



**Synthesis, screening and use of  
parasitic O-glycans and mimetics for  
improved C-type lectin receptor  
targeting**

**Julie Thanh PHAM**

2018





**Synthesis, screening and use of  
parasitic O-glycans and mimetics for  
improved C-type lectin receptor  
targeting**

**Julie Thanh PHAM**

Glycotechnology Laboratory, CIC BiomaGUNE

2018



The research was carried out within the international training network Immunoshape in the Glycotechnology Laboratory at the Centre for Cooperative Research in Biomaterials, CIC BiomaGUNE, San Sebastián, País Vasco, Spain

This project received funding from the European Union's Horizon 2020 research and innovation program under the Marie Curie-Sklodowska grant agreement No 642870



## RESUMEN DE TESIS

La superficie de cada célula está cubierta por una densa capa de glicanos que están implicados en varios eventos fisiológicos como la interacción huésped-patógeno, la diferenciación y el tráfico celular, y la señalización intracelular y extracelular. Los glicanos unidos a proteínas formando glicoproteínas, se pueden dividir en dos categorías mayoritarias: los N- u O-glicanos, según los aminoácidos a los que estén unidos. Tal vez una consecuencia de la heterogeneidad de la biosíntesis de los O-glycanos, sea la causa del escaso conocimiento de su función biológica. Los más investigados, los O-glicanos  $\alpha$ -GalNAc o los de tipo mucina están localizados en las proteínas mucina. Estas glicoproteínas se encuentran en las mucosas de las vías respiratorias, los tractos urogenitales y gastrointestinales donde pueden estar unidos a la membrana o ser secretados. Sirven como agente de adhesión y barrera física contra patógenos externos, protegiendo las superficies epiteliales donde son producidos. De los ocho subtipos de estructuras que constituyen las bases de los O-glicanos de mucinas, las estructuras base 1-4 son las más comunes y se encuentran en una gran variedad de glicoproteínas y mucinas. Según el organismo, el tejido y el estado de desarrollo, estas estructuras base son elongadas y modificadas con GlcNAc, Gal, Fuc, y a menudo están sialiladas e incluso a veces sulfatadas. La elongación más característica es la repetición de la unidad  $\text{GlcNAc}\beta 1\text{-3Gal}\beta 1\text{-4}$  (poliLacNAc, poliLN), que también puede estar fucosilada. Estos glicanos tipo-mucina reciben mucha atención debido a su presencia en cáncer y enfermedades inflamatorias y autoinmunes. En particular, el antígeno Tn su versión sialilada (antígeno STn) y la estructura base 1 están considerados como sellos distintivos de cáncer, dado que se ha demostrado repetidamente que están sobre expresados en carcinomas de varios órganos (como el pulmón, el colon, el estómago y el páncreas).

Cada vez hay más evidencias que demuestran que los carbohidratos en las superficies de las células como los O-glicanos pueden influir en el sistema inmune a través de la interacción con proteínas que se unen a glicanos (GBPs) tales como las lectinas y los receptores tipo Toll. Estos receptores son los responsables de desencadenar una serie de eventos de señalización cuyo efecto final es marcar la respuesta inmune. Un

elemento crucial del sistema inmune son los receptores de lectinas tipo-C (CLRs). Estas GBPs se unen a glicanos de una forma altamente específica dependiente del calcio (aunque no siempre). Actúan como receptores de reconocimiento de patrones (PPRs) y constituyen la primera línea de defensa contra la invasión de patógenos, incluyendo virus (por ejemplo el VIH y el Ebola), bacterias y levaduras (por ejemplo el *Aspergillus fumigatus* y la *Candida albicans*). A través de la interacción con patrones moleculares asociados a patógenos (PAMPs) específicos que están en la superficie de los patógenos, los CLRs median la respuesta inmune innata, asegurando la captura del patógeno, su neutralización y su destrucción. Además, los CLRs presentes sobre las células presentadoras de antígenos (APCs) tales como los macrófagos macrophages y las células dendríticas (DCs) proporcionan un puente con el sistema inmune adaptativo, habilitando en la superficie de la APC la presentación de los fragmentos antigénicos. Esto produce la polarización de los linfocitos T (los linfocitos T colaboradores [Th1, Th2, and Th17], los linfocitos T reguladores y los linfocitos T citotóxicos) y la producción de citoquinas a través una serie de eventos de señalización. Según el tipo de linfocito T activado y las citoquinas resultantes, se consigue la estimulación o la supresión del sistema inmune. Es obvio que los CLRs tienen un papel central en la modulación del sistema inmune y por tanto se ha demostrado que están implicados en patologías que involucran al sistema inmune como el cáncer, las alergias y las enfermedades inflamatorias y autoinmunes. En consecuencia, recientemente se han concentrado esfuerzos en terapias basadas en glicanos al presentar ventajas como una inmunogenicidad reducida, la oportunidad de enlazar con varios CLRs a la vez, y la sencilla modificación de su farmacocinética.

Una pista prometedora la ofrece el helminto *Schistosoma mansoni* cuyo glicoma se distingue por su rico conjunto de ligandos ajenos a los vertebrados tales como la xilosa, la alta manosa y las estructuras poli-fucosiladas. Se ha demostrado que estos ligandos interactúan con CLRs para desencadenar la respuesta inmune tras una infección, lo que asegura la co-supervivencia del helminto y del huésped. Sin embargo, el mecanismo exacto de cómo lo consigue el *S.mansoni* sigue siendo desconocido. Se ha demostrado que los O-glicanos de *S.mansoni* contienen patrones antigénicos de la



misma forma que muchas terapias nuevas potenciales basadas en glicanos para tratar enfermedades que involucran al sistema inmune. Aun así, la mayor parte de los estudios sólo se centran en los N-glicanos o en los patrones antigénicos del helminto y muy pocos describen la relevancia de los O-glicanos y de sus interacciones con CLRS.

En un intento de resolver una de las mayores particularidades del mecanismo por el que el *S.mansoni* evade el sistema inmune del huésped, nos planteamos examinar el papel biofuncional de los O-glicanos del helminto, en particular con respecto a su unión con CLRs. En concreto, nuestro objetivo es obtener una colección de O-glicanos y miméticos basados en las dos bases estructurales predominantes de O-glicanos observados en las fases infecciosas de la esquistosomiasis causada por el *S.mansoni*, la base estructural de mucina 2 y el la base específica de *S.mansoni*. Con este fin desarrollamos la síntesis química de la base estructural de mucina 2 y de la base específica de *S.mansoni*. Mientras que la síntesis de la primera ya había sido descrita, la síntesis de la base estructural específica de *S.mansoni* era novedadosa y, sobretodo, de manera crucial necesitaba optimizar la última glicosilacin. Así, el empleo de una galactosa protegida con grupos benzoilos mejoró el rendimiento de la reacción del 40 al 65% al reducir la indeseada migración de grupos acilo previamente observada. Ambas bases se obtuvieron como glucósidos aminopropilos y se desprotegieron parcialmente para dar lugar a sustratos aptos para modificaciones enzimáticas, a la que vez que mantenían los grupos cromóforos para poder purificarlos más fácilmente por HPLC-UV.

Hoy en día, el empleo complementario de glicosiltransferasas en la síntesis de carbohidratos (mediante síntesis químico-enzimática) se realiza de manera rutinaria ya que estas enzimas presentan importantes ventajas como la alta regio- y estéreo-especificidad además de ser fáciles de usar. Para las elongaciones enzimáticas, propusimos la optimización de la enzima bacteriana LgtA previamente descrita, para de una manera asequible instalar las fracciones  $\beta$ -1,3 GlcNAc presentes en nuestras estructuras objetivo. Mediante el cuidadoso rediseño del vector de ADN de la enzima, se consiguió obtener LgtA\_X de manera fácil de expresar, purificar y en grandes

cantidades. La enzima obtenida era activa y se usó en conjunto con GalT1 y el mutante C342T&Y289L para la construcción de los epítomos LN y LDN deseados.

Estudios de actividad enzimática empleando reacciones en solución y en una superficie de microrray, también han determinado una nueva especificidad  $\beta$ -1,6 de LgtA\_X sobre las bases estructurales de los O-glicanos. Como resultado, se consiguieron químicamente siete estructuras elongadas en la rama  $\beta$ -1,6, es decir, un total de nueve O-glicanos. Para remediar el carácter asimétrico de la colección, se investigó la instalación química de un disacárido sintético en la rama  $\beta$ -1,3 de las bases estructurales de los O-glicanos como sitio apto para elongaciones enzimáticas en esta rama. Estas investigaciones revelaron la importancia del tiol elegido como grupo anomérico para evitar una reacción de transferencia desfavorable. A pesar de no haber sido concluida, el resto de la síntesis se prevé sin problemas y los pasos optimizados representan un avance considerable en la síntesis de O-glicanos a escala preparativa.

Debido a su versatilidad, los microarrays de glicanos emergieron como una herramienta indispensable para los estudios de interacciones entre glicanos y GBP y para el estudio de las uniones a CLR. Antes de examinar estas interacciones, los O-glicanos obtenidos necesitaron desproteger por hidrogenación los grupos protectores que quedaban. Este paso no fue tan fácil como se esperaba pero después de optimizarse adecuadamente, se consiguieron un total de 8 O-glicanos, 3 de tipo mucina y 5 de tipo *S.mansoni*. Se evaluó por primera vez la interacción de estos glicanos con 3 CLRS: DC-SIGN, DC-SIGN R y MGL. En general, se observó que los O-glicanos mostraban las especificidades por las lectinas previamente referenciadas. Sin embargo, una diferencia interesante de especificidad fue observada entre el CRD y el ECD de DC-SIGN R.

Cada vez se han documentado más modificaciones exitosas de glicanos directamente en la superficie de microarrays. En particular, transformaciones enzimáticas realizadas *on-chip* han permitido la construcción rápida con un elevado rendimiento de colecciones de glicanos en pequeñas cantidades. Basado en trabajos previos, se llevó a

cabo la fucosilación de los O-glicanos *on-chip* empleando la enzima bacteriana HP-FucT de *H.pylori*. Después de haber establecido las condiciones óptimas de reacción, la enzima se utilizó en conjunto con el donador natural GDP-fucosa o el donador C-6 GDP-azido-fucosa para dar lugar a un array fucosilado o azido-fucosilado con buenos rendimientos. Se observó una mejora de las interacciones glicanos-CLR para ambos arrays con DC-SIGN, en concordancia con la especificidad previamente descrita para ese CLR. En el caso de MGL, la fucosilación no cambió el perfil general de unión de la lectina, sugiriendo que los ligandos que contienen fucosa son tolerados por la lectina hasta un cierto punto.

Sin embargo, se obtuvo una calidad baja de fluorescencia para los arrays fucosilados y azido-fucosilados debido a un lavado excesivo de las placas después de las repetidas exposiciones a reacciones enzimáticas. Esto afectó a la cuantificación del efecto de las elongaciones enzimáticas sobre la especificidad de CRLs y por tanto comprometió los estudios de glicomiméticos generados por química "click".

En este trabajo, hemos desarrollado una metodología para el desarrollo de una librería de O-glicanos basados en el glicoma de helminto. Esto ha facilitado el estudio de la biofuncionalidad de los O-glicanos de *S.mansoni* ofreciendo un nuevo panorama químico a explorar en la búsqueda de compuestos inmunomoduladores para el desarrollo de terapias basadas en glicanos para tratar enfermedades en las que el sistema inmune se ve comprometido.



## **Acknowledgements**

I would like to thank my supervisor Dr. Niels Christian Reichardt for the opportunity to work in the field of glycoscience and the platform managers at the CIC biomaGUNE for their help.

I would also like to express my gratitude to Prof. Marcelo Guerin, Dr David Albesa-Jove and Alberto Marina for their time and mentoring during the work on the crystal structure of LgtA\_X.

I also thank Prof. Frank Fieschi and Dr. Michel Thépaut at the Institut de Biologie Structurale (Grenoble, France) for their time and mentoring on the recombinant expression of CLRs.

Thank you also to the previous and current members of the Glycotechnology Lab, for their support, guidance and their companionship. To Cristina Diez who truly was the best project manager and to Nerea Ruiz who, despite my frustrations at sugars, shrewdly supplied me with sweets in my most despairing times. Thank you also to Anna Cioce for her precious time and help spent on the microarray part of this thesis and for her friendship. I am also very grateful to Dr Alvaro Hernandez, whose invaluable tutelage enabled my development as biochemist.

Merci à Florian qui a été là pendant tous les hauts et surtout les bas de cette thèse. Finally, my unending thanks go to my Phamily for their continuous faith in me, moral support and prayers.



## Abbreviations

Ac	Acetyl
Ac <sub>2</sub> O	Acetic anhydride
ACN	acetonitrile
AcOH	Acetic acid
APC	Antigen presenting cell
aq	aqueous
Ar	Aryl
Bn	Benzyl
Boc	tert-Butoxycarbonyl
BSA	Bovine serum albumin
Bz	Benzoyl
cat	catalytic
Cbz	Benzyloxycarbonyl
CHCl <sub>3</sub>	Chloroform
CLR	C-type lectin receptor
CRD	Carbohydrate recognition domain
CSA	Camphorsulphonic
CTLD	C-type lectin domain
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DC	Dendritic cell
DCM	Dichloromethane
DC-SIGN	Dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
DC-SIGN R	
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMPU	1,3-Dimethyl-3,4,5,6-tetrahydro-2(1H)-pyrimidone
DMSO	dimethylsulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ECD	Extracellular domain
EDTA	Ethylenediaminetetraacetic
ELISA	Enzyme-linked immunosorbent assay
eq	equivalent
ER	Endoplasmic reticulum
ESI	Electrospray ionization
Et	Ethyl
Et <sub>2</sub> O	Diethyl ether
Et <sub>3</sub> N	Triethylamine
EtOAc	Ethyl acetate
FCC	Flash column chromatography
Fmoc	9-Fluorenylmethoxycarbonyl
FPLC	Flash performance liquid chromatography
Fuc	Fucose
FucT	Fucosyltransferase
FucZ	6-Azido-L-fucose

GAG	Glycosaminoglycan
Gal	Galactose
GalNAc	N-acetylgalactosamine
GalNAcT	N-Acetylgalactosaminyl transferase
GalT	Galactosyltransferase
GBP	Glycan binding protein
GDP	Guanosine diphosphate
Glc	Glucose
GlcNAc	N-Acetylglucosamine
GlcNAcT	N-Acetylglucosaminyl transferase
GT	Glycosyltransferase
His	hexahistidine
HR	High resolution
HSQC	Heteronuclear single quantum coherence
Hz	Hertz
IL	Interleukin
Im	Imidazole
IMAC	Immobilized Metal ion affinity chromatography
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
ITO	Indium tin oxide
Lac	Lactose, Gal $\beta$ 1-4Glc
LB	Luria-Bertani broth
LC-MS	Liquid chromatography mass spectrometry
LDN	LacdiNAc, GalNAc $\beta$ 1-4GlcNAc
LDNF	GalNAc $\beta$ 1-4(Fuca $\alpha$ 1-3)GlcNAc
LeA	Lewis A, Gal $\beta$ 1-3(Fuca $\alpha$ 1-4)GlcNAc
LeB	Lewis B, (Fuca $\alpha$ 1-2)Gal $\beta$ 1-3(Fuca $\alpha$ 1-4)GlcNAc
Lev	Levulinoyl
LeX	Lewis X, Gal $\beta$ 1-4(Fuca $\alpha$ 1-3)GlcNAc
LeY	Lewis Y, (Fuca $\alpha$ 1-2)Gal $\beta$ 1-4(Fuca $\alpha$ 1-3)GlcNAc
LG	Leaving group
LiAlH <sub>4</sub>	Lithium aluminium hydride
LiOH	Lithium hydroxide
LN	LacNAc, Gal $\beta$ 1-4GlcNAc
LOS	Lipooligosaccharide
LPS	Lipopolysaccharide
M	Molar
<i>m/z</i>	Mass to charge ratio
MALDI TOF-	
MS	Matrix assisted laser desorption/ionization time-of-flight mass spectrometry
Man	Mannose
Me	Methyl
MeOH	Methanol
MGL	Macrophage Galactose lectin
MHC	Major histocompatibility complex
MS	Molecular sieves



MW	molecular weight
NaOMe	Sodium methoxide
NGP	Neighbouring group participation
NHS	N-hydroxysuccinimide
NIS	N-Iodosuccinimide
NMR	Nuclear magnetic resonance
n-Pr	n-Propyl
o-,m-,p- Tol	ortho-, metha-, para- Toly
PAMP	Pathogen associated molecular pattern
PBS	Phosphate buffer saline
PG	Protecting group
Ph	Phenyl
PhCH(OMe) <sub>2</sub>	Benzaldehyde dimethyl acetal
Phth	Phthalamide
Piv, Piv	Pivaloyl
PMB	p-Methoxybenzyl
PMP	p-Methoxyphenyl
PNGase	peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase
pNP	para-nitrophenyl
PPR	Pattern recognition receptor
Pr	Propyl
p-TsOH	p-Toluenesulphonic acid
Py	Pyridine
RFU	Relative Fluroescence intensity
RT	Room temperature
SAR	Structure activity relationship
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TBAF	Tetrabutylammonium
TBDMS	tert-Butyldimethylsilyl fluoride
TBS	tert-Butyldimethylsilyl
TCA	Trichloroacetimidate
TDS	Dimethylthexylsilyl
TES	Triethylsilyl
TFA	Trifluoroacetic acid
Th	T helper cell
THF	Tetrahydrofurane
TLC	Thin layer chromatography
TMS	Trimethylsilyl
TMSOTf	Trimethylsilyl trifluoromethanesulfonate
Tn	GalNAc $\alpha$ -Ser/Thr
Treg	Regulatory T cell
Troc	2,2,2-Trichloroethoxycarbonyl
TRX	thioredoxin
UDP	Uridine diphosphate
UPLC	Ultra Performance Liquid Chromatography
UV	Ultraviolet



## Table of contents

1. Introduction.....	1
1.1    Biological importance of O-glycans.....	1
1.1.1    O-glycans.....	2
1.1.2    Biosynthesis of mucin-type O-glycans.....	3
1.1.3    Mucin O-glycan analysis.....	4
1.2    C-type Lectin receptors.....	5
1.2.1    Structural characteristics of CLRs.....	5
1.2.2    CLRs in the immune system.....	6
1.3    Parasitic glycans.....	8
1.3.1 <i>Schistosoma mansoni</i> glycans.....	8
1.3.2    Glycan mimickry for immune evasion.....	11
1.4    Synthesis of O-glycans.....	13
1.4.1    Carbohydrate chemistry.....	13
1.4.2    Mucin O-glycans synthesis.....	15
1.4.3    Enzymatic Glycosylations.....	17
1.5    Glycan microarray.....	19
1.5.1    Principles of glycan microarray.....	19
1.5.2    Applications of glycan microarrays.....	21
References.....	23
2. Scope and objectives of thesis.....	29
3. Results and discussion.....	33
3.1    Synthesis of O-glycan cores.....	33
3.1.1    Synthetic strategy.....	33
3.1.2    Synthesis of the GalNAc building block.....	35
3.1.3    Glycosylation to mucin core 1.....	39
3.1.4    Divergent synthesis to mucin2 and <i>S.mansoni</i> cores.....	40
3.1.5    Partial deprotection to enzymatic substrates.....	43
3.2    Recombinant expression of glycosyltransferases.....	50
3.2.1    LgtA: A GlcNAcT from <i>Neisseria meningitidis</i> .....	50
3.2.2    Enzymatic activity and selectivity.....	52
3.2.3    Towards $\beta$ -1,3 elongated O-glycan structures.....	58
3.2.4    Studies directed towards obtaining a crystal structure of LgtA.....	61
3.2.5    Expression of GalT and GalNT.....	63

3.3	Solution-phase enzymatic transformations .....	64
3.3.1	Solubility of O-glycans in enzymatic transformations.....	64
3.3.2	LacNAc and LDN .....	65
3.3.3	Lewis X and LDN-F .....	68
3.3.4	Deprotection of compounds .....	70
3.4	O-glycans and mimetics against CLRs .....	73
3.4.1	On-chip attachments and experiments.....	73
3.4.2	Effect of in-situ fucosylation on CLR binding.....	78
	References.....	93
4	Conclusions and outlook .....	99
5	Experimental .....	103
5.1	General materials and instrumentation.....	104
5.2	Chemical Synthesis.....	105
5.3	Enzymatic transformations .....	118
5.3.1	Expression and purification of LgtA_X and LgtA_H.....	118
5.4	General procedure for elongation using GalT1 or DM GalT-1 .....	120
5.5	General procedure for elongation using LgtA_X.....	120
5.6	General procedure for deprotection of compounds .....	124
5.7	Microarray.....	128
6	Appendix .....	130

# **1 Introduction**



# 1. Introduction

## 1.1 Biological importance of O-glycans

At the surface of all cells is a dense layer of glycans involved in many physiological events such as host-pathogen interaction, cell differentiation and trafficking, and intracellular and intercellular signaling.<sup>1</sup> Glycan structures attached to proteins, forming glycoproteins, not only contribute to the protein structure and but also indirectly act as signalling stimuli for immune receptors (Figure 1).<sup>2</sup> An estimated 50% of mammalian proteins are glycosylated representing over 10 million glycans on the surface membrane of cells. This post-translational modification is cell-type and developmentally specific, and its composition and structural variability is controlled by the availability of substrates and enzymes in subcellular compartment. As a result, the diversity and complexity of glycan structures represent a daunting enigma as to their exact physiological role.<sup>3</sup> Nonetheless, they promise exciting leads into the elucidation of glycan-immune system interactions and the development of glycan-based therapeutics for the treatment of cancer, inflammatory and autoimmune diseases.<sup>4,5</sup>

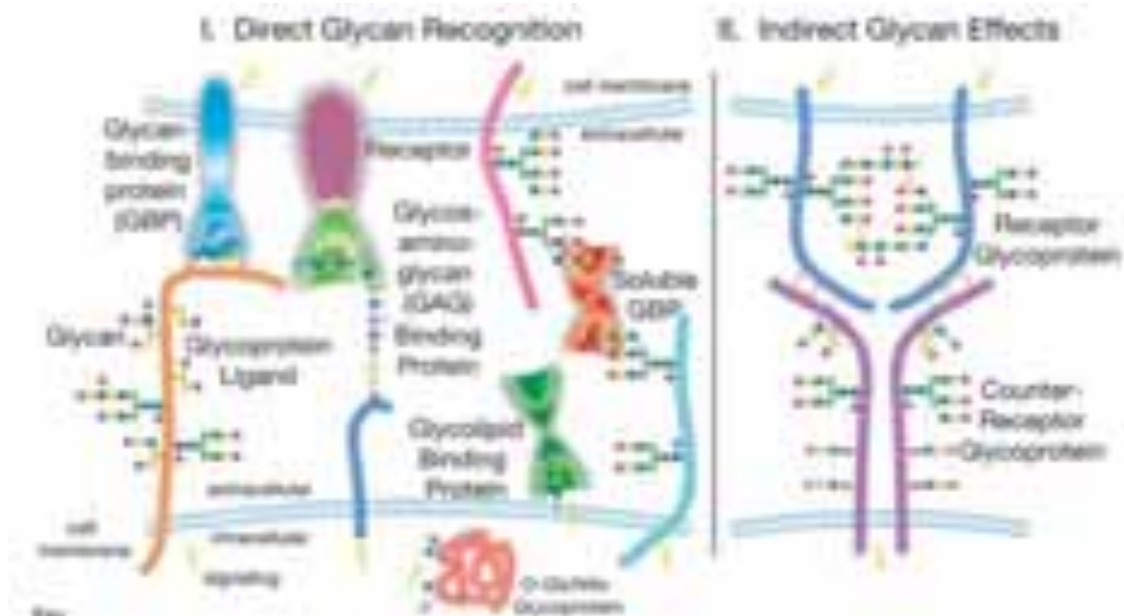
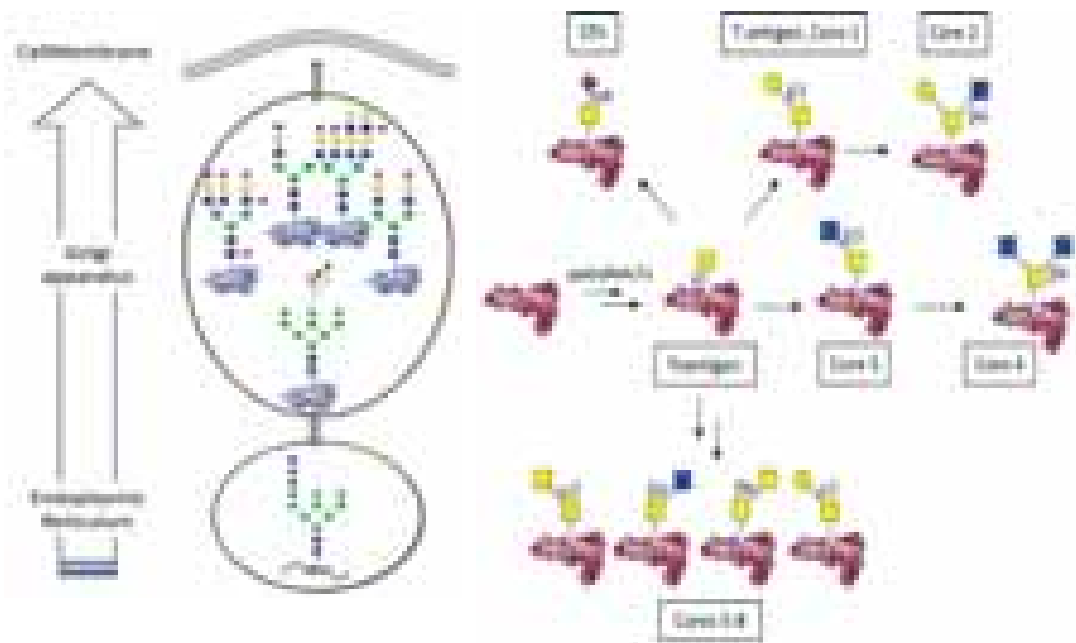


Figure 1. Types of glycans and their effects in biology (Taken from The Challenge and Promise of Glycomics<sup>3</sup>)

Glycoproteins display two main categories of glycans: N- and O-glycans. N-glycans are attached to the amide group of asparagine residues via bloc transfer of a preformed

oligosaccharide unit (10 monosaccharide residues) from the endoplasmic reticulum (ER) to the Golgi apparatus. Sequential addition or removal of single sugar residues, in a defined process regulated by different enzymes which vary depending on the cell type or developmental stage will result in an array of different N-glycans. As a result, all N-glycans share the same core containing 3 mannoses (Man) and 2 N-acetylglucosamines (GlcNAc) which, in mammalian cells, are typically adorned with terminal N-acetylglucosamine, galactose (Gal), sialic acid and core as well as terminal fucose (Fuc).<sup>6</sup>In contrast, O-glycans can be any glycan such mannose, xylose, N-acetylglucosamine and N-acetylgalactosamine (GalNAc), linked to the hydroxyl group of serine, threonine, tyrosine or hydroxylysine. Unlike the N-glycan biosynthesis, O-glycan biosynthesis is initiated by the transfer of a single monosaccharide to the folded protein before undergoing further elongations (Figure 2).<sup>7</sup> Exactly where the biosynthesis occurs in the Golgi also remains unknown. As a result, the heterogeneity of the O-glycans exceeds that of the N-glycans.



**Figure 2. N-glycan vs O-GalNAc glycan biosynthesis**

### 1.1.1 O-glycans

Perhaps as a consequence of the variety of O-glycans, their biological functions remain poorly understood.  $\alpha$ -GlcNAc on the protein  $\alpha$ -Synuclein is suspected to reduce the tendency of this protein to aggregate, thus slowing the progress of Parkinson's disease.<sup>8</sup>  $\alpha$ -Fucose O-glycans found on hepatoma cell lines were suggested to be essential in mediating circulatory clearance of glycoproteins by the liver.<sup>9</sup> However, the most researched are the  $\alpha$ -GalNAc, or mucin-type,

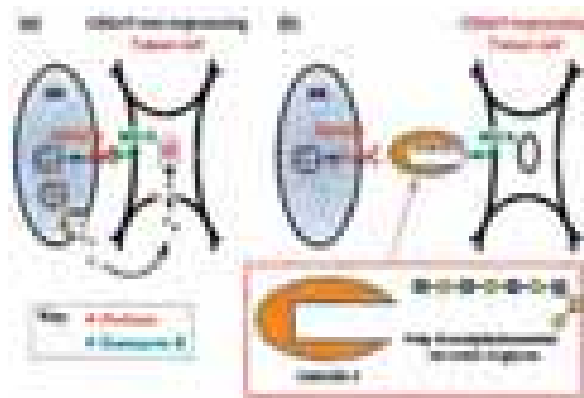


O-glycans which are located on the mucin proteins. These glycoproteins are found in the mucosal sites such as the airways, urogenital and gastrointestinal tracts where they are secreted or membrane-bound. They act as an adhesive agent and physical barrier against external pathogens, protecting the epithelial surfaces from which they are produced.<sup>10</sup> In a healthy intestinal tract, they take on the role of “nightclub bouncers”, allowing healthy bacteria to proliferate but denying access to harmful pathogens.<sup>11</sup> This is thought to be a consequence of the pseudo-bilayer aspect of the mucus. The superficial layer is densely populated by intestinal microbiota whereas the inner layer is bacteria-free, thus shielding the underlying immune cells. Patients suffering from the inflammatory bowel disease ulcerative colitis show a decreased mucus layer thickness, accompanied by altered O-glycosylation and increased penetration of mucus barrier by bacteria.<sup>12</sup> Enriched in serine and threonine to which the first O-GalNAc is attached, mucin O-glycans can make up to 80% by weight of the mucin protein. Considering such dense glycodecoration, it is not surprising that cell-surface mucin O-glycans may play key roles in interactions with the environment.<sup>13</sup>

### **1.1.2 Biosynthesis of mucin-type O-glycans**

Eight core subtypes constitute the bases of the mucin O-glycans (Figure 2).<sup>14</sup> The first GalNAc is transferred in the Golgi from UDP-GalNAc to Ser or Thr of the completely folded protein by polypeptide N-acetyl- $\alpha$ -galactosaminyltransferases (ppGalNAcTs) of which 20 have been identified. These enzymes are conserved across species, are differentially expressed and regulated over tissue and time, and are specific for the sites of attachment of the GalNAc to the Ser/Thr. The first GalNAc (also called the Tn antigen in aberrant glycosylation) is then extended by the enzyme C1GalT-1 to yield core 1 (T antigen) or by C2GnT to yield core 3 (Figure 2). Under the action of the C2GnT and C4GnT, these cores can then be made into cores 2 and 4 respectively. Cores 5-8 also stem from the first  $\alpha$ -GalNAc under the action of different enzymes which remain to be identified. Cores 1 and 2 are found on a variety of glycoproteins and mucins across different cells and tissues and are consequently the most prevalent cores in mammalian cells. For example, the synthesis of core 2 O-glycans is regulated during activation of lymphocytes, cytokine stimulation, and embryonic development. Cores 3-4 are less common, found only in secreted mucins of certain mucin-secreting tissues, such as bronchi, colon, and salivary glands, and cores 5-8 have an even more restricted occurrence (intestinal mucin, adenocarcinoma tissue, ovarian cyst mucin, bovine submaxillary mucin and human respiratory mucin).<sup>7</sup> Typically, these cores are found as elongated structures, modified by GlcNAc, Gal, Fuc, often sialylated and sometimes sulphated, depending on the organism, tissue

and developmental stage. The most characteristic elongation is the repeating GlcNAc  $\beta$ -1,3Gal- $\beta$ 1,4 (polyLacNAc, polyLN) units, which can also be fucosylated. However the mucin-type glycans have generally received more attention for their prevalence in cancer, inflammatory and autoimmune diseases. Cancer cells are notably characterized by their aberrant glycosylation, with the overexpression of truncated O-glycan structures often observed. Specifically, the Tn antigen and its sialylated version (STn antigen) and T antigen represent hallmarks of carcinomas of several organs including lung, colon, stomach, and pancreas.<sup>15,16,17</sup> In contrast, the C3GnT and C2GnT responsible for cores 3 and 4 synthesis were observed to be downregulated in metastatic prostate cancer cell lines suggesting these cores to be tumour suppressive. Yet overexpression of core 2 glycans appeared to promote tumour metastasis by subsidiary interaction with the glycan binding protein galectin 3 thus evading destruction by natural killer cells.<sup>18</sup>



**Figure 3. Core 2 O-glycans tumour metastasis promotion by evasion of NK cell attack. (a) Normal NK cell activation and subsequent destruction of tumour cell not expressing core 2 O-glycans (b) Evasion of the tumour cell from the NK cell attack by interaction of core 2 O-glycans with galectin-3 (taken from Tsuboi *et. al Trends in Molecular Medicine*<sup>18</sup>)**

### 1.1.3 Mucin O-glycan analysis

Despite their relevance in biology, O-glycans have been less studied than N-glycans due to 3 main challenges. The first is the lack of amino acid consensus which complicates sequencing studies and makes it difficult to predict sites of O-glycosylation. The second is the heterogeneity of O-glycosylation which severely complicates the analytical task and no general detection or isolation method currently exists to resolve this. Finally, the lack of a universal enzyme capable of cleaving intact O-glycans from the peptide, like the PNGases for N-glycans, contributes to the difficulty of O-glycan analysis. A commercial O-glycanase is able to cleave core 1 structures but its use is limited by its strict substrate specificity. As a consequence,

chemical release including reductive  $\beta$ -elimination, non reductive  $\beta$ -elimination and hydrazinolysis more conventionally used but despite best efforts still entails drawbacks. Where degradation (peeling) is avoided, O-glycans are unreactive to fluorescent or UV tagging thus compromising quantitative high-resolution separation and analysis. The major limitation in mass spectrometry (MS) coupled analyses are the mass profiles associated with the nature of carbohydrates. For example, hexoses such as galactose, glucose and mannose all have the same mass. Procedures may also suffer from low yields of released O-glycans.<sup>19,20,21</sup>

Increasing evidence shows that cell-surface carbohydrates such as the O-glycans can influence the immune system by interacting with glycan-binding proteins (GBPs) such as lectins and toll-like receptors.<sup>22</sup> These in turn are responsible for triggering a series of signaling events with the overall effect of dictating the immune response. Thus, O-glycans appear to have important roles in shaping the immune system.

## **1.2 C-type Lectin receptors**

Pivotal in regulating the immune response, lectins constitute a large family of receptors found in microorganisms, animals and plants, which specifically bind carbohydrates.<sup>23</sup> They are involved in a variety of biological processes including cell-cell contact, cellular trafficking and signalling and have received particular interest for their role in shaping the immune system. For example, selectins promote migration of leukocytes to the site of injury in the inflammation process whereas collectins are specialized in pathogen recognition.<sup>24</sup> Animal lectins are either soluble or transmembrane bound and can be classified into several subfamilies such as S-, P- or C-type among others, based on the structural similarities of their carbohydrate-recognition domains (CRDs).<sup>25</sup>

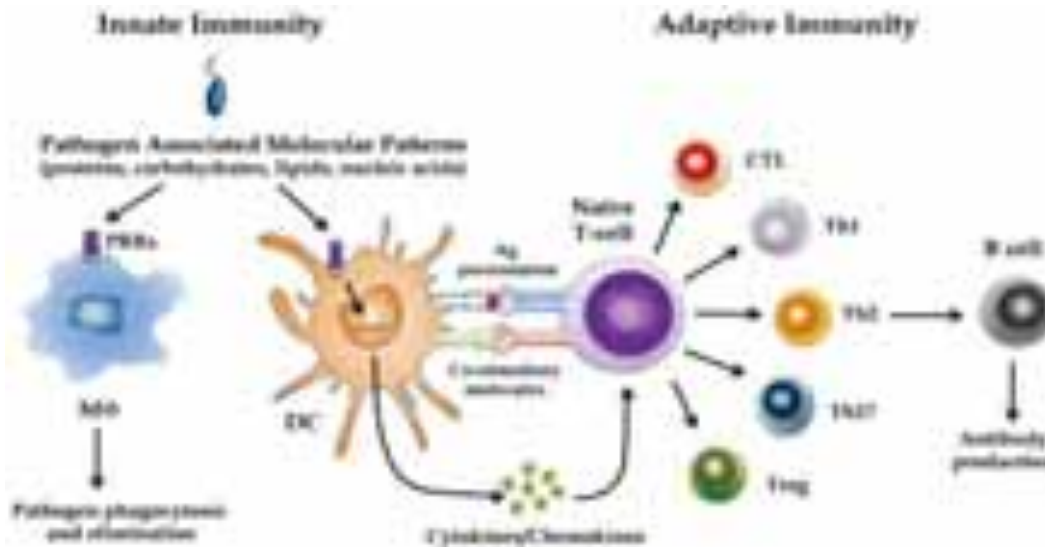
### **1.2.1 Structural characteristics of CLRs**

In the case of C-type lectins, glycan binding is typically  $\text{Ca}^{2+}$  dependent (although not always) and can be first-, second or third order depending on the number of calcium binding sites in a CRD. CLR specificity can be crudely defined by the conserved amino acid residues in the CRD. The "EPN" motif will enable glycans with C-3 and C-4 hydroxyls in equatorial position, ie mannose, glucose, glucosamine or fucose, to chelate to the  $\text{Ca}^{2+}$  with minimal energetic cost. The "QPD" motif will promote binding to galactose or galactosamine via the equatorial-axial C-3 and C-4. Secondary interactions such as hydrogen bonding and pi-stacking in extended or secondary binding sites may increase ligand affinity or the lectin's specificity. For example, the

macrophage galactose binding lectin (MGL) which has the QPD motif is highly specific for GalNAc over Gal residues which modelling studies attributed to complimentary Van der Waals contacts between the 2-acetamido group and His<sup>202</sup>.<sup>26</sup> The CRD being a relatively shallow binding pocket in which typically only terminal glycans or patterns of glycans (epitopes) can be accommodated, absolute assignment of lectin specificity is not possible. Therefore CLRs often display overlapping specificities and may allow for several binding modes. In some cases, CLRs exist as oligomers of their CRDs. The spatial arrangement of the CRD clusters may contribute to subtle differentiation in ligand specificity. Thus, CRDs held together in a fixed geometrical orientation would provide a stricter oligosaccharide specificity by steric hindrance whereas more flexible oligomers could accommodate a wider panel of ligands.<sup>27</sup> Additionally, CLR multivalency has also been shown to increase the receptor's affinity, although this still remains in the millimolar range. Moreover, the multivalency of the ligand presentation is vital in the avidity of the CLR, as the binding affinity has been shown to increase with ligand density.<sup>28</sup> The importance of the multivalent configurations of the CLRs and the glycans is reflected in the emergence of new carrier systems for improved CLR targeting, including dendrimers, nanoparticles, polymers, peptides and other chemical spacers.<sup>29,22</sup>

### **1.2.2 CLRs in the immune system**

C-type lectin receptors are crucial components in the immune system acting as pattern recognizing receptors (PPRs). These constitute the first line of defense against invading pathogens, including viruses (for example HIV and Ebola), bacteria and yeast (for example *Aspergillus fumigatus* and *Candida albicans*).<sup>30,31</sup> By interacting with specific pathogen-associated molecular patterns (PAMPs) at the pathogen surface, CLRs mediate the innate immune response ensuring pathogen capture, neutralization and destruction. Additionally, CLRs present on macrophages and dendritic cells (DCs) provide a bridge to the adaptive immune system (Figure 4). The PAMP endocytosis leads to lysosomal degradation in the antigen presenting cell (APC) which, as its name implies, subsequently presents the antigenic fragments via major histocompatibility complexes (MHCs) at the cell surface. This results in T-cell polarization (T-helper cells [Th1, Th2, and Th17], regulatory T cells, and cytotoxic T cells) and cytokine production through a series of signalling events. Depending on the activated Tcell and the resulting cytokines, immune stimulation or immune suppression is achieved.



**Figure 4. Dendritic cells link innate to adaptive immunity. Different pathogens trigger distinct DC maturation profiles, leading to the polarization of different T-cell subsets. The adaptive immune response is therefore modulated, in some extent, to match the nature of the pathogen. Ag: antigen; CTL: cytotoxic T cell; DC: dendritic cell; Mφ: macrophage. (taken from Protein Kinases, Chapter 6<sup>32</sup>)**

Different pathogens trigger distinct polarization of different T-cell subsets. The adaptive immune response is therefore modulated, to some extent, as a function of the nature of the pathogen. The importance of CLRs is highlighted by the fact that several pathogens take advantage of these receptors to escape intracellular degradation and suppress efficient immune responses.<sup>32</sup> A prime example is illustrated by HIV-1. Viral infection of dendritic cells occurs via interaction of high-mannose structures on the gp120 envelope of the virus with DC-SIGN (Dendritic Cell-Specific Intracellular adhesion molecule-3-Grabbing Non-Integrin, also known as CD209). DC-SIGN is a CLR present on DCs which initially performs its functions well by internalizing the virus for further degradation. However rather than being degraded and presented on MHC complexes, HIV-1 is able to remain in the DCs for a prolonged period of time leading to T cell infection ultimately debilitating the immune system and causing AIDS.<sup>33</sup>

It is now clear that CLRs are pivotal in modulating the immune system and unsurprisingly have therefore been demonstrated to be involved in immune-compromised pathologies including cancer, allergies, inflammatory and autoimmune diseases.<sup>34,35,36,37</sup> While some CLRs are beginning to unravel their complex interactions and related functions, the majority still remain unsolved puzzles. Most efforts towards discovering efficient CLR targeting strategies for therapeutic uses have been focused on antibody-mediated strategies. In recent years, particular efforts have been focused on glycan mimetics as they feature the attractive

advantages of lower immunogenicity, the opportunity to target several CLRs simultaneously and easier tuneability of pharmacokinetics.<sup>38,39</sup>













### 1.3 Parasitic glycans


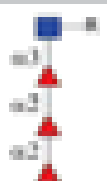
It seems hard to believe that a parasitic worm (or helminth) responsible for schistosomiasis, the world's 2<sup>nd</sup> most socio-economically devastating disease after malaria, may provide cures for inflammatory and autoimmune diseases. Yet *Schistosoma* are indeed receiving increased interest, particularly in their ability to regulate the infected host's immune system. Schistosomiasis is a neglected tropical disease estimated to affect almost 240 million people worldwide according to the World Health Organization. It is a disease that disables more than it kills (adult worms can live up to 40 years in a host) with symptoms including abdominal pain, diarrhea, blood in the stool or urine and in advanced cases liver damage, kidney failure, infertility or and bladder cancer. It is now apparent that the helminths have developed a strategy to evade the host's immune system, to simultaneously ensure their survival and that of its host. Although the mechanism of immune evasion is not clearly established, mounting evidence points to a glycan based strategy. Indeed, studies on the mechanism of action of the drug Praziquantel, the treatment for schistosomiasis, suggested that the triggered immune response was achieved by making the schistosome susceptible to antibody attack through increased presentation of surface antigens.<sup>40</sup>

#### 1.3.1 *Schistosoma mansoni* glycans

Of the four species which constitute the *Schistosoma* genus, *Schistosoma mansoni* is the best studied as it has been shown to modulate the immune response towards allergic tolerance and reduced morbidity in autoimmune diseases such as multiple sclerosis, rheumatoid arthritis, type I diabetes, and inflammatory bowel diseases.<sup>41,42</sup> This is thought to be concomitant with the observed downregulation of the characteristic T<sub>H</sub>2/Treg response and the associated cytokines during chronic infection. It is now widely accepted that *S.mansoni* glycans are immunodominant molecules that interact with CLRs and dictate Tcell differentiation thereby modulating the immune response. Although yet to be fully assigned, *S.mansoni's* glycome stands out by its rich array of ligands foreign to vertebrates, such as xylose, high-mannose and poly-fucosylated structures. The typical helminth glycan modification GalNAc $\beta$ -1,4GlcNAc (termed the LacdiNAc motif, LDN) is also abundantly observed on all glycoconjugates, as well as its fucosylated derivatives (F-LDN, LDN-F, F-LDN-F, LDN-DF and DF-LDN-DF). These are

seldom present in vertebrate glycans and LDN induces the humoral response IgE. Moreover, sialic acid, the most common terminal modification of mammalian glycans that helps maintain immune homeostasis via interaction with sialic acid binding lectins (siglecs), is distinctly absent from the worm's glycome. However, the motifs polyLN and  $\alpha$ 1,3-fucose derivative (Lewis X, Le<sup>X</sup>) are found both in the parasite and mammalian glycoconjugates. (Table 1)

Glycan motif	Glycan structure	Structure in symbols
LacNAc/LN	Gal $\beta$ 1-4GlcNAc $\beta$ 1-	
LewisX/LeX	Gal $\beta$ 1-4(Fuca $\alpha$ 1-3)GlcNAc $\beta$ 1-	
Pseudo Lewis Y/pseudo-LeY	Fuca $\alpha$ 1-3Gal $\beta$ 1-4(Fuca $\alpha$ 1-3)GlcNAc $\beta$ 1-	
LacDiNAc/LDN	GalNAc $\beta$ 1-4GlcNAc $\beta$ 1-	
LDN-F	GalNAc $\beta$ 1-4(Fuca $\alpha$ 1-3)GlcNAc $\beta$ 1-	
F-LDN	Fuca $\alpha$ 1-3GalNAc $\beta$ 1-4GlcNAc $\beta$ 1-	
F-LDN-F	Fuca $\alpha$ 1-3GalNAc $\beta$ 1-4(Fuca $\alpha$ 1-3)GlcNAc $\beta$ 1-	
LDN-DF	GalNAc $\beta$ 1-4(Fuca $\alpha$ 1-2Fuca $\alpha$ 1-3)GlcNAc $\beta$ 1-	
F-LDN-DF	Fuca $\alpha$ 1-3GalNAc $\beta$ 1-4(Fuca $\alpha$ 1-2Fuca $\alpha$ 1-3)GlcNAc $\beta$ 1-	
DF-LDN-DF	Fuca $\alpha$ 1-2Fuca $\alpha$ 1-3GalNAc $\beta$ 1-4(Fuca $\alpha$ 1-2Fuca $\alpha$ 1-3)GlcNAc $\beta$ 1-	
DF-LDN-TF	Fuca $\alpha$ 1-2Fuca $\alpha$ 1-3GalNAc $\beta$ 1-4(Fuca $\alpha$ 1-2Fuca $\alpha$ 1-2Fuca $\alpha$ 1-3)GlcNAc $\beta$ 1-	
F-GlcNAc	Fuca $\alpha$ 1-3GlcNAc $\beta$ 1-	

<b>DF-GlcNAc</b>	Fuca1-2Fuca1-3GlcNAcβ1-	
<b>TF-GlcNAc</b>	Fuca1-2Fuca1-2Fuca1-3GlcNAcβ1-	

**Table 1. Terminal glycan motifs in *S.mansoniglycoconjugates***

The complexity of the schistosome glycome doesn't solely arise from the helminth's unique array of glycans but was also shown to alter as a function of the worm's developmental stage.<sup>43</sup> Regulation of *S. mansoni* glycan expression could be observed by extensive mass chromatography analysed of glycan products of individual stages of the worm's life. This was further supported by an analysis of gene expression corresponding to various glycosyltransferases responsible for the glycan synthesis, which were also shown to be modulated according the maturity of the worm.<sup>44,45</sup> Thus, *S.mansoni's* O-glycans were seen to contrast between predominant of polyLacNAc and polyLeX structures in the cercarial stage and polyLDN and polyLDNF in the eggs stage (Figure 6).

Additional diversity may also arise as a result of the parasite's need to infect two hosts during its lifecycle and therefore need to adapt to both. Schistosome eggs hatch releasing miracidia which swim and infect the specific intermediate snail host, *Biomphalaria Gilbrata* in the case of *S. mansoni*, where they mature until they are released as cercariae into the water. The free-swimming cercariae then infect the definitive human host by skin penetration, during which they lose their tail, becoming schistosomulae. The schistosomulae migrate throughout several tissues via the portal vessels where they feed on blood and grow until they become mature adults. Male and female worm pairs migrate to the blood vessels in the lower intestine where they lay eggs, the majority of which is excreted with the feces.



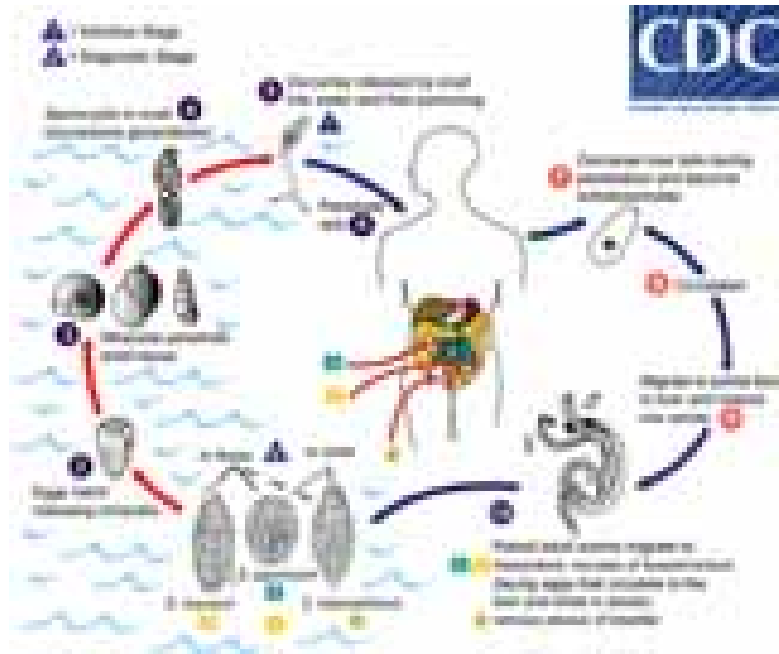
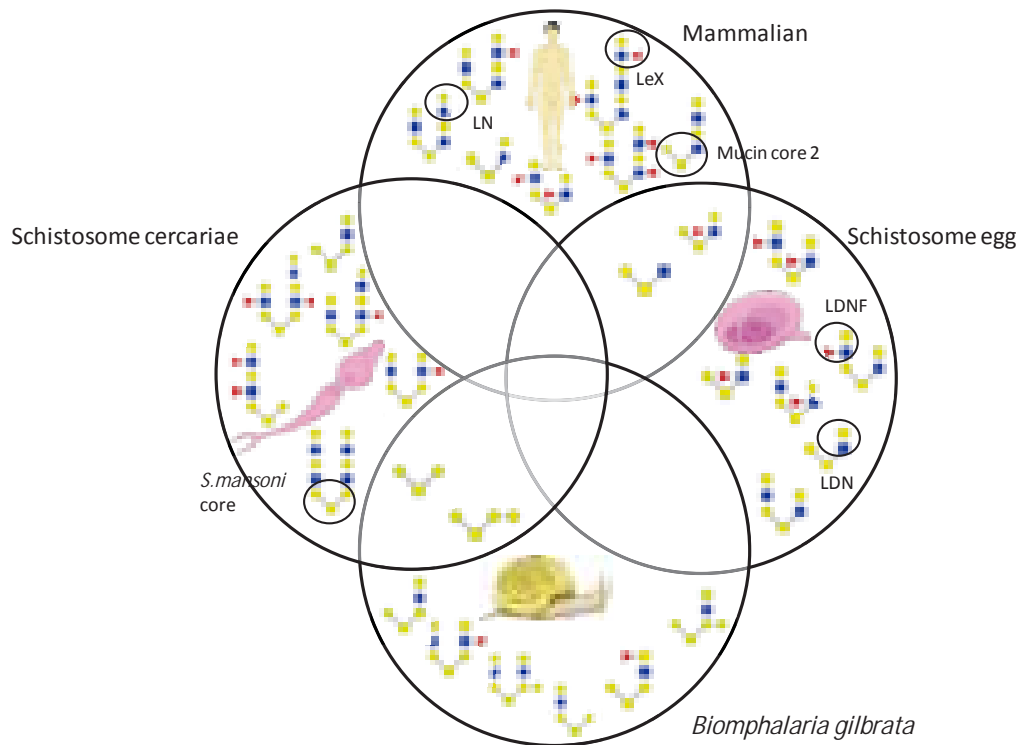


Figure 5. Life cycle of schistosomes (Public Health Image Library, <http://www.dpd.cdc.gov/dpdx>)

### 1.3.2 Glycan mimickry for immune evasion

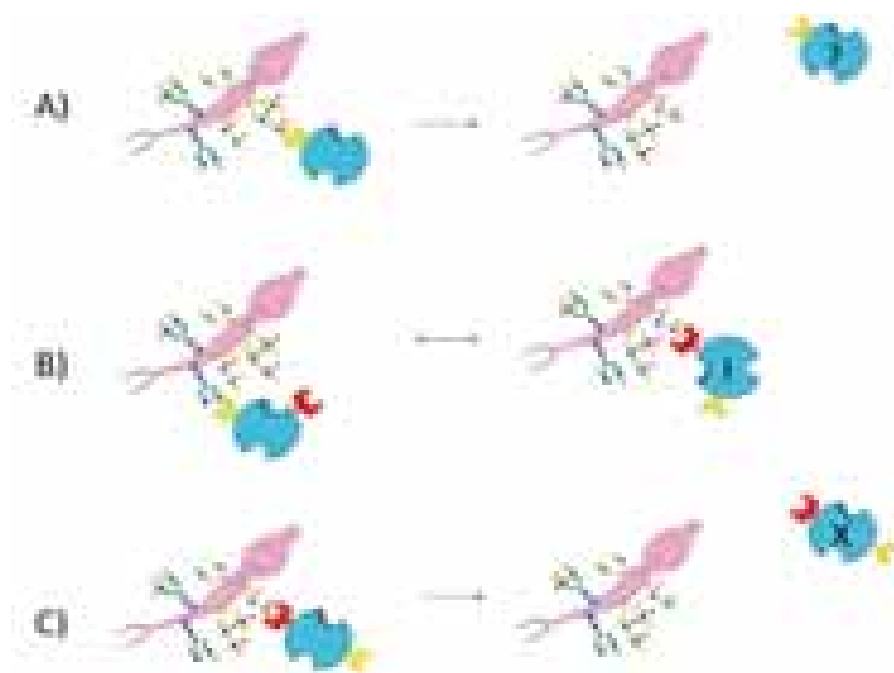
Interesting similarities can be observed when comparing *S. mansoni* O-glycans to those of its hosts. The *S. mansoni* specific core, which is predominant in the cercarial extractions of O-glycans, is reminiscent of the 4-O-methylated core reported to be characteristic of several snails, including the intermediate water snail *B. gilbrata*.<sup>46,47</sup> Moreover, this core is typically observed to be extended with a single hexose and such a structure has also been speculated to occur.<sup>48</sup> In a study examining glycan consensus between snail and parasite, immunocytochemical and Western blot analyses revealed that LDN and F-LDN epitopes are present in both snail hemocytes and parasite miracidia and primary sporocytes. Specifically, the tegument and larval transformation products of these life stages, both of which arise during parasitic development in *B. gilbrata*, were seen to bear these epitopes. Additionally, hemocytes were shown to bind and display larval glycoconjugates thereby modulating various hemocyte functions including protein synthesis/secretion, stress response and MAPK (erk) signaling. However, cercarial O-glycans extracts were seen to consist of the *S. mansoni* core displaying typical mammalian motifs of polyLN and especially polyLeX. On the other hand, the O-glycan extracts of *S. mansoni* eggs were shown to be predominantly based on the mucin 2 core but these O-glycans contrasted strikingly to mammalian O-glycans by their LDN/LDNF

decoration, as opposed to LN/LeX.<sup>49</sup> It almost appears as if *S.mansoni* purposefully uses O-glycan cores of a previous host to display the future host's terminal glycan motifs and begs the question to what purpose?



**Figure 6. Venn diagram showcasing examples of O-glycan structures from *S.mansoni* and its hosts**

In view of these similarities, glycan mimicry is strongly suspected to be either endeavoured or acquired by the parasite as a strategy to subvert the host's immune system. The exact mechanism of such mimicry on *S.mansoni* is unknown however the hypothesis proposed by Yoshino *et al.* is 3-fold: it could be an invisibility cloak, a temporary distraction or a repellent/shield.<sup>50</sup>



**Figure 7. Possible reasons for glycan mimicry exhibited by *S. Mansoni*: A) evasion, B) temporary distraction, C) repulsion**

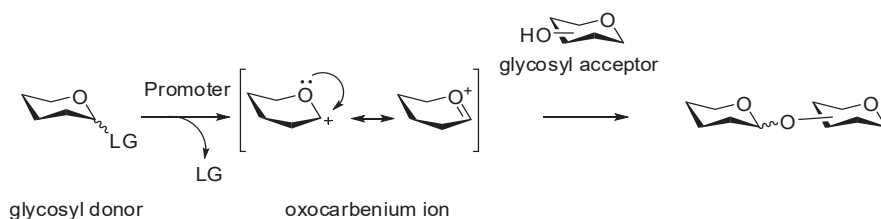
Naturally, isolation of *S. mansoni*'s glycans would greatly advance the understanding of the worm's strategy to target the CLRs and hijack the host's immune system. Typically, glycan extracts from the SEA and cercarial content are used for their demonstrated antigenic properties but also because they are easier to obtain.<sup>51</sup> Within these extracts, the N-glycans and glycolipids have been more investigated as systematic analysis is easier (see 1.1.3).<sup>52,53</sup> On the other hand, O-glycans remain challenging and have therefore received much less attention. Recently, a modified hydrazinolysis enabled van Diepen *et al.* to obtain O-glycans from different *S. mansoni* life stages with limited degradation.<sup>54,49</sup> Multi-dimensional HPLC-separation of the labelled O-glycans, antidody- and MS- based profiling revealed that the O-glycans contain known antigenic motifs and demonstrated them to be equally as valuable as N-glycans and glycolipids.<sup>48, 55</sup> However, isolation of pure compounds in sufficient amounts for additional studies remains challenging.

## 1.4 Synthesis of O-glycans

### 1.4.1 Carbohydrate chemistry

Oligosaccharide synthesis requires elaborate methodologies to achieve the chemo-, regio- and stereoselectivity displayed in the linkage of carbohydrate building blocks. The formation of a

glycosidic bond linking a specific hydroxyl group of a monosaccharide to the anomeric carbon of another (glycosylation) is inherently challenging given the polyhydroxyl nature of carbohydrates. The general mechanistic pathway of glycosylation involves the activation of a leaving group at anomeric carbon of a glycosyl donor using a promoter to form an intermediate electrophilic oxocarbenium ion which can be attacked by a nucleophilic, free hydroxyl group of the glycosyl acceptor to form a new, covalent, glycosidic bond.<sup>56</sup>

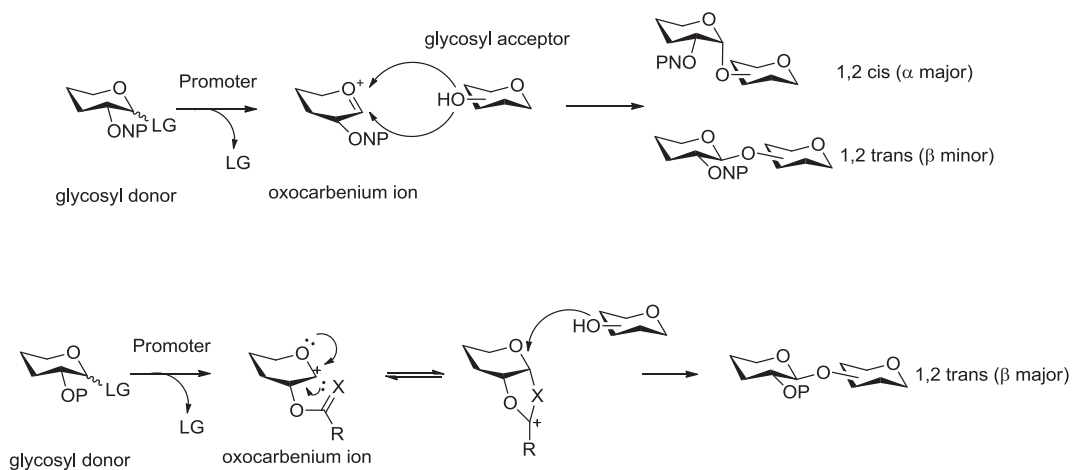


### Scheme 1. Chemical glycosylation

In order to achieve regioisomeric control, elaborate series of intricate protection and deprotection steps allow chemists to selectively functionalize desired positions of carbohydrates for future glycosylations.<sup>57</sup> The order in which these are carried out on a monosaccharide is based on the relative reactivities of the hydroxyl groups (which itself depends on their configuration) in the chemical transformation. For example arylidene acetals are selectively formed between 1,3 diols whilst acetonides are preferentially formed with 1,2 diols. The hydroxyl nucleophilicity which decreases in the order anomeric, primary, secondary and equatorial over axial groups is also an exploitable property.<sup>58</sup> Therefore a judicious synthetic design affords control of regiochemistry by chemoselectively masking and unmasking specific hydroxyls for glycosylation. Protecting groups can be broadly defined as either permanent or temporary. Permanent protecting groups, for example benzyl and naphthyl ethers, are typically installed at the early stages of the synthesis and, being unaffected by all other chemical reactions (orthogonal), remain throughout the synthesis only to be removed in the final steps. As a result, the hydroxyl group on which they are installed is unavailable until the permanent protecting group is removed. On the other hand, temporary protecting groups such as silyl ethers, acetates or 2,2,2-trichloroethyl carbonates are easier to remove, requiring milder reaction conditions. As they are easier to remove, temporary protecting groups are less stable to other reaction conditions, making their orthogonality less broad. Regardless of the type, the protecting groups employed should be easily manipulated as oligosaccharide synthesis is almost always long and laborious owing to its convergent nature and therefore subject to overall low-yields.

Similarly, the anomeric leaving group can vary according to the desired reactivity or required condition reactions. For example, thioglycosides typically employed for their general longer shelf-life whereas the much less stable trichloroacetimidates (TCA) are consequently more reactive. Routine donors include glycosyl halides, trichloroacetimidates, *N*-phenyl trifluoroacetimidates, sulfoxides, phosphates and thioglycosides. TCA are perhaps used even more often due to their ease of preparation, high reactivity, mild activation conditions (catalytic amounts of lewis acid) and compatibility with acid and base labile protecting groups.

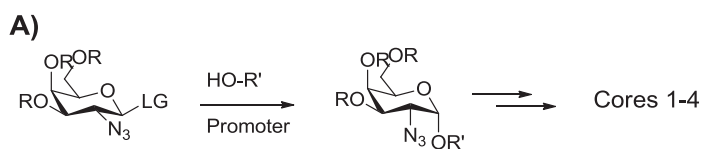
Additional stereochemical concerns of axial ( $\alpha$ ) or equatorial ( $\beta$ ) linkages also contribute to the complexity of the synthesis. The stereochemical outcome of the new bond will depend on the direction of attack of the nucleophilic donor to yield either  $\alpha$ - 1,2 trans, or  $\beta$ - 1,2 cis linkage. Stereochemical control can be achieved by varying several parameters. When a  $\beta$ -linkage is desired, typical strategies involve the use of neighbouring group participation (NGP) in which a protecting group of C-2 will dictate a preferential equatorial  $S_N2$  attack of the acceptor by blocking the axial site of attack. This is recurrently used with groups containing available lone pairs such as naphtyl, troc, acetyl etc. which can form an intermediate acetal or oxazolidine species under acidic conditions. In the absence of NGP, an  $S_N1$  pathway occurs and the  $\alpha$ -anomer is predominantly observed due to the reported anomeric effect in which the unshared electron pair of the endocyclic oxygen and the  $\sigma^*$  orbital for the axial (exocyclic) C–O bond hyperconjugate thus lowering the overall energy of the molecule.<sup>59</sup>



**Scheme2. Stereochemistry of glycosylation**

### 1.4.2 Mucin O-glycans synthesis

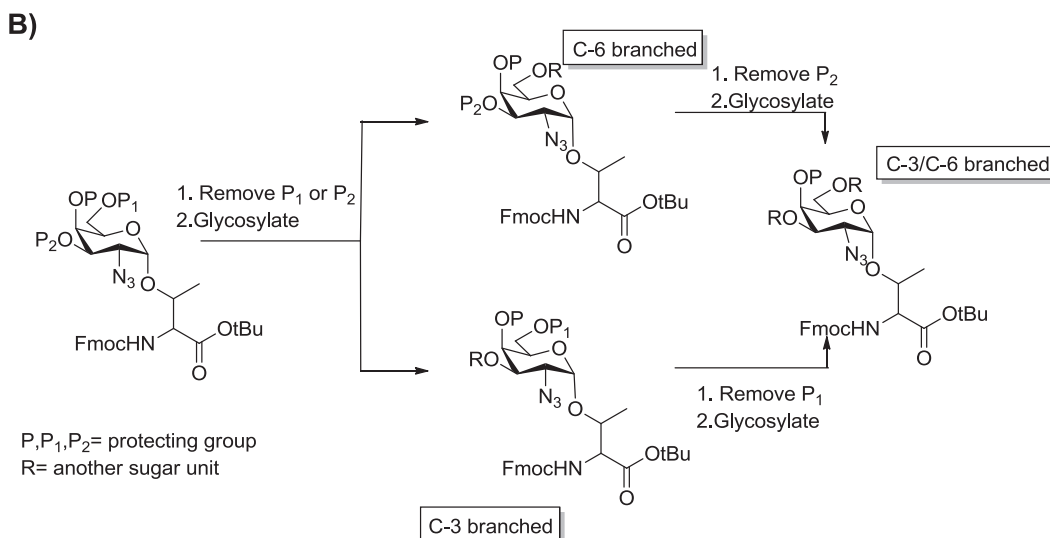
Given the importance of mucins, several methods have been described to access them including, solid-phase peptide synthesis, chemoenzymatic synthesis and native chemical ligation.<sup>60,61</sup> Historically, the most challenging step of O-glycan synthesis was the  $\alpha$ -glycosidic linkage between the non-reducing end of the GalNAc monosaccharide and the side-chain hydroxy groups of L-serine, L-threonine or an unnatural linker. Defined in the final steps of the synthesis of the cores 1-4, the  $\alpha$ -selectivity of Schmidt's trichloroacetimidate and the Norberg's thiomethyl-based strategies relied on the non-participating N-azido protecting group in the GalNAc residue (Figure 8, A).<sup>62,63</sup> The improved strategy developed by Mathieux *et al.* involved an orthogonally protected galactosamine building block already containing the amino acid residue as the starting point.<sup>64</sup> Termed the cassette-method, this strategy exploits the structural similarities of cores 1-4 thus enabling access to all cores via a common precursor (Figure 8, B). More recently, nitroglycals were used in base-catalysed reactions for a Michael-type addition of serine or threonine to form O-glycosides (Figure 8, C).<sup>65</sup> Finally, ionic liquids been studied for one-pot glycosylation reactions under mild conditions.<sup>66</sup>

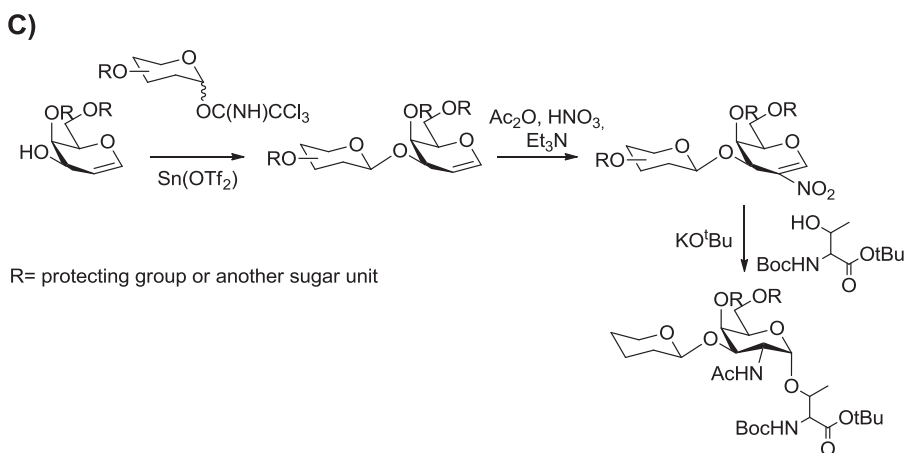


R= Ac or another sugar unit

LG = OC(NH)CCl<sub>3</sub> or SME

R' = OCH<sub>2</sub>CH(NHFmoc)COOBz  
or O(CH<sub>2</sub>)<sub>2</sub>(C<sub>6</sub>H<sub>4</sub>)NHCOCF<sub>3</sub>





**Figure 8. General O-glycan synthetic strategies**

The interesting biological relevance of the T antigen, the Tn antigen (or core 1) and the core 2 (see 1) is reflected by their prevalence in synthetic literary reports compared to the other O-glycan cores.<sup>67,68,69</sup> With the notable exception of Xia *et al.*, reports on the chemical synthesis of O-glycans are limited to the synthesis of the cores.<sup>70,71</sup> The syntheses of more complex oligosaccharides have been enzymatically obtained (see 1.4.3).

In contrast, despite the *S.mansoni* O-glycans being elucidated and shown to be immunogenic, there is no literature to date describing their synthesis.

### 1.4.3 Enzymatic Glycosylations

As elegant as chemical advances have been towards facilitating the tediously long synthesis of polysaccharides, they remain inferior to nature's abilities. Enzymes that are able to catalyze the formation of glycosidic bonds with high specificity and selectivity are called glycosyltransferases and are extensively used in carbohydrate chemistry. They present notable advantages in oligosaccharide synthesis including their high regio- and stereo-specificity, their simplicity of use and their more sustainable aspect by virtue of the fact that all reactions and additives are aqueous, thus having less of an environmental burden. Glycosyltransferases can be separated into 2 classes based on their reaction mechanism: inverting, which is a  $S_N2$  type reaction, or retaining, a  $S_N1$  type reaction. The enzymes use activated sugar donors which in nature are mainly nucleoside di-phosphates (Leloir donors) although mono-phosphates (CMP-sialic acid) and non-activated donors also exist.<sup>72</sup>



**Scheme 3. Inverting vs retaining mechanism of glycosyltransferases**

The complimentary use of glycosyltransferases in carbohydrate chemistry (chemoenzymatic synthesis) is almost routine nowadays and the growing market is a testimony to their power. The enzymes used originate from a range of different organisms, including fungal, plant and human, depending on the desired target and the enzyme activity. As a consequence, an abundance of methods to chemoenzymatically synthesize carbohydrates have been reported.<sup>73,74</sup>

Yet as efficient as glycosyltransferases may be, their use heavily relies on the availability of the enzyme and sugar nucleotide donors. In some cases, such as the highly fucosylated structures in *S. Mansoni*, no enzyme with this function has been isolated yet, making these structures unobtainable other than by chemical synthesis. In other cases, the cost of the enzymatic elongation may not compensate the synthesis of the target compound(s). Recombinant human enzymes (for example B3GnT2) are the most pertinent when investigating the role of glycans in the autoimmune system but are generally poorly expressed by *E.coli*, making their isolation and large scale application in synthesis limited.<sup>75</sup> Although some ingenious solutions have been proposed to circumvent this obstacle, such as repurposing glycosyl hydrolases (the enzymes responsible for cleaving sugars from a protein surface) towards synthesis or engineering plant glycosylation machinery for production of helminth antigens, these solutions remain relatively scarce.<sup>76,77</sup> Mammalian, plant and parasitic glycosyltransferases are often difficult to express or isolate. In contrast, several examples exist of bacterial and fungal glycosyltransferases as tools towards the synthesis of oligosaccharides. The simpler expression in eukaryotic systems such as *E.coli* or *P.pastoris* together with their often broader substrate selectivity make these enzymes the tools of choice. Examples include the fucosyltransferases (FUTs) from the nematode *Candida elegans* and the  $\beta$ 1,3 glucosaminyltransferase ( $\beta$ 3-GlcNAcT) LgtA from *Neisseria meningitidis*. Used in tandem, recombinant glycosyltransferases can be used to

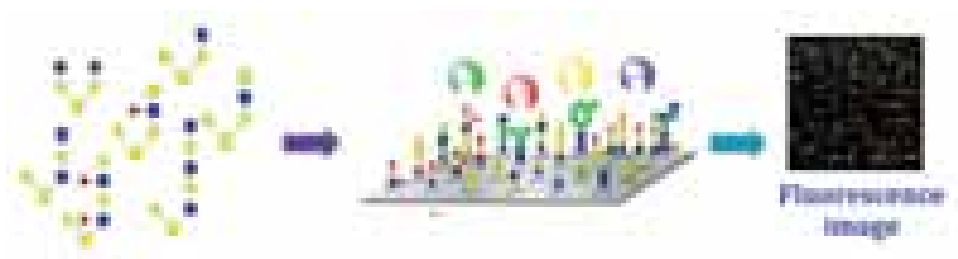


generate elaborate oligosaccharides.<sup>68</sup> This is especially well illustrated in the construction of polyLacNAc structures which alternates a  $\beta$ 1,4 galactosyltransferase ( $\beta$ 1,4-GalT) and a  $\beta$ 1,3-GlcNAcT.<sup>78</sup> While the enzymatic synthesis of O-glycans from a GalNAc building block has been described, for example in the synthesis of a P-selectin glycoprotein ligand-1 (PSGL-1) fragment containing a core 2-based glycan capped with a sialyl Lewis X (sLe<sup>x</sup>) motif, a chemo-enzymatic approach is generally used.<sup>79,80</sup> This approach involves the chemical synthesis of cores which are elongated enzymatically for complex oligosaccharide structures. This is due to the lack of commercially available enzymes catalyzing the reactions leading to the cores structures.

## 1.5 Glycan microarray

### 1.5.1 Principles of glycan microarray

Since their first appearance in literature in 2002, glycan microarrays emerged as a powerful tool to probe glycan-protein interactions.<sup>81,82,83</sup> Its strength lies in the ability to perform multiple analyses of glycan-protein binding events in parallel while only using small amounts of sample. The premise is simple: glycans are attached in a spatially-defined environment to a compatible solid surface and are then exposed to receptors. The binding events are then analyzed, typically by fluorescence where the intensity reflects the binding strength. Glycan microarrays have found broad applications in rapid analysis of the glycan binding properties of proteins, quantification of glycan-protein interactions, pathogen detection, and rapid characterization of carbohydrate-processing enzymes. With the interest in GBPs rapidly growing, numerous methodologies for constructing microarrays have also emerged and have been comprehensively described.<sup>84,85</sup>



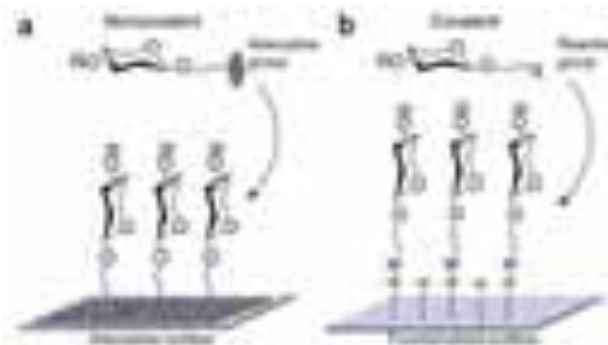
**Figure 9. Principles of glycan arrays**

Despite being conceptually simple, glycan microarrays are technically challenging with many parameters affecting the assay quality. A considerable advantage of glycan microarrays is the minute quantities needed for both the precious carbohydrates (femtomolar range) and the analyte (microgrammes). Other than synthesis, glycans can be sourced from biological extracts

which entails a lot of work in the treatment of the biological extract, isolation of the glycan pool, separation (if possible) of individual glycans, characterization and functionalization for immobilization, at the end of which only small amounts are commonly obtained. The Consortium of Chemical Glycomics (CFG) contains over 600 mammalian glycan targets which it publicly provides as a screening service to investigators as part of its mission to elucidate protein-carbohydrate interactions in cell-cell communication (<http://functionalglycomics.org>). However, this removes the flexibility of modifying certain assay parameters such as the glycan concentration or presentation on the surface. Moreover, the arrays are focused on mammalian structures and rarer glycan structures such as those found in *S.mansoni* are not available.

Of course, dispensing such small amounts (glycan microspots average of 100-200 microns in diameter) requires automation. Printing of glycan solutions can be distinguished as contact or non-contact. Contact printing involves a set of tips being dipped in the ligand solutions and transferring them by direct contact to the surface. The amount of solution transferred is consequently strongly dependent on the contact time of pin type. In contrast, non-contact printing is typically carried out using a piezoelectric printer which dispenses the drop out of a capillary by controlled electric signals. This generally affords spots of more homogenous size and morphology. However, only 4-8 tips can be used simultaneously making it slower than contact printing.

An additional factor to consider when designing microarrays is the method by which glycans are attached to the chip surface. This also falls into two categories -covalent or non-covalent- which both present contrasting advantages and disadvantages. Non-covalent immobilization relies on hydrophobic or charge-based interactions but is intrinsically subject to partial removal of material by washing. Examples include lipid-derivatized glycans onto nitrocellulose-coated glass slides and fluorine-derivatized glycans onto fluoroalkylsilane slides. Covalent immobilization relies on glycans reacting with an activated surface to form a covalent bond. Free reducing sugars can be directly attached onto hydrazide- or aminoxy- modified glass slides but generally glycans tend to be functionalized with spacers (for example maleimides, amines or thiols). N-hydroxysuccinimide (NHS) ester- or epoxy- activated glass slides which react with amino-spacers are most frequently used for their high stability.



