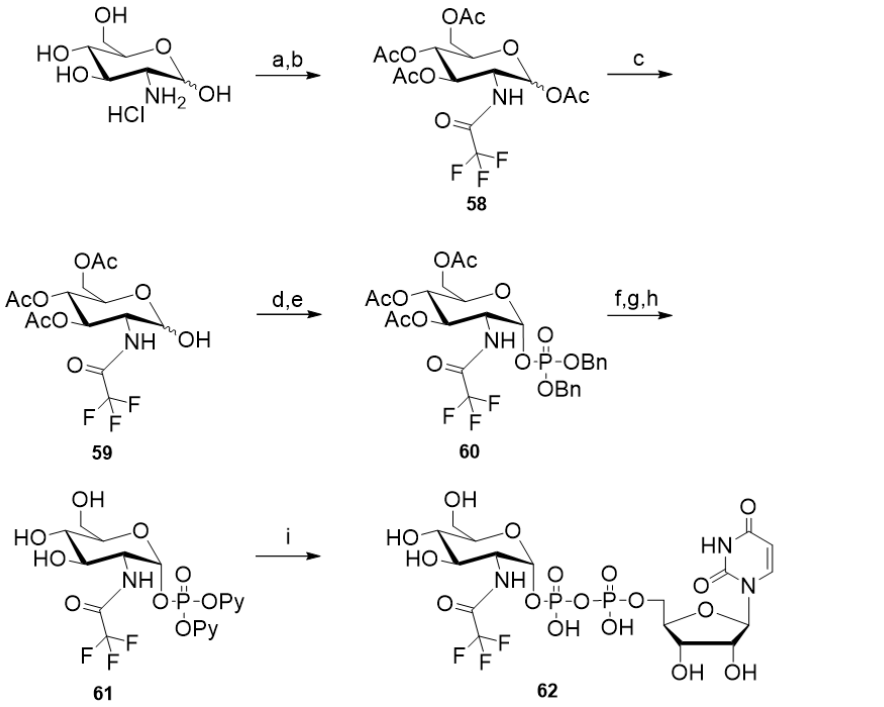
_N_-Benzyloxycarbonyl-5-amino-pentyl 2-trifluoroacetamido-2-deoxy-β-D-glucopyranoside (19)



19

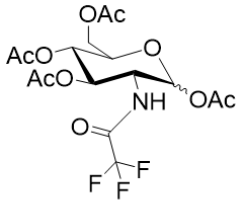
N-Benzyloxycarbonyl-5-amino-pentyl-2-deoxy-2-trifluoroacetamido-3,4,6-tri-O-acetyl-β-D-glucopyranoside (0.61 mmol, 380 mg) was dissolved in MeOH (10 mL which was basified to pH 8 with 1M NaOH(aq)). The resulting mixture was stirred at room temperature for 1 h. The reaction mixture was neutralized with Amberlyst resin H^+ form washed with MeOH, after TLC analysis indicated full conversion of the starting material. The resin was filtered, and the filtrate was concentrated *in vacuo* (white solid, 290 mg, 96% yield). ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.53 – 7.21 (m, 5H, Ar, Bn), 5.08 (s, 2H, CH_2, Bn), 4.57 (d, *J* = 8.6 Hz, 1H, H-1), 3.91 (d, *J* = 12.3 Hz, 1H, H-6a), 3.80 – 3.70 (m, 2H, H-2, H-6b), 3.64 – 3.58 (m, 2H, (CHH a, CHH d), 3.56 (d, *J* = 7.3 Hz, 2H, (CHH a, CHH b), 3.47 – 3.42 (m, 2H, H-3, H-5), 3.16 – 3.04 (m, 2H, (CH_2 e)), 1.59 – 1.39 (m, 4H, (CHH b, CHH d)), 1.37 – 1.18 (m, 2H, (CH_2 c)). ^{13}C NMR (150 MHz, $CDCl_3$) derived from HSQC δ 128.23, 100.33, 75.94, 73.18, 70.56, 69.81, 66.70, 61.72, 60.7, 56.61, 40.26, 28.46, 28.36, 22.26. ESI TOF-MS m/z calculated for $C_{21}H_{29}F_3N_2O_8Na$ $[M + Na]^+$ = 517.1774, found 517.1698.

Synthesis of UDP-GlcNTFA



a) TFA_2O(1.5 eq.), Et_3N(1.5 eq.), MeOH, 0 °C; b) Ac_2O, Py, 0 °C; c) $BnNH_2$(1.3 eq.), THF; d) $(iPr)_2NP(OBn)_2$(1.5 eq.), tetrazole(2 eq.), DCM, -78 °C to r.t.; e) mCPBA(5 eq.), DCM, -78 °C to r.t.; f) Pd/C, H_2, MeOH; g) NaOMe(cat), MeOH; h) Dowex 50WX-8 Py; i) UMP-morpholidate, tetrazole, $(Oc)_3N$, Py;

1-3-4-6-tetra-*O*-acetyl-2-*N*-trifluoroacetyl-D-glucosamine (58)

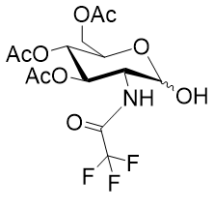


58

Glucosamine HCl (55.7 mmol, 10 g) was dissolved in MeOH (70 mL) and Et_3N (83.7 mmol, 11.65 mL) was added. The mixture was cooled (0 ºC) and placed under an Argon atmosphere. TFA_2O (83.7 mmol. 11.8 mL) was added dropwise and the reaction mixture was stirred for 16 h. The reaction mixture was concentrated *in vacuo* and co-evaporated from toluene (2x100 mL). The residue was dissolved in pyridine (60 mL) and cooled (0 ºC). Ac_2O (60 mL) was added dropwise and the solution was stirred for 4 h. The reaction mixture was concentrated *in vacuo* and subsequently diluted with 1 N HCl (aq, 100 mL) on ice. The aqueous phase was extracted with DCM (3x100 mL) and the combined organic phase was washed with H_2O (100 mL) and brine (100 mL), dried ($MgSO_4$) filtrated and the filtrate was concentrated *in vacuo*. The residue was chromatographed on silica gel (DCM:EtOAc = 2:1) and fractions

containing carbohydrate were pooled and concentrated *in vacuo* (EtOAc:MeOH:H2O = 7:2:1, R_f = 0.77, white foamy solid, 2.94 g, 12% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.81 (d, J = 8.8 Hz, 1H), 5.34 (dd, J = 10.8, 9.5,1H), 5.20 (m, 1H), 4.52 (dd, J = 10.6, 9.2 Hz, 1H), 4.24 (d, J = 12.6 Hz, 1H), 4.09 (d, J = 10.4 Hz, 1H), 4.05 (d, J = 12.4 Hz, 1H), 3.99 (m, 1H), 2.09 (s, 3H), 2.06 (s, 3H), 2.05 (s, 3H), 1.95 (s, 3H). NMR in agreement with literature.[4]

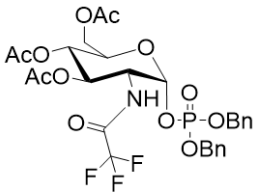
3-4-6-tri-*O*-acetyl-2-*N*-trifluoroacetyl-D-glucosamine (59)



59

1-3-4-6-tetra-*O*-acetyl-2-*N*-trifluoroacetyl-D-glucosamine (4.51 mmol, 2.0 g) was dissolved in THF (15 mL) and benzylamine (5.87 mmol, 677 µl) was added. The reaction mixture was stirred at room temperature for 2.5 h. TLC analysis indicated complete consumption of the starting material. The reaction mixture was concentrated *in vacuo* and diluted with DCM (50 mL). The organic phase was washed with 1 M HCl (aq, 3x50 mL), H_2O (50 mL) and brine (50 mL), dried ($NaSO_4$), filtrated and the filtrate was concentrated *in vacuo*. The residue was chromatographed on silica gel (EtOAc:petrol = 2:1) and fractions containing carbohydrate were pooled and concentrated *in vacuo* (EtOAc:petrol = 2:1, R_f = 0.4 and 0.5, white foam, 1.08 g, 46% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.65 (d, J = 9.3 Hz, 1H), 5.41-5.32 (m, 2H), 5.20-5.10 (m, 1H), 4.34-4,10 (m, 5H), 2.10 (s, 3H), 2.04 (s, 3H), 2.01 (s, 3H). NMR in agreement with literature.[5]

3-4-6-tri-*O*-acetyl-2-*N*-trifluoroacetyl-α-D-glucosamine-1-dibenzylphosphate (60)

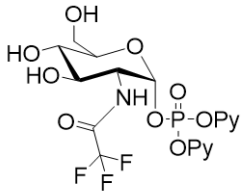


60

3-4-6-tri-*O*-acetyl-2-*N*-trifluoroacetyl-D-glucosamine (2.7 mmol, 1.08 g) and tetrazole (5.4 mmol, 378 mg) were dissolved in DCM (15 mL) under N_2 atmosphere. The stirred solution was cooled (– 78 ºC) and $(iPr)_2NP(OBn)_2$ (4.0 mmol, 1.35 mL) was added dropwise. The reaction mixture was allowed to warm to room temperature after 1 h. The reaction was again cooled (– 78 ºC) and mCPBA (13.5 mmol, 3.0 g) was added. The reaction was slowly allowed to warm to room temperature and washed with 10% Na_2SO_3 (aq. 2x100 mL), sat.aq. $NaHCO_3$

(2x100 mL), H_2O (100 mL) and brine (100 mL). The organic phase was dried ($NaSO_4$), filtrated and filtrate was concentrated *in vacuo*. The residue was chromatographed on silica gel (EtOAc:petrol = 4:6) and fractions containing carbohydrate were pooled and concentrated *in vacuo* (hexane:EtOAc = 1:1, R_f = 0.25, white foam, 1.29 g, 72% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.41-7.30 (m, 10H), 6.88 (d, J = 8.9, 1H), 6.06 (m, 1H), 5.70 (dd, 6.0, 3.3 Hz, 4H), 5.25 (m, 1H), 4.38-3.87 (m, 4H), 3.59-3.49 (m, 1H), 2.04 (s, 3H), 2.03 (s, 3H), 2.02 (s, 3H). NMR in agreement with literature.[6]

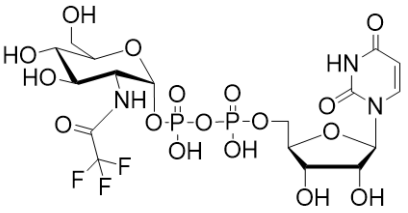
_N_-Trifluoroacetyl-α-D-glucosamine-1-phosphate dipyridinium salt (61)



61

3-4-6-tri-*O*-acetyl-2-*N*-trifluoroacetyl-α-D-glucosamine-1-dibenzylphosphate (1.95 mmol, 1.29 g) was dissolved in MeOH (20 mL) and 500 µl of sat.aq. $NaHCO_3$ was added followed by $Pd(OH)_2$/C (20% Degussa type). The reaction mixture was stirred in a 1 bar H_2 atmosphere overnight. TLC analysis indicated complete consumption of the starting material. The reaction mixture was filtrated through a Celite pad and the filtrate was concentrated *in vacuo*. The residue was dissolved in MeOH (20 mL) and a catalytic amount of 0.1N NaOMe solution in MeOH was added. The reaction mixture was quenched with Amberlyst 15 H^+ form resin to pH 6 after TLC analysis indicated completion of the reaction. The resin was removed by filtration and the filtrate was concentrated *in vacuo*. The mixture was re-dissolved in a minimal amount of MeOH and precipitated in icecold Et_2O. The resulting white solid was dissolved in H_2O and loaded on Dowex 50-WX-8 pyridinium form ion exchange resin. The product was eluded with H_2O and fractions containing carbohydrate were collected, pooled and concentrated (white powder, 500 mg, 50% yield). ^{1}H NMR (600 MHz, D_2O) δ 5.50 (dd, J = 6.1, 2.7 Hz, 1H), 4.05 (d, J = 10.6, 2.8 Hz, 1H), 3.94-3.76 (m, 4H), 3.59-3.49 (m, 1H). NMR in agreement with literature.[5]

Uridine diphosphate *N*-trifluoroacetylglucosamine (62)



62

N-Trifluoroacetyl-α-glucosamine-1-phosphate dipyridinium salt (0.98 mmol, 500 mg) was dissolved in pyridine (3 mL) and trioctylamine (1.96 mmol, 0.87 mL) was added. The mixture was then co-evaporated from pyridine (2x3 mL). UMP-morpholidate (1.56 mmol, 1.07 g) was added to the resulting oil and the mixture was co-evaporated again from pyridine (2x3 mL). The residue was diluted with 10 mL pyridine and subsequently tetrazole (3.91 mmol., 273 mg) was added. The reaction mixture was stirred at room temperature for 2.5 days. The mixture was concentrated *in vacuo* to an oil and diluted with H_2O (50 mL). The solution was washed with Et_2O (2x 50 mL) and concentrated *in vacuo*. The residue was loaded on a Dowex 1 Cl^- form resin and eluded with 2 L of 50 mM LiCl (aq) followed by 800 mL of 100 mM LiCl (aq) to remove impurities. Product was eluded by 500 mM LiCl (aq) and subsequently lyophilized. The resulting solid was dissolved in a minimal amount of H_2O and loaded on a Bio-Gel P2 Size exclusion column. Fractions containing carbohydrate were collected, pooled and lyophilized (247 mg, white crystalline powder, 25% yield). ^{1}H NMR (600 MHz, D_2O) δ 7.95 (d, J = 8.1 Hz, 1H), 5.96 (d, J = 4.5 Hz, 1H), 5.95 (d, *J* = 8.1 Hz, 1H), 5.59 (dd, *J* = 7.1, 3.3 Hz, 1H), 4.36 – 4.34 (m, 2H), 4.27 (m, 1H), 4.09 (dt, *J* = 10.6, 3.0, 2H), 3.97-3.77 (m, 4Hz), 3.57 (dd, J = 10.2, 9.1 Hz, 1H). NMR in agreement with literature.[5]

General procedure for the transfer of β1,3-*N*-trifluoroacetylglucosaminoside with Hpβ3GlcNAcT

UDP-*N*-trifluoroacetylglucosamine (1.5 eq.), CIAP (1 U/ µL) and Hpβ3GlcNAcT (10 µg/ µmol) were added to a solution of 10 mM of the appropriate acceptor in TRIS buffer (TRIS 100 mM, $MnCl_2$ 10 mM, pH 7.0). The reaction mixture was incubated for 16 h with gentle shaking at 37ºC. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was lyophilized when no more starting material was observed. The residue was re-dissolved in a minimal amount of MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

General procedure for the transfer of β1,4-galactoside with B4GalT1

UDP-galactose (1.5 eq.), CIAP (1 U/ µL) and B4GalT1 (10 µg/µmol) were added to a solution of 10 mM of the appropriate acceptor in TRIS buffer (TRIS 50 mM, BSA 0.1 wt%, $MnCl_2$ 10 mM, pH 7.0). The reaction mixture was incubated for 16 h with gentle shaking at 37ºC. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was lyophilized when no more starting material was observed. The residue was re-dissolved in a minimal amount of MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

General procedure for the transfer of α1,2-fucoside with FUT1

GDP-Fucose (1.5 eq.), CIAP (1 U/ µL) and FUT1 (10 µg/µmol) were added to a solution of 10 mM of the appropriate acceptor in TRIS buffer (TRIS 50 mM, $MnCl_2$ 10 mM, pH 7.0). The reaction mixture was incubated for 16 h with gentle shaking at 37ºC. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was lyophilized when no more starting material was observed. The residue was re-dissolved in a minimal amount of MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

General procedure for the transfer of α1,3-fucoside with FUT5

GDP-Fucose (0.9 eq.), CIAP (1 U/ µL) and FUT5 (10 µg/µmol) were added to a solution of 10 mM of the appropriate acceptor in TRIS buffer (TRIS 50 mM, $MnCl_2$ 10 mM, pH 7.0). The reaction mixture was incubated for 16 h with gentle shaking at 37ºC. Reaction progress was monitored by LC-ESI-Q-TOF-MS. Batches of 0.1 eq GDP-Fucose were added ever 2h when no GDP-fucose could be detected and starting material was still present. The reaction mixture was lyophilized when no more starting material was observed. The residue was re-dissolved in a minimal amount of MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

General procedure for the transfer of α1,3-fucoside with FUT9

GDP-Fucose (0.9 eq.), CIAP (1 U/ µL) and FUT9 (10 µg/µmol) were added to a solution of 10 mM of the appropriate acceptor in TRIS buffer (TRIS 50 mM, $MnCl_2$ 10 mM, pH 7.0). The reaction mixture was incubated for 3 h with gentle shaking at 37ºC. Reaction progress was monitored by LC-ESI-Q-TOF-MS. Batches of 0.1 eq GDP-Fucose were added every 1 h when no GDP-fucose could be detected and starting material was still present. The reaction mixture was lyophilized when no more starting material was observed. The residue was re-dissolved in a minimal amount of MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

General procedure for the transfer of α1,4-*N*-acetylgalactosaminoside with BoGT6a

UDP-*N*-acetylgalactosamine (1.5 eq.), CIAP (1 U/ µL) and BoGT6a (10 µg/µmol) were added to a solution of 10 mM of the appropriate acceptor in TRIS buffer (TRIS 50 mM, pH 7.0). The reaction mixture was incubated for 16 h with gentle shaking at 37ºC. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was lyophilized when no more starting material was observed. The residue was re-dissolved in a minimal amount of MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

General procedure for the transfer of α1,4-galactoside with GTB

UDP-galactose (1.5 eq.), CIAP (1 U/ µL) and GTB (50 µg/µmol) were added to a solution of 10 mM of the appropriate acceptor in TRIS buffer (TRIS 50 mM, pH 7.0). The reaction mixture was incubated for 16 h with gentle shaking at 37ºC. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was lyophilized when no more starting material was observed. The residue was re-dissolved in a minimal amount of MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

General procedure for the transfer of β1,3-*N*-acetylglucosaminoside with B3GNT2

UDP-*N*-acetylglucosamine (1.5 eq.), CIAP (1 U/ µL) and B3GNT2 (10 µg/µmol) were added to a solution of 10 mM of the appropriate acceptor in HEPES buffer (HEPES 50 mM, KCl 25 mM, $MgCl_2$ 2 mM, DTT 1 mM, pH 7.3). The reaction mixture was incubated for 16 h with gentle shaking at 37ºC. Reaction progress was monitored by LC-ESI-Q-TOF-MS as described above. The reaction mixture was lyophilized when no more starting material was observed. The residue was re-dissolved in a minimal amount of MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

General procedure for TFA deprotection

TFA protected oligosaccharide was dissolved in a solution of 25% NH_3 in water (100 µL/mg carbohydrate). The solution was shaken for 3-4 h at 37 ºC in a closed container. For α1,3-fucosylated GlcNTFA the reaction time was extended to 5 days. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction was evaporated under N_2 flow when no more starting material was observed. No further purification was applied.

General procedure for Boc protection

Oligosaccharide was dissolved in a 1:1 mixture of MeOH in MQ water to a concentration of 3.5 mM. The solution was basified using $NaHCO_3$ to pH 7 and heated to 50 ºC, subsequently 10 eq. of Boc_2O were added. The reaction was constantly sonicated at 50 ºC for 3h and progress was monitored by LC-ESI-Q-TOF-MS. An additional 10 eq. of Boc_2O and $NaHCO_3$ were added when starting material was still observed. MeOH was evaporated under N2 flow upon completion of the reaction. The reaction was diluted with MQ water and lyophilized. The resulting solid was redissolved in a minimal amount of MQ water and loaded on a P2 size exclusion column. Fractions containing carbohydrate were pooled and lyophilized to a fluffy white solid.

General procedure for Boc deprotection

Boc protected oligosaccharide was dissolved in a solution of 10% TFA in MQ water (100 µL/mg carbohydrate). The solution was shaken for 1 h at 37 ºC, reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was lyophilized when no more starting material was observed. The residue was redissolved in minimal MQ water, loaded on a P2 Biogel size exclusion column, and eluded with MQ water. Fractions containing carbohydrate were collected, pooled, and lyophilized to obtain a white fluffy solid.

Procedure for TFA protection on TFA and Boc protected oligosaccharide

Oligosaccharide was dissolved in 10 µL/mg MeOH and 2µL/mg of TEA and TFAOMe were added. The reaction was shaken at room temperature for 1 h. The crude reaction mixture was loaded on a Hypersep PGC SPE cartridge which was equilibrated with 5x1 mL of MQ water. The cartridge was washed with 1 mL of MQ water followed by 5% Acetonitrile in MQ water. Carbohydrate was eluded with 1 mL 1:1 Acetonitrile/MQ water. Acetonitrile was removed under N_2 flow, and the reaction was subsequently freeze dried. Product was used without further purification.

General procedure for *N*-acetylation

Oligosaccharide was dissolved in 1:1 MeOH/MQ water (40 µL/mg carbohydrate) and acetic anhydride (20 µL/mg carbohydrate for each acetylation) was added. The solution was sonicated for 1 h at 37 ºC, reaction progress was monitored by LC-ESI-Q-TOF-MS. If starting material was observed another 20 µL/mg carbohydrate of MQ water and Acetic anhydride were added. The reaction mixture was lyophilized when no more starting material was observed. The residue was re-dissolved in minimal MQ water and loaded on a P2 Biogel size exclusion column. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

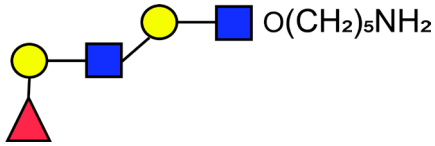
General procedure for Cbz deprotection

Oligosaccharide was dissolved in MQ water (100 µL/mg carbohydrate) and 10 wt% Pd/C was added. The reaction was placed in a hydrogen atmosphere and vigorously stirred overnight. Reaction progress was monitored by LC-ESI-Q-TOF-MS. If no more starting material was observed the reaction mixture is loaded on a Vivaspin 500, 10 kDa MWCO Polyethersulfone spinfilter and centrifuged for 3x30 min at 3000 RPM, washing with 300 µL MilliQ. The solution was lyophilized to obtain the solid product.

NMR nomenclature

Monosaccharides are labelled by their abbreviated name followed by a ranking number. N-Acetylglucosamine is abbreviated by "GlcNAc", glucosamine is abbreviated by "GlcN", galactose is abbreviated by "Gal", N-acetylgalactosamine is abbreviated by "GalNAc" and fucose is abbreviated by "Fuc". N-Acetylglucosamine derivatives are labelled by "GlcN" followed by the abbreviated name of the protecting group (e.g. GlcNBoc for N-tert-butyloxycarbonylglucosamine). Ranking of the monosaccharides is conducted from the reducing towards the non-reducing end. ^{1}H signals that are ambiguous have not been assigned and are reported as n/a. ^{1}H signals of the pentylamino linker are labelled from the non-glycosylated terminus towards the glycosylated terminus by a letter ranking from a-e. Carbon signals are reported for the C1 carbon. J-couplings and integrates are reported when no significant peak overlap is observed.

NMR and MS data glycan library



1

^{1}H (600 MHz, D_2O): δ (ppm)

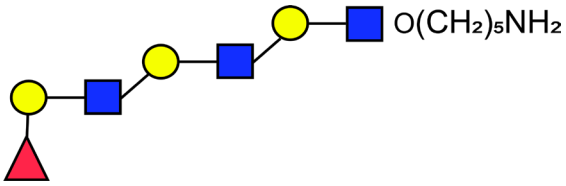
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.53 (d, *J* = 7.6 Hz, 1H)	3.72	3.59	n/a	n/a	n/a	2.03 (6H)
Gal	4.47 (d, *J* = 8.0 Hz, 1H)	3.59	3.71	4.16 (d, *J* = 3.4 Hz, 1H)	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.4 Hz, 1H)	3.81	3.47	3.97	n/a	n/a	2.03 (6H)
Gal (2)	4.56 (d, *J* = 7.8 Hz, 1H)	3.67	3.69	3.88	n/a	n/a	-
Fuc	5.32 (d, *J* = 3.1 Hz, 1H)	3.82	3.90	3.82	4.23 (q, *J* = 6.6 Hz, 1H)	1.23 (app. d, *J* = 6.6 Hz, 3H)	-

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.61 (1H), 3.91 (1H)	1.68 (2H)	1.40 (2H)	1.61 (2H)	2.99 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Gal (2)	Fuc
C1	100.99	102.85	102.73	100.22	99.34

ESI TOF-MS m/z calculated for $C_{39}H_{69}N_3O_{25}Na$ $[M + Na]^+$ = 1002.4118, found 1002.4144



2

^{1}H (600 MHz, D_2O): δ (ppm)

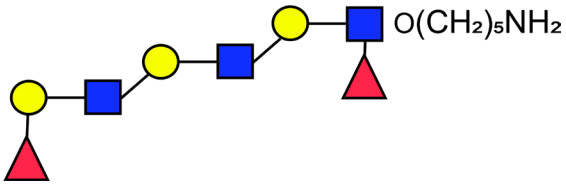
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.53 (d, *J* = 7.5 Hz, 1H)	3.70	3.59	n/a	n/a	n/a	2.04 (9H)
Gal	4.47 (d, *J* = 7.6 Hz, 2H)	3.59	3.72	4.16 (d, *J* = 3.5 Hz, 2H)	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.3 Hz, 2H)	3.81	3.59	3.72	n/a	n/a	2.04 (9H)
Gal (2)	4.47 (d, *J* = 7.6 Hz, 2H)	3.59	3.72	4.16 (d, *J* = 3.5 Hz, 2H)	n/a	n/a	-
GlcNAc (3)	4.71 (d, *J* = 8.3 Hz, 2H)	3.81	3.47	3.72	n/a	n/a	2.04 (9H)
Gal (3)	4.56 (d, *J* = 7.2 Hz, 1H)	3.68	3.71	3.88	n/a	n/a	-
Fuc	5.32 (d, *J* = 3.2 Hz, 1H)	3.81	n/a	3.82	4.23 (q, *J* = 6.7 Hz, 1H)	1.24 (d, *J* = 6.7 Hz, 3H)	-

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.60 (1H), 3.91 (1H)	1.64 (2H)	1.39 (2H)	1.59 (2H)	2.93 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Gal (2)	GlcNAc (3)	Gal (3)	Fuc
C1	100.97	102.86	102.69	102.86	102.69	100.30	99.31

ESI TOF-MS m/z calculated for $C_{53}H_{92}N_4O_{35}Na$ $[M + Na]^+$ =1367.5440, found 1367.5428



3

^{1}H (600 MHz, D_2O): δ (ppm)

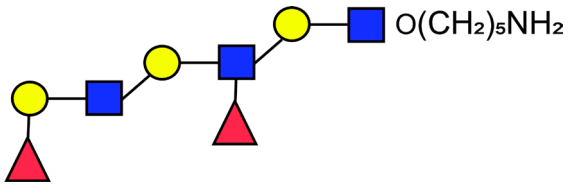
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.52 (d, *J* = 7.2 Hz, 1H))	3.86	3.56	3.90	n/a	n/a	2.03 (s, 9H)
Fuc	5.09 (d, *J* = 3.7 Hz, 1H)	3.69	3.78	3.90	4.80	1.14 (3H)	-
Gal	4.44 (d, *J* = 8.0 Hz, 1H)	3.52	3.71	4.10 (d, *J* = 3.4 Hz, 1H)	n/a	n/a	-
GlcNAc(2)	4.70 (d, *J* = 8.3 Hz, 1H)	3.81	3.48	3.72	n/a	n/a	2.03 (s, 9H
Gal (2)	4.48 (d, *J* = 7.8 Hz, 1H	3.60	3.71	4.15 (d, *J* = 3.4 Hz, 1H)	n/a	n/a	-
GlcNAc (3)	4.70 (d, *J* = 8.3 Hz, 1H)	3.81	3.58	3.72	n/a	n/a	2.03 (s, 9H
Gal (3)	4.55 (d, *J* = 7.6 Hz, 1H)	3.68	3.65	3.88	n/a	n/a	-
Fuc (2)	5.31 (d, *J* = 3.0 Hz, 1H)	3.81	n/a	n/a	4.24	1.22 (3H)	-

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.59 (1H), 3.89 (1H)	1.60 (2H)	1.39 (2H)	1.68 (2H)	2.98 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Fuc	Gal	GlcNAc(2)	Gal (2)	GlcNAc (3)	Gal (3)	Fuc (2)
C1	100.95	98.89	102.02	102.78	103.14	102.78	100.61	99.65

ESI TOF-MS m/z calculated for $C_{59}H_{102}N_4O_{39}Na$ $[M + Na]^+$ = 1513.6019. found 1513.6028.



4

^{1}H (600 MHz, D_2O): δ (ppm)

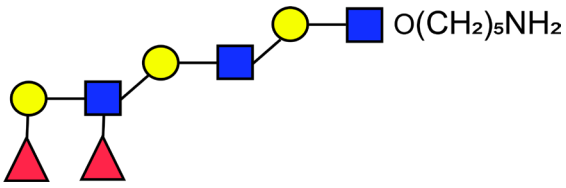
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.70	3.47	3.68	3.82	n/a	n/a	1.95 (9H)
Gal	4.46 (d, *J* = 7.7 Hz, 2H)	3.54	3.74	4.16	n/a	n/a	-
GlcNAc (2)	4.70	3.88	3.69	n/a	n/a	n/a	1.95 (9H)
Fuc	5.12	3.71	3.90	3.77	4.11	1.32 (3H)	-
Gal (2)	4.46 (d, *J* = 7.7 Hz, 2H)	3.54	3.74	4.16	n/a	n/a	-
GlcNAc (3)	4.70	3.47	3.68	3.82	n/a	n/a	1.95 (9H)
Gal (3)	4.46 (d, *J* = 7.7 Hz, 2H)	3.54	3.69	3.84	n/a	n/a	-
Fuc (2)	5.32	3.82	n/a	n/a	4.22	1.22 (3H)	-

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.58 (1H), 3.89 (1H)	1.60 (2H)	1.40 (2H)	1.68 (2H)	2.99 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Fuc	Gal (2)	GlcNAc (3)	Gal (3)	Fuc (2)
C1	102.66	102.72	100.98	99.47	102.72	102.66	102.72	99.58

ESI TOF-MS m/z calculated for $C_{59}H_{102}N_4O_{39}Na$ $[M + Na]^+$ = 1513.6019, found 1513.6042.



5

^{1}H (600 MHz, D_2O): δ (ppm)

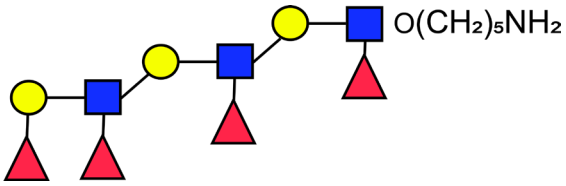
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc (1)	4.52 (d, *J* = 7.8 Hz, 1H)	3.71	3.59	n/a	n/a	n/a	2.12 – 1.96 (m, 9H)
Gal	4.47 (d, *J* = 6.9 Hz, 1H)	3.58	3.71	4.16	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.3 Hz, 1H))	3.81	3.59	3.72	n/a	n/a	2.12 – 1.96 (m, 9H)
Gal (2)	4.47 (d, *J* = 6.9 Hz, 1H))	3.58	3.71	4.16	n/a	n/a	-
GlcNAc (3)	4.70 (d, *J* = 8.3 Hz, 1H)	3.96	3.47	3.86	n/a	n/a	2.12 – 1.96 (m, 9H)
Fuc	5.12 (d, *J* = 4.0 Hz, 1H))	3.70	3.91	3.81	4.87	1.24 (6H)	
Gal (3)	4.52 (d, *J* = 8.3 Hz, 1H)	3.67	3.64	3.86	n/a	n/a	-
Fuc (2)	5.28 (d, *J* = 3.5 Hz, 1H)	3.79	n/a	n/a	4.87	1.24 (6H)	-

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.60 (1H), 3.91 (1H)	1.68 (2H)	1.40 (2H)	1.60 (2H)	2.98 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc (1)	Gal (1)	GlcNAc (2)	Gal (2)	GlcNAc (3)	Fuc	Gal (3)	Fuc (2)
C1	100.40	102.61	102.38	102.61	102.38	98.34	100.25	99.06

ESI TOF-MS m/z calculated for $C_{59}H_{102}N_4O_{39}Na$ $[M + Na]^+$ = 1513.6019, found 1513.6098.



6

^{1}H (600 MHz, D_2O): δ (ppm)

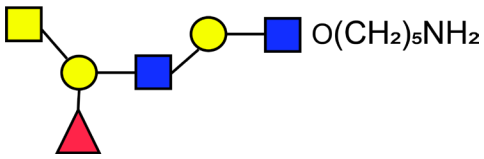
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.53 (d, *J* = 7.7 Hz, 1H)	3.87	3.58	n/a	n/a	n/a	2.00 (3H)
Fuc	5.10 (d, *J* = 4.1 Hz, 1H)	3.68	3.89	3.78	4.81	1.15 (d, *J* = 6.5 Hz, 6H)	-
Gal	4.45 (d, *J* = 9.7 Hz, 1H)	3.52	3.69	4.10 (d, *J* = 3.8 Hz, 2H)	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.2 Hz, 2H)	3.96	3.57	3.85	n/a	n/a	2.04 (6H)
Fuc (2)	5.12 (d, *J* = 4.0 Hz, 2H)	3.69	3.90	3.78	4.81	1.15 (d, *J* = 6.5 Hz, 6H)	
Gal (2)	4.44 (d, *J* = 8.5 Hz, 1H)	3.51	3.70	4.10 (d, *J* = 3.8 Hz, 2H)	n/a	n/a	-
GlcNAc (3)	4.71 (d, *J* = 8.2 Hz, 2H)	3.96	3.45	3.85	n/a	n/a	2.04 (6H)
Fuc (3)	5.12 (d, *J* = 4.0 Hz, 2H)	3.69	3.90	3.83	4.25	1.27 (d, *J* = 6.6 Hz, 3H)	-
Gal (3)	4.52 (d, *J* = 7.3 Hz, 1H)	3.65	3.85	3.85	n/a	n/a	-
Fuc (4)	5.28 (d, *J* = 3.5 Hz, 1H)	3.79	n/a	3.81	4.88	1.24 (d, *J* = 6.6 Hz, 3H)	-

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.59 (1H), 3.90 (1H)	1.67 (2H)	1.39 (2H)	1.59 (2H)	2.99 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Fuc	Gal	GlcNAc (2)	Fuc (2)	Gal (2)	GlcNAc (3)	Fuc (3)	Gal (3)	Fuc (4)
C1	100.86	98.56	101.65	102.40	98.56	101.65	102.40	98.56	100.41	99.31

ESI TOF-MS m/z calculated for $C_{71}H_{122}N_4O_{47}Na$ $[M + Na]^+$ = 1805.7358, found 1805.7261



7

^{1}H (600 MHz, D_2O): δ (ppm)

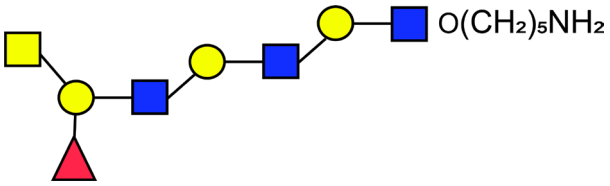
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.53 (d, *J* = 7.6 Hz, 1H)	3.71	3.60	n/a	n/a	n/a	2.05 (9H)
Gal	4.47 (d, *J* = 7.8 Hz, 1H)	3.59	3.71	4.16 (d, *J* = 3.3 Hz, 1H)	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.4 Hz, 1H)	3.81	3.45	3.69	n/a	n/a	2.05 (9H)
Gal (2)	4.62 (d, *J* = 7.7 Hz, 1H)	3.91	3.99	4.26 (d, *J* = 3.6 Hz, 1H)	n/a	n/a	-
Fuc	5.36 (d, *J* = 4.1 Hz, 1H)	3.79	3.73	3.84	4.34	1.26 (d, *J* = 6.5 Hz, 3H)	-
GalNAc	5.19 (d, *J* = 3.8 Hz, 1H)	4.24	3.92	4.00	n/a	n/a	2.05 (9H)

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.61 (1H), 3.91 (1H)	1.67 (2H)	1.41 (2H)	1.61 (2H)	2.99 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Gal (2)	Fuc	GalNAc
C1	101.15	102.96	102.81	100.02	98.73	91.35

ESI TOF-MS m/z calculated for $C_{47}H_{82}N_4O_{30}Na$ $[M + Na]^+$ = 1205.4912, found 1205.4895



8

^{1}H (600 MHz, D_2O): δ (ppm)

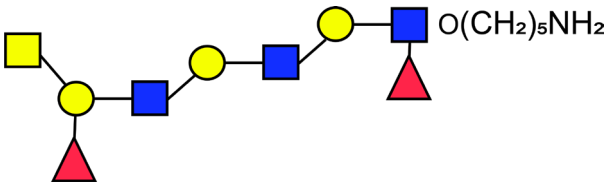
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.53 (d, *J* = 7.7 Hz, 1H)	3.72	3.59	n/a	n/a	n/a	2.06 (12H)
Gal	4.47 (d, *J* = 7.6 Hz, 2H)	3.59	3.71	4.15 (d, *J* = 3.5 Hz, 2H)	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.4 Hz, 2H)	3.81	3.58	3.71	n/a	n/a	2.06 (12H)
Gal (2)	4.47 (d, *J* = 7.6 Hz, 2H)	3.59	3.71	4.15 (d, *J* = 3.5 Hz, 2H)	n/a	n/a	-
GlcNAc (3)	4.71 (d, *J* = 8.4 Hz, 2H)	3.81	3.45	3.71	n/a	n/a	2.06 (12H)
Gal (3)	4.62 (d, *J* = 7.6 Hz, 1H)	3.89	3.99	4.23	n/a	n/a	-
Fuc	5.36 (d, *J* = 4.1 Hz, 1H)	3.78	3.72	3.83	4.32 (q, *J* = 6.6 Hz, 1H)	1.26 (d, *J* = 6.6 Hz, 3H)	-
GalNAc	5.19 (d, *J* = 3.9 Hz, 1H)	4.23	3.90	4.00	n/a	n/a	2.06 (12H)

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.61 (1H), 3.89 (1H)	1.67 (2H)	1.40 (2H)	1.59 (2H)	2.99 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Gal (2)	GlcNAc (3)	Gal (3)	Fuc	GalNAc
C1	100.89	102.86	102.78	102.86	102.78	100.03	98.69	91.14

ESI TOF-MS m/z calculated for $(C_{61}H_{106}N_5O_{40}Na)/2$ $[M + Na + H]^{2+}/2$ = 785.8156, found 785.8055



9

^{1}H (600 MHz, D_2O): δ (ppm)

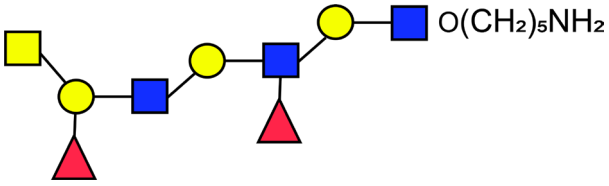
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.53 (d, *J* = 9.0 Hz, 1H)	3.80	3.87	3.94	3.58	n/a	2.04 (12H)
Fuc	5.10 (d, *J* = 4.2 Hz, 1H)	3.68	3.89	3.79	4.81	1.16 (app.d, *J* = 6.6 Hz, 3H)	-
Gal	4.44 (d, *J* = 7.9 Hz, 1H)	3.52	3.71	4.10 (d, *J* = 2.7 Hz, 1H)	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.7 Hz, 2H)	3.81	3.45	3.72	n/a	n/a	2.04 (12H)
Gal (2)	4.47 (d, *J* = 7.5 Hz, 1H)	3.60	3.72	4.15 (d, *J* = 2.5 Hz, 1H)	n/a	n/a	-
GlcNAc (3)	4.70 (d, *J* = 8.7 Hz, 2H)	3.81	3.59	3.72	n/a	n/a	2.04 (12H)
Gal (3)	4.61 (d, *J* = 7.5 Hz, 1H)	3.91	4.00	4.23	n/a	n/a	-
Fuc (2)	5.36 (d, *J* = 4.1 Hz, 1H)	3.82	3.72	n/a	4.32	1.26 (app.d, *J* = 6.7 Hz, 3H)	-
GalNAc	5.18 (d, *J* = 3.8 Hz, 1H)	4.22	3.91	4.01	n/a	n/a	2.04 (12H)

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.60 (1H), 3.91 (1H)	1.59 (2H)	1.39 (2H)	1.67 (2H)	2.98 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Fuc	Gal	GlcNAc (2)	Gal (2)	GlcNAc (3)	Gal (3)	Fuc (2)	GalNAc
C1	100.86	98.52	101.67	102.59	102.52	102.59	99.89	98.49	91.16

ESI TOF-MS m/z calculated for $C_{67}H_{115}N_5O_{44}Na$ $[M + Na]^+$ = 1706.6813, found 1706.6783.



10

^{1}H (600 MHz, D_2O): δ (ppm)

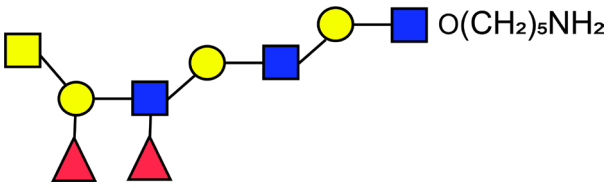
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.51 (d, *J* = 6.4 Hz, 1H)	3.73	3.58	3.86	n/a	n/a	2.04 (12H)
Gal	4.46 (d, *J* = 7.8 Hz, 2H)	3.55	3.72	4.12	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.0 Hz, 1H)	3.94	3.82	n/a	n/a	n/a	2.04 (12H)
Fuc	5.12	3.70	3.89	3.78	4.32	1.36 (3H)	-
Gal (2)	4.46 (d, *J* = 7.8 Hz, 2H)	3.55	3.72	4.12	n/a	n/a	-
GlcNAc (3)	4.71 (d, *J* = 6.9 Hz, 1H)	3.83	3.44	3.71	n/a	n/a	2.04 (12H)
Gal (3)	4.61 (d, *J* = 7.7 Hz, 1H)	3.90	4.00	4.24	n/a	n/a	-
Fuc (2)	5.36 (d, *J* = 4.2 Hz, 1H)	3.81	3.72	n/a	4.32	1.24 (3H)	-
GalNAc	5.18 (d, *J* = 3.8 Hz, 1H)	4.25	3.91	4.00	n/a	n/a	2.04 (12H)

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.59 (1H), 3.89 (1H)	1.60 (2H)	1.40 (2H)	1.66 (2H)	2.98 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Fuc	Gal (2)	GlcNAc (3)	Gal (3)	Fuc (2)	GalNAc
C1	100.87	102.55	102.79	98.66	102.55	102.79	100.09	98.59	91.13

ESI TOF-MS m/z calculated for $C_{67}H_{115}N_5O_{44}Na$ $[M + Na]^+$ = 1716.6813, found 1716.6725



11

^{1}H (600 MHz, D_2O): δ (ppm)

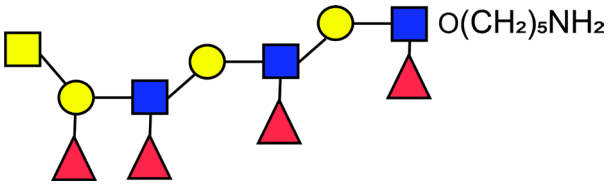
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc (1)	4.52 (d, *J* = 7.7 Hz, 1H)	3.72	3.58	n/a	n/a	n/a	2.04 (12H)
Gal	4.47 (d, *J* = 7.1 Hz, 2H)	3.59	3.71	4.16 (d, *J* = 4.1 Hz, 2H)	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.2 Hz, 2H)	3.81	3.58	3.73	n/a	n/a	2.04 (12H)
Gal (2)	4.47 (d, *J* = 7.1 Hz, 2H)	3.59	3.71	4.16 (d, *J* = 4.1 Hz, 2H)	n/a	n/a	-
GlcNAc (3)	4.70 (d, *J* = 8.2 Hz, 2H)	3.81	3.43	3.73	n/a	n/a	2.04 (12H)
Fuc	5.14 (d, *J* = 4.0 Hz, 1H)	3.70	3.92	3.79	4.88	1.29 (dd, *J* = 13.6, 6.5 Hz, 3H)	
Gal (3)	4.58 (d, *J* = 7.5 Hz, 1H)	3.87	3.95	4.21	n/a	n/a	-
Fuc (2)	5.31 (d, *J* = 4.2 Hz, 1H)	3.79	3.71	3.85	4.34 (q, *J* = 6.6 Hz, 1H)	1.33 (d, *J* = 7.0 Hz, 3H)	-
GalNAc	5.20 (d, *J* = 3.8 Hz, 1H)	4.25	3.93	4.00	n/a	n/a	2.04 (12H)

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.61 (1H), 3.92 (1H)	1.59 (2H)	1.40 (2H)	1.67 (2H)	2.98 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc (1)	Gal (1)	GlcNAc (2)	Gal (2)	GlcNAc (3)	Fuc	Gal (3)	Fuc (2)	GalNAc
C1	100.96	102.91	102.72	102.91	102.72	98.36	100.33	98.84	91.21

ESI TOF-MS m/z calculated for $C_{67}H_{115}N_5O_{44}Na$ $[M + Na]^+$ = 1716.6813, found 1716.6898.



12

^{1}H (600 MHz, D_2O): δ (ppm)

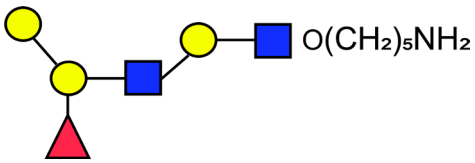
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.53 (d, *J* = 8.1 Hz, 1H)	3.89	3.58	3.92	n/a	n/a	2.04 (12H)
Fuc	5.10 (d, *J* = 3.9 Hz, 1H)	3.70	3.89	3.79	4.81	1.16 (d, *J* = 6.5 Hz, 6H)	-
Gal	4.46 (d, *J* = 7.8 Hz, 2H)	3.52	3.69	4.10	n/a	n/a	-
GlcNAc (2)	4.72 (d, *J* = 8.6 Hz, 2H)	3.96	3.58	3.87	n/a	n/a	2.04 (12H)
Fuc (2)	5.13 (d, *J* = 4.0 Hz, 1H)	3.70	3.89	3.79	4.81	1.16 (d, *J* = 6.5 Hz, 6H)	
Gal (2)	4.46 (d, *J* = 7.8 Hz, 2H)	3.52	3.69	4.10	n/a	n/a	-
GlcNAc (3)	4.72 (d, *J* = 8.6 Hz, 2H))	3.96	3.45	3.87	n/a	n/a	2.04 (12H)
Fuc (3)	5.15 (d, *J* = 3.9 Hz, 1H)	3.70	3.89	3.79	4.34	1.31 (d, *J* = 6.5 Hz, 3H)	-
Gal (3)	4.58 (d, *J* = 7.6 Hz, 1H)	3.89	3.95	4.20	n/a	n/a	-
Fuc (4)	5.31 (d, *J* = 4.1 Hz, 1H)	3.79	3.71	3.81	4.87	1.29 (d, *J* = 6.4 Hz, 3H)	-
GalNAc	5.20 (d, *J* = 3.8 Hz, 1H)	4.26	3.95	4.00	n/a	n/a	2.04 (12H)

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.56 (1H), 3.92 (1H)	1.68 (2H)	1.40 (2H)	1.59 (2H)	2.99 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Fuc	Gal	GlcNAc (2)	Fuc (2)	Gal (2)	GlcNAc (3)	Fuc (3)	Gal (3)	Fuc (4)	GalNAc
C1	101.02	98.66	101.73	102.19	98.66	101.73	102.19	98.46	100.40	98.87	91.19

ESI TOF-MS m/z calculated for $C_{79}H_{135}N_5O_{52}Na$ $[M + Na]^+$ = 2008.7971, found 2008.8035



13

^{1}H (600 MHz, D_2O): δ (ppm)

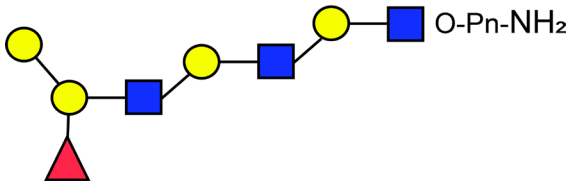
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.53 (d, *J* = 7.5 Hz, 1H)	3.71	3.59	n/a	n/a	n/a	2.05 (6H)
Gal	4.47 (d, *J* = 7.8 Hz, 1H)	3.59	3.71	4.16 (d, *J* = 3.2 Hz, 1H)	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.4 Hz, 1H)	3.82	3.45	3.70	n/a	n/a	2.05 (6H)
Gal (2)	4.64 (d, *J* = 7.7 Hz, 1H)	3.92	4.01	4.30 (d, *J* = 2.9 Hz, 1H)	n/a	n/a	-
Fuc	5.34 (d, *J* = 4.2 Hz, 1H)	3.81	3.72	3.86	4.30	1.25 (d, *J* = 6.6 Hz, 3H)	-
Gal(3)	5.25 (d, *J* = 2.5 Hz, 1H)	3.90	3.99	4.20	n/a	n/a	-

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.60 (1H), 3.91 (1H)	1.67 (2H)	1.40 (2H)	1.59 (2H)	2.99 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Gal (2)	Fuc	Gal (3)
C1	101.16	102.93	102.82	100.14	98.66	93.03

ESI TOF-MS m/z calculated for $C_{45}H_{79}N_3O_{30}Na$ $[M + Na]^+$ = 1164.4646, found 1164.4850



14

^{1}H (600 MHz, D_2O): δ (ppm)

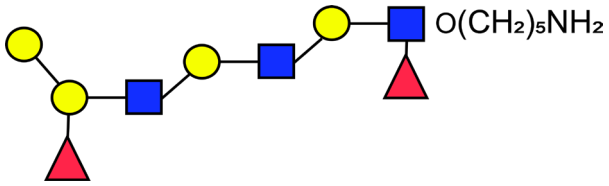
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.53 (d, *J* = 7.6 Hz, 1H)	3.72	3.59	n/a	n/a	n/a	2.05 (9H)
Gal	4.47 (d, *J* = 7.6 Hz, 2H)	3.59	3.71	4.16 (d, *J* = 3.5 Hz, 2H)	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.4 Hz, 2H)	3.81	3.58	3.71	n/a	n/a	2.05 (9H)
Gal (2)	4.47 (d, *J* = 7.6 Hz, 2H)	3.59	3.72	4.16 (d, *J* = 3.4 Hz, 2H)	n/a	n/a	-
GlcNAc (3)	4.71 (d, *J* = 8.4 Hz, 2H)	3.81	3.44	3.71	n/a	n/a	2.05 (9H))
Gal (3)	4.63 (d, *J* = 7.7 Hz, 1H)	3.93	3.99	4.28	n/a	n/a	-
Fuc	5.34 (d, *J* = 4.1 Hz, 1H)	3.79	3.71	3.82	4.31 (q, *J* = 7.0 Hz, 1H)	1.25 (d, *J* = 6.5 Hz, 3H)	-
Gal (4)	5.25 (d, *J* = 2.7 Hz, 1H)	3.88	3.99	4.20	n/a	n/a	-

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.61 (1H), 3.89 (1H)	1.68 (2H)	1.40 (2H)	1.60 (2H)	2.98 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Gal (2)	GlcNAc (3)	Gal (3)	Fuc	Gal (4)
C1	100.98	102.83	102.65	102.83	102.65	100.06	98.64	92.89

ESI TOF-MS m/z calculated for $C_{59}H_{102}N_4O_{40}Na$ $[M + Na]^+$ = 1529.5968, found 1529.5941



15

^{1}H (600 MHz, D_2O): δ (ppm)

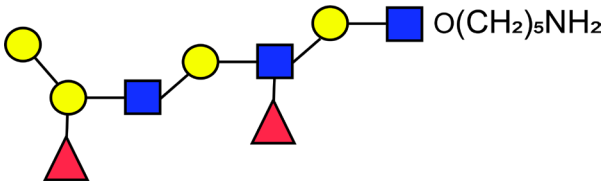
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.53 (d, *J* = 8.1 Hz, 1H)	3.87	3.58	3.89	n/a	n/a	2.04 (12H)
Fuc	5.10 (d, *J* = 4.1 Hz, 1H)	3.68	3.89	3.79	4.80	1.16 (d, *J* = 6.5 Hz, 3H)	-
Gal	4.44 (d, *J* = 8.0 Hz, 1H)	3.52	3.71	4.10 (d, *J* = 2.7 Hz, 1H)	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.1 Hz, 2H)	3.81	3.45	3.71	n/a	n/a	2.04 (12H)
Gal (2)	4.48 (d, *J* = 7.8 Hz, 1H)	3.60	3.72	4.15 (d, *J* = 3.3 Hz, 1H)	n/a	n/a	-
GlcNAc (3)	4.70 (d, *J* = 8.1 Hz, 2H)	3.81	3.59	3.71	n/a	n/a	2.04 (12H)
Gal (3)	4.63 (d, *J* = 7.5 Hz, 1H)	3.92	4.00	4.29	n/a	n/a	-
Fuc (2)	5.33 (d, *J* = 4.1 Hz, 1H)	3.81	3.71	n/a	4.30	1.24 (d, *J* = 6.6 Hz, 3H)	-
Gal (4)	5.29 (d, *J* = 3.0 Hz, 1H)	3.88	3.91	3.98	n/a	n/a	2.04 (12H)

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.60 (1H), 3.91 (1H)	1.58 (2H)	1.38 (2H)	1.63 (2H)	2.93 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Fuc	Gal	GlcNAc (2)	Gal (2)	GlcNAc (3)	Gal (3)	Fuc (2)	Gal (4)
C1	100.74	98.54	101.88	102.75	102.58	102.75	99.90	98.70	93.09

ESI TOF-MS m/z calculated for $C_{65}H_{112}N_4O_{44}Na$ $[M + Na]^+$ = 1675.6547, found 1675.6488



16

^{1}H (600 MHz, D_2O): δ (ppm)

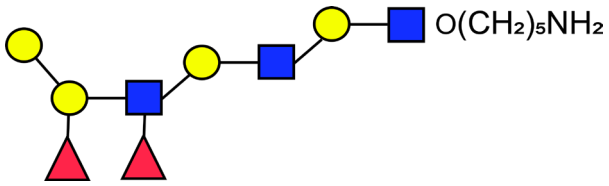
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.53 (d, *J* = 7.6 Hz, 1H)	3.69	3.57	3.83	n/a	n/a	2.05 (9H)
Gal	4.46 (d, *J* = 7.9 Hz, 2H)	3.56	3.72	4.12	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.3 Hz, 2H)	3.96	3.86	3.59	n/a	n/a	2.05 (9H)
Fuc	5.13 (d, *J* = 3.5 Hz, 1H)	3.69	3.90	3.78	4.32	1.16(3H)	-
Gal (2)	4.46 (d, *J* = 7.9 Hz, 2H)	3.56	3.72	4.12	n/a	n/a	-
GlcNAc (3)	4.70 (d, *J* = 8.3 Hz, 2H)	3.81	3.44	3.70	n/a	n/a	2.05 (9H)
Gal (3)	4.63 (d, *J* = 7.6 Hz, 1H)	3.94	4.01	4.29	n/a	n/a	-
Fuc (2)	5.34 (d, *J* = 4.2 Hz, 1H)	3.80	3.72	3.83	4.32	1.24 (3H)	-
Gal (4)	5.25 (d, *J* = 2.8 Hz, 1H)	3.89	3.99	4.20	n/a	n/a	-

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.59 (1H), 3.89 (1H)	1.60 (2H)	1.40 (2H)	1.67 (2H)	2.99 (2H)

from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Fuc	Gal (2)	GlcNAc (3)	Gal (3)	Fuc (2)	Gal (4)
C1	101.08	102.17	102.65	98.86	102.17	102.65	100.04	98.75	92.98

ESI TOF-MS m/z calculated for $C_{65}H_{112}N_4O_{44}Na$ $[M + Na]^+$ = 1675.6547, found 1675.6602



17

^{1}H (600 MHz, D_2O): δ (ppm)

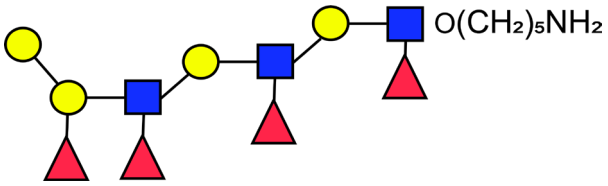
	H1	H2	H3	H4	H5	H6	NAc
GlcNAc (1)	4.53 (d, *J* = 7.5 Hz, 1H)	3.74	3.58	n/a	n/a	n/a	2.04 (9H)
Gal	4.47 (d, *J* = 7.5 Hz, 2H)	3.60	3.71	4.16 (d, *J* = 4.1 Hz, 2H)	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.2 Hz, 2H)	3.79	3.59	3.72	n/a	n/a	2.04 (9H)
Gal (2)	4.47 (d, *J* = 7.5 Hz, 2H)	3.60	3.71	4.16 (d, *J* = 4.1 Hz, 2H)	n/a	n/a	-
GlcNAc (3)	4.71 (d, *J* = 8.2 Hz, 2H)	3.79	3.46	3.72	n/a	n/a	2.04 (9H)
Fuc	5.14 (d, *J* = 4.1 Hz, 1H)	3.71	3.91	n/a	4.87	1.30 (d, *J* = 6.3 Hz, 3H)	
Gal (3)	4.60 (d, *J* = 7.5 Hz, 1H)	3.92	3.98	4.28	n/a	n/a	-
Fuc (2)	5.30 (d, *J* = 4.1 Hz, 1H)	3.82	3.70	n/a	4.34 (q, *J* = 6.7 Hz, 1H)	1.33 (d, *J* = 6.7 Hz, 3H)	-
Gal (4)	5.25 (d, *J* = 3.4 Hz, 1H)	3.90	3.93	3.99	n/a	n/a	-

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.61 (1H), 3.92 (1H)	1.61 (2H)	1.41 (2H)	1.67 (2H)	2.99 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc (1)	Gal (1)	GlcNAc (2)	Gal (2)	GlcNAc (3)	Fuc	Gal (3)	Fuc (2)	Gal (4)
C1	101.11	102.90	102.62	102.90	102.62	98.23	100.29	98.88	92.94

ESI TOF-MS m/z calculated for $C_{65}H_{112}N_4O_{44}Na$ $[M + Na]^+$ = 1675.6547, found 1675.6452.



18

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.54 (d, *J* = 8.1 Hz, 1H)	3.86	3.57	3.91	n/a	n/a	2.03 (12H)
Fuc	5.11 (d, *J* = 4.0 Hz, 1H)	3.69	3.90	3.78	4.81	1.16 (d, *J* = 6.6 Hz, 6H)	-
Gal	4.45 (d, *J* = 7.9 Hz, 2H)	3.50	3.70	4.10 (app.t, *J* = 4.0 Hz, 2H)	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 7.9 Hz, 2H)	3.95	3.58	3.85	n/a	n/a	2.03 (12H)
Fuc (2)	5.13 (d, *J* = 4.1 Hz, 1H)	3.69	3.90	3.78	4.81	1.16 (d, *J* = 6.6 Hz, 6H)	
Gal (2)	4.45 (d, *J* = 7.9 Hz, 2H)	3.50	3.70	4.10 (app.t, *J* = 4.0 Hz, 2H)	n/a	n/a	-
GlcNAc (3)	4.71 (d, *J* = 7.9 Hz, 2H)	3.95	3.44	3.85	n/a	n/a	2.03 (12H)
Fuc (3)	5.14 (d, *J* = 4.1 Hz, 1H)	3.69	3.90	3.78	4.34	1.30 (d, *J* = 6.6 Hz, 3H)	-
Gal (3)	4.60 (d, *J* = 7.7 Hz, 1H)	3.89	3.97	4.27 (d, *J* = 2.7 Hz, 1H)	n/a	n/a	-
Fuc (4)	5.30 (d, *J* = 4.3 Hz, 1H)	3.79	3.71	3.81	4.88	1.26 (d, *J* = 6.5 Hz, 3H)	-
Gal (4)	5.26 (d, *J* = 3.4 Hz, 1H)	3.91	3.98	n/a	n/a	n/a	2.03 (12H)

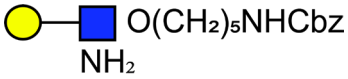
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e
3.60 (1H), 3.91 (1H)	1.68 (2H)	1.40 (2H)	1.59 (2H)	2.99 (2H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Fuc	Gal	GlcNAc (2)	Fuc (2)	Gal (2)	GlcNAc (3)	Fuc (3)	Gal (3)	Fuc (4)	Gal (4)
C1	100.76	98.42	101.63	102.15	98.42	101.63	102.15	98.42	100.05	98.73	92.75

ESI TOF-MS m/z calculated for $C_{77}H_{132}N_4O_{52}Na$ $[M + Na]^+$ = 1967.7705, found 1967.7683

NMR and MS data intermediates



20

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6
$GlcNH_2$	4.38 (d, *J* = 8.3 Hz, 1H)	2.64	3.51	3.59	3.56	3.79, 3.96
Gal	4.42 (d, *J* = 7.8 Hz, 1H)	3.52	3.61	3.91	3.64	3.72

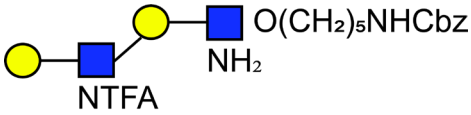
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.61 (1H) , 3.90 (1H)	1.61 (2H)	1.35 (2H)	1.47 (2H)	3.09 (2H)	5.08 (2H)	7.46 – 7.36 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	$GlcNH_2$	Gal
C1	102.38	103.05

ESI TOF-MS m/z calculated for $C_{25}H_{40}N_2O_{12}Na$ $[M + Na]^+$ = 583.2479, found 583.2481.

Yield: 89.2 % over 2 steps, 180.5 µmol, 101.1 mg



21

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6
$GlcNH_2$	4.72 (d, *J* = 8.5 Hz, 1H)	3.01	3.82	3.64	n/a	n/a
Gal	4.44 (d, *J* = 7.8, 1H)	3.59	3.74	4.19 (d, *J* = 3.1 Hz, 1H)	n/a	n/a
GlcNHTFA	4.84 (d, *J* = 7.9 Hz, 1H)	3.91	3.82	3.64	n/a	n/a
Gal (2)	4.50 (d, *J* = 7.8, 1H)	3.56	3.68	3.94	n/a	n/a

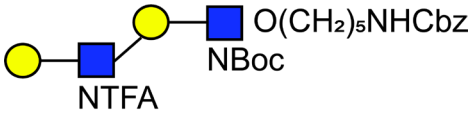
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.68 (1H) , 3.92 (1H)	1.64 (2H)	1.36 (2H)	1.51 (2H)	3.14 (2H)	5.09 (2H)	7.47 – 7.32 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	$GlcNH_2$	Gal	GlcNHTFA	Gal (2)
C1	98.84	103.04	102.04	102.92

ESI TOF-MS m/z calculated for $C_{39}H_{61}F_3N_3O_{22}$ $[M + H]^+$ = 980.3699, found 980.3383.

Yield: 96.4 % over 2 steps, 34.41 µmol, 33.7 mg.



22

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	t-Bu
GlcNHBoc	4.44 (d, *J* = 9.1 Hz, 1H)	3.33	3.51	3.61	n/a	3.77, 3.93	1.39 (9H)
Gal	4.42 (d, *J* = 7.9 Hz, 1H)	3.56	3.70	4.15	n/a	n/a	-
GlcNHTFA	4.82 (d, *J* = 7.9 Hz, 1H)	3.87	3.82	3.60	n/a	n/a	-
Gal (2)	4.47 (d, *J* = 7.8 Hz, 1H)	3.50	3.65	3.90	n/a	n/a	-

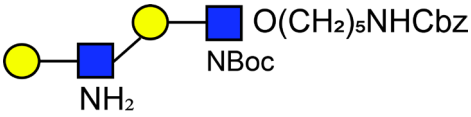
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.55 (1H) , 3.87 (1H)	1.54 (2H)	1.33 (2H)	1.47 (2H)	3.11 (2H)	5.08 (2H)	7.48 – 7.34 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNHBoc	Gal	GlcNHTFA	Gal (2)
C1	101.51	103.05	102.73	103.03

ESI TOF-MS m/z calculated for $C_{44}H_{68}F_3N_3O_{24}Na$ $[M + Na]^+$ = 1102.4043, found 1102.4007.

Yield: 66.5 %, 22.88 µmol, 24.7 mg



23

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	t-Bu
GlcNHBoc	4.44 (d, *J* = 9.1 Hz, 1H)	3.33	3.52	3.61	n/a	n/a	1.39 (9H)
Gal	4.43 (d, *J* = 7.9 Hz, 1H)	3.56	3.70	4.17	n/a	n/a	-
$GlcNH_2$	4.79 (d, *J* = 7.9 Hz, 1H)	2.91	3.82	3.75	3.60	n/a	-
Gal (2)	4.50 (d, *J* = 7.8 Hz, 1H)	3.50	3.65	3.99	n/a	n/a	-

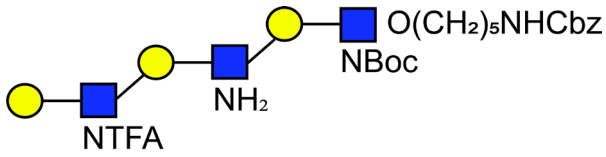
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.55 (1H) , 3.87 (1H)	1.54 (2H)	1.33 (2H)	1.47 (2H)	3.10 (2H)	5.12 (2H)	7.51 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNHBoc	Gal	$GlcNH_2$	Gal (2)
C1	101.44	103.01	101.82	102.47

ESI TOF-MS m/z calculated for $C_{42}H_{70}N_3O_{23}$ $[M + H]^+$ = 984.4400, found 984.4400.

Yield: Quantitative, 22.88 µmol, 22.5 mg



24

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	t-Bu
GlcNHBoc	4.44 (d, *J* = 8.5 Hz, 1H)	3.34	3.61	3.51	n/a	n/a	1.37 (9H)
Gal	4.41 (d, *J* = 7.9 Hz, 1H)	3.56	3.70	4.16	n/a	n/a	-
$GlcNH_2$	4.60 (d, *J* = 8.2 Hz, 1H)	2.73	3.51	3.62	n/a	n/a	-
Gal (2)	4.49 (d, *J* = 7.9 Hz, 1H)	3.64	3.78	4.16	n/a	n/a	-
GlcNHTFA	4.80	3.87	3.76	3.60	n/a	n/a	-
Gal (3)	4.47 (d, *J* = 7.8 Hz, 1H)	3.53	3.65	3.91	n/a	n/a	-

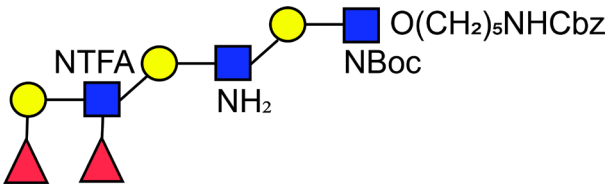
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.56 (1H), 3.88 (1H)	1.54 (2H)	1.34 (2H)	1.47 (2H)	3.11 (2H)	5.09 (2H)	7.49 – 7.34 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNHBoc	Gal	$GlcNH_2$	Gal (2)	GlcNHTFA	Gal (3)
C1	101.54	102.94	104.22	102.61	102.01	102.82

ESI TOF-MS m/z calculated for $C_{56}H_{90}F_3N_4O_{33}$ $[M + H]^+$ = 1403.5440, found 1403.5561.

