



Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	Synthesis and evaluation of multivalent ligands for macrophage galactose C-type lectin. Development of new chemical tools for immunological research and structural investigation of the lectin with GalNAc and derivatives
Author(s)	Gabba, Adele
Publication Date	2019-09-30
Publisher	NUI Galway
Item record	http://hdl.handle.net/10379/15779

Downloaded 2024-04-26T21:48:11Z

Some rights reserved. For more information, please see the item record link above.



**Synthesis and evaluation of multivalent ligands for
macrophage galactose C-type lectin. Development of new
chemical tools for immunological research and structural
investigation of the lectin with GalNAc and derivatives**

By Adele Gabba MSc



Thesis Presented to
The National University of Ireland Galway
For the degree of
Doctor of Philosophy

Under the supervision and direction of
Prof. Paul V. Murphy
National University of Ireland,
Galway.

Date: 30-09-2019

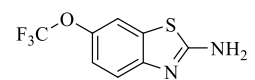
I declare that the following presented work is original and at the best of my knowledge was not copied from any other research. I worked extensively on it alongside talented scientists to produce this innovative work, and the contribution of my collaborators is appropriately acknowledged in the publication. Furthermore, this innovative work has not been previously submitted to any other academic institution.

Adele Gabba

For my professors Roberto Sala and Eliana Frigerio.

For Ida Steiner-Colombo.

For a beginning



Acknowledgments

I clearly remember my last first-year exam; 22nd of September 2010, organic chemistry. Professor Massimo Sisti announced to me that the outcome of the examination was successful. The mark was 26/30. He could tell from my facial expression that I was astonished and moved at the same time. But he couldn't figure out the reason behind my almost tears. He asked me: "What's wrong young lady? Did you expect a higher mark?"

To me, 26/30 is a great score, and it is definitely not the lowest grade I have ever accepted in my academic career.

I was touched because that day I managed to pass all my first-year exams and I couldn't believe it.

That day is so deeply marked in my memories, because for the very first time in twenty years of my life, I realised I was able to achieve something.

Now, almost ten years later, I'm the first one in my family to receive a doctoral degree and I can proudly say, I am not surprised anymore.

In these years I learnt that I like chemistry when it follows laws written by others, but I fall in love with the chemistry that doesn't bend the knee in front of anyone, the Chemistry that keeps me awake at night and pushes me out of bed in the morning so I can dash to the lab to check if the molecule I was chasing for months is eventually mine.

I think everyone has a different idea of what "a successful PhD" means.

For some, the success is measured in how many publications have been accepted, for others in the number of citations and the h-index, for still others, success is counted in funded grants or the number of awards won. All the above are, without doubt, massive achievements.

Nonetheless, in my opinion, a successful PhD is when you start a doctoral degree because you are passionate about a topic, and after few years you are more motivated, more curious and more enthusiastic than when you started. I am still in love with what I do, and I don't see this changing in the near future.

I'm really privileged to have had a positive PhD experience and every second of my life I'm aware that several people around me played a decisive role in it.

I won't say that I will thank them all in the next few lines because it would be a lie, but I'll do my best (with my broken English) to say thank you to the main players!

First and foremost, I'd like to express my deep gratitude to Professor Paul V. Murphy. He gave me what I was craving most: the freedom to fail and pliantly of opportunities! Thanks Paul, you somehow always knew when I needed help and when I just needed more time to figure it out by myself. Thanks to your gentle personality I was never afraid to ask questions, to say "I don't know", "I made a mistake" or "I focused on public engagement this month, I have zero lab work done to show you". You thought me to take pride in my work, I promise I won't forget it.

Thank you to Irish Research Council for being the first founding agency to believe in my potential and practically paying the bills.

Thank you to the NUI Galway crew! Not only my group members, but to all the School of Chemistry fellas and staffs! You are the craic! I probably would need a separate file to properly acknowledge how much they helped me! Amelie, Tanya, Anu, Ben, Max, Lisa, Deb, Antony, Fiach, Richard (Busa brother), Jack, Syl, Ines, Ashis, Marwah, Jimi, Eirini, Jogula, Karen, Megan, Owen, Styliana, Augusto, Ana, Nora, Luis, Yannick, thank you, you are my Irish family.

Thank you to the Dortmund crew: Prof. Westerlind, thank you for not simply hosting me, but welcoming me in your lab. Jin, you are an excellent mentor, I wouldn't have made it without your support, thank you. Thank you Besterlindas, Vanessa and Iris, you never let me eat alone once in my time there, I had great conversation and funny dance time with you! Thank you, Sandra and Manuel! Thanks to EMBO for believing in the project and supporting it.

Thank you to the Boston crew: thank you to Prof. Groopman for hosting me in your lab, thank you to Dr. Birrane for sharing with me what you know, but most of all thank you for having something extraordinarily nice to say the times I needed it most, thank you Özgür (Ozgaaaar) for your guidance and for teaching me patience (and also how to work in a sterile environment) and Yuva for the animated conversations about the (im)perfect Nature and for sharing with me Tatte calories at the end of a working day!

Thank you to my friends outside the Boston lab, Agnese, Luca, Martina. My English vocabulary is too poor to correctly express my feelings for you. You are the most exceptional people I have ever meet. I open up with you more than with anyone else. I miss you every day.

Thank you to the extended MITaly crew and climbing mates, Arturo, Lorenzo, Giampi, Bart, Saras, Francesco, Carlo and Andrea.

Thank you to the Bilbao crew, in particular to Dr. Ana Arda. Prof. Barbero, you are a legend. Thank you FEBS for supporting this project.

Thank you to Cian, my normal, practical, modest, flatmate. Thanks for four intense years of debates about practically every single imaginable subject; I know you have a soft heart behind all those muscles and a damn fine brain!

Thank you to Francesco Civati, (he definitely saw the worse part of this PhD!). Fre, thank you for supporting me no matter what, thank you for teaching me to be kind with myself and for being by my side at every single failure. You are more than my husband, you are my best friend, my most trusted colleague and the family I chose for the rest of my days.

Thank you to Rosaria Gabba, my aunt. I know you still don't understand what exactly I am doing for living (neither do I) but still, you call me once a week to ask how things are and you always like to remind me that I have tough skin! I love it!

Thank you to my little sister and her lovely husband.

I'd like to thank Dr. Andrea Erxleben and Professor Franck Fieschi for taking the time to read and correct my thesis, but most of all for the smart questions the enjoyable discussion during my thesis defence.

Eventually, I'd like to thank you, dear reader. I hope you will find in this thesis some useful content and a few ideas for your research project, don't forget to have fun!

Table of content

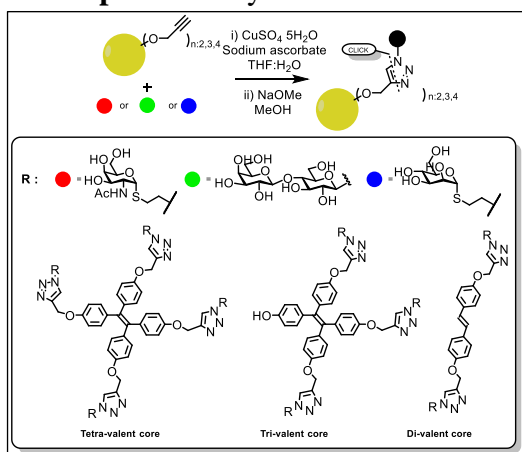
Abstract	xiii
Symbols and Abbreviations	xv
1 Human MGL (CLEC10A, CD301) orthologue structure and selectivity: a review..	19
1.1 Abstract.....	19
1.2 Introduction.....	21
1.3 Human MGL: three isoforms, shared affinity.....	21
1.4 Extracellular fragment oligomerisation	24
1.5 Glycan binding.....	25
1.6 MGL expression	32
1.7 How to target MGL: binding mode and selective ligands overview	34
1.8 Multivalent inhibitors and effectors.....	38
1.9 Concluding remarks	46
1.10 Bibliography	47
2 Synthesis of multivalent ligands based on tetraphenyl ethylene geometry.....	53
2.1 Abstract.....	53
2.2 Introduction.....	54
2.3 Results and discussion	55
2.4. Papers.....	90
2.5 Bibliography	125
3. Synthesis and evalion of mucin 1 and mucin 4 (neo)peptide library containing a ligand for hMGL. Towards the development of candidate vaccines.....	129
3.1. Abstract.....	129
3.2. Introduction: basic immunology principles	132
3.3. The mucin family.....	135
3.4. MGL based vaccine development.....	136

3.5.	Results and discussion.....	138
3.6.	Conclusion.....	157
3.7.	Bibliography.....	158
4.	Synthesis and study of a peptide presenting a glycocluster with potential to assemble to a coiled-coil.....	163
4.1.	Abstract.....	163
4.2.	Introduction: coiled-coils.....	165
4.3.	Sequence analysis.....	166
4.4.	Scaffold synthesis.....	168
4.5.	Circular Dichroism.....	171
4.6.	Conclusion.....	176
4.7.	Bibliography.....	177
5.	Expression and crystallisation of hMGL carbohydrate recognition domain.....	179
5.1.	Abstract.....	179
5.2.	Results.....	180
5.3.	Discussion.....	195
5.4.	Conclusion.....	200
5.5.	Bibliography.....	201
6.	Experimental section.....	203
6.1.	Synthesis: general experimental conditions.....	203
6.2.	Small molecule experimental procedure and characterisation.....	204
6.3.	Peptide synthesis and characterisation.....	254
6.4.	Protein: experimental procedure and characterisation.....	286
6.5.	Isothermal titration calorimetry methodology.....	293
6.6.	Bibliography.....	294

Abstract

Synthetic chemistry can provide powerful tools to explore the correlation between an antigen's chemical structure and the type of immune response triggered. This thesis work has focused on the design, synthesis and evaluation of multivalent molecules which target a relevant protein of the immune system, the human macrophage galactose C-type lectin (hMGL).

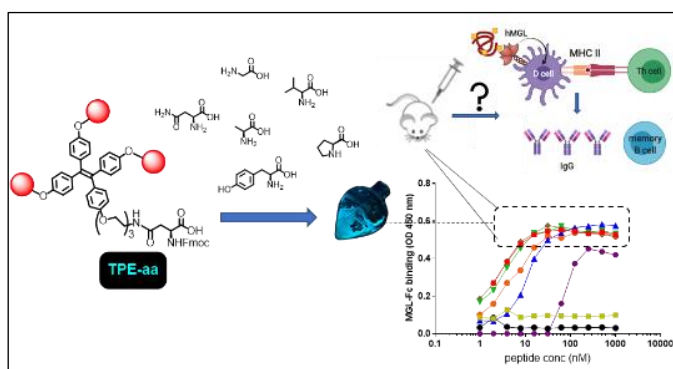
In **chapter 2** the synthesis of multivalent glycomimetics is described as well as measurements



of their affinity for hMGL. The main objective was to learn more about the structure activity relationship with regard to ligands based on tetraphenyl ethylene (TPE) for hMGL and other lectins. To achieve this divalent, trivalent and tetravalent ligand, based TPE geometry as well as trans-stilbene presenting α -mannopyranoside, β -lactose and α -thio-GalNAc as headgroups, were synthesised. The identified

synthetic route to such compounds is reproducible and scalable. The thermodynamics of the interaction of α -thio-GalNAc derivatives were studied using the extracellular domain of hMGL and hMGL CRD, revealing insights into the possible interaction of these compounds with the lectins.

In **chapter 3** the synthesis of an extensive anti-cancer synthetic vaccine candidate library based on MUC1/MUC4 peptide backbone is described. The main objective of this work was to enhance antigen presentation by hMGL via the major histocompatibility complex II. To

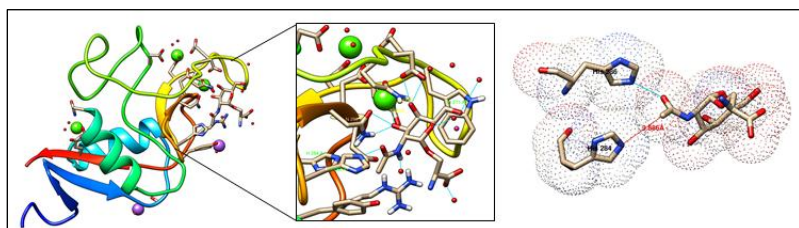


achieve this objective mucin epitopes were decorated with a trivalent glycocluster selective for hMGL. Through solid phase synthesis synthetic peptides, including those presenting the glycocluster were obtained. These were found to bind hMGL from 10 to 200-

fold better than the natural peptides in an ELISA performed in the S. van Vliet laboratory in Amsterdam. The next step of this project will be to identify peptides which induce the desired signalling cascade and test them on mouse model for the production of long-lasting antibodies IgG. This is the first reported attempt to target a receptor with neoglycosylated peptides bearing a fragment specifically designed to selectively target hMGL.

In **chapter 4** the synthesis of a (neo)peptide containing a glycocluster with potential to give a trimeric coiled-coil sequence is described. The glycocluster is based on the tetraphenylethene residue. The peptide was successfully synthesised and purified. While some evidence for helix formation was obtained by CD spectroscopy, the refolding of the peptide after heating/denaturing was not observed and may have occurred partially.

In **chapter 5** the first structural elucidation of hMGL carbohydrate binding mode is reported using X-ray crystallography. hMGL is the only lectin on human DCs that selectively recognise



carbohydrate antigens present on many cancer cell lines and was proven to have the ability to lead the immune system to produce permanent antibody

against cancer. The aim of this work was to understand the structural basis of the hMGL/carbohydrate interaction and to allow a more reliable in silico ligand design. To achieve this the protein was expressed in bacterial cell lines. As the expression resulted in insoluble protein, a refolding screen optimisation was carried out, which facilitated recovery of the protein in high purity and good yield for crystallisation screening. As a result of screening and optimisation, suitable crystals for diffraction were obtained. This led to **elucidation of the first crystal structure of hMGL carbohydrate recognition domain (CRD)**, binding with several monovalent GalNAc derivatives. The structure offers a possible explanation of hMGL binding mode and selectivity.

Symbols and Abbreviations

α	alpha	Dd	Doublet of doublets (spectral)
Å	angstrom	ddd	Doublet of doublets of doublets (spectral)
aa	amino acid	DEPT	Distortionless Enhancement by Polarisation Transfer
Ac	acetyl	DIC	<i>N,N</i> -diisopropylcarbodiimide
APC	antigen presenting cell	DIPEA	diisopropylethylamine
APT	attached proton test	DMAP	4-dimethylaminopyridine
aq.	aqueous	DMF	dimethylformamide
arom.	aromatic	DMSO	dimethyl sulfoxide
β	beta	DNA	deoxy ribonucleic acid
BCR	B-cell receptor	DPS	Diphenyl sulfoxide
Boc	<i>tert</i> -butyloxycarbonyl	dt	Doublet of triplets (spectral)
Bn	Benzyl	EDTA	Ethylenediaminetetraacetic acid
br d	Broad doublet (spectral)	ELISA	enzyme linked immunosorbent assay
br s	Broad singlet (spectral)	Eq.	equivalents
BSA	bovine serum albumin	ES- HRMS	High-Resolution Mass Spectrometry - Electrospray Ionization
Bu	butyl	ESI-MS	electron spray ionization mass spectrometry
Bz	Benzoyl	Et	ethyl
c	concentration	AC	affinity chromatography
C°	Celsius degrees	Fmoc	<i>N</i> -(9H-fluoren-9-yl)-methoxycarbonyl
calc.	calculated	Fuc	L-fucose
cat.	catalytic	Gal	D-galactose
Cy	Cyclohexane	Gal-3	galectin-3
ConA	ConcavallinA	GalNAc	<i>N</i> -acetyl-D-galactosamine
COSY	correlated spectroscopy	gg	gauche-gauche
CRD	carbohydrate recognition domain	Glc	D-glucose
CTL	cytotoxic T-lymphocyte	GlcNAc	<i>N</i> -acetyl-D-glucosamine
CuAAC	Copper(I)-catalyzed Azide-Alkyne Cycloaddition	grad.	gradient
Cy	Cyclohexyl	h	hour
δ□	Chemical shift in ppm downfield from TMS	HATU	<i>O</i> -(1H-7-azabenzotriazol-1-yl)-1,1,3,3- tetramethylunronium hexafluorophosphate
d	Doublet (spectral), day	HBr	hydrobromic acid
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene	HCl	hydrochloric acid
DC	dendritic cell	HBTU	<i>O</i> -(1H-benzotriazol-1-yl)-1,1,3,3- tetramethylunronium hexafluorophosphate
DCC	<i>N,N</i> -dicyclohexylcarbodiimide	HMBC	heteronuclear multiple bond correlation
DCM	dichloromethane	HOAt	7-aza-1-hydroxybenzotriazole
HOBt	1-hydroxybenzotriazole	NBS	<i>N</i> -bromosuccinimide
HPLC	high performance liquid chromatography	Neu5Ac	<i>N</i> -acetylneuraminic acid

HR	high resolution	NHS	<i>N</i> -hydroxysuccinimide
HSQC	heteronuclear single quantum coherence	NIS	<i>N</i> -iodosuccinimide
Hz	hertz	NMP	<i>N</i> -methylpyrrolidone
IC50	50% inhibition concentration	NMR	nuclear magnetic resonance
Ig	immunoglobuline	NSCLC	non-small cell lung cancer
IgG	immunoglobuline G	OTf	trifluoromethanesulfonate (triflate)
IL	interleukin	<i>p</i>	<i>para</i>
<i>iPr</i>	isopropyl	PBS/PBST	phosphate buffered saline/PBS+Tween
IR	Infrared (spectroscopy)	PCR	polymerase chain reaction
ITC	isothermal titration calorimetry	PEG	polyethylene glycol
<i>J</i>	coupling constant	PG	protecting group
Lac	lactose	Ph	phenyl
LacNAc	<i>N</i> -acetyllactosamine	pH	potential of hydrogen
Le^x/Le^a	Lewis x/Lewis a	Phth	phthalimide
m	multiplatt, mouse	PMP	<i>para</i> -methoxy benzylidene
<i>m</i>	meta	ppm	parts per million
M⁺, M⁻	Mass of the molecular ion (mass spectrometry)	<i>p</i>-TsOH	<i>para</i> -toluenesulfonic acid
M	molarity, mega	Py.	Pyridine
mAb	monoclonal antibody	PyBOP	benzotriazol-1-ylxytris(pyrrolidino)phosphonium hexafluorophosphat quartett
MALDI	matrix assisted laser desorbition ionisation	q	quartary
Man	D-mannose	R	residue
mb	millibar	R_f	retention factor
MD	molecular dynamics	RNA	ribonucleic acid
Me	methyl	RP	reverse phase
MGL	macrophage c-type lectin	R_t	retention time
MHC	major histocompatibility complex	r.t.	Room temperature
min	minutes	RuAAC	Ruthenium catalysed Azide-Alkyne Cycloaddition
miRNA	micro ribonucleic acid	s	singlett
MPL	monophosphoryl lipid A	sLe^x/sLe^a	sialyl-Lewis x/Sialyl-Lewis a
mRNA	messenger ribonucleic acid	SNP	single nucleotide polymorphism
ml, µl	milliliter, microliter	SPPS	solid phase peptide synthesis
mol	Mole	ST	sialyl-Thomsen Friedenreich
mM, µM	Millimolar, micromolar	STD	saturation transfer difference
MS	mass spectrometry	ST_N	sialyl-Thomsen Nouveau
NaN₃	sodium azide	Su	succinimide
NaOMe	Sodium methoxide	TLC	thin layer chromatography
t	triplett	TLR	Toll-like receptor
T/TF	Thomsen-Friedenreich-antigen	TMS	tetramethylsilyl
TACA	tumor associated carbohydrate antigen		

TBAF	<i>Tert</i> -butylammonium fluoride	T_N	Thomsen-Nouveau-antigen
TBAI	Tetrabutylammonium iodide	TOCSY	total correlation spectroscopy
TBS	<i>tert</i> -butyldimethylsilyl	TOF	time of flight
TBTU	<i>O</i> -(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate	Troc	trichloroethoxycarbonyl
TCR	T-cell receptor	Trt	trityl
TEG	triethylene glycol	TT	tetanus toxoid
<i>tert</i>	tertiary	UV	ultraviolet
TES	triethylsilyl	VNTR	variable number of tandem repeats
TFA	trifluoroacetic acid	α	specific optical rotation
TfOH	Trifluoromethanesulfonic acid	δ	chemical shift
THF	tetrahydrofuran	λ	wavelength
TIPS	triisopropyl silane		

Amino acid codes

Ala, A	Alanine
Arg, R	Arginine
Asn, N	Asparagine
Asp, D	Aspartate
Cys, C	Cysteine
Gln, Q	Glutamine
Glu, E	Glutamate
Gly, G	Glycine
His, H	Histidine
Ile, I	Isoleucine
Leu, L	Leucine
Lys, K	Lysine
Met, M	Methionine
Phe, F	Phenylalanine
Pro, P	Proline
Ser, S	Serine
Thr, T	Threonine
Trp, W	Tryptophan
Tyr, Y	Tyrosine
Val, V	Valine

1 Human MGL (CLEC10A, CD301) orthologue structure and selectivity: a review

1.1 Abstract

Among the many C-type lectin receptors (CLRs) expressed on dendritic cells (DCs) the human macrophage galactose C-type lectin (hMGL, CLEC10A) is the only one known to selectively recognise derivatives of N-Acetylgalactosamine (GalNAc) such as the tumour associated antigen TN. hMGL can internalise exogenous and endogenous antigens, traffic them in defined DCs compartment and, through mechanism not fully explored, induce their presentation to immune system cells via the MHC-I or MHC-II.^{1,2,3}

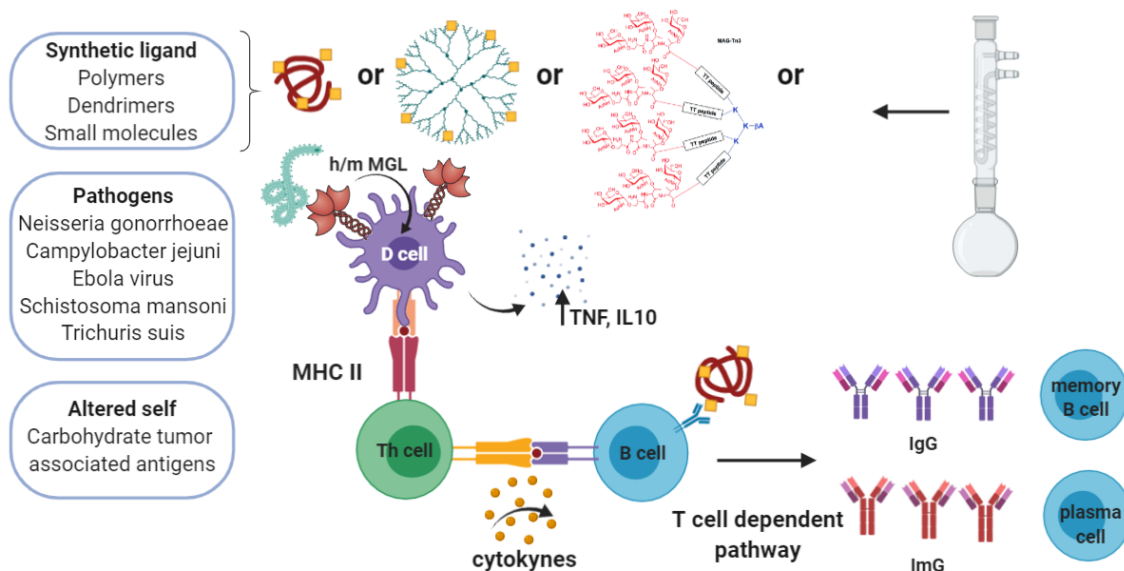


Figure 1. Graphical abstract. Pathogens, cancer cells and synthetic molecules bearing galactosamine moieties, are recognised by MGL and trigger different biological processes.

The type of immune response promoted by MGL seems highly dependent from the antigen structure,⁴ but, despite the tremendous advancements in the design of selective potent ligands for MGL,^{5,6,7} internalisation, trafficking and signalling exploration still mainly rely on the usage of monoclonal antibodies. The structure activity relationship that governs the immune outcome after the engagement of this receptor is still unknown.

I believe that to answer significant questions on the role of MGL, an intersectoral knowledge in chemistry, immunology and cell biology is needed. Comprehensive reviews about MGL's ability to modulate innate and adaptive immunity have been published in the literature.^{8,9,10,11} In these reviews, the type of chemical structures used to engage MGL are discussed only superficially.

Chapter1

In this introductory chapter I wish to offer, from a chemical prospective, a practical overview of the type of ligands and mechanism that can be exploited to target MGL and, in simple term, the type of immune outcome different classes of molecules produce. Furthermore, to encourage chemistry research groups to attempt protein expression, evaluate its binding affinity, and explore this receptor role in the sophisticated immune apparatus, I wish to provide an overview about the existing methodology for the expression of the human orthologue of MGL and structural basis of glycan recognition known. I hope to hand out an initial tool to help more research groups in the scientific community to dive in the fascinating exploration of this not well studied receptor and inspire the chemical synthesis of novel ligands.

1.2 Introduction

Human macrophage galactose-type lectin (MGL, CLEC10A, DC-ASGPR or CD301. UNIPROT entry Q8IU9) is an endocytic receptor exclusively expressed on certain dendritic cells subsets¹² and macrophages. Among the C-type lectin receptor family, this is the only known receptor to selectively recognise with high affinity terminal N-acetylgalactosamine (GalNAc) derivatives. Terminal GalNAc is found rarely in a healthy body. When they arise, their presence is often observed on members of the mucin family, (MUC1, MUC2 and MUC4) expressed on tumour epithelial cells. The ability of MGL to recognise selectively this aberrant glycans can thus, potentially, be exploited for immunological purposes. Furthermore, hMGL binds the envelope protein GP2 of EBOLA virus¹³ and other pathogens like *Neisseria gonorrhoeae*,¹⁴ *Campylobacter jejuni*,¹⁵ *Schistosoma mansoni*,¹⁶ *Trichuris suis*.¹⁵

Orthologues of MGL are present in the human, mouse and rat. It is worth noting that while all the orthologues, to a certain extent, recognise terminal glycans, which have the pyranose C-3 and C-4 hydroxyls respectively equatorial and axial, and bind them in calcium dependent manner, only the mouse MGL-2 shares GalNAc selectivity with the human orthologues, while mouse MGL-1 and rat MGL binds Lewis X and Lewis A, figure 2.

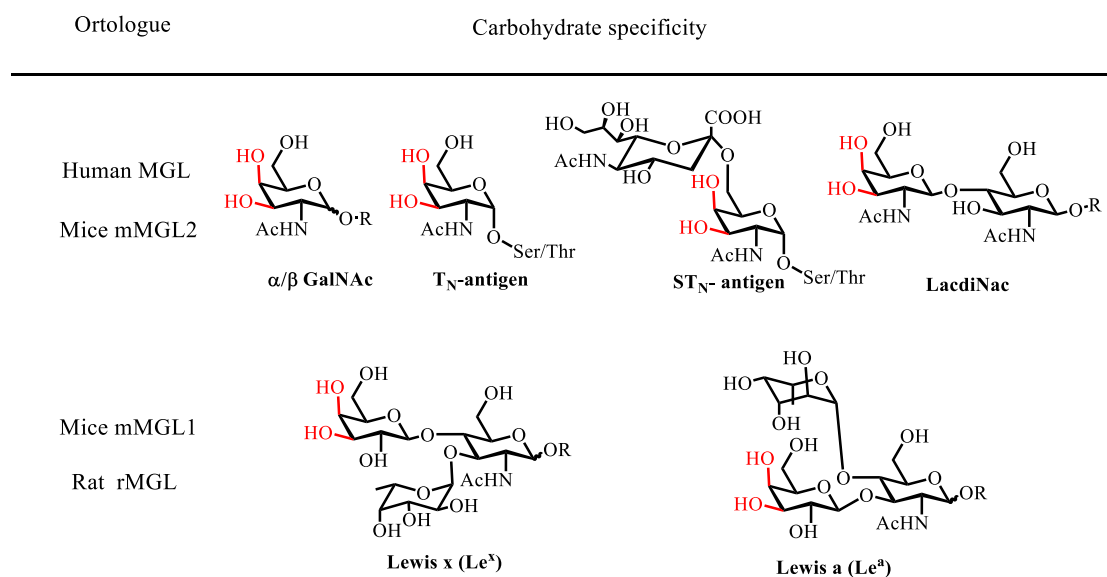


Figure 2. Structure of carbohydrates recognised by different MGL orthologues. Due to the QPDxW sequence in the CRD all the orthologues, at certain extent, recognise terminal glycan with C3 and C4 hydroxyl groups equatorial and axial respectively in a calcium dependent manner.

1.3 Human MGL: three isoforms, shared affinity.

In humans, MGL is a 30kDa protein with two possible glycosylation sites at Asn78 and Asn73. Evidence of post translational modifications are reported when the protein is expressed in human cell lines (HEK293T cells).¹⁷ The first 39 amino acids of hMGL are located in the cytoplasmic region of the cell and contain a Tyr-Glu-Asn-Phe (YENF) internalization signal, followed by a transmembrane region of the protein, which are amino acids 40-60 as numbered according to the UniProt accession code Q8IUN9. The extracellular portion, amino acids 61-316, contains in total 255 amino acids, and a long coiled-coil motive, amino acids 85-176, is present. Amino acids 188- 305 form the carbohydrate recognition domain (CRD). There is only one MGL gene in humans, but due to a post-transcriptional modification, this single gene can code for multiple proteins. This process is commonly defined as gene-splicing and the protein arising from it are called isoforms.¹⁸ Isoforms share high protein sequence similarity, but as in the case of hMGL, the modification of few residue can lead to different proprieties. ¹⁹ The extracellular portion of the protein is present as three major isoforms identified with accession numbers NP_878910.1, NP_006335.2 and NP_001316999.1 of the National Center for Biotechnology Information. Screening of tissue cDNA libraries with polymerase chain reaction primers reveals that the most common form of mRNA encodes isoform 3.²⁰ This study only explored the percentage of hMGL isoforms at mRNA level, this ratio can be different at expression level and the different isoforms ratios could vary from cell subset to cell subset.²¹ Currently no comprehensive study about hMGL isoforms at mRNA and expression level is reported in the literature. Eventually, it is quite common encounter journal articles and review in which the isoform is not specified it is useful knowing that is possible cross-check this information on the National Center for Biotechnology Information website where publications associated to isoforms 1,2 and 3 are reported.

1.3.1. Isoforms sequence comparison

The sequence comparison of the three main isoforms is shown in figure 3.

Isoform 3 is the shortest of the isoforms, it is coded from the most common form of mRNA, and was firstly cloned from DCs.²² Isoform1, first identified by Valladeau et al.²³, is cloned by immature dendritic cells and is the longest isoform due to an insertion of 27 amino acids in the neck region.

Consensus	MTRTYENFQY	LENKVKVQGF	KNGPLPLQSL	LQRLCSGPCH	LLLSLGLGLL
Conservation	██████████	██████████	██████████	██████████	██████████
hMGL_ISO3	1 MTRTYENFQY	LENKVKVQGF	KNGPLPLQSL	LQRLCSGPCH	LLLSLGLGLL
hMGL_ISO1	1 MTRTYENFQY	LENKVKVQGF	KNGPLPLQSL	LQRLCSGPCH	LLLSLGLGLL
hMGL_ISO2	1 MTRTYENFQY	LENKVKVQGF	KNGPLPLQSL	LQRLCSGPCH	LLLSLGLGLL
Consensus	LLVVICVVGF	QNSKFQRDLV	TLRTDFSNFT	SNTVAEIQAL	TSQGSLEET
Conservation	██████████	██████████	██████████	██████████	██████████
hMGL_ISO3	51 LLVVICVVGF	QNSKFQRDLV	TLRTDFSNFT	SNTVAEIQAL	TSQGSLEET
hMGL_ISO1	51 LLVVICVVGF	QNSKFQRDLV	TLRTDFSNFT	SNTVAEIQAL	TSQGSLEET
hMGL_ISO2	51 LLVVICVVGF	QNSKFQRDLV	TLRTDFSNFT	SNTVAEIQAL	TSQGSLEET
Consensus	IASLKAEVEG	FKQERQAVHSEML
Conservation	██████████	████████████████████
hMGL_ISO3	101 IASLKAEVEG	FKQERQAVHSEML
hMGL_ISO1	101 IASLKAEVEG	FKQERQAGVS	ELQEHTTQKA	HLGHCPHCPS	VCVPVHSEML
hMGL_ISO2	101 IASLKAEVEG	FKQERQAVHSEML
Consensus	LRVQQLVQDL	KKLTCCVATL	NNN...ASTE	GTCCPVNWVE	HQDSCYWFSSH
Conservation	██████████	██████████	██████████	██████████	██████████
hMGL_ISO3	124 LRVQQLVQDL	KKLTCCVATL	NNN...ASTE	GTCCPVNWVE	HQDSCYWFSSH
hMGL_ISO1	151 LRVQQLVQDL	KKLTCCVATL	NNN...ASTE	GTCCPVNWVE	HQDSCYWFSSH
hMGL_ISO2	124 LRVQQLVQDL	KKLTCCVATL	NNNGEEASTE	GTCCPVNWVE	HQDSCYWFSSH
Consensus	SGMSWAEAEK	YCQLKNAHLV	VINSREEQNF	VQKYLGSAYT	WMGLSDPEGA
Conservation	██████████	██████████	██████████	██████████	██████████
hMGL_ISO3	171 SGMSWAEAEK	YCQLKNAHLV	VINSREEQNF	VQKYLGSAYT	WMGLSDPEGA
hMGL_ISO1	198 SGMSWAEAEK	YCQLKNAHLV	VINSREEQNF	VQKYLGSAYT	WMGLSDPEGA
hMGL_ISO2	174 SGMSWAEAEK	YCQLKNAHLV	VINSREEQNF	VQKYLGSAYT	WMGLSDPEGA
Consensus	WKWVDGTDYA	TGFQNWKPGQ	PDDWQGHGLG	GGEDCAHFHP	DGRWNDDVCQ
Conservation	██████████	██████████	██████████	██████████	██████████
hMGL_ISO3	221 WKWVDGTDYA	TGFQNWKPGQ	PDDWQGHGLG	GGEDCAHFHP	DGRWNDDVCQ
hMGL_ISO1	248 WKWVDGTDYA	TGFQNWKPGQ	PDDWQGHGLG	GGEDCAHFHP	DGRWNDDVCQ
hMGL_ISO2	224 WKWVDGTDYA	TGFQNWKPGQ	PDDWQGHGLG	GGEDCAHFHP	DGRWNDDVCQ
Consensus	RPYHWVCEAG	LGQTSQESH			
Conservation	██████████	██████████			
hMGL_ISO3	271 RPYHWVCEAG	LGQTSQESH			
hMGL_ISO1	298 RPYHWVCEAG	LGQTSQESH			
hMGL_ISO2	274 RPYHWVCEAG	LGQTSQESH			

Figure 3. Map of hMGL isoforms. An insertion of 27 amino acids is present in the coiled-coil region of isoform 1 and an insertion of three amino acids is present in the CRD of isoform 2. The QPDxW sequence is unchanged, presumably the glycan selectivity among the three isoforms is identical but a comprehensive study is still missing.

The investigation of the oligomerisation state of the neck region using LOGICOIL²⁴ and COILS²⁵ suggests that these 27 amino acids do not form a coiled coil and disturb the coiled coil formation of the rest of the neck, figure 4.

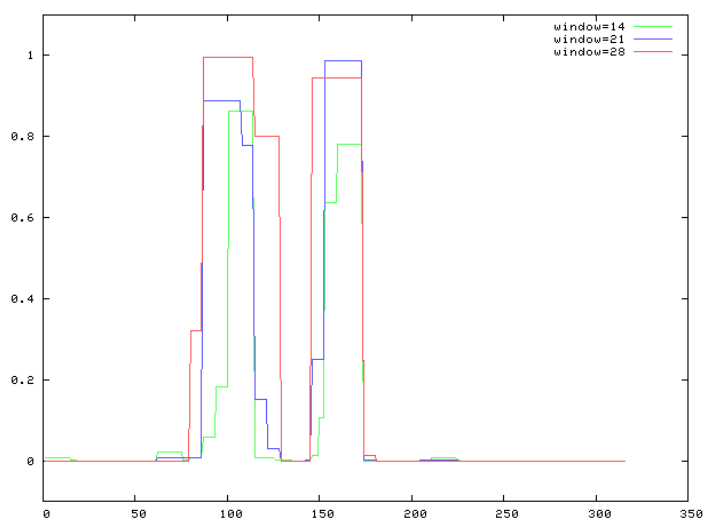


Figure 4. Prediction of coiled coil formation for the hMGL isoform 1. The probability of coiled coil formation drops to zero for the 27 amino acid insertion in isoform 1. On the x-axis: hMGL isoform 1 sequence as reported in figure 3. On y-axis: normalised similarity score. The image was generated with COILS²⁵, a program that compares a sequence to a database of known parallel two-stranded coiled-coils and derives a similarity score. Similar result was obtained with LOGICOIL²⁴, an algorithm which predicts the oligomeric state probabilities of a coiled-coil sequence. LOGICOIL currently allows to classify antiparallel dimers, parallel dimers, trimers and tetramers.

Expression in bacterial and mammalian cells and, of needed, refolding and purification of this isoform from inclusion body is extremely challenging due to this insertion (oral knowhow of the “MGL community”). Isoform2, cloned by macrophages,²⁶ has an insertion of 3 amino acids , Gly-Glu-Gly, in the N terminal region of the CRD while the coiled coil region is unaltered from isoform 3.

Is worth noting, that isoform 3 and 1 have extremely similar carbohydrate recognition profiles, according to their study with glycan arrays with monovalent ligands and they also show identical profiles in ELISA-based assays with multivalent presentation.²⁷ A comprehensive study which compares the binding profiles of all the human isoforms is still missing though. The different isoforms recognition profiles may differ in studies conducted on a more complex glycan matrix, for instance in histochemistry assay or *in vivo*.

1.4 Extracellular fragment oligomerisation

MGL oligomerisation state can follow two different paradigms: A) MGL could behave in a similar manner to mannose binding lectins, the latter having a short tail, in which case among the CRDs there are non-covalent interactions believed to be the main cause of oligomerisation. Other examples of plant lectins which follow paradigm A are Concanavalin A (ConA) and *Dioclea grandiflora* lectin (DGL) as shown in figure 26, chapter 2. Paradigm B), MGL could behave like other lectins in which the neck domain is fundamental to induce oligomerisation, such as DC-SIGN.²⁸

Cross linking experiments performed with the extracellular portion of isoform 3 purified by affinity chromatography on galactose-Sepharose, revealed the formation of a trimer cross-linked polypeptide with a molecular weight of ca 77 kDa.²⁰ The experiment was carried out using bissulfosuccinimidyl suberate cross-linking reagent. This molecule presents two N-hydroxysulfosuccinimide esters speared by a six-carbon aliphatic chain. The two ends of the molecule react with an SN₂ type mechanism with primary amines in the protein sequences, lysine side chains of N-terminus, which are spatially close to each other, so presumably located in the neck region. In the reaction the N-hydroxysulfosuccinimides act as the leaving groups to form a covalent bond. Furthermore, Jégouzo et al. expressed the amino acid sequence corresponding to the neck domain of isoform 3. Gel filtration, calorimetry and circular dichroism studies on the neck protein fragment, show that the neck domain function assembles into a coiled coil α -helical domain independently from the CRD. Gel-filtration analysis of the

neck domain at different concentration (50-400 $\mu\text{g/mL}$) revealed the presence of a large oligomer that eluted at an apparent molecular weight of 38 kDa. This value is larger than expected value corresponding to a trimer expected from the cross-linking analysis (38 kDa vs 29.8 kDa) but anyway supports the idea that the hMGL extracellular domain oligomerise through the neck domain.²⁰ In this scenario, three CRDs would behave as independent units and their orientation dynamically changes in response to the type of glycans presented to the lectin at the cell surface. On the other hand, gel filtration chromatography of the hMGL CRD didn't present peaks corresponding to an oligomer mass.

Another study which disclosed information about the oligomerisation state of the protein was performed by Napolitano et al.²⁹ When performing immune fluorescence experiments on iDC, it was noticed that hMGL is heterogeneously distributed on the plasma membrane. Only after addition of a selective ligand, in this specific case a MUC1 peptide bearing nine T_N antigen moieties and anti MGL antibody, did MGL cluster and show a spotted image. Western blot analysis of the immunoprecipitated complexes, hMGL-MUC1 peptide and hMGL-antibody, in non-reducing condition, revealed two bands corresponding to trimers and dimers.²⁹

Eventually, plate based binding study³⁰ and histochemical assays⁶, do not show a remarkable difference in the binding profiles of multivalent and monovalent ligands to the CRD and the extended CRD suggesting that the proteins behave in a similar fashion. Data reported in the literature suggests that hMGL extracellular domain is an oligomer in its native form and after binding multivalent ligands may clusters on the cell surface. Size exclusion analysis suggest that the CRD on its own is a monomer.

1.5 Glycan binding

Although we will see in section 1.5.2 that glycan orientation in the binding pocket can in some cases bend this rule, typically the QPDxW motif in lectins CRD is known to be the determinant for the binding of glycans with C-3 and C-4 hydroxyls respectively equatorial and axial.³¹ While the EPN motif is found in protein CRD which bind glycans with C-3 and C-4 hydroxyls both in equatorial position, figure 5.

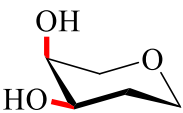

CRD characteristic sequence	C3/C4 Hydroxyl position	Glycan selectivity
Gln-Pro-Asp		Galactose GalNAc
Glu-Pro-Asn		Mannose Glucose L-Fucose GlcNAc

Figure 5. Structure of pyranose C-3 & C-4 hydroxyl groups equatorial & axial as well as in the equatorial equatorial arrangement. Lectins which contain the sequence QPDxW in the CRD will selectively bind glycans with C3 & C4 hydroxyl groups in equatorial/axial position as galactose or GalNAc, while lectins containing the sequence EPN will selectively bind glycans with C3/C4 hydroxyl groups in equatorial/equatorial position.

Since over 30 years is possible altering protein glycan selectivity by mutating the QPD sequence to EPN and vice versa.³¹ All human isoform of MGL, the two-mouse isoform and the rat MGL present the QPD sequence in their CRD, chapter 1, figure 3 and chapter 5 figure 20, and as expected bind Galactose derivatives.

1.5.1. Calcium binding sites

An indispensable component for the binding is the presence of a Ca^{2+} ion. All human isoforms of MGL are widely known to bind galactose derivatives in a calcium dependent manner.³² In the absence of calcium in solution or, after addition of calcium chelating agent, e.g. ethylenediaminetetraacetic acid (EDTA), the binding is lost. In galactose binding proteins, the selectivity determinant amino acids glutamine, Q, and aspartic acid, D, from the QPD sequence play a role in holding the calcium in place while the ligand coordinates via O-3 and O-4. For instance, figure 6 shows the calcium ion involved in the glycan binding of the asialoglycoprotein receptor ASGPR,³³ the closest cousin of MGL with 73% of sequence homology. Gln239 OE1 and Asp241 OD1 coordinate the calcium ion with Asn 264 OD1, Asp 265 OD1 and Glu252 OE2 (for amino acid nomenclature see chapter 4, figure 6).

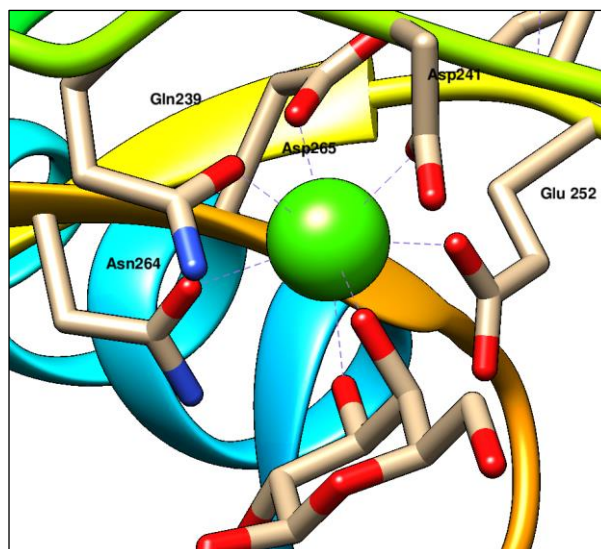


Figure 6. Calcium ion of Asialoglycoprotein receptor involved in the binding. PDB code 5JPV

The involvement of the ligands O-3 and O-4 in the coordination with the calcium ion, elucidate the affinity results arising from glycan array of the Consortium for Functional Glycomics;³⁴ GalNAc derivatives, where the 3-OH and 4-OH groups are removed or functionalised, lose their binding ability towards MGL.³⁴ On the other hand, the functionalisation of the 6-OH is tolerated. For instance hMGL binds sialyl-Tn (Neu5Aca2-6 GalNAca Ser/Thr) with a Kd of circa 179 μ M compared to a Kd 63 μ M of T_N antigen, and ST_N antigen in breast carcinoma.^{35,36}

Furthermore, in other C-type lectin, the calcium-ion coordination is believed to enhance the glycan CH polarisation and consequently enhance CH- π interaction between the bottom face of the sugar and aromatic residue Trp/Tyr/Phe, beneath it.³⁷ Molecular dynamic simulation (MD) of hMGL binding³⁸ reveal that CH- π interaction are possible between hydrogens on atoms C3, C4, C5 and C6³⁹ of the glycan and the Trp 281 of MGL only if the glycan assumes a specific orientation in the binding pocket. Possible binding orientations will be discussed in section 1.5.2.

The extracellular calcium-dependency of hMGL is clear: the physical presence of calcium is a prerequisite for the protein to be able to bind carbohydrates. The intracellular role of it in terms of signal transduction still needs to be largely explored. In a recent review, Cote et al described extensively the importance of calcium for the ligand endocytosis and signalling of the CLEC receptors.³² The receptor internalisation triggered by ligand binding event, allows the transfer of Ca²⁺ in the cytosol.

1.5.2. Binding mode of C-type lectins and galactoside derivative

Performing a search of C-type lectins crystal structures deposited on protein data bank (PDB), it is possible to evince that C-type lectins mainly bind galactoside derivatives with two possible orientations. These two binding modes were firstly defined by Marcelo et al. as binding mode A and binding mode B.³⁸ Protein sequence analysis reveals that the QPD sequence, not-only is the discriminant for galactoside selectivity, but also defines the orientation of the glycan in the binding pocket. In fact, galactose binding C-type lectins which have QPD sequence followed by an aromatic residue, typically bind in mode A. This glycan orientation is characterised by the presence of an aromatic residue at the “bottom” of the binding pocket, which permits CH- π interaction with the glycan, figure 7 left.

A number of lectins that recognise galactosidase but lack the QPD sequence and an aromatic residue in the binding pocket is also reported in the literature. In these proteins, the glycan has the C6 pointing 180° away from where, in binding mode A, the aromatic residue was located, figure 7 right.

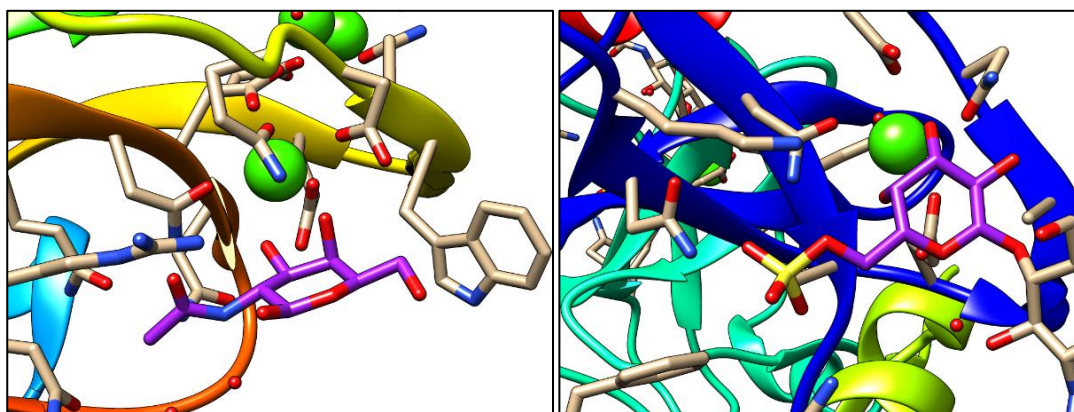


Figure 7. PDB data bank crystal structures of two different C-Type lectin binding modes. Left: binding mode A CEL-I complexed with GalNAc (PDB code 1WMZ). **Right:** a C-type lectin expressed by Langerhans cells, complexed with a sulfated galactoside (PDB code 3P5I).

A broad selection of examples of the two possible binding modes is reported by Marcelo et al.,³⁸ to cite two examples, CEL-I binds in mode A, while human Langerin binds in mode B, figure 7. Interestingly, human Langerin presents the EPN motif and typically binds glycans with C-3 and C-4 hydroxyls both in equatorial position. But in the presence of the sulfated galactoside in figure 8, the calcium coordination and binding are possible by a 180 ° turn of the glycan to fulfil the orientation of mode B. This orientation permits the creation of hydrogen bonds between the sulfate residue on C-6 and protein side chains, figure 8.

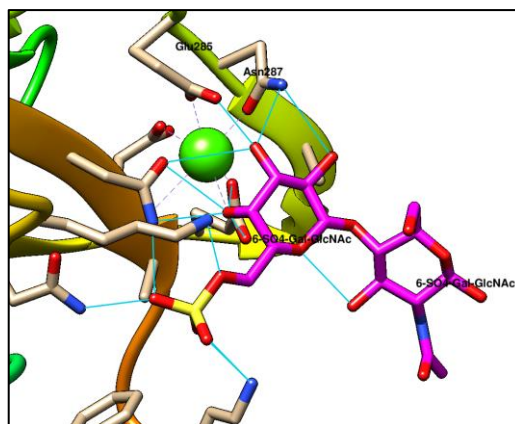


Figure 8 Human Langerin complexed with a sulfated galactoside (PDB code 3P5I).

In human Langerin specific case, it appears that the QPD and EPN rule is broken to permit the creation of strong protein/carbohydrate interaction.

Both binding modes allow coordination of 3-OH and 4-OH with the calcium ion. Understanding which mode is predominant for this specific protein-carbohydrate interaction is fundamental for ligand design.

1.5.3. Binding mode of hMGL: STD NMR and molecular dynamic simulation studies.

A fundamental contribution in understanding the binding mode of hMGL arises from Cañada and co-workers.³⁸ Saturation transfer difference (STD) NMR was used to investigate the protein-carbohydrate interaction in solution between hMGL and of methyl 1 alpha-D-galactopyranose, **FC1**, and methyl 2-(acetylamino)-2-deoxy-alpha-D-galactopyranose, **FC2**, figure 9. STD NMR method is an outstanding tool to explore how these carbohydrates interact with hMGL. The STD NMR method permits the identification of ligand protons that are in close contact with the protein. Protons involved in the binding receive a higher degree of saturation, and as a result stronger STD NMR signal. On the other hand, protons that are either less or not involved in the binding process reveal no STD NMR signals.⁴⁰ The experimental STD investigation revealed different signal pattern of **FC1** versus **FC2**, figure 9. To rationalise these differences, molecular dynamics (MD) simulations of the protein in the process of binding carbohydrates in binding mode A and B were run. The homology model of hMGL was based on the ASGPR structure coordinates. Selected simulation from molecular dynamics runs, were analysed with a program, which predicts the saturation of each proton depending on the geometry of the complex, Corcema-ST (complete relaxation and conformational exchange matrix-saturation transfer).⁴¹ This theoretical study revealed that while the **FC1** derivative STD signals could be explained solely by a binding mode A, a combination of binding mode A:B in

ratio 1:4 was necessary to rationalise the **FC2** signals. The two most populated H-bond obtained from the 20 ns of MD simulations in water at 298K in binding mode B correspond to **FC2** OH-4 with OE2 of Glu 280 and GalNAc OH-3 with OD2 of Asp 269 with 98% of population. In this scenario, a favourable CH- π interaction between the NAc methyl group and Trp281 at the bottom of the binding pocket could justify its high STD resonance.

While for binding mode A, **FC2** OH-3 with OE2 of Glu 280 100% population, GalNAc OH-4 and OH-6 with OD2 of Asp 269 with 58% and 54% respectively, Marcelo et al. Supporting information.³⁸

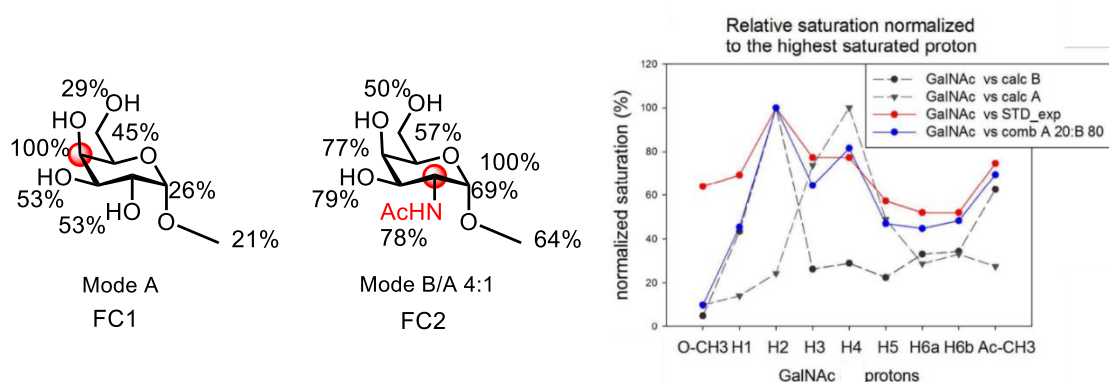


Figure 9. Left: STD NMR epitope mapping of methyl 1 alpha-D-galactopyranose and methyl 2-(acetylamino)-2-deoxy-alpha-D-galactopyranose. Right: graphical representation of the relative saturation of each proton of methyl 2-(acetylamino)-2-deoxy-alpha-D-galactopyranose for all four conditions, mode A, mode B, combination A and B (0,2A:0,8B) predicted with CORCEMA-ST versus experimental. Adapted with permission from Chem. - A Eur. J. 2014, 20, 16147–16155.³⁸

From the study reported by Cañada and co-workers it appears that despite, the hMGL sequence unites it with the binding mode A class of proteins, binding mode B is the most favourable and largely abundant in solution.

1.5.4. hMGL glycan selectivity

Human MGL has increased affinity for methyl 2-(acetylamino)-2-deoxy- α -D-galactopyranose versus methyl 1 α -D-galactopyranose of over 70-fold.³⁸ The structural reason of this preference was never investigated for hMGL, extensive studies were performed for the the asialo glycoprotein receptor (ASGPR), the closest cousin of MGL.⁴²

ASGPR, present this key sequence QPD in the CRD and it is known to have an exquisite selectivity for terminal α/β GalNAc derivatives including T_N antigen and its derivatives.¹⁰ ASPGR has increased affinity for GalNAc over Galactose of 14-folds compared to the 70-fold of hMGL.⁴² Mutagenesis studies revealed that for the ASGPR the key amino acid to enhance GalNAc selectivity is the His 256 (His 284 for hMGL homology). From crystallographic data it is possible to see that His 256 and His 202 in mutated version of mannose binding protein, have direct Van der Waals contact with the methyl group of the N-acetyl substituent of the sugar, figure 10.^{43,42,33}

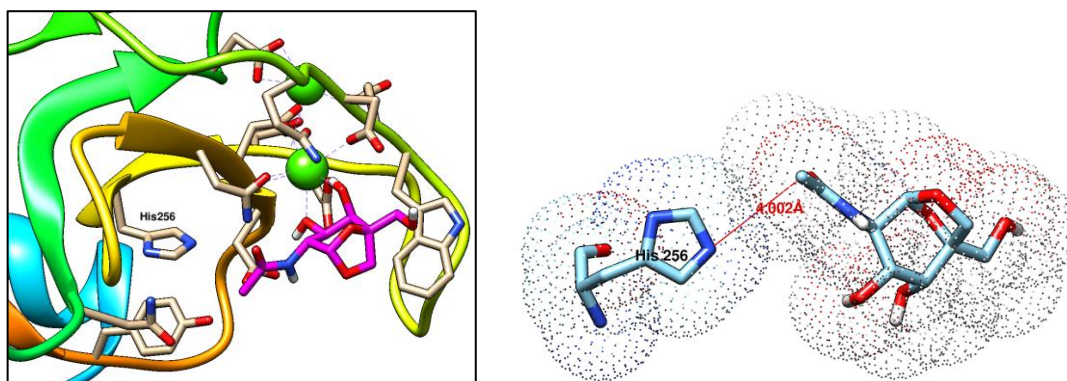


Figure 10. Carbohydrate binding site of ASGPR & 2.0 Å Van der Waals dot surface representation of the GalNAc binding site. On the left ligand in the binding pocket of ASGPR. On the right: Van der Waals contacts His 256 of ASGPR with GalNAc derivative (PDB code 5JQ1).

From the sequence comparison between MGL and ASGPR it is possible see that all human isoforms, figure 3, and mouse and rat MGL, figure 20 chapter 5, share this His residue with the ASGPR. The homologue histidine residue His 284, numbering from hMGLisoform3, could have similar Van der Waals interaction with the acetyl group, which would partially explain this selectivity difference. Although, the interaction with this His residue would be allowed only assuming a binding mode A. On the contrary if binding mode B is the most occurrent one, favourable CH- π interaction between the NAc methyl group and Trp281, numbering from hMGLisoform3, at the bottom of the binding pocket could justify the enhanced GalNAc selectivity.

Chapter1

The first crystal structure elucidation of hMGL CRD provided in chapter 5 offers a possible explanation of the binding mode of hMGL.

1.6 MGL expression

Depending on the construct, MGL size vary between 15 and 30 kDa. All isoform expressed in bacterial cells, that typically enables milligrams of protein to be obtained at a moderate cost, yield in misfolded aggregate of the protein. E. coli microenvironment differs from the mammalian cell one in terms of pH, cofactors, salts concentration etc.⁴⁴

Table 1. Overview of the most common conditions for MGL expression and purification. The isoform and length are indicated for each entry were possible.

Protein	Cell line	Vector	Purification	Denaturation	Refolding buffer	Reference
mMGL CRD	BL21(DE3) pLysS	pET8c	Galactose sepharose	NH ₄ OH 2M	RedGlu 1mM	Sato et al. ⁴⁵
					OxGlu 0.1mM	
					Maleic Acid pH 6.2 25 mM	
					CaCl ₂ 20mM	
					NaCl 500 mM	
PMSF 1mM						
					NaN ₃ 0.02%	
hMGL	BL21 (DE3)	pET8c	Galactose sepharose	NH ₄ OH 2M	MOPS pH 7 20 mM CaCl ₂ 20mM NaCl 500 mM	Suzuki et al. ²²
mMGL CRD Iso1	and BL21/DE3	pT5T	Galactose sepharose	Guanidine HCl Triton X-100 0.5%	Tris-HCl pH.8 25mM CaCl ₂ 25 mM NaCl 1.25 M	Iobst et al. ⁴³
hMGL	HEK293T	EBB	myc-tag	---	---	Nollau et al. ¹⁷
hMGL	NSO	---	Polyhistidine-tag	---	---	R and D Systems
hMGL extracellular	CHO	---	myc-tag	---	---	Mortezai et al. ³⁶
hMGLiso3 and CRD	CHO	pRc/ CMV and	Fc-tag			Van Vliet et al. ³⁴
hMGL extracellular	BL21 (DE3)	pET30b	GalNAc Gelfiltration	Guanidine 6M 0,01 % □mercaptoethanol	Tris-HC pH.8 25mM CaCl ₂ 100 mM NaCl 150m M	Maalej et al ⁴⁶

These variations can be problematic for protein self-folding. Additionally, for protein expression, protease-deficient host strains are used. These strains carry mutations which eliminate the production of proteases and yield in high level of protein expression. At high concentration, hydrophobic stretches in the protein are available for interaction with similar regions. All of these factors lead to protein aggregation as form of inclusion bodies.⁴⁷

Inclusion body formation can be a deterrent for inexperienced researchers and can considerably slow down the expression process. In general, finding appropriate refolding conditions mostly requires trial and error approaches and is therefore time-consuming. Furthermore, rapid and inexpensive analysis to check the folding pattern are not always available. Understanding which isoform was used in biological investigations can be challenging. Isoform 3 or the protein carbohydrate recognition domain are the most frequently used. The longest version of hMGL, isoform 1 presents more challenges in the folding procedure.

In table 1 the known experimental conditions for the expression, refolding and purification of human and mouse MGL are reported with the respective references.

It is worth mentioning that in bacterial expression posttranslational modification are not present and no glycans are present on the reported glycosylation sites Asn 78 and Asn 73.

1.7 How to target MGL: binding mode and selective ligands overview

1.7.1. Multivalency

As anticipated earlier in this chapter, hMGL binds selectively structures which terminate in α/β GalNAc and these include the T_N antigen and, LacdiNAc. A moderate preference for α derivatives over β was reported.⁵ However, the interaction with monovalent ligand is relatively weak, table 2. A common strategy to transform a monovalent ligand into a lead glycomimetic drug is to exploit multivalency⁴⁸, designing a molecule which presents multiple copies of the receptor-binding element which can simultaneously bind with the receptor. Multivalent molecules can present an higher avidity and specificity for hMGL this occur due to different binding mechanisms which are highly dependent from the ligand architecture.⁴⁹ Furthermore the CRD could oligomerize in physiological condition,²⁹ giving space to a multitude of possible binding modes.

Below we report possible binding mechanism of protein-multivalent ligand and 4 classes of compounds which were used to target MGL. The author does not intend to give a comprehensive overview of the state of art but wish to highlight structural features, which chemists can exploit to target MLG.

Table 2 . Summary of the binding affinity of hMGL with monovalent ligand

Entry	Ligand	Binding affinity	Biophysical Method	Reference
1	α Methyl GalNAc	$K_i = 12 \pm 1 \mu\text{M}$	STD NMR	Marcelo et al. ³⁸
3	Gal	$IC_{50} = 2500 \mu\text{M}$	ELISA	André et al. ³⁰
4	α Methyl Galactopyranoside	$K_d = 900 \pm 300 \mu\text{M}$	STD NMR	Marcelo et al. ³⁸
5	GalNAc β 1-4(6S)GlcNAc-Sp8	2.8 $FU_5 / FU_{GalNAc\beta-Sp8}$	Glycan array analysis	Jégouzo et al. ²⁰
6	Ser-T _N Antigen	$K_d = 63 \pm 13 \mu\text{M}$	SPR	Mortazeiet al. ³⁶
7	Neu5Gc T _N	$K_d = 179 \pm 65 \mu\text{M}$	STD NMR	Mortazeiet al. ³⁶

1.7.2. Chelate effect

When the geometry of the multivalent ligand can adequately match the of the geometry of the receptor, then the carbohydrates units of the ligand can bind simultaneously more than one CRD and this can lead to a higher functional affinity. The binding events which follow the first interaction between the ligand and receptor, are intramolecular and thus entropically more favoured. To exploit the chelate effect it is fundamental to understand the entropy role in polyvalent interactions.⁵⁰ An optimal multivalent ligand for hMGL would ideally have a scaffold with low flexibility, but which can meet the geometry requirements for the binding of three CRDs simultaneously, figure 11 b.

1.7.3. Receptor clustering/ Cross linking

If the ligand alters the receptor orientation, proximity or concentration on the cell membrane during the binding process, we are in the presence of a different multivalent binding mode called receptor clustering, figure 11 c. This binding mechanism differs from the chelate effect because it causes the movement of the receptor on the cell's membrane. MGL clustering, may have been observed by Napoletano et al. after treatment of DCs with a multivalent MUC1 glycopeptides bearing nine Tn antigens.²⁹ Receptor clustering mediate the initiation of signalling and ligand processing, for instance different internalisation pathways for ligands bearing different Tn antigen density was observed.⁹ Receptor-clustering was reported to be relevant for signalling to T cells⁵¹ and recently Kiessling and co-workers observed differential signalling and trafficking behaviour of polymers that target Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin (DC-SIGN),⁵² a C-type lectin receptor expressed in the same subset on DCs of MGL.³ The differential trafficking appears to be correlated with different receptor clustering induced by the polymers 3D shapes. Napoletano et al. observed that MUC1 derived peptides were presented by MHC-I or MHC-II in a peptide dependent manner²⁹ after internalisation by hMGL, receptor clustering could be the cause behind this. The oligomerisation of hMGL can allow multidimensional cross-linking polymer formation, figure 11 f. The cross-linking ability of the protein depends on the protein concentration at the cell surface.⁵⁰ To investigate a possible cross link mechanism several experimental designs are available.⁵³

1.7.4. Statistical effect or concentration effect

Ligand binding enhanced by statistical effect does not require rigorous structural features as for the chelate effect. In an intuitive fashion, the rebinding of a second multivalent ligand unit is favoured by the high local concentration of these units. For instance, the transitory encounter complex *i* in the model d) figure 11, could dissociate to give multivalent ligand and receptor unbound, or can proceed to the form *ii*. The later outcome has higher probability to happen due to the multivalent nature of the ligand and the high local concentration of the glycan unit.

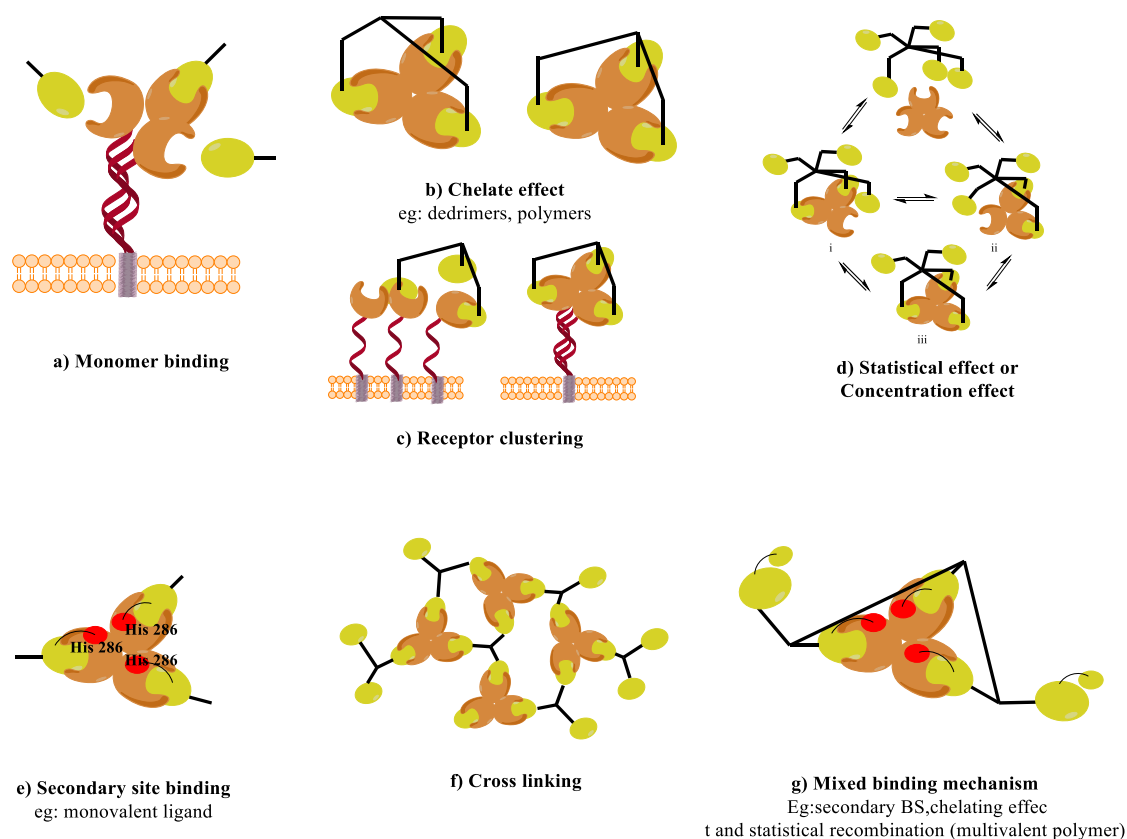


Figure 11. Possible mechanism of ligand binding with MGL: (a) Monovalent ligands bind a single CRD. Multivalent ligands possess multiple copies of a recognition element presented from the same scaffold, which allows a variety of additional mechanistic options. (b) Chelate effect requires the ligand to meet strict geometrical parameters, in case of hMGL triangle of ca 50Å (c) Statistical effects. Rebinding of the multivalent ligand is favoured by the high local concentration of binding elements. (d) Secondary binding site. His 286 is directly involved in hydrogen bonding the NAc residue and can interact with the backbone of MUC1 peptides bearing Tn Antigens. (e) Cross linking. Ligands can bind to a CRD of different trimers. (f) Mixed binding mechanism.

1.7.5. Subsite binding

Lectins are known to have enhanced interaction with ligand due to a secondary recognition site separated from the primary carbohydrate recognition site, figure 11e.⁵⁴ The ability His 286, numbering from hMGL isoform 3, to engage the backbone of T_N glycopeptides through water molecule was recently observed via microarray study and molecular dynamic simulation.²⁷ Furthermore, additional contact with the backbone of MUC1 glycopeptides was described.^{38,55}

1.7.6. Mixed binding mechanism

Due to the nature of multivalent ligands and the trimeric nature of the protein, it is possible to design a ligand which simultaneously target more than one mechanism at the time. For instance, the ligand reported in figure 11g, has the right geometry and an appropriate flexibility to chelate the protein, but also has more than three repetitive units which can lead to statistical recombination. Each single unit was also designed to exploit the interaction with the secondary binding site.

1.8 Multivalent inhibitors and effectors

The coupling of C-type lectin receptors, CLRs, to different signal transduction modules is influenced not only by the receptor but also by the nature, density, size, and architecture of the ligand, which can affect the rate of receptor internalization and trafficking to different intracellular compartments.⁹ CLRs in targeting strategies often relies on the use of individual monoclonal antibodies selective for these receptors. This strategy has several downsides: a) monoclonal antibodies are highly immunogenic – in vivo they can induce an immune reaction against themselves which may decrease the efficiency of the approach or induce severe autoimmune complications; b) they are not cost-effective; and c) they do not offer the structural tunability, which by contrast, is intrinsic to CLRs' natural ligands: glycans. In fact, the type of immune response is not only influenced by receptor interplay, but the antigen structure and nature, intracellular or extracellular, can determine a different immune outcome even when it engages a single receptor. Multivalent ligands designed to target MGL can have a higher affinity compared to a natural ligand, and so prevent the binding of this latter. In this case it is possible to say that these ligands work as inhibitors. Other ligands, the effectors, binds to the protein and can be endocytosed into the cells and may cause a signalling cascade, for brief immunology overview see chapter 2.

Glycans are indispensable for receptor recognition and binding but, apart from unique examples,⁵⁶ **carbohydrates on their own are not presented by the MHC-I or MHC-II complex on APCs**, and therefore can't evoke an immune response.⁵⁷

You will see shortly that a solution to this issue adopted by the scientific community is to conjugate carbohydrates on immunoenhancing elements like immunogenic peptides.

1.8.1. Small molecules and low number of recognition elements molecules

Synthetic scaffold with a maximum of four recognition elements were synthesised and tested for targeting hMGL by the Murphy group.^{6,5} The molecular weight of these clusters ranges from 900 Da to 2000Da.

Carbohydrates profiling with hMGL revealed a distinctive selectivity for GalNAc, mainly α .³⁴ The monovalent α **S-GalNAc** thioglycosides was found to have up to 50 times fold higher inhibitory power for hMGL CRD bound on crypt-associated cells of murine jejunum compared to **GalNAc**, figure 12.⁶ A fourfold preference for the α anomer over the β anomer was also found when comparing thioglycosides with the same core and valency.⁵ Thioglycosides are accepted by most biological systems, they are less susceptible to acid or enzymatic cleavage and due to the longer C-S bond, they can occupy different spaces in the binding pocket of MGL.^{58,59}

Due to its enhanced affinity and the easy access to the α anomer of GalNAc thioglycosides⁶⁰, mainly this sugar head was adopted in the synthesis of multivalent glycoclusters.

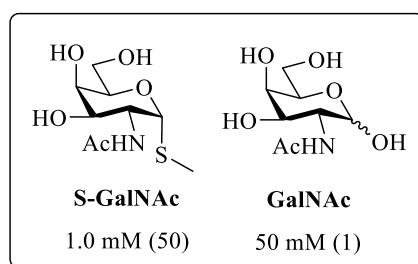


Figure 12. Structure and IC₅₀-values of monosaccharide in lectin histochemical assays with biotinylated lectins applied to sections of fixed murine jejunum, Crypt-associated cells. The relative affinity for MGL is reported in brackets.

A remarkable increase in affinity is observed for ligands having tetraphenyl ethylene (TPE) as central core. Compounds **6** and **2** share the same formal valency, but the TPE derivative **6** has inhibitory ability for the full domain hMGL up to 20 times higher than **2**. For the tetravalent compound **1**, the difference with the valency homologue **4** is up to 25 times. This might be due to the interaction of the ligand as aggregated form. TPE core decorated with different pyranoses can form a large aggregate, 20nm-150 nm, in the concentration range from 20 μ M to 540 μ M in aqueous media.⁶¹

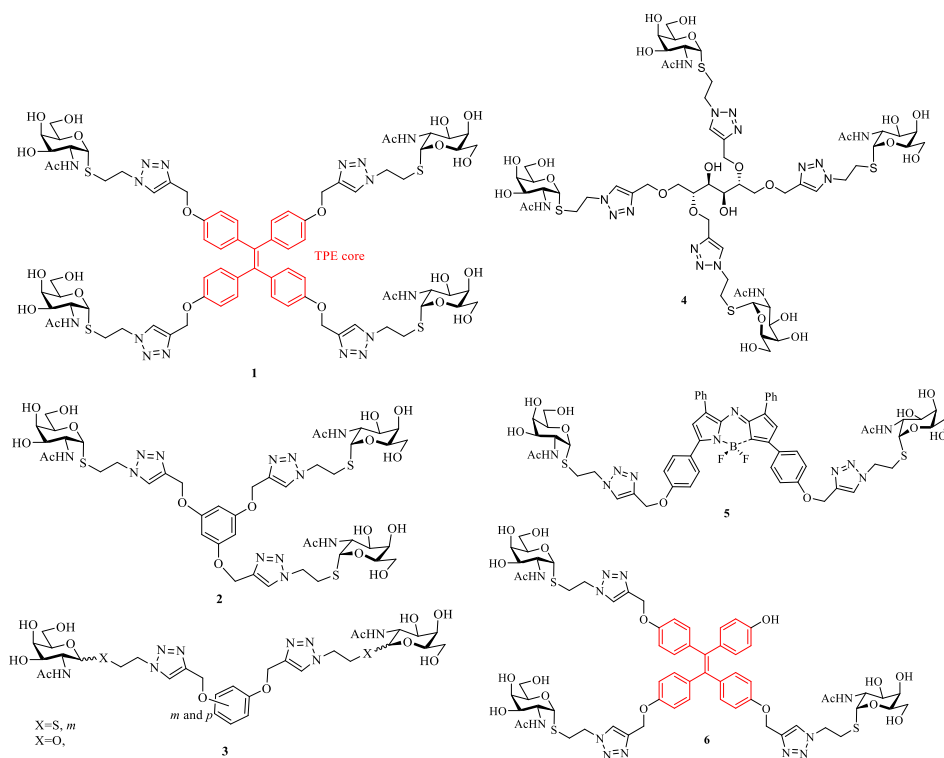


Figure 13. Structure of multivalent glycoclusters

1.8.2. Dendrimers

Dendrimers are repetitive branched molecules with a common core.

Leclerc and co-workers developed a dendrimer construct to target MGL⁶²: protein mimetic are combined to the geometrical requirements for a dendrimer structure to give multiple antigenic glycopeptide (MAG). These dendrimers are based on the (Lys)₂-Lys-Ala core, figure 14,⁶³ with multiple arms carrying a peptide epitope with T_N antigen. The MAG-T_N3 is based on a multiple antigenic peptide (MAP) and it was proved to induce IgG antibody production against the tumor associated T_N antigen in non-human primates, which were able to mediate antibody-dependent cytotoxicity against a panel of T_N antigen positive human tumour cell lines.⁶⁴ MAG-T_N3 is currently under evaluation as a therapeutic vaccine for breast cancer patients, in a phase I clinical trial.^{65,66}

Freire et al. show that GalNAc bearing dendrimers target mouse and human MGL on DCs and induce potent Th2 and B cell responses in mice.⁶⁷ Van Vliet et al. found that the combined triggering of TLR2 and MGL with dendrimer lysine molecules based on glycosylated CD4⁺T cell epitope from TT (sequence: QYIKANS*KFIGIT*EL) or poliovirus (sequence:

KLFAVWKIT*YKDT*) carbohydrates induce secretion of IL-10 and TNF- α by DCs.⁶⁸ In literature are present several dendrimers GalNAc functionalised which were never explored as hMGL ligands,⁶⁹ but proved to bind GalNAc selective plant lectins as SBA.⁷⁰ These dendrimers are based on the polyamidoamine (PAMAM) that is however known for its toxicity and ability to trigger nonspecific interaction,⁷¹ therefore less desirable when planning in vivo-applications.

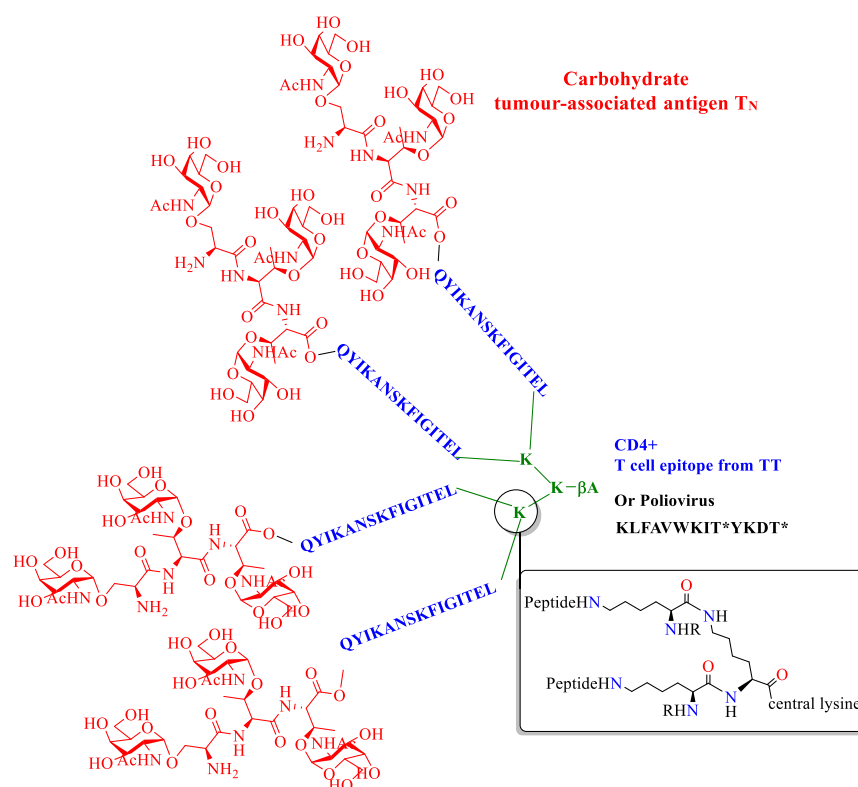


Figure 14. MAG 3 Tn Dendrimer structure base on (Lys)2-Lys-Ala core.

1.8.3. Polymers

In literature there are many GalNAc bearing polymers but only few of them were used to target hMGL.

The synthesis of glycopolymers can be done with two different approaches: 1) through the polymerisation of glycomonomers,^{72,73,74} or 2) through the post-polymerisation functionalisation of a scaffold polymer obtained from commercially available or easily obtainable monomers.^{75,76,77}

An example of the second class of polymers was used by Irimura and co-workers to demonstrate the participation of GalNAc residues in antigen uptake and presentation by bone marrow derived DCs on mMGL2 which lead to CD4⁺T cell activation. A polyacrylamide 60-mer (PAA) with β/α -GalNAc residues was used.⁷⁸ Binding and competition assay on DCs by flow cytometry analysis with the same GalNAc polymer, seems to suggest that MGL binding is only slightly influenced by the polymer density or pattern.⁴

The work of Tenaka et al is another example of post polymerisation functionalisation. The GalNAc insertion was performed on alkyne-functional polymer scaffold exploiting a copper catalysed cycloaddition reaction.^{7,79} Surface plasmon resonance analysis of the linear polymers with different degree of polymerisation (DP_n), revealed that increasing the linear chain length, only moderately lowers the overall K_D of the linear polymer with the lowest K_D registered of 169 nM for n=100.

The K_D was also largely unchanged when the architecture of the polymer was changed; for instance, **GP8** is a tetra-antennary dendrimeric polymer and **GP9** is an octa antennary polymer build on lactose scaffold with a K_D of 157 and 240 nM.

Is noteworthy that even if, in some cases, the binding is not strongly influenced by the architecture of the multivalent polymers, there are evidences that the geometry of the ligand influences the antigen processing.⁴

1.8.4. Peptide mimetic

Solid phase peptide synthesis allows to obtain large libraries of short peptides modifying glycosylation sites and density in short time. The author previously anticipated that the ligand geometry is fundamental, not only for targeting a specific binding mode, but also to modulate the type of immune response. We must keep in mind that the structure of the chosen peptide antigen will define the cellular response, after binding with MGL, the large T_N-bearing glycoproteins are trapped in an endolysosomal /HLAII compartment, while smaller glycopeptides colocalized both in HLAII and HLAI compartments⁹, figure 15.

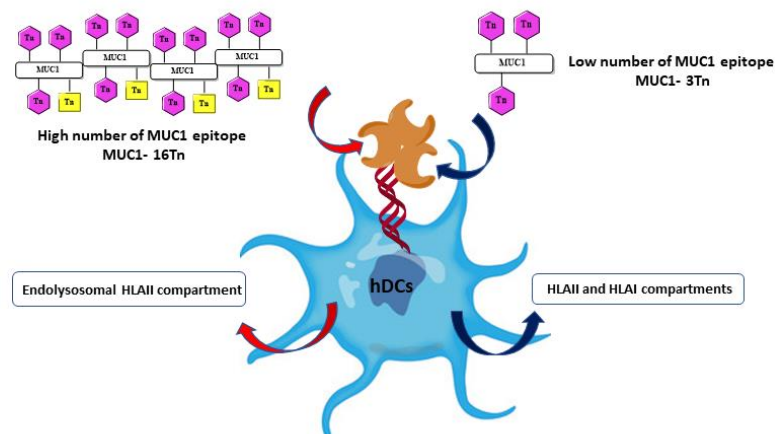


Figure 15. Different type of MGL induced immune responses: the ligand geometry modulates the type of immune response.

1.8.1.1 Mucin 1 (MUC1)

MUC1, HGVTSPDTRPAPGSTAPPA, is a heavy glycosylated protein located on epithelial cell surface, composed by a multiple tandem repeat of the same epitope. In one single tandem repeat there are five possible glycosylation sites on Ser and Thr, which, normally, are decorated with elongated O-glycans. These latter are truncated during malignant transformation and replaced by much shorter glycans, typically T antigen and its derivatives. Aberrant MUC1 glycoform carrying T_N antigens is often overexpressed in cancer cells, examples were reported for breast⁸⁰, pancreas^{81,82}, prostate,⁸³ ovarian⁸⁴ and colon carcinoma.⁸⁵ MGL can distinguish the glycosylation pattern of healthy mucins on epithelial cells from tumorigenic.⁸⁶ The ability of MGL to recognise this aberrant glycans offers an attractive opportunity to design glycomimetics of mucins protein to attempt to prime the immune system. Among all the member of the mucin family, MUC1 derivatives are, without doubt, the most studied as MGL ligand.

Napolitano et al. demonstrated that MGL activated by a MUC1_{9Tn}, a MUC1-60mer peptide (VTSAPDTRPAPGSTAPPAHG) n=3 enzymatically functionalised with 9 T_N antigen moieties, oligomerizes and subsequently activates the extracellular signal-regulated kinase 1,2 (ERK1,2) and nuclear factor- κ B(NF- κ B) pathways and induces phenotypic and functional DC maturation that then enables the DCs to initiate a strong CD8⁺ T-cell immune response.²⁹

Always Napolitano et al. confirmed that DCs were binding to the same MUC1 peptide MGL and this interaction favoured the T_N antigen processing in human histocompatibility leukocyte Ag (HLA) class I and II compartments.⁴

Also, a single MUC1 tandem repeat, was reported to bind hMGL. In 2017 Artigas et al. performed an extensive screening on MUC1 glycopeptides based on two sequences GVT*S*APDT*RPAPGS*T*APPAHGVT and PPAHGVT*S*APDT*RPAPGS*T*A with different grade of glycosylation on all the MGL orthologues: the two murine versions, mMGL1 and mMGL2, and the human one. The glycans used in this study are reported in figure 16. The human orthologues bind to mono-glycosylated peptides with similar avidity, while the inclusion of a GalNAc core at both the GVTS and GSTA regions highly enhance recognition.⁵⁵ Isoform 3 of hMGL seems to favour Thr-based Tn antigen compared to the Ser-linked T_N antigen.²⁷

Strikingly, as previously observed by Marcelo et al. through STD NMR, the glycopeptides carrying Tn antigen in the immunodominant PDTR region show positive binding contribution of the amino acids adjacent to the sugar moiety to the hMGL engagement.³⁸ Mono-glycosylate MUC1 peptide were reported having an inhibition constant in the low μM range (APDT*RPAPGS $K_i=35\pm 5 \mu\text{M}$, HGVT SAPDT*RPAPGSTAPPA $K_i=23\pm 2 \mu\text{M}$, determined by STD-NMR at 298 K).³⁸

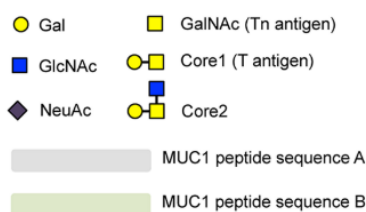


Figure 16. Glycans used to decorate the two MUC1 sequences by Artigas et al.

The peptide $\beta\text{Ala-GVTS*APDTRPAPGST*APPAHGVT-NH}_2$ was reported to bind to CD1c⁺ DCs and to enhance the cytokines secretion through TLR 7/8.¹²

Atomic force microscopy³⁵ and optical tweezer⁸⁷ were used to quantify intermolecular interactions between MGL and breast cancer associate MUC1T_N/ST_N/ST glycopetides. These studies highlight, once again, that MGL tolerate functionalisation at the C6 but requires mandatory C3/C4 hydroxyls in equatorial/axial conformation: intermolecular bonds between MGL and MUC1(T_N) or MUC1(ST_N) were observed, but not with MUC1(ST).⁸⁷ The mutation of His 284 into a Thr residue, translates in a general loss in affinity for MUC1 peptides, MD calculation suggest its interaction with MUC1 backbone.²⁷

A less studied member of the mucin family in relationship of MGL binding is Mucin 2 (MUC2). Microarray study of MUC2 peptides, Ac-PTTTPLK-NH₂, showed that hMGL isoform 3 prefers

di-glycosylated peptides over mono and three, with remarkable increase in fluorescence for T*T* or T*XT*. ⁸⁸

1.8.1.2 Not glycosylated peptides

Recently Eggink et al. used a branched non-glycosylated peptide to develop an innovative immunotherapeutic strategy for ovarian cancer.¹ The peptides **sv6D** and **svL4** were reported to bind in a calcium dependent manner with MGL, K_D of $0.15 \pm 0.02 \mu\text{M}$ and $0.24 \pm 0.04 \mu\text{M}$ respectively, figure 17. These peptides present branched structures based dendrimeric lysin, see figure 14. Docking study and competition study with GalNAc-PAA suggest that the peptides bind to MGL through the same glycan binding site. In vivo study of a murine ovarian cancer model, showed that sv6D effectively suppressed development of ascites both as monotherapy and in combination with the chemotherapeutic drug paclitaxel or the immunotherapeutic antibody against the receptor PD-1.¹

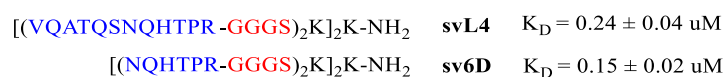


Figure 17. Peptide mimetic structure. In black the dendrimeric lysine core, in red glycine reach sequence used as spacer, in blue the GalNAc mimetic epitope. In sv6D the c-terminal sequence was truncated without affinity lost.

1.9 Concluding remarks

MGL has the ability to trigger an immune response by ligand binding, internalisation and presentation. This receptor can be targeted to achieve an immunotherapeutic response and vaccine development.¹

The type of response initiated is highly dependent of the ligands structure. A lot of questions still need to be answered on this topic. A deeper understanding of the existing correlation between ligand chemical structure and type of immune response initiated by the receptor, can arise tremendous progress in selective drugs design.

Synthetic organic chemistry is an excellent resource to fine tune ligand geometry and glycan density. The variety of binding mechanism which can be exploited to improve protein-ligand affinity allows to work in an enormous chemical space.