**Figure 10. Non-covalent vs covalent glycan attachments**

Fluorescence-based detection methods are the most widely used because of their high sensitivity and throughput as well as the availability of fluorescence detectors such as a high-resolution microarray scanner. More recently, glycan microarrays have been coupled to other detection methods such as surface plasmon resonance (SPR) and MALDI-TOF mass spectrometry (MS), providing additional information such as association and dissociation constants and enzymatic activity.<sup>86</sup> For example, a new platform equipped with a hydrophobic layer and an indium-tin-oxide (ITO) layer developed in our group enabled the study of eight glycosyltransferases activity, the assignment of the specificity of a fucosyltransferase by on-chip product fragmentation by MS and a lectin binding analysis by fluorescence.<sup>87</sup>

### 1.5.2 Applications of glycan microarrays

Glycan microarrays have been used to study a variety of glycan-associated recognition events including the detection of pathogens for diagnosis and the characterization of glycan-processing enzymes.<sup>88,89,90</sup> They have been especially instrumental in the characterization of lectin specificity and affinity. For example the homologous CLRs DC-SIGN and DC-SIGN R were shown to have distinct ligand specificity. Thus, while DC-SIGN was observed to bind various high-mannose and fucose containing ligands, DC-SIGN R only bound mannose-presenting glycans. This further advances the understanding of CLRs roles in immunity.<sup>91</sup> Carbohydrate profiling also identified an exclusive specificity for terminal  $\alpha$ - and  $\beta$ -linked GalNAc residues by the CLR MGL. As these are abundantly found on *S.mansoni* glycans, this might provide an insight into the mechanism of parasitic immune evasion.<sup>92</sup> More recently, a library of mannose- and fucose-based glycomimetics was screened against a panel of CLRs which revealed that only Dectin-2 interacted with  $\beta$ -fucosides. This gave valuable indications for the design and optimization of dectin-2 selective antagonists.<sup>93</sup>

Glycan microarrays also played a particularly significant role in profiling *S.manson*'s complex glycome. In 2015, parasite natural glycan extracts from HPLC fractionation were immobilized on epoxysilane-coated glass slides and assayed against a panel of known antibodies. This led to the discovery of key antigenic glycan motifs in GSL, N- and O-glycans, and further advanced research towards glycan vaccine candidates and glycan-based diagnostics.<sup>48,94</sup> Moreover, interesting results were reported by van Diepen *et al.* regarding the antibody responses to *S.manson*'s O-glycans. In particular, higher IgG responses were observed against relatively large and more complex cercarial and egg-derived O-glycans in schistosome-infected children than against the previously reported N- and GSL- glycans.<sup>48,52</sup> To date, no additional work was reported on the relevance of *S.manson*i O-glycans, probably as a consequence of the lengthy and arduous task that is O-glycan release.

As a result of their versatility, glycan microarrays have emerged as an indispensable tool for glycan-GBP interaction studies and for CLR targeting.

## References

1. S.Weinbaum, J.M.Tarbell, E.R. Diamino, *Annu. Rev. Biomed. Eng.*, **2007**, 9, 121–167
2. P.K.Chaffey, X.Guan, C.Chen, Y.Ruan, X.Wang, A.H.Tran, T.N.Koelsch, Q.Cui, Y.Feng, Z.Tan, *Biochem.*, **2017**, 56, 2897-2906
3. R.D.Cummings, J.M.Pierce, *Chemistry & Biology*, **2014**, 21, 1-15
4. J.E.Hudak, C.R.Bertozi, *Chemistry & Biology*, **2014**, 21, 16-37
5. D.Feng, A.S.Shaikh, F.Wang, *ACS Chem. Biol.*, **2016**, 11, 850-863
6. P. Stanley, N.Taniguchi, M.Aebi, *Essentials of Glycobiology, 3rd Edition Chap.9*
7. I. Brockhausen, P. Stanley, *Essentials of Glycobiology, 3rd Edition Chap.10*
8. Y.E. Lewis, A. Galesic, P.M. Levine, C.A. De Leon, N. Lamiri, C.K. Brennan, M. R. Pratt, *ACS Chem. Biol.*, **2017**, 12, 1020–1027
9. P. Van den Steen, P.M. Rudd, R.A. Dwek, G. Opdenakker, *Critical Reviews in Biochemistry and Molecular Biology*, **1998**, 33, 151-208
10. L.E. Tailford, E.H. Crost, D. Kavanaugh, N. Juge, *Front. Genet.*, **2015**, 6, 1-18
11. S.S. Dhanisha, C. Guruvayoorappan, S. Drishya, P. Abeesh, *Critical Reviews in Oncology/Hematology*, **2018**, 122, 98-122
12. K. S. B. Bergstrom, L. Xia, *Glycobiol.*, **2013**, 23 1026–1037
13. E.Tian, K.G.Ten Hagen, *Glycoconj. J.*, **2009**, 26, 325-334
14. Y. Mechref, M.V. Novotny, *Chem. Rev.*, **2002**, 102, 321–369
15. I. Brockhausen, *EMBO reports* , **2006**, 7, 599-604,
16. W-L.Ho, W-M. Hsu, M-C.Huang, K.Kadomatsu, A.Nakagawara, *Journal of Hematology& Oncology*, **2016**, 9:100
17. S.Holst, M.Wuhrer, Y.Rombouts, *Adv. Cancer Res.*, **2015**, 162, 203-256
18. S.Tsuboi, S.Hatakeyama, C.Ohyama, M.Fuduka, *Trends in Molecular Medecine*, **2012**, 18, 224-232
19. P.H.Jensen, D.Kolarich, N.H.Packer, *FEBS Journal*, **2010**, 277, 81-94
20. S.Mugalapati, V.Koppolu, T.S.Raju, *Biochem.*, **2017**, 56, 1218-1226

21. K.Yamada, S.Hyodo, M.Kinoshita, T.Hayakawa, K.Takehi, *Anal.Chem.*, **2010**, 82, 7436-7443
22. R.J.E. Li, S.J. van Vliet, Y. van Kooyk, *Current Opinion in Biotechnology*, **2018**, 51, 24-31
23. W.I.Weis, K.Drickamer, *Annu.Rev.Biochem.*, **1996**, 65, 441-473
24. W.I.Weis, M.E.Taylor, K.Drickamer, *Immunological Reviews*, **1998**, 163, 19-34
25. K.Drickamer, M.E.Taylor, *Current Opinion in Structural Biology*, **2015**, 34, 26-34
26. S.A.F.Jégouzo, A.Quintero-Martinez, X.Ouyang, A. dos Santos, M.E.Taylor, K.Drickamer, *Glycobiol.*, **2013**, 23, 853-864
27. B. Belardi, C.R. Bertozzi, *Chemistry & Biology*, **2015**, 20, 983-993
28. T.K. Dam, C.F. Brewer, *Glycobiol.*, **2010**, 20, 270-279
29. J. Tanaka, A.S. Gleinich, Q. Zhang, R. Whitfield, K.Kempe, D.M.Haddleton, T.P.Davis, S.Perrier, D.A.Mitchell, P.Wilson, *Biomacromolecules*, **2017**, 18, 1624–1633
30. J.C. Paulson, O.Blixt, B.E.Collins, *Nat.Chem.Bio.*, **2006**, 2, 238-248
31. J.C. Howing, G.J. Wilson, G.D. Brown, *Cell. Microbiol.*, **2014**, 16, 185-194
32. B. Neves, M. Lopes, M. Cruz, Maria, *Protein Kinases, Chap. 6*, **2012**, 123-164
33. Y. van Kooyk, T.B.H. Geijtenbeek, *Nature Reviews Immunology*, **2003**, 3, 697-709
- 34.** I. M. Dambuza, G.D. Brown, *Current Opinion in Immunology*, **2015**, 32, 21–27
35. N.A. Barrett, A. Maekawa, O.M. Rahman, K.F. Austen, Y.Kanaoka, *J. Immunol.*, **2009**, 182, 1119–1128
36. S.T. Chen, Y.L. Lin, M.T. Huang, M.F. Wu, S.C. Cheng, H.Y.Lei, C.K. Lee, T.W. Chiou, C.H. Wong, S.L. Hsieh, *Nature*, **2008**, 453, 672–676
37. Y. van Kooyk, *Biochem. Soc. Trans*, **2008**, 36, 1478-1481
38. T. Johannssen, B. Lepenies, *Trends in Biotechnology*, **2017**, 334-346
39. H. Cai, R. Zhang, J. Orwenyo, J. Giddens, Q. Yang, C. LaBranche, D.C. Montefiori, L-X. Wang, *ACS Cent. Sci.*, **2018**, 4, 582-589
40. N. Reimers, A.Homann, B.Höschler, K.Langhans, R.A.Wilon, C.Pierrot, J.Khalife, C.G.Grevelding, I.W.Chalmers, M.Yazdanbakhsh, K.F.Hoffmann, C.H.Hokke, H.Haas, G.Schramm, *PLOS Neglected Tropical Diseases*, **2016**,9, 1-20
41. M.I. Araujo, B.S.Hoppe, M.MedeirosJr, E.M.Carvalho, *Mem.Inst. Oswaldo Cruz, Rio de Janeiro*, **2004**, 99, 27-32
42. L.M. Kuijk, I.van Die, *Life*, **2010**, 62, 303-312

43. C.H.Smit, A. Homann, V.P. van Hensbergen, G. Schramm, H. Haas, A. van Diepen, C.H. Hokke, *Glycobiol.*, **2015**, 25, 1465-1479
44. S.J. Parker-Manuel, A.C.Ivens, G.P.Dillon, R.A.Wilson, *PLoS.Negl.Trop.Dis.*, **2011**, 5, e1274
45. J.M.Fitzpatrick, E.Peak, S.Perally, I.W.Chalmers, T.P.Yoshino, A.C.Ivens, K.F.Hoffmann, *PLoS.Negl.Trop.Dis.*, **2009**, 3, e543
46. H.Stepan, M.Pabst, F.Altmann, H.Geyer, R.Geyer, E.Staudacher, *Glycoconj. J.*, **2012**, 29, 189-198
47. E. Staudacher, *Molecules*, **2015**, 20, 10622-10640
48. A. van Diepen, A-J van der Plas, R.P. Kozak, L. Royle, D.W. Dunne, C.H. Hokke, *Int.J.Parasit.*, **2015**, 45, 465-475
49. C.H. Smit, A. van Diepen, D.L. Nguyen, M. Wuhrer, K.F.Hoffman, A.M. Deelder, C.H. Hokke, *Mol. Cell.Proteomics*, **2015**, 14, 1750-1769
50. <sup>1</sup>T.P. Yoshino X-J. Wu, L.A. Gonzalez, C.H. Hokke, *Experimental Parasitology*, **2013**, 133, 28–36
51. <sup>1</sup>T.M.Kariuki, I.O.Farah, R.A.Wilson, *Parasit. Immunol.*, **2008**, 30, 554-562
52. A. van Diepen, C.H. Smit, L. van Egmond, N.B. Kabatereine, A.Pinot de Moira, *PLoS. Negl.Trop. Dis.*, **2012**, 6, e1922
53. M.Mandalasi, N. Dorabawila, D.F. Smith, J. Heimbürg-Molinaro, R.D. Cummings, *Glycobiol.*, **2013**, 23, 877-892
54. <sup>1</sup> R.P. Kozak, L. Royle, R.A. Gardner, D.L. Fernandes, M. Wuhrer, *Anal. Biochem.*, **2012**, 423, 119-128
55. <sup>1</sup>K.H.Khoo, S.Sarda, X.Xu, J.P.Caulfield, M.R. McNeil, S.W. Homans, H.R. Morris, A. Dell, *J.Biol.Chem.*, **1995**, 270, 17114-17123
- 56.** B.G. Davis, A.J. Fairbanks, *Carbohydrate Chemistry*, Oxford University Press, **2002**
- 57.** P.G. Wuts, T.W. Greene, *Greene's protective groups in organic synthesis*, John Wiley & Sons, **2006**
58. C. Pedersen, J. Olsen, A. Brka, M. Bols, *Chem. Eu. J.*, **2001**, 17, 7080-7806
59. I. Cumpstey, *Org. Biomol. Chem.*, **2012**, 10, 2503-2508
60. L.A. Marcaurelle, C.R. Bertozzi, *Glycobiol.*, **2012**, 12, 69-77
61. R. Gutiérrez Gallego, G. Dukziak, U. Kragl, C. Wandrey, J.P. Kamerling, J.F.G. Vliegthart, *Biochimie*, **2003**, 85, 275–286

62. W. Kinzy, R.R. Schmidt, *Carbohydr. Res.*, **1987**, 164, 265–276
63. P.J. Garegg, M. Haraldsson, H. Lonn, T. Norberg, *Glycoconjugate J.*, **1987**, 4, 231–238
64. N. Mathieux, H. Paulsen, M. Meldal, K. Bock, *J. Chem. Soc., Perkin Trans. 1*, **1997**, 2359–2368
65. J. Geiger, B.G. Reddy, G.A. Winterfeld, R. Weber, M. Przybylski, R.R. Schmidt, *J. Org. Chem.*, **2007**, 72, 4367–4377
66. M.C. Galan, A.P. Corfield, *Biochem. Soc. Trans.*, **2010**, 38, 1368–1373
67. C. Brocke, H. Kunz, *Synthesis*, **2004**, 4, 525–542
68. M. Hollinger, F. Abraha, S. Oscarson, *Carbohydr. Res.*, **2011**, 346, 1454–1466
69. D. Benito-Alifonso, R.A. Jones, A-T. Tran, H. Woodward, N. Smith, M.C. Galan, *Beilstein J. Org. Chem.*, **2013**, 9, 1867–1872
70. J. Xia, J.L. Alderfer, K.L. Matta, *Bioorg. Med. Chem. Lett.*, **2000**, 10, 2485–2847
71. J. Xia, J. Xue, R.D. Locke, E.V. Chandrasekaran, T. Srikrishnan, K.L. Matta, *J. Org. Chem.*, **2006**, 71, 3696–3706
72. L.L. Lairson, B. Henrissat, G.J. Davies, S.G. Withers, *Annu. Rev. Biochem.*, **2008**, 77, 521–555
73. 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.G. Wang, *Chem. Sci.*, **2015**, 6, 5652–5661
74. O. Blixt, N. Razi, *Methods in enzymology*, **2006**, 415, 137–153
75. S. Koizumi, *Handbook of Carbohydrate Engineering, Chap. 10*
76. P.M. Danby, S.G. Withers, *ACS Chem. Biol.*, **2016**, 11, 1784–1794
77. R.P. Wilbers, L.B. Westerhof, K. van Noort, K. Obieglo, N.N. Driessen, B. Everts, S.I. Gringhuis, G. Schramm, A. Goverse, G. Smant, J. Bakker, H.H. Smits, M. Yazdanbakhsh, A. Schots, C.H. Hokke, *Sci. Rep.*, **2017**, 7, 45910
78. W. Peng, J. Pranskevich, C. Nycholat, M. Gilbert, W. Wakarchuk, J.C. Paulson, N. Razi, *Glycobiol.*, **2012**, 22, 1453–1464
79. <sup>1</sup>A. Leppänen, P. Mehta, Y-B. Ouyang, T. Ju, J. Helin, K.L. Moore, I. van Die, W.M. Canfield, R.P. McEver, R.D. Cummings, *J. Biol. Chem.*, 1999, 274, 24838–24848.
80. K.M. Koeller, M.E. Smith, R-F. Huang, C-H. Wong, *J. Am. Chem. Soc.*, **2000**, 122, 4244–4245
81. S. Park, I. Shin, *Angew. Chem. Int. Ed.*, **2002**, 41, 3180–3182



82. D.N. Wang, S.Y. Liu, B.J. Trummer, C. Deng, A.I. Wang, *Nat. Biotechnol.*, **2002**, 20,275-281
83. S. Fukui, T. Feizi, C. Galustian, A.M. Lawson, W.G. Chai, *Nat. Biotechnol.*, **2002**, 20, 1011-1017
84. L. Wang, R.D. Cummings, D.F. Smith, M. Huflejt, C.T.Campbell, J.C. Gildersleeve, J.Q. Gerlach, M. Kilcoyne, L. Joshi, S. Serna, N-C. Reichardt, N. Parera Pera, R.J. Pieters, W. Eng, L.K. Mahal, *Glycobiology*, **2014**, 24, 507-517
85. E-H. Song, N.LB. Pohl, *Current Opinion in Chem. Biol.*, **2009**, 13, 626-632
86. S.Yan, S.Serna, N-C.Reichardt, K.Paschinger, I.B.H.Wilson, *J.Bio.Chem.*, **2013**, 288, 21015-21028
87. A.Beloqui, J.Calvo, S.Serna, S.Yan, I.B.H. Wilson, M. Martin-Lomas, N.C. Reichardt, *Angew.Chem.Int.Ed.*, **2013**, 52, 7477-7481
88. J.Y. Hyun, J. Pai, I. Shin, *Acc. Chem. Res.*, **2017**, 50, 1069-1078
89. I. Shin, S. Park, M-r. Lee, *Chem. Eur. J.*, **2005**, 11, 2894-2901
90. P-H.Liang, C-Y.Wu, W.A. Greenberg, C-H. Wong, *Current Opinion in Chemical Biology*, **2008**, 12, 86-92
91. Y. Guo, H. Feinberg, E. Conroy, D.A. Mitchell, R. Alvarez, O. Blixt, M.E. Taylor, W.I.Weis, K. Drickamer, *Nature Structural & Molecular Biology*, **2004**, 11, 591-598
92. S.J. van Vliet, E. van Liempt, E. Saeland, C.A. Aarnoudse, B. Appelmelk, T. Irimura, T.B.H. Geijtenbeek, O. Blixt, R. Alvarez, I. van Die, Y. van Kooyk, *Int. Immunol.*, **2005**, 17, 661-669
93. L. Medve, S. Achilli, S. Serna, F. Zucotto, N. Varga, M. Thépaut, M. Civera, C. Vives, F. Fieschi, R. Reichardt, A. Bernardi, *Chem. Eur. J.*, **2018**
94. M.L. Mickum, N.S. Prasanphanich, X. Song, N. Dorabawila, M. Mandalasi, Y. Lasanajak, A. Luyai, W.E. Secor, P.P. Wilkins, I. van Die, D.F. Smith, A.K. Nyame, R.D. Cummings, C.A. Rivera-Marrero, *Infect.Immun.*, **2016**, 84, 1371-1386



## **2 Scope and Objectives**



## 2. Scope and objectives of thesis

It is clear that the glycans of *S.mansoni* play an important role in the helminth's infection and survival mechanism and mounting evidence suggests that they contribute to the parasite's evasion of the host's immune system by hijacking C-type lectin receptors (CLRs). Therefore, their immunomodulatory properties present an interesting potential as leads for the development of carbohydrate based drugs for treatment of immuno-compromised patients with autoimmune diseases, allergies or cancer. However, most studies focus on the parasites' *N*-glycans or the antigenic motifs alone and only few describe the relevance of the *O*-glycans and their interactions with CLRs.

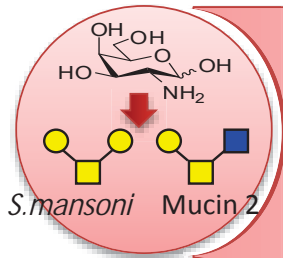
The international training network Immunoshape sought to investigate CLR selectivity and improve targeting for therapeutic applications by developing multivalent glycan mimetics. To this end, we aimed to provide structures missing from the existing glycan libraries and to propose new immunomodulatory glycomimetics with high potential for CLR targeting. Specifically, we aimed to provide a library of *O*-glycans and mimetics based on the two predominant *O*-glycan cores observed in the infectious stages of schistosomiasis caused by *S.mansoni*, the mucin 2 core and the *S.mansoni* core. Synthetic procedures were designed to enable the rapid scale-up of hits for the functional solution-phase assays and the diversification and simplification of structures for the lead generation.

For this, we endeavoured the chemical synthesis of the mucin 2 core and the *S.mansoni* specific core. As the synthesis of *S.mansoni* core was novel, we had to evaluate and optimize the necessary procedures. We first focused on synthesizing the cores functionalized with an aminopentyl linker at the reducing end for microarray studies of the *O*-glycan structures. The use of a Trichloroethylcarbamate (Troc) for the protection of the hexosamines functions will allow a smooth deprotection with zinc dust or LiOH and permit the tagging with mono and difluoro-acetamide or <sup>13</sup>C-labeled acetic anhydride for the preparation of specific probes for NMR and mass spectrometry based glycan analysis.

A library of *O*-glycans containing the characteristic LN, Le<sup>x</sup>, LDN, LDN-F epitopes would be obtained by enzymatic elongation using recombinant glycosyltransferases commercially available or prepared in house. In particular, we aimed to optimize the expression and purification of the bacterial acetylglucosaminyltransferase LgtA and evaluate the suitability of the enzyme for the synthesis of parasitic *O*-glycans structures. Additionally, mutant enzymes

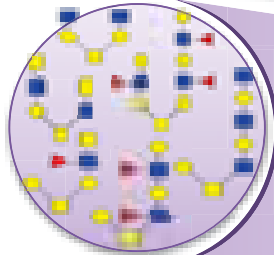
such as HP-FucT from *H. pylori* or bovine DMGalT would be evaluated for the insertion of unnatural sugars units containing an azido-functionality into the O-glycans. This functionality would allow rapid generation of glycomimetics by copper-click catalyzed reaction with various alkynes as previously described for the synthesis of sialic acid mimetics.

Employing microarrays, the library of parasitic O-glycans and mimetics would be assayed against a panel of fluorescently labelled CLRs provided by the Immunoshape network partners. As well as confirming known interactions between specific epitopes and CLRs, additional information on CLR specificity was anticipated to be revealed. Ligands displaying the highest affinity would be scaled up synthetically and labelled isotopically for an NMR study of carbohydrates - CLR binding interactions.



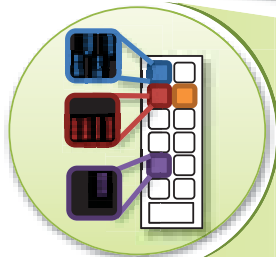
## Chemical synthesis

- Synthesis of *S.mansoni* core
- Synthesis of mucin core 2



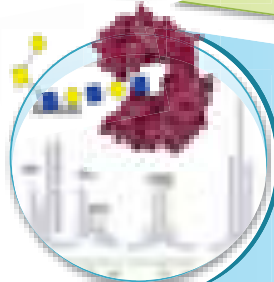
## Enzymatic elongation

- Optimize expression and purification of LgtA
- Generate a library of compounds using recombinant GTs with LN, LDN, LeX and LDNF
- Explore derivitization towards glycomimetics



## Microarray studies

- Print library of parasitic O-glycans on-chip
- Evaluate the library of compounds against CLRs
- Identify strongest binder



## Solution assays

- Scale up synthesis of strongest lectin binder identified in microarrays
- NMR studies of protein-glycan interactions





## **3 Results and Discussion**

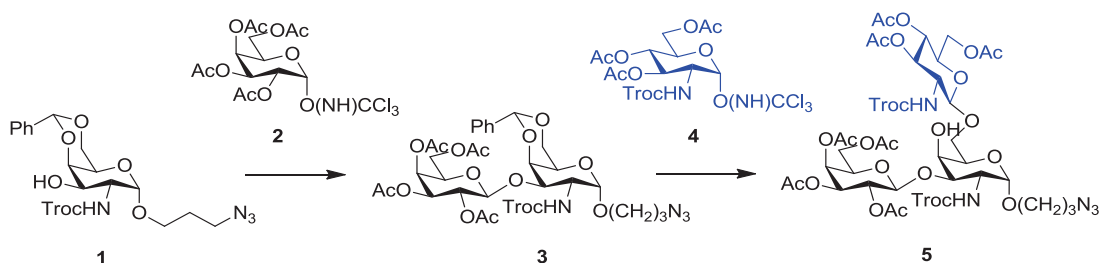


### 3. Results and discussion

#### 3.1 Synthesis of O-glycan cores

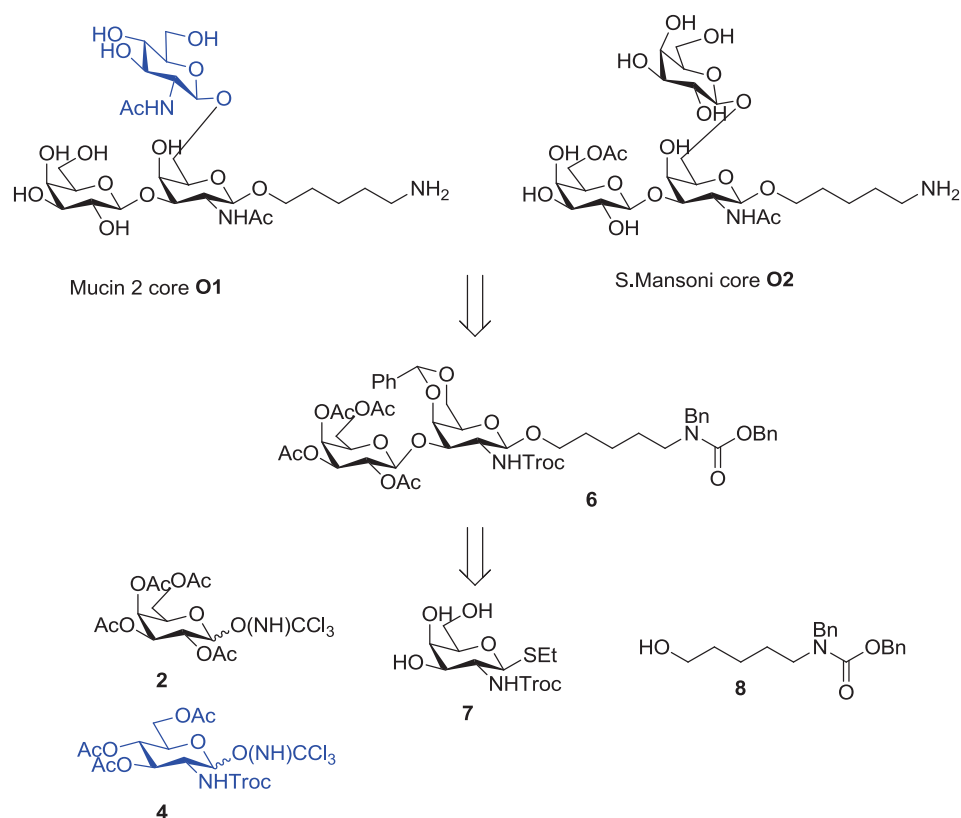
##### 3.1.1 Synthetic strategy

Of the *S.mansoni* core and the mucin core 2 O-glycans, only the mucin core 2 has been previously synthesized. Convergent strategies have efficiently been used to synthesize the core 2 as  $\alpha$ -amino acid glycoside, a fluorinated derivative, a nitrophenyl glycoside and as an aminopropyl glycoside.<sup>1,2,3</sup> The latter was illustrated by Benito-Alifonso *et al.* in the synthesis of the core 2.<sup>4</sup> The 3-azidopropyl- $\alpha$ -linked galactosamine intermediate **1** was galactosylated using the trichloroacetimidate donor **2** to yield the disaccharide **3**. After deprotection of the benzylidene acetal, **3** was then regioselectively glycosylated with the imidate **4** to yield the protected trisaccharide core **5** (Scheme 4).



**Scheme 4. General strategy for the synthesis of the core 2 reported by the Galan group**

Considering the conserved structural features (Gal- $\beta$ 1,3GalNAc) between the mucin core 2 **O1** and the *S.mansoni* core **O2**, a similar synthetic strategy could be applied to obtain both cores from the common disaccharide intermediate **6** (Scheme 5). Therefore the synthesis was envisaged to stem from 3 monosaccharide building blocks: **7** and the donors **2** and **4**.



#### Scheme 5. Retrosynthesis of mucin core 2 and *S.mansoni* core

Due to the stability of thioglycosides under many protecting group manipulations, the ethylthiogalactoside **7** was chosen as a precursor for our synthesis, allowing in addition a late-stage functionalisation of the reducing end. Thus, the insertion of an aminopentyl linker or the attachment of an amino acid could be considered for the immobilization to microarray surfaces and on proteins. The aminopentyl linker is a standard linker used for glycan immobilization but requires protection to avoid the terminal amine from reacting during synthetic manipulations. In the case of the Galan group, an azido-functionality was used. Our synthesis employed benzyl and Cbz protecting groups (seen in **8**) as these would facilitate purification of intermediates by UV-based chromatography. As regioisomers may arise during enzymatic elongation, the chromophores would be especially useful in separation by HPLC-UV.<sup>5</sup> Hydrogenation of these groups would yield final compounds for conjugation.

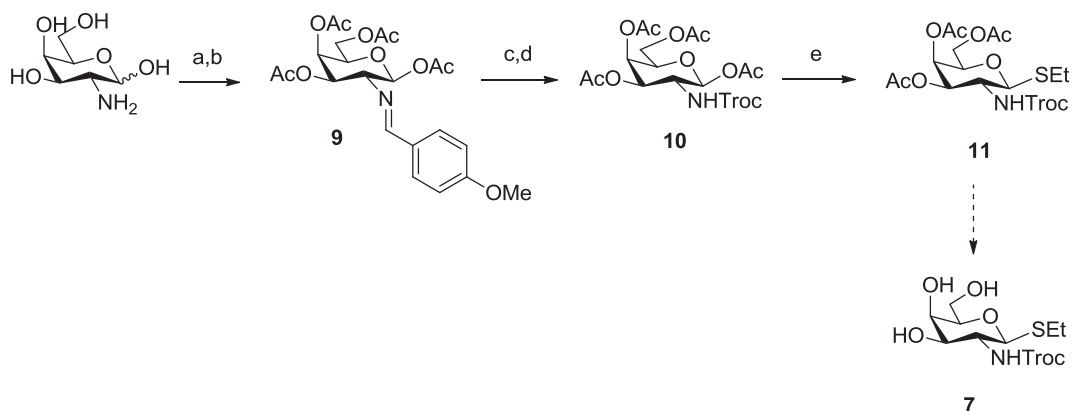
The 2,2,2 trichloroethylcarbamate (Troc) group was chosen as the amine protecting group of the hexosamines for several reasons. Firstly, it protects the amine without significantly affecting the reactivity of the molecule. Indeed, an electron withdrawing azido-functionality at C2 such as that reported by Hollinger *et al.* would likely reduce the nucleophilicity of the

neighbouring 3-OH and reduce the overall reactivity of the molecule.<sup>3,6,7</sup> Considering glycosylation at this position was necessary at the first GalNAc moiety of both cores, this was to be avoided. Secondly, the Troc group was chosen for its neighbouring group participation during which oxazoline formation ensures high  $\beta$ -selectivity upon glycosylation.<sup>8</sup> This stereoschemistry is observed in the mucin core 2 target **O1** as GlcNAc $\beta$ -1,6GalNAc. Thirdly, the relatively small size of the Troc group favours higher glycosylation yields than when using a bulkier protecting group such as N-phthalimide.<sup>8</sup> Finally, its removal by reductive elimination was reported to be facile, mild, and selective.<sup>9</sup> As such, this would facilitate tagging of hexosamines of selected, potent, glycomimetics with fluoro-acetamide or <sup>13</sup>C-labeled acetic anhydride for extended NMR and mass spectrometry based glycan analysis.<sup>10</sup>

Regioselective protection of 4-OH and 6-OH via benzylidene acetal formation would afford a suitable intermediate for glycosylation at 3-OH and yield the first disaccharide building block **6.4**. Glycosylation in 6-OH using either a GlcNAc donor or a Gal donor would afford the fully protected mucin core 2 and *S.mansoni* core respectively. Finally, deprotection of the compounds followed standard procedures including Troc removal, reacetylation, ester hydrolysis and hydrogenation of benzyl and CBz groups in the linker.

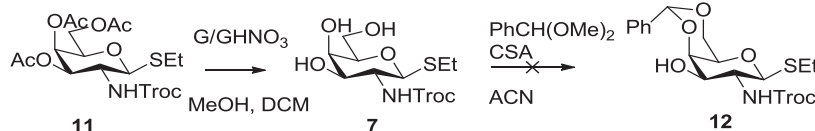
### 3.1.2 Synthesis of the GalNAc building block

To obtain the targeted triol **7**,  $\beta$ -thioglycoside **11** can be formed from **10** but the donor should have the more reactive  $\beta$ -configuration for the reaction to proceed smoothly.<sup>8,11</sup> This was achieved by initial protection of the amine moiety galactosamine as an imine using p-anisaldehyde in 1M NaOH(aq) in 91% yield (Scheme 6).<sup>12</sup> DMAP catalyzed acetylation with acetic anhydride in pyridine yielded **9** in 53% yield. Compared to the study on glucosamine by Appelt *et al.* this route afforded 13% lower yield.<sup>13</sup> The lower reactivity of the axial 4-OH of galactose probably also affects the yield.<sup>14</sup> An alternative acetylation condition using sodium acetate only afforded 16% yield. The imine of **9** was then successfully hydrolyzed in 5M HCl(aq) to yield the free amine (90% yield) which was reacted with TrocCl in DCM (75%). Finally, the obtained carbamate **10** was converted to the thiogalactosamine **11** using ethanethiol in 56% yield, 100%  $\beta$ .



**Scheme 6.** Synthesis of the N-Troc protected thiogalactoside *Erreur ! Source du renvoi introuvable.*; **a)** p-anisaldehyde, 1M NaOH(aq), 91%; **b)** Ac<sub>2</sub>O, DMAP, pyridine, 53%; **c)** 5M HCl(aq), 90%; **d)** TrocCl, Et<sub>3</sub>N, pyridine, 75%; **e)** ethanethiol, BF<sub>3</sub>.Et<sub>2</sub>O, DCM, 56%

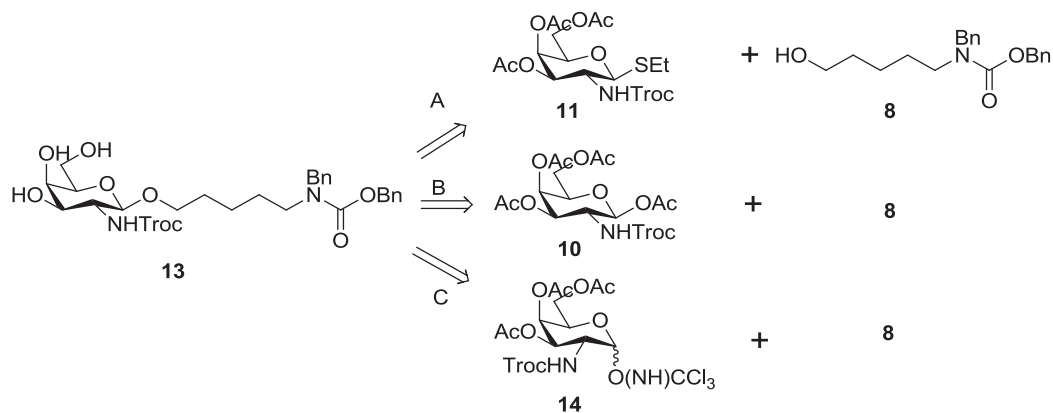
The thioglycoside **11** was then deacetylated. Preservation of the Troc protecting group required careful balancing of the basic conditions needed for deacetylation and those detrimental to the Troc group.<sup>8</sup> Conditions described by Ellervik using a guanidine nitrate (G/GHNO<sub>3</sub>) solution afforded the cleanest deprotection of the hydroxyls but came with the problems of resulting guanidinium salts.<sup>15</sup> Upon carrying out the following benzylidene reaction using conventional conditions i.e. benzaldehyde dimethyl acetal in acetonitrile with catalytic amounts of camphor sulfonic acid, the reaction did not proceed (Scheme 7).



**Scheme 7.** Deacetylation of the thiol **11** and failed benzylidene acetal formation to **12**

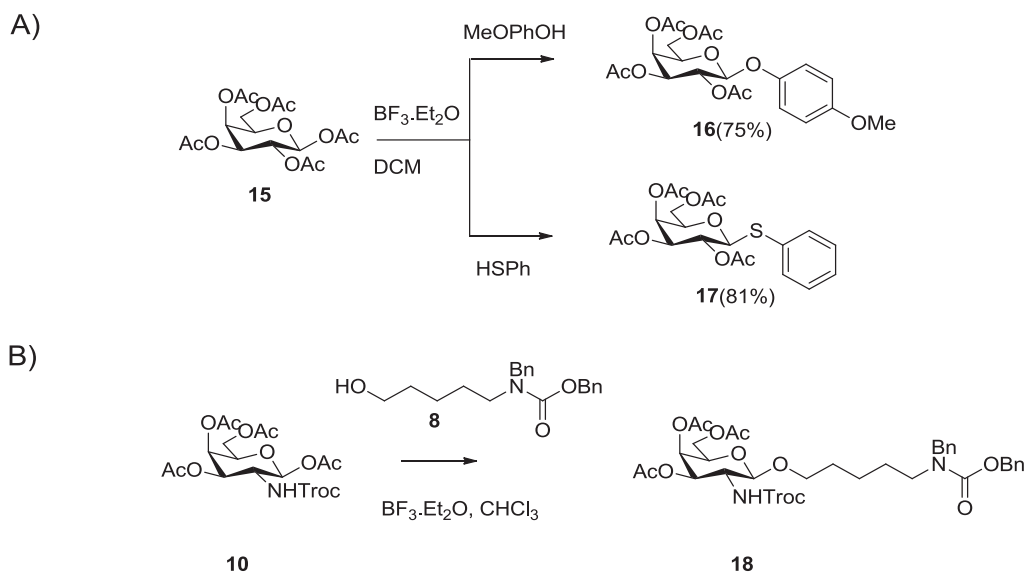
This was due to the basic conditions conferred by the partially soluble guanidinium salts which were carried over from the deacetylation reaction. As a consequence, the acidic conditions required for acetal formation were not achieved and the reaction did not proceed. Considering the large excess of guanidine used for the deacetylation, a counter excess of camphor sulfonic acid to ensure acidic conditions for the acetalation reaction did not seem like a viable choice. To remove the guanidinium salts, a suitable solvent in which the compound **11** was soluble but the salts were not was necessary, in order to filter the compound from the salts. The polarity of **11** meant acetonitrile was the only solvent in which the compound was fully soluble while the salts remained only partially soluble. Unfortunately, even partial removal of the salts hindered the following acetal reaction.

The synthesis was therefore revised to include the insertion of the linker at an earlier stage, under the hypothesis that the new target intermediate **13** might improve in solubility in organic solvents and consequently be separable from the guanidium salts. Three synthetic pathways were considered to obtain the new building block **13** (Scheme 8).



**Scheme 8. Retrosynthesis of the new building block 13 containing the linker**

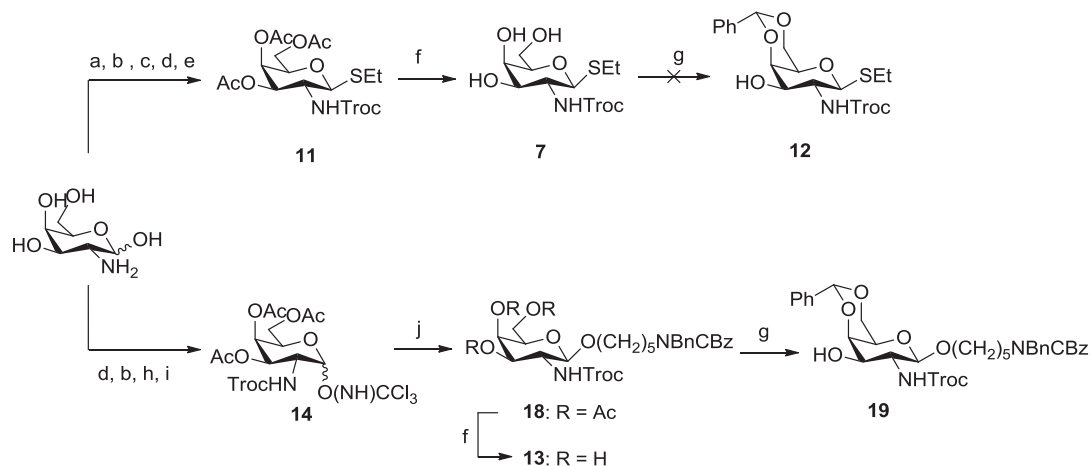
In theory, **13** could be obtained from **11** under standard glycosylation conditions using NIS and TMSOTf (Scheme 8, A). However, the synthesis could conceivably be shortened by inserting of the linker directly from the tetraacetylated **10** - thus bypassing the thiointermediate **11** (Scheme 8, B). Such a procedure was reported successfully for Ohlsson and Magnusson in the synthesis of galabiosyl donors.<sup>16</sup>  $\beta$ -pentacetylated galactose **15** was directly transformed into the 4-methoxyphenyl  $\beta$ -galactopyranoside **16** and under the same conditions as those used for the  $\beta$ -thiogalactopyranoside **17** in 75% yield. However, only a poor 25% yield of target material **18** was obtained when the same strategy was applied using **10**. (Figure 11, B).



**Figure 11. Insertion of the linker 8 from the peracetylate 10**

The success of the reaction was therefore a function of the donor and could be explained by the difference in nucleophilicity between the thiol and the alcohol. Therefore we considered placing the reaction under more forceful conditions by microwaving the reaction at 100°C for 10 mins and increasing the donor equivalents from 1.2 to 2. As excess acid was not previously conducive to successful reaction, the equivalents used for the microwave reaction was lowered from 5 to 1.4. Unfortunately, this method was not viable either as the target material was obtained in a crude yield of 13%. Despite our best efforts, purification of the product was complicated by several by-products of similar polarity which were observed by TLC. It was noted that although a study by Khamsi *et al.* also showed that yields as high as 70% can be achieved for this type of reaction, they are achieved as a 1:2 mix of  $\alpha/\beta$ .<sup>17</sup> Therefore this route was abandoned.

The use of trichloroacetimidate donor **14** also afforded a shorter synthetic route as the synthesis of **14** only requires 4 synthetic steps compared to the 5 steps for the synthesis of the thiodonor **11**, bypassing the circuitous imine formation (Scheme 9).<sup>18</sup> The donor **14** was therefore reacted with linker **8** to the desired **18** in a satisfactory 70% with high  $\beta$ -selectivity. Upon deacetylation using the guanidine/guanidium nitrate solution, the resulting triol **13** was soluble in DCM and could be separated from the insoluble guanidium salts by filtration. The triol **13** was then successfully protected to **19** using  $\text{PhCH}(\text{OMe})_2$ , CSA in ACN in 70% yield.



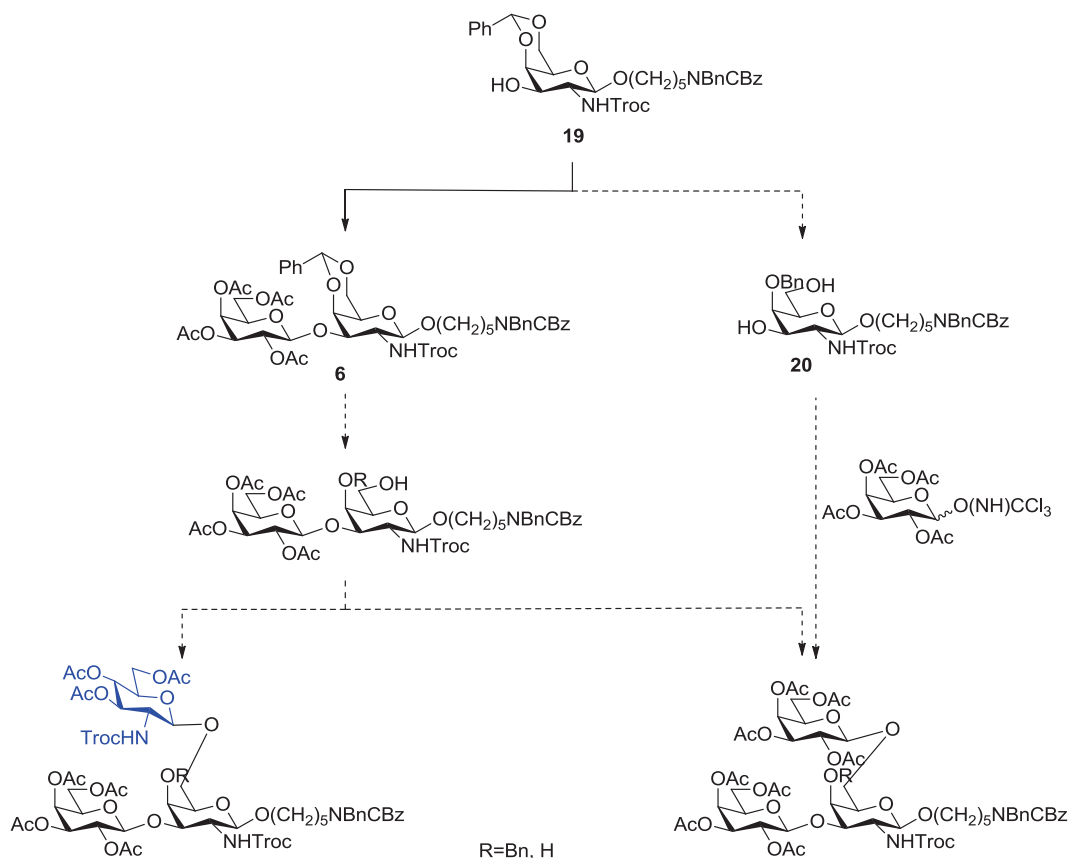
**Scheme 9.**Thiois TCA donor route and outcome of subsequent benzylidenereaction. Reagent and conditions: a) *p*-anisaldehyde, 1M NaOH; b)  $\text{Ac}_2\text{O}$ , pyridine; c) 5M HCl(aq),  $\text{H}_2\text{O}$ ; d) TrocCl,  $\text{Et}_3\text{N}$ , pyridine; e) ethanethiol, 1M  $\text{BF}_3 \cdot \text{Et}_2\text{O}$ , DCM; f) G/ $\text{GHNO}_3$ ;g)  $\text{PhCH}(\text{OMe})_2$ , CSA, MeCN;h)  $\text{N}_2\text{H}_4 \cdot \text{OAc}$ , DMF; i) trichloroacetonitrile, DBU, DCM; j) benzyl(5-hydroxypentyl)carbamate, TMSOTf, DCM, -40°C



The synthesis of the O-glycan cores was therefore redesigned to stem from the donor **14**. While this strategy provided a shorter synthesis overall, it lacks the flexibility of functionalizing the reducing end of the cores with different moieties, eg. for the synthesis of mass spectrometry standards. Changing the ethanethiol for a more hydrophobic group such as phenylthiol may aid with the dissolution of a triol resulting from deacetylation and thus allow for successful benzylidene acetal protection. However at this stage we decided to sacrifice the flexibility for the completion of the synthesis of glycans which could be conjugated to the microarray for screening with CLRs.

### 3.1.3 Glycosylation to mucin core 1

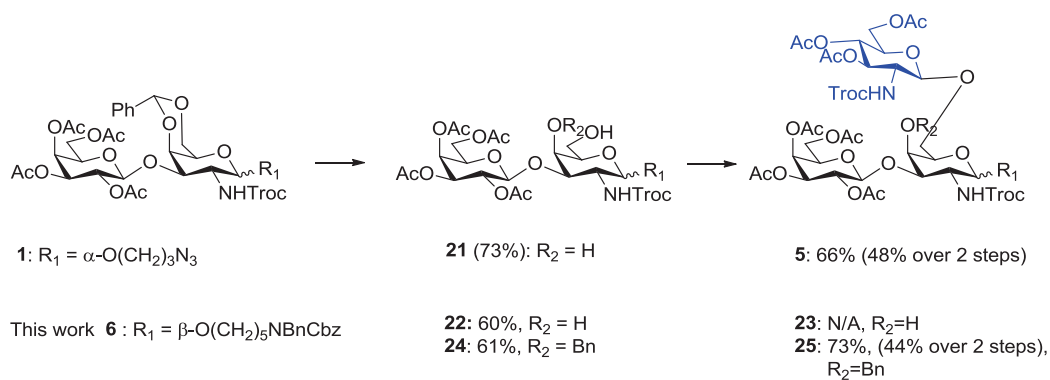
From **19**, the benzylidene acetal could be regioselectively opened to give the 3-, 6- diol **20** using borane and catalytic amounts of trifluoromethanesulfonate.<sup>19</sup> We reasoned that bi-glycosylation using an excess of sufficient galactose donor might afford the *S.mansoni* core (Scheme 10). As the synthesis of this core remained unreported, it was not surprising that this reaction had no literary precedents. However, the mucin core 2 would not be obtained this way. Instead, 3-glycosylation of the protected GalNAc glycoside using a galactose donor is more commonly reported as it affords the intermediate core 1 as a stepping stone towards core 2. As the *S.mansoni* core also shares the common feature of core 1, **19** was glycosylated using the galactose donor **2** to yield the disaccharide **6** in 70% yield. This is in line with the 76% reported by the Galan group.



**Scheme 10.** Possible synthetic pathways towards the cores from 19

### 3.1.4 Divergent synthesis to mucin2 and *S.mansoni* cores

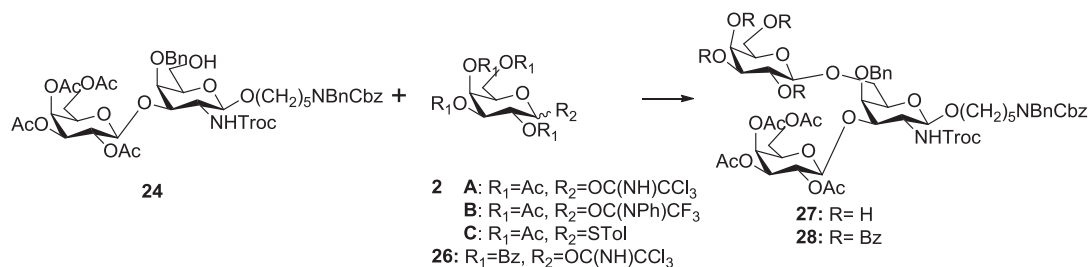
From the synthetic core 1 intermediate, Benito-Alifonso *et al.* described the hydrolysis of the benzylidene acetal using tosic acid in MeOH in 73%. The resulting diol **21** was then differentially glycosylated at C-6 over C-4 with a GlcNAc donor in 66% (Scheme 11). In contrast, hydrolysis of our compound **6** to **22** only reached 60% yield under the same reaction conditions and the following glycosylation to the target **23** also gave a poor outcome. Low reaction temperatures such as the ones used for this reaction ( $-78^{\circ}\text{C}$ ) have been described to drive the chemo- and regio-selective glycosylation of the primary 6-OH hydroxyl over the secondary 4-OH.<sup>2</sup> However, this was not observed in our trial as a mix of mono- and bi-substituted compounds (observed by MALDI-TOF MS), inseparable by flash column chromatography, was obtained.



### Scheme 11. Comparison of yields for the C-6 glycosylation of the mucin core 2

Therefore, the benzylidene acetal of **6** was regioselectively opened to C-4 producing alcohol **24** in 61% on yield average using  $\text{BH}_3 \cdot \text{THF}$  and TMSOTf at  $0^\circ\text{C}$ . Glycosylation of **24** under catalytic TMSOTf using the donor **4** afforded the mucin core 2 **25** in a 73% yield carrying an additional benzyl group as chromophore to assist UV-based HPLC separation. The synthesis of the mucin core 2 was achieved in 45% over two steps from **6**, in a similar yield described by Benito-Alifonso *et al.* who obtained the core 2 **5** in 48%.

For the synthesis of the *S. mansoni* core, the glycosylation of **24** using the donor **2** afforded **27** in a low yield of 40% yield (Table 2, entry 1). Upon MALDI TOF MS and NMR analysis, an acyl migration to the free 6-OH of the acceptor was observed to occur on approximately 30% of the material, thus reducing the relative quantities of acceptor available for reaction. In an effort to reduce the observed side-reactions, the effects of three parameters on the glycosylation yield were investigated: the leaving group, reaction temperature and time, under the hypotheses that the rate of acylation might be slowed down for glycosylation to complete effectively.



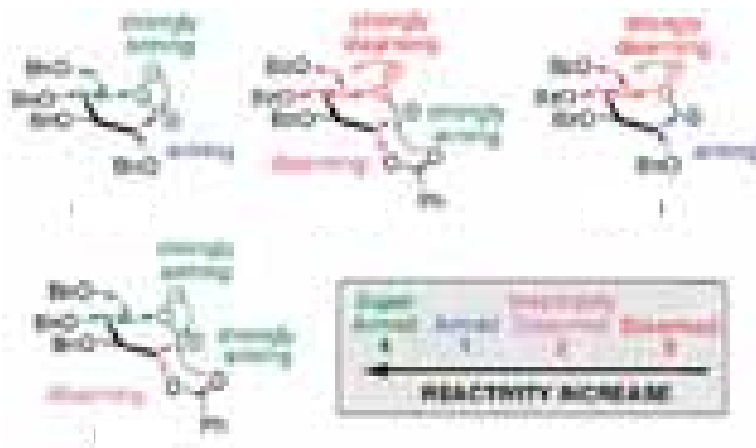
Entry	Donor	Conditions	Conversion(%) <sup>a</sup>
1	2A	TMSOTf (10-20%), $-40^\circ\text{C}$ , 1-3hrs	40
2	2A	TMSOTf (10-20%), $-40^\circ\text{C} \rightarrow 24^\circ\text{C}$ , 1-2hrs	<65 <sup>b</sup>
3	2A	TMSOTf (10-20%), $0^\circ\text{C} \rightarrow 24^\circ\text{C}$ , 1hr	>22
4	2A	TMSOTf (20%), $+10^\circ\text{C}$ , 0,5hr	<32
5	2A	$\text{BF}_3 \cdot \text{Et}_2\text{O}$ (10%), $-40^\circ\text{C} \rightarrow 24^\circ\text{C}$ , 0,5hr	<17

6	2B	TMSOTf (10-20%), -40°C, 1-3hrs	<30
7	2B	TMSOTf (10-20%), -40°C →24°C, 1-2hrs	<40
8	2C	TMSOTf (10-20%), NIS(1,5eq.) -40°C, 1-3hrs	<39
9	2C	TMSOTf (10-20%), NIS(1,5eq.) -40°C →24°C, 1-2hrs	N/A
10	2C	TMSOTf (10-20%), NIS(1,5eq.) 0°C →24°C, 1hr	<31
11	2C	TMSOTf (10-20%), NIS(1,5eq.) +10°C →24°C, 1-2hrs	<55 <sup>b</sup>
12	2C	BF3.Et2O(10%), NIS(1,5eq.) +10°C →24°C, 0,5hr	0
13	26	TMSOTf (10-20%), -40°C, 1-2hrs	65

**Table 2.** Glycosylation to *S.mansoni* core. a: isolated yield; b: purity unsatisfactory by NMR

All were monitored by LCMS, MALDI and TLC and evaluated by NMR however none resulted in improvement. To limit acyl migration, a benzoylated donor **26** (entry 13) was used as a bulkier protecting group would be less prone to migration.<sup>20</sup> Indeed, the reaction was observed to proceed to 65% with no detectable trace of migration, therefore improving the final yield of the deprotected target **O2** compound by 30%.

The modest yield of 65% could be due to the presence of electron-withdrawing protecting groups in the donor, making it a mildly disarmed donor. Indeed, the optimal glucose donor was described to combine the electron-donating properties of benzyl protecting groups in positions O4 and O6 with a benzoyl protecting group for an overall O2/O5 cooperative effect. The overall result provides stability of the glycosyl donor cation with an overall effect of "superarming" the donor.<sup>21</sup>



**Figure 12.** Cooperative arming and disarming effects in glycosyl donors of the gluco-series<sup>21</sup>

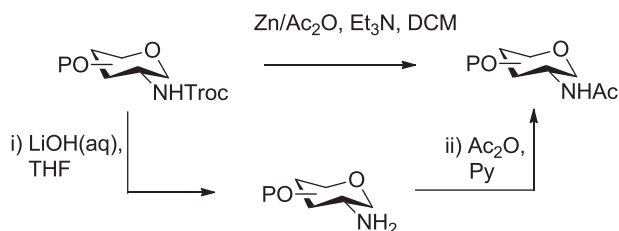
In a similar way, the benzoyl groups of **26** may have a disarming effect on the donor. Unfortunately, arming the galactose donor with EDG such as benzyl ethers is disadvantageous

to the synthesis of the cores as deprotection of the benzyl groups on the galactose arm is not orthogonal to that of the 4-OBn GalNAc and the linker. Therefore, debenzylation would lead to loss of all chromophores of the *S.mansoni* core making subsequent purifications by HPLC solely dependent on mass (something we are not equipped to do efficiently).

As sufficient material **27** was obtained, we pursued the synthesis with this intermediate. However deprotection of **28** was expected to follow the same protocol and afford similar yields. The latter would therefore be a better choice in for preparative synthesis of the *S.mansoni* aminopentyl O-glycosides.

### 3.1.5 Partial deprotection to enzymatic substrates

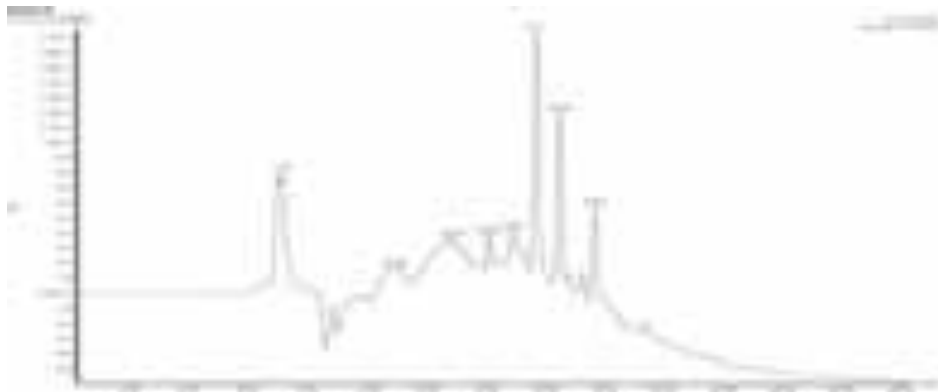
Having assembled both protected cores in satisfactory yields, the deprotection sequence to obtain compounds suitable for enzymatic reaction was undertaken. This involved the Troc removal, acetylation of the amine and deacetylation of the hydroxyl groups. As the benzyl groups did not protect key exploitable positions for enzymatic elongation, they were not targeted by the deprotection sequence at this stage of the synthesis. Moreover, they would be useful for the diode array-HPLC purification of compound mixtures which might result from enzymatic elongations. Typically, LiOH(aq) or a zinc amalgam is used to deprotect the Troc.



**Figure 13. Zinc/AcOH vs LiOH deprotection of *N*-Troc to the acetamide**

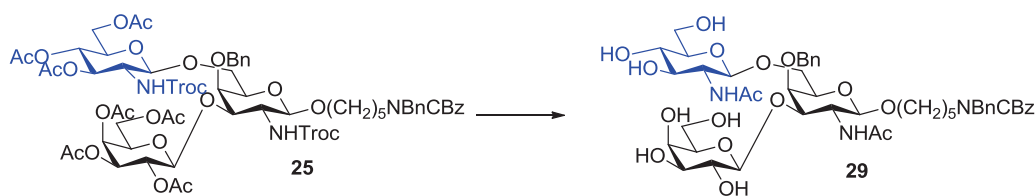
Although the use of Zinc/Ac<sub>2</sub>O system would allow for a shorter deprotection pathway-*via* the generation of the desired acetamide thus removing the reacetylation step, Tran *et al.* reported partial reduction by the metal of the azide functionality in their compounds. Mindful of the benzyl and CBz groups in our compounds which were also sensitive to reduction, the LiOH hydroxide protocol was adopted. This 3-step procedure involving Troc removal by a 1M aqueous lithium hydroxide (LiOH) solution in THF, acetylation of amine and hydroxyl functions by acetic anhydride in pyridine and deacetylation of the acetates by sodium methoxide in MeOH and is generally performed as a one-pot reaction. A key principle of the carbohydrate functional group interconversion is for the tandem steps to be relatively effortless and high yielding. Yet this was not observed to be the case at first as the final yield obtained for the

partially deprotected mucin core 2 **29** was 16% after purification by HPLC. This was much lower than the 64% reported by Tran *et al.* which included an additional azide reduction step, but is less surprising when considering the HPLC chromatogram. Indeed, this showed several peaks of which only one was the target material **29**, indicating the occurrence of adverse reactions.



**Figure 14. HPLC trace of the partial deprotection of the mucin core 2**

The other peaks showed unidentified masses by MALDI TOF and NMR for the most part. On the other hand, the isolated product was seen to contain a mass of -14 *m/z*. In the absence of mechanistic rationale for the deprotection with LiOH, a formaldehyde side-product is suspected to occur. A closer inspection into the deprotection sequence was therefore necessary. Six deprotection protocols were assessed on an analytical scale (to preserve material) and each was monitored by MALDI-TOF (Table 3).

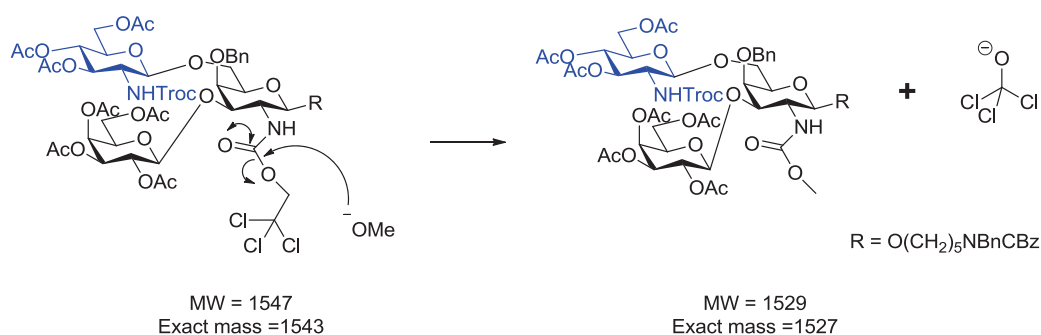


Entry	Reagents and conditions	Observations	Reference
<b>1</b>	i) 0.5 M NaOMe, MeOH, 1hr, RT ii) Ac <sub>2</sub> O, Py, RT, 18h iii) 0.5M NaOMe, MeOH, RT	Methanolysis	-
<b>2</b>	i) 1M LiOH(aq), THF, RT, 18h ii) Ac <sub>2</sub> O, Py, 18h iii) 0.5M NaOMe, MeOH	Target material Suspected formamide(s) Other impurities	[4]
<b>3</b>	i) Zn/Ac <sub>2</sub> O, Et <sub>3</sub> N, DCM sonication 3h ii) 0.5M NaOMe, MeOH	Partial reaction Suspected formamide(s) Impurities	[22],[23]
<b>4</b>	i) Zn/AcOH, Et <sub>3</sub> N, DCM ii) 0.5M NaOMe, MeOH	Partial reaction Suspected formamide(s) Impurities	[24]
<b>5</b>	i) Cd/AcOH, Et <sub>3</sub> N, DCM ii) NaOMe, MeOH	Partial reaction Impurities	[25]

	i) TBAF, THF, $\Delta$		
6	ii) Ac <sub>2</sub> O, Py, Rt, 18h	Target material	[26], [27]
	iii) NaOMe, MeOH		

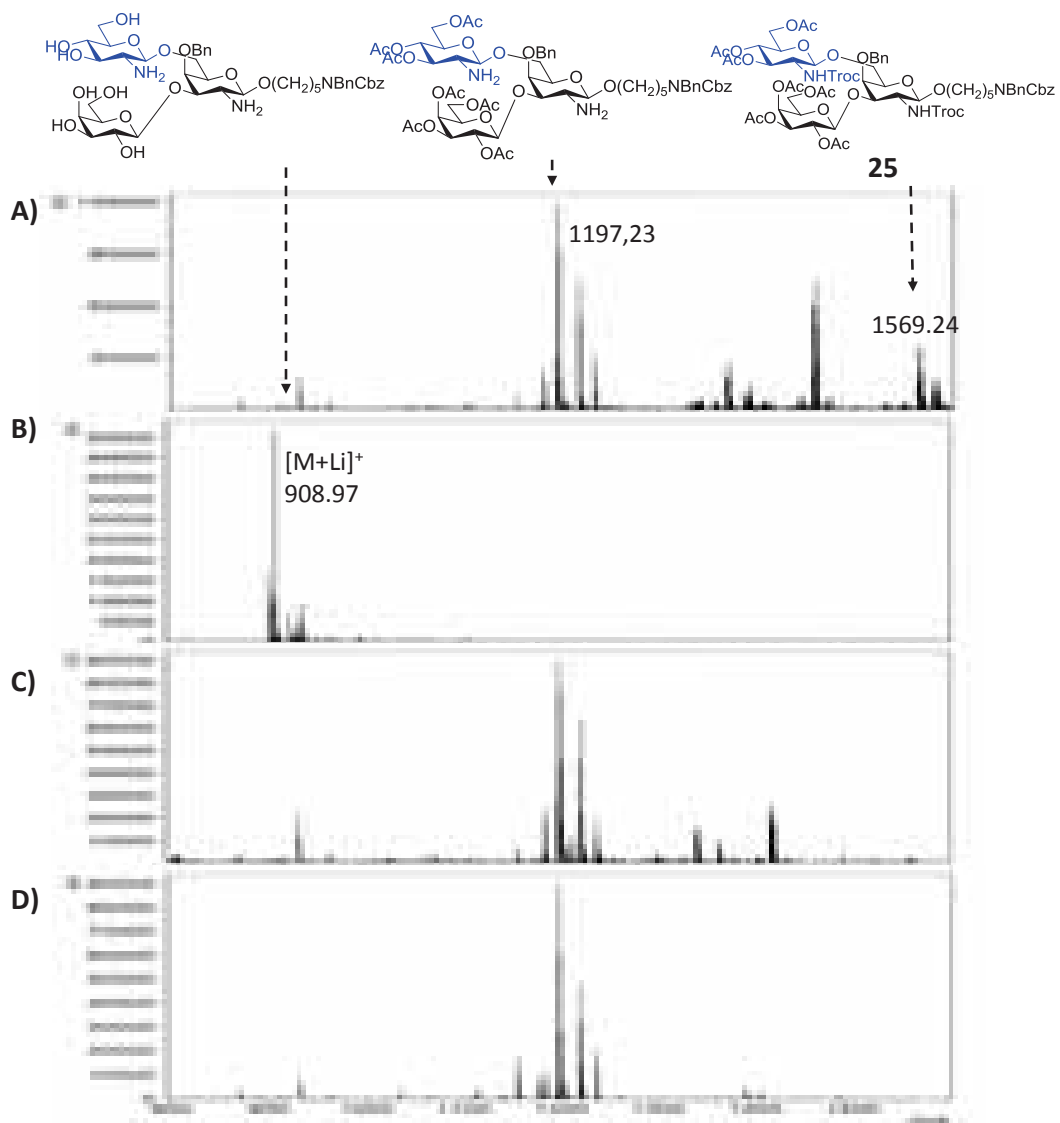
**Table 3. Deprotection of *N*-Troc trials**

We first confirmed the incompatibility of the Troc group with base and in particular with NaOMe (MeOH)(entry 1). After the first step, a mass of  $[M+Na]^+=1158$   $m/z$  was seen which was in line with methanolysis of one the trichloroethoxy group of the Troc carbamate (Scheme 12). This was anticipated given the electron-donation of the amine lone pair into the carbonyl group. Consequently, this method was not viable towards obtaining our target material and was discarded.



**Scheme 12. Proposed mechanism for methanolysis of *N*-Troc**

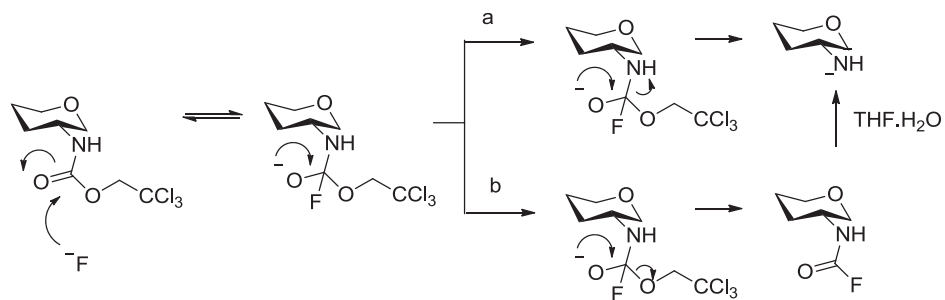
As previously mentioned, the LiOH method was accompanied by a suspected formation of the formamide instead of the targeted acetamide. This was also observed to a minor extent after the first step which was carried out in THF and at room temperature for 18h and where the deacetylated product was observed (entry 1). Also noticeable was the lithium adduct ( $[M+Li]^+=908.97$   $m/z$ ) in the MALDI spectrum, rather than the conventional sodium or potassium. In comparison to the zinc and cadmium conditions (entries 3-5), the LiOH conditions appeared much cleaner and more efficient. Indeed, partial acetylation, deprotection and even degradation were suspected to occur in entries 3-5 respectively. Cadmium was trialled as a substitute for zinc as it had been described as milder and consequently less prone to induce degradation.<sup>25</sup> Considering the outcome of the first step for entries 3-5, Zn and Cd based conditions were avoided. In comparison, the use of TBAF under reflux in THF yielded a clean spectrum of our intermediate compound.



**Figure 15. Comparative MALDI-TOF spectra of different Troc removal conditions. A) Cd:AcOH, B) LiOH; C) Zn:AcOH; D) TBAF, reflux**

The mechanism proposed by Jacquemard *et al.* includes a nucleophilic attack of the fluoride anion to give a tetrahedral intermediate. The latter could evolve in two ways: a) the amide was considered as the leaving group which led to the free amine; b) the alcohol was considered as the leaving group forming the carbamoyl fluoride.<sup>28</sup> Subsequent hydrolysis by the trace amounts of water in the THF would then provide the free amine.<sup>27</sup>

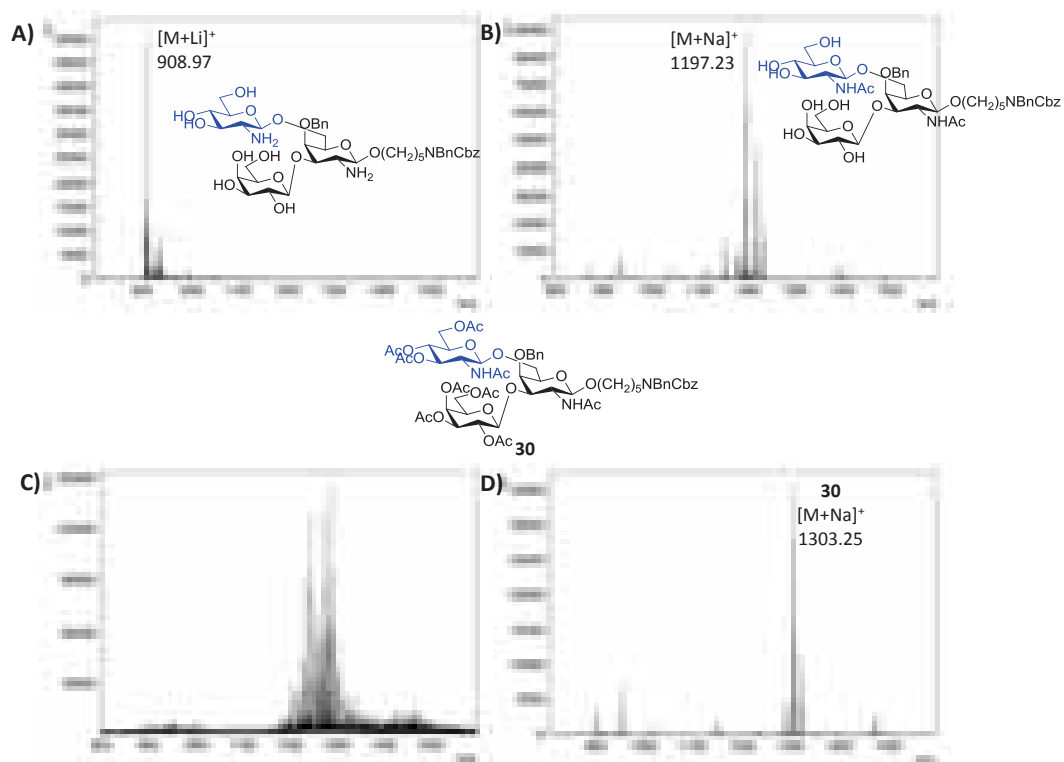




**Scheme 13. Proposed mechanism for the deprotection of Troc by TBAF**

The presence of  $-42$   $m/z$  masses was strongly suspected to be due to deacetylated hydroxyl groups, a by-product which could be remedied in the following acetylation step. The first step of both the LiOH and the TBAF conditions appearing cleanest by MALDI and LC-MS, these conditions were further investigated.

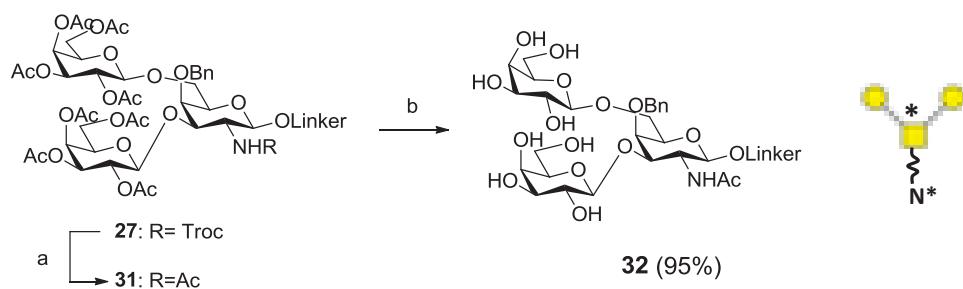
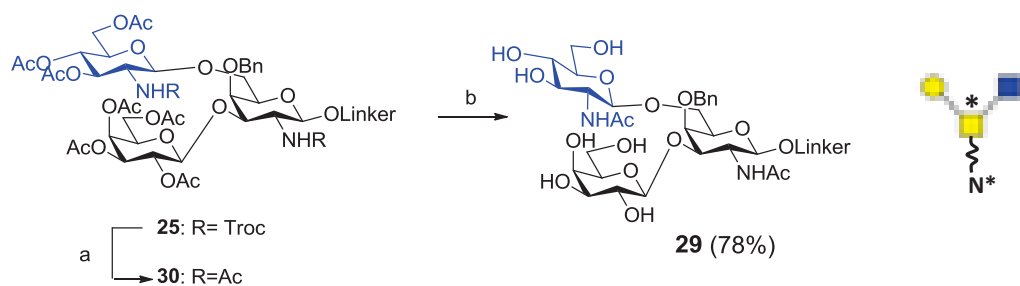
The following step involved reacetylation of the free amines in the case of the TBAF method, and also of the hydroxyls in the case of the LiOH method. After overnight reaction in excess acetic anhydride in pyridine, the reaction from the LiOH deprotection was unsuccessful. Despite the target material mass peak being observed ( $[M+Li]^+ = 1287$ ) it was found among many other  $\Delta 14$   $m/z$  derivatives (Figure 16). Addition of pyridine and a catalyst (DMAP) did not improve the MALDI TOF spectrum appearance.



**Figure 16. MALDI TOF MS of LiOH and TBAF deprotection of the Troc in the mucin core 2. A) 1M LiOH(aq), THF, 24h; B) 1M TBAF, THF, reflux, 16h; C) 0.5M NaOMe, MeOH; D) 0.5M NaOMe, MeOH**

If our hypothesis of *in situ* formylation occurring were correct, subsequent deacetylation using NaOMe would not be basic enough conditions to remove any formamide or even potential anhydride formed. This would explain complex aspect of the HPLC (Figure 14). Indeed, microwaving the crude with excess NaOMe (30eq.) at 120°C did not afford any change in the MALDI-TOF spectrum. In contrast, acetylation of the TBAF trial afforded a much cleaner spectrum. These conditions were subsequently employed for the partial deprotection of the O-glycans.

Carrying out the full TBAF protocol on a preparative scale, the mucin core 2 compound **25** was stirred in THF containing 1M TBAF for 1 hour after which full conversion was observed by MALDI TOF. The crude reaction mixture was reacylated using excess acetic anhydride in pyridine overnight to yield **30** in 78% after FCC. Finally, the acyl groups were quantitatively removed using base under standard NaOMe conditions to yield partially protected core **29** in 78% yield. The *S. mansoni* compound **27** was deprotected to **31** then **32** in 95% in the same way. Therefore the TBAF conditions were much higher yielding than the LiOH conditions employed by Benito-Alifonso *et al.* where the deprotection ranged between 60-70%.



Linker= (CH<sub>2</sub>)<sub>5</sub>NBnCBz

**Scheme 14.** Partial deprotection of the core 2 and *S. mansoni* core. Reagents and conditions: a) i) 1M TBAF, THF, reflux, 1hr; ii) Ac<sub>2</sub>O, pyridine, 18h; b) 0.5M NaOMe, MeOH, 1hr

With both the parasitic and the mucin O-glycan core in hand, we turned our attention to the enzymatic elongations for the generation of a library of parasitic-type O-glycans.