Yield: 74.7 % over 2 steps, 8.55 µmol, 12.0 mg.



25

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	t-Bu
GlcNHBoc	4.47 (d, *J* = 8.5 Hz, 1H)	3.36	3.64	3.55	n/a	n/a	1.42 (s, 9H)
Gal	4.46 (d, *J* = 7.9 Hz, 1H)	3.59	3.72	4.15	n/a	n/a	-
$GlcNH_2$	4.71 (d, *J* = 8.3 Hz, 1H)	3.80	3.60	3.73	n/a	n/a	-
Gal (2)	4.46 (d, *J* = 7.9 Hz, 1H)	3.59	3.72	4.15	n/a	n/a	-
GlcNHTFA	4.72	3.97	3.83	3.46	n/a	n/a	-
Fuc	5.12	3.70	3.92	3.82	4.87	1.24 (d, *J* = 6.2 Hz, 3H)	
Gal (3)	4.52(d, *J* = 7.8 Hz, 1H)	3.66	3.65	3.87	n/a	n/a	-
Fuc (2)	5.29 (d, *J* = 3.1 Hz, 1H)	3.79	n/a	n/a	4.26	1.27 (d, *J* = 6.6 Hz, 3H)	-

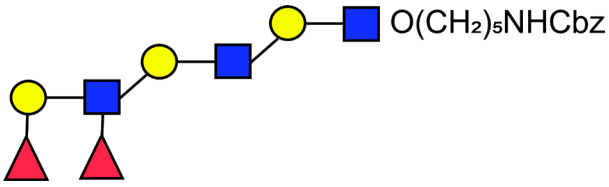
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.58 (1H), 3.88 (1H)	1.56 (2H)	1.36 (2H)	1.50 (2H)	3.13 (2H)	5.09 (2H)	7.49 – 7.34 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNHBoc	Gal	$GlcNH_2$	Gal (2)	GlcNHTFA	Fuc	Gal (3)	Fuc (2)
C1	101.75	102.70	102.66	102.70	102.66	98.60	100.00	99.22

ESI TOF-MS m/z calculated for $C_{68}H_{110}F_3N_4O_{41}$ $[M + H]^+$ = 1695.6598, found 1695.7198

Yield: 62.0 % over 2 steps, 1.06 µmol, 1.8 mg.



26

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc (1)	4.52 (d, *J* = 8.3 Hz, 1H)	3.72	3.59	n/a	n/a	n/a	2.12 – 1.96 (m, 9H)
Gal	4.46 (d, *J* = 7.9 Hz, 1H)	3.58	3.72	4.16	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.3 Hz, 1H)	3.81	3.58	3.72	n/a	n/a	2.12 – 1.96 (m, 9H)
Gal (2)	4.46 (d, *J* = 7.9 Hz, 1H)	3.58	3.72	4.16	n/a	n/a	-
GlcNAc (3)	4.71	3.96	3.46	n/a	n/a	n/a	2.12 – 1.96 (m, 9H)
Fuc	5.12	3.69	3.92	3.81	4.88	1.24 (app.d. J = 7.0 Hz, 3H)	
Gal (3)	4.52(d, *J* = 7.8 Hz, 1H)	3.66	3.64	3.86	n/a	n/a	-
Fuc (2)	5.29 (d, *J* = 3.1 Hz, 1H)	3.79	n/a	n/a	4.82	1.42 (app.d. J = 7.3 Hz, 3H)	-

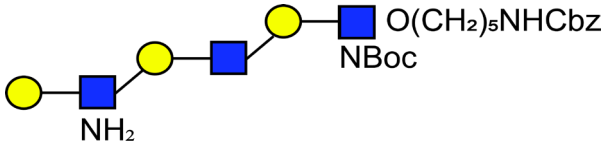
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.56 (1H), 3.86 (1H)	1.57 (2H)	1.30 (2H)	1.48 (2H)	3.11 (2H)	5.09 (2H)	7.50 – 7.36 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc (1)	Gal (1)	GlcNAc (2)	Gal (2)	GlcNAc (3)	Fuc	Gal (3)	Fuc (2)
C1	100.62	102.92	102.59	102.92	102.59	98.56	100.62	99.29

ESI TOF-MS m/z calculated for $C_{67}H_{109}N_4O_{41}$ $[M + H]^+$ = 1625.6567, found 1625.6308.

Yield: 75.5 % over 3 steps, 0.80 µmol, 1.3 mg.



27

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	t-Bu	NAc
GlcNHBoc	4.48 (d, *J* = 8.1 Hz, 1H)	3.36	3.64	3.56	n/a	n/a	1.41 (9H)	-
Gal	4.46 (d, *J* = 7.7 Hz, 2H)	3.60	3.73	4.16	n/a	n/a	-	-
GlcNAc	4.71 (d, *J* = 8.4 Hz, 1H)	3.80	3.59	3.74	n/a	n/a	-	2.04 (3H)
Gal (2)	4.53 (d, *J* = 7.9 Hz, 1H)	3.68	3.81	4.19 (d, *J* = 3.2 Hz, 1H)	n/a	n/a	-	-
$GlcNH_2$	4.63 (d, *J* = 8.2 Hz, 1H)	2.75	3.55	3.66	n/a	n/a	-	-
Gal (3)	4.46 (d, *J* = 7.7 Hz, 2H)	3.54	3.67	3.94	n/a	n/a	-	-

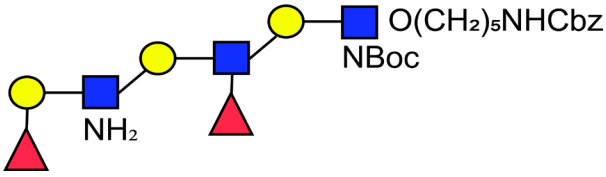
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.58 (1H), 3.89 (1H)	1.57 (2H)	1.36 (2H)	1.50 (2H)	3.13 (2H)	5.13 (2H)	7.48 – 7.40 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNHBoc	Gal	GlcNAc	Gal (2)	$GlcNH_2$	Gal (3)
C1	101.68	102.86	102.79	102.44	104.39	102.86

ESI TOF-MS m/z calculated for $C_{56}H_{93}N_4O_{33}$ $[M + H]^+$ = 1349.5722, found 1349.5552.

Yield: 83.2 % over 2 steps, 2.37 μmol, 3.2 mg.



28

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	t-Bu	NAc
GlcNHBoc	4.48 (d, *J* = 8.1 Hz, 1H)	3.36	3.64	3.56	n/a	n/a	1.41 (9H)	-
Gal	4.46 (d, *J* = 7.7 Hz, 2H)	3.60	3.73	4.16	n/a	n/a	-	-

GlcNAc	5.00 (d, J = 8.3 Hz, 2H)	3.18	3.54	3.81	n/a	n/a	-	2.04 (3H)
Fuc	5.12 (d, J = 3.9 Hz, 1H)	3.70	3.90	3.78	4.78	1.20 (6H)	-	-
Gal (2)	4.53 (d, J = 7.9 Hz, 1H)	3.68	3.81	4.19 (d, J = 3.2 Hz, 1H)	n/a	n/a	-	-
$GlcNH_2$	5.00 (d, J = 8.3 Hz, 2H)	3.18	3.66	3.81	n/a	n/a	-	-
Gal (3)	4.46 (d, J = 7.7 Hz, 2H)	3.54	3.67	3.94	n/a	n/a	-	-
Fuc (2)	5.31 (d, J = 3.4 Hz, 1H)	3.82	n/a	n/a	4.78	1.20 (6H)	-	-

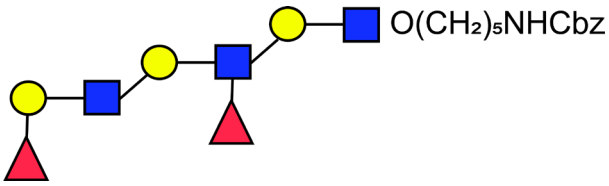
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.58 (1H), 3.89 (1H)	1.57 (2H)	1.36 (2H)	1.50 (2H)	3.13 (2H)	5.13 (2H)	7.48 – 7.40 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNHBoc	Gal	GlcNAc	Fuc	Gal (2)	$GlcNH_2$	Gal (3)	Fuc (2)
C1	100.69	102.31	100.06	98.68	102.31	100.06	101.48	99.36

ESI TOF-MS m/z calculated for $C_{68}H_{113}N_4O_{41}$ $[M + H]^+$ = 1641.6880, found 1641.6678.

Yield: 77.2 % over 2 steps, 1.83 μmol, 3.0 mg.



29

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.70	3.47	3.68	3.82	n/a	n/a	1.95 (9H)
Gal	4.46 (d, J = 7.7 Hz, 2H)	3.54	3.74	4.16	n/a	n/a	-
GlcNAc (2)	4.70	3.88	3.69	n/a	n/a	n/a	1.95 (9H)
Fuc	5.12	3.71	3.90	3.77	4.11	1.32 (3H)	-
Gal (2)	4.46 (d, J = 7.7 Hz, 2H)	3.54	3.74	4.16	n/a	n/a	-
GlcNAc (3)	4.70	3.47	3.68	3.82	n/a	n/a	1.95 (9H)
Gal (3)	4.46 (d, J = 7.7 Hz, 2H)	3.54	3.69	3.84	n/a	n/a	-
Fuc (2)	5.32	3.82	n/a	n/a	4.22	1.22 (3H)	-

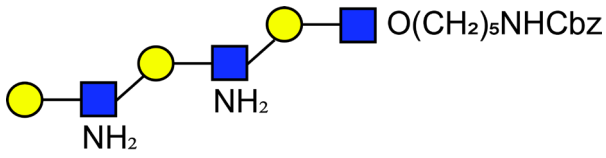
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.58 (1H), 3.89 (1H)	1.57 (2H)	1.36 (2H)	1.50 (2H)	3.13 (2H)	5.13 (2H)	7.48 – 7.40 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Fuc	Gal (2)	GlcNAc (3)	Gal (3)	Fuc (2)
C1	102.66	102.72	100.98	99.47	102.72	102.66	102.72	99.58

ESI TOF-MS m/z calculated for $C_{67}H_{108}N_4O_{41}Na$ $[M + Na]^+$ = 1647.6387, found 1647.6323.

Yield: 57.4 % over 2 steps, 1.05 µmol, 1.7 mg.



30

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.52 (app. t, J = 7.9 Hz, 3H)	3.69	3.58	3.82	n/a	n/a	2.00 (3H)
Gal	4.52 (app. t, J = 7.9 Hz, 3H)	3.69	3.82	4.19 (d, J = 3.2 Hz, 2H)	n/a	n/a	-
$GlcNH_2$	4.67 (d, J = 8.2 Hz, 2H)	2.80	3.58	3.69	n/a	n/a	-
Gal (2)	4.52 (app. t, J = 7.9 Hz, 3H)	3.69	3.82	4.19 (d, J = 3.2 Hz, 2H)	n/a	n/a	-
$GlcNH_2$ (2)	4.67 (d, J = 8.2 Hz, 2H)	2.80	3.58	3.69	n/a	n/a	-
Gal (3)	4.46 (d, J = 7.8 Hz, 1H)	3.56	3.67	3.94	n/a	n/a	-

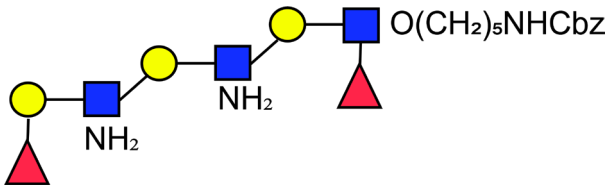
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.57 (1H), 3.88 (1H)	1.48 (2H)	1.32 (2H)	1.55 (2H)	3.12 (2H)	5.11 (2H)	7.48 – 7.39 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	$GlcNH_2$	Gal (2)	$GlcNH_2$ (2)	Gal (3)
C1	100.96	102.71	104.08	102.71	104.08	103.04

ESI TOF-MS m/z calculated for $C_{51}H_{84}N_4O_{31}Na$ $[M + Na]^+$ = 1271.5017, found 1271.5101.

Yield: 57.8 % over 4 steps, 1.36 µmol, 1.7 mg.



31

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.53 (d, *J* = 7.8 Hz, 1H)	3.89	3.58	3.82	n/a	n/a	2.00 (3H)
Fuc	5.11	3.69	3.78	3.89	n/a	n/a	-
Gal	4.51 (d, *J* = 7.7 Hz, 2H)	3.67	3.88	4.19	4.71	1.22	-
$GlcNH_2$	5.00 (d, *J* = 8.3 Hz, 2H)	3.18	3.52	3.82	n/a	n/a	-
Gal (2)	4.51 (d, *J* = 7.7 Hz, 2H)	3.67	3.88	4.19	n/a	n/a	-
$GlcNH_2$ (2)	5.00 (d, *J* = 8.3 Hz, 2H)	3.18	3.65	3.82	n/a	n/a	-
Gal (3)	4.55 (d, *J* = 7.8 Hz, 1H)	3.66	3.87	3.88	n/a	n/a	-
Fuc (2)	5.31 (d, *J* = 3.4 Hz, 1H)	3.81	n/a	n/a	4.83	1.18 (3H)	-

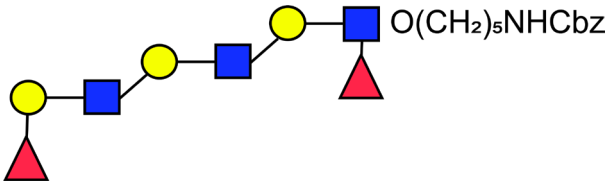
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.56 (1H), 3.87 (1H)	1.48 (2H)	1.32 (2H)	1.55 (2H)	3.12 (2H)	5.11 (2H)	7.48 – 7.39 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Fuc	Gal	$GlcNH_2$	Gal (2)	$GlcNH_2$ (2)	Gal (3)	Fuc (2)
C1	100.52	98.59	101.31	100.09	101.31	100.09	102.59	99.45

ESI TOF-MS m/z calculated for $C_{63}H_{105}N_4O_{39}$ $[M + H]^+$ = 1541.6356, found 1541.6983.

Yield: 61.8 % over 2 steps, 0.84 µmol, 1.3 mg.



32

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.52 (d, *J* = 7.2 Hz, 1H))	3.87	3.56	3.92	n/a	n/a	2.02 (s, 9H)
Fuc	5.09 (d, *J* = 3.7 Hz, 1H)	3.69	3.78	3.88	4.81	1.14 (3H)	-
Gal	4.44 (d, *J* = 8.0 Hz, 1H)	3.53	3.71	4.10 (d, *J* = 3.4 Hz, 1H)	n/a	n/a	-
GlcNAc(2)	4.70 (d, *J* = 8.3 Hz, 1H)	3.82	3.47	3.72	n/a	n/a	-2.02 (s, 9H
Gal (2)	4.48 (d, *J* = 7.8 Hz, 1H	3.60	3.71	4.15 (d, *J* = 3.4 Hz, 1H)	n/a	n/a	-
GlcNAc (3)	4.70 (d, *J* = 8.3 Hz, 1H)	3.82	3.58	3.72	n/a	n/a	-2.02 (s, 9H
Gal (3)	4.55 (d, *J* = 7.6 Hz, 1H)	3.67	3.65	3.88	n/a	n/a	-
Fuc (2)	5.31 (d, *J* = 3.0 Hz, 1H)	3.80	n/a	n/a	4.24	1.22 (3H)	-

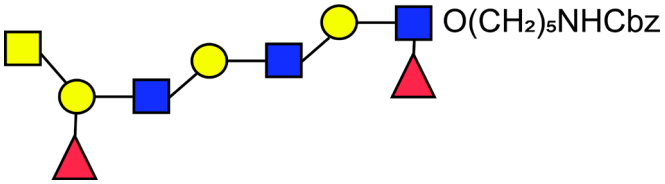
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.55 (1H), 3.87 (1H)	1.54 (2H)	1.29 (2H)	1.48 (2H)	3.11 (2H)	5.11 (2H)	7.49 – 7.39 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Fuc	Gal	GlcNAc(2)	Gal (2)	GlcNAc (3)	Gal (3)	Fuc (2)
C1	100.86	98.61	101.75	102.70	102.84	102.70	100.22	99.30

ESI TOF-MS m/z calculated for $C_{67}H_{109}N_4O_{41}$ $[M + H]^+$ = 1625.6567, found 1625.6541.

Yield: 88.1 %, 0.74 µmol, 1.2 mg.



33

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.51 (d, *J* = 9.0 Hz, 1H)	3.80	3.87	3.94	n/a	n/a	2.04 (12H)
Fuc	5.09 (d, *J* = 4.2 Hz, 1H)	3.68	3.89	3.79	4.83	1.15 (3H)	-
Gal	4.44 (d, *J* = 7.9 Hz, 1H)	3.52	3.71	4.10 (d, *J* = 2.7 Hz, 1H)	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.7 Hz, 2H)	3.82	3.45	3.71	n/a	n/a	2.04 (12H)
Gal (2)	4.48 (d, *J* = 7.5 Hz, 1H)	3.58	3.72	4.15 (d, *J* = 2.5 Hz, 1H)	n/a	n/a	-
GlcNAc (3)	4.70 (d, *J* = 8.7 Hz, 2H)	3.82	3.58	3.71	n/a	n/a	2.04 (12H)
Gal (3)	4.61 (d, *J* = 7.5 Hz, 1H)	3.91	4.00	4.23	n/a	n/a	-
Fuc (2)	5.36 (d, *J* = 4.1 Hz, 1H)	3.81	3.72	n/a	4.30	1.24 (3H)	-
GalNAc	5.18 (d, *J* = 3.8 Hz, 1H)	4.25	3.91	3.99	n/a	n/a	2.04 (12H)

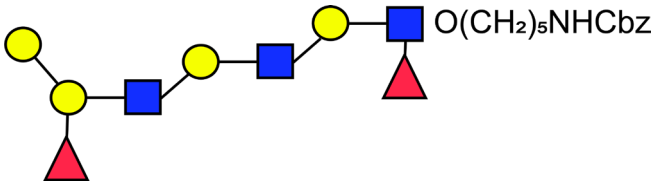
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.55 (1H), 3.87 (1H)	1.54 (2H)	1.33 (2H)	1.48 (2H)	3.11 (2H)	5.11 (2H)	7.48 – 7.40 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Fuc	Gal	GlcNAc (2)	Gal (2)	GlcNAc (3)	Gal (3)	Fuc (2)	GalNAc
C1	100.62	98.63	101.89	102.64	102.78	102.64	100.05	98.37	91.17

ESI TOF-MS m/z calculated for $C_{75}H_{122}N_5O_{46}$ $[M + H]^+$ = 1828.7361, found 1828.7609.

Yield: 61.9 %, 0.17 µmol, 0.4 mg.



34

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.52 (d, *J* = 7.8 Hz, 1H)	3.85	3.57	3.94	n/a	n/a	2.01 (12H)
Fuc	5.09 (d, *J* = 4.0 Hz, 1H)	3.68	3.89	3.78	4.83	1.16 (d, *J* = 6.5 Hz, 3H)	-
Gal	4.44 (d, *J* = 7.9 Hz, 1H)	3.51	3.71	4.10 (d, *J* = 2.7 Hz, 1H)	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.4 Hz, 2H)	3.81	3.45	3.71	n/a	n/a	2.01 (12H)
Gal (2)	4.48 (d, *J* = 7.9 Hz, 1H)	3.59	3.71	4.15 (d, *J* = 2.5 Hz, 1H)	n/a	n/a	-
GlcNAc (3)	4.70 (d, *J* = 8.4 Hz, 2H)	3.81	3.58	3.71	n/a	n/a	2.01 (12H)
Gal (3)	4.63 (d, *J* = 7.7 Hz, 1H)	3.93	4.29	4.00	n/a	n/a	-
Fuc (2)	5.33 (d, *J* = 4.2 Hz, 1H)	3.80	3.72	3.82	4.30	1.24 (d, *J* = 6.5 Hz, 3H)	-
Gal (4)	5.25 (d, *J* = 2.7 Hz, 1H)	3.89	3.91	3.98	n/a	n/a	2.01 (12H)

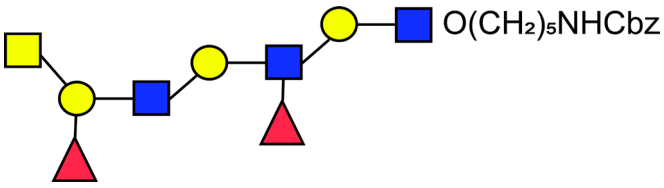
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.55 (1H), 3.87 (1H)	1.54 (2H)	1.31 (2H)	1.48 (2H)	3.11 (2H)	5.11 (2H)	7.47 – 7.38 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Fuc	Gal	GlcNAc (2)	Gal (2)	GlcNAc (3)	Gal (3)	Fuc (2)	Gal (4)
C1	100.91	98.86	101.85	102.76	103.08	102.76	100.16	98.88	92.98

ESI TOF-MS m/z calculated for $C_{73}H_{118}N_4O_{46}Na$ $[M + Na]^+$ = 1809.6915, found 1809.6746.

Yield: 83.6 %, 0.17 μmol, 0.3 mg.



35

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.51 (d, *J* = 6.4 Hz, 1H)	3.73	3.58	3.86	n/a	n/a	2.04 (12H)
Gal	4.46 (d, *J* = 7.8 Hz, 2H)	3.55	3.72	4.12	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.0 Hz, 1H)	3.94	3.82	n/a	n/a	n/a	2.04 (12H)
Fuc	5.12	3.70	3.89	3.78	4.32	1.36 (3H)	-
Gal (2)	4.46 (d, *J* = 7.8 Hz, 2H)	3.55	3.72	4.12	n/a	n/a	-
GlcNAc (3)	4.71 (d, *J* = 6.9 Hz, 1H)	3.83	3.44	3.71	n/a	n/a	2.04 (12H)
Gal (3)	4.61 (d, *J* = 7.7 Hz, 1H)	3.90	4.00	4.24	n/a	n/a	-
Fuc (2)	5.36 (d, *J* = 4.2 Hz, 1H)	3.81	3.72	n/a	4.32	1.24 (3H)	-
GalNAc	5.18 (d, *J* = 3.8 Hz, 1H)	4.25	3.91	4.00	n/a	n/a	2.04 (12H)

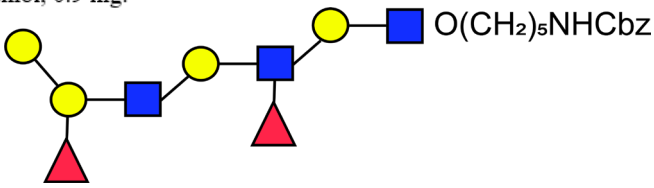
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.57 (1H), 3.88 (1H)	1.55 (2H)	1.32 (2H)	1.38 (2H)	3.12 (2H)	5.13 (2H)	7.48 – 7.40 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Fuc	Gal (2)	GlcNAc (3)	Gal (3)	Fuc (2)	GalNAc
C1	100.87	102.55	102.79	98.66	102.55	102.79	100.09	98.59	91.13

ESI TOF-MS m/z calculated for $C_{75}H_{122}N_5O_{46}$ $[M + H]^+$ = 1828.7361, found 1828.8913

Yield: 91.6 %, 0.49 µmol, 0.9 mg.



36

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.51 (d, *J* = 7.6 Hz, 1H)	3.69	3.57	3.83	n/a	n/a	2.03 (9H)
Gal	4.46 (d, *J* = 7.9 Hz, 2H)	3.56	3.72	4.12	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.3 Hz, 2H)	3.96	3.86	3.59	n/a	n/a	2.03 (9H)
Fuc	5.13 (d, *J* = 3.5 Hz, 1H)	3.69	3.90	3.78	4.32	1.15(3H)	-
Gal (2)	4.46 (d, *J* = 7.9 Hz, 2H)	3.56	3.72	4.12	n/a	n/a	-
GlcNAc (3)	4.70 (d, *J* = 8.3 Hz, 2H)	3.81	3.44	3.70	n/a	n/a	2.03 (9H)
Gal (3)	4.63 (d, *J* = 7.6 Hz, 1H)	3.94	4.01	4.29	n/a	n/a	-
Fuc (2)	5.34 (d, *J* = 4.2 Hz, 1H)	3.80	3.72	3.83	4.32	1.24 (3H)	-
Gal (4)	5.25 (d, *J* = 2.8 Hz, 1H)	3.89	3.99	4.20	n/a	n/a	-

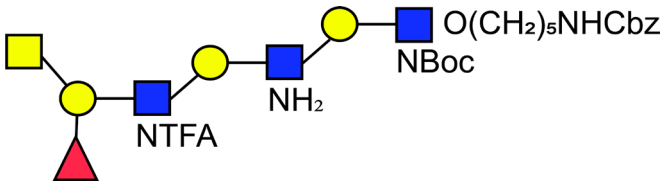
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.57 (1H), 3.88 (1H)	1.54 (2H)	1.31 (2H)	1.38 (2H)	3.11 (2H)	5.12 (2H)	7.48 – 7.40 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Fuc	Gal (2)	GlcNAc (3)	Gal (3)	Fuc (2)	Gal (4)
C1	100.93	102.46	102.70	98.74	102.46	102.70	100.09	98.68	93.04

ESI TOF-MS m/z calculated for $C_{73}H_{118}N_4O_{46}Na$ $[M + Na]^+$ = 1809.6915, found 1809.6808

Yield: 66.8 %, 0.22 µmol, 0.4 mg.



37

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	t-Bu	NAc
GlcNHBoc	4.47 (d, *J* = 8.5 Hz, 1H)	3.36	3.68	3.55	n/a	n/a	1.41 (s, 9H)	-
Gal	4.45 (d, *J* = 7.9 Hz, 1H)	3.58	3.72	4.18	n/a	n/a	-	-
$GlcNH_2$	5.02 (d, *J* = 8.3 Hz, 1H)	3.16	3.63	3.85	n/a	n/a	-	-
Gal (2)	4.53 (d, *J* = 8.0 Hz, 1H)	3.72	3.90	4.20	n/a	n/a	-	-
GlcNHTFA	4.83	3.86	3.72	3.50	n/a	n/a	-	-
Gal (3)	4.62 (d, *J* = 7.8 Hz, 1H)	3.90	4.00	4.25	n/a	n/a	-	-
Fuc	5.36 (d, *J* = 4.2 Hz, 1H)	3.80	3.73	n/a	4.33	1.27 (d, *J* = 6.6 Hz, 3H)	-	-
GalNAc	5.18 (d, *J* = 3.9 Hz, 1H)	4.25	3.91	3.99	n/a	n/a	-	2.04 (3H)

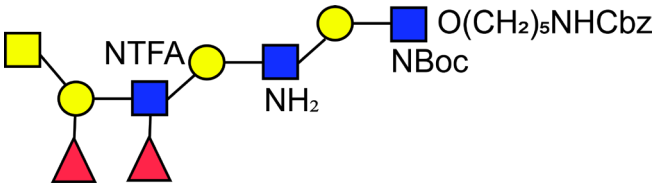
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.58 (1H), 3.89 (1H)	1.58 (2H)	1.35 (2H)	1.49 (2H)	3.14 (2H)	5.11 (2H)	7.47 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNHBoc	Gal	$GlcNH_2$	Gal (2)	GlcNHTFA	Gal (3)	Fuc	GalNAc
C1	101.40	102.91	102.10	102.42	99.96	100.17	98.49	91.31

ESI TOF-MS m/z calculated for $C_{70}H_{113}F_3N_5O_{42}$ $[M + H]^+$ = 1752.6812, found 1752.6591

Yield: 64.5 % over 2 steps, 0.46 µmol, 0.8 mg.



38

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	t-Bu	NAc
GlcNHBoc	4.45 (d, *J* = 8.7 Hz, 1H)	3.34	3.62	3.52	n/a	n/a	1.40 (s, 9H)	-
Gal	4.43 (d, *J* = 7.9 Hz, 1H)	3.57	3.70	4.15 (d, *J* = 3.2 Hz, 1H)	n/a	n/a	-	-
$GlcNH_2$	5.00 (d, *J* = 8.5 Hz, 1H)	3.15	3.63	3.86	n/a	n/a	-	-
Gal (2)	4.51 (d, *J* = 7.8 Hz, 1H)	3.71	3.89	4.18 (d, *J* = 3.1 Hz, 2H)	n/a	n/a	-	-
GlcNHTFA	4.81	3.99	3.83	3.48	n/a	n/a	-	-
Fuc	5.02 (d, *J* = 4.0 Hz, 1H)	3.68	3.88	3.78	4.87	1.29 (dd, *J* = 13.6, 6.6 Hz, 3H)		
Gal (3)	4.57 (d, *J* = 7.6 Hz, 1H)	3.88	3.94	4.18 (d, *J* = 3.1 Hz, 2H)	n/a	n/a	-	-
Fuc (2)	5.29 (d, *J* = 4.4 Hz, 1H)	3.78	3.70	3.83	4.34	1.34 (d, *J* = 6.7 Hz, 3H)	-	-
GalNAc	5.18 (d, *J* = 3.7 Hz, 1H)	4.24	3.92	3.99	n/a	n/a	-	2.02 (3H)

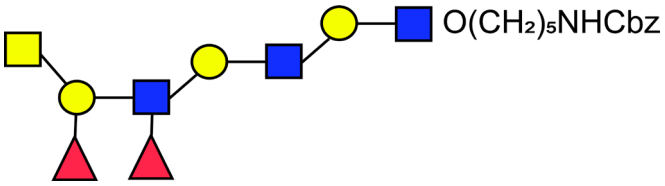
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.58 (1H), 3.89 (1H)	1.55 (2H)	1.34 (2H)	1.49 (2H)	3.14 (2H)	5.10 (2H)	7.45 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNHBoc	Gal	$GlcNH_2$	Gal (2)	GlcNHTFA	Fuc	Gal (3)	Fuc (2)	GalNAc
C1	101.17	102.70	102.10	102.17	99.92	98.49	99.99	98.37	91.08

ESI TOF-MS m/z calculated for $C_{76}H_{112}F_3N_5O_{46}Na$ $[M + Na]^+$ = 1910.6428, found 1910.4825

Yield: 56.5 %, 0.26 µmol, 0.5 mg.



39

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc (1)	4.51 (d, *J* = 7.7 Hz, 1H)	3.69	3.58	n/a	n/a	n/a	2.04 (12H)
Gal	4.47 (d, *J* = 8.1 Hz, 2H)	3.59	3.71	4.15	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.7 Hz, 2H)	3.82	3.58	3.73	n/a	n/a	2.04 (12H)
Gal (2)	4.47 (d, *J* = 8.1 Hz, 2H)	3.59	3.71	4.15	n/a	n/a	-
GlcNAc (3)	4.70 (d, *J* = 8.7 Hz, 2H)	3.82	3.43	3.73	n/a	n/a	2.04 (12H)
Fuc	5.14 (d, *J* = 4.0 Hz, 1H)	3.70	3.92	3.79	4.88	1.29 (dd, *J* = 13.6, 6.6 Hz, 3H)	
Gal (3)	4.58 (d, *J* = 7.6 Hz, 1H)	3.88	3.94	4.20	n/a	n/a	-
Fuc (2)	5.31 (d, *J* = 4.2 Hz, 1H)	3.79	3.84	3.85	4.34	1.34 (d, *J* = 6.9 Hz, 3H)	-
GalNAc	5.20 (d, *J* = 3.9 Hz, 1H)	4.25	3.92	4.00	n/a	n/a	2.04 (12H)

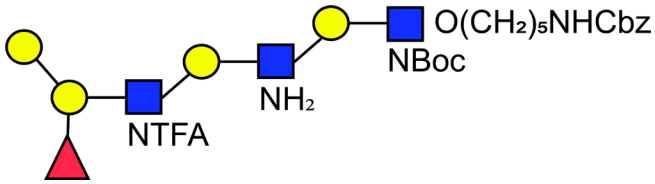
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.56 (1H), 3.86 (1H)	1.55 (2H)	1.30 (2H)	1.49 (2H)	3.12 (2H)	5.09 (2H)	7.48 – 7.39 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc (1)	Gal (1)	GlcNAc (2)	Gal (2)	GlcNAc (3)	Fuc	Gal (3)	Fuc (2)	GalNAc
C1	100.96	102.82	102.75	102.82	102.75	98.32	100.22	98.80	91.14

ESI TOF-MS m/z calculated for $C_{75}H_{121}N_5O_{46}Na$ $[M + Na]^+$ = 1850.7181, found 1850.7427

Yield: 42.3 % over 3 steps, 0.11 μmol, 0.2 mg.



40

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	t-Bu
GlcNHBoc	4.47 (d, *J* = 8.0 Hz, 1H)	3.35	3.67	3.54	n/a	n/a	1.41 (s, 9H)
Gal	4.45 (d, *J* = 7.9 Hz, 1H)	3.59	3.71	4.17 (d, *J* = 3.5 Hz, 1H)	n/a	n/a	-
$GlcNH_2$	5.01 (d, *J* = 8.4 Hz, 1H)	3.13	3.58	3.89	n/a	n/a	-
Gal (2)	4.53 (d, *J* = 7.9 Hz, 1H)	3.69	3.86	4.20 (d, *J* = 3.2 Hz, 2H)	n/a	n/a	-
GlcNHTFA	4.84 (d, *J* = 8.2 Hz, 1H)	3.90	3.71	3.50	n/a	n/a	-
Gal (3)	4.64 (d, *J* = 7.6 Hz, 1H)	3.94	4.00	4.29 (d, *J* = 2.9 Hz, 1H)	n/a	n/a	-
Fuc	5.34 (d, *J* = 4.2 Hz, 1H)	3.80	3.70	n/a	4.31 (q, *J* = 6.9 Hz, 1H)	1.24 (d, *J* = 6.9 Hz, 3H)	-
Gal (4)	5.25 (d, *J* = 2.5 Hz, 1H)	4.25	3.89	3.98	n/a	n/a	-

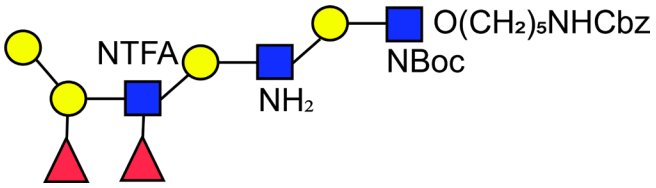
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.59 (1H), 3.88 (1H)	1.58 (2H)	1.37 (2H)	1.50 (2H)	3.14 (2H)	5.11 (2H)	7.47 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNHBoc	Gal	$GlcNH_2$	Gal (2)	GlcNHTFA	Gal (3)	Fuc	Gal (4)
C1	101.43	103.03	100.22	102.47	102.05	100.00	98.68	93.05

ESI TOF-MS m/z calculated for $C_{68}H_{110}F_3N_4O_{42}$ $[M + H]^+$ = 1711.6547, found 1711.6508

Yield: 41.5 % over 2 steps, 0.88 µmol, 0.15 mg.



41

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	t-Bu
GlcNHBoc	4.47 (d, *J* = 8.7 Hz, 1H)	3.35	3.64	3.53	n/a	n/a	1.41 (s, 9H)
Gal	4.45 (d, *J* = 7.9 Hz, 1H)	3.57	3.71	4.16 (d, *J* = 3.7 Hz, 1H)	n/a	n/a	-
$GlcNH_2$	5.01 (d, *J* = 8.4 Hz, 1H)	3.17	3.65	3.87	n/a	n/a	-
Gal (2)	4.53 (d, *J* = 7.9 Hz, 1H)	3.73	3.89	4.20 (d, *J* = 3.2 Hz, 1H)	n/a	n/a	-
GlcNHTFA	4.81	3.97	3.85	3.48	n/a	n/a	-
Fuc	5.04 (d, *J* = 3.9 Hz, 1H)	3.71	3.89	3.78	4.90	1.25 (d, *J* = 6.5 Hz, 3H)	
Gal (3)	4.60 (d, *J* = 7.7 Hz, 1H)	3.91	3.99	4.27	n/a	n/a	-
Fuc (2)	5.29 (d, *J* = 4.2 Hz, 1H)	3.81	3.70	3.83	4.45 (q, *J* = 7.9 Hz, 1H)	1.30 (d, *J* = 6.5 Hz, 3H)	-
Gal (4)	5.25 (d, *J* = 3.4 Hz, 1H)	3.91	3.99	n/a	n/a	n/a	-

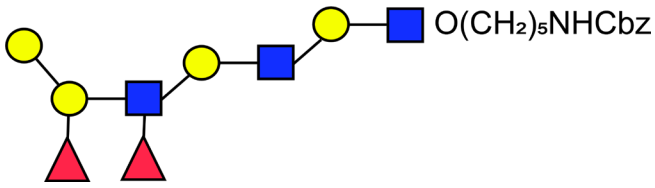
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.58 (1H), 3.89 (1H)	1.57 (2H)	1.34 (2H)	1.50 (2H)	3.13 (2H)	5.10 (2H)	7.45 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNHBoc	Gal	$GlcNH_2$	Gal (2)	GlcNHTFA	Fuc	Gal (3)	Fuc (2)	Gal (4)
C1	101.42	102.91	99.90	102.44	101.65	98.81	100.24	98.85	92.80

ESI TOF-MS m/z calculated for $C_{74}H_{119}F_3N_4O_{46}Na$ $[M + Na]^+$ = 1879.6945, found 1879.6887

Yield: 42.5 %, 0.09 µmol, 0.07 mg.



42

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc (1)	4.51 (d, *J* = 7.7 Hz, 1H)	3.72	3.58	n/a	n/a	n/a	2.00 (3H)
Gal	4.47 (d, *J* = 8.1 Hz, 2H)	3.60	3.72	4.15	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.5 Hz, 2H)	3.80	3.63	3.73	n/a	n/a	2.03 (6H)
Gal (2)	4.47 (d, *J* = 8.1 Hz, 2H)	3.60	3.72	4.15	n/a	n/a	-
GlcNAc (3)	4.70 (d, *J* = 8.5 Hz, 2H)	3.80	3.48	3.73	n/a	n/a	2.03 (6H)
Fuc	5.13 (d, *J* = 4.0 Hz, 1H)	3.68	3.91	n/a	4.89	1.25 (d, *J* = 6.3 Hz, 3H)	
Gal (3)	4.59 (d, *J* = 7.7 Hz, 1H)	3.89	3.94	4.27	n/a	n/a	-
Fuc (2)	5.29 (d, *J* = 4.3 Hz, 1H)	3.80	3.71	n/a	4.33 (q, *J* = 6.5 Hz, 1H)	1.29 (d, *J* = 6.6 Hz, 3H)	-
Gal (4)	5.25 (d, *J* = 3.4 Hz, 1H)	3.90	3.92	4.00	n/a	n/a	-

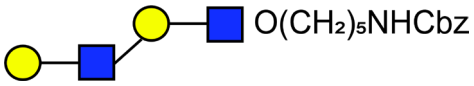
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.56 (1H), 3.86 (1H)	1.55 (2H)	1.31 (2H)	1.49 (2H)	3.12 (2H)	5.11 (2H)	7.48 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc (1)	Gal (1)	GlcNAc (2)	Gal (2)	GlcNAc (3)	Fuc	Gal (3)	Fuc (2)	Gal (4)
C1	101.01	102.91	102.65	102.91	102.65	98.52	100.31	98.78	92.91

ESI TOF-MS m/z calculated for $C_{73}H_{118}N_4O_{46}Na$ $[M + Na]^+$ = 1809.6915, found 1809.6851

Yield: 34.1 % over 3 steps, 0.03 μmol, 0.05 mg.



^{1}H (600 MHz, D_2O): δ (ppm) **44**

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.51 (d, *J* = 8.1 Hz, 1H)	3.71	n/a	n/a	n/a	n/a	2.02 (6H)
Gal	4.46 (d, *J* = 8.2 Hz, 1H)	3.59	3.73	4.16	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.5 Hz, 1H)	3.81	n/a	n/a	n/a	n/a	2.02 (6H)
Gal (2)	4.48 (d, *J* = 8.0 Hz, 1H)	3.60	3.69	3.93	n/a	n/a	-

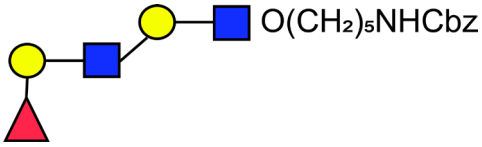
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.55 (1H) , 3.87 (1H)	1.55 (2H)	1.31 (2H)	1.49 (2H)	3.12 (2H)	5.11 (2H)	7.48 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Gal (2)
C1	101.20	103.02	102.55	102.84

ESI TOF-MS m/z calculated for $C_{41}H_{66}N_3O_{23}$ $[M + H]^+$ = 968.4087, found 968.4476.

Yield: 70.1 % over 3 steps, 37.52 µmol, 36.3 mg.



^{1}H (600 MHz, D_2O): δ (ppm) **45**

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.51 (d, *J* = 7.6 Hz, 1H)	3.72	3.58	n/a	n/a	n/a	2.03 (6H)
Gal	4.46 (d, *J* = 7.9 Hz, 1H)	3.60	3.74	4.15 (d, *J* = 3.3 Hz, 1H)	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.4 Hz, 1H)	3.84	3.47	3.97	n/a	n/a	2.03 (6H)
Gal (2)	4.55 (d, *J* = 7.7 Hz, 1H)	3.67	3.69	3.87	n/a	n/a	-
Fuc	5.31 (d, *J* = 3.1 Hz, 1H)	3.81	3.90	3.86	4.21	1.23 (app. d, *J* = 6.6 Hz, 3H)	-

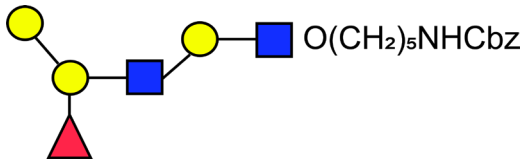
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.57 (1H) , 3.87 (1H)	1.54 (2H)	1.32 (2H)	1.48 (2H)	3.13 (2H)	5.10 (2H)	7.47 – 7.39 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Gal (2)	Fuc
C1	101.02	102.90	102.85	100.28	99.41

ESI TOF-MS m/z calculated for $C_{47}H_{76}N_3O_{27}$ $[M + H]^+$ = 1114.4666, found 1114.4825.

Yield: 59.7 %, 4.94 µmol, 5.5 mg.



46

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.51 (d, *J* = 7.6 Hz, 1H)	3.70	3.58	n/a	n/a	n/a	2.03 (6H)
Gal	4.46 (d, *J* = 7.9 Hz, 1H)	3.58	3.72	4.15 (d, *J* = 3.3 Hz, 1H)	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.4 Hz, 1H)	3.82	3.42	3.70	n/a	n/a	2.03 (6H)
Gal (2)	4.63 (d, *J* = 7.6 Hz, 1H)	3.93	4.00	4.29 (d, *J* = 2.9 Hz, 1H)	n/a	n/a	-
Fuc	5.33 (d, *J* = 4.1 Hz, 1H)	3.81	3.72	3.86	4.29	1.24 (d, *J* = 6.6 Hz, 3H)	-
Gal(3)	5.25 (d, *J* = 2.8 Hz, 1H)	3.89	3.99	4.20	n/a	n/a	-

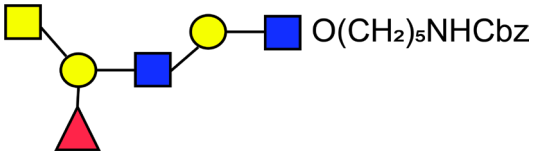
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.57 (1H) , 3.87 (1H)	1.54 (2H)	1.31 (2H)	1.48 (2H)	3.12 (2H)	5.11 (2H)	7.49 – 7.39 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Gal (2)	Fuc	Gal (3)
C1	100.98	102.88	102.84	100.11	98.76	93.03

ESI TOF-MS m/z calculated for $C_{53}H_{86}N_3O_{32}$ $[M + H]^+$ = 1276.5195, found 1276.5139.

Yield: 77.1 %, 1.18 µmol, 1.5 mg.



47

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.51 (d, *J* = 7.7 Hz, 1H)	3.71	3.58	n/a	n/a	n/a	2.03 (9H)
Gal	4.46 (d, *J* = 7.8 Hz, 1H)	3.60	3.72	4.15 (d, *J* = 2.7 Hz, 1H)	n/a	n/a	-

GlcNAc (2)	4.70 (d, J = 8.4 Hz, 1H)	3.81	3.45	3.70	n/a	n/a	2.03 (9H)
Gal (2)	4.61 (d, J = 7.7 Hz, 1H)	3.90	3.99	4.23 (d, J = 3.1 Hz, 1H)	n/a	n/a	-
Fuc	5.36 (d, J = 4.1 Hz, 1H)	3.80	3.73	3.86	4.31	1.25 (d, J = 6.5 Hz, 3H)	-
GalNAc	5.18 (d, J = 3.8 Hz, 1H)	4.23	3.91	4.01	n/a	n/a	2.03 (9H)

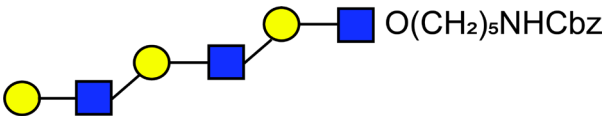
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.57 (1H) , 3.87 (1H)	1.55 (2H)	1.31 (2H)	1.48 (2H)	3.11 (2H)	5.11 (2H)	7.48 – 7.39 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Gal (2)	Fuc	GalNAc
C1	100.90	102.92	102.77	100.04	98.52	91.19

ESI TOF-MS m/z calculated for $C_{61}H_{97}N_3O_{41}Na$ $[M + Na]^+$ = 1317.5460, found 1317.5631.

Yield: 67.2 %, 1.37 µmol, 1.8 mg.



48

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.51 (d, J = 7.7 Hz, 1H)	3.72	3.57	3.81	n/a	n/a	2.00 (3H)
Gal	4.47 (d, J = 7.6 Hz, 2H)	3.58	3.74	4.16	n/a	n/a	-
GlcNAc (2)	4.71 (d, J = 8.0 Hz, 2H)	3.81	3.59	3.74	n/a	n/a	2.04 (6H)
Gal (2)	4.47 (d, J = 7.6 Hz, 2H)	3.58	3.74	4.16	n/a	n/a	-
GlcNAc (3)	4.71 (d, J = 8.0 Hz, 2H)	3.81	3.59	3.74	n/a	n/a	2.04 (6H)
Gal (3)	4.48 (d, J = 7.8 Hz, 1H)	3.55	3.71	3.93	n/a	n/a	-

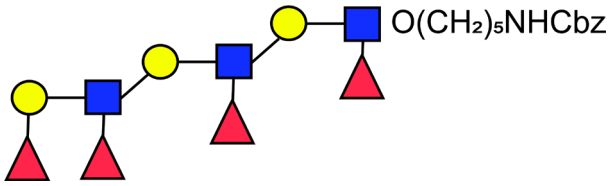
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.55 (1H) , 3.87 (1H)	1.55 (2H)	1.32 (2H)	1.49 (2H)	3.11 (2H)	5.11 (2H)	7.49 – 7.38 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Gal (2)	GlcNAc (3)	Gal (3)
C1	100.88	102.87	102.69	102.87	102.69	102.87

ESI TOF-MS m/z calculated for $C_{55}H_{89}N_4O_{33}$ $[M + H]^+$ = 1333.5409, found 1333.5374.

Yield: 60.8 % over 2 steps, 11.26 µmol, 15.0 mg.



49

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.71 (d, *J* = 8.7 Hz, 2H)	3.87	3.56	n/a	n/a	n/a	2.00 (3H)
Fuc	5.09 (d, *J* = 4.0 Hz, 1H)	3.68	3.89	3.78	4.81	1.15 (d, *J* = 6.5 Hz, 6H)	-
Gal	4.44 (d, *J* = 7.9 Hz, 2H)	3.50	3.68	4.15 (d, *J* = 3.4 Hz, 2H)	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.7 Hz, 2H)	3.96	3.57	3.86	n/a	n/a	2.04 (6H)
Fuc (2)	5.12 (d, *J* = 4.1 Hz, 2H)	3.69	3.90	3.78	4.81	1.15 (d, *J* = 6.5 Hz, 6H)	
Gal (2)	4.44 (d, *J* = 7.9 Hz, 2H)	3.50	3.68	4.15 (d, *J* = 3.4 Hz, 2H)	n/a	n/a	-
GlcNAc (3)	4.71 (d, *J* = 8.7 Hz, 2H)	3.96	3.45	3.86	n/a	n/a	2.04 (6H)
Fuc (3)	5.12 (d, *J* = 4.1 Hz, 2H)	3.69	3.90	3.83	4.25	1.27 (d, *J* = 6.6 Hz, 3H)	-
Gal (3)	4.71 (d, *J* = 8.7 Hz, 2H)	3.65	3.86	3.86	n/a	n/a	-
Fuc (4)	5.28 (d, *J* = 3.4 Hz, 1H)	3.80	n/a	3.81	4.88	1.24 (d, *J* = 6.6 Hz, 3H)	-

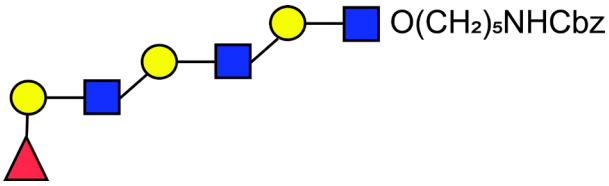
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.55 (1H) , 3.88 (1H)	1.54 (2H)	1.31 (2H)	1.48 (2H)	3.11 (2H)	5.11 (2H)	7.47 – 7.38 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Fuc	Gal	GlcNAc (2)	Fuc (2)	Gal (2)	GlcNAc (3)	Fuc (3)	Gal (3)	Fuc (4)
C1	100.54	98.66	101.71	102.42	98.60	101.71	102.42	98.60	100.10	99.35

ESI TOF-MS m/z calculated for $C_{79}H_{128}N_4O_{49}Na$ $[M + Na]^+$ = 1939.7545, found 1939.7749.

Yield: 67.3 %, 0.68 µmol, 1.3 mg.



50

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.50 (d, *J* = 7.6 Hz, 1H)	3.71	3.57	3.81	n/a	n/a	2.00 (3H)
Gal	4.47 (d, *J* = 7.6 Hz, 2H)	3.58	3.71	4.15 (d, *J* = 3.4 Hz, 2H)	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.4 Hz, 2H)	3.81	3.58	3.72	n/a	n/a	2.04 (6H)
Gal (2)	4.47 (d, *J* = 7.6 Hz, 2H)	3.58	3.71	4.15 (d, *J* = 3.4 Hz, 2H)	n/a	n/a	-
GlcNAc (3)	4.70 (d, *J* = 8.4 Hz, 2H)	3.81	3.46	3.72	n/a	n/a	2.04 (6H)
Gal (3)	4.55 (d, *J* = 7.7 Hz, 1H)	3.67	3.71	3.88	n/a	n/a	-
Fuc	5.31 (d, *J* = 3.0 Hz, 1H)	3.80	n/a	3.82	4.22 (q, *J* = 6.7 Hz, 1H)	1.23 (d, *J* = 6.5 Hz, 3H)	-

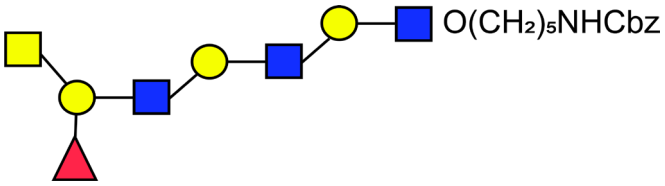
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.57 (1H) , 3.87 (1H)	1.54 (2H)	1.31 (2H)	1.49 (2H)	3.12 (2H)	5.11 (2H)	7.47 – 7.38 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Gal (2)	GlcNAc (3)	Gal (3)	Fuc
C1	100.93	102.86	102.76	102.86	102.76	100.26	99.30

ESI TOF-MS m/z calculated for $C_{61}H_{99}N_4O_{37}$ $[M + H]^+$ = 1479.5988, found 1479.6383.

Yield: 76.9 %, 1.73 μmol, 2.6 mg.



51

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.51 (d, *J* = 7.6 Hz, 1H)	3.71	3.57	3.81	n/a	n/a	2.02 (12H)
Gal	4.47 (d, *J* = 7.6 Hz, 2H)	3.60	3.72	4.15 (d, *J* = 3.4 Hz, 2H)	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.8 Hz, 2H)	3.81	3.58	3.71	n/a	n/a	2.02 (12H)
Gal (2)	4.47 (d, *J* = 7.6 Hz, 2H)	3.60	3.72	4.15 (d, *J* = 3.4 Hz, 2H)	n/a	n/a	-
GlcNAc (3)	4.71 (d, *J* = 8.8 Hz, 2H)	3.81	3.45	3.71	n/a	n/a	2.02 (12H)
Gal (3)	4.61 (d, *J* = 7.2 Hz, 1H)	3.91	3.91	4.23	n/a	n/a	-
Fuc	5.36 (d, *J* = 3.0 Hz, 1H)	3.79	3.72	3.82	4.33 (q, *J* = 6.7 Hz, 1H)	1.26 (d, *J* = 6.5 Hz, 3H)	-
GalNAc	5.18	4.24	3.92	4.00	n/a	n/a	2.02 (12H)

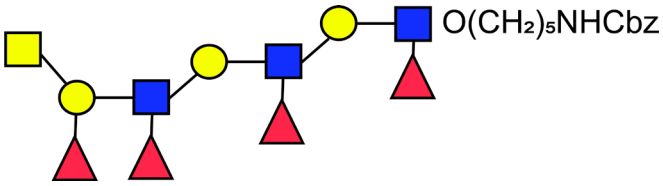
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.55 (1H) , 3.86 (1H)	1.55 (2H)	1.31 (2H)	1.48 (2H)	3.12 (2H)	5.11 (2H)	7.48 – 7.38 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Gal (2)	GlcNAc (3)	Gal (3)	Fuc	GalNAc
C1	100.98	102.82	102.76	102.82	102.76	100.08	98.62	91.23

ESI TOF-MS m/z calculated for $C_{69}H_{111}N_5O_{42}Na$ $[M + Na]^+$ = 1704.6601, found 1704.6687

Yield: 91.5 %, 0.53 µmol, 0.9 mg.



52

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.56 (d, *J* = 7.6 Hz, 1H)	3.87	3.59	3.92	n/a	n/a	2.00 (9H)
Fuc	5.07 (d, *J* = 3.9 Hz, 1H)	3.66	3.86	3.75	4.81	1.13 (d, *J* = 6.6 Hz, 6H)	-
Gal	4.44 (d, *J* = 8.0 Hz, 2H)	3.49	3.68	4.09	n/a	n/a	-
GlcNAc (2)	4.69 (d, *J* = 8.7 Hz, 2H)	3.93	3.55	3.84	n/a	n/a	2.00 (9H)
Fuc (2)	5.11 (d, *J* = 4.0 Hz, 1H)	3.66	3.86	3.75	4.81	1.13 (d, *J* = 6.6 Hz, 6H)	
Gal (2)	4.44 (d, *J* = 8.0 Hz, 2H)	3.49	3.68	4.09	n/a	n/a	-
GlcNAc (3)	4.69 (d, *J* = 8.7 Hz, 2H)	3.93	3.44	3.84	n/a	n/a	2.00 (9H)
Fuc (3)	5.13 (d, *J* = 3.9 Hz, 1H)	3.68	3.90	3.77	4.25	1.28 (d, *J* = 6.6 Hz, 3H)	-
Gal (3)	4.53 (d, *J* = 8.0 Hz, 1H)	3.43	3.62	3.72	n/a	n/a	-
Fuc (4)	5.29 (d, *J* = 4.1 Hz, 1H)	3.77	3.70	3.81	4.88	1.26 (d, *J* = 6.6 Hz, 3H)	-
GalNAc	5.18 (d, *J* = 3.6 Hz, 1H)	4.23	3.94	4.00	n/a	n/a	2.14 (3H)

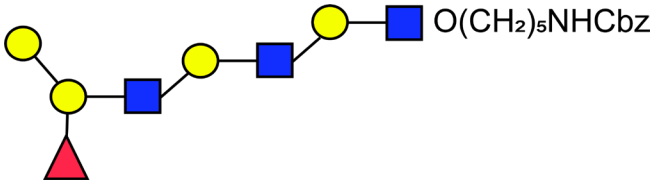
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.55 (1H) , 3.88 (1H)	1.54 (2H)	1.31 (2H)	1.48 (2H)	3.11 (2H)	5.09 (2H)	7.46 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Fuc	Gal	GlcNAc (2)	Fuc (2)	Gal (2)	GlcNAc (3)	Fuc (3)	Gal (3)	Fuc (4)	GalNAc
C1	100.54	98.53	101.80	102.45	98.53	101.80	102.45	98.53	100.10	99.35	91.52

ESI TOF-MS m/z calculated for $(C_{87}H_{141}N_5O_{54}Na_2)/2$ $[M + 2\ Na]^{2+}/2$ = 1082.9118, found 1082.9960

Yield: 65.6 %, 0.24 μmol, 0.5 mg.



53

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.51 (d, *J* = 7.7 Hz, 1H)	3.71	3.57	3.81	n/a	n/a	2.04 (9H)
Gal	4.47 (d, *J* = 7.6 Hz, 2H)	3.59	3.72	4.15 (d, *J* = 3.4 Hz, 2H)	n/a	n/a	-
GlcNAc (2)	4.70 (d, *J* = 8.4 Hz, 2H)	3.81	3.58	3.71	n/a	n/a	2.04 (9H)
Gal (2)	4.47 (d, *J* = 7.6 Hz, 2H)	3.59	3.72	4.15 (d, *J* = 3.4 Hz, 2H)	n/a	n/a	-
GlcNAc (3)	4.70 (d, *J* = 8.4 Hz, 2H)	3.81	3.44	3.71	n/a	n/a	2.04 (9H)
Gal (3)	4.63 (d, *J* = 7.7 Hz, 1H)	3.93	3.99	4.28	n/a	n/a	-
Fuc	5.33 (d, *J* = 4.1 Hz, 1H)	3.79	3.71	3.82	4.33 (q, *J* = 6.7 Hz, 1H)	1.24 (d, *J* = 6.3 Hz, 3H)	-
Gal (4)	5.25 (d, *J* = 2.7 Hz, 1H)	3.88	3.98	4.20	n/a	n/a	-

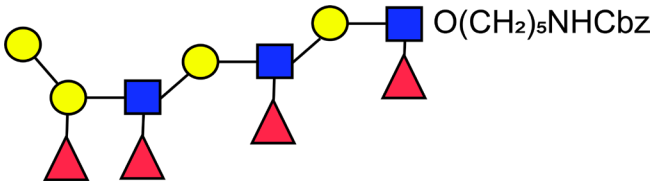
CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.55 (1H) , 3.86 (1H)	1.55 (2H)	1.32 (2H)	1.48 (2H)	3.12 (2H)	5.11 (2H)	7.49 – 7.38 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Gal	GlcNAc (2)	Gal (2)	GlcNAc (3)	Gal (3)	Fuc	Gal (4)
C1	101.01	102.96	102.84	102.96	102.84	100.17	98.76	93.02

ESI TOF-MS m/z calculated for $C_{69}H_{108}N_4O_{42}Na$ $[M + Na]^+$ = 1663.6336, found 1663.6418

Yield: Quantitative, 0.58 μmol, 1.0 mg.



54

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NAc
GlcNAc	4.51 (d, *J* = 7.8 Hz, 1H)	3.86	3.58	n/a	n/a	n/a	2.00 (9H)
Fuc	5.09 (d, *J* = 3.7 Hz, 1H)	3.70	3.90	3.79	4.82	1.15 (d, *J* = 6.5 Hz, 6H)	-
Gal	5.29 (d, *J* = 4.2 Hz, 2H)	3.51	3.70	4.09 (d, *J* = 4.0 Hz, 2H)	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.3 Hz, 2H)	3.96	3.58	3.86	n/a	n/a	2.00 (9H)
Fuc (2)	5.13 (d, *J* = 4.1 Hz, 1H)	3.70	3.90	3.79	4.82	1.15 (d, *J* = 6.5 Hz, 6H)	
Gal (2)	5.29 (d, *J* = 4.2 Hz, 2H)	3.51	3.70	4.09 (d, *J* = 4.0 Hz, 2H)	n/a	n/a	-
GlcNAc (3)	4.71 (d, *J* = 8.3 Hz, 2H)	3.96	3.44	3.86	n/a	n/a	2.00 (9H)
Fuc (3)	5.14 (d, *J* = 3.7 Hz, 1H)	3.70	3.90	3.79	4.33	1.29 (d, *J* = 6.5 Hz, 3H)	-
Gal (3)	4.60 (d, *J* = 7.5 Hz, 1H)	3.90	3.98	4.27 (d, *J* = 2.8 Hz, 1H)	n/a	n/a	-
Fuc (4)	5.29 (d, *J* = 4.2 Hz, 1H)	3.81	3.70	3.84	4.89	1.25 (d, *J* = 6.6 Hz, 3H)	-
Gal (4)	5.25 (d, *J* = 3.8 Hz, 1H)	3.90	3.97	n/a	n/a	n/a	2.14 (3H)

CH_2 a	CH_2 b	CH_2 c	CH_2 d	CH_2 e	Bn (CH_2)	Bn (C_6H_5)
3.56 (1H) , 3.87 (1H)	1.54 (2H)	1.29 (2H)	1.48 (2H)	3.11 (2H)	5.11 (2H)	7.48 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	GlcNAc	Fuc	Gal	GlcNAc (2)	Fuc (2)	Gal (2)	GlcNAc (3)	Fuc (3)	Gal (3)	Fuc (4)	Gal (4)
C1	100.67	98.47	101.49	102.30	98.47	101.49	102.30	98.47	100.10	98.68	92.63

ESI TOF-MS m/z calculated for $C_{85}H_{139}N_4O_{54}$ $[M + H]^+$ = 2079.8254, found 2102.8002

Yield: 58.8 %, 0.24 µmol, 0.7 mg.

Glycan microarray

Results of plant lectins

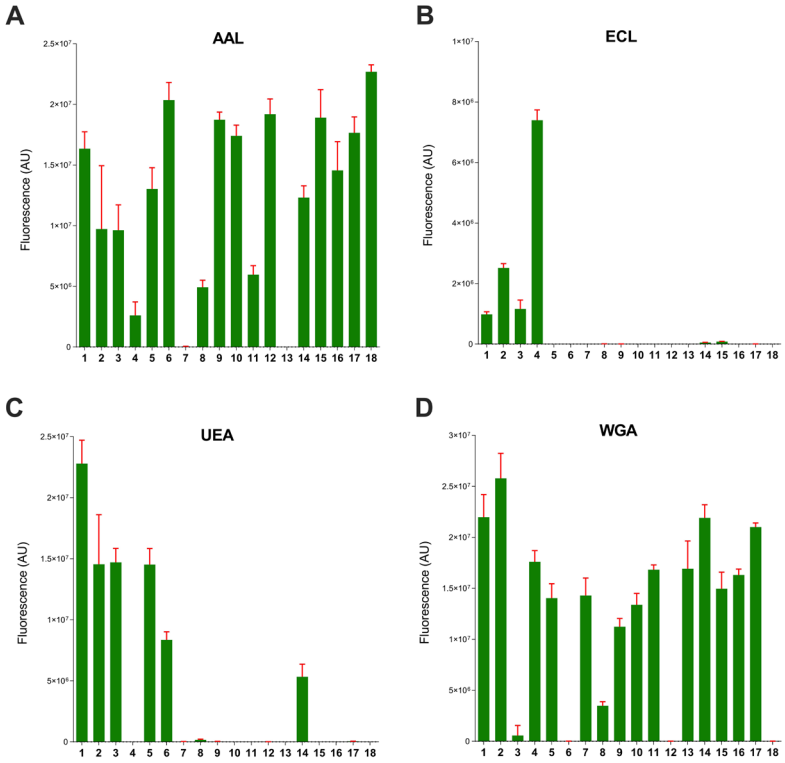


Figure S1. Microarray results of the synthetic ABO glycan library at 100 µM with lectins AAL (5 µg/mL), ECL (20 µg/mL), UEA (20 µg/mL), and WGA (20 µg/mL). Bars represent the mean ± SD (n=4).

Results of blood group antibodies

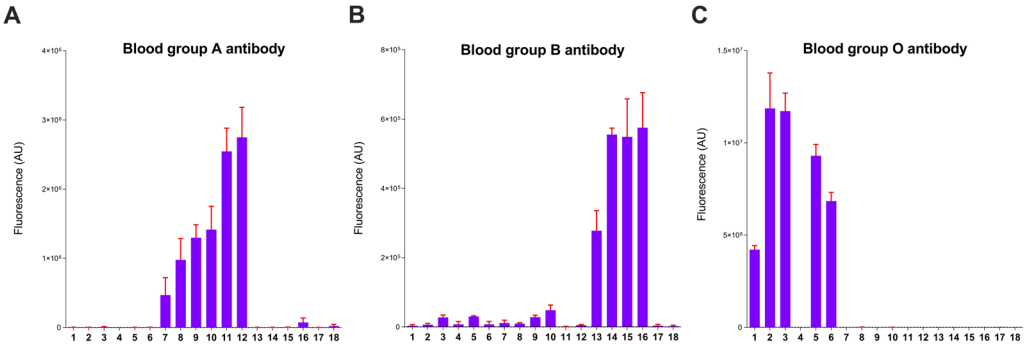


Figure S2. Microarray results of the synthetic ABO glycan library at 100 µM with anti-bloodgroup A, B, and O antibodies (10 µg/mL). Bars represent the mean ± SD (n=4).

Methods Glycan microarray

Glycan array printing

The synthetic ABO glycans (100 µM in sodium phosphate (250 mM), pH 8.5 buffer) were printed as replicates of 6 on activated glass slides (Nexterion Slide H, Schott Inc) by piezoelectric non-contact printing (sciFLEXARRAYER S3, Scienion Inc) with a drop volume of ~400 pL and 1 drop per spot at 50 % relative humidity. Each slide has 24 subarrays (3x8). After printing, the slides were incubated overnight in a saturated NaCl chamber (providing a 75% relative humidity environment), after which the remaining activated esters were quenched with ethanolamine (50 mM) in TRIS (100 mM), pH 9.0. Slides were rinsed with DI water, dried by centrifugation, and stored in a desiccator at room temperature.

Microarray validation and screening

Sub-arrays (6x18 spots) were incubated with biotinylated lectins (*Aleuria aurantia* lectin (AAL; Vector Labs B-1395), *Erythrina cristagalli* lectin (ECL; Vector Labs B-1145), *Ulex europaeus* agglutinin I (UEA; Vector Labs B-1065) and wheat germ agglutinin (WGA; Vector Labs B-1025) at the indicated concentrations premixed with Streptavidin-AlexaFluor635 (5 µg/mL; ThermoFisher Scientific, S32364) in TSM binding buffer (20 mM Tris Cl, pH 7.4, 150 mM NaCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 0.05% Tween, 1% BSA) for 1 h followed by washing. Wash steps involved 4 successive washes of the whole slides with TSM wash buffer (20 mM Tris Cl, pH 7.4, 150 mM NaCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 0.05% Tween-20) - TSM buffer (20 mM Tris Cl, pH 7.4, 150 mM NaCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$) - 2x deionized H_2O with 5 min soak times.