1.10 Bibliography

- (1) Eggink, L. L.; Roby, K. F.; Cote, R.; Kenneth Hooper, J. An Innovative Immunotherapeutic Strategy for Ovarian Cancer: CLEC10A and Glycomimetic Peptides. *J. Immunother. Cancer* **2018**, *6* (1), 1–16. <https://doi.org/10.1186/s40425-018-0339-5>.
- (2) Ward, S. E.; O’Sullivan, J. M.; Drakeford, C.; Aguila, S.; Jondle, C. N.; Sharma, J.; Fallon, P. G.; Brophy, T. M.; Preston, R. J. S.; Smyth, P.; et al. A Novel Role for the Macrophage Galactose-Type Lectin Receptor in Mediating von Willebrand Factor Clearance. *Blood* **2018**, *131* (8), 911–916. <https://doi.org/10.1182/blood-2017-06-787853>.
- (3) van Vliet, S. J.; Gringhuis, S. I.; Geijtenbeek, T. B. H.; van Kooyk, Y. Regulation of Effector T Cells by Antigen-Presenting Cells via Interaction of the C-Type Lectin MGL with CD45. *Nat. Immunol.* **2006**, *7* (11), 1200–1208. <https://doi.org/10.1038/ni1390>.
- (4) Napoletano, C.; Rughetti, A.; Agervig Tarp, M. P.; Coleman, J.; Bennett, E. P.; Picco, G.; Sale, P.; Denda-Nagai, K.; Irimura, T.; Mandel, U.; et al. Tumor-Associated Tn-MUC1 Glycoform Is Internalized through the Macrophage Galactose-Type C-Type Lectin and Delivered to the HLA Class I and II Compartments in Dendritic Cells. *Cancer Res.* **2007**, *67* (17), 8358–8367. <https://doi.org/10.1158/0008-5472.CAN-07-1035>.
- (5) André, S.; O’Sullivan, S.; Koller, C.; Murphy, P. V.; Gabius, H. J. Bi- to Tetraivalent Glycoclusters Presenting GlcNAc/GalNAc as Inhibitors: From Plant Agglutinins to Human Macrophage Galactose-Type Lectin (CD301) and Galectins. *Org. Biomol. Chem.* **2015**, *13* (14), 4190–4203. <https://doi.org/10.1039/c5ob00048c>.
- (6) Kaltner, H.; Manning, J. C.; García Caballero, G.; Di Salvo, C.; Gabba, A.; Romero-Hernández, L. L.; Knospe, C.; Wu, D.; Daly, H. C.; O’Shea, D. F.; et al. Revealing Biomedically Relevant Cell and Lectin Type-Dependent Structure-Activity Profiles for Glycoclusters by Using Tissue Sections as an Assay Platform. *RSC Adv.* **2018**, *8* (50), 28716–28735. <https://doi.org/10.1039/c8ra05382k>.
- (7) Tanaka, J.; Gleinich, A. S.; Zhang, Q.; Whitfield, R.; Kempe, K.; Haddleton, D. M.; Davis, T. P.; Perrier, S.; Mitchell, D. A.; Wilson, P. Specific and Differential Binding of N-Acetylgalactosamine Glycopolymers to the Human Macrophage Galactose Lectin and Asialoglycoprotein Receptor. *Biomacromolecules* **2017**, *18* (5), 1624–1633. <https://doi.org/10.1021/acs.biomac.7b00228>.
- (8) van Vliet, S. J.; Saeland, E.; van Kooyk, Y. Sweet Preferences of MGL: Carbohydrate Specificity and Function. *Trends Immunol.* **2008**, *29* (2), 83–90. <https://doi.org/10.1016/j.it.2007.10.010>.
- (9) Zizzari, I. G.; Napoletano, C.; Battisti, F.; Rahimi, H.; Caponnetto, S.; Pierelli, L.; Nuti, M.; Rughetti, A. MGL Receptor and Immunity: When the Ligand Can Make the Difference. *J. Immunol. Res.* **2015**, *2015*, 1–8. <https://doi.org/10.1155/2015/450695>.
- (10) van Kooyk, Y.; Ilarregui, J. M.; van Vliet, S. J. Novel Insights into the Immunomodulatory Role of the Dendritic Cell and Macrophage-Expressed C-Type Lectin MGL. *Immunobiology* **2015**, *220* (2), 185–192. <https://doi.org/10.1016/j.imbio.2014.10.002>.
- (11) Cornelissen, L. A. M.; Van Vliet, S. J. A Bitter Sweet Symphony: Immune Responses to Altered o-Glycan Epitopes in Cancer. *Biomolecules*. 2016, pp 1–19. <https://doi.org/10.3390/biom6020026>.
- (12) Heger, L.; Balk, S.; Lühr, J. J.; Heidkamp, G. F.; Lehmann, C. H. K.; Hatscher, L.; Purbojo, A.; Hartmann, A.; Garcia-Martin, F.; Nishimura, S. I.; et al. CLEC10A Is a Specific Marker for Human CD1c+dendritic Cells and Enhances Their Toll-like Receptor 7/8-Induced Cytokine Secretion. *Front. Immunol.* **2018**, *9* (APR), 1–16. <https://doi.org/10.3389/fimmu.2018.00744>.
- (13) Takada, A.; Fujioka, K.; Tsuiji, M.; Morikawa, A.; Higashi, N.; Ebihara, H.; Kobasa, D.; Feldmann, H.; Irimura, T.; Kawaoka, Y. Human Macrophage C-Type Lectin Specific for Galactose and N-Acetylgalactosamine Promotes Filovirus Entry. *J. Virol.* **2004**, *78* (6), 2943–2947. <https://doi.org/10.1128/JVI.78.6.2943>.
- (14) Van Vliet, S. J.; Steeghs, L.; Bruijns, S. C. M.; Vaezirad, M. M.; Blok, C. S.; Arenas Busto, J. A.; Deken, M.; Van Putten, J. P. M.; Van Kooyk, Y. Variation of Neisseria Gonorrhoeae Lipooligosaccharide Directs Dendritic Cell-Induced T Helper Responses. *PLoS Pathog.* **2009**, *5* (10). <https://doi.org/10.1371/journal.ppat.1000625>.
- (15) Klaver, E. J.; Kuijk, L. M.; Laan, L. C.; Kringel, H.; van Vliet, S. J.; Bouma, G.; Cummings, R. D.; Kraal, G.; Van Die, I. Trichuris Suis-Induced Modulation of Human Dendritic Cell Function Is Glycan-Mediated. *Int. J. Parasitol.* **2013**, *43* (3–4), 191–200. <https://doi.org/10.1016/j.ijpara.2012.10.021>.
- (16) van Liempt, E.; van Vliet, S. J.; Engering, A.; García Vallejo, J. J.; Bank, C. M. C.; Sanchez-Hernandez, M.; van Kooyk, Y.; van Die, I. Schistosoma Mansonii Soluble Egg Antigens Are Internalized by Human Dendritic Cells through Multiple C-Type Lectins and Suppress TLR-Induced Dendritic Cell Activation. *Mol. Immunol.* **2007**, *44* (10), 2605–2615. <https://doi.org/10.1016/j.molimm.2006.12.012>.
- (17) Nollau, P.; Wolters-Eisfeld, G.; Mortezaei, N.; Kurze, A. K.; Klampe, B.; Debus, A.; Bockhorn, M.; Niendorf, A.; Wagener, C. Protein Domain Histochemistry (PDH): Binding of the Carbohydrate Recognition Domain (CRD) of

- Recombinant Human Glycoreceptor CLEC10A (CD301) to Formalin-Fixed, Paraffin-Embedded Breast Cancer Tissues. *J. Histochem. Cytochem.* **2013**, *61* (3), 199–205. <https://doi.org/10.1369/0022155412474823>.
- (18) Brett, D.; Pospisil, H.; Valcárcel, J.; Reich, J.; Bork, P. Alternative Splicing and Genome Complexity. *Nat. Genet.* **2002**, *30* (1), 29–30. <https://doi.org/10.1038/ng803>.
- (19) Andreadis, A.; Gallego, M. E.; Bernardo, N. GENERATION OF PROTEIN ISOFORM DIVERSITY BY ALTERNATIVE SPLICING: Mechanistic and Biological Implications. *Annu. Rev. Cell Biol.* **1987**, No. 3, 207–242.
- (20) Jégouzo, S. A.; Quintero-Martínez, A.; Ouyang, X.; Dos Santos, Á.; Taylor, M. E.; Drickamer, K. Organization of the Extracellular Portion of the Macrophage Galactose Receptor: A Trimeric Cluster of Simple Binding Sites for N-Acetylgalactosamine. *Glycobiology* **2013**, *23* (7), 853–864. <https://doi.org/10.1093/glycob/cwt022>.
- (21) Vogel, C.; Marcotte, E. M. Insights into Regulation of Pr1. Vogel, C. & Marcotte, E. M. Insights into Regulation of Protein Abundance from Proteomics and Transcriptomics Analyses. *Nat. Rev. Genet.* **13**, 227–232 (2013). Otein Abundance from Proteomics and Transcriptomics Analyses. *Nat. Rev. Genet.* **2013**, *13* (4), 227–232. <https://doi.org/10.1038/nrg3185>. Insights.
- (22) Suzuki, N.; Yamamoto, K.; Toyoshima, S.; Osawa, T.; Irimura, T. Molecular Cloning and Expression of CDNA Encoding Human Macrophage C-Type Lectin. Its Unique Carbohydrate Binding Specificity for Tn Antigen. *J. Immunol.* **1996**, *156* (1), 128–135.
- (23) Valladeau, J.; Duvert-Frances, V.; Pin, J.-J.; Kleijmeer, M. J.; Ait-Yahia, S.; Ravel, O.; Vincent, C.; Vega, F.; Helms, A.; Gorman, D.; et al. Immature Human Dendritic Cells Express Asialoglycoprotein Receptor Isoforms for Efficient Receptor-Mediated Endocytosis. *J. Immunol.* **2014**, *167* (10), 5767–5774. <https://doi.org/10.4049/jimmunol.167.10.5767>.
- (24) Vincent, T. L.; Green, P. J.; Woolfson, D. N. LOGICOIL - Multi-State Prediction of Coiled-Coil Oligomeric State. *Bioinformatics* **2013**, *29* (1), 69–76. <https://doi.org/10.1093/bioinformatics/bts648>.
- (25) Lupas, A.; Van Dyke, M.; Stock, J. Predicting Coiled Coils from Protein Sequences. *Science* (80-.). **1991**, *252* (5009), 1162–1164. <https://doi.org/10.1126/science.252.5009.1162>.
- (26) Higashi, N.; Fujioka, K.; Denda-Nagai, K.; Hashimoto, S. I.; Nagai, S.; Sato, T.; Fujita, Y.; Morikawa, A.; Tsuiji, M.; Miyata-Takeuchi, M.; et al. The Macrophage C-Type Lectin Specific for Galactose/N-Acetylgalactosamine Is an Endocytic Receptor Expressed on Monocyte-Derived Immature Dendritic Cells. *J. Biol. Chem.* **2002**, *277* (23), 20686–20693. <https://doi.org/10.1074/jbc.M202104200>.
- (27) Marcelo, F.; Supekar, N.; Corzana, F.; van der Horst, J. C.; Vuist, I. M.; Live, D.; Boons, G.-J. P. H.; Smith, D. F.; van Vliet, S. J. Identification of a Secondary Binding Site in Human Macrophage Galactose-Type Lectin by Microarray Studies: Implications for the Molecular Recognition of Its Ligands. *J. Biol. Chem.* **2018**, *jbc.RA118.004957*. <https://doi.org/10.1074/jbc.RA118.004957>.
- (28) dos Santos, Á.; Hadjivasiliou, A.; Ossa, F.; Lim, N. K.; Turgut, A.; Taylor, M. E.; Drickamer, K. Oligomerization Domains in the Glycan-Binding Receptors DC-SIGN and DC-SIGNR: Sequence Variation and Stability Differences. *Protein Sci.* **2017**, *26* (2), 306–316. <https://doi.org/10.1002/pro.3083>.
- (29) Napoletano, C.; Zizzari, I. G.; Rughetti, A.; Rahimi, H.; Irimura, T.; Clausen, H.; Wandall, H. H.; Belleudi, F.; Bellati, F.; Pierelli, L.; et al. Targeting of Macrophage Galactose-Type C-Type Lectin (MGL) Induces DC Signaling and Activation. *Eur. J. Immunol.* **2012**, *42* (4), 936–945. <https://doi.org/10.1002/eji.201142086>.
- (30) André, S.; O’Sullivan, S.; Koller, C.; Murphy, P. V.; Gabius, H. J. Bi- to Tetraivalent Glycoclusters Presenting GlcNAc/GalNAc as Inhibitors: From Plant Agglutinins to Human Macrophage Galactose-Type Lectin (CD301) and Galectins. *Org. Biomol. Chem.* **2015**, *13* (14), 4190–4203. <https://doi.org/10.1039/c5ob00048c>.
- (31) Drickamer, K. Engineering Galactose-Binding Activity into a C-Type Mannose-Binding Protein. *Nature* **1992**, *360* (6400), 183–186. <https://doi.org/10.1038/360183a0>.
- (32) Cote, R.; Lynn Eggink, L.; Kenneth Hooper, J. CLEC Receptors, Endocytosis and Calcium Signaling. *AIMS Allergy Immunol.* **2017**, *1* (4), 207–231. <https://doi.org/10.3934/allergy.2017.4.207>.
- (33) Sanhueza, C. A.; Baksh, M. M.; Thuma, B.; Roy, M. D.; Dutta, S.; Préville, C.; Chrnyk, B. A.; Beaumont, K.; Dullea, R.; Ammirati, M.; et al. Efficient Liver Targeting by Polyvalent Display of a Compact Ligand for the Asialoglycoprotein Receptor. *J. Am. Chem. Soc.* **2017**, *139* (9), 3528–3536. <https://doi.org/10.1021/jacs.6b12964>.
- (34) van Vliet, S. J.; van Liempt, E.; Saeland, E.; Aarnoudse, C. A.; Appelmelk, B.; Irimura, T.; Geijtenbeek, T. B. H.; Blixt, O.; Alvarez, R.; van Die, I.; et al. Carbohydrate Profiling Reveals a Distinctive Role for the C-Type Lectin MGL in the Recognition of Helminth Parasites and Tumor Antigens by Dendritic Cells. *Int. Immunol.* **2005**, *17* (5), 661–669. <https://doi.org/10.1093/intimm/dxh246>.
- (35) Beatson, R.; Maurstad, G.; Picco, G.; Arulappu, A.; Coleman, J.; Wandell, H. H.; Clausen, H.; Mandel, U.; Taylor-Papadimitriou, J.; Sletmoen, M.; et al. The Breast Cancer-Associated Glycoforms of MUC1, MUC1-Tn and Sialyl-Tn, Are Expressed in COSMC Wild-Type Cells and Bind the C-Type Lectin MGL. *PLoS One* **2015**, *10* (5), 1–21.

- <https://doi.org/10.1371/journal.pone.0125994>.
- (36) Mortezaei, N.; Behnken, H. N.; Kurze, A. K.; Ludewig, P.; Buck, F.; Meyer, B.; Wagener, C. Tumor-Associated Neu5Ac-Tn and Neu5Gc-Tn Antigens Bind to C-Type Lectin CLEC10A (CD301, MGL). *Glycobiology* **2013**, *23* (7), 844–852. <https://doi.org/10.1093/glycob/cwt021>.
 - (37) Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N. Carbohydrate-Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **2015**, *137* (48), 15152–15160. <https://doi.org/10.1021/jacs.5b08424>.
 - (38) Marcelo, F.; Garcia-Martin, F.; Matsushita, T.; Sardinha, J.; Coelho, H.; Oude-Vrielink, A.; Koller, C.; André, S.; Cabrita, E. J.; Gabius, H. J.; et al. Delineating Binding Modes of Gal/GalNAc and Structural Elements of the Molecular Recognition of Tumor-Associated Mucin Glycopeptides by the Human Macrophage Galactose-Type Lectin. *Chem. - A Eur. J.* **2014**, *20* (49), 16147–16155. <https://doi.org/10.1002/chem.201404566>.
 - (39) Spiwok, V. CH/ π Interactions in Carbohydrate Recognition. *Molecules* **2017**, *22* (7), 1–11. <https://doi.org/10.3390/molecules22071038>.
 - (40) Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chemie - Int. Ed.* **1999**, *38* (12), 1784–1788. [https://doi.org/10.1002/\(SICI\)1521-3773\(19990614\)38:12<1784::AID-ANIE1784>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q).
 - (41) Moseley, H. N. B.; Curto, E. V.; Krishna, N. R. Complete Relaxation and Conformational Exchange Matrix (CORCEMA) Analysis of NOESY Spectra of Interacting Systems; Two-Dimensional Transferred NOESY. *Journal of Magnetic Resonance, Series B.* 1995, pp 243–261. <https://doi.org/10.1006/jmrb.1995.1129>.
 - (42) Kolatkar, A. R.; Leung, A. K.; Isecke, R.; Brossmer, R.; Drickamer, K.; Weis, W. I. Mechanism of N-Acetylgalactosamine Binding to a C-Type Animal Lectin Carbohydrate-Recognition Domain. *J. Biol. Chem.* **1998**, *273* (31), 19502–19508. <https://doi.org/10.1074/jbc.273.31.19502>.
 - (43) Iobst, S. T.; Drickamer, K. Selective Sugar Binding to the Carbohydrate Recognition Domains of the Rat Hepatic and Macrophage Asialoglycoprotein Receptors. *J. Biol. Chem.* **1996**, *271* (12), 6686–6693. <https://doi.org/10.1074/jbc.271.12.6686>.
 - (44) Rosano, G. L.; Ceccarelli, E. A. Recombinant Protein Expression in Escherichia Coli: Advances and Challenges. *Front. Microbiol.* **2014**, *5* (APR), 1–17. <https://doi.org/10.3389/fmicb.2014.00172>.
 - (45) Sato, M.; Kawakami, K.; Osawa, T.; Toyoshima, S. Molecular Cloning and Expression of cDNA Encoding a Galactose/ N-Acetylgalactosamine-Specific Lectin on Mouse Tumoricidal Macrophages. *J. Biochem.* **1992**, *111* (3), 331–336. <https://doi.org/10.1093/oxfordjournals.jbchem.a123758>.
 - (46) Maalej, M. M.; Forgione, R. E.; Marchetti, R.; Bulteau, F. B.; Thepaut, M. T.; Lanzetta, R.; Laguri, C.; Simorre, J.; Fieschi, F.; Molinaro, A.; et al. The Human Macrophage Galactose-type Lectin, MGL, Recognizes the Outer Core of E. Coli Lipooligosaccharide. *ChemBioChem* **2019**, cbic.201900087. <https://doi.org/10.1002/cbic.201900087>.
 - (47) Hartley, D. L.; Kane, J. F. Properties of Inclusion Bodies from Recombinant Escherichia Coli. *Biochem. Soc. Trans.* **1988**, *16* (2), 101–102. <https://doi.org/10.1042/bst0160101>.
 - (48) Ernst, B.; Magnani, J. L. From Carbohydrate Leads to Glycomimetic Drugs. *Nat. Rev. Drug Discov.* **2009**, *8* (8), 661–677. <https://doi.org/10.1038/nrd2852>.
 - (49) Kiessling, L. L.; Gestwicki, J. E.; Strong, L. E. Synthetic Multivalent Ligands in the Exploration of Cell-Surface Interactions. *Current Opinion in Chemical Biology.* 2000, pp 696–703. [https://doi.org/10.1016/S1367-5931\(00\)00153-8](https://doi.org/10.1016/S1367-5931(00)00153-8).
 - (50) Mammen, M.; Choi, S.; Whitesides, G. M. ChemInform Abstract: Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *ChemInform* **2010**, *30* (9), no-no. <https://doi.org/10.1002/chin.199909293>.
 - (51) Gaus, K. T Cell Receptor Clustering - A Mechanism of Signal Transduction. *Biophys. J.* **2017**, *112* (3), 37a. <https://doi.org/10.1016/j.bpj.2016.11.235>.
 - (52) Jarvis, C. M.; Zwick, D. B.; Grim, J. C.; Alam, M. M.; Prost, L. R.; Gardiner, J. C.; Park, S.; Zimdars, L. L.; Sherer, N. M.; Kiessling, L. L. Antigen Structure Affects Cellular Routing through DC-SIGN. *Proc. Natl. Acad. Sci.* **2019**, *116* (30), 14862–14867. <https://doi.org/10.1073/pnas.1820165116>.
 - (53) Dam, T. K.; Brewer, C. F. Carbohydrate-Lectin Cross-Linking Interactions: Structural, Thermodynamic, and Biological Studies. *Methods Enzymol.* **2003**, *362* (1989), 455–486. [https://doi.org/10.1016/S0076-6879\(03\)01031-0](https://doi.org/10.1016/S0076-6879(03)01031-0).
 - (54) Schubert Wright, C. Structural Comparison of the Two Distinct Sugar Binding Sites in Wheat Germ Agglutinin Isolectin II. *J. Mol. Biol.* **1984**, *178* (1), 91–104. [https://doi.org/10.1016/0022-2836\(84\)90232-8](https://doi.org/10.1016/0022-2836(84)90232-8).
 - (55) Artigas, G.; Monteiro, J. T.; Hinou, H.; Nishimura, S. I.; Lepenies, B.; Garcia-Martin, F. Glycopeptides as Targets for

- Dendritic Cells: Exploring MUC1 Glycopeptides Binding Profile toward Macrophage Galactose-Type Lectin (MGL) Orthologs. *J. Med. Chem.* **2017**, *60* (21), 9012–9021. <https://doi.org/10.1021/acs.jmedchem.7b01242>.
- (56) Avci, F. Y.; Kasper, D. L. How Bacterial Carbohydrates Influence the Adaptive Immune System. *Annu. Rev. Immunol.* **2010**, *28* (1), 107–130. <https://doi.org/10.1146/annurev-immunol-030409-101159>.
- (57) Wilson, R. M.; Danishefsky, S. J. A Vision for Vaccines Built from Fully Synthetic Tumor-Associated Antigens: From the Laboratory to the Clinic. *J. Am. Chem. Soc.* **2013**, *135* (39), 14462–14472. <https://doi.org/10.1021/ja405932r>.
- (58) Doyle, L. M.; O’Sullivan, S.; Di Salvo, C.; McKinney, M.; McArdle, P.; Murphy, P. V. Stereoselective Epimerizations of Glycosyl Thiols. *Org. Lett.* **2017**, *19* (21), 5802–5805. <https://doi.org/10.1021/acs.orglett.7b02760>.
- (59) Witczak, Z. J.; Chhabra, R.; Chen, H.; Xie, X. Q. Thiosugars II. A Novel Approach to Thiodisaccharides. The Synthesis of 3-Deoxy-4-Thiocellobiose from Levoglucosenone. *Carbohydr. Res.* **1997**, *301* (3–4), 167–175. [https://doi.org/10.1016/S0008-6215\(97\)00100-6](https://doi.org/10.1016/S0008-6215(97)00100-6).
- (60) Knapp, S.; Myers, D. S. α -GlcNAc Thioconjugates. *J. Org. Chem.* **2001**, *66* (10), 3636–3638. <https://doi.org/10.1021/jo010088e>.
- (61) Donnier-Maréchal, M.; Abdullayev, S.; Bauduin, M.; Pascal, Y.; Fu, M.-Q.; He, X.-P.; Gillon, E.; Imberty, A.; Kipnis, E.; Dessein, R.; et al. Tetraphenylethylene-Based Glycoclusters with Aggregation-Induced Emission (AIE) Properties as High-Affinity Ligands of Bacterial Lectins. *Org. Biomol. Chem.* **2018**, *16* (45), 8804–8809. <https://doi.org/10.1039/C8OB02035C>.
- (62) Freire, T.; Zhang, X.; Dériaud, E.; Ganneau, C.; Vichier-Guerre, S.; Azria, E.; Launay, O.; Lo-Man, R.; Bay, S.; Leclerc, C. Glycosidic Tn-Based Vaccines Targeting Dermal Dendritic Cells Favor Germinal Center B-Cell Development and Potent Antibody Response in the Absence of Adjuvant. *Blood* **2010**, *116* (18), 3526–3536. <https://doi.org/10.1182/blood-2010-04-279133>.
- (63) Tam, J. P. Synthetic Peptide Vaccine Design: Synthesis and Properties of a High-Density Multiple Antigenic Peptide System (Solid-Phase Peptide Synthesis/Peptide Antigen/Antipeptide Antibody). *Biochemistry* **1988**, *85* (August), 5409–5413.
- (64) Lo-man, R.; Vichier-guerre, S.; Perraut, R.; De, E.; Benmohamed, L.; Diop, O. M.; Livingston, P. O.; Bay, S.; Leclerc, C. A Fully Synthetic Therapeutic Vaccine Candidate Targeting Carcinoma-Associated Tn Carbohydrate Antigen Induces Tumor-Specific Antibodies in Nonhuman Primates. *Cancer Res.* **2004**, No. 29, 4987–4994.
- (65) Ganneau, C.; Simenel, C.; Emptas, E.; Courtiol, T.; Coïc, Y. M.; Artaud, C.; Dériaud, E.; Bonhomme, F.; Delepierre, M.; Leclerc, C.; et al. Large-Scale Synthesis and Structural Analysis of a Synthetic Glycopeptide Dendrimer as an Anti-Cancer Vaccine Candidate. *Org. Biomol. Chem.* **2017**, *15* (1), 114–123. <https://doi.org/10.1039/C6OB01931E>.
- (66) Laubretton, D.; Bay, S.; Sedlik, C.; Artaud, C.; Ganneau, C.; Dériaud, E.; Viel, S.; Puaux, A. L.; Amigorena, S.; Gérard, C.; et al. The Fully Synthetic MAG-Tn3 Therapeutic Vaccine Containing the Tetanus Toxoid-Derived TT830-844 Universal Epitope Provides Anti-Tumor Immunity. *Cancer Immunol. Immunother.* **2016**, *65* (3), 315–325. <https://doi.org/10.1007/s00262-016-1802-0>.
- (67) Freire, T.; Zhang, X.; Dériaud, E.; Ganneau, C.; Vichier-Guerre, S.; Azria, E.; Launay, O.; Lo-Man, R.; Bay, S.; Leclerc, C. Glycosidic Tn-Based Vaccines Targeting Dermal Dendritic Cells Favor Germinal Center B-Cell Development and Potent Antibody Response in the Absence of Adjuvant. *Blood* **2010**, *116* (18), 3526–3536. <https://doi.org/10.1182/blood-2010-04-279133>.
- (68) van Vliet, S. J.; Bay, S.; Vuist, I. M.; Kalay, H.; Garcia-Vallejo, J. J.; Leclerc, C.; van Kooyk, Y. MGL Signaling Augments TLR2-Mediated Responses for Enhanced IL-10 and TNF- Secretion. *J. Leukoc. Biol.* **2013**, *94* (2), 315–323. <https://doi.org/10.1189/jlb.1012520>.
- (69) Liu, J.; Zhou, J.; Luo, Y. siRNA Delivery Systems Based on Neutral Cross-Linked Dendrimers. *Bioconjug. Chem.* **2012**, *23* (2), 174–183. <https://doi.org/10.1021/bc200433s>.
- (70) Thomas, B.; Pifferi, C.; Daskhan, G. C.; Fiore, M.; Berthet, N.; Renaudet, O. Divergent and Convergent Synthesis of GalNAc-Conjugated Dendrimers Using Dual Orthogonal Ligations. *Org. Biomol. Chem.* **2015**, *13* (47), 11529–11538. <https://doi.org/10.1039/c5ob01870f>.
- (71) Labieniec-Watala, M.; Watala, C. PAMAM Dendrimers: Destined for Success or Doomed to Fail? Plain and Modified PAMAM Dendrimers in the Context of Biomedical Applications. *Journal of Pharmaceutical Sciences.* 2015, pp 2–14. <https://doi.org/10.1002/jps.24222>.
- (72) Albertin, L.; Stenzel, M. H.; Barner-Kowollik, C.; Foster, L. J. R.; Davis, T. P. Well-Defined Diblock Glycopolymers from RAFT Polymerization in Homogeneous Aqueous Medium. *Macromolecules* **2005**, *38* (22), 9075–9084. <https://doi.org/10.1021/ma051310a>.
- (73) Albertin, L.; Kohlert, C.; Stenzel, M.; Foster, J. R.; Davis, T. P. Chemoenzymatic Synthesis of Narrow-Polydispersity Glycopolymers: Poly(6-O-Vinyladipolyl-D-Glucopyranose). *Biomacromolecules* **2004**, *5* (2), 255–260.

- <https://doi.org/10.1021/bm034199u>.
- (74) Parry, A. L.; Clemson, N. A.; Ellis, J.; Bernhard, S. S. R.; Davis, B. G.; Cameron, N. R. “Multicopy Multivalent” Glycopolymer-Stabilized Gold Nanoparticles as Potential Synthetic Cancer Vaccines. *J. Am. Chem. Soc.* **2013**, *135* (25), 9362–9365. <https://doi.org/10.1021/ja4046857>.
 - (75) Zhang, Q.; Anastasaki, A.; Li, G. Z.; Haddleton, A. J.; Wilson, P.; Haddleton, D. M. Multiblock Sequence-Controlled Glycopolymers via Cu(0)-LRP Following Efficient Thiol-Halogen, Thiol-Epoxy and CuAAC Reactions. *Polym. Chem.* **2014**, *5* (12), 3876–3883. <https://doi.org/10.1039/c4py00320a>.
 - (76) Basuki, J. S.; Esser, L.; Duong, H. T. T.; Zhang, Q.; Wilson, P.; Whittaker, M. R.; Haddleton, D. M.; Boyer, C.; Davis, T. P. Magnetic Nanoparticles with Diblock Glycopolymer Shells Give Lectin Concentration-Dependent MRI Signals and Selective Cell Uptake. *Chem. Sci.* **2014**, *5* (2), 715–726. <https://doi.org/10.1039/c3sc52838c>.
 - (77) Godula, K.; Bertozzi, C. R. Synthesis of Glycopolymers for Microarray Applications via Ligation of Reducing Sugars to a Poly(Acryloyl Hydrazide) Scaffold. *J. Am. Chem. Soc.* **2010**, *132* (29), 9963–9965. <https://doi.org/10.1021/ja103009d>.
 - (78) Denda-Nagai, K.; Aida, S.; Saba, K.; Suzuki, K.; Moriyama, S.; Oo-puthinan, S.; Tsuiji, M.; Morikawa, A.; Kumamoto, Y.; Sugiura, D.; et al. Distribution and Function of Macrophage Galactose-Type C-Type Lectin 2 (MGL2/CD301b). *J. Biol. Chem.* **2010**, *285* (25), 19193–19204. <https://doi.org/10.1074/jbc.M110.113613>.
 - (79) Slavin, S.; Burns, J.; Haddleton, D. M.; Becer, C. R. Synthesis of Glycopolymers via Click Reactions. In *European Polymer Journal*; Elsevier Ltd, 2011; Vol. 47, pp 435–446. <https://doi.org/10.1016/j.eurpolymj.2010.09.019>.
 - (80) Zaretsky, J. Z.; Barnea, I.; Aylon, Y.; Gorivodsky, M.; Wreschner, D. H.; Keydar, I. MUC1 Gene Overexpressed in Breast Cancer: Structure and Transcriptional Activity of the MUC1 Promoter and Role of Estrogen Receptor Alpha (ER α) in Regulation of the MUC1 Gene Expression. *Mol. Cancer* **2006**, *5*, 1–14. <https://doi.org/10.1186/1476-4598-5-57>.
 - (81) Hinoda, Y.; Ikematsu, Y.; Horinouchi, M.; Sato, S.; Yamamoto, K.; Nakano, T.; Fukui, M.; Suehiro, Y.; Hamanaka, Y.; Nishikawa, Y.; et al. Increased Expression of MUC1 in Advanced Pancreatic Cancer. *J. Gastroenterol.* **2003**, *38* (12), 1162–1166. <https://doi.org/10.1007/s00535-003-1224-6>.
 - (82) Moniaux, N.; Andrianifahanana, M.; Brand, R. E.; Batra, S. K. Multiple Roles of Mucins in Pancreatic Cancer, a Lethal and Challenging Malignancy. *British Journal of Cancer.* 2004, pp 1633–1638. <https://doi.org/10.1038/sj.bjc.6602163>.
 - (83) O’Connor, J. C.; Julian, J.; Lim, S. D.; Carson, D. D. MUC1 Expression in Human Prostate Cancer Cell Lines and Primary Tumors. *Prostate Cancer Prostatic Dis.* **2005**, *8* (1), 36–44. <https://doi.org/10.1038/sj.pcan.4500762>.
 - (84) Deng, J.; Wang, L.; Chen, H.; Li, L.; Ma, Y.; Ni, J.; Li, Y. The Role of Tumour-Associated MUC1 in Epithelial Ovarian Cancer Metastasis and Progression. *Cancer and Metastasis Reviews.* 2013, pp 535–551. <https://doi.org/10.1007/s10555-013-9423-y>.
 - (85) Kim, Y. S.; Yang, U.-S.; Frieria, A.; Itzkowitz, S.; Kjeldsen, T.; Hakomori, S.-I. Expression of Tn, Sialosyl Tn, and T Antigens in Human Pancreas. *Gastroenterology* **2017**, *100* (6), 1691–1700. [https://doi.org/10.1016/0016-5085\(91\)90671-7](https://doi.org/10.1016/0016-5085(91)90671-7).
 - (86) Saeland, E.; Van Vliet, S. J.; Bäckström, M.; Van Den Berg, V. C. M.; Geijtenbeek, T. B. H.; Meijer, G. A.; Van Kooyk, Y. The C-Type Lectin MGL Expressed by Dendritic Cells Detects Glycan Changes on MUC1 in Colon Carcinoma. *Cancer Immunol. Immunother.* **2007**, *56* (8), 1225–1236. <https://doi.org/10.1007/s00262-006-0274-z>.
 - (87) Hadjialirezaei, S.; Picco, G.; Beatson, R.; Burchell, J.; Stokke, B. T.; Sletmoen, M. Interactions between the Breast Cancer-associated MUC1 Mucins and C-Type Lectin Characterized by Optical Tweezers. *PLoS One* **2017**, *12* (4), 1–20. <https://doi.org/10.1371/journal.pone.0175323>.
 - (88) Marcelo, F.; Supekar, N.; Corzana, F.; Van Der Horst, J. C.; Vuist, I. M.; Live, D.; Boons, G. J. P. H.; Smith, D. F.; Van Vliet, S. J. Identification of a Secondary Binding Site in Human Macrophage Galactose-Type Lectin by Microarray Studies: Implications for the Molecular Recognition of Its Ligands. *J. Biol. Chem.* **2019**, *294* (4), 1300–1311. <https://doi.org/10.1074/jbc.RA118.004957>.

2 Synthesis of multivalent ligands based on tetraphenyl ethylene geometry

2.1 Abstract

In the following chapter, the synthesis and the derivatisation of multivalent glycoclusters based on the tetraphenylethylene (TPE) scaffold will be described. Previous research from the Murphy group,^{1,2} showed that, among several multivalent ligands, the tetraphenylethylene scaffold, when presenting multiple glycan residues had exquisite ability to enhance the protein-multivalent carbohydrate affinity.^{1,2} In this project were investigated: i) the degree of multivalency as well as geometrical requirements when using tetraphenylethylene scaffolds to target lactose, N-acetyl galactosamine (GalNAc) and mannose binding lectins; ii) the development of a facile, scalable and reproducible synthesis of multivalent glycoclusters; iii) the binding affinity of various GalNAc derivatives based on TPE for hMGL. A selection of the synthesised glycoclusters were tested in the laboratories of Prof. Hans-Joachim Gabius^{3,4} and in the laboratory of Dr. Sandra van Vliet (manuscript in preparation). Furthermore, the thermodynamics of the binding interaction for the glycoclusters with hMGL extracellular domain isoform3 and hMGL CRD were investigated using isothermal titration calorimetry (ITC).

2.2 Introduction

Based on results already published by the Murphy group carried out in collaboration with the group of Hans-Joachim Gabius, the initial focus of this project was to investigate how the valency of the GalNAc based glycoclusters influences their ability to target lectins. More specifically, the tetravalent tetraphenylethylene derivative **AG143**, figure 1, was previously synthesised by Dr. Shane O'Sullivan and proved to have high binding affinity, IC_{50} 10 nM, over 400 000 times higher than GalNAc, for human macrophage galactose C-type lectin (h-MGL).¹

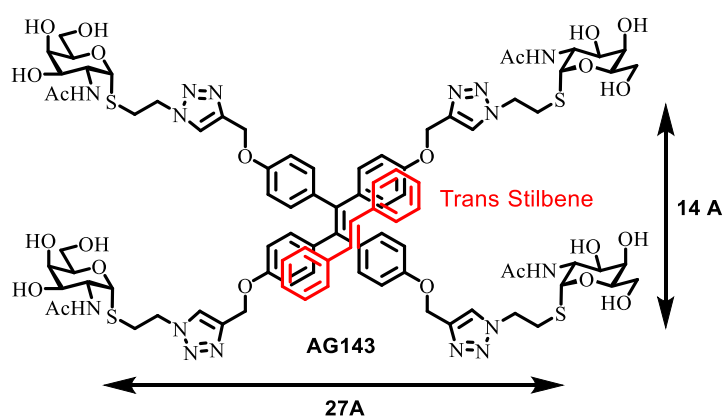


Figure 1. Chemical structure of tetraphenylethylene based multivalent GalNAc: estimated distances between anomeric carbon of the GalNAc residues in extended conformations of glycoclusters. The diagonal distance is ~30Å. In red the divalent trans stilbene geometry.

The libraries of molecule tested previously by the Murphy group revealed that the enhanced affinity for hMGL, as expected, did not scale proportionally with the increasing number of glycan epitopes, and that tetraphenylethylene scaffold has an ideal combination of valency and geometry to ensure high binding affinity for human and plant lectins over other glycocluster. This trend was observed not only with **AG143**, but also with the tetravalent GlcNAc analogue, which binds wheat germ agglutinin (WGA) almost 9000 times better than the monovalent ligand GlcNAc, while trivalent and divalent GlcNAc derivatives bind the same protein between 900 and 1600 times better.² In the published investigation, to exclude the possibility that the tetraphenylethylene scaffold and its appendages, aside from the carbohydrate residues are responsible for a non-specific interaction with a protein subsite, the tetravalent glycocluster **AG143** was tested not only with (WGA), which is GlcNAc selective, but also with *Dolichos Biflorus* agglutinin (DBA), which is GalNAc

selective.¹ The absence of cross reactivity between the two plant lectins, ensures that the binding arises by the carbohydrate headgroup

In regards of hMGL targeting, more recent studies from the group, showed that the α -S-GalNAc residue is a better sugar headgroup to target hMGL. In fact, methyl 1, 2-(acetylamino)-2-deoxy-1-thio- α -D-galactopyranose (**1**) has 50-fold increased affinity for hMGL compared to regular D-GalNAc.³ Up to fivefold preference for the α anomer over the β anomer was also found when comparing thioglycosides **SOS132** with the same core and valency, figure 2.¹

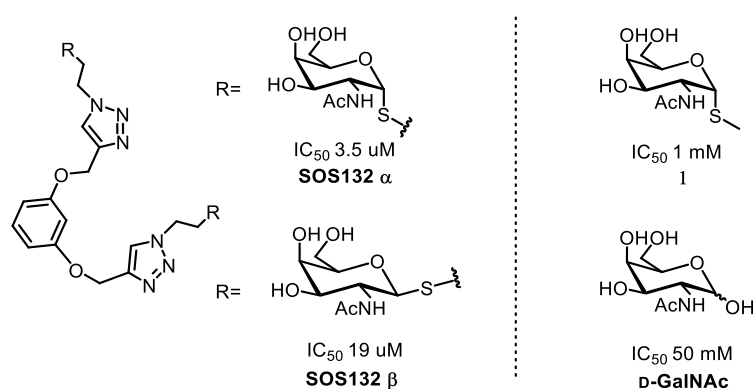


Figure 2. IC_{50} values of different ligands. On the left: SOS132 α has over 5-fold increase affinity for hMGL compared to the β analogue: IC_{50} were obtained for blocking binding of biotinylated extracellular domain hMGL to surface-immobilized (neo)glycoprotein. On the right **1** has 50-fold increase affinity for hMGL compared to GalNAc, the IC_{50} were obtained in lectin histochemical assays with biotinylated hMGL carbohydrate recognition domain applied to sections of fixed murine jejunum of Crypt-associated cells.

2.3 Results and discussion

With these aspects in mind, I planned the synthesis of a small collection of multivalent glycoclusters, figure 3, based on the tetraphenylethylene scaffold geometry. The synthesis of the divalent compounds based on stilbene scaffold was also planned. The divalent molecules could be potentially as potent as the tetravalent one due to its geometrical characteristics, and their synthesis requires less steps, a feature which makes it a more appealing target for the chemical industry.

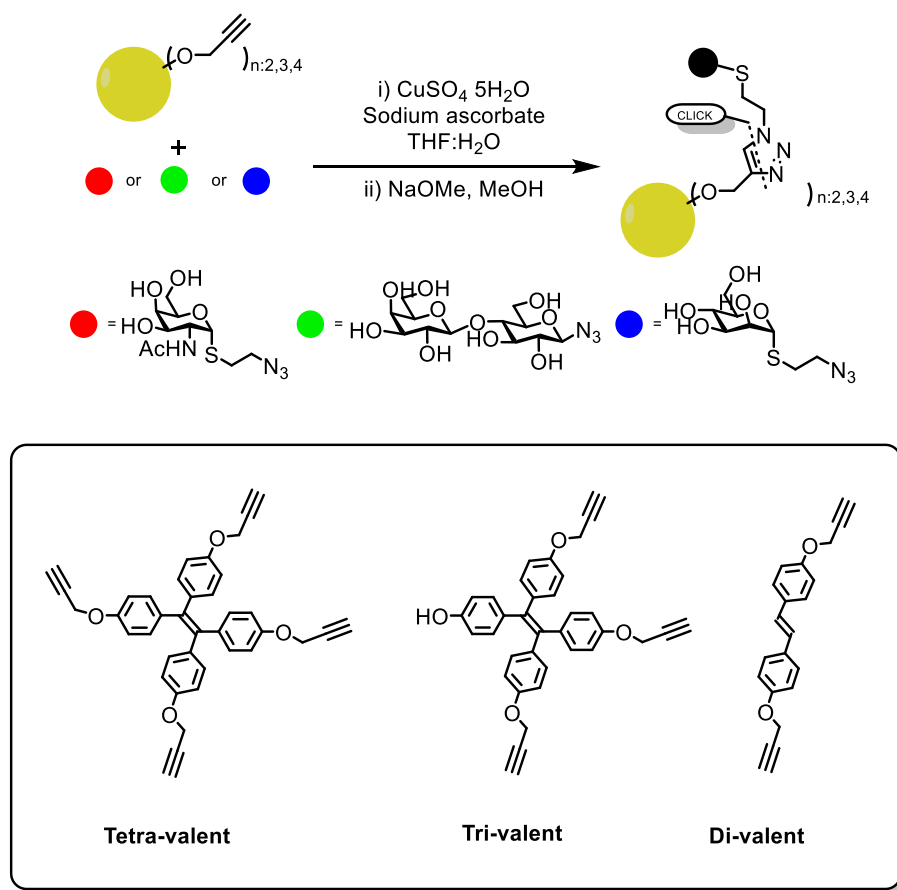


Figure 3. Chemical structure of multivalent glycoclusters: tetravalent, trivalent and divalent propargylated scaffolds were reacted with GalNAc, lactose and mannose azides via the regioselective copper catalysed alkyne azide cycloaddition.

The synthetic approach used, involves the connection of two synthons via a regioselective 1,3 dipolar cycloaddition to give a 1,2,3 triazole, figure 4. For safety reasons, the azide functional group was located on the monovalent synthon, the sugar moiety, while the propargyl group on the multivalent core.

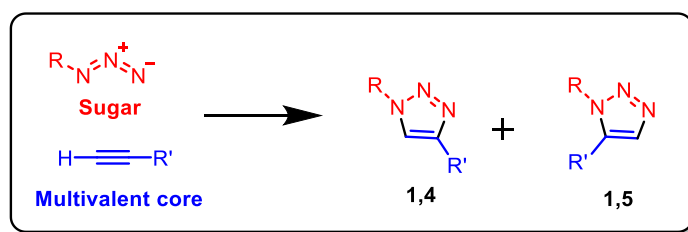


Figure 4. A schematic overview of 1,3 dipolar cycloaddition between an azide and a terminal alkyne: the chosen catalyst determines the regioisomer product. A Cu(I) catalyst typically gives the 1,4 isomers, while a Ru based catalyst gives the 1,5.

Copper promoted alkyne azide 1,3 dipolar cycloaddition,⁵ also called a CLICK reaction,⁶ is massively employed in total synthesis due to the fact it gives the desired product in high yield, it is scalable, reproducible and usually not substrate dependent. It can be carried out in the presence of water and the modulation of the catalyst identity allows a high degree of regioselectivity. All these features are particularly true for the 1,4 regioisomer which is obtained with a Cu(I) catalyst. The combination of CuSO₄ and sodium ascorbate was chosen. In fact, the reducing agent always guarantees the presence of Cu(I) and gives reproducible results independently from the amount of oxygen present in the solvent which is temperature dependent.

Another interesting aspect in the choice of a triazole ring formation during the synthesis, it that triazole is known as bio-isostere of the amide functional group, figure 5.^{7, 8}

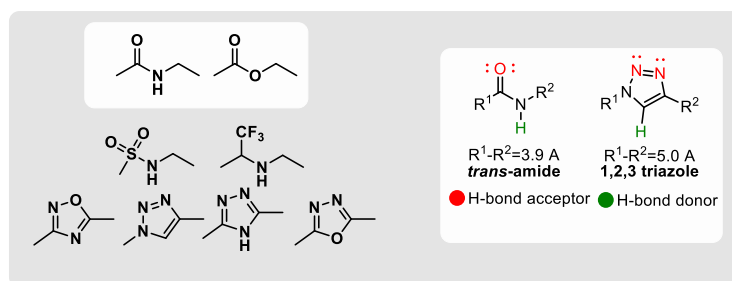


Figure 5. Bioisosterism between 1,4 tiazole and trans amide: on the left groups which mimic amide and ester functional group, on the right geometrical characteristic of trans-amide and 1,2,3 triazole.

2.3.1. Synthesis

Multivalent alkynes

The synthesis of the dialkyne **AG089** started with the reaction of 4-methoxybenzaldehyde with commercially available diethyl (4-methoxybenzyl)phosphonate⁹, to give **AG085** as a single trans-isomer in 92% yield (MP 211–213°C). The Z isomer is an oil at room temperature and the two isomers are reported to have distinct NMR chemical shift: the alkenyl hydrogens of the E isomers are located more downfield, 6.63 ppm in deuterated chloroform, compared to the one of the Z isomer, 6.45 ppm, figure 6 and 7.^{10,11}

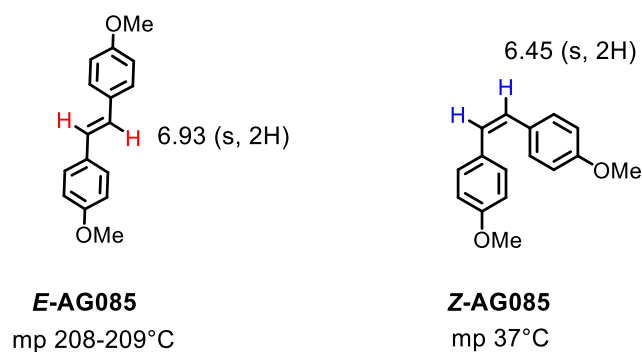


Figure 6. Structural and physical data for E- and Z- AG085 reported in literature.

Proton NMR spectra were recorded in deuterated chloroform on a 500 MHz instrument.

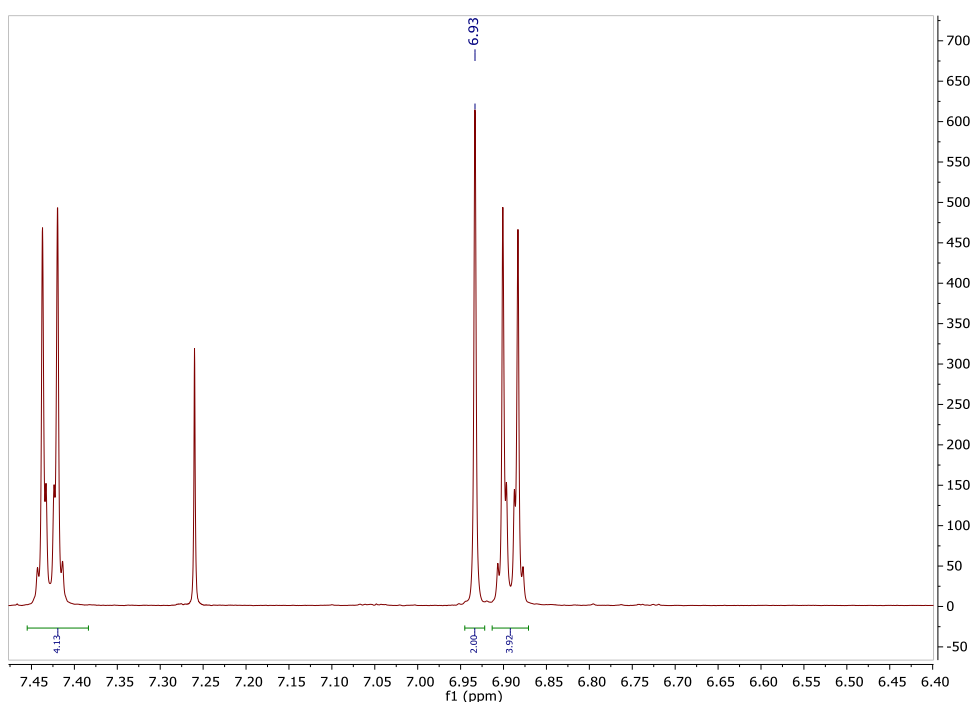
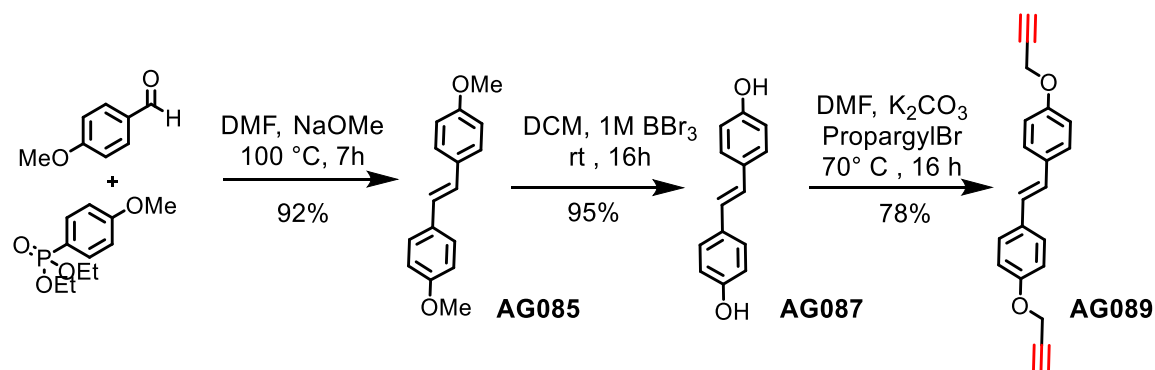


Figure 7. Proton NMR spectrum of E-AG085. The spectrum was recorded in deuterated chloroform on a 500 MHz instrument. No signal for the Z isomer was observed at δ 6.45 ppm in the crude product.

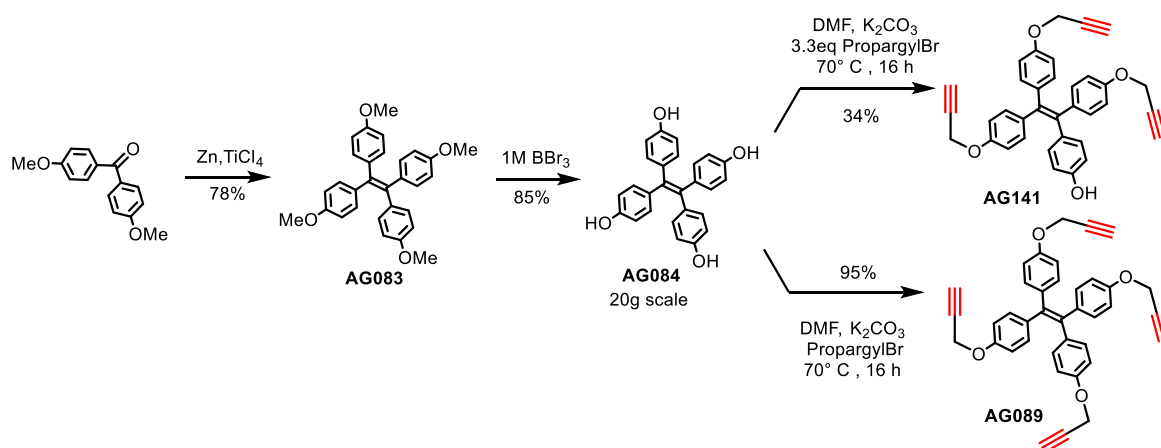
The E isomer was the only one ever observed, even after irradiation with the UV lamp available in the laboratory, no signals of the Z isomer were detected by proton NMR spectroscopy. The methyl groups were removed with 1M solution of boron tribromide and the resulting phenol OH groups were propargylated upon reaction with propargyl bromide in basic conditions, scheme 1.



Scheme 1. Synthesis of the dipropargylated stilbene: only the trans isomer was observed. It appears to be stable, with no traces of the cis isomer being observed by NMR. The overall yield was 68% in three steps.

For the synthesis of the trivalent and tetravalent cores, I decided to form the tetraphenylethylene scaffold using the McMurry reaction of 4,4'-dimethoxybenzophenone to give **AG083**. As previously reported, the treatment with 1M boron tribromide in CH_2Cl_2 gave compound **AG084**,¹² scheme 2.

The tripropargylated product was eventually obtained in sufficient amounts after direct propargylation of the core **AG084** with stoichiometric amount of propargyl bromide. The inevitable side products have a slightly different retention factor, which allowed their separation via column chromatography.

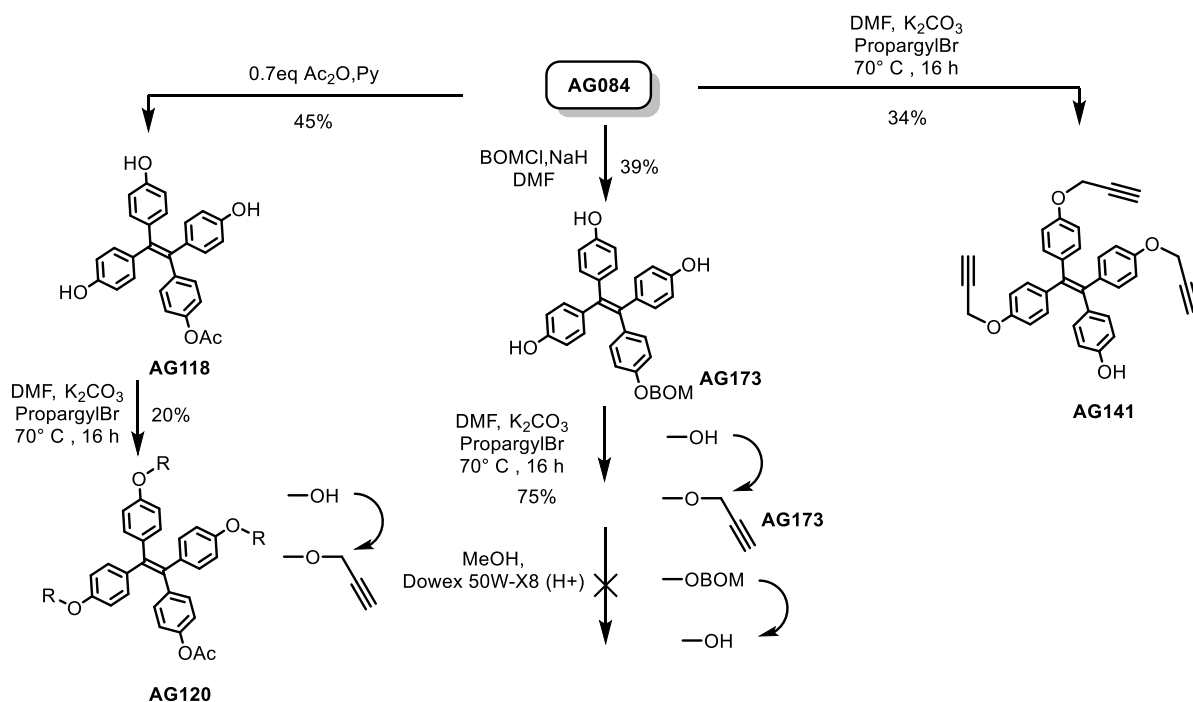


Scheme 2. Synthesis of the tetravalent and trivalent: the direct propargylation of the tetraphenylethylene derivatives **AG084** turned out to be the more direct and efficient strategy. Overall yield to **AG141** 23% in three steps.

In an early stage of the project, the selective mono-protection of one hydroxyl group was performed by reacting with 0.7 equivalents of acetic anhydride to give compound **AG118** in 45% yield. As perhaps could be predicted by an experienced synthetic chemist, the subsequent attempt at propargylation yielded mainly the tetra-propargylated compound **AG089** arising from the deacetylation product of **AG118** caused by the basic environment

arising from the presence of K_2CO_3 , and subsequent propargylation, while providing only 20% of the desired **AG120**. An attempt was also made via the use of benzylozylmethyl ether (BOM), a protecting group stable in basic condition, which failed in the final cleavage step, scheme 3.

Direct propargylation of **AG084** resulted in the most efficient synthesis, even if the yield was only ~20%. An alternative route to obtain a trisubstituted tetraphenyl ethylene scaffold might be a McMurry reaction that yields a heterodimeric product, between 4,4'-dimethoxybenzophenone and 4-methoxybenzophenone. This approach can be considered for the synthesis of the trivalent ligand on its own, but, as will be discussed in chapter 3, the presence of the phenol offers a versatile opportunity to prepare conjugates to a trivalent compound based on tetraphenyl ethylene.



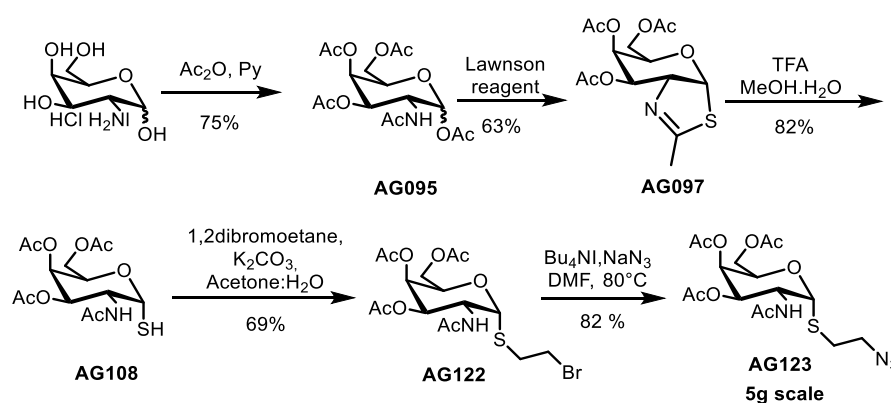
Scheme 3. Attempted synthesis of the trisubstituted TPE: the acetyl group of **AG118** is cleaved in basic environment and high temperature leading to the tetrapropargylated TPE. The overall yield to **AG120** was 6% in four steps. The direct propargylation of the TPE derivatives **AG084** turned out to be the more direct and efficient.

Azide derivatives of carbohydrates

The formation of the α -GalNAc thiol **AG118** was accomplished according to the previous treatment of the thiazoline **AG097** with TFA in wet methanol, scheme 4.¹³

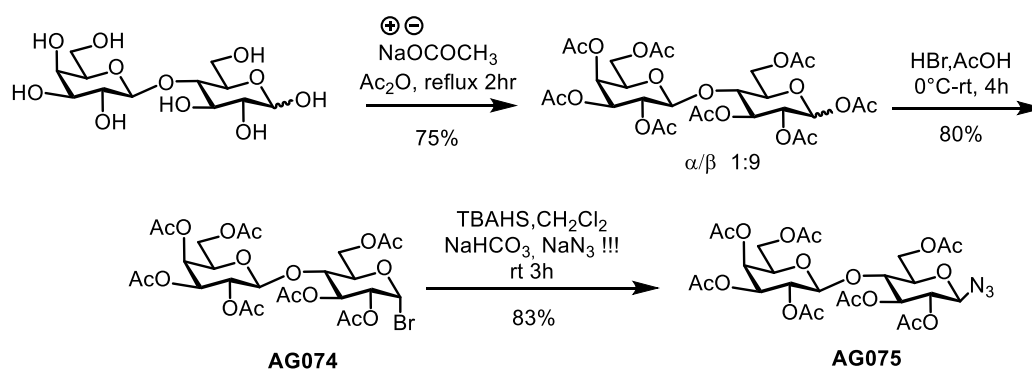
The brominated compound **AG122** was then converted by sodium azide in the presence of tetrabutylammonium iodide in DMF to give the corresponding azide **AG123**.¹

The final azide can be obtain in a higher yield without need of purification using tetrabutylammonium hydrogensulfate and sodium azide with dichloromethane as solvent, the condition reported by Tropper et al.¹⁴ Please note, that this reaction is not suitable for scale up due to the usage of dichloromethane as solvent which could yield in the production of highly explosive di- and tri- azidomethane.



Scheme 4. Synthesis of the azide derivative **AG123**: Overall yield 22% in five steps.

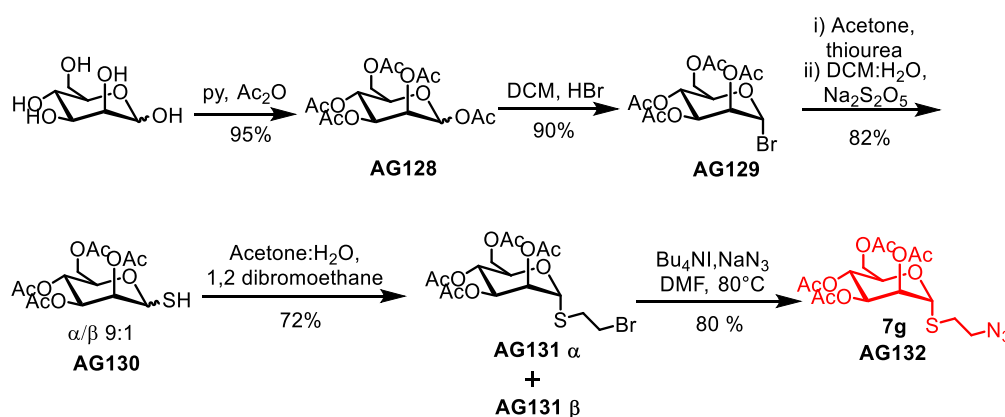
For the synthesis of the lactose azide **AG075**, hydrogen bromide in 33 % acetic acid was used to introduce a bromine onto the anomeric carbon of the peracetylated lactose. An $\text{S}_{\text{N}}2$ reaction led to displacement of the bromine atom with azide, which was carried out in the presence of a phase transfer catalyst, and yielded compound **AG075** as single anomer.¹⁴ Please note this reaction condition must not be scaled up over 100 mg without taking appropriate precautions in the use of sodium azide, scheme 5.



Scheme 5. Synthesis of lactose azide sugar head: Overall yield 50 % in three steps.

To obtain the mannose azide **AG132** peracetylation of D-mannose yielded to the compound **AG128** as a mixture of anomers, which was then converted in the respective α -bromide using HBr.

The treatment of bromide **AG129** with thiourea gave the thiouronium salt, which upon hydrolysis yielded the glycosyl-thiols in 73% yield (α/β 9:1). This mixture was reacted with 1,2 dibromoethane in presence of potassium carbonate to give a separable mixture of brominated products. The α anomer was isolated after chromatography and was converted into the desired azide **AG132**, scheme 6.



Scheme 6. Synthesis of mannose azide AG132. The synthesis was carried on with α/β anomer mixture until the bromides formation, the α anomer was then isolated by column chromatography and converted in the desired azide. Overall yield 36% in five steps.

To confirm the identity of the α anomer we measured the ^{13}C - ^1H coupled spectra, figure 8. The $^1\text{J } ^{13}\text{C1-H1}$ geminal coupling constants are known to be ~ 170 Hz for the α -anomer of hexopyranoses whereas the corresponding value for the β -anomer is 160 Hz.^{15,16,17} Even if nothing is reported in literature about the geminal ^{13}C - ^1H coupling constant of glycosyl

thiols, in the Murphy group, extensive study about the anomerisation of this type of glycans were performed¹⁸ and therefore the geminal coupling constant of α and β anomers were measured. Unpublished data of Lisa Doyle and Iwanowicz Karolina confirm that glycosyl thiols follow the same trend of the reported glycosides (<http://hdl.handle.net/10379/7420>).

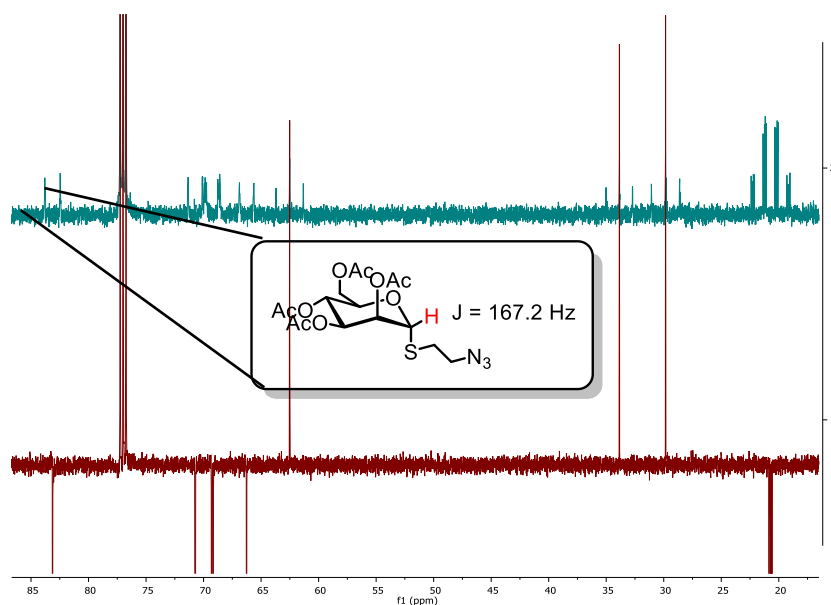


Figure 8. ^{13}C - ^1H coupled spectra of AG132. The $^1\text{J } ^{13}\text{C}1\text{-}^1\text{H}1$ geminal coupling constant is close to 170 Hz, confirming the identity of the α -anomer.

Multivalent ligand assembling

With all the synthons in hand, the assembling of the multivalent ligands was started. The 1,4 regioisomer was obtained in the presence of Cu(I) catalyst. In anticipation of industrial scale up, the combination of CuSO_4 and sodium ascorbate was chosen. In fact, the reducing agent always guarantees the presence of Cu(I) and gives reproducible results independently from the amount of oxygen present in the solvent which can vary throughout the year due to temperature changes.

To ensure the complete functionalisation of the core branches, all the reactions were conducted with an excess of glycan. It is worth mentioning that, in all reaction steps, the fluorescence propriety of the tetra phenyl ethylene is conserved, therefore it is possible to easily follow the product formation and starting material consumption via TLC at 360 nm.

A typical procedure involved using $\text{Cu(II)SO}_4 \cdot 5\text{H}_2\text{O}$ (0.6 equivalents) and a slight excess of sodium ascorbate (0.65 equivalents), the reactions were stirred at 40 °C for 20 h. A faster method was tested in small scale for all the products, which involved the use of the same

solvent and catalyst system followed by of microwave irradiation at 50°C, 120 W for 15-30 min.¹⁹ The yields are as good as the traditional heating, but this method has limitation in terms of reaction scale: the biggest microwave reactor available in the group allows 5 ml scale.

After purification by flash chromatography the various multivalent ligands were deprotected using 1M sodium methoxide solution. The deacetylation reactions were carried out at room temperature until complete disappearance of the acetylated product by TLC, typically 2 h. The reactions were neutralized at pH 7 with Amberlite IR- 120 H⁺. In the specific cases of divalent ligands, **AG104** and **AG139**, the deprotected products were not soluble in methanol. In these cases, the neutralisation was performed adding acetic acid to reach neutral pH.

Final purification by reverse phase column chromatography, gave the desired compounds in good yields, figure 9.

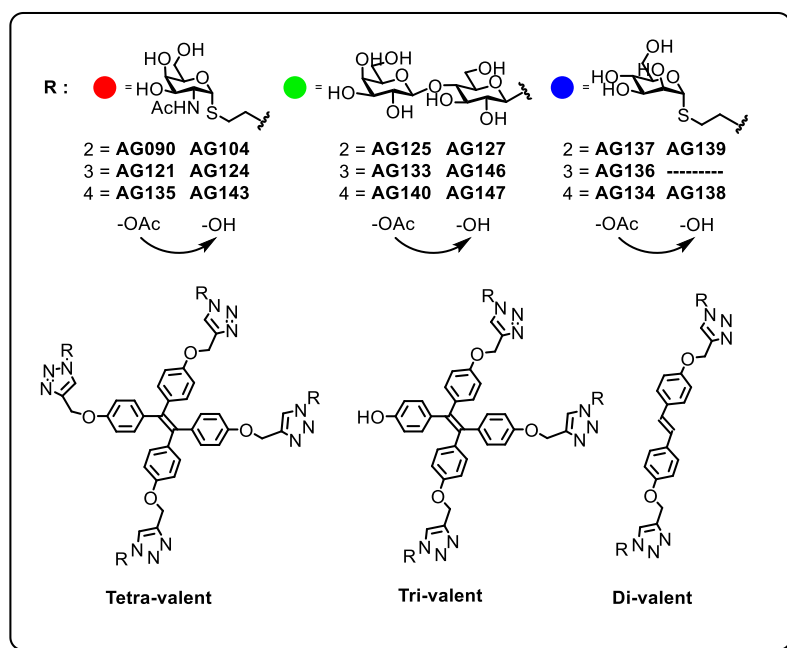


Figure 9. Chemical Structures of the small library of multivalent glycoclusters.

2.3.2. Stilbene and tetraphenyl ethylene fluorescence properties.

During the synthesis, the fluorescence properties of the tetraphenylethylene derivatives were noticed. A strong light blue emission was observed using the 360 nm lamp, which, I have to confess, made every single TLC analysis a matter of amazement and pride among my co-workers, figure 10.

The TPE core is an atypical molecule which is faintly fluorescent in diluted solution while is “turned-on” in the solid state or as aggregate. This phenomenon, introduced for the first time in 2001,²⁰ is known as aggregation-induced emission (AIE) and is the opposite of aggregation-caused quenching (ACQ), a phenomenon observed in classical luminophore systems. The emission of tetraphenyl ethylene derivatives is also observed at low concentration using increasing percentage of water as solvent, the highly hydrophobic nature of the tetraphenyl ethylene derivatives to induce molecules aggregation in the presence of aqueous solvent.²¹

Despite the decoration of the tetraphenyl ethylene core with hydrophilic functional groups, it still showed fluorescence, figure 10. The same fluorescence was reported when the tetraphenylethylene is functionalised with glucose, fucose and galactose.²² As might be expected though, the percentage of water required to turn-on the fluorescence for the glycoclusters is higher than the one required to activate the unfunctionalized tetraphenyl ethylene alone and differs from glycan to glycan.²²

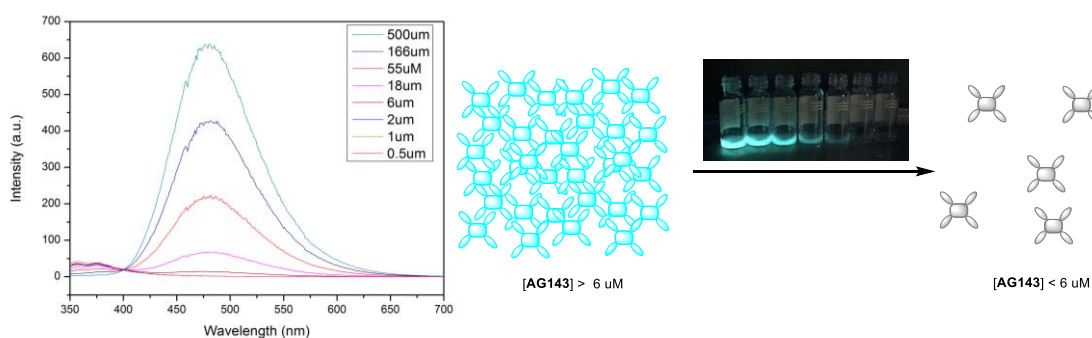


Figure 10. Emission spectra of AG143 at different concentration: the fluorescence intensity increase proportionally to the concentration. At higher concentration the intermolecular rotation of the phenyl rings is inhibited and excited species of **AG143** can't deactivate via intermolecular rotation. The measurement was performed on an Agilent Cary Eclipse Fluorescence Spectrophotometer, Ex 370 nm.

The aggregation of tetraphenyl ethylene derivatives causes the restriction of intramolecular rotation inducing the relaxation of tetraphenyl ethylene * excited state via fluorescence. The bulky phenyl rings are twisted up to -60° out of the ethane plane. Therefore, in diluted solution, the phenyl rotors have enough space to more freely rotate annihilating the excited tetraphenylethylene * species, figure 11. At high concentration and in aggregate form, the free rotation is not possible and the excited state decays through fluorescence emission. Aggregation caused quenching is not observed in tetraphenyl ethylene derivatives, precisely because their non-planar nature, does not allow π - π stacking in the solid state.

Stilbene and its derivatives, including the one synthesised in this PhD project, at the contrary, do not present fluorescence proprieties in regular working condition. Stilbene is a weak emitter at low concentration and the emission is quenched on increasing the concentration. The absence of emitting proprieties, in the past was exclusively attributed to the E-Z photoisomerization, while, recently, the Tang group has offered a different mechanistic insight to stilbene derivatives. In fact they proved by proton NMR analysis, that E-Z isomerisation does not happen in normal photoluminescent measurements.²³ The different photoluminescent behaviour of tetraphenyl ethylene and stilbene may arises from the differences in their chemical structure: stilbene derivatives are almost planar, with π electrons highly delocalised. The chemical structures of stilbene inhibits the phenyl ring rotation but allows the π - π stacking between planar molecules and the excited states of the aggregates decay via non-radiative pathways, figure 11.²⁴

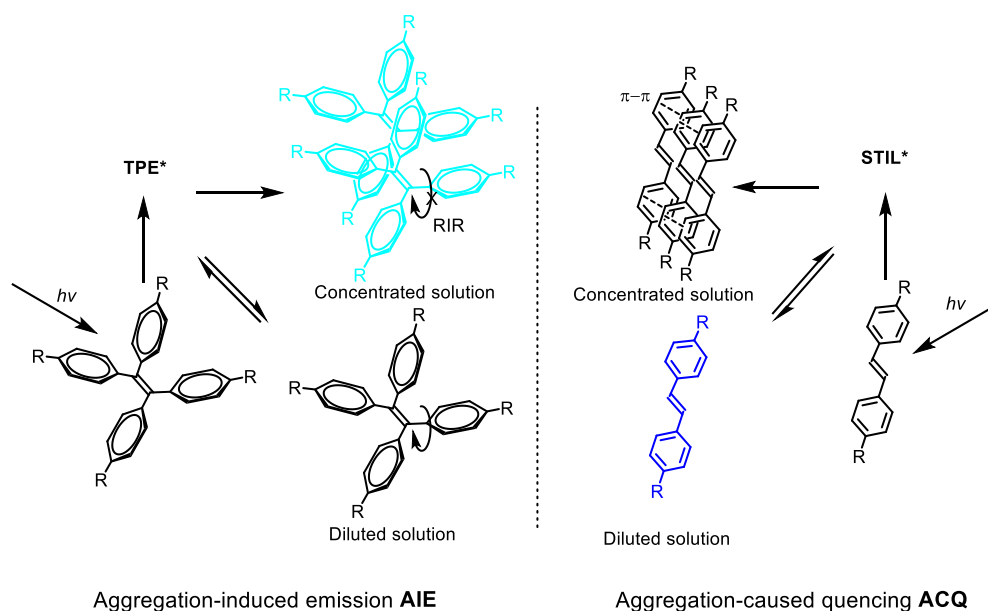


Figure 11. Tetraphenyl ethylene and stilbene possible cause fluorescence relaxation.

In the past the tetraphenylethylene scaffold decorated with lactose moieties was used to target Cholera toxin protein.¹² This protein forms a β pentamer which bind galactose, lactose and fucose molecules, figure 12.²⁵

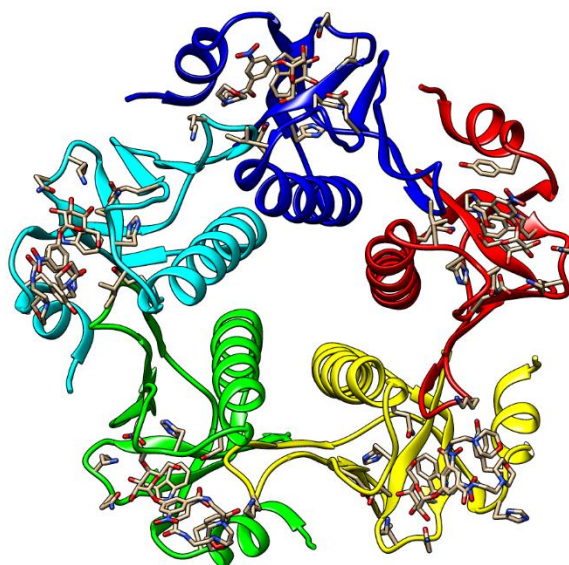


Figure 12. Crystal structure of cholera toxin beta-pentamer with galactose derivative BMSC-0011. PDB code: 1JR0
25

When the lactose bearing ligand was mixed with cholera toxin, its fluorescence was “turned-on”. This phenomenon is caused by the increase in local concentration due to the binding on multiple CRD of the cholera toxin. When the phenyl rings have limited freedom to disperse energy rotating, a strong emission, inversely proportional to the rotation ability is observed.²⁶

In the work of X. M. Hu et al. , Cholera toxin protein structure was well known, and the assay was used to prove the binding event. On the other hand, no fluorescence change was observed when the multivalent lactose was titrated with bovine serum albumin (BSA), a protein with no affinity for lactose, indicating the fluorescence change is associated with interaction of lactose into the CRD, rather than a non-specific interaction.

It was envisaged that the fluorescence properties of the molecules could be used to gain additional information about its oligomerisation state after glycan binding: if hMGL carbohydrate recognition domain behaves as an “individual” domain, the protein would simply bind the GalNAc moieties on each molecule but the binding event would not induce vicinity of tetraphenylethylene molecules (figure 13 A). On the other hand, if the protein carbohydrate recognition domain are interacting among each other, ligand addition will

induce formation of cross-linked structures, the addition of protein would increase ligand local concentration and therefore increase fluorescence signal (figure 13 B).

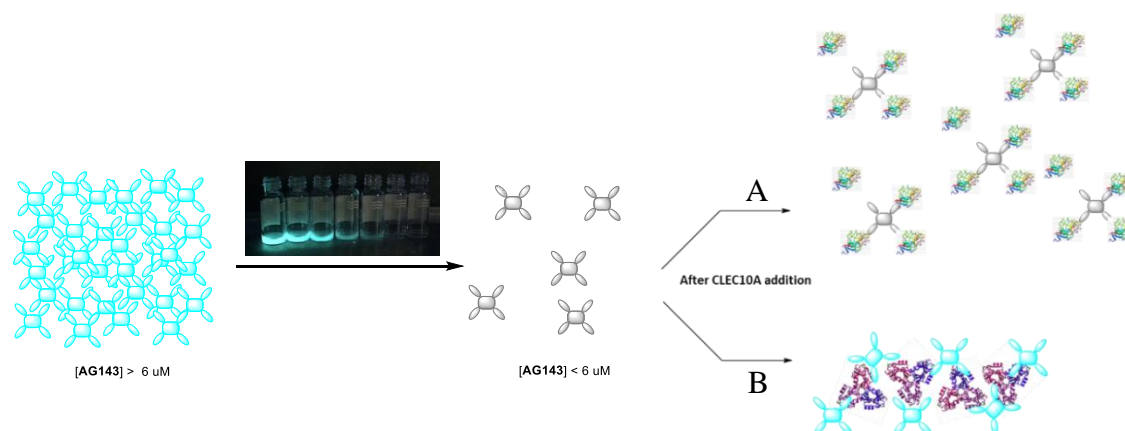


Figure 13. Fluorescence proprieties of tetraphenylethylene derivatives at different concentration. At low concentration, AG143 can disperse energy via free rotation of the aromatic rings. Below $6 \mu M$ AG143 doesn't show measurable fluorescence. If CLEC10A CRD is an oligomer, as cholera toxin, the addition of CLEC10A to a non-fluorescent solution of AG143 will "turn -on" the fluorescence. The binding of AG143 to an oligomeric form of CLEC10A will induce the molecule of AG143 to be spatially closer than in the unbound state.

Furthermore, aggregation of tetraphenylethylene based ligands to a supramolecular structure with a formal valency higher than a simple monomer unit (i.e. one tetraphenyl ethylene scaffold bearing up to four carbohydrates) could be responsible for high affinity and provide ideal geometrical requirements to bind to hMGL variants. In such a scenario the improved affinity could be caused by the enhancement of local glycan concentration which, by statistical recombination effect, makes the ligand more likely to bind the lectin. Alternatively, it could also be that the aggregate may contain a spacing between sugar units that facilitate effective face to face interactions with the lectin (chelate type binding) which leads to much higher affinities. The aggregation state of the tetraphenylethylene can be evaluated by fluorescence spectroscopy but also with transmission electron microscopy (TEM) at low μM concentration. In figure 14 it is possible to observe the increase size of the aggregate increasing the concentration of the solutions. While at $5 \mu M$ and $10 \mu M$ only aggregates of size smaller than 20 nm are visible, already at $50 \mu M$ big lumps of **AG143** aggregate can be observed. At $250 \mu M$ concentration a dense net of aggregate molecule was observed.

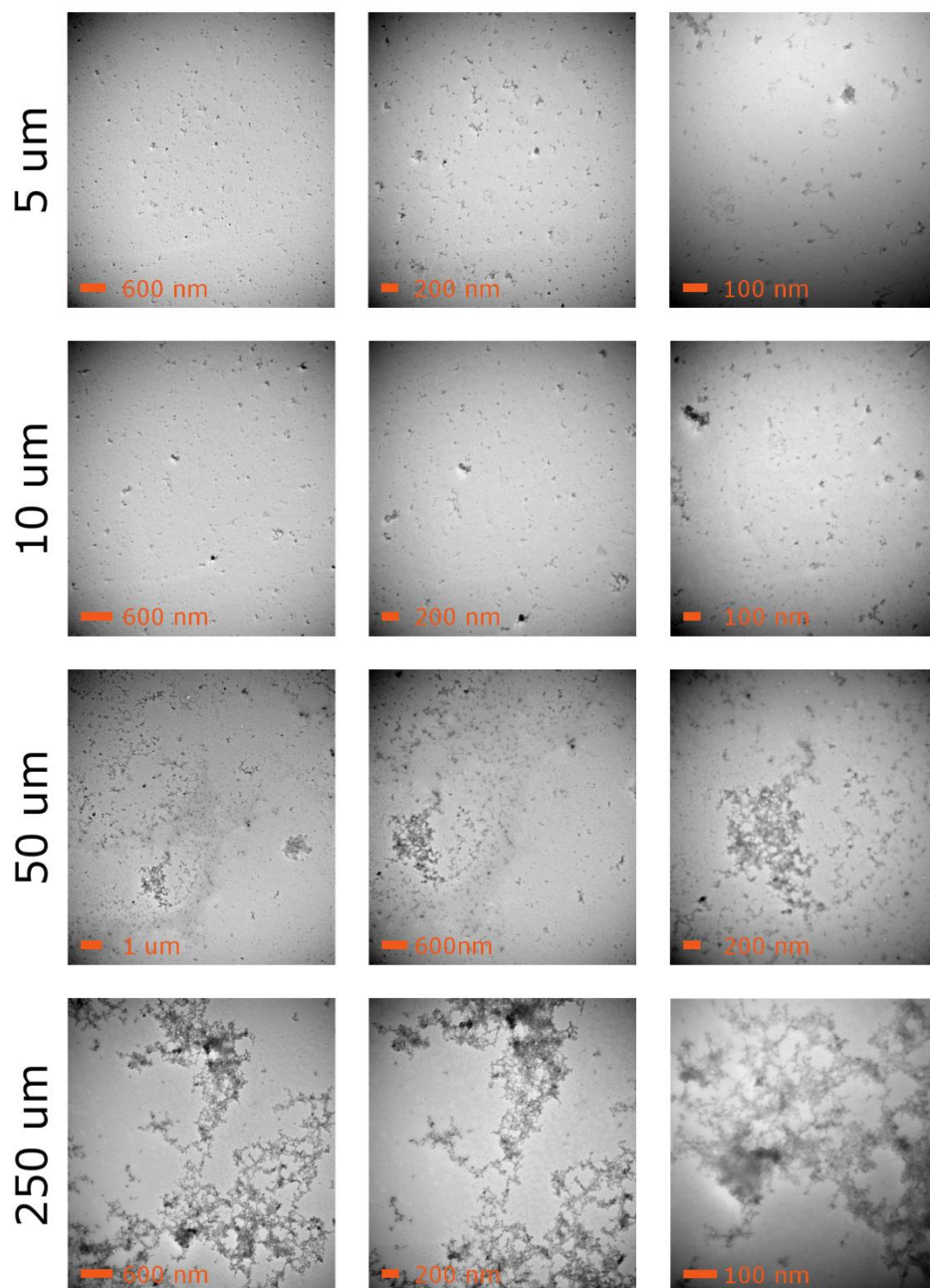


Figure 14. Transmission electron microscopy images of AG143 at different concentration in water.

First, the highest concentration at which the ligand does not present measurable fluorescent signal was identified, figure 15. An aqueous buffer solution of AG143 at different concentration containing calcium, was used to perform the measurements. 6 μM was identified as highest concentration which does not show significant measurable fluorescence.

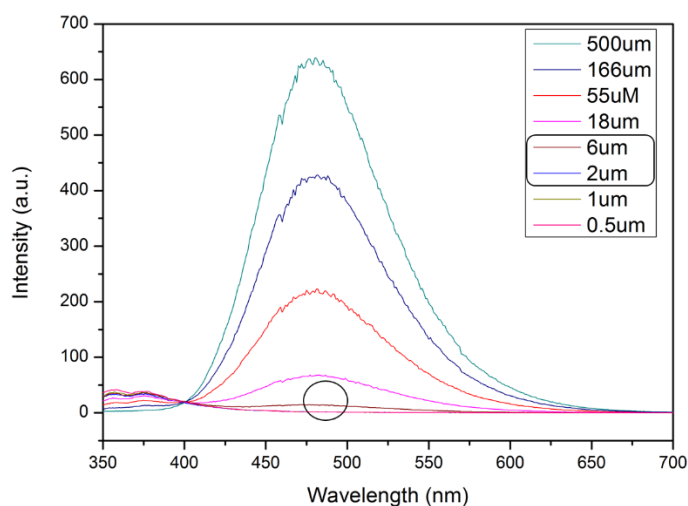


Figure 15. Fluorescence spectra of AG143 (500 μM – 0.5 μM in 20mM Tris s pH 7.5, 2 mM CaCl_2 , 150 mM NaCl). Excitation at 330 nm. The fluorescence is below the limit of detection at 6 μM , the titrations were performed at this concentration to be sure to be able to detect the change in fluorescence.

An array of samples containing 4 μM **AG143** and hMGL CRD at concentration between 0 to 25 μM were prepared and incubated for one hour at room temperature before measuring the fluorescence signal, figure 16. As a negative control, we performed the same experiment using bovine serum albumin at the same concentration of hMGL CRD.

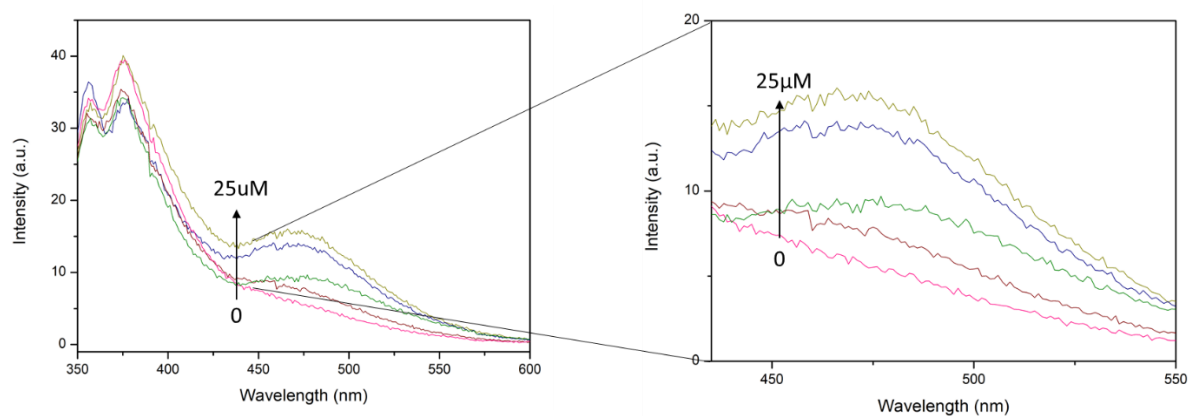


Figure 16 . Fluorescence spectra of AG143 (4 μM in 20mM Tris s pH 7.5, 2 mM CaCl_2 , 150 mM NaCl) with different amounts of CLEC10A 0, 5, 10, 15 and 25 μM .