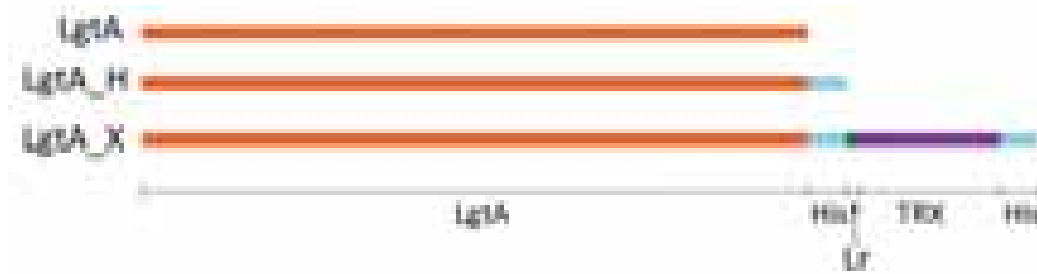
## 3.2 Recombinant expression of glycosyltransferases

### 3.2.1 LgtA: A GlcNAcT from *Neisseria meningitidis*

The library of parasitic O-glycans can be constructed by sequential action of glycosyltransferases. These enzymes catalyze the formation of glycosidic bonds between a glycosyl donor and acceptor in a highly regio- and chemo- specific way. The first enzymatic transformation we considered was the installation of GlcNAc as this was the only possible choice for the *S.mansoni* core. Despite being highly active and commercially available, the human recombinant human  $\beta$ -1,3-N-acetylglucosaminyltransferase 2 (B3GnT2) is not well suited for the preparative scale chemoenzymatic synthesis of glycomimetics. Indeed, its cost and high substrate selectivity towards lacto-N-neotetraose (LNnT) and 2,6-branched N-glycan structures greatly restricted its application.<sup>29</sup> Instead, alternative bacterial have been expressed as they are readily expressed in high amounts in *E.coli*.<sup>30</sup> Their broader substrate selectivity make them also more versatile as tools for the chemoenzymatic synthesis of parasitic O-glycans.<sup>31,32,33</sup>

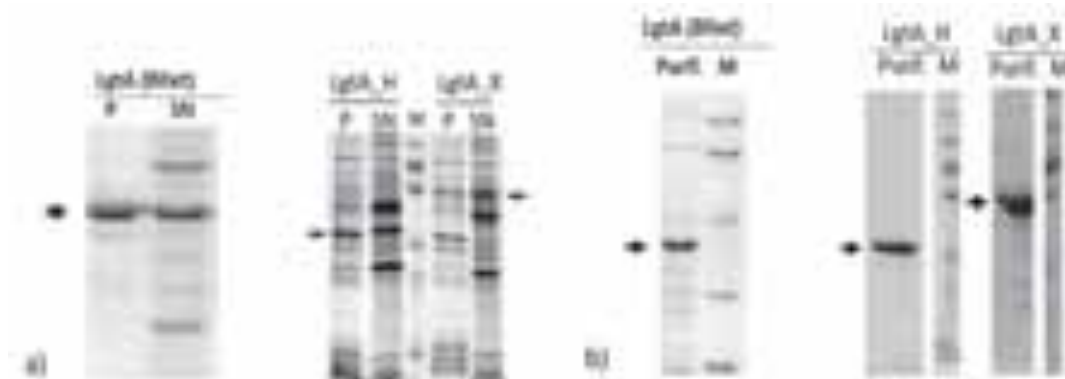
In 1999, Blixt *et al.* described the recombinant expression of the  $\beta$ 1,3GnT LgtA from *N.meningitidis* in *E.coli* for the construction of polyLN structures.<sup>34</sup> Only about 50% of the expressed enzyme corresponded to the soluble, correctly folded and functional form, with the rest of the unfolded protein trapped in inclusion bodies. It was also obtained in only partially pure form. We sought to improve the enzyme expression to obtain an affordable and readily available source of enzyme for our chemoenzymatic synthesis. We therefore had to improve the purification for the enzyme and increase its solubility. To this end, two DNA constructs were designed: *LgtA\_H* and *LgtA\_X*. (Figure 17)

The first construct, *LgtA\_H*, contained an additional hexahistidine (His) tag in the C-terminal of the amino acid sequence described by Blixt *et al.* for ease of purification by Immobilized Metal Ion Affinity Chromatography (IMAC). The second construct, *LgtA\_X*, was designed to address the problem of solubility. Thus, *LgtA\_X* contained an *E. coli* thioredoxin (TRX) domain followed by a second 6xHistag in the C-terminal of *LgtA\_H*. Nucleotide sequence of both genes *lgtA\_H* and *lgtA\_X* were optimized for expression in the heterologous host *E. coli*.



**Figure 17.** General cloning strategies of LgtA, LgtA\_H and LgtA\_X. Orange bars represent DNA sequence identical to *N.meningitidis* LgtA, blue bars represent DNA sequences coding for His tags, purple bar represents DNA sequence coding for a TRX domain and the green bar represents DNA sequence coding for a flexible linker region

Plasmids were transformed into BL215(DE3) chemically competent *E.coli* and the transformants were selected on LB-Agar plates containing an antibiotic. The bacteria were grown at 37°C and induced using 1mM IPTG for 18h at 16°C. After centrifugation and cell lysis, a pellet and a supernatant were obtained which were analyzed by SDS-PAGE. The majority of the LgtA\_H and LgtA\_X protein was seen to be in the supernatant and this was therefore purified from all other proteins by IMAC FPLC and concentrated.



**Figure 18.** a) Relative amount of soluble/insoluble protein were visualized by SDS-PAGE. Black arrows indicate the position of LgtA, LgtA\_H or LgtA\_X. Pellet fractions (P), Supernatant (SN), Molecular weight marker (M). b) Purified proteins LgtA\_H and LgtA\_X after IMAC and LgtA were visualized by SDS-PAGE

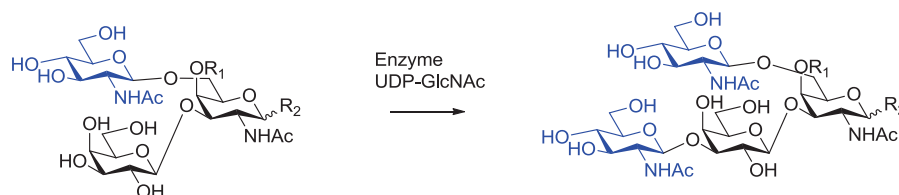
The isolated proteins were >95% pure as estimated by SDS-PAGE after purification by FPLC, a purity which is superior to that reported by Blixt previously. Although significantly purer, the LgtA\_H construct still suffered from poor yield as ca. 50% could be seen to remain in the bacterial pellet by SDS-PAGE analysis, owing to the low protein solubility. In contrast, all of the protein content was observed to be removed from the bacterial pellet for LgtA\_X. Therefore the TRX domain does indeed assist with the protein solubility. However, proteins precipitated upon concentration and above a concentration of 2.3mg/ml protein precipitation was

observed, unless 1mM DTT was added prior to concentration by centrifugation, after which a maximum of 10mg/ml can be obtained.

Overall, the use of the LgtA\_X construct increased enzyme solubility, improving the expression yield and facilitated the purification of the protein. However, the inherent instability of the enzymes made their storage a problem as precipitation was observed over time (hours). This was also observed when the protein was lyophilized and could not subsequently be resolubilized to the initial concentration. Protein precipitation could be slowed by storing the purified protein at -20°C in the FPLC elution buffer which contained 500mM imidazole without any observed effect on the enzymatic activity. In accordance with the LgtA construct, addition of 50% glycerol also stabilized the protein. However, both these additives are disadvantageous for MALDI monitored enzymatic reactions as they prevent drying of the matrix. Nevertheless, the protein shelf-life was estimated to be around 2 weeks at +4°C.

### 3.2.2 Enzymatic activity and selectivity

The substrate and co-factor minimal requirements of LgtA had already been documented by Blixt. While Naruchi *et al.* reported the enzyme to work best at pH 10, this was not the case for our construct.<sup>35</sup> Similarly, rates of conversion observed on the mucin core 2 **29** were lower than the reported values (Table 4). While this could be attributed to reaction conditions, Naruchi *et al.* observed substrate discrimination by the enzyme: when attached to threonine, the mucin core 2 was not elongated by LgtA whereas a 19% conversion was observed for the serine-attached glycan. This was suggested to be due to an unfitted conformational change of the mucin 2 *O*-glycan core against the enzyme active site by presence of the  $\gamma$ -methyl of the threonine residue. In light of this, our linker may somehow cause a similar phenomenon which could explain the difference between the described conversion and our maximum (Table 4, entry 1, 9% observed by MALDI TOF MS).



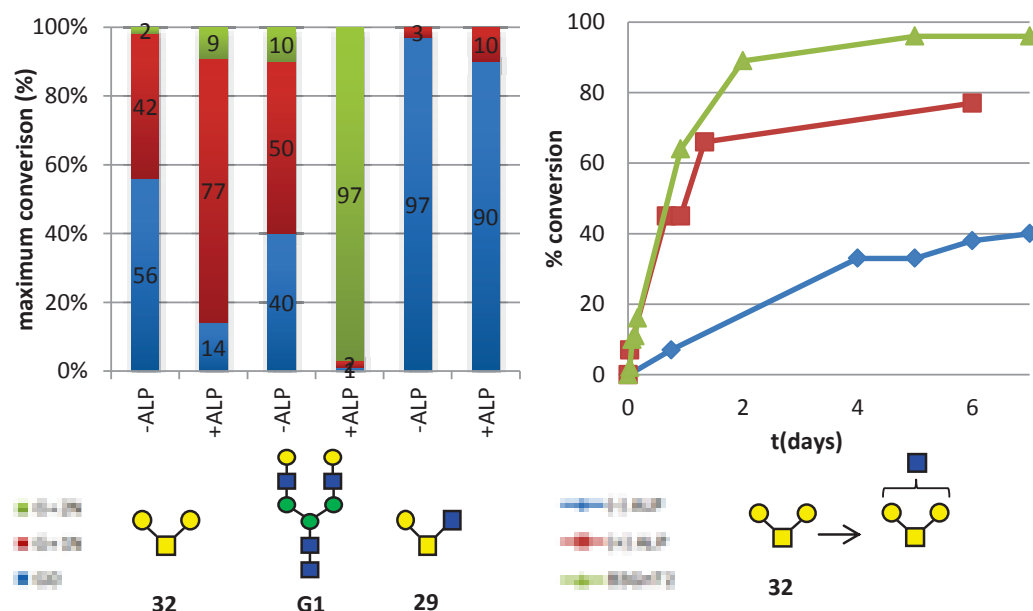
Entry	Substrate	Reaction conditions	Yield /%
1	<b>29</b> R1=Bn R2= $\beta$ -O(CH <sub>2</sub> ) <sub>5</sub> NBnCBz	LgtA_X, 25mM Tris, 20 mM MnCl <sub>2</sub> , 5 mM MgCl <sub>2</sub> , 1mM DTT, 2mM EDTA, UDP-GlcNAc (4eq), pH 7.5	9 <sup>a</sup>

<b>2</b> <sup>129</sup>	R1=H R2= $\alpha$ -OCH <sub>2</sub> CH(NHFmoc)CO <sub>2</sub> H(Ser)	LgtA (30mU/mL), 100mM glycine-NaOH, 10mM MnCl <sub>2</sub> , 10mM MgCl <sub>2</sub> , UDP-GlcNAc (5eq), pH 10	19 <sup>a,b</sup>
<b>3</b> <sup>129</sup>	R1=H R2= $\alpha$ -OCH(CH <sub>3</sub> )CH(NHFmoc)CO <sub>2</sub> H(Thr)	LgtA (30mU/mL), 100mM glycine-NaOH, 10mM MnCl <sub>2</sub> , 10mM MgCl <sub>2</sub> , UDP-GlcNAc (5eq), pH 10	0 <sup>a,b</sup>
<b>4</b> <sup>128</sup>	R1=H R2= $\alpha$ -O-pNP	100 mM sodium cacodylate pH 7.5, 4 mM MnCl <sub>2</sub> , 50mM ATP, UDP-[ <sup>3</sup> H]GlcNAc	9.2 <sup>c</sup>

**Table 4. Reported yields of LgtA and LgtA\_X with the mucin core 2.a= observed by MALDI, b= observed by HPLC-MS, c= observed by liquid scintillation counting and relative to conversion of lactose**

The importance of the anomeric substitution was additionally supported in the donor studies performed by Blixt where the enzymatic activity of a substrate (relative to lactose, in %) varied drastically depending on the anomeric group. LgtA was observed to transform acceptors bearing hydrophobic aglycon better than those without. This was illustrated for lacNAc where the relative activity for conversion of Gal $\beta$ -1,4GlcNAc (100) was far lower than for Gal $\beta$ -1-4GlcNAc $\beta$ -O-pNP (945). It was also reported for Gal(5.3) < Gal $\beta$ -O-pNP(16) < Gal $\beta$ -S-pNP(63) < Gal $\alpha$ -O-pNP(102).

We also observed that an excess of alkaline phosphatase (ALP, in conjunction with its cofactor MgCl<sub>2</sub>) improved conversion and rate of conversion.



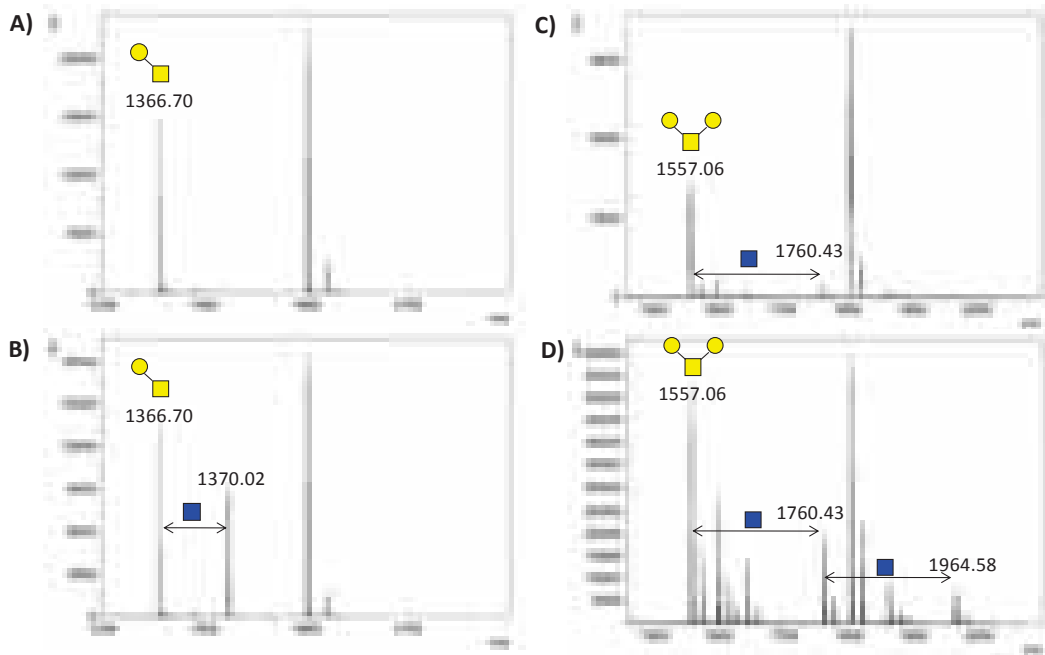
**Figure 19 Effect of ALP on LgtA\_X activity observed for N-glycan G1, *S.mansoni* core 32 and mucin core 2 29 and effect of ALP on LgtA\_X rate of activity on 32 compared to B3GnT2**

Using microarray technique, we were able to perform high-throughput comparison of LgtA\_X against B3GnT2. Regardless of the reaction phase, LgtA\_X was not observed to accept UDP-GalNAc as a donor, unlike reported by Blixt and Guan *et al.*<sup>122,36</sup> This difference could stem from the difference between LgtA\_X and LgtA, the difference in the linker at the reducing end (compare to aglycon containing sugars) or be a result of the reaction environment (solution phase vs on-chip). The acceptor specificity was also assessed by Blixt for both enzymes using a panel of N-glycans and liquid scintillation counting. Our observations on-chip were in line with previously reported results which are:

- LgtA\_X only modifies substrates containing terminal galactose
- However, substrates containing fucose proximal to reaction sites, ieLe<sup>X</sup> and LDNF epitopes, are not substrates for either enzyme.
- In contrast to LgtA\_X, B3GnT2 is able to modify substrates with terminal Gal and GalNAc

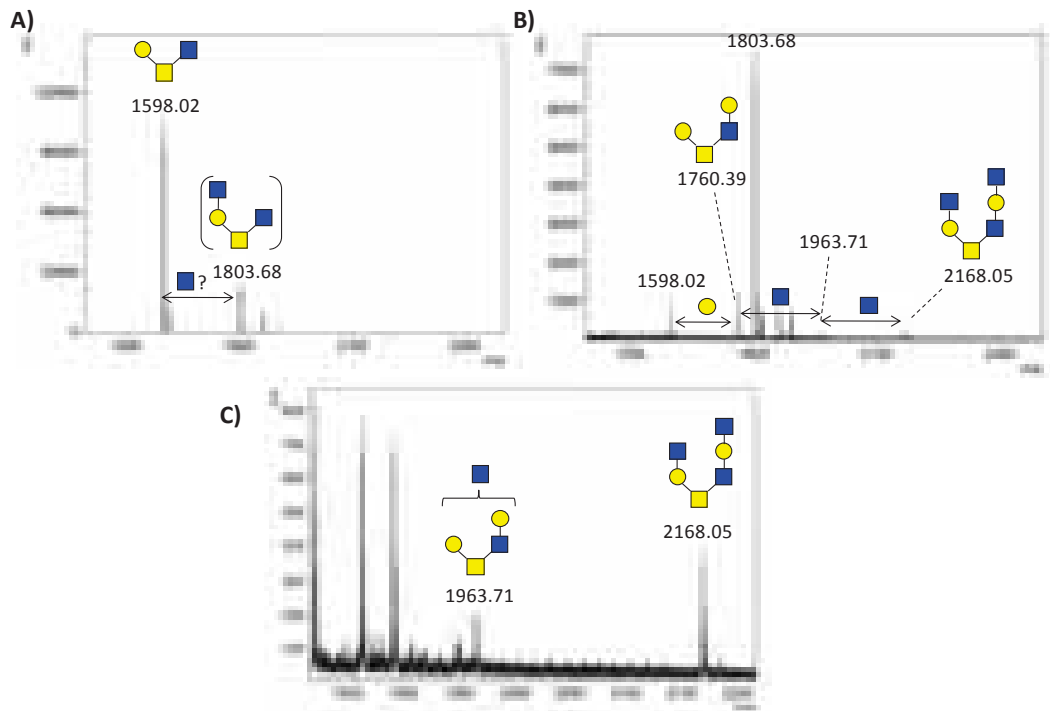
However, additional acceptor specificity was observed when evaluating the enzymes with the O-glycan cores. On-chip evaluation of B3GnT2 with core 1 **G2**, core 2 **O1** and the *S.mansoni* core **O2** revealed a regioselectivity for the  $\beta$ 1,6 arm of O-glycans (Figure 20). This was first suspected by the absence of product for the core 1 substrate and the observation of a single monomer formation for the *S.mansoni* core **O2** mass by MALDI-TOF MS. In contrast, LgtA\_X was observed to convert core 1 **G2** (41%) and a bi-substituted product (7%) was observed to be concomitantly formed with a monomer (56%) for the *S.mansoni* core.





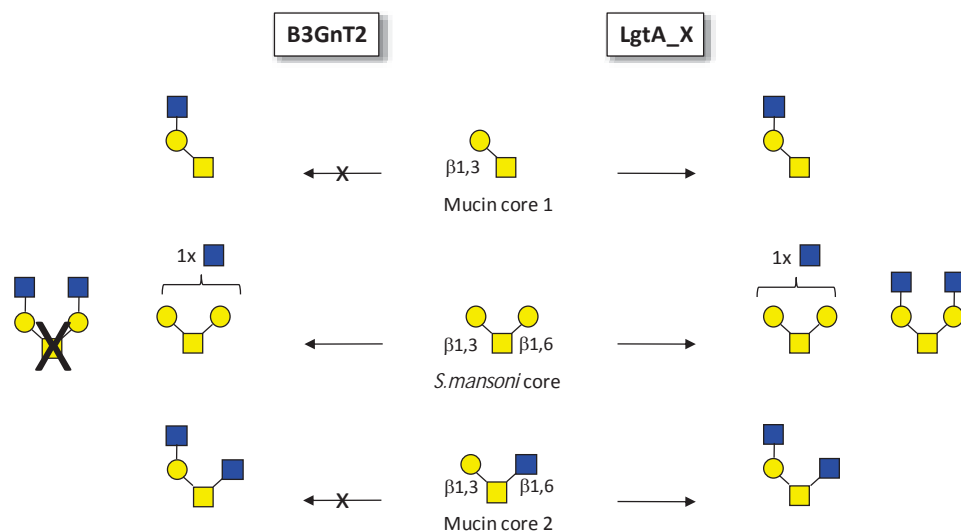
**Figure 20 On-chip evaluation of LgtA\_X vs B3GnT2 specificity. MALDI TOF-MS of A) B3GnT2 activity with core 1 G2; B) LgtA\_X activity with G2; C) B3GnT2 activity with *S.mansoni* core; D) LgtA\_X activity with *S.mansoni* core (contaminated with O1) Surface functionalization peak: 1803 m/z**

Unfortunately, the corresponding product mass for elongation on the mucin core 2 **O1** was identical to the recurring 1803 m/z mass which was concomitant with surface functionalization (Figure 21). This made the direct evaluation of the mucin core 2 as a substrate for B3GnT2 or LgtA\_X impossible by microarray. To reveal the enzymatic activity (or lack thereof) on the mucin core 2, an elongation by GalT-1 was carried out on the mucin core 2 to yield the galactosylated intermediate (1760 m/z). After incubation with the LgtA\_X a newly formed bi-substituted product (2167 m/z) could be observed for the mucin core 2 along with the monomer product (1963 m/z). The prior elongation with GalT therefore enabled us to reveal the activity of LgtA\_X on the mucin core 2 and confirm the  $\beta$ -1,3 and  $\beta$ -1,6 (Gal-GalNAc) activity of the enzyme (Figure 21.).



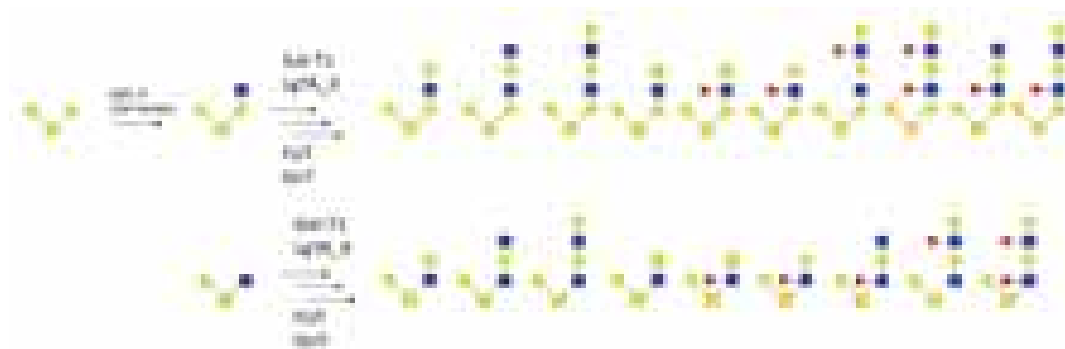
**Figure 21. LgtA\_X activity on the mucin core 2. MALDI-TOF MS of A) LgtA\_X on the mucin core 2; B) LgtA\_X after elongation of the mucin core 2 by GalT; C) peaks showing monomer and dimer formation after two elongations**

In the absence of enzymatic activity on core 1 and considering the exclusive formation of a mono-substituted product for the *S.mansoni* core, B3GnT2 showed preference for Gal $\beta$ -1,6GalNAc linkages. This enzymatic selectivity was further confirmed in solution where no conversion was observed for the mucin core **29** after incubation with B3GnT2 (monitored by MALDI-TOF MS).



**Scheme 16. Schematic summary of enzymatic activities observed on O-glycans**

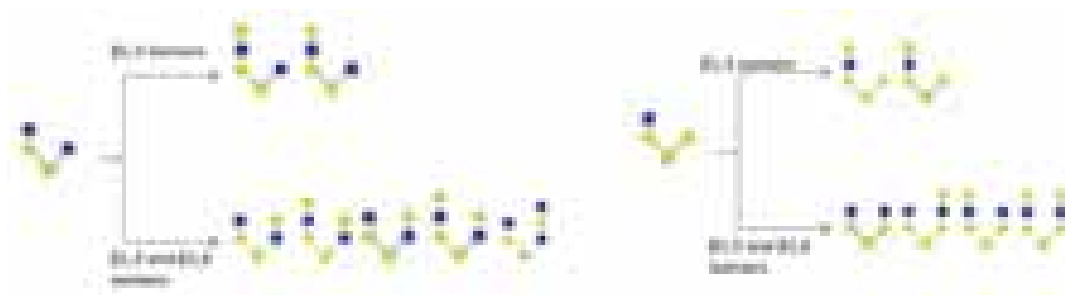
Thus, differential enzymatic activities of both enzymes could be used to generate pure regioisomers accordingly. B3GnT2 could be used to generate pure  $\beta 1,6$  elongated compounds while LgtA\_X could be complimentary in regioselectivity to B3GnT2 and suited to provide bi-substituted compounds. In order to unequivocally determine the regioselectivity of LgtA\_X, the *S.mansoni* core **32** was incubated with the enzyme on a preparative scale (5mg). The degree of conversion observed on-chip however was not mirrored in the solution conditions where conversion to a bi-substituted product **33** was limited to 10% as analysed by MALDI-TOF MS and LC-MS. Unfortunately this was not significant enough for collection by HPLC. However, the monomer product was successfully isolated and unequivocally characterized by NMR as the pure  $\beta$ -1,6 regioisomer **34**. Therefore, elongation on the  $\beta 1,3$  arm of O-glycans was highly disfavoured regardless of the GlcNAcT employed. As a consequence, the number of compounds targetable by solution phase preparative synthesis was restricted to those illustrated in Scheme 16.



**Scheme 16. New library of  $\beta$ -1,6 elongated structures**

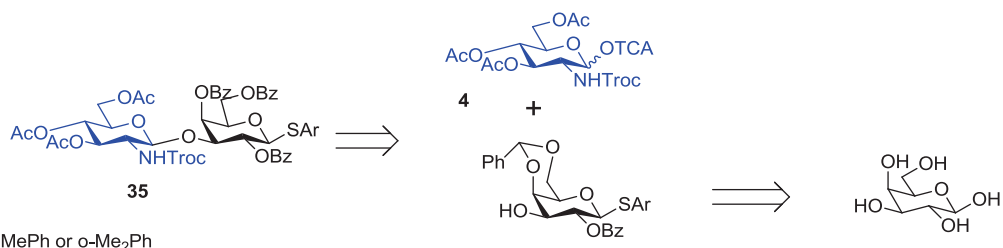
### 3.2.3 Towards $\beta$ -1,3 elongated O-glycan structures

Although the regioselectivity of LgtA\_X is partly circumvented on chip, it does not allow access to the  $\beta$ -1,3 elongated O-glycan structures. To access these, a synthetic precursor could be envisaged. Conveniently, this is common to both mucin 2 and *S.mansoni* cores and regardless of the epitope-LN or LDN- desired. Elaborating on the cassette-strategy, enzymatically challenging  $\beta$ -1,3 elongated compounds can be achieved by inserting a GlcNAc $\beta$ -1,4Gal residuesynthetically. In this way, not only are the  $\beta$ -1,3 elongated compounds made accessible but so are the bi-substituted compounds.



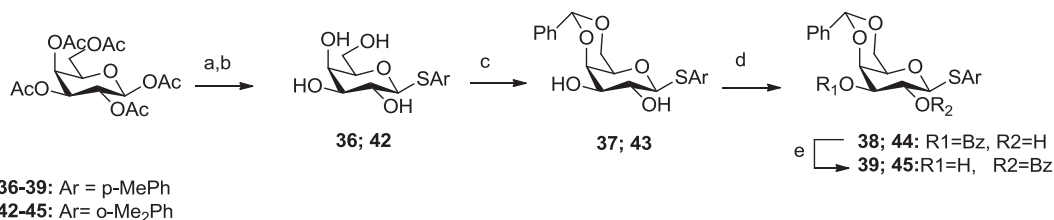
**Scheme 17. Chemoenzymatic synthesis of  $\beta$ -1,3 structures. A) Synthetic precursor; B) accessible structures by enzymatic elongations**

The new synthesis including the GlcNAc $\beta$ -1,4Gal residue could be envisaged by replacing the previous galactose donor **2** by a donor **35**. The latter can be synthesized as described by retrosynthetic analysis in Scheme 18.



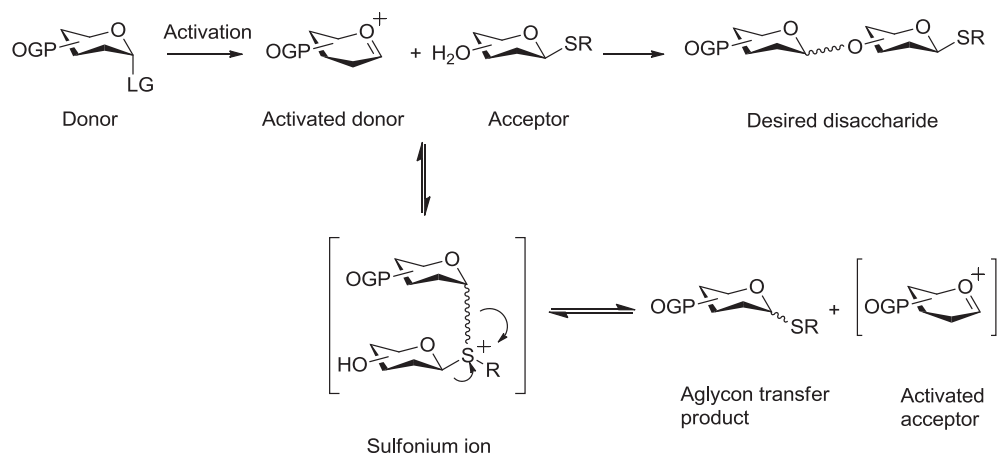
### Scheme 18. Retrosynthesis of the protected GlcNAcβ-1,4Gal residue

Our proposed synthetic strategy was carried out in solution and makes use of donor **4**, previously employed in the synthesis of **25**. Benzoyl protecting groups were targeted for the galactose moiety in view of previous results (see 3.1.4). Peracetylated galactose was converted to a thiogalactoside using tolylthiol and deacetylated to **36** (Scheme 19). Subsequent protection of the 4- and 6- positions as a benzylidene acetal yielded **37** in 80% yield. The 3-OH being more reactive to benzoylation than the desired 2-OH, we sought to ensure full conversion to the 3-OH intermediates **38** before performing an acyl migration in acetone and NaOH.<sup>37</sup> Thus, 2-OH isomers **39** can be obtained in 48% yield over two steps.



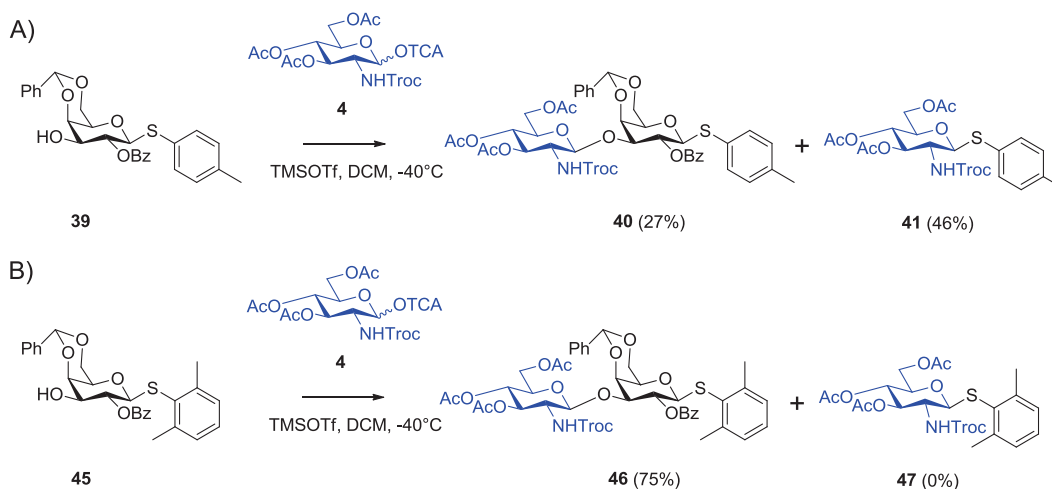
### Scheme 19. Synthesis of the galactose moiety of GlcNAcβ-1-4Gal residue **38**. a) S-Ar, BF<sub>3</sub>·Et<sub>2</sub>O; b) NaOMe, MeOH; c) PhCH(OMe)<sub>2</sub>, cat. CSA, ACN; d) BzCN, DMAP, DCM, -50°C; e) acetone, NaOH, 0°C

At this point, the importance of the choice of thiol must be highlighted. When glycosylation between the thiotolyl acceptor **39** and the imidate **4** was trialed, only a poor 27% yield of **40** was obtained (Scheme 21). NMR analysis of the isolated side-products revealed the product **41**, owing to an aglycon transfer, the mechanism of which is shown in Scheme 20.



**Scheme 20. Aglycon transfer mechanism**

Aglycon transfer is well reported and occurs due to the sulfur atom reacting with an activated glycosyl donor, forming a sulfonium ion. Cleavage of the acceptor-thiol bond then leads to transfer product.<sup>38</sup> Using the 2,6 dimethylphenyl (DMP) thiol was shown to avoid aglycon transfer. The dimethyl groups confer steric hindrance, thus preventing electrophilic attack of the donor oxonium ion on the acceptor.<sup>39</sup> Indeed, when **42**, **43**, **44**, **45** (prepared as shown in Scheme 19) was used, the disaccharide **46** was obtained in 75% yield.



**Scheme 21. Glycosylation of the thioarylglycosides leading to aglycon transfer**

Unfortunately we were unable to pursue the synthesis due to time constraints but subsequent hydrolysis of the benzylidene acetal and benzylation of positions C-4 and C-6 of are **46** expected to proceed smoothly, yielding donor **35**.

### 3.2.4 Studies directed towards obtaining a crystal structure of LgtA

Despite considerable studies on LgtA, no crystal structure has been reported. A bioinformatics analysis using BLAST and EXPASY of LgtA's DNA did not reveal any sequence similarity to any reported  $\beta$ -1,3 GlcNAcT structures. It is highly suspected to be part of the GtB class of enzymes, with an inverting mechanism however there is no absolute evidence. Taking into account the large amounts of pure enzyme required for crystallization studies, the lack of a crystal structure might be explained by the difficulties related to producing sufficient quantities of pure LgtA. In the scope of our work, elucidation of the protein crystal structure would allow insight into the enzymatic mechanism, hence accounting for its regioselectivity. Moreover, LgtA is suspected to be involved in the pathogenesis of *N.meningitidis* by contributing to the synthesis of the lipooligosaccharides responsible for the bacterial evasion of the host immune system.<sup>40</sup> A crystal structure including a natural substrate in the binding site (the HOLO state) would be an asset for the rational design of inhibitors for this important target. Additionally, mutations made to the enzyme active site could be envisaged in the hope of broadening the substrate or even donor specificity.

Crystallization experiments were undertaken under the guidance of Marcelo Guerin, CIC biomaGUNE (Spain). Both constructs were expressed on a large scale (4L) to yield 40 and 60mg of LgtA\_H and LgtA\_X, respectively. Although LgtA\_X was higher yielding than LgtA\_H, the additional flexible linker region of LgtA\_X was a potential disadvantage for successful crystal formation. Hence, both constructs were evaluated in parallel. These were concentrated to the highest concentration and with minimal additives possible.

	LgtA_X	LgtA_H
<b>APO</b>	8.4mg/ml 20mM Tris, 150mM NaCl, 5mM Imidazole, 1mM DTT, 10%glycerol	6.6mg/ml 20mM Tris, 300mM NaCl, 5mM Imidazole, 1mM DTT
<b>HOLO</b>	7.8mg/ml 20mM Tris, 150mM NaCl, 5mM Imidazole, 1mM DTT, 5mM Lactose, 5mM UDP, 5mM MgCl <sub>2</sub>	6.6mg/ml 20mM Tris, 300mM NaCl, 5mM Imidazole, 1mM DTT, 5mM Lactose, 1mM UDP, 5mM MgCl <sub>2</sub>
<b>Crystallization</b>	Structure screens I & II	Structure screens I & II

<b>conditions</b>	Morpheus®	Morpheus®
<b>tested</b>	JCSG+	JCSG+
	Grid Screen™ Ammonium Sulfate	Grid Screen™ Ammonium Sulfate
		PACT
	384 ie 768 potential crystal spots	480 ie 960 potential crystal spots

**Table 5. Summary of crystallization experiment conditions of crystallization performed on LgtA\_X and LgtA\_H**

LgtA\_H appeared to be much less stable at higher concentrations than LgtA\_X as protein precipitation was observed, once again demonstrating the enhanced solubility effect of the added TRX domain in LgtA\_X. The minimal salt concentration required to avoid protein precipitation (300mM NaCl) was twice as high as for LgtA\_X, and addition of glycerol did not improve protein stability.

To increase the chances of protein crystal formation, conditions for the APO form (the protein conformation without substrate in the active site) and the HOLO form of each construct were implemented. For the HOLO form, addition of 5mM UDP caused LgtA\_X to precipitate and the protein solution was reduced from 8.23mg/ml to 7.8 mg/ml. Therefore only 1mM UDP was added to the HOLO reaction of LgtA\_H and no precipitation was observed. MgCl<sub>2</sub>, the fastest enzyme cofactor after MnCl<sub>2</sub>, was chosen as it ensures slower reaction rates which are favourable to protein structures obtained with the substrate in the active site. Additionally, it circumvents the precipitation caused by the fastest cofactor MnCl<sub>2</sub>.

A broad screening of crystallization conditions using commercially available crystallization kits by the sitting-drop vapour-diffusion method (SDVD). The SDVD method involves the equilibration of a drop of protein mixture against a reservoir of precipitant cocktail. In so doing, the protein concentration increases in the drop and under the right conditions leads to crystal formation.<sup>41</sup> Each screening kit contained 96 different conditions. Applied to both the HOLO and the APO reactions of each construct, a total of 768 and 960 conditions were screened for LgtA\_X and for LgtA\_H, respectively. These were stored and monitored over time in a Bruker Crystal Farm at 21°C. In most cases, amorphous precipitation was observed. In the cases where crystal formation could be observed, these samples sent for evaluation to a synchrotron radiation source but unfortunately these were identified as salt crystals in all cases. Therefore no crystals were obtained for the recombinant protein constructs. In general,



bacterial enzymes such as LgtA display inherent instability which can severely hamper their crystallization attempts also reflected by:

1. The need to store in LgtA, LgtA\_H and LgtA\_X in high glycerol content (50%)
2. The superior stability of LgtA\_X over LgtA\_H in solution
3. The expression of the analogous recombinant *H. pylori* enzyme as a maltose binding protein fusion protein to increase its solubility<sup>42</sup>

Gel filtration of LgtA\_X showed the protein to elute as a broad peak suggesting heterogenic conformational states. This was in accordance to typical states of glycosyltransferases reported but decreases chances of protein crystal formation.<sup>43,44</sup> A possible solution for overcoming the low solubility and consequently the difficulties encountered in the protein crystallization might be the use of a stabilizing or a chaperon protein or the inclusion of its transmembrane domain. If so, one solution could be to express LgtA as a fusion protein with a GalT domain such as LgtB, a  $\beta$ -1,4GalT made by *N.meningitidis*. Since both of these enzymes are naturally employed by the bacteria for the synthesis of LacNAc epitopes in lipooligosaccharides, the complex may be more stable overall.

### 3.2.5 Expression of GalT and GalNT

Although *N.meningitidis* also possesses a  $\beta$ -4 galactosyltransferase ( $\beta$ -4 GalT, LgtB) useful for construction of LacNAc epitopes, a bovine Gal-T1 has been more frequently used.<sup>45,46</sup> The native enzyme catalyzes the transfer of a galactose residue to the 4-position of a terminal N-acetylglucosamine and its C342T mutant is readily expressed in *E.coli* in good yields.<sup>47</sup> Moreover, its Y289L mutant which transfers N-acetylgalactosamine (rather than galactose, making it a GalNT) has been applied towards the synthesis of the antigenic LDN structures.<sup>48,49</sup> Therefore both enzymes appeared as valuable tools for the construction of a library of O-glycans.

Briefly, *E.coli* cell lines containing the desired enzyme gene were seeded in LB agar plates containing the appropriate antibiotic and incubated at 37°C overnight. A single colony was inoculated in liquid LB and incubated overnight. The culture was diluted and the bacterial growth was maintained at 37°C until it reached an OD<sub>600nm</sub> value of 0.6-0.7 at which point protein expression was induced for 4 hours. The culture was centrifuged and the bacterial

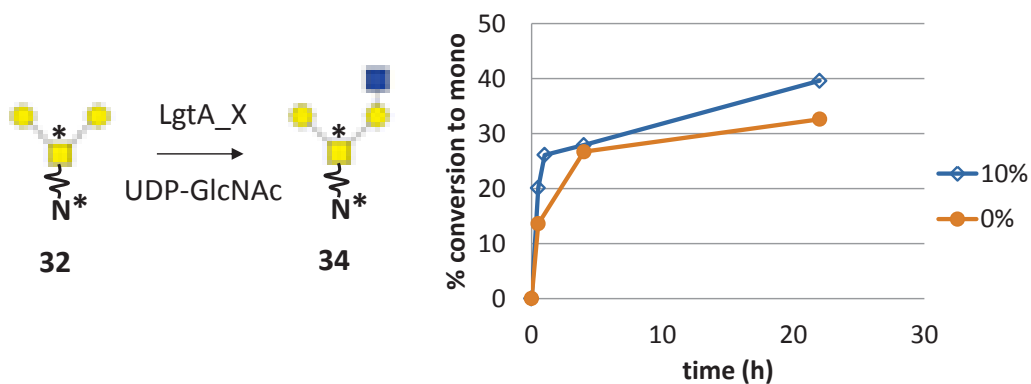
pellet was washed, sonicated and redissolved in guanidine until it reached an OD<sub>280nm</sub> of approximately 2.0. The protein solution was filtered, diluted ten times in folding buffer and kept at 4°C overnight after which the buffer was exchanged by dialysis and the protein solution was centrifuged until satisfactory concentration was obtained.

### 3.3 Solution-phase enzymatic transformations

#### 3.3.1 Solubility of O-glycans in enzymatic transformations

With all the enzymatic tools in place, our attention was turned to the generation of the asymmetric library based on the mucin core 2 **29** and the *S.mansoni* core **32**. Although the remaining benzyl groups provide chromophores for easier purification by HPLC-UV, their hydrophobic contribution had not been anticipated and resulted in poor solubility of all glycans in aqueous reactions. As a consequence, enzymatic reactions were often slow as a high dilution was necessary for a satisfactory solubilisation of the compounds (less than 1mM). This implied reactions progressing very slowly as the probability of collision rate of substrate with enzyme decreased. In the circumstances where recombinant enzymes were expressed in-house and were high yielding- and therefore overall more affordable- a large excess of enzyme was used to increase the probability of collision substrate with enzyme. In the cases of expensive commercial enzymes, notably recombinant B3GnT2 and B3GnT6, or enzymes with low expression yield (eg. CeFUT6), the amount of enzyme employed in the reactions was adjusted carefully.

Alternatively, the compounds were found to be readily soluble in DMSO or acetic acid, however neither are suitable for enzymatic reactions. In some cases however, minor quantities of DMSO can be tolerated by the enzyme and this was seen to be the case for LgtA\_X. When compounds were dissolved to 5mM in DMSO and added to the enzymatic mixture to a maximum of 10% v/v DMSO/water, the reaction was observed to proceed marginally better than with none. Indeed, 40% conversion of the *S.mansoni* core **32** to the monomeric product **34** was observed in 10% v/v after 20h whereas only 30% was observed in 0%v/v.



**Figure 22. Effect of DMSO on LgtA\_X activity on 32**

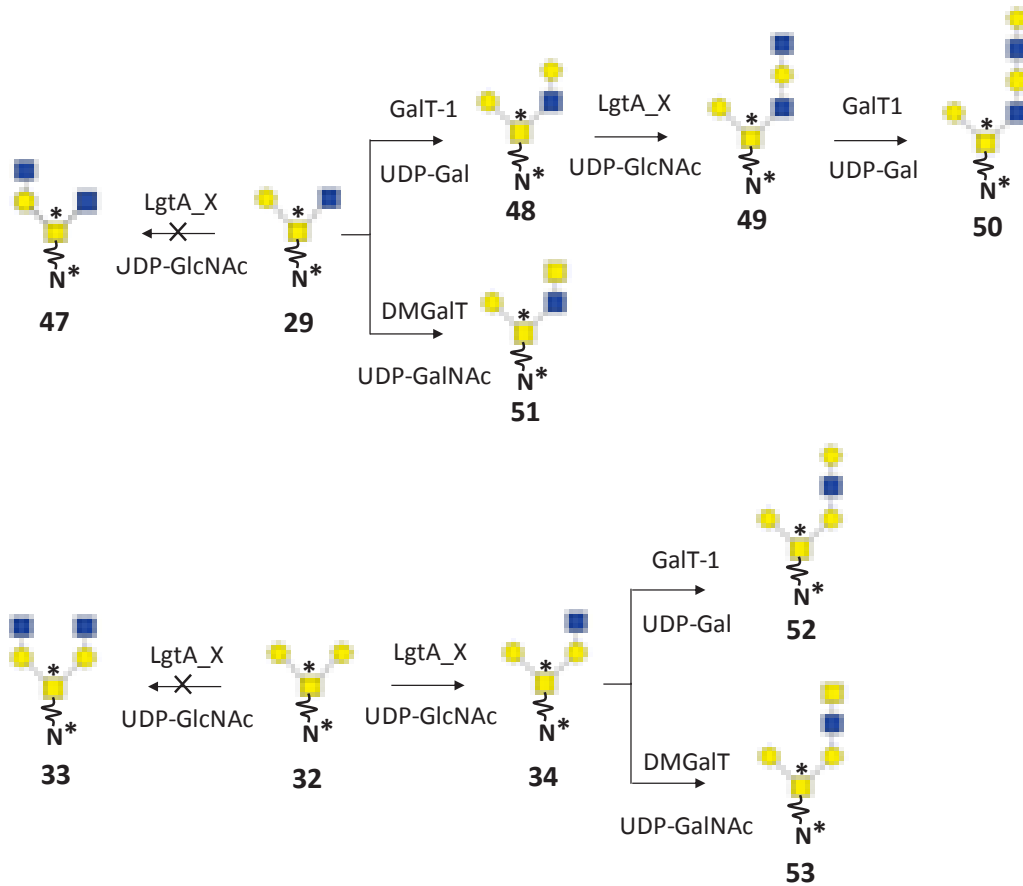
The increased proportion of the compound in solution and the added detergent-like effect of DMSO on LgtA\_X were speculated to be the reasons this was observed. This held true only when the DMSO solution of compound was added to the aqueous reaction mixture as addition in the other way resulted in a bi-phasic mixture. The latter suggests the formation of a colloid by the compounds in the DMSO. The content of DMSO was limited to 10% v/v as higher volumes were not tolerated by the HPLC equipment during injection.

The solubility also affected characterization of the obtained compounds. For NMR analysis, the high concentrations needed were obtained by dissolving compounds in a mix of solvents. In the case of the mucin2 glycans, the compounds were relatively soluble in a mix of MeOD:D<sub>2</sub>O whereas *S.mansoni* glycans were more soluble in a mix of CDCl<sub>3</sub>:MeOD:D<sub>2</sub>O. Although heating initially improved the solubility of the samples, all formed gels over time.

In light of the hurdles caused by the chromophores and the regioselectivity of LgtA\_X, the benzyl and CBz groups were not essential for the regioselective synthesis of the  $\beta$ 1,6 O-glycans. However, they were useful in the purification by HPLC-UV of enzymatic mixtures where full conversion was not obtained and the product needed separating from the starting material. Purification by HPLC-MS would provide an efficient way of purifying compound lacking chromophores.

### 3.3.2 LacNAc and LDN epitopes

Eventually, both cores could be elongated in a sequential way to obtain structures depicted in Scheme 22. Although a maximum 9% and 10% conversion were observed for compounds **33** and **47**(3.2.2), this was not observed upon scale up and were therefore not isolated in useable amounts.



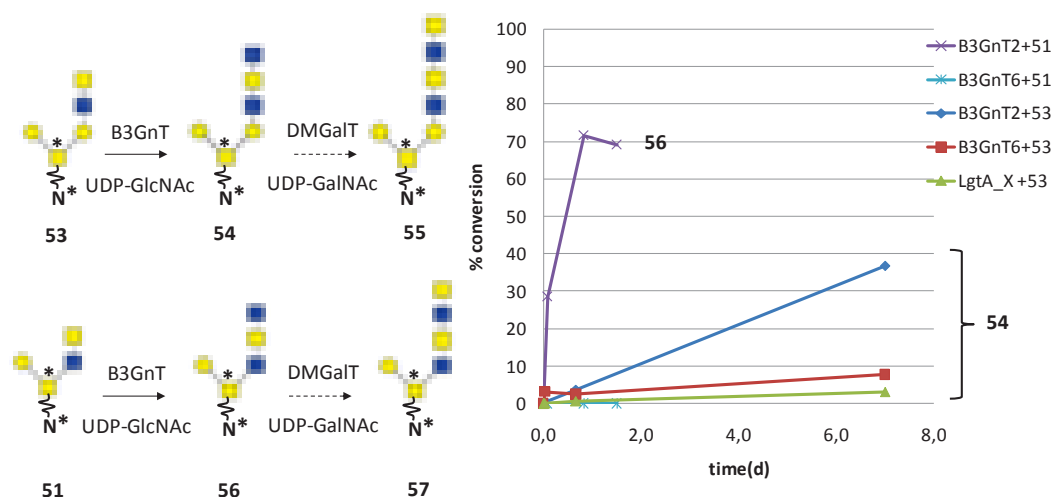
**Scheme 22. Structures of O-glycans obtained by enzymatic elongation in solution**

As previously mentioned, LacNAc structures are abundantly observed in healthy human mucin-type O-glycans and as consequence their chemoenzymatic synthesis has been described. As these structures were missing from the consortium of the research network, we synthesized **48**, **49**, and **50** as a representative sample. Moreover, the *S. mansoni* egg glycan profile was shown to contain tandem repeats of polyLe<sup>x</sup> structures as well polyLDN and highly fucosylated polyLDN residues. The LDN (and fucosylated LDN) structures are considered trademarks of helminth antigens because of their rare expression in human cells and of their recognition by the immune system as molecular pattern for helminths. Indeed, a BSA-LDN neoglycoconjugate reportedly showed differential binding between the soluble lectins galectin-1 and galectin-3. Further studies showed that galectin-3 was able to mediate recognition and phagocytosis of LDN-coated nanoparticles by macrophages whereas galectin-1 was unable to recognize LDN as a ligand, consistent with the galectin's LacNAc specificity.<sup>50</sup> Despite their demonstrated antigenic profile, O-glycans are less investigated, due to the lack of an enzymatic universal method for intact O-glycan release (compare to N-glycans which can be released by PNGase).<sup>51</sup>

As such, LDN epitopes are commonly screened as N-glycan termini, as protein conjugates or as the isolated disaccharide.<sup>52,53</sup> The chemoenzymatic synthesis of compound **51** is therefore novel. Unfortunately, its regioisomer and the dimer are inaccessible via our strategy for reasons mentioned above (see 3.2.2)

LacNAc and Le<sup>x</sup> were shown to be the main structures decorating the *S.mansoni* core.<sup>54</sup> Within our time and material constraints, we were able to synthesize **52** and **53** enzymatically. Although LDN epitopes are not predominantly observed in the cercarial glycan content, **53** was also synthesized for its bioisosteric trait of **52** and for its unrefuted antigenic properties. Given the importance of presentation in glycan-CLR binding affinities, we were also curious to see whether presentation of the LDN epitope on the *S.mansoni* core would differ to its presentation on the mammalian mucin core 2.

The polyLDN content observed on the mucin core 2 in the *S.mansoni* glycome consists of one epitope per branching, ie a single epitope on both the  $\beta$ 1,3 and  $\beta$ 1,6 arms of the core, fucosylated to various extents. In contrast, linear repeats of LDN were seen to occur on the helminth's N-glycans.<sup>55</sup> Considering our regiolimitations for the mucin core 2, we sought to create an atypical linear LDN repeat on the O-glycans. To this end, we considered three  $\beta$ 3GnTs: LgtA\_X, B3GnT2 and B3GnT6. These were incubated with **51** and **53** on an initial analytical scale (nmolar) and the conversions were monitored by MALDI-TOF MS (Figure 23).



**Figure 23. Proposed polyLDN construction and evaluation of 3 B3GnTs for the formation of the GalNAc $\beta$ -1,4GlcNAc bond**

In accordance with the microarray studies (see 3.2.2), LgtA\_X was unable to perform the reaction and no conversion was observed. The human B3GnT2 and B3GnT6 on the other hand

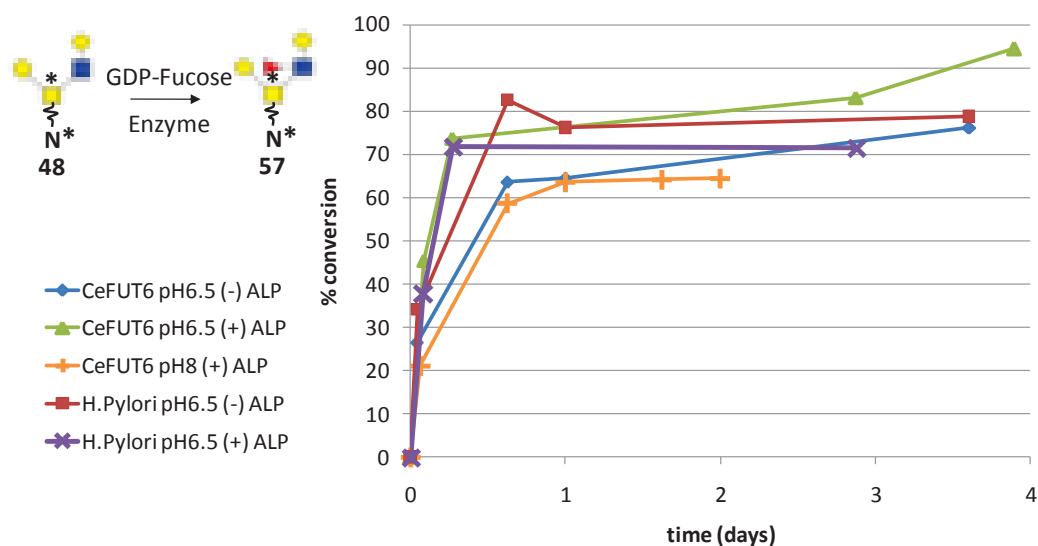
showed promising results. In particular, B3GnT2 was observed to perform better than B3GnT6. This was observed in the 37% conversion to **54** compared to 8% for B3GnT6 after 7 days of incubation and in particular in the 70% conversion to **56** compared to 0% for B3GnT6 after 2 days. This was peculiar given that B3GnT2 is described to initiate and elongate polyLacNAc structures whereas B3GnT6 is involved in the biosynthesis of the O-glycan core 3 (GlcNAc $\beta$ 1-3GalNAc $\alpha$ -Ser/Thr). Thus B3GnT6 was initially expected to perform better on the GalNAc acceptor moiety than B3GnT2 yet the opposite selectivity was observed. A marked difference was also observed between the two O-glycan substrates as B3GnT2 showed a maximal 70% conversion of **51** to **56** after 2 days whereas only 38% was observed for the transformation of **53** to **54**. This was rationalized as the difference in the reaction conditions each transformation was carried out in. Indeed, more enzyme was used on **56** (1.5 times more) than for **54**.

Nonetheless, encouraged by these results, 1.2mg of **51** was incubated with B3GnT2. However, only a maximum of 8% of **56** was observed. The initial satisfactory conversions were therefore attributed to the analytical scale on which this was performed, where the enzyme was used in large excess compared to the substrate. Extrapolating the conditions used for the analytical scale to the preparative, 60ug of enzyme would be required to obtain 70% conversion in 2days. As this represented a substantial cost, the synthesis was not further pursued. While reports of LDN construction typically only involve the formation of a singleGalNAc $\beta$ 1-4GlcNAc linkage using a GalNAcT, reports on the synthesis of linear polyLDN structures remain scarce. One successful example exploited using Chinese Hamster Ovary cell lines however, the degree of LDN polymerization was not controllable and the protocol was specific for *N*-glycans.<sup>56</sup> Therefore, until a specific enzyme catalyzing the formation of GlcNAc $\beta$ -1,3GalNAc bonds is isolated, *S.mansoni* and mucin core 2 derived compounds carrying poly-LDN residues are not accessible via enzymatic routes.

### 3.3.3 Lewis X and LDN-F epitopes

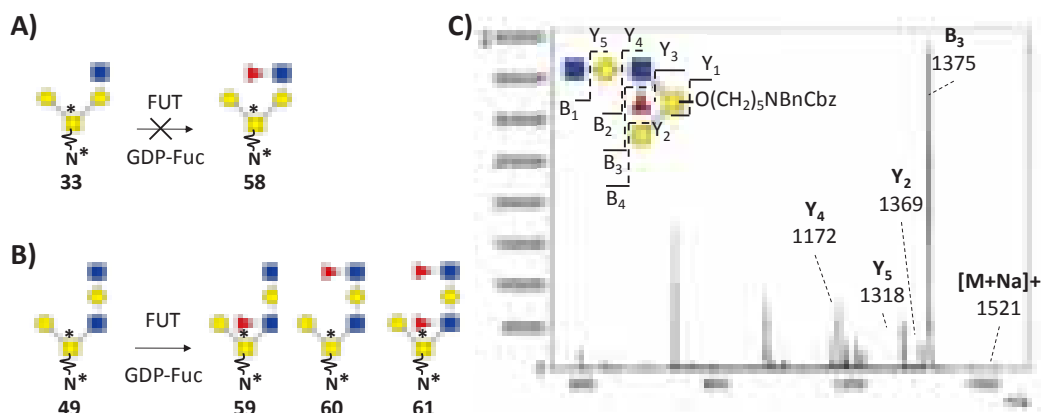
The enzymatic  $\alpha$ -1,3 fucosylation of the GlcNAc residue on the LacNAc moieties has been described previously.<sup>57</sup> Two recombinant enzymes were considered for our synthesis, a fucosyltransferase from the nematode *C.elegans* (CeFUT6, expressed in-house) and a commercial bacterial enzyme from *H.pylori* (HP-FucT). The primary activity of CeFUT6 is generation of the distal Fu $\alpha$ -1,3GlcNAc in the chitobiose moiety of N-glycans but it can also be applied to the synthesis of LeX structures.<sup>58</sup> A comparison of both enzymes by MALDI-TOF MS using tetrasaccharide **48** as a substrate showed similar activity (Figure 24). In view of the

accelerated rates ALP had on LgtA\_X (see 3.2.2), we also compared the rates of both enzymes in the presence or absence of ALP. CeFUT6 showed slightly better conversions with ALP whereas no significant difference was observed for HP-FucT. We also tested the effect of ALP pH8 on the activity of CeFUT6 as ALP was reported to work best at higher pH.<sup>59</sup> A comparable activity to conditions in pH6.5 without ALP was observed which could be attributed to the fact that pH 8 was too high for optimal CeFUT6 activity.



**Figure 24. Comparison of CeFUT6 vs *H.pylori* to make compound 57**

Additionally, we confirmed that LacNAc (Gal $\beta$ 1-4GlcNAc) was an absolute requirement for formation of the Fuc $\alpha$ -1,3GlcNAc bond. Thus, no conversion was observed for **33** when incubated with CeFUT6 (Figure 33, A). Moreover, fucosylation of **44** using either enzyme led to a singly fucosylated compound **54**, identified by fragmentation by MALDI-TOF MS (Figure 25, B), C)).



**Figure 25. Acceptor specificity of FUTs. A) Fucosylation trial of 34 to 58. B) Fucosylation trial of 49 leading to possible formation of 59 60 and 61. C) MALDI ToF/ToF of compound of  $m/z$  1521. Diagnostic fragment ions confirming the shown location of fucose residue are Y2 and Y4. Assignment of fragments according to Domon and Costello<sup>60</sup>**

In addition, HP-FucT had been reported to accept a broader spectrum of donors, including the C-6 azido surrogate (FucZ). This property was subsequently used as a bioorthogonal tag to detect cell-lines containing LacNAc by coupling a fluorescent dye.<sup>61</sup> Similarly, this strategy could be applied towards the generation of glycomimetics. Thus, azido-fucosylated compounds could be further derivitized by copper-catalyzed cycloaddition with a collection of alkynes to afford a library of novel glycomimetics with potentially improved affinity towards CLRs. A similar strategy employing alkyne and azide functionalized sialic acid derivatives was employed by Paulson for the preparation of sialoside glycomimetics to target Siglecs, a family of sialic acid binding lectins primarily expressed in white blood cells (see 3.4).<sup>62</sup> Hence, HP-FucT appeared more advantageous than CeFUT6 due to its laxer donor specificity.

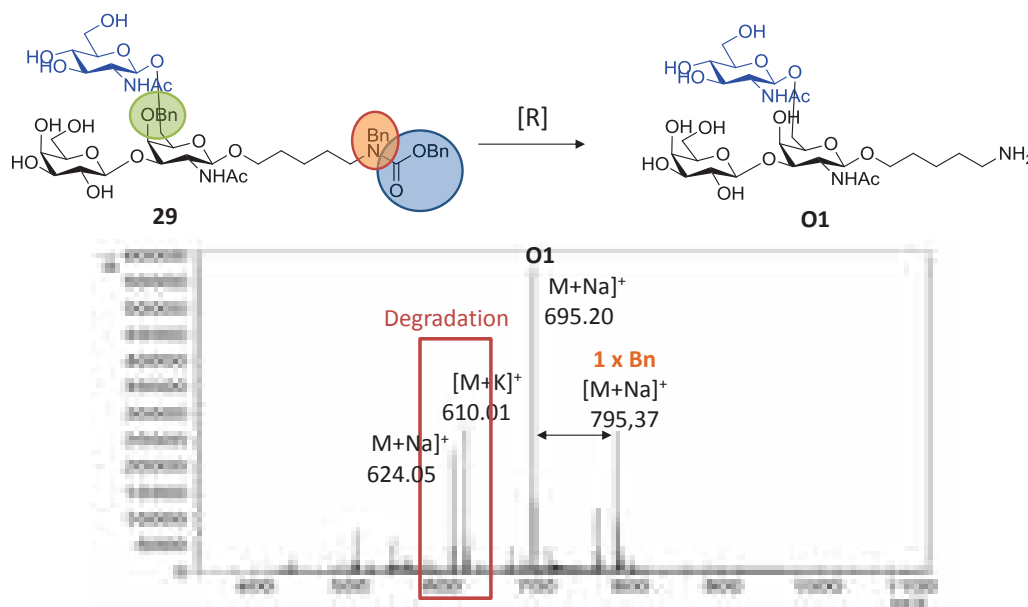
Regardless of the enzyme and donor used, fucosylation was planned after final deprotection of the compounds by hydrogenation, as the harshness of this last step was expected to degrade the labile glycosidic  $\alpha$ -1,3 fucose bond. Additionally, the azide functionality would be preserved in this way.

### 3.3.4 Deprotection of compounds

Once in hand, select O-glycans required the final deprotection step to reveal the free amine necessary for attaching the compounds to the microarray slide and for CLR interaction studies. The remaining 4-O-benzyl, N-benzyl and N-Cbz groups have all been reported to be efficiently removed by flow-chemistry hydrogenation using the H-cube<sup>®</sup>.<sup>63,64</sup> However, this step was found not to be as trivial originally expected in our O-glycans. The N-benzyl was predictably



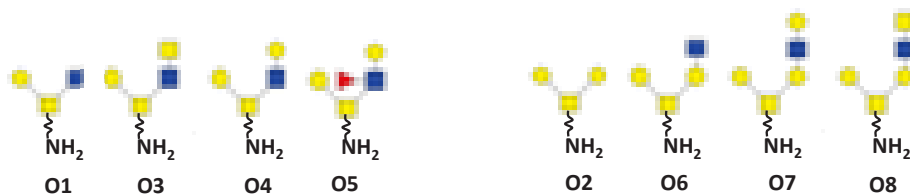
more difficult to deprotect owing to the proximal lone-pair of the nitrogen and therefore repeated exposures to the hydrogenation system (Pd/C (10%) cartridge, atmospheric H<sub>2</sub>(g), 1%v/v trifluoroacetic acid [TFA] in MeOH) were needed in order to observe full consumption of the N-benzyl peak.<sup>65</sup> The TFA prevented the resulting amine from poisoning the catalyst and consequently being retained in the Pd/C cartridge. As a result of several exposures to the acidic hydrogenation conditions, the N-benzyl peak was indeed observed to diminish but incurred concomitant formation of a peak with 85 Dalton mass difference.



**Figure 26. MALDI-TOF spectrum of flow-hydrogenation of mucin core 2**

This seemed to correspond to a cleavage of the linker from the sugar moiety. This was unexpected as it was not seen in previous work from our laboratory on N-glycans. Nonetheless, some form of degradation of the target material clearly occurred as only a maximum of 50% yield was only ever achieved. It was therefore also necessary to optimize the final deprotection step. Several trials were made to circumvent the harsh reaction conditions. A careful balance was needed between the need for harsher conditions for full removal of the challenging N-benzyl and the degradation observed as a consequence. To this end, several parameters were varied including the type, strength and quantity of acid (0.1-2% TFA or acetic acid), the temperature (25-70°C), the flow rate (1-2ml/min), the H<sub>2</sub>(g) pressure (atmospheric-50bar) and the catalyst (10%Pd/C or 20% Pd(OH)<sub>2</sub>/C).

As the variation of hydrogenation parameters on the flow chemistry device did not lead to satisfactory or reproducible results, the reaction was carried out in solution. Conditions involving *in situ* generation of hydrogen with ammonium formate or triethylsilane (TES) under microwave irradiation offered no improvements as neither a clean nor full conversion was observed by MALDI-TOF MS).<sup>66,67,68</sup> Finally, we resorted to conventional hydrogenation using H<sub>2</sub>(g). A first trial in MeOH was run in which full conversion to the N-benzyl species was observed. However, considerable amounts of methylated species (*ca.* 40% estimated by MALDI) were also observed (probably on the amine), as a consequence of the solvent system used. This was recurrently observed in previous work from our group and can be avoided by addition of acid. Finally a procedure reported in 2:2:2:50 THF:H<sub>2</sub>O:MeOH:AcOH for the deprotection of an oligosaccharide as a vaccine towards *C.difficile*, including an identically protected linker, was employed for the hydrogenation of our compounds.<sup>69</sup> Dissolved in a mix of H<sub>2</sub>O/MeOH/AcOH, hydrogenation of benzylated precursors **29**, **32-34**, **48-53** with Pd/C(10%) under atmospheric pressure of H<sub>2</sub>(g) afforded complete and clean conversion of the following compounds overnight.



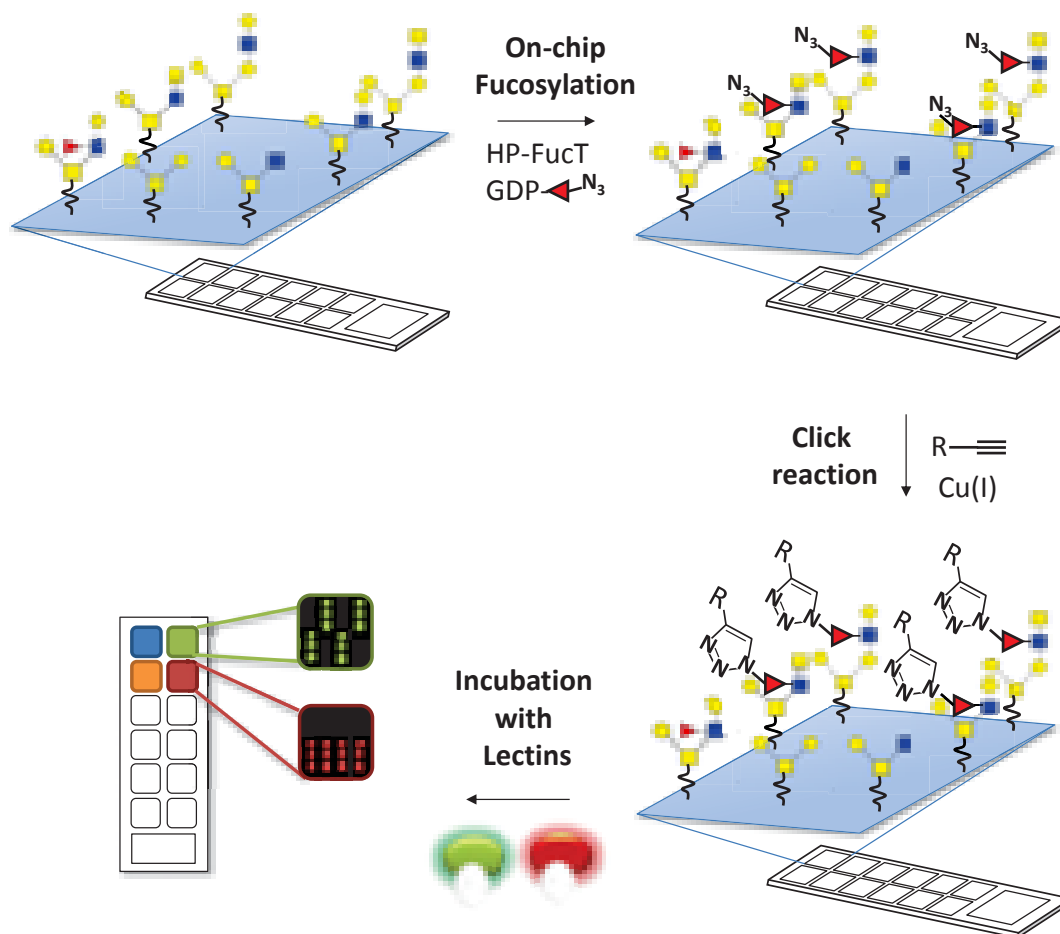
**Figure 27. Final obtained compounds for microarray experiments**

Unfortunately, hydrogenation of **49** and **50** failed as no product was obtained after purification by graphite and no material remained to repeat the deprotection. Glycan **O4** was fucosylated in solution using CeFUT6 to yield **O5** in 73% yield. Considering the expensive amount of the unnatural donor FucZ needed for each compound, we considered using microarray technology to rapidly generate fucose and azidofucose compounds in parallel.

## 3.4 O-glycans and mimetics against CLRs

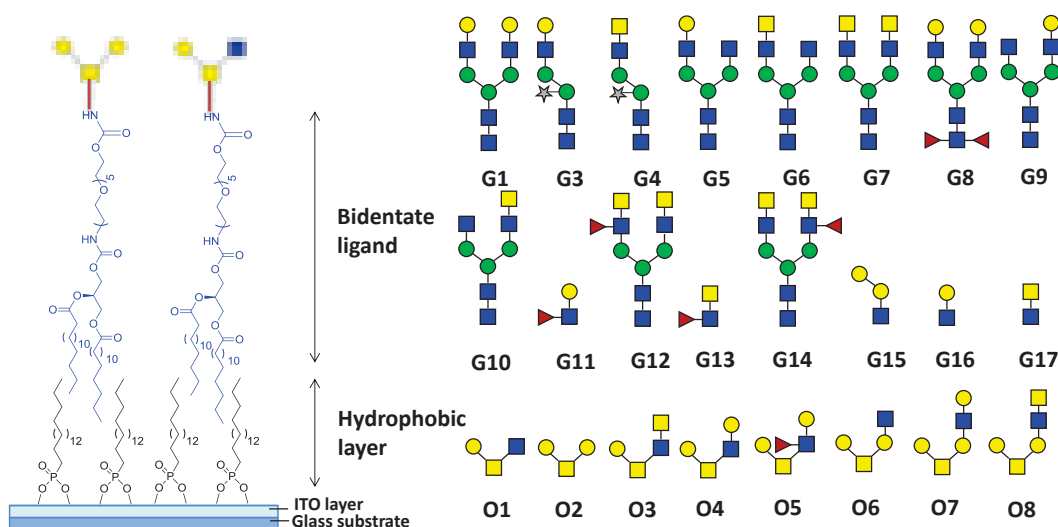
### 3.4.1 On-chip attachments and experiments

An increasing number of reports have documented successful transformations carried out directly on the surface. In particular, enzymatic transformations performed on chip have enabled rapid and high-throughput construction of glycan libraries in minute quantities.<sup>58,70,71</sup> In an elegant study reported by Cory *et al.*, Siglec ligands containing sialic acid and their alkyne-mimetics were printed on chip. After undergoing copper-catalyzed cycloaddition using different azide-substrates, a new library of siglec glycomimetics was rapidly generated and, upon assaying against Siglec 7, led to the identification of high affinity ligands.<sup>72</sup> Adopting a similar strategy, we endeavoured to perform the final enzymatic step- the fucosylation- of our O-glycans on-chip (Figure 28). The use of HP-FucT with GDP-fucose (GDP-Fuc) or GDP-6-azidofucose (GDP-FucZ) would not only generate fucosylated and azido-fucosylated O-glycan mimetics in parallel but would also decrease the need of the expensive azido-donor by several orders of magnitude compared to the solution phase synthesis of individual compounds. Copper-catalyzed (click) reactions between the O-glycans containing FucZ and various alkynes would afford a library of novel glycomimetics to probe against CLRs.



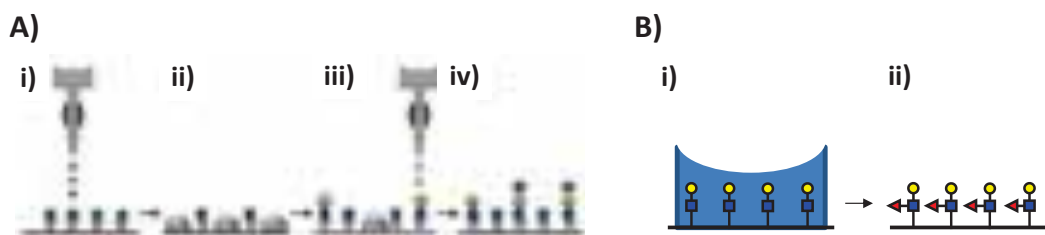
**Figure 1. Microarray strategy for on-chip fucosylation and azido-fucosylation and screening of glycomimetics by interaction with fluorescently labeled CLR**

To this end, the parasitic *O*-glycans were printed alongside a set of N-glycans available from our laboratory (Figure 29) onto NHS activated indium tin oxide (ITO) slides. These slides are composed of activated bidentate lipids (to which the glycans are attached) embedded in a hydrophobic layer. The latter is conjugated to an ITO surface which is on top of a glass slide (Figure 29). The choice of surface allowed for MALDI-TOF MS based analysis of surface transformations (enzymatic and chemical) thus by-passing the traditional need for labelling strategies. The transparent ITP surface was also compatible with fluorescence imaging with a slide scanner hence enabling analysis of CLR-glycan interactions.<sup>73</sup>



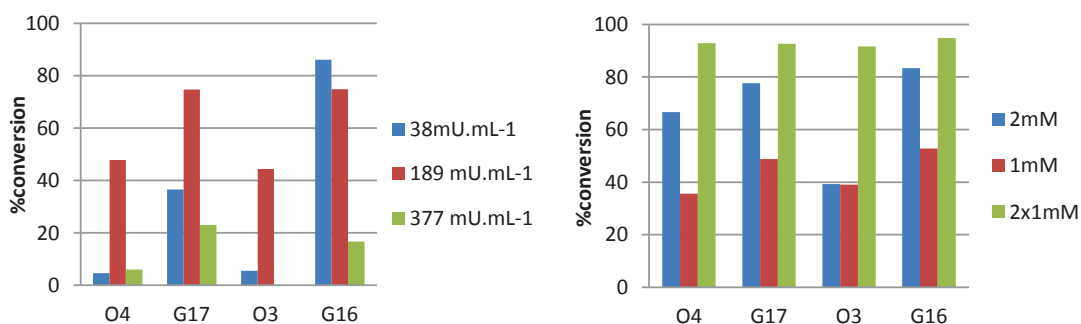
**Figure 29. Glycans printed on activated ITO slides**

In order to explore further miniaturization of the array, we explored spotwise elongation and click chemistry which makes use of the printer to deposit stock solutions of reactions directly onto the printed individual glycan drops (Figure 30).<sup>74</sup> For the enzymatic assay, the presence of an additive was necessary in the stock reaction solution containing the HP-FucT, the donor, cofactors and reaction buffer to avoid rapid drying of spots during incubation as this may result in enzyme precipitation and inhomogeneous reaction. DMSO was chosen over glycerol as the printed spots were observed to better retain their morphology before and after incubation, thus avoiding inhomogeneous dilution of the enzymatic mixture over a given spot. From a stock solution containing the enzymatic reaction with 2% DMSO we printed droplets with a volume of 333 pL on top of each glycan spot and incubated the slides at 37°C. Both GDP-fucose and GDP-6-azido fucose donors were screened in parallel. The conversion for the droplet-based reactions on-chip was evaluated by MALDI-TOF MS as a relative conversion with respect to substrate rather than absolute measures as products may have different ionization efficiencies relative to the substrates. Unfortunately, detection of glycans by MALDI TOF MS was challenging after incubation and washing of the slide. We suspect this to be a consequence of the DMSO which might facilitate removal of the hydrophobically attached glycans during washing of the slide. Nonetheless, only limited enzyme activity was observed for the natural fucose donor (where glycans were found) and no conversion was observed for the GDP-azido-donor.