Mouse monoclonal anti-blood group A IgM antibody, clone HE-193 (Sigma-Aldrich SAB4700674), mouse monoclonal anti-blood group B IgM antibody, clone HEB-29 (Sigma-Aldrich SAB4700676), and mouse monoclonal anti-blood group O (H) IgM antibody, clone 87-N (Fitzgerald Industries International 10R-H103a) at the indicated concentrations were premixed with biotinylated anti-mouse IgM (5 µg/mL; Life Technologies Europe BV 13-5790-82) and Streptavidin-AlexaFluor635 (5 µg/mL) in TSM binding buffer and incubated for 2 h with the sub-arrays, followed by washing as described above.

Evaluation of various galectins binding was performed by incubation of the sub-arrays with premixes of recombinant human galectin 1, 3, 4, 7, 8, and 9 (R&D Systems 1152-GA, 1154-GA, 1227-GA, 1339-GA, 1305-GA, and 2045-GA respectively) at the indicated concentrations, its corresponding goat biotinylated anti-galectin antibody (3 µg/mL; R&D Systems BAF1152, BAF1154, BAF1227, BAF1339, BAF1305, and BAF2045 respectively), and Streptavidin-AlexaFluor635 (5 µg/mL) in TSM binding buffer for 2 h, followed by washing as described above.

All incubation and wash steps were performed at room temperature. Washed arrays were dried by centrifugation and immediately scanned for fluorescence on a GenePix 4000 B microarray scanner (Molecular Devices). The detection gain was adjusted to avoid saturation of the signal, whereby the same settings were used for each experiment to allow comparison. The data were processed with GenePix Pro 7 software and further analyzed using our home written Microsoft Excel macro. After removal of the lowest and highest value of the six replicates, the mean fluorescent intensities and standard deviations (SD) were calculated (n=4). Data were fitted using Prism software Version 8.3.0 (GraphPad Software, Inc). Bar graphs represent the mean ± SD for each compound. The highest possible protein concentration was employed at which good responsiveness was observed to achieve an appropriate dynamic range.

References

[1] W. Wang, T. Hu, P. A. Frantom, T. Zheng, B. Gerwe, D. S. del Amo, S. Garret, R. D. Seidel, P. Wu, Proc. Natl. Acad. Sci. 2009, 106, 16096–16101.

[2] W. Peng, J. Pranskevich, C. Nycholat, M. Gilbert, W. Wakarchuk, J. C. Paulson, N. Razi, Glycobiology 2012, 22, 1453–1464.

[3] A. R. Prudden, L. Liu, C. J. Capicciotti, M. A. Wolfert, S. Wang, Z. Gao, L. Meng, K. W. Moremen, G.-J. Boons, Proc. Natl. Acad. Sci. 2017, 114, 6954–6959.

[4] A. Forman, R. Pfoh, A. Eddenden, P. L. Howell, M. Nitz, Org. Biomol. Chem. 2019, 17, 9456–9466.

[5] R. F. Sala, S. L. MacKinnon, M. M. Palcic, M. E. Tanner, Carbohydr. Res. 1998, 306, 127–136.

[6] L. L. Morais, H. Yuasa, K. Bennis, I. Ripoche, F.-I. Auzanneau, 2006, 84, 10.

CHAPTER 3

Chemo-enzymatic Synthesis of I-branched Polylactosamines

Gaël M. Vos, Roosmarijn van der Woude, Yunfei Wu, Kelley W. Moremen, Robert P. de Vries, Geert-Jan Boons

Introduction

Poly-*N*-acetyl lactosamines (polyLacNAcs) are glycan substructures that are composed of repeating units of type 2 LacNAc disaccharides [Galβ1,4GlcNAcβ1-]$_n$. These substructures form backbones of many complex carbohydrates and can be presented as linear and branched architectures. The topology of polyLacNAc chains also forms the basis of the Ii blood group system classification.[1] The i antigen is a linear β1,3-linked polyLacNAc chain that enzymatically can be converted into an I antigen by *N*-acetyllactosaminide β1,6-*N*-acetylglucosaminyltransferase GCNT2, which introduces a β-linked GlcNAc moiety at the 6-position of internal galactosides of i antigens (Fig. 1A). This branching point can be extended by glycosyltransferases in a similar manner as the main glycan chain to form I-branched multi-antennary glycans.[2] I-branched structures are found on *N*- and *O*-linked glycoproteins, glycolipids and unconjugated human milk oligosaccharides (HMOs) (Fig. 1B).[2–6]

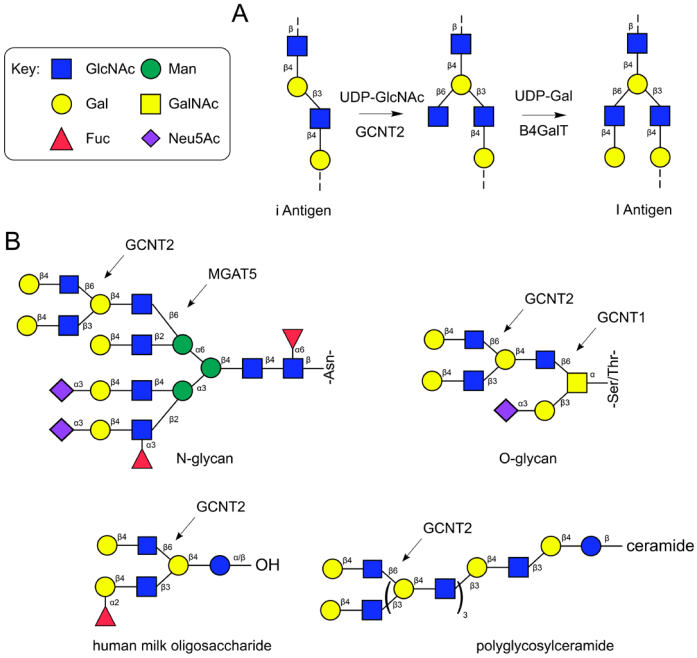


Figure 1: A) Biosynthetic pathway of the Ii blood group system; B) Overview of glycan classes carrying I-antigens. β1,6 GlcNAc modifications are denoted with the appropriate enzyme responsible for the transfer. Oligosaccharides represent only one of the possible structures.

Developing fetuses and neonates solely express i antigens until several months postpartum, after which I antigen expression is initiated.[7] Almost all adults express I antigens, whilst the i_{adult} phenotype, characterized by the absence of I antigen on red blood cells, is extremely rare (1 in 4400).[8] Interestingly, the delay in biosynthesis of I antigens in neonates correlates with a similar delay of blood group A expression. This delay in biosynthesis has been linked to the low incidence of hemolytic disease of the fetus and newborn (HDFN).[7,9] Destruction of fetal red blood cells by HDFN is caused by maternal alloimmunization through antibodies crossing the placental membrane during pregnancy. Infant red blood cells only express i antigen which have lower antigenicity compared to I-branched carrying red blood cells in adults that allows multivalent presentation of ABO antigens.[10]

In adults, reports of rare mutations/deletions in GCNT2 that prevent the biosynthesis of the I antigen are associated with development of congenital cataracts.[11–15] Moreover, alterations in GCNT2 expression and thus I-branch biosynthesis have been linked to cancer progression and metastasis.[4,16–21] I-branched *N*-glycans have also been implicated in regulation of B cell activation through modulation of galectin 9 binding.[22–24]

In I^+ adults, both linear and I-branched glycan structures are ubiquitously expressed.[25,26] Despite the prevalence of such structural motifs, it has been difficult to investigate how I-branching influences recognition by glycan binding proteins. Moreover, the distinction between I-branched and elongated linear polyLacNAc chains is difficult to resolve by conventional glycomic analysis due to identical glycan compositions. It is to be expected that well-defined I-branched glycan standards will be valuable tools to develop mass spectrometry-based methodologies that can distinguish linear from I-branched structures. Furthermore, such compounds will also be valuable to develop glycan microarrays to examine selectivities of glycan binding proteins.

I-branches have been installed using recombinant glycosyl transferases but these approaches lack regioselective and cannot provide isomeric compounds.[22,27] In an alternative approach, linear polyLacNAc structures were synthesized having an unnatural galactoside with a C-6 aldehyde that was employed to generate I-branching glycomimetics.[28] No method has been described yet that can install I-branches in a regioselective manner on polyLacNAc chains. To address this deficiency, we explored the regioselective introduction of I-branching sites using the enzyme GCNT2 in combination with polyLacNAc chains that are modified by traceless blocking groups that exerts regioselective control over GCNT2 activity. It was found that a glucosamine residue modified by *tert*-butyloxycarbamate (Boc) or 1,3-fucoside can function as temporary blocking group for I-branching allowing the preparation of isomeric structures. The methodology was used to prepare a library of well-defined I-branched glycans with complementary topologies.

Results and discussion

Hexasaccharide **3**, which contains two internal galactosides, was prepared to investigate a possible regioselectivity of the I-branching enzyme GCNT2 (Fig. 2A). This glycan structure is based on a lactose core that is present on HMOs and polyglycosylceramides, which are glycan classes that are often modified I-branches.[1,29] Benzyl lactoside **1** was used as the starting material because its anomeric center is locked in one anomeric configuration and furthermore the benzyl moiety increases sensitivity for MS and UV absorption based detection thereby facilitating identification and purification. Benzyl glycoside **1**, which was generated by a four-step procedure involving Koenigs-Knorr glycosylation of lactose with benzyl alcohol, was enzymatically extended to lacto-neo-*N*-tetraose **2** and then to lacto-neo-*N*-hexaose **3** by the alternating action of β1,3-*N*-acetylglucosaminyltransferase 2 (B3GNT2) and β1,4-galactosyltransferase 1 (B4GalT1). Treatment of **3** with the I-branching enzyme GCNT2 generated biantennary octasaccharide **4** when two or more equivalents of UDP-GlcNAc were employed. The addition of only one equivalent of UDP-GlcNAc led to a mixture of bi-antennary octasaccharide, mono-antennary heptasaccharide and starting material, which

indicates that GCNT2 does not possess inherent regioselectivity for this type of substrate. Control over I-branch activity of GCNT2 is necessary for the enzymatic synthesis of well-defined I-branched polyLacNAc chains. To address this challenge, a library of target polyLacNAc oligosaccharides was designed and synthesized that carry I-branches of various length on the proximal, distal or both non-terminal galactosides of a hexasaccharide HMO scaffold.

Figure 2: I-branching of 3 reveals no inherent selectivity of GCNT2 on oligosaccharides containing multiple I-branch acceptor sites.

Enzymatic synthesis of proximal I-branched polyLacNAc chains

PolyLacNAc chains that have an I-branch on the proximal galactoside were prepared by exploiting the inherent selectivity of GCNT2 for internal galactosides.[22] Thus, a compound such as **2** has only one acceptor site for this enzyme and therefore could readily be converted into common precursor **5**. The latter derivative has an asymmetrical architecture in which the main chain terminates as galactoside and the I-branch as a GlcNAc residue (Fig. 3). We exploited this asymmetry to selectively extend the terminal galactoside of main by a β1,3-linked GlcNAc moiety by treatment with B3GNT2 in the presence of UDP-GlcNAc to give compound **6**. Asymmetrical derivative **13** was readily obtained by further extension of both arms of **6** with a β1,4-galactoside by employing B4GalT1 and UDP-Gal. Alternatively, treatment of compound **5** with B4GalT1 in the presence of UDP-Gal gave symmetrical derivative **7** which could readily be converted into **14** that has two LacNAc moieties at the I-branch.

To allow further selective extension of the I-branch to give compounds such as **12**, it was necessary to temporarily deactivate the main chain from LacNAc extension. α2,3 linked sialic acid was selected as a blocking group because it directly occupies the acceptor site and can be readily removed by acid hydrolysis. Sialylation of **5** was achieved using ST3 β-galactoside α2,3-sialyltransferase 4 (ST3Gal4) and CMP-Neu5Ac to give in compound **9**. The I branch of **9** could selectively be extended by a LacNAc moiety to give **11** by subsequent treatment with B3GNT2 and B4GalT1. Next, the sialoside of the main chain of **11** was removed by treatment with aqueous acetic acid to generated **12**, which is an isomer of **13**. Both arms of **12** could be further extended by a LacNAc moiety to give derivative **15**.

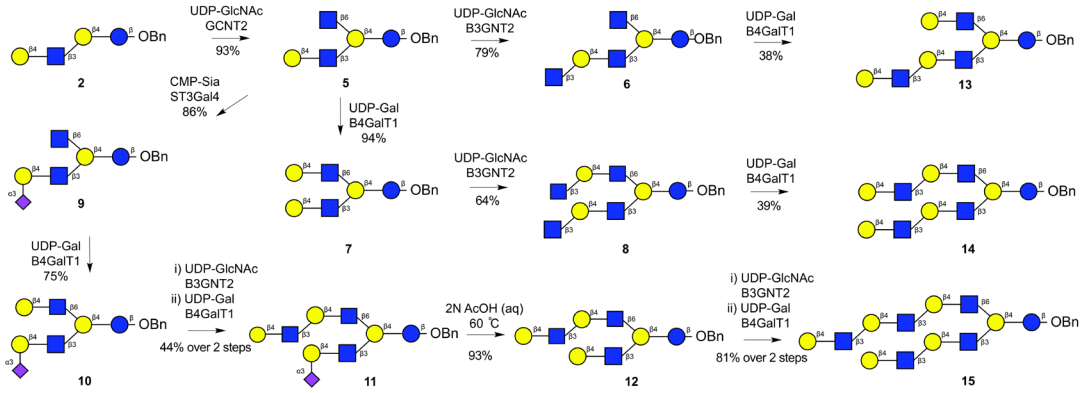


Figure 3: Enzymatic synthesis of compounds **13**, **14** and **15**.

Selective I-branch activity imposed by *N*-modified glucosamine derivatives

The approach described above makes it possible to prepare oligo-LacNAc chains having an I-branch of different length on the proximal lactose moiety. To prepare compounds having I-branch at one of the LacNAc moieties, we pursued a strategy in which a linear oligo-LacNAc chain such as **22** was prepared having a tert-butyloxycarbamate (Boc) modified glucosamine moiety at the central LacNAc position (Fig. 4). It was found that the Boc moiety blocks I-branching by GCNT2 at the proximal galactoside moiety to give selective formation of compound **23**. Key precursor **22** could readily be prepared by exploiting the finding that various GlcNAc transferases can employ UDP-GlcNTFA as unnatural sugar nucleotide donor to give oligosaccharides having a GlcNTFA moiety.[30] At an appropriate stage of synthesis, the TFA moiety can selectively be removed under mild basic conditions to give $GlcNH_2$ that can be further functionalized by for example a Boc masking group. Previously, we demonstrated that $GlcNH_2$ and GlcNBoc are resistant to α1,3-fucosylation to give access to polyLacNAc chains having various patterns of Lewis^x epitopes.[31] Therefore, we were compelled to investigate how these modifications influence the regioselective activity of GCNT2. For this purpose, we prepared model hexasaccharides **17**, **19** and **22** were synthetized containing various modifications on the internal LacNAc moiety (Fig. 4).

Thus, lactoside **1** was extended by a GlcNTFA moiety by treatment with UDP-GlcNTFA in the presence of Hpβ3GlcNAcT.[27,32] The GlcNTFA moiety of the resulting product was tolerated as an acceptor substrate for B4GalT1 resulting in the formation of LacNTFA containing tetrasaccharide **16**. Consecutive LacNAc extension of **17** using B3GNT2 and B4GalT1 gave hexasaccharide **17**. Aminolysis of the TFA protecting group of **17** was accomplished by treatment with aqueous ammonia to give amine **19** that was readily converted into **22** by reaction with Boc_2O in the presence of sodium bicarbonate.

Treatment of **17** with I-branching enzyme GCNT2 resulted in the modification of the two internal galactosides to give biantennary **18** demonstrating the TFA moiety does not exert control over the regioselectivity. A similar treatment of **19** led to full conversion to heptasaccharide **20**, however, extended reaction times still resulted in the slow formation of the undesired octasaccharide **21**. This residual activity suggests that GCNT2 still has some activity for the second I-branching site which may become problematic when blocking multiple sites on longer polyLacNAc chains are required. Gratifyingly, treatment of hexasaccharide **22** with GCNT2 led to the selective formation of I-branched heptasaccharide **23** and in this case no formation of octasaccharide was observed. Even a large excess of enzyme and sugar nucleotide did not result in progression of the reaction supporting that *N*-Boc fully deactivates one of the I-branching sites.

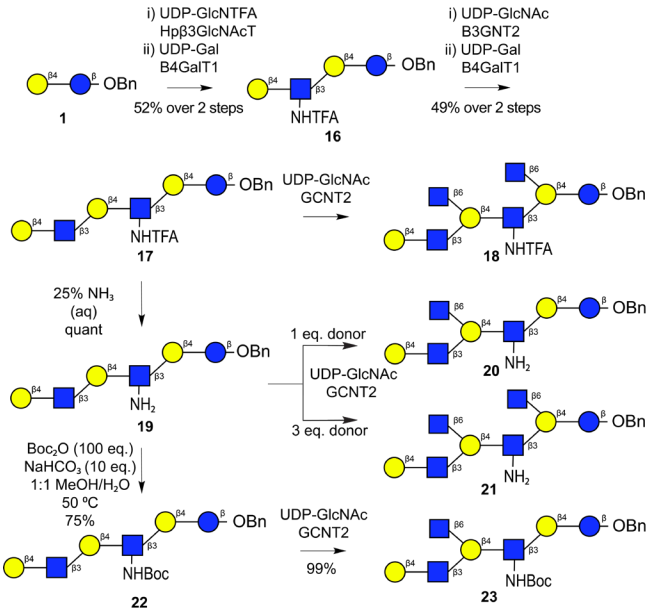


Figure 4: Chemo-enzymatic synthesis of **23**; Only one of the I-branching sites is blocked by GlcNBoc.

NMR analysis reveals selective blocking of the reducing end galactoside for I-branching by *N*-Boc

The position of I-branch regioselectivity imposed by *N*-modified glucosamine was confirmed by a combination of ^{1}H, COSY, NOESY, TOCSY and HSQC NMR experiments on compound **23** (Fig. 5). GlcNBoc **C** was identified by a large up field shift of the H2 signal relative to the other glucosamines (COSY, not shown). NOESY couplings 1/6 of GlcNBoc **C** and couplings 4/5 of GlcNAc **E** with neighboring galactoses revealed the H4 signals of galactose **B** and **D**, respectively. The I-branching modification resulted in a substantial increase in NOESY transfer between galactoside **D** H4 and H6/H6' which is absent in galactoside without modification on the C6 position. Additional NOESY couplings 2/3 could be detected from **D** H6/H6' to GlcNAc **G**-H1. These NOESY couplings reveals that I-branching GlcNAc **G** is connected to galactose **D**. The assignment of galactoside **D** as the I-branched point was confirmed by a downshift of H6 in the HSQC spectrum. Thus, it is revealed that the GlcNBoc modification blocks GCNT2 activity on the neighboring reducing end galactoside, while not interfering with I-branch activity on the neighboring non-reducing end galactoside.

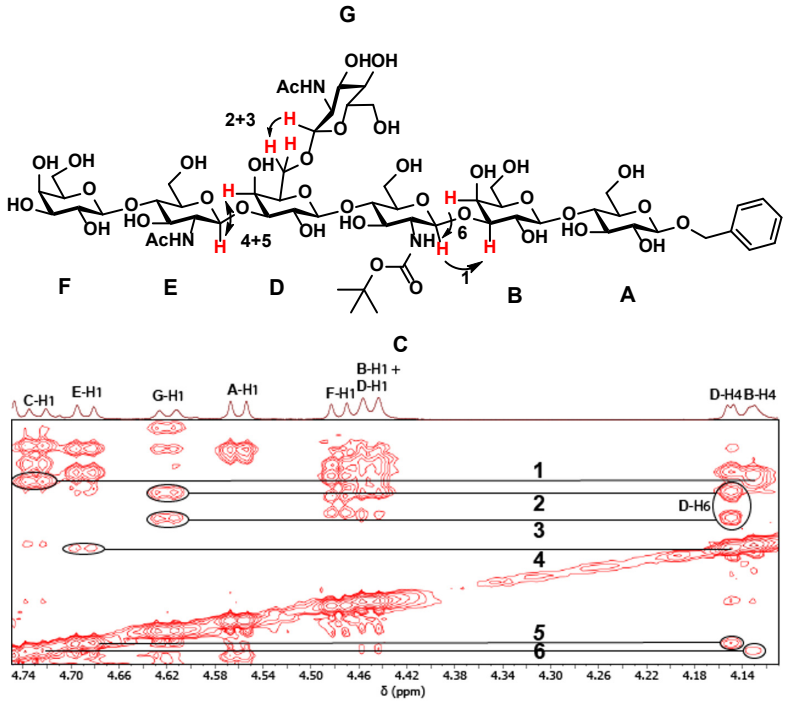


Figure 5: NOESY NMR spectrum of compound **23**; NOESY transfer signals indicative for the structural characterization of **23** are noted by numbers 1-6 and are between the red annotated protons.

N-modified glucosamines only block the neighboring galactoside

To act as a selective blocking group and control I-branching on longer polyLacNAc chains, it is necessary that GlcNBoc does not significantly influence GCNT2 activity on other I-branching sites. To investigate this property, hexasaccharide **25** was prepared that places the GlcNBoc moiety on the terminal position (Fig. 6). The later compound was prepared by treatment of **2** with UDP-GlcNTFA in the presence of Hpβ3GlcNAcT followed by galactosylation of the resulting product with B4GalT1 to give hexasaccharide **24** that was subjected to exchange of TFA by Boc by standard manipulations. Treatment of **25** with UDP-GlcNAc in the presence of GCNT2 resulted only in modification of the galactoside of the lactose moiety to give the expected heptasaccharide **26**. Hydrolysis of the Boc masking group followed by acetylation of the resulting free amine gave target compound **27**. The successful preparation of this derivative supports that GlcNBoc only affects the neighboring reducing end galactoside and does not influence other I-branching sites. Because only one acceptor site is blocked by GlcNBoc, it is expected that this synthetic methodology is suitable to control I-branching on larger polyLacNAc chains.

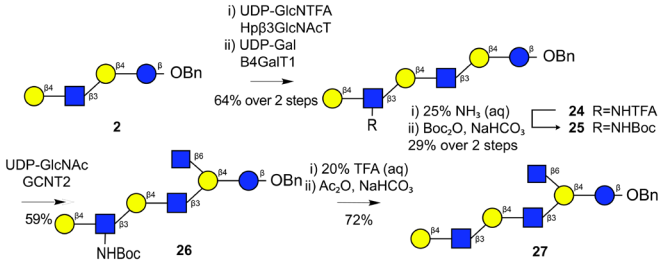


Figure 6: Chemo-enzymatic synthesis of **27** shows that Boc only blocks the reducing-end neighboring galactose.

Reversal of I-branch selectivity by a fucosyl blocking group

It is suggested that α1,3 fucosylated chains are restricted towards I-branching by GCNT2.[33] Therefore, we explored whether α1,3-fucosides can be exploited as a potential I-branch directing group. Moreover, GlcNBoc is known to block fucosylation, and can direct α1,3-fucosyl transferase activity to an unmodified LacNAc residue (*e.g.* **22**→**28**, Fig. 7A).[31] Thus, subsequent removal of a Boc masking group and acetylation of the resulting amine was expected to give a precursor (*e.g.* **29**) that allows regioselective installation of an I-branch at an alternative site.[31]

To investigate the proposed strategy, hexasaccharide **22** was selectively fucosylated by treatment with α1,3-fucosyltransferase 9 (FUT9) and GDP-fucose to form compound **28**. The Boc protecting group prevented FUT9 activity on the inner LacNBoc, whilst activity on lactose was not observed when GDP-fucose was limited to 1.1 equivalents. The latter regioselectivity is explained by a lower activity of FUT9 for lactose aided by a preference of this enzyme for distal LacNAc moieties.[34] After fucosylation, it was necessary to reactivate the I-branch position blocked by GlcNBoc by removal of Boc followed by *N*-acetylation of the resulting amine using Ac_2O and sodium bicarbonate to give compound **29**. It was observed that the use of 20% aqueous TFA resulted in the hydrolysis of Boc as well as the fucoside, however, selective Boc removal could be accomplished by lowering the TFA concentration to 2%. Exposure of **29** to the I-branching enzyme GCNT2 in the presence of UDP-GlcNAc resulted in the selective modification of the proximal lactose moiety to give compound **30** demonstrating that the α1,3-fucoside of **29** directs the regioselectivity of I-branching which was confirmed by detailed NMR experiments (Fig. 7B). Unsubstituted polyLacNAc **27** could be generated by hydrolysis of the fucoside of **30** by exposure to 20% aqueous TFA for 8 hours to give clean formation of heptasaccharides **27**. Regio-isomer **23** could readily be prepared by exposure of **22** to GCNT2 in the presence of UDP-GlcNAc highlighting that the combined use of GlcNBoc and fucoside can provide isomeric I-branched structures from a single structure.

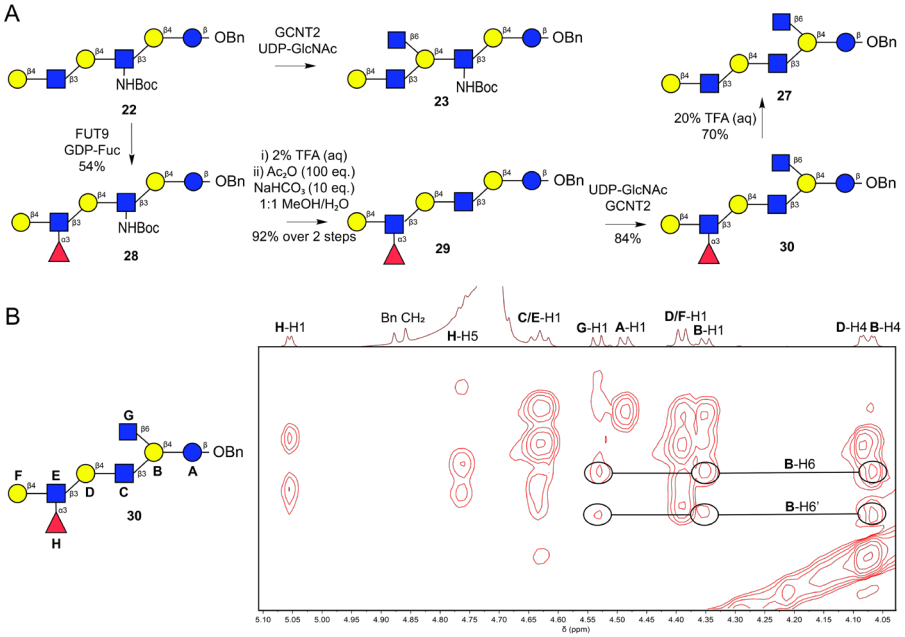


Figure 7: A) Chemo-enzymatic synthesis of **27** using α1,3-fucose as a traceless blocking group. B) NOESY NMR spectrum of compound **30**. NOESY correlation observed between **G**-H1 and **B**-H6/H6' are indicative of a proximal

Selective enzymatic extensions of I-branches

To selectively extend the I-branches of compound **4** with several LacNAc moieties, it was necessary to block the terminal LacNAc moiety of the main chain from enzymatic modification. This was achieved by α2,3-sialylation because it occupies the C3 position on the galactoside blocking further extension (Fig. 8A). Furthermore, sialosides can readily be removed under mild acid conditions without affecting other moieties. [35,36] As expected, ST3Gal4 could install an α2,3-sialoside on the terminus of the main chain of glycan **4**, and subsequent treatment with B4GalT1 in the presence of UDP-Gal resulted in extension of the I-branches to give compound **31**. The terminal LacNAc moieties of **31** could be further extended by alternating action of B3GNT2 and B4GalT1 resulted in the formation of compound **32**. Repeating the enzymatic module gave compound **33** having the I-branches extended by three LacNAc repeating units. Treatment of compounds **31**-**33** with 2N AcOH (aq) at 60 ºC for 18 h resulted in the hydrolysis of the sialosides providing unsubstituted polyLacNAc derivatives **34**-**36**.

Selective extension of an I-branch could also be performed on compound **23** which has a GlcNBoc moiety to give access to derivatives **38-40** (Fig. 8B). Thus, treatment of **23** with ST3Gal4 in the presence of CMP-Neu5Ac resulted in the clean formation of sialoside **37**. The terminal GlcNAc moiety of the latter compound could be converted into LacNAc by treatment with B4GalT1 to give **38** which was followed additional enzymatic treatment with B3GNT2 and B4GalT1 to install additional LacNAc moieties. The Boc protecting groups could readily be cleaved using 20% aqueous TFA at room temperature for 1 h to give amines that were acetylated under standard conditions. Interestingly, the sialosides were stable under these conditions, however, could readily be removed by heating in aqueous acetic acid to provide target compounds **38-40**. Collectively, the results show that the I-branches can be selectively modified without affecting the main chain structure making it possible to prepare multiple glycans from a single I-branched precursor.

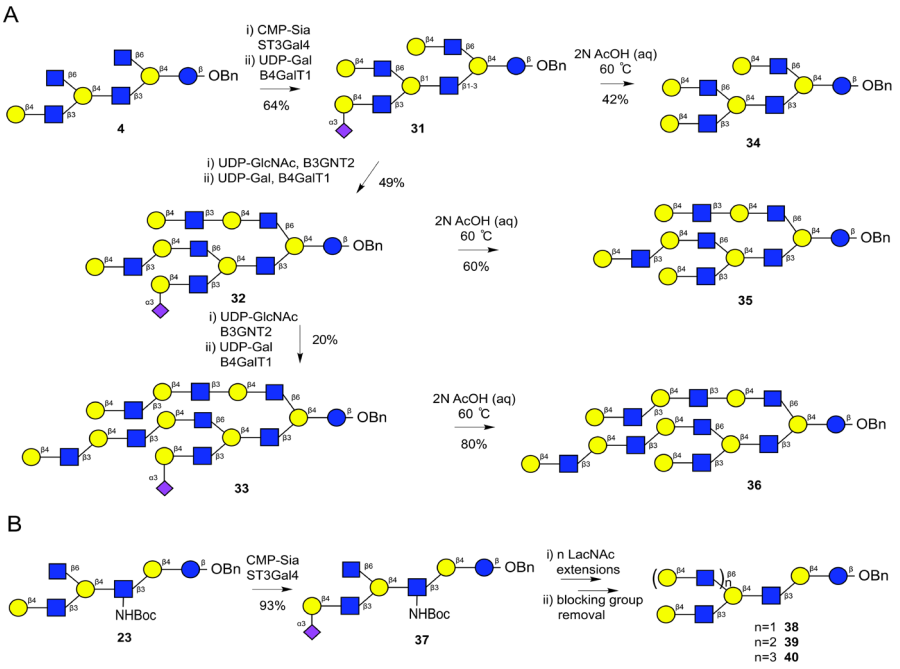


Figure 8: Sialic acid is used as a temporary protecting group on the main chain of I-branched oligosaccharides to allow enzymatic extension of I-branches on oligosaccharides. A) Symmetric I-branched polyLacNAc pathway to compounds **34**, **35** and **36**; B) Pathway to compounds **38**, **39** and **40**.

Conclusions

A chemo-enzymatic methodology is described that can exert regioselective control over I-branching enzyme activity on polyLacNAcs scaffolds using traceless blocking groups. Glucosamine residues modified by *tert*-butyloxycarbamate and α1,3 fucose were found to function as temporal blocking groups for I-branching. A library of well-defined I-branched glycans with complementary polyLacNAc topologies could be prepared by this methodology. It is anticipated that I-branched polyLacNAcs made using this methodology can be derivatized with glycan epitopes such as blood group antigens to study the effects of multivalent glycan presentation. We also expect that this methodology can also be incorporated in the synthesis of complex glycans, such as I-branched *N*-glycans, *O*-glycans and glycolipids. Such glycans may provide useful reference glycans that could be used to develop analytical methodology is able to resolve polyLacNAc topology.

References

[1] L. Cooling, Immunohematology 2020, 26, 133–155.
[2] T. R. Henion, G. A. Schwarting, J Cell Physiol 2014, 229, 471–478.
[3] E. Zdebska, R. Krauze, J. Kościelak, Carbohydrate Research 1983, 120, 113–130.
[4] J. Mikami, Y. Tobisawa, T. Yoneyama, S. Hatakeyama, K. Mori, Y. Hashimoto, T. Koie, C. Ohyama, M. Fukuda, Cancer Sci 2016, 107, 359–368.
[5] F. G. Hanisch, G. Uhlenbruck, J. Peter-Katalinic, H. Egge, J. Dabrowski, U. Dabrowski, Journal of Biological Chemistry 1989, 264, 872–883.
[6] G. P. Reddy, C. Allen Bush, Analytical Biochemistry 1991, 198, 278–284.
[7] W. L. Marsh, Br J Haematol 1961, 7, 200–209.
[8] W. J. Jenkins, W. L. Marsh, J. Noades, P. Tippett, R. Sanger, R. R. Race, Vox Sanguinis 1960, 5, 97–106.
[9] F. J. Grundbacher, Nature 1964, 204, 192–194.
[10] M. Fukuda, A. Dell, M. N. Fukuda, J Biol Chem 1984, 259, 4782–4791.
[11] H. Yamaguchi, Y. Okubo, Proceedings of the Japan Academy 1972, 48, 625–628.
[12] G. Borck, N. Kakar, J. Hoch, K. Friedrich, J. Freudenberg, G. Nürnberg, R. Yilmaz, S. Daud, D. M. Baloch, P. Nürnberg, J. Oldenburg, J. Ahmad, C. Kubisch, Hum Genet 2012, 131, 209–216.
[13] L. C. Yu, Y. C. Twu, C. Y. Chang, M. Lin, Blood 2001, 98, 3840–3845.
[14] B. Irum, S. Y. Khan, M. Ali, M. Daud, F. Kabir, B. Rauf, F. Fatima, H. Iqbal, A. O. Khan, S. Al Obaisi, M. A. Naeem, I. A. Nasir, S. N. Khan, T. Husnain, S. Riazuddin, J. Akram, A. O. Eghrari, S. A. Riazuddin, PLoS One 2016, 11, e0167562.
[15] E. Pras, J. Raz, V. Yahalom, M. Frydman, H. J. Garzozi, E. Pras, J. F. Hejtmancik, Invest Ophthalmol Vis Sci 2004, 45, 1940–1945.
[16] H. Zhang, F. Meng, S. Wu, B. Kreike, S. Sethi, W. Chen, F. R. Miller, G. Wu, Cancer Res 2011, 71, 4846–4856.
[17] J. G. Sweeney, J. Liang, A. Antonopoulos, N. Giovannone, S. Kang, T. S. Mondala, S. R. Head, S. L. King, Y. Tani, D. Brackett, A. Dell, G. F. Murphy, S. M. Haslam, H. R. Widlund, C. J. Dimitroff, Nat Commun 2018, 9, 3368.
[18] M. Nonaka, M. N. Fukuda, C. Gao, Z. Li, H. Zhang, M. I. Greene, D. M. Peehl, T. Feizi, M. Fukuda, Journal of Biological Chemistry 2014, 289, 16478–16486.
[19] C. J. Dimitroff, PNAS 2019, 116, 13729–13737.
[20] C.-C. Chao, P.-H. Wu, H.-C. Huang, H.-Y. Chung, Y.-C. Chou, B.-H. Cai, R. Kannagi, FEBS Lett 2017, 591, 1902–1917.
[21] K. Nakamura, K. Yamashita, H. Sawaki, M. Waraya, H. Katoh, N. Nakayama, H. Kawamata, H. Nishimiya, A. Ema, H. Narimatsu, M. Watanabe, Anticancer Res 2015, 35, 1411–1421.
[22] A. R. Prudden, L. Liu, C. J. Capicciotti, M. A. Wolfert, S. Wang, Z. Gao, L. Meng, K. W. Moremen, G.-J. Boons, PNAS 2017, 114, 6954–6959.
[23] N. Giovannone, J. Liang, A. Antonopoulos, J. Geddes Sweeney, S. L. King, S. M. Pochebit, N. Bhattacharyya, G. S. Lee, A. Dell, H. R. Widlund, S. M. Haslam, C. J. Dimitroff, Nat Commun 2018, 9, 3287.

[24] A. Cao, N. Alluqmani, F. H. M. Buhari, L. Wasim, L. K. Smith, A. T. Quaile, M. Shannon, Z. Hakim, H. Furmli, D. M. Owen, A. Savchenko, B. Treanor, Nat Commun 2018, 9, 3288.
[25] D. H. van den Eijnden, A. H. Koenderman, W. E. Schiphorst, Journal of Biological Chemistry 1988, 263, 12461–12471.
[26] K. Sasaki, K. Kurata-Miura, M. Ujita, K. Angata, S. Nakagawa, S. Sekine, T. Nishi, M. Fukuda, Proc Natl Acad Sci U S A 1997, 94, 14294–14299.
[27] W. Peng, J. Pranskevich, C. Nycholat, M. Gilbert, W. Wakarchuk, J. C. Paulson, N. Razi, Glycobiology 2012, 22, 1453–1464.
[28] D. Laaf, H. Steffens, H. Pelantová, P. Bojarová, V. Křen, L. Elling, Advanced Synthesis & Catalysis 2017, 359, 4015–4024.
[29] H. Karlsson, L. Johansson, H. Miller-Podraza, K.-A. Karlsson, Glycobiology 1999, 9, 765–778.
[30] L. Liu, A. R. Prudden, C. J. Capicciotti, G. P. Bosman, J.-Y. Yang, D. G. Chapla, K. W. Moremen, G.-J. Boons, Nature Chemistry 2019, 11, 161–169.
[31] I. A. Gagarinov, T. Li, N. Wei, J. Sastre Toraño, R. P. de Vries, M. A. Wolfert, G.-J. Boons, Angewandte Chemie International Edition 2019, 58, 10547–10552.
[32] Y. Li, M. Xue, X. Sheng, H. Yu, J. Zeng, V. Thon, Y. Chen, M. M. Muthanaa, P. G. Wang, X. Chen, Bioorg Med Chem 2016, 24, 1696–1705.
[33] A. Leppänen, R. Niemelä, O. Renkonen, Biochemistry 1997, 36, 13729–13735.
[34] S. Nishihara, H. Iwasaki, M. Kaneko, A. Tawada, M. Ito, H. Narimatsu, FEBS Letters 1999, 462, 289–294.
[35] W. Zhu, X. Chen, L. Yuan, J. Wu, J. Yao, Molecules 2020, 25, 5141.
[36] H.-P. Buscher, J. Casals-Stenzel, R. Schauer, European Journal of Biochemistry 1974, 50, 71–82.

Supplementary information

Materials and methods

^{1}H NMR are recorded on a 600 MHz Avance Neo or 400 MHz Agilent NMR. Chemical shifts are recorded in parts per million (ppm) relative to TMS or the residual solvent peak. NMR signals are represented as: chemical shift, multiplicity (s = singlet, d = doublet, app.t. = apparent triplet, t = triplet, dd = doublet of doublets, m = multiplet), J coupling and integration. Assignment of ^{1}H NMR signals was conducted based on ^{1}H-NMR, 2D COSY (COSYGPSW), 2D TOCSY (MLEVPHSW), 2D NOESY (NOESYPHSW) and 2D HSQC (HSQCEDETGPSISP). ^{13}C NMR signal assignment was extracted from 2D HSQC spectra. Mass spectra are recorded by ESI on a Bruker micrOTOF-QII system as m/z. Reaction progress for unprotected carbohydrates was determined with a SeQuant Zic HILIC guard column (20x2.1 mm) coupled to a Bruker micrOTOF-QII MS system running a gradient of 90% Acetonitrile:H_2O to 50% Acetonitrile:H_20 over 5 minutes followed by isocratic 50% Acetonitrile:H_2O until product was detected. MilliQ water was purified by a Millipore Synergy water purification system. Size exclusion was conducted using Bio-Gel P2, P4 or P6 gel polyacrylamide beads from BioRad, eluding with MilliQ water or 20 mM Ammonium Bicarbonate for compounds containing a primary amine. Collumn chromatography was performed on silica gel G60. Thin Layer Chromatography (TLC) was conducted on Silicagel 60 F254 (EMD Chemicals Inc.) on prefabricated glass slides with visualization by UV and, spraying with 10% H_2SO_4 in ethanol, p-Anisaldehyde stain or Hennesian's stain and subsequent charring of the TLC plate. UDP-Galactose, UDP-*N*-Acetylglucosamine and CMP-Neu5Ac were bought from Roche Diagnostics. Calf intestine alkaline phosphatase(CIAP) was obtained from Invitrogen. GDP-Fucose was prepared using L-fucokinase/GDP-fucose pyrophosphorylase.[1] Glycosyltransferases were expressed and purified according to published protocols.[2,3] Reagents purchased from commercial sources were used without further purification. UDP-GlcNTFA was synthesized following the protocol described in chapter 2.

General procedure for the transfer of β1,3-*N*-Acetylglucosamine with β_3GNT2

To a solution of 10 mM acceptor oligosaccharide in HEPES buffer (HEPES 50 mM, KCl 25 mM, $MgCl_2$ 2 mM, DTT 1 mM, pH 7.3) was added UDP-*N*-acetylglucosamine (1.5 eq.), CIAP (1 U/ µL) and β_3GNT2 (10 µg/µmol). The reaction was incubated at 37ºC for 16 h with gentle shaking. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was lyophilized if no more starting material was observed otherwise 5 µg/µmol β_3GNT2 was added. The residue was re-dissolved in minimal MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

General procedure for the transfer of β1,3 *N*-trifluoroacetylglucosamine with Hpβ3GlcNAcT

To a solution of 10 mM acceptor oligosaccharide in TRIS buffer (TRIS 100 mM, $MnCl_2$ 10 mM, pH 8) was added UDP-*N*-trifluoroacetylglucosamine (1.5 eq.), CIAP (1 U/ µL) and Hpβ3GlcNAcT (10 µg/µmol). The reaction was incubated at 37ºC for 16 h with gentle shaking. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was lyophilized if no more starting material was observed otherwise 5 µg/µmol Hpβ1,3GlcNAcT was added. The residue was re-dissolved in minimal MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

General procedure for the transfer of β1,6-*N*-Acetylglucosamine with GCNT2

To a solution of 10 mM acceptor oligosaccharide in TRIS buffer (TRIS 100 mM, $MnCl_2$ 10 mM, pH 8) was added UDP-*N*-acetylglucosamine (1.5 eq.), CIAP (1 U/ µL) and GCNT2 (10 µg/µmol). The reaction was incubated at 37ºC for 16 h with gentle shaking. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was lyophilized if no more starting material was observed otherwise 5 µg/µmol GcNT2 was added. The residue was re-dissolved in minimal MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

General procedure for the transfer of β1,4-galactoside with B4GalT1

To a solution of 10 mM acceptor oligosaccharide in TRIS buffer (TRIS 50 mM, $MnCl_2$ 10 mM, BSA 0.1 wt%, pH 7.3) was added UDP-galactose (1.5 eq.), CIAP (1 U/ µL) and B4GalT1 (10 µg/µmol). The reaction was incubated at 37ºC for 16 h with gentle shaking. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was lyophilized if no more starting material was observed otherwise 5 µg/µmol B4GalT1 was added. The residue was re-dissolved in minimal MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

General procedure for the transfer of β1,4-galactoside with B4GalT4

To a solution of 10 mM acceptor oligosaccharide in TRIS buffer (TRIS 50 mM, $MnCl_2$ 10 mM, BSA 0.1 wt%, pH 7.3) was added UDP-galactose (1.5 eq.), CIAP (1 U/ µL) and B4GalT4 (10 µg/µmol). The reaction was incubated at 37ºC for 16 h with gentle shaking. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was lyophilized if no more starting material was observed otherwise 5 µg/µmol B4GalT1 was added. The residue was re-dissolved in minimal MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

General procedure for the transfer of α2,3-*N*-acetylneuraminic acid with ST3Gal4

To a solution of 10 mM acceptor oligosaccharide in TRIS buffer (TRIS 50 mM, BSA 0.1 wt%, pH 7.3) was added CMP-Neu5Ac (1.5 eq.), CIAP (1 U/ µL) and ST3Gal4 (10 µg/µmol). The reaction was incubated at 37ºC for 16 h with gentle shaking. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was lyophilized if no more starting material was observed otherwise 5 µg/µmol ST3Gal4 was added. The residue was re-dissolved in minimal MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

General procedure for *N*-acetylneuraminic acid hydrolysis

Sialoside was dissolved in 2N AcOH (aq) (200 µL/mg carbohydrate) and the solution was incubated for 16 h at 60 ºC. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was lyophilized when no more starting material was observed. The residue was re-dissolved in minimal MilliQ water and loaded on a Biogel size exclusion column loaded with a resin of the appropriate pore size. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

Procedure for removal of *N*-trifluoroacetamide group

Oligosaccharide was dissolved in a 25 wt% aqueous NH_3 solution and the solution was incubated for 3 h at 37 ºC. Reaction progress was monitored by LC-ESI-Q-TOF-MS. NH_3 was removed under N_2 stream and the reaction mixture was lyophilized when no more starting material was observed. The residue was re-dissolved in minimal MilliQ water and loaded on a P2 Biogel size exclusion column and eluded with 10 mM NH_4HCO_3(aq). Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

Procedure for removal of *N*-tert-butyloxycarbonyl group

Oligosaccharide was dissolved in a 2 v/v% aqueous TFA solution and the solution was incubated for 4 h at 37 ºC. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was diluted with MiliQ water and lyophilized when no more starting material was observed. If no fucose was present on the starting material 20 v/v% aqueous TFA was used and the reaction was completed in 1 h. The residue was re-dissolved in minimal MilliQ water and loaded on a P2 Biogel size exclusion column MilliQ water. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

Procedure for fucose hydrolysis

Fucoside was dissolved in a 20 v/v% aqueous TFA solution and the solution was incubated for 8 h at 37 ºC. Reaction progress was monitored by LC-ESI-Q-TOF-MS. The reaction mixture was diluted with MiliQ water and lyophilized when no more starting material was observed. The residue was re-dissolved in minimal MilliQ water and loaded on a P2 Biogel size exclusion column MilliQ water. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

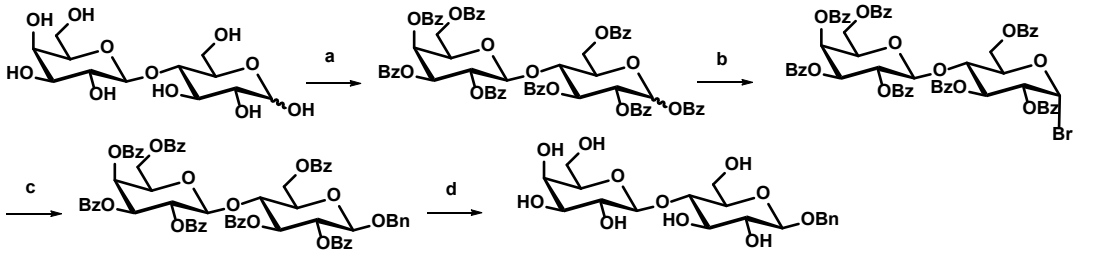
Procedure for installation of *N*-tert-butyloxycarbonyl

Oligosaccharide was dissolved in 1:1 MeOH:H_2O (100 µL/mg carbohydrate) and $NaHCO_3$ (10 eq.) and Boc anhydride (100 eq.) were added. The solution was sonicated for 1 h at 50 ºC, reaction progress was monitored by LC-ESI-Q-TOF-MS. If starting material was observed another 10 eq. of $NaHCO_3$ and Boc anhydride were added. The reaction mixture was lyophilized when no more starting material was observed. The residue was re-dissolved in minimal MilliQ water and loaded on a P2 Biogel size exclusion column. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

Procedure for selective installation of *N*-acetamide

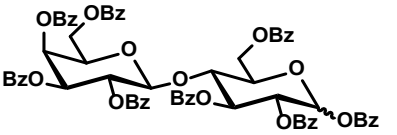
Oligosaccharide was dissolved in 1:1 MeOH:H_2O (40 µL/mg carbohydrate) and acetic anhydride (20 µL/mg carbohydrate) was added. The solution was sonicated for 1 h at 37 ºC, reaction progress was monitored by LC-ESI-Q-TOF-MS. If starting material was observed another 20 µL/mg carbohydrate of H_2O and acetic anhydride were added. The reaction mixture was lyophilized when no more starting material was observed. The residue was re-dissolved in minimal MilliQ water and loaded on a P2 Biogel size exclusion column. Carbohydrate containing fractions were identified by LC-ESI-Q-TOF-MS, pooled and lyophilized.

Synthesis of (β)-O-benzyllactose



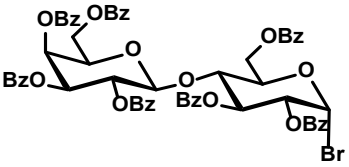
a) BzCl (10 eq), Py; b) HBr (33% in AcOH); c) BnOH (2 eq), AgOTf (1.2 eq), Tol, Ar, - 78 °C to r.t.; d) NaOMe(cat.) in MeOH;

1,2,3,6,2',3'4',6'-Octa-*O*-benzoyllactose



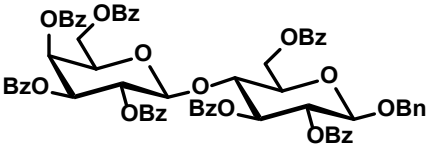
To a vigorously stirred suspension of lactose (14.6 mmol, 5 g) in dry pyridine cooled on ice was added benzoyl chloride (146 mmol, 17 mL) in small portions. The reaction was stirred for 20 min at 0 ºC until TLC analysis showed completion of the reaction. The reaction mixture was concentrated and dissolved in DCM (50 mL) and washed with 1M HCl (aq, 50 mL) and brine (50 mL). The organic phase was dried ($NaSO_4$), filtrated and the filtrate was concentrated *in vacuo* (EtOAc:petrol = 1:1, R_f = 0.66, white foam). Material is used without further purification.

2,3,6,2',3'4',6'-Hepta-*O*-benzoyl-α-bromolactose



Crude Octa-*O*-benzoyllactose was dissolved in DCM (50 mL) and a 33% solution of HBr in AcOH (18 mL) was added. The reaction mixture was stirred for 4 h until TLC analysis showed complete consumption of the starting material. The mixture was diluted with DCM (200 mL) and washed with H_2O (100 mL), sat.aq. bicarbonate (100 mL) and brine (100 mL). The organic phase was dried ($NaSO_4$), filtrated and the filtrate was concentrated *in vacuo* (EtOAc:petrol = 1:1, R_f = 0.43, white foam). Material was used without further purification.

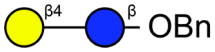
2,3,6,2',3'4',6'-Hepta-*O*-benzoyl-β-benzyllactose



Crude hepta-*O*-benzoyl-α-bromolactose (8.65 mmol, 9.8 g) was dissolved in dry toluene and benzyl alcohol (17.3 mmol, 1.83 mL) was added. The solution was cooled (-78 ºC) and placed under an argon atmosphere. AgOTf (10.38 mmol, 2.67 g) was added, and the stirred solution was allowed to warm to room temperature. The reaction mixture was concentrated, and the resulting residue was diluted with DCM (100 mL). The resulting solution was washed

with H_2O (100 mL) and filtered. The filtrate was washed with H_2O (100 mL) and brine (100 mL). The organic phase was dried ($NaSO_4$), filtrated and the filtrate was concentrated *in vacuo* (EtOAc:petrol = 1:2, R_f = 0.2, white foam). ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.09 – 7.10 (m, 56H, Ar, Bn; Ar, Bz; Ar, residual BnOH, 5.78 – 5.66 (m, 3H, H-3, H-2',H-3'), 5.53 (dd, *J* = 9.8, 7.9 Hz, 1H, H-2), 5.36 (dd, *J* = 10.3, 3.4 Hz, 1H, H-4'), 4.87 – 4.81 (m, 2H, H-1',H-6'a), 4.72 – 4.68 (m, 6H, H-1; CH_2, Bn; CH_2, residual BnOH, 4.60 – 4.49 (m, 3H, H6-ab, H6'-b), 4.25 (app.t, *J* = 9.5 Hz, 1H, H-4), 3.89 – 3.83 (m, 1H, H-5'), 3.81 – 3.75 (m, 1H, H-5). ^{13}C NMR (150 MHz, $CDCl_3$) δ 140.86, 136.43, 133.49, 133.36, 133.22, 133.15, 133.11, 129.96, 129.85, 129.74, 129.71, 129.65, 129.57, 129.52, 128.66, 128.61, 128.53, 128.49, 128.30,128.28, 128.22, 128.19, 127.87, 127.76, 127.61, 126.95, 100.97, 99.08, 76.03, 73.00, 72.90, 71.76, 71.69, 71.36, 70.45, 69.88, 67.50, 65.35, 62.39, 61.05.

(β)-O-Benzyllactose (1)



1

Crude Hepta-O-benzoyl-(β)-benzyllactose (7.1 g, 6.36 mmol) is dissolved in MeOH with 5% DCM (50 mL) and a catalytic amount of 0.5N NaOMe in MeOH is added to the stirred solution (pH>9). After 3h TLC indicated consumption of the starting material. The solution was neutralized to pH 6 with MeOH washed Amberlyst 15 hydrogen form resin. The resin was filtered off and the filtrate was concentrated. The crude material was dissolved in a small amount of MeOH and precipitated in icecold Et_2O. Solids were collected by vacuum and dried by air flow (EtOAc:MeOH:H2O = 7:2:1, R_f = 0.28, white semi-crystalline powder, 1.41 g, 22.3% yield over 4 steps)

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6
Glc	4.56 (d, *J* = 8.0 Hz, 1H)	3.37	3.67	3.67	3.62	3.83; 4.01
Gal	4.44 (d, *J* = 7.9 Hz, 1H)	3.56	3.73	3.94 (d, *J* = 3.2 Hz, 1H)	3.62	3.77

Bn (CH_2)	4.96 (d, *J* = 11.6 Hz, 1H); 4.79 (d, *J* = 11.0 Hz, 1H)
Bn (C_6H_5)	7.54 – 7.41 (m, 5H)

^{13}C (150 MHz, D_2O) derived from HSQC: δ (ppm)

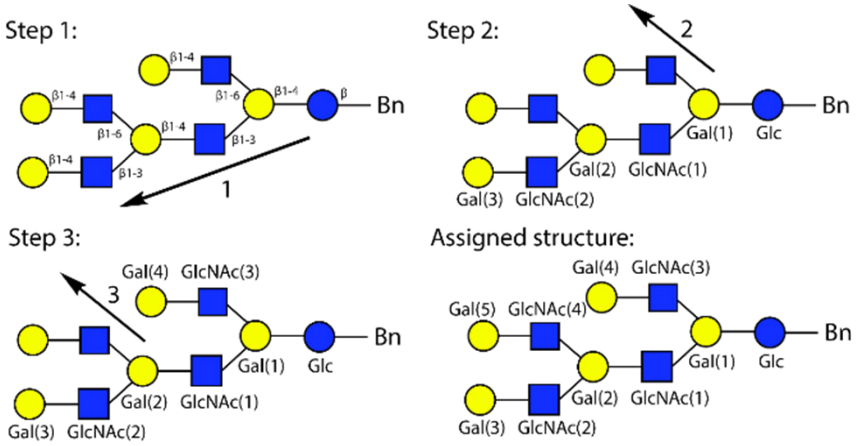
	C1	C2	C3	C4	C5	C6
Glc	101.02	72.80	72.40	78.33	74.67	60.01
Gal	103.08	70.92	75.32	68.37	74.67	60.93

Bn (CH_2)	71.48
Bn (C_6H_5)	128.60

ESI TOF-MS m/z calculated for $C_{19}H_{28}O_{11}Na$ $[M + Na]^+$ = 455.1529, found 455.1584

NMR assignment nomenclature

Carbohydrates are referred to as the abbreviation of the monosaccharide composition combined with a rank number (e.g. Gal(2)), no rank is given when only one instance is present on the glycan. Ranks are assigned starting from the reducing end glucose to the non-reducing end, following the β1,3 GlcNAc on I-branching points. The I-branches are ranked after the main chain, starting at the first β1,6 GlcNAc from the reducing end. The complete I-branch is assigned before following the previous assigned main chain towards the next I-branch sites (see below).



NMR & MS data

2

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6
Glc	4.56 (d, *J* = 8.0 Hz, 1H)	3.32	3.60	3.69	n/a	n/a
Gal	4.44 (d, *J* = 7.9 Hz, 1H)	3.55	3.70	4.16 (d, *J* = 3.3 Hz, 1H)	n/a	n/a
GlcNAc	4.71 (d, *J* = 8.4 Hz, 1H)	3.78	3.56	3.69	n/a	2.04 (s, 3H)
Gal2	4.49 (d, *J* = 7.8 Hz, 1H)	3.45	3.84	3.93 (d, *J* = 3.4 Hz, 1H)	n/a	n/a

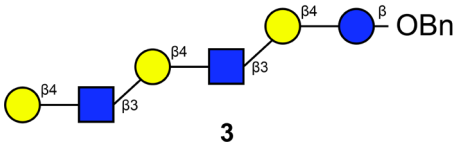
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.6 Hz, 1H)	7.53 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)
C1	101.15	102.68	102.75	102.68

ESI TOF-MS m/z calculated for $C_{33}H_{52}NO_{21}$ $[M + H]^+$ = 798.3032, found 798.3039

Yield: 49.1% over 2 steps, 19.4 μmol, 15.5 mg



3

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.55 (d, *J* = 8.0 Hz, 1H)	3.35	3.61	n/a	n/a	n/a	-
Gal (1)	4.43 (d, *J* = 7.9 Hz, 1H)	3.57	3.72	4.14 (app t, *J* = 3.0 Hz, 2H)	n/a	n/a	-
GlcNAc(1)	4.69 (d, *J* = 8.2 Hz, 2H)	3.79	n/a	n/a	n/a	n/a	-
Gal (2)	4.49 (d, *J* = 8.2 Hz, 1H)	3.68	3.72	4.14 (app t, *J* = 3.0 Hz, 2H)	n/a	n/a	-
GlcNAc (2)	4.69 (d, *J* = 8.2 Hz, 2H	3.79	n/a	n/a	n/a	n/a	2.02 (s, 3H)
Gal (3)	4.47 (d, *J* = 8.1 Hz, 1H)	3.54	3.78	3.93	n/a	n/a	-

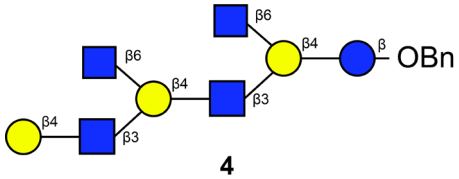
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.6 Hz, 1H)	7.53 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)
C1	101.19	102.89	102.75	102.89	102.75	102.89

ESI TOF-MS m/z calculated for $C_{47}H_{75}N_2O_{31}$ $[M + H]^+$ = 1163.4354, found 1163.4818

Yield: 89.7% over 2 steps, 17.5 µmol, 20.3 mg



4

^{1}H (400 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.55	3.36	3.61	3.68	n/a	3.79, 3.96	-
Gal (1)	4.42 (d, *J* = 7.5 Hz, 1H)	3.56	3.69	4.11 (d, *J* = 3.3 Hz, 1H)	n/a	3.75, 3.90	-
GlcNAc (1)	4.68	3.78	3.71	3.55	n/a	3.76, 3.87	2.06 – 1.97 (m, 12H)
Gal (2)	4.45 4.38 (d, *J* = 7.6 Hz, 1H)	3.53	3.68	4.13 (d, *J* = 3.3 Hz, 1H)	n/a	3.75, 3.90	-
GlcNAc (2)	4.68	3.78	3.71	3.55	n/a	3.76, 3.87	2.06 – 1.97 (m, 12H)
Gal (3)	4.45 (d, *J* = 7.6 Hz, 2H)	3.53	3.64	3.90	n/a	3.66, 3.83	-

GlcNAc (3)	4.59 (d, J = 8.3 Hz, 2H)	3.72	3.45	3.58	n/a	3.76, 3.87	2.06 – 1.97 (m, 12H)
GlcNAc (4)	4.59 (d, J = 8.3 Hz, 2H)	3.70	3.45	3.57	n/a	3.76, 3.87	2.06 – 1.97 (m, 12H)

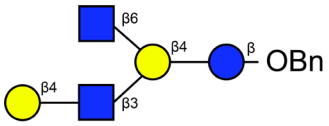
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.6 Hz, 1H)	7.48 – 7.34 (m, 5H)

^{13}C from HSQC (100 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)	GlcNAc (4)
C1	101.13	102.98	102.59	102.98	102.59	102.98	101.04	101.04

ESI TOF-MS m/z calculated for $C_{63}H_{100}N_4O_{41}Na$ $[M + Na]^+$ = 1591.5761, found 1591.6441

Yield: 72.1%, 6.2 µmol, 9.8 mg



5

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.54 (d, J = 8.0 Hz, 1H)	3.33	3.59	n/a	n/a	3.79, 3.96	-
Gal	4.40 (d, J = 7.9 Hz, 1H)	3.55	3.68	4.12 (d, J = 3.2 Hz, 1H)	n/a	3.72, 3.88	-
GlcNAc	4.68 (d, J = 8.3 Hz, 1H)	3.78	3.56	3.72	n/a	n/a	2.05 – 2.00 (s, 6H)
Gal2	4.46 (d, J = 7.8 Hz, 1H)	3.52	3.66	3.91 (d, J = 3.8 Hz, 1H)	n/a	n/a	-
GlcNAc(2)	4.59 (d, J = 8.5 Hz, 1H)	3.67	3.53	3.45	n/a	n/a	2.05 – 2.00 (s, 6H)

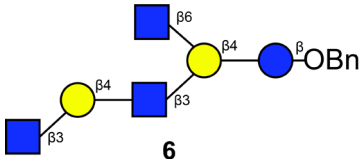
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.75 (d, J = 11.8 Hz, 1H)	7.49 – 7.36 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)
C1	100.95	102.97	102.76	102.84	101.02

ESI TOF-MS m/z calculated for $C_{41}H_{65}N_2O_{26}$ $[M + H]^+$ = 1001.3826, found 1001.3737

Yield: 93.1%, 17.5 µmol, 17.6 mg



6

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.54 (d, *J* = 8.0 Hz, 1H)	3.33	3.52	3.78	n/a	n/a	-
Gal	4.40 (d, *J* = 7.9 Hz, 1H)	3.55	3.68	4.12 (d, *J* = 3.5 Hz, 1H)	n/a	3.80, 3.96	-
GlcNAc	4.67 (d, *J* = 8.5 Hz, 1H)	3.78	3.55	3.70	n/a	n/a	2.11 – 1.97 (s, 9H)
Gal (2)	4.44 (d, *J* = 7.9 Hz, 1H)	3.57	3.71	4.13 (d, *J* = 3.5 Hz, 1H)	n/a	n/a	-
GlcNAc (2)	4.66 (d, *J* = 8.5 Hz, 1H)	3.74	3.42	3.55	n/a	n/a	2.11 – 1.97 (s, 9H)
GlcNAc (3)	4.58 (d, *J* = 8.5 Hz, 1H)	3.67	3.53	3.44	n/a	n/a	2.11 – 1.97 (s, 9H)

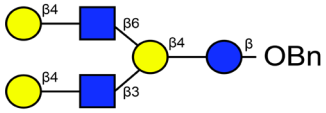
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, *J* = 11.6 Hz, 1H); 4.74 (d, *J* = 11.8 Hz, 1H)	7.49 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	GlcNAc (3)
C1	100.98	102.97	102.77	102.86	102.77	101.05

ESI TOF-MS m/z calculated for $C_{49}H_{77}N_3O_{31}Na$ $[M + Na]^+$ = 1226.4439, found 1226.4378

Yield: 78.6%, 4.3 µmol, 5.3 mg



7

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.52 (d, *J* = 8.0 Hz, 1H)	3.33	3.57	3.78	n/a	3.77, 3.94	-
Gal	4.39 (d, *J* = 7.8 Hz, 1H)	3.53	3.67	4.11 (d, *J* = 3.5 Hz, 1H)	n/a	3.80, 3.95	-
GlcNAc	4.66 (d, *J* = 8.5 Hz, 1H)	3.76	3.55	3.69	n/a	3.80, 3.92	2.04 – 1.97 (s, 6H)
Gal (2)	4.44 (d, *J* = 8.5 Hz, 1H)	3.50	3.63	3.88	n/a	3.72	-
GlcNAc (2)	4.59 (d, *J* = 7.7 Hz, 1H)	3.71	3.57	3.75	n/a	3.79, 3.95	2.04 – 1.97 (s, 6H)
Gal (3)	4.43 (d, *J* = 8.9 Hz, 1H)	3.49	3.62	3.87	n/a	3.72	-

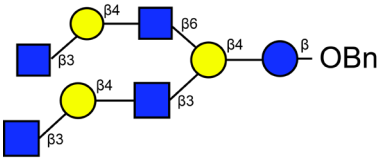
Bn (CH_2)	Bn (C_6H_5)
4.91 (d, *J* = 11.6 Hz, 1H); 4.74 (d, *J* = 11.8 Hz, 1H)	7.41 (app. dq, *J* = 21.7, 7.7, 7.0 Hz, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)
C1	100.96	102.95	102.65	102.83	100.91	102.83

ESI TOF-MS m/z calculated for $C_{47}H_{74}N_2O_{31}Na$ $[M + Na]^+$ = 1185.4173, found 1185.4113

Yield: 93.8%, 5.2 µmol, 6.0 mg



8

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.57 (d, *J* = 8.0 Hz, 1H)	3.37	3.61	n/a	n/a	3.81, 3.98	-
Gal	4.43 (d, *J* = 7.9 Hz, 1H))	3.58	3.71	4.14	n/a	3.82, 3.98	-
GlcNAc	4.70 (d, *J* = 8.6 Hz, 1H)	3.80	3.58	3.73	n/a	3.84, 3.97	2.06 – 2.03 (s, 12H)
Gal (2)	4.47 (d, *J* = 8.2 Hz, 1H)	3.59	3.73	4.16	n/a	3.75	-
GlcNAc (2)	4.69 (d, *J* = 8.4 Hz, 1H)	3.76	3.45	3.57	n/a	3.90	2.06 – 2.03 (s, 12H)
GlcNAc (3)	4.63 (d, *J* = 7.8 Hz, 1H)	3.74	3.61	3.72	n/a	3.83; 3.99	2.06 – 2.03 (s, 12H)
Gal (3)	4.46 (d, *J* = 8.7 Hz, 1H)	3.59	3.73	4.15	n/a	3.75	-
GlcNAc (4)	4.69 (d, *J* = 8.4 Hz, 1H)	3.76	3.45	3.57	n/a	3.90	2.06 – 2.03 (s, 12H)

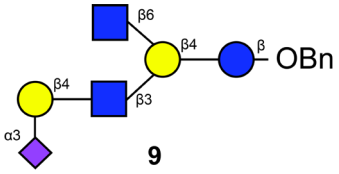
Bn (CH_2)	Bn (C_6H_5)
4.95 (d, *J* = 11.6 Hz, 1H); 4.77 (d, *J* = 11.6 Hz, 1H)	7.50 – 7.40 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	GlcNAc (3)	Gal (3)	GlcNAc (4)
C1	101.00	103.00	102.86	102.95	102.86	100.99	102.95	102.86

ESI TOF-MS m/z calculated for $C_{63}H_{100}N_4O_{41}Na$ $[M + Na]^+$ = 1591.5761, found 1591.5739

Yield: 63.7%, 3.3 µmol, 5.2 mg



9

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	H7	H8	H9	NHAc
Glc	4.54 (d, *J* = 8.0 Hz, 1H)	3.34 (td, *J* = 7.9, 2.4 Hz, 1H)	3.60	3.57	n/a	3.81, 3.96	-	-	-	-
Gal (1)	4.40 (d, *J* = 7.9 Hz, 1H)	3.55	3.68	4.12 (d, *J* = 3.3 Hz, 1H)	3.81	3.83, 3.98	-	-	-	-
GlcNAc (1)	4.67 (d, *J* = 8.3 Hz, 1H)	3.78	3.56	3.71	n/a	3.86, 3.94	-	-	-	2.02 (s, 9H)