After addition of increased amount of hMGL CRD, the sample fluorescence slightly increased, reaching values which are maximum three time higher compared to the ligand on its own. Protein fluorescence in the absence of ligand was analysed at 5, 10, 15 and 25

μM : hMGL on its own does not present fluorescence in the same wavelength range of AG143.

Despite the fact the increase in fluorescence intensity is linearly proportional to the protein concentration, figure 17, the increased fluorescence intensity is not comparable to the increase in fluorescence observed with Cholera toxin titration: the addition of 2.5-fold excess of hMGL to AG143 results in an increase of fluorescence of 1.5-fold, while adding 2.5-fold excess of cholera toxin protein to the lactose derivative resulted in 6-fold increase in fluorescence.

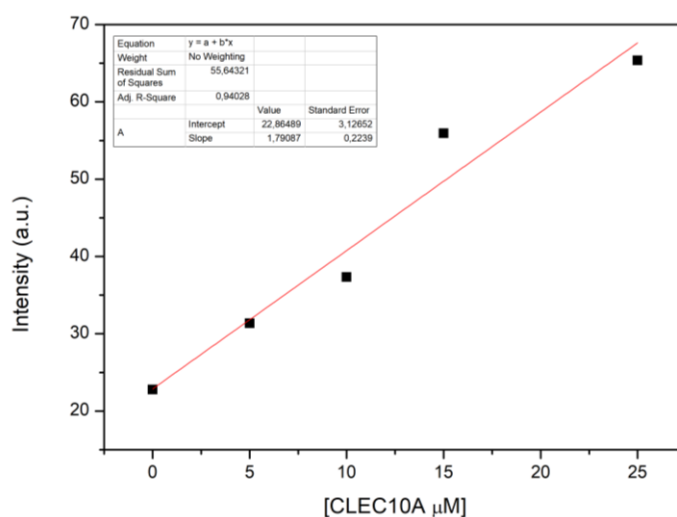


Figure 17. Fluorescence intensity at 475 nm VS AG143 (4 μM in 20mM Tris s pH 7.5, 2 mM CaCl_2 , 150 mM NaCl) plus CLEC10A at concentration of 0, 5, 10, 15 and 25 μM .

It is possible that fluorescence enhancement in the presence of the protein could be caused by the inhibition of phenyl rotation due to interaction the phenyl residue with the protein, figure 18. In this case, due to steric hinderance or hydrophobic interaction between the phenyl rings and some amino acid side chains of the protein, the sample would emit even if the protein was in a monomeric form. It is possible to imagine that the extent of this interaction is highly sequence dependent, so even using a monovalent protein with high GalNAc affinity as negative control does not provide a definitive answer.

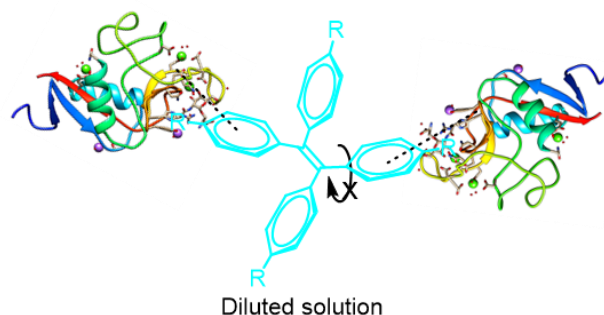


Figure 18. Interaction of CLEC10A backbone with phenyl rings of the tetraphenyl ethylene core.

A possible example of this interaction is visible in the preliminary ELISA study, figure 20, where the negative control **AG142** shows a principle of binding already at $1\mu\text{M}$ concentration, below that where fluorescence can be detected and also below where dynamic light scattering can detect the presence of aggregates.

Diffusion Ordered Spectroscopy (DOSY) experiments of the CRD ($50\mu\text{M}$) carried out in group of Jesus Jimenez Barbero in Bilbao appears to suggest that the CRD is a monomer in solution. After adding **AG143** $5\mu\text{M}$ the diffusion coefficient did not show significant changes indicating the CLEC10A is still a monomer. Further work is being carried out currently in Bilbao on full interpretation of all the NMR data obtained.

To validate the efficacy of this method we plan to perform the same type of experiment with a protein with defined oligomerisation status. Galectin-3 for instance can be expressed as monomer, or, with slightly longer sequence, as a tetramer. The hMGL extracellular domain, which gives a trimer could also be investigated by DOSY and fluorescence experiments.

2.3.3. Investigation of multivalent glycocluster binding with h-MGL isoform

3

The ELISA described below were performed by Joost van der Horst in Dr. Sandra Van Vliet laboratory.

To test how the modification of the tetraphenylethylene and stilbene with sugars and the degree of valency influenced the affinity of the compounds for hMGL, binding inhibition assay were performed. An ELISA plate was coated with polyacrylamide coupled-GalNAc polymers²⁷ followed by the incubation with hMGL which was tagged with a region of immunoglobulin heavy-chain, Fc. The protein has high affinity for the GalNAc polymer and can be detected on the plate surface via goat anti-human Fc-peroxidase.

Upon the incubation with high affinity ligands, the protein is released from the plate surface and removed in the washing steps. The arising profiles are reported in figure 20. The tetravalent GalNAc derivative **AG143** was used as positive control. The tetravalent lactose derivative **AG140** was tested to check if the tetraphenylethylene core did not have any unspecific binding properties. The tetraphenylethylene functionalised with lactose didn't show unspecific binding but **AG142**, not-functionalised core figure 19, shows partial Fc-MGL binding inhibition at 1 μ M concentration.

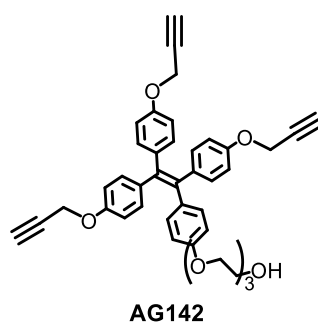


Figure 19. Structure of AG142. Full synthesis description in chapter 3.

This may suggest that when the hydrophobic phenyl rings are “naked” they may interact mildly with the peptide backbone of the protein. The trivalent GalNAc **AG124** inhibited the binding at single digit nM concentration as the positive control **AG143**. **AG104** inhibited the binding only above 100 nM concentration.

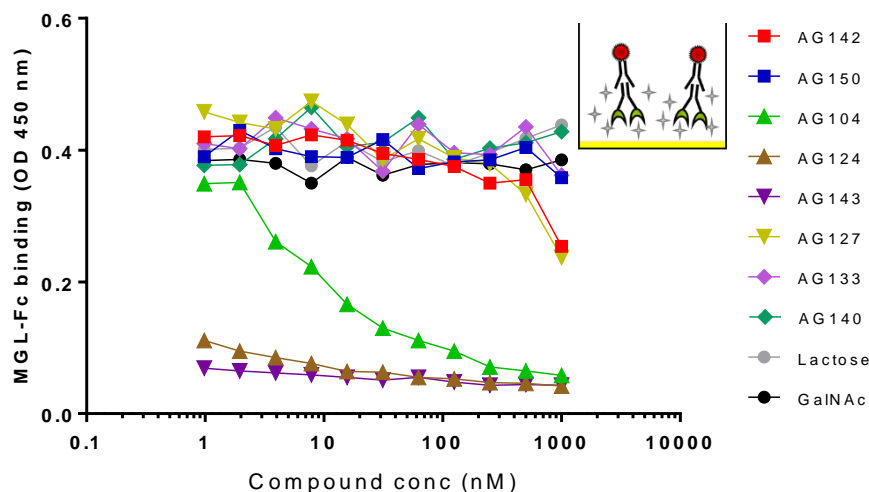


Figure 20. ELISA inhibition curve

2.3.4. Lectin histochemistry assay: getting closer to an *in vivo* model.

Despite the fact multivalent GalNAc glycoclusters showed high affinity for hMGL in plate-based assay, predicting the *in vivo* efficiency of such compounds can be challenging. Lectins are often categorised depending on the selectivity for a monosaccharide. This simplistic categorisation allowed the creation of the widely known plate-based inhibition assays in which, the specific monovalent ligand, e.g. polyacrylamide coupled-GalNAc polymer in the above-mentioned ELISA, is used as first ingredient in the sandwich. These assays offer a cheap and fast way to collect excellent preliminary data for the development of a project but predicting the *in vivo* efficiency of a synthetic ligand based on this crude approach can be challenging. In the papers reported in section 2.3, Hans-Joachim Gabius and co-workers developed an histochemistry assays to explore the inhibition effect of multivalent glycoclusters in a complex model. In fact, tissues present natural glycome complexity and lectins with nominally identical monosaccharide specificity have the ability to recognize fine differences in more complex structures.³ As it is shown in our recent article, *RSC Adv.*,**2018**,8,28716-28735, even if a lectin has identical affinity for monosaccharide substrate, it can give quite different staining patterns on tissue sections, which are related to the distribution of more complex oligosaccharides or glycopeptides.

2.3.5. Binding and protein stability

As part of this thesis work, hMGL was expressed on bacteria cell lines, chapter 5. Thermal folding analysis on the obtained protein, revealed that addition of ligands to the CRD of hMGL can improve the overall stability of the protein. A correlation with protein stability and tight binding was observed. For instance, the inflection temperature of the sample containing the tetravalent derivative **AG143**, among the best hMGL binder, is over 6 °C higher than the free CRD, while lactose derivatives which do not bind hMGL, leave the stability of the protein unaltered, figure 21.

This effect can be explained with stabilisation of the flexible glycine rich loop in hMGL CRD by the ligands. The same type of behaviour was observed recently by protein NMR by Marcelo and co-workers.²⁷

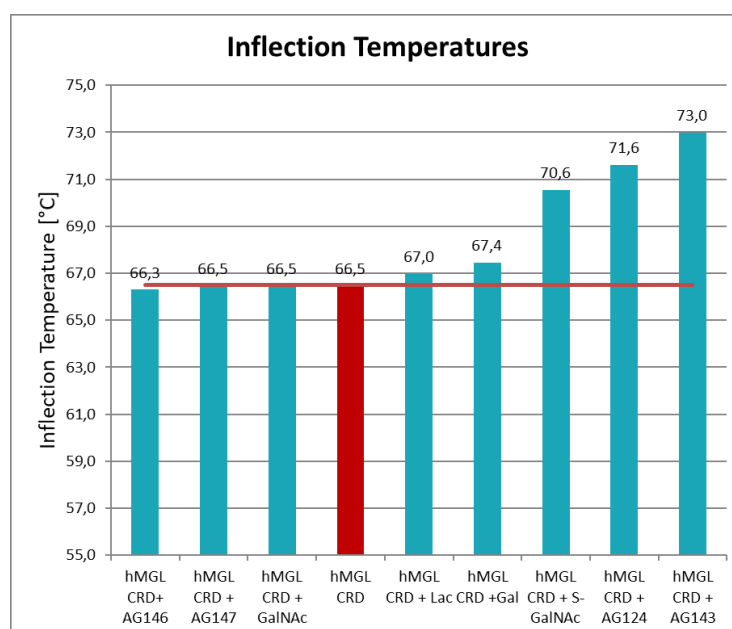


Figure 21. Inflection temperature of hMGL CRD with different ligands. Protein concentration 1mg/mL plus 10 equivalents of ligand in buffer containing calcium.

2.3.6. Isothermal titration calorimetry

The plate-based inhibition assay and histochemistry results taught us that our multivalent glycoclusters, including the trivalent compound have a high affinity for hMGL. It was thus considered worthwhile to obtain more information about the thermodynamics of the binding including the measurement of the direct binding constant. Isothermal titration calorimetry (ITC), is a technique that directly determines thermodynamic-binding parameters.²⁸ Protein and ligand do not need labelling and ITC is an excellent technique to define the thermodynamic parameters of multivalent interaction such as the entropy of binding (ΔS), the enthalpy of binding (ΔH) and the association constant (K_a).

The association constant (K_a) can also be calculated with microscale thermophoresis upon appropriate labelling of the protein polyhistidine tag with a fluorescent dye.²⁹

-ITC instrument design

In a forward ITC experiment the ligand is gradually titrated with an automated syringe into a sample cell containing the protein solution. Beside the sample cell, there is a reference cell filled with water. The calorimeter needs to keep these two cells at the exact same temperature. After the addition of each ligand aliquot, if a binding event occurs, heat is absorbed or released. The heat is measured by the calorimeter. In the reported example, figure 22, the sample cell becomes warmer than the reference cell after each addition, the reaction is exothermic, causing decreasing peak intensity in the raw data. After several ligand additions, the protein becomes more and more saturated, until eventually there is an excess of ligand in the cell and no binding event can further occur. The area of each peak is integrated and plotted against the molar ratio of the ligand, figure 23, producing an isotherm which can be fitted into a binding model and allows the calculation of thermodynamic parameters.

By definition, in a reverse ITC experiment, the ligand is loaded into the sample cell and the protein is titrated via the syringe.

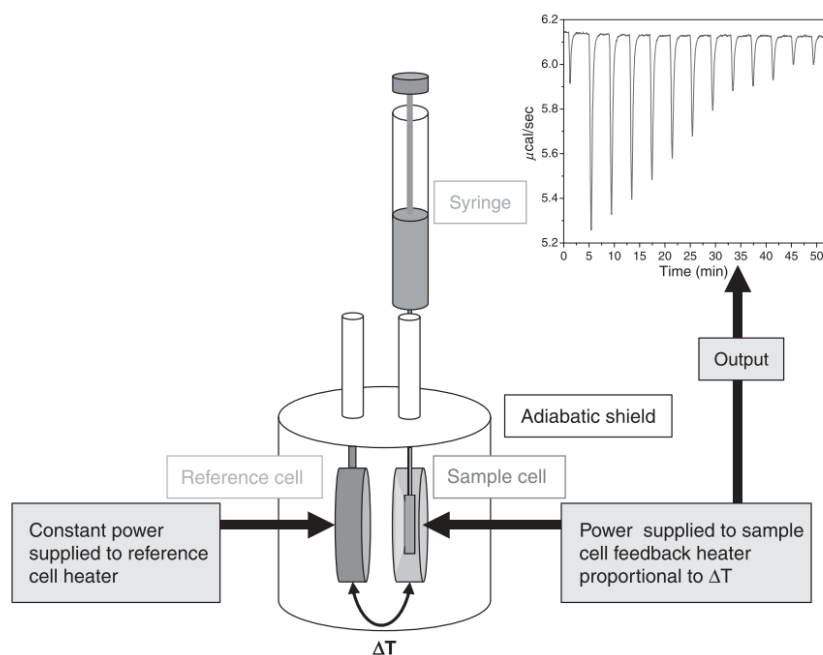


Figure 22. Schematic representation of a power compensation ITC instrument. Image from *Methods Cell Biol.* 2008, 84, 79–113 with copyright permission.²⁸

-Method

In this thesis work, the monovalent and multivalent ligands were titrated into a solution of the lectin, as is typical in most ITC experiments. Injections of 8 μL of ligand solution were added from a computer-controlled microsyringe at an interval of 240–300 seconds into the sample solution of hMGL (cell volume of 1.34 mL) with 330 rpm stirring. The glycan concentration varied from 0.07 to 1 mM. The molar concentrations of multivalent carbohydrate analogues, not their epitope equivalence was used. The ITC analysis of the divalent glycocluster **AG104** was not obtained because it was not sufficiently soluble in the protein buffer. The high percentage of DMF that would have been required for its dissolution is not compatible with ITC analysis and with the stability of the lectin. A common issue in multivalent ligand-protein ITC experiments, is the stoichiometry-dependent precipitation of the reactants due to noncovalent cross-linking that can occur between them. The precipitation influences the quality of the analysis. To avoid cross-linked complex formation or significantly reduced it, a strategy is recommended to keep the protein concentration as low as possible in the limit of the heat change detection. Several successful examples of this strategy are reported by Brewer and co-workers.^{30,31,32}

Solutions of hMGL (expression and purification protocol in chapter 5), 8 and 10 μM , were used for the analysis. Size exclusion analysis of the protein-ligand solution arising from the titration, was consistent with the absence of high molecular weight aggregates.

In table 2. are reported the thermodynamic binding parameters obtained from the measurements.

Binding data for the multivalent sugars were fitted to a theoretical titration curve using a single site binding model. From the theoretical curve the following were obtained: the enthalpy change in cal/mol; ΔH , the association constant in M^{-1} ; K_a , ($K_d=1/K_a$) the number of binding sites per lectin monomer, N ; and the entropy change in cal/mol/deg. The change in free energy was calculated from the equation:

$$\Delta G = \Delta H - T\Delta S = -RT \ln K_a$$

The analysis of the thermogram is a curve fitting process in which a nonlinear regression procedure is used to fit a model to the data.²⁸The model is a mathematical description of the chemical interaction that is taking place in the calorimeter and in which the dependent variable (e.g., heat or heat rate) is defined as a function of the independent variable (e.g. moles of titrant added) and one or more model parameters. In the case of binding experiments, the model is formed from the equilibrium constant and mass balance equations. Nonlinear regression is used along with the model equations to determine the best values of the fitting parameters (e.g., K , ΔH , and n). The goal is to model the experimental data within expected experimental error, using the simplest model, and a model that makes sense in the light of what is already known about the system (e.g. stoichiometry).

The nonlinear regression analysis of ITC data is an iterative process, among the models present in the analysis software provided by MicroCal® ITC Origin™, the model which fits best the experimental curves obtained is the one-site (n identical sites).

$$K = \frac{\theta}{(1 - \theta) \cdot [L]}$$

Figure 23. shows the raw ITC data for a ligand into protein titration on the left, while, on the right, are shown integrated heat data, squared markers, and a nonlinear regression fit to a one-site binding model. The curve corresponds to the theoretical heat produced for 1:3

complex formation between the ligand and the protein ($N=0.36$). The parameters K_a , ΔH , ΔS , and N used for the fit are listed in the box. The nonlinear regression analysis shown was performed using the raw data and the one-site reaction model in the Origin 7 ITC software package provided by the ITC manufacturer, in this case MicroCal.

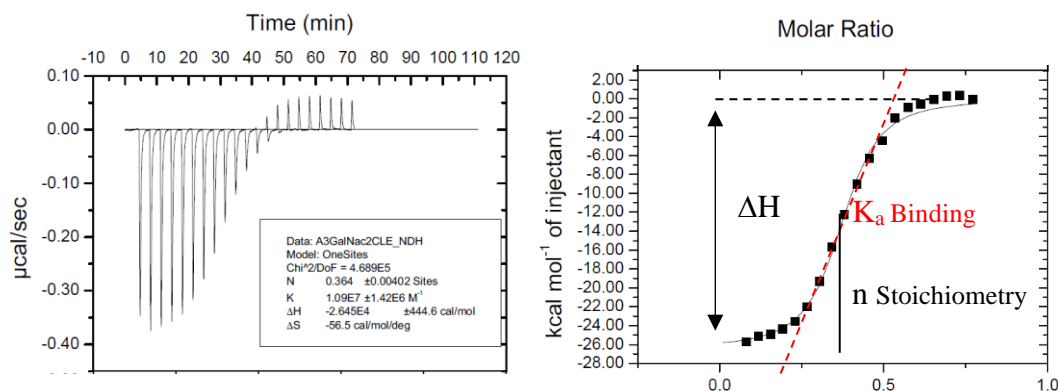


Figure 23. On the left raw data from ITC titration. On the right integration of the raw data over the ligand-protein ratio.

Microcalorimetry provides a total picture of binding thermodynamics. Even if an interaction is characterised by the same ΔG , the determination of both the entropic and enthalpic contribution can potentially give information about the binding mechanism. In addition to the spontaneity of the binding, (table 1. which most of the readers remember from an early undergraduate physical chemistry course, or, for readers like me, had the pleasure to refresh it in preparation of the first-year chemistry tutorial), the enthalpy value is an indication of changes in interactions such as hydrogen bonding and van der Waals forces, while $-T\Delta S$, is indication of changes due to hydrophobic interactions and/or conformational changes.³³

For instance, a binding event which has a thermodynamic profile similar to entry 1 in table 1, is spontaneous at all temperature and is favourable for both hydrogen and hydrophobic interaction, while in entry 2 the binding is dominated by hydrophobic interactions and in entry 3 by hydrogen bonding.

As will be further discussed below, in the presence of multivalent ligands, also the relative $-T\Delta S$ and ΔH values of the multivalent ligand compared to the monovalent ligand offer insight in the binding mode.

Table 1. Summary of the spontaneity and the signs of enthalpy and entropy terms and SAR

Entry	ΔH	$-T\Delta S$	$\Delta G = \Delta H - T\Delta S$	structure–activity relationship (SAR)
1	-	-	The process is favored by both enthalpy and entropy and is spontaneous at all temperatures	Favorable hydrogen and hydrophobic interaction
2	-	+	The process is favored by enthalpy but opposed by entropy. It is spontaneous only at temperatures where $T\Delta S < \Delta H$	Binding dominated by hydrophobic interaction
3	+	-	The process is opposed by enthalpy but favored by entropy. It is spontaneous only at temperatures where $T\Delta S > \Delta H$	Good hydrogen bonding with unfavorable conformational change
4	+	+	The process is opposed by both enthalpy and entropy and will not occur spontaneously.	

-Data analysis and interpretation

From multiple titration pictured in figure 24 and 25, thermodynamic data were obtained.

The values are reported in table 2. For detail about the experimental design see chapter 6.

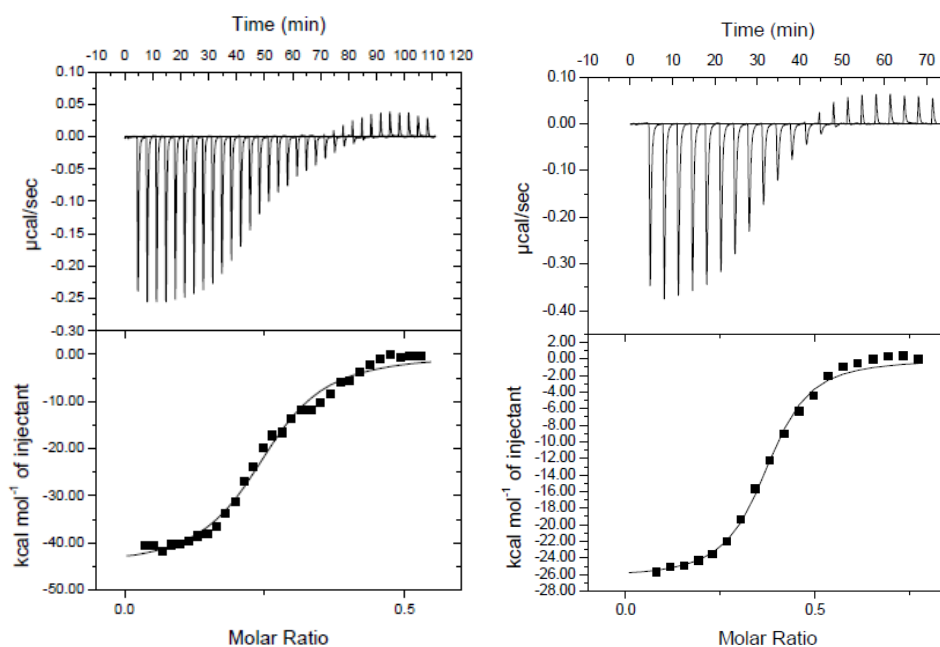


Figure 24. Forward ITC analysis. Examples of binding curves: on the left AG143 titrated into 10 μM (20 mM Tris pH 7.5, 2 mM CaCl_2 , 150 mM NaCl) hMGL carbohydrate recognition domain, on the right, AG124 was titrated into 10 μM (20 mM Tris pH 7.5, 2 mM CaCl_2 , 150 mM NaCl) hMGL carbohydrate recognition domain.

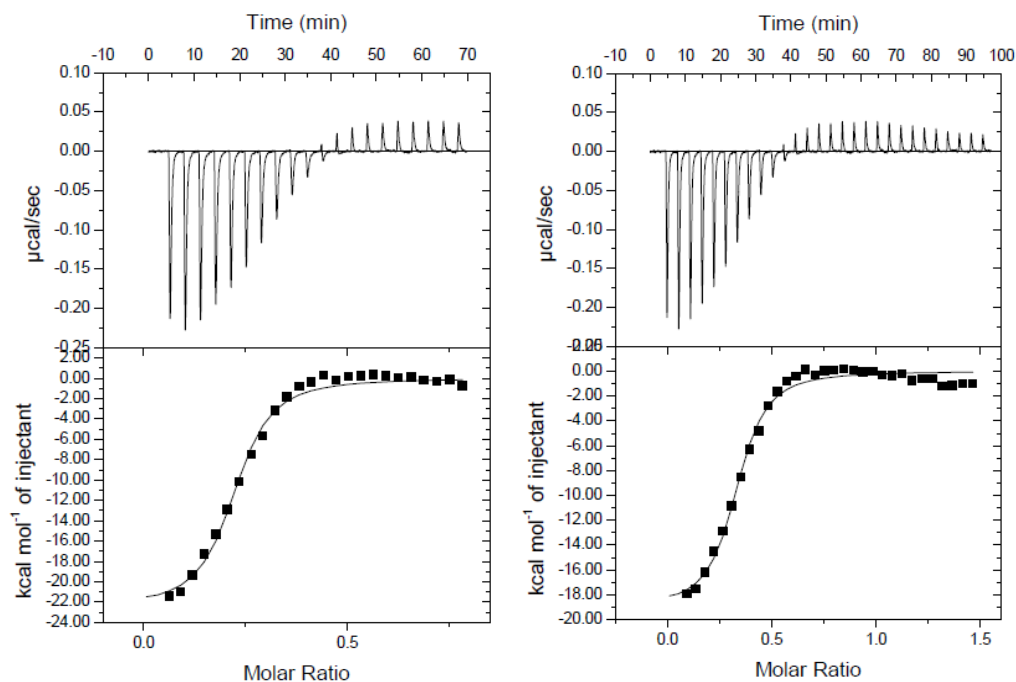


Figure 25. Forward ITC analysis. Examples of binding curves: on the left AG143 titrated into 8 μM (20 mM Tris pH 7.5, 2 mM CaCl_2 , 150 mM NaCl) hMGL isoform 3, on the right, AG124 titrated into 8 μM (20 mM Tris pH 7.5, 2 mM CaCl_2 , 150 mM NaCl) hMGL isoform 3.

Table 2. Thermodynamic binding parameters for hMGL carbohydrate recognition domain (CRD) and extracellular domain hMGL isoform 3 multivalent and monovalent sugars at 25 $^\circ\text{C}$.

Ligand(N° epitope)	Protein	K_d ^a	$-\Delta G$ (kcal/mol) ^b	$-\Delta H$ (kcal/mol) ^c	$-T\Delta S$ (kcal/mol) ^d	N^e
AG143(4-GalNAc)	CRD	136 nM	9.33	45.1	35.8	0.25
AG124(3-GalNAc)	CRD	91.7 nM	9.57	26.4	16.8	0.36
GalNAc(1-GalNAc)	CRD	13.2 μM	6.6	8.7	2	1.05
Ser-T _N (1-GalNAc)	CRD	3.5 μM	7.4	12.2	4.7	1.08
AG143(4-GalNAc)	Extra	103 nM	9.5	22.5	13	0.22
AG124(3-GalNAc)	Extra	134 nM	9.33	19	9.7	0.33
GalNAc(1-GalNAc)	Extra	16.9 μM	6.5	12	5.4	1.06
Ser-T _N (1-GalNAc)	Extra	9.2 μM	6.9	20	12.7	0.98

^a Errors in K_d are 2-6%

^b Errors in ΔG are less than 1%

^c Errors in ΔH are 2-5%

^d Errors in $T\Delta S$ are 2-7%

^e Errors in N are less than 2%

The first four rows of table 2, represent the data obtained from the ITC experiments with the truncated hMGL, which has the carbohydrate recognition domain only. GalNAc has a K_d of 13.2 μM , in good agreement with the K_d calculated recently by STD NMR for methyl 2-(acetylamino)-2-deoxy- α -D-galactopyranose.³⁴ The presence of the serine residue on

the anomeric position increases the affinity for the carbohydrate recognition domain of ca 4-fold. **AG143** bears four GalNAc residues but has a K_d that is much lower than the monovalent GalNAc K_d multiplied by four, ca 100-fold lower. The same is true for the trivalent glycocluster **AG124** which has a K_d 144-fold lower than GalNAc. Considering the numbers of epitopes of each multivalent ligand, the trivalent **AG124** is more potent than the tetravalent. The GalNAc relative affinity per number of epitopes is 24-fold more potent for **AG143** and 47-fold more potent in for **AG124**. The last four rows of the same table represent the data obtained from the ITC experiments with the extracellular fragment of MGL, isoform 3. Similar K_d values and relative ratios are observed when the ligands are titrated with the hMGL extracellular domain.

-Determination of ligands functional valance

In a multivalent glycocluster, the number of glycans or carbohydrate headgroups, which interact with the protein can differ from the number of glycan physically present in the chemical structure. The number of headgroups that interact with the protein is called functional valency, while the actual number of epitopes present in a structure is defined as structural valence.

The functional valence is calculated following the equation.³⁵

$$\text{Functional valence} = 1/N$$

The N values for the multivalent glycoclusters reveals that the structural valence coincides with the functional valence: **AG143** carries four GalNAc and its functional valence is exactly four for the hMGL CRD ($1/0.25=4$) and little over four for the hMGL extracellular domain iso 3 ($1/0.22=4.5$). The small discrepancy may be caused by a low degree of aggregation non-detectable by size exclusion. For the trivalent **AG124** the structural valence with the hMGL CRD is 2.8 and 3.0 for the hMGL extracellular domain iso 3.

Even if the K_d values and the functional valence values of multivalent and monovalent ligands are similar for hMGL CRD and extracellular domain hMGL isoform 3, a more detailed analysis of the thermodynamic binding parameters revealed significant differences in the binding mechanism.

-Macroscopic binding enthalpy: ΔH increases in direct proportion to the valency for *hMGL CRD* but the increase is not in direct proportion for *hMGL isoform3*

(i) *hMGL carbohydrate recognition domain (CRD)*

Comparing my data set with others present in literature,^{36,37} I can state that the observed macroscopic ΔH of the multivalent ligands scale proportionally to the number of binding epitopes. The macroscopic binding enthalpy is close to binding enthalpy of GalNAc multiplied by the number of glycocluster epitopes or headgroups.

The ΔH value for **AG143** is -45.1 kcal/mol, which is approximately four time more negative than the ΔH value for monovalent GalNAc, -8.7 kcal/mol. The **AG124** ΔH value is -26.4 kcal/mol, three times more negative than the GalNAc one. Small differences may arise due to the sugar headgroup not being 100% identical to GalNAc.

(ii) *Extracellular domain hMGL isoform 3*

ΔH of the multivalent ligands again scale to the number of binding epitopes but the increase is lower than the microscopic GalNAc ΔH multiplied by the number of epitopes, table 2. The ΔH values grow from -12 kcal/mol for the monovalent GalNAc, to -19 kcal/mol for the trivalent **AG124**, and -22.5 kcal/mol for the tetravalent **AG143**. The expected value for **AG124** would be -36 kcal/mol instead of -19 kcal/mol, **AG143** would be -48 kcal/mol instead of 22.5 kcal/mol.

-Macroscopic binding entropy: $T\Delta S$ does not increase in proportion to the valency for interactions with the *hMGL CRD* but does increase in proportion to valency for interactions with the *hMGL isoform3*.

(i) *hMGL carbohydrate recognition domain*

While the ΔH values scale proportionally, $T\Delta S$ was significantly more negative than if it had been predicted based on proportionally scaling according to the number of epitopes in the carbohydrates. For instance, the $T\Delta S$ value for **AG143** is -35.8 kcal/mol instead of a predicted -8 kcal/mol value and for **AG124** is -16.8 kcal/mol instead of a predicted -6 kcal/mol.

(ii) *Extracellular domain hMGL isoform 3*

In this case and contrasting with that above, $T\Delta S$ values for the multivalent glycocluster binding scale proportionally to the $T\Delta S$ values for the monovalent GalNAc. The $T\Delta S$ value

for **AG143** is -13 (kcal/mol) which is close to three times the $T\Delta S$ value of GalNAc. The $T\Delta S$ value for **AG124** is -9.7 (kcal/mol), close to the predicted -10.8 (kcal/mol), twice the GalNAc entropic contribution, table 2.

Discussion of ITC data: what are these differences in thermodynamic parameters telling us?

-hMGL CRD

From sequence comparison with other C-Type lectins and from the crystal structure presented in chapter 5, it is possible to evince that the hMGL CRD has a single binding site. From the functional valence we could deduce that more than one protein interacts simultaneously with the tested multivalent ligands, the thermodynamic profiles arising from the analysis reflect this situation. Thus, when I titrated the multivalent ligands with the h-MGL CRD, we observed ΔH scaling proportionally to the value of the monovalent GalNAc, while the $T\Delta S$ is much more negative than predicted when compared to the monovalent ligand. A great contribution to the interpretation of thermodynamic binding parameter as applied to multivalent carbohydrates binding to lectins was given by the Brewer group's work and insights can be gained from their analysis. According to the Brewer group analysis, this trend is characteristic of the binding of a multivalent ligand to **separate receptor molecules with multiple binding sites**.³⁷

Well studied examples with this thermodynamic profile are Concanavalin A (ConA) and *Dioclea grandiflora* lectin (DGL). These proteins have four carbohydrate binding sites, figure 26.

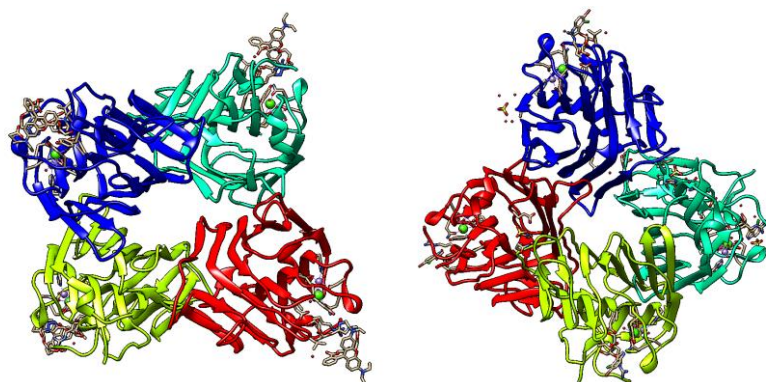


Figure 26. Structure of ConA on the left PDB: 4P9W , and DGL on the right PDB: 2JE9.

When titrated with tetravalent ligand binding involves the interaction of the ligand with one carbohydrate recognition domain from four different protein molecule.³⁵³⁸³⁹ For instance in Brewer and co-workers analysed a collection of multivalent molecules, figure 27, and reported that **FB10** has an affinity enhancement of 35-fold compare to **TriMan** for ConA, table3. For **FB10**, but the same trend is observed for the other multivalent ligands, the ΔH is -53 (kcal/mol) which is circa four times the ΔH of the **TriMan** -14.7 (kcal/mol). On the contrary **FB10** T ΔS is -43.3 (kcal/mol), much more negative than the T ΔS of **TriMan**. Therefore, the enhanced affinity is not to be attributed to **FB10** binding to single molecules of ConA that possess four binding sites, but to the binding of four separate molecule of ConA to **FB10**. This means that the observed K_a value for **FB10** is actually the average of the four microscopic K_a values at each of its four epitopes, since each epitope is involved in binding to a separate lectin molecule.

Table 3. Thermodynamic binding parameters for concanavalin A with multivalent sugars at 27 °C. Table adapted from *JBC*, 2000 19, (12),14223–14230.³⁵

	K_a ^a ($M^{-1} \times 10^{-4}$)	$-\Delta G$ (kcal/mol) ^b	$-\Delta H$ (kcal/mol) ^c	$-T\Delta S$ (kcal/mol) ^d	N^e
TriMan	39	7.6	14.7	7.1	1
FB7	286	8.8	23.1	14.3	0.53
FB8	250	8.7	26.2	17.5	0.53
FB9	420	9.0	29.0	20.0	0.51
FB10	1350	9.7	53.0	43.3	0.26

^a Errors in K_a are range from 1 to 7%.

^b Errors in ΔG are less than 1%.

^c Errors in ΔH are 1–4%.

^d Errors in T ΔS are 1–7%.

^e Errors in N are less than 2%

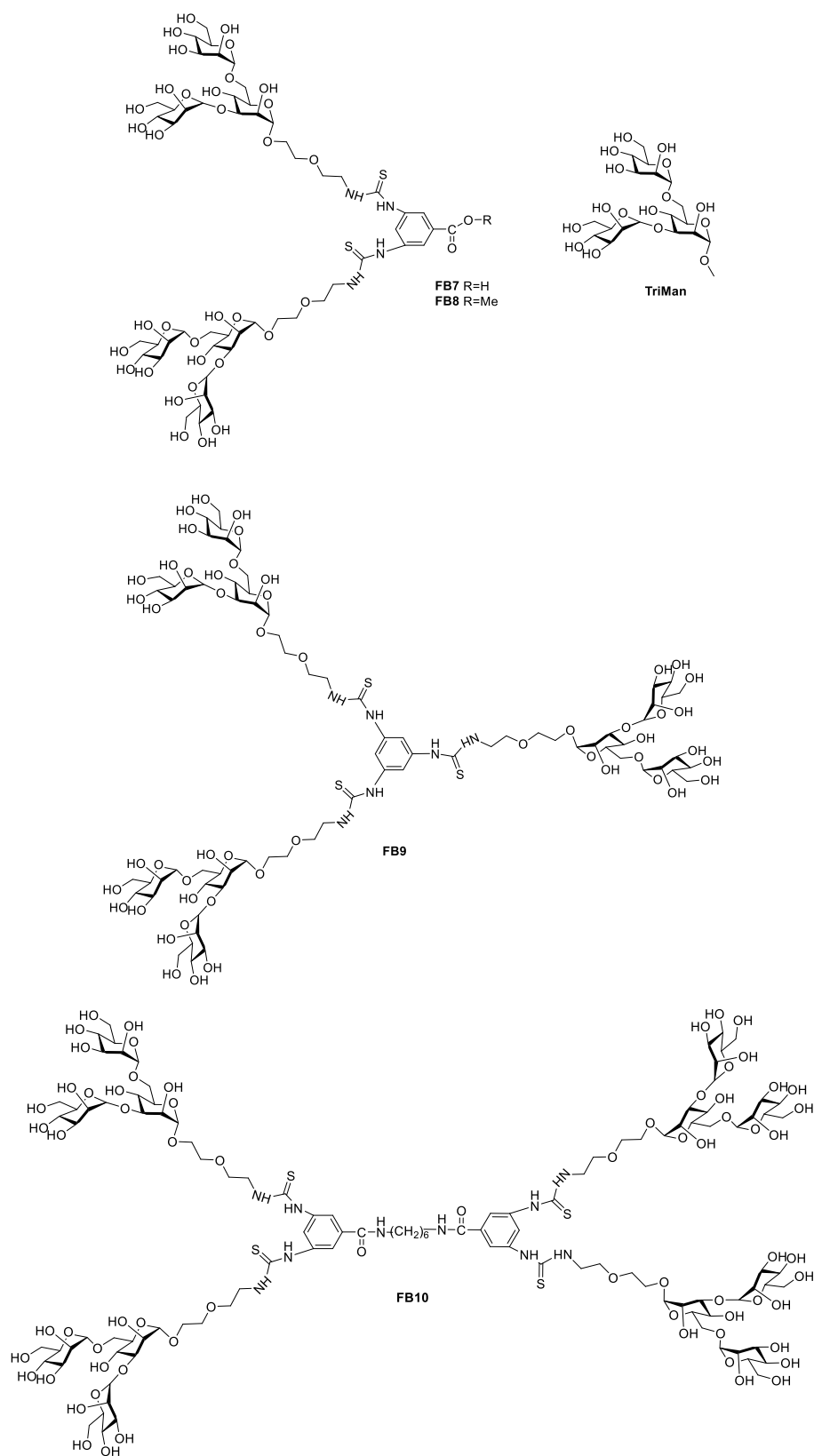


Figure 27. Structure of multivalent mannose derivatives tested for ConA and DGL. Adapted from *J. Biol. Chem.* 2000, 275, 14223–14230.³⁵

The thermodynamic profiles of **AG124** and **AG143** reported in table 2, resemble the one of **FB10** and bwith ConA and DGL rather than the one reported by Shahzad-ul-Hussan et al.⁴⁰ for the monovalent lectin microvirin. When microvirin is titrated with a glycocluster presenting three Man $\alpha(1-2)$ Man domains the Kd of the single unit Man $\alpha(1-2)$ Man is 47 μM , while that of the trivalent mannopyranoside is 6.25 μM , only seven-fold better. The binding stoichiometry is equal to 1:1. The ΔH for the monovalent mannoside is -7.3 ± 0.3 kcal mol⁻¹, not deviating significantly to the ΔH for the trivalent ligand, -6.3 ± 0.4 kcal mol⁻¹. The ΔS goes from negative in the monovalent ligand, -4.7 cal mol⁻¹, to positive for the trivalent ligand, 2.9 cal mol⁻¹. The binding is favoured by an entropic effect, most probably due to the statistical recombination model.

The different binding behaviour is probably to be attributed to the geometry of the tetraphenylethylene scaffold and, consequently, the special arrangement of the glycan epitopes, allow the interaction with more than one molecule of hMGL CRD. On the other hand, the trivalent mannoside synthesised by Shahzad-ul-Hussan et al. only permits the interaction of the trivalent compound with one molecule of Microvirin.

It is worth keeping in mind that for multivalent ligands, after the binding of the first recognition element, the loss of translational entropy is complete and subsequent binding interactions can proceed without paying an additional translational entropy penalty.⁴¹

In the case of the study with the hMGL carbohydrate recognition domain, trivalent **AG124** has a Kd 150-fold higher than GalNAc. The ΔH scale directly proportionally with the numbers of epitope, therefore, the Kd value for the multivalent glycoclusters is actually the average of the four microscopic Kd at each epitope, since each epitope is involved in binding to a separate hMGL CRD molecule. For this reason, the increase of the Kd require a more favourable T ΔS contribution compared to the monovalent GalNAc.

-hMGL isoform 3

Even if structural evidences are not reported in the literature, the MGL extracellular domain is believed to be a trimer,⁴² therefore it can behave as a single receptor with three carbohydrate binding sites. In the extracellular domain hMGL isoform 3 analysis, both ΔH and T ΔS scale proportionally to the numbers of epitope. The entropic contribution is characteristic of the binding of a multivalent ligand to a single multivalent receptor molecule.³⁷ A similar thermodynamic trend is found for a tailor made trimer of the receptor vancomycin with a trimeric ligand, figure 28. When the monovalent ligands **R'tL'** and **L'**

are analysed in the presence of the trivalent protein RtV₃, their ΔH are both -50.2 kJ/mol while the ΔH for the trivalent ligands **R'_tL'₃'** is -167 kJ/mol, three times higher. The same trend is observed for -T ΔS : ca 18.4 kJ/mol for the monovalent ligands and three times higher for the trivalent one. In this specific example the enhancement in binding reaches a K_d in the femto molar range for the trivalent ligand versus in the micromolar range for the monovalent ligand, table 4.⁴³

Table 4. Thermodynamic binding parameters for Vancomycin with multivalent sugars at 25 °C. Table adapted from Science 1998, 280, 708-711.⁴³

Receptor	Ligand	Receptor/ ligand	K _d (M)	ΔG (kJ/mol)	ΔH (kJ/mol)	-T ΔS (kJ/mol)
R _t V ₃	R' _t L' ₃	1:1	4 X 10 ⁻¹⁷	-94	-167	73
R _t V ₃	R' _t L'	1:3	1.1 X 10 ⁻⁶	-33	-50.2	17.2
R _t V ₃	L'	1:3	2.7 X 10 ⁻⁶	-31	-50.2	18.4
Vancomycin	R' _t L' ₃	3:1	0.34 X 10 ⁻⁶	-36.8	-73.2	36.4
Vancomycin	L'	1:1	1.6 X 10 ⁻⁶	-33.0	-50.2	17.2

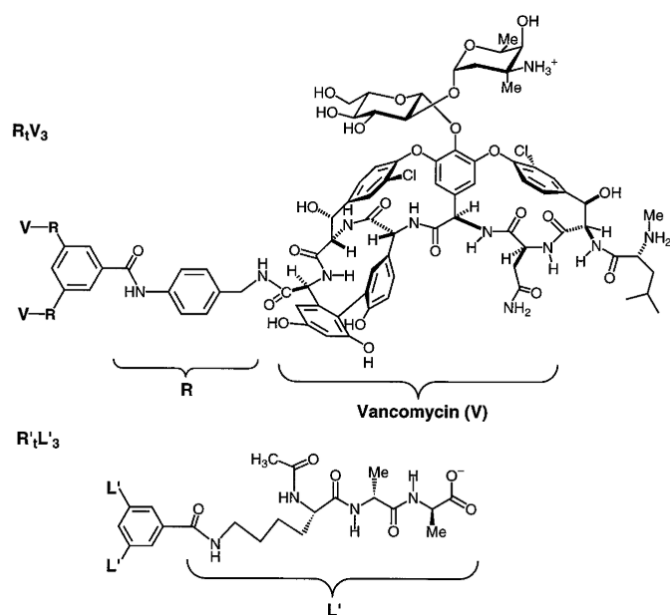


Figure 28. Structures of the trivalent derivatives of vancomycin, R_tV₃, and trivalent, R'_tL'₃ ligand. From Science, 1998, 280, 708-711.⁴³

The vancomycin engineered receptor/ligand couple is an ideal example of perfect geometrical fitting of a multivalent molecule in the multiple binding pocket of a receptor. Among other examples with similar thermodynamic profile, it is possible to find a trivalent glycocluster binding to the hepatic asialoglycoprotein with a K_d in the nM range compared to the mM range of the monovalent molecule.⁴⁴

Binding affinity of the tetravalent **AG143** for the protein is 164-fold higher than for GalNAc, while for **AG124** 126-fold. This modest increase in affinity, compared to the previously reported examples in the vancomycin area as well as for the asialoglycoprotein, require an explanation; in fact, while for multivalent ligands is observed direct proportionality for the $T\Delta S$, not the same is observed for ΔH which increases, but is not three and four times higher than for GalNAc. The magnitude of the enthalpic penalty may have its origins in the fact that tetraphenylethylene dimension does not permit face to face interaction or chelation, as mentioned in section 2.1 the distances between the carbohydrate headgroups are 14 Å by 23 Å while theoretical calculations show that the distances among CRDs in the trimeric protein are about 50 Å³ and this would be a minimum distance required for face to face interactions. The glycocluster ligands studied herein and based on TPE do not have large enough distances between the ligands to enable face to face or chelation type binding to the full-length trimer receptor. This could justify the decreased affinity of **AG124** and **AG142** compared to vancomycin chelating ligands: the ligands are too small to fit perfectly in the binding pocket and this translates in an enthalpy cost, K_d in the nano molar range rather than femto molar. This model though, does not explain the functional balance observed in the ITC experiments.

The future crystallisation of the extracellular domain of MGL and the thermodynamic analysis of mono and di-functionalised tetraphenylethylene glycocluster would provide further opportunities for thermodynamic analysis by ITC and further advancement in understanding the binding mode.

2.4. Papers

2.4.1. RSC Adv.2018,8,28716



RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: RSC Adv., 2018, 8, 28716

Revealing biomedically relevant cell and lectin type-dependent structure–activity profiles for glycoclusters by using tissue sections as an assay platform†

Herbert Kaltner,^a Joachim C. Manning,^a Gabriel Garcia Caballero,^a Claudia Di Salvo,^b Adele Gabba,^b Laura L. Romero-Hernández,^b Clemens Knospe,^c Dan Wu,^d Harrison C. Daly,^d Donal F. O'Shea,^d Hans-Joachim Gabius^a and Paul V. Murphy^{ib}*^b

The increasing realization of the involvement of lectin–glycan recognition in (patho)physiological processes inspires envisioning therapeutic intervention by high-avidity/specificity blocking reagents. Synthetic glycoclusters are proving to have potential for becoming such inhibitors but the commonly used assays have their drawbacks to predict *in vivo* efficacy. They do not represent the natural complexity of (i) cell types and (ii) spatial and structural complexity of glycoconjugate representation. Moreover, testing lectins in mixtures, as present *in situ*, remains a major challenge, giving direction to this work. Using a toolbox with four lectins and six bi- to tetraivalent glycoclusters bearing the cognate sugar in a model study, we here document the efficient and versatile application of tissue sections (from murine jejunum as the model) as a platform for routine and systematic glycocluster testing without commonly encountered limitations. The nature of glycocluster structure, especially core and valency, and of protein features, *i.e.* architecture, fine-specificity and valency, are shown to have an influence, as cell types can differ in response profiles. Proceeding from light microscopy to monitoring by fluorescence microscopy enables grading of glycocluster activity on individual lectins tested in mixtures. This work provides a robust tool for testing glycoclusters prior to considering *in vivo* experiments.

Received 23rd June 2018
Accepted 24th July 2018

DOI: 10.1039/c8ra05382k

rsc.li/rsc-advances

Introduction

The realization of the unsurpassed capacity of glycans to store biological information in a minimum of space enables elucidating the chain of molecular events, from oligosaccharide synthesis and establishing the spatial aspects of their presentation on glycoconjugates and membranes to actual cellular effects, to become a staple of current research.¹ Shaping a paradigm for this area, these efforts disclosed that reading of sugar-encoded messages by tissue lectins is an efficient means to facilitate molecular bridging. When occurring between cells and also between cells and the extracellular matrix (*in trans*), it can lead to adhesion.^{2–5} Cross-linking of constituents within

a membrane (*in cis*) will form so-called glycoconjugate-lectin lattices that can trigger ensuing signalling.^{6–10} In each case, the intimate interplay of structural and topological parameters on both sides minimises stabilisation of random contacts, thus underlying the functional pairing of a lectin with not just any glycoconjugate, but with its matching counterreceptor(s). Hereby, glycan-encoded messages are read and then translated into particular post-binding responses. This theme of mutually specific protein(lectin)–glycan recognition appears to have a broad physiological significance.^{11–14}

Naturally, these emerging insights give strong incentive to search for compounds that will block this type of interaction if clinically unfavourable, and mimetics of polyvalent glycans termed glycoclusters have already been found to be particularly active.^{15–18} Besides their design and synthesis, a central challenge to master toward a prospect of biomedical application is the experimental set-up to measure inhibitory capacity. Optimally, it should come as closely as possible to the *in vivo* situation (in animals and man) with its diversity on the levels of glycans presented by glycoconjugates^{19–22} and of cell types, at the same time allowing for wide-scale testing in a robust manner with a clear read-out. The experimental approach to meet this challenge presented herein is a logical extension of classical lectin histochemistry.

^aInstitute of Physiological Chemistry, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Veterinärstr. 13, 80539 Munich, Germany

^bSchool of Chemistry, National University of Ireland Galway, University Road, Galway, Ireland. E-mail: paul.v.murphy@nuigalway.ie

^cInstitute of Anatomy, Histology and Embryology, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Veterinärstr. 13, 80539 Munich, Germany

^dDepartment of Chemistry, Royal College of Surgeons of Ireland (RCSI), 123 St. Stephen's Green, Dublin 2, Ireland

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8ra05382k



Given their binding to certain glycans, labelled lectins from plants have acquired popularity as reliable and sensitive tools to map the distribution of their ligands in sections of any type of tissue.^{23–28} What commonly is a routine specificity control, *i.e.* the inhibition of lectin binding to cells in a tissue section by a sugar (mono- to oligosaccharide),^{24,27,29} has the attractive potential to be adapted to an assay for glycocluster efficacy, fulfilling the requirements for physiological relevance listed above. By assembling a toolbox of (i) seven mono- to tetravalent *N*-acetyl-D-galactosamine (GalNAc)-presenting compounds and (ii) four lectins that share nominal specificity to a monosaccharide (GalNAc), among them a human protein that binds to GalNAc presented by mucins (for an overview on mucins, please see³⁰), and then working with these two sets to systematically monitor staining and intensity profiles in sections of a model tissue, *i.e.* murine jejunum, by light microscopy, we provide the proof-of-principle case study to document this assay's applicability. In fact, it opens the route to activity grading for any compound and any lectin on standard tissue specimens, offering the possibility for having natural binding partners act on both sides. Considering the occurrence of endogenous lectins in mixtures *in vivo*, a highly relevant situation for glycocluster testing, we also succeeded in adapting the assay to determine the individual behaviour of lectins when tested in combination. Toward this end fluorescence microscopy at different wavelengths was performed.

Results and discussion

The toolbox: 1. glycoclusters and control compounds

The panel of glycocluster-based inhibitors comprises six bi- to tetravalent compounds (1–6), all presenting GalNAc with α -anomeric linkages as thioglycosides (Fig. 1). The parameters of scaffold structure and valency are independently varied within this group. Compounds 3 and 5 differ in valency but share scaffold structure; the three pairs of compounds, *i.e.* 1 and 6, 2 and 3 as well as 4 and 5, share valency but differ in scaffold structure. The group of test compounds is completed by adding an the α -GalNAc derivative 7 and two control compounds (8, 9) with the same bulky cores of compound 5 and 6, but presenting a non-cognate sugar, GlcNAc.

The synthesis of glycoclusters 1, 2 and 5 has been described previously.³¹ The route to obtaining trivalent compound 3 with its tetraphenylethylene core is shown in Scheme 1. It commenced from the tetraphenylethylene derivative 10.³¹ Monoacetylation using acetic anhydride in limiting quantity (0.7 equiv.) followed by propargylation of the remaining phenols gave 11. Next, the copper-catalysed azide-alkyne cycloaddition (CuAAC) between the known azide 12 was performed, followed by deacetylation of the product to give 3, in which three of four positions are occupied by a sugar headgroup (in comparison to tetravalent 5).

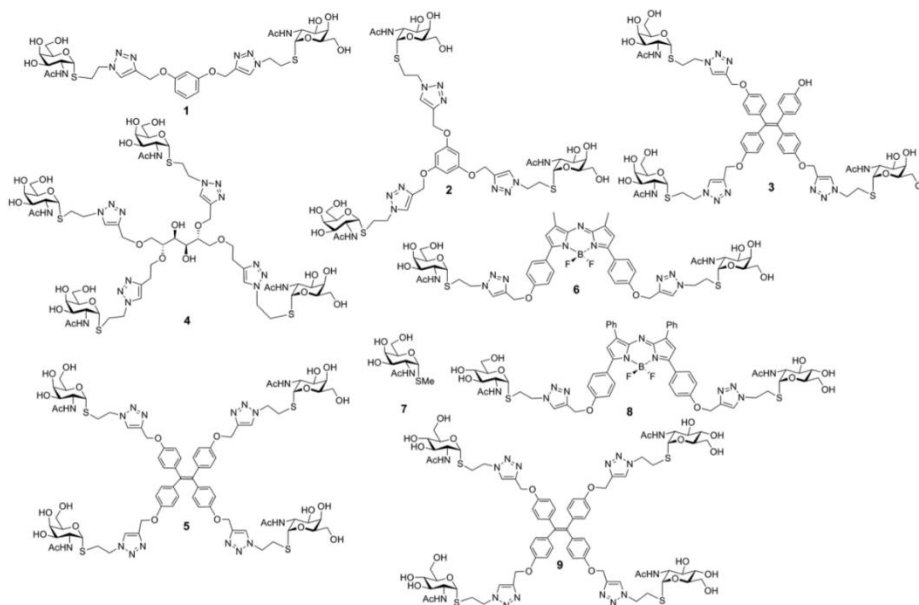
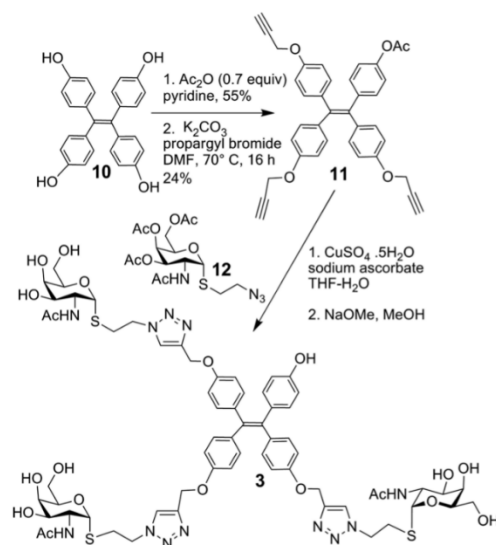
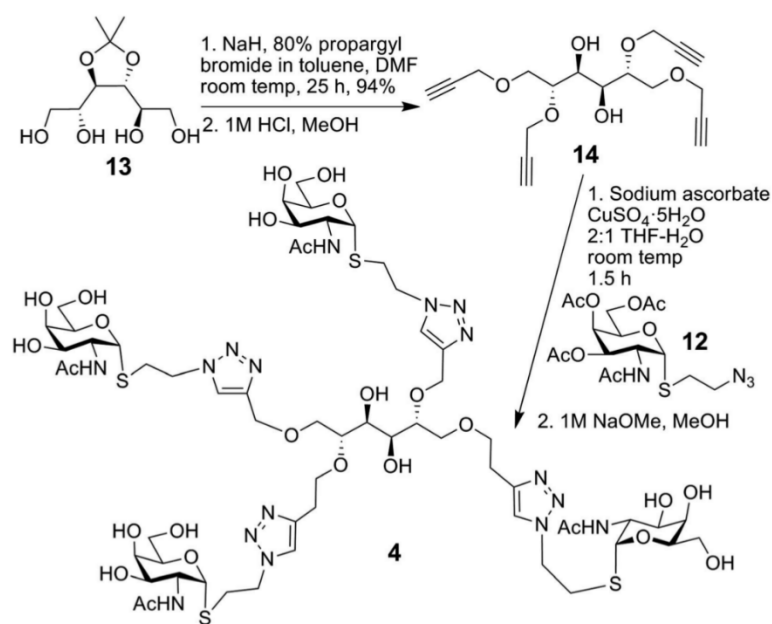


Fig. 1 Structures of glycoclusters 1–6 (with GalNAc as headgroup) and 8, 9 (with GlcNAc) as well as of the monovalent α -anomeric thioglycoside derivative 7.



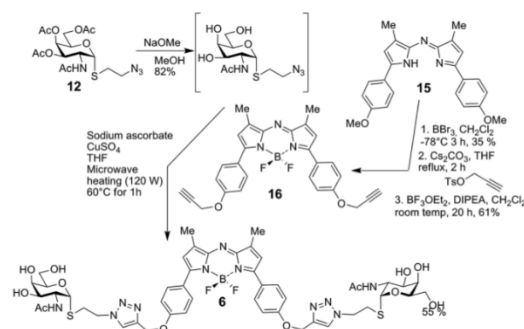


Scheme 1 Synthesis of 3.

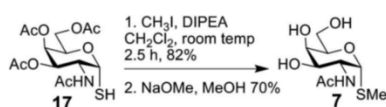


Scheme 2 Synthesis of 4.





Scheme 3 Synthesis of 6.



Scheme 4 Synthesis of 7.

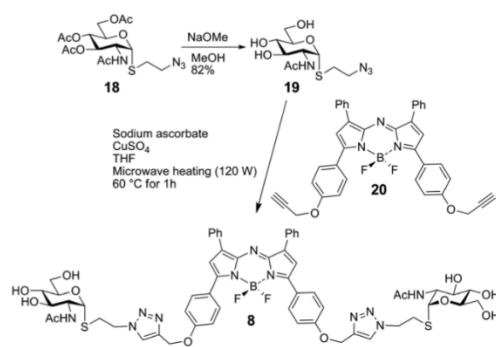
The synthesis of compound 4, using conjugation sites on *D*-mannitol to implement tetravalency, started with 13 and is summarised in Scheme 2.³² Tetrapropargylation was followed by removal of the acetonide group, resulting in 14. Next, CuAAC with 12 and subsequent deacetylation generated 4.

The bivalent compound 6 was prepared as summarised in Scheme 3. Firstly, 12³¹ was deacetylated, and then, with heating, the CuAAC reaction using dialkyne 16³³ produced 6. The dialkyne 16 was prepared in three steps from known 15.³⁴ Following the preparation of the GalNAc-presenting glyco-clusters 1–6, a monovalent α -anomeric GalNAc derivative was prepared as a probe to measure anomeric preference of lectin

binding, when set in relation to results with free GalNAc as standard. In detail, methyl α -thioglycoside 7 was obtained in two steps from glycosyl thiol 17³⁵ (Scheme 4). In addition to this panel of compounds with cognate sugar, an *N*-acetyl-*D*-glucosamine (GlcNAc)-bearing bivalent analogue of compound 6 was synthesized. Its availability allows to perform controls that exclude interaction of glycoclusters with bulky core with lectins by carbohydrate-independent molecular stickiness. This bivalent GlcNAc derivative 8³³ was prepared *via* 18–20,³³ as summarised in Scheme 5.

As further control compound for specificity testing, we included compound 9, a GlcNAc-presenting analogue of compound 5. It had proven to be bioactive as potent inhibitor of two plant lectins with this monosaccharide specificity, *i.e.* WGA and GSA-II,³⁶ here served as additional control to trace carbohydrate-independent signal generation.

In order to characterize spatial aspects of valency on the tetraphenylethylene, *D*-mannitol and BF₂-azadipyromethene cores, distances between the hydroxyl groups were calculated by molecular modelling based on crystal structures obtained from



Scheme 5 Synthesis of 8.



the Cambridge Crystallographic Database in three cases^{34,37} (Fig. 2). In fact, the nature of the core will constrain the maximum spacing possible between GalNAc residues in the most extended conformations. As shown in Fig. 3, these distances in the tetraphenylethylene derivative **3** can likely be larger than for the D-mannitol-based **4** (29 Å vs. 20 Å in Fig. 3). The BF₂-azadipyrromethene scaffold facilitates a spacing that could extend to ~28 Å between its two GalNAc residues, again longer than for the D-mannitol derivative **4** (Fig. 3) and also

longer than for **1**, the latter estimated at ~22 Å in our earlier work.³¹ The D-mannitol- and tetraphenylethylene-based glycoclusters can thus be considered as cornerstones for the spectrum of degree of compactness and of distance between sugar headgroups.

In summary, the panel of glycoclusters covers different aspects of valency and inter-headgroup spacing for the same sugar used as lectin ligand. The effect of a glycocluster on this activity can likely vary between lectins. As consequence, it will best be measured, if GalNAc-specific receptors with different characteristics in terms of spatial presentation of contact sites for the sugar are selected. Of note, spatial valency on both sides of the recognition process can cooperate *in situ* so that its complementarity can result in a Lego-like association with beneficial effects on affinity by lowering off-rates^{38,39} and the

Open Access Article. Published on 14 August 2018. Downloaded on 11/11/2018 8:13:56 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

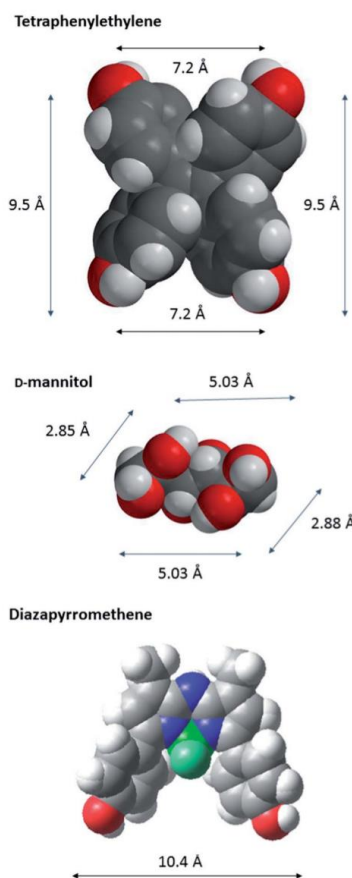


Fig. 2 CPK models of the core scaffolds tetraphenylethylene (a), D-mannitol (b), and BF₂-azadipyrromethene (c), on which GalNAc-containing glycoclusters studied herein were built. The distances between O-I atoms are shown; those in D-mannitol are between O atoms at C-1, -2, -5 and -6. The models are based on data derived for crystal structures obtained from the Cambridge Crystallographic Database.³⁷ The CPK models were generated in Macromodel (<http://www.schrodinger.com>).

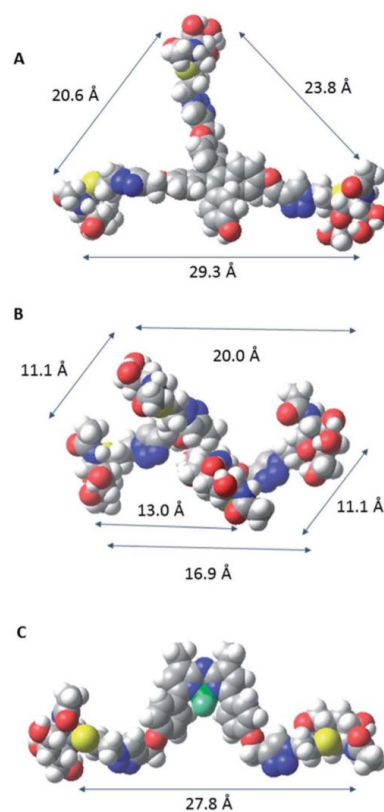


Fig. 3 CPK models of selected conformers of compounds **3** (a), **4** (b) and **6** (c). Distances between anomeric carbons of the GalNAc residues presented by each core in these conformers are shown. The CPK models were built in Macromodel.

inhibitory potency.^{40,41} Besides spatial aspects, the fine-specificities of lectins may well differ. Since more than a dozen folds with diverse molecular architecture and quaternary structure have developed the capacity to act as lectins, "lectins considered 'identical' in terms of monosaccharide specificity possess the ability to recognize fine differences in more complex structures"⁴² and protein engineering facilitates custom-made tailoring, this is possible: the group of selected lectins is presented in the next section.

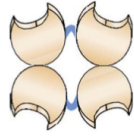
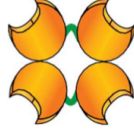


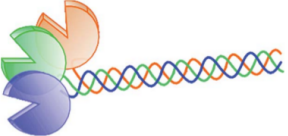
The toolbox: 2. lectins

The panel of lectins comprises plant, invertebrate and human proteins that differ in fine-specificity and valency (Table 1). In detail, two tetravalent leguminous lectins (*Dolichos biflorus* agglutinin (DBA) and soybean agglutinin (SBA)), a hexameric lectin from the albumen gland of the edible (or vineyard) snail (*Helix pomatia* (HPA)) and the human macrophage galactose (-binding C)-type lectin (MGL) in two engineered forms, *i.e.* as

free monomeric carbohydrate recognition domain (CRD) and as CRD extended by the α -helical coiled-coil stalk of the transmembrane receptor that confers its trimerisation. Table 1 presents schemes of the quaternary structures and listings of specificity profiles (for details on fine-specificity, please see;⁴³⁻⁵⁴ for details on structural aspects, please see^{48,50,55-60}). As documented in this table, the relative degrees of affinity to natural GalNAc-containing glycans are not identical. As a consequence, extents of cell binding are likely to differ for these lectins, and this has for example been indeed revealed by array analysis in the cases of Chinese hamster ovary (CHO) cells and three of its glycosylation mutants (Lec1, Lec2, Lec8) for DBA, SBA and HPA.^{61,62}

In summary, the panel of lectins covers five proteins with well-characterised biochemical characteristics (Table 1). A human lectin is deliberately included to document feasibility to work with endogenous (tissue) lectins. The effect of glycocluster presence on ligand binding in a lectin histochemical model

Table 1 Molecular characteristics of the four tested lectins

Acronym	Quaternary structure (number of binding sites) ^a	Glycoligand
DBA	Tetramer (4) 	α -GalNAc, GalNAcz3GalNAcz3Gal β 4Gal β 4Glc > A-tetrasaccharide; GalNAcz4Gal/GalNAcz β 3Gal (P-like)/GalNAcz β 4Gal (asialo CAD)/T(F) disaccharide only weakly active
SBA	Tetramer (4) 	α/β -GalNAc, GalNAcz3Gal(β 6Glc); GalNAcz4Gal, GalNAcz β 3/4Gal more active than T _n antigen, α 2fucosylation of A-disaccharide reduces activity, T(F) disaccharide only weakly active
HPA	Hexamer (6) 	α -GalNAc > β -GalNAc/ α -GlcNAc, GalNAcz3GalNAc/GalNAcz3/4Gal/GalNAcz β 4Gal as active as T _n (GalNAcz β 3 only weakly active), (s/su)T(F) antigens active, GlcNAcz3/4-terminated LacNAc; highly potent precipitinogen/sensor for low-density presentation of T _n antigen
	CRD monomer (1) 	
MGL	(CRD + stalk) trimer (3) 	T _n and s/suT _n antigens (T _n presentation by Tyr active), core 5/6 mucin-type O-glycans, β 4-linked GalNAc in LacdiNAc or in chains of GD2/GM2 (not GD3/GM3) gangliosides

^a Estimated/measured distances between contact sites for the ligand: DBA, SBA about 55 Å or 68 Å; HPA, hexamer arranged as two flat surfaces about 100 Å apart, each formed by a trimer with inter-contact site distance of 25 Å; MGL trimer about 50 Å.



study will best be measured in sections of a tissue, to which all tested proteins can bind, and this then yielding strong signals, most preferably for more than a single cell type. Known binding properties of gastrointestinal tissue to this group of proteins when used as histochemical probes^{63–67} guided us to select murine jejunum.

The toolbox: 3. murine jejunum

Main aspects of the structure of jejunal villi and crypts (of Lieberkuehn) are illustrated in Fig. 4, where a schematic drawing is combined with microphotographs of sections at two levels of

magnification. Running the lectin histochemical protocol without the incubation step with a labelled sensor for GalNAc-containing glycoconjugates led to no generation of a signal, that is to no change in appearance of the haemalaun-stained tissue (Fig. 5a). If generated, the signals will thus be entirely dependent on the presence of a labelled lectin.

In order to record stainings for each protein and to determine optimal conditions for applying glycoclusters as inhibitors for lectin binding systematic titrations of lectin concentrations were performed for each lectin. Staining was routinely assessed as the intensity of the signal and as percentage of positive cells. An example for the correlation

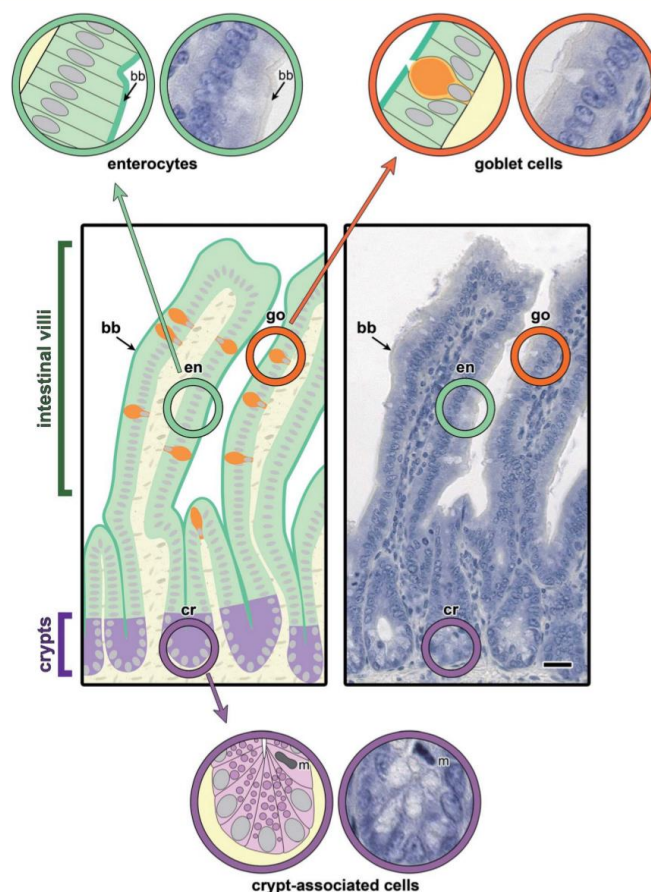


Fig. 4 Illustration of the morphological aspects in a longitudinal section of fixed murine jejunum by a combination of schematic drawings with colour coding and microphotographs at two levels of magnification. Representative regions in intestinal villi (en, enterocytes; go, goblet cells; bb, brush border) and also crypt-associated cells (incl. Paneth cells, enteroendocrine cells, precursors of enterocytes and goblet cells and mitotic cells) at the basis of the crypts (cr) were marked by colour-coded circles, these areas shown in enlarged representation.

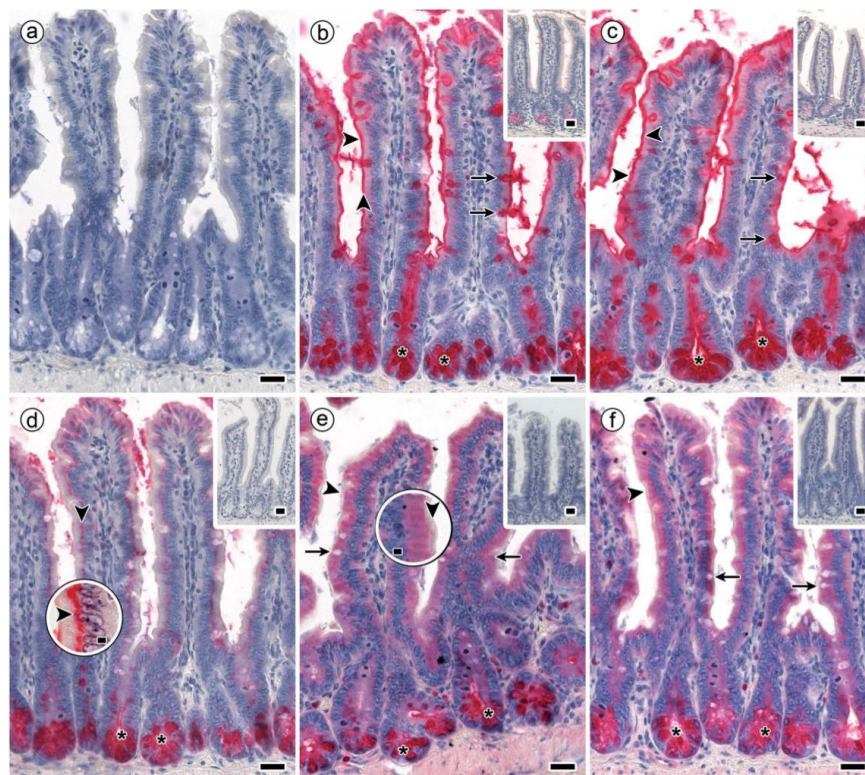


Fig. 5 Staining profiles by biotinylated DBA, SBA, HPA and the two forms of MGL in longitudinal sections through fixed murine jejunum. Incubations were done with concentrations of lectins that lead to a strong signal with minimal background (100% level). (a) Negative control by omission of the incubation step with first-step reagent (labelled lectin) with no evidence for lectin-independent signal generation (blank value as shown in Fig. 4) (b and c) strong reactivity for DBA (b) and SBA (c) was found in the brush border of surface enterocytes from intestinal villi (arrowheads), goblet cells (arrows) and in crypt-associated cells, especially in the deep parts of the crypts (asterisks). (d) Binding of biotinylated HPA was detected supranuclearly in surface enterocytes from intestinal villi (arrowheads) (high-level magnification of this region is shown in an inserted circle above the respective area) and in the deep parts of the crypts (asterisks). (e and f) MGL-dependent positivity (similar for the two protein forms) in crypt-associated cells, most prominently in the crypts' deep parts (asterisks) and comparatively weaker in the cytoplasm of surface enterocytes from intestinal villi. No staining was seen in the brush border (arrowheads; higher magnification is shown in an inserted circle above the respective area). Insets in panels b to e show extent of blocking lectin binding by co-incubation of the lectins with free GalNAc at 200 mM. Concentration of biotinylated lectins used were $1 \mu\text{g mL}^{-1}$ for DBA, $3 \mu\text{g mL}^{-1}$ for SBA, $1.5 \mu\text{g}$ for HPA and $4 \mu\text{g mL}^{-1}$ for both engineered forms of MGL. Scale bars are $20 \mu\text{m}$ (a–f) and $5 \mu\text{m}$ (circles in d and e).

between lectin concentration and staining intensity/number of positive cells is shown in ESI Fig. S1† for SBA. For each lectin, signal increases reached plateau values, revealing that binding was saturable (for details on titration ranges, please see Experimental section). These titration series thus led to defining experimental conditions to reach signals of strong intensity as starting point for testing glycoclusters.

The signals were present in different regions of the tissue sections. As summarized in Table 2, surface enterocytes and crypt-associated cells together with goblet cells, here primarily

apically targeting the mucus, are the main sites for lectin binding. Binding to intestinal villi exhibited differences, whereas signal intensity of crypt-associated cells was rather uniform (Table 2). Exemplary microphotographs illustrate the staining profile for each type of lectin (Fig. 5b–f). They also document (by insets) reduction of staining by inhibition of lectin binding by the cognate monosaccharide used at high concentration (Fig. 5b–f). In sum, staining was dependent on the lectin, saturable, inhibitable by cognate sugar and different with respect to cell and lectin types.



Table 2 Binding of lectins to sections of murine jejunum

	DBA	SBA	HPA	MGL	
				CRD	CRD + stalk
Intestinal villi					
Surface enterocytes					
Cytoplasm	+(3) ^c	+(3)	–	+++ (8)	+++ (8)
Subapical cytoplasm	+(3)	+(3)	–	+++ (8)	+++ (8)
Supranuclear cytoplasm	–	++(4)	+++ (8)	+++ (8)	+++ (8)
Brush (striated) border	+++ (8)	+++ (8)	+/(–)(2)	–	–
Goblet cells	+++ (8)	+++ (8)	–	–	–
Lamina propria					
Immune cells	–	–	–	–	–
Fibroblasts	–	–	–	–	–
Smooth muscle cells	–	–	–	–	–
Endothelial cells ^a	–	–	–	–	–
Crypts of Lieberkuhn					
Crypt-associated cells ^b					
Cytoplasm	+++ (8)	+++ (8)	+++ (8)	+++ (8)	+++ (8)
Subapical cytoplasm	+++ (8)	+++ (8)	+++ (8)	+++ (8)	+++ (8)
Supranuclear cytoplasm	+++ (8)	+++ (8)	+++ (8)	+++ (8)	+++ (8)
Goblet cells	+++ (8)	+++ (8)	–	–	–

^a Capillaries of the lamina propria. ^b Incl. Paneth cells, enteroendocrine cells, precursor of enterocytes and goblet cells, and mitotic cells. ^c For details on categories of semiquantitative assessments in terms of intensity of lectin-dependent signal (– to +++)/number of positive cells (0–100%), please see section on lectin histochemistry in Experimental.

Having equipped the toolbox with panels of synthetic glycoclusters and lectins, also selected a tissue for this model study, whose sections will present carbohydrate-inhibitable signals in distinct cell types, and defined parameters of the histochemical protocol, systematic titrations with the panel of glycocompounds shown in Scheme 1 were performed at a certain lectin concentration to determine relative activities in each case for each type of lectin. Specimens of four animals were processed, in each case at least with duplicates obtained as serial sections in parallel and examined independently by two observers.

Glycoclusters as inhibitors of tissue staining by lectins:

1. light microscopy

The systematic titrations of glycocompound concentration at the constant concentration of a lectin revealed bioactivity of GalNAc, as free monosaccharide or as part of glycoclusters. In other words, lectin binding to GalNAc-containing glycans of cellular glycoconjugates was impaired by the presence of GalNAc in the solution, as free monosaccharide or derivative thereof (compound 7) or as part of a glycocluster (compounds 1–6), in a concentration-dependent manner, as seen in the insets of Fig. 5b–f. The extent of distinctly non-uniform responses depended on the nature of the test compound and the lectin. Exemplary illustrations of how staining profiles are affected by presence of increasing concentrations of the inhibitor are presented for the pairs 3/DBA and 2/SBA (Fig. 6). As shown in this figure, staining (in terms of signal intensity and percentage of positive cells) was reduced by the presence of

a GalNAc-presenting compounds. Conjugation of the cognate sugar did obviously not impair its lectin-binding activity. Expressed in terms of the concentration that caused 50% inhibition (IC₅₀) of signal intensity and cell positivity, respective values for inhibitory activity are compiled for each compound and tissue constituent in Table 3.

This summary teaches a series of lessons, starting with the broad range of concentrations of sugar needed to diminish the signal within this lectin panel. Proceeding from the arrest induced by the α -S-glycoside 7, the chemical design of bi- to tetravalent glycoclusters with α -GalNAc as headgroup is revealed to be capable to lead to up to marked increases (up to 10 000 fold for MGL) of inhibitory capacity of GalNAc (Table 3). Graphically, this point is underscored by comparatively documenting staining profiles by the five labelled proteins in the presence of a constant concentration of tetravalent compound 5 and of bivalent compound 6 (Fig. 7). Since these two types of core are especially bulky and prone to engage in hydrophobic interactions, we tested β -GlcNAc-presenting homologues 8 and 9 to exclude the concern of carbohydrate-independent (non-specific) binding. By substituting cognate GalNAc by its non-cognate epimer GlcNAc, the resulting glycoclusters should lose their capacity to inhibit binding of the lectins. As shown in ESI Fig. S2,† the two compounds with this non-cognate sugar have no inhibitory capacity.

Considering inhibitory activity, compound core structure (for 1 vs. 6, 2 vs. 3 and 4 vs. 5), valency of the core and the protein, also – but less so – valency of the protein in the two different forms of MGL appear to matter. Already bivalency



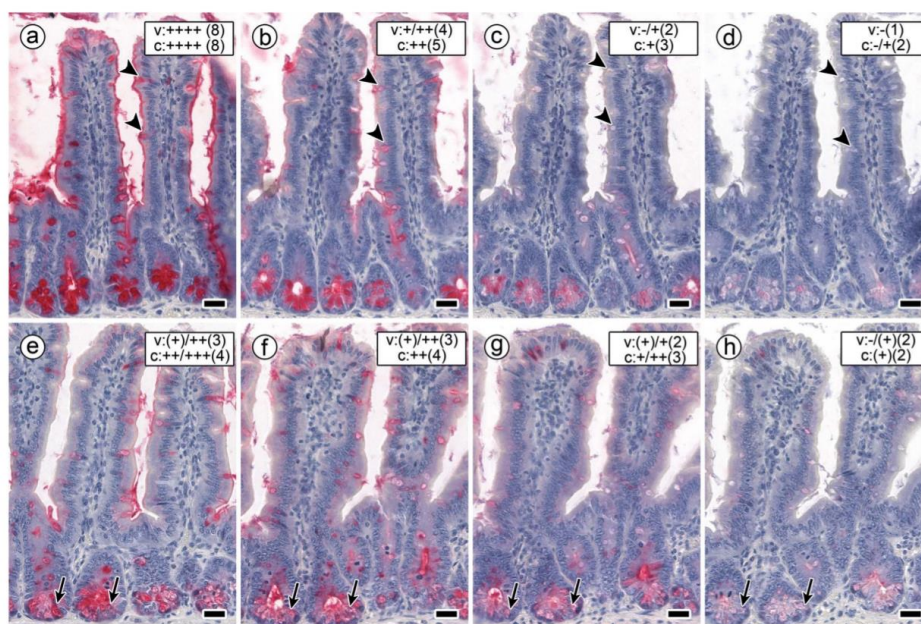


Fig. 6 Staining profiles by biotinylated DBA (a–d, arrowheads mark staining of goblet cells in villi) and SBA (e–h, arrows mark staining of crypt-associated cells) in sections of fixed murine jejunum in the presence of increasing concentrations of cognate sugar (GalNAc) assessed in terms of staining intensity (percentage of positive cells). The signal for DBA binding remained at the 100% level in the presence of 1 mM GalNAc presented by compound **3** (a). When raising the sugar concentration stepwisely to 2 mM (b), 5 mM (c) and 10 mM (d), respectively, scaffold-presented GalNAc (by compound **3**) led to a notable degree of inhibition. SBA-dependent staining intensity and percentage of positive cells were seen to be reduced more strongly at comparatively low concentrations of 0.001 mM (e), 0.05 mM (f), 0.1 mM (g) and 0.5 mM (h) of GalNAc-presenting compound **2**. Semiquantitative grading of staining intensity/percentage of positive cells is given in the rectangular box in the top-right area of each microphotograph (v, surface enterocytes of intestinal villi; c, crypt-associated cells). Intensity of staining in sections is grouped into the following categories: –, no staining; (+), weak but significant staining; ++ medium staining; +++ strong staining; +++++, very strong staining. Percentage of positive cells is expressed in the following eight categories: (1) 0% (no positive cells), (2) < 5% (few positive cells), (3) 5–20%, (4) 21–40%, (5) 41–60%, (6) 61–80%, (7) 81–95% (few negative cells), (8) up to 100% (no negative cells). Concentration of biotinylated DBA/SBA: $1 \mu\text{g mL}^{-1}/3 \mu\text{g mL}^{-1}$. Scale bars are $20 \mu\text{m}$.



brings about marked activity increases related to free GalNAc, as for example the case of HPA blocking by compound **6** (but not **1**) attests. Overall, structural design of each type of glycocluster accounts for differences in inhibitory potency on this hexavalent lectin (Table 3), indicative of the assumed mutual complementarity. The mechanism(s) underlying the inhibition of lectin binding will involve intra- and/or intermolecular bridging, and, “depending on the size and geometry of proteins and ligands, several binding modes may be possible and effective in parallel and in varying proportions.”⁶⁸

The comparison of the data sets for the two leguminous lectins discloses large disparities in affinity already at the monosaccharide level, fully in line with published data.^{43,45} These results emphasise that simple extrapolations between related proteins are not possible, a strong caveat when considering activity on members of families of tissue lectins

such as C-type lectins or galectins. Since they are found to be co-expressed in tissues by immunohistochemical monitoring with non-cross-reactive antibodies, as for example accomplished for galectins,^{69–71} the individual testing of each protein will in the first step be mandatory, as documented herein, then taking the methodological step to testing lectins in mixtures.

In addition to glycocluster and lectin parameters, the assay also provides information on cell types. As shown in Fig. 7 (listed in more detail in Table 3), the reactivity profiles of two cell populations, *i.e.* surface enterocytes and crypt-associated cells, exhibit disparities. In common binding assays using a certain glycoprotein as ligand, occurrence of such varying response profiles would remain undetectable, because cell type-dependent glycome differences are not established. Having glycome complexity present in this assay is thus

Table 3 IC₅₀-values of glycoclusters (given in mM based on GalNAc) and the cognate monosaccharide in lectin histochemical assays with biotinylated lectins applied to sections of fixed murine jejunum

Compound/cell type	DBA	SBA	HPA	MGL	
				CRD	CRD + stalk
1					
Surface enterocytes	n.i. ^a	10.0 (0.05) ^b	>10.0 (<0.5)	>20.0 (<1.25)	>20.0 (<1.25)
Crypt-associated cells	n.i. ^a	20.0 (0.1)	>20.0 (<0.25)	>20.0 (<2.5)	>20.0 (<2.5)
2					
Surface enterocytes	1.0 (25)	0.02/0.01 (50/100)	5.0 (1.0)	1.0 (25)	1.0 (25)
Crypt-associated cells	2.0 (25)	0.01 (200)	5.0 (1.0)	1.0 (50)	2.0 (25)
3					
Surface enterocytes	2.0 (12.5)	0.005/0.002 (200/500)	0.1/0.05 (50/100)	0.05 (500)	0.1 (250)
Crypt-associated cells	5.0 (10)	0.005/0.002 (200/500)	0.5/0.2 (10/25)	0.05 (1000)	0.1 (500)
4					
Surface enterocytes	1.0 (25)	0.005 (200)	1.0 (5.0)	0.05/0.02 (500/1250)	0.1/0.05 (250/500)
Crypt-associated cells	5.0 (5)	0.01 (200)	5.0 (1.0)	0.05/0.02 (1000/2500)	0.1/0.05 (500/1000)
5					
Surface enterocytes	2.0 (12.5)	0.001/0.0005 (1000/2000)	0.05/0.02 (100/250)	0.01/0.005 (2500/5000)	0.005 (5000)
Crypt-associated cells	5.0 (10)	0.001 (2000)	0.02 (250)	0.01/0.005 (5000/10 000)	0.005/0.002 (10 000/25 000)
6					
Surface enterocytes	20 (1.4)	0.1 (10)	0.1 (50)	0.2 (125)	0.1 (250)
Crypt-associated cells	>20 (<2.5)	0.2 (10)	0.2 (25)	0.2 (250)	0.2/0.1 (250/500)
7					
Surface enterocytes	20 (1.25)	0.02 (50)	0.5 (10.0)	5.0 (5)	10.0 (2.5)
Crypt-associated cells	>20 (<2.5)	0.05 (40)	2.0 (2.5)	1.0 (50)	2.0 (25)
D-GalNAc					
Surface enterocytes	25.0	1.0	5.0	25.0	25.0
Crypt-associated cells	50.0	2.0	5.0	50.0	50.0

^a n.i.: not inhibitory; tested up to 20 mM. ^b Numbers in parentheses denote relative inhibitory potency using activity of free GalNAc as standard set to 1.



certainly an advantage, and the detection of cell type-dependent disparities underscores the merit of this system. What is more, tissue sections as test platform also offer a means to address the mentioned issue that tissue lectins can be co-expressed *in situ*.