**Figure 30. Spotwise (A) vs chamber (B) enzymatic elongations. A) i) droplet placement on individual spots, ii) incubation and washing step, iii) spotwise incubation with second glycosyltransferase, and iv) finished high-density array.<sup>74</sup>; B) i) incubation in gasket and washing, ii) finished high-density array**

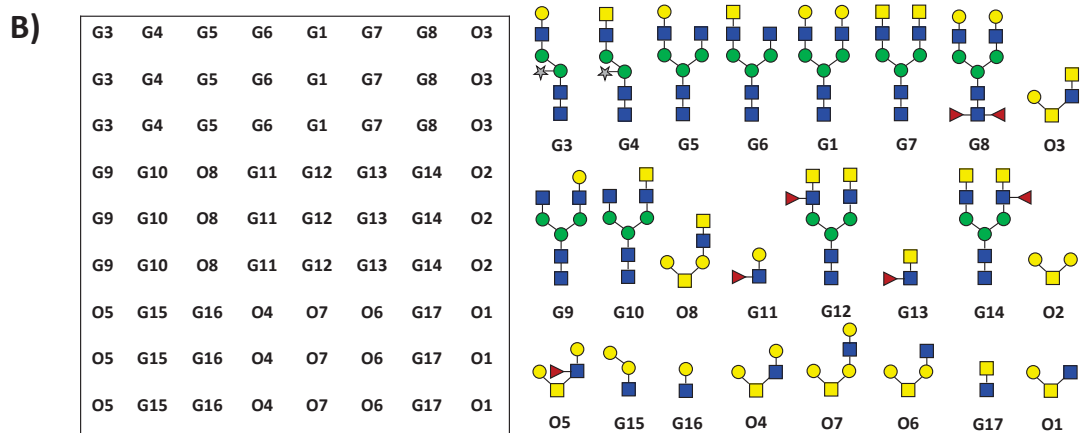
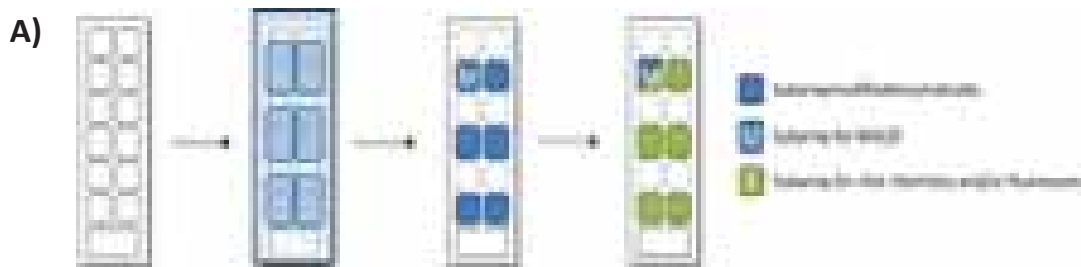
Therefore instead of the droplet based enzymatic reactions that showed unfavourable reaction kinetics, enzymatic transformations were carried out employing a silicone gasket to create wells containing single subarrays on the flat slide. As the subarrays were sealed, no DMSO or glycerol additive was necessary. Donor and enzyme concentrations were screened to determine optimal reaction conditions. These revealed that increasing enzymatic concentration had no effect past  $189 \text{ mU.ml}^{-1}$  (Figure 31, A). Using GDP-fucose, we also observed that repeated exposures ( $\times 16\text{h}$ ,  $37^\circ\text{C}$ ) to freshly prepared enzymatic solution containing  $1\text{mM}$  donor were conducive to higher conversions than a single exposure to the enzymatic solution containing  $2\text{mM}$  donor (Figure 31, B). In the case of the fucose donor, only two incubation cycles were necessary for conversion between  $65\text{-}90\%$  to be obtained whereas 3 cycles were required for the FucZ donor. Therefore enzymatic reactions performed using the gasket were observed to progress significantly better than using the spotwise method.



**Figure 31. Effects on representative ligands of A) enzyme concentration; B) Fucose donor concentration**

Having set up optimal reaction conditions on the surface, we designed the following 24 glycan array which we fucosylated both with the natural and unnatural azido donors separately. The printing workflow is depicted in Figure 32. Although 12 potential reaction fields are possible on a given slide, the following enzymatic elongation step made use of a  $2 \times 4$  hybridization gasket.

The small volume capacity of this gasket (53 $\mu$ l) was advantageous as it reduced quantities of the expensive FucZ used per reaction. However it physically blocked half of the fields. Thus after enzymatic transformations only 6 subarrays were available for further investigations, one of which was set aside to monitor chemical changes (enzymatic and click reactions) by MALDI TOF MS. Hence, only 5 fields were available for further chemistry or fluorescence assays per slide. Click chemistry and ultimately fluorescence assays could be carried out using a 2x8 incubation-chambers gasket (min. 100 $\mu$ l) which overlaid the geometry of our slide perfectly and therefore preserved the remaining fields. This in turn implied one slide per click reaction.



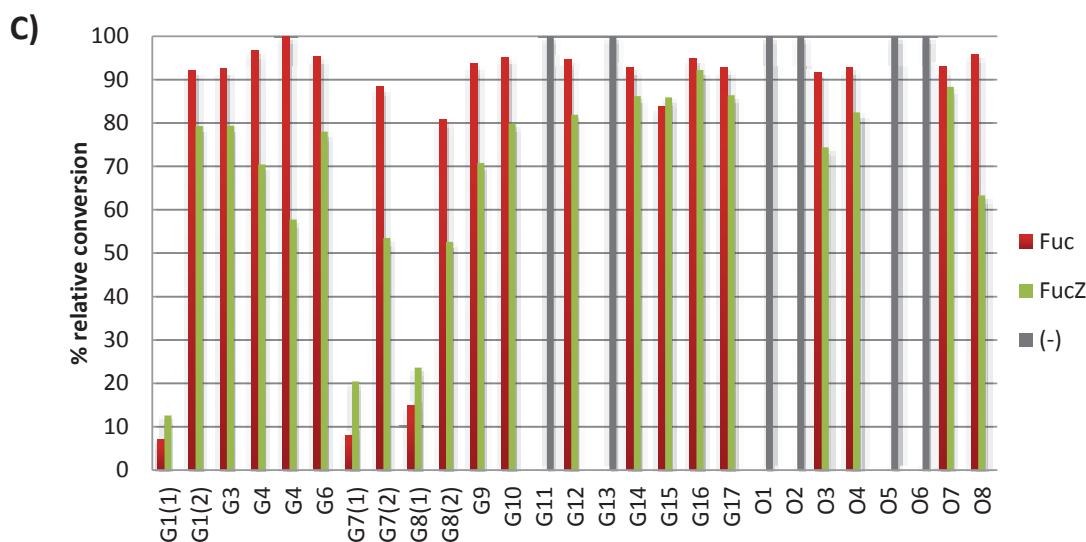


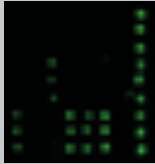
Figure 32. On-chip experiments. A) Workflow and fates of subarrays on the slide; B) The illustrated compounds were printed as 12 subarrays onto NHS-ITO slides in the pattern shown; C) Observed conversion by MALDI for each glycan after fucosylation using GDP-Fuc in 2x1mM cycles (red) and GDP-FucZ in 3x1mM cycles (green) at 37°C overnight. Substrates which were unreactive to enzymatic elongation are shown in grey.

Yields for enzymatic on-chip conversions were determined by MALDI-ToF MS and plotted as ratios of product to the sum of product and remaining starting material plotted for every structure included on the array. The yields showed some structure dependant variability but were all together above 80% for the fucose addition and above 75% for the azidofucose addition (Figure 32, C). Fucosylated are henceforth denoted as **G#F** or **O#F** when GDP-fucose was used and **G#Z** or **O#Z** when GDP-FucZ was used. With the fucosylated and azido-fucosylated libraries in hand, we returned our attention to CLR assays before undertaking further derivitization of the array by click chemistry, for a more informed decision on the choice of alkyne.

### 3.4.2 Effect of in-situ fucosylation on CLR binding

To evaluate the biofunctionality of the printed glycans, the array was incubated with the plant lectin, peanut agglutinin (PNA) specific for Gal $\beta$ -1,3GalNAc residues. As expected, only the *O*-glycans **O1-O8** containing the lectin ligand showed fluorescence (Table 6). The native array was then screened against the fluorescently tagged recombinant human CLRs which had been made available a network partner: DC-SIGN(ECD), DC-SIGN R(ECD), DC-SIGN R(CRD), and MGL. Functions and specificities of the selectins are summarized in Table 6.

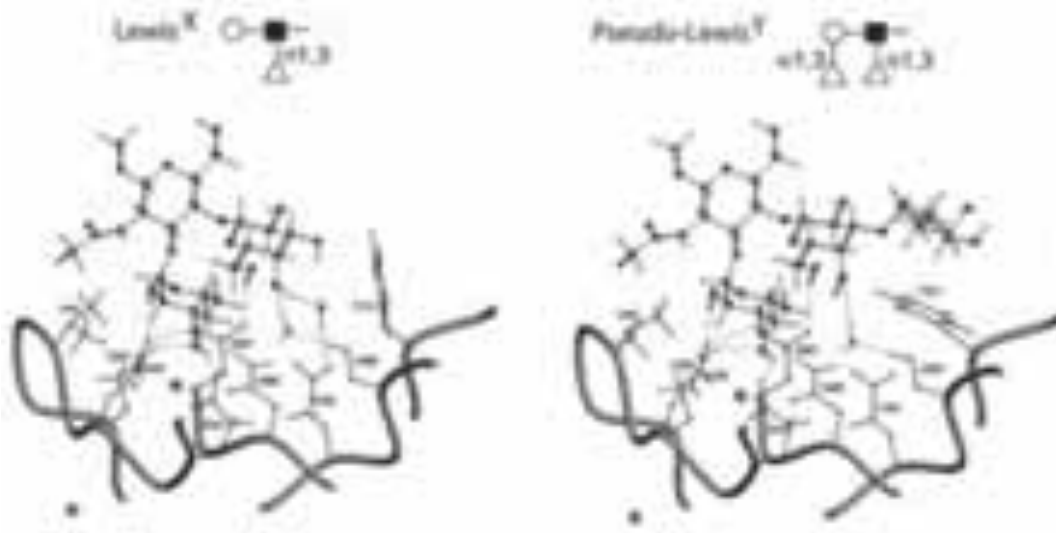


Lectin	Ligandspecificity	Functions
PNA	Gal $\beta$ 1,3GalNAc 	unknown
DC-SIGN	Mannan, LeX, Lea, Ley, Leb, SLea, ManLam	Pathogen recognition/presentation Caption and transmission of HIV-1
DC-SIGN R	Mannan, Lea, Ley, Leb	Caption and destruction of HIV-1
MGL	Terminal GalNAc	Pattern Recognition Receptor

**Table 6. Lectins screened on the glycan array**

### 3.4.2.1 DC-SIGN

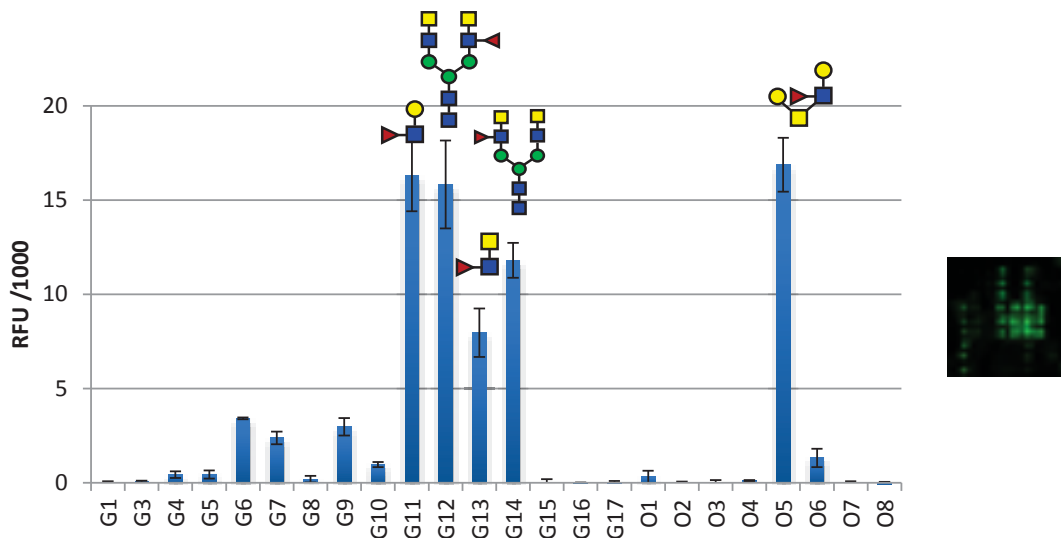
DC-SIGN (also CD209) is one of the most studied CLRs to date, alongside the mannose receptor, including reports on inhibition studies, glycan array and STD-NMR binding studies, crystal structure elucidation and development of specific glycomimetics targeting it.<sup>75,76,77</sup> It is found on DCs and macrophages as a tetramer and has an overlapping glycan specificity with Langerin, recognizing mannose, fucose and GlcNAc, but unlike Langerin it has been shown to bind Lewis a,b,x and Y (Fuc $\alpha$ 1-2Gal $\beta$ 1-4(Fuc $\alpha$ 1-3)GlcNAc). In fact, crystallographic, NMR and modelling experiments have revealed the details of interactions of LeX in the CRD of DC-SIGN which involves chelation of the 3- and 4-OH of the  $\alpha$ 1,3 linked fucose residue by Ca<sup>2+</sup>. According to these reports, the 2-OH of fucose would interact with the proximal Val<sup>351</sup> via strong Van der Waals interactions and the terminal galactose was proposed to contact Phe<sup>313</sup> in a secondary binding site. This model was challenged by Meyer *et al.* when inspecting the interaction between pseudo-Lewis Y (Fuc $\alpha$ 1-3Gal $\beta$ 1-4(Fuc $\alpha$ 1-3)GlcNAc, a Lewis antigen specific to schistosomes) and DC-SIGN to propose a certain degree of flexibility of the secondary binding, allowing Phe<sup>313</sup> to change orientation to accommodate the terminal fucose.



**Figure 33. Models of interaction of DC-SIGN with Le<sup>X</sup> and pseudo-Le<sup>Y</sup>. Calcium ions are represented by gray spheres.<sup>78</sup>**

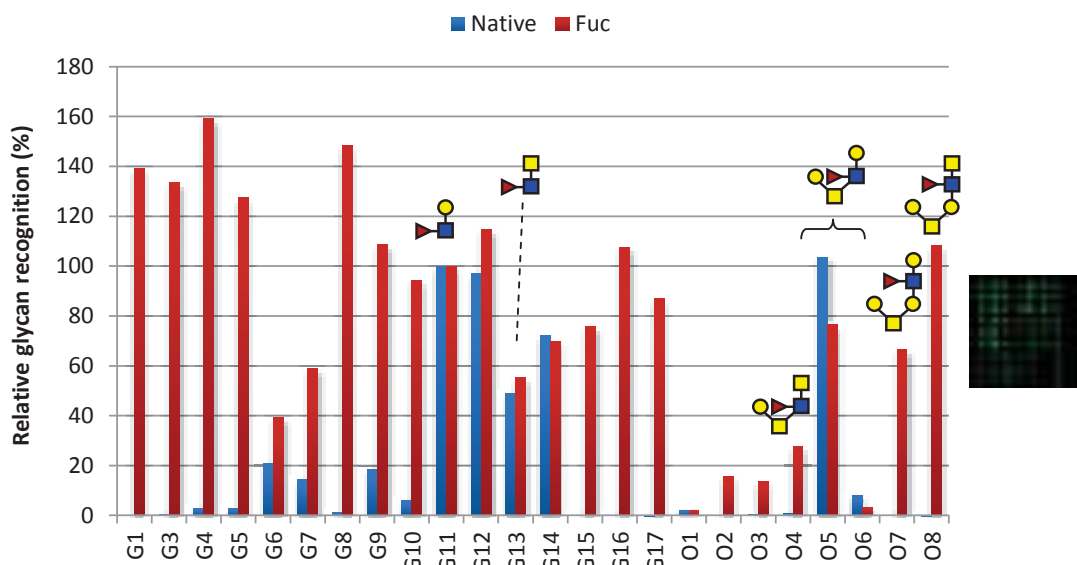
Although additional proof is needed, flexibility in a secondary binding site could therefore allow for different binding modes and consequently entail different signalling pathways which could be exploited by *S.mansoni* to ensure survival.<sup>78</sup> *S.mansoni* N-glycans bearing Le<sup>X</sup> and LDNF epitopes have been shown to bind to DC-SIGN.<sup>79</sup> These N-glycans compose the soluble egg antigens which was shown to inhibit DC activation.<sup>80</sup>

In view of these results, we wondered whether a presentation of the antigenic epitopes on *O*-glycans would somehow induce a different binding to those previously reported on *N*-glycans. A first examination of screening the native array with DC-SIGN confirmed the lectin's affinity for fucose and its specificity for presentation as a terminal epitope, as only glycans containing Le<sup>X</sup> epitopes showed high fluorescence intensity (RFU>10000). A preferential binding of **G11** over **G13** was noted, suggesting specificity for Le<sup>X</sup> over LDNF. However, this was not fully reflected in the comparison between the N-glycans **G12**, **G14** bearing LDNF and **O5** bearing Le<sup>X</sup>. The relatively similar fluorescence intensities of these compounds emphasize the importance of structural context of glycan elements in molecular recognition. This was additionally emphasized by direct comparison of **G12**, **G14** with **G13** as the N-glycans showed 1.5-2 times stronger fluorescence intensity than the trisaccharide.



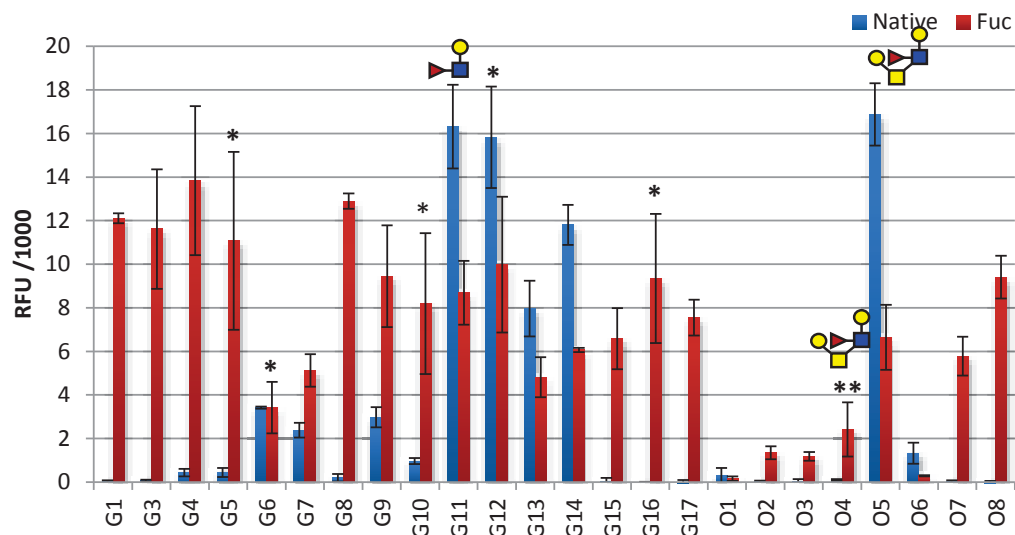
**Figure 34.** Fluorescence intensity after incubation of the native array with DC-SIGN. Each bar in the histogram represents the average of fluorescence from 3 spots and the standard deviation as an error bar

From these results and literary precedents, fucosylation of our array was anticipated to make all elongated glycans ligands susceptible to binding to the CLR and this was indeed observed to be the case (Figure 35). Using **G11** as an internal array standard, we observed that the fucosylated N-glycans were the best binders as they showed marginally higher fluorescence intensity. Although O-glycans **O7F** and **O8F** also showed a marked increase in fluorescence, we noted that it was not as high as for the N-glycans. The preferential binding of LeX over LDNF was also observed to be generally conserved in the N-glycans whereas the trend was reversed in the O-glycans, with **O8F** being 40% higher than **O7F** and **O5** being 60% higher than **O3F**. The absence of binding of **O3F** contrasted interestingly with the *S.mansoni* analog **O8F**.



**Figure 35. Relative fluorescence intensities of the native array (blue bars) and the fucosylated array (red bars) after incubation with DC-SIGN and normalized to G11**

The presence of glycans unreactive to transformations performed on-chip allows for the evaluation of data reliability. Thus, **G11** and **G13** were both unreactive to enzymatic transformations and were observed to keep a relative fluorescence ratio of 2:1. However, **O5** was also unreactive to enzymatic elongation yet a variation of 20% relative to **G11** was observed. Upon closer inspection of the data, several discrepancies were apparent. Notably, the signal for **O4F** was expected to be similar to, if not match, its internal reference **O5** given the 90% fucosylation observed by MALDI-TOF MS (Figure 35). However this was not observed to be the case, yielding barely any significant fluorescence (**O4F**= 2416 RFU). While this could be considered as the effect of the residual **O4**, a closer inspection into the degree of variability of the data revealed some glycan profiles to be disputable. The degree of variability of a given glycan's fluorescence intensity is given by its coefficient of variation (%CV). This is measured by calculating the standard deviation of the average fluorescence intensities for replicates spots divided by the mean and multiplied by hundred ( $100 \times \text{standard deviation}/\text{mean}$ ). If the %CV is high (>20–40%), the results for binding may not be reliable and require a closer inspection, and if the %CV is >50% the data should be disregarded. The average of the spots is normally taken over 4 out of 6 glycan replicates, the 2 outliers having been discarded for a truer representation of the sample.<sup>81</sup> Unfortunately samples in our array were printed in triplicates as better spot-to-spot reproducibility was anticipated. This could explain the inconsistencies observed in our dataset and the suboptimal %CV in some instances (Figure 36).

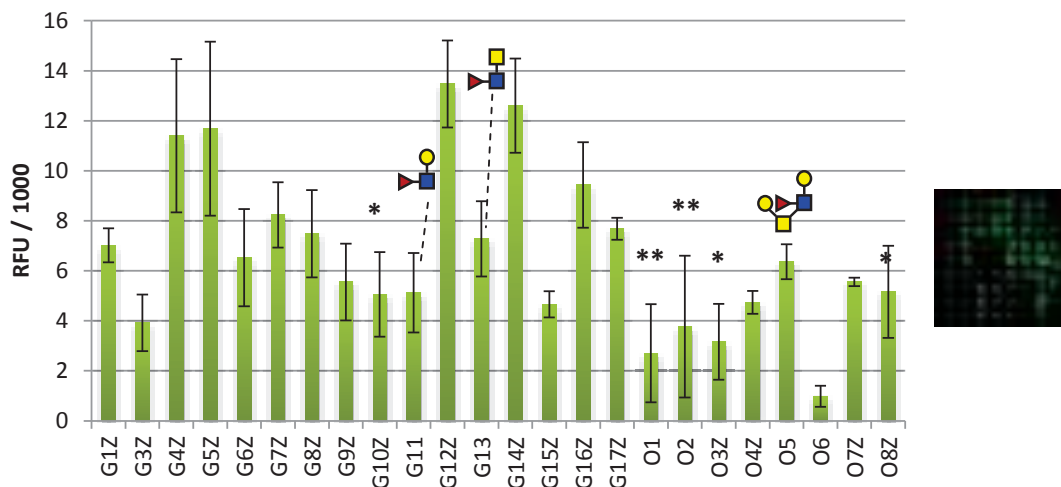


**Figure 36. Fluorescence intensities for the native array (blue bars) and the fucosylated array (red bars) after incubation with DC-SIGN (10µg/ml). Each bar in the histogram represents the average of fluorescence from 3 spots and the standard deviation as an error bar, \* denotes 30%<CV<50, \*\* denotes %CV>50**

We also noted a decrease in overall intensity of fluorescent signals in the fucosylated array compared to the native array. This was attributed to the additional two washing steps performed after enzymatic elongation, during which some glycan material may be removed. Nevertheless, fucosylation undeniably made most of the array suitable ligands for DC-SIGN and revealed interesting aspects of the lectin's specificity.

Before examining how triazole side-chains from click-reactions could affect the ligand binding affinity, we evaluated the minimal change in the fucose, the C-6 azide. Overall the azido-fucosylated array displayed a similar profile to naturally-fucosylated array. However, the quality of the fluorescence was visibly lower as smearing was observed. The overall fluorescence intensity was also observed to decrease compared to the native array and the fucosylated array. Notably, the internal standard **G11** was reported at 5123 RFU compared to 8693 RFU in the fucosylated array and 16322 RFU in the native array (Figure 36). This put the accuracy of the binding into question. Critically, glycan standards **G11** and **G13** should maintain their relative fluorescence intensities regardless of array modifications and while a 2:1 ratio was roughly observed for the native and the naturally fucosylated array, the ratio was calculated at 0.7:1 in the azido-fucosylated array. Faced with this unexplained change in lectin preferential binding, it was difficult to assess whether the **G16Z** was indeed a better binder than **G11**, or whether **O7Z** and **O8Z** did indeed bind similar to **O5**. We noted though that as

with fucosylation, the azido-fucosylation of the array made a certain number of the compounds ligands for the lectin DC-SIGN to varying extents.



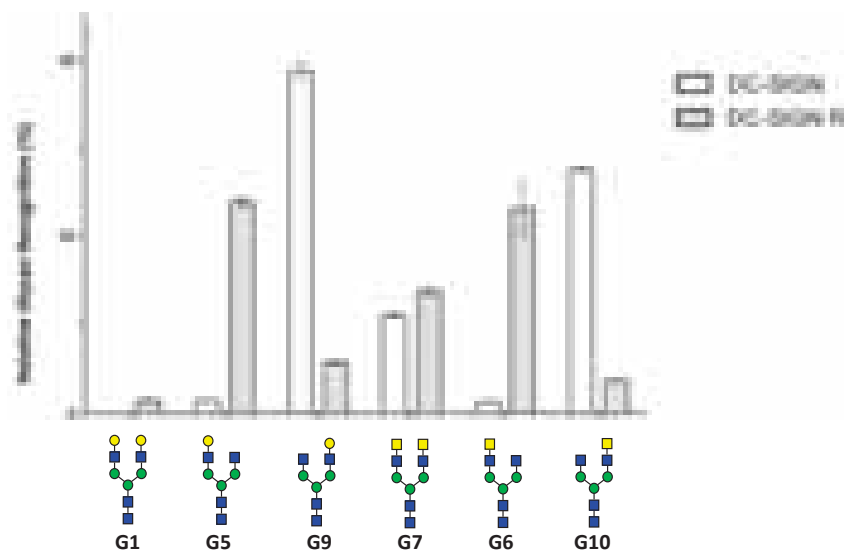
**Figure 37. Fluorescence intensities for the azido-fucosylated array after incubation with DC-SIGN (10µg/ml).** Each bar in the histogram represents the average of fluorescence from 3 spots and the standard deviation as an error bar, \* denotes 30<%CV<50, \*\* denotes %CV>50

Considering the high degree of uncertainty observed in the azido-fucose array, it was deemed unreasonable to pursue click chemistry on the array. Semi-successful printing of the glycans on a slide was omitted as the main reason for error as the slides were systemically checked by MALDI-TOF MS post-printing. Although arrays are supposedly robust to slide washes, a difference in washing buffers used may be responsible for some material loss.<sup>73</sup> Indeed, an aqueous solution containing acetonitrile (0.05%v/v) and TFA (0.1%v/v) was used to wash after each transformation performed on the slide. In investigations carried out by our laboratory, this was found to be optimal for the removal of any precipitated enzyme, the presence of which could affect the fluorescence studies. A repeated exposure to the solvent (3 enzymatic and 1 lectin for the azido-fucose array) therefore probably partially removed some glycan material and resulted in an inhomogeneous sample distribution, a correspondingly skewed lectin binding profile and a lower fluorescence intensity. Further investigations into obtaining a compromise between high enzymatic conversions and less slide washes should therefore be carried out to obtain a reproducible binding experiment before investigating the effects of different glycomimetics.

### 3.4.2.2 DC SIGN R

Albeit 77% homologous to DC-SIGN, DC-SIGNR (also known as L-SIGN, CD299) has been less intensively studied than DC-SIGN and its function remains unknown although it has

been shown to be implicated in the spread or persistence of several viruses.<sup>82</sup> It is located on endothelial cells which are the cells that line liver cells and all blood and lymphatic vessels.<sup>83</sup> Endothelial cells therefore represent an additional barrier that *S.mansoni* larva must cross to begin their migration towards the intestine and liver and a barrier to their energy supply of erythrocytes (red blood cells) and glucose.<sup>84</sup> They also represent the additional challenge to *S.mansoni* eggs which must cross the intestine wall where they are deposited by the adult worms to penetrate the organ in order to be eventually excreted. Interestingly, only 50% of the eggs laid are eventually excreted, with the other 50% being recirculated towards the liver, intestines and urinary tract where they become trapped, leading to granulomous responses causing the main pathology of schistosomiosis. DC-SIGNR could therefore be a vital target for the worm's survival. It has been shown to bind both schistosome egg antigens (SEA) and glycosphingolipids and to mediate the internalization of SEAs.<sup>85</sup> However its binding profile contrasts to DC-SIGN notably by its specificity in binding neoglycoconjugates carrying Lewis a,b and y but excluding Lewis X. This arises from a loss of complimentary forces such as Van der Waals and hydrogen bonding in the primary binding site due to substitution of valine<sup>351</sup> by serine<sup>363</sup>.<sup>86</sup> It has also been described to bind mannose-containing glycans as those found in the N-glycans on HIV-1's envelope protein.<sup>87</sup> Recently, contrasting specificities for isomeric N-glycans **G5/G9** and **G6/G10** was observed for DC-SIGN and DC-SIGN R (Figure 39).<sup>49</sup> In particular, DC-SIGN R showed preferable binding to the biantennary 3-monogalactosylated **G5** but not the 6-monogalactosylated isomer **G9**. The same behaviour was observed for the isomeric pair **G6/G10** displaying terminal GalNAc, **G6** being bound with higher affinity than **G10**. The complete opposite was observed to be true for DC-SIGN. A similar binding was observed for both lectins towards the bis-GalNAc **G7** whereas the bis-Gal **G1** showed less binding affinity.



**Figure 38. Comparison of %RFU values of monogalactosylated and N-glycans with terminal single GalNAc residues in DC-SIGN and DC-SIGN R binding<sup>49</sup>**

Considering the potential DC-SIGNR has in determining successful evolution of *S.mansoni* in its definitive host, we decided to investigate this CLR. Although Lewis X and LDNF epitopes have not been previously reported to be strong ligands for DC-SIGNR, we proposed to query whether our novel presentation on parasitic *O*-glycan cores would affect binding affinity. Firstly, striking contrasts in binding specificities were observed on our array between the two available constructs of DC-SIGN R, the CRD and the ECD (Figure 39). Indeed the CRD appeared to bind the regioisomers **G5** and **G9** indiscriminately but not the GalNAc isomers **G6** and **G10** whereas all four of these ligands were bound by the ECD. This initially suggested a specificity of the CRD which excluded terminal GalNAc residues. Yet monofucosylated N-glycans **G12** and **G14** were observed to bind suggesting that the presence of fucose enables ligand binding to the CRD. This was additionally supported by the observed binding of **G8**, **G11** and **O5**. The LDNF epitope **G13** was not observed to bind in the CRD suggesting the importance of ligand presentation. In comparison, the ECD specificity generally seemed to overlap with the CRD's although some disparities were seen. For example, *O*-glycans **O7**, **O8**, **O5** and **G15** were observed to bind in the CRD whereas only **O1** and **O6** were bound in the ECD.



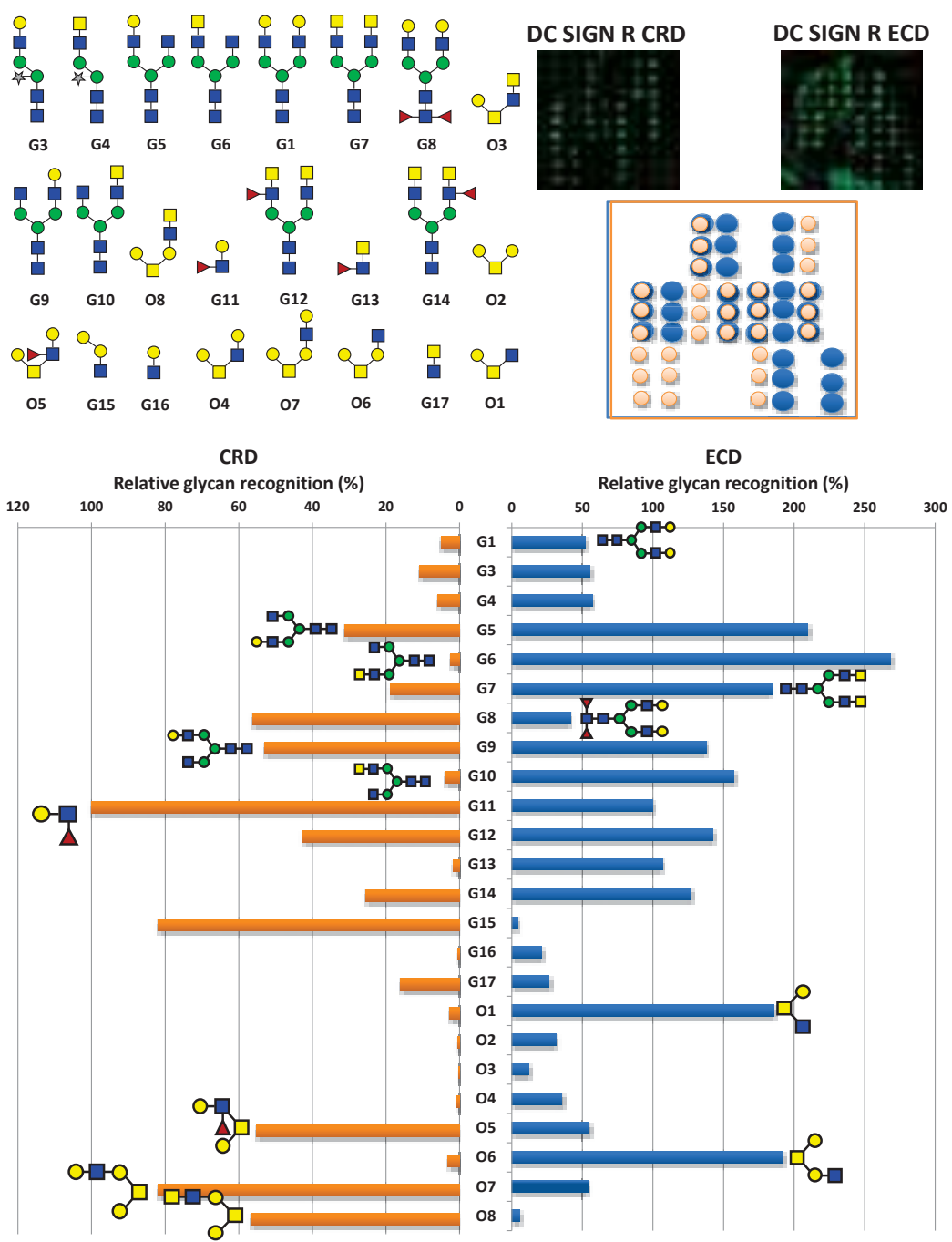


Figure 39. Comparison of two constructs of DC-SIGN R. Schematic representation of glycans in printed order on array. Fluorescence scans of DC-SIGN R CRD (10ug/ml), ECD (50ug/ml) and pictogram representation of the overlay of both images. Fluorescence intensities for the native array after incubation with each DC-SIGN R construct: the CRD (orange bars) and the ECD (blue bars) .

For a more detailed comparison, the fluorescent intensities were normalized to **G11**. This was the best binder in the CRD(8684 RFU) and only two smaller O-glycans, **G15** and **O7** displayed similar binding affinities. In contrast, **G11** was a poor ligand for binding in the ECD (2645 RFU). Indeed, the N-glycans constituted the majority of the ECD binding profile and only the O-

glycans **O6** and **O1** displayed significant binding which was attributed to the terminal GlcNAc moiety. The overall specificity of the ECD was as previously reported, with **G5** and **G6** being preferentially bound over **G9** and **G10** respectively. Furthermore, the bis-GalNAc **G7** was also preferentially bound over **G1**, in accordance with previous results although we noted that monogalactosylated isomers **G5** and **G9** were better recognized than the GalNAc isomers **G6** and **G10**. Fucosylated glycans only showed mild fluorescence intensities in comparison suggesting that fucose reduced binding of ligands in the lectin ECD.

Interestingly, binding strength of the ligands changed for the CRD. Thus, **G9** relative fluorescence was higher than **G5**, while GalNAc isomers **G6** and **G10** showed similar fluorescence intensities. The **G1-G7** relationship was conserved but to a lesser extent than in the ECD. As mentioned above, the O-glycan binding profile also changed remarkably and binding of **G15**, **O7**, **O8** and **O5** suggested a loss in GlcNAc specificity initially observed in the ECD. Moreover, we noted that the LacNAc epitope of **O7** favoured stronger binding than its GalNAc analog **O8** by 20%. It was also interesting to note that both of these O-glycans based on the *S.mansoni* core showed binding whereas the mucin core 2 based O-glycans **O3** and **O4** showed none. Finally, unlike in the ECD, fucosylation enabled ligand recognition. This was evident by comparing **O4** to **O5**, **G12** and **G14** to **G1**, and **G8** to **G1**.

These first differential binding profiles showcase the intricacy in glycan-lectin interaction studies. They also emphasize the importance of structural and spatial considerations of both the ligand and the lectin. The CRD can be considered as the minimal binding entity with a defined glycan binding profile. The ECD construct of DC-SIGN R used for our studies consists of a tetramer of CRDs with a number of neck repeat domains. Therefore, tetramerization can alter the lectin's overall specificity and affinity as a function of the spatial arrangements of the CRD binding sites, making them more or less accessible. Although a crystal structure of DC-SIGN R provided a rationale for the lectin's interaction with fucose, it was based on the CRD and did not include the effect of multimerization of the CRDs.<sup>86</sup> The variability of the binding affinities from the single domain to the multimeric lectins adds yet another layer of complexity in the targeting of CLRs. The ECD construct might resemble most to the natural presentation of the lectin on the cell-surface and therefore seemed most pertinent to our CLR targeting studies. However, whereas only a 10ug/ml concentration was needed for all other lectins investigated, substantially larger amounts of DC-SIGN R ECD were required for appreciable binding data. This was first observed for the native array where even using 50ug/ml at laser power 80 only afforded 7093 RFU for the highest binding glycan **G6**. As a consequence, incubation of the

lectin with the fucosylated arrays led to unsatisfactory binding intensities. Critically, control spots **G11** and **G13** were observed to disappear after fucosylation. As with DC-SIGN, we suspected glycans to be removed during the washes of the slides after enzymatic elongation. Combined with the absence of strong binders in the array and the low degree of labelling for the lectin, this made the analysis of the fucosylated arrays inconclusive.

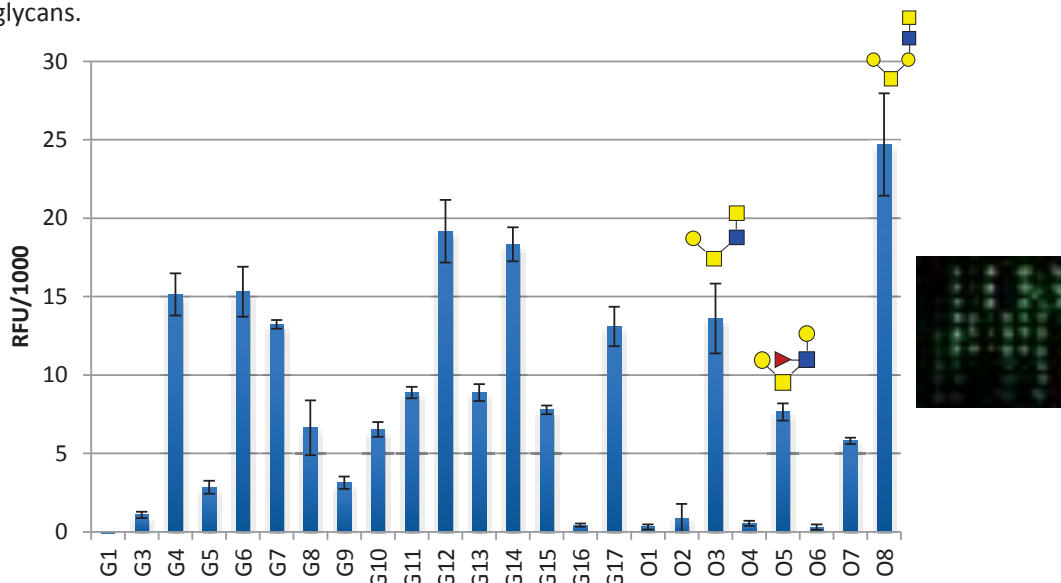
Using the ECD with a higher degree of labelling might significantly improve the fluorescence data. As for DC-SIGN, a compromise between high enzymatic conversions and less slide washes was also needed to obtain a reproducible binding experiment.

### 3.4.2.3 Macrophage galactose Lectin (MGL)

MGL (CD301) was of substantial interest to our studies as it was shown to be a pattern recognition receptor for *S.mansoni* glycans due to its strong reported interaction with helminth glycans containing LDN/LDNF structures.<sup>53</sup> This makes it highly susceptible to hijacking by the parasite for immune evasion. Studies have reported soluble egg antigens to be internalized into DCs expressing MGL and confocal laser scanning microscopy revealed colocalization of the soluble egg antigens with MHC-II in the lysosomal compartments suggesting antigen processing and presentation. However, no signalling function has been identified for the lectin so far although it has been suggested to play a role in downregulating effector T cell function.<sup>88</sup> This would also explain how adenocarcinoma cells which overexpress the Tn-antigen at the cell surface ensure survival and proliferation, targeting MGL in the same way HIV-1 targets DC-SIGN for infection as opposed to routing to the lysosomal compartments for degradation.<sup>89,90</sup>

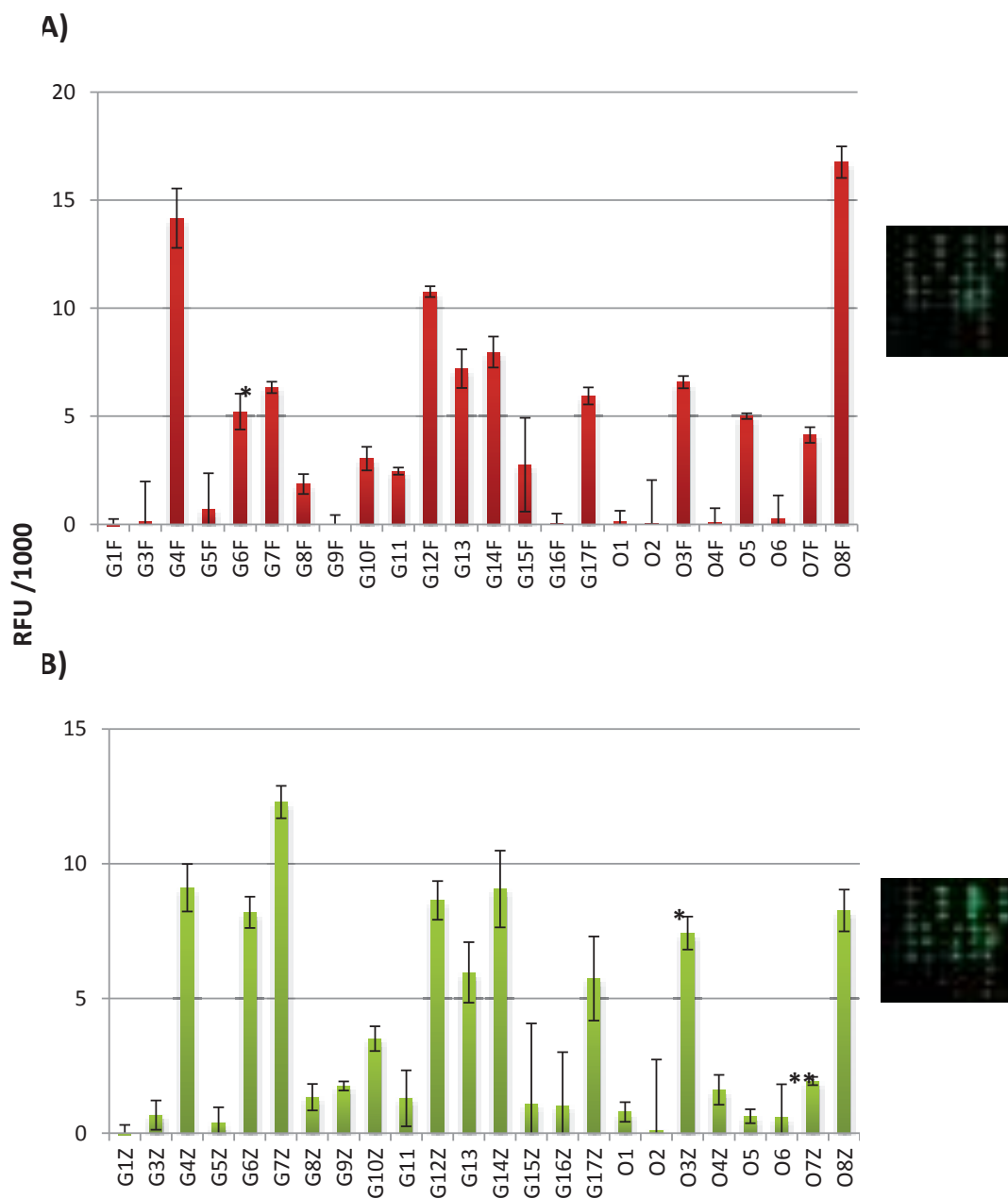
MGL is expressed as a trimer on DCs and macrophages in the small intestine, lymph node and skin and displays high specificity for terminal GalNAc residues. Based on a molecular model, the QPD motif of MGL's CRD was rationally envisaged to accommodate binding to galactosamine and galactose through chelation of the 3- and 4-OH to the calcium. The lectin's specificity was proposed to originate from complementary interactions of the 2-acetamido group in GalNAc in the binding site although this remains to be determined.<sup>91</sup> In line with these findings, differential affinity was observed for the LDN containing compounds depending on their presentation. Thus **O3** showed similar interaction as the minimal binding epitope **G17** but **O8** bound nearly 200 times stronger than both of these glycans. The N-glycans containing terminal GalNAc residues (**G4**, **G6**, **G7**, **G10**, **G12**, **G14**) also showed appreciable binding although less than **O8**.

While GalNAc specificity was definitely confirmed, mild binding was observed for some compounds not presenting terminal GalNAc. Indeed, the LeX epitope **G11** and compound **O5** also displayed the same degree of binding as the LDNF epitope **G13**. This contrasts to the negligible binding observed by van Vliet *et al.* We initially considered the hypothesis that the fucose in **O5** was responsible for the observed fluorescence, as the non-fucosylated analog **O4** showed no binding. Moreover, monofucosylated N-glycans **G12** and **G14** showed approximately 30% more binding compared to the non-fucosylated **G7**, and the LeX epitope **G11** showed 95% more binding than the LacNAc epitope **G16**. On other hand, the LDNF epitope **G13** was observed to bind 68% less than the LDN epitope **G17**. This was in general accordance with the 75% reported by van Vliet *et al.* Also surprisingly, **O7** and **G15** displayed mild binding to the lectin despite the latter not containing any GalNAc residue or resembling any of the O-glycans.



**Figure 40.** Fluorescence intensities for the native array after incubation with MGL (10ug/ml). Each bar in the histogram represents the average of fluorescence from 3 spots and the standard deviation as an error bar

Clearly fucosylation was an interesting aspect to probe. Incubation of the lectin with the fucosylated array showed no difference in the overall binding profile.



**Figure 41. Fluorescence intensities for the fucosylated and the azido-fucosylated array after incubation with MGL (10µg/ml). Each bar in the histogram represents the average of fluorescence from 3 spots and the standard deviation as an error bar, \* denotes 30%<CV<50, \*\* denotes %CV>50**

While this could be thought as **O5** being a unique outlier ligand, it was concerning that **O4F** did display fluorescence despite the quasi quantitative enzymatic conversion was reported (98.9%)observed by MALDI-TOF MS. Additional discrepancies were observed in the data. Notably, **G11** binding was observed to decrease to 34% respective to **G13** whereas it was

initially observed to be equal. Considering the high 44%CV associated with **G11**, the data could be considered as misrepresented. As with incubation with previous glycan arrays, the overall fluorescence intensity of the glycan array was observed to have decreased after fucosylation. Thus, **G11** initially showed significant binding of 8885 RFU yet this decreased to 2462RFU in the fucosylated array and 1225 RFU in the azido-fucosylated array. The azido-fucosylated array displayed even lower fluorescence intensities and consequently even higher %CV (>50%) such that the results were inconclusive. As in the DC-SIGN experiments, we postulate that excessive washes of the slide due to the several enzymatic cycles for satisfactory conversion were the reason for the observed decay in fluorescence and the subsequent deterioration of data quality. However, it was clear that the lectin specificity remained unchanged after fucosylation (or azido-fucosylation) as the glycans displaying terminal GalNAc remained the best binders.

Overall, some interesting features were observed from the array. Although quantitative evaluations were not possible owing to a lack of reproducibility, the effect of fucosylation and azido-fucosylation of the glycan array was observed for DC-SIGN and MGL. Future work should include efforts to remedy the previously mentioned issues for a more robust and therefore more reliable analysis.

## References

1. N. Mathieux, H. Paulsen, M. Meldal, K. Bock, *J. Chem. Soc. Perkin Trans. 1*, **1997**, 2359-2368
2. J. Xia, J. Xue, R.D. Locke, E.V. Chandrasekaran, T. Srikrishnan, K.L. Matta, *J. Org. Chem.*, **2006**, 71, 3696-3706
3. M. Hollinger, F. Abraha, S. Oscarson, *Carb. Res.*, **2011**, 346, 1454-1466
4. D. Benito-Alifonso, R.A. Jones, A-T. Tran, H. Woodward, N. Smith, and M. C. Galan, *Bellstein J. Org. Chem.* **2013**, 9, 1867-1872
5. B. Echeverria, J. Etxebarria, N. Ruiz, A. Hernandez, J. Calvo, M. Habegger, D. Reusch, N-C. Reichardt, *Anal. Chem.*, **2015**, 87, 11460-11467
6. S.D. Cao, Z.H. Gan, R. Roy, *Carb. Res.*, **1999**, 318, 78-81
7. Z. Zhang, I.R. Ollmann, X-S. Ye, R. Wischnat, T. Baasov, C-H. Wong, *J. Am. Chem. Soc.*, **1999**, 121, 734-753
8. U. Ellervik, G. Magnusson, *Carb. Res.*, **1996**, 280, 251-260
9. T.B. Windholz, D.B.R. Johnston, *Tetrahedron Lett.*, **1967**, 2555
10. L. P. Calle, B. Echeverria, A. Franconetti, S. Serna, M. C. Fernández-Alonso, T. Diercks, F. J. Cañada, A. Ardá, N.-C. Reichardt, J. Jiménez-Barbero, *Chemistry-A European Journal*, **2015**, 21, 11408-11416
11. J-C.Lee, C-Y.Wu, J.V. Apon, G. Siuzdak, C-H. Wong, *Angew. Chem. Int. Ed.*, **2006**, 45, 2753-2757
12. F. Safoura, *Res.J.Chem.Sci.*, **2014**, 4, 25-28
13. H.R. Appelt, J.S. Oliveira, R.C.V. Santos, O.E.D. Rodrigues, M.Z. Santos, E.F. Heck, L.C. Rosa, *International J. Carbohyd.Chem.*, **2013**, 3, ID 320892
14. C. Pedersen, J. Olsen, A. Brka, M. Bols, *Chem. Eu. J.*, **2001**, 17, 7080-7806
15. U. Ellervik, G. Magnusson, *Tetrahedron. Lett.*, **1997**, 38, 1627-1628
16. J. Ohlsson, G. Magnusson, *Carbohydr. Res.*, **2000**, 329, 49-55
17. J. Khamsi, R.A. Ashmus, N.S. Schocker, K. Michael, *Carbohydr. Res.*, **2012**, 357, 147-150
18. B.Sun, A. V. Punkin, G.M. Visser, H. Zuilhof, *Tetrahedron Lett.*, **2006**, 47, 7371-7374
19. C-R Shie, Z-H. Tzeng, C-C Wang, S-C. Hung, *J. Chin. Chem. Soc.*, **2009**, 56, 510-523
20. M. U. Roslund, O. Aitio, J. Wärna, H. Maaheimo, D.Yu. Murzin, R. Leino, *JACS*, **2008**, 130, 8769-8772
21. L.K. Mydock A. V. Demchenko, *Org. Lett.*, Vol. 10, **2008**
22. X. Zhu, T. Haag, R.R. Schmidt, *Org. Biomol. Chem.* **2004**, 2, 31-33

23. H-W.Yeh, T-S Lin, H-W.Wang, H-W.Cheng, D-Z.Liu, P-H. Liang, *Org. Biomol. Chem.* **2015**, 13, 11518-11528
24. T.B. Windholz, D.B.R. Johnston, *Tetrahedron Lett.*, **1967**, 27, 2555-2557
25. H. Hancock, I.J. Galpin, B.A. Morgan, *Tetrahedron Lett.* **1982**, 23, 249
26. C-y. Huang, N. Wang, K. Fujiki, Y. Otsuka, M. Akamatsu, Y. Fujimoto, K. Fukase, *J. Carb. Res.*, **2010**, 29, 289-298
27. H. Liu, Y. Zhang, R. Wei, G. Andolina, X. Li, *JACS*, **2017**, 139, 13420-13428
28. U. Jacquemard, V. Beneteau, M. Lefoix, S.Routier, J. Merour, G.Coudert, *Tet.* **2004**, 60, 10039-10047
29. A. Seko, K. Yamashita, *Glycobiology*, **2005**, 15, 943-951
30. S. Koizumi, *Handbook of Carbohydrate Engineering, Chap.10*
31. M.D. Leipold, E. Vinogradov, C. Whitfield, *J. Biol. Chem.*, **2007**, 282, 26786-26792
32. D.J. Namdjou, H.M. Chen, E. Vinogradov, D. Brochu, S.G. Withers, W.W. Wakarchuk, *Chembiochem.*, **2008**, 9, 1632-1640
33. 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. G. Wang, *Chem. Sci.*, **2015**,6, 5652-566
34. O. Blixt, I. van Die, T. Norberg, D.H. van den Eijnden, *Glycobiology*, **1999**,9, 1061-1071
35. K.Naruchi, T.Hamamoto, M.Kuroguchi, H.Hinou, H. Shimizu, T.Matsushita, N.Fujitani, H.Kondo, S-I.Nishimura, *J. Org. Chem.*, **2006**, 71, 9609-9621
36. W.Guan, L.Ban, L.Cai, L.Li, W.Chen, X.Liu, M.Mrksich, P.G.Wang, *Bioorg.Chem.Lett.*, 21, **2001**, 5025-5028
37. G.J.F.Chittenden, J.G.Buchanan, *Carbohydr.Res.*, **1969**, 11, 379-385
38. H.M. Christensen, S. Oscarson, H.H. Jensen, *Carbohydrate Research*, **2015**, 408, 51-95
39. Z.Li, J.C.Gildersleeve, *JACS*, **2006**, 128, 1612-11619
40. J.K.Davies, *Pathogenic Neisseria*, Caister Academic Press, **2014**
41. A. McPherson, J.A. Gavira, *Acta Cryst.*, **2014**, F70, 2-20
42. S.M. Logan, E. Altman, O. Mykytczuk, J-R. Brisson, V. Chandan, F. St. Michael, A. Masson, S. Leclerc, K. Hiratsuka, N. Smirnova, J. Li, Y. Wu, W.W. Wakarchuk, *Glycobiology*, **2005**, 15, 721-733
43. Seko Akira, *Trends in Glycoscience and Glycotechnology*, 18, 2006, 209-230
44. J-E.Park, K-Y.Lee, S-I. Do, S-S. Lee, *Journal of Biochemistry and Molecular Biology*, Vol. 35, No. 3, July 2002, pp. 330-336
45. J.E. Park, K-Y. Lee, S-I. Do, S.S. Lee, *J. Biochem. Mol. Biol.*, **2002**, 35, 330-336



46. 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, 379-383
47. B. Ramakrishnan, P.V. Balaji, P.K. Qasba, *J. Mol. Biol.*, **2002**, 318, 491–502
48. B. Ramakrishnan, P.K. Qasba *J. Biol. Chem.*, **2002**, 277, 9, 20833–9
49. B. Echeverria, S. Serna, S. Achilli, C. Vives, J. Pham, M. Thépaut, C. Hokke, F. Fieschi, N-C. Reichardt, *ACS Chem. Biol.*, **2018**, 13, 2269–2279
50. T.K.van den Berg, H.Honing, N.Franke, A. van Remoortere, W.E.C.M.Schiphorst, F-T.Liu, A.M.Deelder, R.D.Cummings, C.H.Hokke, I. van Die, *J.Immunol.***2004**, 173, 1902-1907
51. S.Mugalapati, V.Koppolu, T.S.Raju, *Biochem.*, **2017**, 56, 1218-1226
52. B. Tefsen, C.M.W. van Stijn, M. van den Broek, H. Kalay, J.C. Knol, C.R. Jimenez, I. van Die, *Carb. Res.*,**2009**, 344, 1501-1507
53. S.J. van Vliet, E. van Liempt, E. Saeland, C.A. Aarnoudse, B. Appelmelk, T. Irimura, T.B.H. Geijtenbeek, O. Blixt, R. Alvarez, I. van Die, I. van Kooyk, *Int. Immunol.*, **2005**, 17, 661–669
54. A. van Diepen, A-J van der Plas, R.P. Kozak, L. Royle, D.W. Dunne, C.H. Hokke, *Int.J.Parasit.*, **2015**, 45, 465-475
55. M.Wuhrer, C.A.M.Koeleman, A.M.Deelder, C.H.Hokke, *FEBS Journal*, **2006**, 273, 347-361
56. Z.S. Kavar, S.M.Haslam, H.R. Morris, A.Dell, R.D. Cummings, *J. Biol. Chem.* 2005, 280:12810-12819
57. M. Eriksson, S. Serna, M. Maglinao, M.K. Schlegel, P.H. Seeberger, N-C. Reichardt, B. Lepenies, *ChemBioChem.*,**2014**, 15, 844-851
58. S.Yan, S.Serna, N-C.Reichardt, K.Paschinger, I.B.H.Wilson, *J.Biol.Chem.***2013**, 288, 21015-21028
59. M.H. Ross, J.O. Ely, J.G. Archer, *J.Biol.Chem.*, 1951, 192 561-568
60. B. Domon and C.E. Costello. *Glycoconjugate J.* **1988**, 5, 397-409
61. T.Zheng, H.Jiang, M.Gros, D.Soriano del Amo, S.Sundaram, G.Lauvau, F.Marlow, Y.Liu, P.Stanley, P.Wu, *Angew.Chem.Int.Ed.Engl.***2011**, 50, 4113-4118
62. C.D. Rillahan, E. Schwartz, R. McBride, V.V. Fokin, J.C. Paulson, *Angew. Chem. Int. Ed.*, **2012**, 51, 11014-11018
63. P.J.Cossar, L.Hizartzidis, M.I.Simone, A.McCluskey, C.P. Gordon, *Org. Biomol. Chem.*, 2015, **13**, 7119
64. M.Irfan, T.N.Glasnov, C.O.Kappe, *ChemSusChem*, 2011, **4**, 300–316
65. K.Brzezicka, B.Echeverria, S.Serna, A.vanDiepen, C.H.Hokke, N-C.Reichardt, *ACS Chem.Biol.*,**2015**, 10, 1290-1302
66. M.C. Daga, M. Taddei, G. Varchi, *Tetrahedron Lett.*, **2001**, 5191-5194
67. P.K. Mandal, J.S. McMurray, *J.Org.Chem.*, **2007**, 72, 6599-6601

68. M.R. Carrasco, C.I. Alvarado, S.T. Dashner, A.J. Wong, M.A. Wong, *J.Org.Chem.*, **2010**, 75, 5757-5759
69. M.A.Oberli, M-L. Hecht, P.Bindschädler, A. Adibekian, T.Adam, P.H.Seeberger, *Chem&Biol*, **18**, **2011**, 580-588
70. J. Voglmeir, R. Sardzik, M.J. Weissenborn, S.L. Flitsch, *OMICS*, **2010**, 14, 437-444
71. S. Serna, S. Yan, M. Martin-Lomas, I.B.H. Wilson, N-C. Reichardt, *J. Am. Chem. Soc.*, **2011**, 133, 16495-16502
72. C.D.Rillahan, E.Schwartz, C.Rademacher, R.McBride, J.Rangarajan, V.V.Fokin, J.C.Paulson, *ACS Chem.Biol.***2013**, 8, 1417-1422
73. A. Beloqui, J. Calvo, S. Serna, S. Yan, I.B.H. Wilson, M. Martin-Lomas, N.C. Reichardt, *Angew. Chem. Int. Ed.*, **2013**, 52, 7477-7481
74. S.Serna, J.Etxebarria, N.ruiz, M.Martin-Lomas, N-C.Reichardt, *Chem.Eur.J.***2010**, 16, 13163-13175
75. M.J.Borrok, L.L.Kiessling, *J.Am.Chem.Soc.*, **2007**, 129, 12780-12785
76. V.Porkolab, E.Chabrol, N.Varga, S.Ordanini, I.Sutkeviciuté, M.Thépaut, M-J. Garcia Jimenez, E.Girard, P.M.Nieto, A.Bernardi, F.Fieschi, *ACS Chem.Biol.*,**2018**, 13, 600-608
77. M.Thépaut, C.Guzzi, I.Sutkeviciute, S.Sattin, R.Ribeiro-Viana, N.Varga, E.Chabrol, J.Rojo, A.Bernardi, J.Angulo, P.M.Nieto, F.Fieschi, *J.Am.Chem.Soc.*, **2013**, 135, 2518-2529
78. S.Meyer, E. van Liempt, A.Imberty, Y. van Kooyk, H.Geyer, R.Geyer, I. van Die, *J.Bio.Chem.*,**2005**, 280,37349–37359
79. M.H.J. Meevissen, N.N. Driessen, H.H. Smits, R. Versteegh, S.J. van Vliet, Y. van Kooyk, G. Schramm, A.M. Deelder, H. Haas, M. Yazdanbakhsh, C.H. Hokke, *Int. J. Parasitol.*, **2012**, 42, 269-277
80. E. van Liempt, S.J. van Vliet, A. Engering, J.J. Garcia Vallejo, C.M.C. Bank, M. Sanchez-Hernandez, Y. van Kooyk, I. van Die, *Molecular Immunology*, **2007**, 44, 2605-2615
81. J. Heimburg-Molinaro, X. Song, D.F. Smith, R.D. Cummings, *Curr. Protoc. Protein Sci.*, **2011**, 64, 12.10.1-12.10-29
82. W. van Breedam, S.Pöhlmann, H.W.Favoreel, R.J. de Groot, H.J. Nauwynck, *FEMS Microbiol.Rev.*,**2014**, 38, 598-632
83. Alberts B, Johnson A, Lewis J, et al. *Molecular Biology of the Cell. 4th Ed.* New York: Garland Science; 2002. Blood Vessels and Endothelial Cells.
84. P.J.Skelly, A. A. Da'dara, X-H. Li, W. Castro-Borges, R.A. Wilson, *PLoS Pathog.*, **2014**, 10, e1004246
85. M.H.J. Meevissen, M. Yazdanbakhsh C.H. Hokke, *Experimental Parasitology*, 132, **2012**, 14–21
86. Y.Guo, H.Feinberg, E.Conroy, D.A.Mitchell, R.Alvarez, O.Blixt, M.E.Taylor, W.I.Weis, K.Drickamer, *Nature Structural &Molecular Biology*,**2004**, 11, 591-598

87. S.Pöhlmann, E.J.Soilleux, F.Baribaud, G.J.Leslie, L.S.Morris, J.Trowsdale, B.Lee, N.Coleman, R.W.Doms, *PNAS*, **2001**, 98, 2670-2675
88. E. van Liempt, S.J. van Vliet, A. Engering, J.J. Garcia Vallejo, C. M.C. Bank, M. Sanchez-Hernandez, Y. van Kooyk, I. van Die, *Molecular Immunol.*, **2007**, 44, 2605-2615
89. S.J. van Vliet, E. van Liempt, E. Saeland, Y. van Kooyk, *Trends in Immunol.*, **2008**, 29, 83-90
90. J.J. Garcia Vallejo, Y. van Kooyk, *Immunity*, **2015**, 983-985
91. S.A.F. Jégouzou, A. Quintero-Martinez, X. Ouyang, A. dos Santos, M.E. Taylor, K.Drickamer, *Glycobiol.*, **2013**, 23, 853-864



## **4 Conclusions and Outlook**



## 4 Conclusions and outlook

In an effort to unravel one of the many intricacies of the puzzle by which *S.mansoni* evades the host immune system, we undertook the synthesis of an O-glycan library inspired from structures previously isolated from the helminth. We described the chemoenzymatic synthesis of parasitic O-glycans based on the mucin core 2 and *S.mansoni* specific core. The library of O-glycans initially targeted was limited to elongations on Gal $\beta$ -1,6GalNAc arm of the cores owing to an unanticipated acceptor specificity of the recombinant glycosyltransferase LgtA\_X. Synthesized as aminopentyl glycosides, eight O-glycans were obtained and were printed on-chip alongside N-glycans available from our laboratory. The array was enzymatically modified using HP-FucT in conjunction with either the fucose or the 6-azido-fucose donor and arrayed against a selection of CLRs.

Our first investigations focused on the synthesis of the targeted mucin core 2 and the *S.mansoni* core which were both synthesized using a convergent strategy. Originally targeted as thioglycosides, insertion of the aminopentyl linker was necessary to increase the solubility of the deprotected triol intermediate in DCM. This was essential for the subsequent benzylidene acetal protection step. Unlike previous reports, regio- and chemo-selective glycosylation of positions C-6 over C-4 was not high yielding as an inseparable mix of di- and tri-substituted compounds was observed. Therefore C-4 was regioselectively protected to afford unequivocal glycosylation at C-6 and with the added benefit of an additional chromophore for purification by HPLC-UV. The mucin core 2 compound was subsequently easily obtained in 73% yield.

We described for the first time the synthesis of the *S.mansoni* specific O-glycan core and derived glycan structures. A key point was the optimization of the final glycosylation in which the yield suffered from an acyl migration to the acceptor. Changing the donor from an acetate- to a benzoyl -protected galactose improved the reaction yield from 40 to 65%. We have therefore made the synthesis of this novel core easier for one skilled in the art and therefore more exploitable for further biochemical investigations.

For our enzymatic elongations towards LN and LDN epitopes we proposed to optimize the bacterial enzyme LgtA, for an affordable way to install the  $\beta$ -1,3 GlcNAc moieties present in our target structures. By careful redesign of the DNA vector, we were able to obtain LgtA\_X which was easily expressed, purified and could be obtained in large quantities (12mg/L). The

enzyme was active and was used in the construction of LN and LDN epitopes. Using microarray and solution-phase activity studies, we established the  $\beta$ -1,6 specificity of the enzyme on O-glycan cores. However, this construct was unstable for long-term storage and suffered from high propensity to precipitate over time. This not only affected overall enzymatic yields but also prevented the formation of a stable enzymatic crystal for crystal structure experiments. Additional work into preventing precipitation would allow for optimal usage of the enzyme. In so doing, perhaps a crystal structure could be obtained of the protein and pave the way towards engineering an enzyme suitable for GlcNAc $\beta$ -1,3GalNAc linkages and construction of antigenic polyLDN epitopes.

As a consequence of the regioisomeric specificity of LgtA\_X towards Gal $\beta$ -1,6GalNAc linkages, the development of the library of compounds was severely restricted. We addressed this issue by proposing a synthetic alternative to generate cores compatible with subsequent enzymatic elongations. To this end, a novel disaccharide donor was designed. Preliminary evaluations of the synthesis of the disaccharide revealed that importance in the choice of anomeric thiol chosen as leaving group. Indeed the tolyl thioglycoside was observed to undergo an unfavourable aglycon transfer by 30%. Using the bulkier 2,6 dimethylphenyl thioglycoside successfully circumvented this side-reaction. Despite not being carried out to synthetic completion, the remaining synthetic steps towards this new donor were anticipated to proceed smoothly. This presents significant value in the preparative scale synthesis of O-glycans and in the exploration of the full potential of the glycan collection made by *S.mansoni*.

The library of O-glycans was therefore revised to target asymmetric  $\beta$ -1,6 lelongated structures. This was additionally challenged by the poor solubility of the compounds in aqueous enzymatic solutions which resulted in overall lower yields in the reactions. This was due to the hydrophobic contribution of the chromophores remaining in the partially deprotected compounds. In light of the Gal $\beta$ -1,6GalNAc regioisomeric preference of LgtA\_X, these could be removed in future work to make the glycans more water soluble and thus yield higher conversions of the asymmetric glycans.

Once target compounds were obtained, the deprotection step towards suitable substrates for microarray printing proved more challenging than expected. Crucially, removal of the electron rich N-benzyl was accompanied by degradation which seemed to correspond to the cleavage of the linker. To tackle this problem, several hydrogenation conditions were investigated before finally selecting an acid-based solution under atmospheric pressure of hydrogen. Finally



we obtained a total of 8 O-glycans, 3 mammalian mucin-type and 5 *S.mansoni* type. To the best of our knowledge, the synthesis of the 5 *S.mansoni* O-glycans (**O2**; **O6-O8**) is previously unreported.

Using microarray technology, we described for the first time of the specificity of selected CLRs MGL, DC-SIGN and DC SIGN R with parasitic O-glycans and put the results in context to previous knowledge on CLR specificity. Overall, our O-glycans respected the specificity previously described, where available. For example, the GalNAc specificity was confirmed for MGL as the **O8** appeared as the strongest binder. However, unusually binding was observed for some glycans displaying Le<sup>x</sup> epitope notably **O5**. In the case of DC-SIGN R, a striking differential binding profile was observed between the CRD and the ECD of the lectin with the native glycan array, emphasizing the importance in considering the structural context of glycan-CLR interactions.

Using HP-FucT with the natural fucose donor or the unnatural C6-azidofucose donor, the array was successfully fucosylated directly on the slide and we reported optimal conditions to obtain high conversion yields (60-98%) by MALDI. In the case of the fucose donor, only two exposures to the reaction were necessary whereas three were needed for the azido-fucose donor. The newly-generated azido-fucose glycomimetics were targeted for rapid on-chip generation of glycomimetics by click-chemistry with a selection of alkynes, as previously described for Siglec-7 targeting. However, we noticed that excessive washing of the slides due to repeated exposure to enzymatic mixtures resulted in a loss of fluorescence signal. This probably reflected the removal of the glycan material from the surface due to the washes. Therefore the fucosylated and azido-fucosylated data suffered from lack of reproducibility and reliability. This made quantification of the effect of enzymatic elongation on CLR specificity challenging and therefore compromised the studies towards click-generated glycomimetics. Additional investigations ensuring high reproducibility of the azido-fucosylation step are therefore needed. These could include evaluation of different washing solutions for minimal material removal from the surface, printing a higher number of glycan replicates for a more accurate sample population and using CLRs constructs yielding higher fluorescence intensities.

Nevertheless, interesting features were observed. In the case of DC-SIGN, both fucosylation and azido-fucosylation of the native array were observed to increase glycan-CLR interactions, in accordance with the previously described CLR specificity. In the case of MGL, fucosylation did not change the overall binding profile of the lectin suggesting that fucose-containing

ligands are tolerated by the enzyme to an extent. Both of these observations suggest the existence of exploitable chemical space for CLR targeting.

The full understanding of the *S.manson's* glycan mimickry to target CLRs remains a complex puzzle. In this work we have developed a methodology for the development of an O-glycan library based on the helminth's glycome. This facilitates the studies into the biofunctionality of *S.manson's* O-glycans and offers new chemical landscape to explore in the search for immunomodulatory compounds for the development of glycan-based therapeutics for the treatment of immune-compromised diseases.

## **5 Experimental**



## 5.1 General materials and instrumentation

**Chemicals** and Amberlite®IR 120 (H) were purchased from Sigma-Aldrich or Acros Organics and were used without further purification. All organic **solvents** were purchased from PanreacAppliChem and used without further purification. Organic solvents were dried over activated 4 Å or 3 Å molecular sieves. All solvents were concentrated using rotary evaporation. **Thin layer chromatography** was carried out using Merck aluminium sheets Silica Gel 60 F<sub>254</sub> and visualized by UV irradiation(254nm) or by staining with vanillin. All aqueous solutions were prepared from nanopure water produced with a Diamond UV water purification system (Branstead International).**Anhydrous reactions** were performed in flame-dried or oven-dried glassware under a positive pressure of dry argon. **Uridine 5'-diphospho-glucosamine disodium salt (UDP-GlcNAc)**, uridine 5'-diphosphogalactose disodium salt (UDP-Gal), uridine 5'-diphospho-2-acetamido-2-deoxy- $\alpha$ -D-galactosamine disodium salt (UDP-GalNAc) and guanosine 5'-diphospho- $\beta$ -L-fucose sodium salt (GDP-Fuc), were purchased from Carbosynth. Alkaline Phosphatase was purchased from Sigma Aldrich, Spain.

**Recombinant C-type lectin receptors** were expressed in *E.coli* and provided by Immunoshape partners Prof. Franck Fieschi and S.Achilli from IBS, Grenoble (FR). Lectins were labeled with Alexafluor555 (MGL) or Cy3 (DC SIGN and DC-SIGN R).

**Purifications** of compounds were performed using Merck 62 Å 230–400 mesh silica gel or on a Biotage SP4 automated flash chromatography system, (Biotage AB) employing prepacked silica cartridges for flash chromatography, SampliQ high performance graphitized carbon cartridges from Agilent Technologies or C18 Sep-Pak Cartridges from Waters (Milford) for SPE. Compound mixtures from enzymatic elongation were separated by preparative HPLC on a Waters autopurification HPLC system including: a Waters 2767 Sample Manager, a Waters System Fluidics Organizer, a Waters 2545 Binary Gradient Module, a Waters 515 HPLC pump, a Waters 2996 Photodiode Array Detector and a Zspray™ SQ Detector 2. Analytical separation was performed using a Thermo Scientific C18 250x4.6mm with 5  $\mu$ m particle size. Samples higher than 5 mg were separated on a Phenomenex Gemini RP C18 10x250mm column with 55  $\mu$ m particle size. The samples were dissolved in a maximum 3:7 ACN:H<sub>2</sub>O with a 20%v/v maximum DMSO content and were eluted with a flow rate of 16 ml/min with a maximum of 20 mg/ml per injection, 1ml/injection. Samples lower than 5 mg were separated on a Waters XBridge C18 10x100mm column with 5  $\mu$ m particle size and eluted at 4 mL/min. Pooled fractions containing target glycan material were lyophilized on an ALPHA-2-4 LSC freeze-dryer from Christ, Osterode, Germany.

Protein crystallization experiments were conducted under the guidance of Prof. Marcelo E. Guerin, Dr David Albesa-Jove and Alberto Marina in the CICbioGUNE, Spain. Crystallization screening was done using Structure Screens I and II, Morpheus<sup>®</sup>, PACT, JCSG+ (Molecular Dimensions Ltd) and Grid Screen<sup>™</sup> Ammonium Sulfate (Hampton Research) was performed using a Cartesian Technologies workstation and a Mosquito<sup>®</sup> liquid handling robot (TTP Labtech).

All **NMR** spectra were acquired on Bruker 500 MHz spectrometer and chemical shifts ( $\delta$ ) are given in parts per million (ppm) relative to the residual signal of the solvent used. Splitting patterns are designated as s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet. Coupling constants (J) are reported in Hertz (Hz). Continuous flow-hydrogenation reactions were carried out on **H-Cube**<sup>®</sup> reactor from ThalesNano Nanotechnology Inc.