Gal (2)	4.54 (d, J = 7.9 Hz, 1H)	3.55	3.94	4.10 (dd, J = 9.9, 3.1 Hz, 1H)	3.55	3.72	-	-	-	-
GlcNAc (2)	4.59 (d, J = 8.5 Hz, 1H)	3.67	3.52	3.44	n/a	3.73, 3.90	-	-	-	2.02 (s, 9H)
Neu5Ac	-	-	-eq 2.74 (dd, J = 12.4, 4.6 Hz, 1H) -axial 1.78 (app.t, J = 12.1 Hz, 1H)	3.66	3.83	3.61	n/a	n/a	3.62, 3.83	2.02 (s, 9H)

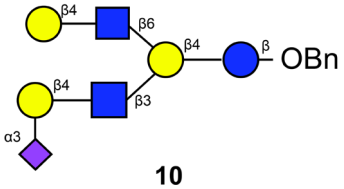
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.7 Hz, 1H)	7.47 – 7.38 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Neu5Ac
C1	100.88	102.98	102.77	102.46	100.97	-
C3	-	-	-	-	-	39.61

ESI TOF-MS m/z calculated for $C_{52}H_{82}N_3O_{34}$ $[M + Na]^+$ = 1292.4780 found 1292.4681

Yield: 86.0%, 4.3 µmol, 5.5 mg



10

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	H7	H8	H9	NHAc
Glc	4.54 (d, J = 8.8 Hz, 1H)	3.34 (app.t, J = 8.3 Hz, 1H)	3.59	3.61	n/a	3.79, 3.96	-	-	-	-
Gal (1)	4.40 (d, J = 7.9 Hz, 1H)	3.55	3.69	4.12 (d, J = 3.2 Hz, 1H)	3.80	3.81, 3.97	-	-	-	-
GlcNAc (1)	4.67 (d, J = 8.4 Hz, 1H)	3.77	3.55	3.71	n/a	n/a	-	-	-	2.02 (s, 9H)
Gal (2)	4.54 (d, J = 8.8 Hz, 1H)	3.55	3.94	4.10 (dd, J = 9.9, 3.1 Hz, 1H)	3.56	n/a	-	-	-	-
GlcNAc (2)	4.61 (d, J = 7.8 Hz, 1H)	3.72	3.59	n/a	n/a	3.81, 3.97	-	-	-	2.02 (s, 9H)

Gal (3)	4.44 (d, J = 8.1 Hz, 1H)	3.52	3.64	3.90	n/a	n/a	-	-	-	-
Neu5Ac	-	-	-eq 2.74 (dd, J = 12.4, 4.6 Hz, 1H) -axial 1.78 (app. t, J = 12.2 Hz, 1H)	3.64	3.82	3.61	n/a	n/a	3.62, 3.83	2.02 (s, 9H)

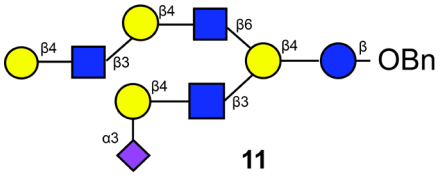
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H), 4.74 (d, J = 11.7 Hz, 1H)	7.48 – 7.36 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	Neu5Ac
C1	100.85	102.96	102.72	102.41	100.95	102.86	-
C3	-	-	-	-	-		39.54

ESI TOF-MS m/z calculated for $C_{58}H_{90}N_3O_{39}$ $[M-H]^-$ = 1452.5152, found 1452.4596

Yield: 75.2%, 3.2 µmol, 4.7 mg



11

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	H7	H8	H9	NHAc
Glc	4.56 (d, J = 8.0 Hz, 1H)	3.36	3.61	3.60	n/a	3.81, 3.98	-	-	-	-
Gal (1)	4.46 (d, J = 7.9 Hz, 1H)	3.58	3.73	4.12 (d, J = 3.3 Hz, 1H)	n/a	n/a	-	-	-	-
GlcNAc (1)	4.70 (d, J = 8.3 Hz, 2H)	3.80	3.58	3.73	n/a	n/a	-	-	-	2.06 – 1.99 (m, 12H)
Gal (2)	4.56 (d, J = 8.0 Hz, 1H)	3.57	3.96	4.12	n/a	3.84, 3.98	-	-	-	-
GlcNAc (2)	4.63 (d, J = 7.8 Hz, 1H)	3.74	3.62	3.69	n/a	3.65, 3.89	-	-	-	2.06 – 1.99 (m, 12H)
Gal (3)	4.43 (d, J = 7.9 Hz, 1H)	3.57	3.71	4.13	n/a	n/a	-	-	-	-
GlcNAc (3)	4.70 (d, J = 8.3 Hz, 2H)	3.80	3.58	-	n/a	n/a	-	-	-	2.06 – 1.99 (m, 12H))
Gal (4)	4.48 (d, J = 7.9 Hz, 1H)	3.54	3.67	3.93	n/a	n/a	-	-	-	-

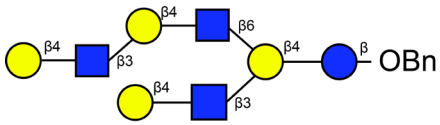
Neu5Ac	-	-	2.76 – eq (dd, J + 12.5, 4.6 Hz, 1H). 1.80 -axial (app t, *J* = 12.2 Hz, 1H)	3.70	3.85	3.63	n/a	n/a	n/a	2.06 – 1.99 (m, 12H)

Bn (CH_2)	Bn (C_6H_5)
4.94 (d, J = 11.7 Hz, 1H); 4.77 (d, J = 11.7 Hz, 1H)	7.56 – 7.33 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)	Gal (4)	Neu5Ac
C1	100.92	102.67	102.84	102.48	100.90	102.84	102.67	102.84	-
C3	-	-	-	-	-	-	-	-	39.64

ESI TOF-MS m/z calculated for $C_{72}H_{114}N_4O_{49}Na$ $[M + Na]^+$ = 1841.6450, found 1841.5840
Yield: 44.1% over 2 steps, 1.4 µmol, 2.6 mg



12

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.56 (d, *J* = 8.0 Hz, 1H)	3.37	3.61	3.60	n/a	3.82, 3.98	-
Gal (1)	4.46 (d, *J* = 7.9 Hz, 1H)	3.57	3.71	4.15 (d, J = 3.3 Hz, 2H)	n/a	n/a	-
GlcNAc (1)	4.70 (d, J = 8.3 Hz, 2H)	3.79	3.59	3.80	n/a	n/a	2.03(m, 9H)
Gal (2)	4.48 (d, *J* = 7.9 Hz, 1H)	3.55	3.68	3.94	n/a	n/a	-
GlcNAc (2)	4.63 (d, *J* = 7.8 Hz, 1H)	3.73	3.61	3.73	n/a	3.83, 3.98	2.03(m, 9H)
Gal (3)	4.46 (d, *J* = 7.9 Hz, 1H)	3.57	3.71	4.15 (d, J = 3.3 Hz, 2H)	n/a	n/a	-
GlcNAc (3)	4.70 (d, J = 8.3 Hz, 2H)	3.79	3.59	3.80	n/a	n/a	2.03(m, 9H))
Gal (4)	4.48 (d, *J* = 7.9 Hz, 1H)	3.55	3.68	3.94	n/a	n/a	-

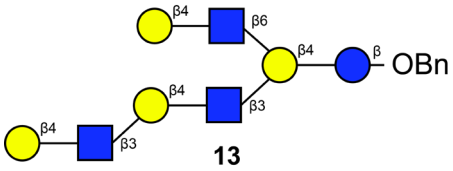
Bn (CH_2)	Bn (C_6H_5)
4.94 (d, J = 11.7 Hz, 1H); 4.77 (d, J = 11.7 Hz, 1H)	7.56 – 7.33 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)	Gal (4)
C1	100.82	102.78	102.68	102.78	100.98	102.78	102.68	102.78

ESI TOF-MS m/z calculated for $C_{61}H_{97}N_3O_{41}Na$ $[M + Na]^+$ = 1550.5495, found 1550.5429

Yield: 93.4%, 1.3 µmol, 2.0 mg



13

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.53 (d, J = 8.0 Hz, 1H)	3.33	3.58	n/a	n/a	3.76, 3.94	-
Gal (1)	4.40 (d, J = 7.8 Hz, 1H)	3.53	3.69	4.11 (d, J = 3.2 Hz, 1H)	n/a	3.79, 3.95	-
GlcNAc (1)	4.67 (dd, *J* = 8.5, 3.6 Hz, 2H)	3.77	3.70	3.55	n/a	n/a	2.00(m,9H)
Gal (2)	4.44 (app. t, *J* = 8.4 Hz, 3H)	3.53	3.69	4.13 (d, J = 3.4 Hz, 1H)	n/a	n/a	-
GlcNAc (2)	4.67 (dd, *J* = 8.5, 3.6 Hz, 2H)	3.77	3.70	3.55	n/a	n/a	2.00(m,9H)
Gal (3)	4.44 (app. t, *J* = 8.4 Hz, 3H)	3.53	3.63	3.90	n/a	n/a	-
GlcNAc (3)	4.60 (d, J = 7.7 Hz, 1H)	3.71	3.70	3.57	n/a	n/a	2.00(m,9H)
Gal (4)	4.44 (app. t, *J* = 8.4 Hz, 3H)	3.53	3.63	3.90	n/a	n/a	-

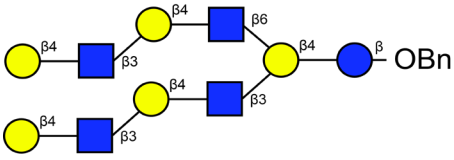
Bn (CH_2)	Bn (C_6H_5)
4.91 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.7 Hz, 1H)	7.51 – 7.35 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)	Gal (4)
C1	100.93	102.95	102.75	102.91	102.75	102.91	100.97	102.91

ESI TOF-MS m/z calculated for $C_{61}H_{98}N_3O_{41}$ $[M + H]^+$ = 1529.5754, found 1529.5524

Yield: 38.1%, 1.6 µmol, 2.4 mg



14

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.53 (d, *J* = 8.0 Hz, 1H)	3.33	3.58	3.60	n/a	3.78, 3.95	-
Gal (1)	4.40 (d, *J* = 7.9 Hz, 1H)	3.54	3.69	4.14 – 4.10 (m, 3H)	n/a	3.80, 3.96	-
GlcNAc (1)	4.67 (dd, *J* = 8.6, 3.2 Hz, 3H)	3.78	3.56	3.71	n/a	n/a	2.06 – 1.97 (m, 12H)
Gal (2)	4.47 – 4.41 (m, 4H)	3.54	3.69	4.14 – 4.10 (m, 3H)	n/a	n/a	-

GlcNAc (2)	4.67 (dd, J = 8.6, 3.2 Hz, 3H)	3.78	3.58	3.71	n/a	n/a	2.06 – 1.97 (m, 12H)
Gal (3)	4.47 – 4.41 (m, 4H)	3.52	3.64	3.90	n/a	n/a	-
GlcNAc (3)	4.60 (d, J = 7.8 Hz, 1H)	3.72	3.56	3.70	n/a	n/a	2.06 – 1.97 (m, 12H)
Gal (4)	4.47 – 4.41 (m, 4H)	3.54	3.69	4.14 – 4.10 (m, 3H)	n/a	n/a	-
GlcNAc (4)	4.67 (dd, J = 8.6, 3.2 Hz, 3H)	3.78	3.56	3.71	n/a	n/a	2.06 – 1.97 (m, 12H)
Gal (5)	4.47 – 4.41 (m, 4H)	3.52	3.64	3.90	n/a	n/a	-

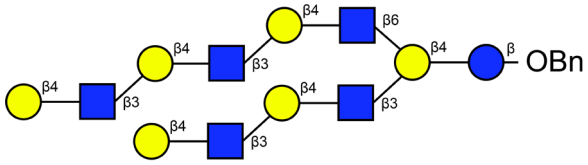
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H), 4.74 (d, J = 11.9 Hz, 1H)	7.48 – 7.36 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc(1)	Gal (2)	GlcNAc(2)	Gal (3)	GlcNAc(3)	Gal (4)	GlcNAc(4)	Gal (5)
C1	100.93	102.80	102.73	102.80	102.73	102.80	100.92	102.80	102.73	102.80

ESI TOF-MS m/z calculated for $C_{75}H_{120}N_4O_{51}Na$ $[M + Na]^+$ = 1915.6817, found 1915.6342

Yield: 39.0%, 1.3 µmol, 2.5 mg



15

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.54 (d, J = 8.0 Hz, 1H)	3.33 (t, J = 8.5 Hz, 1H)	3.59	n/a	n/a	n/a	-
Gal (1)	4.40 (d, J = 7.9 Hz, 1H)	3.54	3.69	4.11 (d, J = 3.1 Hz, 1H)	n/a	n/a	-
GlcNAc (1)	4.67 (dd, J = 8.4, 3.5 Hz, 4H)	3.78	3.56	3.70	n/a	n/a	2.03 – 1.99 (m, 15H)
Gal (2)	4.48 – 4.41 (m, 5H)	3.55	3.69	4.15 – 4.12 (m, 3H)	n/a	n/a	-
GlcNAc (2)	4.67 (dd, J = 8.4, 3.5 Hz, 4H)	3.78	3.56	3.70	n/a	n/a	2.03 – 1.99 (m, 15H)
Gal (3)	4.48 – 4.41 (m, 5H)	3.51	3.65	3.90	n/a	n/a	-
GlcNAc (3)	4.60 (d, J = 7.9 Hz, 1H)	3.72	3.57	3.69	n/a	n/a	2.03 – 1.99 (m, 15H)
Gal (4)	4.48 – 4.41 (m, 5H)	3.55	3.69	4.15 – 4.12 (m, 3H)	n/a	n/a	-
GlcNAc (4)	4.67 (dd, J = 8.4, 3.5 Hz, 4H)	3.78	3.56	3.70	n/a	n/a	2.03 – 1.99 (m, 15H)
Gal (5)	4.48 – 4.41 (m, 5H)	3.55	3.69	4.15 – 4.12 (m, 3H)	n/a	n/a	-
GlcNAc (5)	4.67 (dd, J = 8.4, 3.5 Hz, 4H)	3.78	3.56	3.70	n/a	n/a	2.03 – 1.99 (m, 15H)
Gal (6)	4.48 – 4.41 (m, 5H)	3.51	3.65	3.90	n/a	n/a	-

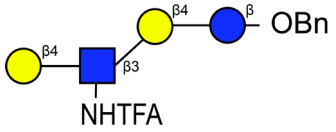
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H), 4.74 (d, J = 11.7 Hz, 1H)	7.46 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)	Gal (4)	GlcNAc (4)	Gal (5)	GlcNAc (5)	Gal (6)
C1	100.95	102.80	102.73	102.80	102.73	102.80	100.93	102.80	102.73	102.80	102.73	102.80

ESI TOF-MS m/z calculated for $C_{89}H_{143}N_5O_{61}Na_2$ $[M + 2Na]^{2+}/2$ = 1151.9019, found 1151.9017

Yield: 80.5% over 2 steps, 1.0 µmol, 2.3 mg



16

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6
Glc	4.47	3.26	3.53	3.69	n/a	n/a
Gal	4.33	3.49	3.62	4.08	n/a	n/a
GlcNTFA	4.74	3.78	3.53	3.69	n/a	n/a
Gal2	4.38	3.45	3.84	3.58	n/a	n/a

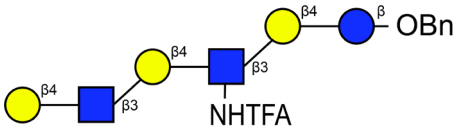
Bn(CH_2)	Bn(Ar)
4.93 (d, J = 11.6 Hz, 1H) 4.75 (d, J = 12.5 Hz, 1H)	7.61 – 7.24 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNTFA	Gal (2)
C1	101.17	102.93	102.02	102.93

ESI TOF-MS m/z calculated for $C_{33}H_{48}F_3NO_{21}Na$ $[M + Na]^+$ = 874.2569, found 874.3101

Yield: 52.1% over 2 steps, 22.9 µmol, 20.0 mg



17

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.54 (d, J = 7.9 Hz, 1H)	3.34	3.62	3.59	n/a	n/a	-
Gal (1)	4.41 (d, J = 7.9 Hz, 1H)	3.57	3.72	4.15 (d, J = 3.3 Hz, 1H)	n/a	n/a	-
GlcNTFA	4.82	3.87	3.61	3.76	n/a	n/a	-
Gal (2)	4.46 (d, J = 8.2 Hz, 1H)	3.56	3.72	4.15 (d, J = 3.3 Hz, 1H)	n/a	n/a	-
GlcNAc (1)	4.69 (d, J = 8.5 Hz, 1H)	3.80	3.74	3.59	n/a	n/a	2.03 (s, 3H)
Gal (3)	4.46 (d, J = 8.1 Hz, 1H)	3.54	3.86	3.93	n/a	n/a	-

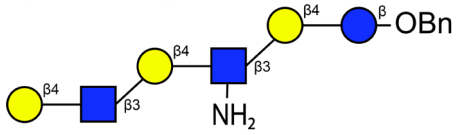
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.6 Hz, 1H)	7.53 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNTFA	Gal (2)	GlcNAc (1)	Gal (3)
C1	101.17	102.93	102.02	102.93	102.81	102.93

ESI TOF-MS m/z calculated for $C_{47}H_{72}F_3N_2O_{31}$ $[M + H]^+$ = 1217.4071, found 1217.4541

Yield: 48.8% over 2 steps, 11.2 µmol, 13.6 mg



19

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.55 (d, *J* = 7.9 Hz, 1H)	3.35	3.61	3.59	n/a	n/a	-
Gal (1)	4.43 (d, *J* = 7.9 Hz, 1H)	3.59	3.72	4.15 (d, J = 3.3 Hz, 1H)	n/a	n/a	-
$GlcNH_2$	4.71 (d, *J* = 8.4 Hz, 1H)	2.85	3.62	3.51	n/a	n/a	-
Gal (2)	4.49 (d, *J* = 8.2 Hz, 1H)	3.68	3.82	4.17 (d, J = 3.3 Hz, 1H)	n/a	n/a	-
GlcNAc (1)	4.69 (d, *J* = 8.5 Hz, 1H)	3.80	3.73	3.59	n/a	n/a	2.02 (s, 3H)
Gal (3)	4.47 (d, *J* = 8.1 Hz, 1H)	3.54	3.81	3.93	n/a	n/a	-

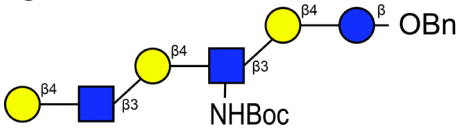
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.6 Hz, 1H)	7.53 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNBoc	Gal (2)	GlcNAc (1)	Gal (3)
C1	101.17	103.03	102.92	102.97	102.75	103.03

ESI TOF-MS m/z calculated for $C_{45}H_{73}N_2O_{30}$ $[M + H]^+$ = 1121.4248, found 1121.4559

Yield: quant., 11.2 µmol, 12.6 mg



22

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc/NHBoc
Glc	4.55 (d, *J* = 8.0 Hz, 1H)	3.34	3.61	3.59	n/a	3.74, 3.88	-
Gal (1)	4.44 (d, *J* = 8.0 Hz, 1H)	3.63	3.74	4.15 (d, J = 3.3 Hz, 1H)	n/a	3.65, 3.84	-
GlcNBoc	4.72 (d, *J* = 8.4 Hz, 1H)	3.46	3.69	3.56	n/a	3.74, 3.88	1.43 (s, 9H))
Gal (2)	4.45 (d, *J* = 7.8 Hz, 1H)	3.63	3.77	4.15 (d, J = 3.3 Hz, 1H)	n/a	3.65, 3.84	-
GlcNAc(1)	4.69 (d, *J* = 8.3 Hz, 1H)	3.79	3.73	3.59	n/a	3.74, 3.88	2.02 (s, 3H)
Gal (3)	4.47 (d, *J* = 7.9 Hz, 1H)	3.54	3.81	3.93	n/a	3.56, 3.79	-

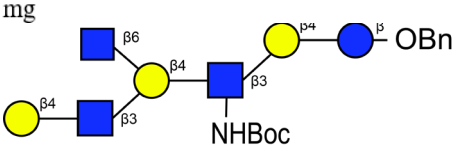
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.6 Hz, 1H)	7.53 – 7.34 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNBoc	Gal (2)	GlcNAc (1)	Gal (3)
C1	101.06	102.95	102.54	102.95	102.75	102.95

ESI TOF-MS m/z calculated for $C_{50}H_{80}N_2O_{32}Na$ $[M + Na]^+$ = 1243.4592, found 1243.5207

Yield: 75.4%, 8.5 µmol, 10.5 mg



23

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc/NHBoc
Glc	4.46 (d, *J* = 7.9 Hz, 1H)	3.36	3.63	3.59	n/a	3.82, 3.99	-
Gal (1)	4.46 (d, *J* = 7.9 Hz, 2H)	3.63	3.77	4.13 (d, *J* = 3.1 Hz, 1H)	n/a	n/a	-
GlcNBoc	4.74 (d, *J* = 7.8 Hz, 1H)	3.47	3.69	3.57	n/a	n/a	1.46 (s, 9H)
Gal (2)	4.46 (d, *J* = 7.9 Hz, 2H)	3.59	3.71	4.15 (d, *J* = 3.2 Hz, 1H)	n/a	3.83, 3.97	-
GlcNAc (1)	4.70 (d, *J* = 8.5 Hz, 1H)	3.81	3.74	3.59	n/a	n/a	2.11 – 2.00 (m, 6H)
Gal (3)	4.57 (d, *J* = 8.0 Hz, 1H)	3.55	3.67	3.94	n/a	n/a	-
GlcNAc (2)	4.63 (d, *J* = 8.5 Hz, 1H)	3.68	3.46	3.59	n/a	3.76, 3.91	2.11 – 2.00 (m, 6H)

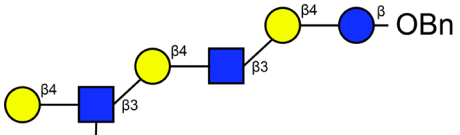
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.6 Hz, 1H)	7.46 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNBoc	Gal (2)	GlcNAc (1)	Gal (3)	GlcNAc (2)
C1	101.01	102.87	102.54	102.87	102.77	102.46	100.92

ESI TOF-MS m/z calculated for $C_{58}H_{94}N_3O_{37}$ $[M + H]^+$ = 1424.5566, found 1424.5590

Yield: 98.8.%, 8.4 µmol, 12 mg



24

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.57 (d, *J* = 8.0 Hz, 1H)	3.37	3.62	3.60	n/a	n/a	-
Gal (1)	4.44 (d, *J* = 7.9 Hz, 1H))	3.60	3.73	4.16 (d, *J* = 3.3 Hz, 1H)	n/a	n/a	-
GlcNAc	4.71 (d, *J* = 8.4 Hz, 1H)	3.80	3.74	3.59	n/a	n/a	2.04 (s, 3H)
Gal (2)	4.47 (d, *J* = 7.8 Hz, 1H)	3.60	3.75	4.19 (d, *J* = 3.2 Hz, 1H)	n/a	n/a	-
GlcNTFA	4.84 (d, *J* = 8.2 Hz, 1H)	3.91	3.82	3.63	n/a	n/a	-
Gal (3)	4.50 (d, *J* = 7.8 Hz, 1H)	3.57	3.68	3.94	n/a	n/a	-

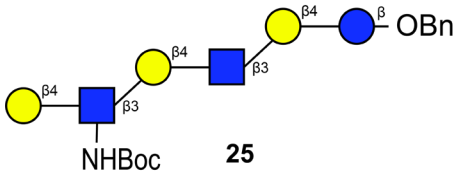
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.6 Hz, 1H)	7.51 – 7.39 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc	Gal (2)	GlcNTFA	Gal (3)
C1	101.05	102.93	102.73	102.93	102.04	102.93

ESI TOF-MS m/z calculated for $C_{47}H_{72}F_3N_2O_{31}$ $[M + H]^+$ = 1217.4071, found 1217.4326

Yield: 63.6% over 2 steps, 1.4 µmol, 1.7 mg



25

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc/NHBoc
Glc	4.57 (d, *J* = 8.0 Hz, 1H)	3.37	3.62	3.65	n/a	n/a	-
Gal (1)	4.44 (d, *J* = 7.9 Hz, 1H)	3.60	3.72	4.16 (d, *J* = 3.4 Hz, 1H)	n/a	n/a	-
GlcNAc	4.71 (d, *J* = 8.3 Hz, 1H)	3.80	3.78	3.60	n/a	n/a	2.04 (s, 3H)
Gal (2)	4.50 (d, *J* = 7.4 Hz, 1H)	3.60	3.79	4.15 (d, *J* = 3.0 Hz, 1H)	n/a	n/a	-
GlcNBoc	4.74 (d, *J* = 8.6 Hz, 1H)	3.45	3.70	3.58	n/a	n/a	1.46 (s, 9H)
Gal (3)	4.49 (d, *J* = 7.6 Hz, 1H)	3.55	3.66	3.94	n/a	n/a	-

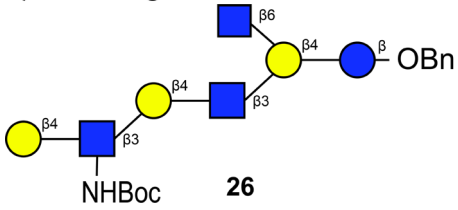
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.6 Hz, 1H)	7.52 – 7.38 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc	Gal (2)	GlcNBoc	Gal (3)
C1	100.95	102.79	102.72	102.79	102.51	102.79

ESI TOF-MS m/z calculated for $C_{50}H_{81}N_2O_{32}$ $[M + H]^+$ = 1221.4773, found 1221.4816

Yield: 29.3% over 2 steps, 0.4 µmol, 0.5 mg



26

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc/NHBoc
Glc	4.57 (d, *J* = 8.1 Hz, 1H)	3.37	3.61	3.65	n/a	n/a	-
Gal (1)	4.43 (d, *J* = 7.6 Hz, 1H)	3.59	3.72	4.15 (d, *J* = 3.4 Hz, 1H)	n/a	n/a	-
GlcNAc	4.71 (d, *J* = 8.3 Hz, 1H)	3.79	3.75	3.60	n/a	n/a	2.03 (s, 3H)
Gal (2)	4.50 (d, *J* = 7.5 Hz, 1H)	3.66	3.77	4.14 (d, *J* = 3.0 Hz, 1H)	n/a	n/a	-
GlcNBoc	4.74 (d, *J* = 8.8 Hz, 1H)	3.48	3.70	3.58	n/a	n/a	1.46 (s, 9H)
Gal (3)	4.49 (d, *J* = 7.8 Hz, 1H)	3.55	3.66	3.94	n/a	n/a	-
GlcNAc (2)	4.62 (d, *J* = 8.4 Hz, 1H)	3.71	3.48	3.60	n/a	n/a	2.06 (s, 3H)

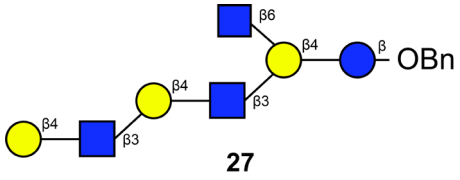
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.6 Hz, 1H)	7.52 – 7.38 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc	Gal (2)	GlcNBoc	Gal (3)	GlcNAc (2)
C1	100.89	103.22	102.59	102.93	102.75	102.93	101.02

ESI TOF-MS m/z calculated for $C_{58}H_{94}N_3O_{37}$ $[M + H]^+$ = 1424.5566, found 1424.5381

Yield: 58.5%, 140 nmol, 0.2 mg



27

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.57 (d, *J* = 8.0 Hz, 1H)	3.36	3.60	3.59	n/a	n/a	-
Gal (1)	4.43 (d, *J* = 7.8 Hz, 1H)	3.59	3.72	4.15 (d, *J* = 3.4 Hz, 1H)	n/a	n/a	-
GlcNAc (1)	4.71 (d, *J* = 8.3 Hz, 2H)	3.81	3.72	3.57	n/a	n/a	2.05(s, 9H)
Gal (2)	4.48 (d, *J* = 7.7 Hz, 1H)	3.59	3.74	4.17 (d, *J* = 3.5 Hz, 1H)	n/a	n/a	-
GlcNAc (2)	4.71 (d, *J* = 8.3 Hz, 2H)	3.81	3.72	3.57	n/a	n/a	2.05(s, 9H)
Gal (3)	4.49 (d, *J* = 7.7 Hz, 1H)	3.54	3.72	3.91	n/a	n/a	-
GlcNAc (3)	4.62 (d, *J* = 8.5 Hz, 1H)	3.69	3.45	3.55	n/a	n/a	2.05(s, 9H)

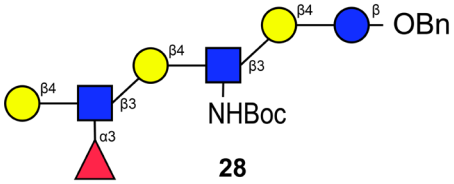
Bn (CH_2)	Bn (C_6H_5)
4.95 (d, J = 11.6 Hz, 1H); 4.78 (d, J = 11.6 Hz, 1H)	7.53 – 7.39 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)
C1	100.97	102.99	102.67	102.85	102.67	102.85	101.03

ESI TOF-MS m/z calculated for $C_{55}H_{88}N_3O_{36}$ $[M + H]^+$ = 1366.5147, found 1366.5135

Yield: 72.4% from **15**, 2.1 µmol, 2.9 mg



28

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc/NHBoc
Glc	4.56 (d, *J* = 8.1 Hz, 1H)	3.35	3.61	3.59	n/a	n/a	-
Gal (1)	4.46 (d, *J* = 7.5 Hz, 2H)	3.61	3.72	4.14 (d, J = 3.3 Hz, 1H)	n/a	n/a	-
GlcNBoc	4.73 (d, *J* = 8.4 Hz, 1H)	3.45	3.69	3.58	n/a	n/a	1.45 (s, 9H)
Gal (2)	4.46 (d, *J* = 7.5 Hz, 2H)	3.61	3.78	4.14 (d, J = 3.3 Hz, 1H)	n/a	n/a	-
GlcNAc	4.71	3.96	3.89	3.58	n/a	n/a	2.02(s, 3H)
Gal (3)	4.46 (d, *J* = 7.5 Hz, 2H)	3.51	3.73	3.90	n/a	n/a	-
Fuc	5.13	3.69	3.91	3.80	4.83	1.19 (d, *J* = 6.6 Hz, 3H)	-

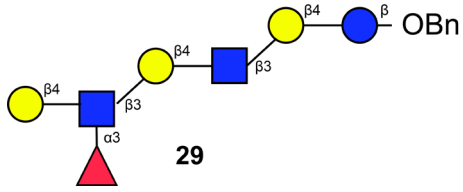
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.6 Hz, 1H)	7.46 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNBoc	Gal (2)	GlcNAc	Gal (3)	Fuc
C1	100.94	102.93	102.68	102.93	102.68	102.93	98.12

ESI TOF-MS m/z calculated for $C_{56}H_{91}N_2O_{36}$ $[M + H]^+$ = 1367.5351, found 1367.5088

Yield: 54.4%, 3.7 µmol, 5 mg



29

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.57 (d, *J* = 8.0)	3.33	3.61	3.59	n/a	n/a	-
Gal (1)	4.44 (d, *J* = 8.2 Hz, 1H)	3.59	3.74	4.14 (d, J = 3.3 Hz, 1H)	n/a	n/a	-
GlcNAc (1)	4.71 (d, *J* = 6.9 Hz, 1H)	3.82	3.71	3.58	n/a	n/a	2.02(s, 6H)
Gal (2)	4.47 (d, *J* = 7.9 Hz, 2H)	3.58	3.58	4.14 (d, J = 3.3 Hz, 1H)	n/a	n/a	-
GlcNAc (2)	4.72 (d, *J* = 7.6 Hz, 1H)	3.96	3.89	3.58	n/a	n/a	2.02(s, 6H)
Gal (3)	4.47 (d, *J* = 7.9 Hz, 2H)	3.46	3.73	3.91	n/a	n/a	-
Fuc	5.14 (d, *J* = 4.0 Hz)	3.69	3.92	3.80	4.85	1.19 (d, *J* = 6.6 Hz, 3H)	-

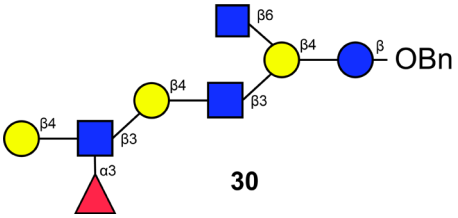
Bn (CH_2)	Bn (C_6H_5)
4.95 (d, J = 11.7 Hz, 1H); 4.78 (d, J = 11.6 Hz, 1H)	7.51 – 7.40 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	Fuc
C1	101.09	102.98	102.79	102.98	102.79	102.15	98.69

ESI TOF-MS m/z calculated for $C_{53}H_{85}N_2O_{35}$ $[M + H]^+$ = 1309.4933, found 1309.4722

Yield: 91.9% over 2 steps, 3.4 μmol, 4.4 mg



30

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.57 (d, *J* = 8.1 Hz, 1H)	3.38	3.62	3.59	n/a	n/a	-
Gal (1)	4.44 (d, *J* = 7.9 Hz, 1H)	3.59	3.70	4.15 (d, *J* = 3.5 Hz, 1H)	n/a	n/a	-
GlcNAc (1)	4.71 (d, *J* = 8.6 Hz, 2H)	3.78	3.73	3.58	n/a	n/a	2.02(s, 6H)
Gal (2)	4.47 (d, *J* = 7.9 Hz, 2H)	3.69	3.74	4.17 (d, *J* = 3.4 Hz, 1H)	n/a	n/a	-
GlcNAc (2)	4.72 (d, *J* = 9.0 Hz, 1H)	4.01	3.90	3.59	n/a	n/a	2.02(s, 6H)
Gal (3)	4.47 (d, *J* = 7.9 Hz, 2H)	3.47	3.73	3.92	n/a	n/a	-
GlcNAc (3)	4.62 (d, *J* = 8.5 Hz, 1H)	3.70	3.47	3.54	n/a	n/a	2.06 (s, 3H)
Fuc	5.14 (d, *J* = 4.0 Hz, 1H)	3.71	3.92	3.80	4.85	1.19 (d, *J* = 6.7 Hz, 3H)	-

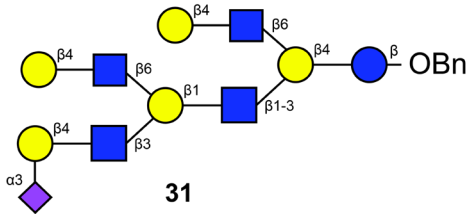
Bn (CH_2)	Bn (C_6H_5)
4.95 (d, J = 11.7 Hz, 1H); 4.78 (d, J = 11.6 Hz, 1H)	7.51 – 7.40 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)	Fuc
C1	101.00	102.71	102.66	102.83	102.66	101.94	101.18	98.59

ESI TOF-MS m/z calculated for $C_{61}H_{98}N_3O_{40}$ $[M + H]^+$ = 1512.5727, found 1512.5567

Yield: 83.6% over 2 steps, 2.9 μmol, 4.3 mg



31

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	H7	H8	H9	NHAc
Glc	4.53 (d, *J* = 8.0 Hz, 1H)	3.33	3.58	3.68	n/a	3.79, 3.96	-	-	-	-
Gal (1)	4.40 (d, *J* = 7.9 Hz, 1H)	3.56	3.69	4.01 (d, J = 3.3 Hz, 1H)	n/a	3.81, 3.96	-	-	-	-
GlcNAc (1)	4.67 (d, J = 8.4 Hz, 2 H)	3.78	3.70	3.56	n/a	3.84, 3.93	-	-	-	2.08 – 1.98 (m, 15H)
Gal (2)	4.44 (dd, *J* = 7.9, 3.3 Hz, 3H)	3.53	3.68	4.13 (d, J = 3.3 Hz, 1H)	n/a	3.81, 3.96	-	-	-	-
GlcNAc (2)	4.67 (d, J = 8.4 Hz, 2 H)	3.78	3.70	3.56	n/a	3.84, 3.93	-	-	-	2.08 – 1.98 (m, 15H))
Gal (3)	4.46 (d, J = 7.8 Hz, 1H)	3.53	3.64	3.90	n/a	3.61, 3.73	-	-	-	-
GlcNAc (3)	4.61 (d, *J* = 7.7 Hz, 1H)	3.72	3.70	3.58	n/a	3.82, 3.97	-	-	-	2.08 – 1.98 (m, 15H))
Gal (4)	4.44 (dd, *J* = 7.9, 3.3 Hz, 3H)	3.53	3.64	3.90	n/a	3.61, 3.73	-	-	-	-
GlcNAc (4)	4.62 (d, *J* = 7.7 Hz, 1H)	3.70	3.70	3.57	n/a	3.81, 3.96	-	-	-	2.08 – 1.98 (m, 15H)
Gal (5)	4.42 (dd, *J* = 7.9, 3.3 Hz, 3H)	3.54	3.64	3.90	n/a	3.61, 3.73	-	-	-	-
Neu5Ac	-	-	2.84 – eq 1.80 -axial (app t, *J* = 12.1 Hz, 1H)	3.66	3.85	3.62	n/a	n/a	n/a	2.08 – 1.98 (m, 15H)

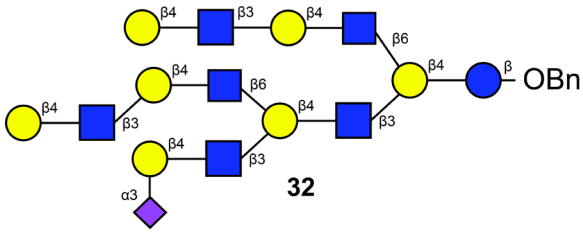
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.75 (d, J = 11.6 Hz, 1H)	7.48 – 7.34 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)	Gal (4)	GlcNAc (4)	Gal (5)	Neu5Ac
C1	100.91	102.84	102.66	102.84	102.66	102.84	100.83	102.84	100.83	102.84	-
C3	-	-	-	-	-	-	-	-	-	-	34.89

ESI TOF-MS m/z calculated for $C_{86}H_{137}N_5O_{59}Na_2$ $[M + 2Na]^{2+}/2$ = 1114.8835, found 1114.8812

Yield: 63.5% over 2 steps, 4.0 µmol, 8.8 mg



32

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	H7	H8	H9	NHAc
Glc	4.54 (d, *J* = 8.0 Hz, 2H)	3.32	3.59	n/a	n/a	3.78, 3.96	-	-	-	-
Gal (1)	4.48 – 4.38 (m, 6H)	3.54	3.70	4.15 – 4.07 (m, 5H)	n/a	n/a	-	-	-	-
GlcNAc (1)	4.68 (d, *J* = 8.6 Hz, 4H)	3.78	3.72	3.56	n/a	n/a	2.06 – 1.98 (m, 18H)	-	-	2.07 – 1.99 (m, 21H)
Gal (2)	4.48 – 4.38 (m, 6H)	3.56	3.70	4.15 – 4.07 (m, 5H)	n/a	n/a	-	-	-	-
GlcNAc (2)	4.68 (d, *J* = 8.6 Hz, 4H)	3.78	3.72	3.56	n/a	n/a	2.06 – 1.98 (m, 18H)	-	-	2.07 – 1.99 (m, 21H)
Gal (3)	4.54 (d, *J* = 8.0 Hz, 2H)	3.52	3.89	4.15 – 4.07 (m, 5H)	n/a	n/a	-	-	-	-
GlcNAc (3)	4.61 (d, *J* = 8.0 Hz, 1H)	3.55	3.69	3.57	n/a	n/a	2.06 – 1.98 (m, 18H)	-	-	2.07 – 1.99 (m, 21H)
Gal (4)	4.48 – 4.38 (m, 6H)	3.56	3.70	4.15 – 4.07 (m, 5H)	n/a	n/a	-	-	-	-
GlcNAc (4)	4.68 (d, *J* = 8.6 Hz, 4H)	3.78	3.72	3.56	n/a	n/a	2.06 – 1.98 (m, 18H)	-	-	2.07 – 1.99 (m, 21H))
Gal (5)	4.48 – 4.38 (m, 6H))	3.51	3.64	3.90	n/a	n/a	-	-	-	-
GlcNAc (5)	4.62 (d, *J* = 8.0 Hz, 1H)	3.55	3.69	3.57	n/a	n/a	2.06 – 1.98 (m, 18H)	-	-	2.07 – 1.99 (m, 21H))
Gal (6)	4.48 – 4.38 (m, 6H)	3.56	3.70	4.15 – 4.07 (m, 5H)	n/a	n/a	-	-	-	-
GlcNAc (6)	4.68 (d, *J* = 8.6 Hz, 4H)	3.78	3.72	3.56	n/a	n/a	2.06 – 1.98 (m, 18H)	-	-	2.07 – 1.99 (m, 21H)
Gal (7)	4.48 – 4.38 (m, 6H)	3.51	3.64	3.90	n/a	n/a	-	-	-	-

Neu5Ac	-	-	2.74 – eq 1.77 - axial (app t, *J* = 12.1 Hz, 1H)	3.66	3.85	3.62	n/a	n/a	n/a	2.07 – 1.99 (m, 21H))

Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.8 Hz, 1H); 4.75 (d, J = 11.3 Hz, 1H)	7.47 – 7.39 (m, 5H)

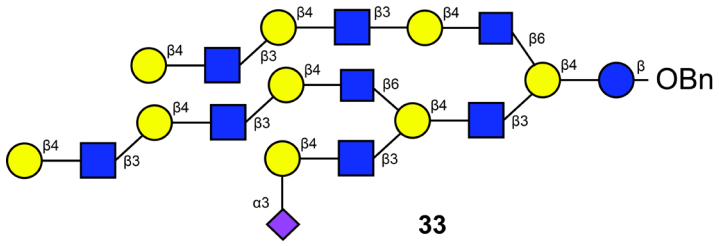
^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)	Gal (4)	GlcNAc (4)
C1	101.24	102.78	102.57	102.78	102.57	102.78	101.29	102.78	102.57
C3	-	-	-	-	-	-	-	-	-

	Gal (5)	GlcNAc (5)	Gal (6)	GlcNAc (6)	Gal (7)	Neu5Ac
C1	102.78	101.29	102.78	102.57	102.78	-
C3	-	-	-	-	-	34.10

ESI TOF-MS m/z calculated for $C_{114}H_{183}N_7O_{79}Na_2$ $[M + 2Na]^{2+}/2$ = 1480.0157, found 1480.0058

Yield: 49.3% over 2 steps, 1.33 μmol, 3.8 mg



33

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	H7	H8	H9	NHAc
Glc	4.54 (d, *J* = 8.0 Hz, 1H)	3.34 (t, *J* = 7.8 Hz, 1H)	3.59	n/a	n/a	n/a	-	-	-	-
Gal (1)	4.40 (d, *J* = 7.8 Hz, 1H)	3.54	3.69	4.15 – 4.10 (m, 6H)	n/a	n/a	-	-	-	-
GlcNAc (1)	4.67 (d, *J* = 8.3 Hz, 6H)	3.78	3.71	3.56	n/a	n/a	-	-	-	2.04 – 2.00 (m, 27H)
Gal (2)	4.47 – 4.41 (m, 8H)	3.56	3.69	4.15 – 4.10 (m, 6H)	n/a	n/a	-	-	-	-
GlcNAc (2)	4.67 (d, *J* = 8.3 Hz, 6H)	3.78	3.71	3.56	n/a	n/a	-	-	-	2.04 – 2.00 (m, 27H)
Gal (3)	4.54 (d, *J* = 8.0 Hz, 2H)	3.52	3.89	4.15 – 4.10 (m, 6H)	n/a	n/a	-	-	-	-
GlcNAc (3)	4.60 (d, *J* = 7.9 Hz, 2H)	3.71	3.68	3.58	n/a	n/a	-	-	-	2.04 – 2.00 (m, 27H)

Gal (4)	4.47 – 4.41 (m, 8H)	3.56	3.69	4.15 – 4.10 (m, 6H)	n/a	n/a	-	-	-	-
GlcNAc (4)	4.67 (d, *J* = 8.3 Hz, 6H)	3.78	3.71	3.56	n/a	n/a	-	-	-	2.04 – 2.00 (m, 27H)
Gal (5)	4.47 – 4.41 (m, 8H)	3.56	3.69	4.15 – 4.10 (m, 6H)	n/a	n/a	-	-	-	-
GlcNAc (5)	4.67 (d, *J* = 8.3 Hz, 6H)	3.78	3.71	3.56	n/a	n/a	-	-	-	2.04 – 2.00 (m, 27H)
Gal (6)	4.47 – 4.41 (m, 8H)	3.52	3.64	3.91	n/a	n/a	-	-	-	-
GlcNAc (6)	4.62 (d, *J* = 7.8 Hz, 1H)	3.70	3.68	3.57	n/a	n/a	-	-	-	2.04 – 2.00 (m, 27H)
Gal (7)	4.47 – 4.41 (m, 8H)	3.56	3.69	4.15 – 4.10 (m, 6H)	n/a	n/a	-	-	-	-
GlcNAc (7)	4.67 (d, *J* = 8.3 Hz, 6H)	3.78	3.71	3.56	n/a	n/a	-	-	-	2.04 – 2.00 (m, 27H)
Gal (8)	4.47 – 4.41 (m, 8H)	3.56	3.69	4.15 – 4.10 (m, 6H)	n/a	n/a	-	-	-	-
GlcNAc (8)	4.67 (d, *J* = 8.3 Hz, 6H)	3.78	3.71	3.56	n/a	n/a	-	-	-	2.04 – 2.00 (m, 27H)
Gal (9)	4.47 – 4.41 (m, 8H)	3.52	3.64	3.91	n/a	n/a	-	-	-	-
Neu5Ac	-	-	2.74 – eq 1.77 -axial (app t, *J* = 12.1 Hz, 1H)	3.66	3.85	3.62	n/a	n/a	n/a	2.04 – 2.00 (m, 27H)

Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.8 Hz, 1H); 4.75 (d, J = 11.3 Hz, 1H)	7.47 – 7.39 (m, 5H)

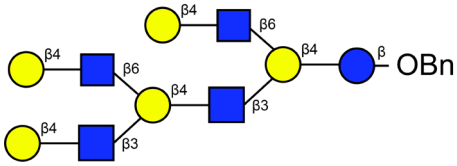
^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc(1)	Gal (2)	GlcNAc(2)	Gal (3)	GlcNAc(3)	Gal (4)	GlcNAc(4)
C1	100.83	102.85	102.96	102.85	102.96	102.85	102.96	102.85	102.96
C3	-	-	-	-	-	-	-		

	Gal (5)	GlcNAc (5)	Gal (6)	GlcNAc (6)	Gal (7)	GlcNAc (6)	Gal (7)	GlcNAc (7)	Gal (8)	Neu5Ac
C1	102.85	102.96	102.85	102.96	102.85	102.96	102.85	102.96	102.85	-
C3			-	-	-	-	-	-	-	36.24

ESI TOF-MS m/z calculated for $C_{142}H_{229}N_9O_{99}Na_2$ $[M + 2Na]^{2+}/2$ = 1845.1479 found 1845.1245

Yield: 20.3% over 2 steps, 0.27 μmol, 0.8 mg



34

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.54 (d, *J* = 8.0 Hz, 1H)	3.34	3.58	3.68	n/a	3.79, 3.96	-
Gal (1)	4.40 (d, *J* = 7.9 Hz, 1H)	3.56	3.69	4.11 (d, J = 3.3 Hz, 1H)	n/a	3.81, 3.96	-
GlcNAc (1)	4.67 (d, J = 8.4 Hz, 2 H)	3.78	3.70	3.56	n/a	3.84, 3.93	2.06 – 1.97 (m, 12H)
Gal (2)	4.44 (dd, *J* = 7.9, 3.3 Hz, 3H)	3.53	3.68	4.13 (d, J = 3.3 Hz, 1H)	n/a	3.81, 3.96	-
GlcNAc (2)	4.67 (d, J = 8.4 Hz, 2 H)	3.78	3.70	3.56	n/a	3.84, 3.93	2.06 – 1.97 (m, 12H)
Gal (3)	4.46 (d, J = 7.8 Hz, 1H)	3.53	3.64	3.90	n/a	3.61, 3.73	-
GlcNAc (3)	4.61 (d, *J* = 7.7 Hz, 1H)	3.72	3.70	3.58	n/a	3.82, 3.97	2.06 – 1.97 (m, 12H)
Gal (4)	4.44 (dd, *J* = 7.9, 3.3 Hz, 3H)	3.53	3.64	3.90	n/a	3.61, 3.73	-
GlcNAc (4)	4.62 (d, *J* = 7.7 Hz, 1H)	3.70	3.70	3.57	n/a	3.81, 3.96	2.06 – 1.97 (m, 12H)
Gal (5)	4.44 (dd, *J* = 7.9, 3.3 Hz, 3H)	3.53	3.64	3.90	n/a	3.61, 3.73	-

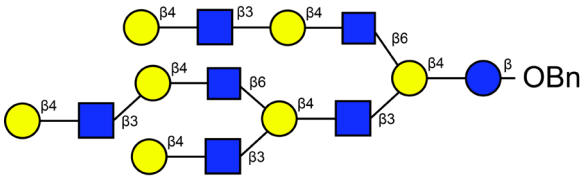
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.6 Hz, 1H)	7.48 – 7.34 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)	Gal (4)	GlcNAc (4)	Gal (5)
C1	100.91	102.84	102.66	102.84	102.66	102.84	100.83	102.84	100.83	102.84

ESI TOF-MS m/z calculated for $C_{75}H_{120}N_4O_{51}Na$ $[M + Na]^+$ = 1915.6817, found 1915.6422

Yield: 41.5%, 1.7 µmol, 3.3 mg



35

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.54 (d, *J* = 8.0 Hz, 1H)	3.34	3.59	n/a	n/a	3.78, 3.96	-
Gal (1)	4.40 (d, *J* = 7.9 Hz, 1H)	3.54	3.69	4.11 (d, *J* = 3.2 Hz, 1H)	n/a	n/a	-
GlcNAc (1)	4.67 (dd, *J* = 8.3, 2.7 Hz, 4H)	3.78	3.72	3.56	n/a	n/a	2.06 – 1.98 (m, 18H)
Gal (2)	4.43 (m, 3H)	3.56	3.69	4.13 (d, *J* = 3.4 Hz, 3H)	n/a	n/a	-
GlcNAc (2)	4.67 (dd, *J* = 8.3, 2.7 Hz, 4H)	3.78	3.72	3.56	n/a	n/a	2.06 – 1.98 (m, 18H)
Gal (3)	4.45 (m, 3H)	3.51	3.64	3.90	n/a	n/a	-
GlcNAc (3)	4.60 (d, *J* = 7.7 Hz, 1H)	3.71	3.69	3.57	n/a	n/a	2.06 – 1.98 (m, 18H)
Gal (4)	4.43 (m, 3H)	3.56	3.69	4.13 (d, *J* = 3.4 Hz, 3H)	n/a	n/a	-
GlcNAc (4)	4.67 (dd, *J* = 8.3, 2.7 Hz, 4H)	3.78	3.72	3.56	n/a	n/a	2.06 – 1.98 (m, 18H)
Gal (5)	4.45 (m, 3H)	3.51	3.64	3.90	n/a	n/a	-
GlcNAc (5)	4.61 (d, *J* = 7.3 Hz, 1H)	3.70	3.69	3.57	n/a	n/a	2.06 – 1.98 (m, 18H)
Gal (6)	4.43 (m, 3H)	3.56	3.69	4.13 (d, *J* = 3.4 Hz, 3H)	n/a	n/a	-
GlcNAc (6)	4.45 (m, 3H)	3.78	3.72	3.56	n/a	n/a	2.06 – 1.98 (m, 18H)
Gal (7)	4.45 (m, 3H)	3.51	3.64	3.90	n/a	n/a	-

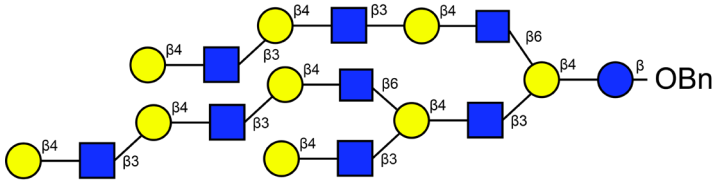
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, *J* = 11.6 Hz, 1H), 4.74 (d, *J* = 11.6 Hz, 1H)	7.47 – 7.38 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)	Gal (4)
C1	100.97	102.85	102.70	102.85	102.70	102.85	100.87	102.85

	GlcNAc (4)	Gal (5)	GlcNAc (5)	Gal (6)	GlcNAc (6)	Gal (7)
C1	102.70	102.85	100.87	102.85	102.70	102.85

ESI TOF-MS m/z calculated for $C_{103}H_{166}N_6O_{71}Na_2$ $[M + 2Na]^{2+}/2$ = 1334.4679, found 1334.4757 Yield: 60.1%, 0.8 µmol, 2.1 mg



36

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.54 (d, *J* = 8.0 Hz, 1H)	3.34 (t, *J* = 7.8 Hz, 1H)	3.59	n/a	n/a	n/a	-
Gal (1)	4.40 (d, *J* = 7.8 Hz, 1H)	3.54	3.69	4.15 – 4.10 (m, 6H)	n/a	n/a	-
GlcNAc (1)	4.67 (d, *J* = 8.3 Hz, 6H)	3.78	3.71	3.56	n/a	n/a	2.04 – 2.00 (m, 24H)
Gal (2)	4.47 – 4.41 (m, 8H)	3.56	3.69	4.15 – 4.10 (m, 6H)	n/a	n/a	-
GlcNAc (2)	4.67 (d, *J* = 8.3 Hz, 6H)	3.78	3.71	3.56	n/a	n/a	2.04 – 2.00 (m, 24H)
Gal (3)	4.47 – 4.41 (m, 8H)	3.52	3.64	3.91	n/a	n/a	-
GlcNAc (3)	4.60 (d, *J* = 7.9 Hz, 2H)	3.71	3.68	3.58	n/a	n/a	2.04 – 2.00 (m, 24H)
Gal (4)	4.47 – 4.41 (m, 8H)	3.56	3.69	4.15 – 4.10 (m, 6H)	n/a	n/a	-
GlcNAc (4)	4.67 (d, *J* = 8.3 Hz, 6H)	3.78	3.71	3.56	n/a	n/a	2.04 – 2.00 (m, 24H)
Gal (5)	4.47 – 4.41 (m, 8H)	3.56	3.69	4.15 – 4.10 (m, 6H)	n/a	n/a	-
GlcNAc (5)	4.67 (d, *J* = 8.3 Hz, 6H)	3.78	3.71	3.56	n/a	n/a	2.04 – 2.00 (m, 24H)
Gal (6)	4.47 – 4.41 (m, 8H)	3.52	3.64	3.91	n/a	n/a	-
GlcNAc (6)	4.62 (d, *J* = 7.8 Hz, 1H)	3.70	3.68	3.57	n/a	n/a	2.04 – 2.00 (m, 24H)
Gal (7)	4.47 – 4.41 (m, 8H)	3.56	3.69	4.15 – 4.10 (m, 6H)	n/a	n/a	-
GlcNAc (7)	4.67 (d, *J* = 8.3 Hz, 6H)	3.78	3.71	3.56	n/a	n/a	2.04 – 2.00 (m, 24H)
Gal (8)	4.47 – 4.41 (m, 8H)	3.56	3.69	4.15 – 4.10 (m, 6H)	n/a	n/a	-
GlcNAc (7)	4.67 (d, *J* = 8.3 Hz, 6H)	3.78	3.71	3.56	n/a	n/a	2.04 – 2.00 (m, 24H)
Gal (8)	4.47 – 4.41 (m, 8H)	3.56	3.69	4.15 – 4.10 (m, 6H)	n/a	n/a	-
GlcNAc (8)	4.67 (d, *J* = 8.3 Hz, 6H)	3.78	3.71	3.56	n/a	n/a	2.04 – 2.00 (m, 24H)
Gal (9)	4.47 – 4.41 (m, 8H)	3.52	3.64	3.91	n/a	n/a	-

Bn (CH_2)	Bn (C_6H_5)
4.92 (d, *J* = 11.6 Hz, 1H), 4.74 (d, *J* = 11.7 Hz, 1H)	7.47 – 7.37 (m, 5H)

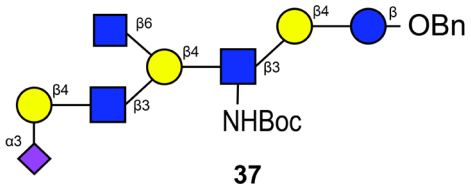
^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)	Gal (4)	GlcNAc (4)
C1	100.93	102.79	102.70	102.79	102.70	102.79	100.82	102.79	102.70

	Gal (5)	GlcNAc (5)	Gal (6)	GlcNAc (6)	Gal (7)	GlcNAc (7)	Gal (8)	GlcNAc (8)	Gal (9)
C1	102.79	102.79	102.79	100.87	102.79	102.70	102.79	102.70	102.79

ESI TOF-MS m/z calculated for $C_{131}H_{212}N_8O_{91}Na_2$ $[M + 2Na]^{2+}/2$ = 1699.6001, found 1699.5932Yield:

Yield: 80.0%, 0.2 µmol, 0.7 mg



37

^1H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	H7	H8	H9	NHAc/NHBoc
Glc	4.57 (d, *J* = 8.0 Hz, 1H)	3.36	3.61	3.59	n/a	3.74, 3.88	-	-	-	-
Gal (1)	4.46 (d, *J* = 7.9 Hz, 2H)	3.54	3.75	4.15 (d, J = 3.3 Hz, 1H)	n/a	3.65, 3.84	-	-	-	-
GlcNBoc	4.75 (d, *J* = 8.4 Hz, 1H)	3.49	3.69	3.59	n/a	3.74, 3.88	-	-	-	1.40 (s, 9H)
Gal (2)	4.46 (d, *J* = 7.9 Hz, 2H)	3.70	3.63	4.15 (d, J = 3.3 Hz, 1H)	n/a	3.65, 3.84	-	-	-	-
GlcNAc (1)	4.75 (d, *J* = 8.4 Hz, 1H)	3.81	3.76	3.59	n/a	3.74, 3.88	-	-	-	2.12 – 2.01 (m, 9H)
Gal (3)	4.57 (d, *J* = 7.9 Hz, 1H)	3.59	3.97	4.13	n/a	3.56, 3.79	-	-	-	-
GlcNAc (2)	4.64 (d, *J* = 8.4 Hz, 1H)	3.69	3.47	3.60	n/a	3.74, 3.88	-	-	-	2.12 – 2.01 (m, 9H)
Neu5Ac	-	-	2.78 – eq 1.80 - axial (app t, *J* = 12.1 Hz, 1H)	3.70	3.85	3.63	n/a	n/a	n/a	2.12 – 2.01 (m, 9H)

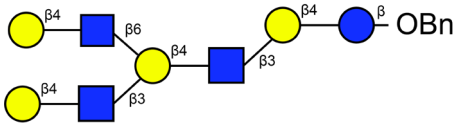
Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.6 Hz, 1H)	7.46 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNBoc	Gal (2)	GlcNAc (1)	Gal (3)	GlcNAc (2)
C1	101.01	102.87	102.54	102.87	102.77	102.46	100.92

ESI TOF-MS m/z calculated for $C_{69}H_{111}N_4O_{45}$ $[M + H]^+$ = 1715.6520, found 1715.6564

Yield: 93%, 7.8 µmol, 13.4 mg



38

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.53 (d, *J* = 8.0 Hz, 1H)	3.32	3.59	3.56	n/a	3.79, 3.96	-
Gal (1)	4.41 (d, *J* = 7.8 Hz, 1H)	3.56	3.70	4.12 (d, J = 3.3 Hz, 1H)	n/a	3.79, 3.96	-
GlcNAc (1)	4.68 (d, *J* = 8.3 Hz, 1H)	3.78	3.71	3.56	n/a	3.83, 3.93	2.05 – 2.00 (m, 9H)
Gal (2)	4.43 (d, *J* = 7.9 Hz, 1H)	3.54	3.69	4.13 (d, J = 3.3 Hz, 1H)	n/a	3.81, 3.94	-
GlcNAc (2)	4.67 (d, *J* = 8.1 Hz, 1H)	3.76	3.71	3.56	n/a	3.83, 3.93	2.05 – 2.00 (m, 9H)
Gal (3)	4.45 (app. t, *J* = 8.4 Hz, 2H)	3.52	3.65	3.90	n/a	3.72	-
GlcNAc (3)	4.62 (d, *J* = 7.5 Hz, 1H)	3.70	3.69	3.57	n/a	3.80, 3.95	2.05 – 2.00 (m, 9H)
Gal (4)	4.45 (app. t, *J* = 8.4 Hz, 2H)	3.52	3.65	3.90	n/a	3.72	-

Bn (CH_2)	Bn (C_6H_5)
4.92 (d, J = 11.6 Hz, 1H); 4.74 (d, J = 11.6 Hz, 1H)	7.46 – 7.37 (m, 5H)

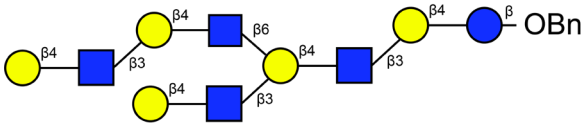
^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)	Gal (4)
C1	100.86	102.78	102.55	102.78	102.55	102.78	100.78	102.78

ESI TOF-MS m/z calculated for $C_{61}H_{97}N_3O_{41}Na$ $[M + Na]^+$ = 1550.5495, found 1550.5330

LacNAc extension yield: 37.7% over 2 steps, 2.9 µmol, 5.5 mg

Deprotection and acetylation (1 µmol sm) yield: 80.0% over 3 steps, 0.8 µmol, 1.2 mg



39

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.53 (d, *J* = 8.1 Hz, 1H)	3.32	3.60	3.56	n/a	n/a	-
Gal (1)	4.47 – 4.39 (m, 5H)	3.55	3.70	4.13 (m, 3H)	n/a	n/a	-
GlcNAc (1)	4.68 (d, *J* = 8.2 Hz, 3H)	3.78	3.56	3.71	n/a	n/a	2.05 – 2.00 (m, 12H)
Gal (2)	4.47 – 4.39 (m, 5H)	3.56	3.70	4.13 (m, 3H)	n/a	n/a	-
GlcNAc (2)	4.68 (d, *J* = 8.2 Hz, 3H)	3.78	3.56	3.71	n/a	n/a	2.05 – 2.00 (m, 12H)
Gal (3)	4.47 – 4.39 (m, 5H)	3.52	3.64	3.90	n/a	n/a	-
GlcNAc (3)	4.61 (d, *J* = 7.6 Hz, 1H)	3.70	3.57	3.69	n/a	n/a	2.05 – 2.00 (m, 12H)
Gal (4)	4.47 – 4.39 (m, 5H)	3.56	3.70	4.13 (m, 3H)	n/a	n/a	-
GlcNAc (4)	4.68 (d, *J* = 8.2 Hz, 3H)	3.78	3.56	3.71	n/a	n/a	2.05 – 2.00 (m, 12H)
Gal (5)	4.47 – 4.39 (m, 5H)	3.52	3.64	3.90	n/a	n/a	-

Bn (CH_2)	Bn (C_6H_5)
4.92 (d, *J* = 12.1 Hz, 1H), 4.74 (d, *J* = 11.8 Hz, 1H)	7.49 – 7.37 (m, 5H)

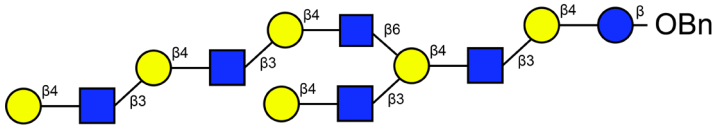
^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)	Gal (4)	GlcNAc (4)	Gal (5)
C1	100.97	102.78	102.68	102.78	102.68	102.78	100.73	102.78	102.68	102.78

ESI TOF-MS m/z calculated for $C_{75}H_{120}N_4O_{51}Na$ $[M + Na]^+$ = 1915.6817, found 1915.6233

LacNAc extension yield: 68.4% over 2 steps, 1.3 μmol, 3.0 mg

Deprotection and acetylation (1 μmol sm) yield: 68.0% over 3 steps, 0.7 μmol, 1.3 mg



40

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	NHAc
Glc	4.54 (d, *J* = 7.9 Hz, 1H)	3.33 (d, *J* = 8.3 Hz, 1H)	3.59	3.57	n/a	n/a	-
Gal (1)	4.44 – 4.39 (m, 4H)	3.56	3.69	4.13	n/a	n/a	-
GlcNAc (1)	4.68 (d, *J* = 8.6 Hz, 4H)	3.78	3.56	3.71	n/a	n/a	2.06 – 1.97 (m, 15H)
Gal (2)	4.44 – 4.39 (m, 4H)	3.56	3.69	4.13	n/a	n/a	-
GlcNAc (2)	4.68 (d, *J* = 8.6 Hz, 4H)	3.78	3.56	3.71	n/a	n/a	2.06 – 1.97 (m, 15H)
Gal (3)	4.46 (d, *J* = 7.7 Hz, 2H)	3.53	3.65	3.90	n/a	n/a	-
GlcNAc (3)	4.62 (d, *J* = 7.8 Hz, 1H)	3.71	3.57	3.69	n/a	n/a	2.06 – 1.97 (m, 15H)
Gal (4)	4.44 – 4.39 (m, 4H)	3.56	3.69	4.13	n/a	n/a	-
GlcNAc (4)	4.68 (d, *J* = 8.6 Hz, 4H)	3.78	3.56	3.71	n/a	n/a	2.06 – 1.97 (m, 15H)
Gal (5)	4.44 – 4.39 (m, 4H)	3.56	3.69	4.13	n/a	n/a	-
GlcNAc (5)	4.68 (d, *J* = 8.6 Hz, 4H)	3.78	3.56	3.71	n/a	n/a	2.06 – 1.97 (m, 15H)
Gal (6)	4.46 (d, *J* = 7.7 Hz, 2H)	3.53	3.65	3.90	n/a	n/a	-

Bn (CH_2)	Bn (C_6H_5)
4.92 (d, *J* = 12.2 Hz, 1H), 4.74 (d, *J* = 11.2 Hz, 1H)	7.50 – 7.37 (m, 5H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	Glc	Gal (1)	GlcNAc (1)	Gal (2)	GlcNAc (2)	Gal (3)	GlcNAc (3)
C1	101.08	102.90	102.78	102.90	102.78	102.90	100.79

	Gal (4)	GlcNAc (4)	Gal (5)	GlcNAc (5)	Gal (6)
C1	102.90	102.78	102.90	102.78	102.90

ESI TOF-MS m/z calculated for $C_{89}H_{143}N_5O_{61}Na_2$ $[M + 2Na]^{2+}/2$ = 1151.9019, found 1151.9087

LacNAc extension yield: 75.0% over 2 steps, 0.7 µmol, 1.7 mg

Deprotection and acetylation (0.7 µmol sm) yield: 76.9% over 3 steps, 0.5 µmol, 1.1 mg

CHAPTER 4

Oxidative Release of *O*-glycans under Neutral Conditions for the Analysis of Glycoconjugates having Base Sensitive Substituents

Gaël M. Vos, Julia Weber, Igor R. Sweet, Kevin C. Hooijschuur,
Javier Sastre Toraño, Geert-Jan Boons

Introduction

The most prominent post-translational modification of proteins in terms of complexity and diversity is by complex glycans.[1,2] Almost all naturally occurring protein glycosylation can be classified as either *N*- and *O*-glycans by modification of side chains of Asn and Ser/Thr, respectively. The biosynthesis of *N*-linked glycans occurs by *en bloc* transfer of a dolichol-linked $Glc_3Man_9GlcNAc_2$ to an Asn-X-Ser/Thr sequon of newly synthesized polypeptides, which is catalyzed by an oligosaccharide transferase complex.[3] Subsequent trimming of the transferred oligosaccharide results in the formation of a $Man_3GlcNAc_2$ core structure that, after transport to the Golgi, is modified into complex structures. *O*-glycosylation is initiated by twenty polypeptide GalNAc-transferases that attach an α-linked *N*-acetyl-galactosamine (GalNAc) residue to specific Ser/Thr residues of proteins.[4,5] The resulting GalNAc moieties are then elaborated into various core structures which can be extended by type 1 and type 2 oligo-*N*-acetyl lactosamine moieties and capped by several forms of sialic acid and histo-blood group antigens. *O*-glycosylation can result in considerable structural diversity which is regulated in a developmental, spatial, and temporal manner.[6] Mucins and mucin-like proteins, which are densely modified by *O*-glycans, are important for tissue lubrication and form a physical barrier that can inhibit pathogen infection.[7,8] They also mediate interactions between cells and their local environment. Changes in *O*-glycosylation is associated with many diseases such as cancer, inflammation and enteric infections.[9-11]

Analysis of glycans of glycoconjugates in complex biological samples is important for linking specific structures to biological properties and for biomarker discovery.[12,13] A critical step in glycomics is the release of glycans from glycoconjugates for analysis by mass spectrometry.[14] *N*-glycans can be enzymatically released under mild conditions by peptide-*N*4-(*N*-acetyl-*α*-glucosaminyl) asparagine amidase (PNGase)[15] or various endo-*β*-*N*-acetylglucosaminidases.[16] No enzymes are available for the release of *O*-glycans and therefore these are commonly detached by base mediated *β*-elimination to produce reducing glycans.[17,18] A major problem of this approach is that the acyclic forms of the reducing glycans can undergo peeling reactions by removal of the acidic C-2 proton resulting in the elimination of C-3 linked substituents.[19–21] Chemical modifications have been explored to reduce peeling by, for example, reduction by sodium cyanoborohydride to give alditols that are resistant to peeling after conversion.[22,23] Reductive amination can achieve a similar result while allowing the incorporation of chemical entities that may facilitate purification and/or ionization.[24–26] These methods, however, still suffer from peeling and loss of base labile substituents such as acetyl esters and sulfates. A mild *O*-glycan release method that does not result in peeling and is compatible with labile substituents is urgently needed for the analysis of *O*-glycomes.