Of note, these lectins can exhibit overlapping carbohydrate specificities, as known from several lectin families.^{72,73} When co-expressed, they can obviously cooperate with additive or antagonistic impact on the functional outcome, as (first) case studies on galectins are revealing.^{74–77} Thus, being able to determine binding profiles of individual lectins and effect of a glycocluster on them at the same time in mixtures of tissue lectins would be ideal, and there is a histochemical way to accomplish this. Different modes of protein labelling facilitate to perform these measurements. In detail, different fluorescent dyes can be used for protein labelling (SBA, MGL) or a two-step procedure with a fluorescent probe to detect biotinylated proteins (DBA, HPA) in the tissue is performed.

Glycoclusters as inhibitors of tissue staining by lectins: 2. fluorescence microscopy

Both procedures led to the typical staining profiles of the tested lectins, as shown for each lectin in ESI Fig. S3† together with the essential specificity control, here presence of compound 5. The comparison of the staining distribution in the absence (panels a–d) and presence of compound 5 (panels e–h) unveils a rather similar signal pattern, as obtained by processing with light microscopical assessment. This comparative analysis built the basis to proceed to perform double and triple staining.

When testing mixtures, competition between lectins for binding sites can ensue, as is the case (patho)physiologically. The resulting staining profiles then inform about the extent of this competition. An example of staining after testing a binary mixture is presented in ESI Fig. S4† in the case of the two leguminous lectins. Presentation of the microphotographs recorded in each channel and of the (photo)merged files documents the merits of this procedure, in the absence of inhibitor (Fig. S4a–c†)

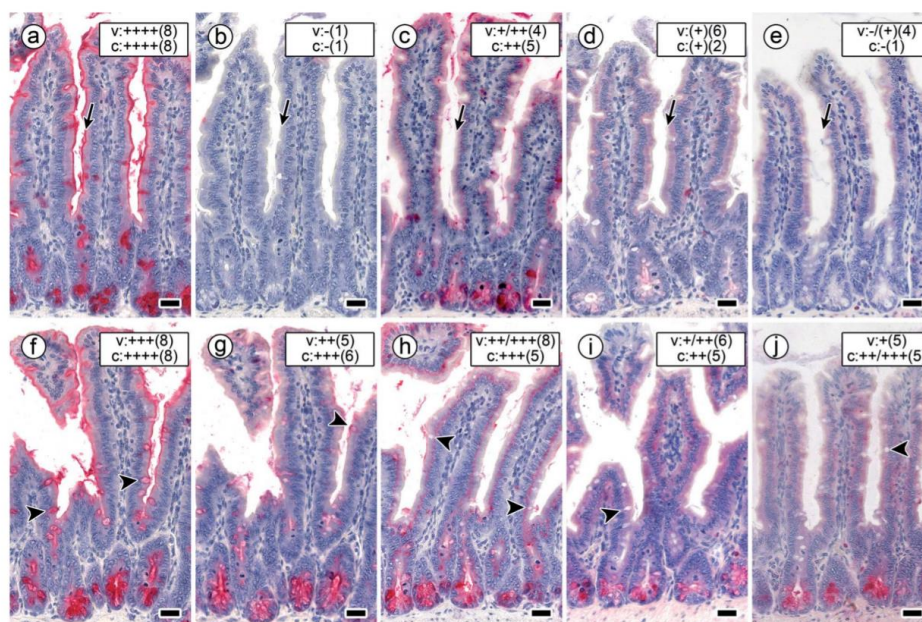


Fig. 7 Staining profiles by the four labelled lectins in longitudinal sections of fixed mouse jejunum in the presence of a constant concentration of cognate sugar (0.02 mM in compound 5; arrows mark staining of the brush border). a–e Signal intensity and percentage of positive cells remained at the 100% level after applying biotinylated DBA (a), whereas both parameters dropped to zero levels when using biotinylated SBA (b). This set-up reduced HPA-dependent binding to nearly 50% level (c). Binding of both forms of MGL was strongly inhibited (d and e). The concentration of 0.2 mM GalNAc presented by the bivalent compound 6 (f–j; arrowheads mark staining of goblet cells) had a slight effect on DBA staining in surface enterocytes (f). This concentration and this type of GalNAc presentation significantly reduced the intensity of the staining patterns of the other lectins applied (for details, please see Table 3). Binding of SBA (g) and HPA (h) was subject to a reduction by about 50% in surface enterocytes of intestinal villi (SBA: brush border, HPA: supranuclearly) and by a less diminished extent of staining intensity in crypt-associated cells. (i and j) The degree of inhibition was strongest for both forms of MGL so that the remaining staining intensity in surface enterocytes of intestinal villi and crypt-associated cells was categorized as weak to medium (for details on categories, please see legend to Fig. 6). Concentration of biotinylated lectins used were $1 \mu\text{g mL}^{-1}$ for DBA, $3 \mu\text{g mL}^{-1}$ for SBA, $1.5 \mu\text{g}$ for HPA and $4 \mu\text{g mL}^{-1}$ for the two engineered forms of MGL. Scale bars are 20 μm .



and in the presence of compound 5 (Fig. S4d–ff). When adding fluorescein isothiocyanate (FITC)-labelled MGL to this binary mixture, a solution containing three lectins is established. Binding of each of these lectins to GalNAc-containing glycoconjugates can now be monitored independently at three wavelengths, and the effect of presence of an inhibitor on each lectin in the mixture is readily accessible.

The obtained results of this test case with the three lectins using three-colour staining the tetraivalent compound 5 as inhibitor are presented in Fig. 8. Its panels show measurements without/with inhibitor at each wavelength (Fig. 8a–c and d–f) and the staining profile after (photo)merging the three individual files (Fig. 8g and h). Should it be the aim to determine individual staining profiles separately, in this setting precluding competition among lectins for binding sites, then processing of serial sections is advised. Here, the inherent slight changes of spatial presentation of cellular

constituents, as given in Fig. S5,† needs to be taken into account.

Using the test platform of the tissue sections with their glycome complexity, any combination of lectins hereby becomes amenable to assaying, that is proteins of the same family and from different families and especially tissue lectins such as C-type lectins, galectins or siglecs. Since the distributions of binding sites can well overlap when comparing this property of activity between lectins, this assay is a means to assess the extent of overlap, in itself biomedically relevant information. The described fluorescence microscopical monitoring goes beyond measuring inhibition of lectin binding to a glycoprotein. It facilitates detection of cell type-dependent differences, thus extends in the tissue context what array analysis with printed vesicles or cultured cells can offer.

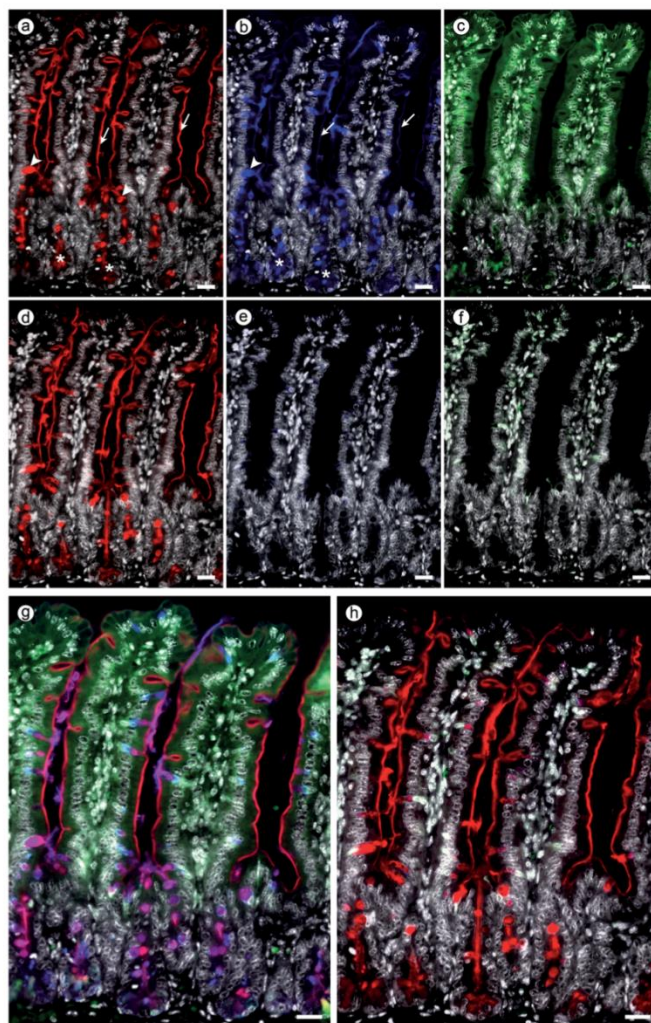


Fig. 8 Three-colour fluorescence staining profiles by DBA, SBA and MGL (CRD + stalk) in longitudinal sections of fixed murine jejunum in the absence or presence of cognate sugar (GalNAc) presented by compound **5** and their (photo)merging. Accessible binding sites for the three lectins were visualised by applying a mixture of biotinylated DBA, Alexa Fluor®-555-labelled SBA (colour assignment to blue) and FITC-labelled MGL (CRD + stalk) (green). Localization of biotinylated DBA became detectable by the second-step reagent Alexa Fluor®-647-conjugated streptavidin ($5 \mu\text{g mL}^{-1}$; colour assignment to red). DAPI was used for staining of nuclei (colour assignment to white). (a–c) Binding of DBA (a) and SBA (b), respectively, was seen in the brush border (arrows) of intestinal enterocytes, in goblet cells (arrowheads) and crypt-associated cells (asterisks). MGL binding was mainly confined to the cytoplasm of surface enterocytes in intestinal villi (c). (d–f) No inhibition of DBA-dependent staining (d) was seen with scaffold (compound **5**)-presented GalNAc (0.1 mM), whereas binding of labelled SBA (e) and of MGL (f) was completely inhibited. The three staining patterns were (photo)merged so that binding sites for each lectin in the same section were visualised in the absence (g) and in the presence of cognate sugar (h). The regionally similar DBA- and SBA-dependent signal presence led to overlap (colour change to magenta) in the brush border of intestinal villi, goblet cells and crypt-associated cells (g). No significant overlap, in contrast, was seen with labelled MGL (g). In the presence of compound **5** applied at 0.1 mM GalNAc, DBA-dependent staining was expectably rather insensitive to inhibitor presence, when compared to SBA and MGL (h). Concentration of biotinylated DBA was $1 \mu\text{g mL}^{-1}$, that of Alexa Fluor®-555-labelled SBA was $2 \mu\text{g mL}^{-1}$ and FITC-labelled MGL was $16 \mu\text{g mL}^{-1}$. DAPI was used at $0.5 \mu\text{g mL}^{-1}$. Scale bars are $20 \mu\text{m}$.

Conclusions

In conclusion, the emerging broad significance of protein(lectin)-glycan recognition is fueling the interest in developing glycoclusters as inhibitors of lectin binding. Since spatial and structural parameters are being disclosed to matter for achieving this type of functional pairing, the target of a therapeutic glycocluster, assaying of these synthetic products should cover these factors of cellular glycomes to the highest possible extent. Taking steps toward this aim is thus timely.

We here introduce a versatile platform for routine and rigorously controlled application, *i.e.* tissue sections, and illustrate the influence of parameters on the side of glycoclusters (especially core structure and valency) and lectins, especially architecture, fine-specificity and valency, on the experimental read-out, and this on two cell populations. The presented data strongly argue against the possibility for simple extrapolations. The pertinent question on glycocluster activity on lectins in mixtures, a common situation for tissue lectins, is answered by fluorescence microscopical monitoring using differently labelled proteins.

Whereas chemical engineering of nanoparticle surfaces and of cells programs tools for measuring inhibition of their bridging by lectins,^{78–80} our assay examines all aspects of lectin binding. It takes advantage of having cell and matrix glycomes available in tissue sections, and their processing can for example ensure to maintain glycolipid presence.^{81,82} In comparison to cultured cells, the tissue ensures analysis of cells within their microenvironment. This favourable aspect is also a part of the *in vivo* studies with labelled neoglycoproteins presenting natural and synthetic glycans up to custom-made microheterogeneity used for targeting and imaging.^{83–89} On the side of the proteins, mimicking the natural occurrence of lectins in mixtures, in terms of constituents and relative concentrations, will enable to trace currently not well-defined consequences of competition for binding sites and cross-reactivity of glycoclusters among lectins. Given the assay's reliability and sensitivity as well as its use with any type of tissue, its application is advised, prior to considering testing glycoclusters *in vivo*, for example to minimise such overlaps in target binding by iterative structural refinements including headgroup chemistry.

Experimental

General methods for synthesis

¹H-NMR spectra were recorded (25 °C) at 400 MHz or 500 MHz; ¹³C-NMR spectra were recorded at 100 MHz or 125 MHz. Chemical shift (δ) data in ppm are report with multiplicities indicated as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet); coupling constants (*J*) given in hertz (Hz). Chemical shifts (δ) are reported relative to internal standard Me₄Si in CDCl₃ (δ 0.0) for ¹H and CDCl₃ (δ 77.0) for ¹³C. ¹H NMR spectral signals were assigned with the aid of COSY, ¹³C NMR spectral signals using DEPT, gHSQCAD and/or gHMBCAD. NMR spectroscopic data for known compounds was in good agreement with those reported in the literature, as referenced. High resolution ESI mass spectra were measured in positive and/or negative mode using a Waters LCT Mass Spectrometer or were

acquired using a microTOF-Q spectrometer interfaced to a Dionex UltiMate 3000 LC in positive and negative modes as required. MicroTof control 3.2 and HyStar 3.2 software were used to carry out the analysis for data obtained with the microTOF-Q. TLC was performed on aluminium sheets pre-coated with silica gel and spots visualized by UV and/or charring with H₂SO₄-EtOH (1 : 20) or cerium molybdate, unless otherwise stated. Chromatography was carried out with silica gel 60 (0.040–0.630 mm) and using a stepwise solvent polarity gradient correlated with TLC mobility, unless otherwise stated. CH₂Cl₂, MeOH, toluene and THF reaction solvents were used as obtained from a Pure Solv™ Solvent Purification System. Unless it is specified, all reagents were used as received without further purification. All reactions involving air-sensitive reagents were performed under nitrogen in oven-dried glassware using syringe-septum cap technique. Optical rotations were determined at the sodium D line at 20 °C using a Schmidt and Haensch UniPol L1000. The IR spectra were recorded using thin film with a PerkinElmer Spectrum 100 FT-IR Spectrometer with an ATR attachment.

4-(1,2,2-Tris(4-(prop-2-yn-1-yloxy)phenyl)vinyl)phenyl acetate 11

To a solution of **10** (1.37 g, 3.45 mmol) at 0 °C in pyridine (27.5 mL), Ac₂O (228 μ L, 2.41 mmol) was added over 2 h. The reaction mixture was allowed to attain room temp and then stirred for a further 3 h. The solvent was removed under diminished pressure and chromatography (CH₂Cl₂-MeOH, 96.6 : 3.4) gave the intermediate monoacetate (827 mg, 55%); ¹H NMR (500 MHz, methanol-*d*₄) δ 6.99 (d, *J* = 8.8 Hz, 2H, *o*-Ar), 6.85–6.76 (m, 8H), 6.55–6.47 (m, 6H), 2.22 (s, 3H, OAc), ¹³C NMR (126 MHz, methanol-*d*₄) δ = 169.8(C, C=O), 155.6(C), 155.5 (C), 155.4 (C), 148.8 (C), 142.5 (C), 139.9 (C), 137.3 (C), 135.5 (C), 135.3 (C), 132.3 (CH), 132.2 (CH, 2 \times), 132.0 (CH), 120.3 (CH), 114.1 (CH), 110.0 (C), 19.5(CH₃, OAc); ESI-HRMS calcd for C₂₈H₂₁O₅ 437.1389, found *m/z* 437.1391 [M – H][–]. To this intermediate (1.7 g, 3.9 mmol) in dry DMF (78 mL) was added anhydrous potassium carbonate (6.42 g, 46.4 mmol); propargyl bromide (80% in toluene, 2.72 mL, 31.0 mmol) was then added to the mixture and it was stirred for 14 h at 70 °C under an inert atmosphere. The reaction was then cooled to room temp and diluted with CH₂Cl₂ (300 mL) and was then washed with satd NH₄Cl. The aqueous layer was again washed with further portions of CH₂Cl₂ (3 \times 150 mL). The combined organic layers were washed with water and dried over Na₂SO₄. The solvent was removed under diminished pressure and column chromatography (cyclohexane-EtOAc, 85 : 15 to 91 : 9) gave the title compound (520 mg, 24%); ¹H NMR (500 MHz, chloroform-*d*) δ 7.00 (d, *J* = 8.7 Hz, 2H, *m*-Ar), 6.96–6.91 (m, 6H, Ar), 6.83 (d, *J* = 8.7 Hz, 2H, *o*-Ar), 6.74–6.68 (m, 6H, Ar), 4.61 (t, *J* = 2.6 Hz, 6H, CH₂), 2.50 (t, *J* = 2.5 Hz, 3H, propargylic), 2.25 (s, 3H, OAc), ¹³C NMR (126 MHz, chloroform-*d*) δ = 169.2(C, C=O), 156.1 (C), 156.0 (C), 148.9 (C), 141.6 (C), 139.5 (C), 138.2 (C), 137.1 (C), 137.0 (C), 137.0 (C), 132.5 (CH), 132.5 (CH), 132.3 (CH), 120.7 (CH), 114.1 (CH), 114.1(CH), 114.0(CH), 75.5 (CH), 75.4 (CH), 75.4 (CH), 55.8 (CH₂, 2 \times), 55.8 (CH₂), 21.2 (CH₃, OAc); ESI-HRMS calcd for C₃₇H₂₈O₅Na 575.1834, found *m/z* 575.1832



[M + H]⁺. Note: the tetrapropargylated compound was the main by-product as a result of partial deacetylation during the propargylation reaction.

4-(1,2,2-Tris(4-((1-(2-(2-deoxy-2-acetamido-D-galactopyranosylthio)ethyl)-1H-1,2,3-triazol-4-yl)methoxy)phenyl)vinyl)phenol 3

To **11** (124 mg, 0.22 mmol) in degassed THF-H₂O (1 : 1, 17.2 mL) were added azide **12** (300 mg, 0.69 mmol), sodium ascorbate (39 mg, 0.198 mmol) and Cu₂SO₄ · 5H₂O (4942 mg, 0.198 mmol). The reaction mixture was stirred under an inert atmosphere in a microwave reactor (120 W) at 50 °C for 30 min. Tetrahydrofuran was then removed under diminished pressure followed by the dilution with CH₂Cl₂. This mixture was then washed with water and the aqueous layer was re-extracted with a further portion of CH₂Cl₂. The combined organic layers were dried over Na₂SO₄ and the solvent was removed under diminished pressure. Chromatography of the residue (CH₂Cl₂-MeOH, 96.5 : 3.5 to 95.5 : 4.5) gave the fully acetylated intermediate (411 mg, 70%) as a colourless solid; ¹H NMR (500 MHz, chloroform-*d*) δ 7.67 (s, 1H, triaz.), 7.67 (s, 1H, triaz.), 7.66 (s, 1H, triaz.), 7.00 (d, *J* = 8.7 Hz, 2H, *m*-Ar), 6.93–6.91 (m, 6H, Ar), 6.83 (d, *J* = 8.7 Hz, 2H, *o*-Ar), 6.73–6.68 (m, 6H, Ar), 5.97 (d, *J* = 7.8 Hz, 1H, NH), 5.92 (d, *J* = 8.1 Hz, 1H, NH), 5.71 (d, *J* = 8.3 Hz, 1H, NH), 5.68–5.63 (m, 3H, H-1), 5.39 (dd, *J* = 3.2, 1.2 Hz, 3H, H-4), 5.14 (s, 2H, CH₂), 5.14 (s, 2H, CH₂), 5.13 (s, 2H, CH₂), 5.01 (dt, *J* = 11.8, 3.3 Hz, 3H, H-3), 4.78–4.69 (m, 3H, H-2), 4.65–4.55 (m, 6H, CH₂), 4.52 (t, *J* = 6.8 Hz, 3H, H-5), 4.20–4.06 (m, 6H, CH₂), 3.25–3.12 (m, 3H, CH₂), 3.12–3.00 (m, 3H, CH₂), 2.25 (s, 3H, ArOAc), 2.16 (s, 3H, NHCH₃), (s, 6H NHCH₃), 2.01 (s, 3H, OAc), 2.00 (s, 6H, OAc), 1.99 (s, 6H, OAc), 1.99 (s, 3H, OAc), 1.97 (s, 3H, OAc), 1.97 (s, 6H, OAc). To a suspension of this intermediate (145 mg, 0.078 mmol), which was precooled to 0 °C, in dry methanol (7 mL), was added freshly prepared NaOMe solution in MeOH until the solution pH was 9. The resulting mixture was stirred for 2 h at room temp and then Amberlite IR-120 H⁺, which had been washed with methanol, was added and the mixture stirred gently for 10 min until the solution was neutralized. The solution containing the product was separated from the solid materials by decanting and the resin was washed with a warm mixture of water-acetonitrile which was added to the methanol solution. The volatile solvents were removed under diminished pressure until a minimal amount of water remained. Reverse phase column chromatography gave compound **3** (84 mg, 75%) as a colourless solid. ¹H NMR (500 MHz, DMSO-*d*₆) δ 9.33 (s, 1H, ArOH) 8.19 (s, 3H, triazo.), 7.80 (d, *J* = 6.9 Hz, 3H, NH), 6.85 (m, 6H, Ar), 6.79 (m, 6H, Ar), 6.73 (d, *J* = 8.4 Hz, 2H, *m*-Ar), 6.51 (d, *J* = 8.4 Hz, 2H, *o*-Ar), 5.51 (d, *J*_{1,2} = 5.2 Hz, 3H, H-1), 5.02 (s, 6H, CH₂), 4.80–4.65 (m, 12H, OH), 4.55 (ddt, *J* = 27.1, 13.9, 6.8 Hz, 6H, CH₂), 4.16 (ddd, *J*_{2,3,2-NH,2-1} = 11.8, 6.9, 5.2 Hz, 3H, H-2), 3.91 (t, *J*_{5,6,5-6'} = 6.0 Hz, 3H, H-5), 3.72 (brs, 3H, H-4), 3.53 (t, *J*_{6,5,6'-6} = 5.9 Hz, 6H, H-6, H-6'), 3.49 (dd, *J*_{3,2,3-4} = 11.8, 2.9 Hz, 3H, H-3), 3.02 (dt, *J* = 13.8, 6.9 Hz, 3H, CH₂), 2.92 (dt, *J* = 13.7, 6.8 Hz, 3H), 1.79 (s, 9H, NHAc). ¹³C NMR (126 MHz, dmsol) δ = 170.3(C, NHAc), 156.8 (C, Ar), 156.1 (C, *m*-Ar), 142.9 (C, triaz.), 142.8 (C), 139.0 (C), 137.1 (C), 132.5 (CH, *m*-Ar), 132.5 (CH, Ar), 125.2 (CH, triaz.), 114.4 (CH, *o*-Ar), 114.3 (CH, Ar), 84.8 (CH, H-1),

72.7 (CH, H-5), 68.5 (CH, H-4), 67.9 (CH, H-3), 61.3 (CH₂), 61.2 (CH₂, C-6), 50.6 (CH, C-2), 49.8 (CH₂), 30.3 (CH₂), 23.1 (CH₃, NHAc); ESI-HRMS calcd for C₆₅H₇₉ N₁₂O₁₉S₃ 1427.4747, found *m/z* 1427.4752 [M – H][–]. Note: for the reverse phase chromatography, a minimum volume of the aqueous residue was loaded onto the column and elution involved passing three column volumes of 1 : 4 CH₃CN–H₂O followed by 1 : 1 CH₃CN–H₂O.

1,2,5,6-Tetra-O-(propargyl)-D-mannitol 14

Sodium hydride (1.0 g, 26.3 mmol, 60% in mineral oil) was added to a cooled solution of **13**²² (0.49 g, 2.2 mmol) in dry DMF (30 mL). The suspension was stirred for 15 min and then 80% propargyl bromide in toluene (2.9 mL, 26.3 mmol) was added. The resulting mixture was stirred at room temp for 25 h, and after this time MeOH was added at 0 °C. The mixture was diluted with EtOAc, then washed with brine (3 × 20 mL), dried over Na₂SO₄, filtered and the solvent was removed. Column chromatography (9 : 1 cyclohexane–EtOAc) gave the propargylated intermediate as a yellow oil (0.77 g, 94%); [α]_D²⁰ = +6 (c 0.8, CHCl₃); ¹H NMR (500 MHz, CDCl₃) δ 4.38 (dd, 2H, *J*_{gem} = 16.0 Hz, *J*_{H,H} = 2.4 Hz, CH_{2a}C≡CH-2), 4.35 (dd, 2H, *J*_{gem} = 16.0 Hz, *J*_{H,H} = 2.4 Hz, CH_{2b}C≡CH-2), 4.20 (dd, 2H, *J*_{gem} = 15.9 Hz, *J*_{H,H} = 2.4 Hz, CH_{2a}C≡CH-1), 4.15 (dd, 2H, *J*_{gem} = 15.9 Hz, *J*_{H,H} = 2.4 Hz, CH_{2b}C≡CH-1), 4.14 (dd, 2H, *J*_{3,2} = 3.6 Hz, *J*_{3,1} = 1.5 Hz, H-3), 3.88 (dddd, 2H, *J*_{2,3} = 3.6 Hz, *J*_{2,1b} = 6.3 Hz, *J*_{2,1a} = 3.2 Hz, *J*_{H,H} = 1.5 Hz, H-2), 3.84 (dd, 2H, *J*_{1a,1b} = 10.6 Hz, *J*_{1a,2} = 3.2 Hz, H-1a), 3.64 (dd, 2H, *J*_{1b,1a} = 10.6 Hz, *J*_{1b,2} = 6.3 Hz, H-1b), 2.44 (t, 2H, *J*_{H,H} = 2.4 Hz, CH₂C≡CH-2), 2.42 (t, 2H, *J*_{H,H} = 2.4 Hz, CH₂C≡CH-1), 1.39 (s, 6H, CH₃); ¹³C NMR (126 MHz, CDCl₃) δ 110.1 (C(CH₃)₂), 80.1 (CH₂C≡CH-2), 79.7 (CH₂C≡CH-1), 78.1 (C-2), 78.0 (C-3), 74.8 (CH₂C≡CH-1), 74.7 (CH₂C≡CH-2), 70.0 (C-1), 58.7 (CH₂C≡CH-1), 58.0 (CH₂C≡CH-2), 27.2 (CH₃); ESI-HRMS calcd for C₂₁H₂₆NaO₆ 397.1627, found: 397.1663 [M + Na]⁺; hydrochloric acid (0.3 mL of 1.0 M) was added to this intermediate (0.12 g, 0.32 mmol) in MeOH (2.7 mL). The resulting mixture was heated at reflux for 2.5 h and then allowed to attain room temperature. Excess solid NaHCO₃ was added and the mixture stirred for a further 15 min. The solid was filtered off and the solvent was evaporated and column chromatography (cyclohexane–EtOAc, 7 : 3) gave **14** as a pale yellow oil (0.1 g, 94%); ¹H NMR (500 MHz, CDCl₃) δ 4.37 (dd, 2H, *J*_{gem} = 16.0 Hz, *J*_{H,H} = 2.4 Hz, CH_{2a}C≡CH-2), 4.32 (dd, 2H, *J*_{gem} = 16.0 Hz, *J*_{H,H} = 2.4 Hz, CH_{2b}C≡CH-2), 4.20 (d, 4H, *J*_{H,H} = 2.4 Hz, CH₂C≡CH-1), 3.91 (m, 2H, H-3), 3.88 (m, 2H, H-1a), 3.85 (m, 2H, H-2), 3.75 (dd, *J*_{1b,1a} = 10.1 Hz, *J*_{1b,2} = 4.8 Hz, 2H, H-1b), 2.84 (d, 2H, *J*_{OH,3} = 6.2 Hz, OH), 2.46 (t, 2H, *J*_{H,H} = 2.4 Hz, CH₂C≡CH-2), 2.45 (t, 2H, *J*_{H,H} = 2.4 Hz, CH₂C≡CH-1); ¹³C NMR (126 MHz, CDCl₃) δ 80.0 (CH₂C≡CH-2), 79.5 (CH₂C≡CH-1), 78.2 (C-2), 75.0 (CH₂C≡CH-1), 74.8 (CH₂C≡CH-2), 69.8 (C-1), 69.4 (C-3), 58.8 (CH₂C≡CH-1), 58.1 (CH₂C≡CH-2). ES-HRMS calcd for C₁₈H₂₂NaO₆: 357.1314, found: 357.1295 [M + Na]⁺.

(2R,3S,4S,5R)-1,2,5,6-Tetrakis((1-(2-(2-deoxy-2-acetamido-α-D-galactopyranosylthio)ethyl)-1H-1,2,3-triazol-4-yl)methoxy)-hexane-3,4-diol 4 to 14

(0.16 g, 0.37 mmol) and **12** (43 mg, 0.13 mmol) in THF–H₂O (10 mL, 2 : 1), sodium ascorbate (26 mg, 0.13 mmol) and



CuSO₄·5H₂O (33 mg, 0.13 mmol) were added. The resulting mixture was stirred for 1.5 h at room temp. After this time, the solvent was removed and the residue dissolved in CH₂Cl₂, then washed with brine (2 × 10 mL), dried over Na₂SO₄, filtered and the solvent was removed. Column chromatography of the residue (CH₂Cl₂-MeOH, 15 : 1) gave the protected intermediate as a white solid (0.11 g, 57%); [α]_D²⁰: +81 (c 1.1, CHCl₃); ¹H NMR (500 MHz, CDCl₃) δ 7.79 (s, 2H, H-triazole 1), 7.78 (s, 2H, H-triazole 2), 6.82 (d, 2H, *J*_{NH,2 Gal} = 7.6 Hz, NH), 6.76 (d, 2H, *J*_{NH,2 Gal} = 7.6 Hz, NH), 5.75 (d, 2H, *J*_{1,2} = 4.6 Hz, H-1 Gal), 5.74 (d, 2H, *J*_{1,2} = 4.6 Hz, H-1 Gal), 5.41 (s, 4H, H-4 Gal), 5.0 (dd, 2H, *J*_{3,2} = 11.8 Hz, *J*_{3,4} = 3.2 Hz, H-3 Gal), 4.99 (dd, 2H, *J*_{3,2} = 11.8 Hz, *J*_{3,4} = 3.2 Hz, H-3 Gal), 4.80 (d, 2H, *J*_{gem} = 12.8 Hz, CH₂-triazole 2), 4.72 (m, 4H, H-2 Gal), 4.71 (d, 2H, *J*_{gem} = 12.8 Hz, CH₂-triazole 2), 4.71 (d, 2H, *J*_{gem} = 12.8 Hz, CH₂-triazole 1), 4.62 (d, 2H, *J*_{gem} = 12.8 Hz, CH₂-triazole 1), 4.61 (m, 8H, SCH₂CH₂), 4.55 (dd, 4H, *J*_{5,6a} = 7.0 Hz, *J*_{5,6b} = 5.4 Hz, H-5 Gal), 4.15 (dd, 4H, *J*_{6a,6b} = 11.3 Hz, *J*_{6a,5} = 5.4 Hz, H-6a Gal), 4.09 (dd, 4H, *J*_{6b,6a} = 11.3 Hz, *J*_{6b,5} = 7.0 Hz, H-6b Gal), 3.87 (dd, 2H, *J*_{3,2} = *J*_{3,OH} = 6.4 Hz, H-3), 3.80 (m, 2H, H-1a), 3.71 (m, 2H, H-2), 3.69 (m, 2H, H-1b), 3.65 (d, 2H, *J*_{OH,3} = 6.4 Hz, OH), 3.18 (m, 4H, SCH₂CH₂), 3.08 (m, 4H, SCH₂CH₂), 2.16 (s, 12H), 2.0 (s, 6H), 1.99 (s, 6H), 1.99 (s, 12H), 1.98 (s, 6H) (each OCOCH₃), 1.98 (s, 6H, NHCOCCH₃); ¹³C NMR (126 MHz, CDCl₃) δ 171.1, 171.1 (each NHCOCCH₃), 170.9, 170.9, 170.6, 170.4, 170.4 (each OCOCH₃), 145.3 (C-4 triazole-2), 145.0 (C-4 triazole-1), 123.7 (×2) (C-5 triazole-1, C-5 triazole-2), 85.2, 84.9 (each C-1 Gal), 78.6 (C-2), 70.0 (C-1), 68.9 (C-3), 68.3 (×2) (C-3 Gal), 67.5 (×2) (C-5 Gal), 67.2 (×2) (C-4 Gal), 64.7 (CH₂-triazole 1), 63.7 (CH₂-triazole 2), 62.2, 62.2 (each C-6 Gal), 49.8, 49.6 (each SCH₂CH₂), 48.4 (×2) (C-2 Gal), 31.3, 31.2 (each SCH₂CH₂), 23.2 (×2) (NHCOCCH₃), 20.9, 20.9, 20.8 (×2), 20.8 (×2) (each OCOCH₃); ESI-HRMS calcd for C₈₂H₁₁₈N₁₆NaO₃₈S₄: 2085.6574, found: 2085.6567 [M + Na]⁺. To this protected intermediate (42 mg, 0.02 mmol) in dry MeOH (2 mL), freshly prepared NaOMe solution in MeOH (0.35 mL of 1 M) was added and the mixture was stirred at room temp for 4 h. Then Amberlite IR-120 H⁺ was added and stirred gently until the pH was 7. The resin was filtered off and the solvent was removed. Reverse phase column chromatography (MeCN-H₂O, 0 : 1 to 1 : 1) gave the title compound 4 as a white solid (24 mg, 76%); [α]_D²⁰: +106 (c 0.8, DMSO); ¹H NMR (500 MHz, DMSO-*d*₆) δ 8.10 (d, 2H, H-triazole 1), 8.0 (d, 2H, H-triazole 2), 7.77 (d, 4H, *J*_{NH,2 Gal} = 6.5 Hz, NH), 5.51 (d, 4H, *J*_{1,2} = 5.1 Hz, H-1 Gal), 4.70 (m, 6H, OH, CH₂-triazole 2), 4.68 (m, 4H, OH), 4.61 (m, 2H, OH), 4.58 (m, 6H, CH₂-triazole 2, OH), 4.56 (m, 4H, CH₂-triazole 1), 4.53 (m, 8H, (SCH₂CH₂)), 4.19 (ddd, 4H, *J*_{2,3} = 11.8 Hz, *J*_{2,NH} = 6.5 Hz, *J*_{2,1} = 5.1 Hz, H-2 Gal), 3.92 (m, 4H, H-5 Gal), 3.85 (m, 2H, H-1a), 3.73 (m, 4H, H-4 Gal), 3.60 (m, 4H, H-2, H-3), 3.58 (m, 2H, H-1b), 3.55 (m, 8H, H-6 Gal), 3.50 (m, 4H, H-3 Gal), 3.02 (m, 4H, SCH₂CH₂), 2.93 (m, 4H, SCH₂CH₂), 1.81 (s, 6H), 1.80 (s, 6H) (each NHCOCCH₃); ¹³C NMR (126 MHz, DMSO-*d*₆) δ 169.9, 169.9 (NHCOCCH₃), 144.4 (C-4 triazole-2), 144.0 (C-4 triazole-1), 124.0 (C-5 triazole-1), 123.8 (C-5 triazole-2), 84.4 (×2) (C-1 Gal), 78.1 (C-2), 72.2, 72.2 (each C-5 Gal), 71.1 (C-1), 68.1 (×3) (C-4 Gal, C-3), 67.5 (×2) (C-3 Gal), 64.0 (CH₂-triazole 1), 63.4 (CH₂-triazole 2), 60.9, 60.9 (each C-6 Gal), 50.0 (×2) (C-2 Gal), 49.3 (×2) (SCH₂CH₂), 29.9, 29.8 (each SCH₂CH₂), 22.7 (×2) (NHCOCCH₃). ESI-HRMS calcd for C₅₈H₉₄N₁₆NaO₂₆S₄: 1581.5306, found: 1581.5365 [M + Na]⁺.

BF₂ chelated 3-methyl-N-(3-methyl-5-(4-(prop-2-yn-1-yloxy)phenyl)-1H-pyrrol-2-yl)-5-(4-(prop-2-yn-1-yloxy)phenyl)-2H-pyrrol-2-imine 16

To 15³⁴ (500 mg, 1.35 mmol) in dry CH₂Cl₂ (20 mL) at -78 °C, was added dropwise BBr₃ (1 M in CH₂Cl₂, 5.4 mmol) and the mixture stirred at -78 °C for 3 h. The reaction mixture was quenched by the slow addition of saturated sodium bicarbonate (30 mL), the organic layer was then separated and the aqueous layer was extracted with CH₂Cl₂ (2 × 40 mL). The combined organic layers were washed with saturated sodium bicarbonate, deionized H₂O, brine, dried over Na₂SO₄, filtered, and the solvent was removed under reduced pressure. Silica gel chromatography (cyclohexane-EtOAc, 3 : 1) gave the demethylated intermediate as a dark blue solid (169 mg, 35%, mp 121–122 °C); ¹H NMR (400 MHz, CD₃OD-*d*₄) δ 7.72 (dd, *J* = 8.8, 1.9 Hz, 4H), 6.89 (dd, *J* = 8.8, 1.9 Hz, 4H), 6.71 (s, 2H), 4.57 (br, 2H), 2.26 (s, 6H) ppm; ¹³C NMR (100 MHz, CD₃OD-*d*₄) δ 159.4, 154.4, 150.6, 140.3, 127.7, 123.8, 115.7, 115.7, 9.9 ppm; ESI-HRMS [M + H]⁺: 358.1541, C₂₂H₂₀N₃O₂ requires 358.1556. To this intermediate (160 mg, 0.45 mmol) in dry THF (25 mL) cesium carbonate (439 mg, 1.35 mmol) was added, and the mixture was stirred at 0 °C under a nitrogen atmosphere for 10 min. Propargyl *p*-toluenesulfonate (473 mg, 2.25 mmol) was then added *via* syringe and the solution was heated at reflux, whilst stirring, for 2 h. The reaction mixture was cooled to room temp and the mixture was partitioned between CH₂Cl₂ (100 mL) and PBS buffer (100 mL, pH 7). The organic phase was washed with water (3 × 100 mL), brine (50 mL), dried over Na₂SO₄, filtered and evaporated to dryness. The crude product was purified by silica gel chromatography, eluting with CH₂Cl₂ to yield the propargylated intermediate compound (107 mg, 55%, mp 133–135 °C) as a dark green solid; ¹H NMR (400 MHz, CDCl₃) δ 7.80 (d, *J* = 7.5 Hz, 1H), 7.08 (d, *J* = 7.4 Hz, 1H), 6.66 (s, 1H), 4.78 (s, 1H), 2.57 (s, 1H), 2.34 (s, 2H) ppm. ¹³C NMR (100 MHz, CDCl₃) δ 158.7, 154.0, 150.0, 141.0, 127.8, 126.2, 116.3, 115.4, 78.2, 76.0, 55.9, 11.4 ppm; ESI-HRMS calcd for C₂₈H₂₄N₃O₂ 434.1869, found *m/z* 434.1869 [M + H]⁺. To this intermediate (100 mg, 0.23 mmol) in dry CH₂Cl₂ (20 mL) DIPEA (0.4 mL, 2.3 mmol) was added and BF₃·OEt₂ (0.4 mL, 3.22 mmol), and the mixture was stirred at room temp under N₂ for 20 h. The reaction mixture was washed with water (3 × 50 mL), brine, dried over Na₂SO₄, filtered and the solvent was removed under reduced pressure. Chromatography (cyclohexane-EtOAc, 4 : 1) gave the title compound 16 as red metallic solid (67 mg, 61%); ¹H NMR (400 MHz, DMSO-*d*₆) δ 8.05 (d, *J* = 9.0 Hz, 4H), 7.14 (d, *J* = 9.0 Hz, 4H), 7.01 (s, 2H), 4.93 (d, *J* = 2.4 Hz, 4H), 3.65 (t, *J* = 2.4 Hz, 2H), 2.33 (s, 6H) ppm; ¹³C NMR (100 MHz, DMSO-*d*₆) δ 160.0, 157.8, 146.0, 143.3, 131.8, 124.3, 122.6, 115.6, 79.3, 79.2, 56.2, 11.6 ppm; ESI-HRMS calcd for C₂₈H₂₃BF₂N₃O₂ 482.1851, found *m/z* 482.1849 [M + H]⁺.

3-Methyl-N-(3-methyl-5-(4-((1-(2-(2-acetamido-2-deoxy-α-D-galactopyranosylthio)ethyl)-1H-1,2,3-triazol-4-yl)methoxy)phenyl)-2H-pyrrol-2-ylidene)-5-(4-((1-(2-(2-acetamido-2-deoxy-α-D-galactopyranosylthio)ethyl)-1H-1,2,3-triazol-4-yl)methoxy)phenyl)-1H-pyrrol-2-amine, BF₂ chelate 6

Compound 16 (5.1 mg, 0.0168 mmol) was dissolved in H₂O (1 mL) and de-O-acetylated 12 (3.7 mg, 7.7 μmol), prepared as



described for **7** (below) from **12**, was dissolved in THF (1 mL) and both solutions were added to a microwave vial and to this sodium ascorbate (7.65 mg, 0.0386 mmol) and copper sulphate pentahydrate (4.19 mg, 0.0168 mmol) were added. The mixture was heated in a microwave reactor (120 W) at 60 °C, for 1 h. The THF was then removed under vacuum and the remaining water was diluted with CH₃CN and reverse phase chromatography (H₂O-CH₃CN, 1 : 1) gave the title compound **6** as a green solid (5.4 mg, 64%); ¹H NMR (500 MHz, DMSO-*d*₆) δ 8.21 (s, 2H, triazole CH), 8.03–7.98 (m, 4H, Ar-H), 7.70 (d, *J* = 7.0 Hz, 2H, NH), 7.18–7.11 (m, 4H), 6.97 (s, 2H, Ar-H), 5.46 (d, *J* = 5.3 Hz, 2H, H-1), 5.21 (s, 4H, CH₂OAr), 4.68–4.61 (m, 4H, OH-4, OH-6), 4.60–4.45 (m, 6H, SCH₂CH₂, OH-3), 4.13 (dt, *J* = 11.9, 6.3 Hz, 2H, H-2), 3.87 (t, *J* = 6.0 Hz, 2H, H-5), 3.68 (t, *J* = 4.0 Hz, 2H, H-4), 3.50 (t, *J* = 5.8 Hz, 4H, H-6), 3.44 (ddd, *J* = 10.8, 6.7, 3.0 Hz, 2H, H-3), 3.00 (dt, *J* = 14.0, 6.9 Hz, 2H SCH₂CH₂), 2.90 (dt, *J* = 13.9, 6.9 Hz, 2H, SCH₂CH₂), 2.28 (s, 6H, CH₃), 1.76 (d, *J* = 1.4 Hz, 6H, NHAc); ¹³C NMR (126 MHz, dmsO) δ 170.21 (C=O), 160.95, 157.70, 155.31, 145.95, 143.04, 142.42 (C), 131.91 (CH), 125.45 (triazole CH), 123.91, 115.48 (each Ar-CH), 84.77 (C-1), 72.67 (C-5), 68.53 (C-4), 67.91 (C-3), 61.85 (CH₂OAr), 61.34 (C-6), 50.43 (C-2), 49.79 (SCH₂CH₂), 30.27 (SCH₂CH₂), 23.06 (NHAc), 11.49 (CH₃); ESI-HRMS: calcd for C₄₈H₅₈BN₁₁O₁₂ F₂S₂ Na 1116.3667; found 1116.3676 [M + Na]⁺

Methyl 2-acetamido-2-deoxy-1-thio-α-D-galactopyranoside **7**

The thiol **16**³⁵ (77 mg, 0.199 mmol) was dissolved in dichloromethane (3 mL) and DIPEA (0.4 mmol, 70 μL) and methyl iodide (0.9 mmol, 56 μL) was added and the mixture was stirred for 2.5 h at room temp. The solvent was removed and flash chromatography (toluene-acetone (6 : 4)) gave the protected intermediate (62 mg, 82%); ¹H NMR (500 MHz, chloroform-*d*) δ 5.64 (d, *J* = 8.8 Hz, 1H, NH), 5.44 (d, *J* = 5.3 Hz, 1H, H-1), 5.38 (dd, *J* = 3.3, 1.3 Hz, 1H, H-4), 5.09 (dd, *J* = 11.7, 3.3 Hz, 1H, H-3), 4.80–4.74 (m, 1H, H-2), 4.55–4.50 (m, 1H, H-5), 4.17–4.06 (m, 2H, H-6), 2.16 (s, 3H, OAc), 2.12 (s, 3H, CH₃), 2.04 (s, 3H, OAc), 2.00 (s, 3H, OAc), 1.97 (s, 3H, NHAc); ¹³C NMR (126 MHz, CDCl₃) δ 171.06, 170.37, 170.26, 170.15 (each C=O), 85.53 (C-1), 68.44 (C-3), 67.35 (C-4), 67.11 (C-5), 61.92 (C-6), 48.38 (C-2), 23.30 (NHAc), 20.74, 20.69, 20.68 (each OAc), 13.15 (CH₃). To this intermediate (55 mg, 0.14 mmol) dissolved in MeOH (2 mL) was added 1 M NaOMe in MeOH and the mixture was left to stir until the pH was ~10 and stirring was continued for 20 min. Amberlite H⁺ resin was then added until the pH was ~6 and was stirred for another 20 min. The mixture was then filtered and solvent removed under reduced pressure to give the title compound as a white solid (25 mg, 70%); ¹H NMR (500 MHz, methanol-*d*₄) δ 5.44 (d, *J* = 5.4 Hz, 1H, H-1), 4.41 (dd, *J* = 11.2, 5.4 Hz, 1H, H-2), 4.14 (ddd, *J* = 6.7, 5.2, 1.3 Hz, 1H, H-5), 3.88 (dd, *J* = 3.3, 1.4 Hz, 1H, H-4), 3.76–3.70 (m, 3H, H-3, H-6a, H-6b), 2.06 (s, 3H, CH₃), 1.96 (s, 3H, NHAc); ¹³C NMR (126 MHz, methanol-*d*₄) δ 172.51 (C=O), 84.82 (C-1), 71.29 (C-5), 68.91 (C-4), 68.08 (C-3), 61.38 (C-6), 50.55 (C-2), 21.14 (NHAc), 11.55 (CH₃); ESI-HRMS calcd for C₉H₁₇NO₅SNa 274.0725, found *m/z* 274.0729 [M + Na]⁺.

2-Azidoethyl 2-acetamido-2-deoxy-1-thio-α-D-glucopyranoside **18**

Thioglycoside **17** (40 mg, 0.092 mmol) was deacetylated as described in the preparation of **7** to give the title compound (24 mg, 85%); ¹H NMR (500 MHz, methanol-*d*₄) δ 5.56 (d, *J* = 5.3 Hz, 1H, H-1), 4.03 (dd, *J* = 11.1, 5.3 Hz, 1H, H-2), 3.96 (ddd, *J* = 10.0, 5.8, 2.4 Hz, 1H, H-5), 3.83 (dd, *J* = 12.0, 2.3 Hz, 1H, H-6a), 3.71 (dd, *J* = 12.0, 5.8 Hz, 1H, H-6b), 3.58 (dd, *J* = 11.1, 8.7 Hz, 1H, H-3), 3.55–3.49 (m, 1H, SCH₂CH₂N₃), 3.45 (dt, *J* = 12.8, 6.7 Hz, 1H, SCH₂CH₂N₃), 3.37–3.31 (m, 1H, H-4), 2.89–2.81 (m, 1H, SCH₂CH₂N₃), 2.73 (dt, *J* = 13.7, 6.6 Hz, 1H, SCH₂CH₂N₃), 1.97 (s, 3H, NHAc); ¹³C NMR (126 MHz, methanol-*d*₄) δ 172.31 (C=O), 83.98 (C-1), 73.05 (C-5), 71.19 (C-3), 71.08 (C-4), 61.20 (C-6), 54.41 (C-2), 50.78 (SCH₂CH₂N₃), 29.77 (SCH₂CH₂N₃), 21.12 (NHAc); ESI-HRMS: calcd. For C₁₀H₁₉N₃O₅S 307.1076; found 307.1070 [M + H]⁺.

3-Phenyl-N-(3-phenyl-5-(4-((1-(2-(2-acetamido-2-deoxy-α-D-glucopyranosylthio)ethyl)-1H-1,2,3-triazol-4-yl)methoxy)phenyl)-2H-pyrrol-2-ylidene)-5-(4-((1-(2-(2-acetamido-2-deoxy-α-D-glucopyranosylthio)ethyl)-1H-1,2,3-triazol-4-yl)methoxy)phenyl)-1H-pyrrol-2-amine, BF₂ chelate **8**

Glucopyranoside **18** (5.56 mg, 0.0181 mmol) in H₂O (1 mL) and **19**³³ (5 mg, 0.082 mmol) in THF (1 mL) were mixed and added to a microwave vial and to this mixture sodium ascorbate (8.2 mg, 0.041 mmol) and copper sulphate pentahydrate (8.2 mg, 0.041 mmol) were added. The reaction was performed in the microwave (120 W) at 60 °C for 1 h. The THF was then removed under vacuum to concentrate the mixture with some water remaining. This was diluted with CH₃CN and reverse phase chromatography (H₂O-CH₃CN, 1 : 1) gave **8** as a green solid (8.6 mg, 86%); ¹H NMR (500 MHz, DMSO-*d*₆) δ 8.29 (s, 2H, triazole CH), 8.19 (dd, *J* = 11.3, 8.1 Hz, 7H, Ar-H), 7.87 (d, *J* = 7.0 Hz, 2H, NH), 7.64 (s, 2H, Ar-H), 7.55 (t, *J* = 7.6 Hz, 4H, Ar-H), 7.49 (t, *J* = 7.3 Hz, 2H, Ar-H), 7.26 (dd, *J* = 7.9, 5.5 Hz, 4H, Ar-H), 7.19–7.14 (m, 1H, Ar-H), 5.47 (d, *J* = 5.3 Hz, 2H, H-1), 5.30 (s, 4H, CH₂OAr), 5.15 (d, *J* = 5.6 Hz, 2H, 4-OH), 4.85 (d, *J* = 5.7 Hz, 2H, 3-OH), 4.68 (t, *J* = 5.7 Hz, 2H, 6-OH), 4.58 (ddt, *J* = 27.7, 13.9, 7.0 Hz, 4H, SCH₂CH₂ triazole), 3.81 (ddd, *J* = 11.8, 6.9, 5.3 Hz, 2H, H-2), 3.78–3.68 (m, 4H, H-5, H-6a), 3.51 (dt, *J* = 11.9, 6.2 Hz, 2H, H-6b), 3.41–3.35 (m, 2H, H-3), 3.18–3.04 (m, 4H, H-4, SCH₂CH₂), 3.00 (dt, *J* = 13.7, 6.8 Hz, 2H, SCH₂CH₂), 1.83 (s, 6H, NHAc); ¹³C NMR (126 MHz, dmsO) δ 170.07 (C=O), 170.07 (C=O), 161.27, 157.84, 148.14, 145.00, 142.46, 142.60 (each Ar-C), 132.27, 130.05, 129.58, 129.35, 129.17, 128.65 (each Ar-H), 125.45, 123.93, 115.60 (each Ar-H), 84.49 (C-1), 74.21 (C-5), 71.42 (C-4), 71.08 (C-3), 61.96 (CH₂OAr), 61.34 (C-6), 54.63 (C-2), 49.82 (SCH₂CH₂), 30.56 (SCH₂CH₂), 23.05 (NHAc); ESI-HRMS: calcd for C₅₈H₆₂BN₁₁O₁₂F₂S₂Na 1240.3980; found 1240.3395 [M + Na]⁺.

Lectins

The two recombinant MGL proteins (CRD and CRD extended by the stalk) were produced, purified and subjected to quality controls as described in detail previously.³¹ SBA was purchased from Sigma-Aldrich (Munich, Germany). Biotinylated DBA was obtained from Enzo Life Sciences (Lörrach, Germany), biotinylated HPA from Sigma-Aldrich. Biotinylation of the MGL proteins and SBA was



performed under activity-preserving conditions with the *N*-hydroxysuccinimide ester derivative of biotin (Sigma-Aldrich) using the protocol used for human galectin-1.^{83,90} Fluorescent labeling was performed using the Alexa Fluor® 555 or the Alexa Fluor® 488 Protein Labeling Kits (Molecular Probes, Thermo Fisher Scientific, Darmstadt, Germany) following the recommendations of the supplier. Maintained lectin activity was ascertained by carbohydrate-inhibitable cell binding in cytofluorometric assays using CHO wild-type and Lec8 mutant cells^{81,84,91} and in the histochemical protocol (please see below).

Tissue samples

Specimens of jejunum obtained from four six-week-old C57BL/6 mice were thoroughly washed, then cut into small pieces and fixed in Bouin's solution (71% (v/v) saturated picric acid, 24% (v/v) formaldehyde solution (37% w/v) and 5% (v/v) glacial acetic acid) for 24 h. Dehydration of specimens was carried out in a series of aqueous ethanol solutions of increasing alcohol concentrations (70%, 80% and 99%) and then in isopropanol, which finally was replaced by xylene. Embedding of the pieces of tissue in paraffin wax at 61 °C followed, thereafter each specimen was cut into serial sections of about five µm thickness that were mounted on Superfrost® plus glass slides (Menzel, Braunschweig, Germany). Tissue specimen of murine jejunum were a kind gift of Dr E. Wolf of the Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Germany; all further processing was done by the authors as described.

Lectin histochemistry

The protocols for processing tissue sections by lectin histochemistry and monitoring staining profiles by light and fluorescence microscopy had been optimized previously.^{69,85,92,93} Routinely, systematic titrations with each labelled lectin were carried out to determine the concentrations, which generate strong signals combined with minimal background staining. The tested concentration ranges for light/fluorescence microscopy were 0.25–1.5/0.5–1 µg mL⁻¹ for DBA, 0.5–4/0.25–4 µg mL⁻¹ for SBA, 0.5–2/1–2 µg mL⁻¹ for HPA, 1–8 µg mL⁻¹ for MGL (CRD) and 1–8/8–32 µg mL⁻¹ for MGL (CRD + stalk), respectively. Assessment of inhibitory capacity of glycoconjugates was determined after a pre-incubation of the mixture containing a glycoconjugate and lectin for one hour at room temperature. Aliquots of these mixtures were then applied to deparaffinized tissue sections for an overnight incubation at 4 °C. The processing of sections used for light or fluorescence microscopy differed slightly.

When running the protocol for light microscopy, sections had been washed in 10 mM 4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid buffer (HEPES, pH 7.5) containing 0.1 mM CaCl₂ and then treated with this buffer containing 1% (w/v) carbohydrate-free bovine serum albumin (BSA), for 1 h at room temperature to saturate sites for non-specific binding of proteins. Following washing steps and after the incubation with lectin-containing solution in the absence/presence of glyco-inhibitors, visualisation of signals was accomplished by applying commercial Vectastain® ABC kit and Vector Red® reagents (Biozol, Eching, Germany).

The route to obtaining profiles of fluorescent staining started by incubating deparaffinized sections in a humid chamber with a 1% (w/v) solution of BSA in HEPES buffer containing 0.1 mM CaCl₂ and 0.1% (w/v) Tween-20 (blocking solution) to saturate non-specific binding of protein. Subsequently, sections were exposed to labelled lectins dissolved in blocking solution, e.g. Alexa Fluor®-555-labelled SBA, biotinylated DBA or HPA, or a ternary mixture containing the fluorescent SBA, biotinylated DBA and FITC-labelled MGL (CRD + stalk), in a humid chamber overnight at 4 °C. After two thorough washing steps (each for 5 min) in HEPES buffer/Tween 20 and a washing step in PBS/Tween-20 (PBS-T) to remove the labelled probes, Alexa Fluor®-647-conjugated streptavidin (Invitrogen, Thermo Fisher Scientific, Darmstadt, Germany) (solutions of 2–10 µg mL⁻¹ were tested, 5 µg mL⁻¹ was found to be optimal) in 1% BSA/PBS-T was applied for 1 h, then 4',6-diamidino-2-phenylindole (DAPI; 0.5 µg mL⁻¹) in PBS-T was added for 5 min at room temperature in a humid chamber in the dark.

Processed sections for light microscopy were counterstained with Mayer's haemalaun, dehydrated and mounted with Eukitt® (Kindler, Freiburg, Germany). For fluorescence microscopy, a mounting medium containing an anti-fading agent (Dako Fluorescence Mounting Medium; Agilent DAKO, Waldbronn, Germany) was used. Controls to exclude lectin-independent staining were performed by omission of the incubation step with lectin-containing solution. Controls of sugar specificity were carried out by using the cognate sugar in titrations. Lack of inhibition by carbohydrate-independent binding of the bulky scaffolds in bivalent compound 6 and tetravalent compound 5 by comparing the activity of GalNAc-presenting compounds 5 and 6 to interfere with lectin binding in parallel with testing the corresponding GlcNAc-presenting compounds 8 and 9 (0.05–2 mM of sugar) on MGL (CRD + stalk) and SBA.

Microphotographs were recorded using an AxioImager.M1 microscope (Carl Zeiss MicroImaging, Göttingen, Germany) equipped with an AxioCam MRm digital camera and the software AxioVision version 4.9. Semi-quantitative analysis by two independent observers was based on independently examining at least ten high-power fields per section. Signal intensity is expressed in the following grading system: –, no staining; (+), very weak but significant staining; +, weak staining; ++, medium staining; +++, strong staining; +++++, very strong staining. The percentage of positive cells is grouped in the categories: (1) 0% (no positive cells), (2) < 5% (few positive cells), (3) 5–20%, (4) 21–40%, (5) 41–60%, (6) 61–80%, (7) 81–95% (few negative cells), (8) up to 100% (no negative cells).

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

Inspiring discussions with Drs B. Friday and A. Leddoz are gratefully acknowledged. Research in this paper was funded in part by Science Foundation Ireland (Grant Number 12/IA/1398) and the



Irish Research Council (Grant Number GOIPG/2016/858). L. L. R.-H. thanks CONACYT-México (Grant 290936) for financial support.

Notes and references

- H.-J. Gabius and J. Roth, *Histochem. Cell Biol.*, 2017, **147**, 111–117.
- F. L. Harrison and C. J. Chesterton, *FEBS Lett.*, 1980, **122**, 157–165.
- W. M. Gallatin, I. L. Weissman and E. C. Butcher, *Nature*, 1983, **304**, 30–34.
- D. Vestweber and J. E. Blanks, *Physiol. Rev.*, 1999, **79**, 181–213.
- R. P. McEver, *Cardiovasc. Res.*, 2015, **107**, 331–339.
- G. L. Nicolson, *Int. Rev. Cytol.*, 1974, **39**, 89–190.
- L. Bhattacharyya, M. I. Khan, J. Fant and C. F. Brewer, *J. Biol. Chem.*, 1989, **264**, 11543–11545.
- C. Boscher, J. W. Dennis and I. R. Nabi, *Curr. Opin. Cell Biol.*, 2011, **23**, 383–392.
- H. Kaltner, S. Toegel, G. García Caballero, J. C. Manning, R. W. Ledeen and H.-J. Gabius, *Histochem. Cell Biol.*, 2017, **147**, 239–256.
- K.-i. Kasai, *Trends Glycosci. Glycotechnol.*, 2018, **30**, E221–E228.
- A. Varki, *J. Clin. Invest.*, 1997, **99**, 158–162.
- E. P. McGreal, J. L. Miller and S. Gordon, *Curr. Opin. Immunol.*, 2005, **17**, 18–24.
- H.-J. Gabius, J. C. Manning, J. Kopitz, S. André and H. Kaltner, *Cell. Mol. Life Sci.*, 2016, **73**, 1989–2016.
- S. Mayer, M. K. Raulf and B. Lepenies, *Histochem. Cell Biol.*, 2017, **147**, 223–237.
- Y. C. Lee and R. T. Lee, *Acc. Chem. Res.*, 1995, **28**, 321–327.
- Y. M. Chabre and R. Roy, in *The Sugar Code. Fundamentals of glycosciences*, ed. H.-J. Gabius, Wiley-VCH, Weinheim, Germany, 2009, ch. 4, pp. 53–70.
- G. C. Daskhan, N. Berthet, B. Thomas, M. Fiore and O. Renaudet, *Carbohydr. Res.*, 2015, **405**, 13–22.
- R. Roy, P. V. Murphy and H.-J. Gabius, *Molecules*, 2016, **21**, 629.
- A. P. Corfield and M. Berry, *Trends Biochem. Sci.*, 2015, **40**, 351–359.
- R. W. Ledeen and G. Wu, *Trends Biochem. Sci.*, 2015, **40**, 407–418.
- A. P. Corfield, *Histochem. Cell Biol.*, 2017, **147**, 119–147.
- J. Kopitz, *Histochem. Cell Biol.*, 2017, **147**, 175–198.
- J. Roth, *Exp. Pathol.*, 1978, **3**(suppl 1), 1–186.
- J. Schrével, D. Gros and M. Monsigny, *Prog. Histochem. Cytochem.*, 1981, **14**(2), 1–269.
- J. Roth, *Biochim. Biophys. Acta*, 1987, **906**, 405–436.
- S. S. Spicer and B. A. Schulte, *J. Histochem. Cytochem.*, 1992, **40**, 1–38.
- J. Roth, *Histochem. Cell Biol.*, 2011, **136**, 117–130.
- J. C. Manning, A. Romero, F. A. Habermann, G. García Caballero, H. Kaltner and H.-J. Gabius, *Histochem. Cell Biol.*, 2017, **147**, 199–222.
- R. Roy, Y. Cao, H. Kaltner, N. Kottari, T. C. Shiao, K. Belkhadem, S. André, J. C. Manning, P. V. Murphy and H.-J. Gabius, *Histochem. Cell Biol.*, 2017, **147**, 285–301.
- A. P. Corfield, *Biochim. Biophys. Acta*, 2015, **1850**, 236–252.
- S. André, S. O'Sullivan, C. Koller, P. V. Murphy and H.-J. Gabius, *Org. Biomol. Chem.*, 2015, **13**, 4190–4203.
- G. J. Miller, K. R. Broberg, C. Rudd, M. R. Helliwell, G. C. Jayson and J. M. Gardiner, *Org. Biomol. Chem.*, 2015, **13**, 11208–11219.
- D. Wu, S. Cheung, R. Daly, H. Burke, E. M. Scanlan and D. F. O'Shea, *Eur. J. Org. Chem.*, 2014, 6841–6845.
- D. Wu and D. F. O'Shea, *Org. Lett.*, 2013, **15**, 3392–3395.
- S. Knapp and D. S. Myers, *J. Org. Chem.*, 2002, **67**, 2995–2999.
- S. André, S. O'Sullivan, H.-J. Gabius and P. V. Murphy, *Tetrahedron*, 2015, **71**, 6867–6880.
- C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr.*, 2016, **B72**, 171–179.
- G. R. Gunther, J. L. Wang, I. Yahara, B. A. Cunningham and G. M. Edelman, *Proc. Natl. Acad. Sci. U. S. A.*, 1973, **70**, 1012–1016.
- D. K. Mandal and C. F. Brewer, *Biochemistry*, 1993, **32**, 5116–5120.
- D. Schwefel, C. Maierhofer, J. G. Beck, S. Seeberger, K. Diederichs, H. M. Moller, W. Welte and V. Wittmann, *J. Am. Chem. Soc.*, 2010, **132**, 8704–8719.
- T. Matsushita, K. Tsuchibuchi, T. Koyama, K. Hatano and K. Matsuoka, *Bioorg. Med. Chem. Lett.*, 2018, **28**, 1704–1707.
- H. Debray, D. Decout, G. Strecker, G. Spik and J. Montreuil, *Eur. J. Biochem.*, 1981, **117**, 41–55.
- M. E. Pereira, E. A. Kabat and N. Sharon, *Carbohydr. Res.*, 1974, **37**, 89–102.
- S. Hammarström, L. A. Murphy, I. J. Goldstein and M. E. Etzler, *Biochemistry*, 1977, **16**, 2750–2755.
- D. A. Baker, S. Sugii, E. A. Kabat, R. M. Ratcliffe, P. Hermentin and R. U. Lemieux, *Biochemistry*, 1983, **22**, 2741–2750.
- B. V. Torres, D. K. McCrumb and D. F. Smith, *Arch. Biochem. Biophys.*, 1988, **262**, 1–11.
- V. Piller, F. Piller and J.-P. Cartron, *Eur. J. Biochem.*, 1990, **191**, 461–466.
- S.-i. Iida, K. Yamamoto and T. Irimura, *J. Biol. Chem.*, 1999, **274**, 10697–10705.
- 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 and Y. van Kooyk, *Int. Immunol.*, 2005, **17**, 661–669.
- J.-F. Sanchez, J. Lescar, V. Chazalet, A. Audfray, J. Gagnon, R. Alvarez, C. Breton, A. Imberty and E. P. Mitchell, *J. Biol. Chem.*, 2006, **281**, 20171–20180.
- O. Oyelaran, Q. Li, D. Farnsworth and J. C. Gildersleeve, *J. Proteome Res.*, 2009, **8**, 3529–3538.
- N. Mortezaei, H. N. Behnken, A. K. Kurze, P. Ludewig, F. Buck, B. Meyer and C. Wagener, *Glycobiology*, 2013, **23**, 844–852.
- G. Artigas, J. T. Monteiro, H. Hinou, S. I. Nishimura, B. Lepenies and F. Garcia-Martin, *J. Med. Chem.*, 2017, **60**, 9012–9021.



- 54 S. Hadjalirezaei, G. Picco, R. Beatson, J. Burchell, B. T. Stokke and M. Sletmoen, *PLoS One*, 2017, **12**, e0175323.
- 55 A. Imberty, F. Casset, C. V. Gegg, M. E. Etzler and S. Perez, *Glycoconjugate J.*, 1994, **11**, 400–413.
- 56 A. Dessen, D. Gupta, S. Sabesan, C. F. Brewer and J. C. Sacchettini, *Biochemistry*, 1995, **34**, 4933–4942.
- 57 M. H. Dao-Thi, T. W. Hamelryck, J. Bouckaert, F. Korber, V. Burkow, F. Poortmans, M. Etzler, G. Strecker, L. Wyns and R. Loris, *Acta Crystallogr.*, 1998, **D54**, 1446–1449.
- 58 T. W. Hamelryck, R. Loris, J. Bouckaert, M. H. Dao-Thi, G. Strecker, A. Imberty, E. Fernandez, L. Wyns and M. E. Etzler, *J. Mol. Biol.*, 1999, **286**, 1161–1177.
- 59 P. J. Coombs, R. Harrison, S. Pemberton, A. Quintero-Martinez, S. Parry, S. M. Haslam, A. Dell, M. E. Taylor and K. Drickamer, *J. Mol. Biol.*, 2010, **396**, 685–696.
- 60 S. A. Jégouzo, A. Quintero-Martínez, X. Ouyang, A. dos Santos, M. E. Taylor and K. Drickamer, *Glycobiology*, 2013, **23**, 853–864.
- 61 H. Tateno, N. Uchiyama, A. Kuno, A. Togayachi, T. Sato, H. Narimatsu and J. Hirabayashi, *Glycobiology*, 2007, **17**, 1138–1146.
- 62 K. T. Pilobello, D. E. Slawek and L. K. Mahal, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 11534–11539.
- 63 M. Watanabe, Z. Takeda, Y. Urano and T. Muramatsu, in *Teratocarcinoma and Embryonic Cell Interactions*, ed. T. Muramatsu, G. Gachelin, A. A. Moscona and Y. Ikawa, Academic Press/Japan Scientific Societies' Press, Tokyo, 1982, pp. 217–228.
- 64 M. Sato and T. Muramatsu, *Differentiation*, 1985, **29**, 29–38.
- 65 Y. Kamada, H. Muramatsu, Y. Arita, T. Yamada and T. Muramatsu, *J. Biochem.*, 1991, **109**, 178–183.
- 66 U. Brinck, R. Bosbach, M. Korabiowska, A. Schauer and H.-J. Gabius, *Histol. Histopathol.*, 1995, **10**, 61–70.
- 67 E. Saeland, S. J. van Vliet, M. Bäckström, V. C. M. van den Berg, T. B. H. Geijtenbeek, G. A. Meijer and Y. van Kooyk, *Cancer Immunol. Immunother.*, 2007, **56**, 1225–1236.
- 68 P. Rohse and V. Wittmann, *Chem. –Eur. J.*, 2016, **22**, 9724–9733.
- 69 J. C. Manning, G. García Caballero, C. Knospe, H. Kaltner and H.-J. Gabius, *J. Anat.*, 2017, **231**, 23–37.
- 70 J. C. Manning, G. García Caballero, F. M. Ruiz, A. Romero, H. Kaltner and H.-J. Gabius, *Trends Glycosci. Glycotechnol.*, 2018, **30**, SE11–SE20.
- 71 J. Nio-Kobayashi, *Trends Glycosci. Glycotechnol.*, 2018, **30**, SE89–SE96.
- 72 T. Angata and E. C. M. Brinkman-van der Linden, *Biochim. Biophys. Acta*, 2002, **1572**, 294–316.
- 73 J. Iwaki and J. Hirabayashi, *Trends Glycosci. Glycotechnol.*, 2018, **30**, SE137–SE153.
- 74 J. Kopitz, C. von Reitzenstein, S. André, H. Kaltner, J. Uhl, V. Ehemann, M. Cantz and H.-J. Gabius, *J. Biol. Chem.*, 2001, **276**, 35917–35923.
- 75 H. Sanchez-Ruderisch, C. Fischer, K. M. Detjen, M. Welzel, A. Wimmel, J. C. Manning, S. André and H.-J. Gabius, *FEBS J.*, 2010, **277**, 3552–3563.
- 76 D. Weinmann, K. Schlangen, S. André, S. Schmidt, S. M. Walzer, B. Kubista, R. Windhager, S. Toegel and H.-J. Gabius, *Sci. Rep.*, 2016, **6**, 39112.
- 77 D. Weinmann, M. Kenn, S. Schmidt, K. Schmidt, S. M. Walzer, B. Kubista, R. Windhager, W. Schreiner, S. Toegel and H.-J. Gabius, *Cell. Mol. Life Sci.*, DOI: 10.1007/s00018-018-2856-2.
- 78 B. Belardi, G. P. O'Donoghue, A. W. Smith, J. T. Groves and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2012, **134**, 9549–9552.
- 79 H.-J. Gabius, *Folia Biol. (Praha)*, 2017, **63**, 121–131.
- 80 Q. Xiao, A.-K. Ludwig, C. Romano, I. Buzzacchera, S. E. Sherman, M. Vetro, S. Vértesy, H. Kaltner, E. H. Reed, M. Moller, C. J. Wilson, D. A. Hammer, S. Oscarson, M. L. Klein, H.-J. Gabius and V. Percec, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E2509–E2518.
- 81 A. Schwarz and A. H. Futerman, *J. Histochem. Cytochem.*, 1997, **45**, 611–618.
- 82 J. W. Celie, R. H. J. Beelen and J. van den Born, *J. Immunol. Methods*, 2005, **298**, 155–159.
- 83 J. C. Rogers and S. Kornfeld, *Biochem. Biophys. Res. Commun.*, 1971, **45**, 622–629.
- 84 S. Kojima, M. Ishido, A. Kubota, T. Kubodera, T. Hellmann, B. Kohnke-Godt, B. Wosgien and H.-J. Gabius, *Biol. Chem. Hoppe-Seyler*, 1990, **371**, 331–338.
- 85 S. André, C. Unverzagt, S. Kojima, X. Dong, C. Fink, K. Kayser and H.-J. Gabius, *Bioconjugate Chem.*, 1997, **8**, 845–855.
- 86 C. Unverzagt, S. André, J. Seifert, S. Kojima, C. Fink, G. Srikrishna, H. Freeze, K. Kayser and H.-J. Gabius, *J. Med. Chem.*, 2002, **45**, 478–491.
- 87 A. Ogura, T. Tahara, S. Nozaki, K. Morimoto, Y. Kizuka, S. Kitazume, M. Hara, S. Kojima, H. Onoe, A. Kurbangalieva, N. Taniguchi, Y. Watanabe and K. Tanaka, *Sci. Rep.*, 2016, **6**, 21797.
- 88 L. Latypova, R. Sibgatullina, A. Ogura, K. Fujiki, A. Khabibrakhmanova, T. Tahara, S. Nozaki, S. Urano, K. Tsubokura, H. Onoe, Y. Watanabe, A. Kurbangalieva and K. Tanaka, *Adv. Sci.*, 2017, **4**, 1600394.
- 89 A. Ogura, S. Urano, T. Tahara, S. Nozaki, R. Sibgatullina, K. Vong, T. Suzuki, N. Dohmae, A. Kurbangalieva, Y. Watanabe and K. Tanaka, *Chem. Commun.*, 2018, **54**, 8693–8696.
- 90 H.-J. Gabius, B. Wosgien, M. Hendrys and A. Bardosi, *Histochemistry*, 1991, **95**, 269–277.
- 91 J. Kopitz, Q. Xiao, A.-K. Ludwig, A. Romero, M. Michalak, S. E. Sherman, X. Zhou, C. Dazen, S. Vértesy, H. Kaltner, M. L. Klein, H.-J. Gabius and V. Percec, *Angew. Chem., Int. Ed.*, 2017, **56**, 14677–14681.
- 92 H. Kaltner, G. García Caballero, F. Sinowatz, S. Schmidt, J. C. Manning, S. André and H.-J. Gabius, *Biochim. Biophys. Acta*, 2016, **1860**, 2298–2312.
- 93 J. C. Manning, G. García Caballero, C. Knospe, H. Kaltner and H.-J. Gabius, *Ann. Anat.*, 2018, **217**, 66–81.



Structural Biology

How altering the modular architecture affects aspects of lectin activity: case study on human galectin-1

Tanja J Kutzner^{2,*}, Adele Gabba^{3,*}, Forrest G FitzGerald^{4,*}, Nadezhda V Shilova^{5,*}, Gabriel García Caballero², Anna-Kristin Ludwig², Joachim C Manning², Clemens Knospe⁶, Herbert Kaltner², Fred Sinowatz⁶, Paul V Murphy^{3,1}, Mare Cudic^{4,1}, Nicolai V Bovin^{5,7,1}, and Hans-Joachim Gabius^{2,1}

²Institute of Physiological Chemistry, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, 80539 Munich, Germany, ³School of Chemistry, National University of Ireland, Galway H91 TK33, Ireland, ⁴Department of Chemistry and Biochemistry, Florida Atlantic University, Boca Raton FL 33431, USA, ⁵Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Laboratory of Carbohydrates, Russian Academy of Sciences, 117997 Moscow, Russia, ⁶Institute of Anatomy, Histology and Embryology, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, 80539 Munich, Germany, and ⁷Centre for Kode Technology Innovation, School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland 1010, New Zealand

¹To whom correspondence should be addressed: Tel: +353 91 492 465, e-mail: paul.v.murphy@nuigalway.ie (Paul V. Murphy); Tel: +1 561 297 4645, Fax: +1 561 297 2759 (Mare Cudic), e-mail: mcudic@fau.edu; Tel: +7 495 330 71 38, Fax: +7 495 330 55 92, e-mail: professorbovin@yandex.ru (Nicolai V. Bovin); Tel: +49 89 2180 2290, Fax: +49 89 2180 992290 (H.-J. Gabius), e-mail: gabius@tph.vetmed.uni-muenchen.de

*These authors have contributed equally to this work.

Received 14 February 2019; Revised 26 April 2019; Editorial decision 26 April 2019; Accepted 11 May 2019

Abstract

Discoveries on involvement of glycan–protein recognition in many (patho)physiological processes are directing attention to exploring the significance of a fundamental structural aspect of sugar receptors beyond glycan specificity, i.e., occurrence of distinct types of modular architecture. In order to trace clues for defining design–functionality relationships in human lectins, a lectin's structural unit has been used as source material for engineering custom-made variants of the wild-type protein. Their availability facilitates comparative analysis toward the stated aim. With adhesion/growth-regulatory human galectin-1 as example, the strategy of evaluating how changes of its design (here, from the homodimer of non-covalently associated domains to (i) linker-connected di- and tetramers and (ii) a galectin-3-like protein) affect activity is illustrated by using three assay systems of increasing degree of glycan complexity. Whereas calorimetry with two cognate disaccharides and array testing with 647 (glyco)compounds disclosed no major changes, galectin histochemical staining profiles of tissue sections that present natural glycome complexity revealed differences between wild-type and linker-connected homo-oligomers as well as between the galectin-3-like variant and wild-type galectin-3 for cell-type positivity, level of intensity at the same site and susceptibility for inhibition by a bivalent glycocompound. These results underscore the strength of the documented approach. Moreover, they give direction to proceed to

- (i) extending its application to other members of this lectin family, especially galectin-3 and
- (ii) then analyzing impact of architectural alterations on cell surface lattice formation and ensuing biosignaling systematically, considering the variants' potential for translational medicine.

Key words: calorimetry, glycan, histochemistry, lectin, sugar code

Introduction

What distinguishes carbohydrates from the other alphabets of life (i.e., nucleotides and amino acids) is their ability to generate oligomers of unsurpassed coding capacity (Laine 1997; Schnaar 2015; Solís et al. 2015; Gabius and Roth 2017). Indeed, this chemical potential is turned into glycome complexity (Ginsburg and Neufeld 1969; Roth 1987; Brockhausen and Schachter 1997; Cummings 2009; Zuber and Roth 2009). As a consequence, proteins and sphingolipids can present a wide diversity of sugar-encoded signals (Buddecke 2009; Schengrund 2015; Corfield 2017; Kopitz 2017; Ledeen et al. 2018; Sandhoff et al. 2018). One route of the flow of this information toward eliciting effects is that these “messages” are “read” and then “translated” by sugar receptors (lectins) (Gabius 2017; Manning, Romero, et al. 2017; Kaltner, García Caballero, et al. 2018). Their target (glycan) specificity and their modular architecture are assumed to be key factors that determine the profile of the functional outcome of glycan–lectin recognition. In fact, evolution has used ancestral lectin domains as source material to generate diversity on the levels of sequence and modular design. Along this line of reasoning, each type of design to present the common carbohydrate recognition domain (CRD) in a lectin family can be postulated to have its own characteristic mission. Intriguingly, not all theoretically possible modes of CRD arrangement are apparently realized in Nature, and, equally puzzling at present, differences in modular display within a lectin family exist for example between vertebrates and invertebrates. The fundamental issue on defining design–functionality relationships and the given open questions prompted us to perform this study.

Looking at adhesion/growth-regulatory galectins (Gals) as the test case, the CRD is presented in three modes in vertebrates (Hirabayashi 1997; Cooper 2002; Kaltner et al. 2017; Manning, García Caballero et al. 2018). As illustrated in Figure 1, the typical cross-linking (lattice-forming) activity of galectins is made possible via different structural means: bivalency is attained by non-covalent or covalent (linker-dependent) association. The formation of CRD oligomers can alternatively involve a second type of module, i.e., an N-terminal tail (NT) with collagen-like repeats. Obviously, the restriction to the three types of design illustrated in Figure 1 poses the question as to why this particular set has phylogenetically become a stable trait in vertebrates. In order to resolve this issue, engineering of protein variants that display a distinct CRD in other types of design than the natural one enables to characterize the impact of altering CRD presentation on lectin properties.

Selecting homodimeric (proto-type) galectin-1 (Gal-1) for proof-of-principle work, the conversion of non-covalent association to covalent connection by short (Gly-Gly), flexible or rigid (α -helical) linkers (Bättig et al. 2004; Bi et al. 2008; Earl et al. 2011; Tribulatti et al. 2012; Vértessy et al. 2015), by disulfide bonding directed by a leucine zipper (van der Leij et al. 2007), by fusion of an immunoglobulin G_1 (IgG₁) F_c part to the CRD (Tsai et al. 2008; Cedeno-Laurent et al. 2010) and by site-specific (Cys130-dependent) self-

conjugation of the Cys2Ser/Cys16Ser/Cys88Ser triple mutant using poly(ethylene glycol) diacrylate (Fetis and Hudalla 2018) led to bio-active homodimers. As predicted by studies using atomic force microscopy or glycodendrimersomes, the non-covalent association of this CRD appears to be more suited for *cis*-cross-linking and transient *trans*-bridging than for establishing firm contacts (Detmann et al. 2000; Zhang et al. 2015). The insertion of a linker opens the route to generate homo-oligomers with the CRD of human Gal-1 beyond the dimer. Hereby, the artificial equivalent of the assumedly anti-microbial tandem-repeat-type oyster galectins, which are tetramers not found in vertebrates (Tasumi and Vasta 2007; Feng et al. 2013; 2015), became available (Kopitz et al. 2017). Equally important, a design switch between human galectins had recently been accomplished: Gal-1's CRD had been combined with the NT of the chimera-type Gal-3 to turn homodimeric Gal-1 into a Gal-3-like Gal-1 variant termed Gal-3NT/1 (Ludwig, Michalak, et al. 2019; Ludwig, Kaltner, 2019). Thus, rational engineering with the Gal-1 CRD has reached the status to facilitate comparative analysis of the wild-type protein versus variants of different design.

The selection of assay systems should consider the documented influence of the mode of glycan presentation, either free in solution (measured for example by frontal affinity chromatography or isothermal titration calorimetry, ITC) or on a solid phase/cell surface, on the extent of binding (Ahmad et al. 2002; Hirabayashi et al. 2002; Leppänen et al. 2005; Stowell et al. 2004; 2008; Song et al. 2009; Iwaki and Hirabayashi 2018). As consequence, protein design–ligand binding relationships were systematically determined in three assay systems: (i) calorimetry to characterize the thermodynamics of binding the canonical ligands lactose (Lac) and N-acetyllactosamine (LacNAc), (ii) glycan microarray monitoring and (iii) galectin histochemistry on sections of two organs, i.e., murine epididymis and jejunum, that present characteristic, physiologically complex glycomes. By applying this experimental strategy, combining protein engineering with comparing activity profiles of the resulting proteins all built with the Gal-1 CRD as lectin part, the presented results provide insights into the effect of changes of modular design on aspects of lectin properties.

Results

The panel of Gal-1-based variants

Natural human Gal-1 is a homodimer non-covalently stabilized by mostly hydrophobic contacts of the CRD interfaces that appears to undergo monomerization from low- μ M concentrations downward (Giudicelli et al. 1997; López-Lucendo et al. 2004; Stowell et al. 2009). Playing a modular puzzle with the human Gal-1 CRD as building block, five variants were engineered by (i) covalent conjugation of two or four Gal-1 CRDs using either the 33-amino-acid linker of human Gal-8 termed 8S (Figure 2, left) or the dipeptide Gly-Gly termed GG (Figure 2, center) to generate homodi- and tetramers (Figure 2, left and center) and (ii) human Gal-3's NT as the

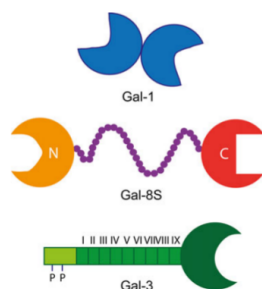


Fig. 1. Illustration of the three types of modular design of vertebrate galectins, i.e., proto-type (non-covalent association of two identical CRDs as in Gal-1), tandem-repeat-type (two different covalently connected CRDs, in human Gal-8 either by the shown short (S) 33-amino-acid linker or by its longer version (L; 74 amino acids)) and chimera-type (a CRD conjugated to an N-terminal tail (NT) composed of a peptide with two (Ser) sites for phosphorylation and non-triple helical collagen-like repeats, nine in human Gal-3) proteins (from top to bottom).

second module (Figure 2, right). As a consequence, fundamental design switches originating from the proto-type structure are established (these basic structures are shown in Figure 1). These five proteins that share presence of the Gal-1 CRD could all be isolated by affinity chromatography on Lac-presenting beads, ascertaining their activity for binding the cognate disaccharide. They constitute the toolbox shown in Figure 2 to proceed to the comparative analysis with the wild-type protein, first by ITC.

Binding properties: ITC

Each protein was processed in the same experimental set-up using the common physiological target of the *ga*(lactose-binding)*lectins*, i.e., LacNAc, and also Lac. As exemplarily shown in Figure 3 for wild-type Gal-1, its GG-linked di- and tetramers as well as its Gal-3-like Gal-3NT/1 variant, injections of ligand-containing solution led to heat release, whose extent successively decreased as ligand concentration reached saturation. The calculated number of binding sites per protein (n) was invariably close to the expected values of 2 (for homodimers), 4 (for homotetramers) and 1 for the Gal-3-like monomer (Table 1), fully in line with completely maintained lectin activity. As commonly reported for human galectins (e.g., by Dam et al. 2005), the binding

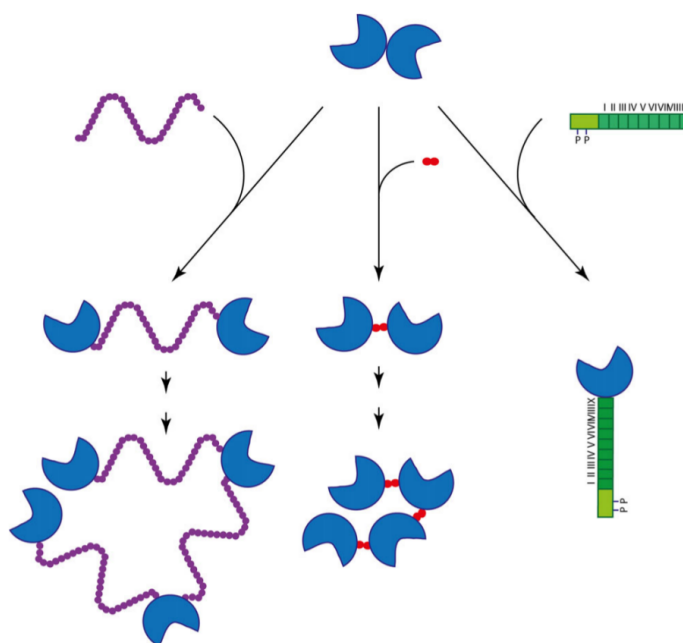


Fig. 2. Illustration of the routes of modular engineering to turn the CRD of Gal-1 into covalently associated homodi- and tetramers by linker insertion (left: 33-amino-acid linker of human Gal-8 shown in Figure 1; center: Gly-Gly) and into a Gal-3-like variant by bringing it together with Gal-3's NT thus termed Gal-3NT/1 (right). Color coding is used as in Figure 1.

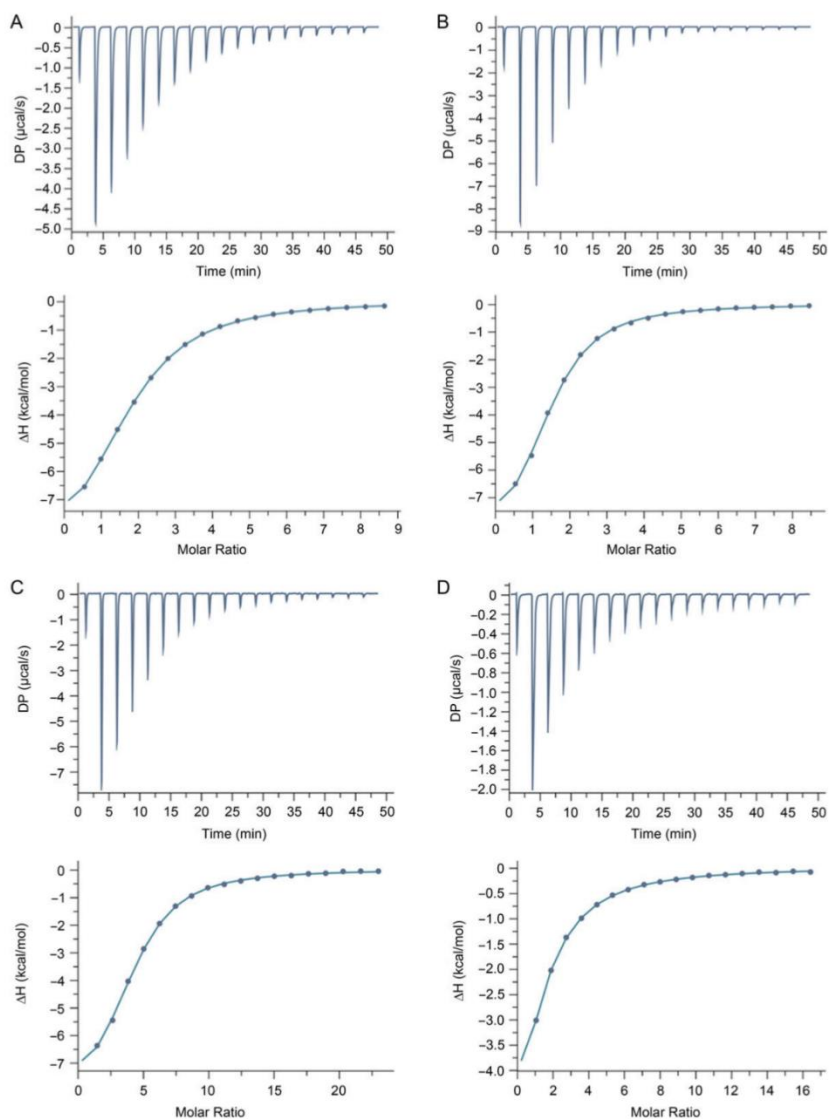


Fig. 3. Illustration of the pair of thermogram (top) and isotherm (bottom) for the titration of galectin-containing solution in phosphate buffer (pH 7.2) containing 5 mM or 150 mM NaCl and 10 mM β -mercaptoethanol with 2 μl aliquots of a 6-mM LacNAc-containing solution in 150 s intervals at 25°C in the cases of wild-type Gal-1 (A), the GG-linked homodimer (B), the GG-linked homotetramer (C) and the Gal-3NT1 variant (D).

process was enthalpically driven with a typical entropic penalty (Table I). This was also the case when using Lac as ligand, albeit at the known lower level of affinity (Supplementary data, Table SI).