**Microwave** irradiation was performed on Biotage Initiator monomode oven, Biotage AB, Uppsala, Sweden

**MALDI-TOF** mass analyses were performed on an Ultraflex<sup>™</sup> III time-of-flight mass spectrometer equipped with a pulsed N<sub>2</sub> laser (337 nm) and controlled by FlexControl 3.3 software (BrukerDaltonics). **For the microarray experiments**, DHB matrix (4mg/mL in water:acetonitrile, 95:10, and 0.002% sodium formate) was printed on top of the immobilized glycans and allowed to crystallize.

**Microarrays** were printed employing a robotic non-contact piezoelectric SciFLEXARRAYER spotter S11 (Scienion). Arrays were compartmentalized using an 8 well hybridization gasket or 12 well incubation gasket (Agilent Technologies). Arrays were imaged on an Agilent G2565BA fluorescence scanner system (Agilent Technologies) at 10  $\mu$ M resolution, using 2 lasers (532nm or 633 nm). Quantification of fluorescence was performed using ProScanArray<sup>®</sup> Express software from Perkin Elmer. A quantification method using an adaptive circle with minimum diameter 50  $\mu$ M and maximum 300  $\mu$ M was employed. Each histogram represents the average of mean fluorescence from 3 replicates after fluorescence background subtraction.

## 5.2 Chemical Synthesis

Monosaccharides **2**<sup>1</sup>, **2B**<sup>2</sup>, **4**<sup>3</sup>, **10**<sup>4</sup>, **11**<sup>4</sup> were synthesized as previously described.

---

<sup>1</sup>Hai Yu, and Xi Chen, *Org. Lett.*, **2006**, 8, 2393-2396

### 3,4,6-tri-*O*-acetyl-2-deoxy-2-((2,2,2-trichloroethoxy carbonyl amino)-D-galactopyranosyl trichloroacetimidate **14**

To a solution of galactose (23.3 mmol) in water (65ml) were added TrocCl(3.8ml, 27.9 mmol) and NaHCO<sub>3</sub> (6.8g, 81.4 mmol). The solution was stirred at RT for 4 hours until a turbid white mixture was obtained. 1M HCl(aq.) was added to quench the reaction and the reaction mixture was concentrated. The dried residue was diluted in pyridine (26ml) and acetic anhydride (13 ml, 140mmol) was added. The mixture was stirred at RT for 4 hours, washed with water and sat. CuSO<sub>4</sub>(aq.), and the combined organic extracts concentrated to afford the peracetylated compound which was re-dissolved in dry DMF (100ml). Hydrazine acetate (2.3g, 24.8mmol) was added under argon and the reaction mixture was stirred at RT for 3hours. The mixture was quenched with water, washed with brine and extracted with EtOAc. The combined organic fractions were concentrated and re-dissolved in dry DCM (40ml) and trichloroacetonitrile (27.6 ml, 27.6 mmol) and DBU (0.657ml, 4.6 mmol) were added at 0°C. After 1hr stirring at RT, full conversion was observed by TLC therefore the reaction was concentrated and the crude purified by silica gel column chromatography (5:25:70 Et<sub>3</sub>N:EtOAc:Hexane) to yield **14** as a pale yellow residue, 78%. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 8.66 (s, 1H, NH), 6.60 (d, J = 3.5 Hz, 1H, H-1), 5.56 (dd, J = 3.3, 1.4 Hz, 1H, H-4), 5.42 (dd, J = 10.9, 3.2 Hz, 1H, H-3), 5.36 (dd, J = 10.8, 3.5 Hz, 1H, H-2), 4.44 (td, 1H, H-5), 4.16 (dd, J = 11.4, 6.7 Hz, 1H, H-6), 4.08 (dd, J = 11.3, 6.7 Hz, 1H, H-6), 2.17 (s, 3H, CH<sub>3</sub>), 2.03 – 2.00 (m, 9H, 3CH<sub>3</sub>). NMR consistent with literary precedents.<sup>5</sup>

### Synthesis of 5- (benzyl (benzyloxycarbonyl)amino) pentyl 3,4,6-tri-*O*-acetyl-2-deoxy-2-((2,2,2-trichloroethoxycarbonylamino) β-D-galactopyranoside **18**

To a solution of **14** (11.8 g, 18.9 mmol) in dry DCM (3ml) on activated molecular sieves and under argon was added a solution of N-benzyl-N-(5-hydroxypentyl)carbamate (7.42 g, 22.7 mmol) in dry DCM (200ml). The solution was placed at 0°C before TMSOTf (0.690 ml, 3.8 mmol) was added dropwise. The reaction was allowed to warm to RT and stirred for 2hours after which TLC showed full consumption of starting material. The reaction mixture was quenched with Et<sub>3</sub>N and filtered through celite. The crude sample was purified by flash column chromatography (15→40% EtOAc:Tol) to yield **18** as a white gum, 11 g, 73%. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 7.41 – 7.23 (m, 9H, Ar), 7.17 (d, J = 8.1 Hz, 1H, Ar), 5.36 (d, J = 3.3 Hz, 1H, H-4), 5.25 – 5.09 (m, 3H, H-3, CH<sub>2</sub>Bn), 4.79 – 4.59 (m, 2H, CH<sub>2</sub>Troc), 4.59 – 4.40 (m, 3H, H-1, CH<sub>2</sub>Bn), 4.14 (qd, J = 11.2, 6.7 Hz, 2H, 2H-6), 3.94 – 3.70 (m, 3H, H-5, H-2, CH<sub>linker</sub>), 3.50 – 3.25 (m, 1H, CH<sub>linker</sub>), 3.23 – 3.12 (m, 2H, CH<sub>2</sub>), 2.14 (s, 3H, OCH<sub>3</sub>), 2.05 (s, 3H, OCH<sub>3</sub>), 1.99 (s, 3H, OCH<sub>3</sub>), 1.59

---

<sup>2</sup>M. Thomas, J-P. Gesson, S. Papot, *J. Org. Chem.*, **2007**, 72, 4262-4264

<sup>3</sup>.Benakli, K.; Zha, C. X.; Kerns, R. J. *J. Am. Chem. Soc.* 2001, 123, 12933

<sup>4</sup>U. Ellervik, G. Magnusson, *Carb. Res.*, **1996**, 280, 251-260

<sup>5</sup> B.Sun, A. V. Punkin, G.M. Visser, H. Zuilhof, *Tetrahedron Lett.*, **2006**, 47, 7371-7374

- 1.43 (m, 4H, 2CH<sub>2linker</sub>), 1.39 – 1.18 (m, 2H, CH<sub>2linker</sub>). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 170.45, 170.37, 156.77, 156.30, 154.38, 137.79, 136.74, 128.57, 128.44, 127.92, 127.81, 127.73, 127.35, 127.20, 101.25, 95.67, 74.30, 70.50, 69.86, 69.73, 67.21, 66.76, 61.48, 52.73, 50.48, 50.37, 47.28, 46.10, 29.04, 28.66, 27.76, 26.99, 22.96, 20.69, 20.65. HRMS (MALDI-Tof): *m/z* calcd. for C<sub>35</sub>H<sub>43</sub>Cl<sub>3</sub>N<sub>2</sub>O<sub>12</sub> [M+Na]<sup>+</sup>: 811.1778, found: 811.1728. [α]<sub>D</sub><sup>20</sup> = -11.8° (c=1, CHCl<sub>3</sub>)

#### Synthesis of 5- (benzyl (benzyloxycarbonyl)amino) pentyl 2-deoxy-2- ((2,2,2-trichloroethoxycarbonylamino) β-D-galactopyranoside 13

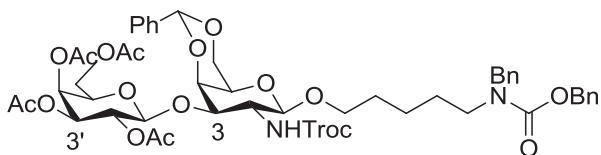
**18**(0.456mmol) was dissolved in G/GHNO<sub>3</sub> (27 ml) and the solution was stirred at RT for 15 mins until full consumption of starting material was observed by TLC. The reaction mixture was concentrated *in vacuo* then washed thoroughly with DCM and filtered through celite. The filtrate was concentrated to yield **13** as a pale pink residue in a quantitative yield. <sup>1</sup>H NMR (500 MHz, MeOD) δ 7.42 – 7.14 (m, 10H, Ar), 5.16 (d, J = 18.8 Hz, 2H, CH<sub>2Bn</sub>), 4.80 – 4.62 (m, 2H, CH<sub>2Bn</sub>), 4.32 (dd, J = 8.8 Hz, 1H, H-1), 3.92 – 3.68 (m, 2H, H-4, CH<sub>Troc</sub>), 3.76 (m, 2H, H-6), 3.68 – 3.53 (m, 2H, H-2, H-3), 3.46 (t, J = 6.1 Hz, 1H, H-5), 3.39 (m, 1H, CH<sub>Troc</sub>), 3.35 – 3.11 (m, 2H, CH<sub>2</sub>), 1.61 – 1.41 (m, 5H, CH<sub>2</sub>), 1.38 – 1.17 (m, 3H, CH<sub>2</sub>).

#### Synthesis of 5- (benzyl (benzyloxycarbonyl)amino) pentyl 4,6-O-benzylidene-2-deoxy-2- ((2,2,2-trichloroethoxycarbonylamino) β-D-galactopyranoside 19

To a solution of **13** (8.55 g, 12.9 mmol) in dry acetonitrile (100ml) under argon were added benzaldehyde dimethyl acetal (3.4 ml, 2.25 mmol) and camphor sulfonic acid (600 mg, 2.6 mmol). The reaction was stirred at RT overnight after which successful conversion was observed by TLC. Et<sub>3</sub>N was added and the reaction mixture was concentrated *in vacuo*. The resulting residue was purified by flash column chromatography (3 → 10% MeOH:DCM) to yield **19** as a white solid, 6.78 g, 70%. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 7.51 (d, J = 5.2 Hz, 2H, Ar), 7.42 – 7.22 (m, 11H, Ar), 7.17 (d, J = 7.0 Hz, 2H, Ar), 5.57 (s, 1H, CH<sub>Ph</sub>), 5.23 – 5.13 (m, 2H, CH<sub>2Ph</sub>), 4.72 (d, J = 11.7 Hz, 2H, CH<sub>2Troc</sub>), 4.50 (m, 3H, H-1, CH<sub>2Bn</sub>), 4.32 (d, J = 12.4 Hz, 1H, H-6<sub>a</sub>), 4.20 (d, J = 3.6 Hz, 1H, H-4), 4.08 (dd, J = 12.4, 1.9 Hz, 1H, H-6<sub>b</sub>), 4.03 – 3.82 (m, 2H, H-3, CH<sub>2</sub>), 3.65 (m, 1H, H-2), 3.5 (s, 1H, H-5), 3.43-3.25 (m, 1H, CH<sub>linker</sub>), 3.25-3.15 (m, 2H, CH<sub>2</sub>), 1.62 – 1.41 (m, 2H, 2CH<sub>2linker</sub>), 1.39 – 1.16 (m, 2H, CH<sub>2linker</sub>). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 152.34, 151.88, 150.76, 150.49, 134.39, 134.36, 134.34, 134.15, 133.47, 133.32, 126.34, 125.72, 125.69, 125.60, 125.43, 125.16, 125.10, 124.98, 124.88, 124.57, 124.54, 124.39, 123.71, 99.78, 99.53, 99.33, 94.60, 75.14, 75.07, 74.49, 70.66, 70.37, 69.68, 69.62, 69.52, 67.75, 67.72, 66.93, 56.43, 51.84, 51.71, 48.87, 47.78, 31.72, 31.33, 30.52, 29.75, 26.02, 25.90. HRMS (MALDI-Tof) *m/z* calcd. for C<sub>36</sub>H<sub>41</sub>Cl<sub>3</sub>N<sub>2</sub>O<sub>9</sub> [M+Na]<sup>+</sup>: 773.1774, found: 773. 1772. [α]<sub>D</sub><sup>20</sup> = -3.5° (c=1, CHCl<sub>3</sub>)

#### 5- (benzyl (benzyloxycarbonyl)amino) pentyl 2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl-(1→3)-4,6-O-benzylidene-2-deoxy-2- ((2,2,2-trichloroethoxycarbonyl amino) β-D-galactopyranoside 6





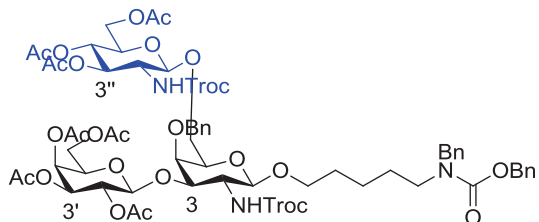
To a solution of **19** (6.26 g, 8.3mmol) and **2** (5.0 g, 10.1 mmol) in dry DCM (120ml) on activated molecular sieves at  $-40^{\circ}\text{C}$  and under argon was added TMSOTf (0.301 ml, 1.7 mmol) and the reaction was stirred at  $-20$  for 1hr as monitored by TLC. Triethylamine was added and the reaction was filtered through celite then concentrated to be purified by FCC (20 $\rightarrow$ 70% EtOAc:Hex) yielding **6** as a white foam, 5.6 g, 62%.  $^1\text{H}$  NMR (500 MHz, Chloroform- $d$ )  $\delta$  7.55 (d,  $J = 7.6$  Hz, 2H, Ar), 7.40 – 7.20 (m, 12H, Ar), 7.16 (d,  $J = 7.3$  Hz, 1H, Ar), 5.57 (s, 1H,  $\text{CH}_{\text{Ph}}$ ), 5.36 (d,  $J = 3.4, 1.2$  Hz, 1H, H-4'), 5.21 (dd, 1H, H-2'), 5.16 (s, 2H,  $\text{CO}_2\text{CH}_{2\text{Bn}}$ ), 4.97 (dd,  $J = 10.4, 3.5$  Hz, 1H, H-3'), 4.90 – 4.80 (m, 2H, H1,  $\text{CHCl}_3$ ), 4.77 (d,  $J = 8.0$  Hz, 1H, H-1'), 4.65 – 4.55 (m, 1H,  $\text{CHCl}_3$ ), 4.47 (s, 3H, H-3,  $\text{NCH}_{2\text{Bn}}$ ), 4.34 – 4.28 (m, 2H, H-6<sub>a</sub>, H-4), 4.15 (ddd, 2H, 2H-6'), 4.06 (dd,  $J = 12.4, 1.8$  Hz, 1H, H-6<sub>b</sub>), 3.92 (s, 1H,  $\text{CH}_{\text{linker}}$ ), 3.87 (td, 1H, H-5'), 3.53 (s, 1H, H-2), 3.44 (s, 1H, H-5), 3.39 (s, 1H,  $\text{CH}_{\text{linker}}$ ), 3.23 (s, 2H,  $\text{CH}_{2\text{linker}}$ ), 2.16 (s, 3H, Ac), 2.04 (d,  $J = 5.7$  Hz, 6H, 2 $\text{CH}_3$ ), 1.97 (s, 3H,  $\text{CH}_3$ ), 1.52 (d,  $J = 28.7$  Hz, 4H,  $\text{CH}_{2\text{linker}}$ ), 1.36 – 1.23 (m, 2H,  $\text{CH}_{2\text{linker}}$ ).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  170.31, 170.08, 169.34, 156.77, 156.22, 154.05, 137.86, 136.86, 136.74, 128.84, 128.55, 128.47, 128.11, 127.93, 127.80, 127.30, 127.19, 126.26, 101.86, 100.69, 99.74, 95.61, 76.01, 74.23, 70.87, 70.82, 69.69, 69.41, 69.19, 68.87, 67.17, 67.08, 66.45, 61.56, 53.82, 50.35, 47.26, 45.95, 29.68, 29.08, 28.81, 27.77, 27.27, 23.31, 23.06, 20.71, 20.56. HRMS  $[\text{M}+\text{Na}]^+$  (MALDI-Tof)  $m/z$  calcd. for  $\text{C}_{50}\text{H}_{63}\text{Cl}_3\text{N}_2\text{O}_{18}$   $[\text{M}+\text{Na}]^+$ : 1103.2724 found: 1103.2716.  $[\alpha]_{\text{D}}^{20} = +13.9^{\circ}$  ( $c=1$ ,  $\text{CHCl}_3$ )

**5- (benzyl (benzyloxycarbonyl)amino) pentyl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-4-O-benzyl-2-deoxy-2- ((2,2,2-trichloroethoxycarbonylamino)  $\beta$ -D-galactopyranoside **24****

To a solution of **6** (5.6 g, 5.2 mmol) in dry DCM (35ml) under Ar at  $0^{\circ}\text{C}$  was added 1M  $\text{BH}_3\cdot\text{THF}$  (20.7 ml, 20.7 mmol) dropwise. The solution was allowed to cool before adding TMSOTf (0.467 ml, 2.59mmol) dropwise and the reaction was stirred at  $0^{\circ}\text{C}$  under Ar for 1.5 hours after which full conversion was observed by TLC. The ice bath was removed and the solution was quenched with a solution of MeOH:Et<sub>3</sub>N (10:1) until effervescence ceased. The reaction mixture was concentrated *in vacuo* and purified by FCC (50  $\rightarrow$ 100% EtOAc:Hex) to yield **24** as a white solid, 3.4 g, 61%.  $^1\text{H}$  NMR (500 MHz, Chloroform- $d$ )  $\delta$  7.43 (d,  $J = 6.8$  Hz, 2H, Ar), 7.39 – 7.13 (m, 13H, Ar), 5.43 – 5.38 (s, 1H, H-4'), 5.27 (dd,  $J = 10.5, 8.0$  Hz, 1H, H-2'), 5.17 (d,  $J = 10.9$  Hz, 2H,  $\text{OCH}_{2\text{Bn}}$ ), 4.99 (dd,  $J = 10.5, 3.4$  Hz, 1H, H-3'), 4.92 (d,  $J = 11.8$  Hz, 1H, 4- $\text{OCH}_{2\text{Bn}}$ ), 4.84 – 4.64 (m, 5H,  $\text{CH}_{2\text{Troc}}$ , 4- $\text{OCH}_{2\text{Bn}}$ , H-1, H-1'), 4.47 (d,  $J = 8.8$  Hz, 3H, H-3,  $\text{NCH}_{2\text{Bn}}$ ), 4.17 (ddq,  $J = 15.9, 11.1, 6.1, 4.7$  Hz, 2H, H-6'), 3.98 – 3.65 (m, 4H, H5', H-4,  $\text{CH}_{\text{linker}}$ , H-6<sub>a</sub>), 3.53 – 3.14 (m, 6H, H-6<sub>b</sub>, H-2, H-5,  $\text{CH}_{\text{linker}}$ ,  $\text{CH}_{2\text{linker}}$ ), 2.14 (s, 3H,  $\text{CH}_3$ ), 2.10 (s, 3H,  $\text{CH}_3$ ), 2.02 (s, 3H,  $\text{CH}_3$ ), 1.99 (s, 3H,  $\text{CH}_3$ ), 1.53 (d,  $J = 18.6$  Hz, 4H, 2 $\text{CH}_{2\text{linker}}$ ), 1.31 (s, 2H,  $\text{CH}_{2\text{linker}}$ ).  $^{13}\text{C}$  NMR (126 MHz, Chloroform- $d$ )  $\delta$  170.41, 170.15, 170.07, 169.51, 138.31, 137.82, 129.09, 128.55, 128.46, 128.35, 128.03, 127.94, 127.79, 127.30, 127.19, 102.45, 99.95, 95.60, 78.73, 74.53, 74.42, 74.22, 73.97,

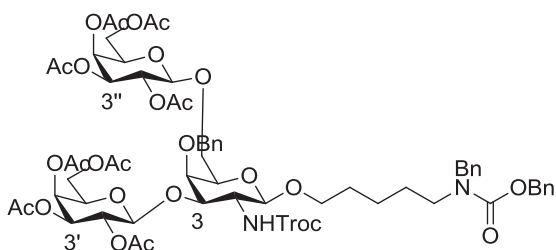
73.85 , 70.74 , 70.68 , 69.88 , 69.53 , 69.04 , 67.20 , 67.09 , 61.79 , 61.27 , 54.89 , 50.37 , 47.24 , 45.97 , 29.12 , 28.75 , 27.78 , 27.19 , 23.28 , 23.07 , 20.72 , 20.63 , 20.56 . HRMS (MALDI-Tof)  $m/z$  calcd. for  $C_{50}H_{61}Cl_3N_2O_{18}$   $[M+Na]^+$  : 1105.2881, found: 1105.2839.  $[\alpha]_D^{20} = -18.4^\circ$  (c=1,  $CHCl_3$ )

**5- (benzyl (benzyloxycarbonyl)amino) pentyl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[3,4,6-tri-O-acetyl-2-deoxy-2- ((2,2,2-trichloroethoxycarbonylamino)- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2- ((2,2,2-trichloroethoxycarbonylamino)  $\beta$ -D-galactopyranoside 25**



To a solution of **24** (420 mg, 0.387 mmol) and with the imidate **4** (291 mg, 0.465 mmol) in dry DCM (5ml) on activated molecular sieves, under argon and at  $-60^\circ C$  was added TMSOTf (10.5  $\mu L$ , 58.1  $\mu mol$ ). The reaction was stirred between  $-40^\circ C$  and  $-20^\circ C$  for 1hr until full reaction was observed as monitored by TLC. Triethylamine was added and the reaction was filtered through celite. The reaction crude was then concentrated *in vacuo* and purified by FCC (40 $\rightarrow$ 60% EtOAc:Hex) yielding **25** as a white solid, 438 mg, 73%.  $^1H$  NMR (500 MHz, Chloroform-d)  $\delta$  7.42 – 7.13 (m, 15H, Ar), 5.98 (s, 0.5H, N-H<sub>Troc</sub>), 5.45 (s, 0.5H, N-H<sub>Troc</sub>), 5.39 (s, 1H, H-4'), 5.30 – 5.13 (m, 4H, H-2', H-3'', OCH<sub>2Bn</sub>), 5.06 (t, J = 9.6 Hz, 1H, H-4''), 4.98 (dd, J = 10.5, 3.5 Hz, 1H, H-3'), 4.91 (d, J = 11.3 Hz, 1H, CH<sub>Troc</sub>), 4.84 – 4.59 (m, 6H, CH<sub>Troc</sub>, 4-OCH<sub>2Bn</sub>, H-1, H-1', H-1'') 4.46 (dt, J = 22.7, 12.8 Hz, 3H, NCH<sub>2Bn</sub>, H-3), 4.27 (dd, J = 12.3, 4.7 Hz, 2H, H-6'\_a, H-6''\_a), 4.12 (ddd, J = 26.1, 11.8, 4.5 Hz, 2H, H-6'\_b, H-6''\_b), 3.96 – 3.25 (m, 10H, H-4, H-5, H-5', H-5'', 2H-6, CH<sub>2linker</sub>, H-2, H-2''), 3.19 (s, 2H, CH<sub>2linker</sub>), 2.16 – 1.95 (m, 21H, 7CH<sub>3</sub>), 1.65 – 1.44 (m, 4H, 2CH<sub>2linker</sub>), 1.38 – 1.28 (m, 2H, CH<sub>2linker</sub>).  $^{13}C$  NMR (126 MHz, CDCl<sub>3</sub>)  $\delta$  170.73, 170.58, 170.22, 170.12, 169.52, 156.32, 154.11, 138.38, 137.83, 136.67, 128.90, 128.58, 128.49, 128.18, 127.97, 127.79, 127.34, 127.23, 102.31, 101.03, 100.07, 95.63, 78.22, 75.09, 74.40, 74.23, 72.18, 71.82, 70.79, 70.70, 69.00, 68.50, 67.24, 67.09, 62.03, 61.18, 56.38, 54.80, 50.39, 47.35, 28.83, 27.81, 27.20, 23.37, 20.75, 20.68, 20.65, 20.59. HRMS (MALDI-Tof)  $m/z$  calcd. for  $C_{65}H_{79}Cl_6N_3O_{27}$   $[M+Na]^+$ : 1566.2928, found: 1566.2886.  $[\alpha]_D^{20} = -6.2^\circ$  (c=1,  $CHCl_3$ )

**5- (benzyl (benzyloxycarbonyl)amino) pentyl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[2,3,4,6-tetra-O-acetyl- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2- ((2,2,2-trichloroethoxycarbonylamino)  $\beta$ -D-galactopyranoside 27**



To a solution of **24** (470 mg, 0.434mmol) and **2**(256 mg, 0.520 mmol) in dry DCM (5ml) on activated molecular sieves, under argon and at  $-60^{\circ}\text{C}$  was added TMSOTf (XX ml, mmol) dropwise. The reaction was stirred between  $-40^{\circ}\text{C}$  and  $-20^{\circ}\text{C}$  for 1hr as monitored by TLC. Triethylamine was added and the reaction was filtered through celite then concentrated *in vacuo* before purification by FCC (40 $\rightarrow$ 100% EtOAc:Hex) to yield **27** as a white foam, 252 mg, 41%.  $^1\text{H}$  NMR (500 MHz, Chloroform-*d*)  $\delta$  7.43 – 7.21 (m, 14H, Ar), 7.16 (d,  $J$  = 7.4 Hz, 1H, Ar), 5.99 (s, 0.5H,  $\text{NH}_{\text{Troc}}$ ), 5.42 – 5.35 (m, 2H, H-4', H-4''), 5.31 (d,  $J$  = 10.4 Hz, 0.5H,  $\text{NH}_{\text{Troc}}$ ), 5.24 (dd,  $J$  = 10.5, 7.9 Hz, 1H, H-2'), 5.21 – 5.11 (m, 3H, H-2',  $\text{OCH}_{2\text{Bn}}$ ), 5.01 – 4.89 (m, 3H, H-3, H-3',  $\text{CH}_{\text{Troc}}$ ), 4.83 – 4.62 (m, 6H, H-1, H-1', 4- $\text{OCH}_{2\text{Bn}}$ ,  $\text{CH}_{\text{Troc}}$ ), 4.58 – 4.40 (m, 4H, H-1'', H-3,  $\text{NCH}_{2\text{Bn}}$ ), 4.21 – 4.09 (m, 4H, H-6', H-6''), 3.96 – 3.61 (m, 6H, 2H-6, H-4, H-5', H-5'',  $\text{CH}_{\text{linker}}$ ), 3.60 – 3.54 (m, 1H, H-5), 3.52 – 3.14 (m, 4H, H-2, 3 $\text{CH}_{\text{linker}}$ ), 2.15 – 1.95 (m, 24H, 8 $\text{CH}_3$ ), 1.52 (d,  $J$  = 26.0 Hz, 4H, 2 $\text{CH}_{2\text{linker}}$ ), 1.36 – 1.26 (m, 2H,  $\text{CH}_{2\text{linker}}$ ).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  170.49, 170.43, 170.34, 170.27, 170.22, 170.19, 169.57, 169.30, 156.35, 154.23, 138.35, 137.96, 129.07, 128.66, 128.58, 128.28, 128.07, 127.97, 127.90, 127.42, 127.30, 102.46, 101.35, 99.88, 78.38, 75.15, 74.36, 70.95, 70.82, 70.78, 70.75, 69.14, 69.06, 67.29, 67.15, 67.03, 61.32, 61.08, 55.02, 50.50, 47.45, 29.81, 29.24, 28.84, 27.94, 27.42, 23.60, 23.28, 20.94, 20.85, 20.79, 20.76, 20.68. HRMS(MALDI-Tof)  $m/z$  calcd. for  $\text{C}_{64}\text{H}_{79}\text{Cl}_3\text{N}_2\text{O}_{27}$  [ $\text{M}+\text{Na}$ ] $^{\dagger}$ : 1435.383, found: 1435.3923.  $[\alpha]_{\text{D}}^{20} = -11.3^{\circ}$  ( $c=1$ ,  $\text{CHCl}_3$ )

**5- (benzyl (benzyloxycarbonyl)amino) pentyl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[2,3,4,6-tetra-O-benzoyl- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside **28****

To a solution of **24** (108 mg, 99.6  $\mu\text{mol}$ ) and **26** (118 mg, 160  $\mu\text{mol}$ ) in dry DCM (5ml) on activated molecular sieves, under argon and at  $-40^{\circ}\text{C}$  was added TMSOTf (2.7  $\mu\text{l}$ , 14.9  $\mu\text{mol}$ ) dropwise. The reaction was stirred between  $-40^{\circ}\text{C}$  and  $-20^{\circ}\text{C}$  for 1hr as monitored by TLC. Triethylamine was added and the reaction was filtered through celite then concentrated *in vacuo* before purification by FCC (0 $\rightarrow$ 50% EtOAc:Tol) to yield **27** as a white foam, 82.8 mg, 65%.  $^1\text{H}$  NMR (500 MHz, Chloroform-*d*)  $\delta$  8.05 – 7.97 (m, 4H, Ar), 7.89 – 7.84 (m, 2H, Ar), 7.72 (d,  $J$  = 7.8 Hz, 2H, Ar), 7.53 (dt,  $J$  = 28.3, 7.4 Hz, 3H, Ar), 7.45 – 7.09 (m, 30H, Ar), 5.96 (d,  $J$  = 3.4 Hz, 1H, H-4''), 5.71 (ddd,  $J$  = 10.0, 7.9, 1.8 Hz, 1H, H''), 5.56 (d,  $J$  = 10.2, 2.3 Hz, 1H, H3''), 5.36 – 5.32 (m, 1H, H-4'), 5.26 – 5.09 (m, 4H, H2',  $\text{OCH}_{2\text{Bn}}$ ), 4.96 – 4.83 (m, 3H), 4.80 – 4.05 (m, 16H), 3.88 – 3.70 (m, 5H), 3.58 – 3.30 (m, 3H,  $\text{CH}_{\text{linker}}$ , H-2, H-5''), 3.26 – 3.05 (m, 2H,  $\text{CH}_{2\text{linker}}$ ), 2.99 (dq,  $J$  = 22.2, 6.9 Hz, 1H,  $\text{CH}_{\text{linker}}$ ), 2.09 (s, 3H,  $\text{CH}_3$ ), 2.04 (s, 3H,  $\text{CH}_3$ ), 1.99 (s, 3H,  $\text{CH}_3$ ), 1.94 (s, 3H,  $\text{CH}_3$ ), 1.51 – 1.36 (m, 2H,  $\text{CH}_{2\text{linker}}$ ), 1.32 – 1.04 (m, 4H, 2 $\text{CH}_{2\text{linker}}$ ).  $^{13}\text{C}$  NMR (130 MHz,  $\text{CDCl}_3$ )  $\delta$  167.04, 166.94, 166.84, 166.25, 162.93, 162.45, 162.09, 136.12, 135.77, 131.63, 131.39, 131.34, 131.31, 128.15, 127.98, 127.93, 127.83, 127.56, 127.53, 127.27, 127.19, 127.14, 126.96, 126.83, 126.79, 126.74, 126.65, 126.51, 126.42, 126.21, 126.12, 126.04, 125.58,

125.45, 101.42, 100.76, 99.00, 78.30, 75.22, 74.33, 71.83, 71.45, 70.94, 70.17, 70.02, 69.31, 68.45, 67.55, 67.34, 62.40, 61.60, 55.63, 51.36, 48.45, 47.09, 30.71, 30.29, 29.56, 29.00, 25.34, 25.07, 22.72, 22.68, 22.65, 22.57. HRMS (MALDI-Tof)  $m/z$  calcd. for  $C_{84}H_{87}Cl_3N_2O_{27}$   $[M+Na]^+$ : 1683.4456, found: 1683.4514

**5- (benzyl (benzyloxycarbonyl)amino) pentyl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[3,4,6-tri-O-acetyl-2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 30**

To a solution of **25** (390 mg, 0.253 mmol) in THF (6ml) was added  $BF_3 \cdot Et_2O$  (1M in THF, 0.607ml, 0.607 mmol) dropwise. The yellow reaction mixture was refluxed for 1.5 hours after which it was placed on ice and quenched with MeOH. The reaction mixture was concentrated to dryness *in vacuo* before being redissolved in anhydrous pyridine (3ml). Acetic acid (0.6 ml, 6.32 mmol) was added and the reaction was left to stir overnight. MeOH was added and the reaction mixture was concentrated *in vacuo*. The crude was purified by FCC (0 $\rightarrow$ 5% MeOH:DCM) to yield **30** as a brown solid, 226mg, 70%.  $^1H$  NMR (500 MHz, Chloroform-d)  $\delta$  7.43 – 7.13 (m, 15H, Ar), 5.39 (d,  $J$  = 3.5 Hz, 1H, H-4'), 5.26 – 5.10 (m, 4H, H-2', H-3'',  $OCH_{2Bn}$ ), 5.08 – 4.98 (m, 2H, H-4'', H-3'), 4.87 (d,  $J$  = 11.4 Hz, 1H, 4- $OCH_{Bn}$ ), 4.82 (d,  $J$  = 8.3 Hz, 1H, H-1), 4.64 (dt,  $J$  = 27.6, 9.0 Hz, 4H, 4- $OC_{HBn}$ , H-1', H-1'', H-3), 4.49 (m,  $J$  = 17.7, 16.6 Hz, 2H,  $NCH_{2Bn}$ ), 4.30 – 4.03 (m, 4H, 2H-6', 2H-6''), 3.97 – 3.71 (m, 5H, H-2'',  $H_{6a}$ , H-4, H-5',  $CH_{linker}$ ), 3.70 – 3.57 (m, 3H,  $H_{6b}$ , H-5'', H-5), 3.52 – 3.25 (m, 3H, H-2,  $CH_{2linker}$ ), 3.24 – 3.11 (m, 2H,  $CH_{2linker}$ ), 2.15 – 1.84 (m, 24H), 1.64 – 1.42 (m, 4H,  $CH_{2linker}$ ), 1.36 – 1.26 (m, 2H,  $CH_{2linker}$ ).  $^{13}C$  NMR (126 MHz,  $CDCl_3$ )  $\delta$  170.88, 170.75, 170.22, 170.12, 169.44, 169.23, 138.52, 137.76, 128.86, 128.60, 128.51, 128.12, 127.72, 127.68, 127.38, 127.27, 101.87, 100.95, 99.53, 75.34, 74.33, 73.63, 72.62, 71.76, 70.88, 70.77, 69.66, 69.26, 68.96, 68.53, 67.23, 67.18, 62.08, 61.17, 54.88, 50.35, 47.32, 28.82, 27.34, 23.66, 23.30, 20.85, 20.79, 20.75, 20.71, 20.68, 20.65, 20.59. HRMS (MALDI-Tof)  $m/z$  calcd. for  $C_{63}H_{81}N_3O_{25}$   $[M+Na]^+$ : 1302.5057, found: 1302.5106,  $[\alpha]_D^{20} = -18.2^\circ$  ( $c=1$ ,  $CHCl_3$ )

**5- (benzyl (benzyloxycarbonyl)amino) pentyl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[2,3,4,6-tetra-O-acetyl- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 31**

To a solution of **27** (154 mg, 0.109mmol) in THF (3ml) was added  $BF_3 \cdot Et_2O$  (1M in THF, 0.163 ml, 0.163 mmol) dropwise. The yellow reaction mixture was refluxed for 1.5 hours after which it was placed on ice and quenched with MeOH. The reaction mixture was concentrated to dryness *in vacuo* before being redissolved in anhydrous pyridine (1 ml). Acetic acid (0.3 ml, 3.27mmol) was added and the reaction was left to stir overnight. MeOH was added and the reaction mixture was concentrated *in vacuo*. The crude was purified by FCC (50 $\rightarrow$ 100% EtOAc:Hex) to yield **31** as a pale yellow solid, 125 mg, 90%.  $^1H$  NMR (500 MHz, Chloroform-d)  $\delta$  7.43 – 7.13 (m, 15H,Ar), 5.38 (dd,  $J$  = 12.7, 3.5, 1.1 Hz, 2H, H-4',H-4''), 5.24 (t,  $J$  = 9.2 Hz, 1H,H-

2'), 5.19 – 5.11 (m, 3H, OCH<sub>2Bn</sub>, H-2''), 5.02 (dd, J = 10.4, 3.4 Hz, 1H, H-3'), 4.95 (dd, J = 10.5, 3.4 Hz, 1H, H-3''), 4.88 (dd, 2H, H-1, 4-OCH<sub>Bn</sub>), 4.75 – 4.63 (m, 3H, H-1', H-3, 4-OCH<sub>Bn</sub>), 4.60 – 4.38 (m, 3H, H-1'', NCH<sub>2Bn</sub>), 4.22 – 4.09 (m, 4H, 2H-6', 2H-6''), 3.96 – 3.55 (m, 7H, H-4, H-5, H-5', H-5'', 2H-6, CH<sub>linker</sub>), 3.43 – 3.28 (m, 2H, H-2, CH<sub>linker</sub>), 3.16 (s, 2H, CH<sub>2linker</sub>), 2.15 – 1.93 (m, 27H, 9CH<sub>3</sub>), 1.54 (s, 4H, 2CH<sub>2linker</sub>), 1.37 – 1.21 (m, 2H, CH<sub>2linker</sub>). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 170.37, 170.31, 170.25, 170.17, 170.10, 170.07, 169.18, 138.26, 137.74, 101.95, 101.18, 99.26, 77.88, 75.12, 74.13, 70.83, 70.58, 69.88, 69.21, 68.91, 67.18, 67.06, 67.02, 61.26, 60.98, 55.21, 50.30, 47.29, 45.91, 29.63, 29.10, 28.71, 27.41, 23.59, 23.12, 20.81, 20.64, 20.61, 20.54. HRMS (MALDI-Tof) *m/z* calcd. for C<sub>63</sub>H<sub>80</sub>N<sub>2</sub>O<sub>26</sub> [M+Na]<sup>+</sup> : 1303.4893, found: 1303.5106, [α]<sub>D</sub><sup>20</sup> = -20.5 (c=1 CHCl<sub>3</sub>)

**5- (benzyl (benzyloxycarbonyl)amino) pentyl β-D-galactopyranosyl-(1→3)-[ 2-deoxy-2-acetamido-β-D-glucopyranosyl-(1→6)]-4-O-benzyl-2-deoxy-2-acetamido-β-D-galactopyranoside 29**

To a solution of **30** (218 mg, 0.170 mmol) in dry MeOH (9 ml) was added 0.5M NaOMe (2.5 ml, 1.24 mmol) RM was stirred at RT for 1 hour after which it was quenched with Amberlite® IR 120 (H). The filtrate was concentrated, purified by Sephadex LH-20 (D=3.5cm, H=45cm, MeOH) and lyophilized to yield **29** as a white powder, 153 mg, 100%. <sup>1</sup>H NMR (500 MHz, Methanol-d<sub>4</sub>) δ 7.45 – 7.40 (m, 2H, Ar), 7.40 – 7.20 (m, 12H, Ar), 7.18 (s, 1H, Ar), 5.15 (d, J = 15.8 Hz, 2H, OCH<sub>2Bn</sub>), 4.98 (d, J = 11.6 Hz, 1H, 4-OCH<sub>Bn</sub>), 4.67 (d, J = 11.5 Hz, 1H, 4-OCH<sub>Bn</sub>), 4.50 (s, 2H, NCH<sub>2Bn</sub>), 4.41 (brs, 1H, H-1), 4.36 (d, J = 8.4 Hz, 1H, H-1''), 4.30 (d, J = 7.6 Hz, 1H, H-1'), 4.06 (brs, 1H, H-2), 4.02 (d, J = 3.1 Hz, 1H, H-4), 3.91 – 3.73 (m, 6H, H6'<sub>a</sub>, H6''<sub>b</sub>, H6<sub>c</sub>, H-3, H-4'), 3.71 – 3.54 (m, 5H, H6'<sub>a</sub>, H6''<sub>b</sub>, H6<sub>c</sub>, H-5, H-2', H-2'', CH<sub>linker</sub>), 3.54 – 3.50 (m, 1H, H-5'), 3.49 – 3.33 (m, 3H, H-3', H-3'', CH<sub>linker</sub>), 3.24 (m, 3H, H-5', CH<sub>2linker</sub>), 1.92 (d, 6H, 2CH), 1.52 (s, 4H, 2CH<sub>2linker</sub>), 1.29 (s, 2H, CH<sub>2linker</sub>). <sup>13</sup>C NMR (126 MHz, MeOD) δ 174.24, 173.63, 158.36, 157.84, 140.36, 139.09, 137.95, 129.55, 129.48, 129.03, 128.88, 128.63, 128.36, 128.28, 107.00, 102.61, 102.42, 81.45, 77.79, 77.30, 76.93, 75.98, 75.65, 74.75, 74.47, 72.57, 71.99, 70.24, 70.12, 69.74, 68.29, 62.82, 62.67, 57.18, 53.64, 51.42, 51.22, 49.85, 47.44, 30.14, 28.85, 28.35, 24.19, 23.28, 23.20. HRMS (MALDI-Tof) *m/z* calcd. for C<sub>49</sub>H<sub>67</sub>N<sub>3</sub>O<sub>18</sub> [M+Na]<sup>+</sup>: 1008.4317, found: 1008.437, [α]<sub>D</sub><sup>20</sup> = -6.3 (c=0,1, MeOH)

**5- (benzyl (benzyloxycarbonyl)amino) pentyl β-D-galactopyranosyl-(1→3)-[ β-D-galactopyranosyl-(1→6)]-4-O-benzyl-2-deoxy-2-acetamido-β-D-galactopyranoside 32**

To a solution of **31** (120 mg, 94 μmol) in dry MeOH (4 ml) was added 0.5M NaOMe (1.5 ml, 750 μmol) RM was stirred at RT for 1 hour after which it was quenched with Amberlite® IR 120 (H). The filtrate was concentrated, purified by Sephadex LH-20 (D=3.5cm, H=45cm, MeOH) and lyophilized to yield **32** as a white powder, 87.5 mg, 99%. <sup>1</sup>H NMR (500 MHz, Methanol-d<sub>4</sub>) δ 7.46 – 7.16 (m, 15H, Ar), 5.15 (d, J = 16.3 Hz, 2H, OCH<sub>2Bn</sub>), 4.99 (d, J = 11.6 Hz, 1H, 4-OCH<sub>Bn</sub>), 4.70 (d, J = 11.6 Hz, 1H, 4-OCH<sub>Bn</sub>), 4.50 (s, 2H, NCH<sub>2Bn</sub>), 4.47 (s, 1H, H-1), 4.30 (d, J = 7.6 Hz, 1H,

H-1'), 4.24 (d,  $J = 7.5$  Hz, 1H, H-1''), 4.08 (dd,  $J = 15.4, 6.2$  Hz, 2H, H-2, H-4), 3.90 (dd,  $J = 10.9, 2.9$  Hz, 1H, H-3), 3.87 – 3.69 (m, 10H, H-4', H-4'', 2H-6, 2H-6', 2H-6'', H-5, CH<sub>linker</sub>), 3.58 (dd,  $J = 9.7, 7.6$  Hz, 1H, H-2'), 3.48 (tdd,  $J = 20.9, 10.2, 7.3$  Hz, 7H, H-2'', H-3', H-3'', H-5', H-5'', H-2'', CH<sub>linker</sub>), 3.22 (dd,  $J = 14.3, 6.9$  Hz, 2H, CH<sub>2linker</sub>), 1.97 – 1.89 (m, 3H, CH<sub>3</sub>), 1.57 – 1.42 (m, 4H, 2CH<sub>2linker</sub>), 1.28 – 1.21 (m, 2H, CH<sub>2linker</sub>). <sup>13</sup>C NMR (126 MHz, MeOD)  $\delta$  174.30, 140.51, 129.59, 129.55, 129.05, 128.36, 107.18, 105.20, 102.59, 81.67, 77.15, 76.96, 76.54, 75.63, 75.13, 74.97, 74.56, 72.63, 72.54, 70.28, 70.09, 69.83, 68.47, 62.85, 62.26, 53.74, 49.85, 30.24, 24.24, 23.27. HRMS (MALDI-Tof)  $m/z$  calcd. for C<sub>47</sub>H<sub>64</sub>N<sub>2</sub>O<sub>23</sub> [M+Na]<sup>+</sup>: 967.4052, found: 967.408,  $[\alpha]_D^{20} = -5.7$ (c=1, MeOH)

### p-Tolyl 4,6-di-O-benzylidene -1-thio- $\beta$ -D-galactopyranoside 37

To a suspension of p-tolyl  $\beta$ -D-thiogalactopyranoside<sup>6</sup> (845mg, 2.95 mmol) in dry acetonitrile were added CSA (137mg, 0.59mmol) and PhCH(OMe)<sub>2</sub> (775  $\mu$ l, 5.17 mmol) under argon. The solution was stirred at room temperature for 18h after which satisfactory conversion was observed by TLC. The solution was quenched with Et<sub>3</sub>N, concentrated *in vacuo* and purified by FCC (0 $\rightarrow$ 5% MeOH:DCM) to yield **37** as a white solid (873mg, 79%). <sup>1</sup>H NMR (500 MHz, Chloroform-d)  $\delta$  7.60 – 7.53 (m, 2H, Ar), 7.42 – 7.32 (m, 5H, Ar), 7.11 (d,  $J = 7.8$  Hz, 2H, Ar), 5.50 (s, 1H, CH<sub>Ph</sub>), 4.45 (d,  $J = 9.3$  Hz, 1H, H-1), 4.38 (dd,  $J = 12.5, 1.6$  Hz, 1H, H-6<sub>a</sub>), 4.20 (dd,  $J = 3.5, 1.2$  Hz, 1H, H-4), 4.02 (dd,  $J = 12.4, 1.8$  Hz, 1H, H-6<sub>b</sub>), 3.69 (dd,  $J = 9.2, 3.5$  Hz, 1H, H<sub>3</sub>), 3.63 (t,  $J = 9.2$  Hz, 1H, H-2), 3.54 (q,  $J = 1.5$  Hz, 1H, H-5), 2.36 (s, 3H, CH<sub>3</sub>). NMR consistent with literary precedents.<sup>6</sup>

### p-Tolyl 3-O-benzoyl-4,6-di-O-benzylidene -1-thio- $\beta$ -D-galactopyranoside 38

To a solution of purged solution of **38** (42mg, 112  $\mu$ mol) on activated molecular sieves in dry DCM under Ar, was added BzCN (15 $\mu$ l, 1.49, 123  $\mu$ mol). The solution was cooled to -50°C and DMAP(2mg, 16.8  $\mu$ mol) was added under Ar. The reaction mixture was stirred at -50°C until full conversion was observed by TLC. The RM was quenched with MeOH and sat.NH<sub>4</sub>Cl(aq), diluted in DCM and filtered through Celite. The filtrate was washed with NH<sub>4</sub>Cl(aq) and dried over MgSO<sub>4</sub> before concentrating. This crude was then purified by FCC to yield **38** as a white solid, m=40.8 mg, 76%. <sup>1</sup>H NMR (500 MHz, Chloroform-d)  $\delta$  8.05 – 8.01 (m, 2H, Ar), 7.62 – 7.57 (m, 2H, Ar), 7.56 – 7.50 (m, 1H, Ar), 7.42 – 7.31 (m, 8H, Ar), 7.08 (d,  $J = 8.0$  Hz, 2H, Ar), 5.49 (s, 1H, CH<sub>Ph</sub>), 5.18 (dd,  $J = 9.8, 3.4$  Hz, 1H, H-3), 4.62 (d,  $J = 9.4$  Hz, 1H, H-1), 4.48 (dd,  $J = 3.4, 1.1$  Hz, 1H, H-4), 4.41 (dd,  $J = 12.4, 1.7$  Hz, 1H, H-6<sub>a</sub>), 4.10 (t,  $J = 9.6$  Hz, 1H, H-2), 4.05 (dd,  $J = 12.4, 1.7$  Hz, 1H, H-6<sub>b</sub>), 3.67 (q,  $J = 1.5$  Hz, 1H, H-5), 2.36 (s, 3H, CH<sub>3</sub>). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>)  $\delta$  130.10, 133.81, 133.06, 128.48, 126.33, 128.93, 129.05, 128.12, 129.96, 100.80, 75.26, 87.79, 73.85, 69.18, 65.74, 69.19, 69.90, 21.08

<sup>6</sup> B. Yang, K. Yoshida, Z. Yin, H. Dai, H. Kavunja, M.H. El-Dakdouki, S. Sungsuwan, S.B. Dulaney, X. Huang, *Angew Chem Int Ed Engl.*, **2012**, 51,10185–10189

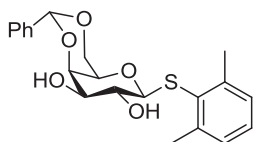
### p-Tolyl 2-O-benzoyl-4,6-di-O-benzylidene -1-thio-β-D-galactopyranoside **39**

To a suspension of **38** (40.8 mg, 85 μmol) in acetone (0.9 ml) at 0°C was added NaOH (0.05M, 0.9ml) and the slurry was vigorously stirred for 20 mins. The resulting precipitate was filtered and purified by FCC to yield **39**, as a white solid, 25mg, 63%. <sup>1</sup>H NMR (500 MHz, Chloroform-d) δ 8.11 – 8.06 (m, 2H, Ar), 7.62 – 7.56 (m, 1H, Ar), 7.47 (dt, *J* = 7.5, 3.7 Hz, 4H, Ar), 7.40 (qt, *J* = 10.8, 3.8 Hz, 5H, Ar), 7.09 (d, *J* = 7.9 Hz, 2H, Ar), 5.53 (s, 1H, CH<sub>Ph</sub>), 5.21 (t, *J* = 9.6 Hz, 1H, H-2), 4.77 (d, *J* = 9.7 Hz, 1H, H-1), 4.42 (dd, *J* = 12.4, 1.6 Hz, 1H, H-6<sub>a</sub>), 4.27 – 4.24 (m, 1H, H-4), 4.06 (dd, *J* = 12.4, 1.7 Hz, 1H, H-6<sub>b</sub>), 3.89 (dd, *J* = 9.5, 3.5 Hz, 1H, H-3), 3.63 – 3.58 (s, 1H, H-5), 2.35 (s, 3H, CH<sub>3</sub>). NMR consistent with literary precedents.<sup>7</sup>

### 2,6-Dimethylphenyl 1-thio-D-galactopyranoside **42**

To a purged solution of β-D-Galactose pentaacetate (3g, 7.69 mmol) in dry DCM (30ml) under Ar was added the 2,6-dimethylbenzenethiol (1.18ml, 8.85 mmol). The solution was stirred at 0°C for 40 mins then BF<sub>3</sub>.Et<sub>2</sub>O (4.7ml, 38.5 mmol) was added dropwise. The solution was stirred at RT for 5 hours until complete conversion was observed by TLC. The solution was quenched with MeOH, sat. NaHCO<sub>3</sub>(aq) and Na<sub>2</sub>CO<sub>3</sub>(s). The organic phase was washed with NaOH(aq), sat. NaCl(aq) and dried over MgSO<sub>4</sub>. The organic phase was filtered and concentrated before purifying by FCC to yield a white solid, 2.52g (69%), white solid. <sup>1</sup>H NMR (500 MHz, Chloroform-d) δ 7.20 – 7.09 (m, 3H, Ar), 5.38 (dd, *J* = 3.4, 1.2 Hz, 1H, H-4), 5.33 (t, *J* = 10.1 Hz, 1H, H-2), 5.00 (dd, *J* = 9.9, 3.5 Hz, 1H, H-3), 4.40 (d, *J* = 10.2 Hz, 1H, H-1), 4.13 – 4.01 (m, 2H, 2H<sub>6</sub>), 3.76 – 3.71 (m, 1H, H-5), 2.53 (s, 6H, 2CH<sub>3</sub>), 2.16 (d, *J* = 23.7 Hz, 6H, 2CH<sub>3</sub>), 1.98 (d, *J* = 9.1 Hz, 6H, 2CH<sub>3</sub>). NMR consistent with literary precedent.<sup>8</sup> The solid was purged and redissolved in dry MeOH (30ml) under Ar 0°C was added 0.5M NaOMe (47ml, 23.5mmol) dropwise. A precipitate was observed to form and the turbid solution was stirred at RT for 1 hour after which full conversion was observed by TLC. The reaction mixture was diluted with MeOH and quenched with Amberlite®IR 120 (H), filtered and concentrated to yield **42** as a white solid. <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ 7.15 – 7.08 (m, 3H, Ar), 5.19 (d, *J* = 5.9 Hz, 1H), 4.84 (d, *J* = 5.7 Hz, 1H), 4.58 (t, *J* = 5.5 Hz, 1H), 4.48 (d, *J* = 4.5 Hz, 1H), 4.12 (d, *J* = 9.6 Hz, 1H, H-1), 3.49 (dt, *J* = 10.7, 6.3 Hz, 1H, H-6<sub>a</sub>), 3.38 (td, *J* = 9.3, 5.8 Hz, 1H, H-2), 3.32 (dt, *J* = 10.9, 5.6 Hz, 1H, H-6<sub>b</sub>), 3.26 (ddd, *J* = 9.0, 5.6, 3.2 Hz, 1H, H-3), 3.18 (q, *J* = 5.7, 5.0 Hz, 1H, H-5), 2.51 (s, 6H, 2 CH<sub>3</sub>). NMR consistent with literary precedent.<sup>8</sup>

### 2,6-Dimethylphenyl 4,6-O-benzylidene-1-thio-β-D-galactopyranoside



<sup>7</sup>Z. Zhang, I.R. Ollmann, X-S. Ye, R. Wischnat, T. Baasov, C-H. Wong, *J. Am. Chem. Soc.*, **1999**, 121, 734-753

<sup>8</sup>Z. Li, J.C. Gildersleeve, *J. Am. Chem. Soc.*, **2006**, 128, 11612-11619

To a purged solution **42** (500mg, 1.67 mmol) in dry acetonitrile (30ml) were added CSA (77.3 mg, 33  $\mu$ mol) and PhCH(OMe)<sub>2</sub> (437  $\mu$ L, 2.92 mmol) under Ar. The mixture was stirred at room temperature for 18h. The mixture was quenched with Et<sub>3</sub>N, concentrated *in vacuo* to yield **43** as a white solid, 527mg, 81%. <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>)  $\delta$  7.38 (ddt, *J* = 16.7, 5.2, 2.3 Hz, 5H, Ar), 7.19 – 7.10 (m, 3H, Ar), 5.49 (s, 1H, CH<sub>Ph</sub>), 4.32 (d, *J* = 9.3 Hz, 1H, H-1), 4.04 (d, *J* = 3.3 Hz, 1H, H-4), 3.93 (dd, *J* = 12.3, 1.8 Hz, 1H, H-6<sub>a</sub>), 3.87 (dd, *J* = 12.2, 1.6 Hz, 1H, H-6<sub>b</sub>), 3.49 – 3.35 (m, 3H, H-2, H-5, H-3), 2.55 (s, 6H, 2CH<sub>3</sub>).

#### 2,6-Dimethylphenyl 4,6-O-benzylidene-3-O-benzoyl-1-thio- $\beta$ -D-galactopyranoside **44**

To a solution of purged solution of **43** (524mg, 1.35mmol) on activated molecular sieves in dry DCM under argon, was added BzCN (176 $\mu$ l, 1.49 mmol). The solution was cooled to -50°C and DMAP(25mg, 0.203mmol) was added under Ar. The reaction mixture was stirred at -50°C until full conversion was observed by TLC. The RM was quenched with MeOH and sat.NH<sub>4</sub>Cl(aq), diluted in DCM and filtered through Celite. The filtrate was washed with NH<sub>4</sub>Cl(aq) and dried over MgSO<sub>4</sub> before concentrating. This crude was then purified by FCC to yield **44** as a white solid, m=549mg, 83%. <sup>1</sup>H NMR (500 MHz, Chloroform-d)  $\delta$  8.11 – 8.06 (m, 2H, Ar), 7.59 – 7.54 (m, 1H, Ar), 7.47 – 7.40 (m, 4H, Ar), 7.38 – 7.34 (m, 3H, Ar), 7.22 – 7.18 (m, 1H, Ar), 7.17 – 7.13 (m, 2H, Ar), 5.47 (s, 1H, CH<sub>Ph</sub>), 5.13 (dd, *J* = 9.6, 3.6 Hz, 1H, H-3), 4.51 (d, *J* = 9.7 Hz, 1H, H-1), 4.45 (d, *J* = 3.6, 1.1 Hz, 1H, H-4), 4.24 (dd, *J* = 12.4, 1.6 Hz, 1H, H-6<sub>a</sub>), 4.19 (t, *J* = 9.7 Hz, 1H, H-2), 3.98 (dd, *J* = 12.4, 1.9 Hz, 1H, H-6<sub>b</sub>), 3.47 (s, 1H, H-5), 2.64 (s, 6H, 2CH<sub>3</sub>). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>)  $\delta$  166.47, 144.55, 137.90, 133.35, 130.54, 129.98, 129.70, 129.27, 128.97, 128.44, 128.37, 128.17, 126.32, 100.80, 90.77, 75.45, 73.86, 69.62, 69.28, 67.85, 53.50, 22.83. HRMS (MALDI-Tof) *m/z* calcd. for C<sub>28</sub>H<sub>28</sub>O<sub>6</sub>S [M+Na]<sup>+</sup>: 515.1503, found: 515.1548, [ $\alpha$ ]<sub>D</sub><sup>20</sup> = +32.2 (c=1, CHCl<sub>3</sub>)

#### 2,6-Dimethylphenyl 2-O-benzoyl-4,6-O-benzylidene-1-thio- $\beta$ -D-galactopyranoside **45**

To a suspension of **44**, (549 mg, 1.12mmol) in acetone (11.2ml) at 0°C was added NaOH (0.05M, 11.2ml) and the slurry was vigorously stirred for 20 mins. The resulting precipitate was filtered and purified by FCC to yield **45**, as a white solid, 304mg, 55.4 % <sup>1</sup>H NMR (500 MHz, Chloroform-d)  $\delta$  8.13 – 8.09 (m, 2H, Ar), 7.61 – 7.53 (m, 3H, Ar), 7.49 – 7.40 (m, 5H, Ar), 7.16 – 7.11 (m, 1H, Ar), 7.09 – 7.06 (m, 2H, Ar), 5.55 (s, 1H, CH<sub>Ph</sub>), 5.46 (t, *J* = 9.7 Hz, 1H, H-2), 4.60 (d, *J* = 10.0 Hz, 1H, H-1), 4.27 – 4.22 (m, 2H, H-4, H-6<sub>a</sub>), 4.00 (dd, *J* = 12.4, 1.9 Hz, 1H, H-6<sub>b</sub>), 3.89 – 3.83 (m, 1H, H-3), 3.39 (q, *J* = 1.5 Hz, 1H, H-5), 2.51 (s, 6H, 2CH<sub>3</sub>). NMR consistent with literary precedent.<sup>4</sup>

#### 2,6-Dimethylphenyl (3,4,6-tri-O-acetyl-2-deoxy-2-((2,2,2-trichloroethoxycarbonylamino)- $\beta$ -D-galactopyranosyl)-(1 $\rightarrow$ 3)- 2-O-benzoyl-4,6-O-benzylidene-1-thio- $\beta$ -D-galactopyranoside **46**

Acceptor **45** (40mg, 81.3  $\mu$ mol) and donor with the imidate **4** (91.5mg, 146  $\mu$ mol) were placed on activated molecular sieves under Ar. Dry DCM (1.5 ml) was added and the reaction mixture was stirred at room temperature for 20 mins before cooling to -40°C. TMSOTf (2.25  $\mu$ l, 12.2  $\mu$ mol) was added dropwise and the reaction was stirred at -20°C for 2 hours before quenching with Et<sub>3</sub>N. The reaction was allowed to warm to room temperature before filtering through



celite and concentrating *in vacuo*. The crude was purified by FCC to yield **46** as a white solid, m= 50 mg, 65 %. H NMR consistent with target structure. <sup>1</sup>H NMR (500 MHz, Chloroform-d) δ 8.13 – 8.08 (m, 2H, Ar), 7.64 – 7.57 (m, 3H, Ar), 7.52 – 7.37 (m, 6H, Ar), 7.20 – 7.00 (m, 3H, Ar), 5.69 (t, J = 9.9 Hz, 1H, H-2'), 5.55 (s, 1H, CH<sub>Ph</sub>), 5.27 (t, 1H, H-3'), 5.08 – 4.96 (m, 2H, H-1', H-4), 4.55 (d, J = 9.9 Hz, 1H, H-1), 4.41 – 4.35 (m, 2H, H-4, CH<sub>Troc</sub>), 4.23 – 4.17 (m, 2H, H-6<sub>a</sub>, H-6'<sub>a</sub>), 4.12 (dd, J = 12.2, 3.9 Hz, 1H, H-6'<sub>b</sub>), 4.05 – 3.96 (m, 2H, H-6<sub>b</sub>, H-3), 3.77 (d, J = 12.1 Hz, 1H, CH<sub>Troc</sub>), 3.65 (d, J = 10.1 Hz, 1H, H-5'), 3.44 (q, J = 9.1 Hz, 1H, H-2'), 3.33 (s, 1H, H-5), 2.42 (s, 6H, 2CH<sub>3</sub>), 1.97 (d, J = 3.2 Hz, 6H, 2CH<sub>3</sub>), 1.92 (s, 3H, CH<sub>3</sub>). <sup>13</sup>C NMR (126 MHz, Chloroform-d) δ 170.67, 170.22, 169.60, 165.43, 153.62, 144.40, 137.81, 133.49, 131.48, 130.13, 129.36, 129.14, 128.66, 128.52, 128.19, 126.57, 101.20, 100.44, 95.50, 88.86, 79.70, 76.37, 73.78, 71.79, 71.09, 69.89, 69.32, 68.75, 61.65, 56.46, 22.58, 20.86, 20.71, 20.62. HRMS (MALDI-Tof) *m/z* calcd. C<sub>43</sub>H<sub>46</sub>Cl<sub>3</sub>NO<sub>15</sub>S for [M+Na]<sup>+</sup>: 976.1550, found: 976.1583, [α]<sub>D</sub><sup>20</sup> = +30,8 (c=1, CHCl<sub>3</sub>)

## 5.3 Enzymatic transformations

### 5.3.1 Expression and purification of LgtA\_X and LgtA\_H

LgtA\_H sequence:

```
MQPLVSVLICAYNVEKYFAQSLAAVVNQTWRNLEILVDDGSTDGTLAIAQRFQEQDGRIRILAQPRNSGLIP
SLNIGLDELAKSGGGEYIARTDADDIAAPDWIEKIVGEMEKDRSIIAMGAWLEVLSEEKDGRLARHHRHGK
IWKKPTRHEDIADFFPFGNPIHNNTMIMRRSVIDGGLRYNTERDWAEDYQFWYDVSKLGLRAYYPEALVKY
RLHANQVSSKYSIRQHEIAQGIQKTARNDFLQSMGFKTRFDSLEYRQIKAVAYELLEKHLPEEDFERARRFLY
QCFKRTDTPPAGAWLDFAADGKMRRFLTMRQYFGLHRLIKNRHHHHHHH
```

LgtA\_X sequence:

```
MQPLVSVLICAYNVEKYFAQSLAAVVNQTWRNLEILVDDGSTDGTLAIAQRFQEQDGRIRILAQPRNSGLIP
SLNIGLDELAKSGGGEYIARTDADDIAAPDWIEKIVGEMEKDRSIIAMGAWLEVLSEEKDGRLARHHRHGK
IWKKPTRHEDIADFFPFGNPIHNNTMIMRRSVIDGGLRYNTERDWAEDYQFWYDVSKLGLRAYYPEALVKY
RLHANQVSSKYSIRQHEIAQGIQKTARNDFLQSMGFKTRFDSLEYRQIKAVAYELLEKHLPEEDFERARRFLY
QCFKRTDTPPAGAWLDFAADGKMRRFLTMRQYFGLHRLIKNRHHHHHHHSGSGMSDKIIHLTDDSFDTD
VLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAA
TKVGALSKGQLKEFLDANLAHHHHHHH
```

*Gene synthesis and plasmids construction*— Truncated *LgtA\_X* and *LgtA\_H* genes, with optimized nucleotide sequence for expression in *E. coli*, were synthesized by Genscript (Piscataway, NJ). Both genes were inserted in-frame into pET-12a plasmids between NdeI and BamHI sites. The new plasmids, henceforth named pET-12a\_LgtA\_X and pET-12a\_LgtA\_H, were transformed into *One Shot® BL21(DE3) Chemically Competent E. coli* (Thermo Fisher Scientific, MA) according to manufacturer's instructions. Transformants were selected on LB Agar Plates containing 100 µg/ml ampicillin and then incubated 24 hours at 37 °C.

*Clone selection and expression optimization*— Three colonies, for each construct, were used to screen for the best expresser. Each colony was used to inoculate 5 ml of LB broth with 100 µg/ml ampicillin and left incubating shaking at 280 rpm at 37 °C. When bacteria reached late-log phase ( $A_{600} = 0.8$ ) expression was induced with 0.5 mM isopropyl-β-d-thiogalactopyranoside (IPTG) for 18 hours at 16 °C. The culture was harvested by centrifugation at  $5,000 \times g$  for 15 min at 4 °C, and the pellet was resuspended in cold Tris-buffered saline (20 mM Tris-HCl pH 7.5, 0.5 M NaCl, containing 5 mM imidazole, and 0.5 mM phenylmethylsulfonyl fluoride). Cells were stored at -20 °C for 18 hours. The frozen cell suspensions were thawed, and mixed with xTractor Buffer (Clontech Laboratories, CA) and DNase I solution (Clontech Laboratories, CA). After 10 min at RT, lysates were centrifuged at  $14,000 \times g$  for 20 min at 4°C. Relative amount of protein in supernatant or pellet was estimated by SDS-PAGE. Similarly, to determine the optimal IPTG concentration, protein expression was induced using 0, 0.1, 0.3, 0.5 or 1 mM IPTG. Bacterial growth and lysate preparation was done as described above. Relative amount of protein in supernatant from the different IPTG concentrations was compared using SDS-PAGE. SDS-PAGE was in all cases done under reducing conditions and visualized by Colloidal Coomassie® Blue stain.

*Protein Expression and Purification*—BL21(DE3) strain harboring the plasmids pET-12a\_LgtA\_X or pET-12a\_LgtA\_H were grown at 37 °C in LB broth supplemented with 100 µg/ml ampicillin to late-log phase ( $A_{600} = 0.8$ ). Protein expression was induced with 0.1 mM IPTG 18 hours at 16 °C. The culture was harvested by centrifugation at  $5000 \times g$  for 15 min at 4 °C, and the pellet was resuspended in cold Tris-buffered saline (20 mM Tris-HCl, 0.5 M NaCl, 0.5 mM PMSF, pH 8) and stored at -20 °C. Frozen pellets were thawed by placing the tube in tap water, and placed immediately on ice. Bacterial lysis was done by sonication and triton X-100 was added to a final 1 % concentration. Lysates were incubated on ice for 30 min before separating soluble proteins by centrifugation at  $20,000 \times g$  for 25 min, at 4 °C. Supernatant was passed through a 0.20 µm filter and diluted in binding/washing buffer (20 mM Tris HCl, 500 mM NaCl, 5 mM Imidazole, pH 7.5) to a final volume of 50 mL. Purification was done using an FPLC (BioRAD, BioLogic DuoFlow) system, and loaded via superloop to a 1ml HisTrap HP column (GE Healthcare). After washing, bound proteins were released by addition of elution buffer (20 mM Tris HCl, 500 mM NaCl, 500 mM Imidazole, pH 7.5). Sample injection, washing, and elution was done at a flow rate of 1ml/min. The purified fraction buffer was made to contain 1mM DTT and concentrated by centrifugation. The buffer was then exchanged to (25 mM TrisHCl, 150 mM NaCl, 1mM DTT pH 7.0) using a PD10 column to remove the imidazole. Finally, the

eluted protein was concentrated by centrifugation to obtain a final protein concentration *ca.* 10mg/ml. Purification of the proteins was confirmed by SDS-PAGE and concentration was determined using a NanoDrop 1000 spectrophotometer (DE, Thermo Scientific) considering the extinction coefficient for LgtA\_X 73590 M<sup>-1</sup> cm, and 59485 M<sup>-1</sup> cm for LgtA\_H. The percentage of purity was determined from the Coomassie® Blue stained SDS-PAGE.

*Enzymatic activity tests*— GlcNAc transferase activity of LgtA\_H and LgtA\_X was confirmed using asialo, bis-galactosylated biantennary complex N-glycan (G1) as acceptor and UDP-GlcNAc as the donor substrate. Reaction conditions were based on those described by Blixt et al.<sup>9</sup> with the following modifications: 2  $\mu$ L of LgtA\_X (at 2.3 mg/ml) were added to a 23  $\mu$ L reaction volume containing 50 mM Tris HCl, pH 8.0, 10 mM MnCl<sub>2</sub>, 5mM MgCl<sub>2</sub>, 5 U of Alkaline Phosphatase (Sigma Aldrich, Spain), 5 equiv. UDP-GlcNAc, 1 eq.G1 (19  $\mu$ g), 0.2  $\mu$ g/ $\mu$ L BSA, and 1 mM 2- $\beta$ -mercaptoethanol. Reactions were left to proceed at 25 °C for 18 hours. Addition of one or two molecules of GlcNAc to G2 was confirmed by an increase in the G2 mass, as seen by MALDI-TOF analysis. The same method was used to confirm the capability of LgtA\_X to use as acceptors the following glycans: G1 (O-glycan core 1), *S.mansoni* core **32** and mucin core 2 **29**.

Optimum temperature and pH were determined after testing the LtgA\_X transferase activity at different temperatures (18 °C, 25 °C, 30 °C, 37 °C, 42 °C, and 60 °C), and pH values (6.5, 7.0, 7.5, 8.0, and 8.8).

#### **5.4 General procedure for elongation using GalT1 or DM GalT-1**

GalT1 and the DMGalT1 were expressed as described previously.<sup>10,11</sup>

To a solution of glycan(1 eq, 1 mM min.) in buffer (50 mM HEPES,20mM MnCl<sub>2</sub>, pH 7.5) were added UDP-Gal or UDP-GalNAc(1.2 eq), ALP (150 U/ $\mu$ mol) and Galt or DMGalT (0.344 mg/ $\mu$ mol). The reaction mixture was incubated at 37°C until satisfactory conversion was observed by MALDI. If necessary, additional donor and enzymes solutions were added to achieve satisfactory conversion. MeOH was added to precipitate the enzyme. The mixture was centrifuged at 4°C and the pellet washed with a solution of 1:1 H<sub>2</sub>O:MeOH. The combined supernatants were concentrated *in vacuo* and the crude purified.

---

<sup>9</sup>O. Blixt, I. van Die, T. Norberg, D.H. van den Eijnden, *Glycobiology*, **1999**,9, 1061-1071

<sup>10</sup>Boeggeman, E. E.; Ramakrishnan, B.; Qasba, P. K. *Protein Expres.Purif.* , 2003, 30, 219–229.

<sup>11</sup>Kawar, Z. S.; Van Die, I.; Cummings, R. D. *J. Biol. Chem.* 2002, 277, 34924–34932

## 5.5 General procedure for elongation using LgtA\_X

A solution of glycan(1 eq) in DMSO was dissolved in buffer (50 mM Tris.HCl, 20mM MnCl<sub>2</sub>, 5mM MgCl<sub>2</sub>, 1mM DTT, 2mM EDTA, pH 7.5) to a final concentration 10% DMSO v/v max. If necessary, the mixture was made more soluble by heating to 37°C. UDP-GlcNAc (1.2 eq), ALP (100 U/μmol) and LgtA\_X (0.344 mg/μmol) were added. The reaction mixture was incubated at room temperature until satisfactory conversion was observed by MALDI. If necessary, additional donor and enzymes solutions were added over time. When satisfactory conversion was obtained, MeOH was added to precipitate the enzyme. The mixture was centrifuged at 4°C and the pellet washed with a solution of 1:1 H<sub>2</sub>O:MeOH. The combined supernatants were concentrated *in vacuo* and the crude purified.

**5- (benzyl (benzyloxycarbonyl)amino) pentyl β-D-galactopyranosyl-(1→3)-[β-D-galactopyranosyl-(1→4)- 2-deoxy-2-acetamido-β-D-glucopyranosyl-(1→6)]-4-O-benzyl-2-deoxy-2-acetamido-β-D-galactopyranoside 48**

Compound **29** (29 mg, 29 μmol) was treated with a mixture **GalT-1** and **UPD-Gal** to yield **48** as white powder (20.5 mg, 61 %). This experiment was repeated 3 times. <sup>1</sup>H NMR (500 MHz, Methanol-d<sub>4</sub>) δ 7.46 – 7.35 (m, 4H, Ar), 7.34 – 7.21 (m, 10H, Ar), 7.18 (s, 1H, Ar), 5.16 (d, J = 14.4 Hz, 2H, OCH<sub>2Bn</sub>), 4.97 (d, J = 11.6 Hz, 1H, 4-OCH<sub>Bn</sub>), 4.68 (d, J = 11.7 Hz, 1H, 4-OCH<sub>Bn</sub>), 4.51 (s, 2H, NCH<sub>2Bn</sub>), 4.37 (m, J = 8.3 Hz, 3H, H-1, H-1'', H-1<sub>Gal</sub>), 4.30 (d, J = 7.5 Hz, 1H, H-1'), 4.06 (brs, 1H, H-2), 4.01 (s, 1H, H-2''), 3.90– 3.44 (m, 24H), 3.37 (brs, 1H, CH<sub>linker</sub>), 3.24 (brs, 2H, CH<sub>2 Linker</sub>), 1.96 – 1.89 (m, 6H, 2CH<sub>3</sub>), 1.52 (brs, 4H, 2CH<sub>2linker</sub>), 1.29 (brs, 2H, CH<sub>2linker</sub>). <sup>13</sup>C NMR (126 MHz, MeOD/D<sub>2</sub>O) δ 139.39, 138.59, 137.32, 129.93, 129.62, 129.58, 129.50, 129.34, 129.22, 128.96, 128.87, 128.79, 128.46, 128.06, 106.40, 104.21, 102.39, 102.18, 81.17, 79.75, 76.66, 76.56, 76.33, 75.84, 75.41, 74.48, 73.93, 73.81, 73.62, 72.18, 72.11, 70.35, 69.85, 69.77, 68.54, 68.45, 62.37, 62.12, 61.15, 56.01, 53.04, 51.16, 47.47, 29.56, 28.36, 27.98, 23.69, 23.23, 23.17. HRMS (MALDI-Tof) *m/z* calcd. for C<sub>55</sub>H<sub>77</sub>N<sub>3</sub>O<sub>23</sub> [M+Na]<sup>+</sup>: 1170.4846, found: 1170.4853, [α]<sub>D</sub><sup>20</sup> = -4.2° (c=0.4, MeOH),

**5- (benzyl (benzyloxycarbonyl)amino) pentyl β-D-galactopyranosyl-(1→3)-[2-deoxy-2-acetamido-β-D-glucopyranosyl-(1→3)-β-D-galactopyranosyl-(1→4)- 2-deoxy-2-acetamido-β-D-glucopyranosyl-(1→6)]-4-O-benzyl-2-deoxy-2-acetamido-β-D-galactopyranoside 49**

Compound **48** (10 mg, 8.71  $\mu$ mol) was treated with a mixture **LgtA\_X** and **UDP-GlcNAc** to yield **49** as a white powder (8.23 mg, 70%). This experiment was repeated once.  $^1\text{H}$  NMR (500 MHz, Methanol- $d_4$ )  $\delta$  7.44 (d,  $J$  = 7.4 Hz, 4H, Ar), 7.41 – 7.20 (m, 10H, Ar), 7.16 (d,  $J$  = 7.0 Hz, 1H, Ar), 5.14 (d,  $J$  = 20.7 Hz, 2H,  $\text{OCH}_{2\text{Bn}}$ ), 4.48 (s, 2H,  $\text{NCH}_{2\text{Bn}}$ ), 4.44 – 4.32 (m, 4H, H-1 $_{\text{GalNAc}}$ , H-1', H-1'', H-1 $_{\text{GlcNAc}}$ ), 4.11 – 4.02 (m, 3H, H-2 $_{\text{GalNAc}}$ , H4, H), 3.93 – 3.34 (m, 21H), 3.24 (dd,  $J$  = 27.2, 8.3 Hz, 2H,  $\text{CH}_{2\text{linker}}$ ), 2.03 (s, 3H,  $\text{CH}_3$ ), 1.94 (q,  $J$  = 7.6, 6.8 Hz, 6H, 2 $\text{CH}_3$ ), 1.59 – 1.39 (m, 4H, 2 $\text{CH}_{2\text{linker}}$ ), 1.26 (q,  $J$  = 25.3, 19.5 Hz, 2H,  $\text{CH}_{2\text{linker}}$ ).  $^{13}\text{C}$  NMR (126 MHz,  $\text{D}_2\text{O}/\text{MeOD}$ )  $\delta$  139.78, 129.79, 129.60, 129.26, 128.77, 106.61, 104.43, 103.99, 102.54, 102.32, 83.09, 80.08, 77.28, 76.91, 76.55, 76.31, 76.08, 75.52, 75.18, 74.06, 73.80, 72.31, 70.01, 68.47, 62.52, 62.21, 61.94, 61.38, 57.12, 56.15, 53.27, 47.79, 29.80, 23.92, 23.28, 23.18. HRMS (MALDI-Tof)  $m/z$  calcd. for  $\text{C}_{63}\text{H}_{90}\text{N}_4\text{O}_{28}$   $[\text{M}+\text{Na}]^+$ : 1373.5636, found: 1373.576

**5- (benzyl (benzyloxycarbonyl)amino) pentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[ $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)-2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)- 2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside **50****

Compound **49** (9.27 mg, 6.86  $\mu$ mol) was treated with a mixture **GalT-1** and **UDP-Gal** to yield **50** as a white powder (5.26 mg, 51%).