Alkaline hypochlorite is an inexpensive and scalable method for oxidatively release of *N*-glycans, *O*-glycans and glycolipids.[27] In the case of *O*-glycans, it produces several products including glycosides of glycolic or lactic acid and reducing sugars that are prone to peeling.[27–29] Here, we employed well-defined synthetic glycopeptides to establish a robust workflow for the oxidative release of *O*-glycans suitable for glycomic analysis, preserving sensitive substituents. Surprisingly, it was found that *O*-glycans can be released by neutralized hypochlorite, which results in the selective formation of lactic/glycolic acid linked *O*-glycan

thereby giving unique information of the parent amino acid (serine/threonine) that is lost by *β*-elimination. It locks the glycan in a closed ring configuration thereby preventing peeling reactions. An additional advantage of the procedure is that the anomeric tag contains a carboxylate that promotes ionization in negative ion mode mass spectroscopy (MS), thereby increasing signal intensities and improving limits of detection with respect to the analysis of free reducing end or reduced *O*-glycans. Moreover, labile modifications such as sialic acid, sulfates, and acetyl esters are maintained during the release procedure. The promise of the approach was demonstrated by the analysis of *O*-glycans from bovine submaxillary mucin (BSM), which identified mono- and di-*O*-acetylated sialoglycans as well as previously undetected tri-*O*-acetylated and sulfated glycans. The use of well-defined glycopeptide standards made it also possible to identify reaction intermediates which in turn allowed us to postulate a reaction mechanism for oxidative *O*-glycan release under neutral conditions.

Results and discussion

Development of the oxidative release workflow.

Recent developments in synthesis of *O*-glycopeptides provide access to well-defined *O*-glycopeptide standards.[30,31] We anticipated that such standards would be useful to establish workflows for the controlled release of *O*-glycans. Furthermore, it was expected that well-defined glycopeptides can also be useful to detect intermediate products which, in turn, can give insights into the reaction mechanism of glycan release.

To develop an oxidative release workflow suitable for *O*-glycomic analysis, glycopeptides **1-4** were prepared. [Weber *et al.*; manuscript in preparation] The compounds all contain a β1,3-linked GalNAc moiety that is prone to peeling (Fig.1). Compounds 1-3 are core-3 glycopeptides, including neutral glycan 1 and structures decorated with sialic acid (2) and fucose (3). Compound 4 is a core-2 glycopeptide carrying 6-sulfated GlcNAc. Glycopeptide 1 was used as an initial standard to investigate *O*-glycan release by hypochlorite. Structures 2-4 were then used to determine the chemical stability of these common glycan substructures under the optimized release conditions.

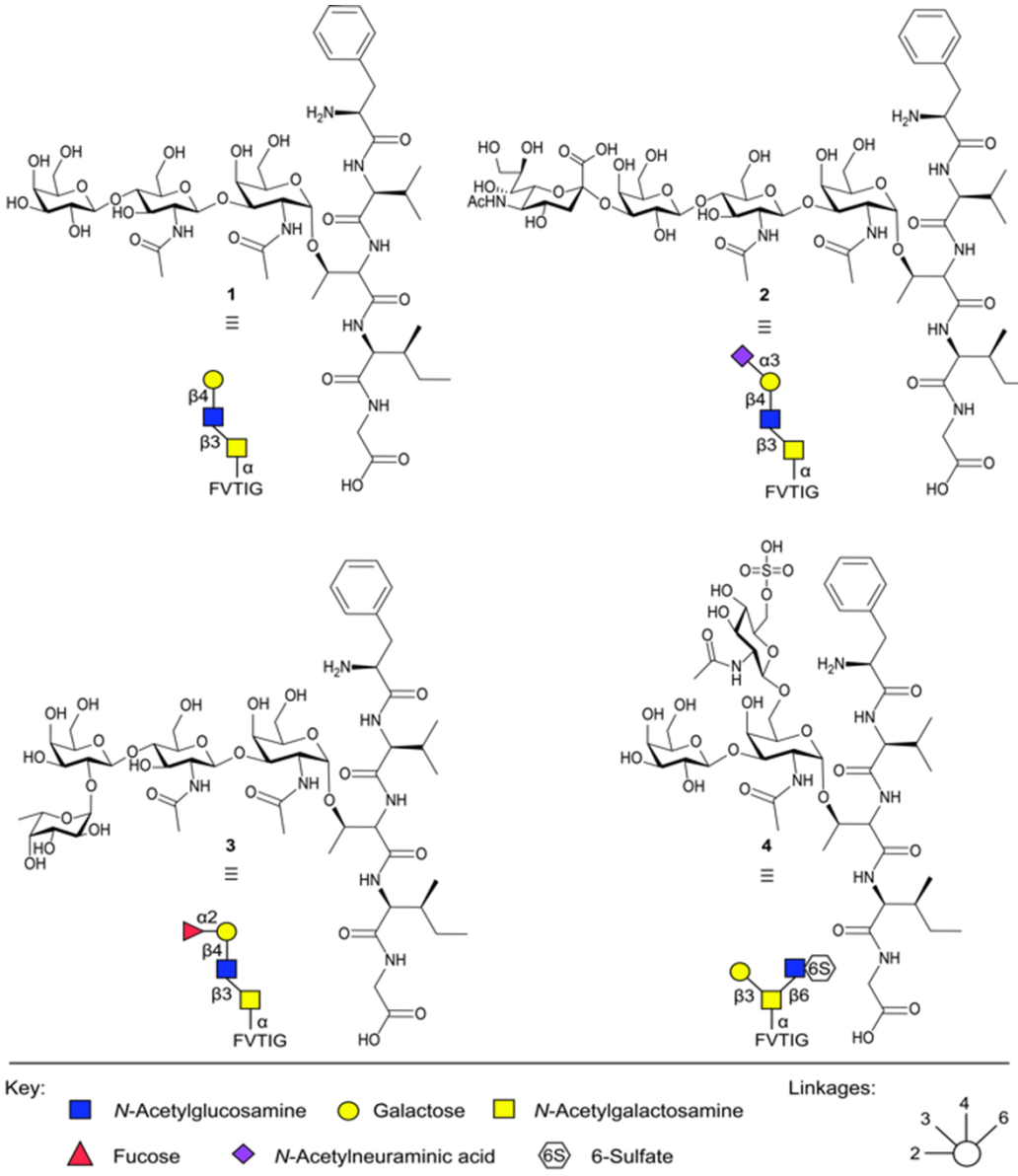


Figure 1: Chemical structures and graphic representation of the O-glycopeptide standards with FVTIG peptide sequence.

First, the effect of pH of hypochlorite on reaction time and product formation was investigated using core 3 glycopeptide **1**. Thus, this compound was subjected to a 3% sodium hypochlorite solution which was acidified with 1M HCl covering a pH range from pH 5 to pH 12, keeping the final hypochlorite concentration constant. The reactions were performed in an ice bath to minimize potential degradation by *β*-elimination of the reactions performed under alkaline conditions. Samples were taken at 10 min time intervals, reactions were quenched with formic acid and samples were purified before analysis by LC-MS. Surprisingly, complete consumption of starting material was observed within 10 min in the pH range 7.5-12, which is much shorter

than previously reported.[27-29] The major release product under the most alkaline condition (pH 12) was free reducing carbohydrate **6** (approximately 68% as relative concentration determined by MS). Both lactone **5** and lactic acid-linked *O*-glycan **7** were also observed (Fig. 2A, B and C).

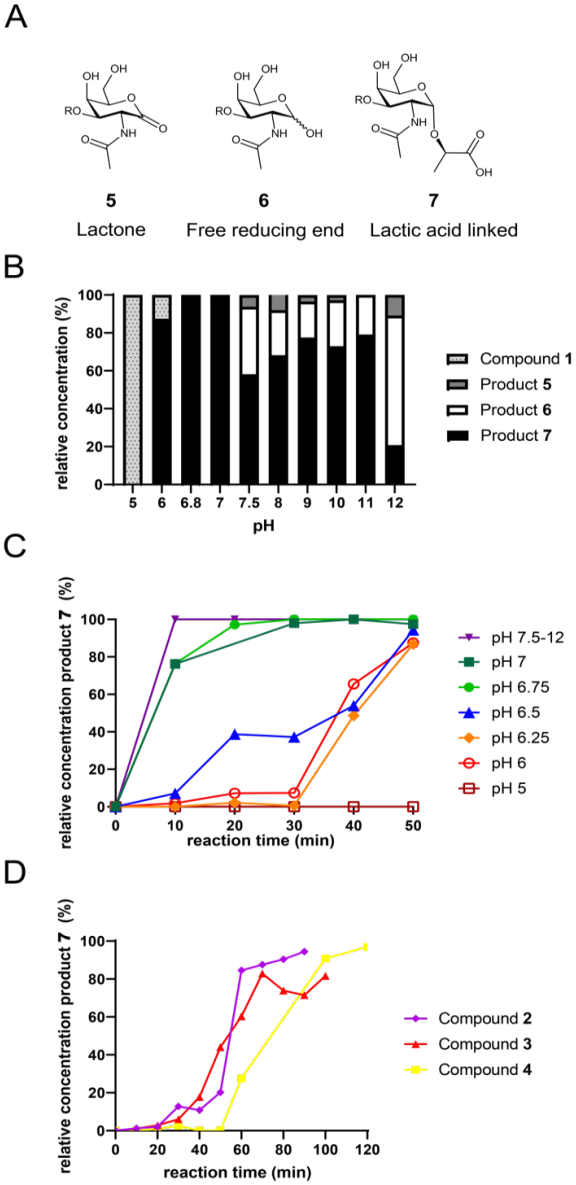


Figure 2: A) Chemical structure of products generated by oxidative release B) Relative concentrations by MS of O-glycopeptide standard **1** and reaction products after 60 min incubation in a 3% hypochlorite solution at 0 ºC and different pH values. C) Relative concentrations by MS of released amounts of **1** over time in a 3% hypochlorite solution at 0 ºC, measured as "product" concentration relative to starting material and observed side products. D) Relative concentrations by MS of released amounts of **2-4** over time in a 3% hypochlorite solution at 0 ºC and pH 6.8, measured as "product" concentration relative to starting material and observed side products.

In the pH range 7.5-11, the major product of the reaction was lactic acid-linked *O*-glycan but the lactone containing product and free reducing glycan were also formed, albeit at lower relative quantities compared to treatment at pH 12. Further acidification to pH 7-5.5 resulted in the clean formation of lactic acid-linked product. The reaction at neutral pH did result in a somewhat slower *O*-glycan release. At a pH below 5.5 no reaction was observed, and only unreacted glycopeptide was detected. Neutralized hypochlorite in a pH range of 6.8-7 resulted in complete *O*-glycan release within 30 min with no detectable side reactions and therefore all further experiments were conducted within this pH range.

Sialic acid, fucose and sulfate containing glycopeptides **2**, **3** and **4** were subjected to hypochlorite neutralized to pH 6.8 to evaluate the stability of common glycan epitopes. The reactions were sampled in 10 min intervals and reaction progress and glycan stabilities were evaluated by LC-MS. All evaluated glycan epitopes were converted to the lactic acid glycosides within 60-90 min and remained stable under the neutralized hypochlorite conditions (Fig. 2D). Only minimal Neu5Ac hydrolysis was observed which might be attributed to the acidic quenching conditions. Lactic acid-linked glycoside was the only release product observed for glycopeptide **4** and accounted for 93% of relative abundancy observed for glycopeptide **2** and 97% for glycopeptide **3**; the remainder was free reducing carbohydrate. Notably, a delay in the consumption of starting material was observed for all substituted glycopeptides. This delay can possibly be explained by the buildup of $Cl_3O_2^-$, which is a more reactive chlorinating agent suggested to be formed in neutralized hypochlorite.[32] Treatment during 60 min with 3% hypochlorite at pH 6.8 and 0 ºC resulted in complete conversion of glycopeptides **1**-**3** to the lactic acid-linked *O*-glycan, while sulfated glycopeptide **4** required a treatment time of 90 min for complete release.

Oxidative release of *O*-glycans from submaxillary mucin.

Bovine submaxillary mucin (BSM) was subjected to the optimized oxidative release conditions. BSM is a commercially available mucin that is relatively well characterized and commonly employed in biomedical research.[27,33–35] It carries sialoglycans modified by acetyl esters[36] that are prone to migration and hydrolysis under alkaline conditions.[37] Thus, BSM was treated with neutralized hypochlorite and samples were taken at 30 min intervals to confirm the previous established release kinetics. Comparable results were obtained at 30- and 60-min intervals for smaller glycans but larger glycans were detected at higher relative abundancies after 60 min treatment. Surprisingly, prolonged exposure to the hypochlorite solution (90 and 120 min) resulted in the formation of previously undetected chlorinated products. Therefore, a 60 min release time was selected at which minimal chlorination was observed (<0.5% for the most abundant *O*-glycan). It resulted in the identification of 275 *O*-glycans, which were either modified by an anomeric lactic or glycolic acid derived from threonine or serine, respectively (Fig. 3A). Gratifyingly, no reducing glycans or other derivatives were observed in notable quantities. Furthermore, most of the abundant *O*-glycans were detected both as lactic and glycolic acid linked products, thereby conforming the *O*-glycan structure.

The released glycans from BSM consisted mostly of sialylated di- and trisaccharides but also included larger compounds such as a low abundant dodecasaccharide. The majority (92% of the total ion abundance) of all detected structures contained sialic acids with Neu5Ac and Neu5Gc

accounting for 73% and 27% of the sialic acid content, respectively. Fucose was detected on 56 *O*-glycans accounting for 3.8% of the total glycan abundancy. Eight di-fucosylated structures were observed (0.7% of the total glycan abundancy) indicating the presence of Lewis^{y} or Lewis^{b} epitopes. Furthermore, 36 structures, accounting for 1% of the total glycan abundancy, contained both a fucoside and sialosides, probably representing sialyl Lewis^{x} or sialyl Lewis^{a} epitopes.

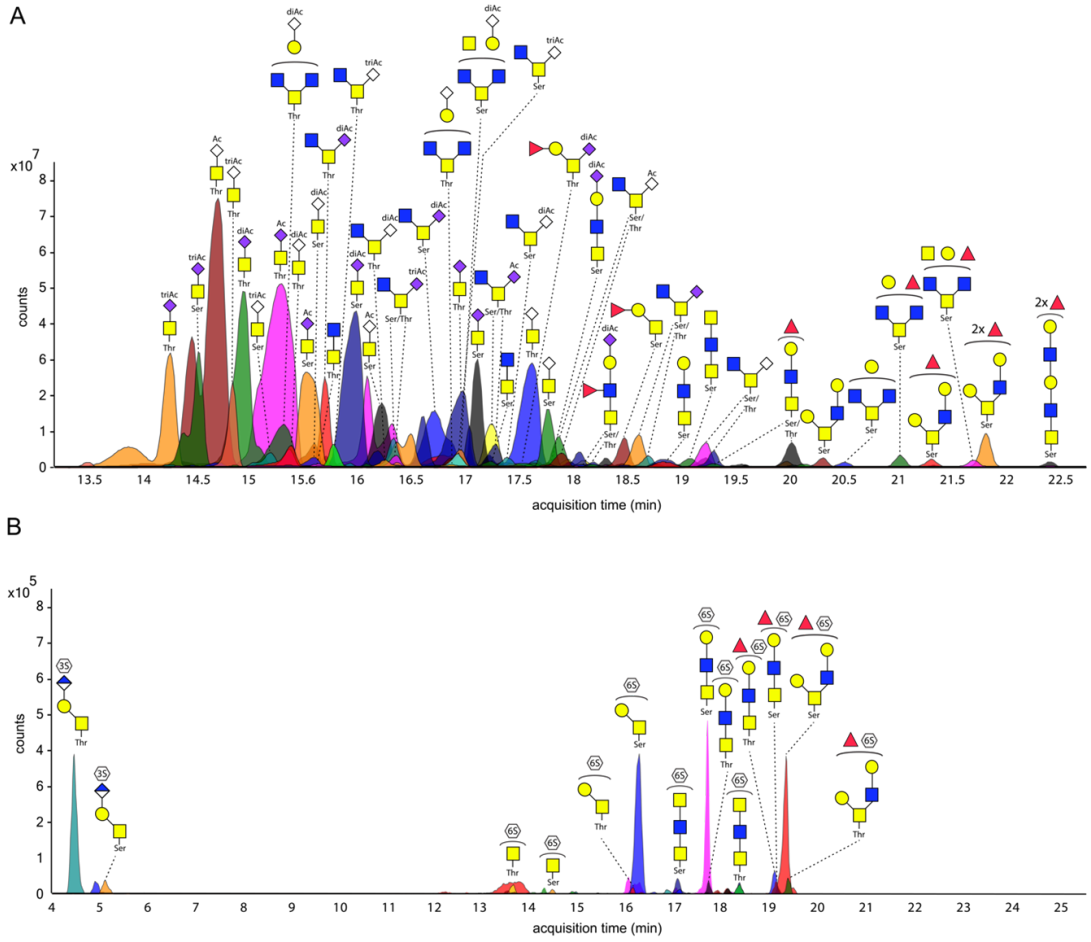


Figure 3: A) Suggested structures of *O*-glycans from Bovine submaxillary mucin (BSM) with >0.1% relative concentration released using neutralized hypochlorite at pH 6.8 for 60 min. B) Suggested structures of sulfated *O*-glycans from Bovine submaxillary mucin (BSM) released using neutralized hypochlorite at pH 6.8 for 60 min.

The presence of acetyl ester and sulfated *O*-glycans in BSM was evaluated to confirm that neutralized hypochlorite is suitable to release such saccharides. Acetyl esters were observed on 84 *O*-glycans as mono-, di- and tri- *O*-acetylated sialic acids. The presence of acetyl esters was detected on 87% of all Neu5Ac and 74% of all Neu5Gc containing ions, supporting high preservation of this functionality. Furthermore, sialic acids having multiple acetyl esters were detected in substantial quantities (Neu5Ac: mono 38%; di 37%; tri 12%) (Neu5Gc: mono 33%; di 25%; tri: 16%). The presence of these di- and tri-*O*-acetylated sialic acids shows that even the very labile acetyl esters on the C-7 and C-8 position are preserved under neutralized conditions.

A total of 14 sulfated structures were detected accounting for 0.4% of the total ion abundancy, which to the best of our knowledge, is the first time that such structures have been detected on BSM (Fig. 3B). Fucosylated and sulfated structures were observed, but none of these compounds contained sialic acid, which is in stark contrast to the highly sialylated unsulfated glycans. Surprisingly, two HNK-1 epitope containing *O*-glycans were detected, composed of a core 1 trisaccharide. The conservation of labile di- and tri-*O*-acetylated sialic acids and sulfated *O*-glycans released from BSM supports the attractiveness of the release method for the analysis of complex mucinous samples. In this respect, previous analysis only provided mono-acetylated derivatives.[27]

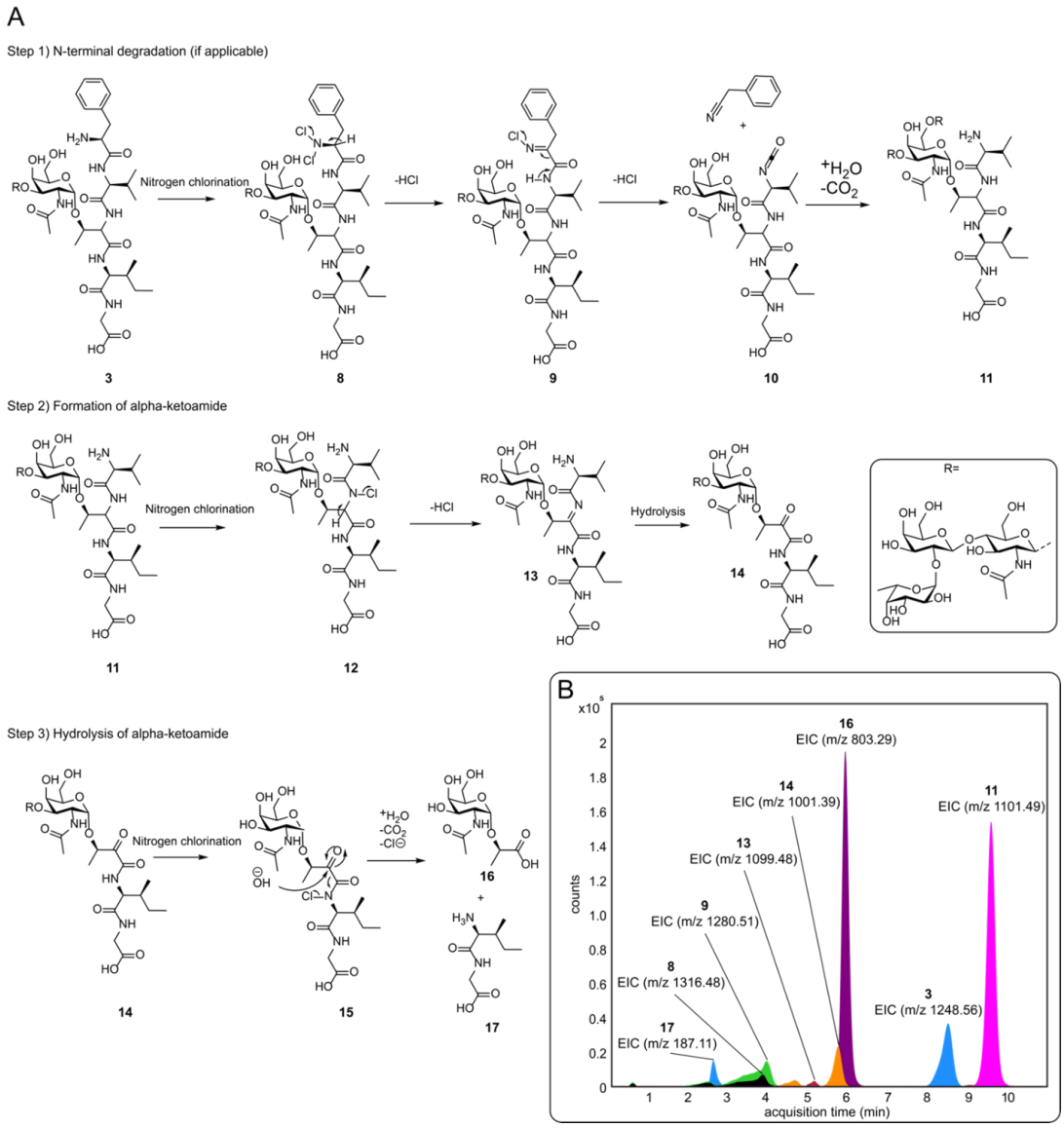


Figure 4: A) Proposed mechanism for the formation of lactic acid-linked *O*-glycan by hypochlorite mediated *O*-glycan release. B) Extracted ion chromatograms of compound **3** with proposed intermediates and reaction products after 50 min incubation with neutralized hypochlorite at pH 6.8, obtained by LC-MS in negative ion mode.

Reaction mechanism for oxidative release of *O*-glycans from glycopeptides and glycoproteins.

The synthetic glycopeptide standards made it possible to identify reactions intermediates. The identified intermediates were consistently observed for glycopeptides **1-4** when released at neutral pH and made it possible to postulate a reaction mechanism (Fig. 4). Intermediate **11** (*m/z* 1101.49) was abundantly detected and consistent of a loss of phenylalanine. The masses of **8** (*m/z* 1316.48) and **9** (*m/z* 1280.51) corresponded with di-chlorinated starting material and mono-chlorinated imine which is most like a result of elimination. Detection of **13**, which has an *m/z* -2 with respect to compound **11** (*m/z* 1099.48) indicates the formation of an imine derivative of **11**. Intermediate **14** (m/z 1001.39) was observed in relatively low abundancies for all standards, corresponding to a loss of the complete *N*-terminal peptide with respect to the glycosylation site and formation of an alpha-ketoamide.

We propose that the initial step is cleavage of the *N*-terminal phenylalanine moiety resulting in a new valine *N*-terminal peptide. The *N*-terminal degradation probably occurs due to *N*-terminal *N*-dichlorination observed as intermediate **8**. This *N*-dichloropeptide is then converted to the *N*-chloroiminopeptide **9** by an elimination reaction. Subsequent degradation of *N*-chloroiminopeptide **9** results in formation of **11** and 2-phenylacetonitrile.[38] Surprisingly, compound **4** did not show significant *N*-terminal degradation which suggests that this step is not required for *O*-glycan release. Further *N*-terminal degradation of valine following this mechanism would generate a cyanomethyl glycoside, which is not observed. Although hydrolysis of the cyanomethyl glycoside could produce **16**, we can exclude this degradation pathway since these cyanomethyl glycosides have been reported as a stable product produced by the alkaline oxidative release of glycolipids.[27] Compound **11** then eliminates to compound **13**, initiated by *N*-chlorination of the amide bond between valine and glycosylated threonine. Hydrolysis of imine **13** results in the formation of alpha-ketoamide **14**. The hydrolysis of **14** takes place at the more nucleophilic ketone moiety and leads to the lactic acid **16**. Detection of dipeptide **17** supports this pathway and indicates a decarboxylative process.

Conclusions

An oxidative O-glycan release method is described using neutralized hypochlorite which preserves labile entities such as acetyl esters and sulfates. It provides glycosides of lactic and glycolic acid, which prevents peeling and increases the sensitivity of detection by mass spectrometry. The latter tag also provides a convenient filtering approach to avoid false positive hits, which is expected to be especially useful in the *O*-glycomic analysis of complex natural samples. We anticipate that neutralized hypochlorite O-glycan release will find broad use as a mild and efficient *O*-glycan release method that better reflects the native glycome.

References

[1] A. Varki, *Glycobiology* **2017**, *27* (1), 3-49.
[2] G.W. Hart, R.J. Copeland, *Cell*. **2010**, *143* (5), 672-676
[3] K.W. Moremen, R.S. Haltiwanger, *Nat. Chem. Biol.* **2019**, *15* (9), 853-864.
[4] E.P. Bennett, U. Mandel, H. Clausen, T.A. Gerken, T.A. Fritz, L.A. Tabak, *Glycobiology*. **2012**, *22* (6), 736-756.
[5] K.T. Schjoldager, Y. Narumatsu, H.J. Joshi, H. Clausen, *Nat. Rev. Mol. Cell Biol*. **2020**, *21* (12), 729-749.
[6] I. Brockhausen, H. Schachter, P. Stanley, *O-GalNAc Glycans*, Cold Spring Harbor Laboratory Press, **2009**.
[7] S.K. Linden, P. Sutton, N.G. Karlsson, V. Korolik, M.A. McGuckin, *Mucosal Immunol.* **2008**, *1* (3), 183–197.
[8] M.E.V. Johansson, J.M.H. Larsson, G.C. Hansson, *Proc. Natl. Acad. Sci.* **2011**, *108*, 4659–4665.
[9] E.A. Grice, J.A. Segre, The skin microbiome. *Nat. Rev. Microbiol.* **2013**, *9* (4), 244-253.
[10] S.K. Linden, P. Sutton, N.G. Karlsson, M.A. McGuckin, **2008**, *1* (3), 183-197.
[11] G.E. Cervoni, J.J. Cheng, K.A. Stackhouse, J. Heimburg-Molinaro, R.D. Cummings, *Biochem. J.* **2020**, *477* (8), 1541-1564
[12] C. Reily, T.J. Stewart, M.B. Renfrow, J. Novak, *Nat. Rev. Nephrol.* **2019**, *15* (6), 346–366.
[13] R. Gupta, F. Leon, S. Rauth, S.K. Batra, M.P. Ponnusamy, *Cells* **2020**, *9* (2), E446.
[14] N. Leymarie, J. Zaia, *Anal. Chem.* **2012**, *84* (7), 3040–3048.
[15] P. Hermentin, R. Doenges, R. Witzel, C.H. Hokke, J.F. Vliegenthart, J.P. Kamerling, H.S. Conradt, M. Nimtz, D. Brazel, *Anal. Biochem.* **1994**, *221* (1), 29-41.
[16] Y. Kimura, T. Tokuda, A. Ohno, H. Tanaka, Y. Ishiguro, *Biochim. Biophys. Acta*. **1998**, *1381* (1), 27-36.
[17] H. Wilkinson, R. Saldova, *J. Proteome Res.* **2020**, *19* (10), 3890-3905.
[18] S.B. Levery, C. Steentoft, A. Halim, Y. Narimatsu, H. Clausen, S.Y. Vakhrushev, *Biochim. Biophys. Acta.* **2015**, *1850* (1) 33-42.
[19] A. Kameyama, W.W. Thet Tin, M. Toyoda, M. Sakaguchi, *Biochem. Biophys. Res. Commun.* **2019**, *513* (1), 186–192.
[20] N.G. Karlsson, C. Jin, M.A. Rojas-Macias, B. Adamczyk, *Trends Glycosci. Glycotechnol.* **2017**, *29* (166), E35–E46.
[21] M. Fukuda, T. Kondo, T.J. Osawa, *Biochem. (Tokyo)* **1976**, *80* (6), 1223–1232.
[22] D.M. Carlson, *J. Biol. Chem.* **1968**, *243* (3), 616–626.
[23] Y. Huang, T. Konse, Y. Mechref, M.V. Novotny, *Rapid Commun. Mass Spectrom. RCM* **2002**, *16* (12), 1199–1204.
[24] L.R. Ruhaak, E. Steenvoorden, C.A.M. Koeleman, A.M. Deelder, M. Wuhrer, *PROTEOMICS* **2010**, *10* (12), 2330–2336.
[25] J.C. Bigge, T.P. Patel, J.A. Bruce, P.N. Goulding, S.M. Charles, R.B. Parekh, *Anal. Biochem.* **1995**, *230* (2), 229–238.

[26] Z. Wang, Z.S. Chinoy, S.G. Ambre, W. Peng, R. McBride, R.P. de Vries, J. Glushka, J.C. Paulson, G.-J. Boons, *Science* **2013**, *341* (6144).
[27] X, Song, H. Ju, Y. Lasanajak, M.R. Kudelka, D.F. Smith, R.D. Cummings, *Nat. Methods* **2016**, *13* (6), 528–534.
[28] Q. Zhang, Z. Li, X. Song, *Front. Chem.* **2020**, *8*, 508.
[29] Y. Zhu, M. Yan, Y. Lasanajak, D.F. Smith, X. Song, *Carbohydr. Res.* **2018**, *464*, 19–27.
[30] S. Wang, C. Chen, M.R. Gadi, V. Saikam, D. Liu, H. Zhu, R. Bollag, K. Liu, X. Chen, F. Wang, P.G. Wang, P. Ling, W. Guan, L. Li, *Nat. Commun.* **2021**, 12, 3573
[31] Z. Xu, Y. Deng, Z. Zhang, W. Ma, W. Li, L. Wen, T. Li, *J. Am. Chem. Soc.* **2022**, *144* (20), 9057-9065.
[32] M. Busch, N. Simic, E. Ahlberg, *Phys. Chem. Chem. Phys.* **2019**, *21* (35), 19342–19348.
[33] G. Zauner, C.A.M. Koeleman, A.M. Deelder, M. Wuhrer, *Biochim. Biophys. Acta* **2012**, *1820* (9), 1420–1428.
[34] G. Yu, Y. Zhang, Z. Zhang, L. Song, P. Wang, W. Chai, *Anal. Chem.* **2010**, *82* (22), 9534–9542.
[35] E.M. Reuven, S.L. Ben-Arye, T. Marshanski, M.E. Breimer, H. Yu, I. Fellah-Hebia, J.-C. Roussel, C. Costa, M. Galiñanes, R. Mañez, T.L. Tourneau, J.-P. Soulillou, E. Cozzi, X. Chen, V. Padler-Karavani, *Xenotransplantation* **2016**, *23* (5), 381–392.
[36] J.P. Kamerling, R. Schauer, A.K. Shukla, S. Stoll, H. Van Halbeek, J.F. Vliegenthart, *Eur. J. Biochem.* **1987**, *162* (3), 601–607.
[37] V. Vandamme-Feldhaus, R.J. Schauer, *Biochem. (Tokyo)* **1998**, *124* (1), 111–121.
[38] T. Stelmaszynska, J.M. Zgliczynski, *Eur. J. Biochem.* **1978**, *92* (1), 301–308.

Supplementary information

Materials and methods

Bovine submaxillary mucin (BSM), trifluoroacetic acid (TFA) and formic acid were obtained from Sigma-Aldrich (Saint Louis, MO, US). Acetonitrile (ACN; LC-MS grade) was purchased from Biosolve B.V. (Valkenswaard, The Netherlands). Hypercarb™ Hypersep™ porous graphitized carbon solid phase extraction (PGC-SPE) cartridges with a bed weight of 25 mg and NaOCl (10-15% active chlorine) were acquired from Thermo Fisher Scientific (Waltham, MA, US). Ultrapure water was produced by a Synergy UV water purification system from Merck Millipore (Burlington, MA, US).

Release of *O*-glycan from glycopeptides 1-4.

The pH of a 15% NaOCl solution was adjusted to the corresponding pH by adding 1M HCl to 5 mL of the hypochlorite solution and adjusting the volume to 8 mL with water. To 50 µL glycopeptide solution (1 mg/mL), 25 µL NaOCl solution was added and the reaction mixture was placed on ice. AT 10 min intervals, a 5 µL volume of the mixture was transferred to a new eppendorf tube and the reaction was quenched with 3 µL 1% formic acid. The mixture was directly freeze dried and then purified with PGC-SPE.

Release of *O*-glycan from BSM.

The pH of a 15% NaOCl solution was adjusted to 6.8 by adding approximately 3 mL of 1M HCl to 5 mL of the hypochlorite solution. To a 50 µL solution of BSM (1 mg/mL), 25 µL of the acidified NaOCl solution (pH 6.8) was added and the reaction mixture was kept on ice. At 30-min intervals, a sample of 300 µL of the reaction mixture was taken and the reaction was quenched with 3 µL 1% (v/v) formic acid. The mixture was directly freeze dried and then purified with PGC-SPE.

PGC-SPE purification of released glycans.

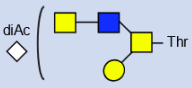
SPE cartridges were equilibrated with water. Aqueous samples were loaded onto the cartridge, the cartridge was washed with 1 mL water and glycans were eluted with 1 mL ACN:0.1% TFA 60%:40% (v/v). The eluent was evaporated under a stream of nitrogen and reconstituted for liquid chromatography (LC)-MS analysis.

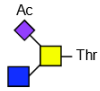
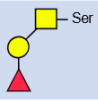
LC-MS analysis.

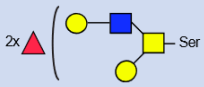
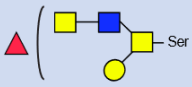
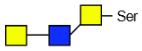
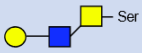
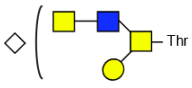
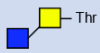
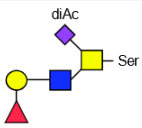
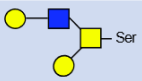
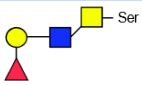
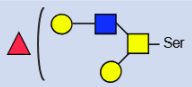
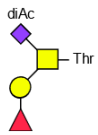
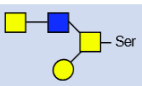
LC-MS was performed on an Agilent Technologies (Santa Clara, CA, US) Infinity 1290 LC system coupled via a dual-source AJS electrospray interface to an Agilent Technologies 6560B ESI Ion Mobility Q-TOF. Glycopeptide standards were analyzed with a SeQuant ZIC-HILIC column (20x2.1 mm, 3.5 µm particles; Merck, Darmstadt, Germany), with 0.1 % (v/v) formic acid as eluent A and ACN as eluent B, using a linear gradient from 90-50% B over 5 min and maintaining 50% B for 8 min. MS was performed in positive ion mode for glycans **1** and **3** and negative ion mode for glycans **2** and **4**. Released *O*-glycans from BSM were separated with a ZIC-HILIC column (150 x 2.1 mm, 3 µm particles) with

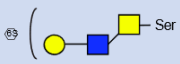
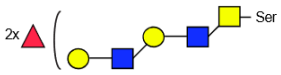
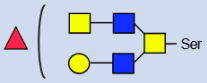
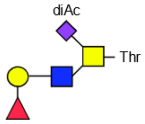
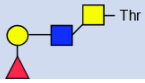
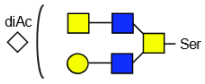
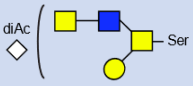
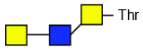
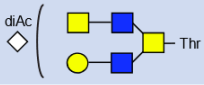
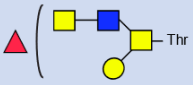
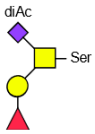
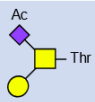
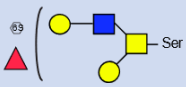
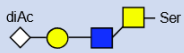
a ZIC-HILIC guard column (20 x 2.1 mm, 3 µm particles; Merck, Darmstadt, Germany) using 5 mM ammonium formate pH 5 as eluent A and ACN as eluent B. Chromatographic separation was achieved using 85% B for 5 min, followed by a linear gradient to 50% B over 25 min at 0.2 mL/min. MS analysis was performed in negative ion mode with a capillary voltage of 3500 V, nozzle voltage of 2000 V, nebulizer pressure of 40 psi, drying gas flow rate of 300 °C at 8L/min and sheath gas temperature of 300 °C at 11 L/min.

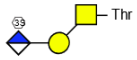
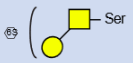
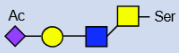
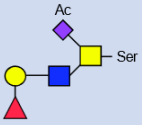
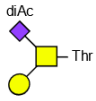
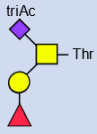
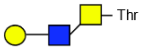
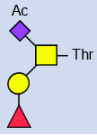
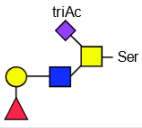
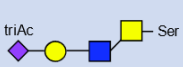
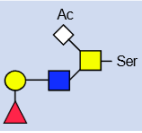
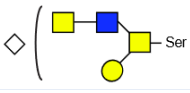
Table I: Structures of *O*-glycans released from BSM wit neutralized hypochlorite. AA=amino acid (T=threonine, S=serine), Hex=hexose, Fuc=fucose, Neu5Ac= N-Acetylneuraminic acid, Neu5Gc= N-Glycolylneuraminic acid,OAc=O-acetylated, HexA=hexosamine, Sulf=sulfated.

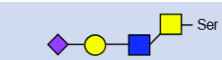
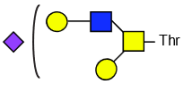
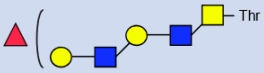
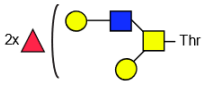
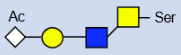
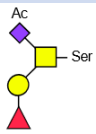
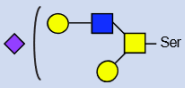
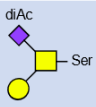
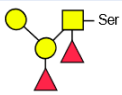
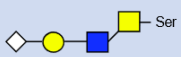
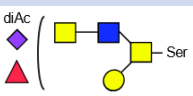
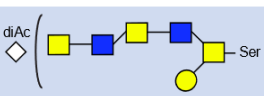
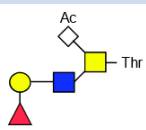
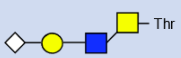
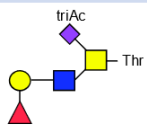
Suggested structure	Observed *m/z* $[M-H]^-$	Calculated *m/z* $[M-H]^-$	Relative abundancy (%)	AA	Hex	HexNAc	Fuc	Neu5Ac	Neu5Gc	OAc	HexA	Sulf
	625.2105	625.2092	15.1	T	0	1	0	1	0	1	0	0
	667.2211	667.2198	10.5	T	0	1	0	1	0	2	0	0
	611.1948	611.1936	7.50	S	0	1	0	1	0	1	0	0
	653.2053	653.2042	5.98	S	0	1	0	1	0	2	0	0
	641.2055	641.2041	5.46	T	0	1	0	0	1	1	0	0
	870.3014	870.2992	4.34	T	0	2	0	1	0	2	0	0
	709.231	709.2304	4.09	T	0	1	0	1	0	3	0	0
	856.2857	856.2836	3.95	S	0	2	0	1	0	2	0	0
	583.1997	583.1986	3.68	T	0	1	0	1	0	0	0	0
	599.1947	599.1935	3.08	T	0	1	0	0	1	0	0	0
	1251.4298	1251.4263	2.67	T	1	3	0	0	1	2	0	0
	569.1838	569.1830	2.47	S	0	1	0	1	0	0	0	0
	695.2161	695.2148	2.32	S	0	1	0	1	0	3	0	0

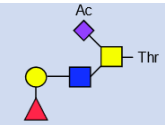
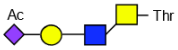
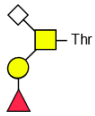
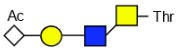
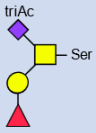
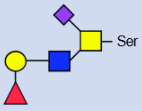
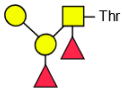
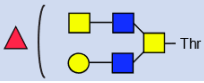
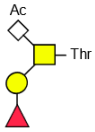
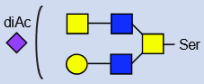
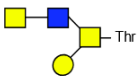
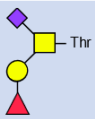
	814.2747	814.2730	1.95	S	0	2	0	1	0	1	0	0
	725.2264	725.2253	1.84	T	0	1	0	0	1	3	0	0
	828.2906	828.2886	1.59	T	0	2	0	1	0	1	0	0
	683.2161	683.2147	1.52	T	0	1	0	0	1	2	0	0
	585.1788	585.1779	1.39	S	0	1	0	0	1	0	0	0
	786.2796	786.2780	1.26	T	0	2	0	1	0	0	0	0
	830.27	830.2679	1.19	S	0	2	0	0	1	1	0	0
	772.2643	772.2624	1.05	S	0	2	0	1	0	0	0	0
	802.2748	802.2729	1.03	T	0	2	0	0	1	0	0	0
	586.1991	586.1983	1.02	S	1	1	1	0	0	0	0	0
	912.3121	912.3098	0.97	T	0	2	0	1	0	3	0	0
	928.3072	928.3047	0.93	T	0	2	0	0	1	3	0	0
	627.1902	627.1885	0.84	S	0	1	0	0	1	1	0	0
	711.2111	711.2097	0.79	S	0	1	0	0	1	3	0	0
	898.2968	898.2942	0.78	S	0	2	0	1	0	3	0	0
	788.2593	788.2573	0.75	S	0	2	0	0	1	0	0	0
	844.2858	844.2835	0.71	T	0	2	0	0	1	1	0	0
	481.1677	481.1670	0.69	S	0	2	0	0	0	0	0	0

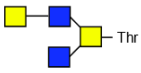
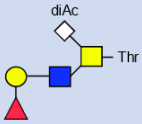
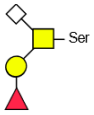
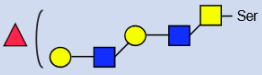
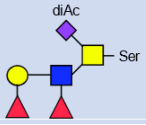
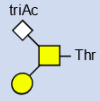
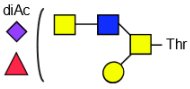
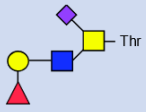
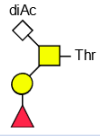
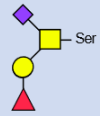
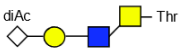
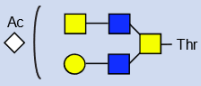
	886.2968	886.2941	0.66	T	0	2	0	0	1	2	0	0
	669.2008	669.1991	0.66	S	0	1	0	0	1	2	0	0
	872.2809	872.2785	0.52	S	0	2	0	0	1	2	0	0
	1097.3909	1097.3884	0.50	S	2	2	2	0	0	0	0	0
	914.2915	914.2891	0.44	S	0	2	0	0	1	3	0	0
	992.36	992.3571	0.37	S	1	3	1	0	0	0	0	0
	684.2477	684.2463	0.33	S	0	3	0	0	0	0	0	0
	643.2211	643.2198	0.28	S	1	2	0	0	0	0	0	0
	1167.4091	1167.4051	0.27	T	1	3	0	0	1	0	0	0
	495.1836	495.1826	0.22	T	0	2	0	0	0	0	0	0
	1164.3977	1164.3943	0.17	S	1	2	1	1	0	2	0	0
	805.2743	805.2726	0.15	S	2	2	0	0	0	0	0	0
	789.2789	789.2777	0.15	S	1	2	1	0	0	0	0	0
	951.3324	951.3305	0.15	S	2	2	1	0	0	0	0	0
	975.3333	975.3306	0.14	T	1	1	1	1	0	2	0	0
	846.3009	846.2992	0.13	S	1	3	0	0	0	0	0	0
	1018.3384	1018.3364	0.13	S	1	2	0	1	0	2	0	0

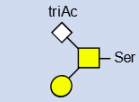
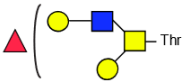
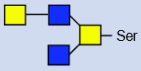
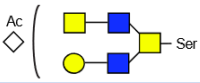
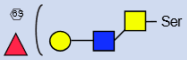
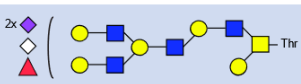
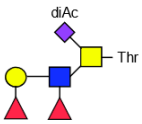
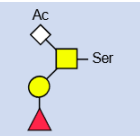
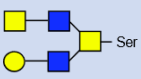
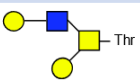
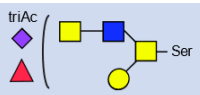
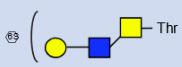
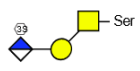
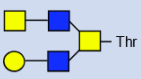
	723.1776	723.1766	0.12	S	1	2	0	0	0	0	0	1
	1300.471	1300.4678	0.11	S	2	3	2	0	0	0	0	0
	1195.4398	1195.4364	0.11	S	1	4	1	0	0	0	0	0
	1178.4123	1178.4100	0.11	T	1	2	1	1	0	2	0	0
	803.2954	803.2933	0.10	T	1	2	1	0	0	0	0	0
	1440.4927	1440.4900	0.10	S	1	4	0	0	1	2	0	0
	1237.4146	1237.4107	0.098	S	1	3	0	0	1	2	0	0
	698.2635	698.2620	0.084	T	0	3	0	0	0	0	0	0
	1454.5076	1454.5057	0.083	T	1	4	0	0	1	2	0	0
	761.2478	761.2464	0.082	T	1	1	0	0	1	0	0	0
	1006.3648	1006.3727	0.077	T	1	3	1	0	0	0	0	0
	961.3174	961.3149	0.077	S	1	1	1	1	0	2	0	0
	787.2627	787.2621	0.076	T	1	1	0	1	0	1	0	0
	1032.353	1032.3520	0.073	T	1	2	0	1	0	2	0	0
	1031.2891	1031.2873	0.069	S	2	2	1	0	0	0	0	1
	745.2523	745.2515	0.066	T	1	1	0	1	0	0	0	0
	1034.3325	1034.3313	0.064	S	1	2	0	0	1	2	0	0

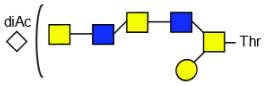
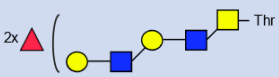
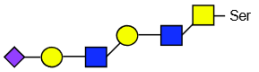
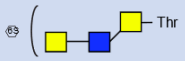
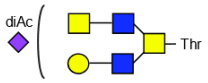
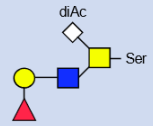
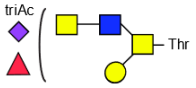
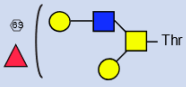
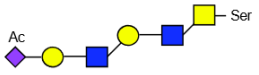
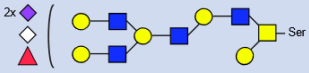
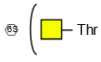
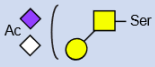
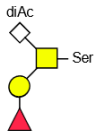
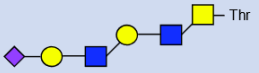
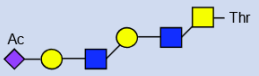
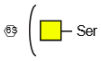
	710.1433	710.1450	0.064	T	1	1	0	0	0	0	1	1
	520.0981	520.0972	0.060	S	1	1	0	0	0	0	0	1
	803.2581	803.2570	0.057	T	1	1	0	0	1	1	0	0
	976.3271	976.3258	0.055	S	1	2	0	1	0	1	0	0
	600.2144	600.2140	0.054	T	1	1	1	0	0	0	0	0
	1122.3852	1122.3837	0.053	S	1	2	1	1	0	1	0	0
	829.2739	829.2727	0.051	T	1	1	0	1	0	2	0	0
	1017.3435	1017.3412	0.046	T	1	1	1	1	0	3	0	0
	657.2359	657.2354	0.044	T	1	2	0	0	0	0	0	0
	933.3217	933.3200	0.041	T	1	1	1	1	0	1	0	0
	1206.4056	1206.4049	0.039	S	1	2	1	1	0	3	0	0
	1060.3482	1060.3470	0.038	S	1	2	0	1	0	3	0	0
	731.2363	731.2358	0.037	S	1	1	0	1	0	0	0	0
	1138.3802	1138.3786	0.036	S	1	2	1	0	1	1	0	0
	1153.393	1153.3895	0.036	S	1	3	0	0	1	0	0	0

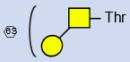
	934.316	934.3152	0.034	S	1	2	0	1	0	0	0	0
	1110.3849	1110.3837	0.032	T	2	2	0	1	0	0	0	0
	1168.4214	1168.4255	0.031	T	2	3	1	0	0	0	0	0
	1111.4054	1111.4041	0.029	T	2	2	2	0	0	0	0	0
	992.3218	992.3207	0.029	S	1	2	0	0	1	1	0	0
	919.3063	919.3043	0.027	S	1	1	1	1	0	1	0	0
	1096.3694	1096.3680	0.027	S	2	2	0	1	0	0	0	0
	1074.3639	1074.3626	0.027	T	1	2	0	1	0	3	0	0
	815.2582	815.2570	0.027	S	1	1	0	1	0	2	0	0
	894.312	894.3090	0.025	S	2	1	2	0	0	0	0	0
	950.3121	950.3101	0.025	S	1	2	0	0	1	0	0	0
	1367.475	1367.4737	0.024	S	1	3	1	1	0	2	0	0
	1643.5718	1643.5694	0.024	S	1	5	0	0	1	2	0	0
	1152.3955	1152.3942	0.023	T	1	2	1	0	1	1	0	0
	964.3268	964.3257	0.023	T	1	2	0	0	1	0	0	0
	1220.4209	1220.4206	0.024	T	1	2	1	1	0	3	0	0

	1136.4007	1136.3994	0.023	T	1	2	1	1	0	1	0	0
	990.3423	990.3414	0.022	T	1	2	0	1	0	1	0	0
	440.1406	440.1404	0.020	S	1	1	0	0	0	0	0	0
	907.3058	907.3043	0.020	T	1	1	1	0	1	0	0	0
	789.2433	789.2413	0.019	S	1	1	0	0	1	1	0	0
	1006.3378	1006.3363	0.019	T	1	2	0	0	1	1	0	0
	1003.3272	1003.3255	0.019	S	1	1	1	1	0	3	0	0
	747.2309	747.2307	0.019	S	1	1	0	0	1	0	0	0
	1080.3735	1080.3731	0.019	S	1	2	1	1	0	0	0	0
	908.3265	908.3247	0.018	T	2	1	2	0	0	0	0	0
	1209.4533	1209.4521	0.018	T	1	4	1	0	0	0	0	0
	949.3164	949.3149	0.017	T	1	1	1	0	1	1	0	0
	1424.4936	1424.4951	0.016	S	1	4	0	1	0	2	0	0
	860.3152	860.3148	0.016	T	1	3	0	0	0	0	0	0
	891.3102	891.3094	0.016	T	1	1	1	1	0	0	0	0

	901.3424	901.3413	0.016	T	0	4	0	0	0	0	0	0
	1194.4035	1194.4048	0.015	T	1	2	1	0	1	2	0	0
	893.29	893.2886	0.015	S	1	1	1	0	1	0	0	0
	1154.411	1154.4099	0.015	S	2	3	1	0	0	0	0	0
	1076.3427	1076.3419	0.015	S	1	2	0	0	1	3	0	0
	1310.4546	1310.4522	0.014	S	1	2	2	1	0	2	0	0
	1094.3536	1094.3524	0.014	T	1	1	0	1	1	1	0	0
	887.2808	887.2782	0.012	T	1	1	0	0	1	3	0	0
	1381.4892	1381.4893	0.012	T	1	3	1	1	0	2	0	0
	1094.3883	1094.3888	0.012	T	1	2	1	1	0	0	0	0
	991.3266	991.3255	0.011	T	1	1	1	0	1	2	0	0
	877.2949	877.2937	0.011	S	1	1	1	1	0	0	0	0
	1048.3489	1048.3469	0.011	T	1	2	0	0	1	2	0	0
	1412.4957	1412.4951	0.011	T	1	4	0	0	1	1	0	0
	948.332	948.3308	0.011	T	1	2	0	1	0	0	0	0

	873.2608	873.2625	0.010	S	1	1	0	0	1	3	0	0
	965.3472	965.3462	0.010	T	2	2	1	0	0	0	0	0
	887.3264	887.3257	0.0099	S	0	4	0	0	0	0	0	0
	1398.4803	1398.4794	0.0097	S	1	4	0	0	1	1	0	0
	869.2373	869.2345	0.0095	S	1	2	1	0	0	0	0	1
	1090.3589	1090.3575	0.0095	T	1	2	0	0	1	3	0	0
	1291.9372 [M-2H]2-	2584.8917	0.0092	T	4	4	1	2	1	0	0	0
	1324.4697	1324.4679	0.0088	T	1	2	2	1	0	2	0	0
	935.3001	935.2992	0.0086	S	1	1	1	0	1	1	0	0
	764.2051	764.2031	0.0086	S	0	3	0	0	0	0	0	1
	1049.3795	1049.3785	0.0084	S	1	4	0	0	0	0	0	0
	819.2885	819.2883	0.0083	T	2	2	0	0	0	0	0	0
	704.2403 [M-2H]2-	704.2385 [M-2H]2-	0.0082	S	1	3	1	1	0	3	0	0
	883.251	883.2501	0.0079	T	1	2	1	0	0	0	0	1
	737.1926	737.1922	0.0076	T	1	2	0	0	0	0	0	1
	696.1263	696.1293	0.0066	S	1	1	0	0	0	0	1	1
	1063.3954	1063.3942	0.0063	T	1	4	0	0	0	0	0	0

	1657.5857	1657.5850	0.0058	T	1	5	0	0	1	2	0	0
	1314.4854	1314.4834	0.0057	T	2	3	2	0	0	0	0	0
	649.22 [M-2H]2-	649.2201 [M-2H]2-	0.0056	S	2	3	0	1	0	0	0	0
	778.2194	778.2188	0.0052	T	0	3	0	0	0	0	0	1
	718.7516 [M-2H]2-	718.7518 [M-2H]2-	0.0052	T	1	4	0	1	0	2	0	0
	1180.391	1180.3892	0.0050	S	1	2	1	0	1	2	0	0
	711.2465 [M-2H]2-	711.2463 [M-2H]2-	0.0043	T	1	3	1	1	0	3	0	0
	1045.3048	1045.3030	0.0042	T	2	2	1	0	0	0	0	1
	670.2241 [M-2H]2-	670.2254 [M-2H]2-	0.0039	S	2	3	0	1	0	1	0	0
	1284.9305 [M-2H]2-	1284.934 [M-2H]2-	0.0039	S	4	4	1	2	1	0	0	0
	372.0602	372.0600	0.0034	T	0	1	0	0	0	0	0	1
	1080.3393	1080.3367	0.0034	S	1	1	0	1	1	1	0	0
	977.3106	977.3098	0.0031	S	1	1	1	0	1	2	0	0
	1313.4643	1313.4630	0.0029	T	2	3	0	1	0	0	0	0
	454.1563	454.1561	0.0028	T	1	1	0	0	0	0	0	0
	677.2337 [M-2H]2-	677.2332 [M-2H]2-	0.0021	T	2	3	0	1	0	1	0	0
	358.0443	358.0444	0.0019	S	0	1	0	0	0	0	0	1

	534.1141	534.1129	0.0013	T	1	1	0	0	0	0	0	1
	773.2485	773.2464	0.00093	S	1	1	0	1	0	1	0	0

CHAPTER 5

Sialic acid *O*-Acetylation Pattern and Glycosidic Linkage Type can be Determined by Ion Mobility-Mass Spectrometry

Gaël M. Vos, Kevin C. Hooijschuur, Zeshi J. Li, John Fjeldsted, Christian Klein, Robert P. de Vries, Javier Sastre Toraño, Geert-Jan Boons

Introduction

Sialic acids are negatively charged nine-carbon monosaccharides that are often part of complex glycans of higher animals.[1] Several pathogenic microorganisms also express sialylated glycoconjugates which are used for molecular mimicry to evade host immune detection.[2,3] Sialoglycans regulate many biological processes and can function as receptors for many pathogens including viruses, bacteria and protozoa. Sialylation of glycans also modulates half-life, immunogenicity, and efficacy of biologicals.[4–8]

N-Acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc) are major forms of sialic acid expressed by mammals. They can be modified by acetyl esters at the 4-, 7-, 8-, and/or 9-position to give as many as fifteen different patterns of *O*-acetylation. Further structural diversity comes from different Neu5Ac and Neu5Gc glycosidic linkage types and the most common ones are α2,3-linked to galactose (Gal), α2,6-linked to Gal and *N*-acetyl-galactosamine (GalNAc) and α2,8-linked to another sialic acid. The resulting glycotopes can be presented at different underlying glycan moieties and can be part of *N*- and *O*-linked glycans, glycolipids as well as free floating glycans such as human milk oligosaccharides. The expression of sialic acids is regulated in a developmental and tissue-specific manner. Furthermore, there are marked differences in the expression of sialoglycans in different species, which likely is due to evolutionary pressure evoked by host–pathogen interactions.[9]

A growing body of literature associates sialic acid *O*-acetylation with various diseases including cancer, immune disorders, and infection.[10–13] *O*-acetylation can preclude recognition by glycan binding proteins such as the Siglec immune-receptors and complement protein factor H and as a result can function as a molecular switch.[14] It also substantially reduces the rate of hydrolysis by several human endogenous sialidases thereby regulating properties of glycoconjugates such as turnover and degradation.[15] In addition, it can block the activity of bacterial sialidases and by this means protect the integrity of the epithelial mucus barrier. *O*-acetylated sialic acids serve also as receptors for many viruses including embecoviruses (family Coronaviridae), toroviruses (Tobaniviridae), and influenza C and D viruses (Orthomyxoviridae).[16] Recent analysis of receptor specificities of these viruses infecting different spiecies of animals and humans demonstrated host-specific patterns of receptor recognition in relationship to both the pattern of acetylation and glycosidic linkage type. It was found that human respiratory viruses uniquely bind 9-*O*-acetylated α2,8-linked disialoside found on glycolipids.[16]

Despite advances, the roles of distinct *O*-acetylated sialosides in health and disease remain difficult to explore. A major hurdle is a lack of convenient experimental approaches to determine exact structures of the sialic acid variants including the pattern of acetylation and glycosidic linkage type.[17] These molecules are chemically labile and prone to acetyl ester migration and hydrolysis complicating isolation and characterization.[18] *O*-acetylated sialosides have been released from glycoconjugates by treatment with acetic acid at elevated temperatures, labelled with 1,2-dihydroxy-4,5-methylendioxybenzol (DMB) and then analyzed by high-performance liquid chromatography (HPLC) or liquid chromatography-mass spectrometry (LC-MS).[19] Although these methods are sensitive, they do not provide information about glycosidic linkage type. In another approach, carboxylic acids of sialosides of *N*-linked glycan were modified as methylamines to increase the stability and then analyzed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS.[20] This approach only provides compositions and cannot assign positions of acetyl esters and

glycosidic linkage type. Soluble hemagglutinin-esterases have been used as lectins for tissue staining to qualitative assess acetyl ester expression.[21–23] Although powerful, virolectins exhibit promiscuous binding behavior and cannot detect all common *O*-acetylation patterns.[16]

Here, we describe a drift tube ion-mobility (IM)-MS approach that can elucidate exact *O*-acetylation patterns as well as glycosidic linkage types of sialosides isolated from complex biological samples. IM can separate gas-phase ions based on their mobility in an electric field through a gas-filled drift cell, resulting in distinctive arrival time distributions (ATD).[24] The arrival times can be converted into collision cross sections (CCS) which are intrinsic values of ions, and in combination with *m/z* values provide molecular descriptors for reliable compound identification. Isomeric glycans can have unique conformational properties and as a result may have distinct CCS values. [25–28] Therefore, IM-MS has the potential to determine exact structures of glycans. The challenge of implementing IM-MS for exact glycan structure determination is the need of standards to determine CCS values of diagnostic fragment ions.

In this study, we employed a large panel of synthetic *O*-acetylated *N*-acetyl and *N*-glycolylneuraminic acids to establish CCS values of diagnostic B_1 and B_3 fragment ions.[16] We demonstrate that the library of CCS values can be employed to determine the exact pattern of *O*-acetylation and glycosidic linkage type of sialosides isolated from complex biological samples. An important aspect of the approach was the implementation of methods that can release *N*- and *O*-glycans without affecting acetyl esters. *N*-linked glycans could be released by treatment with PNGase or ENDO-F2 under neutral or slightly acidic conditions whereas *O*-glycans could be oxidatively cleaved by sodium hypochlorite at pH 6.8. The approach can be employed for the analysis of biotherapeutics for quality control and be used to examine structures of *N*- and *O*-linked sialosides obtained from tissue samples and secreted mucins. The uncovering of sialylation provides insights in the biosynthesis of this class of compounds. Furthermore, analysis of equine upper airway tissue uncovered contrasting sialic acid linkage types of acetylated and non-acetylated sialic acids which provided a rationale for sialic acid binding preferences of equine H7 influenza A viruses.

Results and discussion

IM-MS of *O*-acetylated sialosides

Previously, we developed a methodology to synthesize sialosides that differ in the pattern of *O*-acetylation and glycosidic linkage type.[16] It is based on the chemical synthesis of 2,3-, 2,6- and 2,8-linked sialoglycans having acetyl esters at *C*-4, *C*-7 and *C*-9 (Fig. 1A). These compounds were treated with hemagglutinin-esterases (HE) from bovine coronavirus (BCoV) or mouse hepatitis virus strain S (MHV-S), which cleave acetyl esters on the *C*-4 and *C*-9 position of sialic acid, respectively and in combined with controlled acetyl ester migration from *C*-7 to *C*-9 could readily be diversified to give a large panel of compounds (Fig. 1B, compounds **1**-**27**). In this study, the resulting sialoglycans were employed as standards to establish a library of CCS values of informative fragment ions for compound identification. Each compound was analyzed on an Agilent Technologies 1290 LC system equipped with a SeQuant ZIC-HILIC LC column, connected to an Agilent Technologies 6560B drift tube ion mobility/QTOF MS instrument. Measurements were performed in positive ion mode and in-source collisionally activated dissociation was used by applying a fragmentor voltage of 600V to achieve glycosidic bond fragmentation and produce informative B-ions (Fig. 1C). The fragment ions were analyzed by IM-MS using nitrogen as buffer gas. ATDs of standards were obtained in triplicate using drift tube IM with ion multiplexing and associated demultiplexing[29] combined with high-resolution demultiplexing to improve resolution for the separation of isomeric fragment ions.[30] CCS values of the ions were calculated from their arrival times (Fig. 2 and Table I).[31]

B_1 ions of the three mono-acetylated Neu5Ac types could be resolved with the fragment ions from 4-*O*-acetylated Neu5Ac having the smallest CCS, followed by the fragment ions of 7- and 9-*O*-acetylated Neu5Ac. The B_1 ions for the Neu5Gc derivatives showed a similar trend although the B_1 ions from 4- and 7-*O*-acetylated Neu5Gc had a smaller difference in CCS value. B_1 ions could also distinguish 4,9- and 7,9-di-*O*-acetylated Neu5Ac and Neu5Gc. In the case of Neu5Ac the fragment ion from the 7,9-isomer was the smaller isomer whereas the opposite was observed for Neu5Gc with the fragment ion from the 4,9-isomer having the smaller CCS. As anticipated, the CCS values of B_1 ions derived from α2,3- and α2,6-linked sialosides, having the same acetylation pattern, as well as those from α2,8-linked Neu5Ac derivatives **25**-**27** were in close agreement.