Covalently connecting the two Gal-1 CRDs by bringing in a linker (either GG or 8S) and increasing the number of CRDs to form the tetramer may affect the avidity of consecutive binding processes. As an

Table I. Thermodynamic parameters and Hill coefficients of binding of LacNAc (6 mM) to human galectins at 25°C.

Cell	[Cell] (μM)	K_d ($\times 10^4 \text{ M}^{-1}$)	$-\Delta G$ (kcal/mol)	$-\Delta H$ (kcal/mol)	$-\Delta S$ (kcal/mol)	n (sites/protein)	K_d (μM)	Hill coefficient ^a
Gal-1	110	1.17	5.55	9.81 (± 0.083)	4.26	2.09	85.2 (± 1.52)	1.10
(Gal-1) ₂ -GG	95	1.99	5.88	9.15 (± 0.220)	3.27	1.96	50.1 (± 2.96)	1.09
(Gal-1) ₂ -8S	95	1.10	5.51	10.8 (± 0.122)	5.29	1.98	90.9 (± 1.78)	1.069
(Gal-1) ₄ -GG	50	1.99	5.87	8.78 (± 0.188)	2.91	3.97	50.2 (± 2.75)	1.15
(Gal-1) ₄ -8S	34	1.12	5.52	10.1 (± 0.267)	4.58	4.06	89.5 (± 3.14)	1.09
Gal-3NT/1	70	0.74	5.27	11.8 (± 0.081)	6.53	0.98	136 (± 1.63)	0.980

^aObtained from successive three-point tangent slope at $y = 0$.

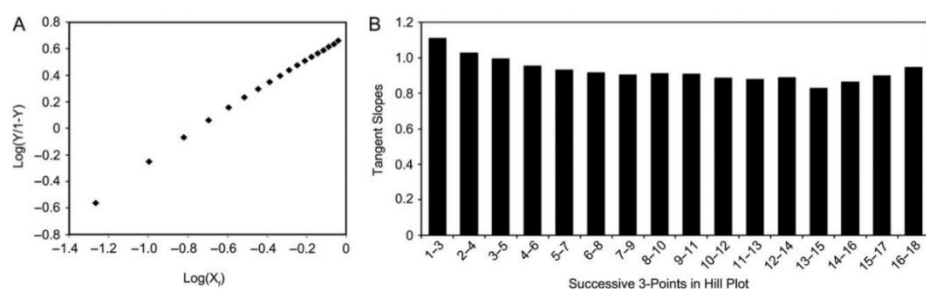


Fig. 4. Illustration of the Hill plot of the ITC data for LacNAc (6 mM) binding to Gal-3NT/1 (please see Figure 3D) at functional valency of 1 (A) and the corresponding bar graph of three-point tangent slope data in the course of the titration (B).

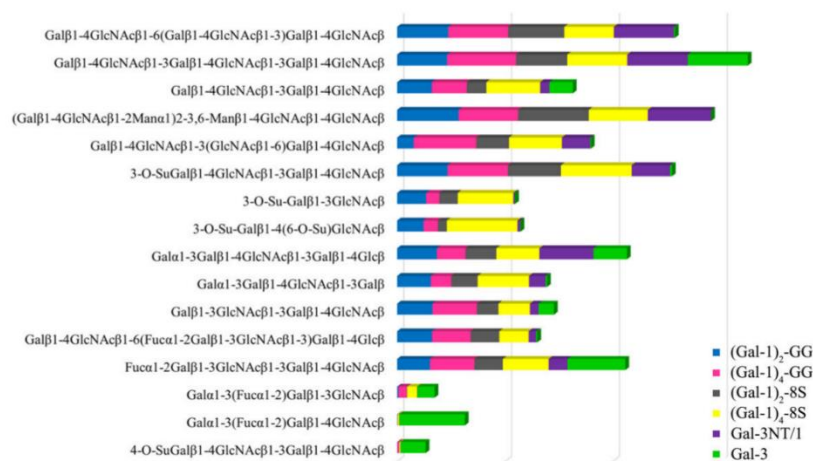
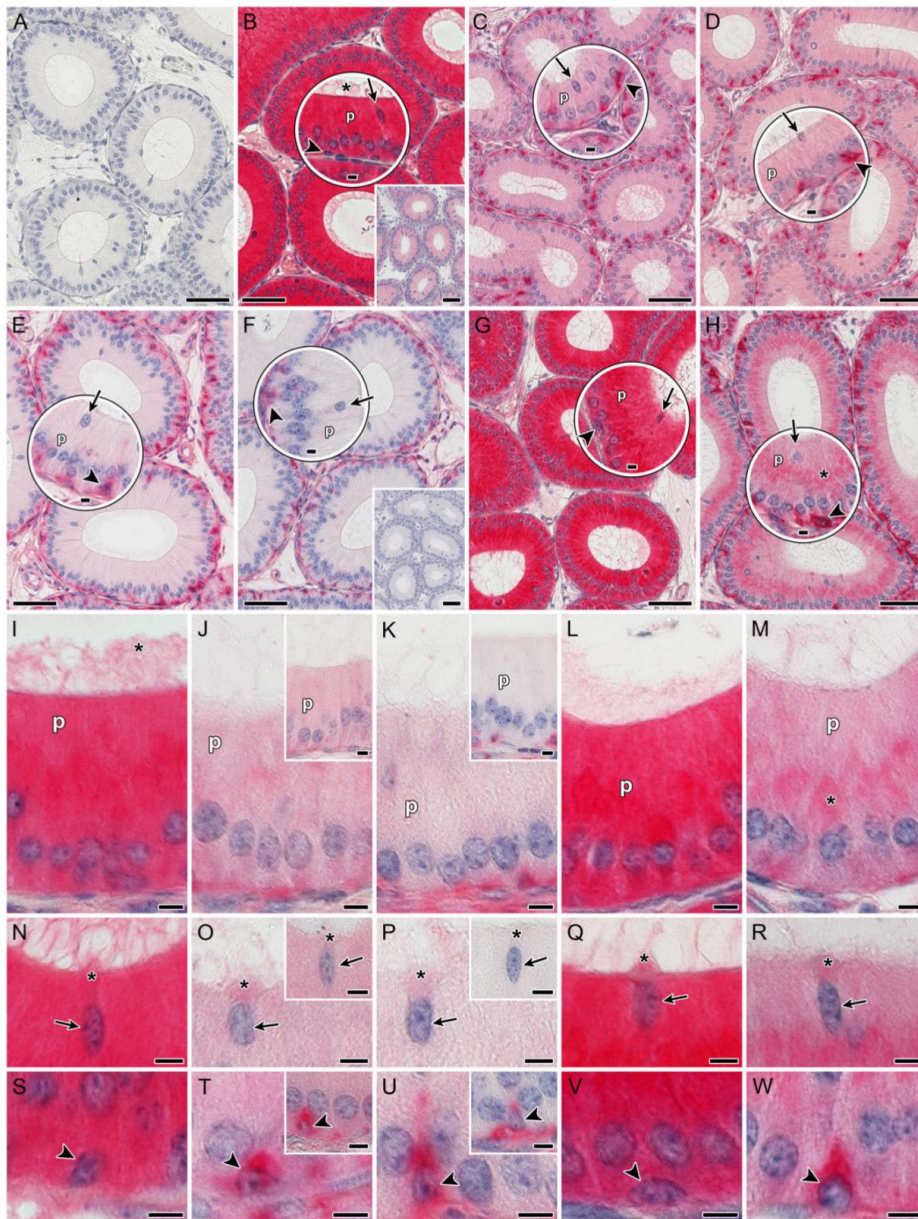


Fig. 5. Illustration of side-by-side comparisons of relative signal intensity of galectin binding to selected glycans within the 647-compound-based array.

indicator for cooperativity of binding of a monovalent ligand to a protein (complex) with at least two binding sites, Hill plots were derived from the ITC data and presented as $\log(\text{concentration of free ligand})$ vs $\log(\text{fraction of ligand-loaded galectin}/\text{fraction of ligand-free galectin})$. Together with the advantage of covering all available data by its logarithmic scaling in this type of plot, any (substantial) deviation from

linearity (with a slope of 1.0) in the Hill plot will signal cooperativity of binding processes. The data obtained for the Gal-3NT/1 variant exemplarily illustrate the linearity of the Hill plot in the case of binding to a monovalent ligand, here LacNAc (Figure 4A). Fittingly, the tangent slopes of successively calculated three-point intervals are around 1.0 throughout the titration (Figure 4B).



Downloaded from <https://academic.oup.com/glycob/article-abstract/29/8/593/5489794> by National University of Ireland, Galway user on 22 August 2019

Applying this type of data processing to each case, the tabulated slope values were obtained. As summarized in Table I, they provide no robust evidence for cooperativity, considering deviations from normality mostly occurring at minimal changes of fractional occupancy, as shown in Supplementary data, Figure S1A,B, and the error brought in by Q subtraction. When running the titrations with this protein and Lac as ligand (independently up to 6 mM and 10 mM), slope values of 1.02 and 1.08, respectively, were obtained (Supplementary data, Figure S1C–F), as all other titrations with Lac (up to 14 mM for the 8S-linked tetramer) led to slope values close to 1 (Supplementary data, Table SI). These results document full loading of the proteins with ligand and a maintained thermodynamics with enthalpic gain as driving force for binding in each case. In addition, titrations with the GG-linked trimer and Lac (6 mM, 10 mM, 14 mM)/LacNAc (6 mM) consistently resulted in n -values close to 3, enthalpically driven thermodynamics and slope values close to 1 (Supplementary data, Tables SI/III). Hill plots therefore indicate no solid evidence for an occurrence of positive cooperativity by the structural remodeling of the dimer to tri- and tetramers. Moreover, computational processing of data with the PEAQ software yielded fitting exclusively for the one-set-of-sites model.

In order to probe into binding properties of this panel for glycans beyond LacNAc when presented on a surface, each protein was biotinylated under activity-preserving conditions, then the proteins' Lac-inhibitable binding to surface-presented glycoprotein (asialofetuin) and to cells was ascertained and binding properties on a glycan microarray tested.

Binding properties: printed glycan array

The glycan array comprised a total of 647 printed substances encompassing glycoconjugates (mono- to oligosaccharides and derivatives, glycopeptides, (lipo)polysaccharides and glycosaminoglycans) and peptides. Its capacity to delineate fine-specificity differences among closely related family members had previously been documented for chicken galectins (García Caballero et al. 2016) so that this test system was applied for this panel of proteins. With the exception of (biotinylated or fluorescent) wild-type Gal-1, all proteins proved active in this setting and gave graded signal profiles of binding, the intensity values recorded at the constant (mass) concentration of 50 µg/mL (please see Figure 5 for a side-by-side comparison; Supplementary data, Figure S2 for a bar graph; for complete

listings of signal intensity, please see Supplementary data, Tables SIV–SIX).

In principle, LacNAc and its oligomers were found to be binding partners for the Gal-1-type CRD. Few major differences occurred between di- and tetramers, here the particularly strong signal intensity of LacNAc sulfated at two sites (3-O-SuGal and 6-O-Su-GlcNAc) when tested with the 8S-linked tetramer (Figure 5; Supplementary data, Figure S2). The broad-panel testing was also a means to answer the question on an influence of Gal-3's NT on ligand binding of the CRD of Gal-1, for example by the tail's back-folding or tendency for aggregation. Affinity to internal (reducing-end) LacNAc units and the low-level recognition of the histo-blood group B tetrasaccharide clearly separate the proteins with the Gal-1 CRD, especially Gal-3-like Gal-3NT/1, from wild-type Gal-3 (Supplementary data, Figure S2). Obviously, these cases of glycans disclose that Gal-1's CRD rather maintains its binding pattern irrespective of the tested alterations of the protein architecture when interacting with surface-immobilized glycans in arrays. In order to take the test system from array-presented glycoconjugates to cellular glycomes, tissue sections are an experimental platform that can trace differences up to the level of cell types. Of special note, in comparison to array-type presentation, the surface of tissue sections, by presenting physiologically complex glycomes, offers the naturally encountered possibility to bridge two structurally different binding partners for the cross-linking galectins.

Binding properties: galectin histochemistry

Glycophenotyping of sections of adult fixed murine epididymis and jejunum by plant, fungal and invertebrate lectins had indicated the suitability of these two organs for the given purpose (Lohr et al. 2010; Kaltner, Manning, et al. 2018). Sections of fixed tissues were processed first by controls to exclude signal generation in the absence of labeled galectin (Figure 6A; Supplementary data, Figure S3A) and to ascertain sensitivity of signal generation to presence of the cognate sugar (Figure 6B; Supplementary data, Figure S3B), then by titrations to determine the optimal concentration that avoids significant background staining for systematic comparisons. The distribution profiles of the galectin-dependent and Lac-inhibitable staining are summarized in Table II for the data on adult murine epididymis.

Fig. 6. Staining profiles obtained with the six biotinylated galectin proteins in cross sections through the initial segment of fixed murine epididymis. (A–H) Microphotographs present overviews with higher-level magnifications of distinct regions (inserted circle above the respective area including principal (p), apical (arrow) and basal cells (arrowhead)). (I–W) Moreover, enlarged views on these distinct cell types (i.e., principal cells (p), I–M; apical cells (arrows), N–R; basal cells (arrowheads), S–W) are given. (A) Negative control by omission of the incubation step with first-step reagent (labeled galectin) to exclude lectin-independent signal generation. (B) Strong binding of Gal-1 in the epithelial lining and comparatively weaker positivity in stereocilia (asterisk). Inset to B shows extent of reduction of galectin binding by co-incubation of labeled Gal-1 with Lac (200 mM). (C–F) Variants stained cytoplasm of epithelial cells, particularly strong in basal cells. In principal and apical cells, intensity was moderate for (Gal-1)₂-GG (C) and for (Gal-1)₂-8S (D) or weak ((Gal-1)₄-GG (E)). No staining was detected in the case of (Gal-1)₄-8S (F). Presence of 10 mM Lac completely inhibited binding of (Gal-1)₄-8S (inset to F; please see also Figure 8A, B). (G) Labeled Gal-3NT/1 generated a staining profile and degree of intensity comparable to Gal-1 (B), whereas binding sites for Gal-3 were detected at moderate intensity of staining supranuclearly (asterisk) in principal cells and at very strong intensity in basal cells (H). Enlarged views of the three main cell types (please see corresponding encircled areas in B–H) document nearly identical staining profiles for principal (p, I–M) and apical cells (arrows, N–R), except for a slightly stronger staining in the supranuclear cytoplasm of apical cells (asterisks). Incubations with labeled Gal-1 or Gal-3NT/1 led to strong and rather homogeneous staining of cytoplasm of principal (I, L) and of apical cells (N, Q). Weak staining intensity of principal and apical cells by the two variants (Gal-1)₂-GG (J, O) and (Gal-1)₂-8S (insets to J and O) as well as at best very weak staining by (Gal-1)₄-GG (K, P) and (Gal-1)₄-8S (insets to K and P) was recorded. Gal-3 binding was detected in apical and basal cytoplasm of both cell types (M, R) and, with moderate intensity, supranuclearly in principal cells (M, asterisk). (S–W) Basal cells were positive after processing of sections with labeled (Gal-1)₂-GG (T), (Gal-1)₂-8S (inset to T), (Gal-1)₄-GG (U), (Gal-1)₄-8S (inset to U) and Gal-3 (W). Processing with labeled Gal-1 (S) or Gal-3NT/1 (V) resulted in staining intensity of basal cells, which was not different from that of principal and apical cells. The following concentrations were applied: Gal-1, (Gal-1)₂-GG, (Gal-1)₂-8S, (Gal-1)₄-GG, (Gal-1)₄-8S, Gal-3NT/1: 0.5 µg/mL; Gal-3: 8.0 µg/mL. Scale bars are 50 µm (A–H, bottom right insets in B, F) or 5 µm (circles in B–H, I–W).

Table II. Distribution and cellular localization of galectin-dependent and Lac-inhibitable staining in sections of fixed adult murine epididymis^a

Type of protein	Gal-1	(Gal-1) ₂ -GG	(Gal-1) ₂ -8S	(Gal-1) ₄ -GG	(Gal-1) ₄ -8S	Gal-3NT/1	Gal-3
Site of staining							
Principal cells							
Stereocilia	++	-/+	-/+	-	-	(+)	-/+
Apical ^b	++++	+/+	+/+	(+)	-/+	++++	+
Supranuclear ^b	++++	+/+	+/+	(+)	-/+	++++	++
Basal ^b	++++	(+)	(+)	(+)	-/+	++++	+
Apical cells	++++	+/+ ^c	+/+ ^c	-/(+) ^c	-/(+) ^c	++++	+/+ ^c
Basal cells	++++	+++	+++	++++	++++	++++	++++
Smooth muscle cells	-	-	-	-	-	-	-
Connective tissue	-	-	-	+	+	-	-

^aIntensity of staining in sections is grouped into the following categories: -, no staining; (+), very weak but above background; +, weak; ++, medium; +++, strong; +++++, very strong.

^bPositivity of given regions of cytoplasm.

^cStaining intensity of supranuclear cytoplasm of apical cells consistently in higher category.

The overall quantitative differences of staining intensity seen between wild-type Gal-1 (yielding up to very strong (++++)) intensity) and its covalently linked di- and tetrameric variants (only reaching up to weak-level (+) intensity) in the cases of principal and apical but not basal cells (in this case, all proteins caused high (++++/++++) signal intensity) are illustrated in Figure 6B–F. High-magnification microphotographs substantiate the cell-type-dependent differences for comparisons of wild-type Gal-1 (Figure 6I, N, S) versus the GG-linked dimer (Figure 6J, O, T) or tetramer (Figure 6K, P, U). The similarity of staining profiles in signal distribution and intensity when testing Gal-1 and the Gal-3NT/1 variant, shown in Figure 6B, G, is documented in more detail in Figure 6I, N, S (Gal-1) and Figure 6L, Q, V (Gal-3NT/1). Since Gal-3 application led to reduced level of staining of principal and apical cells relative to the grade of intensity reached by the Gal-3-like Gal-1 variant (Table II, Figure 6H, M, R, W), the data of the galectin histochemical analysis were in line with the array-based results. In both cases, the nature of the CRD mattered in the two proteins of identical (chimera-type) design.

In order to answer the question on the possibility for observing differences by pairwise analyzing staining profiles of labeled wild-type Gal-1 versus di- to tetramer and Gal-3NT/1 versus wild-type Gal-3 in cell types of other functionalities than reproduction, galectin histochemistry was performed in sections of adult murine jejunum. In principle, this appears to be the case (Supplementary data, Table SX). Strong intensity of staining of apical and supranuclear cytoplasm of surface enterocytes was obtained by labeled Gal-1 and its Gal-3NT/1 variant, whereas covalently linked di- and tetramers (exemplarily shown for the GG-linked dimer) and also wild-type Gal-3 reached comparatively lower intensity levels (Figure 7A, in clockwise representation; Supplementary data, Figure S3B–G for overviews and Figure S3I–L for enlarged details). The same congruence was seen for the neck of the epithelial lining, goblet cell staining revealing disparity among di- and tetramers, whereas the fundus, here Paneth cells, presents similarly strong intensity (Figure 7B, in clockwise representation; Supplementary data, Figure S3B–G for overview and Figure S3N–Q for enlarged details). Again, the profile of staining by the Gal-3NT/1 variant could be distinguished from that of Gal-3 (Supplementary data, Figure S3G, H for overviews, Figure S3L, M, Q, R for enlarged details).

In addition to ascertaining binding by the canonical site of contact for Lac, as shown in the insets of Figure 6B and Supplementary

data, Figure S3B, the inhibition assays were extended from using the cognate disaccharide to include synthetic neoglycoconjugates. They have the added value to serve as molecular rulers. The possibility for differential degrees of susceptibility to the presence of certain types of topological inhibitor presentation was examined with a pair of bi- and tetravalent compounds (Scheme 1). Bivalency of the Lac-presenting compound 1 is based on conjugation of sugar to a backbone with a stilbene residue (the synthetic route to its production shown in Scheme 2), whereas scaffold with a tetraphenylethylene residue established tetravalency for compound 2. The types of central bridging lead to a distance of up to 33 Å for the two sugar units in the extended conformation of compound 1, 18 Å, 28.5 Å and 32 Å separate the sugar headgroup at neighboring and at opposing positions in compound 2, as shown in Supplementary data, Figure S4 (all distances measured between the carbon atoms at the anomeric position of galactose are reported as averages observed during 10-ns molecular dynamics runs). Crystal structures of substances point to the possibility either for a coplanar arrangement between phenyl rings and the alkene in *bis(p-methoxy)-trans*-stilbene (Theocharis et al. 1984) or for a lack of planarity between the phenyl ring and alkene in tetraphenylethene (Li et al. 2017). Conformational flexibility in other parts of the backbone enable sugar headgroups to adopt a range of distances beyond those given in Supplementary data, Figure S4.

Testing sections of both organs systematically by titrations, the bivalent compound was clearly more effective than free Lac and Lac presented by the tetrameric scaffold. When concentrations were normalized for sugar content, reductions of staining when applying the labeled 8S-linked tetramer in a mixture with inhibitor on epididymis sections (Figure 8) and the GG-linked tetramer on jejunum sections (Supplementary data, Figure S5) back this statement. Remarkably, covalent association of the CRD caused increased degree of susceptibility to Lac-dependent inhibition of binding to sites in sections. The conjugation by insertion of the linker thus not only changes cellular aspects of binding profiles, but also affects inhibitory potency of the cognate sugar, on sections of the tested organs processed by fixation especially for the bivalent compound. When testing the inhibitory capacity comparatively on native cell (Chinese hamster ovary (CHO) glycosylation mutant Lec8) surface binding, the signals were reduced rather similarly by the two synthetic compounds (not shown). Examining the natural chimera-type protein, valency of the scaffold affected Gal-3-dependent signal intensity more potently

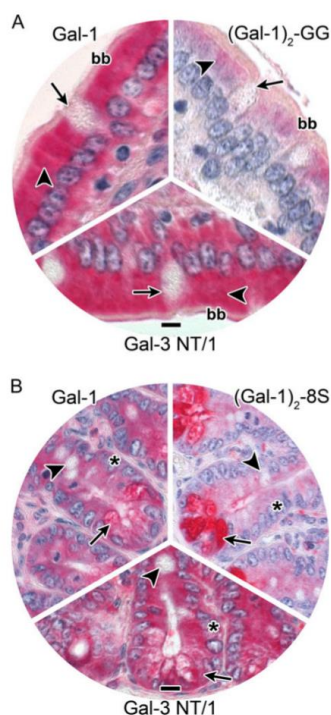


Fig. 7. Enlarged views of discriminatory aspects of the staining patterns of Gal-1, (Gal-1)₂-GG and Gal-3NT/1 in villi enterocytes (**A**) and of Gal-1, (Gal-1)₂-8S and Gal-3NT/1 in the fundus of epithelial lining of crypts (**B**) in sections of fixed murine jejunum. Microphotographs are presented in a clockwise manner. (**A**) Strong cytoplasmatic staining of surface enterocytes by Gal-1 and moderate supranuclear (arrowhead) staining by (Gal-1)₂-GG. The staining profile by Gal-3NT/1 was nearly identical to that of Gal-1. The brush border (bb) was moderately positive with Gal-1 and Gal-3NT/1, weakly with (Gal-1)₂-GG. Contents of goblet cells (arrows) was invariably negative. (**B**) Strong positivity cytoplasmatically in precursors of enterocytes (asterisks), of goblet cells (arrowheads, contents was negative) and of crypt-associated cells (arrows) such as enteroendocrine cells and Paneth cells was obtained with Gal-1, very strong with (Gal-1)₂-8S in crypt-associated cells only. Profile of Gal-3NT/1 staining was similar to that of Gal-1. The concentration of probe was constant at 0.25 µg/mL. Scale bars are 5 µm (**A**) and 10 µm (**B**).

than it did for Gal-1-type proteins, in jejunum up to a 40-fold increase in inhibitory potency for both glycoconjugates (not shown). Again, there is a difference between the two proteins with chimera-type design depending on the nature of the CRD, Gal-3 invariably being more sensitive to the presence of the two synthetic substances than Gal-3NT/1. The nature of origin of the CRD thus appears to matter for more aspects than glycan specificity.

Discussion

The basic structural unit of a lectin is defined by the folding around its contact site for the cognate ligand. The cooperation between

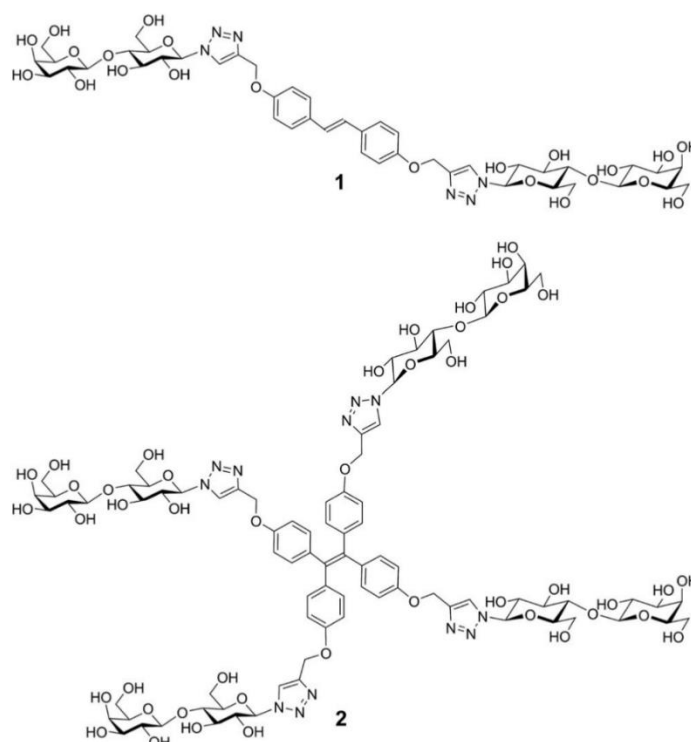
glycan specificity and modular lectin design is assumed to ensure the apparently high selectivity for distinct counterreceptors (cellular glycoconjugates): this functional pairing underlies the accurate and efficient translation of glycan-encoded messages (Gabijs et al. 2016). Altering protein architecture can thus be an approach to delineate an influence of defined architecture variations on lectin activities and to open perspectives for examining translational biomedical applications of such variants of human lectins.

In this study, we used the CRD of human Gal-1 as a common platform to build a set of five variants shown in Figure 2. Two different lengths of linker were deliberately included, because anomalous scaling of diffusion coefficients in measurements on tandem-repeat-type galectins and a linkerless variant of Gal-4 by fluorescence correlation spectroscopy had taught the lesson on “counter-intuitive consequences when simply considering molecular mass increase” (Göhler et al. 2010). Since Gal-3’s trimodular design is unique among galectins (see Figure 1), its CRD also being involved in homotypic interactions (Kuklinski and Probstmeier 1998; Yang et al. 1998; Lepur et al. 2012; Halimi et al. 2014; Ippel et al. 2016; Flores-Ibarra et al. 2018; Xiao et al. 2018), examining the Gal-3-like Gal-1 variant is supposed to probe into consequences of CRD substitution. At the same time, the Gal-1 CRD will attain monovalency, with potential for aggregation different from that of the wild-type protein.

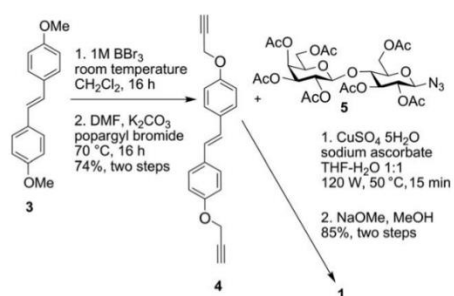
Methodologically, we teamed up three types of assay to characterize binding properties. The first assay, i.e., ITC, informed us about affinity for Lac/LacNAc with enthalpic/entropic contributions and about the occurrence of cooperativity. Overall, the driving force for Lac/LacNAc invariably was the enthalpy gain and affinity values were rather similar. Covalent CRD association did not lead to cooperativity. No indication for cooperativity had been observed for LacNAc binding to homodimeric Gal-1, -2 and -7 and tandem-repeat-type Gal-4 (Dam et al. 2005), also seen with thiodigalactoside as ligand for Gal-4 (Martín-Santamaría et al. 2011). Absence of deviation from linearity in Scatchard plots of binding of radiolabeled wild-type Gal-1 and the homo-oligomers to human (SK-N-MC) neuroblastoma cells adds evidence to a one-set-of-sites binding process (Kopitz et al. 1998; 2017). In contrast, marked negative cooperativity had been reported based on measurements from the perspective of loading of a multi(nona)valent ligand, i.e., the branch ends of the three complex-type *N*-glycans of asialofetuin, with galectin (Dam et al. 2005).

Proceeding from measurements with disaccharides in solution to monitoring specificity on a glycan array, rather similar profiles of signals were detected for the homo-oligomers. This result is in line with respective results obtained with the IgG₁ F_c part–Gal-1 CRD fusion proteins (Tsai et al. 2008; Cedeno-Laurent et al. 2010). It is thus not surprising that an 8S-linked Gal-1 homodimer did not acquire a new biological activity, i.e., to become an eosinophil chemoattractant as tandem-repeat-type Gal-9 is (Sato et al. 2002). Linker insertion therefore does not necessarily reprogram a galectin, although experience with a hexa(Gly)peptide in the place of the natural linker in Gal-8, leading to the statement that “it depends on the hinge” (Levy et al. 2006), advises to be cautious when considering extrapolation. The same applies to CRD conjugation with the NT of Gal-3. In comparison to wild-type Gal-3, the Gal-3NT/1 variant maintains the typical target specificity of Gal-1 to the terminal (non-reducing-end) position of LacNAc oligomers, whereas Gal-3 homes in on internal disaccharide units in polyLacNAc chains.

In order to take analysis from the glycan array a step closer to cellular relevance and to examine the physiological glycan



Scheme 1. Structural illustrations of the stilbene-based bivalent compound **1** and the tetraphenylethylene-based tetravalent compound **2**.



Scheme 2. Route of synthesis of the bivalent compound **1**.

complexity, we have added galectin histochemistry of tissue sections. They present complex glycomes and enable comparative monitoring on the level of various cell types. Testing revealed marked differences in signal intensities. Also, sensitivity toward inhibition with cognate sugar, most effective with a bivalent glycoconjugate, was

enhanced by linker insertion. The example of intramolecular bridging of contact sites in wheat germ agglutinin by a bivalent compound has taught the remarkable lesson that the synthetic scaffolds with conjugated Lac can be molecular rulers (Maierhofer et al. 2007; Schwefel et al. 2010; André et al. 2016). Indeed, spatial features of interplay between the Gal-1 CRDs in the wild-type protein and in the linker-connected variants and contact sites in sections appear to be non-uniform; the same applied to wild-type Gal-3 and Gal-3-like Gal-3NT/1, here likely also involving the individual characteristics of the interplay of CRD (Gal-1 or -3) with itself in intermolecular contacts and/or with the NT. Concerning binding of ligands in sections, similar staining profiles for the wild-type and Gal-3-like Gal-1 proteins were determined, as opposed to wild-type Gal-3. The cases of difference in staining profiles of epididymal principal/apical versus basal cells as well as subapical versus supranuclear cytoplasmatic positivity of surface enterocytes when working with wild-type Gal-1 and its covalently linked variants yet argue against simple extrapolations. Evidently, covalent CRD conjugation can modulate certain aspects of staining profiles in different manners.

In this sense, our report gives direction to broaden the study of variants and to define their properties in ligand recognition up to

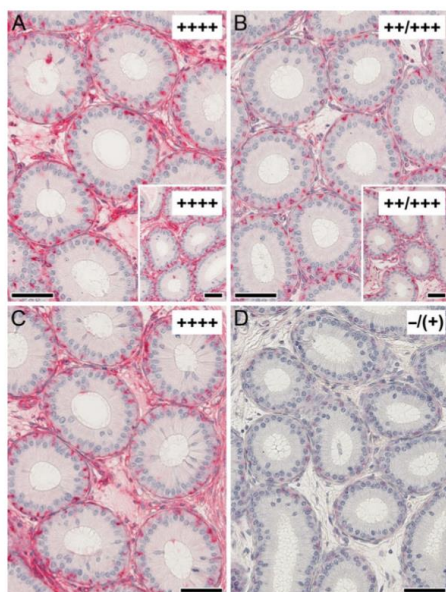


Fig. 8. Effect of increasing concentrations of cognate sugar (Lac) added free in solution or as part of the two glycoconjugates on the staining profile obtained with biotinylated (Gal-1)₄-8S in cross sections through the initial segment of fixed murine epididymis. The signal remained at near to 100% level in the presence of 0.025 mM Lac tested as free sugar (A) or presented either by tetravalent compound **2** (inset to A) or bivalent compound **1** (C). Increasing the sugar concentration to 0.5 mM free Lac (B; please also see complete inhibition by 10 mM Lac in Fig. 6 inset to F) and scaffold-presented Lac by tetravalent compound **2** (inset to B) reduced staining intensity by 1–2 categories in the semiquantitative ranking (from ++++ to ++/+++) and number of positive basal cells by approximately 20%. In stark contrast, presence of compound **1** at this concentration precluded any binding of galectin (D). Symbols for semiquantitative grading of staining intensity are given in the rectangular box in the top-right area of each microphotograph and inset (for correlation of symbols to staining intensity, please see footnote in Table II). Scale bars are 50 μ m. (Gal-1)₄-8S was applied at a concentration of 0.5 μ g/mL.

binding to cell surfaces and eliciting post-binding effects, starting with analyzing the spatial nature of galectin-induced lattices. After all, cross-linking of counterreceptors is assumed to be at the heart of galectin functionality (Brewer 2002; Boscher et al. 2011; Kasai 2018; Sato 2018). Considering antagonist activity of Gal-3 (and also the Gal-3NT1 variant) on Gal-1 as negative neuroblastoma growth regulator (Kopitz et al. 2001; Ludwig, Michalak, et al. 2019) and formation of cross-linked complexes with disparate structural organization together with Lac-bearing glycoclusters by these two proteins (Ahmad et al. 2004), lectin design apparently is a source for variability of functional aspects. Further comparative analyses with wild-type and variant proteins, then also with a Gal-1-like Gal-3 homodimer (Ludwig, Michalak, et al. 2019), can therefore provide hints to understand emerging aspects of expression of galectins as network (Zivicová et al. 2017; Manning, García Caballero, Ruiz, et al. 2017; Manning, García Caballero, Knosp,

2018) and of their emerging, likely clinically relevant networking (Weinmann et al. 2018). Such studies have potential to establish an innovative class of specific antagonist/effector proteins on the platform of human galectin domains with a perspective for testing clinical applicability.

Materials and methods

Protein production and labeling

The wild-type proteins and five variants were obtained by recombinant production, purified to homogeneity by affinity chromatography on home-made Lac-presenting resin and biotinylated under activity-preserving conditions using the *N*-hydroxysuccinimide ester derivative of biotin (Sigma, Munich, Germany), followed by routinely ascertaining maintained lectin activity, as described (Gabius et al. 1991; Kopitz et al. 2017; Ludwig, Michalak, et al. 2019). In addition, the GG-linked homotrimer was produced, purified and characterized as described (Kopitz et al. 2017).

ITC measurements

Titrations were performed in 20 mM phosphate buffer (pH 7.2) containing 5 mM or 150 mM NaCl and 10 mM β -mercaptoethanol at constant temperature for the six proteins using a PEAQ-ITC (Malvern, Westborough, MA, USA) calorimeter, as described (Ludwig, Michalak, et al. 2019). In brief, adding ligand stepwisely in 150 s intervals at 25°C and 750 rpm in 2 μ L aliquots of a 36.4- μ L solution of 6 mM LacNAc/Lac (and 10 mM or 14 mM) to 200 μ L galectin-containing solution (details on protein concentrations listed in Table I and Supplementary data, Table SI) in the calorimetric cell yielded measurements of heat release, respective data were processed by the MicroCal PEAQ-ITC Analysis software using a one-site model, and a fitted off-set parameter was applied to each titration to account for background. In the cases of Lac titrations, fixing the *n*-value at or near theoretical stoichiometry was applied, using lectin concentration of LacNAc titrations for parameter settings. Hill plot data analysis was performed, as applied for human galectins and a nonavalent ligand (asialofetuin) as described (Dam et al. 2005).

Array measurements

Biotinylated galectins were comparatively tested at the constant concentration of 50 μ g/mL in phosphate buffered saline (PBS) containing 0.1% Tween-20, 1% bovine serum albumin and 0.01% Na₂S₂O₃ for 1 h at 37°C in a humidified chamber with the array slide presenting the panel of 647 (glyco)compounds, as described for chicken galectins (García Caballero et al. 2016). Slide surface had been pre-treated with PBS containing 0.1% Tween-20 for 15 min to reduce background by non-specific protein adsorption. After thorough washing to remove unbound labeled protein, probing with fluorescent streptavidin (labeled with AlexaFluor[®]-555 dye; Molecular Probe, Eugene, USA) followed for 45 min at 20°C. After thoroughly washing with PBS-0.001% Tween-20 and then with deionized water to remove the fluorescent sensor protein, slides were scanned on an InnoScan 1100 AL scanner (Innopsys, Carbonne, France) using an excitation wavelength of 543 nm at 10 μ m resolution. The obtained data were processed using ScanArray Express 4.0 software and the fixed 70- μ m diameter circle method as well as Microsoft Excel. Six spots represented each compound on the array (details on the nature of the 647 compounds given in Supplementary data, Tables SIV–IX), and binding data are reported as median relative fluorescence units of replicates. Median deviation was measured as interquartile

range. A signal, whose fluorescence intensity exceeded the background value by a factor of five, was considered to be significant.

Galectin histochemistry

Fresh tissue specimen of four 6-week-old C57 BL/6 mice were fixed in Bouin's solution for 24 h, dehydrated by passing them through a series of solutions of increasing contents of ethanol, then isopropanol and finally xylene prior to embedding in paraffin wax at 61°C. Sections mounted on Superfrost[®] plus glass slides (Menzel, Braunschweig, Germany) were processed by an optimized protocol ensuring minimal background with Vectastain[®] ABC Kit and Vector[®] Red reagents (Biozol, Eching, Germany) for staining, as described (Kaltner, Manning, et al. 2018). In the case of each protein, systematic titrations including blocking by cognate sugar were carried out to compare profiles and identify a concentration for comparative analysis in three to five independent series that covered the following ranges: 0.0625–4 µg/mL for wild-type Gal-1 and the pair of homodimeric variants, up to 1 µg/mL for the pair of homotrimeric variants, up to 0.5 µg/mL for the Gal-3NT/1 protein and 16 µg/mL for wild-type Gal-3. Of note, titrations covered equal mass and molar concentration for the wild-type versus variant pairs. Data acquisition and recording followed a routine protocol (Kaltner, Manning, et al. 2018), and semiquantitative grading of intensity of staining is given in the footnote of Table II.

Glycocompound synthesis

NMR spectra were recorded with 500 MHz and 600 MHz Varian spectrometers. Chemical shifts are reported relative to internal Me₄Si in CDCl₃ (δ 0.0), HOD for D₂O (δ 4.84) or CD₂HOD (δ 3.31) for ¹H and CDCl₃ (77.16) or CD₃OD (49.05) for ¹³C. NMR spectra were processed and analysed using MestReNova software. ¹H NMR signals were assigned with the aid of gCOSY. ¹³C NMR signals were assigned with the aid of APT, gHSQCAD and/or gHMBCAD. Coupling constants are reported in Hertz. Low- and high-resolution mass spectra were measured on a Waters LCT Premier XE Spectrometer, measuring in both positive and/or negative mode as, using MeCN, H₂O and/or MeOH as solvent. Thin layer chromatography (TLC) was performed on aluminium sheets precoated with silica gel 60 (HF254, E. Merck) and spots visualized by UV and charring with H₂SO₄-EtOH (1:20), cerium molybdate, or phosphomolybdic acid stains. Flash chromatography was carried out with silica gel 60 (0.040–0.630 mm; E. Merck or Aldrich) and using a stepwise solvent polarity gradient (starting with the conditions indicated in each case and increasing the polarity as required) correlated with TLC mobility. Chromatography solvents, cyclohexane, EtOAc, CH₂Cl₂ and MeOH were used as obtained from suppliers (Fisher Scientific and Sigma-Aldrich). Solvents for reactions under anhydrous conditions were directly used as obtained from a Pure Solv[™] Solvent Purification System.

To obtain (E)-1,2-bis(4-(prop-2-yn-1-yloxy)phenyl)ethane (4), compound 3 was prepared (Andrus et al. 2002). To 3 (2.50 g, 10.4 mmol) dissolved in CH₂Cl₂ (40 mL), which had been cooled in an ice-salt bath, was added, dropwise, a solution of boron tribromide in CH₂Cl₂ (39 mL of 1.0 M, 39 mmol). The solution was allowed to attain room temperature and was then stirred for 15 h. Water (35 mL) was added slowly, dropwise, with stirring. The organic solvent was removed under reduced pressure, the aqueous phase was extracted with EtOAc (3 × 30 mL), the combined organic phases were dried over Na₂SO₄ and the solvent removed under reduced pressure. Column chromatography (7:3, cyclohexane-

EtOAc) gave the demethylated intermediate (2.1 g, 97%); ¹H NMR (500 MHz, DMSO-d₆) δ 7.33 (d, *J* = 8.7 Hz, 4H, aromatic H), 6.88 (s, 2H, alkene H), 6.72 (d, *J* = 8.6 Hz, 4H, aromatic H); ¹³C NMR (126 MHz, DMSO-d₆) δ = 157.2 (C), 129.0 (C), 127.8 (CH aromatic), 125.6 (CH, alkene), 115.9 (CH aromatic); ESI-HRMS calcd for C₁₄H₁₁O₂ 211.0759, found *m/z* 211.0702 [M-H]⁻. To this intermediate (2.1 g, 9.9 mmol) dissolved in dry DMF (300 mL) was added anhydrous potassium carbonate (11.0 g, 79.3 mmol). Propargyl bromide (80% in toluene, 4.8 mL, 55 mmol) was then added and the mixture was stirred for 14 h at 70°C under a nitrogen atmosphere. The reaction was then cooled to room temperature and was diluted with CH₂Cl₂ (300 mL). This solution was washed with saturated NH₄Cl, with the aqueous layer being re-extracted with a further portion of CH₂Cl₂ (3 × 150 mL). The combined organic layers were then washed with water and dried over Na₂SO₄, solvent was removed at reduced pressure. Column chromatography (7:3, cyclohexane-EtOAc) gave the title compound (2.43 g, 85%); ¹H NMR (500 MHz, CDCl₃) δ 7.44 (d, *J* = 8.7 Hz, 4H, aromatic H), 6.97 (d, *J* = 8.7 Hz, 4H, aromatic H), 6.94 (s, 2H, alkene H), 4.71 (s, 4H, CH₂), 2.53 (s, 2H, alkyne H); ¹³C NMR (126 MHz, chloroform-d) δ = 157.4 (C), 131.3 (C), 127.4 (aromatic CH), 126.5 (alkene CH), 115.1 (aromatic CH), 75.6 (alkyne CH), 55.9 (CH₂); ES-HRMS calcd for C₂₀H₁₆NaO₂ 311.1048, found *m/z* 311.1052 [M + Na]⁺.

To obtain glycocompound 1, compound 5 was prepared as described (Leyden et al. 2009). To 4 (108 mg, 0.42 mmol) dissolved in degassed THF-H₂O (1:1, 6 mL) were added 5 (561 mg, 0.92 mmol), sodium ascorbate (49 mg, 0.25 mmol) and Cu₂SO₄·5H₂O (62 mg, 0.25 mmol). The reaction mixture was stirred under inert atmosphere in a microwave reactor at 50°C (120 W) for 30 min. Tetrahydrofuran was then removed under reduced pressure followed by the dilution of the solution with CH₂Cl₂. This mixture was washed with water. The aqueous layer was re-extracted with a further portion of CH₂Cl₂. The combined organic layers were dried over Na₂SO₄ and the solvent was removed at reduced pressure. Chromatography (95:5, CH₂Cl₂-MeOH) gave the protected intermediate (600 mg, 89%) as a colorless solid; ¹H NMR (500 MHz, chloroform-d) δ 7.78 (s, 2H, triazole H), 7.40 (d, *J* = 8.8 Hz, 4H, aromatic H), 6.94 (d, *J* = 8.8 Hz, 4H, aromatic H), 6.91 (s, 2H, alkene H), 5.83 (d, *J* = 9.1 Hz, 2H, H-1), 5.39 (overlapping signals, 4H), 5.35 (d, *J* = 3.4 Hz, 2H), 5.20 (s, 4H, CH₂), 5.12 (dd, *J* = 10.4, 7.9 Hz, 2H, H-2'), 4.96 (dd, 10.4, 3.4 Hz, 2H, H-3'), 4.52 (d, *J* = 7.9 Hz, 2H, H-1'), 4.47 (d, *J* = 11.0 Hz, 2H, H-6), 4.18–4.06 (overlapping signals, 6H, H-6 and H-6' protons), 3.98–3.83 (overlapping signals, 6H, H-4, H-5, H-5'), 2.15 (s, 6H, OAc), 2.09 (s, 6H, OAc), 2.06 (s, 6H, OAc), 2.05 (s, 6H, OAc), 2.04 (s, 6H, OAc), 1.96 (s, 6H, OAc), 1.82 (s, 6H, OAc); ¹³C NMR (126 MHz, chloroform-d) δ = 170.3, 170.2, 170.1, 170.0, 169.4, 169.1, 169.0 (each C, each OAc), 157.5 (aromatic C), 144.8 (triazole C), 131.0 (aromatic C), 127.5 (aromatic CH), 126.3 (alkene CH), 121.1 (triazole CH), 115.0 (aromatic CH), 101.1 (CH, C-1'), 85.5 (CH, C-1), 75.9 (CH, C-4), 75.6 (CH, C-5), 72.6 (CH, C-3), 70.9 (CH, C-3'), 70.8 (CH, C-2), 70.5 (CH, C-2'), 69.0, 66.6 (CH, C-4'), 61.9 (CH₂), 61.7 (CH₂, C-6), 60.8 (CH₂, C-6'), 20.8 (CH₃, OAc), 20.7 (CH₃, OAc), 20.6 (CH₃, OAc), 20.6 (CH₃, OAc), 20.6 (CH₃, OAc), 20.5 (CH₃, OAc), 20.1 (CH₃, OAc), ES-HRMS calcd for C₂₂H₁₆N₆NaO₃₆ 1633.4981, found *m/z* 1633.4985 [M + Na]⁺. To a suspension of this intermediate (400 mg, 0.25 mmol) in dry MeOH (10 mL, cooled 0°C), freshly prepared 1 M NaOMe in MeOH was added until the solution reached pH 10. The reaction mixture was allowed to attain room temperature and stirred for 15 h. Glacial acetic acid was added

to neutralize (pH = 7) the solution [NOTE: it is possible to use Amberlite IR-120H+ instead of AcOH, but filtration is difficult due to the low solubility of 1]. The solvent was removed under reduced pressure. Reverse phase column chromatography (three volumes of water were flushed through the column, to ensure salt removal, followed by a MeCN–H₂O mixture of 3:2 ratio) gave the title compound 1 as white solid (250 mg, 96%) [NOTE: the compound dissolves in a minimal volume of water with few drops of MeCN added. If the product precipitates, one drop of AcOH can be added and the mixture heated slightly to redissolve solid compound]; ¹H NMR (500 MHz, DMSO-d₆) δ 8.46 (s, 2 H, triazole H), 7.52 (d, J = 8.8 Hz, 4 H, aromatic H), 7.11–7.00 (overlapping signals, 6 H, aromatic H, alkene H), 5.67 (d, J = 9.3 Hz, 2 H, H-1), 5.17 (s, 4 H, CH₂), 4.26 (d, J = 7.9 Hz, 2 H, H-1'), 3.87 (t, J = 9.3 Hz, 2 H, H-2), 3.78 (d, J = 10.6 Hz, 2 H, H-6), 3.70–3.44 (overlapping signals, 16 H), 3.42–3.29 (overlapping signals, 4 H); ¹³C NMR (126 MHz, DMSO-d₆) δ 157.9 (C), 143.1 (C), 130.8 (C), 127.9 (CH, aromatic), 126.4 (CH, aromatic), 124.4 (CH, triazole), 115.3 (CH, aromatic), 104.2 (CH, C-1'), 87.4 (CH, C-1), 80.2 (CH), 78.2 (CH), 76.0 (CH), 75.6 (CH), 73.7 (CH), 72.2 (CH), 71.0 (CH), 68.6 (CH), 61.4 (CH₂), 60.9 (CH₂), 60.5 (CH₂); ES-HRMS calcd for C₄₄H₅₇N₆O₂₂ 1021.3526, found m/z 1021.3566 [M-H]⁻.

The tetraphenylene-based glycoconjugate 2 was prepared, as described and applied to lectin testing previously (Hu et al. 2011; André et al. 2015). Analytical data: ¹H NMR (500 MHz, DMSO-d₆) δ 8.43 (s, 4 H, triazole H), 6.90 (d, J = 8.5 Hz, 8 H, aromatic H), 6.84 (d, J = 8.6 Hz, 8 H, aromatic H), 5.66 (d, J = 9.3 Hz, 4 H, H-1), 5.65–5.61 (overlapping signals, 4 H, OH), 5.20 (br s, 4 H, OH), 5.07 (s, 8 H, CH₂), 4.95 (br s, 4 H, OH), 4.92 (br s, 2 H, OH), 4.75–4.65 (overlapping signals, 8 H, OH), 4.59 (br s, 4 H, OH), 4.26 (d, J = 7.1 Hz, 4 H, H-1'), 3.87 (q, J = 8.9 Hz, 4 H, H-2), 3.78 (t, J = 10.8, 4.9 Hz, 4 H), 3.70–3.45 (overlapping signals, 34 H), 3.40–3.33 (overlapping signals, 8 H); ¹³C NMR (126 MHz, DMSO-d₆) δ 171.1 (C), 156.9 (C), 143.0 (C), 137.0 (C), 132.5 (CH, aromatic), 124.4 (CH, triazole), 114.3 (CH, aromatic), 104.2 (CH, C-1'), 87.4 (CH, C-1), 80.2 (CH), 78.2 (CH), 76.1 (CH), 75.6 (CH), 73.7 (CH), 72.2 (CH, C-2), 71.0 (CH), 68.6 (CH₂), 60.9 (CH₂), 60.5 (CH₂); ES-HRMS calcd for C₈₆H₁₁₁N₁₂O₄₄ 2015.6817, found m/z 2015.6818 [M-H]⁻.

Low-energy conformers of both glycoconjugates were generated by molecular modeling applying structure building using Maestro version 6.0 (Schrodinger Inc., LLC, New York, USA) and energy minimization (OPLS-AA force field, GB/SA continuum solvation model for water (Still et al. 1990)) using MacroModel version 6.0.107 (Schrodinger Inc.), in part as described (Wang et al. 2012). Distances reported are averages obtained from 1000 structures sampled during 10-ns molecular dynamics simulations. The stochastic dynamics method was employed at a simulation temperature of 300 K, a time step of 1.5 fs and an equilibration time of 1.0 ps. No constraints were imposed during the simulation in the case of compound 2. In the case of substance 1, the phenyl rings and alkene were constrained to match a coplanar arrangement.

Inhibitory capacity of glycoconjugates

Systematic titrations with solutions containing free or glycoconjugate-presented Lac (all concentrations normalized to Lac) were performed in 2-fold serial dilutions in the range of 0.025–200 mM (Lac) or to 10 mM Lac (glycoconjugate). Solutions with biotinylated galectin and inhibitor were mixed and incubated for 1 h at room temperature, and pretreated sections were exposed

to aliquots of the mixture in parallel overnight at 4°C together with mock-treated controls (= 100%). All concentrations used for experiments of documented data are given in the legends of respective figures. Cytofluorometric analysis with fluorescent galectins and cells of the CHO glycosylation mutant Lec8 were carried out as described (André et al. 2016; Kopitz et al. 2017).

Supplementary data

Supplementary data is available at *GLYCOBIOLOGY* online.

Acknowledgements

We gratefully acknowledge the excellent input provided by the reviewers, generous support by the Science Foundation Ireland [16/IA/4419 to P.M.V.] (co-funded by the European Regional Development Fund) and the Irish Research Council [GOIPG/2016/858 to P.M.V.] and the National Institutes of Health [R21 CA178754 to M.C.] as well as inspiring discussions with Drs B. Friday, A. Leddoz and A.W.L. Nose.

Conflict of interest statement

None declared.

References

- Ahmad N, Gabius H-J, André S, Kaltner H, Sabesan S, Roy R, Liu B, Macaluso F, Brewer CF. 2004. Galectin-3 precipitates as a pentamer with synthetic multivalent carbohydrates and forms heterogeneous cross-linked complexes. *J Biol Chem*. 279:10841–10847.
- Ahmad N, Gabius H-J, Kaltner H, André S, Kuwabara I, Liu F-T, Oscarson S, Norberg T, Brewer CF. 2002. Thermodynamic binding studies of cell surface carbohydrate epitopes to galectins -1, -3 and -7. Evidence for differential binding specificities. *Can J Chem*. 80:1096–1104.
- Andrus MB, Mendenhall KG, Meredith EL, Soma Sekhar BBV. 2002. Glycolate aldol reactions with boron enolates of bis-4-methoxyphenyl dioxanone. *Tetrahedron Lett*. 43:1789–1792.
- André S, Kaltner H, Kayser K, Murphy PV, Gabius H-J. 2016. Merging carbohydrate chemistry with lectin histochemistry to study inhibition of lectin binding by glycoclusters in the natural tissue context. *Histochem Cell Biol*. 145:185–199.
- André S, O'Sullivan S, Koller C, Murphy PV, Gabius H-J. 2015. Bi- to tetra-valent glycoclusters presenting GlcNAc/GalNAc as inhibitors: from plant agglutinins to human macrophage galactose-type lectin (CD301) and galectins. *Org Biomol Chem*. 13:4190–4203.
- Bi S, Earl LA, Jacobs L, Baum LG. 2008. Structural features of galectin-9 and galectin-1 that determine distinct T cell death pathways. *J Biol Chem*. 283:12248–12258.
- Boscher C, Dennis JW, Nabi IR. 2011. Glycosylation, galectins and cellular signaling. *Curr Opin Cell Biol*. 23:383–392.
- Brewer CF. 2002. Binding and cross-linking properties of galectins. *Biochim Biophys Acta*. 1572:255–262.
- Brockhausen I, Schachter H. 1997. Glycosyltransferases involved in N- and O-glycan biosynthesis. In: Gabius H-J, Gabius S, editors. *Glycosciences: Status and Perspectives*. London-Weinheim: Chapman & Hall. p. 79–113.
- Buddecke E. 2009. Proteoglycans. In: Gabius H-J, editor. *The Sugar Code. Fundamentals of glycosciences*. Weinheim, Germany: Wiley-VCH. p. 199–216.
- Bätzig P, Saudan P, Gunde T, Bachmann MF. 2004. Enhanced apoptotic activity of a structurally optimized form of galectin-1. *Mol Immunol*. 41:9–18.
- Cedeno-Laurent F, Barthel SR, Opperman MJ, Lee DM, Clark RA, Dimitroff CJ. 2010. Development of a nascent galectin-1 chimeric molecule for studying the role of leukocyte galectin-1 ligands and immune disease modulation. *J Immunol*. 185:4659–4672.

- Cooper DNW. 2002. Galectinomics: finding themes in complexity. *Biochim Biophys Acta*. 1572:209–231.
- Corfield AP. 2017. Eukaryotic protein glycosylation: a primer for histochemists and cell biologists. *Histochem Cell Biol*. 147:119–147.
- Cummings RD. 2009. The repertoire of glycan determinants in the human glycome. *Mol Biosyst*. 5:1087–1104.
- Dam TK, Gabius H-J, André S, Kaltner H, Lensch M, Brewer CF. 2005. Galectins bind to the multivalent glycoprotein asialofetuin with enhanced affinities and a gradient of decreasing binding constants. *Biochemistry*. 44:12564–12571.
- Dettmann W, Grandbois M, André S, Benoit M, Wehle AK, Kaltner H, Gabius H-J, Gaub HE. 2000. Differences in zero-force and force-driven kinetics of ligand dissociation from β -galactoside-specific proteins (plant and animal lectins, immunoglobulin G) monitored by plasmon resonance and dynamic single molecule force microscopy. *Arch Biochem Biophys*. 383:157–170.
- Earl LA, Bi S, Baum LG. 2011. Galectin multimerization and lattice formation are regulated by linker region structure. *Glycobiology*. 21:6–12.
- Feng C, Ghosh A, Amin MN, Bachvaroff TR, Tasumi S, Pasek M, Banerjee A, Shridhar S, Wang LX, Bianchet MA et al. 2015. Galectin CvGal2 from the Eastern oyster (*Crassostrea virginica*) displays unique specificity for ABH blood group oligosaccharides and differentially recognizes sympatric *Perkinsus* species. *Biochemistry*. 54:4711–4730.
- Feng C, Ghosh A, Amin MN, Giomarelli B, Shridhar S, Banerjee A, Fernandez-Robledo JA, Bianchet MA, Wang LX, Wilson IB et al. 2013. The galectin CvGal1 from the eastern oyster (*Crassostrea virginica*) binds to blood group A oligosaccharides on the hemocyte surface. *J Biol Chem*. 288:24394–24409.
- Fettis MM, Hudalla GA. 2018. Engineering reactive oxygen species-resistant galectin-1 dimers with enhanced lectin activity. *Bioconjug Chem*. 29:2489–2496.
- Flores-Ibarra A, Vértesy S, Medrano FJ, Gabius H-J, Romero A. 2018. Crystallization of a human galectin-3 variant with two ordered segments in the shortened N-terminal tail. *Sci Rep*. 8:9835.
- Gabius H-J. 2017. How to crack the sugar code. *Folia Biol (Praba)*. 63:121–131.
- Gabius H-J, Manning JC, Kopitz J, André S, Kaltner H. 2016. Sweet complementarity: the functional pairing of glycans with lectins. *Cell Mol Life Sci*. 73:1989–2016.
- Gabius H-J, Roth J. 2017. An introduction to the sugar code. *Histochem Cell Biol*. 147:111–117.
- Gabius H-J, Wosgien B, Hendrys M, Bardosi A. 1991. Lectin localization in human nerve by biochemically defined lectin-binding glycoproteins, neoglycoprotein and lectin-specific antibody. *Histochemistry*. 95:269–277.
- García Caballero G, Kaltner H, Michalak M, Shilova N, Yegres M, André S, Ludwig A-K, Manning JC, Schmidt S, Schnölzer M et al. 2016. Chicken GRIFIN: a homodimeric member of the galectin network with canonical properties and a unique expression profile. *Biochimie*. 128–129:34–47.
- Ginsburg V, Neufeld EF. 1969. Complex heterosaccharides of animals. *Annu Rev Biochem*. 38:371–388.
- Giudicelli V, Lutomski D, Levi-Strauss M, Bladier D, Joubert-Caron R, Caron M. 1997. Is human galectin-1 activity modulated by monomer/dimer equilibrium? *Glycobiology*. 7:viii–x.
- Göhler A, André S, Kaltner H, Sauer M, Gabius H-J, Doose S. 2010. Hydrodynamic properties of human adhesion/growth-regulatory galectins studied by fluorescence correlation spectroscopy. *Biophys J*. 98:3044–3053.
- Halimi H, Rigato A, Byrne D, Ferracci G, Sebban-Kreuzer C, ElAntak L, Guerlesquin F. 2014. Glycan dependence of galectin-3 self-association properties. *PLoS One*. 9:e111836.
- Hirabayashi J, editor. 1997. Recent topics on galectins. *Trends Glycosci Glycotechnol*. 9:1–180.
- Hirabayashi J, Hashidate T, Arata Y, Nishi N, Nakamura T, Hirashima M, Urashima T, Oka T, Futai M, Müller WEG et al. 2002. Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. *Biochim Biophys Acta*. 1572:232–254.
- Hu XM, Chen Q, Wang JX, Cheng QY, Yan CG, Cao J, He YJ, Han BH. 2011. Tetraphenylethylene-based glycoconjugate as a fluorescence “turn-on” sensor for cholera toxin. *Chem Asian J*. 6:2376–2381.
- Ippel H, Miller MC, Vértesy S, Zheng Y, Canada FJ, Suylen D, Umemoto K, Romano C, Hackeng T, Tai G et al. 2016. Intra- and intermolecular interactions of human galectin-3: assessment by full-assignment-based NMR. *Glycobiology*. 26:888–903.
- Iwaki J, Hirabayashi J. 2018. Carbohydrate-binding specificity of human galectins: an overview by frontal affinity chromatography. *Trends Glycosci Glycotechnol*. 30:SE137–SE153.
- Kaltner H, García Caballero G, Ludwig A-K, Manning JC, Gabius H-J. 2018. From glycophenotyping by (plant) lectin histochemistry to defining functionality of glycans by pairing with endogenous lectins. *Histochem Cell Biol*. 149:547–568.
- Kaltner H, Manning JC, García Caballero G, Di Salvo C, Gabba A, Romero-Hernández LL, Knosp C, Wu D, Daly HC, O’Shea DF et al. 2018. Revealing biomedically relevant cell and lectin type-dependent structure-activity profiles for glycoclusters by using tissue sections as assay platform. *RSC Advances*. 8:28716–28735.
- Kaltner H, Toegel S, García Caballero G, Manning JC, Ledeen RW, Gabius H-J. 2017. Galectins: their network and roles in immunity/tumor growth control. *Histochem Cell Biol*. 147:239–256.
- Kasai K-i. 2018. Galectins: quadruple-faced proteins. *Trends Glycosci Glycotechnol*. 30:SE221–SE223.
- Kopitz J. 2017. Lipid glycosylation: a primer for histochemists and cell biologists. *Histochem Cell Biol*. 147:175–198.
- Kopitz J, von Reitzenstein C, André S, Kaltner H, Uhl J, Ehemann V, Cantz M, Gabius H-J. 2001. Negative regulation of neuroblastoma cell growth by carbohydrate-dependent surface binding of galectin-1 and functional divergence from galectin-3. *J Biol Chem*. 276:35917–35923.
- Kopitz J, von Reitzenstein C, Burchert M, Cantz M, Gabius H-J. 1998. Galectin-1 is a major receptor for ganglioside GM1, a product of the growth-controlling activity of a cell surface ganglioside sialidase, on human neuroblastoma cells in culture. *J Biol Chem*. 273:11205–11211.
- Kopitz J, Xiao Q, Ludwig A-K, Romero A, Michalak M, Sherman SE, Zhou X, Dazen C, Vértesy S, Kaltner H et al. 2017. Reaction of a programmable glycan presentation of glycodendrimersomes and cells with engineered human lectins to show the sugar functionality of the cell surface. *Angew Chem Int Ed*. 56:14677–14681.
- Kuklinski S, Probstmeier R. 1998. Homophilic binding properties of galectin-3: involvement of the carbohydrate recognition domain. *J Neurochem*. 70:814–823.
- Laine RA. 1997. The information-storing potential of the sugar code. In: Gabius H-J, Gabius S, editors. *Glycosciences: Status and Perspectives*. London-Weinheim: Chapman & Hall. p. 1–14.
- Ledeen RW, Kopitz J, Abad-Rodríguez J, Gabius H-J. 2018. Glycan chains of gangliosides: functional ligands for tissue lectins (siglecs/galectins). *Progr Mol Biol Transl Sci*. 156:289–324.
- Leppänen A, Stowell S, Blixt O, Cummings RD. 2005. Dimeric galectin-1 binds with high affinity to α 2,3-sialylated and non-sialylated terminal N-acetylglucosamine units on surface-bound extended glycans. *J Biol Chem*. 280:5549–5562.
- Lepur A, Salomonsson E, Nilsson UJ, Leffler H. 2012. Ligand induced galectin-3 protein self-association. *J Biol Chem*. 287:21751–21756.
- Levy Y, Auslender S, Eisenstein M, Vidavski RR, Ronen D, Bershadsky AD, Zick Y. 2006. It depends on the hinge: a structure-functional analysis of galectin-8, a tandem-repeat type lectin. *Glycobiology*. 16:463–476.
- Leyden R, Velasco-Torrijos T, André S, Gouin S, Gabius H-J, Murphy PV. 2009. Synthesis of bivalent lactosides based on terephthalamide, N,N'-diglycosylterephthalamide, and glycothane scaffolds and assessment of their inhibitory capacity on medically relevant lectins. *J Org Chem*. 74:9010–9026.
- Li D, Hu R, Guo D, Zang Q, Li J, Wang Y, Zheng Y-S, Tang BZ, Zhang H. 2017. Diagnostic absolute configuration determination of tetraphenylethene core-based chiral aggregation-induced emission compounds: particular fingerprint bands in comprehensive chiroptical spectroscopy. *J Phys Chem C*. 21:20947–20954.

- Lohr M, Kaltner H, Schwartz-Albiez R, Sinowatz F, Gabius H-J. 2010. Towards functional glycomics by lectin histochemistry: strategic probe selection to monitor core and branch-end substitutions and detection of cell-type and regional selectivity in adult mouse testis and epididymis. *Anat Histol Embryol*. 39:481–493.
- Ludwig A-K, Kaltner H, Kopitz J, Gabius H-J. 2019. Lectinology 4.0: altering modular (ga)lectin display for functional analysis and biomedical applications. *Biochim Biophys Acta*. 1863:935–940.
- Ludwig A-K, Michalak M, Xiao Q, Gilles U, Medrano FJ, Ma H, FitzGerald FG, Hasley WD, Melendez-Davila AM, Liu M et al. 2019. Design-functionality relationships for adhesion/growth-regulatory galectins. *Proc Natl Acad Sci USA*. 116:2837–2842.
- López-Lucendo MF, Solís D, André S, Hirabayashi J, Kasai K-i, Kaltner H, Gabius H-J, Romero A. 2004. Growth-regulatory human galectin-1: crystallographic characterisation of the structural changes induced by single-site mutations and their impact on the thermodynamics of ligand binding. *J Mol Biol*. 343:957–970.
- Maierhofer C, Rohmer K, Wittmann V. 2007. Probing multivalent carbohydrate-lectin interactions by an enzyme-linked lectin assay employing covalently immobilized carbohydrates. *Bioorg Med Chem*. 15: 7661–7676.
- Manning JC, García Caballero G, Knospe C, Kaltner H, Gabius H-J. 2017. Network analysis of adhesion/growth-regulatory galectins and their binding sites in adult chicken retina and choroid. *J Anat*. 231:23–37.
- Manning JC, García Caballero G, Knospe C, Kaltner H, Gabius H-J. 2018. Three-step monitoring of glycan and galectin profiles in the anterior segment of the adult chicken eye. *Ann Anat*. 217:66–81.
- Manning JC, García Caballero G, Ruiz FM, Romero A, Kaltner H, Gabius H-J. 2018. Members of the galectin network with deviations from the canonical sequence signature, 2. Galectin-related protein (GRP). *Trends Glycosci Glycotechnol*. 30:SE11–SE20.
- Manning JC, Romero A, Habermann FA, García Caballero G, Kaltner H, Gabius H-J. 2017. Lectins: a primer for histochemists and cell biologists. *Histochem Cell Biol*. 147:199–222.
- Martín-Santamaría S, André S, Buzamet E, Caraballo R, Fernández-Cureses G, Morando M, Ribeiro JP, Ramirez-Gualito K, de Pascual-Teresa B, Cañada FJ et al. 2011. Symmetric dithiodigalactoside: strategic combination of binding studies and detection of selectivity between a plant toxin and human lectins. *Org Biomol Chem*. 9:5445–5455.
- Roth J. 1987. Subcellular organization of glycosylation in mammalian cells. *Biochim Biophys Acta*. 906:405–436.
- Sandhoff R, Schulze H, Sandhoff K. 2018. Ganglioside metabolism in health and disease. *Progr Mol Biol Transl Sci*. 156:1–62.
- Sato S. 2018. Cytosolic galectins and their release and roles as carbohydrate-binding proteins in host-pathogen interaction. *Trends Glycosci Glycotechnol*. 30:SE129–SE135.
- Sato M, Nishi N, Shoji H, Seki M, Hashidate T, Hirabayashi J, Kasai K-i, Hata Y, Suzuki S, Hirashima M et al. 2002. Functional analysis of the carbohydrate recognition domains and a linker peptide of galectin-9 as to eosinophil chemoattractant activity. *Glycobiology*. 12:191–197.
- Schengrund C-L. 2015. Gangliosides glycosphingolipids essential for normal neural development and function. *Trends Biochem Sci*. 40:397–406.
- Schnaer RL. 2015. Glycans and glycan-binding proteins in immune regulation: a concise introduction to glycobiology for the allergist. *J Allergy Clin Immunol*. 135:609–615.
- Schwefel D, Maierhofer C, Beck JG, Seeberger S, Diederichs K, Möller HM, Welte W, Wittmann V. 2010. Structural basis of multivalent binding to wheat germ agglutinin. *J Am Chem Soc*. 132:8704–8719.
- Solís D, Bovin NV, Davis AP, Jiménez-Barbero J, Romero A, Roy R, Smetana K Jr, Gabius H-J. 2015. A guide into glycosciences: how chemistry, biochemistry and biology cooperate to crack the sugar code. *Biochim Biophys Acta*. 1850:186–235.
- Song X, Xia B, Stowell SR, Lasanajak Y, Smith DF, Cummings RD. 2009. Novel fluorescent glycan microarray strategy reveals ligands for galectins. *Chem Biol*. 16:36–47.
- Still WC, Tempczyk A, Hawley RC, Handrickson T. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc*. 112:6127–6129.
- Stowell SR, Arthur CM, Mehta P, Slanina KA, Blixt O, Leffler H, Smith DF, Cummings RD. 2008. Galectin-1, -2, and -3 exhibit differential recognition of sialylated glycans and blood group antigens. *J Biol Chem*. 283:10109–10123.
- Stowell SR, Cho M, Feasley CL, Arthur CM, Song X, Colucci JK, Karmakar S, Mehta P, Dias-Baruffi M, McEver RP et al. 2009. Ligand reduces galectin-1 sensitivity to oxidative inactivation by enhancing dimer formation. *J Biol Chem*. 284:4989–4999.
- Stowell SR, Dias-Baruffi M, Penttilä L, Renkonen O, Nyame AK, Cummings RD. 2004. Human galectin-1 recognition of poly-N-acetylglucosamine and chimeric polysaccharides. *Glycobiology*. 14:157–167.
- Tasumi S, Vasta GR. 2007. A galectin of unique domain organization from hemocytes of the Eastern oyster (*Crassostrea virginica*) is a receptor for the protistan parasite *Perkinsus marinus*. *J Immunol*. 179:3086–3098.
- Theocharis CR, Jones W, Ramachandra Rao CN. 1984. An unusual photo-induced conformational polymorphism: a crystallographic study of bis(p-methoxy)-trans-stilbene. *J Chem Soc, Chem Commun*. 19:1291–1293.
- Tribulatti MV, Figini MG, Carabelli J, Cattaneo V, Campetella O. 2012. Redundant and antagonistic functions of galectins-1, -3, and -8 in the elicitation of T cell responses. *J Immunol*. 188:2991–2999.
- Tsai C-M, Chiu Y-K, Hsu T-L, Lin I-Y, Hsieh S-L, Lin K-I. 2008. Galectin-1 promotes immunoglobulin production during plasma cell differentiation. *J Immunol*. 181:4570–4579.
- van der Leij J, van den Berg A, Harms G, Eschbach H, Vos H, Zwiers P, van Weeghel R, Groen H, Poppema S, Visser L. 2007. Strongly enhanced IL-10 production using stable galectin-1 homodimers. *Mol Immunol*. 44:506–513.
- Vértessy S, Michalak M, Miller MC, Schnölzer M, André S, Kopitz J, Mayo KH, Gabius H-J. 2015. Structural significance of galectin design: impairment of homodimer stability by linker insertion and partial reversion by ligand presence. *Protein Eng Des Sel*. 28:199–210.
- Wang G-N, André S, Gabius H-J, Murphy PV. 2012. Bi- to tetraivalent glycoclusters: synthesis, structure-activity profiles as lectin inhibitors and impact of combining both valency and headgroup tailoring on selectivity. *Org Biomol Chem*. 10:6893–6907.
- Weinmann D, Kenn M, Schmidt S, Schmidt K, Walzer SM, Kubista B, Windhager R, Schreiner W, Toegel S, Gabius H-J. 2018. Galectin-8 induces functional disease markers in human osteoarthritis and cooperates with galectins-1 and -3. *Cell Mol Life Sci*. 75:4187–4205.
- Xiao Q, Ludwig A-K, Romano C, Buzzacchera I, Sherman SE, Vetro M, Vértessy S, Kaltner H, Reed EH, Möller M et al. 2018. Exploring functional pairing between surface glycoconjugates and human galectins using programmable glycodendrimersomes. *Proc Natl Acad Sci USA*. 115: E2509–E2518.
- Yang R-Y, Hill PN, Hsu DK, Liu F-T. 1998. Role of the carboxyl-terminal lectin domain in self-association of galectin-3. *Biochemistry*. 37:4086–4092.
- Zhang S, Moussodia R-O, Murzeau C, Sun HJ, Klein ML, Vértessy S, André S, Roy R, Gabius H-J, Percec V. 2015. Dissecting molecular aspects of cell interactions using glycodendrimersomes with programmable glycan presentation and engineered human lectins. *Angew Chem Int Ed*. 54:4036–4040.
- Zivicová V, Broz P, Fik Z, Miřková A, Plzák J, Cada Z, Kaltner H, Kucerová JF, Gabius H-J, Smetana K Jr. 2017. Genome-wide expression profiling (with focus on the galectin network) in tumor, transition zone and normal tissue of head and neck cancer: marked differences between individual patients and the site of specimen origin. *Anticancer Res*. 37:2275–2288.
- Zuber C, Roth J. 2009. N-Glycosylation. In: Gabius H-J, editor. *The Sugar Code. Fundamentals of Glycosciences*. Weinheim, Germany: Wiley-VCH. p. 87–110.

2.5 Bibliography

- (1) André, S.; O’Sullivan, S.; Koller, C.; Murphy, P. V.; Gabius, H. J. Bi- to Tetraivalent Glycoclusters Presenting GlcNAc/GalNAc as Inhibitors: From Plant Agglutinins to Human Macrophage Galactose-Type Lectin (CD301) and Galectins. *Org. Biomol. Chem.* **2015**, *13* (14), 4190–4203. <https://doi.org/10.1039/c5ob00048c>.
- (2) André, S.; O’Sullivan, S.; Gabius, H. J.; Murphy, P. V. Glycoclusters as Lectin Inhibitors: Comparative Analysis on Two Plant Agglutinins with Different Folding as a Step towards Rules for Selectivity. *Tetrahedron* **2015**, *71* (38), 6867–6880. <https://doi.org/10.1016/j.tet.2015.07.020>.
- (3) Kaltner, H.; Manning, J. C.; García Caballero, G.; Di Salvo, C.; Gabba, A.; Romero-Hernández, L. L.; Knospe, C.; Wu, D.; Daly, H. C.; O’Shea, D. F.; et al. Revealing Biomedically Relevant Cell and Lectin Type-Dependent Structure-Activity Profiles for Glycoclusters by Using Tissue Sections as an Assay Platform. *RSC Adv.* **2018**, *8* (50), 28716–28735. <https://doi.org/10.1039/c8ra05382k>.
- (4) Kutzner, T. J.; Gabba, A.; FitzGerald, F. G.; Shilova, N. V.; García Caballero, G.; Ludwig, A.-K.; Manning, J. C.; Knospe, C.; Kaltner, H.; Sinowatz, F.; et al. How Altering the Modular Architecture Affects Aspects of Lectin Activity: Case Study on Human Galectin-1. *Glycobiology* **2019**, *29* (8), 593–607. <https://doi.org/10.1093/glycob/cwz034>.
- (5) Meldal, M. P.; Meldal, M.; Tornøe, C. W. Cu-Catalyzed Azide – Alkyne Cycloaddition Cu-Catalyzed Azide - Alkyne Cycloaddition. *Chem. Rev* **2017**, *108* (8), 2952–3015. <https://doi.org/10.1021/cr0783479>.
- (6) Liang, L.; Astruc, D. The Copper(I)-Catalyzed Alkyne-Azide Cycloaddition (CuAAC) “Click” Reaction and Its Applications. An Overview. *Coord. Chem. Rev.* **2011**, *255* (23–24), 2933–2945. <https://doi.org/10.1016/j.ccr.2011.06.028>.
- (7) Bachl, J.; Mayr, J.; Sayago, F. J.; Cativiela, C.; Díaz Díaz, D. Amide-Triazole Isosteric Substitution for Tuning Self-Assembly and Incorporating New Functions into Soft Supramolecular Materials. *Chem. Commun.* **2015**, *51* (25), 5294–5297. <https://doi.org/10.1039/c4cc08593k>.
- (8) Bonandi, E.; Christodoulou, M. S.; Fumagalli, G.; Perdicchia, D.; Rastelli, G.; Passarella, D. The 1,2,3-Triazole Ring as a Bioisostere in Medicinal Chemistry. *Drug Discovery Today*. Elsevier Ltd 2017, pp 1572–1581. <https://doi.org/10.1016/j.drudis.2017.05.014>.
- (9) Andrus, M. B.; Mendenhall, K. G.; Meredith, E. L.; Soma Sekhar, B. B. V. Glycolate Aldol Reactions with Boron Enolates of Bis-4-Methoxyphenyl Dioxanone. *Tetrahedron Lett.* **2002**, *43* (10), 1789–1792. [https://doi.org/10.1016/S0040-4039\(02\)00152-1](https://doi.org/10.1016/S0040-4039(02)00152-1).
- (10) Oh, K.; Kim, S.; Lee, J.; Cho, W.; Lee, T.; Kim, S. Discovery of Diarylacrylonitriles as a Novel Series of Small Molecule Sortase A Inhibitors. *J. Med. Chem.* **2004**, *5* (47), 2418–2421. <https://doi.org/10.1021/jm0498708>.
- (11) Alonso, F.; Riente, P.; Yus, M. Wittig-Type Olefination of Alcohols Promoted by Nickel Nanoparticles: Synthesis of Polymethoxylated and Polyhydroxylated Stilbenes. *European J. Org. Chem.* **2009**, No. 34, 6034–6042. <https://doi.org/10.1002/ejoc.200900951>.
- (12) Hu, X. M.; Chen, Q.; Wang, J. X.; Cheng, Q. Y.; Yan, C. G.; Cao, J.; He, Y. J.; Han, B. H. Tetraphenylethylene-Based Glycoconjugate as a Fluorescence “Turn-on” Sensor for Cholera Toxin. *Chem. - An Asian J.* **2011**, *6* (9), 2376–2381. <https://doi.org/10.1002/asia.201100141>.
- (13) Knapp, S.; Myers, D. S. α -GlcNAc Thioconjugates. *J. Org. Chem.* **2001**, *66* (10), 3636–3638. <https://doi.org/10.1021/jo010088e>.
- (14) Tropper, F. D.; Andersson, F. O.; Braun, S.; Roy, R. Phase Transfer Catalysis as a General and Stereoselective Entry into Glycosyl Azides from Glycosyl Halides. *Synthesis (Stuttg.)*. **2003**, 1992 (07), 618–620. <https://doi.org/10.1055/s-1992-26175>.
- (15) Bock, K.; Lundt, I.; Pedersen, C. Assignment of Anomeric Structure to Carbohydrates through Geminal¹³C-H Coupling Constants. *Tetrahedron Lett.* **1973**, *14* (13), 1037–1040. [https://doi.org/10.1016/S0040-4039\(01\)95898-8](https://doi.org/10.1016/S0040-4039(01)95898-8).

- (16) Bock, K.; Pedersen, C. A Study of ^{13}C Coupling Constants in Hexopyranoses. *J. Chem. Soc. Perkin Trans. 2* **2004**, No. 3, 293. <https://doi.org/10.1039/p29740000293>.
- (17) Yu, B.; Van Ingen, H.; Vivekanandan, S.; Rademacher, C.; Norris, S. E.; Freedberg, D. I. More Accurate ^1J CH Coupling Measurement in the Presence of ^3J HH Strong Coupling in Natural Abundance. *J. Magn. Reson.* **2011**, *215*, 10–22. <https://doi.org/10.1016/j.jmr.2011.09.037>.
- (18) Doyle, L. M.; O’Sullivan, S.; Di Salvo, C.; McKinney, M.; McArdle, P.; Murphy, P. V. Stereoselective Epimerizations of Glycosyl Thiols. *Org. Lett.* **2017**, *19* (21), 5802–5805. <https://doi.org/10.1021/acs.orglett.7b02760>.
- (19) Appukkuttan, P.; Mehta, V. P.; Van Der Eycken, E. V. Microwave-Assisted Cycloaddition Reactions. *Chemical Society Reviews*. 2010, pp 1467–1477. <https://doi.org/10.1039/b815717k>.
- (20) Luo, J.; Xie, Z.; Xie, Z.; Lam, J. W. Y.; Cheng, L.; Chen, H.; Qiu, C.; Kwok, H. S.; Zhan, X.; Liu, Y.; et al. Aggregation-Induced Emission of 1-Methyl-1,2,3,4,5-Pentaphenylsilole. *Chem. Commun.* **2001**, *18*, 1740–1741. <https://doi.org/10.1039/b105159h>.
- (21) Hu, R.; Leung, N. L. C.; Tang, B. Z. AIE Macromolecules: Syntheses, Structures and Functionalities. *Chem. Soc. Rev.* **2014**, *43* (13), 4494–4562. <https://doi.org/10.1039/c4cs00044g>.
- (22) Donnier-Maréchal, M.; Abdullayev, S.; Bauduin, M.; Pascal, Y.; Fu, M.-Q.; He, X.-P.; Gillon, E.; Imberty, A.; Kipnis, E.; Dessein, R.; et al. Tetraphenylethylene-Based Glycoclusters with Aggregation-Induced Emission (AIE) Properties as High-Affinity Ligands of Bacterial Lectins. *Org. Biomol. Chem.* **2018**, *16* (45), 8804–8809. <https://doi.org/10.1039/C8OB02035C>.
- (23) Tseng, N. W.; Liu, J.; Ng, J. C. Y.; Lam, J. W. Y.; Sung, H. H. Y.; Williams, I. D.; Tang, B. Z. Deciphering Mechanism of Aggregation-Induced Emission (AIE): Is E-Z Isomerisation Involved in an AIE Process? *Chem. Sci.* **2012**, *3* (2), 493–497. <https://doi.org/10.1039/c1sc00690h>.
- (24) Hong, Y.; Lam, J. W. Y.; Tang, B. Z. Aggregation-Induced Emission: Phenomenon, Mechanism and Applications. *Chem. Commun.* **2009**, No. 29, 4332–4353. <https://doi.org/10.1039/b904665h>.
- (25) Mertz, J. A.; Mccann, J. A.; Picking, W. D. Fluorescence Analysis of Galactose, Lactose, and Fucose Interaction with the Cholera Toxin B Subunit (1, 2), but Details of the Cholera Toxin B Subunit (CTB) Role in Intoxication Remain Unclear. Target Cell Cytoplasm Where It Exerts Cytopathic Eff. **2010**, *144* (1996), 140–144.
- (26) Pickens, J. C.; Merritt, E. A.; Ahn, M.; Verlinde, C. L. M. J.; Hol, W. G. J.; Fan, E. Anchor-Based Design of Improved Cholera Toxin and E. Coli Heat-Labile Enterotoxin Receptor Binding Antagonists That Display Multiple Binding Modes. *Chem. Biol.* **2002**, *9* (2), 215–224. [https://doi.org/10.1016/S1074-5521\(02\)00097-2](https://doi.org/10.1016/S1074-5521(02)00097-2).
- (27) Marcelo, F.; Supekar, N.; Corzana, F.; Van Der Horst, J. C.; Vuist, I. M.; Live, D.; Boons, G. J. P. H.; Smith, D. F.; Van Vliet, S. J. Identification of a Secondary Binding Site in Human Macrophage Galactose-Type Lectin by Microarray Studies: Implications for the Molecular Recognition of Its Ligands. *J. Biol. Chem.* **2019**, *294* (4), 1300–1311. <https://doi.org/10.1074/jbc.RA118.004957>.
- (28) Freyer, M. W.; Lewis, E. A. Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions. *Methods Cell Biol.* **2008**, *84* (07), 79–113. [https://doi.org/10.1016/S0091-679X\(07\)84004-0](https://doi.org/10.1016/S0091-679X(07)84004-0).
- (29) Jerabek-Willemsen, M.; André, T.; Wanner, R.; Roth, H. M.; Duhr, S.; Baaske, P.; Breitsprecher, D. MicroScale Thermophoresis: Interaction Analysis and Beyond. *J. Mol. Struct.* **2014**, *1077*, 101–113. <https://doi.org/10.1016/j.molstruc.2014.03.009>.
- (30) Dam, T. K.; Roy, R.; Pagé, D.; Brewer, C. F. Negative Cooperativity Associated with Binding of Multivalent Carbohydrates to Lectins. Thermodynamic Analysis of the “Multivalency Effect.” *Biochemistry* **2002**, *41* (4), 1351–1358. <https://doi.org/10.1021/bi015830j>.
- (31) Dam, T. K.; Roy, R.; Pagé, D.; Brewer, C. F. Thermodynamic Binding Parameters of Individual Epitopes of Multivalent Carbohydrates to Concanavalin A as Determined by “Reverse” Isothermal Titration Microcalorimetry. *Biochemistry* **2002**, *41* (4), 1359–1363. <https://doi.org/10.1021/bi015829k>.
- (32) Dam, T. K.; Brewer, C. F. Thermodynamic Studies of Lectin-Carbohydrate Interactions by Isothermal Titration Calorimetry. *Chem. Rev.* **2002**, *102* (2), 387–429. <https://doi.org/10.1021/cr000401x>.

- (33) Freire, E.; Mayorga, O. L.; Straume, M. Isothermal Titration. *Anal. Chem.* **1990**, 62 (18), 950A–959A. <https://doi.org/10.1021/ac00217a002>.
- (34) Marcelo, F.; Garcia-Martin, F.; Matsushita, T.; Sardinha, J.; Coelho, H.; Oude-Vrielink, A.; Koller, C.; André, S.; Cabrita, E. J.; Gabius, H. J.; et al. Delineating Binding Modes of Gal/GalNAc and Structural Elements of the Molecular Recognition of Tumor-Associated Mucin Glycopeptides by the Human Macrophage Galactose-Type Lectin. *Chem. - A Eur. J.* **2014**, 20 (49), 16147–16155. <https://doi.org/10.1002/chem.201404566>.
- (35) Dam, T. K.; Roy, R.; Das, S. K.; Oscarson, S.; Brewer, C. F. Binding of Multivalent Carbohydrates to Concanavalin A and Dioclea Grandiflora Lectin. Thermodynamic Analysis of the “Multivalency Effect.” *J. Biol. Chem.* **2000**, 275 (19), 14223–14230. <https://doi.org/10.1074/jbc.275.19.14223>.
- (36) Rohse, P.; Wittmann, V. Mechanistic Insight into Nanomolar Binding of Multivalent Neoglycopeptides to Wheat Germ Agglutinin. *Chem. - A Eur. J.* **2016**, 22 (28), 9724–9733. <https://doi.org/10.1002/chem.201600657>.
- (37) Dam, T. K.; Talaga, M. L.; Fan, N.; Brewer, C. F. *Measuring Multivalent Binding Interactions by Isothermal Titration Calorimetry*, 1st ed.; Elsevier Inc., 2016; Vol. 567. <https://doi.org/10.1016/bs.mie.2015.08.013>.
- (38) Dam, T. K.; Oscarson, S.; Brewer, C. F. Thermodynamics of Binding of the Core Trimannoside of Asparagine-Linked Carbohydrates and Deoxy Analogs to Dioclea Grandiflora Lectin. *J. Biol. Chem.* **1998**, 273 (49), 32812–32817. <https://doi.org/10.1074/jbc.273.49.32812>.
- (39) Rozwarski, D. A.; Swami, B. M.; Brewer, C. F.; Sacchettini, J. C. Crystal Structure of the Lectin from Dioclea Grandiflora Complexed with Core Trimannoside of Asparagine-Linked Carbohydrates. *J. Biol. Chem.* **1998**, 273 (49), 32818–32825. <https://doi.org/10.1074/jbc.273.49.32818>.
- (40) Shahzad-ul-Hussan, S.; Gustchina, E.; Ghirlando, R.; Clore, G. M.; Bewley, C. A. Solution Structure of the Monovalent Lectin Microvirin in Complex with Man α (1-2)Man Provides a Basis for Anti-HIV Activity with Low Toxicity. *J. Biol. Chem.* **2011**, 286 (23), 20788–20796. <https://doi.org/10.1074/jbc.M111.232678>.
- (41) Wiederschain, G. Glycoscience. Chemistry and Chemical Biology. *Biochem.* **2009**, 74 (11), 1289–1289. <https://doi.org/10.1134/s0006297909110170>.
- (42) Jégouzo, S. A.; Quintero-Martínez, A.; Ouyang, X.; Dos Santos, Á.; Taylor, M. E.; Drickamer, K. Organization of the Extracellular Portion of the Macrophage Galactose Receptor: A Trimeric Cluster of Simple Binding Sites for N-Acetylgalactosamine. *Glycobiology* **2013**, 23 (7), 853–864. <https://doi.org/10.1093/glycob/cwt022>.
- (43) Rao, J.; Lahiri, J.; Isaacs, L.; Weis, R. M.; Whitesides, G. M. A Trivalent System from Vancomycin · D-Ala-d-Ala with Higher Affinity than Avidin Biotin. *Science (80-.)*. **1998**, 280, 708–711. <https://doi.org/10.1126/science.280.5364.708>.
- (44) Lee, Y. C.; Townsend, R. R.; Hardy, M. R.; Lönngren, J.; Arnarp, J.; Haraldsson, M.; Lönn, H. Binding of Synthetic Oligosaccharides to the Hepatic Gal/GalNAc Lectin. Dependence on Fine Structural Features. *J. Biol. Chem.* **1983**, 258 (1), 199–202.