$^1\text{H}$  NMR (500 MHz,  $\text{MeOD}/\text{D}_2\text{O}$ , 273K)  $\delta$  7.50 – 7.10 (m, 15H, Ar), 5.14 (d,  $J$  = 20.9 Hz, 2H,  $\text{OCH}_{2\text{Bn}}$ ), 4.48 (s, 2H,  $\text{NCH}_{2\text{Bn}}$ ), 4.45 – 4.32 (m, 4H, 4H-1 $_{\text{Gal}}$ ), 4.11 – 4.01 (m, 3H), 3.94 – 3.48 (m, 33H), 3.42 (ddt,  $J$  = 9.3, 6.1, 3.6 Hz, 1H), 3.28 – 3.16 (m, 2H), 2.03 (s, 3H), 1.94 (q,  $J$  = 7.7, 6.9 Hz, 6H), 1.56 – 1.39 (m, 4H), 1.33 – 1.18 (m, 2H).  $^1\text{H}$  NMR (500 MHz,  $\text{MeOD}/\text{D}_2\text{O}$ , 323K)  $\delta$  7.46 – 7.15 (m, 15H, Ar), 5.14 (s, 2H,  $\text{OCH}_{2\text{Bn}}$ ), 4.94 (d,  $J$  = 11.7 Hz, 1H, 4- $\text{OCH}_{\text{Bn}}$ ), 4.74 – 4.66 (m, 2H, H1 $_{\text{GlcNAc}}$ , 4- $\text{OCH}_{\text{Bn}}$ ), 4.09 (d,  $J$  = 2.7 Hz, 1H), 4.05 – 3.99 (m, 2H, H2, H-4), 3.96 – 3.49 (m, 34H), 3.43 (dt,  $J$  = 9.7, 3.2 Hz, 2H), 3.22 (d,  $J$  = 7.7 Hz, 2H), 2.02 (s, 3H), 1.95 (d,  $J$  = 4.0 Hz, 6H), 1.53 – 1.43 (m, 4H, 2 $\text{CH}_{2\text{linker}}$ ), 1.36 – 1.23 (m, 2H,  $\text{CH}_{2\text{linker}}$ ).  $^{13}\text{C}$  NMR (126 MHz,  $\text{MeOD}$ )  $\delta$  129.83, 129.61, 129.28, 128.85, 128.54, 106.57, 104.40, 103.91, 102.52, 83.15, 76.86, 76.72, 76.50, 75.98, 75.50, 74.16, 73.52, 72.32, 71.38, 69.93, 68.48, 62.49, 62.24, 62.17, 61.15, 56.50, 56.11, 29.76, 23.88, 23.19. HRMS (MALDI-Tof)  $m/z$  calcd. for  $\text{C}_{69}\text{H}_{100}\text{N}_4\text{O}_{33}$   $[\text{M}+\text{Na}]^+$ : 1535.6164, found: 1535.6237

**5- (benzyl (benzyloxycarbonyl)amino) pentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[2-deoxy-2-acetamido- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)- 2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 51**

Compound **29** (8 mg, 6.72  $\mu$ mol) was treated with a mixture **DMGalT** and **UDP-GalNAc** to yield **51** as a white powder (5.56 mg, 59%).

$^1\text{H}$  NMR (500 MHz, Methanol- $d_4$ )  $\delta$  7.46 – 7.12 (m, 15H, Ar), 5.13 (d,  $J$  = 21.3 Hz, 2H,  $\text{OCH}_{2\text{Bn}}$ ), 4.93 (d,  $J$  = 11.7 Hz, 1H, 4- $\text{OCH}_{\text{Bn}}$ ), 4.51 – 4.45 (m, 3H,  $\text{NCH}_{2\text{Bn}}$ ,  $\text{H-1}_{\text{GalNAc}}$ ), 4.38 (h,  $J$  = 8.0, 7.0 Hz, 2H, H-1, H-1'), 4.32 (d,  $J$  = 8.2 Hz, 1H, H-1''), 4.03 (d,  $J$  = 3.2 Hz, 2H, H-2, H-4), 3.98 – 3.49 (m, 22H), 3.48 – 3.33 (m, 2H), 3.22 (dt,  $J$  = 26.0, 5.5 Hz, 2H,  $\text{CH}_{2\text{linker}}$ ), 2.05 (s, 3H,  $\text{CH}_3$ ), 1.94 (q,  $J$  = 7.8, 7.0 Hz, 6H, 2 $\text{CH}_3$ ), 1.56 – 1.39 (m, 4H, 2 $\text{CH}_{2\text{linker}}$ ), 1.25 (dd,  $J$  = 31.0, 16.6 Hz, 2H,  $\text{CH}_{2\text{linker}}$ ).  $^{13}\text{C}$  NMR (126 MHz, MeOD)  $\delta$  129.82, 129.59, 129.29, 128.83, 128.51, 106.49, 102.94, 102.44, 80.23, 76.80, 76.63, 76.43, 75.78, 75.47, 73.93, 73.76, 72.20, 69.92, 69.02, 62.44, 62.13, 55.86, 53.85, 53.16, 29.68, 23.81, 23.26, 23.18. HRMS (MALDI-Tof)  $m/z$  calcd. for  $\text{C}_{57}\text{H}_{80}\text{N}_4\text{O}_{23}$   $[\text{M}+\text{Na}]^+$ : 1211,5107, found: 1211,51

**5-(benzyl (benzyloxycarbonyl)amino) pentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[ 2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 34**

Compound **32** (12.7 mg,  $\mu$ mol) was treated with a mixture **LgtA\_X** and **UDP-GlcNAc** to yield **34** as a white powder (8.37mg, 57%). This experiment was repeated three times.

$^1\text{H}$  NMR (500 MHz, Methanol- $d_4$ )  $\delta$  7.45 – 7.16 (m, 15H, Ar), 5.16 (d,  $J$  = 14.8 Hz, 2H,  $\text{OCH}_{2\text{Bn}}$ ), 4.99 (d,  $J$  = 11.6 Hz, 1H, 4- $\text{OCH}_{\text{Bn}}$ ), 4.69 (d,  $J$  = 11.6 Hz, 1H, 4- $\text{OCH}_{\text{Bn}}$ ), 4.63 (d,  $J$  = 8.4 Hz, 1H,  $\text{H1}_{\text{GlcNAc}}$ ), 4.51 (s, 2H,  $\text{NCH}_{2\text{Bn}}$ ), 4.46 (s, 1H, H-1), 4.30 (d,  $J$  = 7.5 Hz, 1H, H-1'), 4.23 (s, 1H, H-1''), 4.12 – 4.04 (m, 3H), 3.92 – 3.65 (m, 14H), 3.60 – 3.32 (m, 10H), 3.30 – 3.18 (m, 2H), 1.99 (s, 3H,  $\text{CH}_3$ ), 1.92 (d,  $J$  = 14.5 Hz, 3H,  $\text{CH}_3$ ), 1.52 (s, 4H, 2 $\text{CH}_{2\text{linker}}$ ), 1.29 (s, 2H,  $\text{CH}_{2\text{linker}}$ ).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3:\text{MeOD}:\text{D}_2\text{O}$  15:15:2)  $\delta$  157.41, 139.65, 138.38, 137.09, 129.12, 129.06, 128.78, 128.73, 128.25, 128.08, 128.03, 127.87, 106.22, 104.21, 103.41, 101.82, 83.03, 80.77, 76.91, 76.35, 76.09, 75.25, 75.18, 73.72, 71.99, 70.92, 70.88, 70.02, 69.63, 68.88, 68.08, 62.31, 61.66, 61.41, 56.85, 53.20, 50.96, 47.96, 47.03, 30.26, 29.79, 23.71, 23.27, 23.24, 23.04. HRMS (MALDI-Tof)  $m/z$  calcd. for  $\text{C}_{55}\text{H}_{77}\text{N}_3\text{O}_{23}$   $[\text{M}+\text{Na}]^+$  : 1170.4846, found: 1170.4953,  $[\alpha]_{\text{D}}^{20} = -7.3$  (c=1,  $\text{CH}_3\text{Cl}_3:\text{MeOH}:\text{H}_2\text{O}$  15:15:2)

**5- (benzyl (benzyloxycarbonyl)amino) pentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)-2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 52**

Compound **34** (17 mg, 14.8  $\mu$ mol) was treated with a mixture **LgtA\_X** and **UDP-GlcNAc** to yield **52** as a white powder (5.43 mg, 28%).

$^1\text{H}$  NMR (500 MHz, Methanol- $d_4$ )  $\delta$  7.42 (d,  $J$  = 7.5 Hz, 2H, Ar), 7.38 (s, 3H, Ar), 7.34 – 7.16 (m, 10H, Ar), 5.16 (d,  $J$  = 14.2 Hz, 2H,  $\text{OCH}_{2\text{Bn}}$ ), 4.99 (d,  $J$  = 11.5 Hz, 1H, 4- $\text{OCH}_{\text{Bn}}$ ), 4.69 (d,  $J$  = 11.6 Hz, 1H, 4- $\text{OCH}_{\text{Bn}}$ ), 4.66 (d,  $J$  = 8.2 Hz, 1H, H-1 $_{\text{GlcNAc}}$ ), 4.51 (s, 2H,  $\text{NCH}_{2\text{Bn}}$ ), 4.46 (s, 1H, H-1), 4.38 (d,  $J$  = 7.6 Hz, 1H, H-1 $_{\text{Gal}}$ ), 4.30 (d,  $J$  = 7.5 Hz, 1H, H-1'), 4.22 (d,  $J$  = 7.6 Hz, 1H, H-1''), 4.11 – 4.03 (m, 3H, H-2, H-4), 3.93 – 3.38 (m, 21H), 3.24 (d,  $J$  = 15.4 Hz, 1H,  $\text{CH}_{\text{linker}}$ ), 1.98 (s, 3H,  $\text{CH}_3$ ), 1.93 (d,  $J$  = 13.6 Hz, 3H,  $\text{CH}_3$ ), 1.49 (d,  $J$  = 30.8 Hz, 4H,  $2\text{CH}_{2\text{linker}}$ ), 1.29 (d,  $J$  = 3.9 Hz, 2H,  $\text{CH}_{2\text{linker}}$ ).  $^{13}\text{C}$  NMR (126 MHz, MeOD/ $D_2\text{O}$ )  $\delta$  140.53, 132.39, 129.61, 129.54, 129.06, 128.37, 107.21, 105.18, 105.06, 104.16, 102.63, 83.61, 81.72, 80.49, 77.16, 77.00, 76.49, 76.14, 75.66, 74.83, 74.58, 73.97, 72.64, 72.59, 71.68, 70.32, 69.91, 69.74, 69.10, 62.90, 62.55, 62.21, 61.70, 56.89, 53.74, 51.49, 40.18, 31.62, 30.25, 24.95, 24.25. HRMS (MALDI-Tof)  $m/z$  calcd. for  $\text{C}_{61}\text{H}_{87}\text{N}_3\text{O}_{28}$   $[\text{M}+\text{Na}]^+$ : 1332.537, found: 1332.5435,  $[\alpha]_{\text{D}}^{20} = -4.2^\circ$  ( $c=0.45$  MeOH)

**5- (benzyl (benzyloxycarbonyl)amino) pentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[ 2-deoxy-2-acetamido- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)-2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 53**

Compound **34** (3.8 mg, 3.3  $\mu$ mol) was treated with a mixture **DMGalT** and **UDP-GalNAc** to yield **53** as a white powder (2.15 mg, 49%).  $^1\text{H}$  NMR (500 MHz, Methanol- $d_4$ )  $\delta$  7.42 – 7.11 (m, 15H, Ar), 5.13 (d,  $J$  = 16.4 Hz, 2H,  $\text{OCH}_{2\text{Bn}}$ ), 4.95 (d,  $J$  = 11.6 Hz, 1H, 4- $\text{OCH}_{\text{Bn}}$ ), 4.28 (d,  $J$  = 7.4 Hz, 1H, H-1'), 4.22 – 4.15 (m, 1H, H-1''), 4.09 (s, 1H, H-4), 4.04 – 3.89 (m, 3H), 3.86 – 3.32 (m, 26H), 3.26 – 3.15 (m, 2H), 2.03 – 1.89 (m, 9H,  $3\text{CH}_3$ ), 1.55 – 1.42 (m, 4H,  $2\text{CH}_{2\text{linker}}$ ), 1.33 – 1.26 (m, 2H,  $\text{CH}_{2\text{linker}}$ ).  $^{13}\text{C}$  NMR (126 MHz, MeOD)  $\delta$  129.04, 128.63, 128.56, 128.48, 128.38, 105.99, 103.89, 103.07, 102.42, 101.45, 82.86, 80.57, 79.53, 76.10, 76.06, 75.21, 75.13, 74.95, 74.31, 73.74, 73.32, 71.75, 69.14, 68.81, 68.62, 67.79, 61.34, 55.62, 53.24, 50.86, 49.12, 49.08, 47.22, 28.57, 23.16, 23.13. HRMS (MALDI-Tof)  $m/z$  calcd. for  $\text{C}_{61}\text{H}_{87}\text{N}_3\text{O}_{28}$   $[\text{M}+\text{Na}]^+$ : 1373.5639, found: 1373.5706

## 5.6 General procedure for deprotection of compounds

The glycan was dissolved in a mixture of 0.5ml MeOH, 2ml H<sub>2</sub>O and glacial acetic acid (1 ml/mg). Pd/C(10%) was added to the solution which was then purged with argon and charged with H<sub>2</sub>(g) using a balloon. The solution was stirred overnight under atmospheric pressure of H<sub>2</sub>. The reaction mixture was filtered through Celite and washed with 1:1 MeOH:H<sub>2</sub>O. The MeOH was removed from the combined filtrates by rotary evaporation then quenched with NaHCO<sub>3</sub>(s). The aqueous crude was concentrated *in vacuo* then purified by graphite carbon cartridge using a gradient of 0 → 100% MeOH:H<sub>2</sub>O. The fractions of interest were concentrated *in vacuo* and lyophilized to yield the target material.

### 5-aminopentyl β-D-galactopyranosyl-(1→3)-[ 2-deoxy-2-acetamido-β-D-glucopyranosyl-(1→6)]-2-deoxy-2-acetamido-β-D-galactopyranoside **O1**

Compound **29** (4.02 mg, 4.07 μmol) was hydrogenated following the general procedure described above to yield **O1** as a white powder (2.2 mg, 77 %). <sup>1</sup>H NMR (500 MHz, Deuterium Oxide) δ 4.54 (d, J = 8.5 Hz, 1H, H-1''), 4.47 (d, J = 8.6 Hz, 1H, H-1), 4.44 (d, J = 7.7 Hz, 1H, H-1'), 4.16 (d, J = 3.3 Hz, 1H, H-4), 4.05 (q, 1H, CH<sub>2</sub>), 4.01 – 3.84 (m, 5H), 3.82 – 3.69 (m, 7H), 3.67 – 3.42 (m, 8H), 2.98 – 2.93 (m, 2H, NCH<sub>2</sub>), 2.03 (d, J = 5.4 Hz, 6H, 2CH<sub>3</sub>), 1.64 (dp, J = 21.6, 7.0, 6.4 Hz, 4H, 2CH<sub>2linker</sub>), 1.41 (p, J = 7.5, 7.0 Hz, 2H, CH<sub>2linker</sub>). <sup>13</sup>C NMR (126 MHz, D<sub>2</sub>O) δ 174.63, 174.40, 104.80, 101.45, 101.16, 79.63, 75.82, 74.98, 73.74, 73.48, 72.44, 70.56, 69.88, 69.80, 68.54, 68.23, 60.98, 60.67, 55.46, 51.21, 39.43, 28.03, 23.23, 22.19. HRMS (MALDI-Tof) *m/z* calcd. for C<sub>27</sub>H<sub>49</sub>N<sub>3</sub>O<sub>16</sub> [M+Na]<sup>+</sup> : 694.3009, found: 694.3064

### 5-aminopentyl β-D-galactopyranosyl-(1→3)-[ β-D-galactopyranosyl-(1→6)]-2-deoxy-2-acetamido-β-D-galactopyranoside **O2**

Compound **32** (4.02 mg, 4.25 μmol) was hydrogenated following the general procedure described above to yield **O2** as a white powder (2.68 mg, 100%). <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O) δ 4.53 (d, J = 8.6 Hz, 1H, H-1), 4.46 (dd, J = 9.6, 7.8 Hz, 2H, H-1', H-1''), 4.24 – 4.21 (m, 1H, H-4), 4.10 – 4.05 (m, 1H, CH), 4.04 – 3.98 (m, 1H, H-2), 3.97 – 3.85 (m, 7H), 3.82 – 3.60 (m, 10H), 3.53 (dd, J = 9.9, 7.8 Hz, 2H, H-2', H-2''), 2.92 (t, J = 7.6 Hz, 2H, NCH<sub>2</sub>), 2.03 (s, 3H, CH<sub>3</sub>), 1.68 – 1.57 (m, 4H, 2CH<sub>2linker</sub>), 1.44 – 1.36 (m, 2H, CH<sub>2linker</sub>). <sup>13</sup>C NMR (126 MHz, D<sub>2</sub>O) δ 174.63, 104.85, 103.19, 101.32, 79.71, 75.15, 74.96, 73.50, 72.67, 72.45, 70.71, 70.55, 70.16, 69.06, 68.55,



68.07, 60.96, 51.19, 39.52, 28.13, 27.34, 22.20, 22.16. HRMS (MALDI-Tof)  $m/z$  calcd. for  $C_{25}H_{46}N_2O_{16}$   $[M+Na]^+$ : 653.2744, found: 653.2708

**5-aminopentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[2-deoxy-2-acetamido- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)]-2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6)]-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside O3**

Compound **51** (3.2 mg, 2.69  $\mu$ mol) was hydrogenated following the general procedure described above to yield **O3** a white powder (2.3 mg, 98 %).  $^1H$  NMR (500 MHz, Deuterium Oxide)  $\delta$  4.54 (dd,  $J = 9.8, 7.8$  Hz, 2H, H-1'', H-1<sub>GalNAc</sub>), 4.45 (dd,  $J = 15.9, 8.1$  Hz, 2H, H-1', H-1), 4.15 (d,  $J = 3.2$  Hz, 1H, H-4), 4.05 (q,  $J = 7.5$  Hz, 1H, CH), 4.01 – 3.49 (m, 27H), 2.97 (t, 2H, NCH<sub>2</sub>), 2.08 (s, 3H, CH<sub>3</sub>), 2.02 (d, 6H, 2CH<sub>3</sub>), 1.64 (dp,  $J = 27.2, 6.9, 6.1$  Hz, 4H, 2CH<sub>2linker</sub>), 1.42 (p,  $J = 8.0$  Hz, 2H, CH<sub>2linker</sub>).  $^{13}C$  NMR (126 MHz, D<sub>2</sub>O)  $\delta$  174.69, 174.58, 174.30, 104.75, 101.65, 101.26, 101.11, 79.56, 79.05, 75.26, 74.93, 74.41, 73.44, 72.39, 70.57, 70.51, 69.74, 69.44, 68.50, 68.17, 67.51, 62.41, 60.94, 60.88, 60.04, 54.71, 52.47, 51.15, 39.33, 27.98, 22.14, 22.11. HRMS (MALDI-Tof)  $m/z$  calcd. for  $C_{35}H_{62}N_4O_{21}$   $[M+Na]^+$ : 897.3803, found: 897.3834

**5-aminopentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[ $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)]-2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6)]-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside O4**

Compound **48** (4.57 mg, 3.98  $\mu$ mol) was hydrogenated following the general procedure described above to yield **O4** as a white powder (2.84 mg, 86%).  $^1H$  NMR (500 MHz, Deuterium Oxide)  $\delta$  4.57 (d,  $J = 8.1$  Hz, 1H, H-1''), 4.48 (dd,  $J = 8.2, 3.4$  Hz, 2H, H1, H1<sub>Gal</sub>), 4.44 (d,  $J = 7.7$  Hz, 1H, H-1'), 4.16 (d,  $J = 3.2$  Hz, 1H, H-4), 4.10 – 4.03 (m, 1H, CH), 4.03 – 3.96 (m, 2H, H-6<sub>a</sub>, H-2), 3.95 – 3.49 (m, 24H), 2.96 (t,  $J = 7.7$  Hz, 2H, CH<sub>2linker</sub>), 2.03 (d,  $J = 4.3$  Hz, 6H, 2CH<sub>3</sub>), 1.64 (dp,  $J = 21.3, 7.0, 6.3$  Hz, 4H, 2CH<sub>2linker</sub>), 1.42 (q,  $J = 7.9$  Hz, 2H, CH<sub>2linker</sub>).  $^{13}C$  NMR (126 MHz, D<sub>2</sub>O)  $\delta$  180.51, 174.63, 174.35, 104.80, 102.87, 101.36, 101.16, 79.63, 78.43, 75.35, 74.98, 74.70, 73.49, 72.48, 72.44, 72.36, 70.93, 70.56, 69.77, 69.61, 68.52, 68.22, 61.00, 60.01, 54.98, 51.20, 39.30, 28.00, 26.38, 22.71, 22.23, 22.19, 22.15. HRMS (MALDI-Tof)  $m/z$  calcd. for  $C_{33}H_{59}N_3O_{21}$   $[M+Na]^+$ : 856.3538, found: 856.3587

**Synthesis of 5-aminopentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-{  $\alpha$ -L-fucopyranosyl-(1 $\rightarrow$ 3)-[  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)]-2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6)]-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside O5**

To a solution of **O4** (2mg, 2.4  $\mu$ mol) in buffer (80 mM MES, 30 mM MnCl<sub>2</sub>, 5 mM MgCl<sub>2</sub>, pH 6.5) was added 20 mM GDP-Fuc (144  $\mu$ L, 2.88  $\mu$ mol). ALP (100U) and CeFUT6 (200  $\mu$ L, 1.1

mg/ml) were added and the reaction was incubated at 37°C. After 38h, satisfactory conversion was observed. The reaction was stopped by addition of MeOH. The mixture was centrifuged at 4°C and the pellet washed with a solution of 1:1 H<sub>2</sub>O:MeOH. The combined supernatants were concentrated *in vacuo* and the crude purified to yield **O5** as a white solid, 1.76mg, 75%.

<sup>1</sup>H NMR (500 MHz, Deuterium Oxide) δ 5.12 (d, J = 4.0 Hz, 1H, H<sub>1Fuc</sub>), 4.86 – 4.81 (m, 1H, H-5<sub>Fuc</sub>), 4.57 (d, J = 8.4 Hz, 1H, H-1''), 4.49 – 4.42 (m, 3H, H-1, H-1<sub>Gal</sub>, H-1'), 4.16 (d, J = 3.3 Hz, 1H, H-4), 4.08 – 3.83 (m, 12H), 3.82 – 3.47 (m, 20H), 2.98 (t, J = 7.7 Hz, 2H, NCH<sub>2</sub>), 2.02 (s, 6H, 2CH<sub>3</sub>), 1.64 (dp, J = 30.4, 7.0, 6.4 Hz, 4H, 2CH<sub>2linker</sub>), 1.41 (p, J = 7.6 Hz, 2H, CH<sub>2linker</sub>), 1.18 (d, J = 6.6 Hz, 3H, CH<sub>3Fuc</sub>). <sup>13</sup>C NMR (126 MHz, D<sub>2</sub>O) δ 174.63, 174.10, 104.80, 101.81, 101.14, 98.63, 79.63, 75.30, 74.98, 74.90, 73.44, 73.31, 72.44, 71.87, 71.01, 70.56, 69.79, 69.66, 69.19, 68.54, 68.32, 68.23, 67.66, 66.71, 66.05, 61.49, 60.99, 59.73, 55.66, 51.20, 39.34, 22.15, 15.27. HRMS (MALDI-Tof) *m/z* calcd. for C<sub>39</sub>H<sub>69</sub>N<sub>3</sub>O<sub>25</sub> [M+Na]<sup>+</sup>: 1002.4116, found: 1002.4148

**5-aminopentyl β-D-galactopyranosyl-(1→3)-[2-deoxy-2-acetamido-β-D-glucopyranosyl-(1→3)]-β-D-galactopyranosyl-(1→6)]-2-deoxy-2-acetamido-β-D-galactopyranoside O6**

Compound **34** (10.1 mg, 8.77 μmol) was hydrogenated following the general procedure 4.4 described above to yield **O6** as a white powder (5.88 mg, 78%). <sup>1</sup>H NMR (500 MHz, Deuterium Oxide) δ 4.70 (d, J = 8.4 Hz, 1H, H<sub>1GlcNAc</sub>), 4.52 (d, J = 8.6 Hz, 1H, H-1), 4.44 (d, 2H, H-1', H-1''), 4.22 (d, J = 3.2 Hz, 1H, H-4), 4.15 (d, J = 3.2 Hz, 1H, H-4<sub>GlcNAc</sub>), 4.07 (dd, J = 9.5, 2.5 Hz, 1H, CH<sub>linker</sub>), 4.01 (dd, J = 10.9, 8.5 Hz, 1H, H-2), 3.90 (m, 6H), 3.82 – 3.42 (m, 17H), 2.90 (t, J = 7.6 Hz, 2H, NCH<sub>2</sub>), 2.03 (d, J = 5.1 Hz, 6H, 2CH<sub>3</sub>), 1.61 (dp, J = 13.1, 6.9, 6.1 Hz, 4H, 2CH<sub>2linker</sub>), 1.39 (p, J = 7.5, 7.1 Hz, 2H, CH<sub>2linker</sub>). <sup>13</sup>C NMR (126 MHz, D<sub>2</sub>O) δ 102.75, 101.34, 104.78, 103.20, 68.05, 68.24, 69.15, 51.17, 69.55, 73.35, 60.46, 79.70, 60.87, 55.72, 82.14, 74.86, 70.19, 72.47, 69.79, 73.52, 70.55, 69.65, 75.72, 39.58, 22.25, 27.64, 27.95, 28.01, 22.29. HRMS (MALDI-Tof) *m/z* calcd. for C<sub>25</sub>H<sub>46</sub>N<sub>2</sub>O<sub>16</sub> [M+Na]<sup>+</sup>: 856.3537, found: 856.3554

**5-aminopentyl β-D-galactopyranosyl-(1→3)-[β-D-galactopyranosyl-(1→4)-2-deoxy-2-acetamido-β-D-glucopyranosyl-(1→3)]-β-D-galactopyranosyl-(1→6)]-2-deoxy-2-acetamido-β-D-galactopyranoside O7**

Compound **52** (5.8 mg, 6.96 μmol) was treated with a mixture **GalT-1** and **UPD-Gal** to yield **O7** as a white powder (2.78 mg, 40%). <sup>1</sup>H NMR (500 MHz, Deuterium Oxide) δ 4.72 (d, J = 8.2 Hz, 1H, H-1<sub>GlcNAc</sub>), 4.52 (d, J = 8.6 Hz, 1H, H-1), 4.48 (d, J = 7.8 Hz, 1H), 4.44 (d, J = 7.8 Hz, 2H), 4.22 (d, J = 3.3 Hz, 1H), 4.15 (d, J = 3.4 Hz, 1H, H4), 4.07 (dd, J = 9.5, 2.4 Hz, 1H, H4), 4.03 – 3.50 (m,

25H), 2.87 (t, J = 7.5 Hz, 2H, NCH<sub>2</sub>), 2.03 (d, J = 3.7 Hz, 6H, 2CH<sub>3</sub>), 1.64 – 1.56 (m, 4H, 2CH<sub>2linker</sub>), 1.42 – 1.35 (m, 2H, CH<sub>2linker</sub>). <sup>13</sup>C NMR (126 MHz, D<sub>2</sub>O) δ 104.81, 103.18, 102.80, 102.60, 101.32, 82.20, 79.65, 78.13, 75.30, 74.93, 74.66, 74.52, 73.36, 72.45, 72.13, 70.90, 70.51, 70.06, 69.70, 69.13, 68.49, 68.20, 67.99, 60.96, 60.87, 59.82, 55.18, 51.14, 39.26, 28.05, 26.33, 22.15, 22.06. HRMS (MALDI-Tof) *m/z* calcd. for C<sub>39</sub>H<sub>69</sub>N<sub>3</sub>O<sub>26</sub> [M+Na]<sup>+</sup>: 1018.4065, found: 1018.4135

**5-aminopentyl β-D-galactopyranosyl-(1→3)-[2-deoxy-2-acetamido-β-D-galactopyranosyl-(1→4)-2-deoxy-2-acetamido-β-D-glucopyranosyl-(1→3)-β-D-galactopyranosyl-(1→6)]-2-deoxy-2-acetamido-β-D-galactopyranoside O8**

Compound **53** (2.17 mg, 1.61 μmol) was hydrogenated following the general procedure 4.4 described above to yield **O8** as a white powder (1.13mg, 68 %). <sup>1</sup>H NMR (500 MHz, Deuterium Oxide) δ 4.70 (d, J = 8.1 Hz, 1H, H-1<sub>GlcNAc</sub>), 4.52 (dd, J = 8.5, 3.2 Hz, 2H, H-1, H-1<sub>GalNAc</sub>), 4.44 (dd, J = 7.8, 2.3 Hz, 2H, H-1', H-1''), 4.22 (d, J = 3.3 Hz, 1H, H-4), 4.14 (d, 1H, H-4<sub>GlcNAc</sub>), 4.07 (dd, J = 9.6, 2.5 Hz, 1H, CH<sub>linker</sub>), 4.01 (dd, J = 10.9, 8.5 Hz, 1H, H-2), 3.96 – 3.49 (m, 31H), 2.95 (t, J = 7.6 Hz, 2H, NCH<sub>2</sub>), 2.07 (s, 3H, CH<sub>3</sub>), 2.03 (d, J = 1.2 Hz, 6H, 2CH<sub>3</sub>), 1.69 – 1.57 (m, 4H, 2CH<sub>2linker</sub>), 1.44 – 1.36 (m, 2H, CH<sub>2linker</sub>). <sup>13</sup>C NMR (126 MHz, D<sub>2</sub>O) δ 102.58, 101.49, 104.78, 103.22, 68.00, 68.22, 69.14, 51.13, 67.58, 52.49, 69.68, 69.59, 68.45, 73.34, 79.71, 59.88, 54.99, 60.92, 70.63, 72.27, 75.35, 82.20, 74.84, 59.88, 78.82, 70.16, 72.42, 69.80, 70.51, 74.31, 39.45, 22.24, 27.07, 28.20, 22.28. HRMS (MALDI-Tof) *m/z* calcd. for C<sub>41</sub>H<sub>72</sub>N<sub>4</sub>O<sub>26</sub> [M+Na]<sup>+</sup>: 1059,4331 found: 1059.4458

## 5.7 Microarray

**Hydrophobic ITO slides was prepared according to previously described procedures.**<sup>12,13</sup>

### NHS-activation of hydrophobic ITO

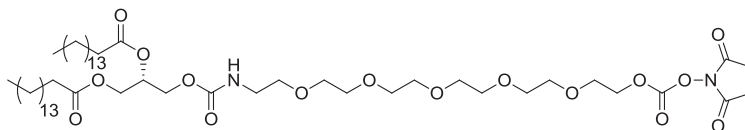
The hydrophobic ITO-slide was washed in acetone and placed in *vibrational vaporization machine (IMAGEPrep®)* machine in which a solution of the NHS-activated linker<sup>14</sup> (Figure 42) (15 mg/mL in 20mL CHCl<sub>3</sub>:MeOH 1:1) was sprayed on the ITO-surface for 3 min. The slide was

<sup>12</sup>A. Vega, P. Thissen, Y.J. Chabal, *Langmuir*, **2012**, 28, 8046–8051

<sup>13</sup>S-P. Pujari, L. Scheres, A.T.M. Marcelis, H. Zuilhof, *Angew. Chem. Int. Ed.*, **2014**, 53, 2-36.

<sup>14</sup>A. Beloqui, J. Calvo, S. Serna, S. Yan, I.B.H. Wilson, M. Martin-Lomas, N.C. Reichardt, *Proteomics*, **2013**, 52, 7477-7481

dried under stream of argon. This procedure is performed twice for each slide. Finally, the slide was sonicated in nanowater for 5 min and dried with steam of argon.



**Figure 42. NHS activated bidentate linker<sup>10</sup>**

### **Ligand immobilization**

Ligand solutions were prepared from stock solutions (1 mM in water) by dilution with sodium phosphate buffer (300mM, pH 8.7) to a final concentration of 50  $\mu$ M. A total amount of 40  $\mu$ L of each glycans solutions were placed into a 384 well source plate (*Scienion, Berlin, Germany*) which was stored at -20°C and reused if necessary. Glycans (50 drops of 246  $\mu$ L) were robotically printed onto NHS-activated hydrophobic-ITO slide with a distance between spots of 409  $\mu$ m in both, x and y axes. 24 glycans were spotted in 3 vertical replicates (8 different glycans/column), establishing the complete 24x12 array. After printing, the slides were placed in a 75 % humidity chamber at 20 °C for 18 hours. The slides were placed in a 50 mM solution of ethanolamine in sodium borate buffer, pH 9.0, for 10mins then washed in water dried under a stream of argon. Successful printing was determined by MALDI-TOF analysis.

### **Enzymatic elongation on-chip**

An enzymatic solution (53  $\mu$ L) containing HP-FucT, 80 mM MES, 30 mM MnCl<sub>2</sub>, 5 mM MgCl<sub>2</sub>, 0.2% BSA, 1mM GDP-Fuc or GDP-FucZ, pH 6.5 was freshly prepared and incubated with the array overnight at 37°C using the hybridization gasket(Agilent Technologies). The gasket was removed and the slide washed in an aqueous solution containing TFA (0.1% v/v) and ACN (0.05%v/v) before being rinsed in water and dried under a stream of argon. MALDI-TOF analysis was performed on a defined well to quantify conversion for each glycan. When necessary, the reaction process was repeated (2-3 times) to obtain satisfactory conversions.

### **Glycan binding analysis**

Labelled lectin solutions (100-200  $\mu$ L) were prepared from a stock solution of labeled lectin to a final concentration of 10  $\mu$ g/mL(Tris 500mM pH=7.5 with 150mM NaCl, 4mM CaCl<sub>2</sub>, 0.5% BSA). These solutions were incubated with the array using a 2x8 incubation-chambers (Agilent

technologies) overnight at 4° C for recombinant human CLRs or 1hr at 24°C for PNA. Solutions containing protein were removed and each subarray was washed with water before being dried under a stream of argon. The fluorescence was imaged by microarray scanner and quantified.

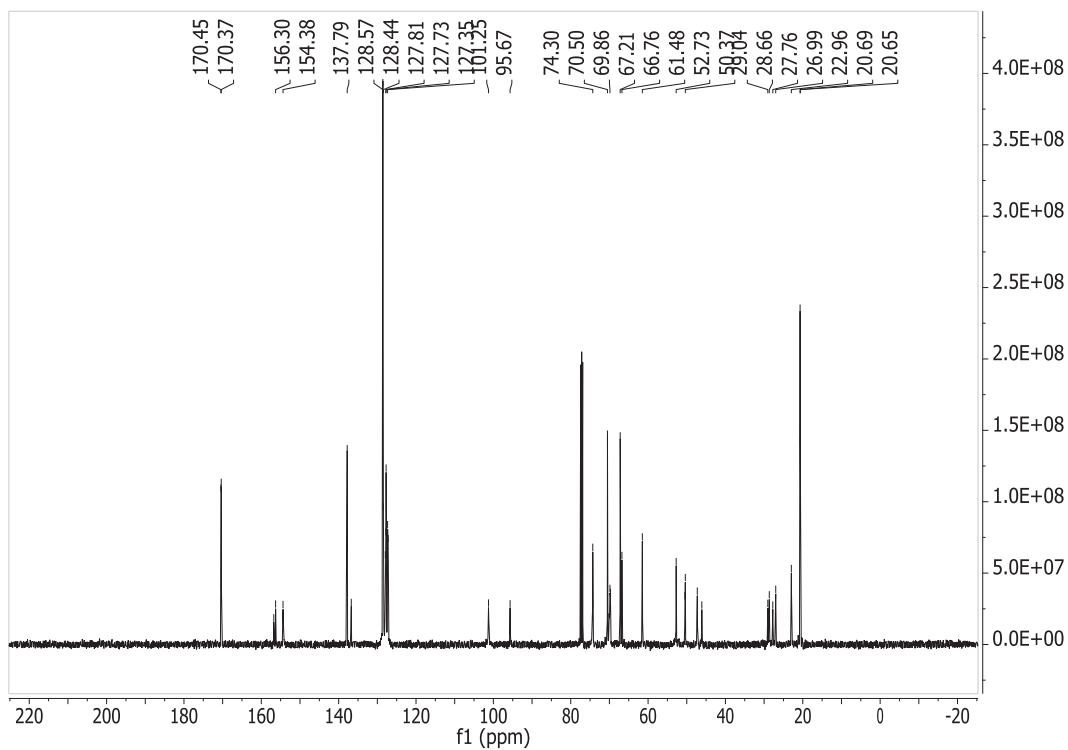
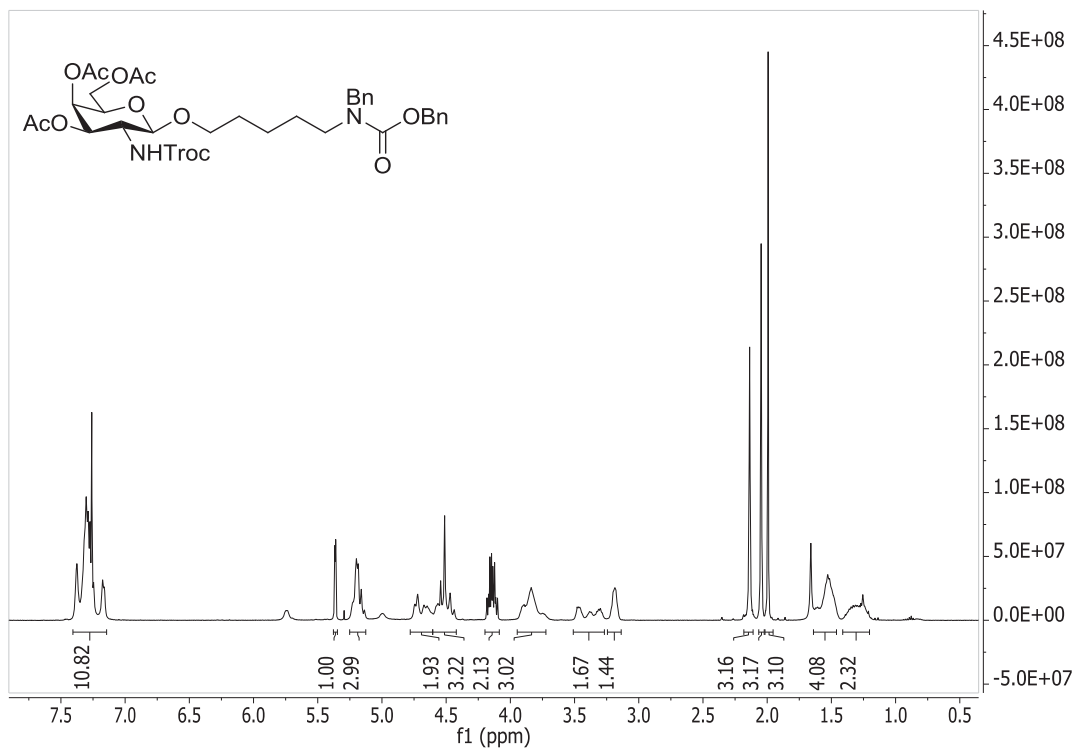


## **6 Appendix**

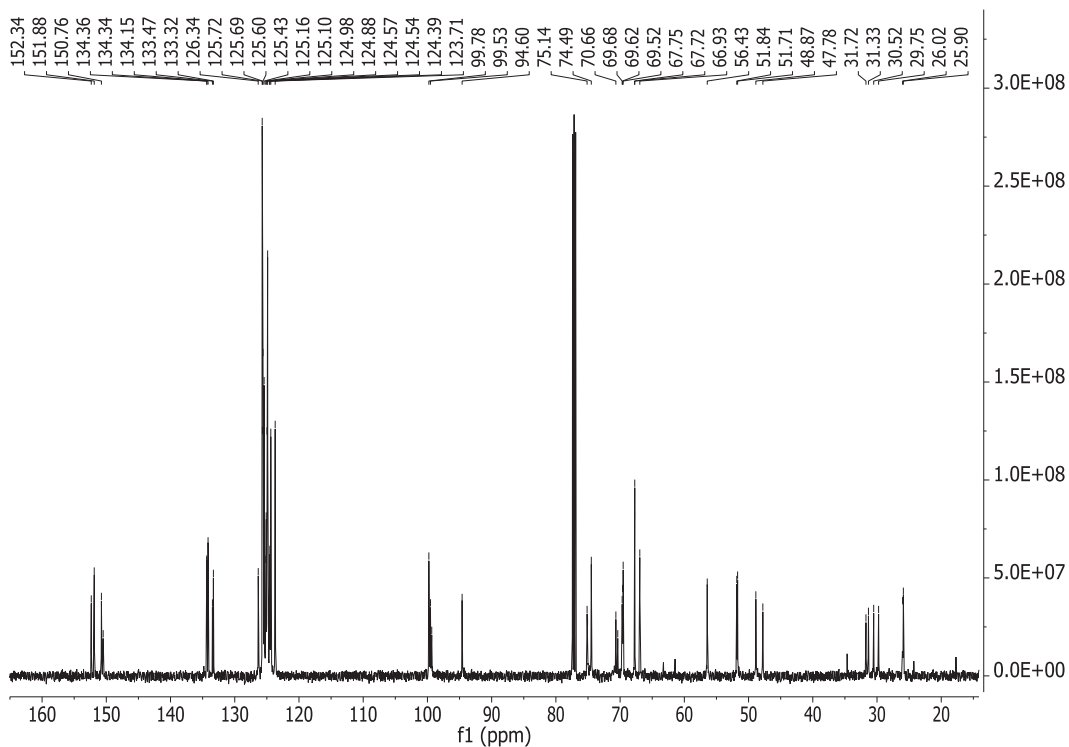
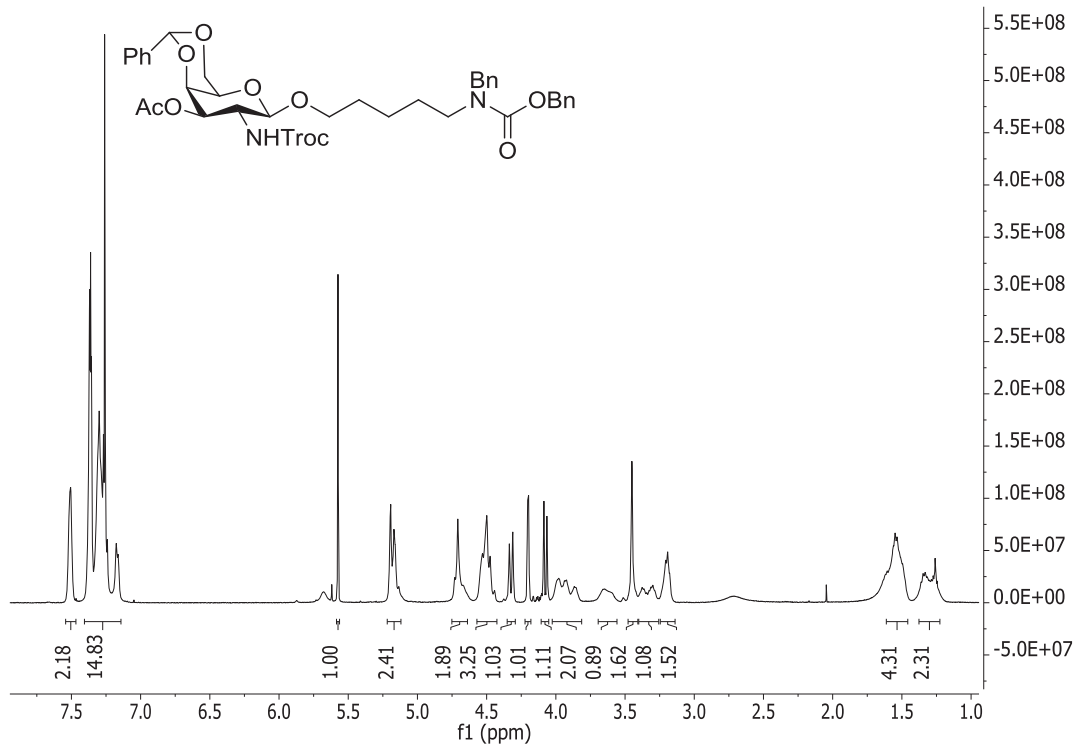




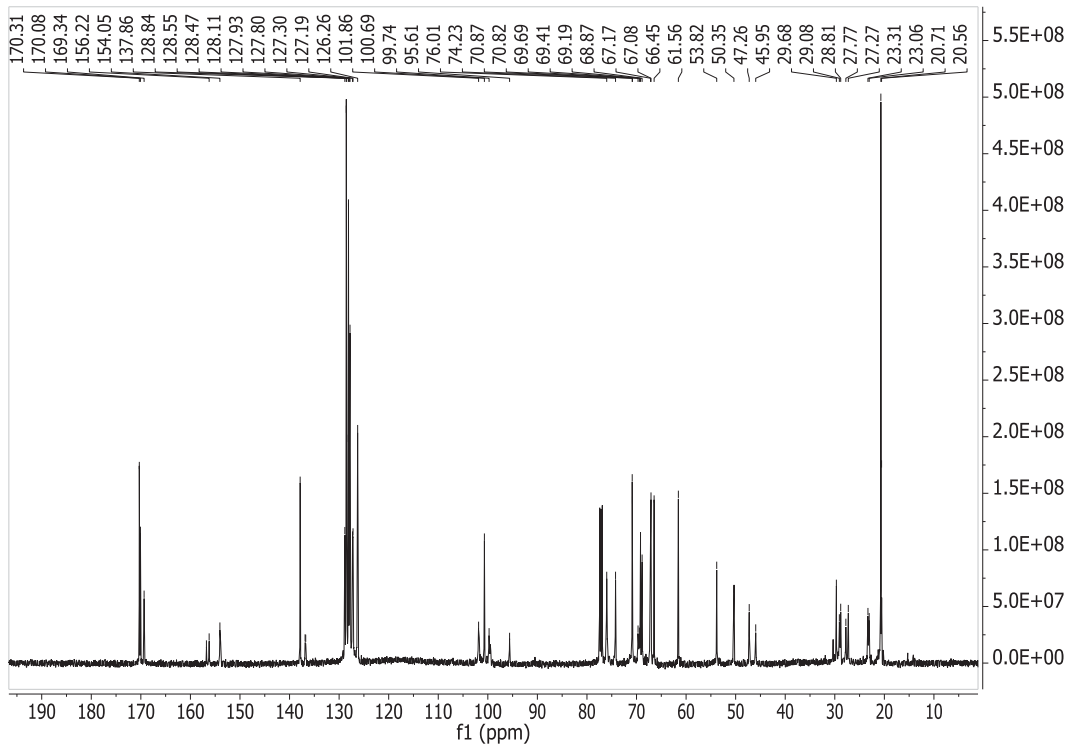
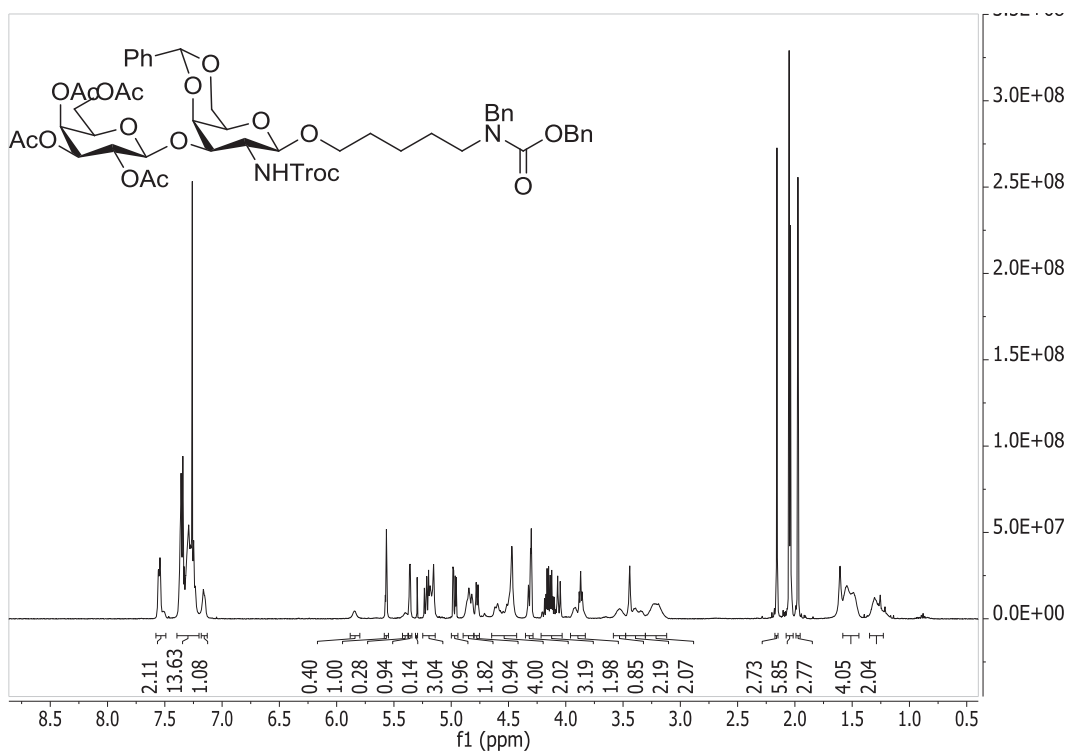
**5- (benzyl (benzyloxycarbonyl)amino) pentyl 3,4,6-tri-*O*-acetyl-2-deoxy-2- ((2,2,2-trichloroethoxy)carbonylamino)  $\beta$ -D-galactopyranoside 18**



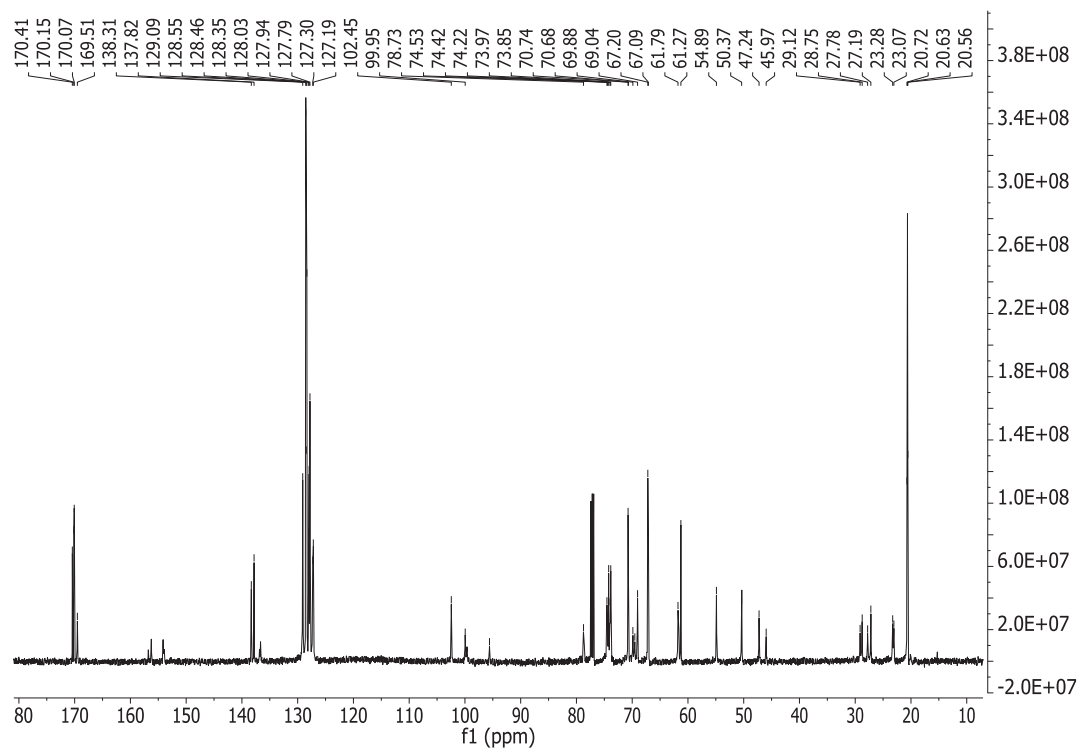
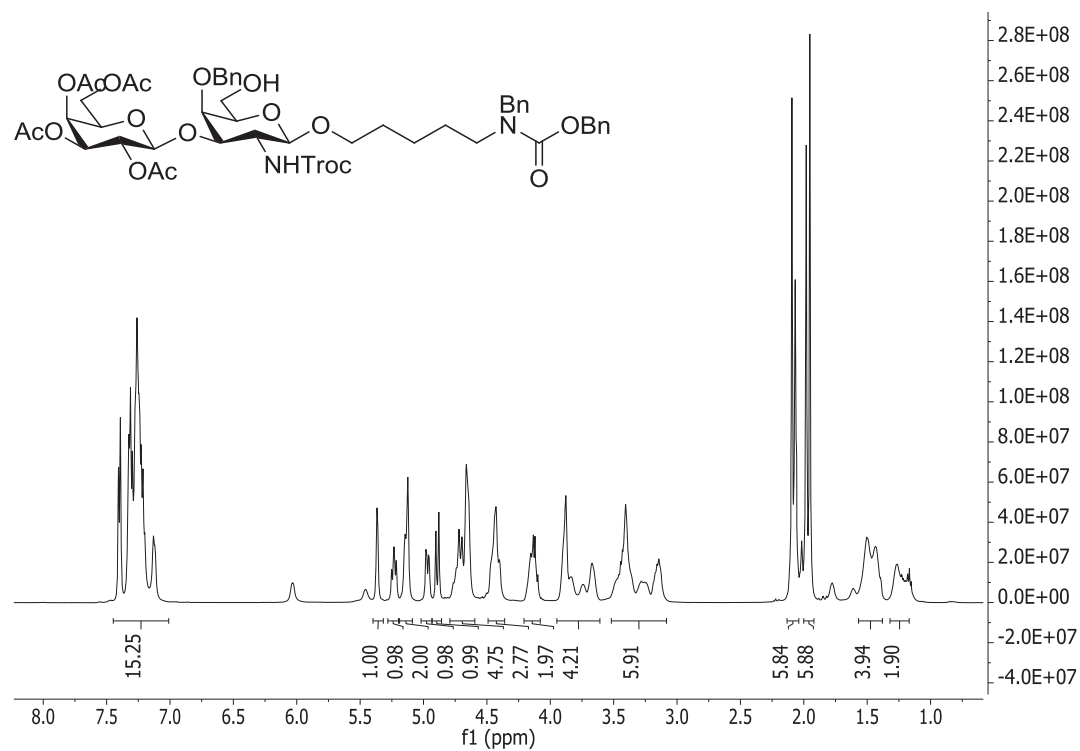
5- (benzyl (benzyloxycarbonyl)amino) pentyl 4,6-O-benzylidene-2-deoxy-2- ((2,2-trichloroethoxy-carbonyl)amino)  $\beta$ -D-galactopyranoside 19



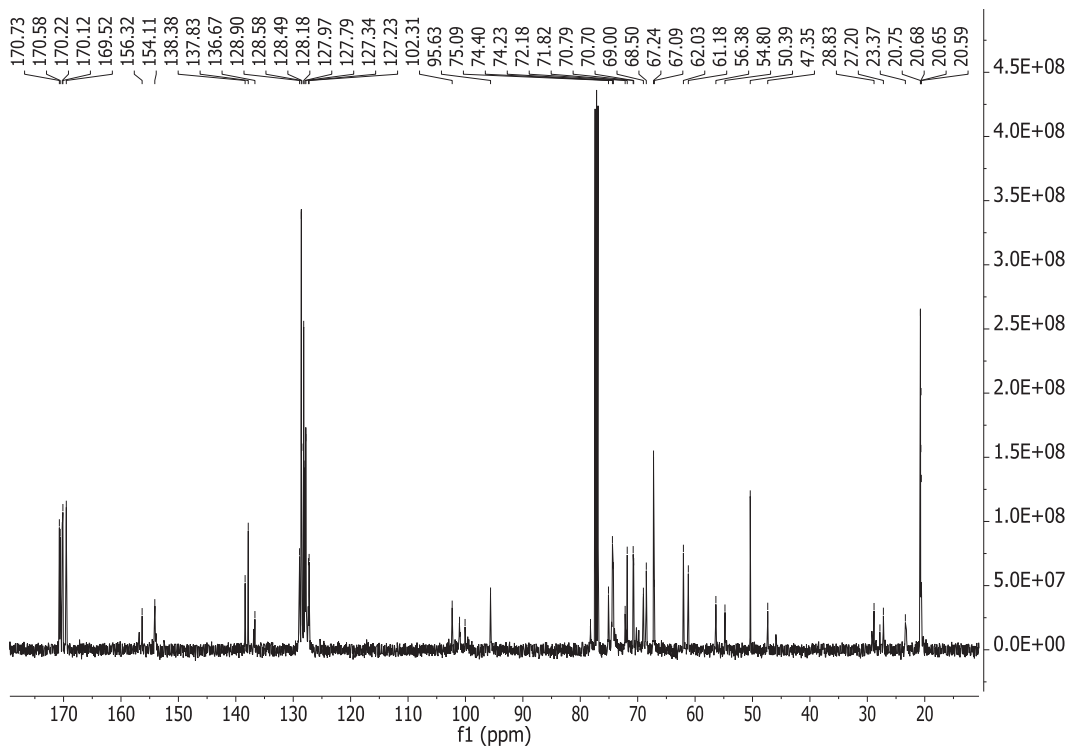
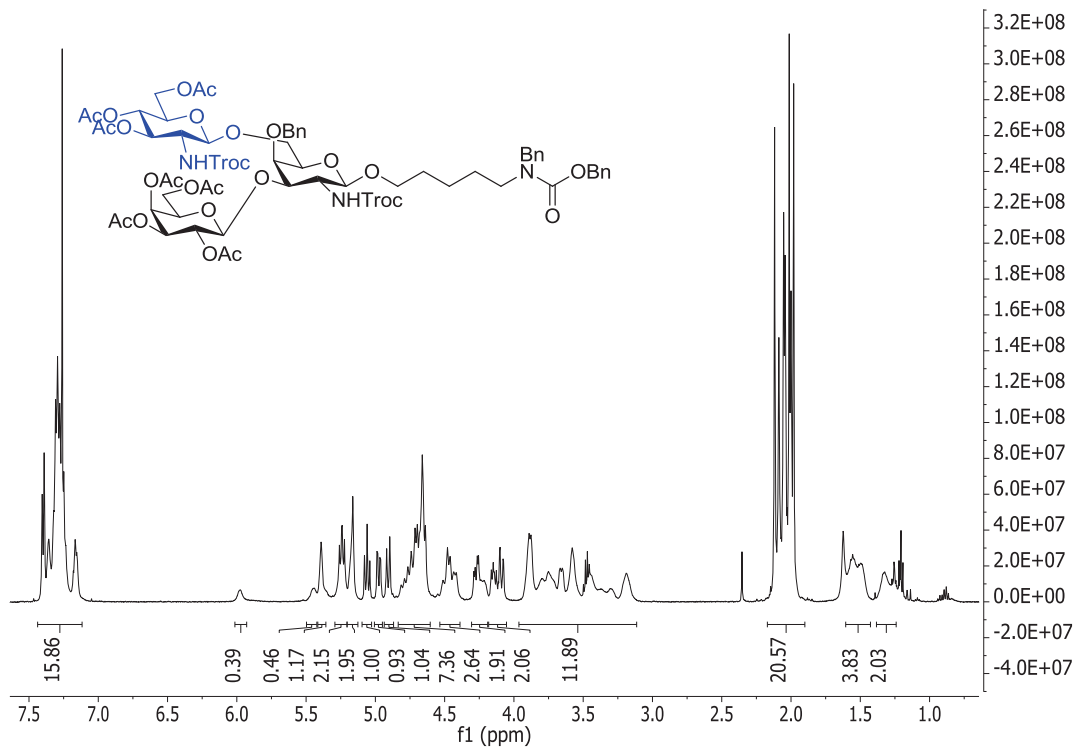
5- (benzyl (benzyloxycarbonyl)amino) pentyl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-4,6-O-benzylidene-2-deoxy-2-((2,2,2-trichloroethoxycarbonyl)amino)  $\beta$ -D-galactopyranoside 6



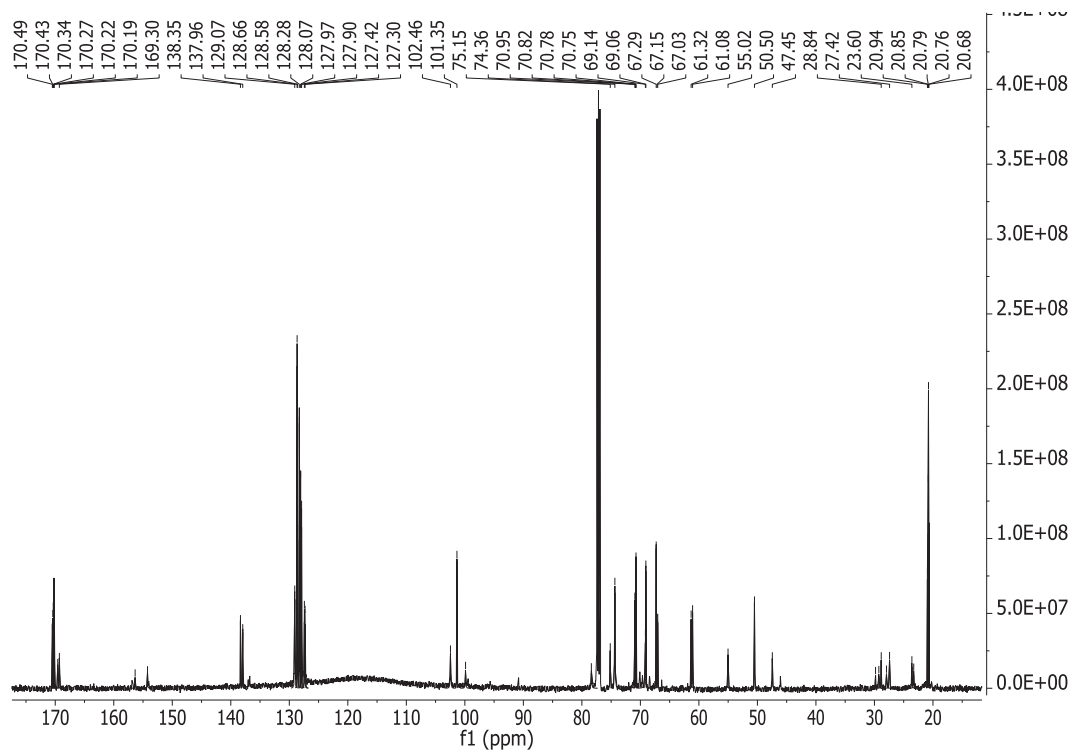
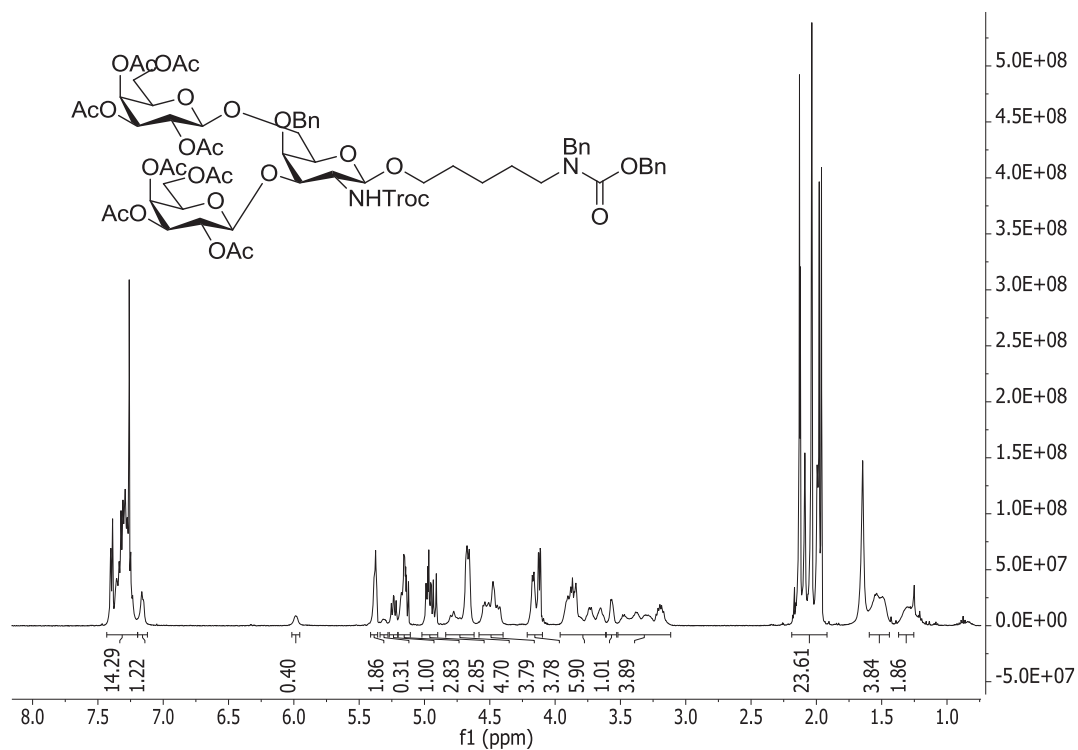
5- (benzyl (benzyloxycarbonyl)amino) pentyl 2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl-  
 (1→3)-4-O-benzyl-2-deoxy-2- ((2,2,2-trichloroethoxycarbonylamino) β-D-galactopyranoside  
 24



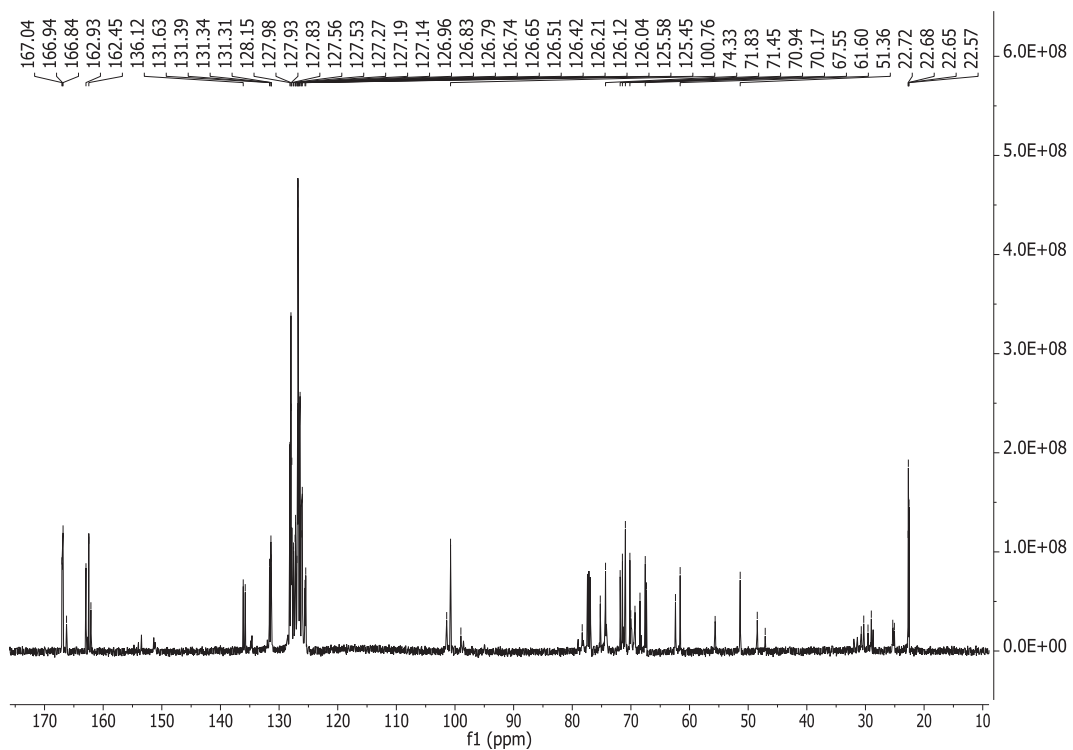
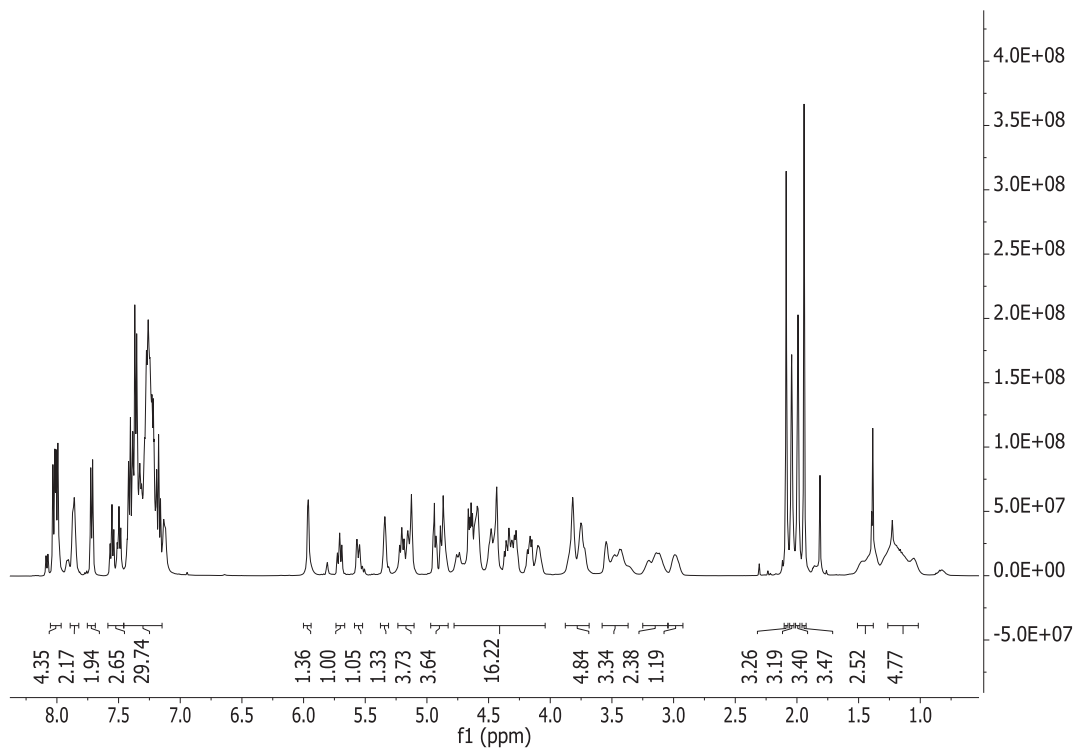
**5- (benzyl (benzyloxycarbonyl)amino) pentyl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[3,4,6-tri-O-acetyl-2-deoxy-2- ((2,2,2-trichloroethoxycarbonylamino)  $\beta$ -D-glucopyranosyl -(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2- ((2,2,2-trichloroethoxycarbonylamino)  $\beta$ -D-galactopyranoside 25**



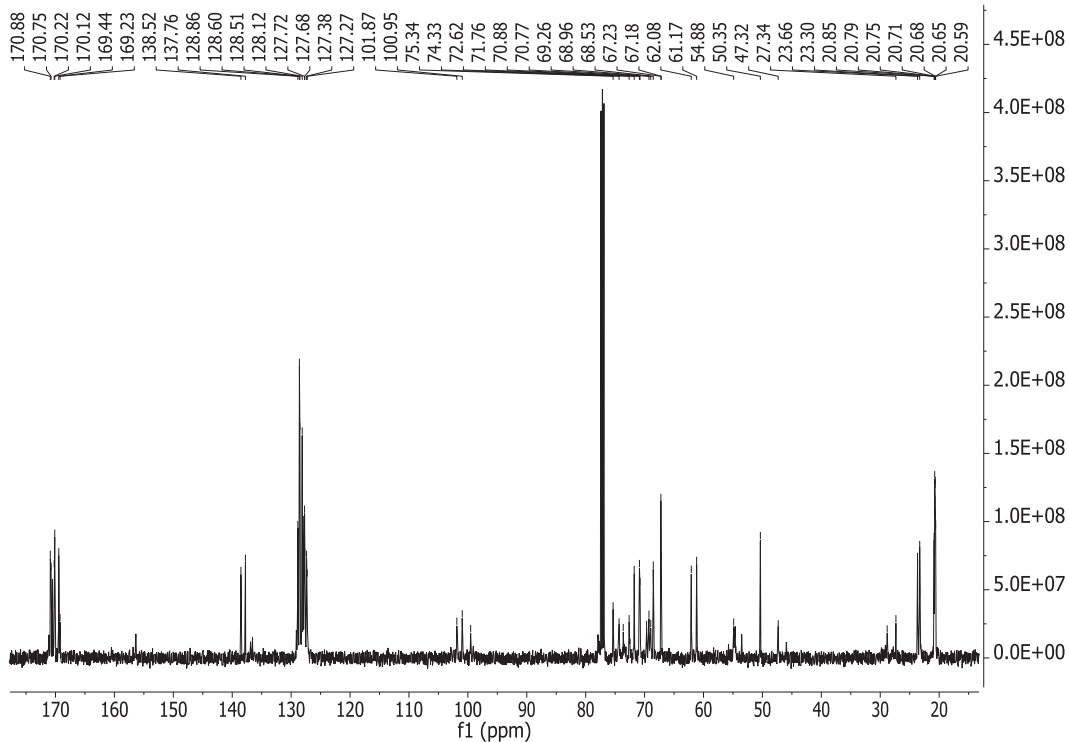
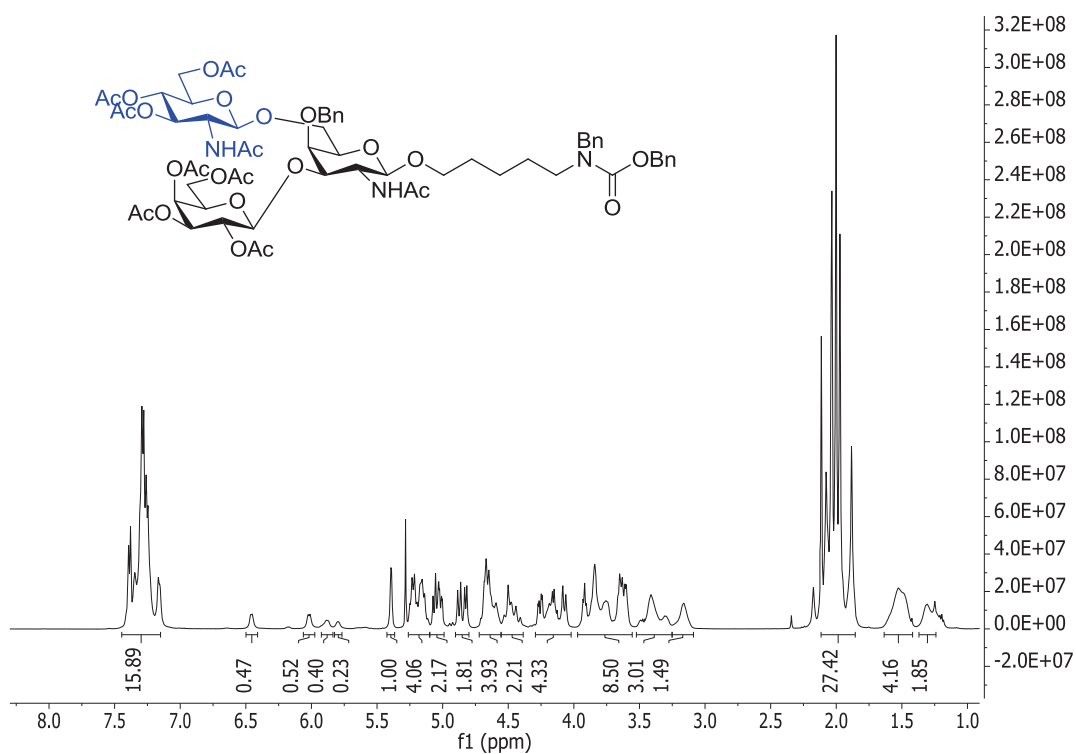
**5- (benzyl (benzyloxycarbonyl)amino) pentyl 2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl-(1→3)-[2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl-(1→6)]-4-O-benzyl-2-deoxy-2-((2,2,2-trichloroethoxycarbonyl)amino) β-D-galactopyranoside 27**



5- (benzyl (benzyloxycarbonyl)amino) pentyl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-  
 [2,3,4,6-tetra-O-benzoyl- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2- acetamido- $\beta$ -D-  
 galactopyranoside 28