B_3 ions also provide informative structural information, and it was found that those of the α2,3-linked structures have longer drift times and thus higher CCS values compared to the corresponding α2,6-linked compounds, making it possible to determine sialic acid glycosidic linkage type. Furthermore, most B_3 ions having the same number of acetyl esters could be separated enabling determination of *O*-acetylation positions. Only the B_3 ions derived from the 7- and 9-*O*-acetylated α2,6-linked Neu5Ac-LacNAc (**2** and **4**) and Neu5Gc-LacNAc (**14** and **16**) could not be sufficiently resolved for direct unambiguous *O*-acetyl position determination. De-*O*-acetylation, which can result in misassignment of structures was minimally observed (<1%).

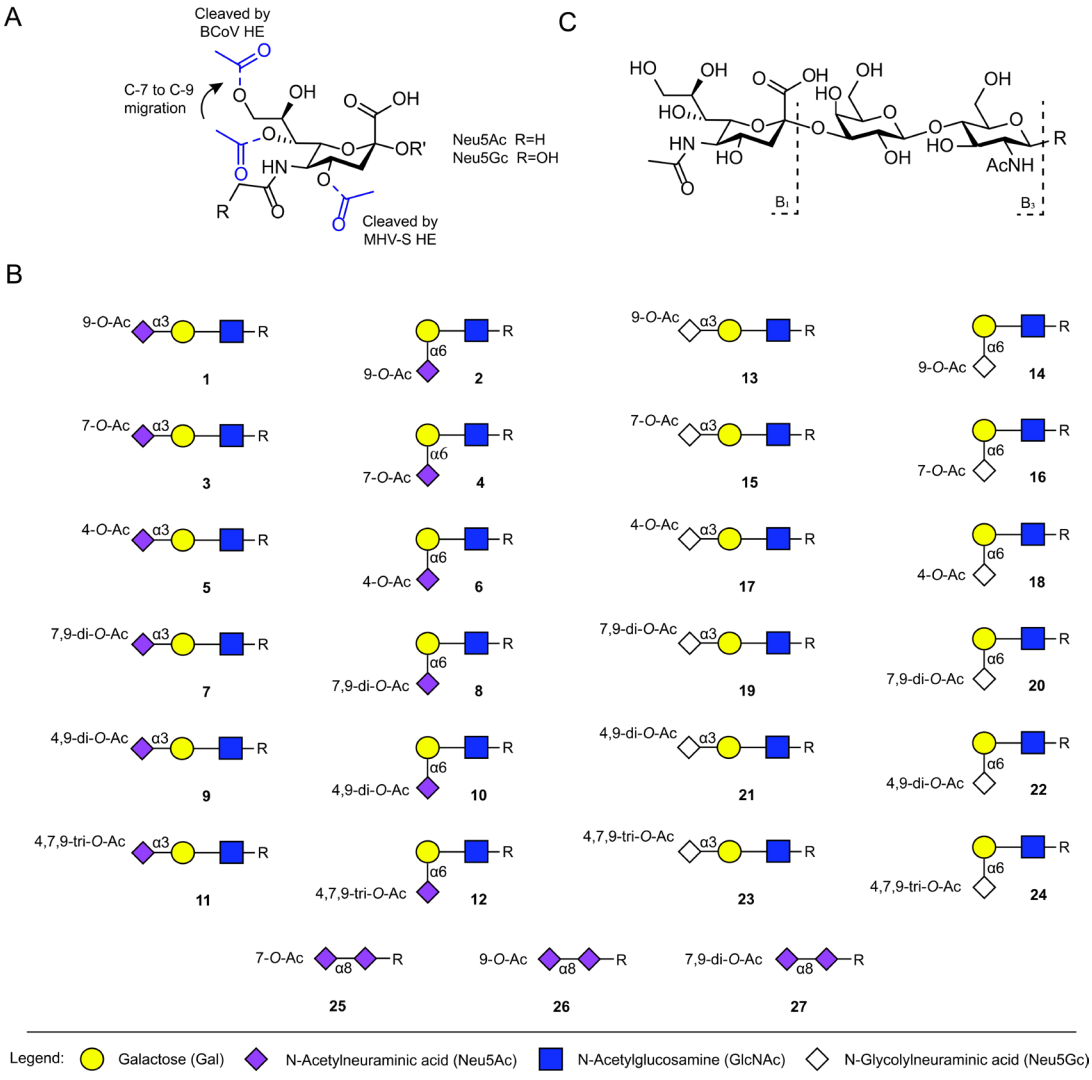


Figure 1: A) Structure of *O*-acetylated Neu5Ac and Neu5Gc. Annotated to the structure are the reactions that enable the controlled synthesis of well-defined O-acetylation patterns. B) Library of synthetic glycan standards containing *O*-Acetylated Neu5Ac (**1**-**12**), Neu5Gc (**13**-**24**) and di-Neu5Ac (**25**-**27**); R= pentylaminobiotin. C) Nomenclature of glycan fragment ions.[32]

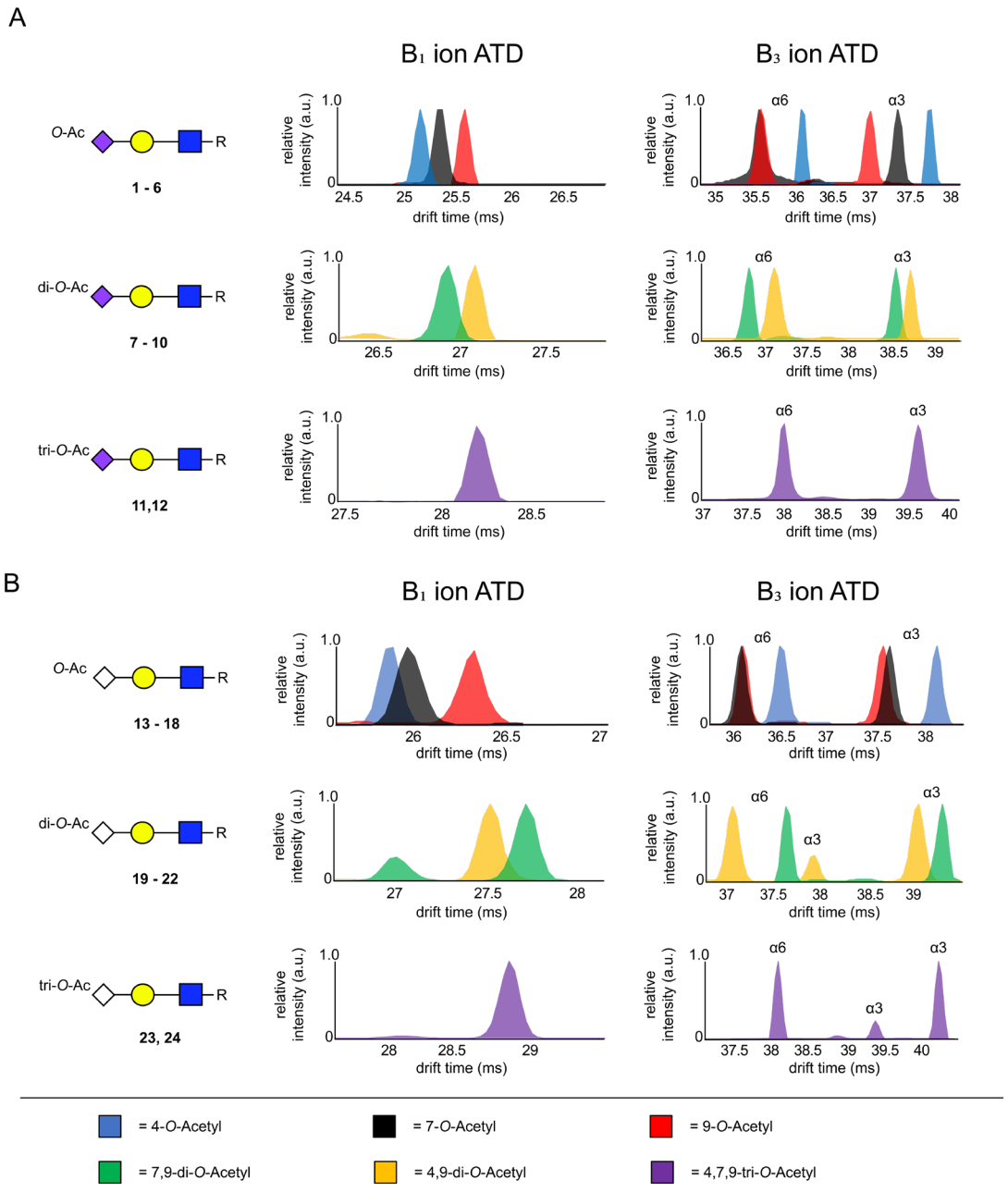


Figure 2: A) Combined ATDs of single charged B_1 and B_3 fragment ions of mono- (B_1 at *m/z* 334.1134, B_3 at *m/z* 699.2455), di- (B_1 at *m/z* 376.1243, B_3 at *m/z* 741.2560) and tri-*O*-Acetylated Neu5Ac (B_1 at *m/z* 418.1348, B_3 at *m/z* 783.2666). B) Combined ATDs of single charged B_1 and B_3 fragment ions of mono- (B_1 at *m/z* 350.1090, B_3 at *m/z* 715.2404), di- (B_1 at *m/z* 392.1192, B_3 at *m/z* 757.2509) and tri-*O*-Acetylated Neu5Gc (B_1 at *m/z* 434.1293, B_3 at *m/z* 799.2615).

Table I: Average CCS values of B_1 and B_3 ions of compounds **1-24** (n=3).

Compound	Sialic acid	Sialic acid linkage	Acetyl ester position	*m/z* B_1-ion	*m/z* B_3-ion	CCS B_1-ion (Å²)	CCS B_3-ion (Å²)
1	Neu5Ac	α2,3	9	334.1134	699.2450	174.77(±0.05)	248.26(±0.08)
2	Neu5Ac	α2,6	9	334.1134	699.2450	174.82(±0.06)	238.48(±0.04)
3	Neu5Ac	α2,3	7	334.1134	699.2450	173.34(±0.02)	250.74(±0.07)
4	Neu5Ac	α2,6	7	334.1134	699.2450	173.23(±0.04)	238.38(±0.11)
5	Neu5Ac	α2,3	4	334.1134	699.2450	171.90(±0.01)	253.59(±0.02)
6	Neu5Ac	α2,6	4	334.1134	699.2450	171.91(±0.08)	242.24(±0.15)
7	Neu5Ac	α2,3	7,9	376.1243	741.2561	183.53(±0.07)	258.75(±0.06)
8	Neu5Ac	α2,6	7,9	376.1243	741.2561	183.36(±0.02)	246.26(±0.08)
9	Neu5Ac	α2,3	4,9	376.1243	741.2561	184.48(±0.04)	259.88(±0.11)
10	Neu5Ac	α2,6	4,9	376.1243	741.2561	184.29(±0.03)	248.73(±0.08)
11	Neu5Ac	α2,3	4,7,9	418.1348	783.2663	191.52(±0.05)	265.83(±0.04)
12	Neu5Ac	α2,6	4,7,9	418.1348	783.2663	191.52(±0.06)	254.13(±0.02)
13	Neu5Gc	α2,3	9	350.1090	715.2401	179.30(±0.11)	253.37(±0.03)
14	Neu5Gc	α2,6	9	350.1090	715.2401	179.30(±0.05)	242.17(±0.12)
15	Neu5Gc	α2,3	7	350.1090	715.2401	176.69(±0.11)	253.37(±0.03)
16	Neu5Gc	α2,6	7	350.1090	715.2401	176.89(±0.06)	241.62(±0.07)
17	Neu5Gc	α2,3	4	350.1090	715.2401	176.30(±0.05)	257.20(±0.03)
18	Neu5Gc	α2,6	4	350.1090	715.2401	176.42(±0.23)	244.93(±0.01)
19	Neu5Gc	α2,3	7,9	392.1192	757.2510	187.02(±0.03)	261.56(±0.09)
20	Neu5Gc	α2,6	7,9	392.1192	757.2510	186.98(±0.07)	247.88(±0.02)
21	Neu5Gc	α2,3	4,9	392.1192	757.2510	188.31(±0.06)	263.43(±0.09)
22	Neu5Gc	α2,6	4,9	392.1192	757.2510	188.25(±0.05)	251.71(±0.05)
23	Neu5Gc	α2,3	4,7,9	434.1293	799.2594	195.12(±0.11)	268.92(±0.02)
24	Neu5Gc	α2,6	4,7,9	434.1293	799.2594	195.22(±0.07)	254.64(±0.06)

O-acetylation of *N*-linked sialosides derived from biologicals

The database of ATDs and CCS values was employed to analyze *O*-acetylated sialosides of two biologicals that are produced in Chinese hamster ovary (CHO) cell lines. These cell lines are commonly employed for the expression of therapeutic proteins and have been reported to modify *N*-linked glycans with *O*-acetylated sialosides.[33] The pattern of *O*-acetylation can, however, differ between CHO cell lines and change during purification or storage.[34] Myozyme (Genzyme) is an alpha-glucosidase employed for enzyme replacement therapy for patients with Pompe disease and has at least 6 confirmed *N*-glycosylation sites. Both mono- and di-*O*-acetylated sialic acids have been reported on the *N*-glycans of Myozyme.[35] Aflibercept (Regeneron) is a fusion protein of the extracellular domain of the human VEGF receptor modified and the Fc region of human IgG1 and used for the treatment of metastatic colorectal cancer. Aflibercept has four *N*-glycosylation sites on the VEGF receptor domain and one on its Fc domain which predominantly carry biantennary *N*-glycans.[36] The recombinant glycoproteins were dialyzed to remove additives and then the *N*-glycans were released enzymatically by treatment with ENDO-F2 under mild acidic conditions

(100 mM NaOAc, pH 4.5) to prevent acetyl ester migration and hydrolysis. ENDO-F2 does not cleave high-mannose-, hybrid- and multi-antennary *N*-glycans and only releases the abundant biantennary complex *N*-glycans. A total of 29 biantennary *N*-glycans were identified for Myozyme and 21 for Aflibercept (supplementary information, Table I and II). The MS data for Myozyme revealed the presence of *O*-acetyl modified sialosides on ~3% of the total *N*-glycan content, which corresponds to ~15% of the sialylated *N*-glycans. For five *N*-glycans, the measured molecular weights indicated the presence of mono- and di-*O*-acetylated sialic acids (Fig. 3B). The CCS of the B_1 fragment ions confirmed the presence of 7- and 9-*O*-mono- and 7,9-di-*O*-acetylated Neu5Ac. B_3 fragment ions could be detected for the majority of the mono-*O*-acetylated *N*-glycans and confirmed these contain 7- or 9-*O*-acetylation. The CCS of B_3 fragment ions related to α2,3-linked sialosides. The CCS of B_1 fragment ions detected for three di-*O*-acetylated *N*-glycans correlated with the 7,9-di-*O*-acetylation pattern in our database. CCS values of B_3 fragment ions could be determined for the two most abundant structures confirming α2,3-linkage types. These observation are in agreement with the fact that CHO cells only express α2,3-linked sialosides.[34,35,37–39]

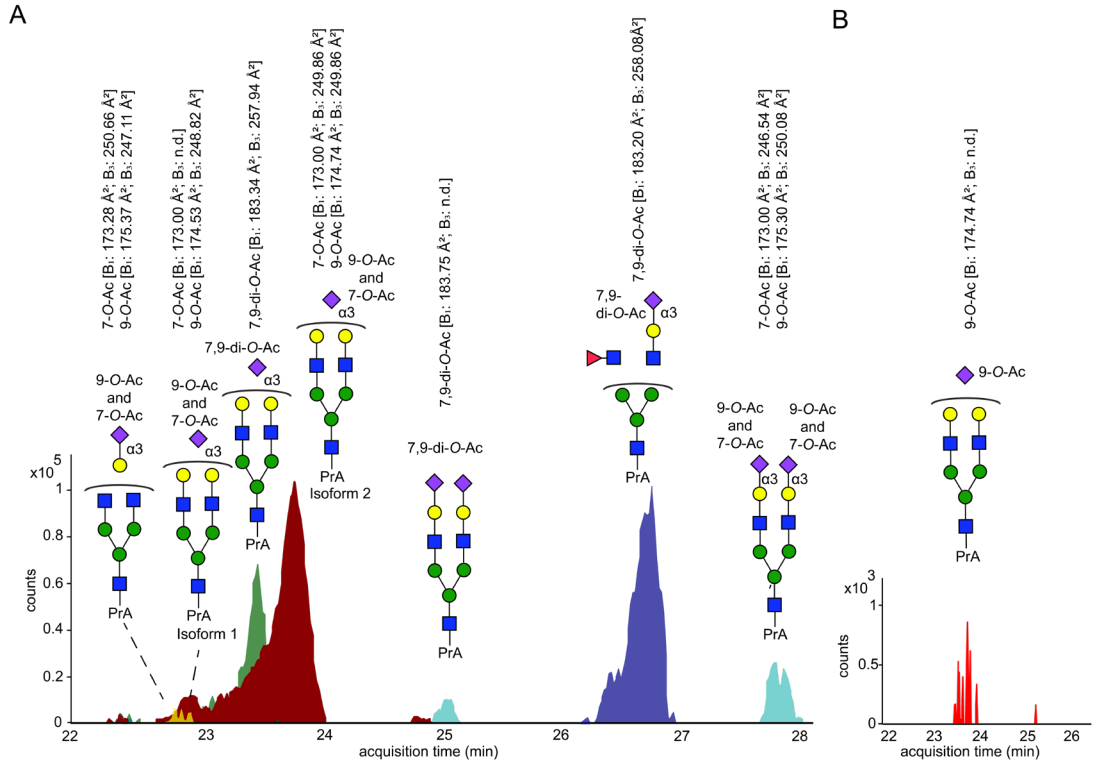


Figure 3: Extracted ion chromatograms of glycans released from Myozyme (A) from left to right at *m/z* 925.8682 ($[M+H+Na]^{2+}$, yellow), *m/z* 1006.8946 ($[M+H+Na]^{2+}$, burgundy), *m/z* 1027.8999 ($[M+H+Na]^{2+}$, green), *m/z* 1173.4476 ($[M+H+Na]^{2+}$,cyan), *m/z* 1030.8934 *m/z* ($[M+2Na]^{2+}$, dark blue) and Aflibercept (B) (at *m/z* 1017.8856 ($[M+2Na]^{2+}$). ENDO F2-released glycans were derivatized with procainamide (PrA) at the reducing end and separated by LC using a ZIC-HILIC (150x4.6 mm, 3.5 µm) column, with a linear gradient from 80% acetonitrile/20% H_2O (0.1% formic acid; %v/v) to 50% acetonitrile/50% H_2O (0.1% formic acid; %v/v) over 30 minutes. The peaks were assigned to structures using high resolution MS and sialic acid linkage and acetyl ester positions were determined by the CCS values of B_1 and B_3 ions; n.d.=not determined.

The MS data showed only one *O*-acetylated *N*-glycan in the Aflibercept sample. Ion mobility analysis of the B_1 ion revealed that this *N*-glycan contains exclusively 9-*O*-acetylated sialic acid (Fig. 3A). Due to the low abundancy, no B_3 fragment ions could be detected and therefore the sialic acid linkage type could not be established. Collectively, the results demonstrate that IM-MS can assign both acetyl ester position and sialic acid linkage type of *N*-glycans released from biologicals using CCS values of B_1 and B_3 fragment ions. Previously, such analysis required a combination of nuclear magnetic resonance and chromatography of hydrolyzed sialic acids.[40]

_O_-acetylation of _N_-linked sialosides from equine tracheal and nasal tissues.

Respiratory viruses, which cause enormous disease burden, commonly employ glycans for cell attachment and/or entry.[41] The relentless pressure of microbial infections at the mucosal interface has driven the evolution of host and pathogen.[42] It has shaped the glycomes of the host and even close related species express substantially different glycans. In turn, pathogens evolved glycan receptor specificities that determine host range and tissue tropism. To understand this co-evolution, it is critical to determine exact structures of glycans of respiratory tissues. Previous tissue staining of equine respiratory tract tissues with the plant lectins SNA and MAH revealed similar levels of α2,3- and α2,6-linked sialic acids.[43] Furthermore, the 4-*O*-acetyl specific lectin MHV-S showed abundant presence of this sialoglycan.[43] Although MHV-S has an obligatory requirement for sialate-4-*O*-acetylation, it tolerates the presence of acetyl esters at C-7 and C-9, and can bind to 2,3- as well as 2,6-linked sialosides. Thus, this lectin cannot determine fine structural details of O-acetylated sialosides. Therefore, *N*-glycans obtained from equine tracheal and nasal tissues were analyzed by the IMS-MS approach.

N-glycans from equine tracheal and nasal tissues were released using PNGaseF in TRIS buffer (100 mM, pH 7.0). MS analysis of the nasal sample revealed that whilst only 4 of the 32 detected *N*-glycans were modified by an *O*-acetylated sialoside, they account for over 30% of the total *N*-glycan content (supplementary information, Table IV). Two di-sialylated *N*-glycans having one or two acetyl esters accounted for more than 90% of the observed *O*-acetylated ions. In tracheal tissue, acetyl esters were observed in 6 of 19 detected *N*-glycans accounting for >50% of the total *N*-glycan content (supplementary information, Table V). In both samples, bi- tri- and tetra-antennary *N*-glycans could be detected, but acetyl esters were only observed on biantennary *N*-glycans. *O*-acetylation of these *N*-glycans was mainly found on Neu5Ac but also *O*-acetylated Neu5Gc was detected (approximately 3% of total *N*-glycan content in trachea). According to the CCS values of the B_1 fragment ions, all *O*-acetylated sialic acids are exclusively 4-*O*-acetylated (Fig. 4). CCS values of the B_3 ions confirmed the *O*-acetylation assignment and revealed the presence of both α2,3- and α2,6-linked Neu5Ac and Neu5Gc derivatives. B_3 ions corresponding to α2,6-linked-Neu4,5diAc were abundantly present, whereas those for α2,3-linked sialosides were only sparsely detected (Fig. 4AB).[44] For mono-acetylated Neu5Gc, only B_3 ions that matched with a α2,6 linkage configuration ($^{DT}CCS_{N2}$: 245.16 Å^2) were detected. In contrast, the B_3 ions of non-acetylated sialosides corresponded almost exclusively to α2,3-linked sialic acid ($^{DT}CCS_{N2}$: 242.99 Å^2). In tracheal tissue, 28% of all Neu5Gc was 4-*O*-acetylated compared to 60% for Neu5Ac. In nasal tissue, no 4-*O*-acetylated Neu5Gc was detected, whilst 47% of all Neu5Ac was modified by 4-*O*-acetylation. Thus, IM-MS analysis indicates that the prevalence of C-4 acetylation

is dependent on sialoside type and is most abundantly present on α2,6-linked Neu5Ac. Although the acetyl transferase for the biosynthesis of 4-*O*-acetylated sialic acids has not been identified yet, it is thought that the transfer occurs at the sugar nucleotide level.[45] The observation that the specific sialosides are preferentially modified by acetyl esters indicates that different sialyl transferases have distinct tolerances for *O*-acetylated CMP-sialic acid.[46]

Surprisingly, it was observed that core-fucosylation greatly reduced the modification by O-acetylated sialosides and only 6-7% of the *O*-acetylated sialosides contained core-fucose *vs.* 67% of non-acetylated structures in the nasal as well as tracheal tissue samples. Core-fucosylation was completely absent when both sialic acids were *O*-acetylated. This observation suggests a possible biosynthetic bifurcation between core-fucosylation and 4-*O*-acetylation in equine tissues. The findings are consistent with the notion that *O*-acetylated sialic acid content is regulated by glycosyltransferase activity.

_O_-acetylation of _N_-linked sialosides from equine α2-macroglobulin

Equine α2-macroglobulin is a serum protein that has inhibitory activity for human influenza A virus (IAV) infections,[40,47] which is associated with 4-O-acetylation of sialic acid on *N*-glycans. It appears that 4-O-acetylation does not increase the affinity for hemagglutinin of IAV but confers resistance to cleavage by viral neuraminidases, and as a result can function as a decoy receptor.[48] Human H3N2 variants can become equine serum resistant by losing binding affinity to 4-O-acetylated sialic acid, indicating that 4-O-acetylation of sialic acid can assert selective pressure and possibly prevents cross-species transmission.[49] IM-MS analysis of *N*-glycans released from equine α2-macroglobulin revealed the presence 50 *N*-glycans (supplementary information, Table III), 16 of which are modified by acetyl esters. Among these *O*-acetylated *N*-glycans are 3 structures that contain *O*-acetylated Neu5Gc which have not previously been detected in equine α2-macroglobulin. CCS values of the B_1 ions of O-acetylated Neu5Ac and Neu5Gc corresponded to the standards containing an acetyl ester at the C-4 position. Analysis of B_3 ions confirmed that both *O*-acetylated and non-acetylated Neu5Ac and Neu5Gc were almost exclusively present in a α2,6-linkage type. Using IM-MS, we observed that equine α2-macroglobulin is exclusively modified by α2,6-linked sialosides and that all observed acetyl esters are at the C-4 position, the detected linkage type is in agreement with previously reported analysis by nuclear magnetic resonance.[40] Additionally, we observe the same 10-fold reduction in core-fucosylation among *O*-acetylated *N*-glycans as we detected in equine upper airway tissues.

Glycan microarray analysis of receptor specificities of influenza A viruses

The IM-MS approach revealed that horse upper airway tissues abundantly express 4-O-acetylated α2,6-linked Neu5Ac receptors.[48] In addition, it is known that neuraminidases of IAVs cannot cleave sialic acids that are modified by a acetyl ester on the *C*-4 position, and the utility of such a receptor may hamper viral egress. Therefore, we expect that HAs of equine IAVs have evolved not to interact with 4-O-acetylated sialoside. Rather, equine IAVs appear to use the less abundant α2,3-linked-Neu5Gc as receptor.[50] The N-glycans of equine α2-macroglobulin are differently modified compared to those of upper airway tissues, and abundantly express α2,6-linked unacetylated Neu5Ac. Equine α2-macroglobulin is a decoy

receptor for many IAVs, including those infecting humans. To gain further insight in receptor specificities of HAs and link these to receptor expression, we screened several HAs on a glycan microarray populated with the O-acetylated sialoglycan shown in Fig 1B. We screened the now extinct, but highly pathogenic equine H7N7 IAV, an equine H3N8 IAV and three human H3N2 IAV.

The 4-O-acetylated sialoforms we observe in equine α2-macroglobulin can act as decoy receptors for human H3N2 IAV strains. These IAV were introduced in the human population in 1968 as the Hong Kong strain.[48] By glycan microarray, we evaluate the ability of this Hong Kong strain, and more recent H3N2 strains, to bind O-acetylated sialosides. We selected and evaluated three time separated human H3N2 IAVs: HK68, NL03 and CH13 (supplementary information Fig. S1). We find that the strong interaction of HK68 to acetylated sialic acids is specific for 4-O-acetylated Neu5Ac and does not depend on linkage type. This interaction is completely lost in later strains, which suggests that the inhibition of more recent human H3N2 IAV by equine α2-macroglobulin is only mediated by the non-acetylated-α2,6-linked sialic acids that act as decoy receptors. The ability of 4-O-acetylated sialic acids detected on equine α2-macroglobulin to act as neuraminidase resistant decoy receptors appears to be lost during H3N2 evolution.

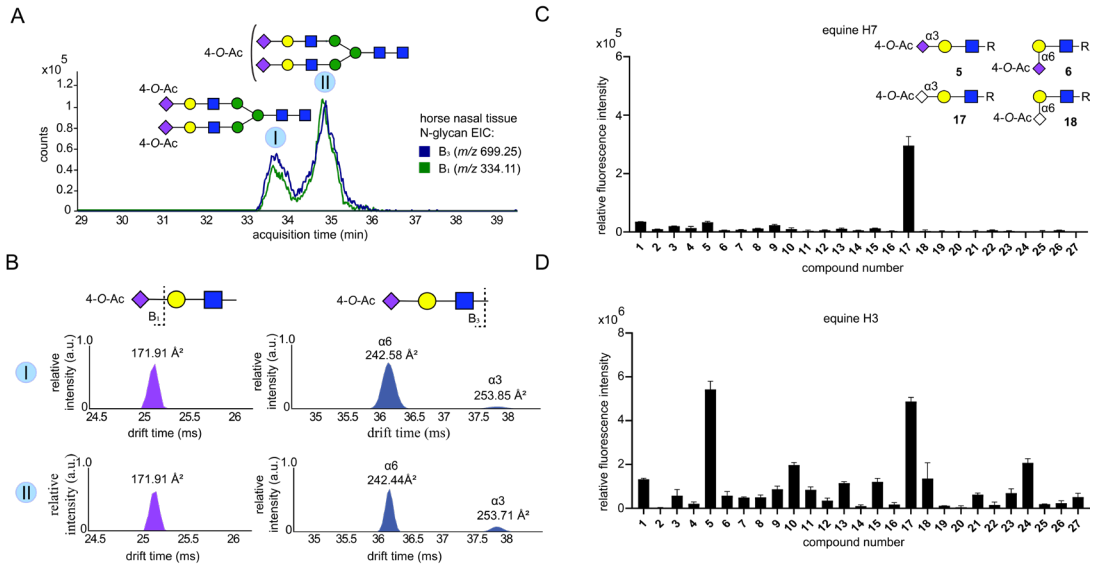


Figure 4: A) Extracted ion chromatograms of B_3 ions of PNGaseF released glycans from equine nasal tissue, analyzed by LC-IM-MS using a ZIC-HILIC column (150x4.6 mm, 3.5 µm particles) and eluting with a linear gradient from 80% acetonitrile/20% H_2O (0.1% formic acid; %v/v) to 50% acetonitrile/50% H_2O (0.1% formic acid) over 30 minutes. B) ATDs of single charged B_1 and B_3 fragment ions of di- (I) and mono-acetylated *N*-glycans (I) released from equine nasal tissue. The *O*-acetylated *N*-glycan structures were assigned using accurate mass values; Sialic acid linkage and acetyl ester positions were determined by the CCS values of B_1 and B_3 ions. C) Glycan microarray analysis of the HA ectodomain of equine H7(A/Equine/New York/49/73 H7N7, GenBank accession no. LC414434).[50] D) Glycan microarray analysis of the HA ectodomain of equine H3(A/Equine/Miami/1/1963 H3N8, GenBank accession no. AAA43105.1).

O-acetylation of *O*-linked sialoglycan from mucins

Next, the scope of the IM-MS approach was extended to the analysis of acetylated *O*-glycans. Glycomics of this compound class is usually performed by base-mediated beta-elimination to release a reducing glycan that is *in-situ* either reductively labeled or reduced to the corresponding alditol.[51] The employed alkaline conditions will hydrolyze acetyl esters making the analysis of acetylated *O*-sialoglycans of *O*-glycans very challenging. In chapter 4, we optimized an oxidative release method for *O*-glycans under neutral conditions, and it was anticipated this approach will preserve *O*-acetylated sialoglycans. In this approach, *O*-glycans are released by hypochlorite that is neutralized to pH 6.8 to prevent degradation of glycan structures. Thus, bovine submaxillary mucin (BSM) was incubated for 60 min with a 3% hypochlorite solution that was neutralized by the addition of 1 M HCl to pH 6.8. The released glycans were isolated by solid phase extraction using porous graphitized carbon and subjected to IM-MS analysis, which identified 65 *O*-glycans (supplementary information Table V). Sialic acids were detected on 39 glycan structures (25 Neu5Ac and 14 Neu5Gc) of which 21 were *O*-acetylated. We found 70% of all Neu5Ac to be *O*-acetylated which is higher than previously reports, although this difference can potentially be attributed to sample variations.[52] Mono-acetylated sialic acids were detected on 14 *O*-glycans corresponding to a relative abundancy of 29%. On 17 *O*-glycans, two acetyl esters could be detected which corresponds to 21% of the glycan abundancy. The oligosaccharides consist mainly of di- and trisaccharides and therefore we could identify the core structures of these *O*-glycans based on compositions. It should, however, be noted that core 3 and core 5 structures could not discriminated, which have both been reported to be present in BSM.[53] It is to be expected that the synthesis of core 3 and core 5 glycan standards will facilitate the development of IM-MS based methodology for these isomeric structures.

IM-MS analysis of the B_1 ions from acetylated sialic acid (at *m/z* 334.11) revealed a doublet of peaks corresponding to 7-*O*-acetyl and 9-*O*-acetyl Neu5Ac (Fig. 5B). The high abundance of 7-*O*-acetylated Neu5Ac compared to the C-9 isomer, such as STn antigen, is in agreement with previous findings that sialic acids in BSM are primarily acetylated on the C-7 hydroxyl, while 9-*O*-acetylation may be acquired via acetyl ester migration.[54] We found differences in the ratio of 7/9-acetylation among different sialylated structures, which can be due to differences in rate of acetyl ester migration or differences in preference of sialyltransferase for *O*-acetylated donors. *O*-glycans biosynthesized by ST6GalNAc-I (STn antigen) had more 7-*O*-acetylation compared to *O*-glycans modified by ST6GalNAc-II (Sialyl core 1 and Sialyl core 3). Additionally, sialyl core 3 was mostly mono-acetylated while STn was abundantly non- and di-*O*-acetylated. This observation is in agreement with a potential preference of ST6GalNAc-II for CMP-Neu5,7-di-Ac. No 4-*O*-acetyl B_1 fragments were observed, which was expected since the 4-sialate-*O*-acetyltransferase has not been identified in bovine, and acetylation on the C-4 hydroxyl cannot be achieved through migration.[54] The B_1 ion of di-*O*-acetylated Neu5Ac corresponded exclusively to acetylation of the C-7 and C-9 hydroxyls (Fig. 5C). B_1 ions of acetylated Neu5Gc were not detected in sufficient quantities to confidently assign acetylation position.

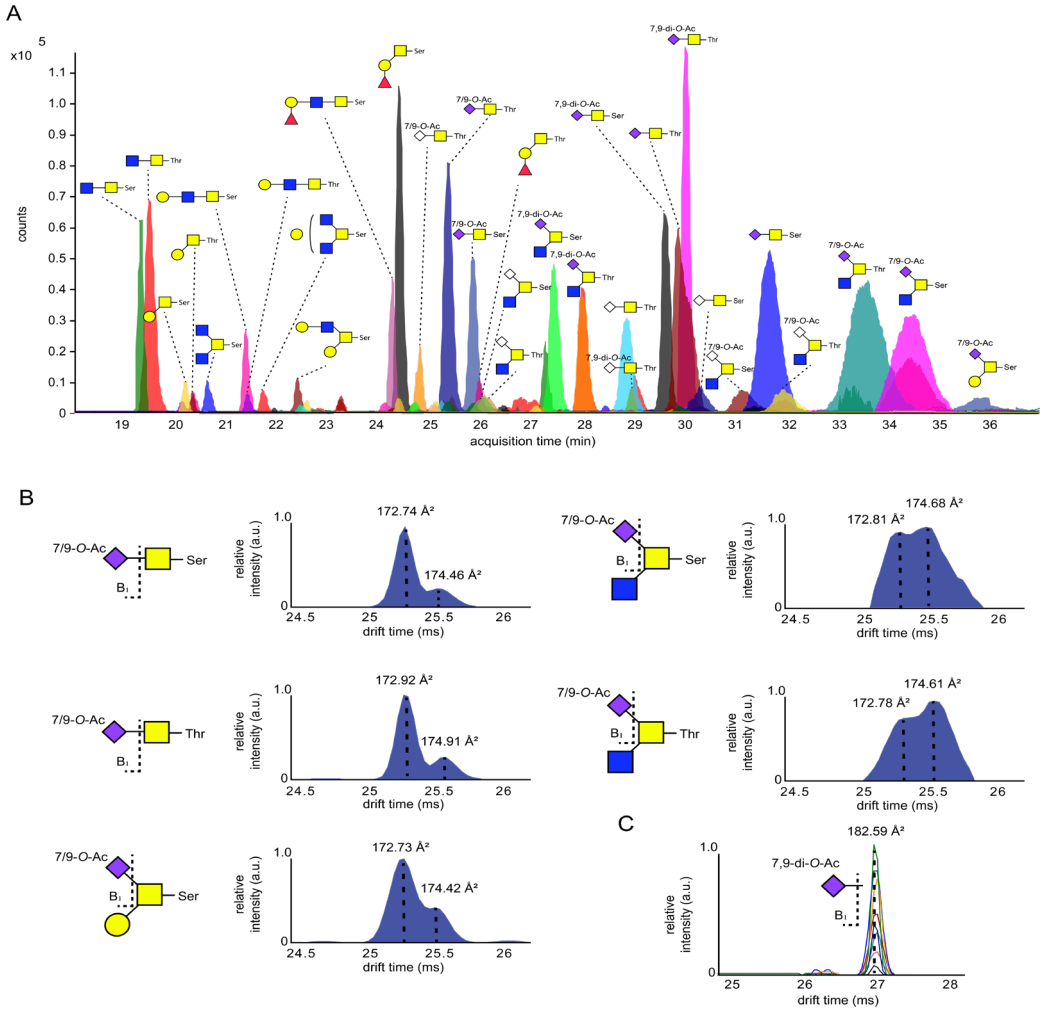


Figure 5: A) Structures of abundant O-glycans (>0.1% relative concentration) of bovine submaxillary mucin (BSM) released with neutralized hypochlorite at pH 6.8. Released *O*-glycans were analyzed by LC-IM-MS using a ZIC-HILIC column (150 x 2.1 mm, 3 µm particles). Chromatographic separation was achieved with an initial isocratic hold of 85% acetonitrile/15% H_2O (0.1% formic acid; %v/v) for 5 min followed by a linear gradient from 85% acetonitrile/15% H_2O (0.1% formic acid; %v/v) to 50% acetonitrile/50% H_2O (0.1% formic acid) over 30 minutes. B) ATDs and CCS of O-acetylated Neu5Ac B_1 ion of selected O-glycan structures of BSM. ATD in blue. Left dashed line corresponds to the B_1 ion from the 7-O-acetylated Neu5Ac standard. Right dashed line corresponds to the B_1 ion from 9-O-acetylated Neu5Ac standard. C) stacked ATDs and CCS of B1 ions from di-O-acetylated Neu5Ac of BSM. Dashed line corresponds to the B_1 ion from the 7,9-diO-acetylated Neu5Ac standard.

Conclusions

We developed an IM-MS method that can determine the exact *O*-acetylation pattern and glycosidic linkage type of glycans obtained from complex biological samples. Well-defined, synthetic *O*-acetylated standards made it possible to develop a database of *m/z* and CCS values for the characteristic B_1 and B_3 fragment ions for both *O*-acetylated Neu5Ac and *O*-acetylated Neu5Gc containing 4-*O*-acetyl, 7-*O*-acetyl or 9-*O*-acetyl and combinations thereof. The CCS values were employed to assign both *O*-acetylation positions as well as sialic acid linkages of *N*-glycans of biologicals and upper airway tissue samples, demonstrating the applicability of the methodology for analysis if complex biological samples. *O*-acetylation positions detected with IM-MS were in agreement with previous described observations and B_3 ions from glycans on CHO expressed biotherapeutics matched the expected α2,3 linkage. We observed that *O*-acetylated sialic acids found on the *N*-glycans of equine upper airway tissues were predominantly α2,6 linked, oppose to the α2,3 linkage of sialic acids devoid of acetyl esters, indicating that 4-*O*-acetylated sialic acid donors can be differently accepted by sialic acid transferases. Using glycan microarray, we show that equine H3 and H7 IAV can interact with 4-*O*-acetylated sialic acids but not to the α2,6 linkage type expressed in equine upper airway tissue. The IM-MS methodology was also successfully applied to the analysis of *O*-glycans derived from mucins, confirming that the developed neutralized hypochlorite workflow limits acetyl ester migration and can preserve labile 7-*O*-acetylated sialic acids.

It is expected that the IM-MS method can be widely applied to many types of samples and glycan classes that contain acetyl esters on sialic acids by using CCS values, without the further need for *O*-acetylated glycan standards.

References

[1] A. Varki, Trends Mol. Med. 2008, 14, 351–360.

[2] A. F. Carlin, S. Uchiyama, Y.-C. Chang, A. L. Lewis, V. Nizet, A. Varki, Blood 2009, 113, 3333–3336.

[3] A. P. Moran, M. M. Prendergast, B. J. Appelmelk, FEMS Immunol. Med. Microbiol. 1996, 16, 105–115.

[4] B. Gao, C. Long, W. Lee, Z. Zhang, X. Gao, D. Landsittel, M. Ezzelarab, D. Ayares, Y. Huang, D. K. C. Cooper, Y. Wang, H. Hara, PLoS ONE 2017, 12, e0180768.

[5] R. Jefferis, Trends Pharmacol. Sci. 2009, 30, 356–362.

[6] M. J. Borrok, S. T. Jung, T. H. Kang, A. F. Monzingo, G. Georgiou, ACS Chem. Biol. 2012, 7, 1596–1602.

[7] Z. Wang, Z. S. Chinoy, S. G. Ambre, W. Peng, R. McBride, R. P. de Vries, J. Glushka, J. C. Paulson, G.-J. Boons, Science 2013, 341.

[8] Eur. J. Pharm. Biopharm. 2016, 100, 94–100.

[9] A. Varki, Proc. Natl. Acad. Sci. U. S. A. 2010, 107, 8939–8946.

[10] N. M. Varki, A. Varki, Lab. Invest. 2007, 87, 851–857.

[11] A. K. Chava, M. Chatterjee, C. Mandal, in Handb. Carbohydr. Eng., CRC Press, 2005.

[12] S. Cavdarli, J. H. Dewald, N. Yamakawa, Y. Guérardel, M. Terme, J.-M. Le Doussal, P. Delannoy, S. Groux-Degroote, Glycoconj. J. 2019, 36, 79–90.

[13] V. S. Mahajan, S. Pillai, Immunol. Rev. 2016, 269, 145–161.

[14] W. X. Shi, R. Chammas, N. M. Varki, L. Powell, A. Varki, J. Biol. Chem. 1996, 271, 31526–31532.

[15] W. H. Yang, P. V. Aziz, D. M. Heithoff, M. J. Mahan, J. W. Smith, J. D. Marth, Proc. Natl. Acad. Sci. U. S. A. 2015, 112, 13657–13662.

[16] Z. Li, Y. Lang, L. Liu, M. I. Bunyatov, A. I. Sarmiento, R. J. de Groot, G. J. Boons, Nat. Chem. 2021, 13, 496–503.

[17] C. J. Gray, L. G. Migas, P. E. Barran, K. Pagel, P. H. Seeberger, C. E. Eyers, G.-J. Boons, N. L. B. Pohl, I. Compagnon, G. Widmalm, S. L. Flitsch, J. Am. Chem. Soc. 2019, 141, 14463–14479.

[18] J. P. Kamerling, R. Schauer, A. K. Shukla, S. Stoll, H. Van Halbeek, J. F. Vliegenthart, Eur. J. Biochem. 1987, 162, 601–607.

[19] S. Hara, M. Yamaguchi, Y. Takemori, K. Furuhata, H. Ogura, M. Nakamura, Anal. Biochem. 1989, 179, 162–166.

[20] Z. Wu, H. Li, Q. Zhang, X. Liu, Q. Zheng, J. Li, Sci. Rep. 2017, 7, 46206.

[21] M. Aamelfot, O. B. Dale, S. C. Weli, E. O. Koppang, K. Falk, Glycoconj. J. 2014, 31, 327–335.

[22] M. A. Langereis, M. J. G. Bakkers, L. Deng, V. Padler-Karavani, S. J. Vervoort, R. J. G. Hulswit, A. L. W. van Vliet, G. J. Gerwig, S. A. H. de Poot, W. Boot, A. M. van Ederen, B. A. Heesters, C. M. van der Loos, F. J. M. van Kuppeveld, H. Yu, E. G. Huizinga, X. Chen, A. Varki, J. P. Kamerling, R. J. de Groot, Cell Rep. 2015, 11, 1966–1978.

[23] C. Biswas, D. Sinha, C. Mandal, Mol. Immunol. 2000, 37, 745–754.

[24] A. B. Kanu, P. Dwivedi, M. Tam, L. Matz, H. H. Hill, J. Mass Spectrom. JMS 2008, 43, 1–22.
[25] A. Marchand, S. Livet, F. Rosu, V. Gabelica, Anal. Chem. 2017, 89, 12674–12681.
[26] J. Sastre Toraño, I. A. Gagarinov, G. M. Vos, F. Broszeit, A. D. Srivastava, M. Palmer, J. I. Langridge, O. Aizpurua-Olaizola, V. J. Somovilla, G.-J. Boons, Angew. Chem. Int. Ed Engl. 2019, 58, 17616–17620.
[27] J. Sastre Toraño, O. Aizpurua-Olaizola, N. Wei, T. Li, L. Unione, G. Jiménez-Osés, F. Corzana, V. J. Somovilla, J. M. Falcon-Perez, G.-J. Boons, Chem. Weinh. Bergstr. Ger. 2021, 27, 2149–2154.
[28] W. Winkler, W. Huber, R. Vlasak, G. Allmaier, Rapid Commun. Mass Spectrom. RCM 2011, 25, 3235–3244.
[29] A. Bilbao, B. C. Gibbons, S. M. Stow, J. E. Kyle, K. J. Bloodsworth, S. H. Payne, R. D. Smith, Y. M. Ibrahim, E. S. Baker, J. C. Fjeldsted, J. Proteome Res. 2021.
[30] J. C. May, R. Knochenmuss, J. C. Fjeldsted, J. A. McLean, Anal. Chem. 2020, 92, 9482–9492.
[31] S. M. Stow, T. J. Causon, X. Zheng, R. T. Kurulugama, T. Mairinger, J. C. May, E. E. Rennie, E. S. Baker, R. D. Smith, J. A. McLean, S. Hann, J. C. Fjeldsted, Anal. Chem. 2017, 89, 9048–9055.
[32] B. Domon, C. Costello, Glycoconj. J. 2005.
[33] W.-X. Shi, R. Chammas, A. Varki, J. Biol. Chem. 1996, 271, 15130–15138.
[34] R. I. Thomson, R. A. Gardner, K. Strohfeldt, D. L. Fernandes, G. P. Stafford, D. I. R. Spencer, H. M. I. Osborn, Anal. Chem. 2017, 89, 6455–6462.
[35] H. Park, S. You, J. Kim, W. Kim, J. Do, Y. Jang, D. Kim, J. Lee, J. Ha, D.-B. Oh, J. I. Kim, H. H. Kim, J. Pharm. Biomed. Anal. 2019, 169, 188–195.
[36] Z. Shen, Y. Wang, H. Xu, Q. Zhang, C. Sha, B. Sun, Q. Li, Int. J. Biol. Macromol. 2021, 180, 494–509.
[37] Z. Shahrokh, L. Royle, R. Saldova, J. Bones, J. L. Abrahams, N. V. Artemenko, S. Flatman, M. Davies, A. Baycroft, S. Sehgal, M. W. Heartlein, D. J. Harvey, P. M. Rudd, Mol. Pharm. 2011, 8, 286–296.
[38] M. Takeuchi, S. Takasaki, H. Miyazaki, T. Kato, S. Hoshi, N. Kochibe, A. Kobata, J. Biol. Chem. 1988, 263, 3657–3663.
[39] E. U. Lee, J. Roth, J. C. Paulson, J. Biol. Chem. 1989, 264, 13848–13855.
[40] K. Hanaoka, T. J. Pritchett, S. Takasaki, N. Kochibe, S. Sabesan, J. C. Paulson, A. Kobata, J. Biol. Chem. 1989, 264, 9842–9849.
[41] A. J. Thompson, R. P. de Vries, J. C. Paulson, Curr. Opin. Virol. 2019, 34, 117–129.
[42] N. M. Koropatkin, E. A. Cameron, E. C. Martens, Nat. Rev. Microbiol. 2012, 10, 323–335.
[43] B. R. Wasik, K. N. Barnard, R. J. Ossiboff, Z. Khedri, K. H. Feng, H. Yu, X. Chen, D. R. Perez, A. Varki, C. R. Parrish, mSphere 2017, 2.
[44] Y. Suzuki, T. Ito, T. Suzuki, R. E. Holland, T. M. Chambers, M. Kiso, H. Ishida, Y. Kawaoka, J. Virol. 2000, 74, 11825–11831.
[45] J. Tiralongo, H. Schmid, R. Thun, M. Iwersen, R. Schauer, Glycoconj. J. 2000, 17, 849–858.

[46] H. H. Higa, J. C. Paulson, J. Biol. Chem. 1985, 260, 8838–8849.
[47] B. Levinson, D. Pepper, G. Belyavin, J. Virol. 1969, 3, 477–483.
[48] T. J. Pritchett, J. C. Paulson, J. Biol. Chem. 1989, 264, 9850–9858.
[49] M. Matrosovich, P. Gao, Y. Kawaoka, J. Virol. 1998, 72, 6373–6380.
[50] C. M. Spruit, X. Zhu, I. Tomris, M. Ríos-Carrasco, A. X. Han, F. Broszeit, R. van der Woude, K. M. Bouwman, M. M. T. Luu, K. Matsuno, Y. Sakoda, C. A. Russell, I. A. Wilson, G.-J. Boons, R. P. de Vries, J. Virol. n.d., 96, e02120-21.
[51] H. Wilkinson, R. Saldova, J. Proteome Res. 2020, 19, 3890–3905.
[52] X. Song, H. Ju, Y. Lasanajak, M. R. Kudelka, D. F. Smith, R. D. Cummings, Nat. Methods 2016, 13, 528–534.
[53] J. Kim, C. Ryu, J. Ha, J. Lee, D. Kim, M. Ji, C. S. Park, J. Lee, D. K. Kim, H. H. Kim, Biomolecules 2020, 10, 636.
[54] G. V. Srinivasan, R. Schauer, Glycoconj. J. 2008 268 2008, 26, 935–944.

Supplementary information

Materials and methods

Aflibercept was purchased from Bayer (Leverkusen, Germany). Myozyme, Endoglycosidase F2 (Endo-F2) and Peptide:N-glycosidase F (PNGaseF) were produced in-house. Vivaspin2 10kDa spinfilters were obtained from Sartorius (Göttingen, Germany). Sodium hypochlorite 15% was acquired from Acros Organics. Trifluoroacetic acid (TFA), ethanol, dimethyl sulfoxide (DMSO), dichloromethane (DCM) and hydrochloric acid (HCl) were purchased from Merck (Darmstadt, Germany). Sodium acetate, sodium cyanoborohydride, tris(hydroxymethyl) aminomethane (TRIS), procainamide, acetic acid (LC-MS grade), formic acid (LC-MS grade), ammonium acetate and bovine submaxillary mucin (BSM) were acquired from Sigma-Aldrich (Saint Louis, MO). Hypercarb 25 mg porous graphitized carbon (PGC) solid phase extraction (SPE) cartridges were obtained from Thermo Fisher Scientific (Waltham, MA). Acetonitrile (LC-MS grade) was purchased from Biosolve B.V. (Valkenswaard, The Netherlands). Ultrapure water was produced by a Synergy UV water purification system from Merck Millipore (Burlington, MA).

General Methods

Release of *N*-glycans from biologicals

N-glycans were released enzymatically from the proteins Aflibercept and myozyme by Endo-F2. Aflibercept/Myozyme solutions (1 mg protein) were loaded on a 10kDa spinfilter and centrifuged till near dryness at 4500 RPM. A 4x 1-ml volume of 100 mM sodium acetate (pH 4.5) was added and the sample was centrifuged again (4500 RPM leaving at least 100 µL solution). The samples were diluted to 2 mg/mL with 100 mM sodium acetate (pH 4.5) and 10 µl EndoF2 (1 mg/mL) was added. The sample was incubated at 37 ºC for 16h under mild agitation and the purified by PGC SPE.

Release of *N*-glycans from equine upper airway tissue

N-glycans were released enzymatically from equine upper airway tissue by PNGaseF. Nasal epithelium/tracheal tissue (~1 cm^3) in ethanol was decanted and washed with 2x 20 mL water for 15 min. To each tissue 1ml 100 mM TRIS buffer (pH 7) and 10 µl PNGaseF (10 mg/mL) was added. Tissues were incubated for 72h at 37 ºC under mild agitation, then the samples were centrifuged for 15 min at 4500 RPM and the supernatant was collected, lyophilized, dissolved in a minimal volume of water and purified by PGC SPE.

Release of *N*-glycans from equine α2-macroglobulin

N-glycans were released from equine α2-macroglobulin by PNGaseF. Equine α2-macroglobulin (1 mg) was dissolved in 1ml 100mM TRIS buffer (pH 7) and 10 µL PNGaseF (10 mg/mL) was added. The reaction was incubated for 24h at 37 ºC under mild agitation and then purified by PGC SPE.

PGC SPE purification and derivatization of released *N*-glycans

The released *N*-glycans were purified by PGC SPE by equilibrating a cartridge with 1 mL of 0.1% TFA and loading the dissolved glycans on the cartridge. The cartridge was washed with 1 mL of 0.05% TFA followed by 1 mL 5%/95% ACN/water. Glycans were eluded with 50%/50% ACN/water and then the eluent was evaporated under a nitrogen flow, yielding N-glycans for derivatization.
The positively chargeable mass label procainamide was attached to *N*-glycan free reducing ends via reductive amination. A 50-µg amount of glycan was dissolved in 120 µL water and mixed with 40 µL labelling solution (32.5 µg/µL procainamide HCl and 75 µg/µL sodium cyanoborohydride in DMSO) and 23 µL acetic acid. The mixture was vortexed and incubated at room temperature for 4 hours to prevent desialylation.[54] The reaction mixture was evaporated under nitrogen flow, redissolved in a small volume of water and extracted with 3x 200 µL DCM to remove residual procainamide. Then the water layer was desalted by PGC SPE by equilibrating the cartridge with 1 mL water, loading the sample in a minimal amount of water, washing first with 1 mL of water and then with 1 mL of 5%/95% ACN/water and eluting the glycans with 60%/40 ACN/water with 0.1% TFA. The eluent was evaporated under nitrogen flow and analyzed by HPLC-IM-MS.

***O*-glycan release from bovine submaxillary mucin**

O-glycans were released from BSM using sodium hypochlorite. The pH of a 15% sodium hypochlorite solution was adjusted to 6.8 by adding approximately 3 mL of 1 M HCl to 5 mL of the hypochlorite solution. BSM was dissolved in water to a concentration of 1 mg/mL, 0.5 mL of the neutralized sodium hypochlorite solution (pH 6.8) was added and the reaction mixture was kept on ice for 60 minutes before quenching with 15 µL 1% formic acid. The mixture was directly freeze dried and then purified with PGC SPE by equilibrating the cartridge with 1 mL water, loading the sample in a minimal amount of water, washing first with 1 mL of water, then with 1 mL of 5%/95% ACN/water and eluting the glycans with 60%/40% ACN/water with 0.1% TFA. The eluent was evaporated under nitrogen flow and the released glycans were analyzed with HPLC-IM-MS.

HPLC-IM-MS analysis of glycans

All standards and released glycans were analyzed with HPLC-IM-MS using an Agilent Technologies 1290 LC system (Santa Clara, CA) coupled via a dual-source AJS electrospray interface to an Agilent Technologies 6560B drift tube ion mobility/QTOF MS instrument.

Synthetic standards were analyzed using a SeQuant ZIC-HILIC LC pre-column (Merck, Darmstadt, Germany). Solutions of standards in 80%/20% ACN/water were injected into the chromatographic system, eluted with 60%/40% ACN/water containing 0.1% formic acid and further analyzed with IM-MS.

The derivatized *N*-glycans derived from biologicals, tissue samples and equine α2-macroglobulin were dissolved in 80%/20% ACN/water, injected into the chromatographic system and separated using a ZIC-HILIC (150x4.6 mm, 3.5 µm) column (Merck, Darmstadt, Germany), with a linear 30 min gradient from 80%/20% ACN/water containing 0.1% formic acid to 50%/50% ACN/water. Both eluent solutions contained either 0.1% formic acid for positive mode or 10 mM ammonium acetate at pH 4.4 for negative mode IM-MS analysis.

Released *O*-glycans were separated on a ZIC-HILIC column (150 x 2.1 mm, 3 µm particles) using 85%/15% ACN/water with 0.1% formic acid for 5 min, followed by a linear gradient to 50%/50% ACN/water with 0.1% formic acid over 30 minutes and then further analyzed by IM-MS.

IM-MS was performed with a transfer capillary voltage of 3500 V, nozzle voltage of 2000 V, nebulizer pressure of 40 psi, nitrogen drying gas at a temperature of 300 °C and a flow rate of 8L/min and a sheath gas at 300 °C with a flow rate of 11 L/min. IM was operated with a transient rate of 16 transients/frame, a trap fill time of 3900 µs, a trap release time of 250 µs, a drift tube entrance voltage of 1400 V and a multiplexing pulsing sequence length of 4 bit. The instrument was modified with a lens that was installed at the exit of the ion transfer capillary. By increasing the voltage applied to the fragmentor, this lens creates a potential difference between the capillary and the front high-pressure funnel in the instrument. The potential difference between the fragmentor voltage and the voltage at the capillary exit determines the voltage for in-source collisionally activated dissociation. The capillary exit and high-pressure funnel entrance are kept at 360V, resulting in a maximum potential difference of 240V that can be applied for ion activation when the fragmentor is set to the maximum of 600V.

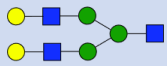
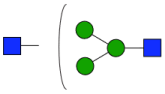
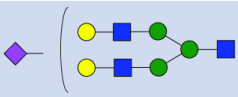
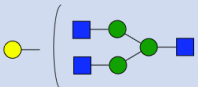
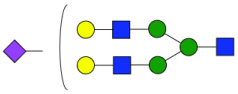
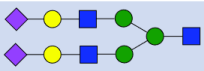
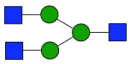
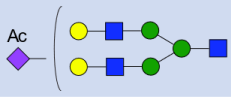
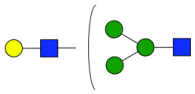
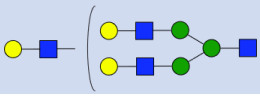
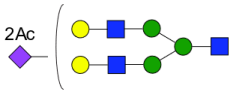
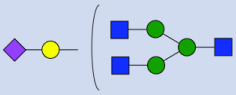
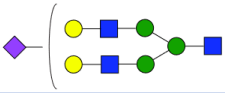
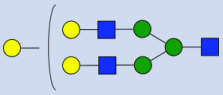
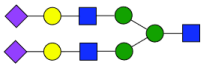
IM-MS data processing

Masses of raw 4 bit multiplexed IM-MS data were recalibrated on reference masses with *m/z* 121 and *m/z* 922 using the IM-MS Data File Reprocessing Utility in the Agilent Technologies Masshunter software. Reprocessed data was demultiplexed using the PNNL Preprocessor software (Pacific Northwest National Laboratory, Richland, WA) using an interpolation of 3 drift bins and a 5 point moving average smoothing.[28] Features were identified with the Agilent Technologies Masshunter IMS browser software using an unbiased isotope model, allowing for single features with a maximum charge state of 5 and a minimal ion intensity of 500. High resolution ATDs were obtained using Agilent Technologies HRdm 2.0 software at processing level high, with an *m/z* width multiplier of 12, saturation check of 0.40 and an IF multiplier of 1.125 with SSS and Post QC enabled.[29]

Data availability

The ion mobility-mass spectrometric source data that support the findings of this study are available in MassIVE with the identifier doi:10.25345/C57S7HX6G (MSV000090864).