3. Synthesis and evalion of mucin 1 and mucin 4 (neo)peptide library containing a ligand for hMGL. Towards the development of candidate vaccines.

3.1. Abstract

Mucins are protein over expressed on epithelial tumour cells.¹ Due to changes in the cell microenvironment, the glycosylation pattern of healthy mucins differs from that of tumour mucins. Aberrantly short glycans are present on mucins, mainly the Thomsen-Friedenreich antigen (T_N-antigen) and its sialylated version (ST_N-antigen).² The human macrophage galactose-type lectin (h-MGL), an endocytic receptor located on the surface of immune system cells (see chapter 1.) is the only known receptor with the ability to bind these glycans and, after internalisation, potentially present them via MHC class I and II pathways to T cells.³ The ability of h-MGL to selectively recognise aberrant glycans makes it a potential target receptor in the development of in vivo dendritic cell (DC) based immunotherapy, such as in vaccine development.

One of the goals in cancer immunology is to avoid tumour metastasis or relapse by priming the immune system to recognize and attach malignant cells.⁴ The fact that tumour cells have a different glycosylation pattern from normal cells makes this challenge feasible. The tumour associated carbohydrate antigens (TACAs) can be used in the design of synthetic antigen-mimic carbohydrates for the development of new immunological strategies.⁴

Aside from exceptional cases,⁵ carbohydrates themselves are unable to induce a T-cell-dependent response and to target the MHC-II, and thus immunoenhancing elements need to be added in the vaccine construct. Several immune enhancing strategies are known,⁴ but the work disclosed in this chapter, is based on the innovative strategy of Horst Kunz and co-workers, with Ulrika Westerlind among them,⁶ who firstly introduced mucin derived epitopes in combination with tetanus toxoid (TT) for the development of potential glycopeptide vaccine candidates.^{7,8}

Kunz and co-workers in the past identified a difluoro-TF-TT candidate, figure 1, able to generate IgG antibodies in the preclinical setting,⁹ proving the feasibility of this approach.

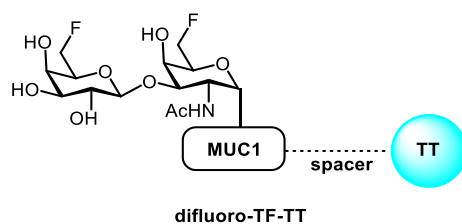


Figure 1. Structure of difluoro-TF-TT vaccine. This vaccine candidate was found to exhibit immunogenic activity in preclinical settings and IgG antibodies induced by difluoro-TF-TT recognised the native TF antigens present on the MCF-7 breast cancer cell lines.⁹

Several issues still need to be fully addressed: a selective immune response must be evoked to avoid the autoimmunity problem; the synthetic vaccine needs to lead to immune memories (memory B cells) and the vaccine candidate needs to overcome the natural tolerance level of the immune system.

Recently, numerous studies directed on DC based immune therapy had encouraging results, which are now shifting the immune therapy strategy from *ex-vivo* culturing to *in-vivo* targeting of DC receptors.¹⁰

Among the most common surface proteins exploited for this purpose, we find DC-SIGN,¹¹ CD205¹² and mannose receptors,¹³ with some of those leading to synthetic vaccine candidates, which were tested in clinical trials phase I and II.^{14,15}

Our strategy, to partially address the issues mentioned above, is to target DCs via h-MGL. Despite h-MGL being the only receptor that selectively recognises GalNAc derivatives in macrophages and dendritic cells, it has not been used classically for DC targeting immune therapy, and very little research has been carried out to investigate its full potential.

The h-MGL is reported to drive Th2-mediated responses via the MHC class II processing and presentation pathway.³ In this work mucin 1 and 4 epitopes were functionalised with a trivalent glycocluster based on the tetraphenylethylene core, which is known to have high affinity for hMGL (see chapter 2). To mimic the tumour glycan microheterogeneity and to exploit multivalency, serine and threonine-based T_N-antigen were inserted in the sequence and further conjugation with different immune stimulants such as immune silent-polymers, P30 T-cell epitope and a peptide sequence from Tetanus toxoid was performed.

The aim of the work described in the first part of this chapter is to identify a synthetic unimolecular multicomponent vaccine, based on a mucin backbone and T_N-antigen like molecules, which are able to induce a robust T-cell-dependent response through the increased antigen uptake (endocytosis) by DCs via the h-MGL receptor. The long term goal

of the project is to test the ability of the candidate to elicit IgG antibodies in humanised mice and to test the ability of the produced antibody to recognise native tumorigenic antigens.^{16,17} Preliminary data about the (neo)peptides binding to h-MGL were collected via ELISA by Sandra Van Vliet and co-worker and these results are briefly discussed in section 3.5.8.

3.2. Introduction: basic immunology principles

Since the first half of the 1980s, the discovery that tumour cells have different glycosylation outline on their surface compared to normal cells⁶ has stimulated the advancement of the area of tumour immunology. The structural identification of these unique tumour fingerprints, the tumour associated carbohydrate antigens (TACAs), allowed synthetic chemists to design and obtain structural analogues that could be presented to the immune system and evoke an adaptive immune response.

Before diving into my PhD project, as a chemist with limited knowledge in immunology, I believed all antigens cause antibody production, and that the presence of antibodies led to immunity. As I have learned during this work, this is a naive belief.

Firstly, due to the microheterogeneity of glycans type on the cell surface and the fact TACAs are also present on healthy cells, often these glycans are processed as self-antigens, not triggering an immune response. Secondly, I have learned that even when antibody production arises that not all antibodies have the same value in term of their “immunity coin”.

In fact, most carbohydrate antigens initialise only a weak T-cell independent response: they bind to receptors located on B cell surfaces, which then differentiate to plasma B cells and produce short lived antibodies, the immunoglobulin M, IgM. This type of response is defined as a T cell-independent response, figure 2, and the antibodies produced from it, IgM, are not lasting antibodies, and we can say they are that are less valuable in terms of ‘immunity coin’.¹⁸

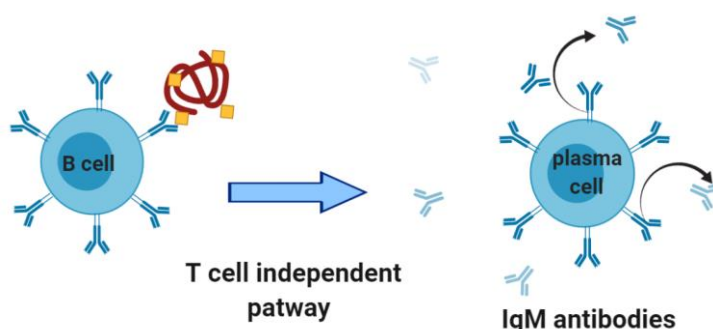


Figure 2. Representation of T cell independent pathway, drawn with BioRender. Most antigens induce a weak T cell independent humoral response. After the engagement of their surface receptors, B cells differentiate to plasma B cells, and produce only short-lived, low-affinity IgM antibodies.

To produce long lasting antibodies, desired in immunology, antigens capable of activating the T cell-dependent pathway are needed. These antigens bind to receptor located on the antigen presenting cells surfaces (APCs), such as dendritic cells, get internalised, digested

and again presented on the APCs surface via the major histocompatibility complexes II (MHC-II). At this stage, T helper cells (T_h), bind to the complex thanks to other proteins located on their surface and start to produce cytokines, a large group of proteins, peptides or glycoproteins. The presence of these molecules activates B cells and other type of T cells, the cytotoxic T cells. It should not be forgotten that B cells have always been around all of this time with the antigen already bound on their surfaces receptors but, it is only when T_h cells present them the same antigen, that potent cytokines are produced and the B cells are willing to differentiate to memory B cells and produce immunoglobulin G antibodies, IgG, figure 3.¹⁹

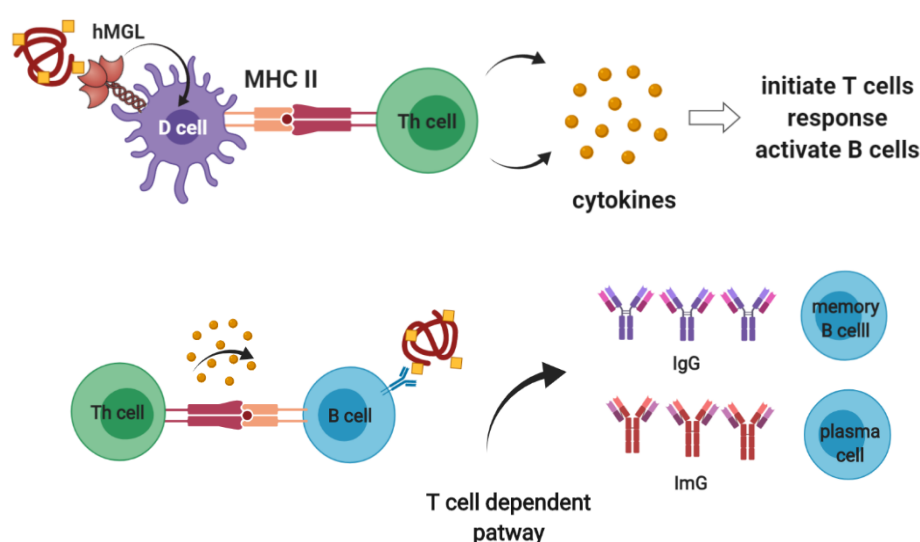


Figure 3. Representation of the T cell dependent pathway, drawn with BioRender. Receptors on antigen presenting cells, can induce antigens internalisation and digestion. In some case, the antigens are re-present on the D cells surface to T cells via MHC II. T cells start production of cytokines which activate B cells and induce them to produce long lasting antibodies, IgG and memory B cells. Examples of receptor targeted for antigen delivery via DCs are, mannose receptor, CD205, DC-SIGN, Dectin-1 and Lox1.

DCs can bind both pathogenic and harmless antigens, internalise them and present them through MHC to T cells. If the presentation is accompanied by a danger signal, the DCs fully mature and induce a lifelong lasting immunity via a T cell-dependent response, while if the antigens processed are harmless, no danger signal is produced, and T cells become tolerogenic.

Dendritic cell based immune therapy gave promising results²⁰ in the past, and it currently seems a great strategy to move arduous and expensive *ex-vivo* culturing to *in-vivo* targeting. The review of Paul J. Tacken et al. offers a comprehensive overview of DC receptors used in the past for targeting studies.¹⁰ Only recently the scientific community gained interest in h-MGL for this purpose. Napolitano et al., in 2007 proved the ability of h-MGL to

internalise MUC1 glycopeptides and, depending on their chemical structure, process them through HLA-I or II. If the MHC-II is initiated, helper T cells bind to the MHC-antigen complex and their activation can cause a class switching of the response from T cell-independent to T cell-dependent.³

Other examples of the use of this receptor for dendritic cell immune strategy include the use of synthetic vaccines based on T_N antigen moieties that are O-linked to tetanus toxoid epitope created by Claude Leclerc's group.²¹ This latter work involved induced IgG antibody production in non-human primates, which were able to mediate antibody-dependent cytotoxicity against a panel of T_N antigen positive human tumour cell lines and in 2016 entered phase I clinical trial under the name TT830-844 MAG-Tn3.^{22,23,24}

The presentation cascade is highly structure dependent and it is often reported that, upon engagement with antigen, h-MGL triggers anti-inflammatory cytokine production, such as that of IL-10,²⁵ which correlates to cancer escape from the immune system sentinels. This feature, which h-MGL shares with several receptors used in a DC vaccination strategy, including DC-SIGN and SIGLECs, seems to oppose the desired antigen presentation to T cells.

Although, it needs to be remembered that the immunity game is quite complex, and it is based on fine equilibrium: even if IL-10 acts as an anti-inflammatory cytokine, it also promotes the growth of CD8⁺ and B cells boosting their responses. The high IL-10 production by h-MGL licensed DCs might thus favour humoral immunity. Clearly, the potential positive effects of MGL triggering on CD8⁺ T and B cells need careful evaluation, before human MGL can be used as a therapeutic target for the induction of regulatory T cell responses.²⁵

Eventually, it is worth mentioning though, that most of TACAs on their own are not presented by APCs by the MHC-II, and therefore cannot activate a T cell like dependent immune response. The candidate vaccine needs to be enriched with elements which will enable its processing via MHC-II, like mucin epitopes.⁶

Our goal is to enhance the likelihood of the mucin glycopeptide processing by dendritic cells, so to maximise the chance of the MHC-II activation. h-MGL receptor is known to process MUC1 glycopeptides in this way, so as part of this thesis work, there was an aim to engage this receptor with a highly selective trivalent glycocluster embedded into peptides.

3.3. The mucin family

Mucins are an extended family of heavily glycosylated proteins which includes 21 members. They can be divided in two main classes, those that get secreted and form gels and those bound to a membrane. Both, mucin 1 (MUC1) and mucin 4 (MUC4), analysed in this chapter belong to the second category. Thanks to their glycosylation pattern they play a fundamental role in innate immunity. In fact, the glycans on health cell mucins have a different chemical structure compared to the glycans of cancer cell mucins. They are typically formed by several tandem repeat of the same epitope and the number of repeats depend of genetic polymorphism and is different in each individual. The epitope repetition allows the usage of little tandem repeat sequences as immune stimulants and the synthesis of relatively short peptide sequences is achievable through chemical synthesis.²⁶

3.3.1. Abnormal glycosylation and the role of MUC1 and MUC4 in cancer

Two members of this large family are mucin-1 (MUC1) and mucin-4 (MUC4). These two proteins share common features: (1) are presented as different polymorphs due to a variable number of identical tandem repeats, (2) have multiple O-glycosylation sites, (3) are overexpressed in many cancer cell lines and (4) the glycans on cancer cell mucins are much shorter than those of normal cell mucins.²⁷ The most recurrent tumour associated carbohydrates found on mucins are the Tn antigen and its derivatives: the Thomsen–Friedenreich-(T)-antigen and the STn antigen, figure 4.

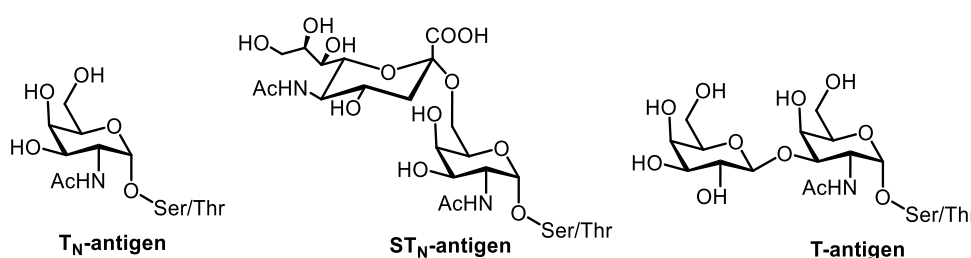


Figure 4. Structure of T_N, ST_N and T antigens

MUC4 is the first mucin to be expressed in lung epithelial tissue during development, it is present also in squamous epithelial cells of the oesophagus, small bronchi, bronchioles and is also reported on the digestive tract.^{28,29} The extracellular domain of MUC4 is composed by variable numbers of tandem repeats (VNTRs) TSSASTGHATPLPVTD, which present multiple possible O-glycosylation sites on

serine and threonine and is predicted to extend >2 μm above the cell surface in the apical region.³⁰

Aberrant MUC4 glycoforms were observed in many inflammatory diseases and several cancers.³¹ Just to cite few examples, MUC4 is up-regulated in high-grade dysplasia and adenocarcinoma of the oesophagus³² and was reported to be overexpressed in 91.4% cases of lung adenocarcinoma.³³

Similarly, MUC1 is expressed in many epithelial cells and the extracellular portion is as well composed by VNTRs of the sequence HGVTSAPDTRPAPGSTAPPA³⁴ with five O-glycosylation sites.

Likewise MUC4 and MUC1 expression are dramatically increased when the cells became malignant in particular in breast,¹ ovarian,^{35,36} pancreatic³⁷ and prostate cancers.³⁸

As anticipated above MUC1-4, not only are overexpressed in tumours cells, but also present an aberrant glycosylation pattern. In fact, due to microenvironmental changes, the activity of glycosyltransferases is altered, and the glycosylation process gets interrupted and leave the protein decorated with shorter glycans compared to the one of mucins on normal cell.

3.4. MGL based vaccine development

Interestingly, our immune system developed proteins which can selectively recognise and bind the aberrant glycans on mucins.

Human macrophage galactose C-type lectin (h-MGL, CLEC10A) is an endocytic receptor located on dendritic cells (DCs) and macrophage surfaces, which selectively binds, in a calcium dependent manner, to terminal $\alpha/\beta\text{GalNAc}$ derivatives including the Tn antigen and derivatives^{39,40} (see chapter 1). The h-MGL was proven to be an optimal receptor for the internalization of MUC1 glycopeptides,^{41,42} the processing is highly structure dependent, large Tn-bearing glycoproteins are trapped in an endo lysosomal compartment, while short peptides are delivered into HLA class I and II compartments.³ Furthermore, hMGL expressed specifically on CD1c+ DCs subset was reported to bind and internalise MUC-1 peptides and enhance TNF α , IL-8, and IL-10 secretion.⁴³

Recently, Eggink et al., found that, branched unglycosylated peptides bind the protein in a calcium dependent manner with high affinity. The multivalent molecules

svL4 and **svD6** are base lysine core in which each amine moiety of a lysine is reacted with the *C terminus* of another lysine until the desired valency is obtained, tetravalent molecules in this specific case. *In vivo* study a murine ovarian cancer model, showed that sv6D, figure 5, effectively suppressed development of ascites both as monotherapy and in combination with the chemotherapeutic drug paclitaxel or the immunotherapeutic antibody against the receptor PD-1.³⁶

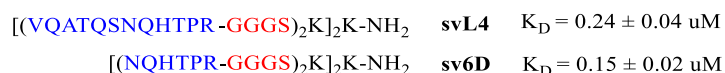


Figure 5. Peptide sequence of svL4 and sv6D.

Even if SV6D showed remarkable activity, no antibodies were detected after *in vivo* mouse model immunisation, indicating the peptide is not processed through the MHC.

With the aim to identify potential synthetic cancer vaccines able to lead to the production of IgG antibody, I synthesised an extensive library of peptides based on the MUC1-4 epitopes. The biotinylated tandem repeats were conjugated with immune stimulant P30 epitope and with a derivative of a potent trivalent glycocluster already shown to bind, in the low nanomolar range, both the extracellular domain of MGL as well as the hMGL carbohydrate recognition domain (see chapter 2).⁴⁴ The synthesis of **AG141** was modified to allow incorporation of the tetraphenylethylene core onto the sided chain of a suitably protected aspartic acid residue, making neoglycopeptide synthesis possible.

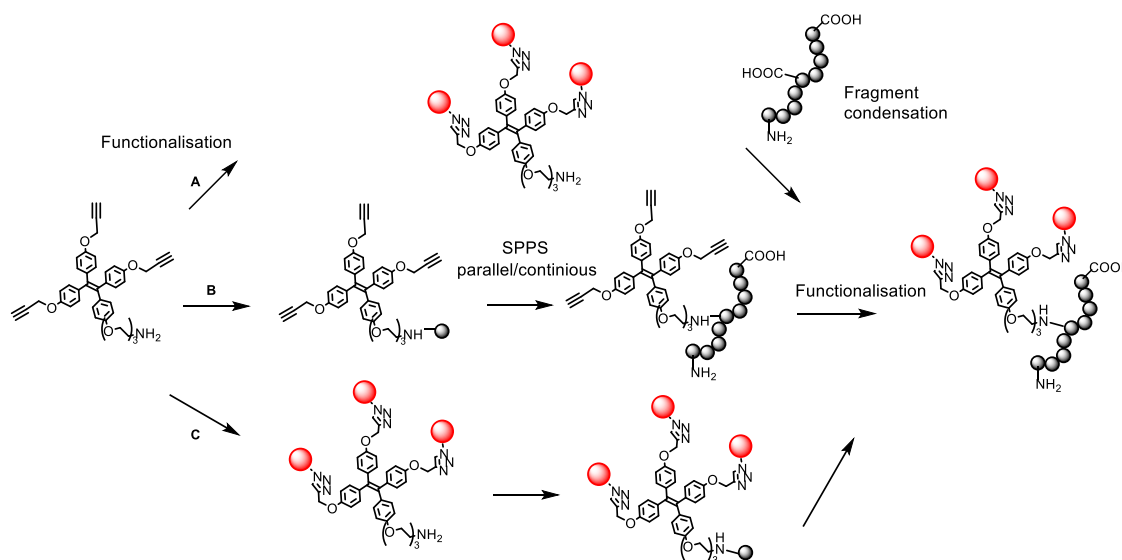
3.5. Results and discussion

3.5.1. (Neo)Glycopeptide synthesis: general approach

Different options were considered when planning the synthesis of the peptide library: a possible approach, scheme **1A**, was to functionalise with GalNAc the core **AG150** and then condense it with previously synthesised MUC1-MUC4 epitopes. This approach is commonly defined as convergent synthesis and used in the synthesis of N-glycoproteins and highly repetitive protein domains.

Typically, convergent synthesis results in low yield due to steric hindrance. The non-planar nature of the tetraphenylethylene residue raised concrete concerns in term of steric hindrance and, as discussed in chapter 2, it has a remarkable ability to aggregate. Furthermore, peptides can be poorly soluble in common glycosylation condition.

A second option was to condense the un-glycosylated tetraphenyl ethylene scaffold to the pre-made peptide and perform the alkyne-azide cycloaddition as last step, scheme **1B**. Copper catalysed cycloaddition is highly selective and can be performed in the presence of other functional moieties. Numerous examples of this approach are reported in literature.⁴⁵



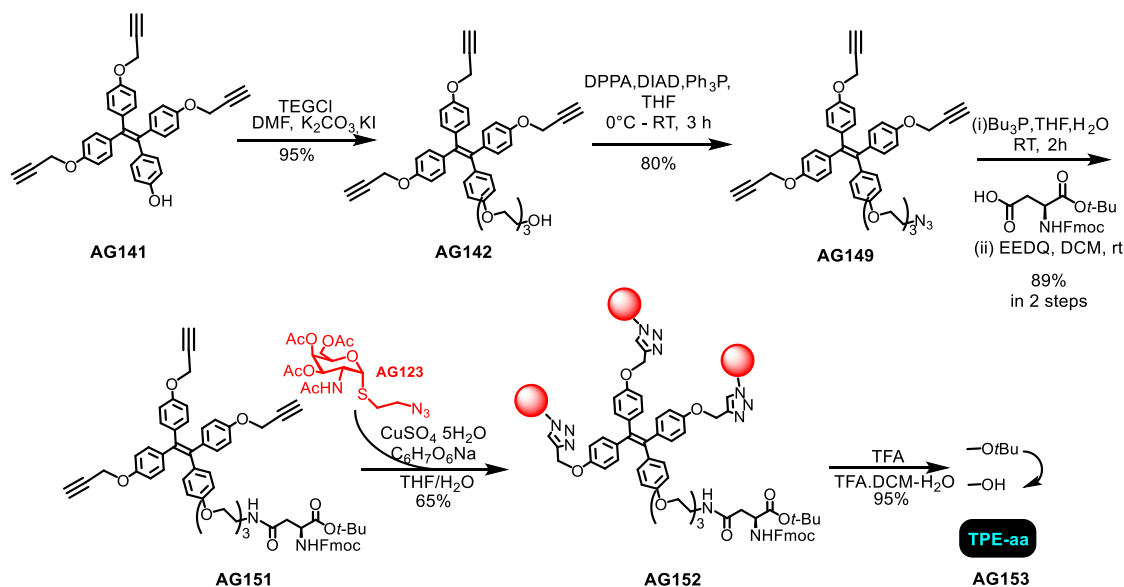
Scheme 1. Possible synthetic approaches to obtain the desired glycopeptides.

Although, due to the fact that the chemical synthesis of the modified amino acid **AG153** and the solid phase synthesis of the MUC1/4 epitope would have been carried out in separate laboratories, I decided to follow a third possible route, scheme **1C**. With this approach the glutamic acid functionalized with the trivalent glycocluster was prepared in

NUI Galway and then inserted in a parallel SPPS protocol well established by the Ulrika Westerlind laboratory in ISAS Dortmund. The Fmoc-SPPS is a commonly used strategy to generate glycopeptides and was used to prepare the glycosylated peptides reported in table 1. In principle, the synthesis could still be extended in the future by enzymatic glycosylation with sialic acid or fucose, furthermore this approach allowed us to insert other glycans such as the Ser/Thr T_N-antigen to ensure microheterogeneity and to explore if the presence of the multivalent scaffold itself significantly increases the affinity of the peptides for hMGL.

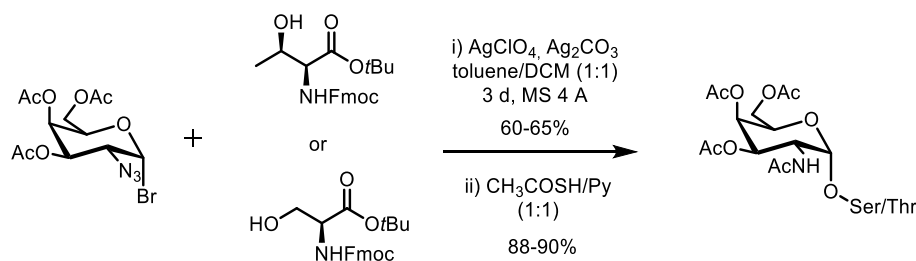
3.5.2. Synthesis of the conjugated amino acid

As anticipated in chapter 2, the trivalent tetraphenylethylene **AG141** presents a free hydroxyl group ideal for functionalisation. **AG141** was reacted with 2-[2-(2-chloroethoxy)ethoxy]ethanol to give **AG142**, scheme 2. This was followed by Mitsunobu reaction which yielded the azide **AG149** which was then reduced to the corresponding amine with tributylphosphine. With **AG150** in our hands, the coupling with NFmoc-Asp-OtBu was performed using EEDQ as coupling agent giving **AG151** in good yield. Subsequently, **AG123**, synthesised as reported in chapter 2 scheme 4, was ‘clicked’ on **AG151** via copper(I)-catalysed 1,3 dipolar cycloaddition.⁴⁴ Final acid promoted hydrolytic cleavage of the *tert*-butyl ester gave us the desired compound **AG153** for solid phase synthesis.



Scheme 2. Synthesis of asparagine derived building block, which was functionalised with the trivalent glycocluster based on tetraphenylethylene (42% yield in 6 steps).

The synthesis of S-T_N antigen and T-T_N antigen was carried out in the Westerlind group following the glycosylation methodology of H. Paulsen and J.-P. Hölck which involves the use of silver perchlorate/carbonate mixtures as glycosylation promoters and a glycosyl bromide as donor, scheme 3.⁴⁶



Scheme 3. Synthesis of serine T_N-antigen and threonine T_N-antigen.

3.5.3. Glycopeptide synthesis on solid phase

In solid phase synthesis, the amino acid chain is elongated step-wise. The first amino acid is connected on an insoluble solid support by a linkage which need to be stable to the chemical conditions of every step of the synthesis.

Typically, the solid support is a polymer, a resin, which can be functionalised with different linkers in which the length of the spacers, its bulkiness and the final functional group on which the first amino acid will be temporally connected, can be fine-tuned.

The first solid phase peptide synthesis was achieved by Merrifield and was based on the N-terminal tert-butyl oxycarbonyl (Boc) protecting group for the peptide amine and required aqueous hydrogen fluoride for release of the peptides from the resin.⁴⁷

A version of this methodology was developed thanks to the introduction of the base promoted removable amine protecting group, the fluorenylmethyloxycarbonyl, Fmoc.^{48,49}

The chemical conditions required to cleave the Fmoc protecting group, 20-40% piperidine in DMF, are non-acidic when compared to the Boc based synthesis and allows the insertion in the synthesis of glycosylated amino acids. In fact, the acetals of the glycosylic bonds are acid sensitive and the piperidine concentration required for the cleavage does not harm base labile connection of O-glycans to the peptide backbone.

The solid phase strategy is based on orthogonal protecting groups: the protecting groups on the amino acid side chains must not be cleaved in the same condition of the protecting group located on the main chain. The reader will immediately understand that the two

different approaches require different sets of amino acids; in the Fmoc strategy the functional groups on the amino acids side chains are protected with acid sensitive protecting groups, while for Boc strategy with base sensitive protecting group, figure 6. In the synthesis involving carbohydrates (glycopeptide synthesis), acetate protecting groups are often used to protect the saccharide OH groups. The acetates are stable in both Boc and Fmoc synthesis conditions and are removed after peptide cleavage from the resin.

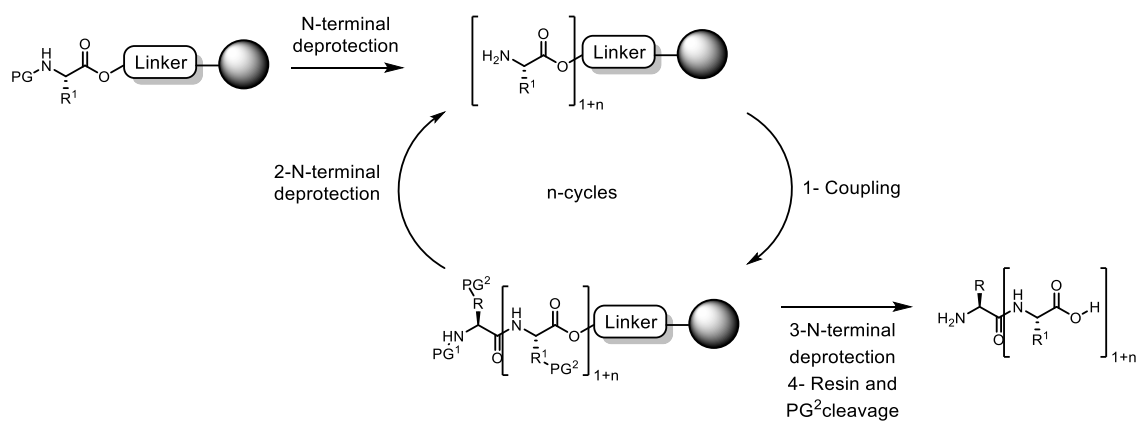
The Fmoc strategy is the one most currently used and the set of amino acids adopted in the synthesis is cheaper than the second one.

Amino acid	Fmoc strategy	Boc strategy
His (H)	Trt	Dnp, Bom
Arg (R)	Mtr, Pmc, Pbf	Tos, Mts
Lys (K)	Boc	2-Cl-Z
Asp (D)	t-Bu	Bzl, Cy
Glu (E)	t-Bu	Bzl, Cy
Trp (W)	Boc	For
Cys (C)	Trt, Acm	4-MeBzl, Acm
Ser (S)	t-Bu	Bzl
Thr (T)	t-Bu	Bzl
Tyr (Y)	t-Bu	2-Br-Z
Asn (N)	Trt	–
Gln (Q)	Trt	–

Figure 6. Common amino acid protecting group for Fmoc and Boc synthesis.

As described in scheme 4, in Fmoc based peptide synthesis, the peptide is attached via the C-terminal carboxyl group to the resin. After the initial Fmoc deprotection, every coupling step is performed at regular interval in a reactor in the presence of the solid support and the peptide grows towards the N-terminus. Each coupling cycle is repeated until the desired sequence is achieved. At the completion of each coupling step, unreacted amino acids and coupling reagents, are washed away by simply filtering off the resin.

The release from the resin requires acid, and typically trifluoroacetic acid TFA is used. Under these conditions the side chain protecting groups are normally also cleaved. Peptide purification then gives the desired product. Hence, the Fmoc solid phase peptide synthesis is extremely robust and both, the scale of the synthesis and the length of the peptide can vary significantly.



Scheme 4. General mechanism of solid phase peptide synthesis. The elongation proceeds from the C-terminus to the N-terminus. 1) attachment of a N α -protected amino acid and 2) removal of the N α protection. These steps are repeated in cycles until the desired chain length is reached and the peptide is 3) deprotected on the N-terminal side and 4) released from the resin.

3.5.4. Resin

The resin choice is fundamental for a successful synthesis. As briefly anticipated, the resin is an insoluble polymer, typically polystyrene (PS), cross-linked with 1% divinylbenzene which physically resembles small pearls.

The resin chosen for the mucin glycopeptide library is a copolymer of polyethylene glycol and polystyrene matrix, which is referred to as a TentaGel®-resin (Rapp Polymere GmbH, Tübingen), figure 7.

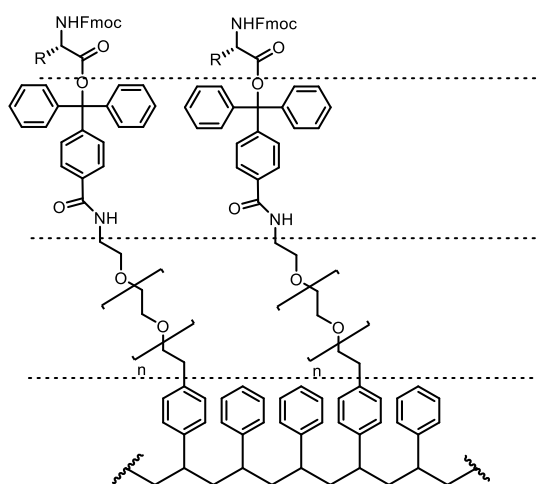
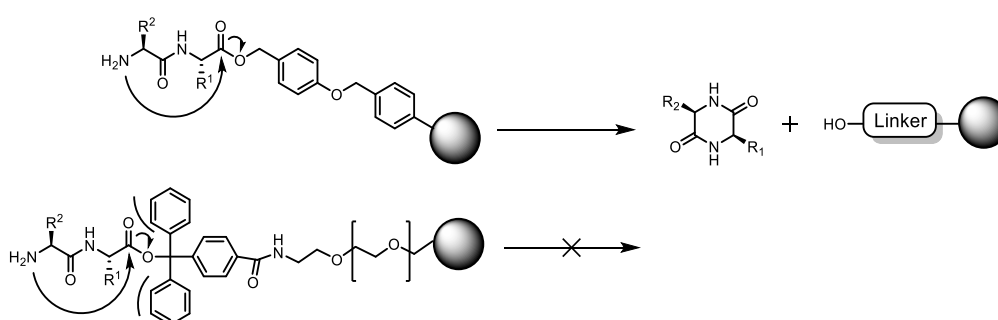


Figure 7. Structure of TentaGel® trityl resin. Figure adapted from Christian Pett PhD thesis, “Synthesis and Development of a Mucin Glycopeptide Microarray System for Evaluation of Protein- Interactions”, Technischen Universität Dortmund, Dortmund, Germany, 2016, doi.org/10.17877/DE290R-16906.

The PEG chain makes the reacting centre more accessible by increasing the polarity of the polymer and allowing polar solvent to swell the resin. Furthermore, TentaGel®-resins are available with different permanent groups on the PEG chain. The one used in the synthesis has a bulky triphenyl residue. The steric hindrance of the three aromatic rings help to prevent diketopiperazine formation, a common side reaction in solid phase synthesis, scheme 5.

Lastly, the resin can be purchased already functionalised with the first C-terminal Fmoc-protected amino acid, this is ideal when performing multiple peptide syntheses having all similar sequences.



Scheme 5. Diketopiperazine formation reaction mechanism. Diketopiperazine formation is inhibited in resins with a bulky linker.

3.5.5. Coupling

Typical coupling reagents are phosphonium, aminium/uronium salts or carbodiimides. Commonly used aminium/uronium salts are O-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), O-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TBTU) and O-(1H-7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU). Probably the most known phosphonium salt is benzotriazol-1-yloxytris(pyrrolidino)phosphonium hexafluorophosphate (PyBOP). Among the carbodiimides there are N, N'-dicyclohexylcarbodiimide (DCC) and N,N'-diisopropylcarbodiimide (DIC), figure 8.

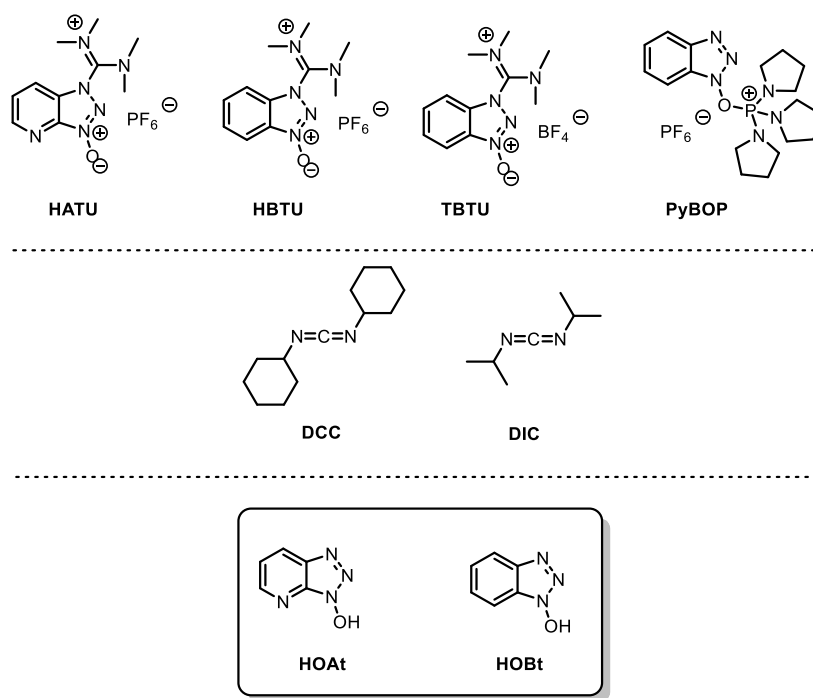


Figure 8. Chemical structures of coupling agents commonly used in peptide synthesis. In the box are shown additives, which can be used to further reduce the tendency for racemization.

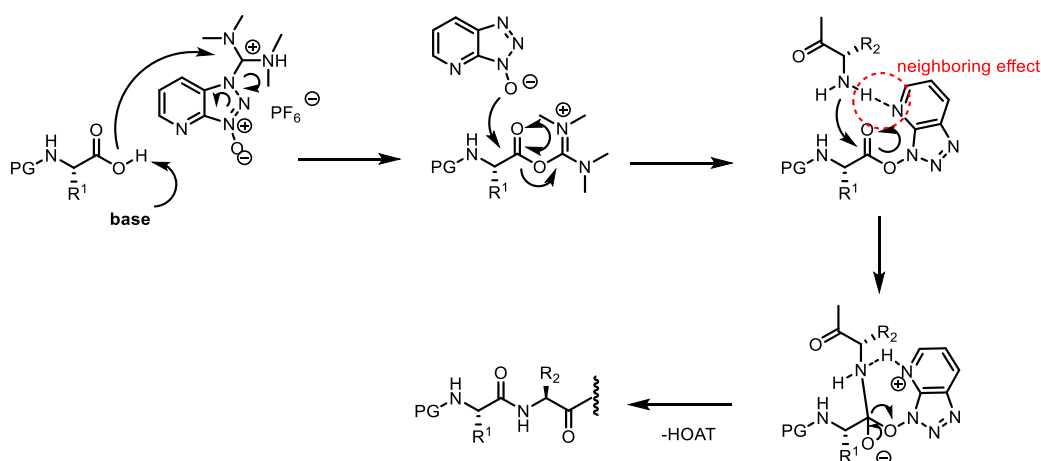
Phosphonium and aminium/uronium salt are more used compared to the carbodiimides due to their superior reactivity and lower tendency of racemization of the $C\alpha$.

To further reduce the tendency of racemization of the $C\alpha$, the benzotriazoles HOBT or HOAt are added as additives to the corresponding aminium/uronium salt. The HATU/HOAt-system is reported to be more effective in terms of reactivity and lower epimerization and is preferred in couplings with sterically demanding amino acids, such as glycosylated amino acids.⁵⁰

In the reaction involving an aminium/uronium salt activated coupling, the carboxylic acid of the entering amino acid is pre-activated with the coupling reagents and DIPEA. The base removes the acidic proton, and the nucleophilic oxygen attacks the electrophilic carbon of the guanidinium moiety to form an O-acylisouronium salt, scheme 6.

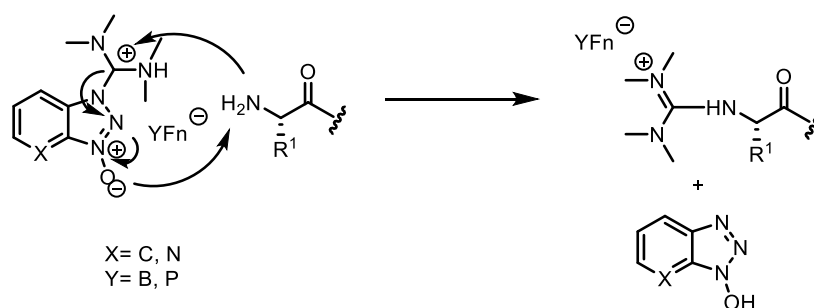
The produced HOAT/HOBT reacts to ultimately give an active ester intermediate.

HOAT coupling efficiency is higher compared to HOBT. The better reactivity is most probably attributed to the neighboring effect of the pyridine nitrogen which coordinates the amine of the approaching peptide as indicated in scheme 6. The final step of the reaction is the replacement of the benzotriazole by the amine of the approaching peptide which happens via the attack at the ester carbonyl group.



Scheme 6. Mechanism involving curly arrows of HATU mediated peptide coupling.

As you will notice in the description of the experimental condition in section 3.5.7, the coupling reagents are added in slight deficit compared to the amino acid. This is to avoid the tetramethylguanidinium capping of N-terminus of the growing peptide by the coupling reagent, scheme 7. The side products of this reaction can be detected by mass spectrometry as the truncation of target +101.



Scheme 7. Mechanism of capping reaction of guanidinium derivatives in case of excess use of of used aminium/uronium reagents

In synthesis with Fmoc protecting groups, when the coupling reagent of choice belongs to the carbodiimide family, DIC is preferable to DCC.

The byproduct of DCC coupling is dicyclohexylurea which in Fmoc synthesis is not dissolved by the solvent and is precipitated.

In fact, CEM patented a microwave assisted solid phase synthesis with all the reagents added in continuous flow. The coupling reagent they recommend is DIC and HOBt for synthesis with Fmoc.⁵¹

When a researcher is carrying out peptide synthesis for the first time, they may be confused by the fact that in different papers or books the HBTU or HATU coupling reagents can be named in interchangeable way as aminium or uronium salt. When these coupling reagents were first introduced, they were described as uronium salts.⁵²

This has persisted even though a few years later, Carpino demonstrated by X-ray diffraction of crystals obtained experimentally, that they structures are in their aminium (guanidinium N-oxide, see figure 9) form.⁵³

Despite the structural differences, both forms are active and can be selectively synthesized following different experimental procedures. It is worth mentioning, that under basic conditions, for instance in the presence of DIPEA, the O-form (uronium salt) isomerizes into the N-form (aminium salt).

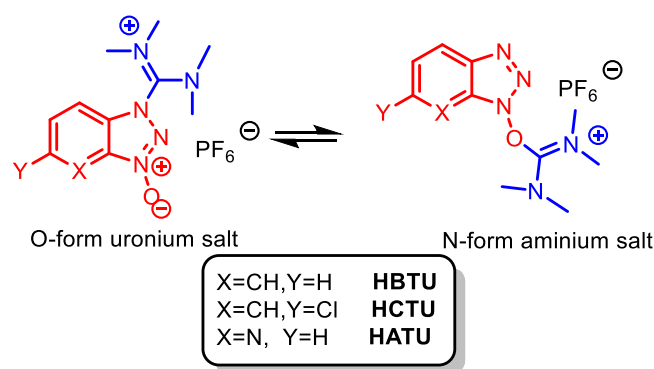
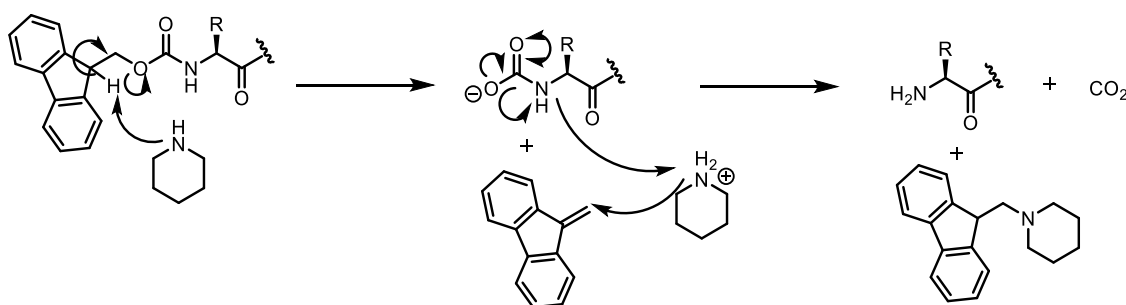


Figure 9. Chemical structure of the two possible form of guanidinium salts. Both the form are able to promote the coupling reaction but under basic condition the N-form is is the prevalent one.

N-Terminus deprotection

Another indispensable step in Fmoc based peptide synthesis is the removal of this protecting group by piperidine, scheme 8. This base is not only for removal of the acidic fluorene proton initializing the cleavage, but also acts as scavenger of the reaction side product benzofulvene, which could otherwise react with the N-terminus.



Scheme 8. Mechanism of Fmoc cleavage by piperidine. Piperidine react with benzofulvene.

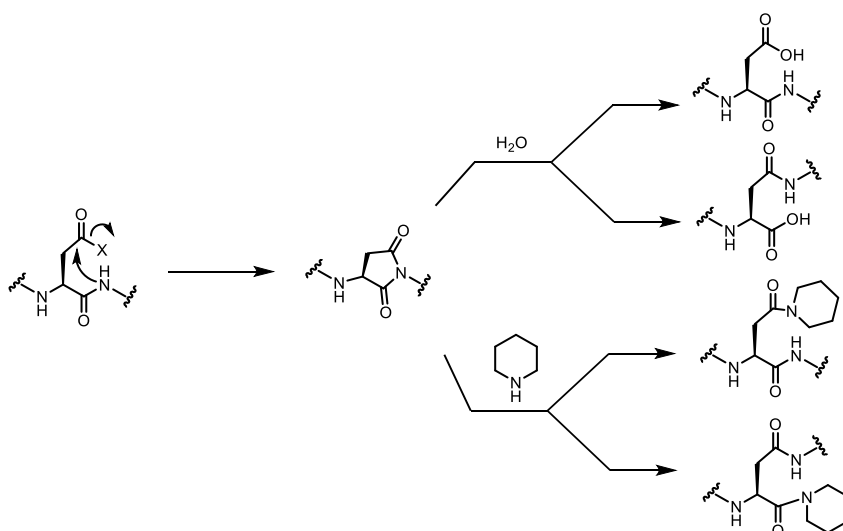
3.5.6. Side reactions in solid phase glycopeptide synthesis

Possible side products solid phase peptide synthesis are summarised in scheme 9 and 10. The ester protecting group of protected aspartates can give cyclic succinimides as intermediates (also called aspartimides). Aspartimide formation is highly sequence dependent, and is often reported that the most problematic sequences are peptides containing Asp(O t-Bu) followed by Asn(Trt) and Gly.⁵⁴

Aspartamide formation was also observed in the synthesis of Val-Lys-Asp-X-Tyr-Ile peptide when X = Arg(Pmc), Asn(Trt), Asp(OtBu), Cys(Acm), Gly, Ser, Thr and Thr(tBu).⁵⁵

Addition of HOBt, ethyl 2-cyano-2-(hydroxyimino)acetate (Oxyma),⁵⁶ or organic acid⁵⁷ in the deprotection solution can reduce the amount of aspartimide formation.

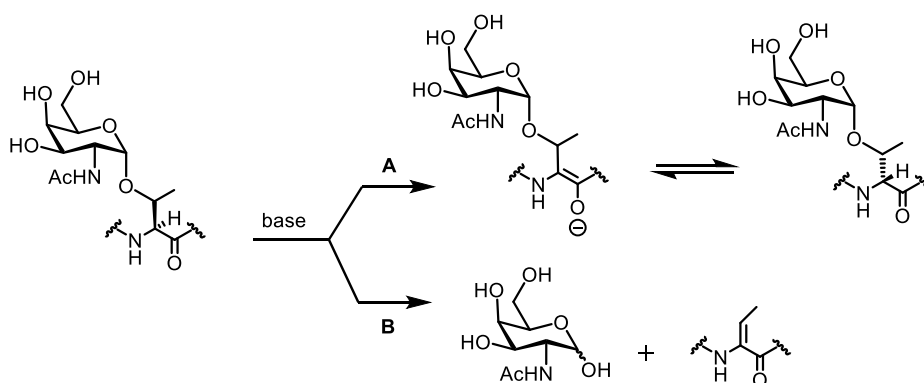
Reaction of water or piperidine with this intermediate, can generate side product α - and β -aspartidyl peptides or α - and β -piperidyl peptides respectively. Succinimide formation is highly sequence dependent and is often observed in the presence of a threonine residues followed by aspartate.



Scheme 9. Mechanism of aspartimide formation. Aspartamide can react with water during peptide cleavage or with piperidine during Fmoc-deprotection.

When the peptide includes glycosylated amino acids, additional side products can be attributed to the cleavage of the glycosidic bond or epimerisation of the stereogenic centres on $C\alpha$ of the amino acids. Even if the formation of this side product is reported to be extremely low, it is worth mentioning.

The use of a strong base, typically with sodium methoxide in methanol (often used for deacetylation of protected carbohydrates), and high pH for a prolonged period of time can cause the removal of the $C\alpha$ -hydrogen of serine and threonine and cause isomerisation of the amino acid, scheme 10 A, or initiate β -elimination of the glycan by an E1cB mechanism, scheme 10 A.



Scheme 10. Possible side products formation induced by high pH.

3.5.7. Synthesis of glycopeptide library based on MUC1 MUC4 epitope

During this thesis work, the MUC1-4 library was synthesized on solid phase, scheme 11, starting with preloaded Tentagel R Trt Fmoc-Ala, Fmoc-Asp and Fmoc-Val with low loading capacity, loading capacities of 0.17 mmol/g or 0.15 mmol/g.

The biotin tag required for peptide immobilisation on streptavidin coated plate needed in subsequent ELISA and for internalisation assay, was inserted either, at the beginning of the synthesis due to the coupling a biotinylated lysine, or at the end of the synthesis by direct coupling with biotin.

When MUC1 and 4 sequences were conjugated with P30 T Cell epitope, the two peptide sequences were spaced with a TEG spacer.

To mimic the microheterogeneity of tumour glycoproteins, the glycosylated serine and threonine were inserted in the synthesis and had different degrees of functionalisation.

To avoid peptide elongation problems due to the steric hindrance of the glycocluster, we coupled the amino acid **AG153** at the end of the peptide synthesis.

Despite the sequence differences and the glycosylation density all the peptides prepared can generally be divided in four classes depicted in figure 10, the positive and negative signs in the brackets indicate the envisioned binding with MGL.

- Class A: negative control (-), peptides which do not present carbohydrate antigens, this class of peptide is not expected to bind with MGL, but if a strong interaction between the peptide backbone and the protein is present, binding will be observed.
- Class B: Positive controls (+), peptides which present glycosylated serine and/or threonine, this class is expected to bind with MGL, the binding affinity is expected to be proportional to the number of glycosylated sites.
- Class C: peptides which present only the trivalent tetraphenylethylene glycocluster. The expected affinity for MGL may be superior to peptides from class B (++) due to the fact the trivalent tetraphenylethylene derivative as a K_d in the nM range for h-MGL.
- Class D: peptides which present glycosylated serine and/or threonine plus the trivalent tetraphenylethylene glycocluster. The expected affinity for MGL should be unchanged (++) from peptides of class C if the tetraphenylethylene based glycocluster occupies the binding site and no secondary interaction between the protein and glycosylated serine or threonine occurs. The affinity should be superior to peptides from class C (+++) if secondary interactions are occurring.

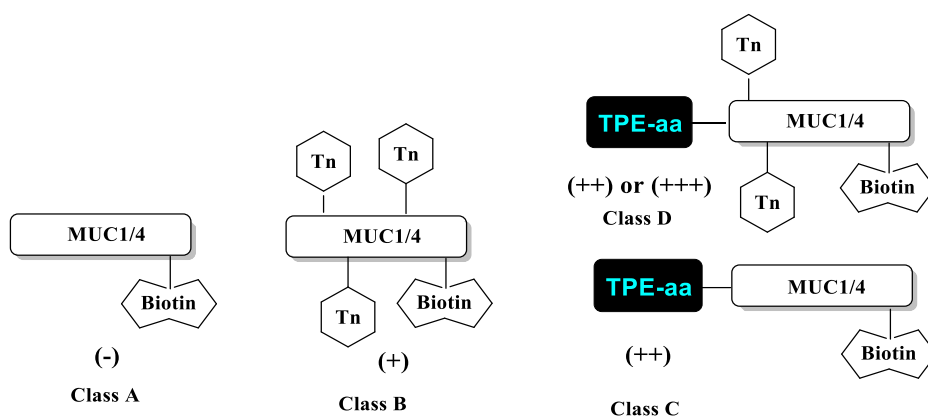


Figure 10. Simplified F various peptide classes synthesised.

The full set of peptides synthesised is reported in table 1. The MUC4 sequence, AS*T*GHAT*PLPVT*D, has four possible glycosylation sites. A shorter version of the MUC4 epitope was also adopted in the synthesis, GHAT*PLPVT*D. In this sequence the last three amino acid were removed, AST, and only two possible T_N-antigen are present (located at threonine, T).

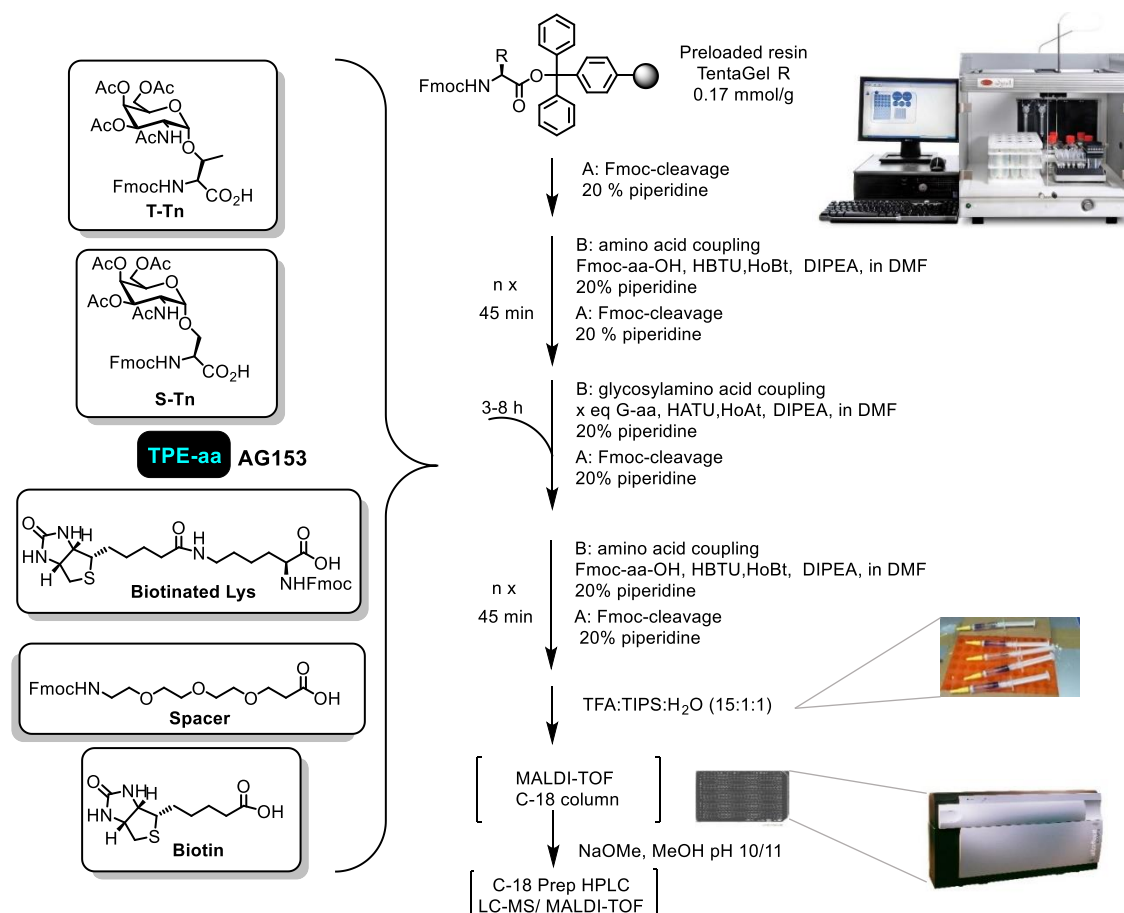
The MUC1 sequence, HGVT*S*APDT*RPAPGS*T*APPA, has five possible glycosylation sites.

An overview the full process of synthesis and purification is depicted in scheme 11. The synthesis started with Fmoc deprotection of the preloaded amino acid on the resin with 20% piperidine in DMF. Each coupling cycle was achieved using 8 equivalents of standard Fmoc amino acid and the HBTU/HOBt system. The capping step, typical in manual synthesis, is here avoided.

After each coupling cycle the Fmoc deprotection is repeated. The biotinylated lysine and the spacer (3 equivalents), scheme 11, were pre-activated manually with 2.9 equivalents HATU/HOAt and 5.8 eq of DIPEA in DMF in a separate vessel, added to the synthesis reactor and stirred for 3 hours before continuing the solid phase protocol.

On a similar fashion, the glycosylates amino acid serine and threonine T_N-antigen building blocks, 1.5 equivalents, were pre-activated manually with 1.4 eq HATU/HOAt and 2.8 equivalents of DIPEA in DMF in a separate vessel, added to the synthesis reactor. Also, **AG153** was coupled using HATU/HOAt system: thus, the multivalent antigen, 2 equivalents, was pre-activated manually using 1.8 eq

HATU/HOAt and 3.6 eq of DIPEA in DMF, added to the reactor and then stirred for 6h.



Scheme 11. Representation of synthesis, purification and characterisation of glycosylated MUC1-MUC4 peptides. AG153 synthesis is described in scheme 2. If the coupling with the TPE-aa was successful, when acidic cleavage cocatalyst is added to the resin, a strong violet color appeared: interesting experimental annotation !

Final Fmoc deprotection were done according to standard conditions, 20% piperidine in DMF.

Peptides 26-30 were capped with an N-terminal acetyl group using a solution of 0.015 M HOBt, 0.5 M Ac₂O, 0.125 M DIPEA in NMP on N-terminus for 45 min. The obtained peptides were then released from the resin, and all acid sensitive side-chain protecting groups were simultaneously removed using TFA/TIPS/H₂O 15:0.9:0.9 in 2.5 h followed by solvent concentration, lyophilization and desalted using a C-18 cartridge. To ensure the presence of the correct peptides, the mass was checked with MALDI-TOF.

For saccharide deprotection, the O-acetyl groups were cleaved by transesterification in methanol using catalytic amounts of NaOMe at pH 9-9.5 for 16-24 h, checked

with MALDI-TOF, to yield an array of glycosylated peptides reported in table 1, which were purified by preparative HPLC and lyophilised again.

All the peptide were characterised with analytical HPLC-MS (see chapter 6 for characterisation).

Table 1. Sequence of the peptides synthesised. Synthesis scale and yields are indicated in the final columns. **X**: amino acid preloaded on the resin; T*: threonine T_N antigen; S*: serine T_N antigen; **TPEaa**: AG153; Sp: TEG spacer, K^B: biotinated lysin.

Glycan	# AG0	Expected Affinity	Sequence	Scale μ mol	%
MUC1	01	(-)	HGVTSPDTRPAPGSTAPP A	12.5	54
	12	(-)	<i>Biotin</i> -FNNFTVSFWLRVPKVSASHLE- <i>Sp</i> -HGVTSPDTRPAPGSTAPP A	12.5	10
	16	(-)	FNNFTVSFWLRVPKVSASHLE- <i>Sp</i> -HGVTSPDTRPAPGSTAPP A	12.5	37
	21	(-)	<i>Biotin</i> -HGVTSPDTRPAPGSTAPP A	12.5	50
	26	(-)	<i>Ac</i> -HGVTSPDTRPAPGSTAPP A	12.5	56
	27	(-)	<i>Ac-Sp</i> -HGVTSPDTRPAPGSTAPP A	12.5	30
	28	(-)	<i>Ac</i> -HGVTSPDTRPAPGSTAPP A	12.5	43
	31	(-)	HGVTSPDTRPAPGSTAPP A	12.5	67
	34	(+)	HGVTSPDTRPAPGSTAPP A	12.5	38
	35	(++)/(+++)	TPEaa -HGVTSPDTRPAPGSTAPP A	7.5	20
	41	(+)	FNNFTVSFWLRVPKVSASHLE- <i>Sp</i> -HGVTSPDTRPAPGSTAPP A	12.5	6
	42	(++)/(+++)	TPEaa -FNNFTVSFWLRVPKVSASHLE- <i>Sp</i> -HGVTSPDTRPAPGSTAPP A	10	1
	50	(++)	TPEaa -HGVTSPDTRPAPGSTAPP A	7.5	29
	51	(++)	TPEaa - <i>Sp</i> -HGVTSPDTRPAPGSTAPP A	7.5	31
52	(-)	HGVTSPDTRPAPGSTAPP A	7.5	48	
P30 T-Cell	25	(-)	<i>Biotin</i> -FNNFTVSFWLRVPKVSASHLE E	12.5	39
	05	(-)	FNNFTVSFWLRVPKVSASHLE E	12.5	53
MUC4 long:	03	(-)	ASTGHATPLPVT D	12.5	78
	14	(-)	<i>Biotin</i> -FNNFTVSFWLRVPKVSASHLE- <i>Sp</i> -ASTGHATPLPVT D	12.5	41
	18	(-)	FNNFTVSFWLRVPKVSASHLE- <i>Sp</i> -ASTGHATPLPVT D	12.5	48
	23	(-)	<i>Biotin</i> -ASTGHATPLPVT D	12.5	55
	30	(-)	<i>Ac</i> -ASTGHATPLPVT A	12.5	76
	33	(++)	TPEaa -ASTGHATPLPVT A	7.5	49
	38	(+)	AS*T*GHAT*PLPVT A	12.5	23
	39	(++)/(+++)	TPEaa -AS*T*GHAT*PLPVT A	7.5	18
	45	(+)	FNNFTVSFWLRVPKVSASHLE- <i>Sp</i> -AS*T*GHAT*PLP V	12.5	3
	46	(++)/(+++)	TPEaa -FNNFTVSFWLRVPKVSASHLE- <i>Sp</i> -AS*T*GHAT*PLP V	10	1
MUC4 short:	04	(-)	GHATPLPVT D	12.5	80
	15	(-)	<i>Biotin</i> -FNNFTVSFWLRVPKVSASHLE- <i>Sp</i> -GHATPLPVT D	12.5	44
	19	(-)	FNNFTVSFWLRVPKVSASHLE- <i>Sp</i> -GHATPLPVT D	12.5	41
	20	(++)	TPEaa -GHATPLPVT D	12.5	21
	24	(-)	<i>Biotin</i> -GHATPLPVT D	12.5	70
	29	(-)	<i>Ac</i> -GHATPLPVT D	12.5	80
	32	(++)	TPEaa -GHATPLPVT D	7.5	44
	36	(+)	GHAT*PLPVT*DK A	12.5	76
	37	(++)/(+++)	TPEaa -GHAT*PLPVT*DK A	7.5	35
	40	(++)/(+++)	TPEaa -FNNFTVSFWLRVPKVSASHLE- <i>Sp</i> -GHAT*PLPVT*DK A	7.5	5
	43	(+)	FNNFTVSFWLRVPKVSASHLE- <i>Sp</i> -GHAT*PLPVT*DK D	12.5	12
	44	(++)/(+++)	TPEaa -FNNFTVSFWLRVPKVSASHLE- <i>Sp</i> -GHAT*PLPVT*DK D	10	4
	47	(+)	<i>Biotin</i> -FNNFTVSFWLRVPKVSASHLE- <i>Sp</i> -GHAT*PLPVT*DK D	12.5	18

3.5.8. ELISA investigation

A preliminary binding study was performed in the laboratory of Sandra Van Vliet in Vrije Universiteit Amsterdam by Joost C. van der Horst. The ELISA sandwich assay, figure 11, involves the incubation of a streptavidin-coated plate with the biotinylated glycopeptides, followed by treatment with human MGL decorated with Fc Tag. The presence of bound

MGL is then detected by anti-human Fc-peroxidase antibody. The results from this study are summarised in figures 12, 13 and 15. To help the reader following the discussion, the observed activity is indicated in the signs in the brackets.

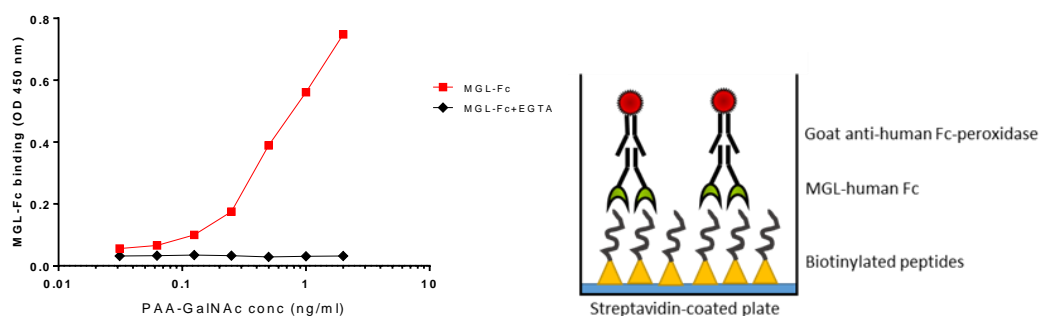
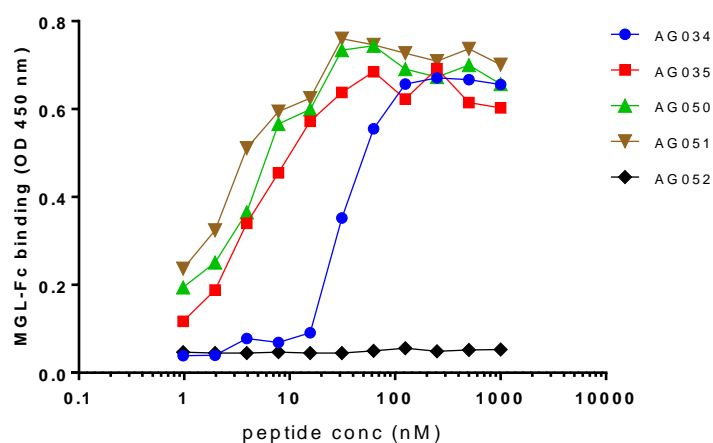


Figure 11. Left: binding curve of polyacrylamide coupled-GalNAc (PAA-GalNAc) in red, positive control and curve of the control in the presence of calcium chelating agent EGTA, black: when the concentration of the GalNAc polymer reaches 0.1 ng/mL start to inhibit the protein binding to the peptide coated surface. Right: graphical representation of the ELISA sandwich assay: on a streptavidin coated plate are linked the biotininated peptides. The protein on the plate surface is detected through the Fc tag with a goat ant-human Fc-peroxidase.

The IC_{50} values arising from the ELISA on MUC1 peptides are reported in figure 12. The peptides bearing the glycocluster **AG153** clearly outperformed the peptide with T_N only, **AG034 (+)**. Comparing **AG035 (++)** with **AG050 (++)** and **AG051 (++)**, is possible to evince that the affinity is conserved even in the absence of T_N antigens in the sequence, is due to attributed to the trivalent glycocluster. A slight improvement in the IC_{50} , from 3.90 for **AG50 (++)**, to 2.65 for **AG051 (++)**, is observed when the spacer separate the MUC1 chain from the glycocluster. No affinity is attributed to the peptide epitope as proved by the absence of any binding for **AG052 (-)**.



Glycan	# AG0	Expected affinity	Observed affinity	IC ₅₀ nM	Sequence
MUC1	34	(+)	(+)	32.12	HGVT SAPDT*RPAPGS*T*APK ^{B/A}
	35	(+)/(+++)	(++)	3.75	TPE ^{aa} -HGVT SAPDT*RPAPGS*T*APK ^{B/A}
	50	(++)	(++)	3.90	TPE ^{aa} -HGVT SAPDTRPAPGSTAPK ^{B/A}
	51	(++)	(++)	2.65	TPE ^{aa} -Sp-HGVT SAPDTRPAPGSTAPK ^{B/A}
	52	(-)	(-)	--	HGVT SAPDTRPAPGSTAPK ^{B/A}

Figure 12. ELISA assay peptides with MUC1 epitope.

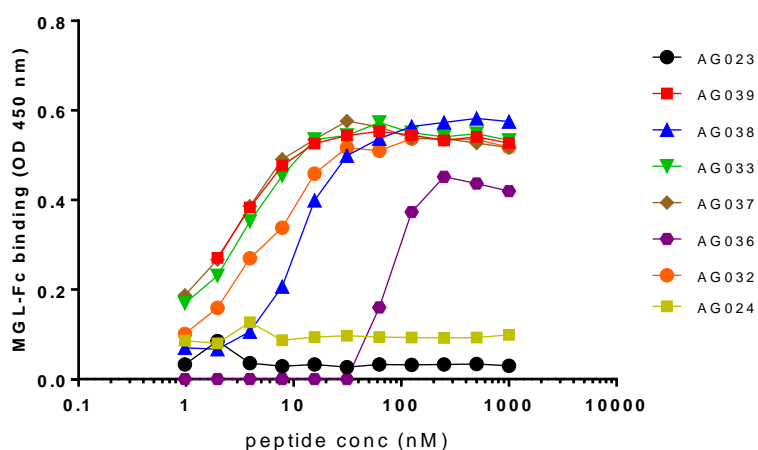
The IC₅₀ values arising from the ELISA on MUC4 peptides are reported in figure 13.

Two different length of mucin 4 sequences were analysed: a long one, ASTGHATPLPVT D, and a short one, GHATPLPVT D. The short MUC4 is lacking three amino acids in the C-terminus side, alanine-serine-threonine.

The short MUC4 sequences bearing the synthetic multivalent antigen, **AG032** (++) and **AG037** (+++) highly outperform the peptide with Tn only, **AG036** (+).

In the MUC4 long glycopeptides is possible noticing an interesting trend: **AG038** (+), bearing only natural antigens, binds to MGL-Fc much better than the respective peptide in the short MUC4 sequence **AG136** (+), IC₅₀ respectively 11.38 nM and 72.60 nM.

The affinity resembles the one of the glyocluster-containing peptides **AG033** (++) and **AG039** (+++), with minor differences.



Glycan	# AG0	Expected affinity	Observed affinity	IC ₅₀ nM	Sequence
MUC4:	23	(-)	(-)	--	Biotin-ASTGHATPLPVT D
LONG	33	(++)	(++)	2.98	TPE ^{aa} -ASTGHATPLPVK ^{B/A}
	38	(+)	(+)	11.38	AS*T*GHAT*PLPVK ^{B/A}
	39	(+)/(+++)	(++)	2.65	TPE ^{aa} -AS*T*GHAT*PLPVK ^{B/A}
MUC4:	24	(-)	(-)	--	Biotin-GHATPLPVT D
SHORT	32	(++)	(++)	5.32	TPE ^{aa} -GHATPLPVT DK ^{B/A}

36	(+)	(+)	72.60	GHAT*PLPVT*DK ^B A
37	(++)/(+++)	(+++)	2.33	TP ^{Eaa} -GHAT*PLPVT*DK ^B A

Figure 13. ELISA assay peptides with MUC4 epitope.

It is possible that the enhancement in affinity of **AG038** compared to **AG036** could arise simply by the peptide valency: the first one bears three antigens while the second one only two. The presentation of three antigens did not caused similar affinity enhancement in the MUC1 glycopeptide **AG034**.

An alternative reasoning of the observed behaviour could be due to the interaction of the MUC4 long peptide backbone with h-MGLsequence.

Recently, water mediated interaction between between carbonyl (CO) of a Thr residue of MUC1 glycopeptides and the polar nitrogen N ϵ of the aromatic ring of His286 were derived from 200-ns MD simulations, figure 14.⁵⁸ This type of water mediate interaction is solely possible if the antigen orientation correspond to a binding mode of type A, see chapter 5.

The glycosylated threonine at the end of the MUC4 long sequence, could have similar interaction with the same histidine or another side chain located nearby the carbohydrate recognition domain. The trend in MUC1 and MUC4 short glycopeptides, suggests that, in peptides carrying the synthetic antigen, that this antigen is mostly responsible for the binding. For this reason no enhancement of the binding is observed for the peptide **AG039**. The importance of MGL His286 for ligand recognition and binding was also shown in this PhD work, which is discussed in chapter 5.

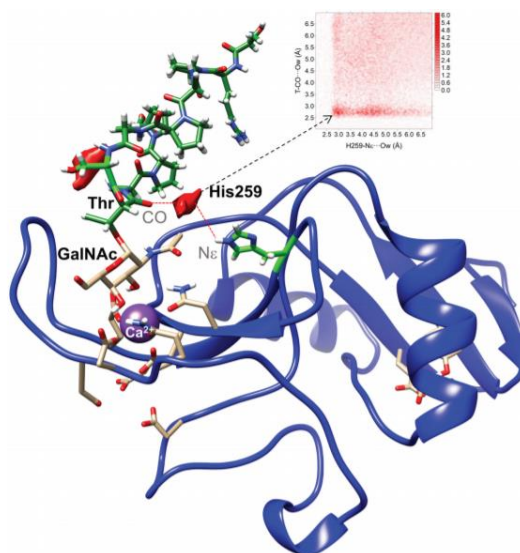
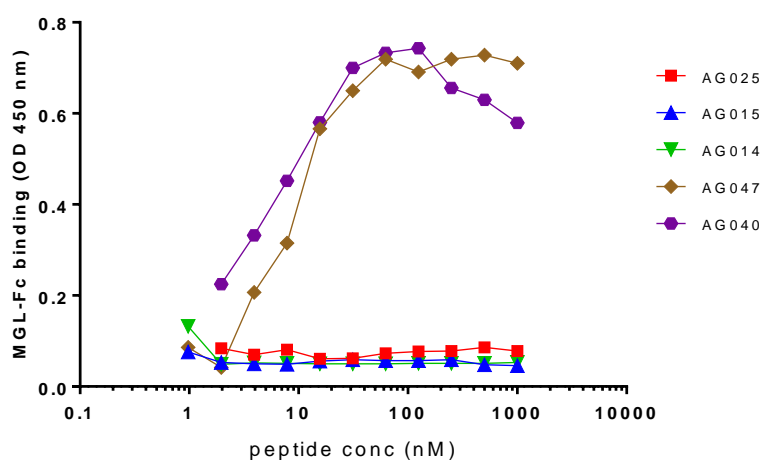


Figure 14. Image from *J. Biol. Chem.* (2019) 294(4) 1300–1311. Water-mediated binding between Tn-glycopeptide MUC1 (H2N-TRPAPGST*APPA-NH2) and His286 of MGL. Shown is the 2D-rdf function derived from 200-ns MD simulations between carbonyl (CO) of the glycosylated Thr and the polar nitrogen N ϵ of the aromatic ring of His286 (top), together with the water density (shown in red) calculated for the MGL-His286/MUC1– derived Tn-glycopeptide complex. No water density was detected in the case of mutant MGL-Thr286/MUC1-derived Tn-glycopeptide complex.⁵⁸

Eventually, the combination of P30 TCell and MUC4 short epitopes, figure 15, does not improve significantly the IC_{50} values compare to the peptides with only MUC4 epitopes, but completely different immunisation outcome could be generated by these sequences in the future.

Furthermore no differences were observed between **AG040 (+)** and **AG047 (+)** with or without **AG153**.



Glycan	#	Expected affinity	Observed affinity	IC_{50} nM	Sequence
MUC4:	14	(-)	(-)	--	<i>Biotin-FNNFTVSFWLRVPKVSASHLE-Sp-ASTGHATPLPVT</i> ^D
MUC4:	15	(-)	(-)	--	<i>Biotin-FNNFTVSFWLRVPKVSASHLE-Sp-GHATPLPVT</i> ^D
	40	(+)/(+++)	(+)	5.60	<i>TP^Eead-FNNFTVSFWLRVPKVSASHLE-Sp-GHAT*PLPVT*DK^BA</i>
	47	(+)	(+)	7.92	<i>Biotin-FNNFTVSFWLRVPKVSASHLE-Sp-GHAT*PLPVT*^D</i>
P30 T-Cell	25	(-)	(-)	--	<i>Biotin-FNNFTVSFWLRVPKVSASHL</i> ^E

Figure 15. ELISA assay peptides with MUC4 epitope.

3.6. Conclusion

With the aim to obtain a glycopeptides able to produce IgG antibody and B memory cells, an extensive library of synthetic antigens comprised of MUC1/MUC4 glycopeptides was synthesised by solid phase peptide synthesis. To enhance the antigen internalisation and presentation to T cells via MHC-II, indispensable for activating the T cell-dependent pathway, we decided to engage DCs and macrophages by aiming to target an endocytic receptor located on their surfaces, hMGL.

To do so, we developed a synthesis which allowed us to locate a trivalent glycocluster known to bind hMGL in the nM range, scheme 2, on the side chain of an asparagine residue, which was then inserted in sequence of MUC1/4 peptides.

Preliminary binding study showed that in case of MUC1 and MUC4-short peptides the presence of the trivalent glycocluster highly enhanced the binding with hMGL compared with the positive controls where only Ser/Thr T_N-antigens were present.

Binding data arising from MUC4-long peptides, suggest a possible interaction of hMGL with their peptide backbones.

Immunisation experiments with the best candidates identified herein will be performed in the future.

3.7. Bibliography

- (1) Zaretsky, J. Z.; Barnea, I.; Aylon, Y.; Gorivodsky, M.; Wreschner, D. H.; Keydar, I. MUC1 Gene Overexpressed in Breast Cancer: Structure and Transcriptional Activity of the MUC1 Promoter and Role of Estrogen Receptor Alpha (ER α) in Regulation of the MUC1 Gene Expression. *Mol. Cancer* **2006**, *5*, 1–14. <https://doi.org/10.1186/1476-4598-5-57>.
- (2) Springer, G. F. T and Tn , General Carcinoma Autoantigens. *Science (80-.)*. **1984**, *224* (4654), 1198–1206.
- (3) Napoletano, C.; Rughetti, A.; Agervig Tarp, M. P.; Coleman, J.; Bennett, E. P.; Picco, G.; Sale, P.; Denda-Nagai, K.; Irimura, T.; Mandel, U.; et al. Tumor-Associated Tn-MUC1 Glycoform Is Internalized through the Macrophage Galactose-Type C-Type Lectin and Delivered to the HLA Class I and II Compartments in Dendritic Cells. *Cancer Res.* **2007**, *67* (17), 8358–8367. <https://doi.org/10.1158/0008-5472.CAN-07-1035>.
- (4) Wilson, R. M.; Danishefsky, S. J. A Vision for Vaccines Built from Fully Synthetic Tumor-Associated Antigens: From the Laboratory to the Clinic. *J. Am. Chem. Soc.* **2013**, *135* (39), 14462–14472. <https://doi.org/10.1021/ja405932r>.
- (5) Avci, F. Y.; Kasper, D. L. How Bacterial Carbohydrates Influence the Adaptive Immune System. *Annu. Rev. Immunol.* **2009**, *28* (1), 107–130. <https://doi.org/10.1146/annurev-immunol-030409-101159>.
- (6) Gaidzik, N.; Westerlind, U.; Kunz, H. The Development of Synthetic Antitumour Vaccines from Mucin Glycopeptide Antigens. *Chem. Soc. Rev.* **2013**, *42* (10), 4421–4442. <https://doi.org/10.1039/c3cs35470a>.
- (7) Kaiser, A.; Gaidzik, N.; Westerlind, U.; Kowalczyk, D.; Hobel, A.; Schmitt, E.; Kunz, H. A Synthetic Vaccine Consisting of a Tumor-Associated Sialyl-TNMUC1 Tandem-Repeat Glycopeptide and Tetanus Toxoid: Induction of a Strong and Highly Selective Immune Response. *Angew. Chemie - Int. Ed.* **2009**, *48* (41), 7551–7555. <https://doi.org/10.1002/anie.200902564>.
- (8) Westerlind, U.; Hobel, A.; Gaidzik, N.; Schmitt, E.; Kunz, H. Synthetic Vaccines Consisting of Tumor-Associated MUC1 Glycopeptide Antigens and a T-Cell Epitope for the Induction of a Highly Specific Humoral Immune Response. *Angew. Chemie - Int. Ed.* **2008**, *47* (39), 7551–7556. <https://doi.org/10.1002/anie.200802102>.
- (9) Hoffmann-Röder, A.; Kaiser, A.; Wagner, S.; Gaidzik, N.; Kowalczyk, D.; Westerlind, U.; Gerlitzki, B.; Schmitt, E.; Kunz, H. Synthetic Antitumor Vaccines from Tetanus Toxoid Conjugates of MUC1 Glycopeptides with the Thomsen-Friedenreich Antigen and a Fluorine-Substituted Analogue. *Angew. Chemie - Int. Ed.* **2010**, *49* (45), 8498–8503. <https://doi.org/10.1002/anie.201003810>.
- (10) Tacke, P. J.; De Vries, I. J. M.; Torensma, R.; Figdor, C. G. Dendritic-Cell Immunotherapy: From Ex Vivo Loading to in Vivo Targeting. *Immunity Rev. Immunol.* **2007**, *7* (10), 790–802. <https://doi.org/10.1038/nri2173>.
- (11) Tacke, P. J.; De Vries, I. J. M.; Gijzen, K.; Joosten, B.; Wu, D.; Rother, R. P.; Faas, S. J.; Punt, C. J. A.; Torensma, R.; Adema, G. J.; et al. Effective Induction of Naive and Recall T-Cell Responses by Targeting Antigen to Human Dendritic Cells via a Humanized Anti-DC-SIGN Antibody. *Blood* **2005**, *106* (4), 1278–1285. <https://doi.org/10.1182/blood-2005-01-0318>.
- (12) Mahnke, K.; Guo, M.; Lee, S.; Sepulveda, H.; Swain, S. L.; Nussenzweig, M.; Steinman, R. M. The Dendritic Cell Receptor for Endocytosis, DEC-205, Can Recycle and Enhance Antigen Presentation via Major Histocompatibility Complex Class II-Positive Lysosomal Compartments. *J. Cell Biol.* **2000**, *151* (3), 673–683. <https://doi.org/10.1083/jcb.151.3.673>.
- (13) Keler, T.; Ramakrishna, V.; Fanger, M. W. Mannose Receptor-Targeted Vaccines. *Expert Opin. Biol. Ther.* **2004**, *4* (12), 1953–1962.
- (14) Apostolopoulos, V.; Pietersz, G. A.; Tsibanis, A.; Tsikkinis, A.; Drakaki, H.; Loveland, B. E.; Piddlesden, S. J.; Plebanski, M.; Pouniotis, D. S.; Alexis, M. N.; et al. Pilot Phase III Immunotherapy Study in Early-Stage Breast Cancer Patients Using Oxidized Mannan-MUC1 [ISRCTN71711835]. *Breast Cancer Res.* **2006**, *8* (3), 1–11. <https://doi.org/10.1186/bcr1505>.
- (15) Karanikas, V.; Hwang, L. A.; Pearson, J.; Ong, C. S.; Apostolopoulos, V.; Vaughan, H.; Xing, P. X.;

- Jamieson, G.; Pietersz, G.; Tait, B.; et al. Antibody and T Cell Responses of Patients with Adenocarcinoma Immunized with Mannan-MUC1 Fusion Protein. *J. Clin. Invest.* **1997**, *100* (11), 2783–2792. <https://doi.org/10.1172/JCI119825>.
- (16) Wu, X.; Yin, Z.; McKay, C.; Pett, C.; Yu, J.; Schorlemer, M.; Gohl, T.; Sungsuwan, S.; Ramadan, S.; Baniel, C.; et al. Protective Epitope Discovery and Design of MUC1-Based Vaccine for Effective Tumor Protections in Immunotolerant Mice. *J. Am. Chem. Soc.* **2018**, *140*, jacs.8b08473. <https://doi.org/10.1021/jacs.8b08473>.
- (17) Yin, Z.; Wu, X.; Kaczanowska, K.; Sungsuwan, S.; Comellas Aragones, M.; Pett, C.; Yu, J.; Baniel, C.; Westerlind, U.; Finn, M. G.; et al. Antitumor Humoral and T Cell Responses by Mucin-1 Conjugates of Bacteriophage Q β in Wild-Type Mice. *ACS Chem. Biol.* **2018**, *13* (6), 1668–1676. <https://doi.org/10.1021/acscchembio.8b00313>.
- (18) Fagarasan, S. T-Independent Immune Response: New Aspects of B Cell Biology. *Science* (80-.). **2002**, *290* (5489), 89–92. <https://doi.org/10.1126/science.290.5489.89>.
- (19) Parker, D. C. T Cell-Dependent B Cell Activation. In *Encyclopedia of Immunobiology*; 2016; Vol. 3, pp 175–178. <https://doi.org/10.1016/B978-0-12-374279-7.09010-X>.
- (20) Sprooten, J.; Ceusters, J.; Coosemans, A.; Agostinis, P.; De Vleeschouwer, S.; Zitvogel, L.; Kroemer, G.; Galluzzi, L.; Garg, A. D. Trial Watch: Dendritic Cell Vaccination for Cancer Immunotherapy. *Oncoimmunology* **2019**, *00* (00), 1–21. <https://doi.org/10.1080/2162402X.2019.1638212>.
- (21) Freire, T.; Lo-Man, R.; Bay, S.; Leclerc, C. Tn Glycosylation of the MUC6 Protein Modulates Its Immunogenicity and Promotes the Induction of Th17-Biased T Cell Responses. *J. Biol. Chem.* **2011**, *286* (10), 7797–7811. <https://doi.org/10.1074/jbc.M110.209742>.
- (22) Lo-man, R.; Vichier-guerre, S.; Perraut, R.; De, E.; Benmohamed, L.; Diop, O. M.; Livingston, P. O.; Bay, S.; Leclerc, C. A Fully Synthetic Therapeutic Vaccine Candidate Targeting Carcinoma-Associated Tn Carbohydrate Antigen Induces Tumor-Specific Antibodies in Nonhuman Primates. *Cancer Res.* **2004**, No. 29, 4987–4994.
- (23) Freire, T.; Zhang, X.; Dériaud, E.; Ganneau, C.; Vichier-Guerre, S.; Azria, E.; Launay, O.; Lo-Man, R.; Bay, S.; Leclerc, C. Glycosidic Tn-Based Vaccines Targeting Dermal Dendritic Cells Favor Germinal Center B-Cell Development and Potent Antibody Response in the Absence of Adjuvant. *Blood* **2010**, *116* (18), 3526–3536. <https://doi.org/10.1182/blood-2010-04-279133>.
- (24) Laubreton, D.; Bay, S.; Sedlik, C.; Artaud, C.; Ganneau, C.; Dériaud, E.; Viel, S.; Puaux, A. L.; Amigorena, S.; Gérard, C.; et al. The Fully Synthetic MAG-Tn3 Therapeutic Vaccine Containing the Tetanus Toxoid-Derived TT830-844 Universal Epitope Provides Anti-Tumor Immunity. *Cancer Immunol. Immunother.* **2016**, *65* (3), 315–325. <https://doi.org/10.1007/s00262-016-1802-0>.
- (25) van Kooyk, Y.; Ilarregui, J. M.; van Vliet, S. J. Novel Insights into the Immunomodulatory Role of the Dendritic Cell and Macrophage-Expressed C-Type Lectin MGL. *Immunobiology* **2015**, *220* (2), 185–192. <https://doi.org/10.1016/j.imbio.2014.10.002>.
- (26) Dekker, J.; Rossen, J. W. A.; Büller, H. A.; Einerhand, A. W. C. The MUC Family: An Obituary. *Trends Biochem. Sci.* **2002**, *27* (3), 126–131. [https://doi.org/10.1016/S0968-0004\(01\)02052-7](https://doi.org/10.1016/S0968-0004(01)02052-7).
- (27) Kufe, D. W. Mucins in Cancer: Function, Prognosis and Therapy. *Nat. Rev. Cancer* **2009**, *9* (12), 874–885. <https://doi.org/10.1038/nrc2761>.
- (28) Degand, P.; Porchet, N.; Aubert, J.; Devisme, L.; Buisine, M.; Copin, M.; Gosselin, B.; Dieu, M. Developmental Mucin Gene Expression in the Gastrointestinal Tract and Accessory Digestive Glands. II. Duodenum and Liver, Gallbladder, and Pancreas. *J. Histochem. Cytochem.* **2011**, *48* (12), 1667–1676. <https://doi.org/10.1177/002215540004801210>.
- (29) Gendler, S. Epithelial Mucin Genes. *Annu. Rev. Physiol.* **2002**, *57* (1), 607–634. <https://doi.org/10.1146/annurev.physiol.57.1.607>.
- (30) Moniaux, N.; Andrianifahanana, M.; Brand, R. E.; Batra, S. K. Multiple Roles of Mucins in Pancreatic Cancer, a Lethal and Challenging Malignancy. *British Journal of Cancer.* 2004, pp 1633–1638. <https://doi.org/10.1038/sj.bjc.6602163>.
- (31) Cabello, C. M.; Bair, W. B.; Lamore, S. D.; Ley, S.; Alexandra, S.; Azimian, S.; Wondrak, G. T.