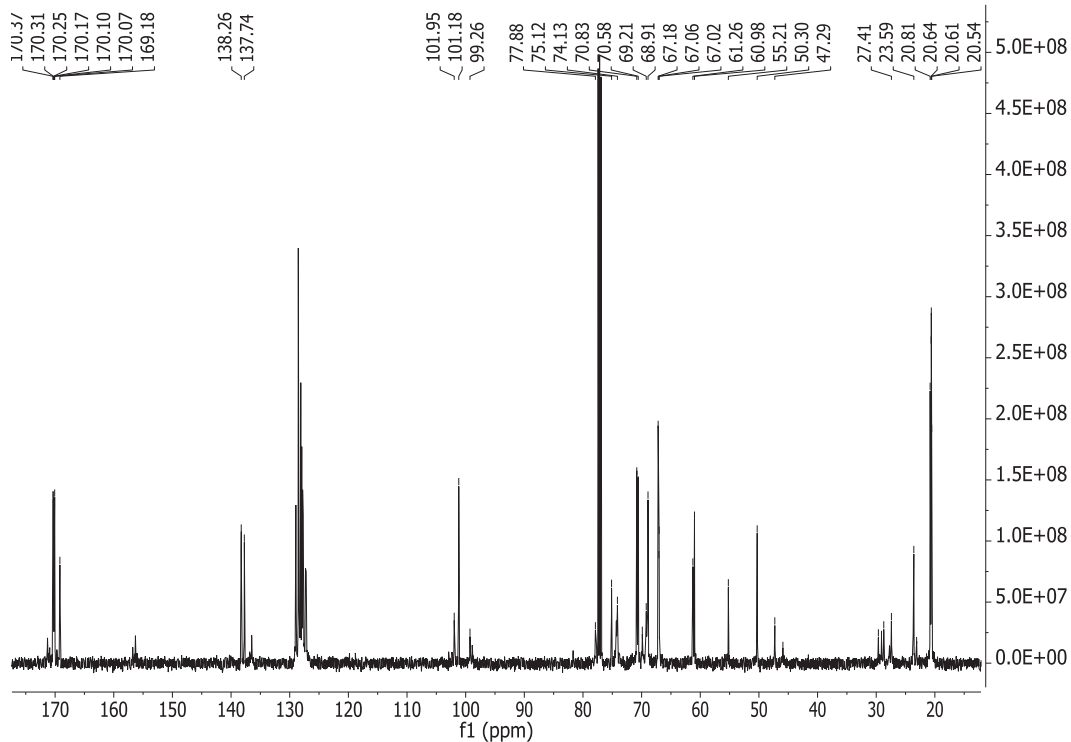
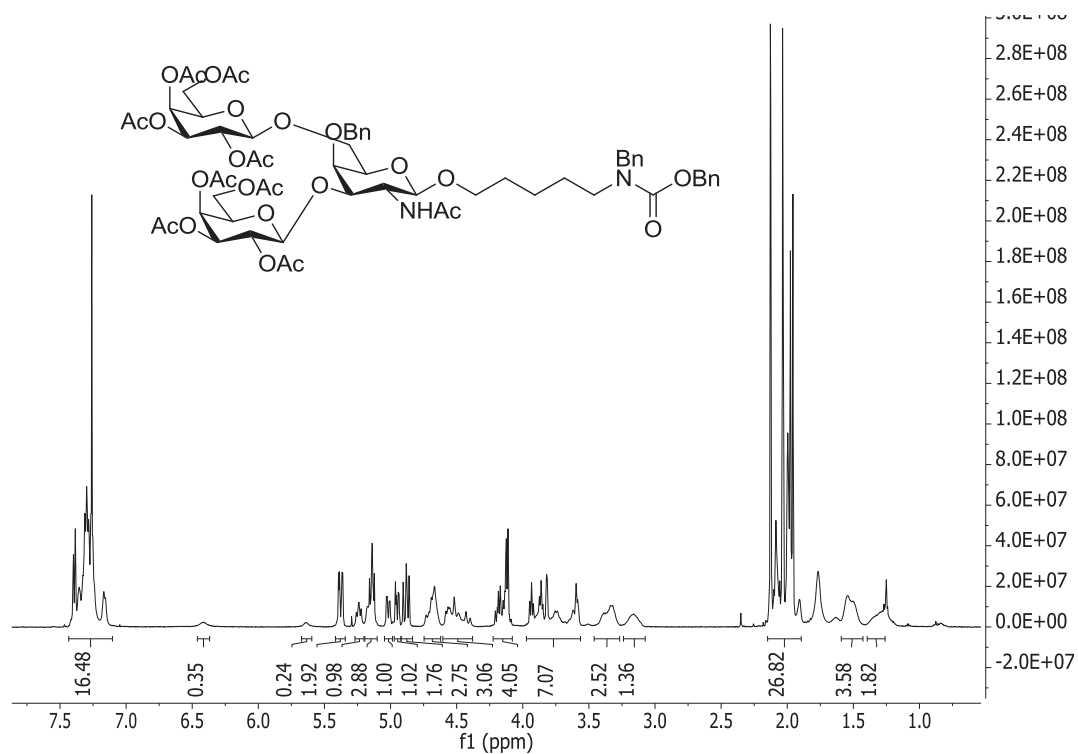


**5- (benzyl (benzyloxycarbonyl)amino) pentyl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[3,4,6-tri-O-acetyl-2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 30**

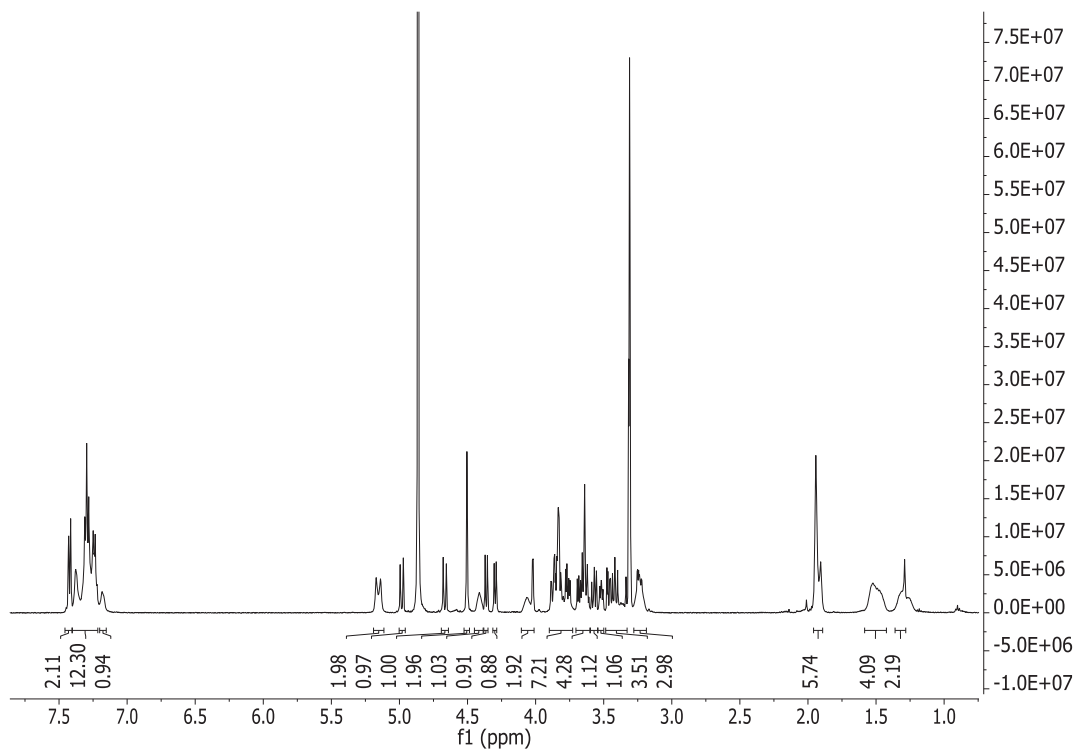
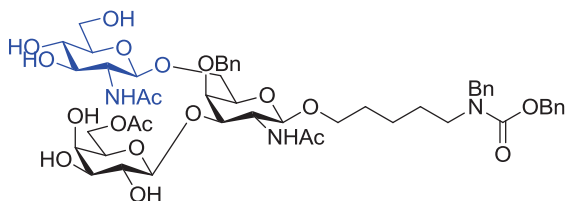


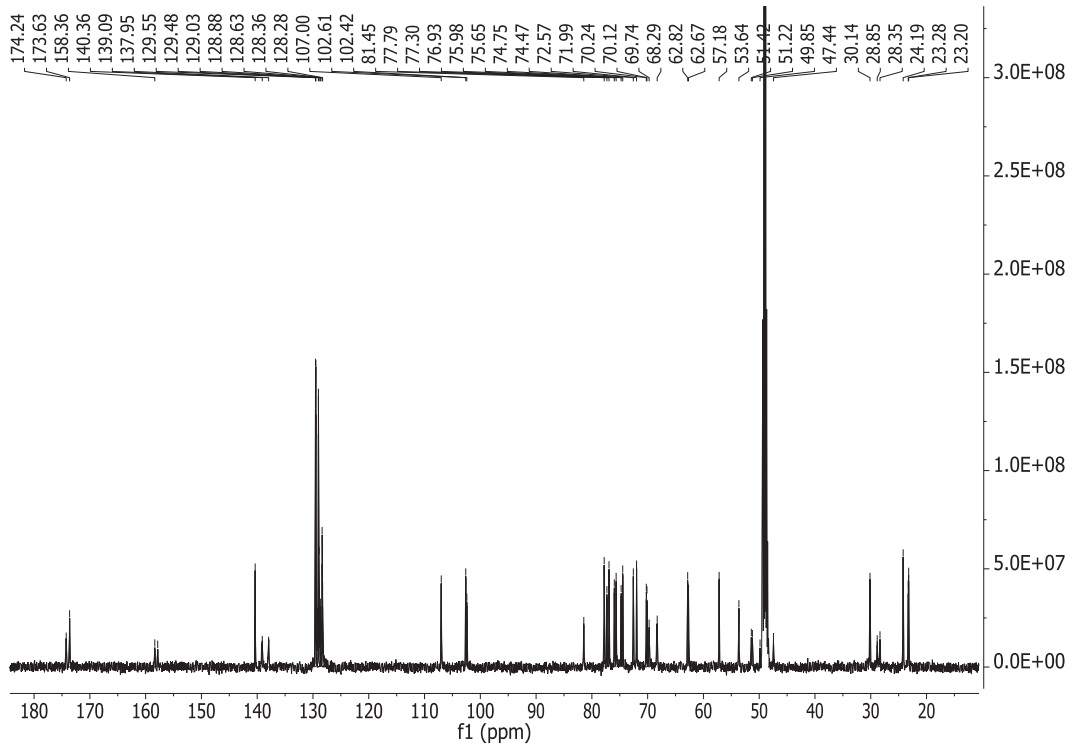


**5- (benzyl (benzyloxycarbonyl)amino) pentyl 2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl-(1→3)-[2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl-(1→6)]-4-O-benzyl-2-deoxy-2-acetamido-β-D-galactopyranoside 31**

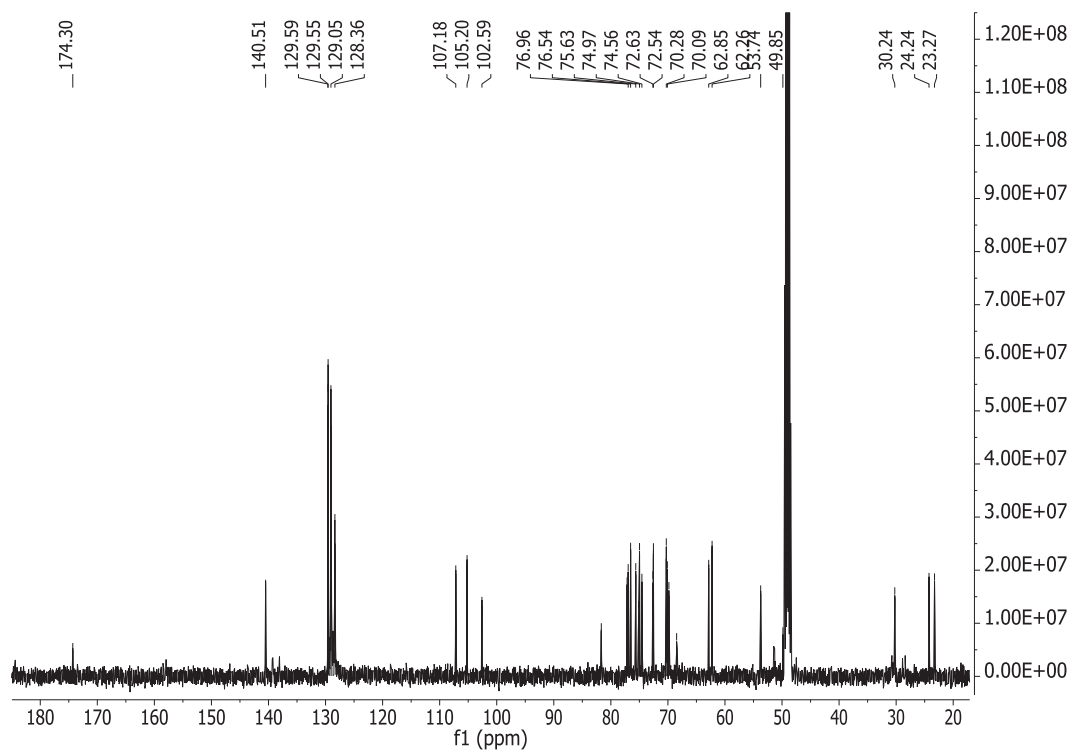
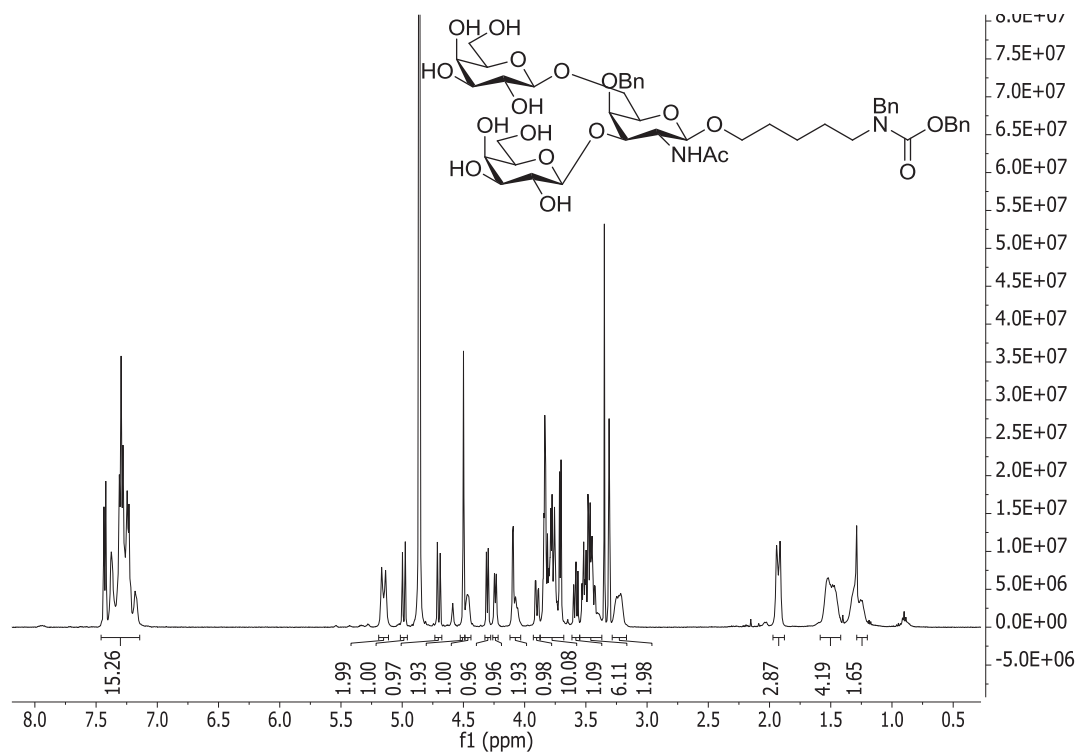


5- (benzyl (benzyloxycarbonyl)amino) pentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[ 2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 29

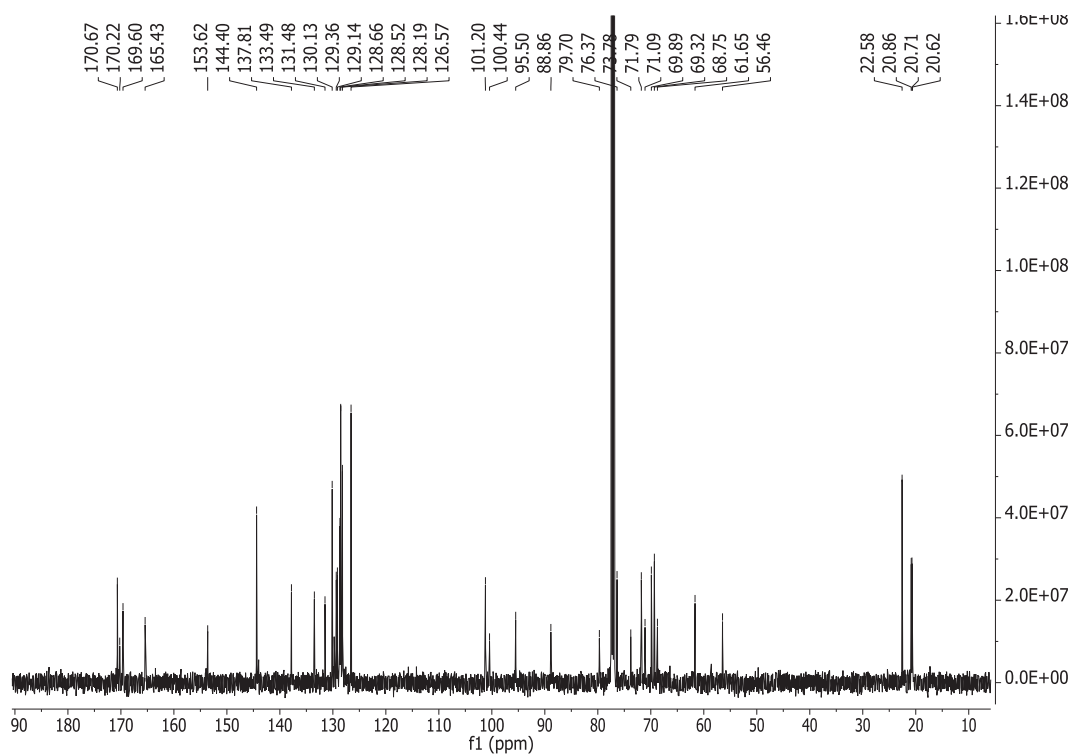
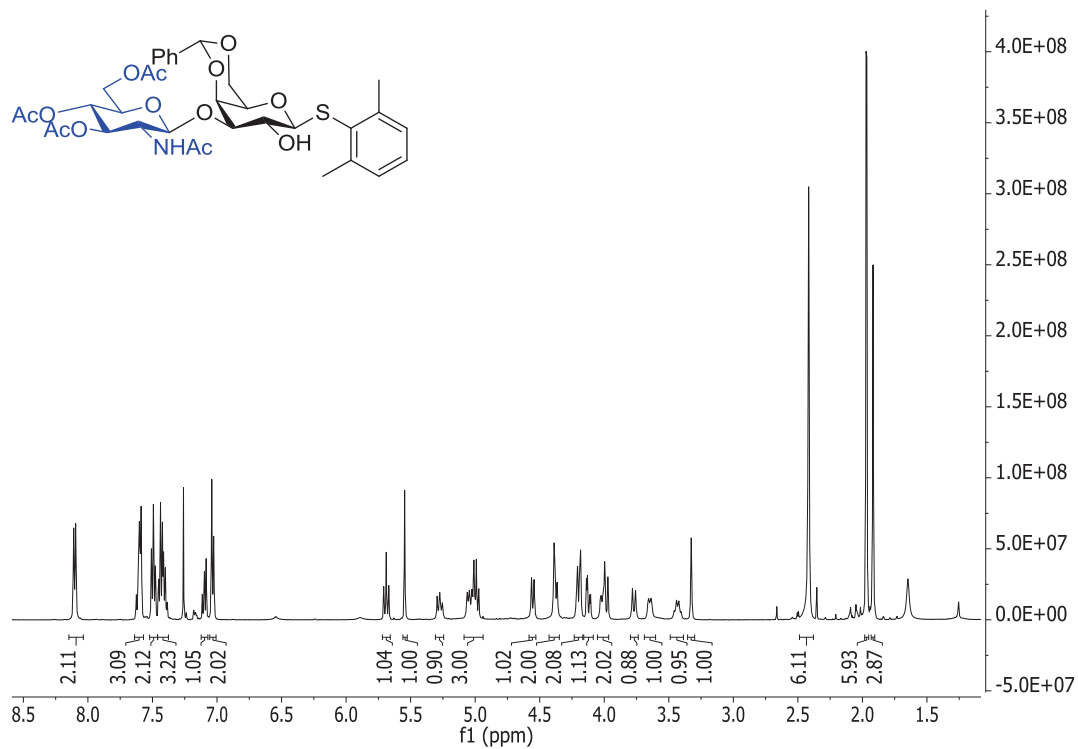




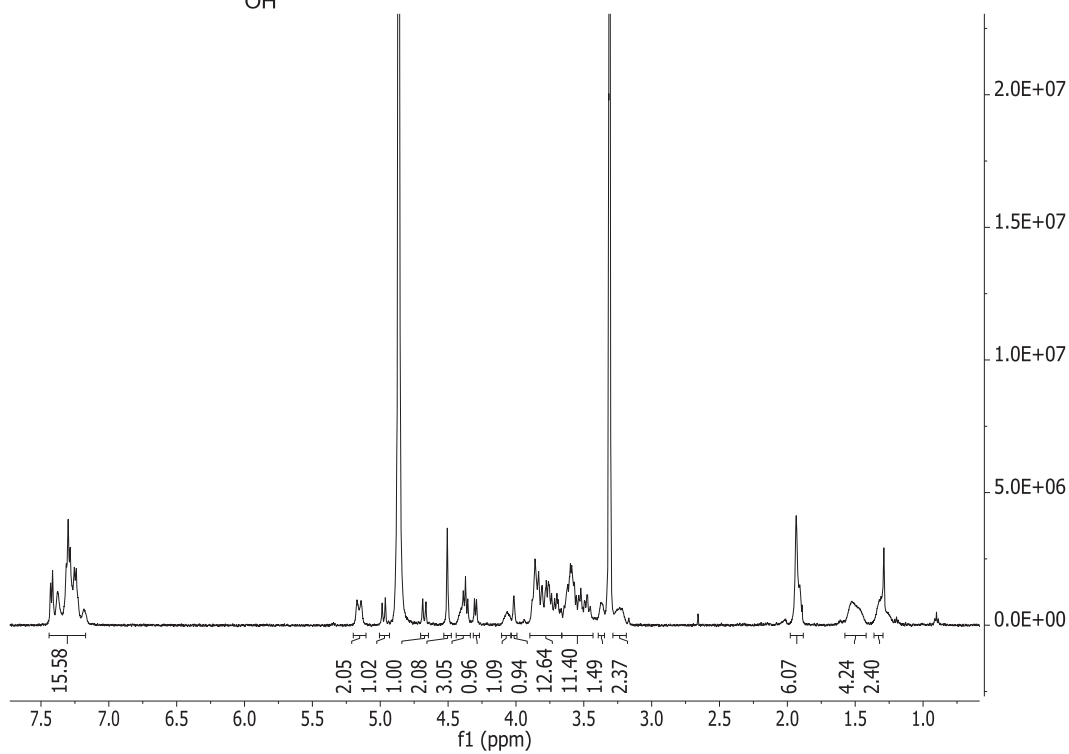
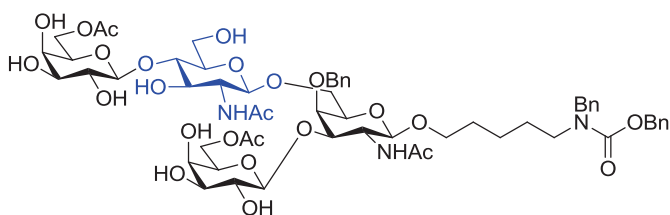
5- (benzyl (benzyloxycarbonyl)amino) pentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 32

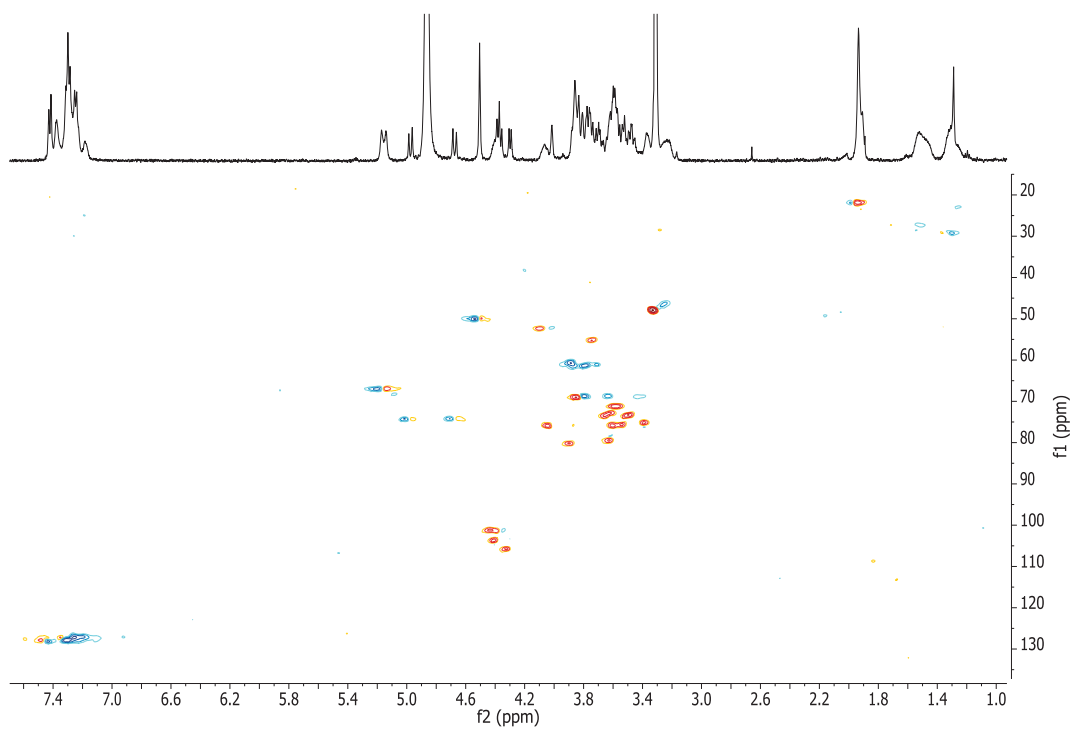
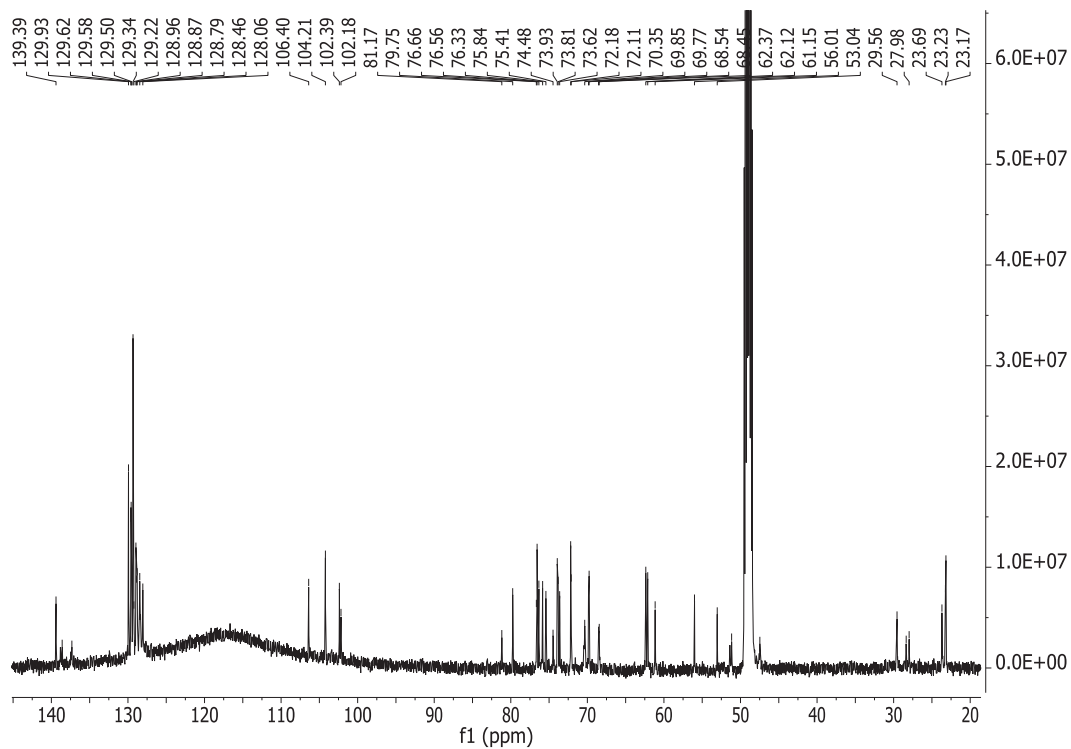


**2,6-Dimethylphenyl (3,4,6-tri-*O*-acetyl-2-deoxy-2-((2,2,2-trichloroethoxycarbonylamino)- $\beta$ -D-galactopyranosyl)-(1 $\rightarrow$ 3)- 2-*O*-benzoyl-4,6-*O*-benzylidene-1- thio- $\beta$ -D-galactopyranoside 46**

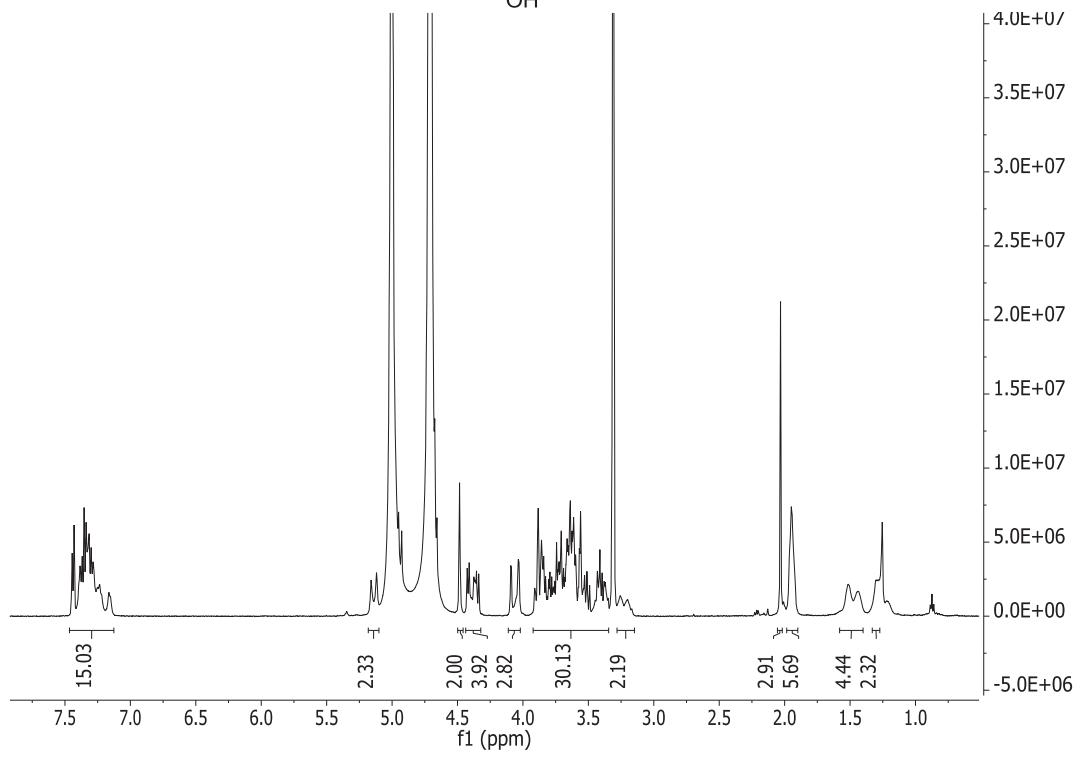
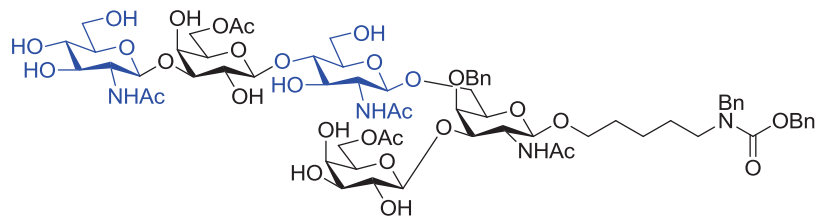


5- (benzyl (benzyloxycarbonyl)amino) pentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[ $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)- 2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 48

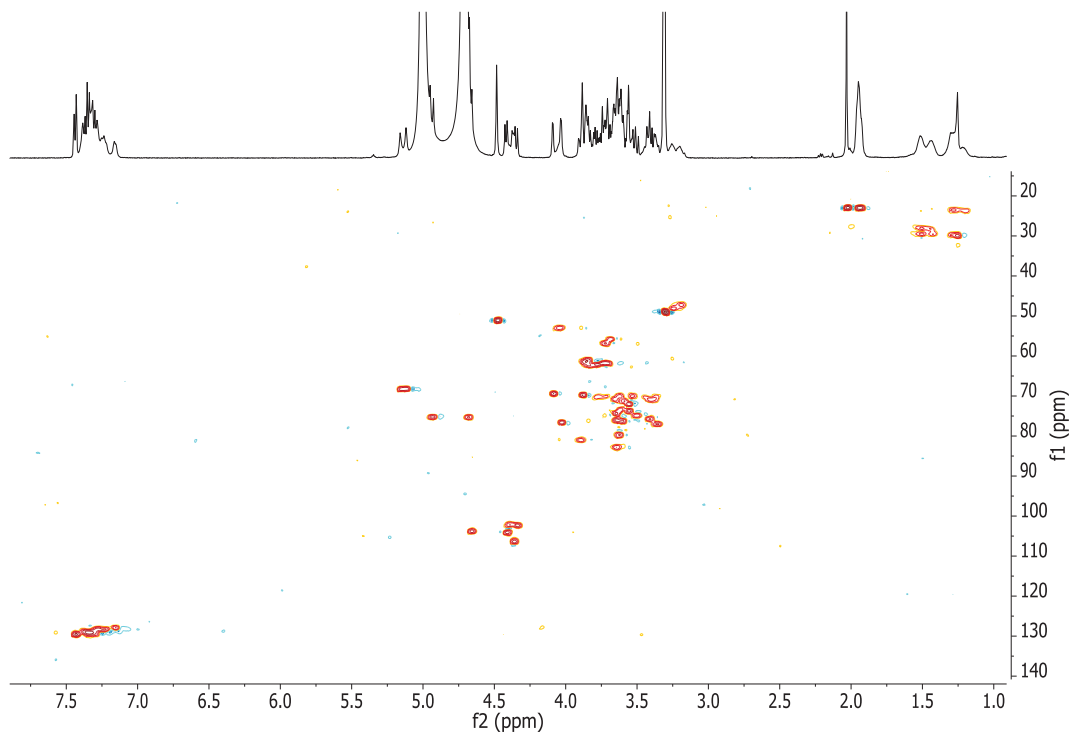
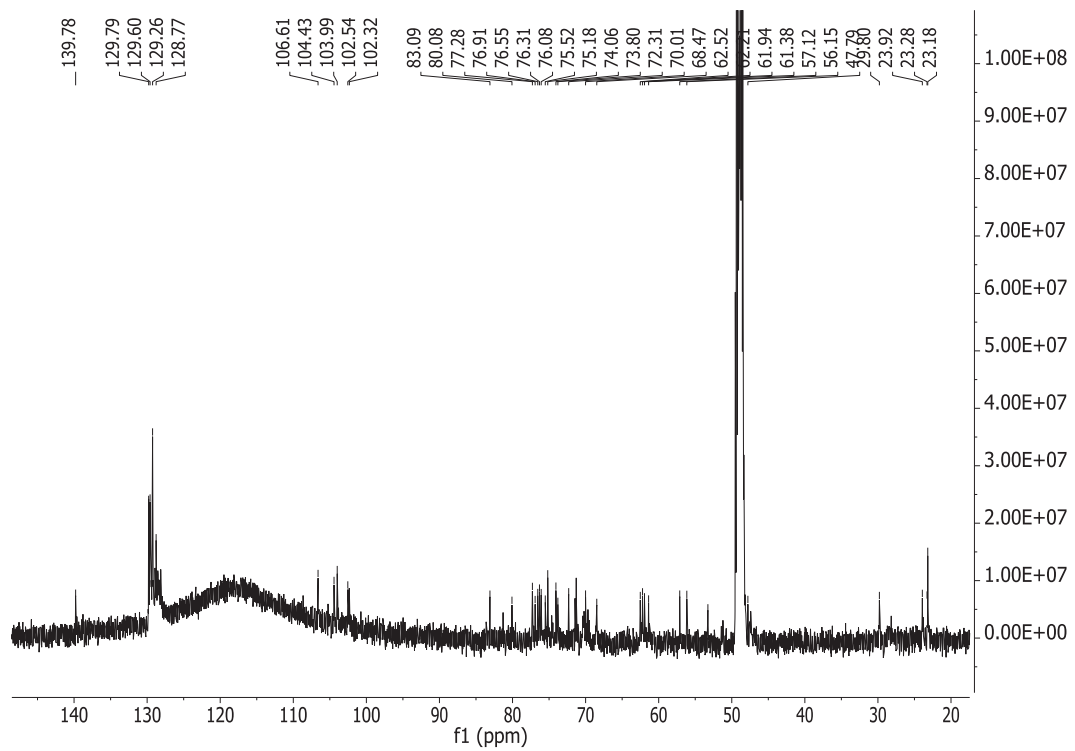




**5- (benzyl (benzyloxycarbonyl)amino) pentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)]- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)- 2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 49**

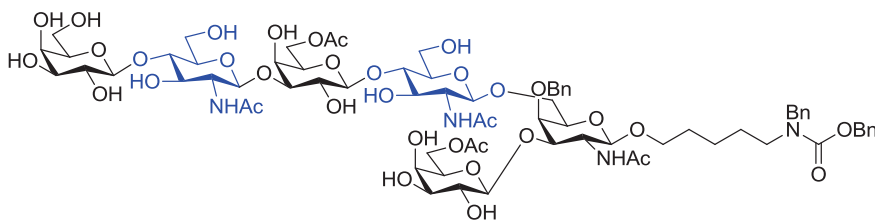




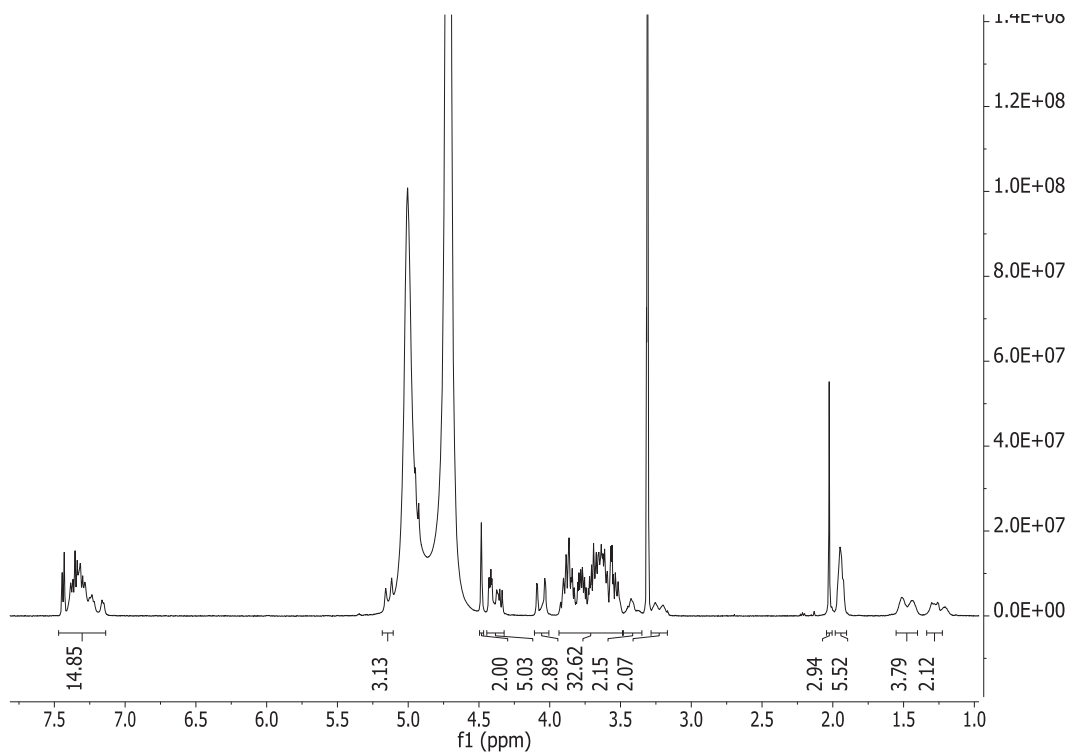


5- (benzyl (benzyloxycarbonyl)amino) pentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[ $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)-2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)- $\beta$ -D-

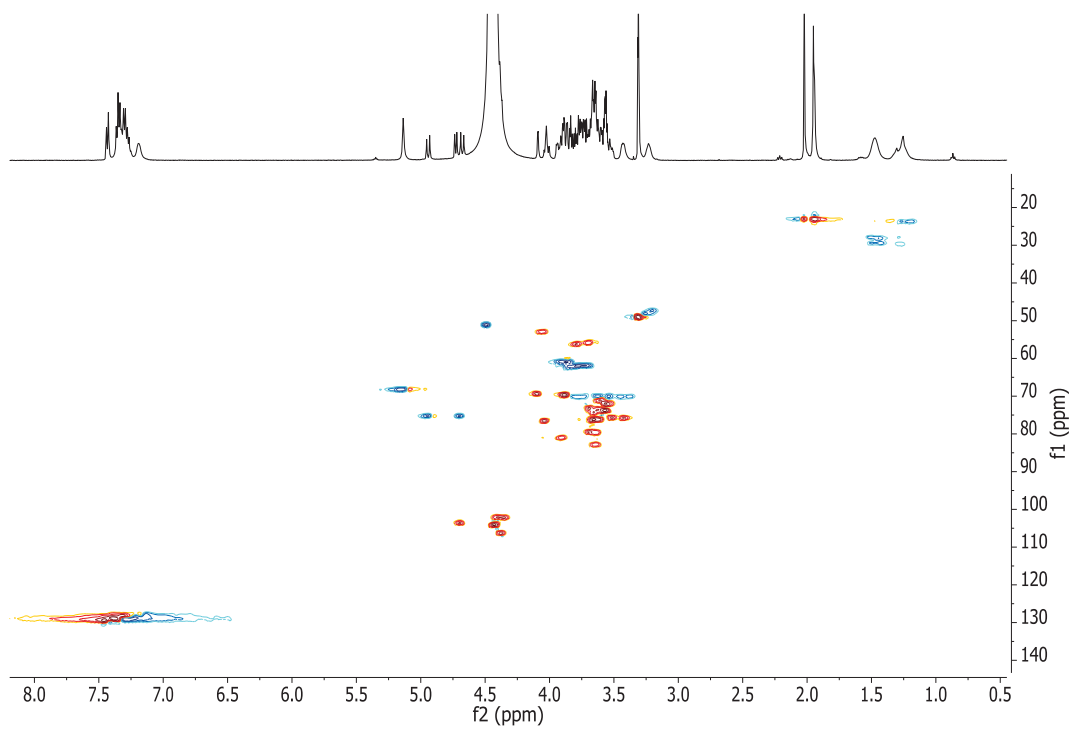
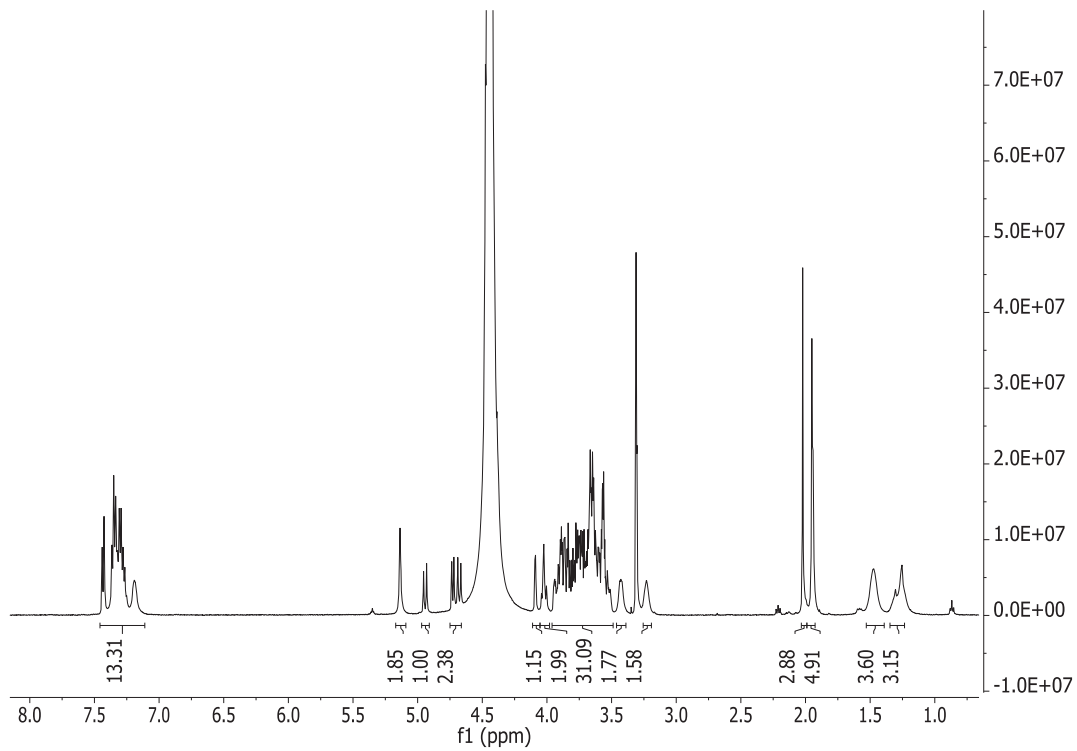
galactopyranosyl-(1→4)- 2-deoxy-2-acetamido-β-D-glucopyranosyl-(1→6)]-4-O-benzyl-2-deoxy-2-acetamido-β-D-galactopyranoside 50

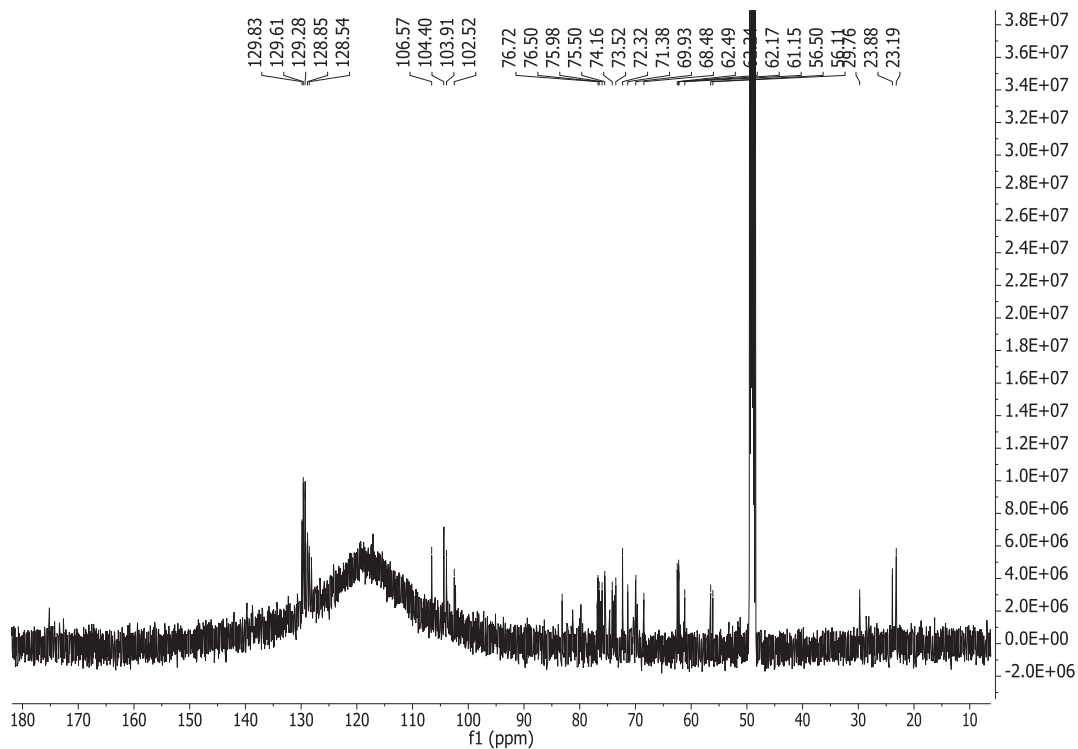


273K NMR

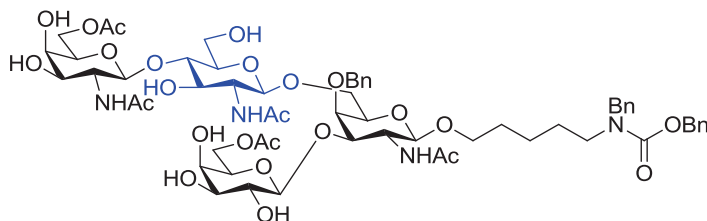


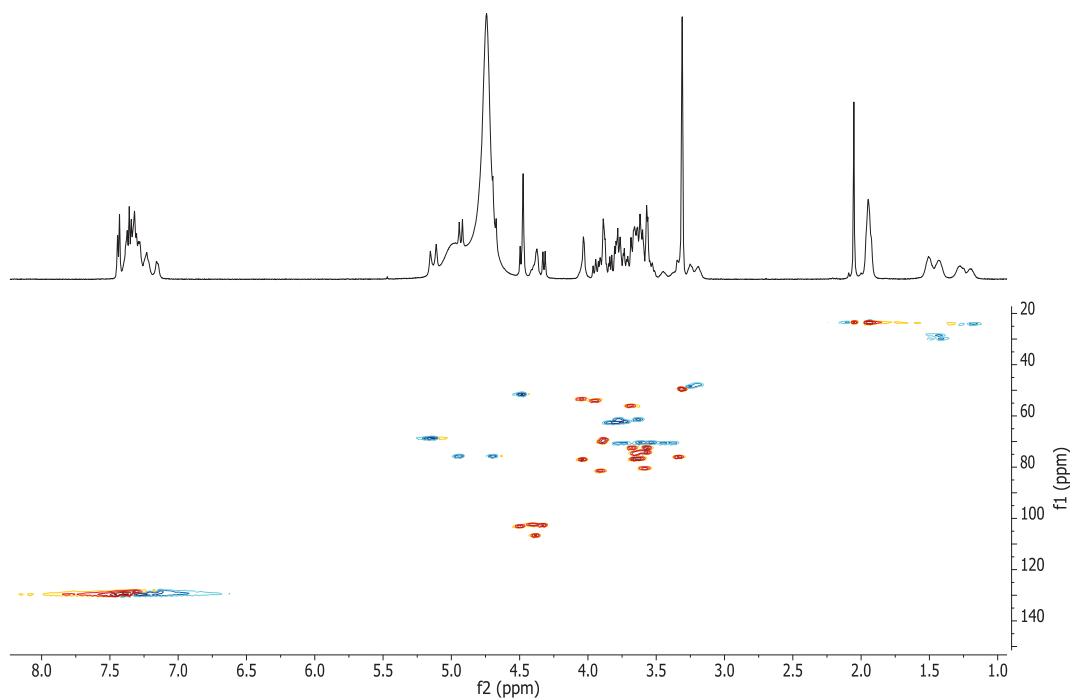
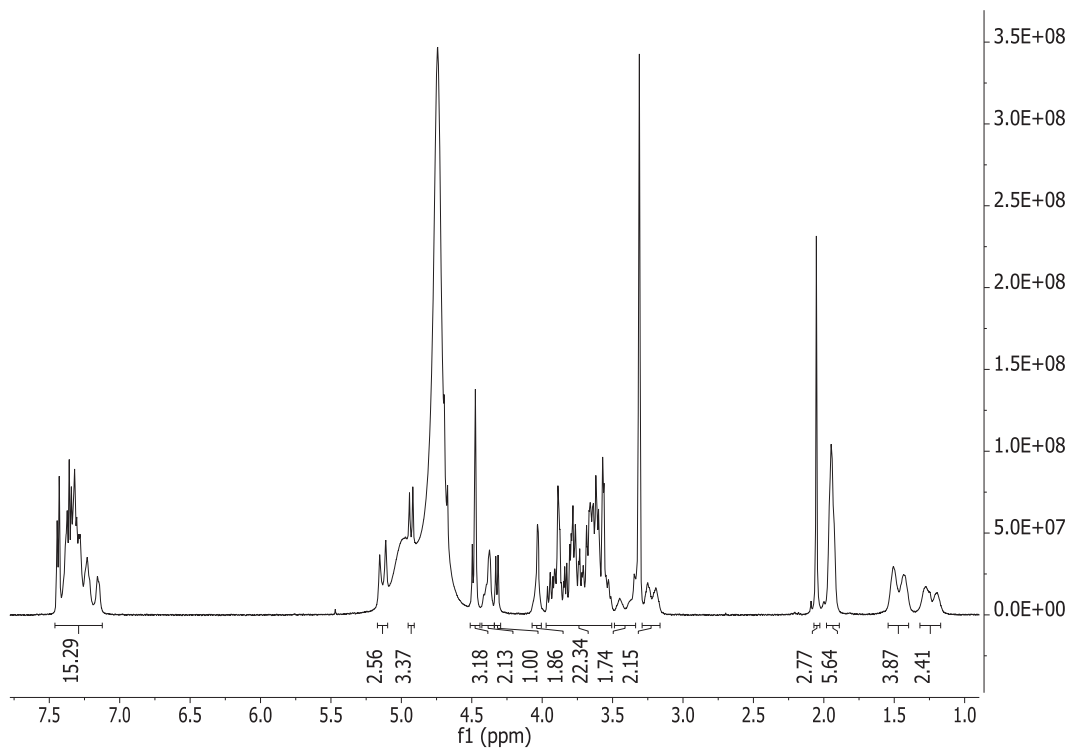
VT NMR 323K



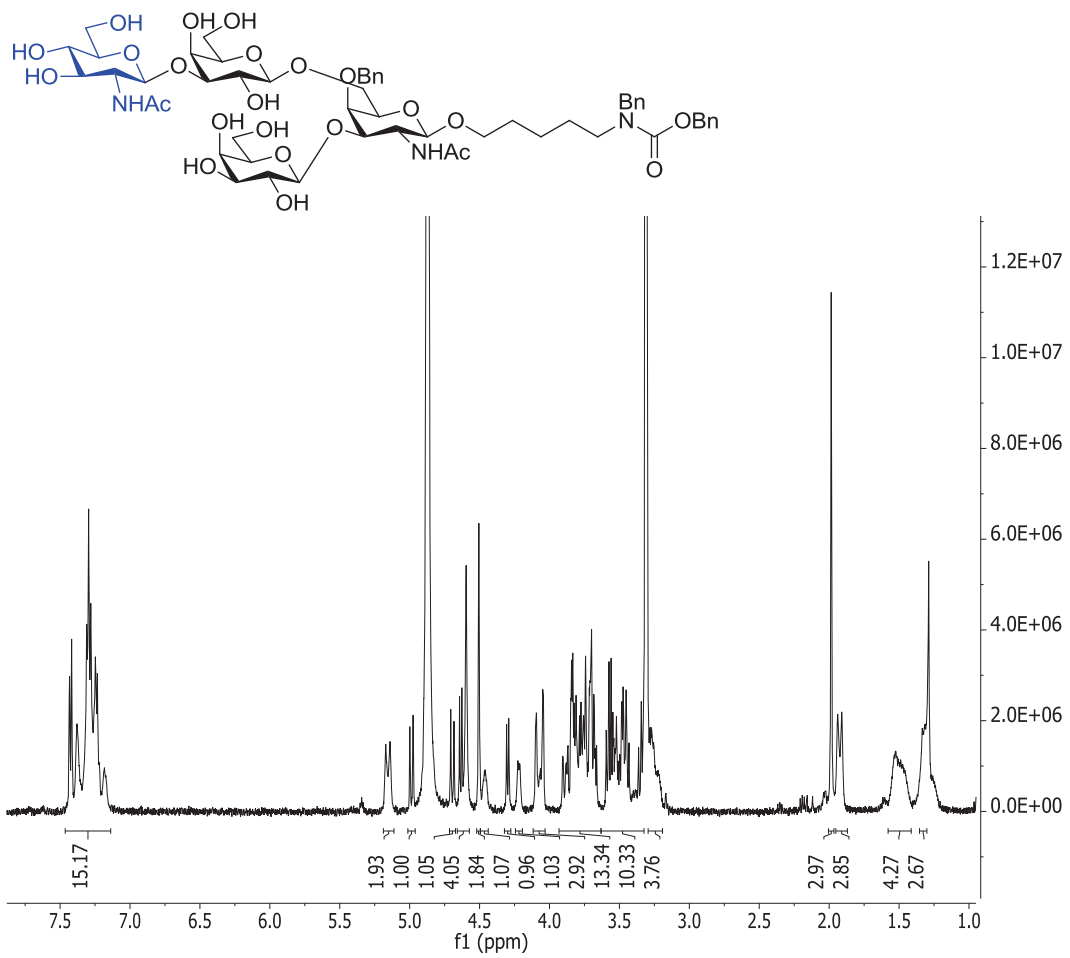


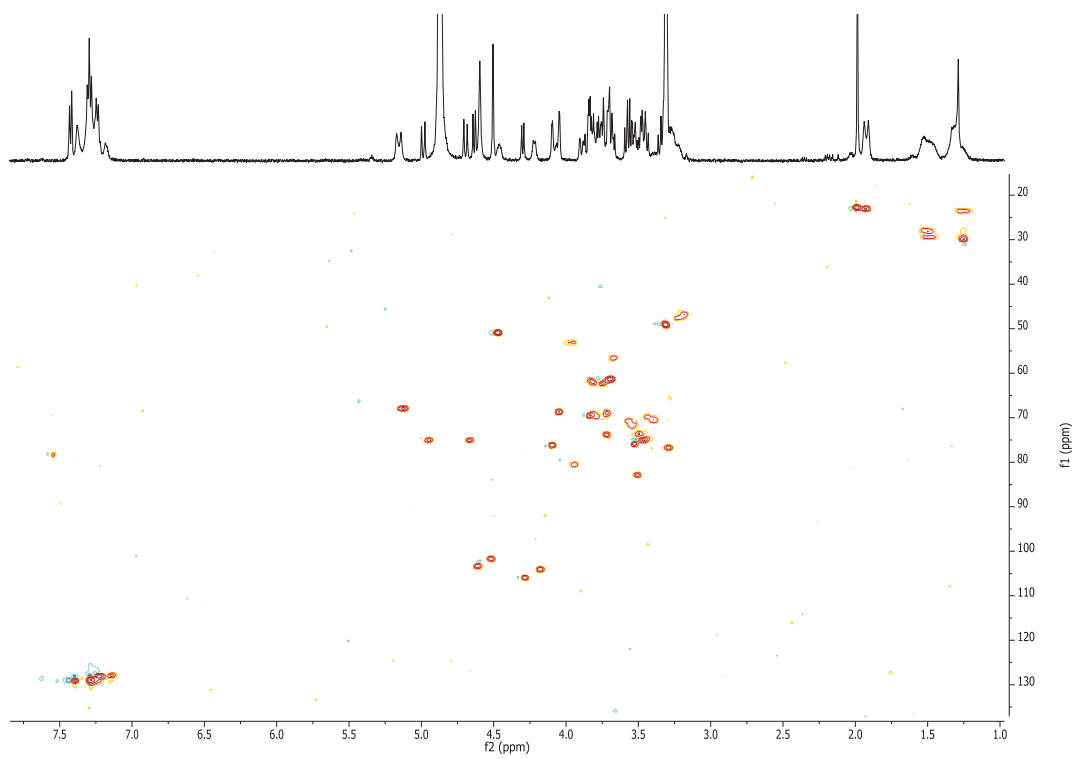
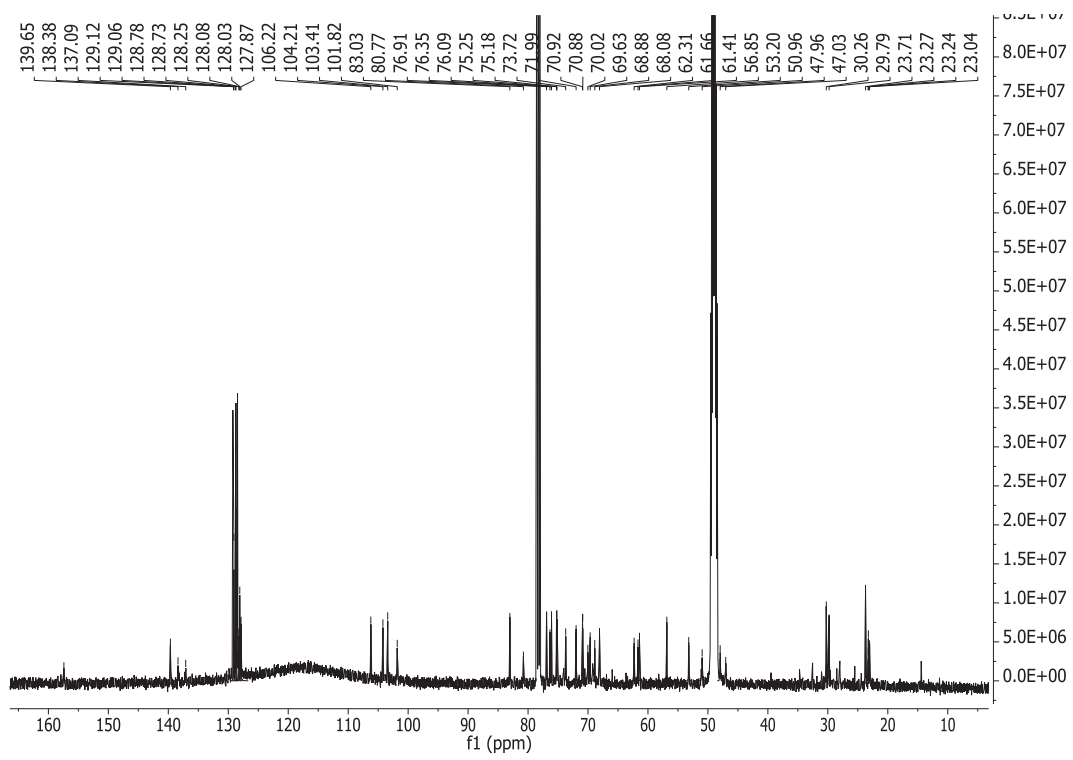
**5- (benzyl (benzyloxycarbonyl)amino) pentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[2-deoxy-2-acetamido- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)- 2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 51**



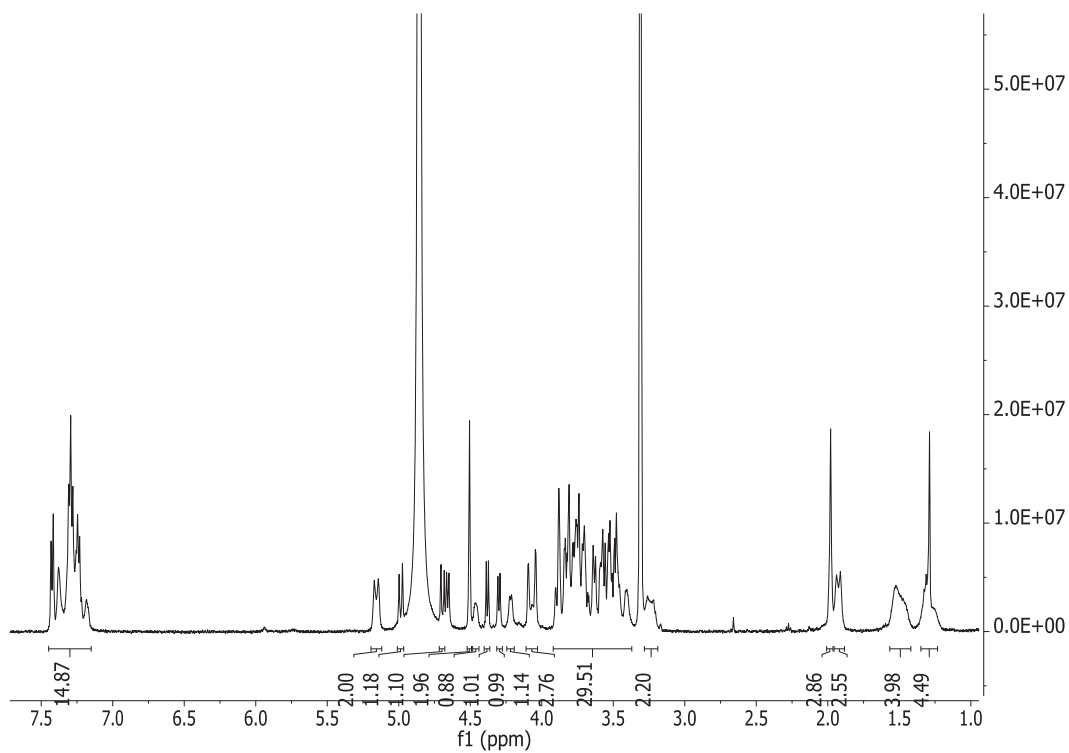
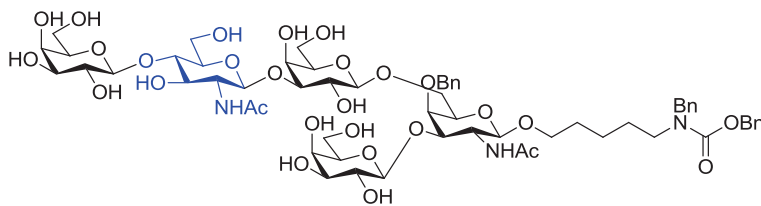


**5- (benzyl (benzyloxycarbonyl)amino) pentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[ 2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 34**

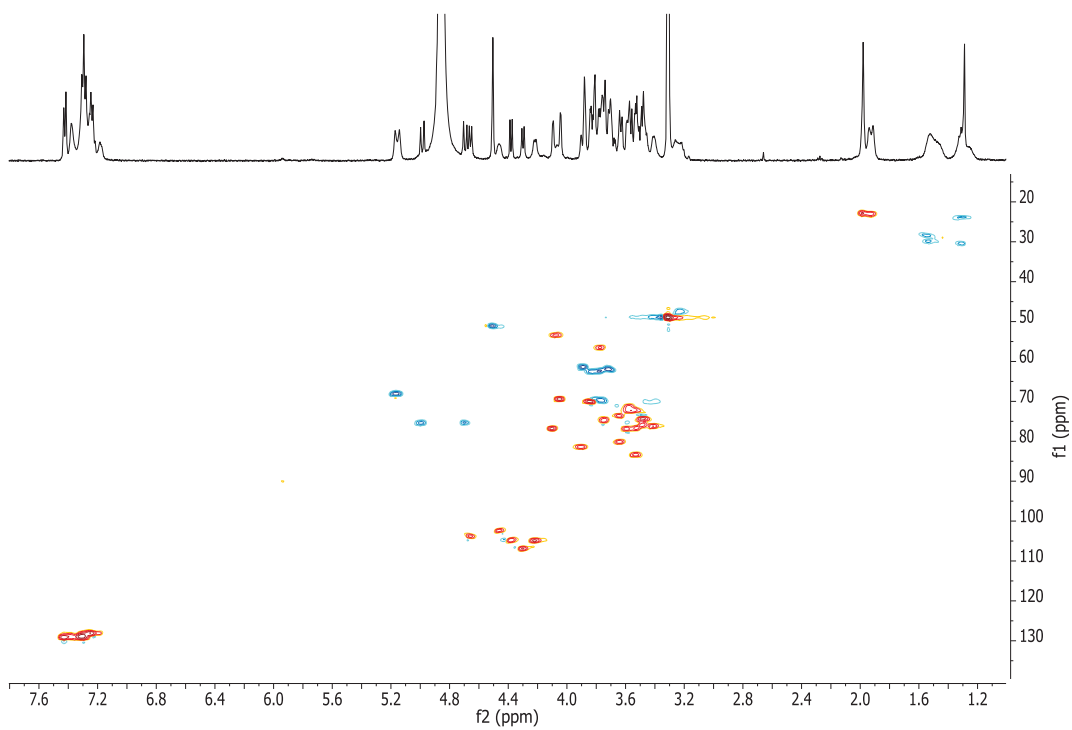
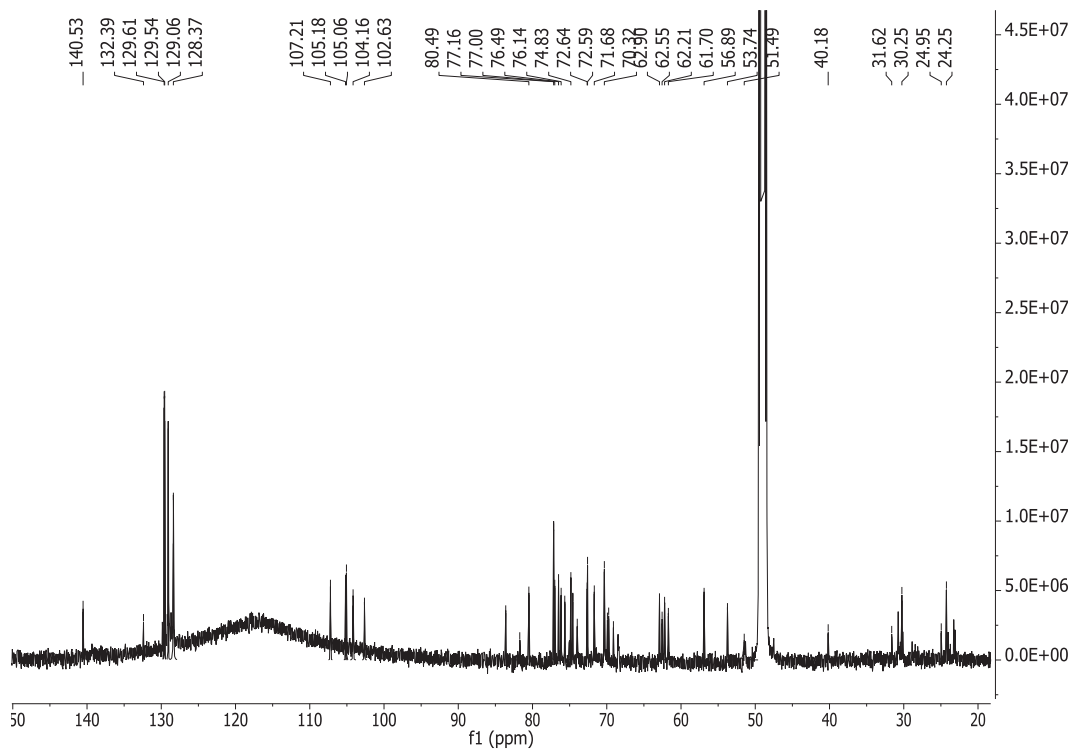




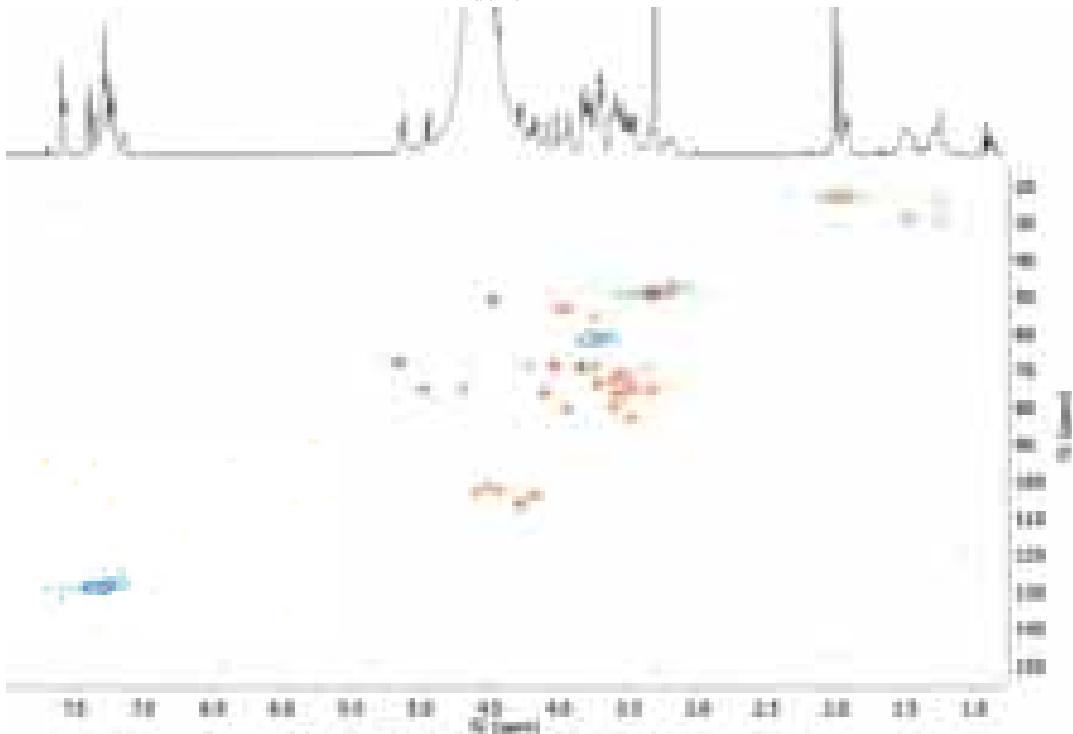
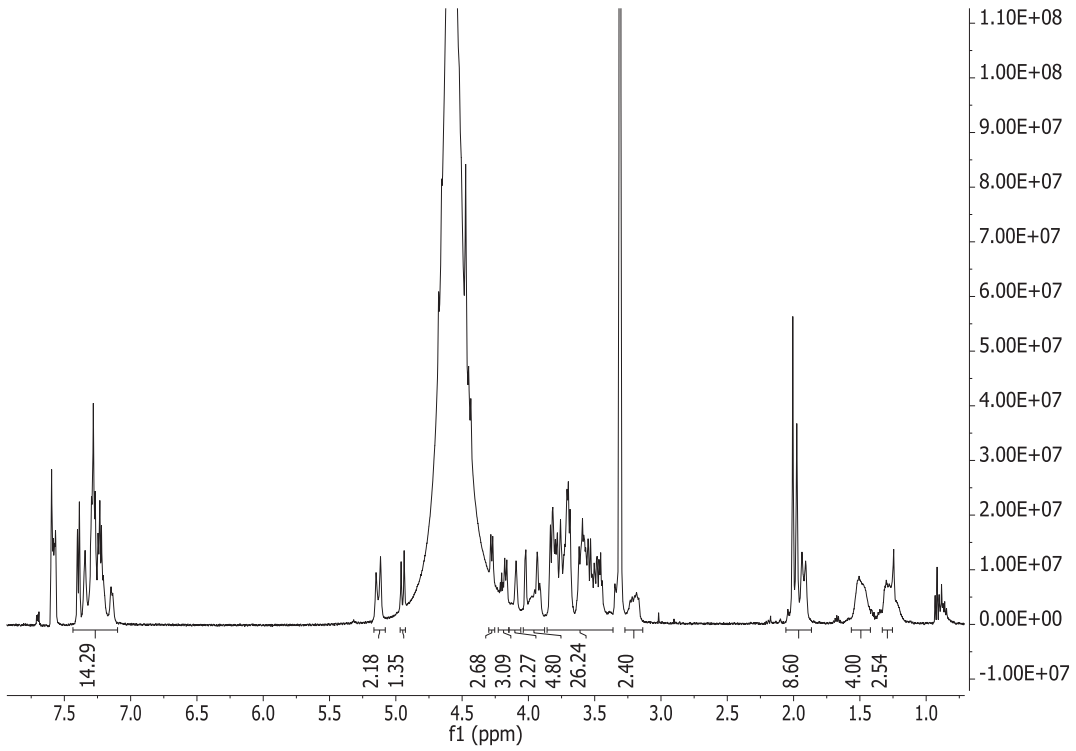
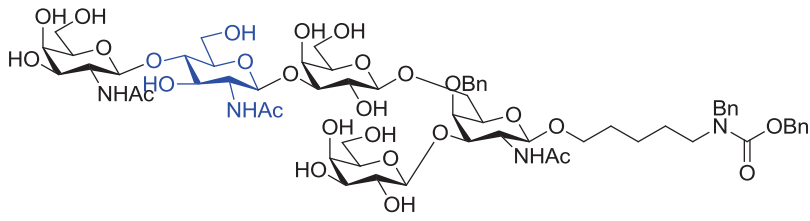
5- (benzyl (benzyloxycarbonyl)amino) pentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)-2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 52