Table I: Structure list of Endo-F2 released and procainamide labelled N-glycans identified in Myozyme

Suggested structure	Observed *m/z*	Calculated *m/z*	Relative abundancy (%)	$^{DT}CCS_{N2}$
	1657.6929 $[M+H]^+$ 1679.6725 $[M+Na]^+$ 840.3425 $[M+H+Na]^{2+}$ 851.3335 $[M+2Na]^{2+}$	1657.6932 $[M+H]^+$ 1679.6743 $[M+Na]^+$ 840.3416 $[M+H+Na]^{2+}$ 851.3328 $[M+2Na]^{2+}$	48.8	371.02 $[M+H]^+$ 374.87 $[M+Na]^+$ 401.32 $[M+H+Na]^{2+}$ 402.20 $[M+2Na]^{2+}$
	1130.5097 $[M+H]^+$ 1152.4908 $[M+Na]^+$ 576.7494 $[M+H+Na]^{2+}$	1130.5082 $[M+H]^+$ 1152.4901 $[M+Na]^+$ 576.7487 $[M+H+Na]^{2+}$	13.4	315.31 $[M+H]^+$ 325.62 $[M+Na]^+$ 363.64 $[M+H+Na]^{2+}$
	996.8823 $[M+2Na]^{2+}$	996.8803 $[M+2Na]^{2+}$	10.5	429.06 $[M+2Na]^{2+}$
	927.4297 $[M+H]^+$ 949.4108 $[M+Na]^+$	927.4288 $[M+H]^+$ 949.4107 $[M+Na]^+$	8.31	283.86 $[M+H]^+$ 292.95 $[M+Na]^+$
	1495.6393 $[M+H]^+$ 759.3156 $[M+H+Na]^{2+}$	1495.6413 $[M+H]^+$ 759.3152 $[M+H+Na]^{2+}$	5.64	357.57 $[M+H]^+$ 386.77 $[M+H+Na]^{2+}$
	985.8916 $[M+H+Na]^{2+}$	985.8894$[M+H+Na]^{2+}$	3.15	443.35 $[M+H+Na]^{2+}$
	1131.4386 $[M+H+Na]^{2+}$	1131.4371 $[M+H+Na]^{2+}$	1.72	473.48 $[M+H+Na]^{2+}$
	1333.5879 $[M+H]^+$ 1355.5695 $[M+Na]^+$ 678.2894 $[M+H+Na]^{2+}$	1333.5884 $[M+H]^+$ 1355.5704 $[M+Na]^+$ 678.2888 $[M+H+Na]^{2+}$	1.25	341.23 $[M+H]^+$ 347.69 $[M+Na]^+$ 378.99 $[M+H+Na]^{2+}$
	1006.8955 $[M+H+Na]^{2+}$ 1017.8853 $[M+2Na]^{2+}$	1006.8942 $[M+H+Na]^{2+}$ 1017.8852 $[M+2Na]^{2+}$	1.07	448.79 $[M+H+Na]^{2+}$ 433.07 $[M+2Na]^{2+}$
	1292.5613 $[M+H]^+$ 1314.5425 $[M+Na]^+$ 657.7761 $[M+H+Na]^{2+}$	1292.5619 $[M+H]^+$ 1314.5429 $[M+Na]^+$ 657.7756 $[M+H+Na]^{2+}$	0.99	339.57 $[M+H]^+$ 341.04 $[M+Na]^+$ 372.57 $[M+H+Na]^{2+}$
	1022.9088 $[M+H+Na]^{2+}$	1022.9077 $[M+H+Na]^{2+}$	0.83	446.81 $[M+H+Na]^{2+}$
	1027.9010 $[M+H+Na]^{2+}$ 1038.8928 $[M+2Na]^{2+}$	1027.8995 $[M+H+Na]^{2+}$ 1038.8904 $[M+2Na]^{2+}$	0.67	453.81 $[M+H+Na]^{2+}$ 439.14 $[M+2Na]^{2+}$
	904.8628 $[M+H+Na]^{2+}$	904.8625 $[M+H+Na]^{2+}$	0.60	429.89 $[M+H+Na]^{2+}$
	985.8877 $[M+H+Na]^{2+}$	985.8894$[M+H+Na]^{2+}$	0.50	441.48 $[M+H+Na]^{2+}$
	921.3686 $[M+H+Na]^{2+}$	921.3681 $[M+H+Na]^{2+}$	0.45	425.06 $[M+H+Na]^{2+}$
	1131.4368 $[M+H+Na]^{2+}$	1131.4371 $[M+H+Na]^{2+}$	0.39	473.75 $[M+H+Na]^{2+}$

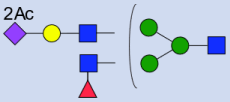
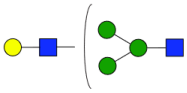
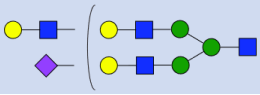
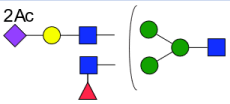
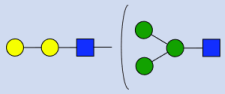
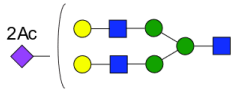
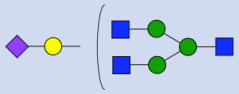
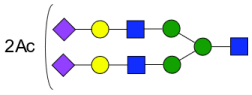
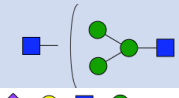
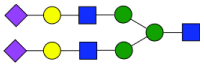
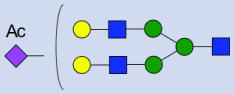
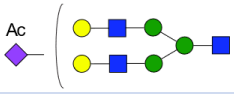
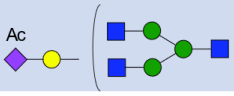
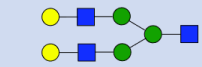
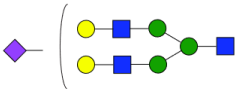
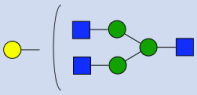
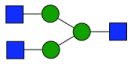
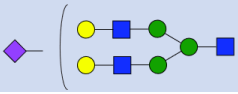
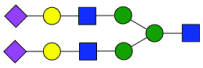
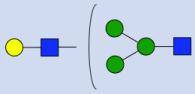
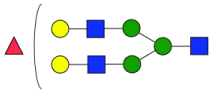
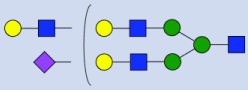
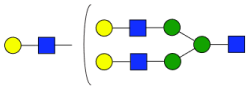
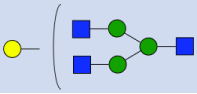
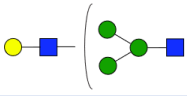
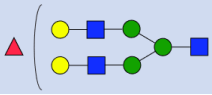
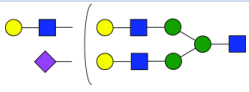
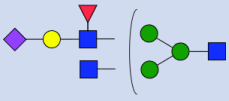
	1030.8908 $[M+2Na]^{2+}$	1030.8930 $[M+2Na]^{2+}$	0.38	453.73 $[M+2Na]^{2+}$
	1292.5600 $[M+H]^{+}$	1292.5619 [M+H]	0.36	339.96 $[M+H]^{+}$
	1168.4550 $[M+H+Na]^{2+}$ 1179.4468 $[M+2Na]^{2+}$	1168.4555 $[M+H+Na]^{2+}$ 1179.4460 $[M+2Na]^{2+}$	0.26	478.76 $[M+H+Na]^{2+}$ 471.27 $[M+2Na]^{2+}$
	1030.891 $[M+2Na]^{2+}$	1030.8930 $[M+2Na]^{2+}$	0.26	454.30 $[M+2Na]^{2+}$
	1454.6111 $[M+H]^{+}$ 738.8023 $[M+H+Na]^{2+}$	1454.6138 $[M+H]^{+}$ 738.8020 $[M+H+Na]^{2+}$	0.17	348.33 $[M+H]^{+}$ 395.93 $[M+H+Na]^{2+}$
	1027.8995 $[M+H+Na]^{2+}$	1027.8995 $[M+H+Na]^{2+}$	0.13	459.77 $[M+H+Na]^{2+}$
	904.8619 $[M+H+Na]^{2+}$	904.8625 $[M+H+Na]^{2+}$	0.12	430.10 $[M+H+Na]^{2+}$
	1173.4461 $[M+H+Na]^{2+}$	1173.4472 $[M+H+Na]^{2+}$	0.08	482.19 $[M+H+Na]^{2+}$
	576.7494 $[M+H+Na]^{2+}$	576.7487 $[M+H+Na]^{2+}$	0.04	363.83 $[M+H+Na]^{2+}$
	1142.4234 $[M+2Na]^{2+}$	1142.4280 $[M+2Na]^{2+}$	0.04	476.01 $[M+2Na]^{2+}$
	1017.8856 $[M+2Na]^{2+}$	1017.8852 $[M+2Na]^{2+}$	0.01	447.47 $[M+2Na]^{2+}$
	1017.8853 $[M+2Na]^{2+}$	1017.8852 $[M+2Na]^{2+}$	0.01	434.91 $[M+2Na]^{2+}$
	925.8692 $[M+H+Na]^{2+}$	925.8678 $[M+H+Na]^{2+}$	0.003	431.99 $[M+H+Na]^{2+}$

Table II: Structure list of Endo-F2 released and procainamide labelled N-glycans identified in Aflibercept

Suggested structure	Observed *m/z* ([M+H]$^+$ unless specified)	Calculated *m/z* ([M+H]$^+$ unless specified)	Relative abundancy (%)	$^{DT}CCS_{N2}$ ([M+H]$^+$ unless specified)
	1657.6941 [M+H]$^+$ 1679.6760 [M+Na]$^+$ 840.3424 [M+H+Na]$^{2+}$ 851.3326 [M+2Na]$^{2+}$	1657.6932 [M+H]$^+$ 1679.6743 [M+Na]$^+$ 840.3416 [M+H+Na]$^{2+}$ 851.3328 [M+2Na]$^{2+}$	51.7	370.86 [M+H]$^+$ 374.35 [M+Na]$^+$ 401.40 [M+H+Na]$^{2+}$ 402.62 [M+2Na]$^{2+}$
	985.8904 [M+H+Na]$^{2+}$ 996.8812 [M+2Na]$^{2+}$	985.8894 [M+H+Na]$^{2+}$ 996.8803 [M+2Na]$^{2+}$	20.1	444.25 [M+H+Na]$^{2+}$ 429.64 [M+2Na]$^{2+}$
	759.3157 [M+H+Na]$^{2+}$ 1495.6386 [M+H]$^+$ 1517.6202 [M+Na]$^+$	759.3152 [M+H+Na]$^{2+}$ 1495.6413 [M+H]$^+$ 1517.6232 [M+Na]$^+$	11.1	387.30 [M+H+Na]$^{2+}$ 357.17 [M+H]$^+$ 364.31 [M+Na]$^+$
	1333.5869 [M+H]$^+$ 1355.5680 [M+Na]$^+$ 678.2890 [M+H+Na]$^{2+}$	1333.5884 [M+H]$^+$ 1355.5704 [M+Na]$^+$ 678.2888 [M+H+Na]$^{2+}$	7.38	340.80 [M+H]$^+$ 378.86 [M+Na]$^+$ 349.28 [M+H+Na]$^{2+}$
	985.8896 [M+H+Na]$^{2+}$	985.8894 [M+H+Na]$^{2+}$	5.89	441.97 [M+H+Na]$^{2+}$
	1131.4379 M+H+Na]$^{2+}$ 1142.4281 [M+2Na]$^{2+}$	1131.4371 [M+H+Na]$^{2+}$ 1142.4280 [M+2Na]$^{2+}$	1.31	473.78 [M+H+Na]$^{2+}$ 479.46 [M+2Na]$^{2+}$
	1292.5594 [M+H]$^+$ 657.7753 [M+H+Na]$^{2+}$	1292.5619[M+H]$^+$ 657.7756 [M+H+Na]$^{2+}$	0.88	327.17 [M+H]$^+$ 364.54 [M+H+Na]$^{2+}$
	913.3710 [M+H+Na]$^{2+}$	913.3706 [M+H+Na]$^{2+}$	0.54	422.52 [M+H+Na]$^{2+}$
	1168.4565 [M+H+Na]$^{2+}$	1168.4555 [M+H+Na]$^{2+}$	0.29	475.03 [M+H+Na]$^{2+}$
	1022.9092 [M+H+Na]$^{2+}$	1022.9077 [M+H+Na]$^{2+}$	0.27	443.73 [M+H+Na]$^{2+}$ 446.95 [M+H+Na]$^{2+}$
	759.3148 [M+H+Na]$^{2+}$ 770.3079 [M+2Na]$^{2+}$	759.3152 [M+H+Na]$^{2+}$ 770.3062 [M+2Na]$^{2+}$ 1495.6413 [M+H]$^+$ 1517.6232 [M+Na]$^+$	0.13	392.11 [M+H+Na]$^{2+}$ 394.34 [M+2Na]$^{2+}$
	1192.5584 [M+H]$^+$	1292.5619 [M+H]$^+$	0.11	339.43 [M+H]$^+$
	913.3700 [M+H+Na]$^{2+}$	913.3706 [M+H+Na]$^{2+}$	0.10	422.86 [M+H+Na]$^{2+}$
	1168.4563 [M+H+Na]$^{2+}$	1168.4555 [M+H+Na]$^{2+}$	0.10	474.48 [M+H+Na]$^{2+}$
	977.8921 [M+H+Na]$^{2+}$	977.8919 [M+H+Na]$^{2+}$	0.09	445.47 [M+H+Na]$^{2+}$

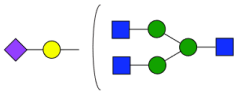
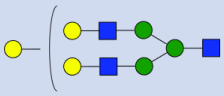
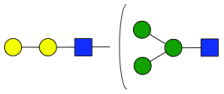
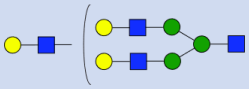
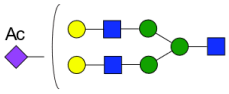
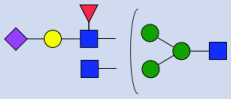
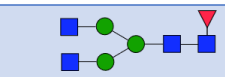
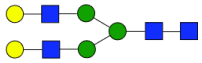
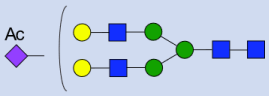
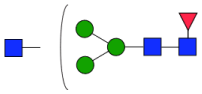
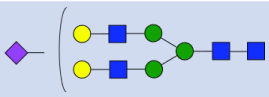
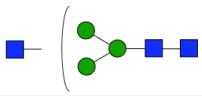
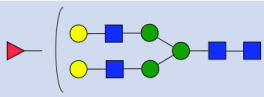
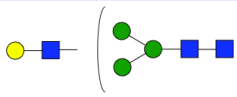
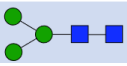
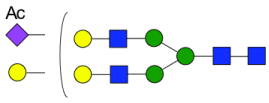
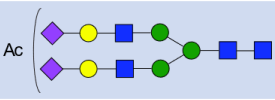
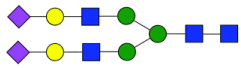
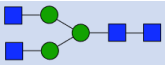
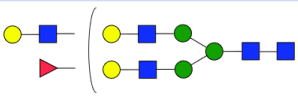
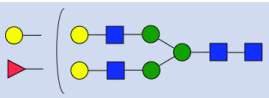
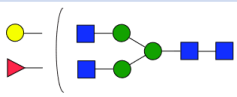
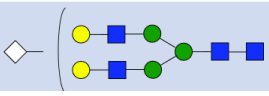
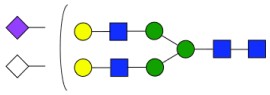
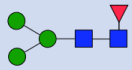
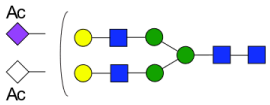
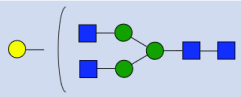
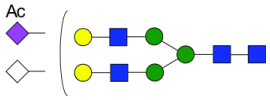
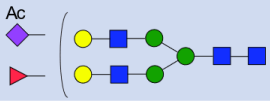
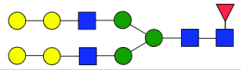
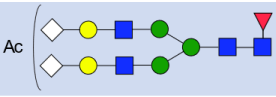
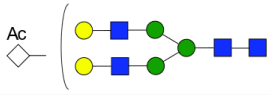
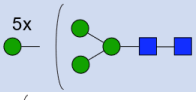
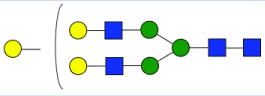
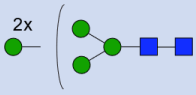
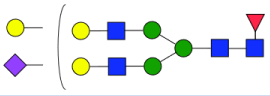
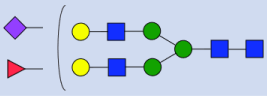
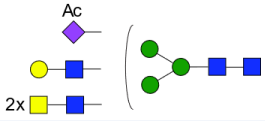
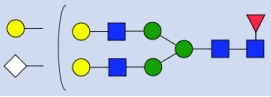
	904.8629 $[M+H+Na]^{2+}$	904.8629 $[M+H+Na]^{2+}$	0.08	432.33 $[M+H+Na]^{2+}$
	921.3636 $[M+H+Na]^{2+}$	921.3681 $[M+H+Na]^{2+}$	0.05	422.74 $[M+H+Na]^{2+}$
	738.8023 $[M+H+Na]^{2+}$	738.8020 $[M+H+Na]^{2+}$	0.04	397.84 $[M+H]^{+}$
	1022.9099 $[M+H+Na]^{2+}$	1022.9077 $[M+H+Na]^{2+}$	0.03	440.64 $[M+H+Na]^{2+}$
	1006.8969 $M+H+Na]^{2+}$	1006.8946 $[M+H+Na]^{2+}$	0.02	449.50 $[M+H+Na]^{2+}$
	977.8891 $[M+H+Na]^{2+}$	977.8919 $[M+H+Na]^{2+}$	0.01	436.48 $[M+H+Na]^{2+}$

Table III: Structure list of PNGAseF-released and procainamide-labelled N-glycans identified in horse α2-macroglobulin

Suggested structure	Observed *m/z*	Calculated *m/z*	Relative abundancy (%)	$^{DT}CCS_{N2}$
	1682.73 $[M+H]^+$	1681.7180 $[M+H]^+$	27.3	383.2 $[M+H]^+$
	941.8852 $[M+H+Na]^{2+}$	941.8814 $[M+H+Na]^{2+}$	17.1	424.6 $[M+H+Na]^{2+}$
	1116.4237 $[M+H+K]^{2+}$	1116.4211 $[M+H+K]^{2+}$	9.15	473.1 $[M+H+K]^{2+}$
	1479.6513 $[M+H]^+$	1479.6459 $[M+H]^+$	6.29	352.7 $[M+H]^+$
	1087.4339 $[M+H+Na]^{2+}$	1087.4291 $M+H+Na]^{2+}$	5.72	465 $M+H+Na]^{2+}$
	1333.5905 $[M+H]^+$	1333.5880 $[M+H]^+$	3.47	338.4 $[M+H]^+$
	1022.9013 $[M+H+K]^{2+}$	1022.8970 $[M+H+K]^{2+}$	3.37	447.4 $[M+H+K]^{2+}$
	767.3019 $[M+H+K]^{2+}$	767.3020 $[M+H+K]^{2+}$	3.19	388 $[M+H+K]^{2+}$
	1168.4607 $[M+K]^+$	1168.4645 $M+K]^+$	2.75	239.9 $M+K]^+$
	1197.4497 $[M+H+K]^{2+}$	1197.4475 $[M+H+K]^{2+}$	1.90	477.0 $[M+H+K]^{2+}$
	1253.9874 $[M+H+Na]^{2+}$	1253.9821 $[M+H+Na]^{2+}$	1.75	507.1 $[M+H+Na]^{2+}$
	1240.9665 $[M+H+K]^{2+}$	1240.9635 $[M+H+K]^{2+}$	1.75	500.4 $[M+H+K]^{2+}$
	1536.6724 $[M+H]^+$	1536.6673 $[M+H]^+$	1.72	367.4 $[M+H]^+$
	1641.7027 $[M+H]^+$	1641.6987 $[M+H]^+$	1.63	371.7 $[M+H]^+$
	1095.9403 $[M+H+Na]^{2+}$	1095.9367 $[M+H+Na]^{2+}$	1.15	450.7 $[M+H+Na]^{2+}$
	933.8866 $[M+H+Na]^{2+}$	933.8839 $[M+H+Na]^{2+}$	0.80	432 $[M+H+Na]^{2+}$
	1095.4225 $[M+H+Na]^{2+}$	1095.4265 $[M+H+Na]^{2+}$	0.77	455 $[M+H+Na]^{2+}$

	1240.9795 $[M+H+Na]^{2+}$	1240.9742 $[M+H+Na]^{2+}$	0.75	497.2 $[M+H+Na]^{2+}$
	1276.569 $[M+H]^{+}$	1276.5665 $[M+H]^{+}$	0.58	322.4 $[M+H]^{+}$
	1290.9747 $[M+H+K]^{2+}$	1290.9715 $[M+H+K]^{2+}$	0.55	503.9 $[M+H+K]^{2+}$
	860.8535 $[M+H+Na]^{2+}$	860.8550 $[M+H+Na]^{2+}$	0.50	410.1 $[M+H+Na]^{2+}$
	1269.9685 $[M+H+K]^{2+}$	1269.9662 $[M+H+K]^{2+}$	0.50	503.6 $[M+H+K]^{2+}$
	1189.4531 $[M+H+K]^{2+}$	1189.4500 $[M+H+K]^{2+}$	0.33	474.5 $[M+H+K]^{2+}$
	1176.9677 $[M+H+Na]^{2+}$	1176.9632 $[M+H+Na]^{2+}$	0.32	475.1 $[M+H+Na]^{2+}$
	1361.986 $[M+Na+K]^{2+}$	1361.9839 $[M+H+K]^{2+}$	0.25	527.0 $[M+H]^{+}$ $[M+H+K]^{2+}$
	1124.4214 $[M+H+K]^{2+}$	1124.4185 $[M+H+K]^{2+}$	0.19	475.0 $[M+H+K]^{2+}$
	981.8846 $[M+H+Na]^{2+}$	981.8812 $[M+H+Na]^{2+}$	0.19	453.6 $[M+H+Na]^{2+}$
	1022.9056 $[M+H+Na]^{2+}$	1022.9078 $[M+H+Na]^{2+}$	0.18	447.4 $[M+H+Na]^{2+}$
	1454.6187 $[M+H]^{+}$	1454.6142 $[M+H]^{+}$	0.15	350.4 $[M+H]^{+}$
	1249.4694 $[M+H+K]^{2+}$	1249.4711 $[M+H+K]^{2+}$	0.13	490.3 $[M+H+K]^{2+}$
	1160.4625 $[M+H+Na]^{2+}$	1160.4580 $[M+H+Na]^{2+}$	0.11	475.8 $[M+H+Na]^{2+}$
	1343.0125 $[M+2Na]^{2+}$	1343.0183 $[M+2Na]^{2+}$	0.098	522.4 $[M+2Na]^{2+}$
	1249.4874 $[M+H+Na]^{2+}$	1249.4819 $[M+H+Na]^{2+}$	0.098	491.3 $[M+H+Na]^{2+}$

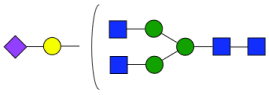
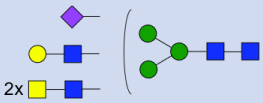
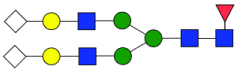
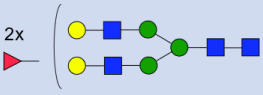
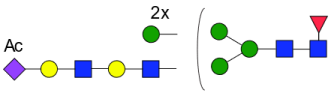
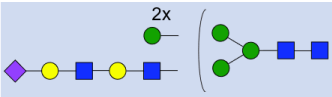
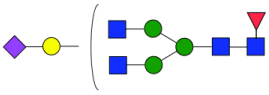
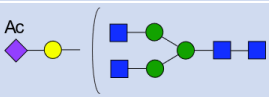
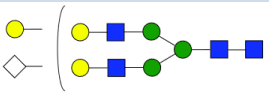
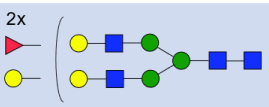
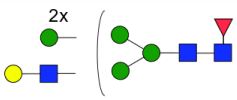
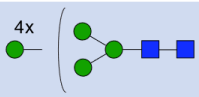
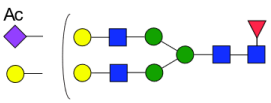
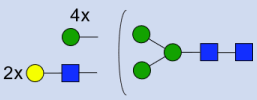
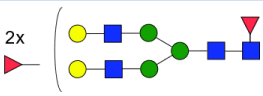
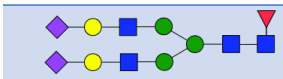
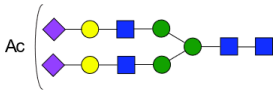
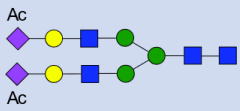
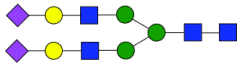
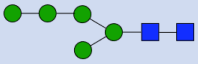
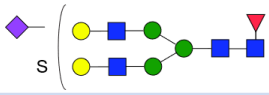
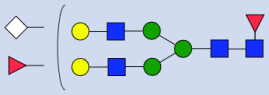
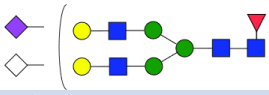
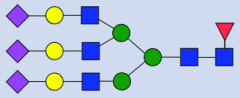
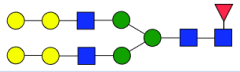
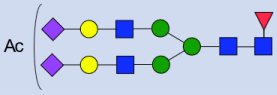
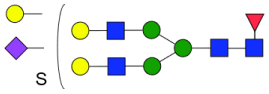
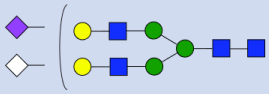
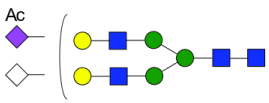
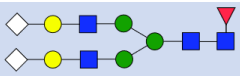
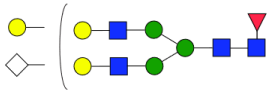
	1014.3875 $[M+H+K]^{2+}$	1014.3894 $[M+H+K]^{2+}$	0.092	448.7 $[M+H+K]^{2+}$
	1322.0071 $[M+2Na]^{2+}$	1322.0130 $[M+2Na]^{2+}$	0.077	518.3 $[M+2Na]^{2+}$
	1340.9801 $[M+Na+K]^{2+}$	1340.9786 $[M+Na+K]^{2+}$	0.062	520.6 $[M+Na+K]^{2+}$
	1095.9278 $[M+H+K]^{2+}$	1095.9260 $[M+H+K]^{2+}$	0.054	461.1 $[M+H+K]^{2+}$
	1351.4971 $[M+H+K]^{2+}$	1351.5029 $[M+H+K]^{2+}$	0.049	511.2 $[M+H+K]^{2+}$
	1257.471 $[M+H+K]^{2+}$	1257.4686 $[M+H+K]^{2+}$	0.047	500.2 $[M+H+K]^{2+}$
	1079.4358 $[M+H+Na]^{2+}$	1079.4316 $[M+H+Na]^{2+}$	0.038	464.3 $[M+H+Na]^{2+}$
	1035.3947 $[M+H+K]^{2+}$	1035.3947 $[M+H+K]^{2+}$	0.036	447.5 $[M+H+K]^{2+}$
	791.9624 $[M+2Na]^{2+}$	791.9652 $[M+2Na]^{2+}$	0.030	551.9 $[M+2Na]^{2+}$
	1179.9523 $[M+2Na]^{2+}$	1179.9569 $[M+2Na]^{2+}$	0.022	488.5 $[M+2Na]^{2+}$
	994.3975 $[M+H+Na]^{2+}$	994.3971 $[M+H+Na]^{2+}$	0.021	438.4 $[M+H+Na]^{2+}$
	900.852 $[M+H+Na]^{2+}$	900.8548 $[M+H+Na]^{2+}$	0.021	421.8 $[M+H+Na]^{2+}$
	1262.4846 $[M+H+Na]^{2+}$	1262.4898 $[M+H+Na]^{2+}$	0.020	492.3 $[M+H+Na]^{2+}$
	1265.9868 $[M+H+Na]^{2+}$	1265.9870 $[M+H+Na]^{2+}$	0.016	516.8 $[M+H+Na]^{2+}$
	1168.9591 $[M+H+K]^{2+}$	1168.9549 $[M+H+K]^{2+}$	0.010	476.6 $[M+H+K]^{2+}$

Table IV: Structure list of PNGaseF released N-glycans identified in horse nasal tissue

Suggested structure	Observed *m/z* ([M-2H]$^{2-}$ unless specified)	Calculated *m/z* ([M-2H]$^{2-}$ unless specified)	Relative abundancy (%)	$^{DT}CCS_{N2}$ ([M-2H]$^{2-}$ unless specified)
	1183.4097	1183.4132	23.6	491.28
	1131.3871	1131.3895	17.8	461.50
	1152.3920	1152.3948	11.8	486.62
	1110.3824	1110.3842	7.81	462.58
	1233.4207 [M-H]$^{-}$	1233.4262 [M-H]$^{-}$	7.57	308.30 [M-H]$^{-}$
	1077.8425	1077.8439	4.65	453.06
	1118.8897	1118.8919	4.49	471.65
	1191.4067	1191.4106	3.29	491.35
	1007.3484 [M-3H]$^{3-}$	1007.3489 [M-3H]$^{3-}$	2.64	571.00 [M-3H]$^{3-}$
	1054.3690	1054.3706	2.20	443.99
	1204.4144	1204.4185	1.98	490.95
	1158.8664	1158.8703	1.39	466.51
	1118.3786	1118.3817	1.14	469.67
	1139.3836	1139.3870	1.10	465.15
	1199.4034	1199.4081	0.95	489.78
	1126.8860	1126.8893	0.92	471.27

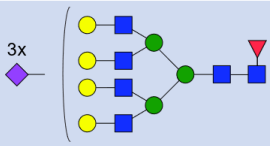
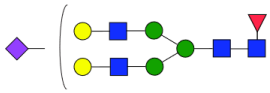
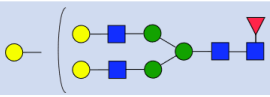
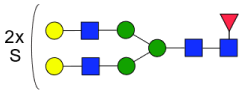
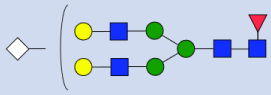
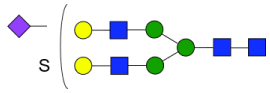
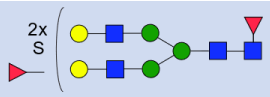
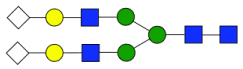
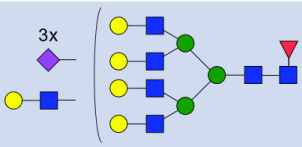
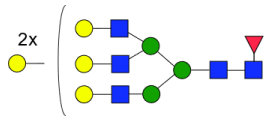
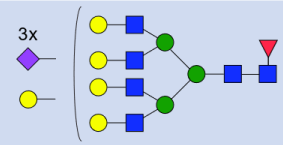
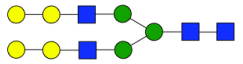
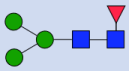
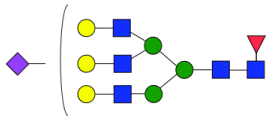
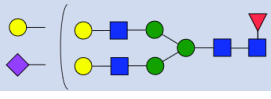
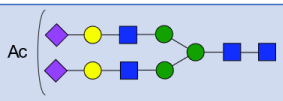
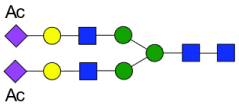
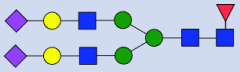
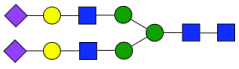
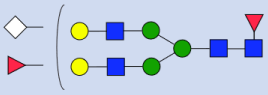
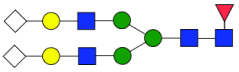
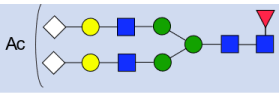
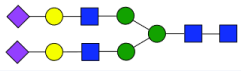
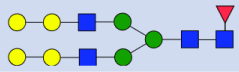
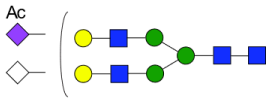
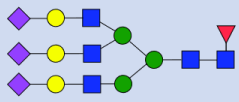
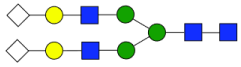
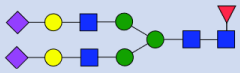
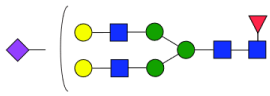
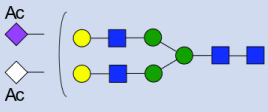
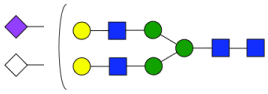
	1129.0564 $[M-3H]^{3-}$	1129.0596 $[M-3H]^{3-}$	0.77	604.13 $[M-3H]^{3-}$
	1037.8642	1037.8655	0.71	453.26
	973.3434	973.3442	0.71	435.65
	972.2746	972.2746	0.57	431.74
	1045.8614	1045.8629	0.51	450.15
	1004.8143	1004.8149	0.48	433.93
	1207.3559	1207.3564	0.48	477.48
	1126.3756	1126.3791	0.46	469.45
	1250.7645 $[M-3H]^{3-}$	1250.7704 $[M-3H]^{-}3$	0.39	636.03$[M-3H]^{3-}$
	1236.9308	1236.9367	0.38	486.83
	1183.0737 $[M-3H]^{3-}$	1183.0772 $[M-3H]^{-}3$	0.33	616.55 $[M-3H]^{3-}$
	981.3410	981.3416	0.26	427.03
	1055.3762 $[M-H]^{-}$	1055.3784 $[M-H]^{-}$	0.22	287.72 $[M-H]^{-}$
	1220.4257	1220.4316	0.19	487.59
	1118.8873	1118.8919	0.17	470.69

Table V: Structure list of PNGaseF released N-glycans identified in horse tracheal tissue

Suggested structure	Observed *m/z* $[M-2H]^{2-}$	Calculated *m/z* $[M-2H]^{2-}$	Relative abundancy (%)	$^{DT}CCS_{N2}$ ($[M-2H]^{2-}$ unless specified)
	1131.3871	1131.3895	24.2	460.95
	1152.3917	1152.3948	21.2	485.70
	1183.4092	1183.4132	15.4	489.60
	1110.3819	1110.3842	10.6	461.69
	1118.8886	1118.8918	5.19	470.53
	1199.4034	1199.4081	4.89	488.03
	1220.4084	1220.4134	2.78	489.39
	1110.3815	1110.3842	2.44	466.47
	1054.3683	1054.3706	1.57	443.99
	1139.3833	1139.3870	1.39	464.78
	1007.3484 $[M-3H]^{3-}$	1007.3489 [M-3H]3-	1.37	569.47$[M-3H]^{3-}$
	1126.3762	1126.3791	1.34	469.18
	1183.4093	1183.4132	1.19	489.39
	1037.8640	1037.8655	1.05	452.56
	1160.3882	1160.3922	1.01	487.97
	1118.3786	1118.3817	0.94	471.02

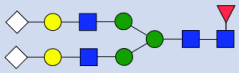
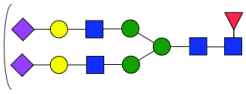
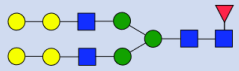
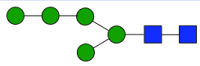
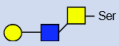
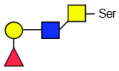
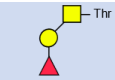
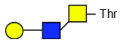
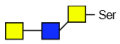
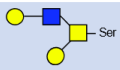
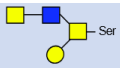
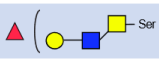
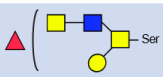
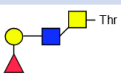
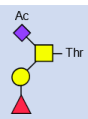
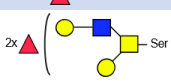
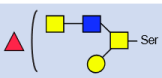
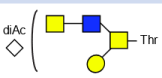
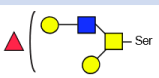
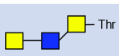
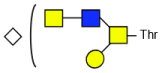
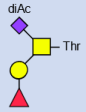
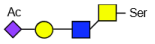
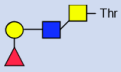
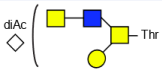
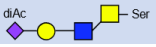
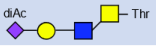
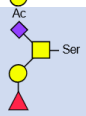
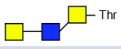
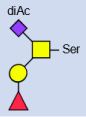
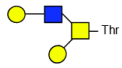
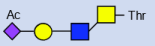
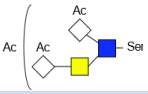
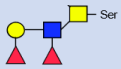
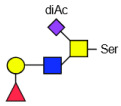
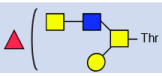
	1199.4024	1199.4081	0.89	486.76
Ac 	1204.4133	1204.4185	0.89	489.19
	1054.3680	1054.3706	0.86	443.86
	1233.4210 $[M-H]^-$	1233.4262 $[M-H]^-$	0.80	308.21 $[M-H]^-$

Table VI: Structure list of neutralized hypochlorite released O-glycans identified in BSM

Suggested structure	Observed *m/z* $[M+Na]^+$	Calculated *m/z* $[M+Na]^+$	Relative abundancy (%)	$^{DT}CCS_{N2}$ (Å^2) $[M+Na]^+$	Amino acid
	838.2691	838.2704	10.1	265.20	S
	607.1944	607.1961	9.53	231.10	T
	691.2159	691.2173	8.17	244.80	T
	852.2841	852.2861	7.91	268.70	T
	593.1794	593.1804	7.88	226.10	S
	610.1949	610.1957	5.37	224.50	S
	649.2056	649.2067	5.10	238.30	T
	519.1793	519.1800	4.95	211.10	T
	316.1017	316.1007	4.60	167.60	T
	505.1634	505.1644	4.13	208.30	S
	677.2002	677.2016	4.09	240.30	S
	880.2801	880.2810	3.34	274.00	S
	623.1895	623.1910	3.05	235.60	T
	635.1896	635.1910	2.90	233.30	S
	894.2955	894.2967	2.67	277.40	T
	894.2949	894.2967	1.78	279.80	T
	667.2157	667.2172	1.75	237.60	S
	813.2736	813.2751	1.67	258.50	S
	854.2633	854.2653	1.13	270.30	S
	868.2792	868.2809	1.06	273.50	T
	302.0851	302.0850	0.79	163.30	S
	811.2574	811.2595	0.58	259.50	T

	624.2102	624.2114	0.52	227.10	T
	681.2309	681.2329	0.51	242.40	T
	812.2531	812.2547	0.50	263.90	S
	708.2419	708.2438	0.50	251.10	S
	826.2683	826.2703	0.46	267.40	T
	707.2101	707.2122	0.44	249.40	T
	829.2676	829.2700	0.43	260.70	S
	464.137	464.1378	0.40	196.80	S
	870.2944	870.2966	0.31	268.70	S
	651.184	651.1859	0.24	234.80	S
	813.273	813.2751	0.24	268.30	S
	478.1523	478.1535	0.24	201.10	T
	1016.3526	1016.3545	0.20	302.40	S
	896.2742	896.2759	0.20	278.70	S
	910.2894	910.2915	0.19	282.00	T
	827.2895	827.2908	0.19	263.20	T
	957.3158	957.3174	0.19	285.40	T
	1121.3855	1121.3858	0.17	313.50	S
	1016.3527	1016.3545	0.16	290.40	S
	1275.4212	1275.4237	0.14	331.60	T
	797.2407	797.2439	0.13	255.80	S
	975.3254	975.3279	0.13	284.40	S
	722.2576	722.2594	0.11	253.20	T

	1191.4014	1191.4025	0.11	316.20	T
	853.2696	853.2701	0.09	275.10	T
	609.1738	609.1753	0.08	224.40	S
	999.3256	999.3280	0.08	295.00	T
	1000.3229	1000.3232	0.07	289.50	S
	827.2873	827.2908	0.07	270.90	T
	1275.4237	1275.4237	0.07	330.50	T
	1042.3311	1042.3338	0.06	295.50	S
	853.2677	853.2701	0.05	266.40	T
	1056.3497	1056.3495	0.05	299.10	T
	665.1996	665.2016	0.03	231.50	T
	708.2422	708.2438	0.03	249.30	S
	769.2493	769.2489	0.03	261.30	T
	943.2992	943.3018	0.02	281.90	S
	722.2584	722.2594	0.02	255.10	T
	985.3118	985.3124	0.01	288.10	S
	843.2836	843.2857	0.010	263.80	T
	1014.3371	1014.3389	0.008	293.20	T
	1223.3926	1223.3950 $[M+H]^+$	0.008	321.40 $[M+H]^+$	S
	959.3312	959.3330	0.007	291.80	S
	1188.3929	1188.3917	0.005	325.30	S
	1030.3699	1030.3701	0.003	304.70	T

Glycan microarray analysis of the HA ectodomains of human H3N2

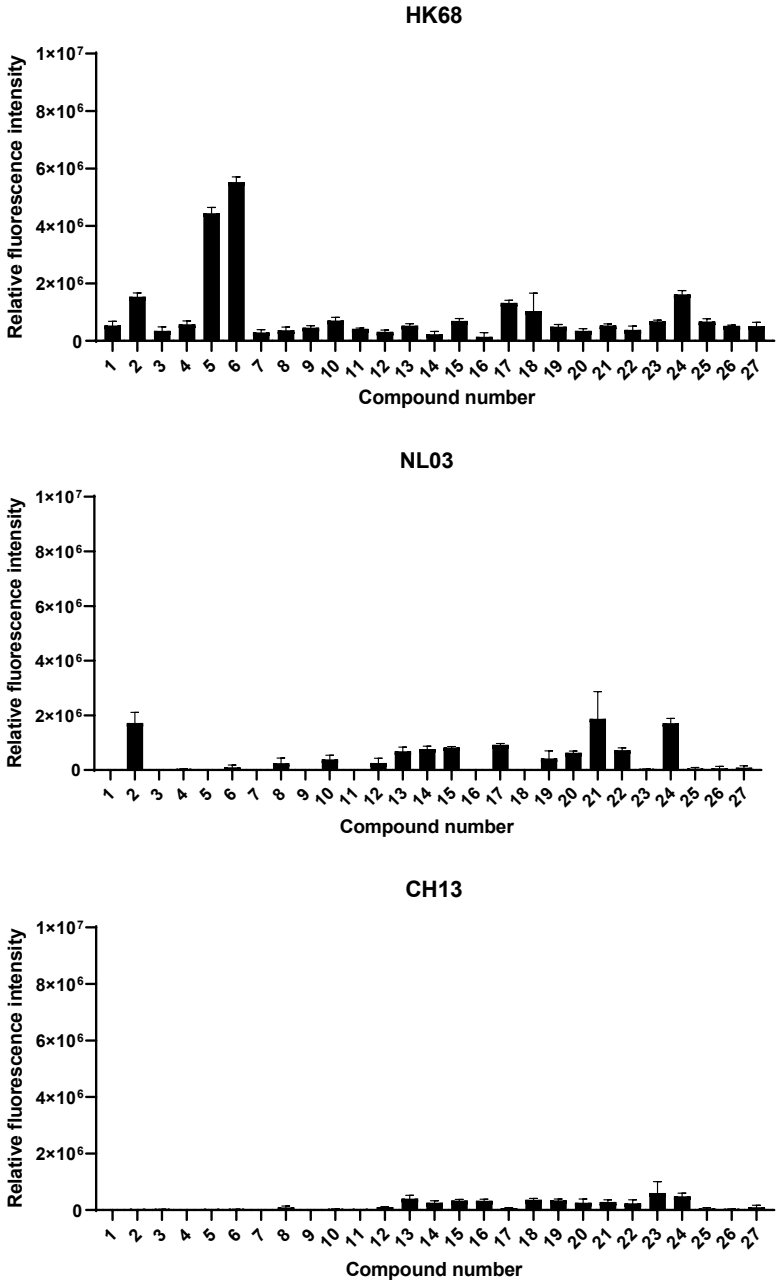


Figure S1: Glycan microarrays of the HA ectodomains of human H3N2.
Top: HA ectodomain of HK68 (A/Hong kong/1/68 H3N2, Genbank accession no. AF358177). Middle: HA ectodomain of NL03 (A/Netherlands/213/03 H3N2, Genbank accession no. AY661035). Bottom: HA ectodomain of CH13 (A/Switzerland/9715293/13 H3N2, Genbank accession no. AU46905.1).

CHAPTER 6

Summary and Future Perspectives

Summary

Glycans are an essential component of all living organisms but are, unlike other bio-polymers such as RNA and proteins, not directly coded in the genome. Instead, the non-template driven biosynthesis of glycans results from a combination of co-operative processes and substrate competition between the many glycosyltransferases, which causes high structural heterogeneity among glycans. The first part of the introduction describes the structural diversity of glycans found in humans. The monosaccharides that form the building blocks of glycans are introduced, together with nomenclature that describes how these monosaccharides are linked to each other. The glycan classes are discussed which are distinguished by their scaffold type. The biosynthesis of *N*- and *O*-linked glycans are highlighted and differences in the core architecture are explained. Glycan epitopes such as Lewis and blood group epitopes are discussed as an important source of isomerism in complex glycans and non-glycan modifications such as sulfates and acetyl esters are introduced.

After establishing the structural diversity of glycans, the focus is shifted in the second section of the introduction to strategies developed for the synthesis of glycans using glycosyltransferases. The different proposed mechanisms of this enzymatic glycosylation reaction are discussed. An overview of the many human glycosyltransferases and the transformations they are involved in is given. This diversity in transferases highlights the complexity of glycan biosynthesis. The enzymatic synthesis of glycans can be complicated if multiple acceptor positions are present that can undergo glycosylation. Reported regioselective enzymatic transformations are discussed as tools to regulate enzymatic glycan synthesis. These type of transformations can be used to not only selectively glycosylate but also desymmetrize glycans and thereby dictate regioselectivity of subsequent enzymatic transformations. However, not enough regioselective enzymatic glycosylation reactions have been uncovered to access all glycans present in the human glycome. Chemo-enzymatic synthesis is a promising solution to this problem: non-natural modifications are introduced in glycans which are used to direct enzymatic glycosylation. Multiple strategies to install these non-natural modifications are discussed, including the use of non-natural sugar nucleotide donors.

The third section of the introduction describes analytical techniques used to characterize glycan structure. Several NMR techniques are explained that are applied to confirm the glycan structure of purified synthetic glycans. Lectin histochemistry is briefly discussed as a method to analyze glycosylation at a tissue level. After this, release methodologies are discussed that cleave *O*- and *N*-glycans from glycoproteins and the challenges that are currently faced with *O*-glycan release are noted. Advanced mass spectrometry-based techniques for the analysis of released glycans are discussed such as multistage mass spectrometry and ion mobility mass spectrometry. These techniques can be used to generate detailed information about the glycan structure. The principles behind the separation of isomeric glycans using ion mobility mass spectrometry are explained as well as the calculation of collision cross sections using the Mason-Schamp equation. The introduction concludes with a outline of challenges that the high degree of possible isomers of a complex glycan bring in the synthesis and analysis of these structures.

A stop-and-go strategy for the chemo-enzymatic synthesis of selectively fucosylated polylactosamine derivatives

The enzymatic introduction of α1,3 fucose to GlcNAc by fucosyltransferases is one of the few enzymatic glycosylation reactions in humans that can occur on both the terminal and

internal LacNAc residues. The promiscuity of this transformation on polyLacNAc scaffolds does not allow for the regioselective synthesis of these type of fucosylated glycans. The biological implications that internal α1,3 fucose can have on the binding of glycan binding proteins is still mostly unknown. This is due to a lack of accessible methodology for the synthesis of synthetic standards and the difficulty to assign the α1,3 fucose position in biological samples. Previous work demonstrated that a chemically synthesized lactosamine hexasaccharide carrying orthogonal N-protecting groups could be manipulated with simple chemistry to allow for regioselective fucosylation. The work described here utilizes a chemo-enzymatic approach to build a similar orthogonally N-protected lactosamine with greatly reduced synthetic effort. Key in this procedure is the enzymatic introduction of the unnatural *N*-protected monosaccharide GlcNTFA in polyLacNAc's. This allows for the formation of *N*-protected lactosamine disaccharide subunits in a linear fashion in four synthetic steps. An orthogonally protected hexasaccharide is created by this method in only eight synthetic steps and forms a common precursor to all possible fucosylation patterns. Simple chemistry is used to (de)activate the common precursor at selected sites to systematically install α1,3 fucose by enzymatic fucosylation. A library of ABO blood group antigens is generated by this methodology that also carry α1,3 fucose residues. These glycans are printed on glass slides to form a glycan microarray that can be used to evaluate binding to the printed glycans. The glycan microarray is used to screen the binding of several galectins to blood group antigens. We find that α1,3 fucose can influence the binding of galectins on the glycan microarray and we observe distinct binding behavior of galectins as a response to this fucosylation.

Chemo-enzymatic synthesis of I-branched polylactosamines

Another internal glycosylation reaction in adult humans is the I-branching of polyLacNAc's. This β1,6 GlcNAcylation installs a branching point on non-terminal galactoses present on glycans. Other glycosyltransferases can extend this branch to generate multi-antennary glycans. These multi-antennary glycans can undergo multivalent interactions and are especially prevalent in human milk oligosaccharides. Currently, no methods have been reported to selectively generate I-branching sites on polyLacNAc scaffolds. The enzymatic installation of *N*-protected lactosamines developed in the previous chapter was able to influence the regioselectivity of the I-branching GlcNAc transferase GCNT2. NMR experiments show that Boc protected lactosamine selectively inhibits the I-branching reaction on the reducing-end neighboring galactose which allows regioselective glycosylation of other internal galactoses. Lewis^x fucosylation by FUT 5 is inhibited by the same unnatural modification and this modification is used to flip the activated and deactivated glycosylation sites. A library of complex polyLacNAc architectures is generated using regioselective I-branching strategies which contains regioisomers with different I-branching patterns.

Neutral oxidative release for the analysis of *O*-glycans with labile substituents from glycoconjugates

O-glycosylation of mucin-type glycoproteins has a great prognostic, diagnostic and therapeutic potential for many human diseases. A critical step in the analysis of *O*-glycans from biological samples is the glycan release from glycoconjugates. However, unlike *N*-glycans, there are no efficient enzymatic release methods. Chemical release methods can damage the glycan structure in several ways such as the peeling reaction and hydrolysis of pH sensitive groups. Recently, it was demonstrated that glycan release using hypochlorite results in the partial formation of a lactic acid-linked product which is inert to peeling and retains information of the amino acid linkage. However, alkaline hypochlorite generates multiple products and is not

compatible with base sensitive modifications. To address these problems, neutral hypochlorite is developed as a mild *O*-glycan release method. First, the side-reaction formation and reaction kinetics of hypochlorite glycan release are investigated with the help of synthetic *O*-glycopeptide standards. Using pure glycopeptides as starting material allowed us to monitor the reaction progress and characterize side product formation. The main product produced by alkaline hypochlorite was free reducing end glycan which is still prone to peeling. Neutralization of hypochlorite results in a slightly slower reaction but generates lactic acid-linked glycan as the only product, suppressing side reactions. Next, the compatibility of neutralized hypochlorite with pH sensitive glycan modifications was investigated using glycopeptide standards functionalized with sulfate, sialic acid or fucose. All reactions resulted in complete conversion to the expected product without major side or degradation products. To assess the ability of neutralized hypochlorite to release *O*-glycans from mucin-type glycoproteins, we subjected bovine submaxillary mucin to neutralized hypochlorite. A large fraction of observed glycans still contained one or multiple acetyl esters on sialic acids which are base labile modifications that are difficult to study with other *O*-glycan release methods. Additionally, sulfated *O*-glycans were detected in bovine submaxillary mucin which were not previously identified. Reaction intermediates could be identified in the analysis of neutralized hypochlorite release of the pure synthetic standards that allowed the proposition of a reaction mechanism.

Sialic acid *O*-acetylation patterns and glycosidic linkage type can be determined by ion mobility-mass spectrometry

Sialic acids can be modified by acetyl esters at the 4-, 7-, 8- and/or 9-position to give as many as fifteen different pattens of *O*-acetylation. These acetyl ester-modified sialic acids are associated with various diseases including cancer, immune disorders, and infection. Furthermore, *O*-acetylation can act as a molecular switch and can substantially influence the rate of hydrolysis of both endo- and exogeneous sialidases. The acetyl esters on sialic acids are base labile and can migrate at mildly basic conditions. Current methods to determine the acetyl ester positions of *O*-acetylated sialic acid require the acid hydrolysis of the sialic acid which causes a loss of information on the structure of the parent glycan and sialic acid linkage type. In this chapter we develop a method to determine both acetyl ester position and sialic acid linkage by ion-mobility mass spectrometry. Key to develop this methodology is a library of synthetic Neu5Ac and Neu5Gc sialosides that contain well-defined *O*-acetylation patterns and sialic acid linkages. Distinct arrival time distributions were observed using ion-mobility mass spectrometry for characteristic fragments, generated using in-source fragmentation. The obtained arrival time distributions were converted to collisional cross sections, which are molecular properties independent on instrument variables and can thus be transferred between labs and instrumentation. The observed collisional cross section values are compiled in a database that is used to determine exact *O*-acetyl position in conjunction with sialic acid linkage of multiple model samples. Acetyl ester positions and sialic acid linkages are determined on the *N*-glycans of CHO expressed biotherapeutics and match reported linkages and *O*-acetyl positions expressed in CHO cells. Moreover, in the *N*-glycans released from horse upper airway tissues, we observe large discrepancies in the sialic acid linkage type between *O*-acetylated and non-acetylated sialic acids that are not easily detectable with previously reported techniques. Lastly, we revisit the acetyl esters found in *O*-glycan content of bovine submaxillary mucin in the previous chapter. Ion mobility is able to confirm that most of the *O*-acetyl esters of this sample were positioned on the C-7 position. The high abundance of acetyl esters on the C-7 position, which is especially prone to migration, supports that neutralized hypochlorite retains the acetyl ester position without significant migration.

Future perspectives

The research described in this thesis focuses on the synthesis and analysis of several sources of isomerism found in complex glycans: namely, Lewis^x fucosylation, polyLacNAc architecture and the position of *O*-acetyl esters present on sialic acids. The methods and strategies developed herein to address regioselective synthesis and exact characterization of glycan structure may inspire future research and help address still-standing challenges in glycan synthesis and analysis. Below, I propose three research projects inspired by my results and experiences.

Keratan sulfates, a new frontier in chemo-enzymatic synthesis

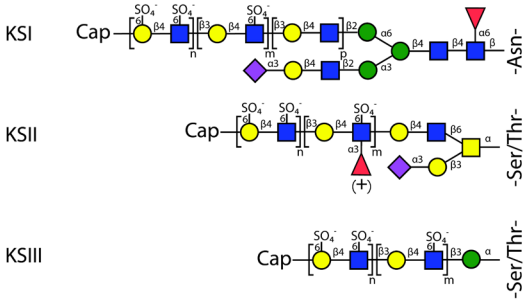


Figure 1: General structure of keratan sulfate

The enzymatic incorporation of unnatural GlcNAc analogs allows for the efficient and controlled synthesis of internal I-branched and α1,3 fucosylated polyLacNAc's. Additional to these two internal modifications that can be found on complex glycans, there is a third modification, sulfation, which is an essential part of keratan sulfates. Keratan sulfates are highly complex proteoglycans abundantly expressed in extracellular matrix of cornea, bone, cartilage, brain, and on surface of epithelial cells (Fig. 1).[1–3] Corneal keratan sulfate (KSI) is attached to asparagine of several core proteins *via* a complex *N*-linked glycan, whereas in cartilage it is attached as *O*-linked glycans *via* a core-2 structure (KSII). A third type that was recently identified on mannose linked to site chain of serine (KSIII). One of the antennae of the *N*- and *O*-glycans is extended by LacNAc repeating units to create an oligomeric structure that is modified by sulfate esters at C-6 positions of galactose and GlcNAc. The LacNAc backbone can be further modified by fucose residues, and at the terminus by α2,3-linked sialic acid.[4] Thus, KS has a modular architecture in which the backbone is composed of several types of LacNAc repeating units having different patterns of sulfation and fucosylation. Current synthetic strategies have relied solely on chemical synthesis and have only achieved tetrasaccharides containing KS epitopes while natural KS can reach lengths of up to 50 LacNAc repeats.[5] A (chemo-) enzymatic approach towards the synthesis of KS can provide glycan standards with architectures found on KS and help uncover the interactions of KS with glycan binding proteins. Three components need to be controlled to be able to selectively synthesize keratan sulfates by (chemo-)enzymatic means: GlcNAc-6-sulfation, Gal-6-sulfation and fucosylation.

The enzymes thought to be responsible for GlcNAc-6-sulfation, CHST2 or CHST6, are only active on terminal GlcNAc residues and this sulfation is probably introduced during formation of the polyLacNAc backbone.[6] A strategy following the biosynthetic route seems most practical: enzymatic sulfation with CHST2 or 6 can be used to incorporate GlcNAc-6-sulfation only on GlcNAc residues where this modification is needed. During the installation of LacNAcs that need to remain unsulfated, the GlcNAc-6-sulfation step can simply be skipped.

The second component that needs to be controlled is fucosylation. A similar problem has been addressed in chapter 2 and provides a potential solution to direct fucosylation on KS. GlcNTFA can be introduced at LacNAc positions where fucosylation needs to be blocked and natural GlcNAc at positions where the polyLacNAc will be fucosylated. GlcNTFA residues can be converted to $GlcNH_2$ after synthesis of the poly-LacNAc chain which are then deactivated towards fucosylation by Hp39-FT, FUT5 or FUT9. After fucosylation, the $GlcNH_2$ can be acetylated to form the natural GlcNAc/KS.

The third, and perhaps most challenging component of keratan sulfates from a synthesis perspective, is Gal-6-sulfation. The enzyme CHST1 or C6ST1 are thought to introduce this type of sulfate after assembly of the complete LacNAc chain.[7–9] Although this enzyme possesses inherent selectivities, it does not allow the selective modification of such chains. A stop-and-go-strategy can be designed to address this limitation. A potential method could employ natural UDP-Gal and unnatural UDP-Gal6X, in which the C-6 position is chemically modified by a blocking group. It is known that the galactosyl transferase B4GalT1 tolerates modifications at C-6 of UDP-Gal.[10] Natural UDP-Gal can be employed for sites where galactose needs to be sulfated whereas a modified UDP-Gal derivative can be introduced in those sites that should remain unmodified. After such an oligosaccharide assembly has been completed, it can be exposed to either CHST1 or C6ST1 and the sulfate donor PAPS. This would only introduce sulfates on the free C-6 hydroxyl residues of galactose. Next, the blocking group X can be converted into hydroxyls to provide the natural glycan. The chemo-enzymatic synthesis of keratan sulfates provides a fitting challenge and could provide synthetic standards that could contribute to understanding the biological role of keratan sulfates.

Solving glycan structures, one step at a time

Discrimination of isomeric glycans using mass spectrometry-based techniques remains a substantial challenge. Approaches that are able to collect an additional axis of information such as CCS in ion mobility and absorption in cold-ion infra-red spectroscopy can be combined with mass spectrometry to distinguish isomers.[11–14] These approaches show great potential in proof-of-concept studies with controlled mixtures of glycan isomers but have yet to be applied in more complex glycomics. The day-to-day implementation of such methods in glycomic studies would need to rely on databases that contain a complete data set on each possible isomer to confidently separate all isomers. However, the diversity and complexity of the human glycome makes the synthesis of homogeneous standards for all glycans unfeasible. For this reason, it will not be possible to obtain a complete set of synthetic standards that could be used to identify all unknown glycans in natural samples. In chapter 5, we demonstrated that information obtained from the fragmentation of small glycan standards can be used to solve the (partial) structure of more complex glycans. Fragmentation of complex glycan can be used to generate substructures that can, in turn, be assigned using information gained from standards that contain the minimal epitope of interest making them relatively accessible.[15] Such an approach could provide sufficient information to solve the complete structure if enough identifying features can be generated and assigned. Outlined in the following two sections are proposals that identify structural features of complex oligosaccharide fragments which might provide valuable information to help determine exact glycan structure.

CCS library for the exact assignment of terminal glycan epitopes

The terminal epitopes present on complex glycans are often important mediators of their interaction with glycan binding proteins.[16] Most terminal epitopes can be expressed on either a type 1 or type 2 LacNAc base, which results in isomeric structures carrying similar molecular features in a different spatial organization.[17] It is challenging to distinguish these isomers by common techniques such as mass spectrometry since advanced and laborious methods are required.[18] Furthermore, the terminal epitope can rarely be assigned based on glycan composition. For example, glycans carrying fucosylated terminal epitopes such as sialyl Lewis^x are often not distinguishable from internally fucosylated glycans. In addition, epitopes carrying monosaccharides with similar masses as those found in the polyLacNAc backbone (Hex and HexNAc), such as the GalNAc or Gal found in blood group A and B epitopes, can easily cause misassignments in the glycan core or the polyLacNAc backbone of the oligosaccharides.

In the work described in chapter 5, we demonstrated how the *O*-acetylated sialyl LacNAc fragments from complex glycans can be used to assign *O*-acetylation position and sialic acid linkage. We anticipated that fragments could be obtained for other epitopes which may allow us to resolve the structure of terminal epitopes. As a preliminary study, human glycosyltransferases were used to generate a complete set of Lewis- and ABO blood group antigens of both type 1 and type 2 LacNAc.

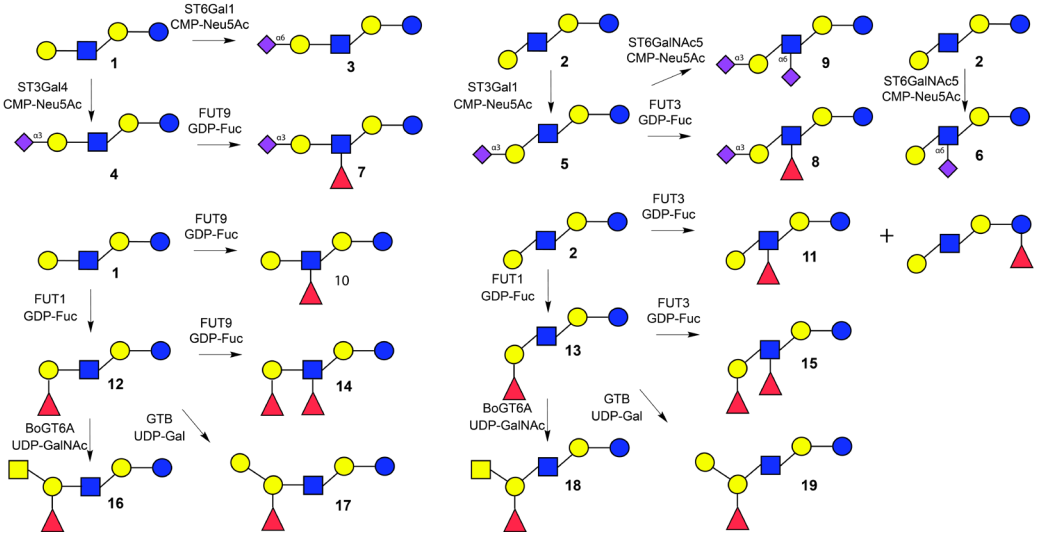


Figure 2: Schematic of the enzymatic synthesis of a complete library of Lewis and blood group containing human milk oligosaccharides.

The human milk oligosaccharides LnNT (**1**) and LNT (**2**) form suitable starting materials for the enzymatic synthesis of a library of glycans since these compounds are commercially available and already have type 1 and 2 LacNAc installed at the terminal position (Fig. 2). Furthermore, the lactose core has a different mass then the terminal LacNAc so possible impurities or side reactions on the lactose do not lead to misinterpretation of the glycan fragments. LnNT was treated with either ST6Gal1 or ST3Gal4 to form **3** and **4** with an α2,6 and α2,3 linked Neu5Ac respectively. The α2,6 linkage to galactose on type 1 LacNAc is not present in humans so

only the α2,3 linked sialic acid **5** was generated by the action of ST3Gal1. Another sialylation reaction can be conducted on **5** and **2** to form disialoglycan **9** and sialoglycan **6**, which are epitopes found only on the type 1 backbone. Glycan **4** could be selectively fucosylated by FUT9 to form the sialyl Lewis^x epitope **7.** This selectivity was obtained because FUT9 is unable to efficiently fucosylated lactose. The type 1 isomer **8** was generated using the enzyme FUT3, since FUT9 does not accept type 1 substrates, FUT3 was fully selective in this transformation.

Next, focus was shifted to the asialo epitopes: fucosylation of **1** using FUT9 generates **10** selectively or generates **12** with FUT1. The type 1 isomers could be generated by FUT3 for compound **11** and FUT1 for compound **13**. Compound **11** was generated as a 1:1 mixture but could be separated by preparative HPLC using HILIC chromatography. Fucoside **12** could again be fucosylated by FUT9 to generate the Lewis^y antigen **14** selectively. Transformation of fucoside **13** to the Lewis^b antigen **15** by FUT3 was also selective due to the presence of the terminal α1,2 fucose. Compounds **12** and **13** were transformed from the blood group O antigen to blood group A and B antigens by the action of BoGT6A and GTB respectively forming compounds **16-19**. This concludes the synthesis of a complete set of Lewis and ABO blood group epitopes on HMO scaffolds.

The HMOs were purified by preparative HPLC and characterized by NMR (see supplementary information) to ensure correct glycan structure and purity. After conformation of the correct stereo- and regiochemistry by NMR, all structures were analyzed by high resolution IM-MS in triplicate. Mild in-source fragmentation (360V) was applied to efficiently generate arrival time distributions of the terminal epitopes. Epitope fragments were solely present as B_i oxonium ions in positive ion mode. The CCS of each epitope B_i fragment was subsequently computed. The arrival time distributions of the isomeric epitopes illustrated that CCS of terminal epitope fragments can indeed be used to discriminate between type 1 and 2 isomers as well as other regional isomers (Fig. 3). The high-resolution post processing provided by HRdm 2.0 allowed us to reach resolutions of between 200 to 350 for most epitopes, which is over three times the normal resolution of DTIMS and essential to distinguish less resolved epitopes such as monofucosides. The CCS values were highly reproducible across the repeated measurements, as the standard deviations did not exceed 0.11 Å^2.

These preliminary results show that isomeric terminal glycan epitope B_i fragments can be distinguished by mild in-source fragmentation of the glycan in combination with IM-MS. Further studies are required to determine if these CCS can be used to assign exact terminal epitope structure of complex glycans in mixtures. A limitation of this strategy in the current IM-MS set-up was that no ion selection can be performed before fragmentation, which prevents the accurate assignment when multiple glycans co-elude. Ion selection in both MS and IM axis would allow the analysis of single isomers, but this is currently not yet feasible in combination with IM-MS fragment analysis. This limitation can be circumvented by complete chromatographical separation of the glycan isomers which can be challenging dependent on sample complexity. Another potential implementation of this terminal glycan epitope analysis can be to screen samples for glycan epitopes. Such a method could be used to determine genetic glycosylation features such as blood group or secretor status.

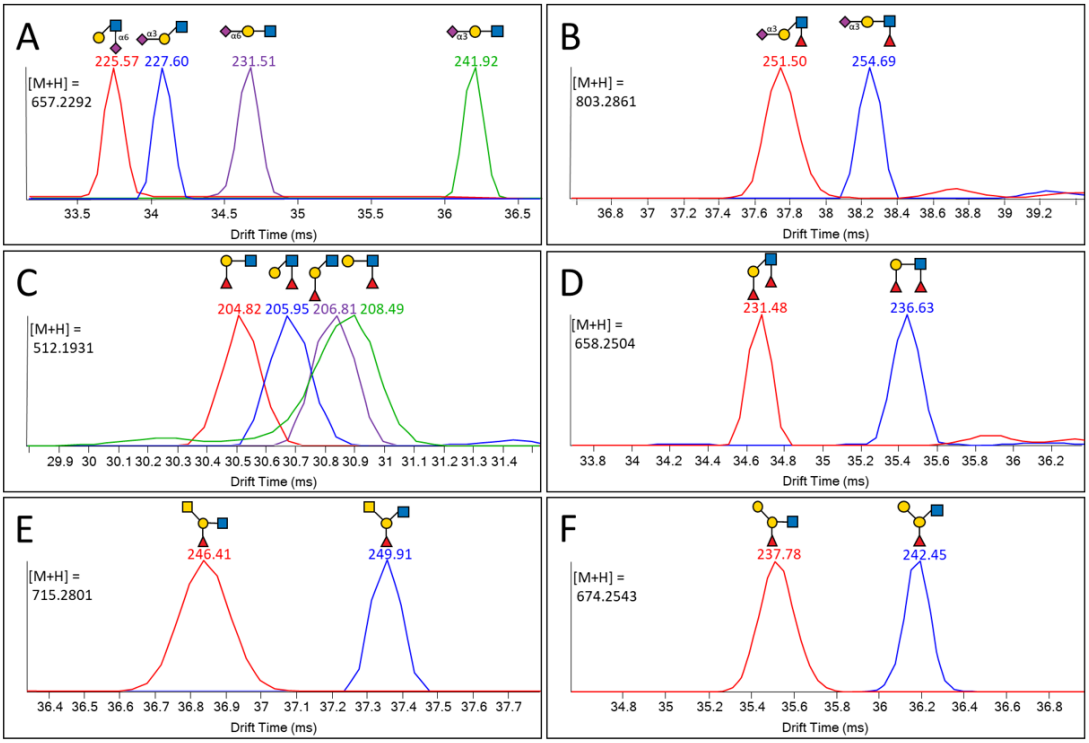


Figure 3: Arrival time distributions of B_i epitope fragments generated by in-source fragmentation of human milk oligosaccharides containing various terminal epitopes with fragment structure and CCS. A) Sialyl-LacNAcs, B) Sialyl-Lewis epitopes, C) Fucosides, D) Difucosides, E) Blood group A, F) Blood group B

Core assignment

Both *O*- and *N*-glycans can be classified by their core, which is based on the common glycan structures found on the reducing end part of the glycan.[19,20] Complex type *N*-glycans can exist as bi-, tri- or tetra-antennary *N*-glycans that all may contain a core fucose. Additional to that, bi-antennary glycans can contain a bisecting GlcNAc which is not enzymatically extended. Furthermore, tri-antennary glycans can be present as two regioisomers leading to a total of 12 possible combinations. This core diversity results in a large increase in the number of possible isomers: there are already 8 possible isomers for a relatively simple *N*-glycan with a composition of N6H7 (Fig. 4A).

O-glycans can be classified as one of eight core structures, all consisting of a combination of GalNAc/GlcNAc and/or galactose. The determination of *O*-glycan core structure can in turn provide useful information on the rest of the glycan structure. For example, a glycan with a composition such as N3H2FS can either be a blood group A containing core 1 *O*-glycan, a polyLacNAc core 3 *O*-glycan or, a multi-antennary core 4 *O*-glycan (Fig. 4B). Assignment of the *O*-glycan core type would directly resolve the rest of the carbohydrate structure or greatly reduce the space of possibilities. However, these core structures cannot be readily discriminated by MS but require an additional axis of separation such as ion-mobility.

Currently, most *N*-glycomics experiments still rely on compositional analysis wherein the core composition of *N*-glycans is mostly guessed. The first observed fucose is often assigned as core fucosylation but can just as well be part of a Lewis epitope. Furthermore, the amount of antenna present on the *N*-glycans is commonly deduced from the amount of LacNAcs observed, which results in a misassignment of polyLacNAc containing *N*-glycans. The correct assignment of *N*-glycan core structure can also be important to, for example, assess cancer-associated glycome features such as bulk tri- and tetra-antennary *N*-glycans, *N*-glycan core fucosylation and, bisecting N-glycans.[21] Because it is challenging to obtain this structural information, it is often not collected.

The structural assignment of the *N*-glycan core may be addressed by IM-MS fragment analysis as has already been demonstrated for several isomeric branched *N*-glycan cores.[22] Pure *N*-glycan standards can be analyzed by IM-MS in combination with mild fragmentation to create a library of CCS and *m/z* encompassing all *N*-glycan core structures. A complete library of *N*-glycan core fragments can then possibly provide information on the important structural features that can be present on the *N*-glycan core without the need for full structural characterization of the *N*-glycome. Focusing analysis on the core region makes the synthesis of all fragments of interest possible. In addition, the branch substitutions can be ignored which limits the complexity of data analysis. Such an IM-MS approach can be a tool in the development of IM-MS based cancer detection methods. The same information on *N*-glycan core structure may find use in glycomics to greatly reduce the potential isomers of a known glycan composition.

O-glycomics has not seen the same developments as *N*-glycomics due to the difficulties associated with glycan release, this problem was partly addressed in chapter 3. The effect of the various *O*-glycan core structures on the glycan binding has yet to be established, partly due to a lack of synthetic standards and practical analytical methods to assess core structure in natural samples.[23] Recently, synthetic methodology for the preparation of *O*-glycan standards has become available.[24] These synthetic standards can now be used to create a library of CCS and *m/z* encompassing all *O*-glycan core structures. This library might then be used to determine *O*-glycan core structure by IM-MS combined with fragmentation. Such a method could provide useful information that can be used to solve exact glycan structure and dissect the biological function of the various *O*-glycan core structures.

Solving the right combination of sub-structures may allow us to solve exact glycan structure without the need for all possible isomeric synthetic standards (Fig. 4C). After the initial proof-of-concept studies, there will be a significant challenge to develop instrumentation that is able to select single isomer glycans and then apply fragmentation before IM-MS analysis of these fragments. For implementation in glycomics, there will also be a need for the development of software that can assign glycan fragment structure and use this information to predict complete glycan structure.

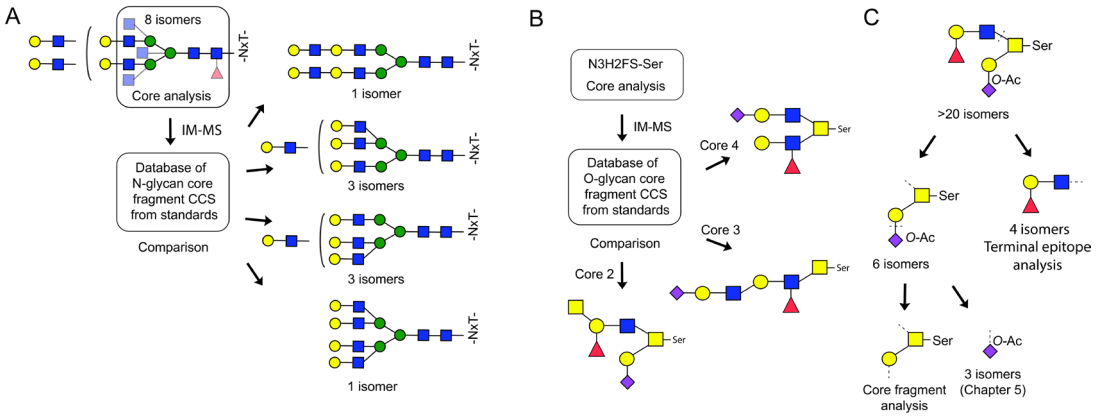


Figure 4: A) Information of the *N*-glycan core architecture can be used to discriminate multi-antennary N-glycans from polyLacNAc containing *N*-glycans. B) Determination of the *O*-glycan core structure reduces the possible *O*-glycans of a given composition and can give information on glycan architecture. C) The exact identification of multiple glycan fragments can be used to solve exact glycan structure of complex glycans.