- Silencing of Germline-Expressed Genes by DNA Elimination in Somatic Cells. *Dev. Cell* **2010**, *46* (2), 220–231. <https://doi.org/10.1016/j.freeradbiomed.2008.10.025>.The.
- (32) Arul, G. S.; Moorghen, M.; Myerscough, N.; Alderson, D. A.; Spicer, R. D.; Corfield, A. P. Mucin Gene Expression in Barrett's Oesophagus: An in Situ Hybridisation and Immunohistochemical Study. *Gut* **2000**, *47* (6), 753–761. <https://doi.org/10.1136/gut.47.6.753>.
- (33) Llinares, K.; Escande, F.; Aubert, S.; Buisine, M. P.; De Bolos, C.; Batra, S. K.; Gosselin, B.; Aubert, J. P.; Porchet, N.; Copin, M. C. Diagnostic Value of MUC4 Immunostaining in Distinguishing Epithelial Mesothelioma and Lung Adenocarcinoma. *Mod. Pathol.* **2004**, *17* (2), 150–157. <https://doi.org/10.1038/modpathol.3800027>.
- (34) Taylor-Papadimitriou, J.; Burchell, J.; Miles, D. W.; Dalziel, M. MUC1 and Cancer. *Biochimica et Biophysica Acta - Molecular Basis of Disease*. 1999, pp 301–313. [https://doi.org/10.1016/S0925-4439\(99\)00055-1](https://doi.org/10.1016/S0925-4439(99)00055-1).
- (35) Deng, J.; Wang, L.; Chen, H.; Li, L.; Ma, Y.; Ni, J.; Li, Y. The Role of Tumour-Associated MUC1 in Epithelial Ovarian Cancer Metastasis and Progression. *Cancer and Metastasis Reviews*. 2013, pp 535–551. <https://doi.org/10.1007/s10555-013-9423-y>.
- (36) Eggink, L. L.; Roby, K. F.; Cote, R.; Kenneth Hooper, J. An Innovative Immunotherapeutic Strategy for Ovarian Cancer: CLEC10A and Glycomimetic Peptides. *J. Immunother. Cancer* **2018**, *6* (1), 1–16. <https://doi.org/10.1186/s40425-018-0339-5>.
- (37) Hinoda, Y.; Ikematsu, Y.; Horinouchi, M.; Sato, S.; Yamamoto, K.; Nakano, T.; Fukui, M.; Suehiro, Y.; Hamanaka, Y.; Nishikawa, Y.; et al. Increased Expression of MUC1 in Advanced Pancreatic Cancer. *J. Gastroenterol.* **2003**, *38* (12), 1162–1166. <https://doi.org/10.1007/s00535-003-1224-6>.
- (38) O'Connor, J. C.; Julian, J.; Lim, S. D.; Carson, D. D. MUC1 Expression in Human Prostate Cancer Cell Lines and Primary Tumors. *Prostate Cancer Prostatic Dis.* **2005**, *8* (1), 36–44. <https://doi.org/10.1038/sj.pcan.4500762>.
- (39) van Vliet, S. J.; Saeland, E.; van Kooyk, Y. Sweet Preferences of MGL: Carbohydrate Specificity and Function. *Trends Immunol.* **2008**, *29* (2), 83–90. <https://doi.org/10.1016/j.it.2007.10.010>.
- (40) van Vliet, S. J.; van Liempt, E.; Saeland, E.; Aarnoudse, C. A.; Appelmelk, B.; Irimura, T.; Geijtenbeek, T. B. H.; Blixt, O.; Alvarez, R.; van Die, I.; et al. Carbohydrate Profiling Reveals a Distinctive Role for the C-Type Lectin MGL in the Recognition of Helminth Parasites and Tumor Antigens by Dendritic Cells. *Int. Immunol.* **2005**, *17* (5), 661–669. <https://doi.org/10.1093/intimm/dxh246>.
- (41) Marcelo, F.; Garcia-Martin, F.; Matsushita, T.; Sardinha, J.; Coelho, H.; Oude-Vrielink, A.; Koller, C.; André, S.; Cabrita, E. J.; Gabius, H. J.; et al. Delineating Binding Modes of Gal/GalNAc and Structural Elements of the Molecular Recognition of Tumor-Associated Mucin Glycopeptides by the Human Macrophage Galactose-Type Lectin. *Chem. - A Eur. J.* **2014**, *20* (49), 16147–16155. <https://doi.org/10.1002/chem.201404566>.
- (42) Mortezaei, N.; Behnken, H. N.; Kurze, A. K.; Ludewig, P.; Buck, F.; Meyer, B.; Wagener, C. Tumor-Associated Neu5Ac-Tn and Neu5Gc-Tn Antigens Bind to C-Type Lectin CLEC10A (CD301, MGL). *Glycobiology* **2013**, *23* (7), 844–852. <https://doi.org/10.1093/glycob/cwt021>.
- (43) Heger, L.; Balk, S.; Lühr, J. J.; Heidkamp, G. F.; Lehmann, C. H. K.; Hatscher, L.; Purbojo, A.; Hartmann, A.; Garcia-Martin, F.; Nishimura, S. I.; et al. CLEC10A Is a Specific Marker for Human CD1c+dendritic Cells and Enhances Their Toll-like Receptor 7/8-Induced Cytokine Secretion. *Front. Immunol.* **2018**, *9* (APR), 1–16. <https://doi.org/10.3389/fimmu.2018.00744>.
- (44) Kaltner, H.; Manning, J. C.; García Caballero, G.; Di Salvo, C.; Gabba, A.; Romero-Hernández, L. L.; Knospe, C.; Wu, D.; Daly, H. C.; O'Shea, D. F.; et al. Revealing Biomedically Relevant Cell and Lectin Type-Dependent Structure-Activity Profiles for Glycoclusters by Using Tissue Sections as an Assay Platform. *RSC Adv.* **2018**, *8* (50), 28716–28735. <https://doi.org/10.1039/c8ra05382k>.
- (45) Ahmad Fuaad, A. A. H.; Azmi, F.; Skwarczynski, M.; Toth, I. Peptide Conjugation via CuAAC “click” Chemistry. *Molecules* **2013**, *18* (11), 13148–13174. <https://doi.org/10.3390/molecules181113148>.
- (46) Paulsen, H.; Holck, J.-P. Synthese Der Glycopeptide O-β-d-Galactopyranosyl-(1→3)-O-(2-

- Acetamido-2-Desoxy- α -D-Galactopyranosyl)-(1 \rightarrow 3)-L-Serin Und -L-Threonin. *Carbohydr. Res.* **1982**, *109*, 89–107.
- (47) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154. <https://doi.org/10.1021/ja00897a025>.
- (48) Carpino, L. A.; Han, G. Y. The 9-Fluorenylmethoxycarbonyl Amino-Protecting Group. *J. Org. Chem.* **1979**, *44* (21), 3739. <https://doi.org/10.1021/jo01335a600>.
- (49) Atherton, E.; Fox, H.; Harkiss, D.; Logan, C. J.; Sheppard, R. C.; Williams, B. J. A Mild Procedure for Solid Phase Peptide Synthesis: Use Of. *J. Chem. Soc. Chemical Commun.* **1978**, *1978*, 537–539.
- (50) I, G. S.; Mulinacci, B.; I, M. C. A.; Chelli, M.; Rovero, P.; Fi, F.; Farmaceutiche, S.; Salerno, U.; Sa, F. Assessment of New 6-Cl-HOBt Based Coupling Reagents for Peptide Synthesis. Part 1: Coupling Efficiency Study. *Lett. Pept. Sci.* **2002**, *9*, 119–123.
- (51) Vanier, G. S. Peptide Synthesis and Applications. *Pept. Synth. Appl.* **2005**, *1047*, 235–249. <https://doi.org/10.1385/1592598773>.
- (52) Abdelmoty, I.; Albericio, F.; Carpino, L. A.; Foxman, B. M.; Kates, S. A. Structural Studies of Reagents for Peptide Bond Formation: Crystal and Molecular Structures of HBTU and HATU. *Lett. Pept. Sci.* **1994**, *1* (2), 57–67. <https://doi.org/10.1007/BF00126274>.
- (53) Carpino, L. A.; Imazumi, H.; Ferrer, F. J.; Zhang, C.; Lee, Y.; Foxman, B. M.; Henklein, P.; Hanay, C.; Clemens, M.; Wenschuh, H.; et al. The Uronium/Guanidinium Peptide Coupling Reagents: *Angew. Chemie Int. Ed.* **2002**, *41* (3), 441–445.
- (54) Yang, Y.; Sweeney, W. V.; Schneider, K.; Thörnqvist, S.; Chait, B. T.; Tam, J. P. Aspartimide Formation in Base-Driven 9-Fluorenylmethoxycarbonyl Chemistry. *Tetrahedron Lett.* **1994**, *35* (52), 9689–9692. [https://doi.org/10.1016/0040-4039\(94\)88360-2](https://doi.org/10.1016/0040-4039(94)88360-2).
- (55) Lauer, J. L.; Fields, C. G.; Fields, G. B. Sequence Dependence of Aspartimide Formation during 9-Fluorenylmethoxycarbonyl Solid-Phase Peptide Synthesis. *Lett. Pept. Sci.* **1995**, *1* (4), 197–205. <https://doi.org/10.1007/BF00117955>.
- (56) Subirós-Funosas, R.; El-Faham, A.; Albericio, F. Use of Oxyma as PH Modulatory Agent to Be Used in the Prevention of Base-Driven Side Reactions and Its Effect on 2-Chlorotriyl Chloride Resin. *Biopolymers* **2012**, *98* (2), 89–97. <https://doi.org/10.1002/bip.21713>.
- (57) Michels, T.; Dölling, R.; Haberkorn, U.; Mier, W. Acid-Mediated Prevention of Aspartimide Formation in Solid Phase Peptide Synthesis. *Org. Lett.* **2012**, *14* (20), 5218–5221. <https://doi.org/10.1021/ol3007925>.
- (58) Marcelo, F.; Supekar, N.; Corzana, F.; Van Der Horst, J. C.; Vuist, I. M.; Live, D.; Boons, G. J. P. H.; Smith, D. F.; Van Vliet, S. J. Identification of a Secondary Binding Site in Human Macrophage Galactose-Type Lectin by Microarray Studies: Implications for the Molecular Recognition of Its Ligands. *J. Biol. Chem.* **2019**, *294* (4), 1300–1311. <https://doi.org/10.1074/jbc.RA118.004957>.

4. Synthesis and study of a peptide presenting a glycocluster with potential to assemble to a coiled-coil

4.1. Abstract

In this chapter the asparagine derivative **AG153** was inserted into a peptide sequence previously reported to form a trimeric coiled-coil sequence.¹

When targeting multivalent lectins, coiled-coil structures can be used as a scaffold to generate new multivalent carbohydrates.

The carbohydrate ligands can, in principle, be placed in different positions on the coiled coil peptide sequence. The nature of the coiled coils (the number of peptides that aggregate together, number of helices, for example) are sequence dependent.² We hypothesised that this multivalent neoglycoprotein could tightly bind to h-MGL and most likely block the receptor, but if internalised by DCs, we expect them to produce a different signalling cascade compared to single strand peptides.

The peptide sequence chosen for the work described in this chapter is based on the previous design of a trimeric self-assembling scaffold for sugar ligand presentation developed in collaboration with the Anna Peacock group in Birmingham, by a past group member, Sinclair Sweeney.¹

Sinclair Sweeney inserted an asparagine residue carrying lactose on the side chain in different position of the coiled-coil sequence and found that sequence tolerates the insertion in terms of folding. Both glycosylated peptides are slightly less folded than the control peptide, 87% α -helix folded, but still consistent with a well folded coiled coil (SS4 78% and SS3 75%) and aggregation to a trimer.¹ Unfortunately, most probably due to the steric hindrance of the lactose, the overall yield of the synthesis was poor. Following work by Claudia Di Salvo from the Murphy group, proved that the insertion of a 10 Å spacer between the lactose and the aspartic acid side chain, allowed better overall yields with similar folding percentage (<https://aran.library.nuigalway.ie/handle/10379/7489>, unpublished data). The glycopeptides prepared by Claudia Di Salvo were obtained in sufficient yield to be tested as inhibitors for galectin 3 and galectin-1, and this work is currently in progress.

As a natural progression of the research done in the past by the Murphy group, I wished to test the possibility to introduce multivalent glycoclusters in a coiled-coil sequence. The

trivalent glycocluster **AG153** above cited, was used as inspiration in the design. The multivalency of the GalNAc residues was envisaged to increase through aggregation of the glycocluster via the coiled coil, although aggregation through the tetraphenylethylene also seemed possible. If a coiled coil trimer resulted it would generate a nonavalent ligand. It was considered of interest to investigate if this approach led to further improvement in affinity for h-MGL. The peptides generate in this chapter may have an exquisite hMGL affinity.

In the past, a fine-tuned ligand presentation through a double helical coiled coil was used to give new ligands for the closest cousin of hMGL, the asialoglycoprotein receptor (ASGPR). Zacco et al. identified among a library of rationally designed coiled coil glycopeptides, two ideal binders of the desired protein, proving that protein targeting via this approach is valuable for drug delivery.³ In contrast with the peptides obtained by Zacco et al. which are internalised and directed to lysosome degradation HepG2 cells,³ previous work of Napoletano et al. led us to believe that such peptides may not be internalised by dendritic cells due to their high molecular weight.⁴ It is known that the intracellular sorting and processing of large glycosylated proteins by DCs appears to be blocked, this was attributed in the past to mannose receptor,⁵ this work will may lead to the discovery of a novel receptor with the same role, MGL.

If glycopeptides or neoglycopeptides can be developed which are not internalised, then they could have potential to be used to block h-MGL as a channel for viral entry, such as Ebola⁶ virus.

Below is reported the synthesis and the folding evaluation by circular dichroism of a self-assembling coiled-coil peptide decorated with the trivalent glycocluster above cited.

4.2. Introduction: coiled-coils

Coiled-coils are peptides in which several α -helices, 2-7, self-coil together forming a super helical motif. An estimated 2–10% of all proteins harbor the coiled-coil motif,⁷ including the major protein of interest in this thesis, hMGL itself, which is believed to have a trimeric coiled-coil neck domain.⁸

Self-assembling structures are important in protein engineering and their main use is in drug delivery.⁹ Undoubted reason for this extensive use in medicinal chemistry are their stability, specificity and tuneability.^{10,11,12}

The typical coiled-coil domain comprises a series of consecutive heptad repeats, named $(abcdefg)_n$, that are readily recognizable by the presence of hydrophobic residues at the 'a' and 'd' positions, as glycine (Gly), alanine (Ala), valine (Val), leucine (Leu), isoleucine (Ile), proline (Pro), phenylalanine (Phe), methionine (Met), and tryptophan (Trp).

The helices formed by multiple heptad repeat are interlocked together every two turns of the helix. Therefore, a coiled-coil has 3.5 amino acids per turn, ($2 \times 3.5 = 7$), compared to 3.6 amino acid per turn in a regular α helix. The distortion arises from the interaction between positions 'a' and 'd' which is attributed to the stability and oligomerisation state of coiled coils: one residue fits into a space generated by four residues of the opposing helix. Crick referred to this packing as a "knobs into holes".¹³

The side chains of the amino acid spacing between 'a' and 'd' are all displaced on one side of the helix, allowing the coexistence of both hydrophobic and hydrophilic residues.¹⁰ The hydrophobic residues are oriented into the centre of the coil to minimize the water repulsion, on the other hand, hydrophilic residues are located in the external side of the coil to maximise interaction with water.

Even if examples of right handed engineered coiled-coil,¹⁴ and natural occurring, as the *Staphylothermus marinus*,¹⁵ are reported in the literature, due to the fact the α -helix is right handed, the majority of heptad repeat coiled coil arrange in a supercoiled coil which is left handed.

4.3. Sequence analysis

The desired coiled coil sequence is based on the heptad repeat (IaAbAcIdEeQfKg)₄.

To ensure the self-assembling of coiled-coil, the sequence had to favour positive interaction, such as hydrophobic, hydrophilic and salt bridge interactions.¹⁶

As mentioned above, in position 'a' and 'd' are place hydrophobic residues which will enhance the “knobs into holes” packing. The hydrophobic amino acid which occupies both these positions is isoleucine (I).

In positions 'b' and 'c' amino acids with small sides chain are ideal to encourage helical formation,¹⁷ alanine (Ala) residues were introduced for this purpose.

In position 'e' and 'g' were selected two amino acid with lateral side chain with opposite charge, glutamic acid (E), carboxylic acid residue, and lysine (K), amine residue, figure 1. These two polar amino acids interact and form interhelical salt bridges, which stabilise the superhelix.

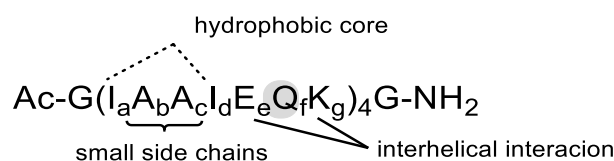


Figure 1. Sequence of the heptad repeat used to create the trivalent coiled coil. In position 'a' and 'd' hydrophobic amino acids, 'b' and 'c' small side chain amino acids, 'e' and 'g' opposite charge amino acids. Position 'f' was elected for the insertion of **AG153**.

As position 'f' is not critical to the formation of the helix, this was the chosen position to incorporate **AG153**. Only looking at the two-dimensional synthetic scheme, it appears that inserting the bulky **AG153** in the first heptad repeat to be synthesised, the sequence closest to the resin, was not the ideal choice. It is indeed true that the steric hindrance can cause low yield but, as possible to see in the three-dimensional representation of the coil figure 2 this repeat, after resin cleavage is the most external one in the supramolecular structure.

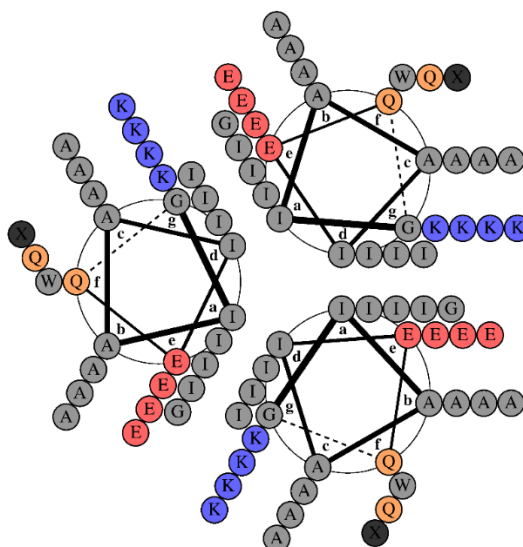


Figure 2. Helical-Wheel diagram of the control sequence designed with DrawCoil 1.0.¹⁸ The Trp was located on the second heptad repeat while amino acid AG153 on the fourth one.

The same position was used to include a tryptophan (Trp) residue for detection purpose in the control sequence, Trp has an adsorption peak at 280 nm ($\epsilon_{280} = 5690 \text{ M}^{-1} \text{ cm}^{-1}$), figure 3. **AG053** presents the tetraphenylethylene glycocluster, which absorb as well at 280 nm, when preparing the peptide solutions for circular dichroism analysis the appropriate molar extinction coefficient of **AG053** was used.

As possible noticing in figure 2, a glycine (G) residue was introduced both at the N and C-termini with the acetylation of the N-terminus to prevent the ionic charge of the amine to disrupt the coiled coil formation.

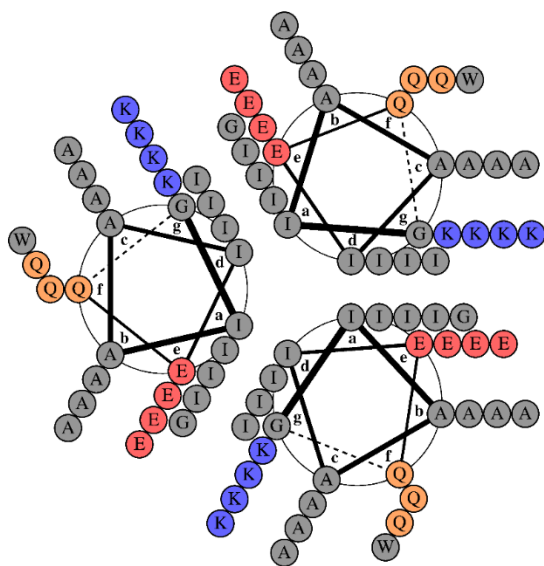


Figure 3. Helical-Wheel diagram of the control sequence designed with DrawCoil 1.0.¹⁸ The Trp was located on the fourth heptad repeat.

4.4. Scaffold synthesis

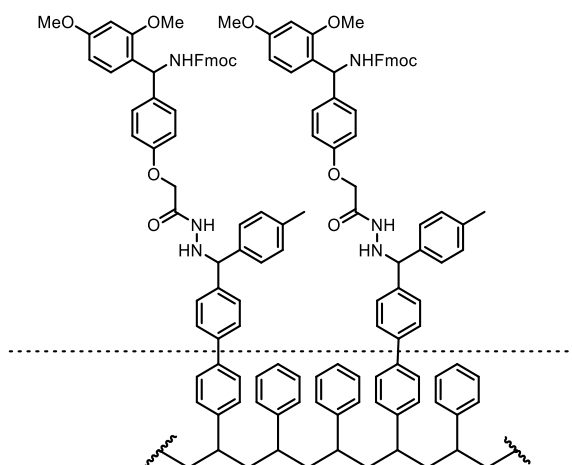


Figure 4. Structure of Rink Amide MBHA resin LL (100-200 mesh)

The peptides were synthesized by stepwise solid-phase peptide synthesis using the Fmoc strategy Rink amide 4-methylbenzhydramine (MBHA), figure 4, which is a low loading resin used for the glycopeptide synthesis. This resin not only is suitable to prevent diketopiperazine formation due to the steric hindrance of the two aromatic groups neighbouring the elongation site, but also, upon the cleavage of the peptide from the resin, it produces an amide on the C-terminus. The amide satisfies hydrogen bonding requirement

in the helix while the carboxylate ionic charges generated after a cleavage from Tenta gel® resin can disrupt coiled-coil formation.

The coupling of the glycosylation site was performed using 2 equivalents of the amino acid building block **AG153**, which was pre-activated manually using 1.8 equivalents HATU/HOAt and 3.6 equivalents of DIPEA, followed by Fmoc deprotection. The standard Fmoc amino acids were coupled automatically on a peptide synthesizer using 8 equivalents of the amino acid and the HBTU/HOBt reagents. Fmoc deprotection was done according to standard conditions, 20% piperidine in DMF. Peptides **AG048** and **AG53** were capped with an *N*-terminal acetyl group using a solution of 0.015 M HOBt, 0.5 M Ac₂O, 0.125 M DIPEA in NMP on *N*-terminus for 45 min. The obtained peptides were then released from the resin, and all acid sensitive side-chain protecting groups were simultaneously removed using TFA/TIPS/H₂O 15:0.9:0.9 in 2.5 h followed by solvent concentration, lyophilization and purification using a C-18 cartridge. For saccharide deprotection, the *O*-acetyl groups were cleaved by transesterification in methanol using catalytic amounts of NaOMe at pH 9-9.5 for 16-24 h, the deprotection was followed by analytical HPLC. The deprotected peptides were purified by preparative HPLC to yield the desired pure products, figure 6, were obtained, **AG048** in 54% yield and **AG053** in 20% yield, figure 5.

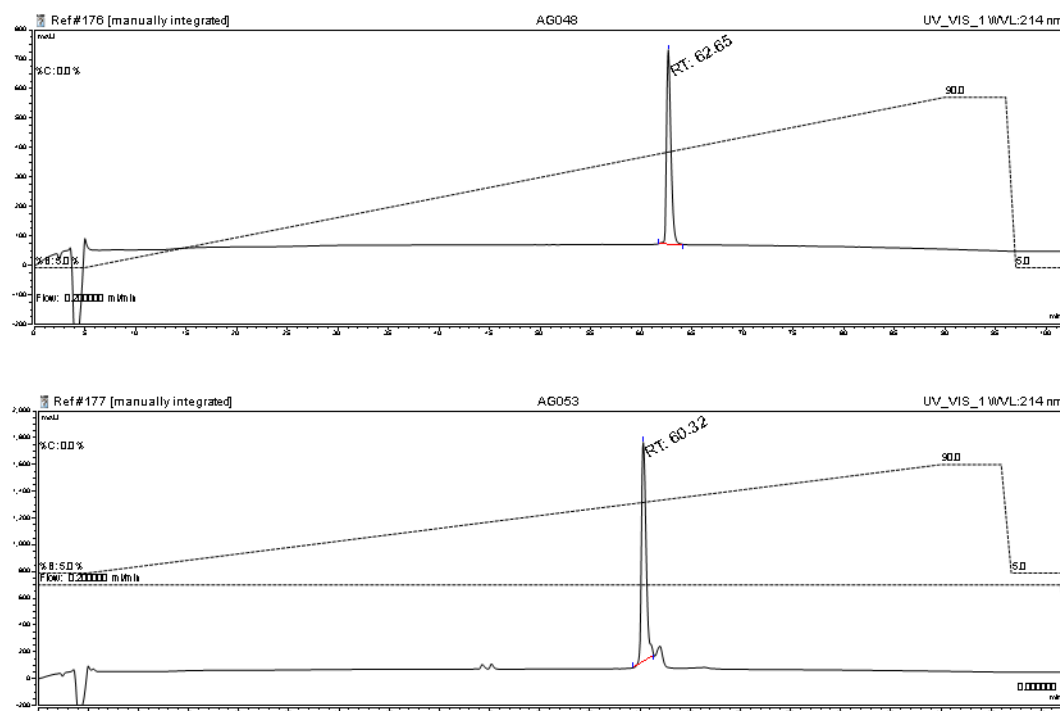


Figure 5. Analytical HPLC chromatograms. On the top AG048 rt:62.65 min, on the bottom AG053 rt:60.32 min.

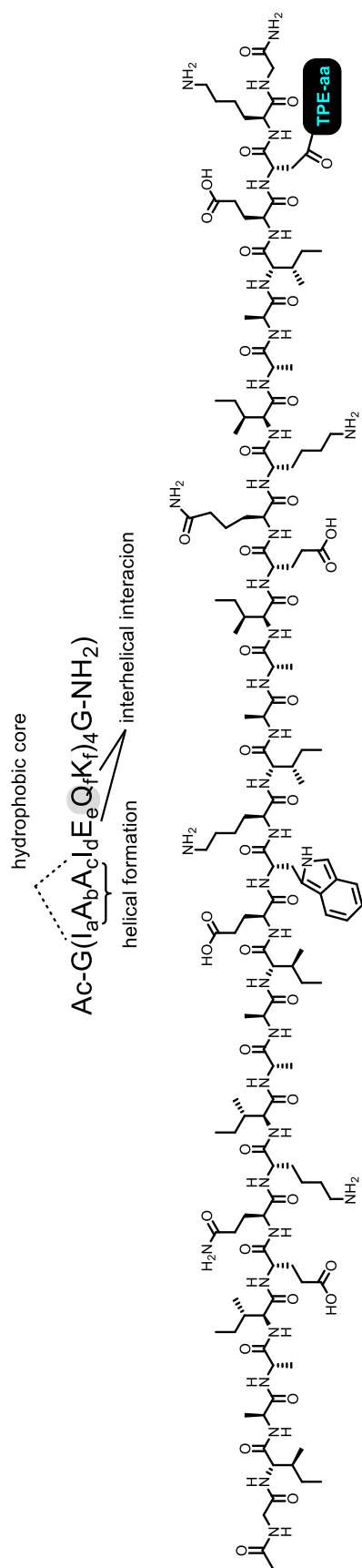


Figure 6. Structure of AG053. The modified amino acid was inserted in the fourth heptad repeat in position 'f'.

4.5 Circular Dichroism

When a molecule contains one or more chiral chromophores, it will absorb to a different extent left-handed circularly polarised light (L-CPL) and right-handed circularly polarised light (R-CPL). The difference of absorption is circular dichroism (CD).

$$\text{Circular dichroism} = \Delta A(\lambda) = A(\lambda)\text{L-CPL} - A(\lambda)\text{R-CPL}$$

In a CD experiment linearly polarised light, a light with oscillation confined to a single plane, is transformed into a sinusoidal wave through either prisms or filters, figure 7.

The filter/prism sums linearly polarised light waves of equal amplitude but out of phase with the other by 90° , to obtain circularly polarised light (CPL).

If the helix spirals clockwise, is defined right-handed, if it spirals anticlockwise is defined left-handed, the two helices are mirror images of each other's.

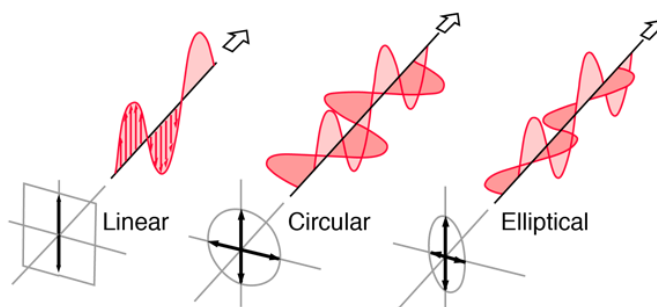


Figure 7. Representation of linearly, circular and elliptical polarised light.

<http://hyperphysics.phy-astr.gsu.edu/hbase/phyopt/polclas.html>

During the measurements of CD spectra of chiral molecule, the two polarised lights, typically from 190 to 270 nm, are passed through the sample, which will absorb one of the helical waves to a greater degree, and the post-sample light travelling to the detector takes on an elliptical form due to the changes in amplitude.

For this reason, the CD data can also be reported, as you will notice below, as the degree of ellipticity (θ), which is defined as the angle between the tangent of the ratio of the minor to major elliptical axis. A circular dichroism signal can be positive or negative, depending on whether L-CPL is absorbed to a greater extent than R- CPL (CD signal positive) or to a lesser extent (CD signal negative).

This spectroscopy technique is used to determine the secondary structure of a protein including α -helix, β -sheet and β -turns which give rise to a typical CD spectrum,¹⁹ figure 8.

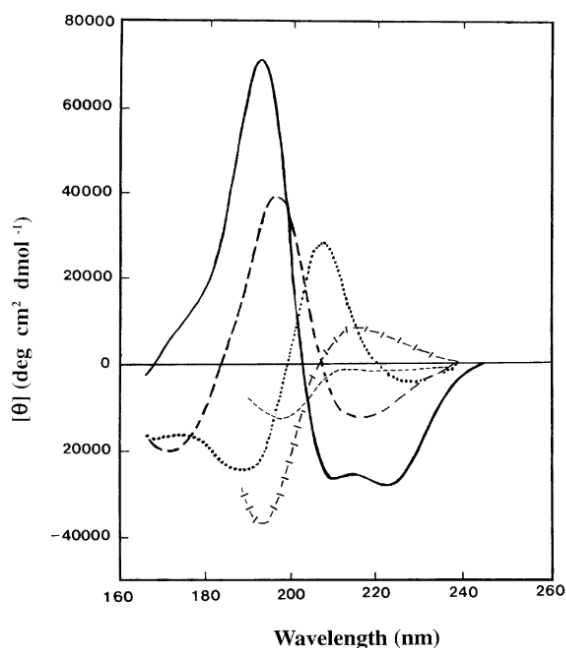


Figure 8. UV CD spectra associated with various types of secondary structure. Solid line, α -helix; long dashed line, anti-parallel β -sheet; dotted line, type I β -turn; cross dashed line, extended 3_1 -helix or poly (Pro) II helix; short dashed line, irregular structure. Figure from *Biochimica et Biophysica Acta* 1751 (2005) 119 – 139.¹⁹

The CD and the thermal folding experiments of the peptide **AG053** and the control sequence were performed in University of Birmingham, by Louise Slope, in the group of Anna Peacock.

From the degree of ellipticity (θ), the molar ellipticity (Θ_m) and the percentage of folding were calculated. The control peptide and **AG053** were subjected to CD study. **AG053** revealed to be particularly sensitive to the presence of HEPES buffer and the measurements needed to be repeated in water.

While the control peptide at both 100 and 30 μ M concentration showed a folding of 95%, measured at 222nm, and the CD spectra associated with α -helix structure with minima at 222 and 208 nm, the presence of the glycocluster significantly disturbed the folding and **AG053** 100 μ M sample reached only 86%, but still α -helix structure is visible, figure 9.

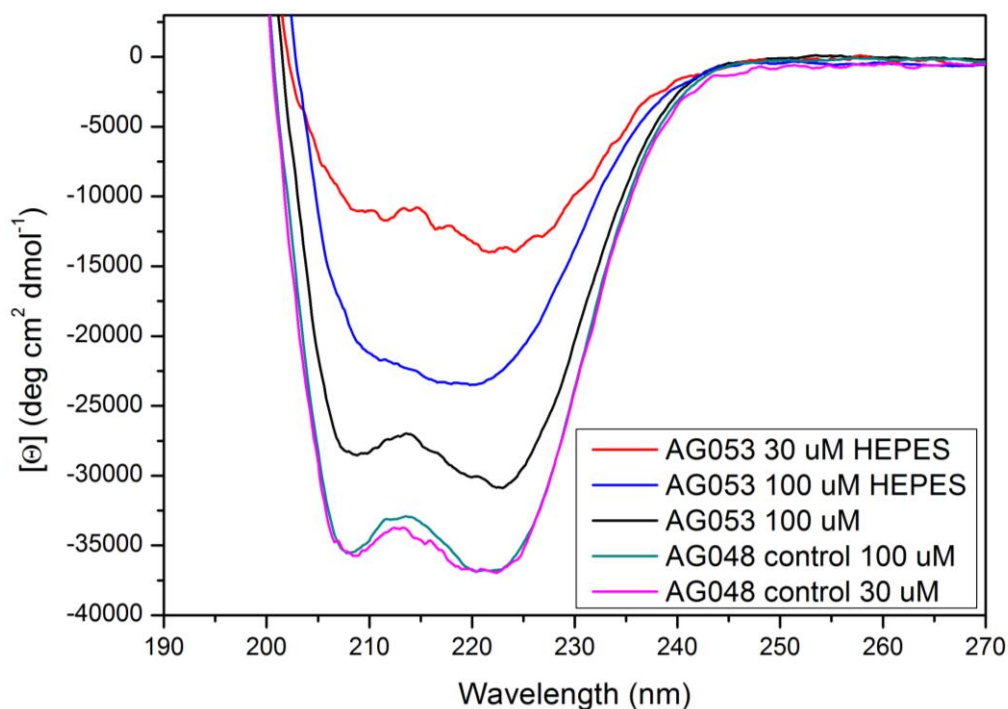


Figure 9. CD analysis of AG053 in HEPES solution and water. Control sequence, in HEPES buffer. AG053 is highly sensible to the presence of HEPES buffer and does not fold in it.

AG053 and the control sequence were also subjected to thermal folding analysis: the samples were subjected to a gradual temperature change from 11°C to 85°C and back. Interestingly, thermal folding analysis of the **AG053** showed that the coil does not return to the original shape after the temperature was increased, figure 10. The folding decreases gradually until the temperature reached 70°C, after this point the folding percentage drop drastically to reach 45%. When the temperature returned to the initial one, the total sample folding observed was only 64%.

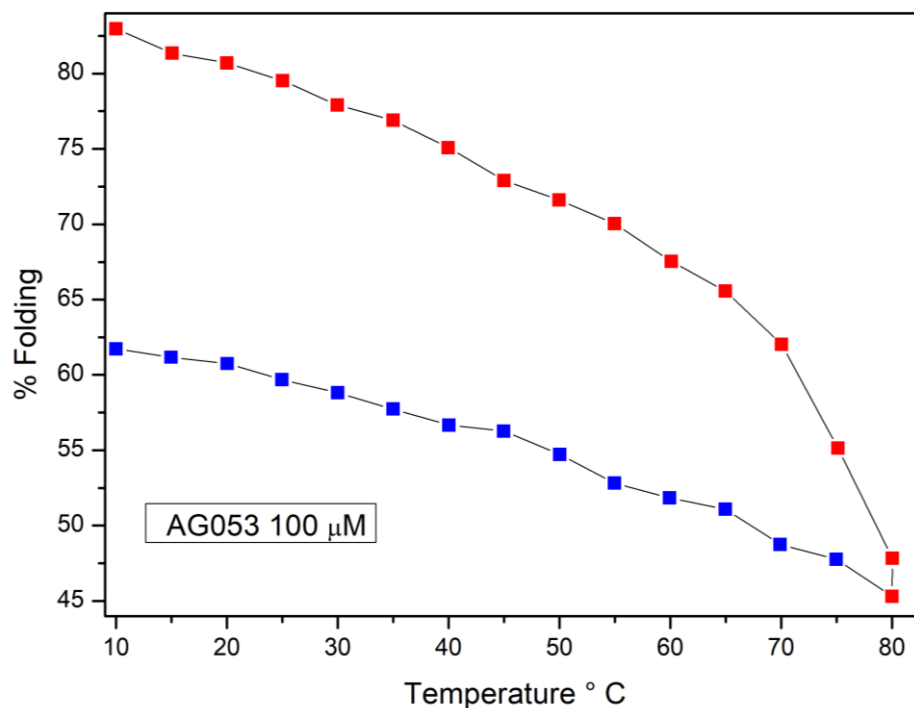


Figure 10. Thermal folding analysis of AG053 100 μM in water. The red dots indicate the temperature increase phase, the blue dots indicate the cooling phase, AG-CC does not return to the original folding status.

The thermal folding analysis of the control sample revealed that this behaviour is to be attributed uniquely to the presence of the glycocluster. In fact, the control sequence folding, which does not display any glycocluster, is practically unchanged after a temperature cycle, figure 11.

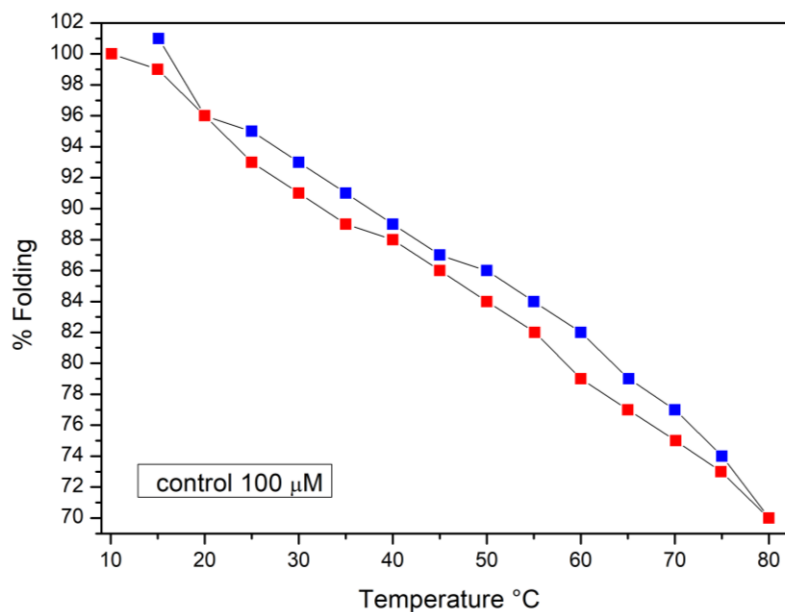


Figure 11. Thermal folding analysis of control sequence in 5 mM HEPES. The red dots indicate the temperature increase phase, the blue dots indicate the cooling phase.

A possible reason for the observed behaviour could be attributed to the tendency of the tetraphenylethylene residue to aggregate in water media. The temperature increase promotes the opening of the self-assembled structure, the single strands are, at this stage, pushed together by the hydrophobic interaction of the aromatic cores.

Analytical ultra-centrifugation (AUC) was used in the past to study the self-assembling characteristics of the coiled-coil sequence carrying lactose by Sinclair Sweeney.¹ The sedimentation equilibrium curves analysis revealed monomer–dimer–trimer coiled-coil stoichiometry. The same type of analysis will be repeated in the future to evaluate the self-assembling characteristic of **AG053**.

4.6. Conclusion

The synthesis of the novel peptide **AG053** was successfully carried out on 12.5 μM scale with an overall yield of 20% during the course of this thesis work. The difference in yield compare to the control peptide, 54%, can be attributed to the difficulties to coupling the bulky tetraphenylethylene based amino acid. The CD data indicates the folding tolerate the addition of the multivalent glycocluster less than the lactosyl analogues previously synthesised in our group. DC data show an alpha-helix of 86% for **AG053** after dissolving in the aqueous buffer compared to the 95% for the control sequence.

The coiled-coil folding showed a strong sensitivity to the presence of buffer in solution: **AG053** is largely unfolded in 5 mM HEPES buffer, while presenting greater folding in pure water at least after initial dissolution. Thermal folding analysis showed that thermal cycling disrupts helix formation in the presence of tetraphenyl ethylene substituent and that folding was slow to reverse, much more so than the control peptide. This research proved that is possible to introduce a multivalent cluster in the trimeric (IAAIEQK)₄ coiled-coil sequence with discrete overall yield. Further research is required to investigate whether refolding of this neoglycopeptide can be achieved after heating, and also to determine if trimers are formed by AUC analysis.

4.7 Bibliography

- (1) Sweeney, S. M.; Bullen, G. A.; Gillis, R. B.; Adams, G. G.; Rowe, A. J.; Harding, S. E.; Tucker, J. H. R.; Peacock, A. F. A.; Murphy, P. V. Coiled Coil Type Neoglycoproteins Presenting Three Lactose Residues. *Tetrahedron Lett.* **2016**, *57* (13), 1414–1417. <https://doi.org/10.1016/j.tetlet.2016.02.005>.
- (2) Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Paula, J. Self-Assembling Cages from Coiled-Coil Peptide Modules. *Science* (80-.). **2013**, *340* (6132), 595–599. <https://doi.org/10.1126/science.1233936>. Self-Assembling.
- (3) Zacco, E.; Hütter, J.; Heier, J. L.; Mortier, J.; Seeberger, P. H.; Lepenies, B.; Kokscha, B. Tailored Presentation of Carbohydrates on a Coiled Coil-Based Scaffold for Asialoglycoprotein Receptor Targeting. *ACS Chem. Biol.* **2015**, *10* (9), 2065–2072. <https://doi.org/10.1021/acscchembio.5b00435>.
- (4) Napoletano, C.; Rughetti, A.; Agervig Tarp, M. P.; Coleman, J.; Bennett, E. P.; Picco, G.; Sale, P.; Denda-Nagai, K.; Irimura, T.; Mandel, U.; et al. Tumor-Associated Tn-MUC1 Glycoform Is Internalized through the Macrophage Galactose-Type C-Type Lectin and Delivered to the HLA Class I and II Compartments in Dendritic Cells. *Cancer Res.* **2007**, *67* (17), 8358–8367. <https://doi.org/10.1158/0008-5472.CAN-07-1035>.
- (5) Hiltbold, E. M.; Vlad, A. M.; Ciborowski, P.; Watkins, S. C.; Finn, O. J. The Mechanism of Unresponsiveness to Circulating Tumor Antigen MUC1 Is a Block in Intracellular Sorting and Processing by Dendritic Cells. *J. Immunol.* **2000**, *165* (7), 3730–3741. <https://doi.org/10.4049/jimmunol.165.7.3730>.
- (6) Takada, A.; Fujioka, K.; Tsuiji, M.; Morikawa, A.; Higashi, N.; Ebihara, H.; Kobasa, D.; Feldmann, H.; Irimura, T.; Kawaoka, Y. Human Macrophage C-Type Lectin Specific for Galactose and N-Acetylgalactosamine Promotes Filovirus Entry. *J. Virol.* **2004**, *78* (6), 2943–2947. <https://doi.org/10.1128/JVI.78.6.2943>.
- (7) Wolf, E.; Kim, P. S.; Berger, B. Multicoil - a Program for Predicting 2-Stranded and 3-Stranded Coiled Coils. *Protein Sci* **1997**, *6* (6), 1179–1189.
- (8) Jégouzo, S. A.; Quintero-Martínez, A.; Ouyang, X.; Dos Santos, Á.; Taylor, M. E.; Drickamer, K. Organization of the Extracellular Portion of the Macrophage Galactose Receptor: A Trimeric Cluster of Simple Binding Sites for N-Acetylgalactosamine. *Glycobiology* **2013**, *23* (7), 853–864. <https://doi.org/10.1093/glycob/cwt022>.
- (9) Fan, T.; Yu, X.; Shen, B.; Sun, L. Peptide Self-Assembled Nanostructures for Drug Delivery Applications. *Journal of Nanomaterials*. 2017. <https://doi.org/10.1155/2017/4562474>.
- (10) McFarlane, A. A.; Orriss, G. L.; Stetefeld, J. The Use of Coiled-Coil Proteins in Drug Delivery Systems. *Eur. J. Pharmacol.* **2009**, *625* (1–3), 101–107. <https://doi.org/10.1016/j.ejphar.2009.05.034>.
- (11) Frezzo, J. A.; Montclare, J. K. Exploring the Potential of Engineered Coiled-Coil Protein Microfibers in Drug Delivery. *Ther. Deliv.* **2015**, *6* (6), 643–646. <https://doi.org/10.4155/tde.15.19>.
- (12) Yu, Y. B. Coiled-Coils: Stability, Specificity, and Drug Delivery Potential. *Adv. Drug Deliv. Rev.* **2002**, *54* (8), 1113–1129. [https://doi.org/10.1016/S0169-409X\(02\)00058-3](https://doi.org/10.1016/S0169-409X(02)00058-3).
- (13) Crick, F. Packing of Alpha-Helices: Simple Coiled-Coils. *Acta Crystallogr.* **1953**, *6*, 986–997.
- (14) Sales, M.; Plecs, J. J.; Holton, J. M.; Alber, T. O. M. Structure of a Designed , Right-Handed Coiled-Coil Tetramer Containing All Biological Amino Acids. *Protein Sci.* **2007**, *16*, 2224–2232. <https://doi.org/10.1110/ps.062702907.model>.
- (15) Stetefeld, J.; Jenny, M.; Schulthess, T.; Landwehr, R.; Engel, J.; Kammerer, R. A. Crystal Structure of a Naturally Occurring Parallel Right-Handed Coiled Coil Tetramer. *Nat. Struct. Biol.* **2000**, *7* (9), 772–776.
- (16) Fletcher, J. M.; Boyle, A. L.; Bruning, M.; Bartlett, Sg. J.; Vincent, T. L.; Zaccari, N. R.; Armstrong, C. T.; Bromley, E. H. C.; Booth, P. J.; Brady, R. L.; et al. A Basis Set of de Novo Coiled-Coil Peptide Oligomers for Rational Protein Design and Synthetic Biology. *ACS Synth. Biol.* **2012**, *1* (6), 240–250. <https://doi.org/10.1021/sb300028q>.
- (17) Pace, C. N.; Scholtz, J. M. A Helix Propensity Scale Based on Experimental Studies of Peptides and

Chapter 4

- Proteins. *Biophys. J.* **1998**, 75 (1), 422–427.
- (18) Grigoryan, G.; Keating, A. E. Structural Specificity in Coiled-Coil Interactions. *Curr. Opin. Struct. Biol.* **2008**, 18 (4), 477–483. <https://doi.org/10.1016/j.sbi.2008.04.008>.
- (19) Kelly, S. M.; Jess, T. J.; Price, N. C. How to Study Proteins by Circular Dichroism. *Biochim. Biophys. Acta - Mol. Basis Dis.* **2005**, 1751, 119–139. <https://doi.org/10.1016/j.bbapap.2005.06.005>.

5. Expression and crystallisation of hMGL carbohydrate recognition domain

5.1. Abstract

hMGL, also known as CLEC10A, is the only known receptor on antigen-presenting cells that can selectively recognise the T_N- and ST_N-antigen cancer biomarkers. As discussed in previous chapters, this receptor causes antigen endocytosis and induces unique cytokine/chemokine production, which activates the maturation of a specific subset of T cells. The type of response triggered is highly dependent on the antigen structure, however, no structure-activity relationship has been identified due to the lack of structural information and the efforts to target it are based on the try and error approach. Advance techniques to investigate macromolecular structures are available,¹ but X-ray crystal structure analysis remains the most reliable method to obtain atomic resolution details of macromolecular structures and the binding mode of ligands.

This work describes the development of an expression, refolding and purification protocol for the production of hMGL CRD and extracellular domain in bacterial cells, *e. coli*, which led to the recovery of the protein in high yield. It also led to determination of appropriate crystallisation conditions for the hMGL CRD and after the X-ray crystal structure determination enabled identification of key protein-carbohydrate interactions that can explain the hMGL selectivity for GalNAc, and the orientation of this residue in the binding pocket.

The findings disclosed in this chapter provide the first step, based on an X-ray crystal structure of the hMGL CRD with and without ligands, to help defining structure-activity relationships for ligands of this lectin. The atomic resolution structural information will help researchers to develop small molecule therapeutics to target hMGL and to engineer novel selective antibodies to detect cancer biomarkers.

5.2. Results

5.2.1. Protein expression

For the experimental details describing the cloning, plasmid preparation, expression, and recovery of the inclusion body see chapter 6.

The expression of the hMGL CRD as well as the full-length isoforms 3 and 1 were successfully achieved using *Escherichia coli* (*E. coli*) C41(DE3) strain as prokaryotic expression system, figure 1. After performing an extensive search of the Protein Data Bank (PDB), it was noticed that the typical coiled-coil tail length in crystal structures has not yet exceeded that of more than forty amino acids. It was envisioned that a longer tail may disrupt the crystal packing and inhibit appropriate crystallisation. In accordance with this theory, truncated versions, with tails that are 30 and 60 amino acids in length, corresponding to the full-length isoform 1 were also expressed.

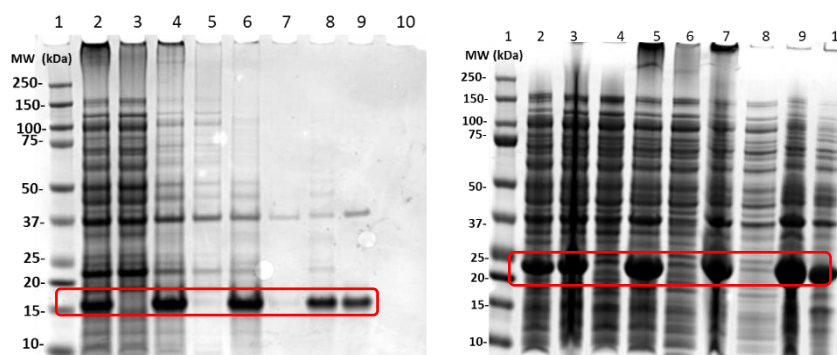


Figure 1. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) of hMGL CRD, ca 15 kDa left, and full-length isoform 3, ca 25 KDa right, both after expression in *E. Coli* C41(DE3). Different lanes show washing steps of the inclusion bodies and the soluble fractions, see experimental.

As mentioned in chapter 1.6, protein expression in bacteria is fast, inexpensive and typically allows for milligrams of product to be obtained using protease-deficient host strains. The main drawbacks of bacterial expression include the lack of post translational modifications seen in higher organisms and proper disulfide bond formation in the cytoplasm. These limitations can be problematic for correct folding, in particular for high molecular weight, multi-domain or disulfide-rich proteins. Additionally, for protein expression, protease-deficient host strains are used. These strains carry mutations which eliminate the production of proteases and yield in high level of protein expression. At high expression levels, hydrophobic portions of folding intermediates or improperly folded

proteins may interact with each other leading to aggregation in the form of inclusion bodies.² hMGL CRD, despite the fact that is not an extremely high molecular weight protein, is recovered as insoluble inclusion bodies, most probably due to the presence of three disulphide bonded cysteine residues in its CRD. Following of protocols reported in the literature, yielded small amounts of correctly folded hMGL. Various experiments were carried out to identify conditions to enable the cells produce more soluble protein. These included use of lower amounts of isopropyl β -D-thiogalactose (IPTG) and longer expression times at reduced temperature. However, the protein could still only be obtained mainly in the insoluble form.

Protein insolubility and formation of inclusion bodies is often a strong deterrent to those working in the structural biology field. It is likely that the difficulties experienced in obtaining large quantities of pure hMGL has been the major bottleneck to obtaining an atomic resolution crystal structure for hMGL.

A typical process to recover the protein from inclusion bodies includes complete protein denaturation in basic conditions, followed by slow exchange of the denaturing agent with a buffer in which the protein is stable and self-folds, figure 2.

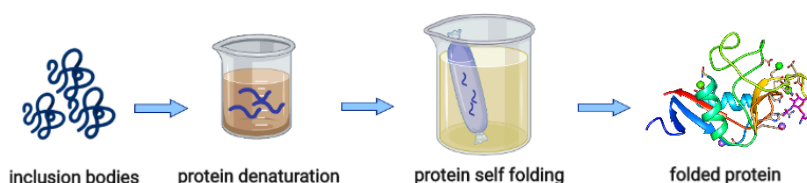


Figure 2. Protein recovery from inclusion bodies. A buffer which allows protein self-folding is required.

Finding appropriate refolding conditions is a trial-and-error approach and is therefore time-consuming. Furthermore, rapid and inexpensive analysis to check the folding pattern are not always available in molecular biology laboratories.

In this study, a systematic approach to determining the optimal refolding condition of hMGL was used. This strategy, is based on the innovative work of Y. Wang et al.^{3 4} The most common hMGL refolding conditions reported in the literature were selected. Parameters, such as, pH, salt concentration, buffer concentration and addition of “folding helpers” (additives) were systematically screened using a 96 well plate where 190 μ L of refolding buffer was prepared, table 1.

	20 mM Maleic Acid pH 4.5	20 mM Maleic Acid pH 5.0	20 mM Maleic Acid pH 6.2	20 mM Mops pH 5.0	20 mM Mops pH 6.0	20 mM Mops pH 7.0	20 mM Tris pH 6.0	20 mM Tris pH 6.5	20 mM Tris pH 7.0	20 mM Tris pH 7.5	20 mM Tris pH 8.0
200 mM NaCl	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
400 mM NaCl	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11
600 mM NaCl	C1	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
800 mM NaCl	D1	D3	D3	D4	D5	D6	D7	D8	D9	D10	D11
200 mM NaCl	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11
400 mM NaCl	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
600 mM NaCl	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
800 mM NaCl	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11

Table 1. Refolding screen. The buffer (3-(N-morpholino) propanesulfonic acid, MOPS, tris(hydroxymethyl)-aminomethane, Tris) and pH are systematically changed as indicated from left to right, while salt concentration and presence of “folding helper”, row E to H contain 400 mM lysine, are changed from top to bottom. This simple method facilitates checking in a limited time various folding condition.

After expression, the inclusion body pellets were solubilised with ammonium hydroxide and 10 μ L of the protein solution were added in each vial of the plate described above to mimic twenty-fold dialysis dilution. The plate was gently mixed at 4°C and the protein folding was assessed at various time points using fluorescence spectroscopy. The ratio of intrinsic fluorescence of tyrosine (Tyr) and tryptophan (Trp) residues as a function of temperature gradient is a convenient way to quickly check the status of protein folding while needing only 10 μ L of sample for the analysis, figure 3. This simple analytical method was recently introduced on the market by NanoTemper Technologies. This approach enabled testing over eighty refolding conditions using less than 5 mg of protein.

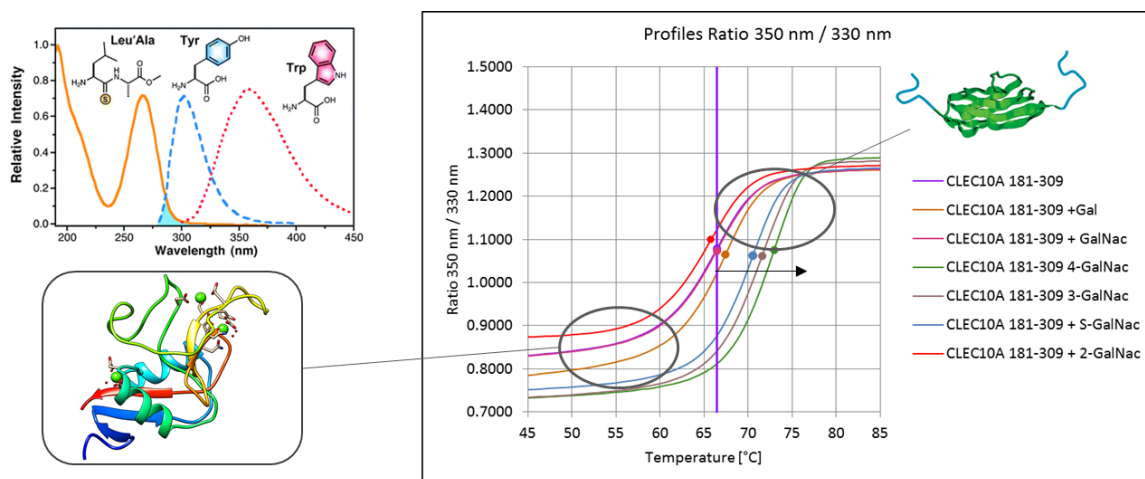


Figure 3. Fluorescence spectroscopy analysis: the sample is applied to a temperature gradient. The change in fluorescence caused by the different special orientation of Tyr and Trp in the folded versus misfolded form are detected.

Refolding was found to occur within 8-16 h using a dialysis buffer containing 20mM Tris pH 7.5, 500 mM NaCl and 25 mM CaCl_2 . The further addition of 400mM of L-arginine in the refolding buffer reduced the refolding time to only 4-6 h. A second systematic refolding screen in which different concentrations of reduced and oxidized glutathione were investigated resulted in further optimisation of this condition. The optimal ratio was identified at 1mM reduced glutathione, 0.1 mM oxidized glutathione. This buffer was subsequently used for the refolding in larger scale of all constructs of hMGL.

5.2.2. Protein purification

All the constructs were purified in a similar fashion, for detail see chapter 6. The poly histidine tag was exploited for the first purification step. In a similar fashion, all the constructs were allowed to bind to Ni-NTA resin. Non-specifically bound proteins were removed with extensive washing steps using a buffer containing 20mM imidazole. Final elution of the desired protein occurred with by eluting with 200mM imidazole. The His tag was then cleaved overnight at 4°C with TEV protease, a cysteine protease that recognizes the cleavage site of Glu-X-X-Y-X-Gln-(Gly/Ser) and cleaves between Gln and Gly/Ser. The efficiency of the His tag removal was followed by SDS-PAGE figure 4.

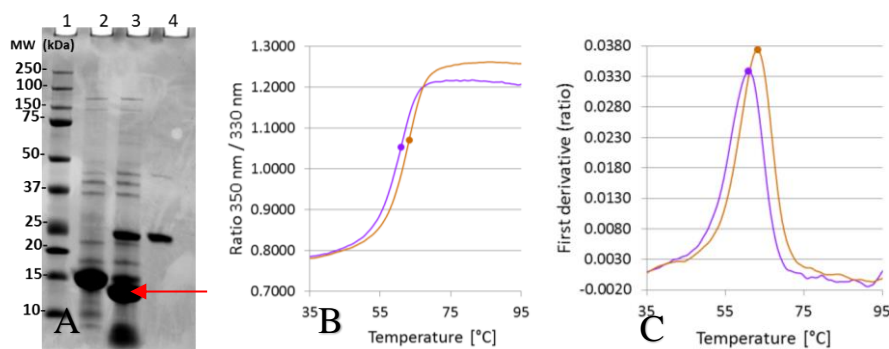


Figure 4. SDS PAGE His Tag cut with TEV & Thermal unfolding analysis of hMGL CRD. A: Lane 1 Molecular weight marker (BioRad), lane 2; 6His-hMGL CRD 16.9 kDa, lane 3; aliquot of enzymatic digestion reaction, hMGL CRD 15.0 kDa, small portion of 6His-hMGL and TEV, lane 4; TEV protease. B: thermal folding curve of hMGL with and without poly-histidine tag. T_m 6His-hMGL = 60.9 °C, purple curve, T_m hMGL = 63.2 °C, yellow curve. C: first derivatives of the thermal folding curves. Measurements performed with 10 μ L sample with Tyco from NanoTemper.

After completion of the His tag removal, the protein was further purified by anion exchange, figure 5, and by size exclusion chromatography figure 6. The purification was carried out with buffers containing 500 mM galactose to prevent the protein sticking to the column matrix packed with Q Sepharose, a polysaccharide containing galactose units. Galactose was removed with a desalting column prior to crystallization and ITC analysis. The final protein was concentrated to 67 mg/ml, supplemented with 5 mM CaCl_2 and stored at -80°C .

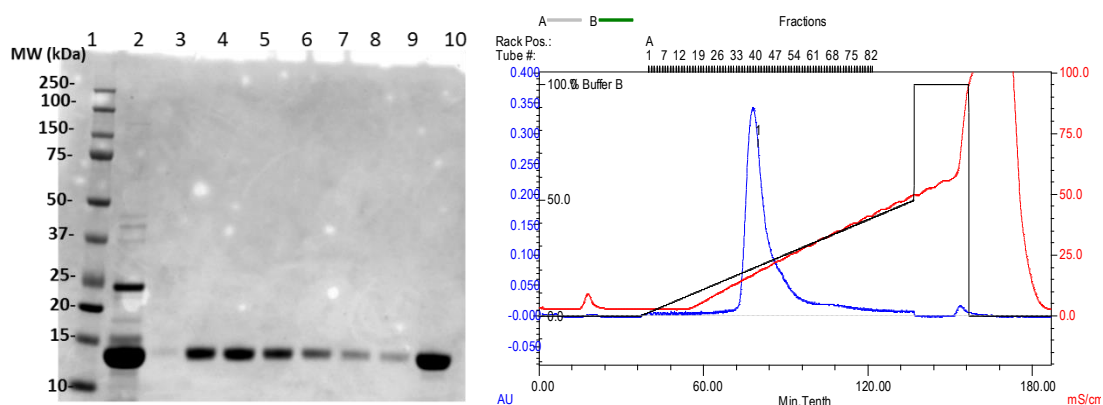


Figure 5. On the left SDS page ion exchange chromatography; on the right chromatogram.

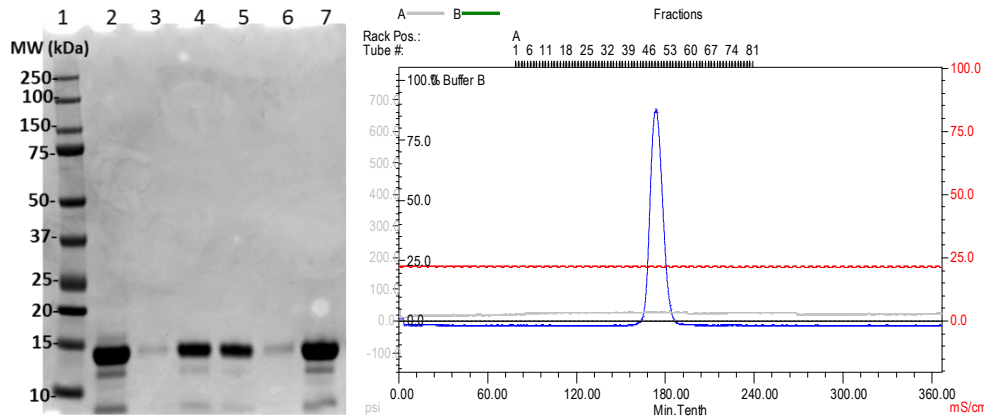


Figure 6. On the left size exclusion chromatography; on the right chromatogram.

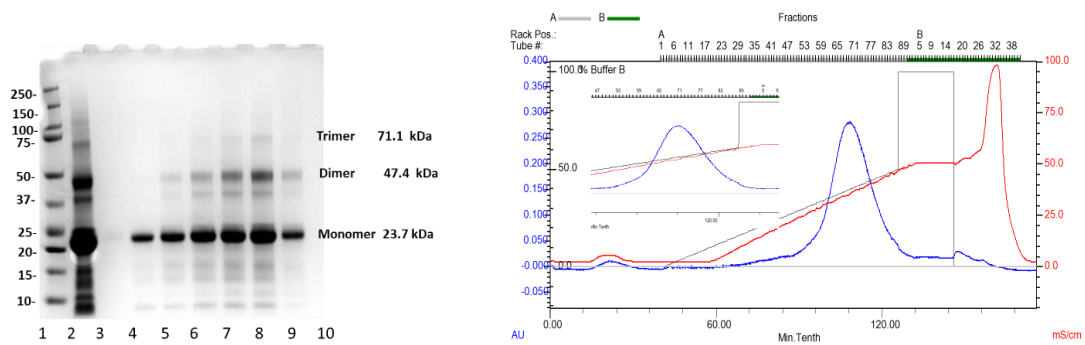


Figure 7. On the left SDS page ion exchange chromatography of hMGL extracellular domain; on the right chromatogram

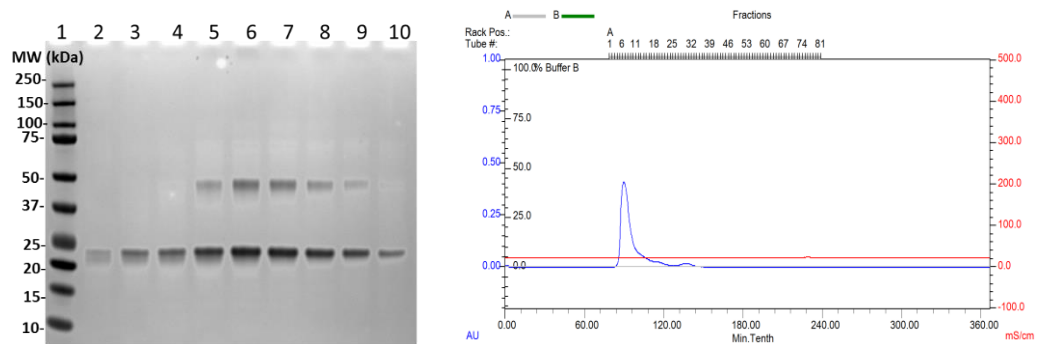


Figure 8 On the left size exclusion chromatography of hMGL extracellular domain; on the right chromatogram

5.2.3. Crystallisation of hMGL carbohydrate recognition domain

The Joint Centre for Structural Genomic (JCSG) Core Suite I, II, III and IV suites (Qiagen) and a PEG smear screen (Rigaku Reagents) were selected for initial crystallization trials performed by the hanging drop vapour diffusion method, figure 7. To co-crystals, the protein and ligand reservoir solutions were premixed to give the desired test concentration. The initial tests occurred with protein concentration between 4 mg/mL and 20 mg/mL and 10 equivalents of ligand. In step 2, 100 μ l of crystallization solution is dispensed in the reservoir of a 96-well crystallization plate. In step 3, 100 nL of protein solution at a concentration of 4-20mg/ml is dispensed in the well. Finally, 100 nL of reservoir solution is dispensed and mixed with the protein solution in the well, step 4. The chambers were sealed with a clear adhesive tape, step 5, and the well was allowed to equilibrate with the reservoir for several days at 20 °C step 6.

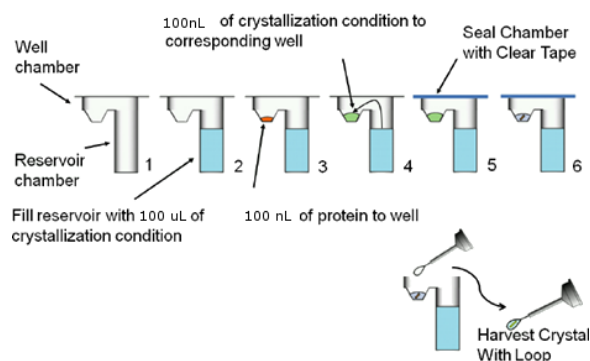


Figure 9. Representation of hanging drop vapour diffusion method. Adapted from HamptonResearch.com.

Crystallisations of the full-length protein, and the truncated versions was attempted., however, no useful crystals were obtained. Even at low concentration, upon mixing with the reservoir solution, the protein formed heavy precipitate. In contrast, hMGL CRD crystallisation or precipitation occurred at concentration 20 mg/ml. Almost 500 crystallisation conditions were tested but, at this protein concentration, only three of these conditions generated crystals suitable for X-ray analysis. Recurrent in the three crystallisation conditions, were calcium chelating agents, which therefore prevented the binding of carbohydrate ligands.



Figure 10. Crystal of hMGL CRD. All the reservoir solutions which yielded the first set of crystals, contained calcium chelating agent (e.g. Sodium citrate) which prevented the binding of carbohydrate ligands, but yielded the crystal structure of the CRD alone.

To find crystallisation conditions which would allow hMGL to bind the GalNAc derivatives and crystallise, a protein concentration of 70 mg/mL was required. At this concentration, a reservoir solution containing 100 mM tris pH 6-7, 200 mM MgCl₂, NaCl 2-3 M yielded needle like crystals, figure 9. The crystallisation was scaled-up to 2 μ L drops to obtain larger crystals, which were then harvested and analysed using synchrotron radiation.



Figure 11. Crystal of hMGL CRD in complex with T_N antigen.

5.2.4. MGL crystals harvesting

With the aim to impede ice formation during the harvesting process and minimise the radiation damage during data collection, the drops containing crystals of appropriate dimension were mixed with the same volume of cryoprotectant solutions (crystallisation buffer with 20-30% glycerol). After cryoprotection the crystals were mounted in loop of appropriate size, flash frozen in liquid nitrogen and transferred to an opposite liquid nitrogen tank for the shipping to synchrotron.

5.2.5. Data processing

The crystals of the free hMGL carbohydrate recognition domain and those which contained the protein bound with different monovalent ligands (GalNAc, methyl 2-(acetylamino)-2-deoxy-1-thio- α -D-galactopyranose, and the serine T_N-antigen) were analysed with synchrotron radiation at ESRF (Grenoble, France). The data collected facilitated determination of the high-resolution crystal structures.

To determine the macromolecular structures by X-ray crystallography the suite of programmes Collaborative Computational Project (CCP4) was used. Data processing and scaling was carried out using HKL2000. Data collection statistics are listed in table 2. Structures were solved by molecular replacement with known atomic coordinate ASGPR (PDB:1DV8⁵), using Phaser.⁶ Ligands, ions and water molecules were only included in the model when unbiased positive difference electron density was clearly visible. All the structures were refined using the program Refmac5⁷ with 5% of unique reflections excluded from refinement calculations for cross validation purposes. Ligand restraints were generated using the CCP4i monomer sketcher,⁸ Coot,⁹ and the PRODRG server.¹⁰ Electron density map inspection and model building were done using Coot. Geometry was validated, and Ramachandran plot was calculated using the Molprobit server. Distances were measured using Coot. The structure analysis and the images were produced using UCSF Chimera.¹¹

Table 2. Crystallographic data

	hMGL CRD	hMGL CRD GalNAc	hMGL CRD S-GalNAc	hMGL CRD Tn Antigen
PDB ID	6PUV	6PY1	6U70	N/A
Root image name	HRV-CLEC 2-1	CLEC-Gal-5	HRV-CLEC- tGalNAc_2_1	HRV-CLEC-Tn- 2
Wavelength (Å)	0.9763	0.9763	0.9863	0.9863
Beamline	ESRF, ID-30B	ESRF, ID-30B	ESRF, ID-30B	ESRF, ID-30B
Detector	Pilatus 3 6M	Pilatus 3 6M	Pilatus 3 6M	Pilatus 3 6M
Space group	P2 ₁ 2 ₁ 2 ₁	P3 ₁ 2	P3 ₁ 2	P3 ₁ 2
Cell constants	a = 28.9 Å	a = 51.8 Å	a = 52.3 Å	a = 52.2 Å
	b = 42.7 Å	b = 51.8 Å	b = 52.3 Å	b = 52.2 Å
	c = 89.9 Å	c = 112.8 Å	c = 113.3 Å	c = 112.7 Å
	α = β =	α = β = 90 °	α = β = 90 °	α = β = 90 °
	γ = 90 °	γ = 120.0°	γ = 120.0°	γ = 120.0°
Resolution (Å)	45.00 – 1.20 (1.22 – 1.20)	50 – 1.7 (1.73 – 1.70)	37.8 – 2.31 (3.37 – 2.31)	42.2 – 2.00 (2.07 – 2.00)
Observed reflections	124,598	172,222	74,405	72,279
Unique reflections	34,612	20,014	8,370	12,784
Completeness (%)	96.8 (99.2)	100.0 (99.9)	99.6 (99.3)	99.6 (98.8)
Redundancy	3.7 (3.6)	8.6 (6.5)	8.9 (8.5)	5.7 (4.9)
Overall <I/σ(I)>	14.0 (2.1)	12.4 (1.6)	14.2 (8.1)	12.1 (5.9)
CC _{1/2}	0.99 (0.82)	0.98 (0.27)	0.99 (0.90)	0.99 (0.89)
R _{pim} (%) ^c	3.7 (25.0)	10.2 (86.7)	6.7 (24.9)	8.0 (28.0)
R _{work} /R _{free} ^d (%)	15.5/18.9	16.5/18.8	18.6/22.8	18.3/21.1
Bond lengths ^e (Å)	0.012	0.013	0.011	0.013
Bond angles ^e (°)	1.722	1.639	1.531	1.666
Ramachandran Plot (%)	97/3/0	96/4/0	98/2/0	98/2/0
Favoured/allowed/outliers				

^aValues in parentheses are for the highest resolution shell.

$${}^b R_{sym} = \frac{\sum |I - \langle I \rangle|}{\sum I}$$

$${}^c R_{pim} = \sqrt{\frac{1}{n-1} \frac{\sum |I - \langle I \rangle|}{\sum I}}$$

Where I is the observed integrated intensity, $\langle I \rangle$ is the average integrated intensity obtained from multiple measurements, and the summation is over all observed reflections.

$${}^d R_{cryst} = \frac{\sum ||F_o| - k|F_c||}{\sum F_o}$$

F_o and F_c are the observed and calculated structure factors, respectively, and k is a scaling factor. The summation is over all measurements. ^d R_{free} is calculated as R_{cryst} using 5% of the reflections chosen randomly and omitted from the refinement calculations. ^eBond lengths and angles are root-mean-square deviations from ideal values.

5.2.6. hMGL crystals analysis

The crystals of the free hMGL carbohydrate recognition domain and those which contained the protein bound with different monovalent ligands (GalNAc, methyl 2-(acetylamino)-2-deoxy-1-thio- α -D-galactopyranose, and the serine T_N-antigen) were analysed with synchrotron X-ray radiation and the data obtained facilitated determination of the high-resolution crystal structures.

The free carbohydrate recognition domain (CRD) produced rod cluster crystal in space group P2₁2₁2₁. The structure, figure 10, contains two α -helices, five long β -strands (β 1- β 6) of 4/5 residues and three shorts, composed only by a couple of amino acids. Aligning the β -strands along an imaginary Y axis the α -helices are located on each side of the axis with a 90° angle in a similar to the structure of the ASGPR CRD.¹² The α -helices are separated by ten amino acids of random coil structure. His 215 and Leu 216 form a short β -strand which is rare for animal lectins.

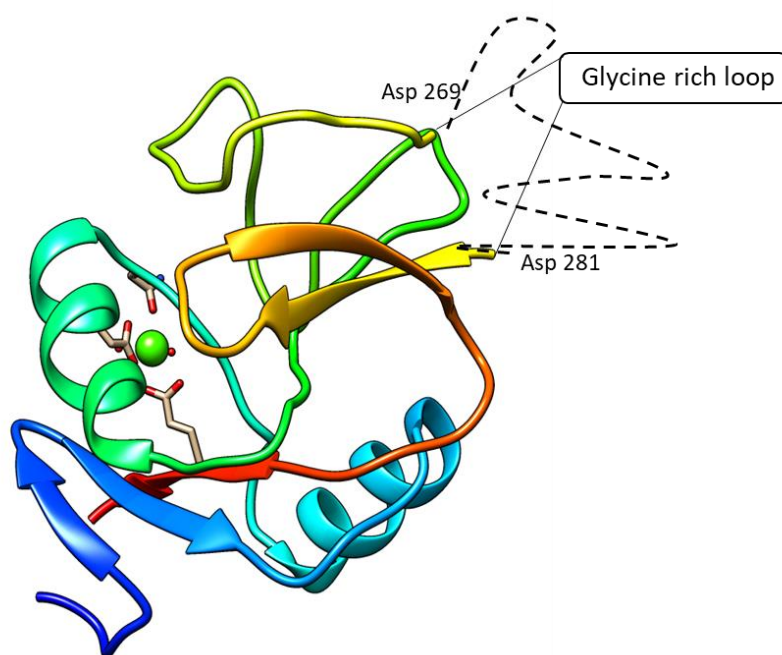


Figure 12. Ribbon diagram of the free hMGL CRD. The two α -helices and the β -strands are shown, the calcium ions in green. Both the N and the C terminus are on the bottom of the image. The tract line represents glycine rich loop, which is probably too flexible to produce consistent electron density in a unique conformation. Figure was generated with UCSF Chimera.¹¹

In the absence of ligand, the region between Asp 269 and Asp 281 is flexible and is missing in electron density maps. In this part are present five glycine residues and for this reason it was named “glycine- rich loop”¹³ Its flexibility suggests the involvement in the sugar binding site.

As indicated by UniProt (accession number Q8IUN9), there are three intrachain disulphide bonds. The first disulphide bond connects cysteine 181 at the N-terminus to Cys 192 in β 2. The second disulphide bond connects Cys 209 to Cys304 and bridges α 1 and β 5. The final disulphide from Cys 282 to Cys 296. The short strand at the N terminus, together with strands β 1 and β 5 form the first β -sheet, while β 4, β 3 and part of β 2 form the second β -sheet.

The structures that were bound to ligands produced needle-like crystals in the space group $P3_12$ and showed a high degree of similarity to the free carbohydrate recognition domain, figure 11. The five β strands and the two α helices were named in the same order as that of the unbound CRD. In the structures of the complex, it was possible to assign with high confidence the position of the glycine rich loop. This loop, which is part of the longest unorganised portion of the CRD, is located between β 2 and β 3 and makes up part of the carbohydrate binding site.

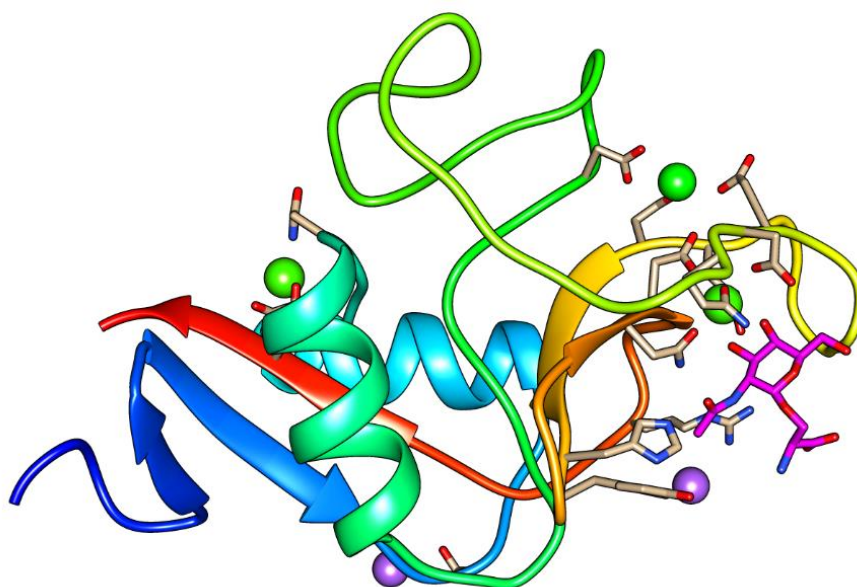


Figure 13. Ribbon diagram of the hMGL CRD binding serine T_N Antigen. Similar structures were observed for GalNAc and methyl 2-(acetylamino)-2-deoxy-1-thio- α -D-galactopyranose. The glycine rich loop is more rigid due to the interaction with the glycan. This figure generated with UCSF Chimera.¹¹

In all the structures that contain the ligand, it is possible to assign with high level of confidence the location of three calcium ions. This is in agreement with what was observed for ASGPR.^{12,14} The first calcium atom is involved in the carbohydrate binding site and coordinates two vicinal hydroxyl groups, the 3-OH and 4-OH groups of the pyranose ring. The coordination at these two specific hydroxyl groups to the calcium atom explains why

any chemical modification at these positions of the GalNAc is not tolerated and with binding capability being lost.¹⁵ As previously reported, this calcium ion and the orientation of the vicinal hydroxyl groups, which are equatorially and axially arranged on the GalNAc residue are necessary for the carbohydrate binding.¹⁶

The second calcium ion in the CRD is located 8.41Å from the first one and through the interactions with Asp243 and Asp281 is effectively pinning together two parts of the long loop, appearing to hold it in place. The third calcium ion is located close to the C-terminus and plays no structural role in ligand recognition.

After molecular replacement, several Fo - Fc electron densities were found to be too big to be assigned as water molecules. The possibility that they are calcium or magnesium ions that are coordinating to histidine residues cannot be excluded.

A closer look at the carbohydrate binding site reveals that the presence of the QPDxW motif on the lectin, can once again, be associated to ligand interactions with the protein via the C3 and C4 hydroxyl groups.¹⁷ For all the GalNAc derivatives, hydrogen bonding is observed between GalNAc 4-OH, OD2/OD1_of Asp-269 and NE2 of Gln-267, for amino acid side chain nomenclature see figure 12. The GalNAc OH-3 hydrogen bonds with Asn 292 OD1/ND2 and with Glu 280 OE1/OE2, figure 12.

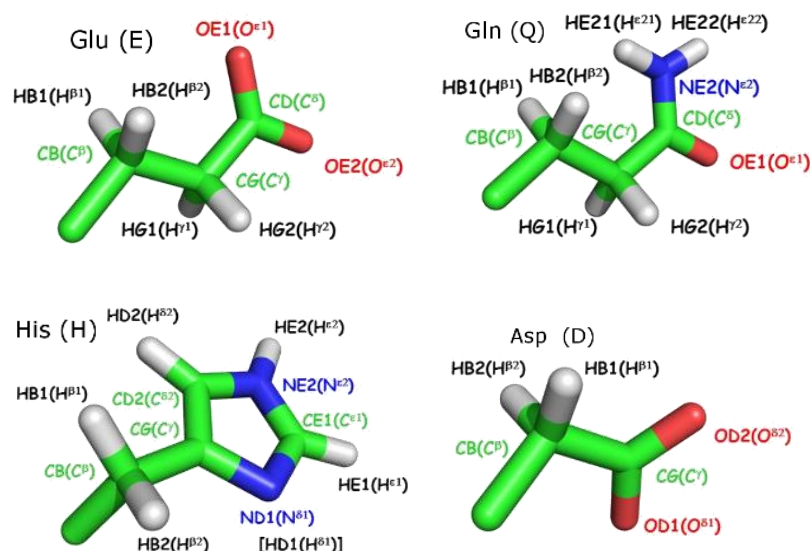


Figure 14. Amino acid side chain common nomenclature.

Of particular interest is the hydrogen bond between the carboxylic acid oxygen of the NAc moiety and NE2 of His286. As in other C-type lectins, the calcium ion coordination is believed to enhance the glycan CH polarisation and consequently enhance the CH- π

interaction between the bottom face of the sugar and aromatic residue, in this specific case Trp271, figure 13.¹⁸

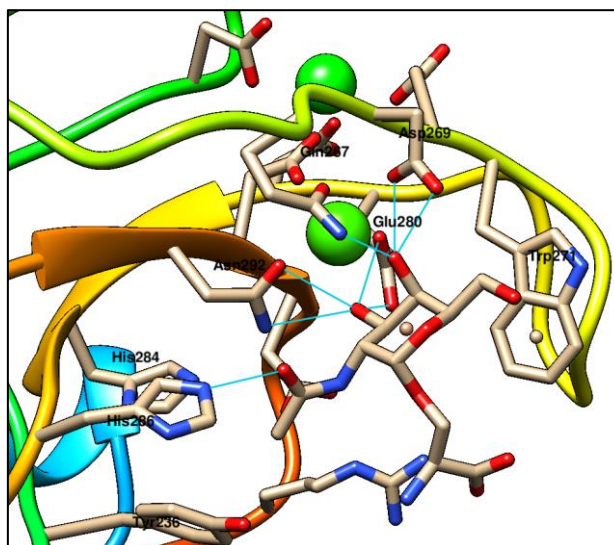


Figure 15. Crystal structure of hMGL CRD in complex with T_N antigen: hydrogen bonds in light blue with distance from the calcium ion shown in green.

Crystallisation in the presence of the multivalent ligands **AG143** and **AG124** was attempted at protein concentration of 70 mg/mL with the same crystallisation conditions. Ratios of protein to ligand of 1:1, 1:3, 1:4 and 1:10 were tested, however, none of the crystals obtained presented electron density for the ligand, figure 14.

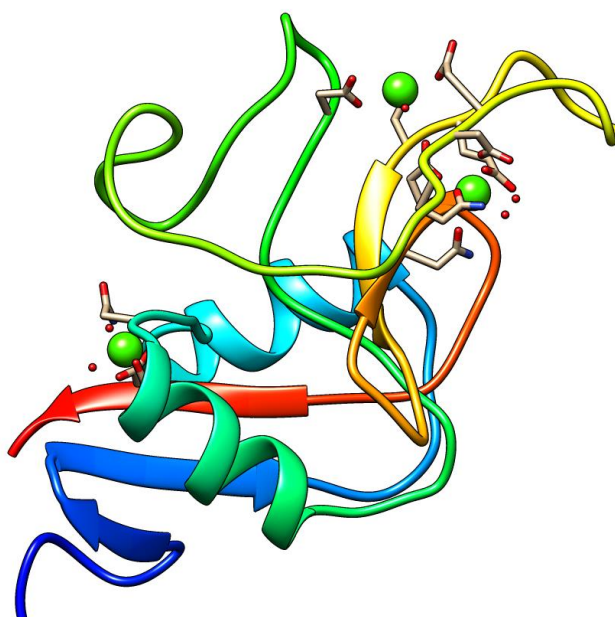


Figure 16. Ribbon diagram of the hMGL CRD crystallised in the presence of AG143.

Comparing ligand bound and unbound protein crystal structures also revealed different crystal packing arrangements. The unbound CRD packs in a less dense manner, while in the packing of hMGL CRD bound with GalNAc, CRDs are arranged in a shamrock shape figure 15.

The same supramolecular arrangement is visible in the absence of the ligand when three calcium ions are present, while no trimeric formation is observed in crystals obtained in the absence of calcium.

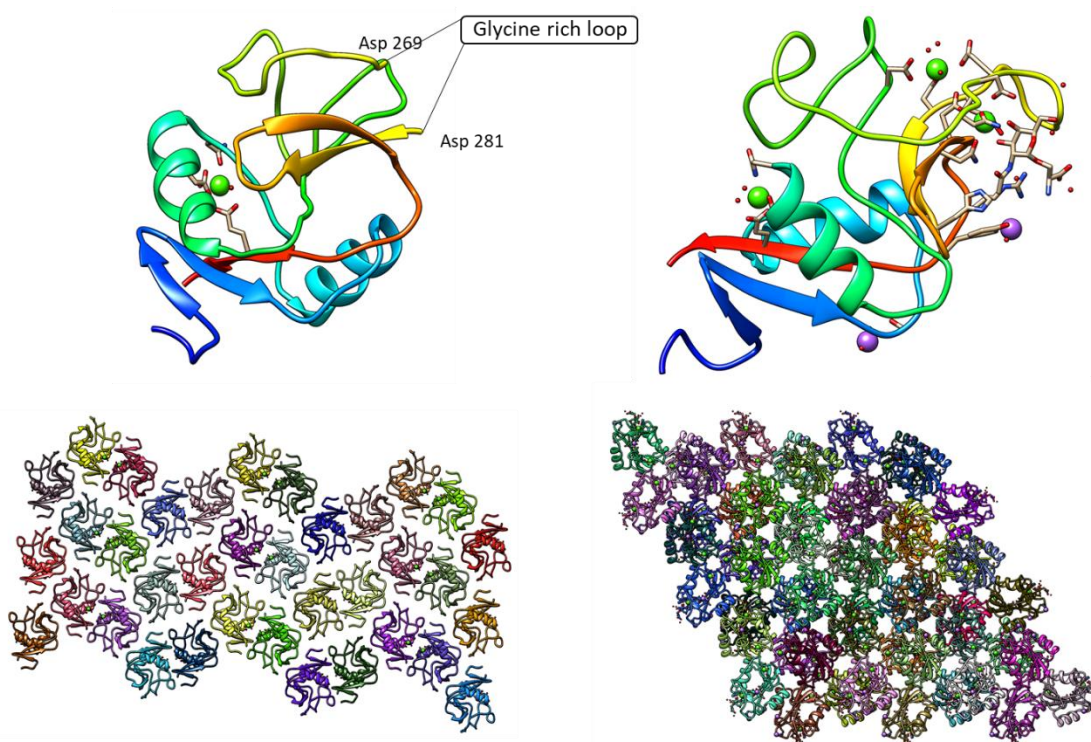


Figure 17. Crystal packing of the CRD of h MGL. On the left: Unbound CRD readily crystallised in the presence of calcium chelating agents in space group $P2_12_12_1$; on the right: CRD bound to T_N antigen in space group, $P3_12$. The supramolecular trimeric formation is observed in the absence of the coiled coil neck domain and seems to be modulated by the presence of calcium ion

5.3. Discussion

Examination of the crystal structure of hMGL CRD in complex with ligands reveals new insights into the preference for GalNAc over galactose.