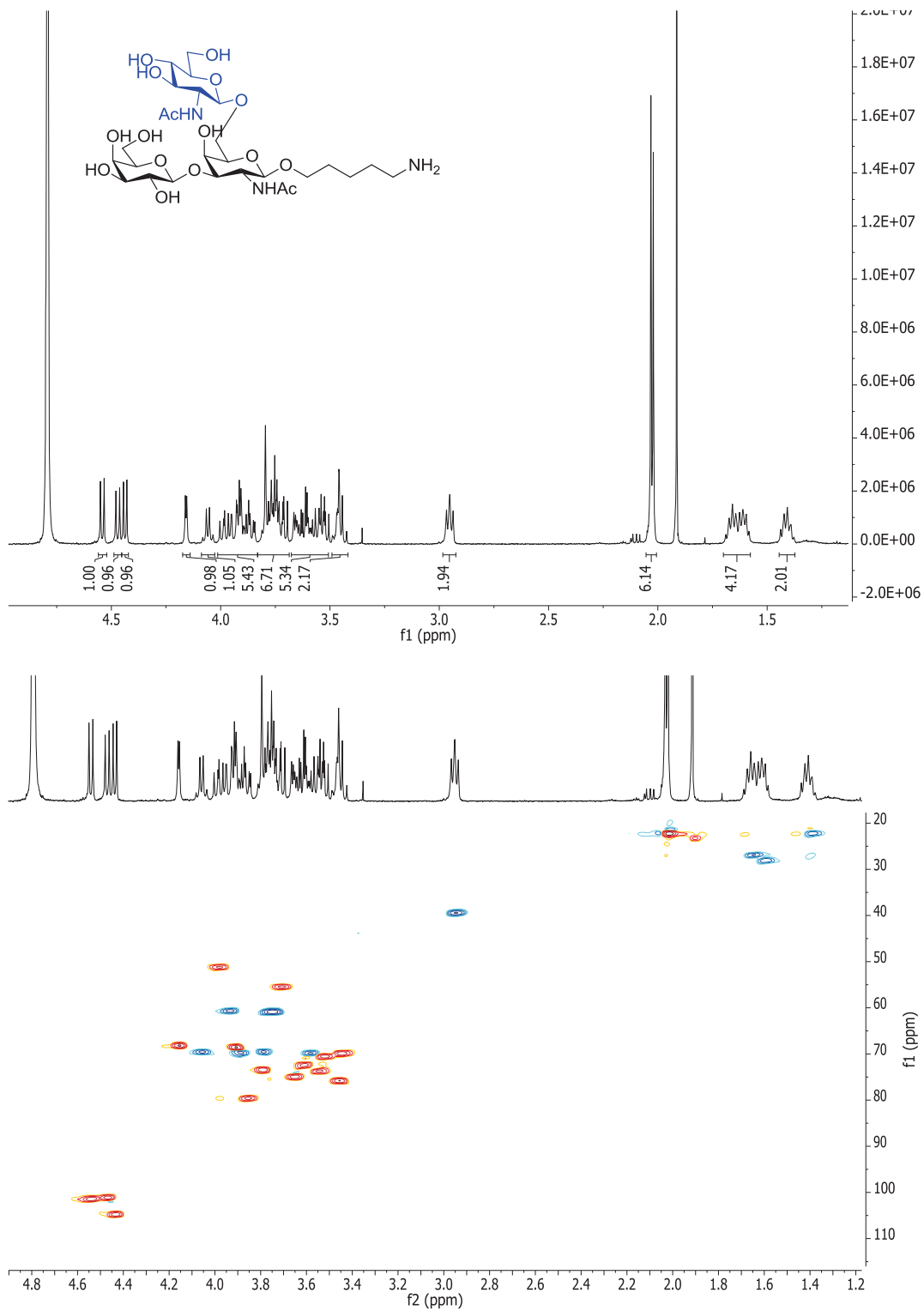


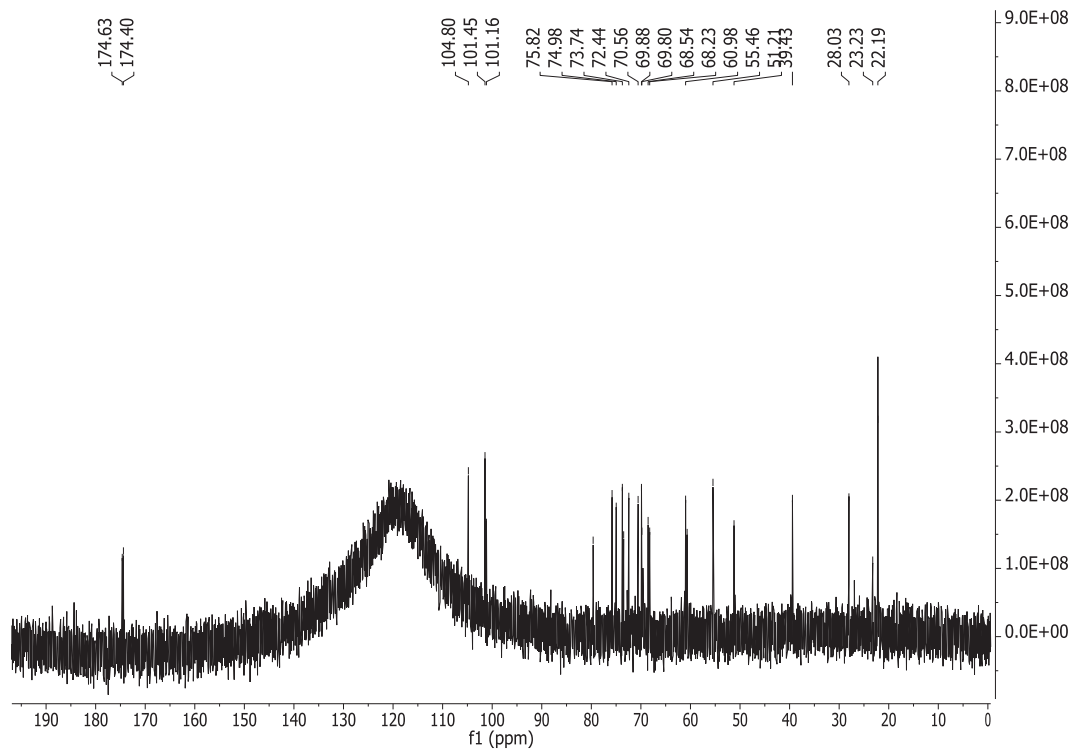


**5- (benzyl (benzyloxycarbonyl)amino) pentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[ 2-deoxy-2-acetamido- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)-2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-4-O-benzyl-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside 53**

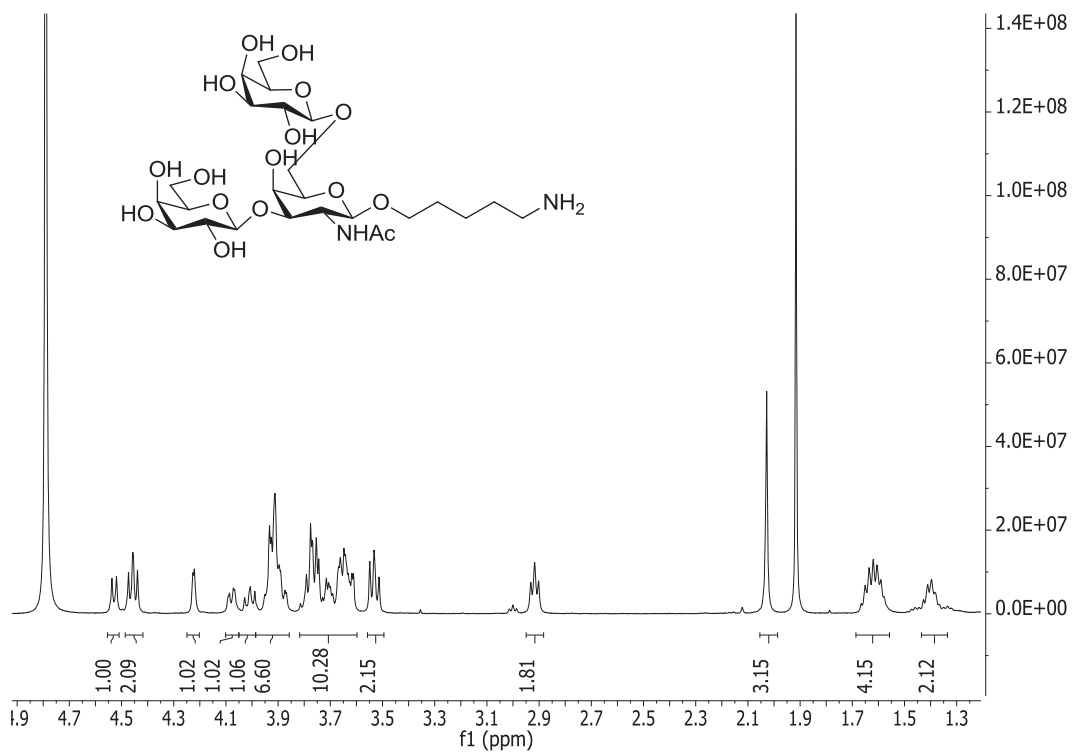


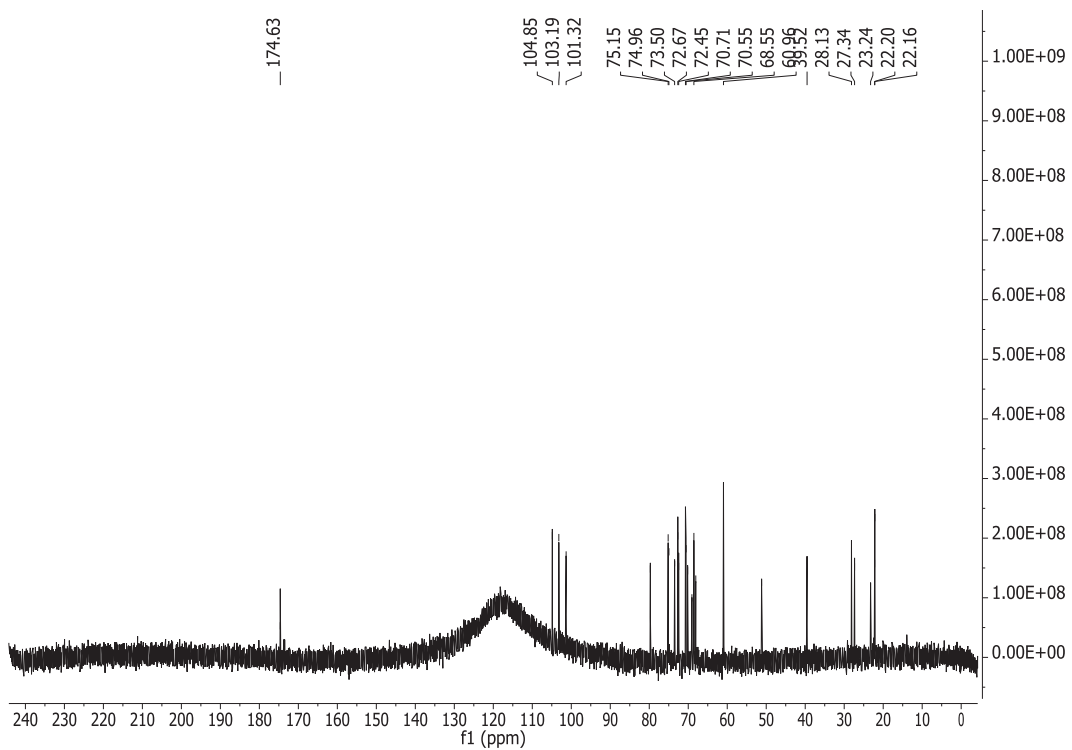
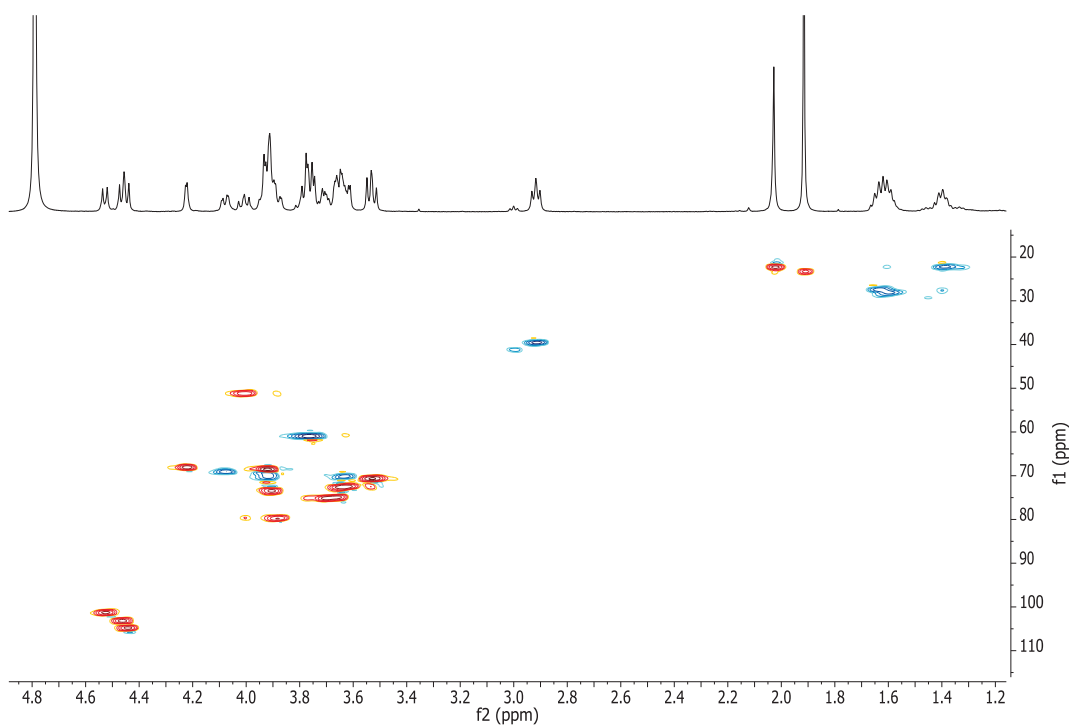
5-aminopentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[ 2-deoxy-2- acetamido - $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6) ]-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside O1



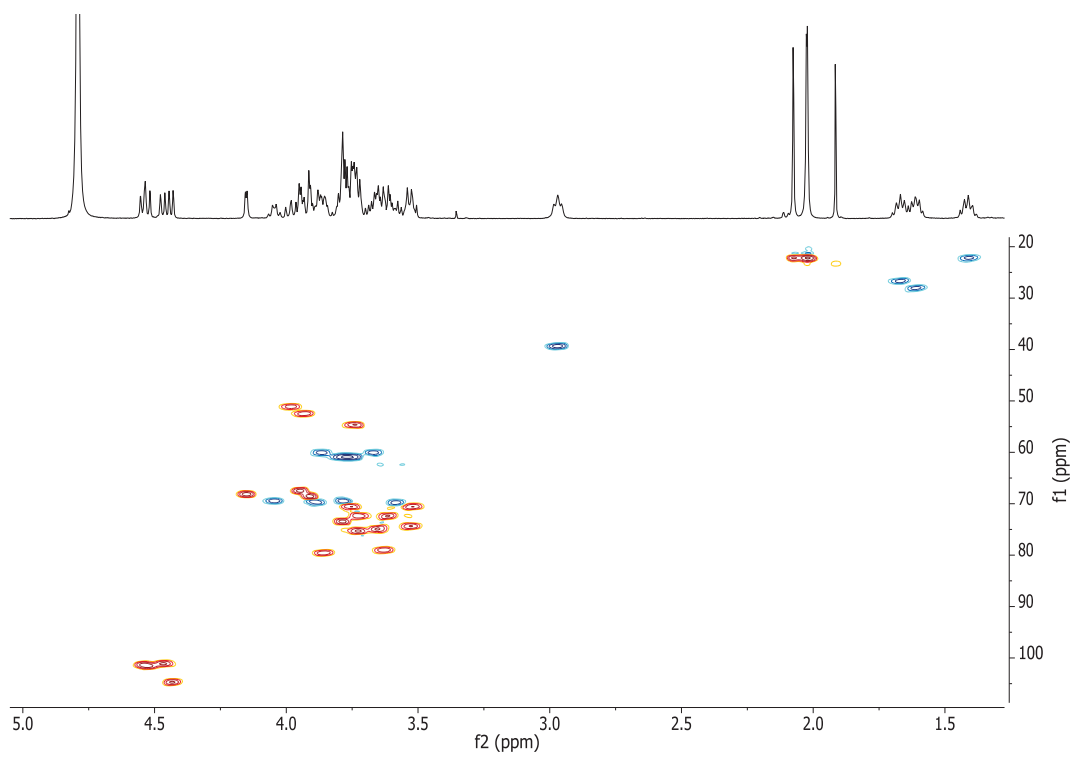
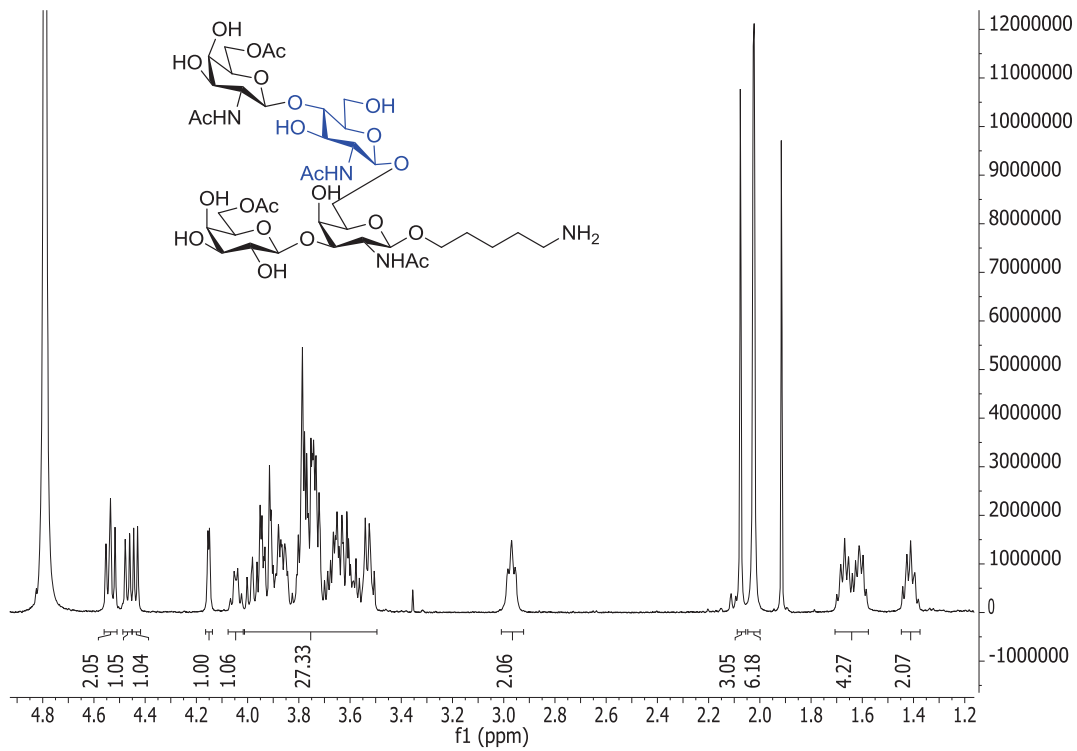


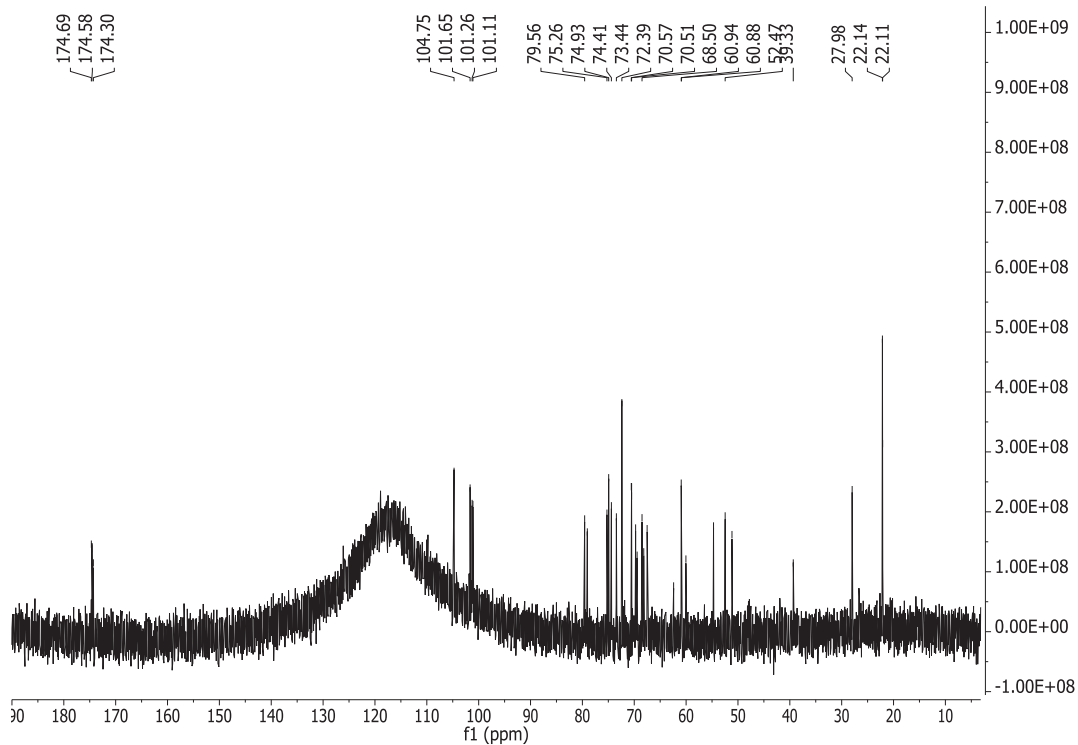
**5-aminopentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside O2**



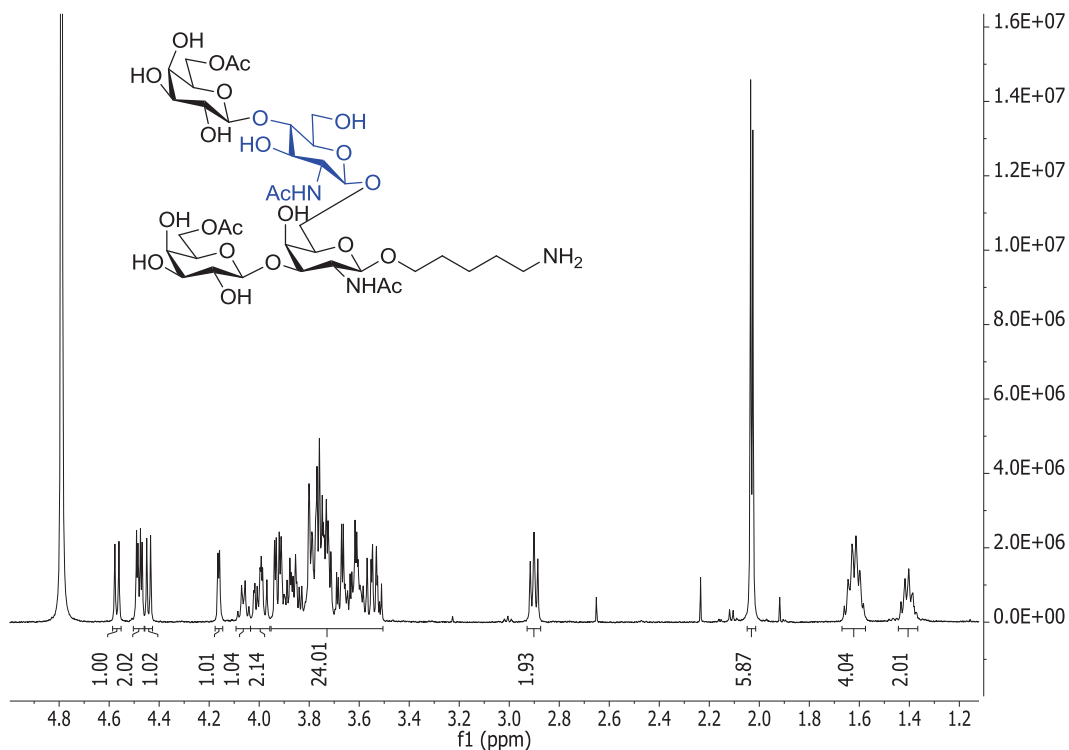


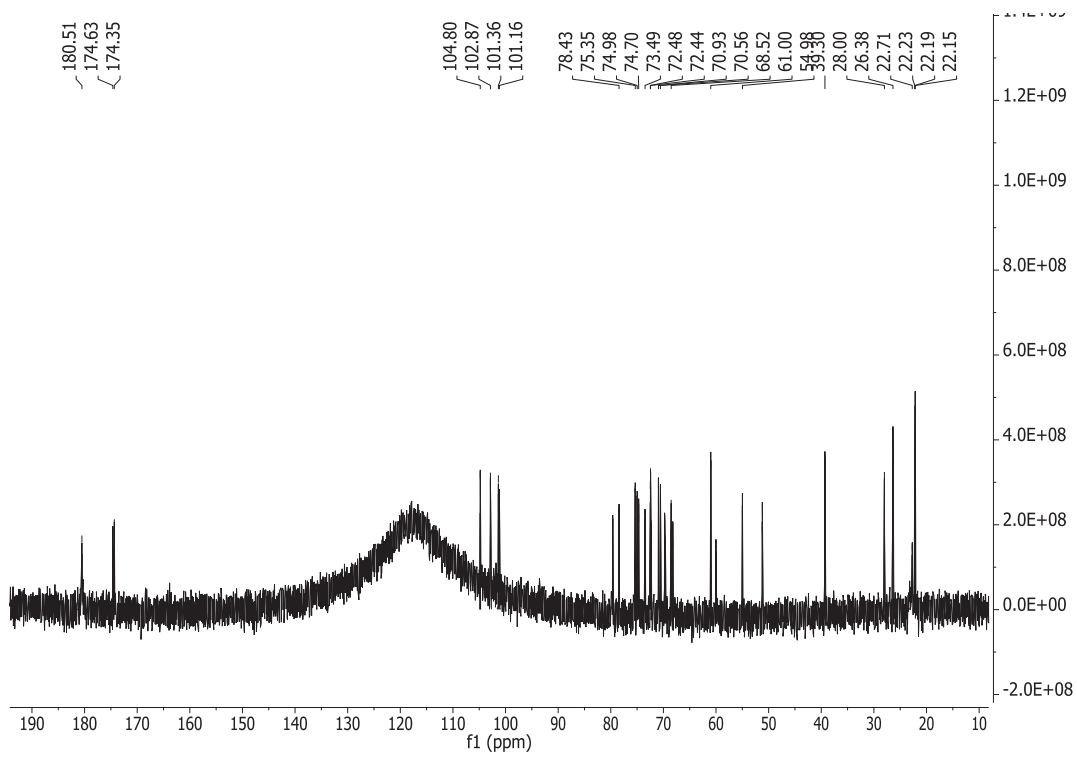
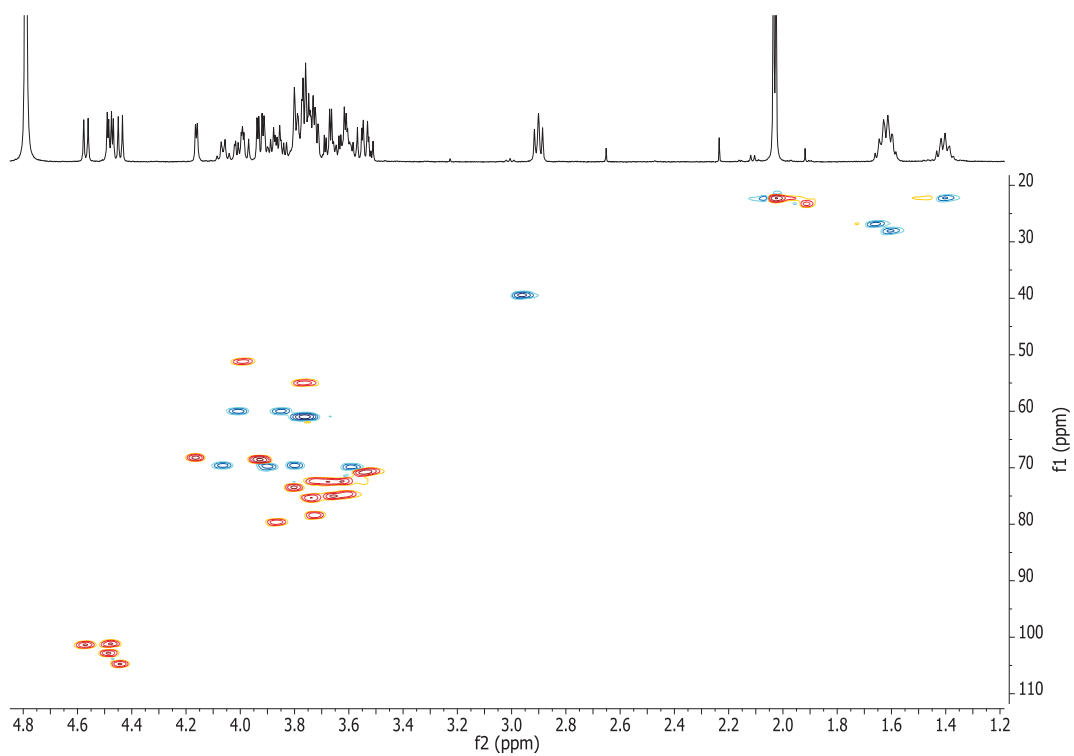
5-aminopentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[2-deoxy-2-acetamido- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)-2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6)]-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside O3





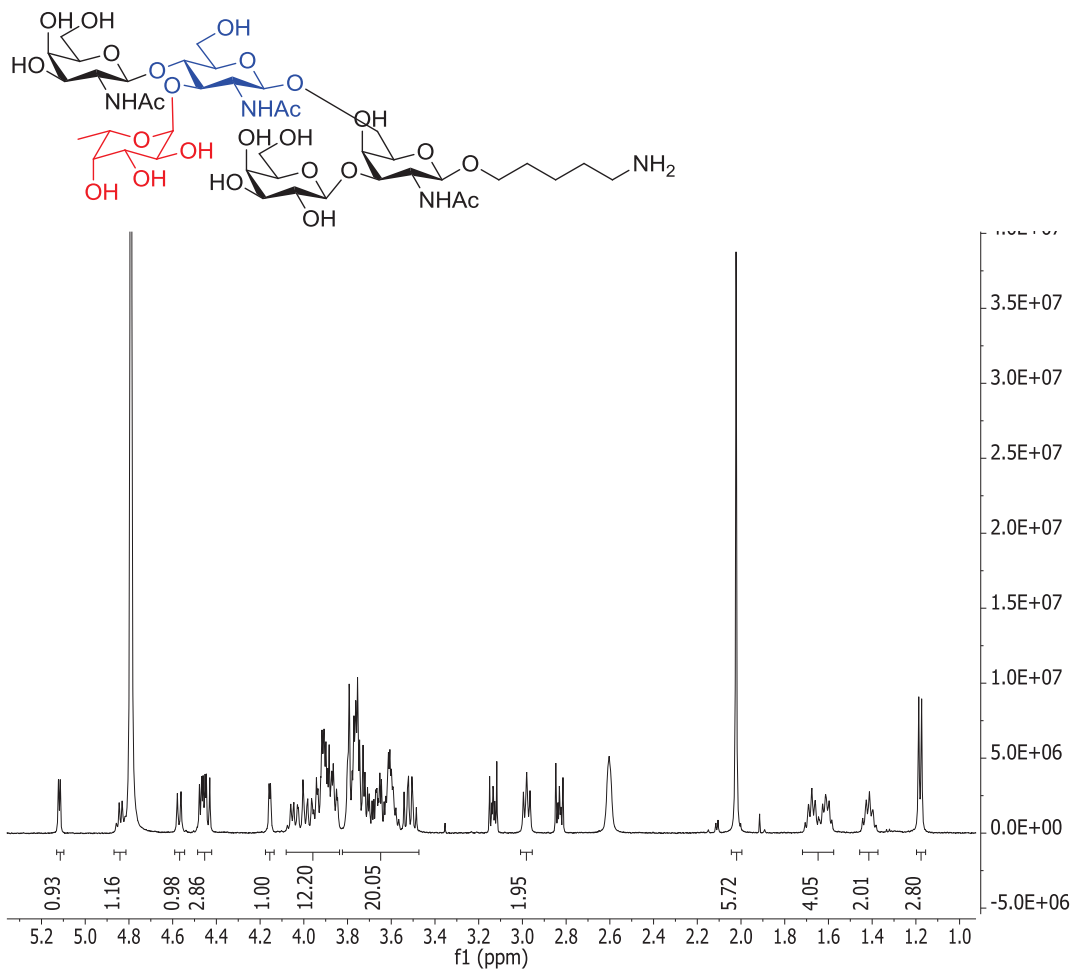
**5-aminopentyl β-D-galactopyranosyl-(1→3)-[β-D-galactopyranosyl-(1→4)-2-deoxy-2-acetamido-β-D-glucopyranosyl-(1→6)]-2-deoxy-2-acetamido-β-D-galactopyranoside O4**

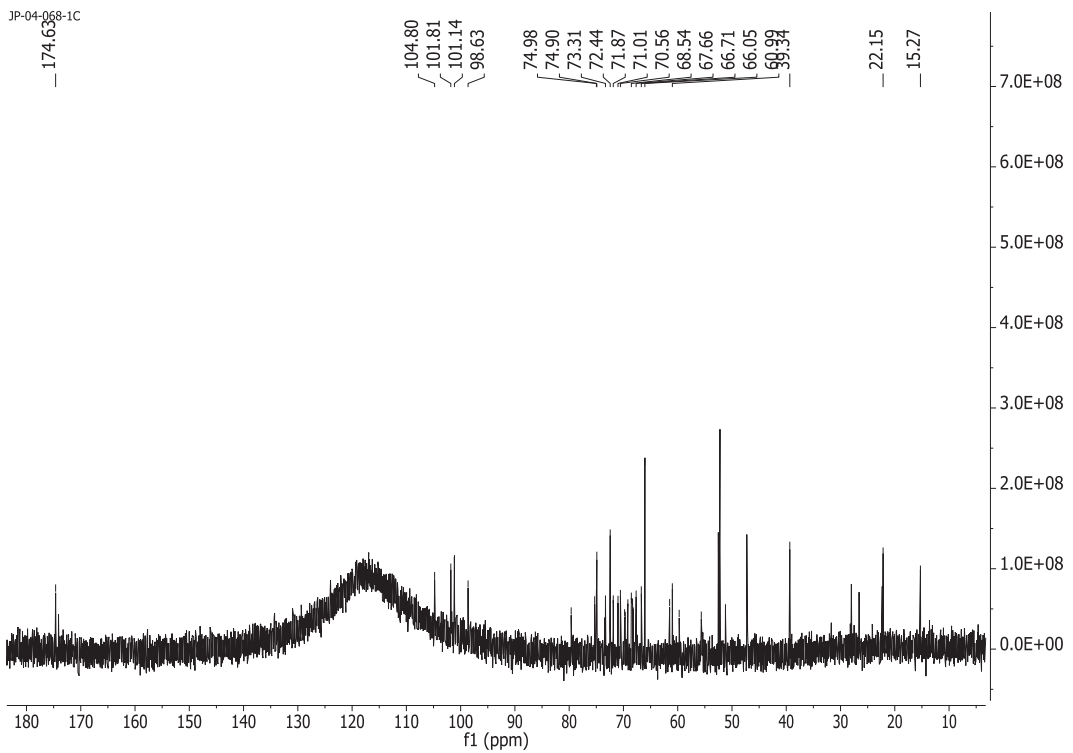
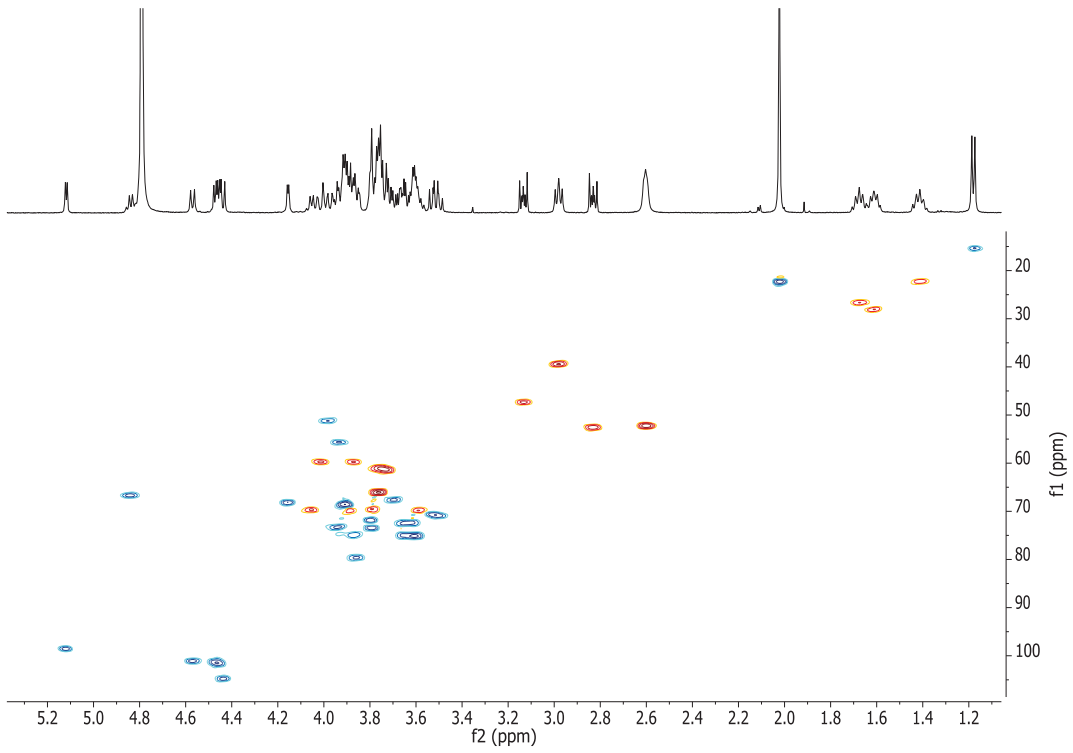




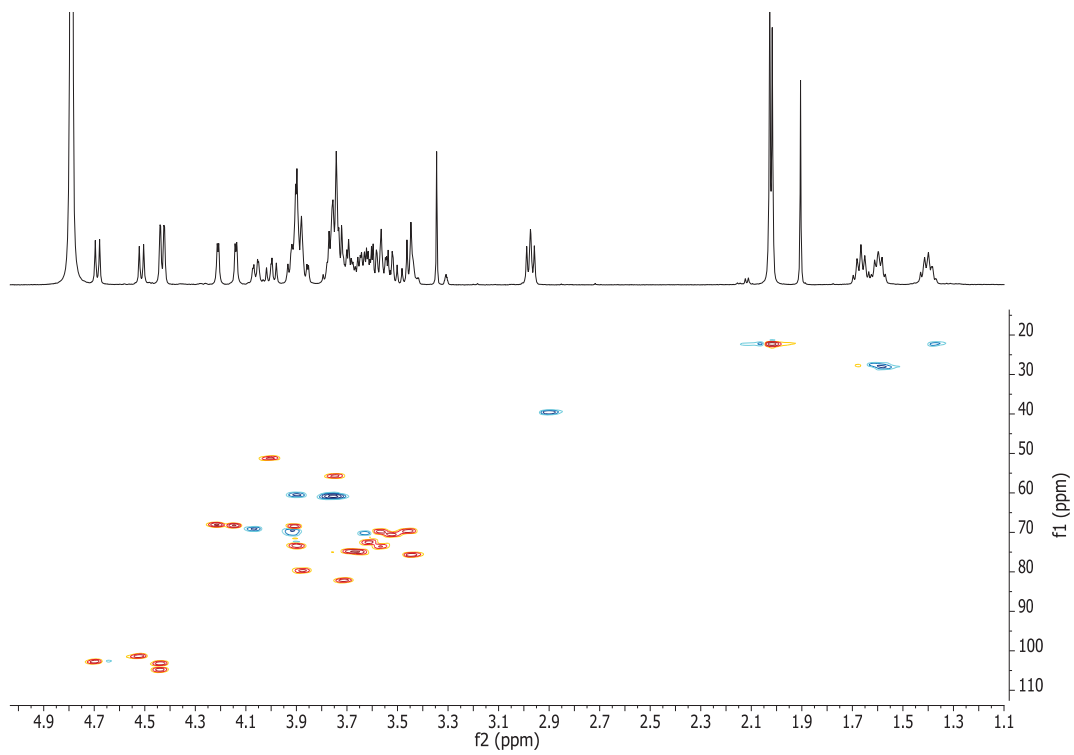
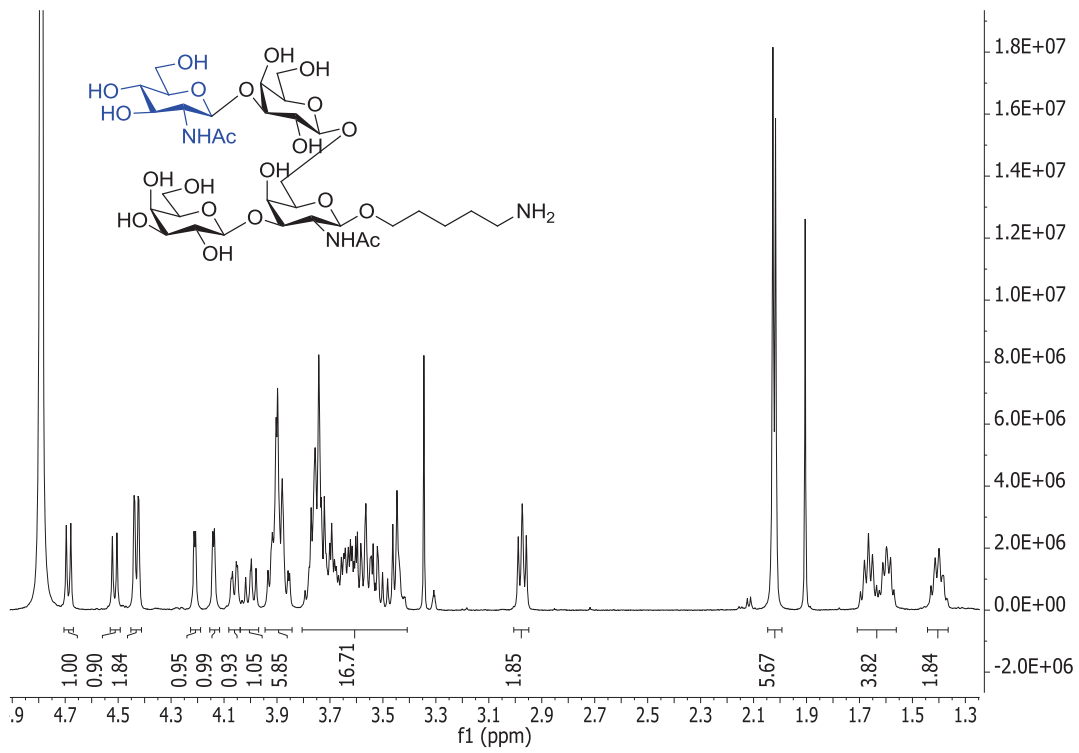


5-aminopentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-{  $\alpha$ -L-fucopyranosyl-(1 $\rightarrow$ 3)-[  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)]- 2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 6)}-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside O5

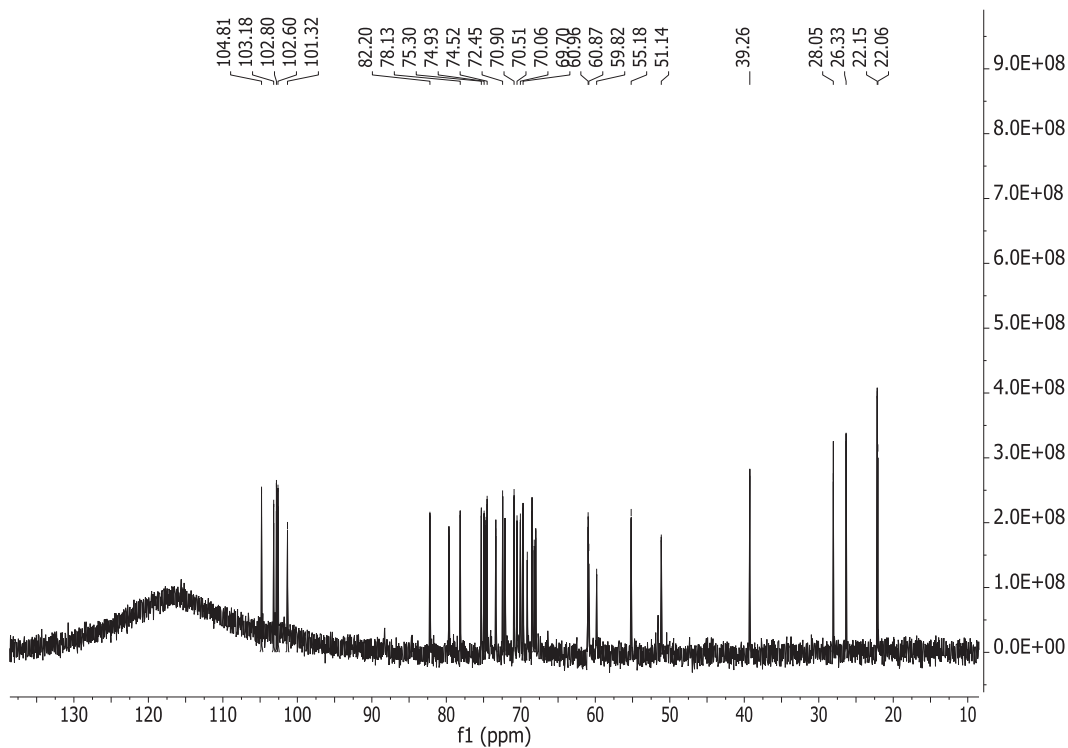
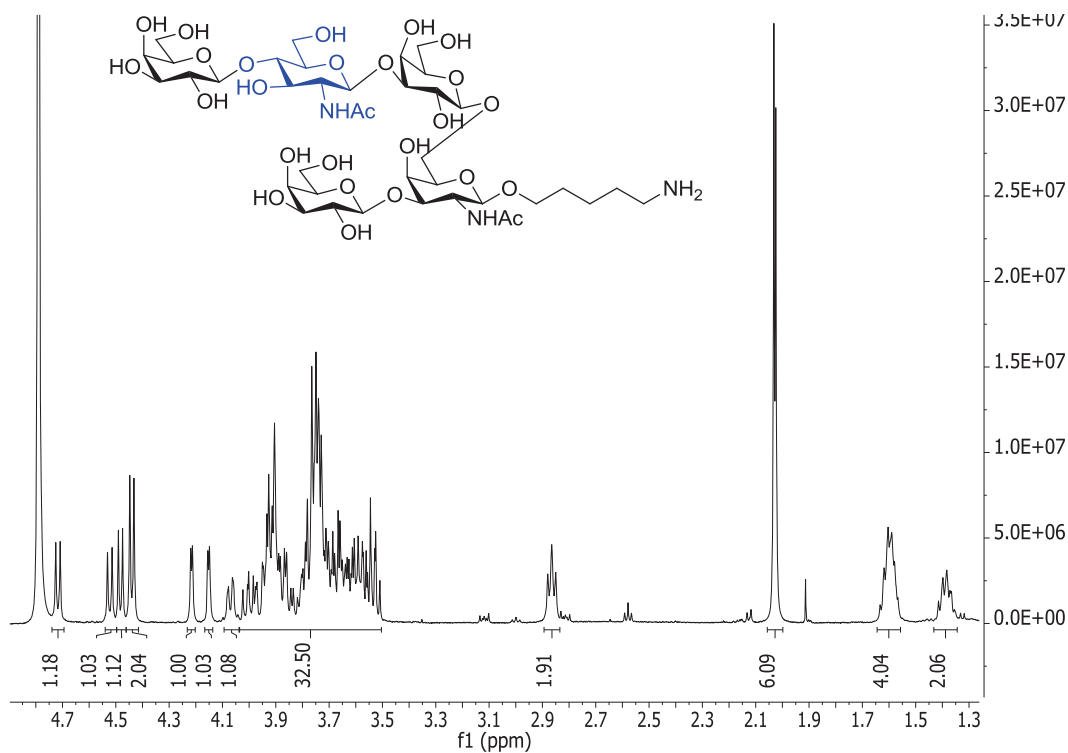


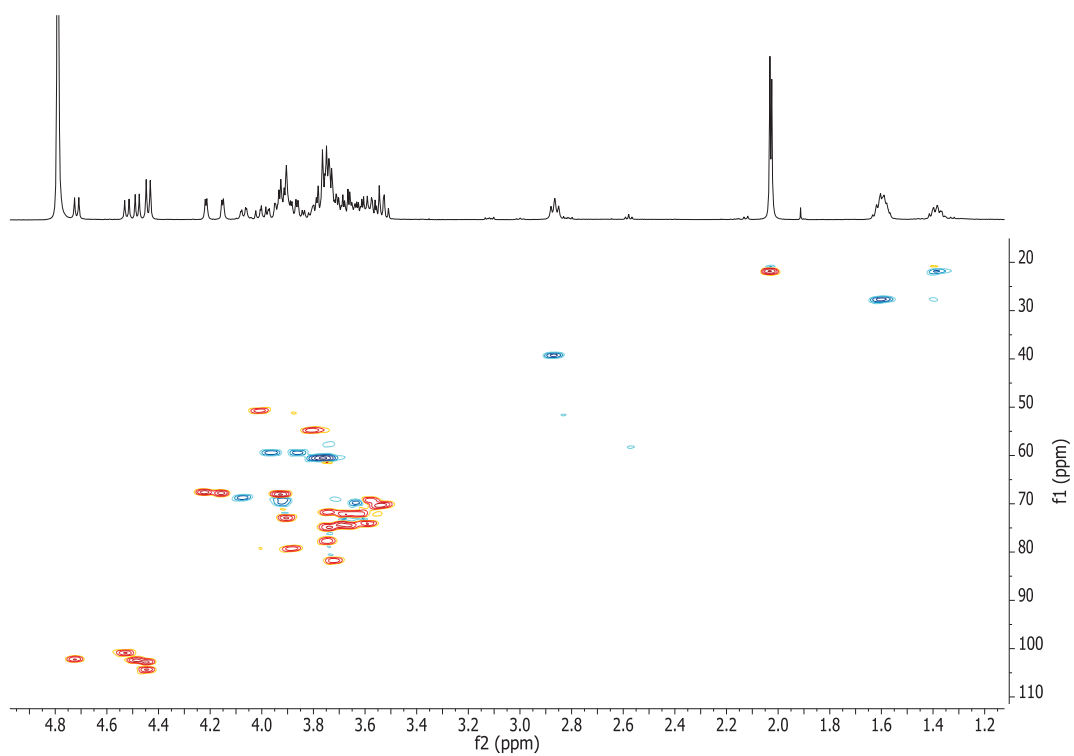


5-aminopentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)]- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside O6

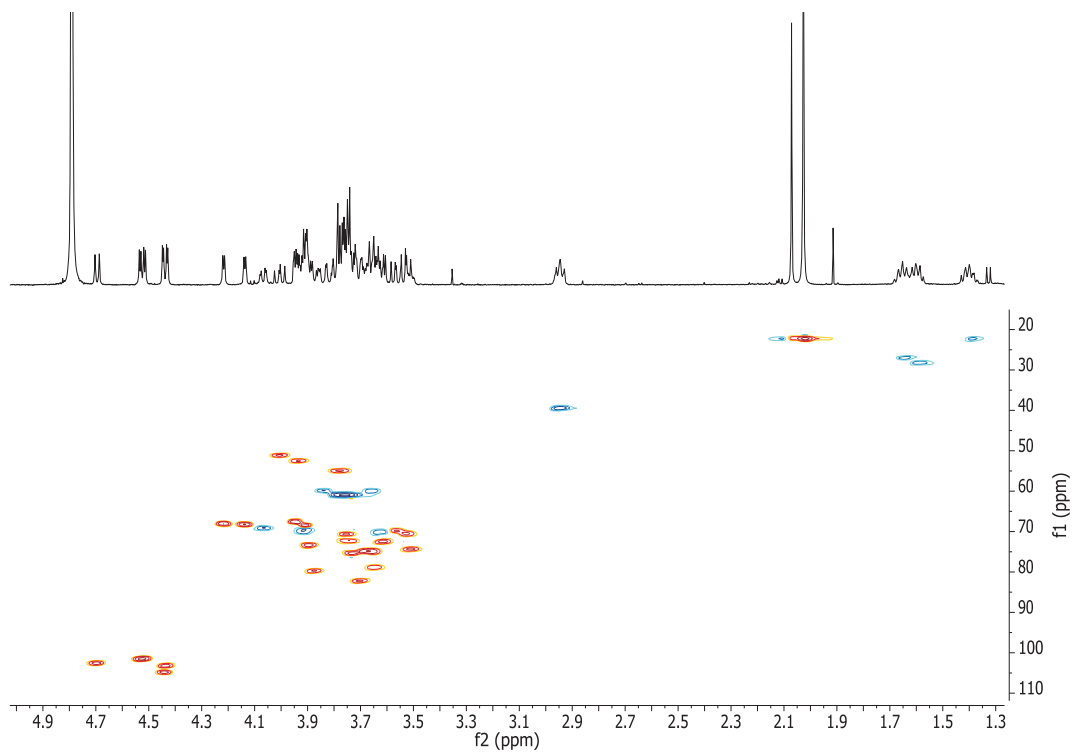
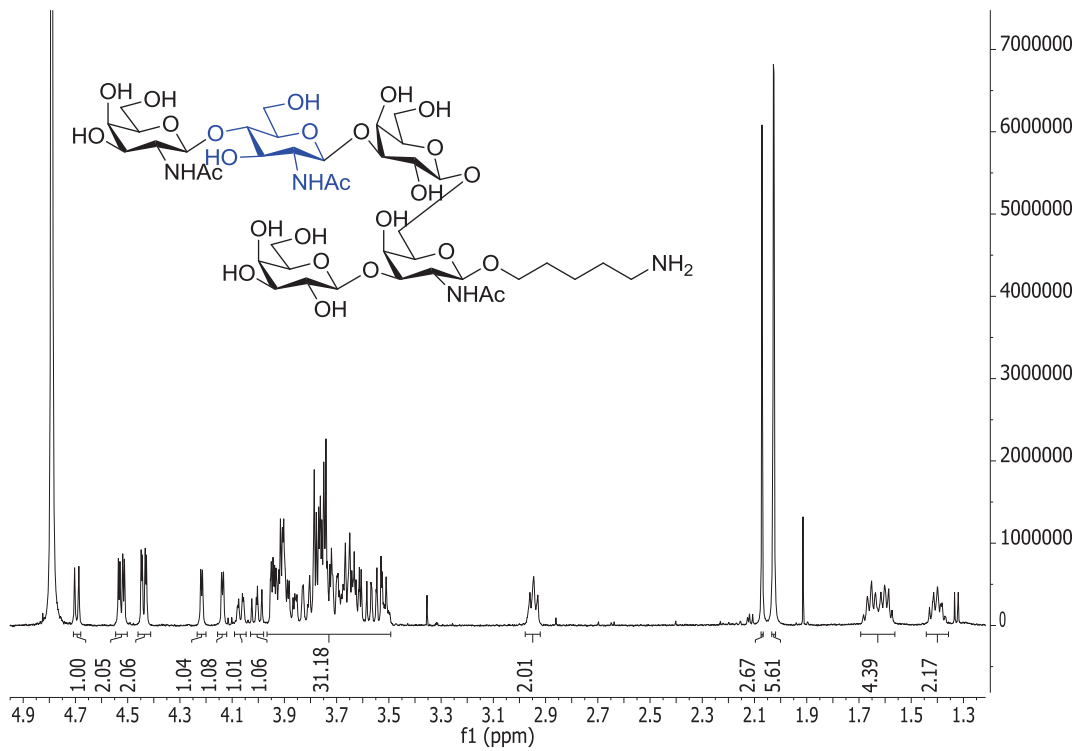


**5-aminopentyl       $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[ $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)-2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)]- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside O7**

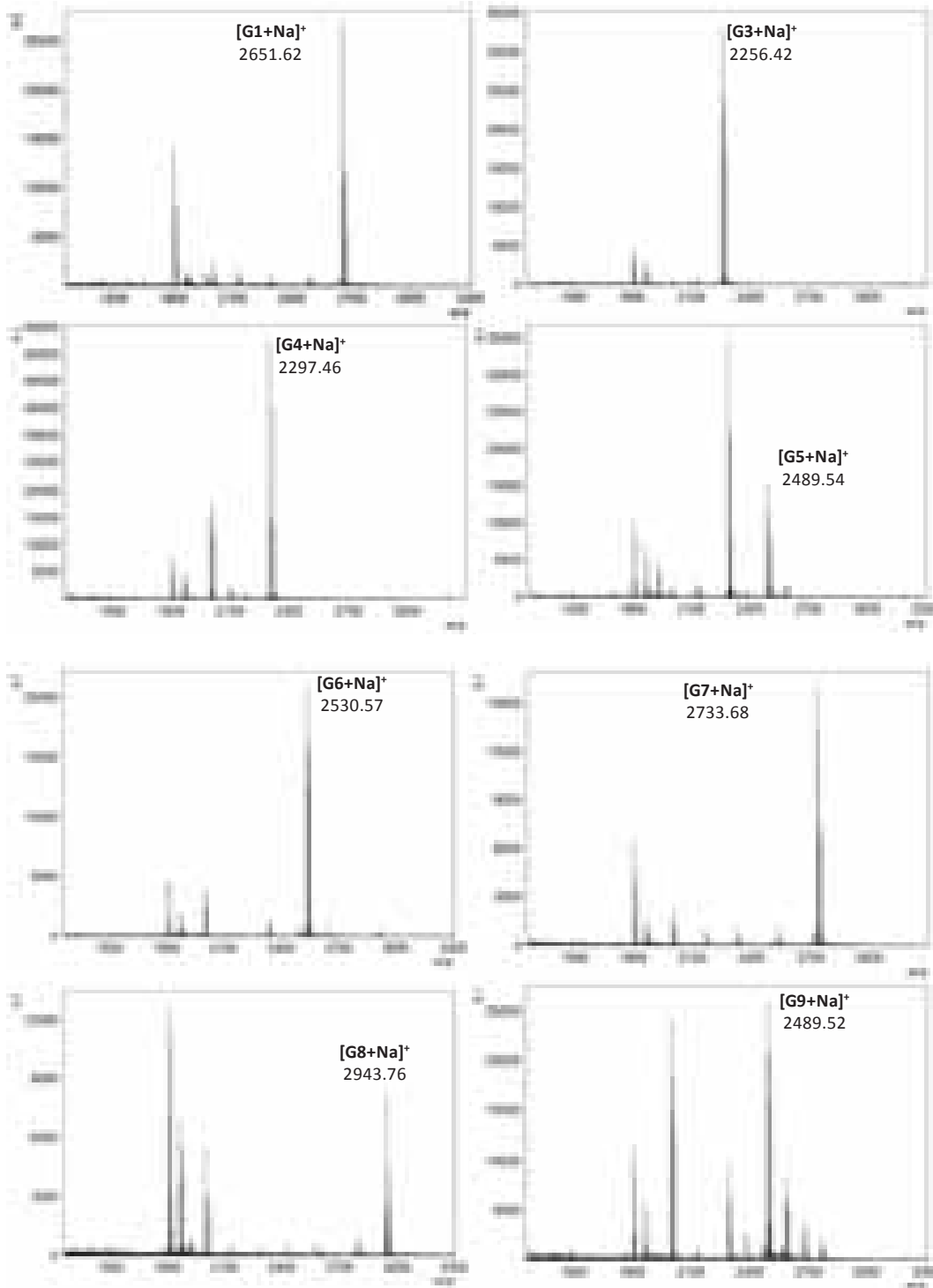


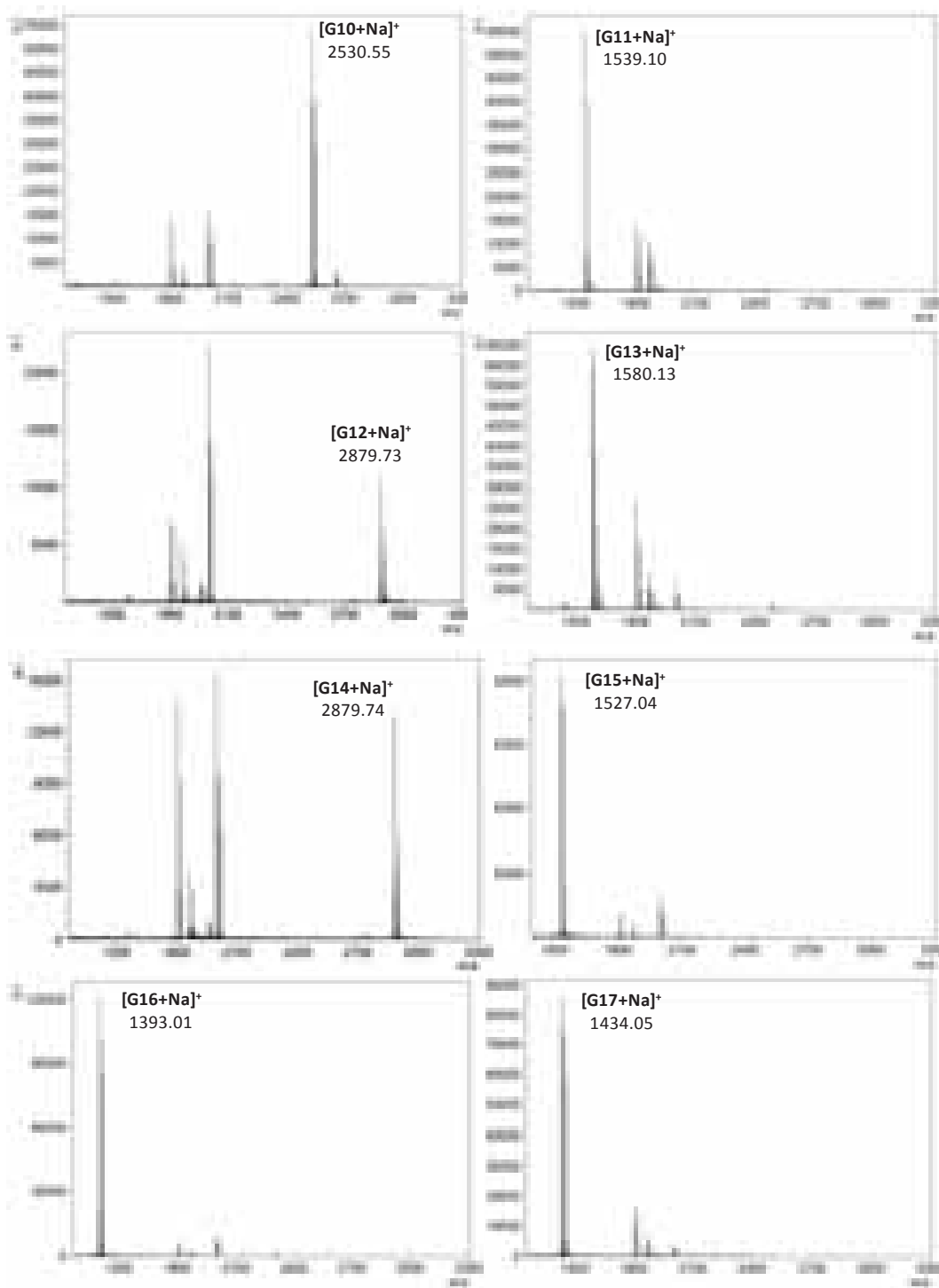


**5-aminopentyl  $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-[2-deoxy-2-acetamido- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)-2-deoxy-2-acetamido- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside O8**

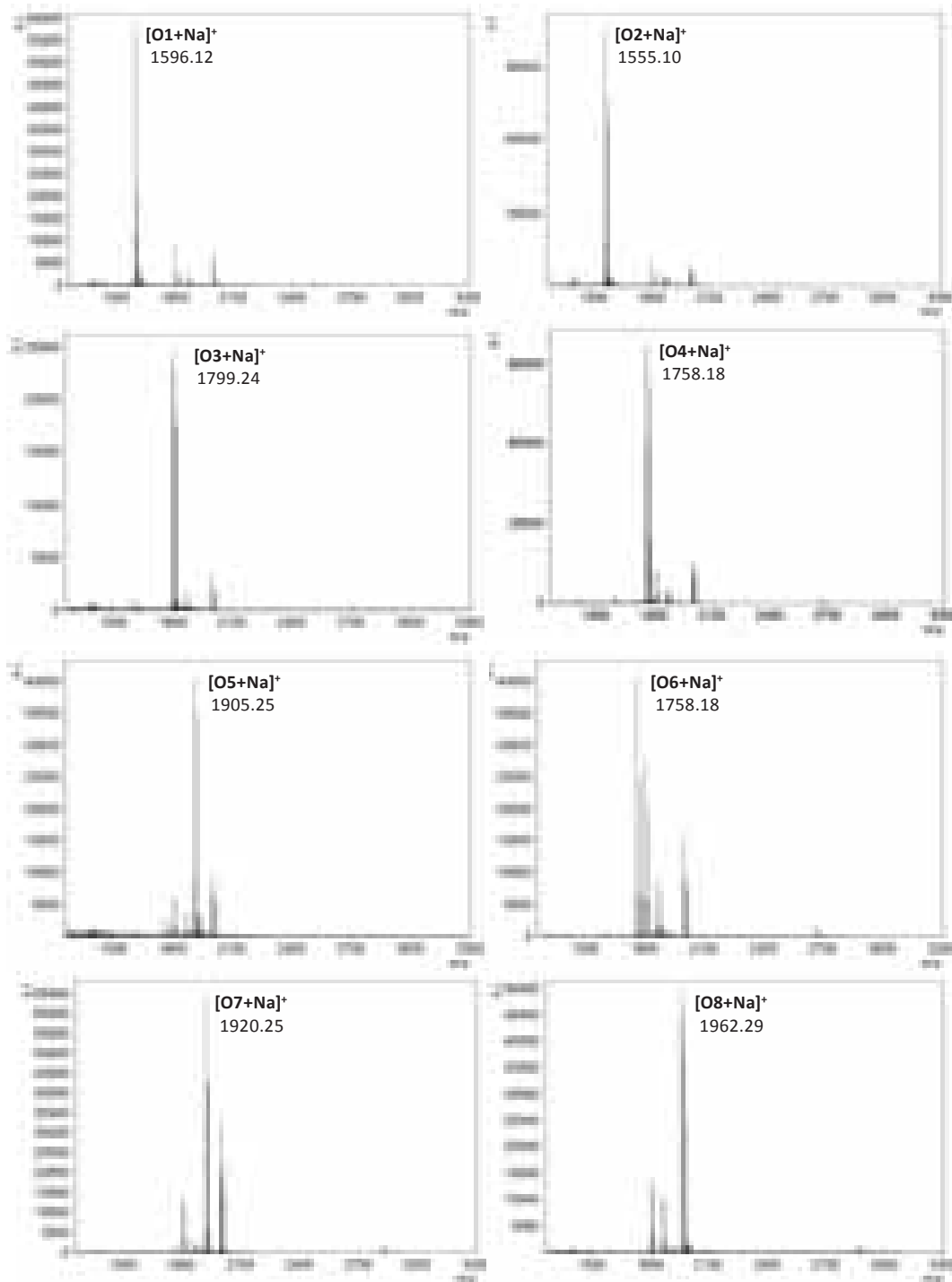


**MALDI-TOF MS of immobilized glycans on surface- surface peaks: 1801.50 and 1997.6826  
m/z**

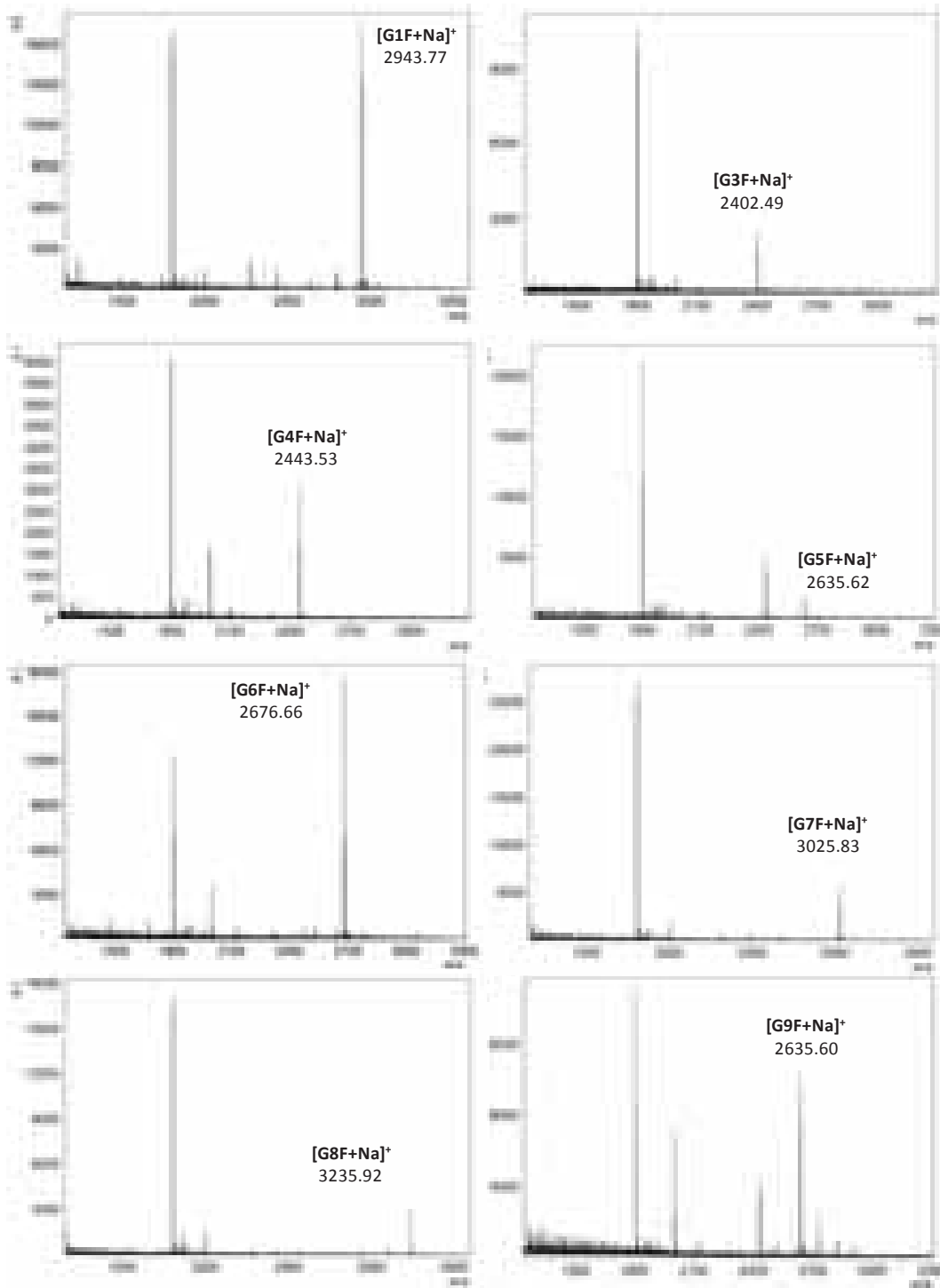


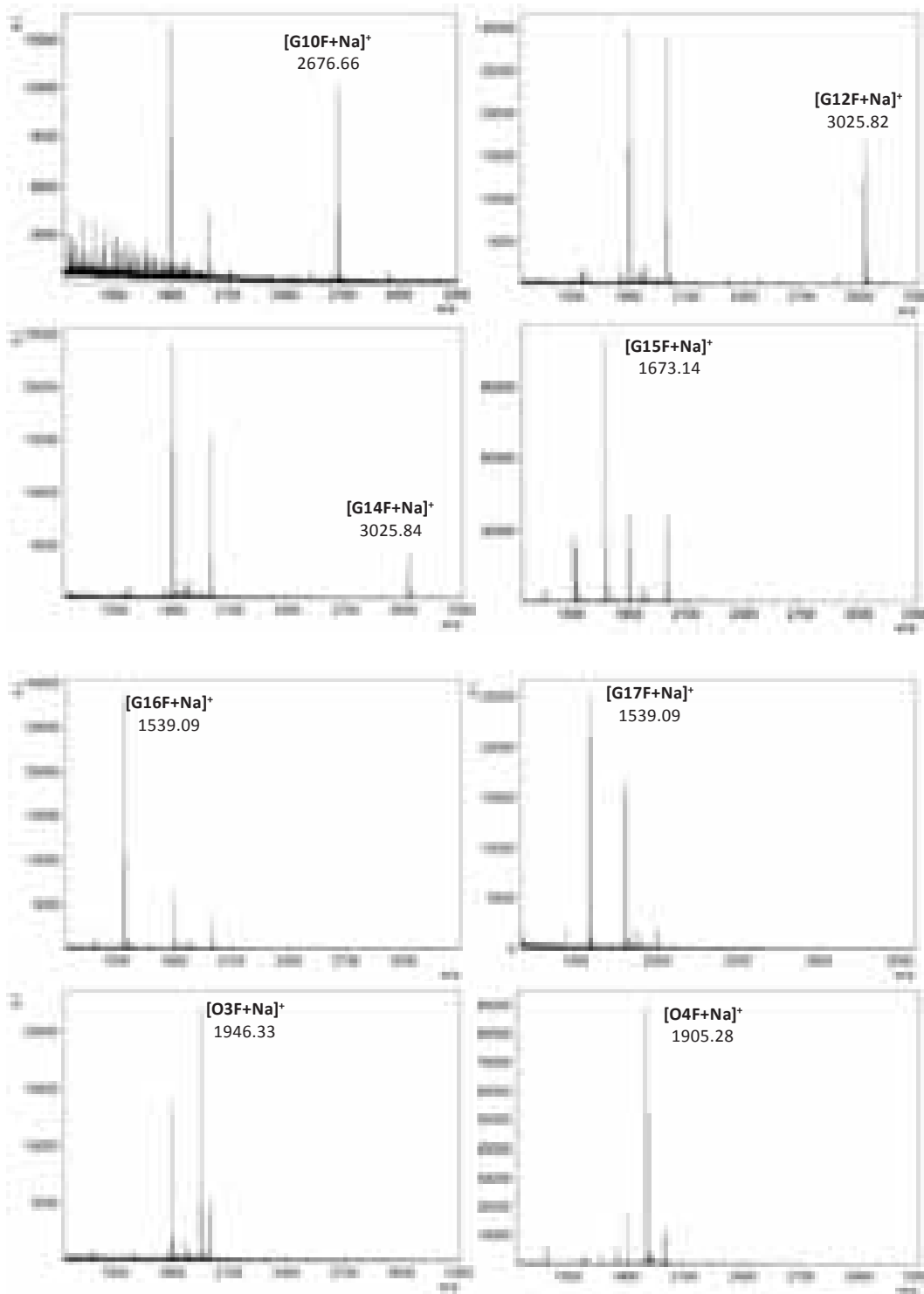


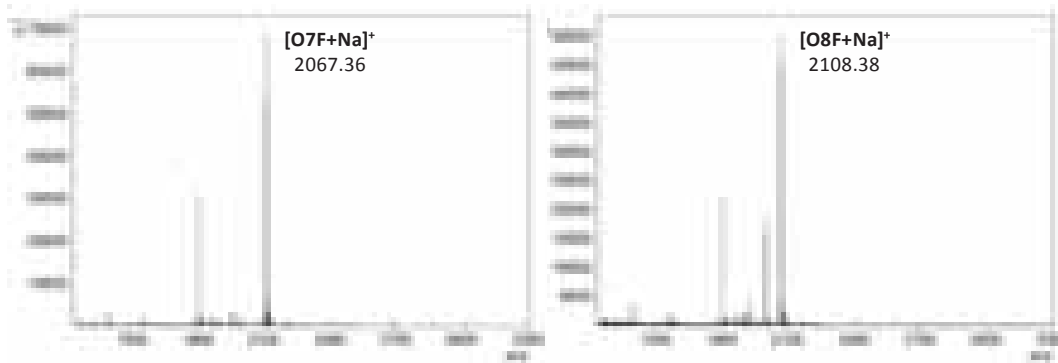




**MALDI-TOF MS of immobilized glycans on surface after treatment with HP-FucT- surface peaks: 1801.50 and 1997.6826  $m/z$**







**MALDI-TOF MS of immobilized glycans on surface after treatment with HP-FucT- surface peaks: 1801.50 and 1997.6826  $m/z$**