References

[1] B. Caterson, J. Melrose, *Glycobiology* **2018**, *28*, 182–206.

[2] S. Puri, Y. M. Coulson-Thomas, T. F. Gesteira, V. J. Coulson-Thomas, *Front. Cell Dev. Biol.* **2020**, *8*, 731.

[3] M. Wendel, Y. Sommarin, D. Heinegård, *J. Cell Biol.* **1998**, *141*, 839–847.

[4] H. U. Choi, K. Meyer, *Biochem. J.* **1975**, *151*, 543–553.

[5] H. S. Hahm, F. Broecker, F. Kawasaki, M. Mietzsch, R. Heilbronn, M. Fukuda, P. H. Seeberger, *Chem* **2017**, *2*, 114–124.

[6] J. R. Grunwell, V. L. Rath, J. Rasmussen, Z. Cabrilo, C. R. Bertozzi, *Biochemistry* **2002**, *41*, 15590–15600.

[7] J. L. Funderburgh, *Glycobiology* **2000**, *10*, 951–958.

[8] H. Hoshino, T. Foyez, S. Ohtake-Niimi, Y. Takeda-Uchimura, M. Michikawa, K. Kadomatsu, K. Uchimura, *J. Histochem. Cytochem.* **2014**, *62*, 145–156.

[9] T. Kumagai, T. Kiwamoto, M. E. Brummet, F. Wu, K. Aoki, Z. Zhu, B. S. Bochner, M. Tiemeyer, *Glycobiology* **2018**, *28*, 406–417.

[10] J. L. Daughtry, W. Cao, J. Ye, J. M. Baskin, *ACS Chem. Biol.* **2020**, *15*, 318–324.

[11] J. Hofmann, A. Stuckmann, M. Crispin, D. J. Harvey, K. Pagel, W. B. Struwe, *Anal. Chem.* **2017**, *89*, 2318–2325.

[12] M. Grabarics, M. Lettow, C. Kirschbaum, K. Greis, C. Manz, K. Pagel, *Chem. Rev.* **2022**, *122*, 7840–7908.

[13] K. Greis, C. Kirschbaum, G. von Helden, K. Pagel, *Curr. Opin. Struct. Biol.* **2022**, *72*, 194–202.

[14] H. Hinneburg, J. Hofmann, W. B. Struwe, A. Thader, F. Altmann, D. V. Silva, P. H. Seeberger, K. Pagel, D. Kolarich, *Chem. Commun.* **2016**, *52*, 4381–4384.

[15] T. L. Peterson, G. Nagy, *Anal. Chem.* **2021**, *93*, 9397–9407.

[16] C. D. Rillahan, J. C. Paulson, *Annu. Rev. Biochem.* **2011**, *80*, 797–823.

[17] P. Stanley, R. D. Cummings, in *Essent. Glycobiol.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2015**.

[18] H.-T. K. Wong, X. Chen, S. Zhang, T.-Y. Lui, D. Hu, T.-W. D. Chan, *J. Am. Soc. Mass Spectrom.* **2022**.

[19] P. Stanley, K. W. Moremen, N. E. Lewis, N. Taniguchi, M. Aebi, *N-Glycans*, Cold Spring Harbor Laboratory Press, **2022**.

[20] I. Brockhausen, H. H. Wandall, K. G. T. Hagen, P. Stanley, *O-GalNAc Glycans*, Cold Spring Harbor Laboratory Press, **2022**.

[21] S. V. Glavey, D. Huynh, M. R. Reagan, S. Manier, M. Moschetta, Y. Kawano, A. M. Roccaro, I. M. Ghobrial, L. Joshi, M. E. O'Dwyer, *Blood Rev.* **2015**, *29*, 269–279.

[22] E. G. Pallister, M. S. F. Choo, I. Walsh, J. N. Tai, S. J. Tay, Y. S. Yang, S. K. Ng, P. M. Rudd, S. L. Flitsch, T. Nguyen-Khuong, *Anal. Chem.* **2020**, *92*, 15323–15335.

[23] H. H. Wandall, M. A. I. Nielsen, S. King-Smith, N. de Haan, I. Bagdonaite, *FEBS J.* **2021**, *288*, 7183–7212.

[24] S. Wang, C. Chen, M. R. Gadi, V. Saikam, D. Liu, H. Zhu, R. Bollag, K. Liu, X. Chen, F. Wang, P. G. Wang, P. Ling, W. Guan, L. Li, *Nat. Commun.* **2021**, *12*, 3573.

Supplementary information

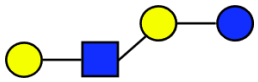
Materials and methods

^{1}H NMR are recorded on a Bruker 600 MHz Avance Neo. Samples are dissolved in 99.9% D_2O. NMR Chemical shifts are recorded in parts per million (ppm) relative to the residual H_2O (δ 4.790) peak. NMR signals are represented as: chemical shift and multiplicity (d = doublet, t = triplet, q = quadruplet, dd = doublet of doublets, app. = apparent) and J coupling if applicable. Assignment of ^{1}H NMR signals was conducted based on ^{1}H-NMR, 2D COSY (COSYGPSW), 2D TOCSY (MLEVPHSW), 2D NOESY (NOESYPHSW) and 2D HSQC (HSQCEDETGPSISP). ^{13}C NMR signal assignment was extracted from 2D HSQC spectra. Reaction progress was determined with a SeQuant Zic HILIC guard column (20x2.1 mm) coupled to a Bruker micrOTOF-QII MS System running a gradient of 90% Acetonitrile:H_2O to 50% acetonitrile:H_20 over 5 minutes followed by isocratic 50% acetonitrile Acetonitrile:H_2O until product was detected. Size exclusion chromatography was conducted using Bio-Gel P2 gel polyacrylamide beads from BioRad, eluting with MilliQ water. Silica gel-coated glass plates were used to locate glycan containing fractions. Glycan was visualized by dipping the plate in a 10% sulfuric acid solution in ethanol followed by charring. Preparative HPLC was conducted on a Shimadzu Prominence HPLC system, consisting of two LC-20AT (L20115), a SIL-20A (L201656), a CBM-20A (L202355) and a FRC-10A (C203755) connected to a Bruker microTOF MS using a 10x250 mm 5 µ Waters Xbridge amide BEH prep column (186006602) in sequence with a Shimadzu SPD-20AV UV detector for visualization.

Mass spectra and Ion mobility spectra are recorded by ESI on an Agilent 6560 Ion mobility Q-TOF LC/MS system connected to an Agilent infinity LC system, consisting of a 1290 Binary pump (DEBAA04032), a 1290 sampler (DEBAP04954) and a 1260 thermostatted column compartment (DEACN24533). CCS of standards were obtained by a 1ul injection of a 1 µg/µL solution in 50% acetonitrile:H_2O without column at an isocratic flow of 0.2 mL/min of 50% acetonitrile:H_2O containing 0.1% formic acid. Agilent ESI source parameters were set at capillary voltage of 3500 V, nozzle voltage of 2000 V, nebulizer pressure of 40 psi, N_2 drying gas heated to 300 °C at 8L/min flow, and sheath gas heated to 300 °C at 11 L/min flow. The Ion Mobility was set at an IM Transient Rate of 16 transients/frame, with a Trap Fill Time of 3500 µs, a Trap Release Time of 250 µs, a Drift Tube Entrance Voltage of 1400 V, and a Multiplexing Pulsing Sequence Length of 4 bit.

MilliQ water was purified by a Millipore Synergy water purification system. UDP-Galactose, UDP-*N*-Acetylglucosamine, UDP-*N*-Acetylgalactosamine and CMP-Neu5Ac were bought from Roche Diagnostics. Calf intestine alkaline phosphatase (CIAP) was obtained from Invitrogen. GDP-Fucose was prepared using L-fucokinase/GDP-fucose pyrophosphorylase as described in chapter 2. Lacto-N-tetraose was a gift from FrieslandCampina and purified by preparative HPLC before use. Reagents purchased from commercial sources were used without further purification.

Synthesis of Lacto-*N*-neotetraose/LNnT (1)



1

Lactose (30 mg, 69.4 mmol) was dissolved in HEPES buffer (50 mM HEPES, 25 mM KCl, 2 mM $MgCl_2$, 1 mM DTT, pH 7.3) at a concentration of 10 mM. To this, UDP-GlcNAc (1.5 eq, 15 mM), B3GNT2 (10% v/v) and CIAP (1% v/v) were added, the mixture was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. The reaction mixture was passed through a Vivaspin® 20 10 kDa spinfilter by centrifugation at 4500 rpm. Filtrate was adjusted to a pH of 7-8 with NaOH (100 mM). UDP-gal (1.5 eq, 15mM), CIAP (1% v/v) and B4GalT1 (2% v/v) was added to the mixture, which was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. After completion, the mixture was freeze dried and **1** was purified using Bio-gel P2 size exclusion chromatography eluding with H_2O. Fractions containing carbohydrate were identified by spotting 0.3 µL of each collected fraction on a TLC plate followed by a dip in 10% H_2SO_4 and charring. Fractions containing carbohydrate were analyzed by LC/MS and fractions with product was observed were pooled and freeze dried. The solids were purified by preparative HPLC (19 mg, 26.9 mmol, 39% over 2 steps).

^1H (600 MHz, D_2O): δ (ppm)

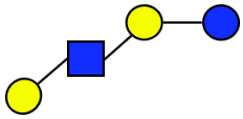
	H1	H2	H3	H4	H5	H6	Ac
Glc(α)	5.23 (d, J=3.8)	3.58	3.84	3.65	n/a	n/a	-
Glc(β)	4.67 (d, J=8.0)	3.28 (app. t)	3.64	3.65	n/a	n/a	-
Gal	4.44 (d, J=7.82)	3.60	3.73	4.16 (d, J=3.33)	n/a	n/a	-
GlcNAc	4.71 (d, J=8.36)	3.81	3.73	3.59	n/a	n/a	2.04
Gal (2)	4.48 (d, J=7.86)	3.54	3.68	3.93	n/a	n/a	-

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.63
Glc(β)	95.65
Gal	102.92
GlcNAc	102.64
Gal (2)	102.81

ESI TOF-MS m/z calculated for $C_{26}H_{45}NO_{21}Na$ $[M + Na]^+$ = 730.2376, found 730.2311

Purification Lacto-*N*-tetraose/LNT (2)



2

Crude Lacto-*N*-tetraose (100 mg) was purified by Bio-gel P2 size exclusion chromatography eluding with H_2O to remove most lactose and lacto-N-triose impurities. Fractions containing only product were pooled and freeze dried (45 mg); 10 mg of the product was additionally purified by preparative HPLC resulting in 8 mg of high purity LNT.

^1H (600 MHz, D_2O): δ (ppm)

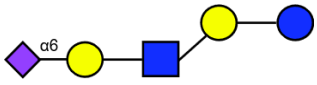
	H1	H2	H3	H4	H5	H6	Ac
Glc(α)	5.23 (d, J=3.7)	3.58	3.83	3.65	n/a	n/a	-
Glc(β)	4.58 (d, J=7.9)	3.29	3.65	3.61	n/a	n/a	-
Gal	4.44 (d, J=7.9)	3.61	3.74	4.16	n/a	n/a	-
GlcNAc	4.74	3.90	3.79	3.58	n/a	n/a	2.03
Gal (2)	4.44 (d, J=7.9)	3.53	3.64	3.92	n/a	n/a	-

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	95.73
Glc(β)	91.75
Gal	103.22
GlcNAc	102.63
Gal (2)	103.22

ESI TOF-MS m/z calculated for $C_{26}H_{45}NO_{21}Na$ $[M + Na]^+$ = 730.2376, found 730.2318

Synthesis of α2,6 Sialyllacto-*N*-neotetraose/ LSTc (3)



3

LNnT (2 mg, 2.8 µmol) was dissolved in 100 mM TRIS (pH 7.4) to a concentration of 10 mM, and CMP-Sia (1.5 eq, 15 mM), bovine serum albumin (0.1% wt/v), CIAP (1% v/v) and ST6Gal1 (20% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. Upon completion, the mixture was freeze dried and **3** was purified using size exclusion chromatography and preparative HPLC as described above (2.4 mg, 86%, 2.4 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

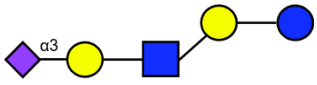
	H1	H2	H3	H4	H5	H6	H7	H8	H9	Ac
Glc(α)	5.23 (d, J=3.80)	3.57	3.83	3.65	n/a	n/a	-	-	-	-
Glc(β)	4.67 (d, J=7.98)	3.28 (app. t)	3.65	3.84	n/a	n/a	-	-	-	-
Gal	4.44 (	3.61	3.73	4.16 (d, J=3.31	n/a	n/a	-	-	-	-
GlcNAc	4.74 (d, J=6.64)	3.81	3.61	3.84	n/a	n/a	-	-	-	2.03
Gal (2)	4.46	3.54	3.68	3.93	n/a	n/a	-	-	-	-
Sia	-	-	1.73 (t, J=12.19), 2.68 (dd, J=4.68/12.42)	3.67	3.81	3.57	n/a	n/a	n/a	2.06

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.68
Glc(β)	95.66
Gal	103.14
GlcNAc	102.52
Gal (2)	103.14

ESI TOF-MS m/z calculated for $C_{37}H_{62}N_2O_{29}Na$ $[M + Na]^+$ = 1021.3330, found 1021.3239

Synthesis of α2,3 Sialyllacto-*N*-neotetraose/LSTd (4)



4

LNnT (4 mg, 5.6 µmol) was dissolved in 100 mM TRIS (pH 7.4) to a concentration of 10 mM, and CMP-Sia (1.5 eq, 15 mM), bovine serum albumin (0.1% wt/v), CIAP (1% v/v) and ST3Gal4 (30% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. Upon completion, the mixture was freeze dried and **4** was purified using size exclusion chromatography and preparative HPLC as described above (3.9 mg, 70%, 3.9 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	H7	H8	H9	Ac
Glc(α)	5.22 (d, J=3.80)	3.58	3.84	3.65	n/a	n/a	-	-	-	-
Glc(β)	4.67 (d, J=7.90)	3.28 (app. t)	3.65	3.80	n/a	n/a	-	-	-	-
Gal	4.45 (d, J=7.90)	3.60	3.73	4.17 (d, J=3.34)	n/a	n/a	-	-	-	-
GlcNAc	4.71 (d, J=8.34)	3.81	3.58	3.75	n/a	n/a	-	-	-	2.05 (6H)

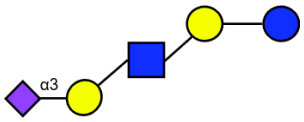
Gal (2)	4.57 (d, J=7.85)	3.57	4.12	3.96 (dd, J=3.14/9.87)	3.87	n/a	-	-	-	-
Sia	-	-	1.80 (t, J=12.11), 2.77 (dd, J=4.62/12.45)	3.70	3.86	3.65	n/a	n/a	n/a	2.05 (6H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.73
Glc(β)	95.69
Gal	102.90
GlcNAc	102.79
Gal (2)	102.51

ESI TOF-MS m/z calculated for $C_{37}H_{62}N_2O_{29}Na$ $[M + Na]^+$ = 1021.3330, found 1021.3226

Synthesis of α2,3 Sialyllacto-*N*-tetraose/LSTa (5)



5

LNT (7.5 mg, 10.6 µmol) was dissolved in 100 mM TRIS (pH 7.4) to a concentration of 10 mM, and CMP-Sia (1.5 eq, 15 mM), bovine serum albumin (0.1% wt/v), CIAP 1% (v/v) and ST3Gal1 (10% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. Upon completion, the mixture was freeze dried and **5** was purified using size exclusion chromatography and preparative HPLC as described before (4 mg, 38%, 4.0 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

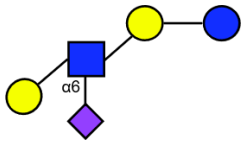
	H1	H2	H3	H4	H5	H6	H7	H8	H9	Ac
Glc(α)	5.23 (d, J=3.76)	3.59 (app. t)	3.84	3.65	n/a	n/a	-	-	-	-
Glc(β)	4.67 (d, J=7.95)	3.28	3.65	3.96	n/a	n/a	-	-	-	-
Gal	4.45 (d, J=7.85)	3.60	3.74	4.16 (d, J=3.33)	n/a	n/a	-	-	-	-
GlcNAc	4.75 (d, J=8.48)	3.90	3.81	3.49	n/a	n/a	-	-	-	2.05 (6H)
Gal (2)	4.52 (d, J=7.76)	3.55	4.09 (dd, J=3.13,9.81)	3.94	n/a	n/a	-	-	-	-
Sia	-	-	1.79 (t, J=12.16), 2.77 (dd, J=4.61,12.45)	3.69	3.85	n/a	n/a	n/a	n/a	2.05 (6H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.75
Glc(β)	95.67
Gal	102.93
GlcNAc	102.43
Gal (2)	103.33

ESI TOF-MS m/z calculated for $C_{37}H_{62}N_2O_{29}Na$ $[M + Na]^+$ = 1021.3330, found 1021.3239

Synthesis of β-Gal(1,3)-[α-Sialyl(2,6)]-β-GlcNAc(1,3)-β-Gal(1,4)-Glc/LSTb (6)



6

LNT (4 mg, 5.6 µmol) was dissolved in 50 mM TRIS (pH 7.5) to a concentration of 10 mM, and CMP-Sia (1.5 eq, 15 mM), bovine serum albumin (0.1 wt%), CIAP (1% v/v) and ST6GalNAc5 (50% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. When no further reaction progress was observed after 2 days, the mixture was freeze dried and **6** was purified using size exclusion chromatography and preparative HPLC as described before (2.6 mg, 46%, 2.6 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	H7	H8	H9	Ac
Glc(α)	5.22 (d, J=3.78)	3.59	3.81	3.65	n/a	n/a	-	-	-	-
Glc(β)	4.67 (d, J=8.03)	3.28	3.65	3.61	n/a	n/a	-	-	-	-
Gal	4.44 (d, J=7.81)	3.59	3.73	4.29 (d, J=3.26)	n/a	n/a	-	-	-	-
GlcNAc	4.70 (d, J=8.52)	3.91	3.81	3.64	3.56	3.77, 3.96	-	-	-	2.05 (6H)
Gal (2)	4.44 (d, J=7.81)	3.53	3.63	3.91	n/a	n/a	-	-	-	-
Sia	-	-	1.70 (t, J=12.41), 2.75 (dd, J=4.68/12.42)	3.68	3.83	n/a	n/a	n/a	n/a	2.05 (6H)

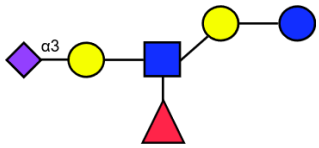
^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.71
Glc(β)	95.69
Gal	103.15

Gal	103.15
GlcNAc	102.59
Gal (2)	103.15

ESI TOF-MS m/z calculated for $C_{37}H_{62}N_2O_{29}Na$ $[M + Na]^+$ = 1021.3330, found 1021.3237

Synthesis of sialyl Lewis^x lacto-*N*-neohexose/SLeXLNnH (7)



7

4 (1 mg, 1 µmol) was dissolved in 100 mM TRIS (pH 7.5) to a concentration of 10 mM, and GDP-Fuc (0.5 eq, 5 mM), $MnCl_2$ (10 mM, from 1 M solution at 1% v/v of reaction volume), CIAP (1% v/v) and FUT9 (10% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. 0.1 eq. of GDP-Fucose was added each hour until the reaction went to completion, after which the mixture was freeze dried. Size exclusion chromatography and preparative HPLC were used to obtain **7** (0.4 mg 31%, 3.1 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

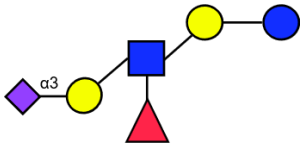
	H1	H2	H3	H4	H5	H6	H7	H8	H9	Ac
Glc(α)	5.23 (d, J=3.82)	3.58	3.83	3.65	n/a	n/a	-	-	-	-
Glc(β)	4.67 (d, J=7.92)	3.28 (app. t, J=8.49)	3.64	3.80	n/a	n/a	-	-	-	-
Gal	4.44 (d, J=7.90)	3.67	3.71	4.17 (d, J=3.32)	n/a	n/a	-	-	-	-
GlcNAc	4.71 (d, J=9.28)	3.97	3.88	3.60	n/a	n/a	-	-	-	2.05 (6H)
Gal (2)	4.53 (d, J=7.79)	3.53	4.09 (dd, J=3.19/ 9.84)	3.94	n/a	n/a	-	-	-	-
Fuc	5.13 (d, J=4.04)	3.68	3.90	3.78	4.83	1.17 (d, J=6.54)	-	-	-	-
Sia	-	-	1.80 (t, J=12.17), 2.77 (dd, J=4.61/12.48)	3.69	3.86	n/a	n/a	n/a	n/a	2.05 (6H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.83
Glc(β)	95.68
Gal	102.68
GlcNAc	102.53
Gal (2)	101.42
Fuc	98.46

ESI TOF-MS m/z calculated for $C_{43}H_{72}N_2O_{33}Na$ $[M + Na]^+$ = 1167.3910, found 1167.3804

Synthesis of sialyl Lewis^a lacto-*N*-hexose/SLeALNH (8)



8

5 (1 mg, 1 µmol) was dissolved in 100 mM TRIS (pH 7.5) to a concentration of 10 mM, and GDP-Fuc (0.5 eq, 5 mM), $MnCl_2$ (10 mM, from 1 M solution at 1% v/v of reaction volume), CIAP (1% v/v) and FUT3 (10% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. 0.1 eq. of GDP-Fucose was added each hour until the reaction went to completion, after which the mixture was freeze dried. Size exclusion chromatography and preparative HPLC were used to obtain **8** (1 mg 87%, 0.9 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

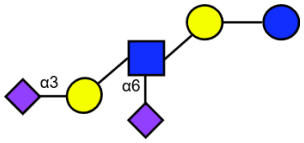
	H1	H2	H3	H4	H5	H6	H7	H8	H9	Ac
Glc(α)	5.23 (d, J=3.78)	3.58	3.84	3.65	n/a	n/a	-	-	-	-
Glc(β)	4.67 (d, J=7.94)	3.28	3.65	n/a	n/a	n/a	-	-	-	-
Gal	4.44 (d, J=7.88)	3.59	3.72	4.16 (d, J=3.32)	n/a	n/a	-	-	-	-
GlcNAc	4.71 (d, J=9.51)	3.94	4.10 (t, J=9.83)	3.73	3.53	n/a	-	-	-	2.05 (6H)
Gal (2)	4.55 (d, J=7.70)	3.51	4.06 (dd, J=3.14/9.78)	3.92	n/a	n/a	-	-	-	-
Fuc	5.02 (d, J=4.00)	3.79	3.88	3.78	4.88 (q, J=6.64)	1.18 (3H, d, J=6.57)	-	-	-	-
Sia	-	-	1.77 (t, J=12.81), 2.77 (dd, J=4.69/12.44)	3.68	3.84	n/a	n/a	n/a	n/a	2.05 (6H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.77
Glc(β)	95.72
Gal	102.96
GlcNAc	102.47
Gal (2)	102.67
Fuc	97.93

ESI TOF-MS m/z calculated for $C_{43}H_{72}N_2O_{33}Na$ $[M + Na]^+$ = 1167.3910, found 1167.3798

Synthesis of Disialyllacto-*N*-tetraose/DSLNT (9)



9

LNT (10 mg, 14.1 µmol) was dissolved in 50 mM TRIS (pH 7.5) to a concentration of 10 mM, and CMP-Sia (3 eq, 30 mM), bovine serum albumin (0.1 wt%), CIAP (1% v/v) and ST3Gal1 (20% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. When no starting material was detected, ST6GalNAc5 (10% v/v) was added, and the reaction was incubated again. Once no more mono-sialylated product was detected, the mixture was freeze dried and **9** was purified using size exclusion chromatography and preparative HPLC as described before (11.7 mg, 64%, 9.0 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

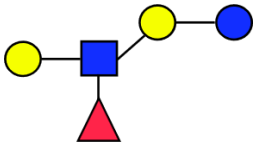
	H1	H2	H3	H4	H5	H6	H7	H8	H9	Ac
Glc(α)	5.22 (d, J=3.77	3.59 (app. t)	3.84	3.65	n/a	n/a	-	-	-	-
Glc(β)	4.67 (d, J=7.95)	3.29	3.65	3.61	n/a	n/a	-	-	-	-
Gal	4.44 (d, J=7.87)	3.60	3.73	4.17 (d, J=3.30)	n/a	n/a	-	-	-	-
GlcNAc	4.71 (d, J=8.44)	3.90	3.80	3.62	3.56	3.99	-	-	-	2.07 (3H)
Gal (2)	4.51 (d, J=7.80)	3.54	4.09 (d, J=9.84)	3.94	n/a	n/a	-	-	-	-
Sia	-	-	1.70 (t, J=12.16), 2.75 (m)	3.70	3.84	n/a	n/a	n/a	n/a	2.03 (6H)
Sia(2)	-	-	1.80 (t, J=12.15), 2.77 (m)	3.68	3.85	n/a	n/a	n/a	n/a	2.03 (6H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.81
Glc(β)	95.75
Gal	102.96
GlcNAc	102.51
Gal (2)	103.46

ESI TOF-MS m/z calculated for $C_{48}H_{79}N_3O_{37}Na$ $[M + Na]^+$ = 1312.4285, found 1312.4184

Synthesis of Lewis^x lacto-*N*-neopentose/LNFPIII (10)



10

LnNT (3 mg, 4.2 µmol) was dissolved in 100 mM TRIS (pH 7.5) to a concentration of 10 mM, and GDP-Fuc (0.5 eq, 5 mM), $MnCl_2$ (10 mM, from 1 M solution at 1% v/v of reaction volume), CIAP (1% v/v) and FUT9 (10% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. 0.1 eq. of GDP-Fucose was added each hour until the reaction went to completion, after which the mixture was freeze dried. Size exclusion chromatography and preparative HPLC were used to obtain **10** (1.7 mg, 48%, 2.0 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

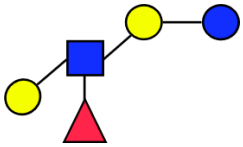
	H1	H2	H3	H4	H5	H6	Ac
Glc(α)	5.23 (d, J=3.77)	3.58 (app. t)	3.83	3.65	n/a	n/a	-
Glc(β)	4.67 (d, J=7.91)	3.28	3.64	3.61	n/a	n/a	-
Gal	4.44 (d, J=7.87)	3.58	3.71	4.16 (d, J=3.38)	n/a	n/a	-
GlcNAc	4.72 (d, J=8.35)	3.97	3.89	3.59	n/a	n/a	2.03 (3H)
Gal (2)	4.47 (d, J=7.83)	3.50 (dd, J=7.80,9.84)	3.66	3.91	n/a	n/a	-
Fuc	5.13 (d, J=4.03)	3.70	3.91	3.81	4.84	(1.18, 3H)	-

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.76
Glc(β)	95.66
Gal	102.81
GlcNAc	102.49
Gal (2)	101.73
Fuc	98.49

ESI TOF-MS m/z calculated for $C_{32}H_{55}NO_{25}Na$ $[M + Na]^+$ = 876.2955, found 876.2888

Synthesis of Lewis^a lacto-*N*-pentose/LNFPII (11)



11

LNT (3 mg, 5.6 µmol) was dissolved in 100 mM TRIS (pH 7.5) to a concentration of 10 mM, and GDP-Fuc (0.5 eq, 5 mM), $MnCl_2$ (10 mM, from 1 M solution at 1% v/v of reaction volume), CIAP (1% v/v) and FUT3 (10% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. 0.1 eq. of GDP-Fucose was added each hour until only limited amounts of starting material were observed, after which the mixture was freeze dried. NMR after size exclusion chromatography revealed a ratio of 1:1 fucosylation of GlcNAc and glucose. FUT5 led to a ratio of 1:9 and FUT6 and FUT 9 were not able to fucosylate type 1 LacNAc. The mono-fucosylated isomers were separated on preparative HPLC to obtain **11** (1.2 mg, 25%, 1.4 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

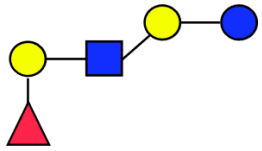
	H1	H2	H3	H4	H5	H6	Ac
Glc(α)	5.22 (d, J=3.78)	3.58	3.84	3.65	n/a	n/a	-
Glc(β)	4.67 (d, 7.88)	3.28 (app. t, J=8.48)	3.64	3.95	n/a	n/a	-
Gal	4.44 (d, J=7.87)	3.60	3.72	4.16 (d, J=3.33)	n/a	n/a	-
GlcNAc	4.71 (d, J=8.54)	3.95	4.09 (t, J=9.73)	3.76	3.55	3.88	2.07 (3H)
Gal (2)	4.51 (d, J=7.66)	3.49	3.63	3.89	n/a	n/a	-
Fuc	5.03 (d, J=4.01)	3.81	3.96	3.90	4.88	(1.18, 3H)	-

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.74
Glc(β)	95.83
Gal	102.90
GlcNAc	102.63
Gal (2)	102.76
Fuc	97.92

ESI TOF-MS m/z calculated for $C_{32}H_{55}NO_{25}Na$ $[M + Na]^+$ = 876.2955, found 876.2281

Synthesis of H type 2 lacto-*N*-neopentose/LNnFPI (12)



12

LNnT (10 mg, 14.1 µmol) was dissolved in 50 mM TRIS (pH 7.5) to a concentration of 10 mM, and GDP-Fuc (1.5 eq, 15 mM), $MnCl_2$ (10 mM, from 1 M solution at 1% v/v of reaction volume), CIAP (1% v/v) and FUT1 (10% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. The products were purified using size exclusion chromatography and preparative HPLC as described before to obtain **12** (7 mg, 58%, 8.2 µmol).

^1H (600 MHz, D_2O): δ (ppm)

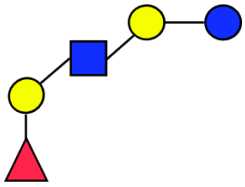
	H1	H2	H3	H4	H5	H6	Ac
Glc(α)	5.23 (d, J=3.77)	3.58	3.81	3.65	n/a	n/a	-
Glc(β)	4.67 (d, J=7.97)	3.28 (app. t)	3.64	3.96	n/a	n/a	-
Gal	4.45 (d, J=7.74)	3.60	3.71	4.15 (d, J=3.34)	n/a	n/a	-
GlcNAc	4.71 (d, J=8.39)	3.82	3.47	3.97	n/a	n/a	2.06 (3H)
Gal (2)	4.55 (d, J=7.79)	3.68	3.69	3.88	n/a	n/a	-
Fuc	5.31 (d, J=3.06)	3.80	3.96	3.83	4.22 (q, J=6.65)	(1.23, 3H)	-

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.73
Glc(β)	95.73
Gal	102.88
GlcNAc	102.75
Gal (2)	100.23
Fuc	99.36

ESI TOF-MS m/z calculated for $C_{32}H_{55}NO_{25}Na$ $[M + Na]^+$ = 876.2955, found 876.2881

Synthesis of H type 1 lacto-*N*-pentose/LNFPI (13)



13

LNT (7.5 mg, 10.6 µmol) was dissolved in 50 mM TRIS (pH 7.5) to a concentration of 10 mM, and GDP-Fuc (1.5 eq, 15 mM), $MnCl_2$ (10 mM, from 1 M solution at 1% v/v of reaction volume), CIAP (1% v/v) and FUT1 (10% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. The products were purified using size exclusion chromatography and preparative HPLC as described before to obtain **13** (8 mg, 89%, 9.4 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

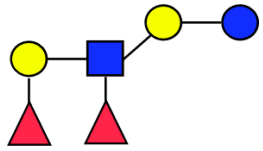
	H1	H2	H3	H4	H5	H6	Ac
Glc(α)	5.22 (d, J=3.80)	3.58	3.83	n/a	n/a	n/a	-
Glc(β)	4.67 (d, J=7.74)	3.28	3.63	3.95	n/a	n/a	-
Gal	4.43 (d, J=7.89)	3.57	3.72	4.15 (d, J=3.32)	n/a	n/a	-
GlcNAc	4.63 (d, J=8.41)	3.80	3.51	3.55	n/a	n/a	2.06 (3H)
Gal (2)	4.65 (d, J=7.84)	3.60	3.84	3.89	n/a	n/a	-
Fuc	5.19 (d, J=4.11)	3.77	3.67	3.74	4.30 (q, J=6.7)	1.24(CH_3)	-

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.76
Glc(β)	95.68
Gal	102.95
GlcNAc	103.23
Gal (2)	100.19
Fuc	99.45

ESI TOF-MS m/z calculated for $C_{32}H_{55}NO_{25}Na$ $[M + Na]^+$ = 876.2955, found 876.2880

Synthesis of Lewis^y lacto-*N*-neohexose/LNnDFHI (14)



14

12 (3 mg, 3.5 µmol) was dissolved in 100 mM TRIS (pH 7.5) to a concentration of 10 mM, and GDP-Fuc (0.5 eq, 5 mM), $MnCl_2$ (10 mM, from 1 M solution at 1% v/v of reaction volume), CIAP (1% v/v) and FUT9 (10% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. 0.1 eq. of GDP-Fucose was added each hour until the reaction went to completion, after which the mixture was freeze dried. Size exclusion chromatography and preparative HPLC were used to obtain **14** (2.3 mg 66%, 2.3 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

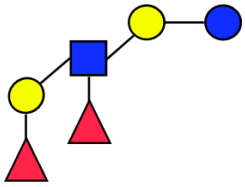
	H1	H2	H3	H4	H5	H6	Ac
Glc(α)	5.22 (d, J=3.76)	3.58	3.85	3.65	n/a	n/a	-
Glc(β)	4.67 (d, J=8.02)	3.28	3.65	3.61	n/a	n/a	-
Gal	4.45 (d, J=7.73)	3.60	3.70	4.15 (d, J=3.30)	n/a	n/a	-
GlcNAc	4.72 (d, J=8.55	3.96	3.46	3.87	n/a	n/a	2.03 (3H)
Gal (2)	4.52 (d, J=7.82)	3.66	3.86	3.86	n/a	n/a	-
Fuc glcnac	5.12 (d, J=3.99)	3.70	3.93	3.81	4.89 (q, J=6.84/7.18)	(1.24, 3H)	-
Fuc(2) gal	5.28 (d, J=3.45)	3.80	3.77	3.83	4.62 (q, J=6.68/6.75)	(1.27, 3H)	-

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.75
Glc(β)	95.72
Gal	102.90
GlcNAc	102.49
Gal (2)	100.18
Fuc	98.56
Fuc(2)	99.37

ESI TOF-MS m/z calculated for $C_{38}H_{65}NO_{29}Na$ $[M + Na]^+$ = 1022.3534, found 1022.3450

Synthesis of Lewis^b lacto-*N*-hexose/LNDFHI (15)



15

13 (1.5 mg, 1.8 µmol) was dissolved in 100 mM TRIS (pH 7.5) to a concentration of 10 mM, and GDP-Fuc (0.5 eq, 5 mM), $MnCl_2$ (10 mM, from 1 M solution at 1% v/v of reaction volume), CIAP (1% v/v) and FUT3 (10% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction progress was monitored with LC-MS. 0.1 eq. of GDP-Fucose was added each hour until the reaction went to completion, after which the mixture was freeze dried. Size exclusion chromatography and preparative HPLC were used to obtain **15** (0.8 mg 44%, 0.8 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

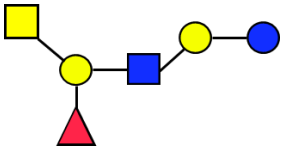
	H1	H2	H3	H4	H5	H6	Ac
Glc(α)	5.23 (d, J=3.75)	3.56	3.82	n/a	n/a	n/a	-
Glc(β)	4.67 (2H, m)	3.26 (app. t, J=8.46)	3.65	3.95	n/a	n/a	-
Gal	4.42 (d, J=7.90)	3.57	3.73	4.15 (m, 2H)	n/a	n/a	-
GlcNAc	4.66 (2H, m)	3.60	n/a	n/a	n/a	n/a	2.07 (3H)
Gal (2)	4.61 (d, J=8.16)	3.85	3.74	4.15 (m, 2H)	n/a	n/a	-
Fuc	5.03 (d, J=3.94)	3.81	3.93	3.83	4.87	1.26 (3H, d, J=6.52)	-
Fuc(2)	5.16 (d, J=4.06)	3.75	3.68	3.77	4.35 (q, J=6.65)	1.28 (3H, d, J=6.77)	-

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.67
Glc(β)	95.72
Gal	102.95
GlcNAc	100.61
Gal (2)	103.27
Fuc	97.78
Fuc(2)	99.49

ESI TOF-MS m/z calculated for $C_{38}H_{65}NO_{29}Na$ $[M + Na]^+$ = 1022.3534, found 1022.3456

Synthesis of A Type 2 lacto-*N*-neohexose (16)



16

12 (0.4 mg, 0.5 µmol) was dissolved in 100 mM TRIS (pH 7) to a concentration of 10 mM, and UDP-GalNAc (1.5 eq, 15 mM), CIAP (1% v/v) and BoGT6A (5% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction process was monitored with LC-MS. Size exclusion chromatography and preparative HPLC were used to obtain **16** (0.2 mg , 40% , 0.2 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

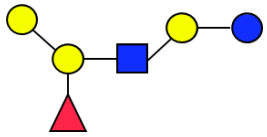
	H1	H2	H3	H4	H5	H6	Ac
Glc(α)	5.23 (d, J=3.80)	3.58	3.83	3.66	n/a	n/a	-
Glc(β)	4.67 (d, J=7.98)	3.29 (app. t, J=8.50)	3.65	3.81	n/a	n/a	-
Gal	4.45 (d, J=7.81)	3.60	3.68	4.15 (d, J=3.31)	n/a	n/a	-
GlcNAc	4.71 (d, J=8.32)	3.80	3.45	3.68	n/a	n/a	2.05 (6H)
Gal (2)	4.62 (d, J=7.67)	3.91	3.76	4.22	n/a	n/a	-
Fuc	5.36 (d, J=4.13)	3.79	3.72	3.83	4.32 (q, J=6.71)	1.25 (d, J=6.63)	-
GalNAc	5.18 (d, J=3.87)	4.24	3.90	4.00	n/a	n/a	2.05 (6H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.95
Glc(β)	95.74
Gal	102.96
GlcNAc	102.68
Gal (2)	99.99
Fuc	98.61
GalNAc	91.22

ESI TOF-MS m/z calculated for $C_{40}H_{68}N_2O_{30}Na$ $[M + Na]^+$ = 1079.3749, found 1079.3710

Synthesis of B Type 2 lacto-*N*-neohexose (17)



17

12 (0.7 mg, 0.8 µmol) was dissolved in 100 mM TRIS (pH 7) to a concentration of 10 mM, and UDP-GalNAc (1.5 eq, 15 mM), CIAP (1% v/v) and GTB (50% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction process was monitored with LC-MS. Size exclusion chromatography and preparative HPLC were used to obtain **17** (0.2 mg , 23% , 0.2 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

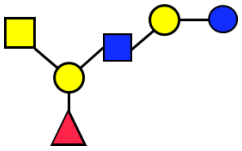
	H1	H2	H3	H4	H5	H6	Ac
Glc(α)	5.23 (d, J=3.93)	3.59 (app. t, J=8.47)	3.83	3.66	n/a	n/a	-
Glc(β)	4.67 (d, J=8.02)	3.29	3.65	3.81	n/a	n/a	-
Gal	4.45 (d, J=7.86)	3.61	3.72	4.15 (d, J=3.31)	n/a	n/a	-
GlcNAc	4.71 (d, J=7.81)	3.82	3.45	3.70	3.98	n/a	2.05 (3H)
Gal (2)	4.64 (d, J=7.65)	3.93	4.30	4.01	n/a	n/a	-
Fuc	5.35 (d, J=4.12)	3.79	3.73	3.82	4.31	1.24 (d, J=6.64)	-
Gal (3)	5.26 (app. s)	3.89	3.99	4.20 (t, J=6.20)	3.75	n/a	-

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.87
Glc(β)	95.64
Gal	102.81
GlcNAc	100.10
Gal (2)	102.75
Fuc	98.67
Gal (3)	92.92

ESI TOF-MS m/z calculated for $C_{38}H_{65}NO_{30}Na$ $[M + Na]^+$ = 1038.3484, found 1038.3449

Synthesis of A Type 1 lacto-*N*-hexose (18)



18

13 (0.25 mg, 0.3 µmol) was dissolved in 100 mM TRIS (pH 7) to a concentration of 10 mM, and UDP-GalNAc (1.5 eq, 15 mM), CIAP (1% v/v) and BoGT6A (5% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction process was monitored with LC-MS. Size exclusion chromatography and preparative HPLC were used to obtain **18** (0.2 mg , 65% , 0.2 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

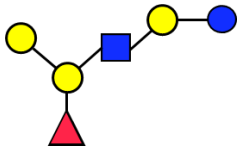
	H1	H2	H3	H4	H5	H6	Ac
Glc(α)	5.23 (d, J=3.77)	3.58	3.85	3.64	n/a	n/a	-
Glc(β)	4.67 (d, J=8.00)	3.28 (app. t)	3.65	3.96	n/a	n/a	-
Gal	4.43 (d, J=7.89)	3.57	3.72	4.15 (d, J=3.33)	n/a	n/a	-
GlcNAc	4.63 (d, J=8.46)	3.83	4.02	3.54	n/a	n/a	2.06 (6H)
Gal (2)	4.71 (d, J=7.67)	3.82	3.95	4.24	n/a	n/a	-
Fuc	5.26 (d, J=4.25)	3.76 (2H)	3.63	3.76 (2H)	4.36	1.25 (3H) (d, J=6.54)	-
GalNAc	5.19 (d, J=3.83)	4.23	3.97 (2H)	3.97 (2H)	n/a	n/a	2.06 (6H)

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	91.91
Glc(β)	95.61
Gal	103.00
GlcNAc	103.22
Gal (2)	100.09
Fuc	99.01
GalNAc	91.10

ESI TOF-MS m/z calculated for $C_{40}H_{68}N_2O_{30}Na$ $[M + Na]^+$ = 1079.3749, found 1079.3704

Synthesis of B Type 1 lacto-*N*-hexose (19)



19

13 (0.25 mg, 0.3 µmol) was dissolved in 100 mM TRIS (pH 7) to a concentration of 10 mM, and UDP-GalNAc (1.5 eq, 15 mM), CIAP (1% v/v) and GTB (50% v/v) were added, after which the solution was incubated at 37 °C with gentle shaking and the reaction process was monitored with LC-MS. Size exclusion chromatography and preparative HPLC were used to obtain **19** (0.1 mg , 31% , 0.2 µmol).

^{1}H (600 MHz, D_2O): δ (ppm)

	H1	H2	H3	H4	H5	H6	Ac
Glc(α)	5.22 (q, 3H, J=4.21-5.30)	3.60 (app. t, J=8.27)	3.83	n/a	n/a	n/a	-
Glc(β)	4.67 (d, J=8.11)	3.29	3.65	3.81	n/a	n/a	-
Gal	4.43 (d, J=7.91)	3.57	3.73	4.15 (d, J=3.32)	n/a	n/a	-
GlcNAc	4.64 (d, J=8.27)	3.83	4.02 (app. t, J=9.5)	3.53	n/a	n/a	2.07 (3H)
Gal (2)	4.73 (d, J=7.72)	3.84	3.96	4.29	n/a	n/a	-
Fuc	5.23 (q, 3H, J=4.21-5.30)	3.77	3.63	3.75	4.36 (q, J=6.72)	1.25 (3H, d, J=6.61)	-
Gal (3)	5.24 (q, 3H, J=4.21-5.30)	3.89	3.95	4.28	3.74	n/a	-

^{13}C from HSQC (150 MHz, D_2O): δ (ppm)

	C1
Glc(α)	92.09
Glc(β)	95.64
Gal	102.91
GlcNAc	103.31
Gal (2)	100.16
Fuc	99.26
Gal (3)	92.95

ESI TOF-MS m/z calculated for $C_{38}H_{65}NO_{30}Na$ $[M + Na]^+$ = 1038.3484, found 1038.3446

Nederlandse samenvatting

Glycanen zijn een essentieel component van alle levende organismes, maar worden -in tegenstelling tot andere componenten zoals RNA en eiwitten- niet direct gecodeerd in het genoom. De biosynthese van glycanen wordt gereguleerd door directe competitie tussen glycosyltransferases. In de inleiding van deze thesis worden de chemische structuur van glycanen in de mens en de variaties die hierbinnen voorkomen beschreven. Dit hoofdstuk introduceert de verschillende monosachariden die de bouwstenen van glycanen vormen en beschrijft hoe deze monosachariden aan elkaar gekoppeld zijn. Hierna worden de glycanen geclassificeerd op basis van hun verankeringstype en wordt een gedetailleerd overzicht gegeven van de biosynthese van *N*- en *O*-gekoppelde glycanen en de verschillen in de opbouw van de kernstructuren. Ook wordt besproken hoe veelvoorkomende terminale glycaan epitopen, zoals Lewis en bloedgroep epitopen, een belangrijke bron van isomerisme vormen in complexe glycanen. Vervolgens worden glycaan modificaties zoals sulfaten en acetyl esters geïntroduceerd.

In het tweede deel van de inleiding wordt de aandacht verlegd naar de synthetische strategieën voor het maken van glycanen met behulp van glycosyltransferase enzymen. De verschillende mechanismes waarmee enzymatische glycosylering zich kan voltrekken worden besproken en er wordt een overzicht gegeven van alle glycosyltransferases die in de mens betrokken zijn bij de synthese van complexe glycanen. De diversiteit aan transferases draagt bij aan de complexiteit van glycaanbiosynthese.

De synthese van complexe glycanen wordt bemoeilijkt door het feit dat glycanen vaak op meerdere posities glycosylering kunnen ondergaan. Er wordt besproken hoe regioselectieve enzymatische transformaties kunnen worden gebruikt om glycanen te glycolyseren op specifieke posities. Zo kan regioselectiviteit verder worden gecontroleerd tijdens opeenvolgende enzymatische glycosyleringsreacties. Er zijn echter niet genoeg regioselective transformaties om alle glycanen te kunnen synthetiseren. Chemo-enzymatische synthese vormt een oplossing voor dit probleem door enzymatische glycosylering te dirigeren met niet-natuurlijke, chemische modificaties. Meerdere methodes voor het installeren van niet-natuurlijke modificaties worden besproken, waaronder het gebruik van niet-natuurlijke glycaan donoren in enzymatische transformaties.

Het derde deel van de inleiding behandelt analytische technieken voor de structurele karakterisering van glycanen. Er wordt beschreven hoe verschillende NMR-technieken worden toegepast om de glycaanstructuur van gezuiverde synthetische glycanen te bevestigen. Lectine histochemie wordt kort besproken als een methode om glycosylering op weefselniveau te bekijken. Daarna worden methodes om *O*- en *N*-glycanen van glycoproteïnen los te koppelen besproken en wordt er gewezen op de uitdagingen die er momenteel nog bestaan bij het vrijmaken van *O*-glycanen.

Geavanceerde massaspectrometrie technieken voor de analyse van vrijgemaakte glycanen, zoals meertrapse massaspectrometrie en ionmobiliteit-massaspectrometrie, worden besproken. Met deze geavanceerde technieken kan informatie worden verkregen over de exacte glycaanstructuur. De principes achter de scheiding van isomere glycanen met behulp van ionmobiliteit-massaspectrometrie worden uitgelegd, evenals de berekening van botsingsdoorsneden met behulp van de Mason-Schamp vergelijking. De inleiding wordt afgesloten met een overzicht van de uitdagingen bij zowel de synthese als analyse van complexe glycanen ten gevolge van hun hoge mate van isomeriteit.

Divergente, chemo-enzymatische synthese van α1,3 gefucosyleerde polylactosamines

De enzymatische koppeling van α1,3 fucose aan GlcNAc door fucosyltransferases is een van de weinige enzymatische glycosyleringsreacties die bij de mens zowel op de terminale als op de interne LacNAc-residuen kan plaatsvinden. Deze is echter niet selectief op polyLacNAc structuren, wat de regioselectieve synthese van dit type glycanen verhinderd. Hierdoor is het effect van interne α1,3 fucose op de binding van glycaanbindende eiwitten meestal nog onbekend. Dit kan worden verholpen met de ontwikkeling van toegankelijke methodologie voor het maken van synthetische standaarden. Eerder werk toonde aan dat een chemisch gesynthetiseerde lactosamine hexasacharide met orthogonale *N*-beschermingsgroepen gemanipuleerd kon worden met eenvoudige chemische transformaties om selectieve fucosylering mogelijk te maken. In dit hoofdstuk wordt beschreven hoe met een chemo-enzymatische benadering een soortgelijk orthogonaal N-beschermd lactosamine gemaakt kan worden met een sterk gereduceerde synthetische inspanning.

De belangrijkste stap in deze procedure is het enzymatisch installeren van het onnatuurlijke *N*-beschermde monosacharide GlcNTFA. De installatie van GlcNTFA maakt het introduceren van *N*-beschermde lactosamine-disachariden mogelijk in slechts vier synthetische stappen. Met in totaal acht stappen ontstaat hiermee een orthogonaal beschermde hexasacharide die een gemeenschappelijk intermediair vormt voor het produceren van glycanen met alle mogelijke fucosyleringspatronen. Eenvoudige chemie wordt vervolgens gebruikt om de gemeenschappelijke intermediair op geselecteerde plaatsen te (de)activeren voor enzymatische α1,3 fucosylering. Een bibliotheek van glycanen met zowel bloedgroep als Lewis antigenen wordt op deze manier gegenereerd uit één gemeenschappelijke intermediair en geprint als glycaanmicroarray. Deze glycaanmicroarray wordt vervolgens gebruikt om de binding van verschillende galectines aan bloedgroepantigenen te bestuderen. We hebben vastgesteld dat α1,3 fucose de binding van galectines op de glycaanmicroarray kan beïnvloeden en we hebben verschillend bindingsgedrag van galectines als reactie op deze fucosylering geobserveerd.

Chemo-enzymatische synthese van vertakte polylactosamines

Bij volwassen mensen kan een interne glycosyleringsreactie optreden die vertakkingen op polyLacNAc structuren aanbrengt. Deze β1,6 GlcNAcyleringsreactie installeert een vertakkingspunt op een glycaan dat door andere glycosyltransferases kan worden verlengd, resulterend in meertakse glycanen. Deze glycanen kunnen multivalente interacties ondergaan en komen veelvuldig voor in menselijke melkoligosachariden. Momenteel zijn er geen efficiënte methodes bekend om selectief vertakkingen te installeren op polyLacNAc structuren. De *N*-beschermde lactosamines uit het vorige hoofdstuk bleken ook de regioselectiviteit van het enzym GCNT2, dat vertakkingen kan introduceren, te kunnen beïnvloeden. NMR-experimenten tonen aan dat Boc-beschermd lactosamine selectief de vertakkingsreactie kan beïnvloeden. Alleen de aan het reducerende uiteinde aangrenzende galactose wordt geblokkeerd, waardoor regioselectieve glycosylering van de andere interne galactoses mogelijk wordt. α1-3 fucosylering door FUT5 wordt geremd door dezelfde onnatuurlijke modificatie en kan worden gebruikt om de geactiveerde en gedeactiveerde glycosylerings-posities om te draaien. Met behulp van regioselectieve vertakkingsreacties is een bibliotheek van complexe polyLacNAc architecturen gegenereerd die regio-isomeren bevat met verschillende vertakkingspatronen.

De loskoppeling van *O*-glycanen met geneutraliseerde hypochloriet

De compositie van *O*-glycosylatie kan bij veel menselijke ziektes een grote voorspellende, diagnostische en therapeutische waarde hebben. Een kritische stap in de analyse van *O*-glycanen uit biologische monsters is het loskoppelen van de glycanen van glycoconjugaten. In tegenstelling tot N-glycanen zijn er echter geen efficiënte enzymatische methodes om deze los te koppelen. Chemische methodes beschadigen vaak de glycaanstructuur op meerdere manieren, zoals de "peeling" reactie en hydrolyse van pH-gevoelige groepen.

Onlangs is aangetoond dat *O*-glycanen met hypochloriet kunnen worden losgekoppeld. Dit resulteert in de gedeeltelijke vorming van een glycaan gekoppeld aan melkzuur die ongevoelig is voor de "peeling" reactie en tegelijkertijd informatie behoudt over het type aminozuur waar de glycaan aan gekoppeld was. Om de nevenreacties die plaatsvinden tijdens het loskoppelen van *O*-glycanen met hypochloriet te kunnen identificeren, hebben wij de reactie uitgevoerd op zuivere, synthetisch geprepareerde *O*-glycopeptides. Deze glycopeptides bevatten veelvoorkomende glycaanmodificaties zoals fucose, siaalzuur en sulfaat. Het zuivere startmateriaal maakte het mogelijk de voortgang van de reactie in detail te bekijken en versimpelde de karakterisering van bijproducten. Verrassend genoeg kon hypochloriet de *O*-glycanen volledig loskoppelen in slechts 10 minuten, wat veel sneller is dan de eerder gerapporteerde 24 uur. Het hoofdproduct van deze reactie was niet het verwachte glycaan gekoppeld aan melkzuur, maar glycanen met een vrij reducerend uiteinde dat gevoelig is voor de "peeling" reactie. Neutralisatie van hypochloriet resulteerde in een iets tragere reactie, maar genereerde enkel het glycaan gekoppeld aan melkzuur zonder nevenreacties.

Met synthetische glycopeptides kon de stabiliteit van pH-gevoelige glycaanmodificaties worden geëvalueerd tijdens glycaan loskoppeling met geneutralizeerde hypochloriet. Alle synthetische glycopeptiden konden volledig worden losgekoppeld, resulterend in volledige omzetting tot de verwachte product zonder nevenreacties. De experimentele condities die we ontwikkelden met synthetische glycopeptiden konden ook de *O*-glycanen van submaxillaire mucine van een koe loskoppelen. Een groot deel van de waargenomen glycanen in dit sample bevatte nog steeds een of meerdere O-acetylesters op siaalzuren. Deze acetylesters zijn moeilijk te bestuderen met andere methodes om O-glycanen los te koppelen, omdat ze niet stabiel zijn onder de basische condities die doorgaans gebruikt worden. Gesulfateerde O-glycanen die nog niet eerder in submaxillaire mucine waren geïdentificeerd konden gedetecteerd worden. Tijdens de testreacties van geneutraliseerd hypochloriet op zuivere synthetische glycopeptides konden we ook tussenproducten identificeren. Deze observaties vormden de basis voor het vaststellen van het reactiemechanisme voor het loskoppelen van glycanen.

De vaststelling van siaalzuur *O*-acetyleringspatronen en glycosidebindingstype met ionmobiliteit-massaspectrometrie

Siaalzuren kunnen op verschillende posities een of meer *O*-acetylesters bevatten. Deze *O*-acetylesters worden in verband gebracht met diverse ziekten en infecties. Daarnaast kan O-acetylering fungeren als een moleculaire schakelaar en wordt de hydrolysesnelheid van zowel endo- als exogene sialidases aanzienlijk door *O*-acetylesters beïnvloed. *O*-acetylesters zijn gevoelig voor basische pH en kunnen hydrolyseren of migreren naar andere posities onder mild basische condities. Met de huidige methoden om de posities

van *O*-acetylesters op een siaalzuur te bepalen moet het siaalzuur onder zure pH van het glycaan worden losgekoppeld. Door loskoppeling gaat informatie over de oorspronkelijke structuur verloren, waardoor het type siaalzuurverbinding niet meer kan worden achterhaald.

In dit hoofdstuk hebben wij een methode ontwikkeld om zowel de O-acetylesterpositie als het type siaalzuurverbinding te bepalen met behulp van ionmobiliteit-massaspectrometrie. Een bibliotheek van synthetische Neu5Ac- en Neu5Gc-bevattende glycanen met goed gedefinieerde *O*-acetylatiepatronen en siaalzuurverbindingen vormde een belangrijke basis voor de ontwikkeling van deze methode. De ionmobiliteit van karakteristieke glycaanfragmenten kon worden bepaald met ionmobiliteit-massaspectrometrie. Hiermee konden we distinctieve aankomsttijdverdelingen vaststellen waarmee deze isomere glycanen konden worden onderscheiden. De verkregen aankomsttijdverdelingen werden gebruikt om botsingsdoorsneden te bepalen. Deze botsingsdoorsneden zijn moleculaire eigenschappen die onafhankelijk zijn van instrumentvariabelen en dus kunnen worden overgedragen tussen laboratoria en instrumenten. De exacte *O*-acetylpositie en het verbindingstype van het siaalzuur konden zonder verder glycaanstandaarden worden bepaald met behulp van de vastgestelde koppelingsdoorsnedes. Deze methode werd toegepast om de *O*-acetylesterposities en siaalzuurverbinding te bepalen op de *N*-glycanen van meerdere biotherapeutica. De gevonden posities en verbindingen kwamen overeen met de gerapporteerde *O*-acetylposities. Ook konden met deze methode grote verschillen worden gevonden tussen het type siaalzuurverbinding van geacetyleerde en niet-geacetyleerde siaalzuren in de *N*-glycanen uit de bovenste luchtwegen van paarden.

Tot slot hebben we de *O*-acetylesters op de *O*-glycanen van submaxillaire mucine opnieuw bekeken met de ontwikkelde ionmobiliteit-massaspectrometrie methode. De meeste *O*-acetylesters die we aantroffen waren gepositioneerd op de C-7 positie, die bijzonder gevoelig is voor migratie. Deze bevinding bevestigt dat geneutraliseerd hypochloriet de *O*-acetylesterpositie behoudt zonder significante migratie.

CHAPTER 7

Appendices

Publications

1. Sastre Toraño J, Gagarinov IA, **Vos GM**, Broszeit F, Srivastava AD, Palmer M, Langridge JI, Aizpurua-Olaizola O, Somovilla VJ, Boons GJ. Ion-Mobility Spectrometry Can Assign Exact Fucosyl Positions in Glycans and Prevent Misinterpretation of Mass-Spectrometry Data After Gas-Phase Rearrangement. Angew Chem Int Ed Engl. 2019

2. Overeem NJ, Hamming PHE, Grant OC, Di Iorio D, Tieke M, Bertolino MC, Li Z, **Vos G**, de Vries RP, Woods RJ, Tito NB, Boons GPH, van der Vries E, Huskens J. Hierarchical Multivalent Effects Control Influenza Host Specificity. ACS Cent Sci. 2020

3. Meijer FA, Doveston RG, de Vries RMJM, **Vos GM**, Vos AAA, Leysen S, Scheepstra M, Ottmann C, Milroy LG, Brunsveld L. Ligand-Based Design of Allosteric Retinoic Acid Receptor-Related Orphan Receptor γt (RORγt) Inverse Agonists. J Med Chem. 2020

4. Xu Y, Hernández-Rocamora VM, Lorent J, Cox R, Wang X, Bao X, Stel M, **Vos GM**, van den Bos R, Pieters RJ, Gray J, Vollmer W, Breukink E. Metabolic labeling of the bacterial peptidoglycan by functionalized glucosamine. *iScience 2022*

5. **Vos GM**, Weber J, Sweet I, Hooijschuur KC, Sastre Toraño J, Boons GJ. Oxidative release of *O*-glycans under neutral conditions for analysis of glycoconjugates having base sensitive substituents. Manuscript submitted

6. **Vos GM**, Hooijschuur KC, Li ZJ, Fjeldsted J, Klein C, de Vries RP, Sastre Toraño J, Boons GJ. Sialic acid *O*-Acetylation Patterns and Glycosidic Linkage Type Determination by Ion Mobility-Mass Spectrometry. Manuscript submitted

7. **Vos GM**, Bosman G, Wu Y, Liu L, Wolfert M, Boons GJ. A Stop-and-Go Strategy for the Chemo-enzymatic Synthesis of selectively Fucosylated Polylactosamine Derivatives. Manuscript in preparation

8. Sastre Toraño J, **Vos GM**, Hooijschuur KC, Vogelaar S, Broszeit F, Fjeldsted J, Klein C, Boons GJ. Exact glycan structure identification by high-resolution fingerprinting and *de novo* sequencing via ion mobility-mass spectrometry. Manuscript in preparation

Extra activities

Conference presentation:
Controlling branching in human milk oligosaccharides, Eurocarb XX (Leiden NL 2019)

Poster presentations:
29th Joint glycobiology meeting (Gent BE 2018)
NWO CHAINS (Velthoven NL 2018)
UIPS (Utrecht NL 2018)
Eurocarb XX(Leiden NL 2019)
Science for life (Utrecht NL 2019)

Acknowledgements

This thesis and the work described in it are a product of the ideas and contributions of many individuals. First of all, I would like to express my gratitude to **Geert-Jan Boons**. This work was only possible through the many talents that you have brought together in the CBDD research group. I am especially grateful for the freedom that I have enjoyed during these years to explore many facets of glycoscience.

Next, I would like to show my appreciation to **Rob de Vries**. You were always willing to discuss ideas, broadening my scientific horizon, and you have been a great help in the exploration of promising research ideas. Many of the biological experiments conducted in this thesis originate from the discussions we had.

I am also indebted to **Kevin Hooijschuur**, who collaborated with me for his master thesis. From the beginning, it was evident that you have a talent in analytical chemistry which is reflected by your contributions to this thesis. The high quality of your work and your dedication during your internship formed the basis of multiple publications. I am confident that you will be able to further develop your academic skills during your PhD and hope you will enjoy your time at the Boon's lab.

When I started my PhD, I was fortunate enough to meet **Frederik Broszeit** and **Zeshi ("Jack") Li**, who both helped me get settled in the lab. Frederik, when I joined you had already spent significant effort in setting up the infrastructure to conduct enzymatic reactions, from which I could profit. You have been more than willing to share your experience in enzymatic synthesis with me and we had countless coffee meetings discussing our research problems and developments. Jack, you were always available for a discussion of scientific ideas and your remarkable knowledge and understanding of carbohydrates was always great way to gauge project viability.

About a year or two after me, **Yunfei ("George") Wu** joined our group. George started his work on projects that took concepts developed in this thesis and took them to the next level. The achievements made in his work on the chemo-enzymatic synthesis of keratan sulfates are inspiring and are some of the most complex molecules made by chemo-enzymatic synthesis to-date. George, I have no worries that you will be able to complete this challenging project and wish you all the best with future endeavors.

At around the same time, **Francesco Palmieri** was shipped over from UGA to Utrecht. Francesco was always available for advice and our different views on most aspects of chemistry often led to interesting discussions and ideas. When these discussions could not be resolved during working hours, we would continue them in the evenings, catalyzed by food and wine. I am sure we will continue to share some more wine bottles for science. On a similar note, I would like to show my appreciation to my neighbors: **Cyril Balsollier**, **Xianke Meng** and **Vito Thijssen**. It was great living in the same building, and I very much enjoyed the get-togethers. Together with Cyril we always made sure there was a steady stream of French delicacies from the mountains to enjoy.

I would like to thank **Javier Sastre Toraño** for his help in using, maintaining (and repairing!) mass spectrometry equipment. Your help was especially important for the chapters of this thesis that focus on glycan analysis. Likewise, I owe my gratitude towards **Justyna Dobruchowska** who not only maintained NMR equipment in tip-top shape but was always willing to go the extra mile to get the perfect NMR data. Sometimes, the scale of enzymatic reactions required

NMR experiments that took significant amounts of time, but Justyna was always willing to accommodate this, which sometimes prevented that I needed to redo a complete synthetic route. Many thanks to **Margreet Wolfert** for your support in getting everything out of the glycan microarray experiments. I greatly appreciate all the effort you put into ensuring the data was correct and reproducible and the help that you provided in the analysis of the data.

The glycopeptides prepared by **Julia Weber** formed integral part of developing neutralized hypochlorite O-glycan release. These compounds required significant effort to prepare, so I am especially grateful that you allowed me to destroy them (for scientific purposes). I am grateful to **Igor Sweet** who helped me to finalize the *O*-glycan release chapter. I would also like to express gratitude towards **Roos van der Woude** and **Gerlof Bosman** for the expression of glycosyltransferases, **Lin Liu** for glycan microarray printing, **Anthony Prudden** and **Kun Yuan** for providing HMO samples, **Sander Herfst** and **Pieter Fraaij** for mucin samples, **Enrico Mastrobattista** for biopharmaceuticals, **Nika Nemanichvili** for tissue samples, and **Koen Giesbert** and **Karin Strijbis** for mucin samples and discussion on the purification of mucins.

I greatly enjoyed my time at Utrecht University and was often inspired by the surrounding research and discussion for which I must thank the many members of the CBDD group: **Ana Gimeno Cardells**, **Anna Ehlers**, **Arwin Brouwer**, **Apoorva Strivastava**, **Balthasar Heesters**, **Barbara Steigenberger**, **Bernd Stahl**, **Christina Bueno Diaz**, **Cindy Spruit**, **Dania Martinez-Alarcon**, **Dirk Rijkers**, **Dushen Chen**, **Ed Moret**, **Elif Uslu**, **Enrico Verpalen**, **Erianne Alvarado Melendez**, **Liangwei ("Frank") Zhang**, **Hanna de Jong**, **Helena Ehren**, **Ilhan Tomris**, **Inez Stoof, Ingrid 't Hart**, **Ivan Gagarinov**, **Jan-Willem Langenbach**, **Jelle Fok**, **John Kruijtzer**, **Jun Ong**, **Kim Bouwman**, **Krishna Desai**, **Laura Tadè**, **Lifeng Sun**, **Lemeng Chao**, **Linda Silvertand**, **Lisa Brandenburg**, **Luca Unione**, **Luuk Stel**, **Margherita Duca**, **Maria Moure Garcia**, **Maria Carrasco**, **Mehman Bunyatov**, **Miha Sovrovic**, **Minglong Liu**, **Niels Ponsé**, **Nino Trattnig**, **Nishant Sewgobind**, **Nives Hribernik**, **Nuria Martinez-Saez**, **Oier Aizpurua**, **Patrycja Lenartowicz**, **Pieter de Saint Aulaire**, **Pouya Zaree**, **Pradeep Chopra**, **Roland Pieters**, **Rosanne van Beek**, **Ryoji Yoshisada**, **Seino Jongkees, Shannon Vogelaar**, **Tim Hogervorst**, **Tim Leenders**, **Tom Wennekes**, **Torben Heise**, **Victor Somovilla**, **Xiufen Liu**, **Xuan Wei**, **Yanyan Liu**, **Yujie Ma**, **Yvette Luijkx** and **Zhiyong Zhang**.

In the following sections, I would like to thank friends and family for their support. Very important to me are my long-lasting friendships with **Ruben van den Bosch**, **Luc ter Horst**, **Bart ("Bert") Spanjersberg,** and **Laurens Steehouder**. During all these years you have made significant and continuous contributions to this body of work by providing a solid framework of friendship and support. I always enjoyed the many parties, weekends away, festivals, and hanging out in general. Also, I would like to thank **Hugo ("Henk") Beks** and **Joshua Nierop** for their continuing friendship and the many amazing people that I have met thanks to you. **Thijmen Sijnesael**, I will always cherish the time we spent the entire concert discussing science and missed the complete show. I was happy to hear that you decided to pursue a PhD at UMC and wish you all the best in this endeavor.

Important in my life was the help and encouragements of my parents, **Paul** and **Marie-Anne**, and my sister **Daan ("Noeknoek")**. Your unconditional support has allowed me to pursue my academic interests for which I am ever grateful. Last, but most important during this time has been the love and support I have received from my partner **Iris Geigenmüller**. I greatly cherish the life we share together and hope we have many adventures to look forward to.