Firstly, I briefly sum-up the state of the art already reported in chapter 1:

i) A literature search of C-type lectins crystal structure revealed that they can bind carbohydrates mainly in two different modes. The easiest way the reader can imagine the two different orientation of the ligand is to pretend that the OH3 and OH4 are glued to the calcium ion; for binding mode A the sugar is orientated so that the C6 is positioned on top of the tryptophan residue. All proteins belonging to this class present the characteristic sequence QPDxW. In mode B the sugar is flipped by 180° (remember: OH-3/4 are glued to the calcium, so it can be flipped only on one axis) so that the C6 is now pointing in the opposite direction. The proteins observed to bind ligands in mode B lack the QPDxW signature sequence and the tryptophan residue in the loop. In figure 16, the C-type lectins CEL-I and human Langerin are reported binding in the two different binding modes.

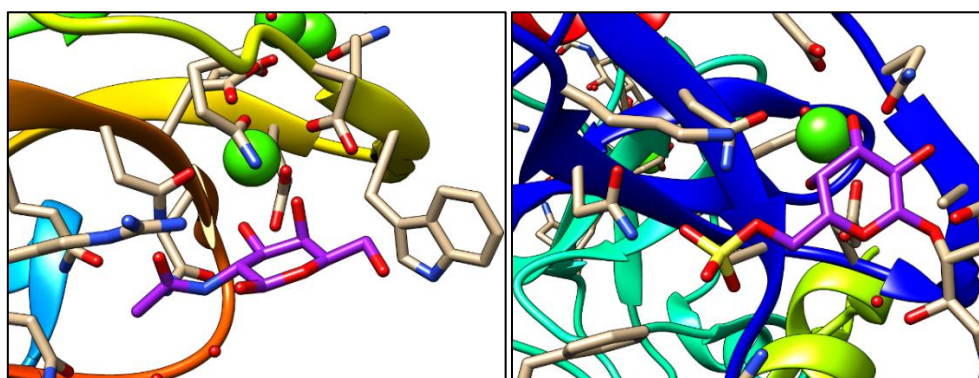


Figure 18. Crystal structures obtained from the PDB of two different C-Type lectin binding modes: on the left binding mode A CEL-I complexed with GalNAc (PDB code 1WMZ), on the right a C-type lectin expressed by Langerhans cells, complexed with a sulphated galactoside (PDB code 3P5I).

ii) hMGL has all the structural features typical of a protein which binds its ligands in mode A. Assuming this is the binding mode, it is possible to partially explain the GalNAc selectivity over Gal, using the closest cousin of hMGL, the asialoglycoprotein receptor (ASGPR). For ASGPR Drickamer and co-workers, in 1990, proposed that His 256 (His 284 hMGL) is responsible for this preference.¹⁷ This single residue was shown to be responsible for a 14-fold increase in affinity for GalNAc over galactose in ASPGR, figure 17. The increased affinity is mainly attributed to van der Waals contacts with the methyl group of the N-acetyl substituent of the sugar.^{19,20,14}

	GalNAc Derivative	Gal Derivative
	Relative binding affinity	
MT-ASGPR	(14)	(1)
hMGL	(70)	(1)

Figure 19. Relative binding affinity of a mannose binding lectin mutated with ASGPR feature and hMGL.^{17,19,20,14,21}

iii) To rationalise the different saturation transfer difference (STD) NMR pattern of methyl α -D-galactopyranose versus methyl 2-(acetylamino)-2-deoxy- α -D-galactopyranose, which had been obtained experimentally, Marcelo and co-workers, built a homology model of hMGL based on the ASGPR and then ran molecular dynamics simulations of the protein in the process of binding carbohydrates in binding mode A and B. Selected structures from molecular dynamics runs, were analysed with a program, which predicts the saturation of each proton depending on the geometry of the complex, Corcema-ST (complete relaxation and conformational exchange matrix-saturation transfer). This theoretical study revealed that while the galactose derivative STD signals could be explained solely by a binding mode A, a combination of binding mode A:B in ratio 1:4 was necessary to rationalise the GalNAc derivative signals, figure 18. In this scenario, a favourable CH- π interaction between the methyl group and the Trp at the bottom of the binding pocket could justify the preference for GalNAc.

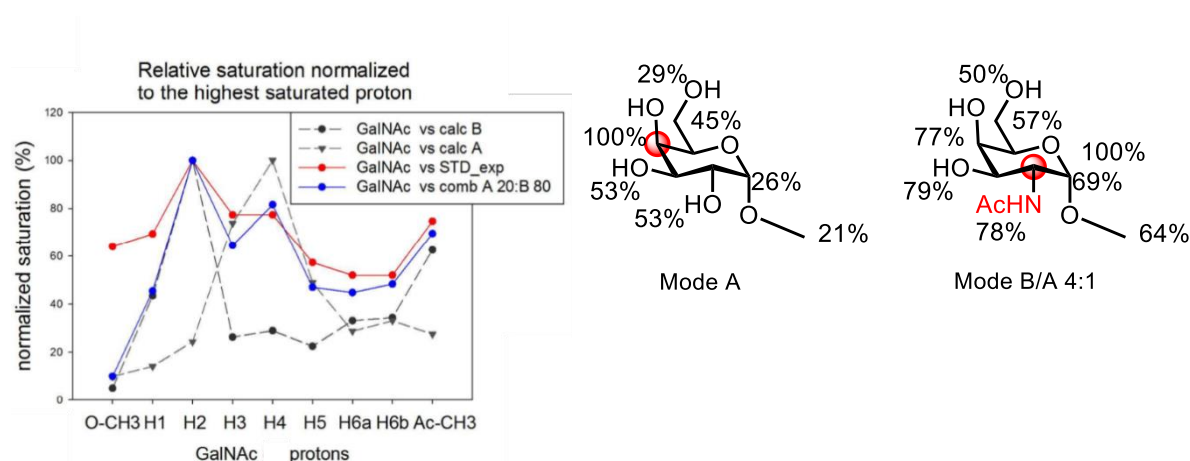


Figure 20. Left: STD NMR epitope mapping of methyl 1 α -D-galactopyranose and methyl 2-(acetylamino)-2-deoxy- α -D-galactopyranose. Right: graphical representation of the relative saturation of each proton of methyl 2-(acetylamino)-2-deoxy- α -D-galactopyranose for all four conditions, mode A, mode B, combination A and B (0,2A:0,8B) predicted with CORCEMA-ST versus experimental. Adapted with permission from Chem. - A Eur. J. 2014, 20, 16147–16155.²¹

As noted in section 5.2.6, binding mode A was exclusively observed in all the crystal structures determined herein. An increase in the relative affinity of hMGL was for methyl 2-(acetylamino)-2-deoxy- α -D-galactopyranose compared with methyl α -D-galactopyranose of over 70-fold,²¹ compared to 14-fold for ASGPR¹⁷ has been reported.

The enhanced affinity of hMGL for GalNAc derivatives therefore could be only partially justified by comparing with the known ASGPR model: the crystal structure of hMGL carbohydrate recognition domain discloses a slightly shorter distance than that observed in the homologue ASGPR between the His284 and the methyl group, 3.9 Å vs 4.0 Å, and similar van der Waals contact with the T_N-antigen. The van der Waals interaction is a relatively weak force ranging from 0.5 to 1 kcal/mol. The improved preference for GalNAc is most probably attributed to His286, which not only makes a van der Waals contact with the methyl group, but also a hydrogen bond with the NAc carbonyl oxygen, a moderate hydrogen bonds, which are the most common, are formed between neutral donors and acceptors are from 3 - 12 kcal/mol.²²

Supporting this theory, early this year, the work of Marcelo *et al.* showed the importance of His286 for glycan binding. When the His286 of MGL isoform 1 and 3, was mutated to threonine, the protein binding ability for the ligand was significantly reduced compared to the native protein. In some cases, mainly for ligand bearing blood group A GalNAc α 1-3(Fuc α 1-2)Gal β the interaction with MGL H286T was completely lost.²³ Binding mode A not only facilitates hydrogen bond formation between the ligand and His286, but also permits a close vicinity of the methyl group and the side chain of Tyr236, 4.0 Å from the centroid, figure 19.

Within these distances, CH₃- π interaction are likely also to occur.^{24,25} This interaction may justify the high STD resonance observed experimentally on H-2.

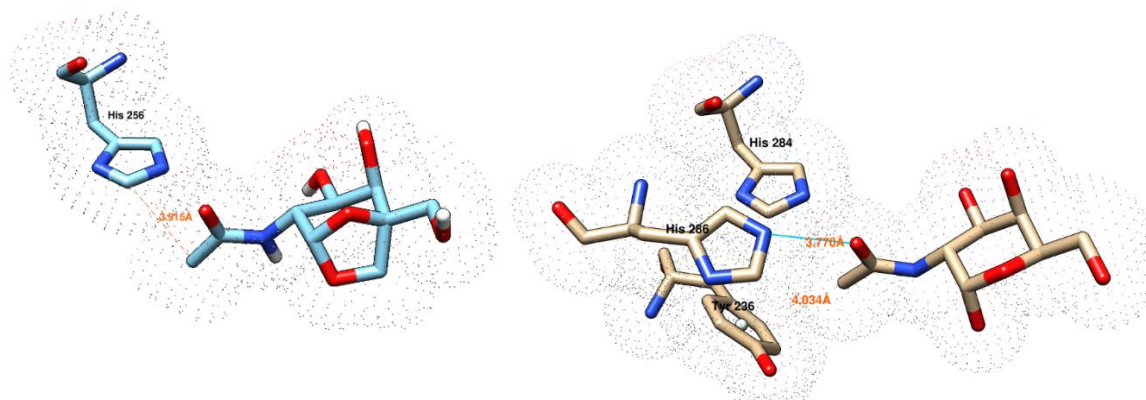


Figure 21. 2.0 Å Van der Waals dot surface representation of the GalNAc binding site. On the left: Van der Waals contacts His 256 of ASGPR with GalNAc derivative (PDB entry: 5JQ1). On the right: Tn antigen Van der Waals contacts of the homologue His 284 and His 286 of hMGL. The NAc carbonyl oxygen functions as a hydrogen bond acceptor of His 286, this is believed to be the driving force in the GalNAc selectivity over Gal.²⁶

Furthermore, the majority of carbohydrate binding proteins having an aromatic residue in the ligand binding loop,¹⁸ are observed binding in a mode so that the polar face of the sugar, in some cases further polarised by calcium interaction, can exploit CH- π interactions with the aromatic side chain.¹⁸ In this specific case, binding mode A permits the interaction of hydrogens on atoms C3, C4, C5 and C6²⁴ with Trp271.

I believe the contribution of His 286 and Tyr 236 could not have been predicted by homology models based on the ASGPR crystal structure, since these two amino acids are not conserved in hMGL and ASGPR, figure 20. Another amino acid not shared between the two proteins is Pro287, despite the fact no direct interaction between Pro287 and the glycans is observed, this proline, which is again a unique feature of hMGL, forces a twist in the hMGL backbone, which places His 286 in an optimal orientation for interaction with the ligand.

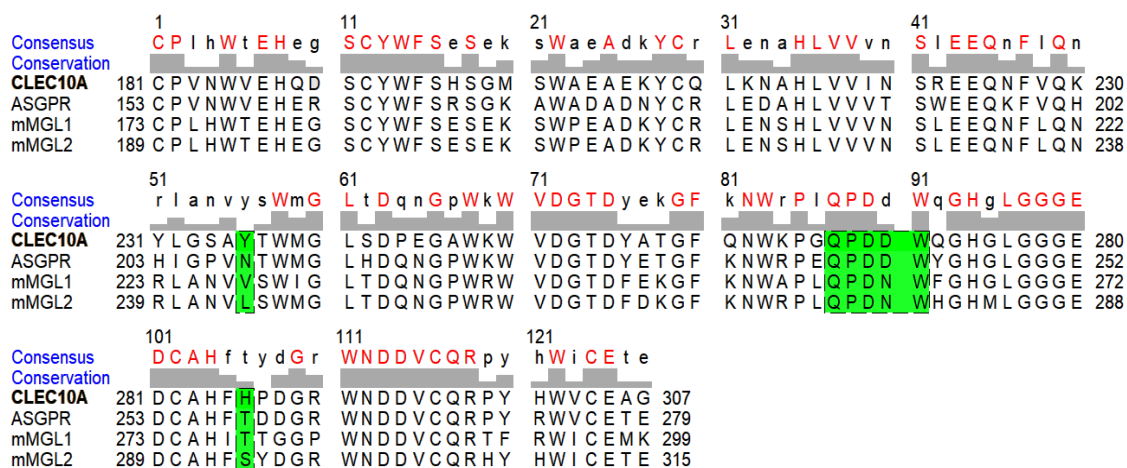


Figure 22. Aligned sequence comparison between human CLEC10A (hMGL), ASGPR-1, mMGL1 and mMGL2. Y236 and H286 are a unique feature of human CLEC10A. Pro287 twists hMGL backbone and allows His286 to interact with the NAc moieties.

The H286 and Y236 are unique features of human MGL: mMGL1 and mMGL2 present the characteristic sequence QPDxW but do not have this residue in the human analogue positions. This suggest that despite mMGL2 having affinity for GalNAc derivatives,²⁷ the molecular basis of its recognition is not the same as the human, and ligands designed to optimise interaction with H286 and Y236 will bind better to the human analogue but may not bind as well with the mouse model.

The possibility that GalNAc derivatives behave differently in solution compared to the crystalline state, cannot be excluded. The scientific community has relied on the ASGPR model when rationalising the behaviour of MGL. I believe it is important address the question as to whether ASGPR is a good model for MGL

To address this question, a collaboration with Filipa Marcelo and Javier F. Canada is currently planned. The coordinates of hMGL from the X-ray crystal structures will be used to predict STD NMR signal with Corcema-ST and to investigate if the experimental STD NMR data can be explained uniquely with the binding mode A.

5.4. Conclusion

In this chapter, the crystal structures of hMGL carbohydrate recognition domain and complexes with carbohydrate ligands and the Tn antigen are reported. The structure reveals the atomic details of the exquisite GalNAc selectivity displayed by this protein. The importance of His 286, which has been long alleged in the scientific community, is explained by the crystal structural evidence provided. CH₃- π interactions with Tyr236 were also observed in crystal structures. In all the crystals investigated, a single binding mode, corresponding to the mode A, was observed. We believe that the repetition of the dynamics simulations and Corcema-ST using hMGL protein coordinate, which considers His286, Pro 287 and Tyr 236 in the sequence, could result in a unique binding model consistent with the one observed in the solid-state and this is currently under investigation. Until the calculations are not repeated though, we cannot exclude the existence of a different binding mode in solution as evidenced by earlier NMR spectroscopy experiments.

These data provide the three-dimensional information necessary for the structure-based design of ligands with higher affinity and specificity for hMGL. It may also be possible and to engineer proteins with high affinity for the cancer biomarker T_N antigen which could be used for detection or therapeutic purposes by targeting cancer cells for drug delivery.

Many questions still need to be answered about the role of hMGL in the immune system, its glycan internalisation/processing and the structural basis of its complete structure. I must admit with a bit of melancholy, this will be a chapter of the life of another fortunate PhD candidate.

5.5. Bibliography

- (1) Ardá, A.; Jiménez-Barbero, J. The Recognition of Glycans by Protein Receptors. Insights from NMR Spectroscopy. *Chem. Commun.* **2018**, *54* (38), 4761–4769. <https://doi.org/10.1039/c8cc01444b>.
- (2) Hartley, D. L.; Kane, J. F. Properties of Inclusion Bodies from Recombinant Escherichia Coli. *Biochem. Soc. Trans.* **1988**, *16* (2), 101–102. <https://doi.org/10.1042/bst0160101>.
- (3) Wang, Y.; Van Oosterwijk, N.; Ali, A. M.; Adawy, A.; Anindya, A. L.; Dömling, A. S. S.; Groves, M. R. A Systematic Protein Refolding Screen Method Using the DGR Approach Reveals That Time and Secondary TSA Are Essential Variables. *Sci. Rep.* **2017**, *7* (1), 1–10. <https://doi.org/10.1038/s41598-017-09687-z>.
- (4) Lee, M. E.; Dou, X.; Zhu, Y.; Phillips, K. J. Refolding Proteins from Inclusion Bodies Using Differential Scanning Fluorimetry Guided (DGR) Protein Refolding and MeltTraceur Web. *Curr. Protoc. Mol. Biol.* **2019**, *125* (1), 1–16. <https://doi.org/10.1002/cpmb.78>.
- (5) Otwinowski, Z.; Minor, W. Processing of X-Ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* **1997**, *276* (January 1993), 307–326. [https://doi.org/10.1016/S0076-6879\(97\)76066-X](https://doi.org/10.1016/S0076-6879(97)76066-X).
- (6) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (4), 658–674. <https://doi.org/10.1107/S0021889807021206>.
- (7) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (4), 355–367. <https://doi.org/10.1107/S0907444911001314>.
- (8) Potterton, E.; Briggs, P.; Turkenburg, M.; Dodson, E. A Graphical User Interface to the CCP4 Program Suite. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **2003**, *59* (7), 1131–1137. <https://doi.org/10.1107/S0907444903008126>.
- (9) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (12 I), 2126–2132. <https://doi.org/10.1107/S0907444904019158>.
- (10) Schüttelkopf, A. W.; Van Aalten, D. M. F. PRODRG: A Tool for High-Throughput Crystallography of Protein-Ligand Complexes. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (8), 1355–1363. <https://doi.org/10.1107/S0907444904011679>.
- (11) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. <https://doi.org/10.1002/jcc.20084>.
- (12) Meier, M.; Bider, M. D.; Malashkevich, V. N.; Spiess, M.; Burkhard, P. Crystal Structure of the Carbohydrate Recognition Domain of the H1 Subunit of the Asialoglycoprotein Receptor. *J. Mol. Biol.* **2000**, *300* (4), 857–865. <https://doi.org/10.1006/jmbi.2000.3853>.
- (13) Kolatkar, A. R.; Weis, W. I.; Chem, B. Structural Basis of Galactose Recognition by C-Type Animal Lectins * Protein A by Changing Three Amino Acids and Inserting A. **1996**, *271* (12), 6679–6685.
- (14) Sanhueza, C. A.; Baksh, M. M.; Thuma, B.; Roy, M. D.; Dutta, S.; Préville, C.; Chrnyk, B. A.; Beaumont, K.; Dullea, R.; Ammirati, M.; et al. Efficient Liver Targeting by Polyvalent Display of a Compact Ligand for the Asialoglycoprotein Receptor. *J. Am. Chem. Soc.* **2017**, *139* (9), 3528–3536. <https://doi.org/10.1021/jacs.6b12964>.
- (15) van Vliet, S. J.; Saeland, E.; van Kooyk, Y. Sweet Preferences of MGL: Carbohydrate Specificity and Function. *Trends Immunol.* **2008**, *29* (2), 83–90. <https://doi.org/10.1016/j.it.2007.10.010>.
- (16) Jégouzo, S. A.; Quintero-Martínez, A.; Ouyang, X.; Dos Santos, Á.; Taylor, M. E.; Drickamer, K. Organization of the Extracellular Portion of the Macrophage Galactose Receptor: A Trimeric Cluster of Simple Binding Sites for N-Acetylgalactosamine. *Glycobiology* **2013**, *23* (7), 853–864. <https://doi.org/10.1093/glycob/cwt022>.
- (17) Drickamer, K. Engineering Galactose-Binding Activity into a C-Type Mannose-Binding Protein.

- Nature* **1992**, *360* (6400), 183–186. <https://doi.org/10.1038/360183a0>.
- (18) Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N. Carbohydrate-Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **2015**, *137* (48), 15152–15160. <https://doi.org/10.1021/jacs.5b08424>.
- (19) Iobst, S. T.; Drickamer, K. Selective Sugar Binding to the Carbohydrate Recognition Domains of the Rat Hepatic and Macrophage Asialoglycoprotein Receptors. *J. Biol. Chem.* **1996**, *271* (12), 6686–6693. <https://doi.org/10.1074/jbc.271.12.6686>.
- (20) Kolatkar, A. R.; Leung, A. K.; Isecke, R.; Brossmer, R.; Drickamer, K.; Weis, W. I. Mechanism of N-Acetylgalactosamine Binding to a C-Type Animal Lectin Carbohydrate-Recognition Domain. *J. Biol. Chem.* **1998**, *273* (31), 19502–19508. <https://doi.org/10.1074/jbc.273.31.19502>.
- (21) Marcelo, F.; Garcia-Martin, F.; Matsushita, T.; Sardinha, J.; Coelho, H.; Oude-Vrielink, A.; Koller, C.; André, S.; Cabrita, E. J.; Gabius, H. J.; et al. Delineating Binding Modes of Gal/GalNAc and Structural Elements of the Molecular Recognition of Tumor-Associated Mucin Glycopeptides by the Human Macrophage Galactose-Type Lectin. *Chem. - A Eur. J.* **2014**, *20* (49), 16147–16155. <https://doi.org/10.1002/chem.201404566>.
- (22) Kim, K.; Friesner, R. A. Hydrogen Bonding between Amino Acid Backbone and Side Chain Analogues: A High-Level Ab Initio Study. *J. Am. Chem. Soc.* **1997**, *119* (52), 12952–12961. <https://doi.org/10.1021/ja971836d>.
- (23) Marcelo, F.; Supekar, N.; Corzana, F.; van der Horst, J. C.; Vuist, I. M.; Live, D.; Boons, G.-J. P. H.; Smith, D. F.; van Vliet, S. J. Identification of a Secondary Binding Site in Human Macrophage Galactose-Type Lectin by Microarray Studies: Implications for the Molecular Recognition of Its Ligands. *J. Biol. Chem.* **2018**, jbc.RA118.004957. <https://doi.org/10.1074/jbc.RA118.004957>.
- (24) Spiwok, V. CH/ π Interactions in Carbohydrate Recognition. *Molecules* **2017**, *22* (7), 1–11. <https://doi.org/10.3390/molecules22071038>.
- (25) Jiménez-Moreno, E.; Jiménez-Osés, G.; Gómez, A. M.; Santana, A. G.; Corzana, F.; Bastida, A.; Jiménez-Barbero, J.; Asensio, J. L. A Thorough Experimental Study of CH/ π Interactions in Water: Quantitative Structure-Stability Relationships for Carbohydrate/Aromatic Complexes. *Chem. Sci.* **2015**, *6* (11), 6076–6085. <https://doi.org/10.1039/c5sc02108a>.
- (26) Marcelo, F.; Supekar, N.; Corzana, F.; Van Der Horst, J. C.; Vuist, I. M.; Live, D.; Boons, G. J. P. H.; Smith, D. F.; Van Vliet, S. J. Identification of a Secondary Binding Site in Human Macrophage Galactose-Type Lectin by Microarray Studies: Implications for the Molecular Recognition of Its Ligands. *J. Biol. Chem.* **2019**, *294* (4), 1300–1311. <https://doi.org/10.1074/jbc.RA118.004957>.
- (27) Singh, S. K.; Streng-Ouwehand, I.; Litjens, M.; Weelij, D. R.; García-Vallejo, J. J.; van Vliet, S. J.; Saeland, E.; van Kooyk, Y. Characterization of Murine MGL1 and MGL2 C-Type Lectins: Distinct Glycan Specificities and Tumor Binding Properties. *Mol. Immunol.* **2009**, *46* (6), 1240–1249. <https://doi.org/10.1016/j.molimm.2008.11.021>.

6. Experimental section

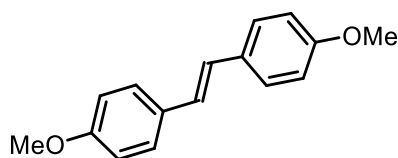
6.1. Synthesis: general experimental conditions

NMR spectra were recorded with 500 & 600 MHz Varian spectrometers. Chemical shifts are reported relative to internal Me₄Si in CDCl₃ (δ 0.0), HOD for D₂O (δ 4.84) or CD₂HOD (δ 3.31) for ¹H and CDCl₃ (77.16) or CD₃OD (49.05) for ¹³C. NMR spectra were processed and analysed using MestReNova software. ¹H NMR signals were assigned with the aid of gCOSY. ¹³C NMR signals were assigned with the aid of APT, gHSQCAD and/or gHMBCAD. Coupling constants are reported as observed in Hertz. Low and high-resolution mass spectra were measured on a Waters LCT Premier XE Spectrometer, measuring in both positive and/or negative mode as, using MeCN, H₂O and/or MeOH as solvent. Thin layer chromatography (TLC) was performed on aluminium sheets precoated with silica gel 60 (HF254, E. Merck) and spots visualized by UV and charring with H₂SO₄-EtOH (1:20), cerium molybdate stain, or phosphomolybdic acid stain. Flash chromatography was carried out with silica gel 60 (0.040-0.630 mm, E. Merck or Aldrich) and using a stepwise solvent polarity gradient correlated with TLC mobility. Chromatography solvents, cyclohexane, EtOAc, CH₂Cl₂ and MeOH were used as obtained from suppliers (Fisher Scientific and Sigma-Aldrich). Anhydrous pyridine was purchased from Sigma Aldrich with further reactions solvents, for anhydrous conditions, being obtained from a Pure Solv™ Solvent Purification System.

6.2. Small molecule experimental procedure and characterisation

AG120¹, **Lactose azide**² and **AG140**³ were synthesised as previously reported. The analytical data are in perfect agreement with what is reported in literature.

(E)-1,2-bis(4-methoxyphenyl)ethane (AG085)⁴



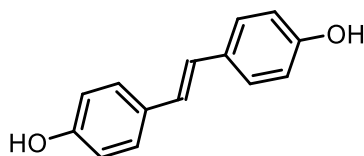
To a solution of diethyl 4-methoxybenzyl phosphonate (10 g, 38.7 mmol) in DMF, NaOMe (6.27 g, 116 mmol) was added. The reaction mixture was stirred at room temperature for 20 min then *p*-anisaldehyde (5.66 mL, 46.5 mmol) was added. The solution was allowed to stir for 30 min at room temperature, then warmed to 100°C for 8 h. The reaction mixture was quenched with water. The filtration of the white precipitate afforded the desired product (8.5 g, 92% yield).

¹H NMR (500 MHz, chloroform-*d*) δ 7.43 (d, *J* = 8.8 Hz, 4H, aromatic H), 6.93 (s, 2H, alkene CH), 6.89 (d, *J* = 8.8 Hz, 4H, aromatic H), 3.83 (s, 6H, OMe).

¹³C NMR (125 MHz, chloroform-*d*) δ = 159.0 (C), 130.5 (C), 127.4 (CH aromatic), 126.2 (CH alkene), 114.1 (CH aromatic), 55.3 (CH₃, OMe).

ES-HRMS calcd for C₁₆H₁₆ NaO₂ 263.1048, found *m/z* 263.1045 [M+Na]⁺

Melting temperature = 213–214°C

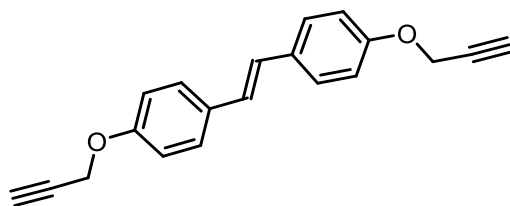
(E)-4,4'-(Ethene-1,2-diyl) diphenol (AG087) ⁴

To AG085 (2.5 g, 10.4 mmol) in CH_2Cl_2 (40 mL), cooled with an ice-salt bath, was added, dropwise, a 1M boron tribromide solution in CH_2Cl_2 (39 mL, 39.0 mmol). The solution was allowed warm to room temperature and was stirred overnight. Water (35 mL) was added slowly, dropwise, with stirring. The organic solvent was removed under reduced pressure, the aqueous phase was extracted with EtOAc (3 x 30 mL), the combined organic extracts were dried over Na_2SO_4 and the solvent removed under reduced pressure. Column chromatography (Cyclohexane-EtOAc 3:7) yielded the desired compound (2.1 g, 97%)

^1H NMR (500 MHz, DMSO-d_6) δ = 7.33 (d, J = 8.7 Hz, 4H, aromatic H), 6.88 (s, 2H, alkene H), 6.72 (d, J = 8.6 Hz, 4H, aromatic H).

^{13}C NMR (125 MHz, DMSO-d_6) δ = 157.2 (C), 129.0 (C), 127.8 (CH aromatic), 125.6 (CH, alkene), 115.9 (CH aromatic).

ES-HRMS calcd for $\text{C}_{14}\text{H}_{11}\text{O}_2$ 211.0759, found m/z 211.0702 $[\text{M-H}]^-$

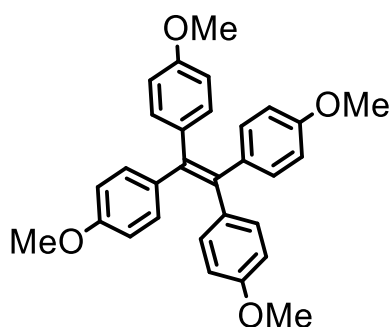
(E)-1,2-Bis(4-(prop-2-yn-1-yloxy)phenyl)ethane (AG089)⁵

To a solution of AG087 (2.1 g, 9.9 mmol) in dry DMF (300 mL) was added anhydrous potassium carbonate (10.96 g, 79.3 mmol). Propargyl bromide (80% in toluene, 4.83 mL, 54.5 mmol) was then added and the mixture was stirred for 14 h at 70 °C under a nitrogen atmosphere. The reaction was then cooled to room temperature and diluted with CH₂Cl₂ (300 mL). This solution was washed with saturated NH₄Cl, with the aqueous layer being re-extracted with a further portion of CH₂Cl₂ (3 x 150 mL). The combined organic layers were then washed with water and dried over Na₂SO₄. The solvent was removed under diminished pressure. Column chromatography (Cyclohexane-EtOAc 7:3) gave the desired compound (2.43 g, 85%).

¹H NMR (500 MHz, chloroform-d) δ 7.44 (d, J = 8.7 Hz, 4H, aromatic H), 6.97 (d, J = 8.7 Hz, 4H, aromatic H), 6.94 (s, 2H, alkene H), 4.71 (s, 4H, CH₂), 2.53 (s, 2H, alkyne H).

¹³C NMR (125 MHz, chloroform-d) δ = 157.4 (C), 131.3 (C), 127.4 (aromatic CH), 126.5 (alkene CH), 115.1 (aromatic CH), 80.3 (C), 75.6 (alkyne CH), 55.9 (CH₂).

ES-HRMS calcd for C₂₀H₁₆NaO₂ 311.1048, found m/z 311.1052[M+Na]⁺

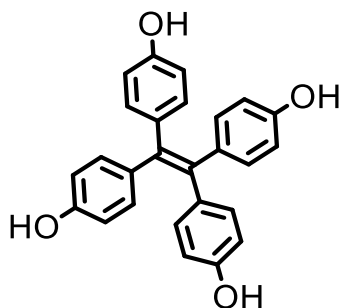
1,1,2,2-tetrakis(4-methoxyphenyl)ethene (AG083)⁶

To a solution of 4,4'-dimethylbenzophenone (5 g, 20.6 mmol) in dry THF (100 mL), zinc powder (6.60 g, 101 mmol) was added. To this stirring suspension titanium tetrachloride (7.50 mL, 68.4 mmol) was added dropwise, slowly and carefully (!!!), at room temperature. The mixture was stirred at reflux for 16 h. To the cooled reaction mixture water (100 mL) was added slowly. The mixture was diluted with CH₂Cl₂ (100 mL) and the layers separated. The aqueous layer was extracted 3 times with CH₂Cl₂ (3 x 50 mL). The combined organic layers were dried over Na₂SO₄ and the solvent removed under reduced pressure. Column chromatography (CH₂Cl₂-cyclohexane 1:1) gave the desired compound (3.63 g, 78%) as a colourless solid. (*Note*: if traces of titanium tetrachloride are present, the solid will be slightly purple, this will not influence the yield of the following step).

¹H NMR (500 MHz, chloroform-d) δ = 6.93 (d, J = 8.8 Hz, 8H, aromatic H), 6.64 (d, J = 8.8 Hz, 8H, aromatic H), 3.74 (s, 12H, OMe).

¹³C NMR (125 MHz, chloroform-d) δ = δ 157.7 (C), 138.3 (C), 136.9 (C), 132.5 (CH, aromatic), 113.0 (CH, aromatic), 55.1 (CH₃).

ES-HRMS calcd for C₃₀H₂₉O₄ 453.2066, found m/z 453.2068 [M+H]⁺

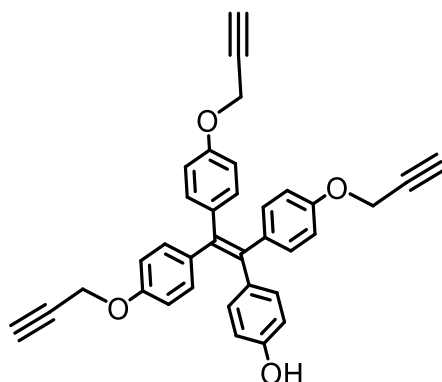
4,4',4'',4'''-(ethene-1,1,2,2-tetrayl)tetraphenol AG084⁶

To a solution of AG083 (3.63 g, 8.0 mmol) in CH_2Cl_2 (30 mL), cooled with an ice-salt bath, was added, dropwise, a 1M boron tribromide solution in CH_2Cl_2 (39 mL, 39.0 mmol). The ice bath was then removed, the solution was allowed to warm to room temperature and stirred overnight. Water (35 mL) was added slowly, dropwise, with stirring. The precipitate was collected by filtration, with the solid washed with clean portions of water. Recrystallisation of this solid from acetone-water (1:1, 100 mL) gave the desired compound (2.96 g, 93%) as a colourless solid.

^1H NMR (500 MHz, $\text{DMSO}-d_6$) δ = 9.19 (s, 4H, OH), 6.68 (d, J = 8.5 Hz, 8H, aromatic H), 6.46 (d, J = 8.5 Hz, 8H, aromatic H).

^{13}C NMR (125 MHz, DMSO) δ = 155.8 (C), 138.1 (C), 135.5 (C), 132.4 (CH, aromatic), 114.9 (CH, aromatic).

ES-HRMS calcd for $\text{C}_{26}\text{H}_{19}\text{O}_4$ 395.1283, found m/z 395.1279 $[\text{M}-\text{H}]^-$

4-(1,2,2-tris(4-(prop-2-yn-1-yloxy)phenyl)vinyl)phenol (AG141/AG178)⁷

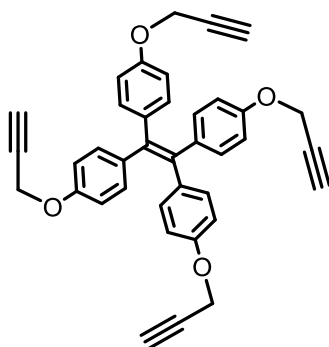
To a solution of AG084 (2.50 g, 6.3 mmol) in dry DMF (50 mL) K_2CO_3 (2.79 g, 20.2 mmol) was slowly added followed by propargylbromide 80% in toluene (2.29 ml, 20.8 mmol). The reaction mixture was heated at 70° C for 16 h, allowed to cool down to room temperature, then diluted with DCM and washed with HCl 1M, brine and water.

The organic phase was dried over Na_2SO_4 and the solvent removed under reduced pressure. Colum chromatography (8.7:1.3 cyclohexane-EtOAc) yielded the titled compound (1.1 g, 34%).

1H (500 MHz, Chloroform-d) δ 6.93 (d, $J = 8.7$ Hz, 6H, aromatic H), 6.87 (d, $J = 7.7$ Hz, 2H, aromatic H), 6.70 (d, $J = 8.0$ Hz, 6H, aromatic H), 6.56 (d, $J = 8.6$ Hz, 2H, aromatic H), 4.62 (s, 6H, CH_2), 3.50 (br s, 1H, OH), 2.50 (s, 3H, alkyne).

^{13}C NMR (125 MHz, Chloroform-d) δ 158.9 (C), 156.1(C), 154.0 (C), 138.8(C), 137.6 (C), 132.9 (CH, aromatic), 132.7 (CH, aromatic), 114.8 (CH, aromatic), 114.2(CH, aromatic), 78.8 (C), 75.5(CH), 55.9 (CH_2).

ES-HRMS calcd for $C_{35}H_{25}O_4$ 509.1753, found m/z 509.1759 $[M-H]^-$

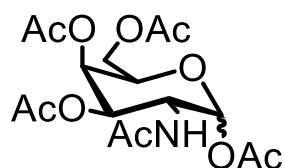
1,1,2,2-tetrakis(4-(prop-2-yn-1-yloxy)phenyl)ethene (AG088) ⁶

To a solution of AG084 (2.96 g, 7.47 mmol) in DMF (240 mL) was added anhydrous potassium carbonate (8.34 g, 60.3 mmol). Propargyl bromide (80% in toluene, 3.70 mL, 33.2 mmol) was then added to the reaction mixture and stirred for 14 h at 70 °C under a nitrogen atmosphere. The reaction was then cooled to room temperature and diluted with CH₂Cl₂. This solution was washed with water, with the aqueous layer being re-extracted with a further portion of CH₂Cl₂. The combined organic layers were then washed with water and dried over Na₂SO₄. The solvent was removed under diminished pressure. Column chromatography (1:1 CH₂Cl₂-cyclohexane) gave the desired compound (2.95 g, 72%) as yellow solid.

¹H (500 MHz, Chloroform-d) δ 6.93 (d, *J* = 8.8 Hz, 8H, aromatic H), 6.70 (d, *J* = 8.8 Hz, 8H, aromatic H), 4.62 (d, *J* = 2.4 Hz, 8H, CH₂), 2.50 (t, *J* = 2.4 Hz, 4H, alkyne).

¹³C NMR (125 MHz, Chloroform-d) δ 155.96 (C), 138.56 (C), 137.41 (C), 132.49 (CH, aromatic), 113.99 (CH, aromatic), 78.60 (C), 75.37 (CH), 55.77 (CH₂).

ES-HRMS calcd for C₃₈H₂₉O₈ 549.2066, found *m/z* 549.2063 [M+H]⁺

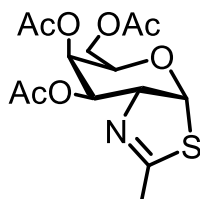
2-Acetamido-2-deoxy-1,3,4,6-tetra-*O*-acetyl- β -D-galactopyranose (AG095)⁸

To a 0 °C cooled solution of D-galactosamine hydrochloride (4.0 g, 18.5 mmol) in pyridine (60 mL), acetic anhydride (15 mL, 158.5 mmol) was added slowly (!!!), portion wise. The reaction was allowed to reach room temperature and stirred overnight. The solvents were then removed by co-evaporation with toluene until a colourless solid was formed. Recrystallisation of this solid from MeOH gave the title compound (5.7 g, 79%) as a colourless solid.

¹H NMR (500 MHz, chloroform-d) δ = 5.70 (d, J = 8.8 Hz, 1H, H-1), 5.37 (dd, J = 3.6, 1.2 Hz, 1H, H-4), 5.38 (d, J = 9.4 Hz, 1H, NH) 5.08 (dd, J = 11.3, 3.3 Hz, 1H, H-3), 4.45 (d_{aprnt}, J = 11.4, 9.4, 8.8 Hz, 1H, H-2), 4.17 (dd, J = 11.3, 6.6 Hz, 1H, H-6a), 4.11 (dd, J = 11.3, 6.5 Hz, 1H, H- 6b), 4.01 (aprntd, J = 6.5, 6.6 ,1.2 Hz, 1H, H-5), 2.17 (s, 3H, AcOMe), 2.13 (s, 3H, AcOMe), 2.05 (s, 3H, AcOMe), 2.02 (s, 3H, AcOMe), 1.94 (s, 3H, NHAc).

¹³C NMR (125 MHz, chloroform-d) δ = 170.7(C), 170.4 (C), 170.2(C) , 170.2(C), 169.5 (C), 93.0 (CH,C-1), 71.9 (CH, C-5), 70.3 (CH, C-3), 66.3 (CH, C-4), 61.3 (CH₂, C-6), 49.8 (CH,C-2), 23.3 (CH₃), 20.9 (CH₃), 20.6 (CH₃), 20.6 (CH₃), 20.6 (CH₃).

ES-HRMS calcd for C₁₆H₂₃NO₁₀Na 412.1220, found m/z 412.1220 [M+Na]⁺

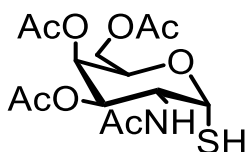
(3aR,5R,6R,7R,7aR)-5-(Acetoxymethyl)-6,7-diacetoxy-2-methyl-5,6,7,7a-tetrahydro-3aH-pyrano[3,2-d]thiazole (AG097)⁹

AG095 (1.00 g, 2.57 mmol) (*note*: scaling-up results in yield lost) was stirred in toluene (12 mL) at 50 °C until the sugar is completely dissolved. Lawessons reagent (890 mg, 2.20 mmol) was added, the reaction was stirred at 80 °C for 2.5 h. The reaction mixture was then cooled to room temperature and neutralised by the addition of NaHCO₃ (120 mg). Chromatography of the residue (3:7-1:1 EtOAc-CH₂Cl₂) gave the title compound (708 mg, 80%) as a yellow solid.

¹H NMR (500 MHz, chloroform-d) δ = 6.25 (d, *J* = 6.3 Hz, 1H, H-1), 5.52 (appnt, *J* = 3.3 Hz, 1H, H-4), 5.21 (dd, *J* = 7.9, 3.3 Hz, 1H, H-3), 4.39 (ddd, *J* = 7.9, 6.3, 1.6 Hz, 1H, H-2), 4.30 (ddd, *J* = 7.4, 5.6, 3.4 Hz, 1H, H-5), 4.23 (dd, *J* = 11.5, 7.4 Hz, 1H, H-6a), 4.10 (dd, *J* = 11.3, 5.5 Hz, 1H, H-6b), 2.27 (d, *J* = 1.6 Hz, 3H, thiazoline CH₃), 2.14 (s, 3H, AcOMe), 2.09 (s, 3H, AcOMe), 2.06 (s, 3H, AcOMe) .

¹³C NMR (125 MHz, chloroform-d) δ = 170.5 (C), 170.2(C), 169.8 (C), 89.0 (CH, C-1), 74.5 (CH, C- 2), 70.7 (CH, C-3), 70.6 (CH, C-5), 66.3 (CH, C-4), 60.9 (CH₂,C-6), 21.5 (CH₃), 20.9(CH₃), 20.7(CH₃), 20.6 (CH₃).

ES-HRMS calcd for C₁₄H₂₀NO₇S 346.0960, found *m/z* 346.0957 [M+H]⁺

2-Acetamido-2-deoxy-3,4,6-tri-*O*-acetyl-1-thio- α -D-galactopyranose (AG108)⁹

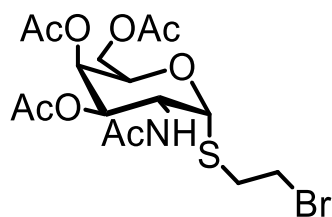
To a 0°C cooled solution of AG097 (545 mg, 1.58 mmol) in MeOH (6 mL), TFA (12 drops) and H₂O (12 drops) were added. The reaction was allowed warm-up at room temperature and stirred for 2 h. The solvent was removed under diminished pressure to give the title compound (573 mg, > 95%) as a colourless solid. (*Note*: disulfide bond formation can result in yield loss. To prevent formation, you can add few drops of 2-mercaptoethanol)

¹H NMR (500 MHz, chloroform-d) δ = 5.87 (dd, J = 6.4, 5.4 Hz, 1H, H-1), 5.70 (d, J = 8.5 Hz, 1H, NH), 5.41 (dd, J = 3.4, 1.3 Hz, 1H, H-4), 5.08 (dd, J = 11.7, 3.3 Hz, 1H, H-3), 4.75 (ddd, J = 11.6, 8.5, 5.1 Hz, 1H, H-2), 4.55 – 4.51 (m, 1H, H-5), 4.17 (dd, J = 11.3, 6.4 Hz, 1H, H-6a), 4.06 (dd, J = 11.3, 6.6 Hz, 1H, H-6b), 2.17 (s, 3H, AcOMe), 2.06 (s, 3H, AcOMe), 2.03 (s, 3H, AcOMe), 2.01 (s, 3H, AcOMe), 1.96 (d, J = 6.6 Hz, 1H, SH).

¹³C NMR (125 MHz, chloroform-d) δ = 171.28 (C), 170.6 (C), 170.4 (C), 170.2 (C), 79.8 (CH, C-1), 67.9 (CH, C-3 or C-5), 67.8 (CH, C-3 or C-5), 67.1 (CH, C-4), 61.57 (CH₂, C-6), 48.6 (CH, C-2), 23.2 (CH₃, NHAc), 20.7 (CH₃), 20.70 (CH₃), 20.68 (CH₃).

ES-HRMS calcd for C₁₄H₂₂NO₈S 364.1066, found m/z 364.1071 [M+H]⁺

2-Bromoethyl 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-1-thio- α -D-galactopyranoside (AG122)¹⁰



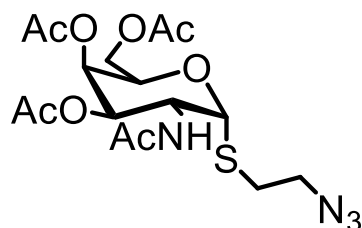
To a solution of AG108 (2.50 g, 6.88 mmol) in an acetone-H₂O (2:1, 30 mL) potassium carbonate (1.13 g, 8.18 mmol) and 1,2 dibromoethane (4.8 mL, 55.5 mmol) were added. This mixture was stirred at room temperature for 3 h. CH₂Cl₂ (50 mL) was added, the layers were separated, and the aqueous layer was re-extracted with CH₂Cl₂ (3 x 30 mL). The combined organic layers were dried over Na₂SO₄ and then concentrated under diminished pressure. Chromatography (Cyclohexane-EtOAc, 3:1) gave the desired compound (2.25 g, 69%) as a colourless solid.

¹H NMR (500 MHz, chloroform-d) δ = 5.76 (d, J = 8.5 Hz, 1H, NH), 5.60 (d, J = 5.3 Hz, 1H, H-1), 5.38 (dd, J = 3.3, 1.3 Hz, 1H, H-4), 5.03 (dd, J = 11.8, 3.3 Hz, 1H, H-3), 4.74 (ddd, J = 11.8, 8.5, 5.3 Hz, 1H, H-2), 4.60 – 4.50 (m, 1H, H-5), 4.13 (dd, J = 11.5, 5.3 Hz, 1H, H-6a), 4.09 (dd, J = 11.5, 7.4 Hz, 1H, H-6b), 3.59 (td, J = 10.0, 5.7 Hz, 1H,), 3.48 (td, J = 10.0, 6.1 Hz, 1H,), 3.10 (ddd, J = 13.9, 10.1, 6.1 Hz, 1H,), 3.01 (ddd, J = 13.9, 10.1, 5.7 Hz, 1H,), 2.16 (s, 3H, AcOMe), 2.07 (s, 3H, AcOMe), 2.01 (s, 3H, AcOMe), 1.97 (s, 3H, NHAc).

¹³C NMR (125 MHz, chloroform-d) δ = 171.1 (C), 170.4 (C), 170.8 (C), 170.4 (C), 85.6 (CH, C-1), 68.2 (CH, C-3), 67.7 (CH, C-5), 67.3 (CH, C-4), 62.2 (CH₂, C-6), 48.5 (CH, C-2), 33.6 (CH₂), 30.8 (CH₂), 23.3 (CH₃, NHAc), 20.7 (CH₃), 20.7 (CH₃), 20.7 (CH₃).

ES-HRMS calcd for C₁₆H₂₄NO₈Na₁SBr 492.0304, found m/z 492.0321 [M+Na]⁺

2-Azidoethyl 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-1-thio- α -D-galactopyranoside (AG123) ¹⁰

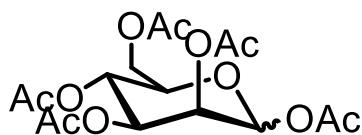


To a solution of AG122 (2.20 g, 4.68 mmol) in DMF (35 mL), tetrabutylammonium iodide (1.75 g, 4.74 mmol) and sodium azide (1.23 g, 18.9 mmol) were added. The reaction was stirred at 80 °C and for 12 h. The reaction was then allowed to cool, after which it was diluted with CH₂Cl₂ (50 mL) and washed with a saturated solution of NH₄Cl (2 x 40 mL) and with water (40 mL). The organic layer was dried over Na₂SO₄ and then concentrated under diminished pressure. Chromatography (EtOAc-cyclohexane, 1:3-1:1) gave the title compound (1.76 g, 87%) as a colourless solid.

¹H NMR (500 MHz, chloroform-d) δ = 5.64 (d, J = 8.5 Hz, 1H, NH), 5.60 (d, J = 5.3 Hz, 1H, H-1), 5.39 (dd, J = 3.5, 1.2 Hz, 1H, H-4), 5.05 (dd, J = 11.7, 3.2 Hz, 1H, H-3), 4.76 (ddd, J = 11.7, 8.5, 5.3 Hz, 1H, H-2), 4.57 – 4.52 (appntm, 1H, H-5), 4.14 (dd, J = 11.4, 5.8 Hz, 1H, H- 6a), 4.08 (dd, J = 11.4, 7.0 Hz, 1H, H-6b), 3.53 (dt, J = 13.4, 6.8 Hz, 1H, CH₂), 3.50 – 3.44 (m, 1H, CH₂), 2.87 (dt, J = 13.7, 6.8 Hz, 1H, CH₂), 2.77 (dt, J = 13.6, 6.6 Hz, 1H, CH₂), 2.15 (s, 3H, AcOMe), 2.05 (s, 3H, AcOMe), 2.00 (s, 3H, AcOMe), 1.97 (s, 3H, NHAc).

¹³C NMR (125 MHz, chloroform-d) δ = 171.0 (C), 170.3 (C), 170.2 (C) , 170.2 (C), 85.3 (CH, C-1), 68.2 (CH, C-3), 67.6 (CH, C-5), 67.5 (CH, C-4), 61.9 (CH₂, C-6), 50.9 (CH₂), 48.4 (CH, C-2), 30.6 (CH₂), 23.27 (CH₃, NHAc), 20.7 (CH₃), 20.6 (CH₃), 20.6 (CH₃).

ES-HRMS calcd for C₁₆H₂₄N₄O₈NaS 455.1213, found m/z 455.1217 [M+Na]⁺

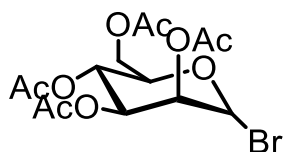
Mannopyranose pentaacetate (AG129)¹¹

To a 0°C cooled solution of D-mannose (2.0 g, 11.1 mmol) in pyridine (30 mL) acetic anhydride (7.3 mL, 77.7 mmol) was added, slowly (!!!) portion-wise. The reaction mixture was allowed to warm-up at room temperature and stirred for 12 h. The reaction mixture was poured onto an ice-water mixture and stirred for 15 mins. CH₂Cl₂ (50 mL) was added and stirred for 5 min. The organic phase was then washed with 1 M HCl (3 x 50 mL), a saturated solution of NaHCO₃ (50 mL), water (50 mL) and the organic phase was dried over Na₂SO₄. The solvent was removed under reduced pressure and purified through column chromatography (1:2 cyclohexane-EtOAc) to give the title compound (4.09 g, 95%), as mixture of anomers (7:1 α:β).

¹H NMR (500 MHz, chloroform-d) δ = 6.09 (d, *J* = 1.9 Hz, 1H, H-1α), 5.86 (d, *J* = 1.2 Hz, 1H, H-1β), 5.49 (d, *J* = 3.3 Hz, 1H, H-2β), 5.38 – 5.34 (overlapping signals, 2H, H-3α and H-4α), 5.32 – 5.29 (m, 1H, H-4β), 5.27 (d, *J* = 2.2 Hz, 1H, H-2α), 5.14 (dd, *J* = 10.0, 3.3 Hz, 1H, H-4β), 4.32 (d, *J* = 5.4 Hz, 1H, H-6aβ), 4.29 (dd, *J* = 12.4, 4.9 Hz, 1H, H-6aα), 4.16 (d, *J* = 2.3 Hz, 1H, H-6bβ), 4.11 (dd, *J* = 12.4, 2.4 Hz, 1H, H-6bα), 4.07 – 4.03 (m, 1H, H-5α), 3.81 (ddd, *J* = 9.9, 5.5, 2.3 Hz, 1H, H-5β), 2.22 (s, 3H, AcOMe), 2.18 (s, 6H, 2 x AcOMe), 2.17 (s, 3H, AcOMe), 2.11 (s, 3H, AcOMe), 2.10 (s, 3H, AcOMe), 2.06 (s, 6H, 2 x AcOMe), 2.01 (s, 6H, 2 x AcOMe).

¹³C NMR (125 MHz, chloroform-d) δ = 170.6 (C), 170.2 (C), 169.9 (C), 169.7 (C), 169.7 (C), 169.5 (C), 169.5 (C), 168.3 (C), 168.0 (C), 90.6 (CH, C-1α), 90.4 (CH, C-1β), 73.3 (CH, C-5β), 70.6 (CH, C-3β), 70.6 (CH, C-5α), 68.7 (CH, C-3α), 68.3 (CH, C-2α), 68.2 (CH, C-2β), 65.5 (CH, C-4α), 65.4 (CH, C-4β), 62.1 (CH₂, C-6α), 62.0 (CH₂, C-6β), 20.8 (CH₃), 20.7 (CH₃), 20.7 (CH₃), 20.7 (CH₃), 20.7 (CH₃), 20.6 (CH₃), 20.6 (CH₃), 20.5 (CH₃).

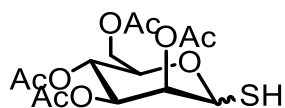
ES-HRMS calcd for C₁₆H₂₂O₁₁Na 413.1060, found *m/z* 413.1059 [M+Na]⁺

2,3,4,6-Tetra-*O*-acetyl- α -mannopyranosyl bromide (AG129i)¹²

To a 0°C cooled solution of AG129 (4.04 g, 10.35 mmol) in dry anhydrous CH₂Cl₂ (20 mL), HBr (33% in AcOH, 20 mL) was added drop-wise. The reaction mixture was allowed to warm to room temperature and stirred for 12 h. The reaction mixture was then poured onto an ice-water mixture and stirred for 15 mins. It was diluted with CH₂Cl₂, then separated and washed with satd. NaHCO₃, brine, dried over Na₂SO₄ and concentrated under reduced pressure to yield a colourless viscous oil (3.83 g, 90%) as the title compound. (*Note*: proceed with the following step immediately, it is not possible to store the bromide)

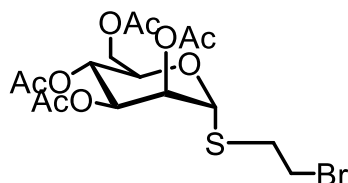
¹H NMR (500 MHz, chloroform-d) δ = 6.30 (d, J = 1.6 Hz, 1H, H-1), 5.72 (dd, J = 10.2, 3.5 Hz, 1H, H-3), 5.45 (dd, J = 3.5, 1.6 Hz, 1H, H-2), 5.37 (appnt, J = 12.3, 10.2 Hz, 1H, H-4), 4.33 (dd, J = 12.5, 5.0 Hz, 1H, H-6a), 4.23 (ddd, J = 10.3, 5.0, 2.1 Hz, 1H, H-5), 4.14 (dd, J = 12.6, 2.2 Hz, 1H, H-6b), 2.18 (s, 3H, AcOMe), 2.11 (s, 3H, AcOMe), 2.08 (s, 3H, AcOMe), 2.01 (s, 3H, AcOMe).

¹³C NMR (125 MHz, chloroform-d) δ = 170.5 (C), 169.6 (C), 169.5 (C), 83.0 (CH, C-1), 72.8 (CH, C-5), 72.1 (CH, C-2), 67.9 (CH, C-3), 65.3 (CH, C-4), 61.4 (CH₂, C-6), 20.7 (CH₃), 20.7 (CH₃), 20.6 (CH₃), 20.6 (CH₃).

2,3,4,6-Tetra-*O*-acetyl-1-thio- α/β -D-mannopyranose (AG130)¹³

To a solution of AG129 (3.82 g, 9.29 mmol) in acetone (48 mL), thiourea (1.2 g, 15.8 mmol) was added. The reaction was heated at 70 °C and stirred for 12h. The solution was allowed to cool to room temperature and the solvent was removed under diminished pressure. The thiouronium salt was used without any further purification. The salt (4.5 g, 9.29 mmol) was suspended in a 3:2 CH₂Cl₂-H₂O mixture (75 mL) and sodium metabisulfite (2.3 g, 12.08 mmol) was added. The reaction mixture was stirred at 70 °C for 2 h. The solution was allowed to cool to room temperature and separated. The aqueous layer was re-extracted with CH₂Cl₂ and the combined organic layer were washed with water, dried over Na₂SO₄, with the solvent then being removed under reduced pressure. The crude α - β anomers were used in the following step without further purification.

2-Bromoethyl 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-1-thio- α - D-mannopyranoside (AG131)



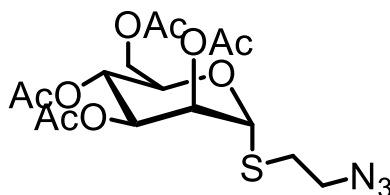
To a solution of AG130 (2.12 g, 5.82 mmol) in acetone–water (2:1, 24 mL) potassium carbonate (930 mg, 6.73 mmol) was added. The mixture was stirred under nitrogen for 15 min. 1,2-dibromoethane (6.3 mL, 72.7 mmol) was then added and the reaction was stirred at room temperature for 3 h. The mixture was extracted twice with CH₂Cl₂ (50 mL). The combined organic layer was dried with anhydrous Na₂SO₄ and the solvent removed under diminished pressure.

Column chromatography (Cyclohexane- EtOAc 3: 1) yielded the desired compound as yellow oil. (*Note*: by TLC is not possible to clearly see two spots, the first anomer to be eluted is the α , do not combine all the test tubes together).

¹H NMR (500 MHz, Chloroform-*d*) δ = 5.33 (dd, J = 3.3, 1.2 Hz, 1H, H-2), 5.31 (d, J = 1.2 Hz, 1H, H-1), 5.27 (appnt, J = 10.0, 9.3 Hz, 1H, H-4), 5.21 (dd, J = 10.0, 3.3 Hz, 1H, H-3), 4.38 (ddd, J = 9.3, 6.4, 2.2 Hz, 1H, H-5), 4.28 (dd, J = 12.2, 6.4 Hz, 1H, C-6a), 4.11 (dd, J = 12.2, 2.2 Hz, 1H, C-6b), 3.60 (td, J = 10.1, 5.7 Hz, 1H), 3.49 (td, J = 10.2, 6.1 Hz, 1H), 3.19 – 2.98 (overlapped signals, 2H), 2.16 (s, 3H), 2.11 (s, 3H), 2.05 (s, 3H), 1.99 (s, 3H).

¹³C NMR (125 MHz, Chloroform-*d*) δ = 170.5 (C), 169.9 (C), 169.7 (C), 169.7 (C), 83.1 8 (CH, C-1, decoupled d, J = 167.2 Hz), 70.7 (CH.C-2), 69.3 (CH, C-5), 69.2 (CH, C-3), 66.3 (CH, C-4), 62.5 (CH₂, C-6), 33.9(CH₂), 29.8 (CH₂), 20.8 (CH₃), 20.7 (CH₃), 20.7 (CH₃), 20.6 (CH₃).

**2-Azidoethyl 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-1-thio- α - D-mannopyranoside
AG132**



To a mixture of AG131 (470 mg, 1.00 mmol) and ammonium chloride (55 mg, 1.0 mmol) in dry DMF, sodium azide (200 mg, 3.0 mmol) was added. The resulting mixture was stirred at 80 °C for 7 h. The reaction mixture was poured into water (20 mL) and extracted twice with ethyl acetate (50 mL). The combined organic layers were washed with water (20 mL). The combined organic layers were dried over Na₂SO₄ and the solvent removed under diminished pressure.

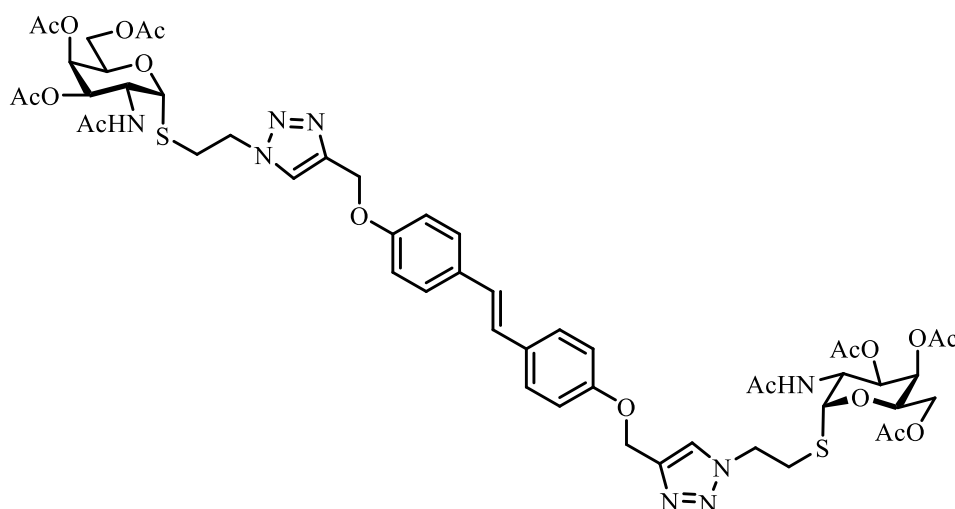
Column chromatography (Cyclohexane-EtOAc, 3: 1) yielded the desired compound as yellow liquid (424.76 mg, 98%).

¹H NMR (500 MHz, Chloroform-*d*) δ = 5.35 (dd, J = 3.3, 1.5 Hz, 1H, H-2), 5.32 (d, J = 1.5 Hz, 1H, H-1), 5.30 (appnt, J = 10.0, 9.3 Hz, 1H, H-4), 5.24 (dd, J = 10.0, 3.3 Hz, 1H, H-3), 4.37 (ddd, J = 9.3, 5.7, 2.3 Hz, 1H, H-5), 4.31 (dd, J = 12.2, 5.7 Hz, 1H, H-6a), 4.12 (dd, J = 12.2, 2.3 Hz, 1H, H-6b), 3.56 - 3.51 (overlapped signals, 2H), 2.88 (dt, J = 14.1, 7.1 Hz, 1H), 2.79 (dt, J = 13.7, 6.7 Hz, 1H), 2.17 (s, 3H, AcOMe), 2.10 (s, 3H, AcOMe), 2.06 (s, 3H, AcOMe), 2.00 (s, 3H, AcOMe).

¹³C NMR (125 MHz, Chloroform-*d*) δ 170.4 (C), 169.8 (C), 169.7 (C), 169.6 (C), 82.9 (CH, C-1), 70.8 (CH), 69.2 (CH), 69.2 (CH), 66.2 (CH), 62.4 (CH₂), 50.8 (CH₂), 30.7 (CH₂), 20.8 (CH₃), 20.6 (CH₃), 20.6 (CH₃), 20.5 (CH₃).

ES-HRMS calcd for C₁₆H₂₄N₃O₉S 434.1233, found m/z 434.1237 [M+H]⁺

(E)-1,2-Bis[4-(1-(ethyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-1-thio- α -D-glucopyranosyl)-1,2,3-triazol-4-ylmethoxy)-phenyl]ethane (AG090)



To a solution of AG089 (50 mg, 0.19 mmol) in degassed THF-H₂O (5mL, 1:1) were added AG123 (185,6 mg, 0.43 mmol), sodium ascorbate (23.2 mg, 0.12 mmol) and Cu₂SO₄ · 5H₂O (299.2 mg, 0.12 mmol). The reaction mixture was stirred under nitrogen for 18h at room temperature. Tetrahydrofuran was removed under diminished pressure and the solution was diluted with CH₂Cl₂. This mixture was then washed with water. The aqueous layer was re-extracted with a further portion of CH₂Cl₂. The combined organic layers were then washed with water and dried over Na₂SO₄ and the solvent removed under diminished pressure. Chromatography of the crude material (49:1 → 19:1 CH₂Cl₂-MeOH) gave the desired compound (161.9 mg, 72 %) as a colourless solid.

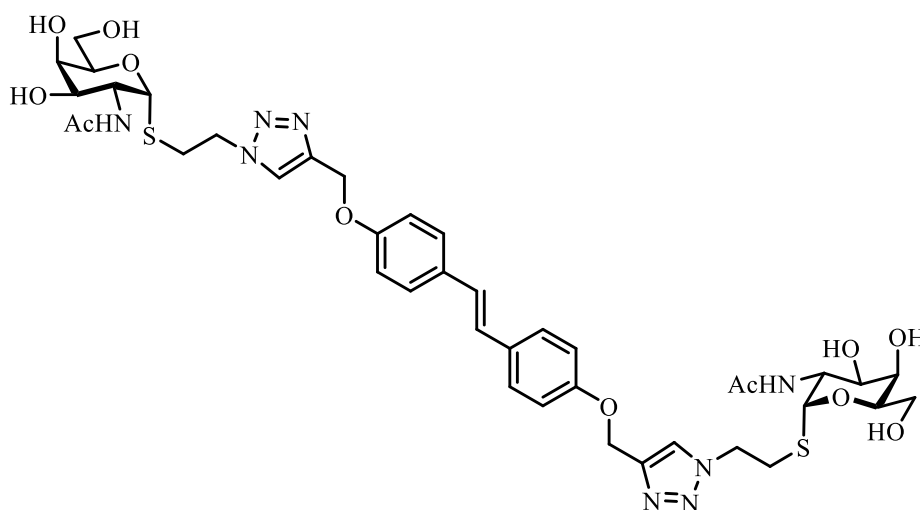
¹H NMR (500 MHz, Chloroform-*d*) δ 7.69 (s, 2H, triazole), 7.42 (d, *J* = 8.7 Hz, 4H, aromatic H), 6.95 (d, *J* = 8.7 Hz, 4H, aromatic H), 6.92 (s, 2H, allylic), 5.68 (d, *J* = 8.1 Hz, 2H, NH), 5.63 (d, *J* = 5.3 Hz, 2H, H-1), 5.35 (d, *J* = 3.0 Hz, 2H, H-4), 5.25 (s, 4H, CH₂), 4.97 (dd, *J* = 11.9, 3.0 Hz, 2H, H-3), 4.72 (ddd, *J* = 11.9, 8.1, 5.3 Hz, 2H, H-2), 4.62 (dt, *J* = 13.5, 6.7 Hz, 2H, CH₂), 4.55 (dt, *J* = 13.9, 6.8 Hz, 2H, CH₂), 4.41 (t, *J* = 6.4 Hz, 2H, H-5), 4.08-4.17 (overlapped signals, 4H, H-6a, H-6b), 3.19 (dt, *J* = 13.7, 6.8 Hz, 2H, CH₂), 3.08 (dt, *J* = 13.8, 6.7 Hz, 2H, CH₂), 2.16 (s, 6H, NHCH₃), 2.01 (s, 6H, AcOCH₃), 1.98 (s, 6H, AcOCH₃), 1.95 (s, 6H, AcOCH₃).

Chapter 6

^{13}C NMR (125 MHz, Chloroform-*d*) δ = 171.1 (C, C=O), 170.4 (C, C=O), 170.3 (C, C=O), 170.1 (C, C=O), 157.5 (C, triazole), 144.3 (C, aromatic), 130.9 (C, aromatic), 127.5 (CH, aromatic), 126.3 (CH, Allylic), 123.2 (CH, triazole), 115.0 (CH, aromatic), 85.3 (CH, C-1), 68.1(CH, C-3) , 67.7(CH, C-5), 67.1 (CH, C-4), 62.1 (CH₂), 62.1 (CH₂, C-6), 49.7(CH₂), 48.7 (CH, C-2), 31.5 (CH₂), 23.3 (CH₃,NHMe), 20.6 (CH₃, 3 x AcOMe).

ES-HRMS calcd for C₅₃H₆₆N₈O₁₈S₂Na 1189.3834, found m/z 1189.3839 [M+Na]⁺

(E)-1,2-Bis[4-(1-(ethyl 2-acetamido-2-deoxy-1-thio- α -D-glucopyranosyl)-1,2,3-triazol-4-ylmethoxy)-phenyl]ethane (AG104)



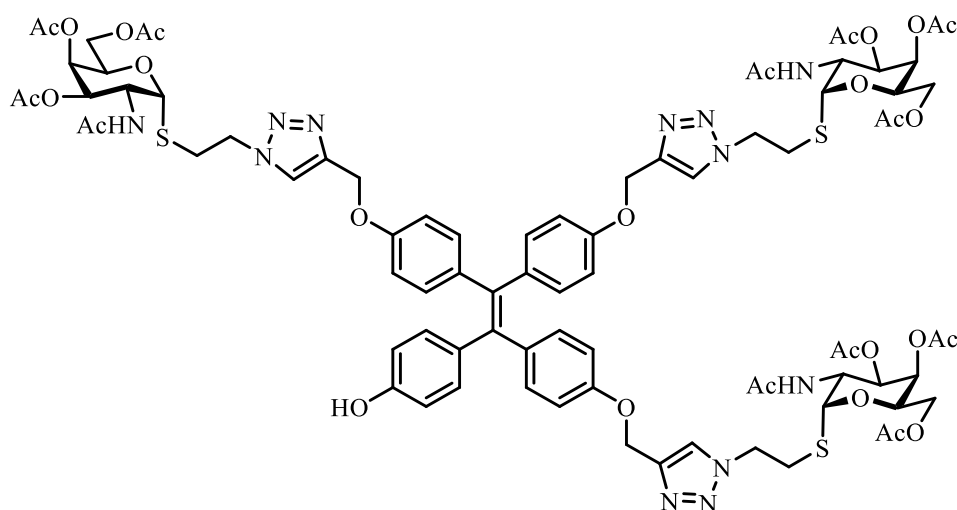
To a suspension of AG90 (135 mg, 0,117 mmol) in MeOH (5mL) Amberseph® 900 OH form (500 mg) was added. The reaction mixture was gently stirred at rt for 3h. The white precipitate was dissolved in warm water, the solution was filtered, and the solvent removed under diminished pressure to give the title compound (87.4 mg, 83%) as white solid.

^1H NMR (500 MHz, DMSO- d_6) δ = 8.21 (s, 2H, triazole), 7.77 (d, J = 7.0 Hz, 2H, NH), 7.49 (d, J = 8.6 Hz, 4H, aromatic H), 7.04 (s, 2H, allylic), 7.02 (d, J = 8.6 Hz, 4H, aromatic), 5.51 (d, J = 5.2 Hz, 2H, H-1), 5.14 (s, 4H, CH₂), 4.72-4.67 (m, 4H, CH₂), 4.63 (d, J = 6.1 Hz, 2H, OH-3), 4.61 – 4.55 (m, 2H, CH₂), 4.56 - 4.50 (m, 2H, CH₂), 4.17 (ddd, J = 11.8, 7.0, 5.2 Hz, 2H, H-2), 3.91 (t, J = 5.8 Hz, 2H, H-5), 3.72 (s, 2H, H-4), 3.54 (t, J = 5.3 Hz, 4H, H-6, H-6'), 3.52 – 3.45 (m, 2H, H-3), 3.03 (dt, J = 13.9, 6.8 Hz, 2H, CH₂), 2.93 (dt, J = 13.9, 6.8 Hz, 2H, CH₂), 1.80 (s, 6H, NHMe).

^{13}C NMR (125 MHz, DMSO- d_6) δ = 170.2 (C), 157.9 (C), 142.9 (C, triazole), 130.8 (C), 127.9 (CH, aromatic H), 126.4 (CH, allylic), 125.2 (CH, triazole), 115.4 (CH, aromatic), 84.8 (CH, C-1), 72.7 (CH, C-5), 68.5 (CH, C-4), 67.9 (CH, C-3), 61.6 (CH₂), 61.4 (CH₂, C-6), 50.5 (CH, C-2), 49.8 (CH₂), 30.3 (CH₂), 23.1 (CH₃, NHMe).

ES-HRMS calcd for C₄₀H₅₁N₈O₁₂S₂ 899.3068, found m/z 899.3062 [M-H]⁻

4-1,2,2-tris [4-(1-(ethyl 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-1-thio- α -D-galactopyranosyl)-1,2,3-triazol-3-ylmethoxy-phenyl)] phenol (AG121)⁷



To a solution of AG120 (124 mg, 0.22 mmol) in degassed THF-H₂O (1:1, 17.2 mL) were added AG123 (300 mg, 0.69 mmol), sodium ascorbate (39 mg, 0.20 mmol) and Cu₂SO₄ · 5H₂O (mg, 0.20 mmol). The reaction mixture was stirred under nitrogen atmosphere in MW at 50°C, 120 W for 30 min. Tetrahydrofuran was then removed under diminished pressure followed by the dilution of the solution with CH₂Cl₂. This mixture was then washed with water. The aqueous layer was re-extracted with a further portion of CH₂Cl₂. The combined organic layers were dried over Na₂SO₄ and the solvent was removed under diminished pressure. Chromatography of the crude material (28:1 → 21:1 CH₂Cl₂-MeOH) gave the title compound (410.9 mg, 70 %) as a colourless solid.

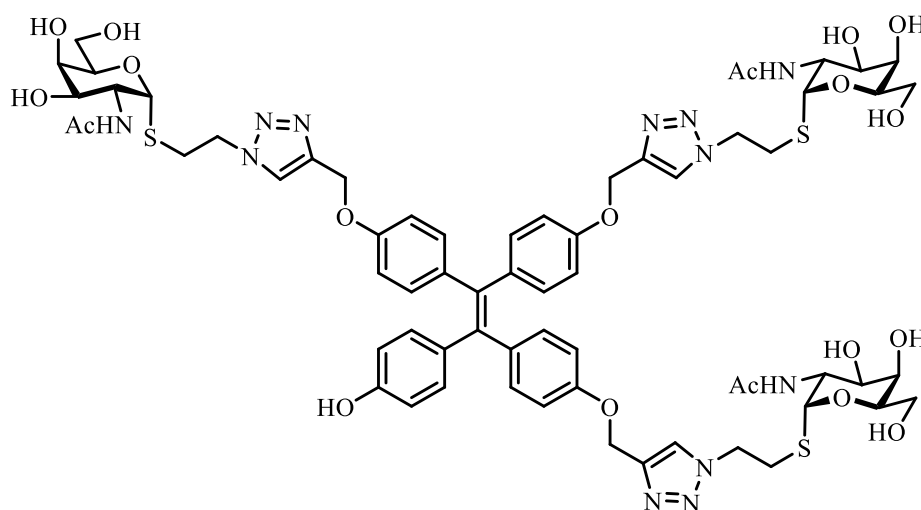
¹H NMR (500 MHz, Chloroform-*d*) δ 7.67 (s, 1H, triazole), 7.67 (s, 1H, triazole), 7.66 (s, 1H, triazole), 7.00 (d, *J* = 8.7 Hz, 2H, aromatic H), 6.93-6.91 (overlapped signals, 6H, aromatic H), 6.83 (d, *J* = 8.7 Hz, 2H, aromatic), 6.73 – 6.68 (overlapped signals, 6H, aromatic H), 5.97 (d, *J* = 7.8 Hz, 1H, NH), 5.92 (d, *J* = 8.1 Hz, 1H, NH), 5.71 (d, *J* = 8.3 Hz, 1H, NH), 5.68 – 5.63 (overlapped signals, 3H, H-1), 5.39 (dd, *J* = 3.2, 1.2 Hz, 3H, H-4), 5.14 (s, 2H, CH₂), 5.14 (s, 2H, CH₂), 5.13 (s, 2H, CH₂), 5.01 (dt, *J* = 11.8, 3.3 Hz, 3H, H-3), 4.78 – 4.69 (overlapped signals, 3H, H-2), 4.65 – 4.55 (overlapped signals, 6H, CH₂), 4.52 (t, *J* = 6.8 Hz, 3H, H-5), 4.20 – 4.06 (overlapped signals, 6H, CH₂), 3.25 – 3.12 (overlapped signals, 3H, CH₂), 3.12 – 3.00 (overlapped signals, 3H, CH₂), 2.25 (s, 3H, AcOMe), 2.16 (s, 3H, NHCH₃), (s, 6H NHCH₃), 2.01 (s, 3H, AcOMe), 2.00 (s, 6H,

AcOMe), 1.99 (s, 6H, AcOMe), 1.99 (s, 3H, AcOMe), 1.97 (s, 3H, AcOMe), 1.97 (s, 6H, AcOMe).

^{13}C NMR (126 MHz, cdCl_3) δ = 171.0 (C), 170.5 (C), 170.4 (C), 170.1 (C), 156.6 (C, aromatic), 144.3 (C, triazole), 138.2 (C), 137.0 (C, aromatic), 132.6 (CH, aromatic), 123.1 (CH, triazole), 114.0 (CH, aromatic), 85.1 (CH, C-1), 68.2 (CH, C-3), 67.6 (CH, C-5), 67.1 (CH, C-4), 62.0 (CH, C-2), 61.9 (CH_2), 49.6 (CH_2), 48.5 (CH_2 , C-6), 31.1 (CH_2), 23.3 (NHAc), 20.7 (CH_3 , AcOMe), 20.7 (CH_3 , AcOMe), 20.6 (CH_3 , AcOMe).

HRMS calcd for $\text{C}_{83}\text{H}_{98}\text{N}_{12}\text{NaO}_{28}\text{S}_3$ 1829.5673, found m/z 1829.5667 $[\text{M}+\text{Na}]^{+\vee}$

4-1,2,2-tris [4-(1-(ethyl 2-acetamido-2-deoxy-1-thio- α -D-galactopyranosyl)-1,2,3-triazol-3-ylmethoxy-phenyl)] phenol (AG124)⁷



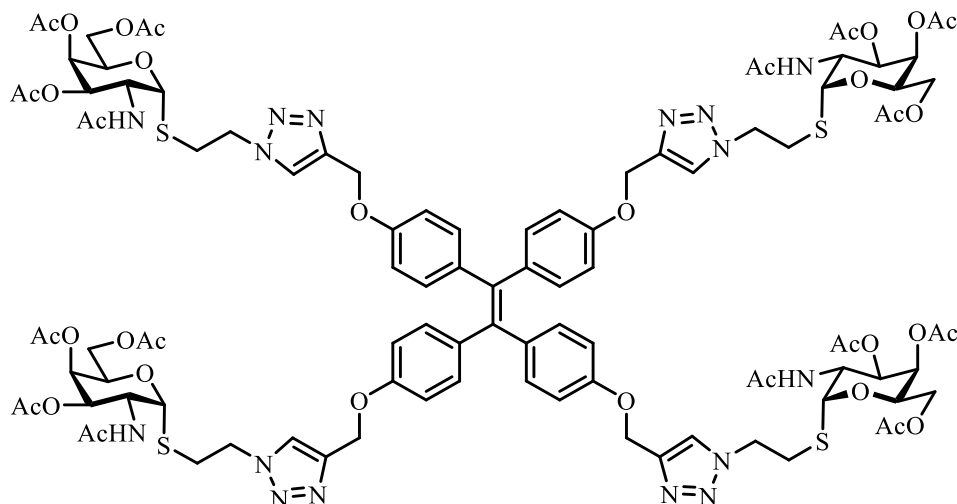
To a 0°C cooled suspension of AG121 (145 mg, 0.078 mmol) in dry methanol (7 mL), a freshly prepared NaOMe solution in MeOH was added until the solution reaches pH=9. The resulting mixture was stirred for 2 h at room temperature, then previously washed Amberlite IR- 120 H⁺ was added and stirred gently for 10 mins, to neutralize (pH = 7) the solution (acetic acid can be added instead). The solid was decanted and the resin was washed with a warm mixture of water-acetonitrile. The solvents were removed to a minimal amount of water. Reverse phase column chromatography was then carried out. The compound, in a minimal amount of water (made acidic and slightly heated, if solid crashes out), was loaded onto the column, with the elution carried out as follows: 3 volumes of CH₃CN-H₂O 2:8 were flushed through the column, followed by a 1:1 CH₃CN-H₂O mixture. This gave the title compound (83.5 mg, 75%) as a colourless solid.

¹H NMR (500 MHz, DMSO-*d*₆) δ = 9.33 (s, 1H, ArOH) 8.19 (s, 3H, triazole), 7.80 (d, *J* = 6.9 Hz, 3H, NH), 6.85 (overlapped signals, 6H, aromatic), 6.79 (overlapped signals, 6H, aromatic), 6.73 (d, *J* = 8.4 Hz, 2H, aromatic), 6.51 (d, *J* = 8.4 Hz, 2H, aromatic), 5.51 (d, *J* = 5.2 Hz, 3H, H-1), 5.02 (s, 6H, CH₂), 4.80 – 4.65 (overlapped signals, 12H, OH), 4.55 (ddt, *J* = 27.1, 13.9, 6.8 Hz, 6H, CH₂), 4.16 (ddd, *J* = 11.8, 6.9, 5.2 Hz, 3H, H-2), 3.91 (t, *J* = 6.0 Hz, 3H, H-5), 3.72 (brs, 3H, H-4), 3.53 (t, *J* = 5.9 Hz, 6H, H-6, H-6'), 3.49 (dd, *J* = 11.8, 2.9 Hz, 3H, H-3), 3.02 (dt, *J* = 13.8, 6.9 Hz, 3H, CH₂), 2.92 (dt, *J* = 13.7, 6.8 Hz, 3H), 1.79 (s, 9H, NHAc)

^{13}C NMR (126 MHz, DMSO- d_6) δ = 170.3(C, NHAc), 156.8 (C), 156.1 (C, aromatic), 142.9 (C, triazole), 142.8 (C), 139.0 (C), 137.1 (C), 132.5 (CH, aromatic), 132.5 (CH, aromatic), 125.2 (CH, triazole), 114.4 (CH, aromatic), 114.3 (CH, aromatic), 84.8 (CH,H-1), 72.7 (CH, H-5), 68.5 (CH, H-4), 67.9 (CH, H-3), 61.3 (CH₂), 61.2 (CH₂,C-6), 50.6 (CH,C-2), 49.8 (CH₂), 30.3(CH₂), 23.1 (CH₃,NHAc).

ES-HRMS calcd for C₆₅H₇₉ N₁₂O₁₉S₃ 1427.4747, found m/z 1427.4752 [M-H]⁻

1,1,2,2-Tetrakis[4-(1-(ethyl 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-1-thio- α -D-galactopyranosyl)-1,2,3-triazol-4-ylmethoxy-phenyl)]ethane(AG135)¹⁰



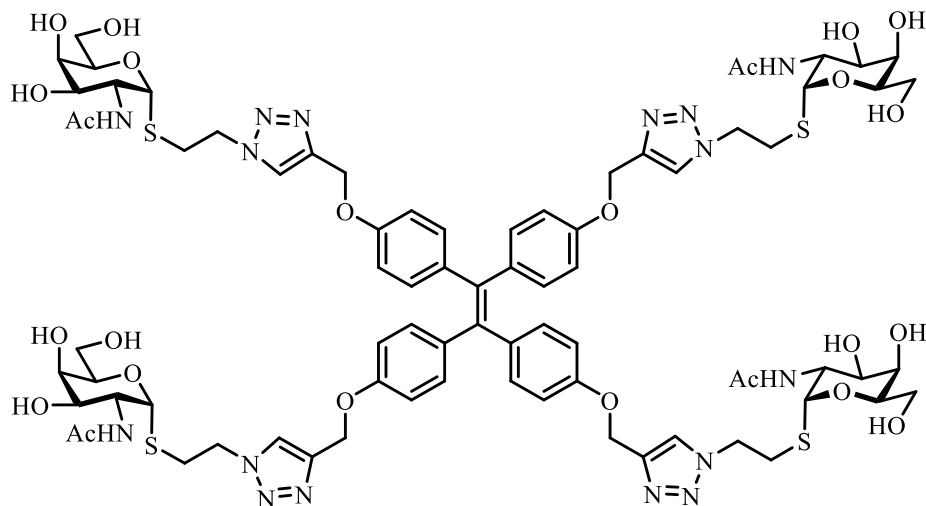
To a solution of and tetrapropargyl tetraphenylethene ether AG085 (46 mg, 0.084 mmol) in degassed THF-H₂O (1:1, 210 mL), sugar AG123 (149 mg, 0.345 mmol), sodium ascorbate (10 mg, 0.050 mmol) and Cu₂SO₄ · 5H₂O (12 mg, 0.048 mmol). The mixture was stirred under nitrogen atmosphere at 40 °C for 12 h. Tetrahydrofuran was then removed under diminished pressure followed by the dilution of the solution with CH₂Cl₂. This mixture was then washed with water. The aqueous layer was re-extracted with a further portion of CH₂Cl₂. The combined organic layers were dried over Na₂SO₄ and the solvent was removed under diminished pressure. Chromatography of the crude material (95:5-93:7 CH₂Cl₂- MeOH) gave the title compound (175 mg, 92%) as a colourless solid.

¹H NMR (500 MHz, Chloroform-*d*) δ = 7.66 (s, 4H, triazole), 6.92 (d, *J* = 8.5 Hz, 8H, aromatic H), 6.71 (d, *J* = 8.7 Hz, 8H, aromatic H), 5.91 (d, *J* = 8.2 Hz, 4H, NH), 5.67 (d, *J* = 5.4 Hz, 4H, H-1), 5.40 (d, *J* = 3.2, 4H, H-4), 5.15 (s, 8H, CH₂), 5.01 (dd, *J* = 12.6, 3.2 Hz, 4H, H-3), 4.74 (ddd, *J* = 12.6, 8.2, 5.4 Hz, 4H, H-2), 4.59 (hept, *J* = 6.9 Hz, 8H, CH₂), 4.52 (appnt, *J* = 7.0, 5.9 Hz, 4H, H-5), 4.15 (dd, *J* = 11.5, 5.9 Hz, 4H, H-6a), 4.11 (dd, *J* = 11.5, 7.0 Hz, 4H, H-6b), 3.18 (dt, *J* = 13.8, 6.8 Hz, 4H, CH₂), 3.08 (dt, *J* = 14.0, 6.9 Hz, 4H, CH₂), 2.17 (s, 12H, AcOMe), 2.01 (s, 12H, AcOMe), 2.00 (s, 12H, AcOMe), 1.97 (s, 12H, NHAc).

^{13}C NMR (126 MHz, Chloroform-*d*) δ = 171.0 (C), 170.5 (C), 170.4 (C), 170.6 (C), 156.5 (C, aromatic), 144.3 (C, triazole), 138.6 (C), 137.2 (C, aromatic), 132.5 (CH, aromatic), 123.1 (CH, triazole), 113.9 (CH, aromatic), 85.1 (CH, C-1), 68.2 (CH, C-3), 67.6 (CH, C-5), 67.1 (CH, C-4), 62.0 (CH₂, C-6), 61.9 (CH₂), 49.6 (CH₂), 48.5 (CH, C-2), 31.1 (CH₂), 23.2 (CH₃, NHAc), 20.7 (CH₃, AcoMe), 20.7 (CH₃, AcoMe), 20.7 (CH₃, AcoMe).

ES-HRMS calcd for C₅₁H₆₂N₈O₁₈S₂Na 1161.3521, found m/z 1161.35030 [M/2+Na]⁺

1,1,2,2-Tetrakis[4-(1-(ethyl 2-acetamido-2-deoxy-1-thio- α -D-galactopyranosyl)-1,2,3-triazol-4-ylmethoxy-phenyl)]ethene (AG143)¹⁰



To a solution of AG135 (220 mg, 0.097 mmol) in dry methanol (20 mL), NaOMe solution in MeOH was added until the solution reaches pH=9.0. The resulting mixture was stirred for 2 h at room temperature. Amberlite IR- 120 H⁺ was added and stirred gently for 10 mins, to neutralize (pH = 7) the solution (acetic acid can be added instead). The resin was removed by filtration and washed with water and acetonitrile. The solvents were removed to a minimal amount of water. Reverse phase column chromatography was then carried out. The compound, in a minimal amount of water (made acidic and slightly heated, if solid crashes out, to return it to solution), was loaded onto the column, with the elution carried out as follows: 3 volumes of water were flushed through the column, to ensure salt removal, followed by a 1:1 MeCN-H₂O mixture. This gave the title compound (159 mg, 93%) as a yellow solid.

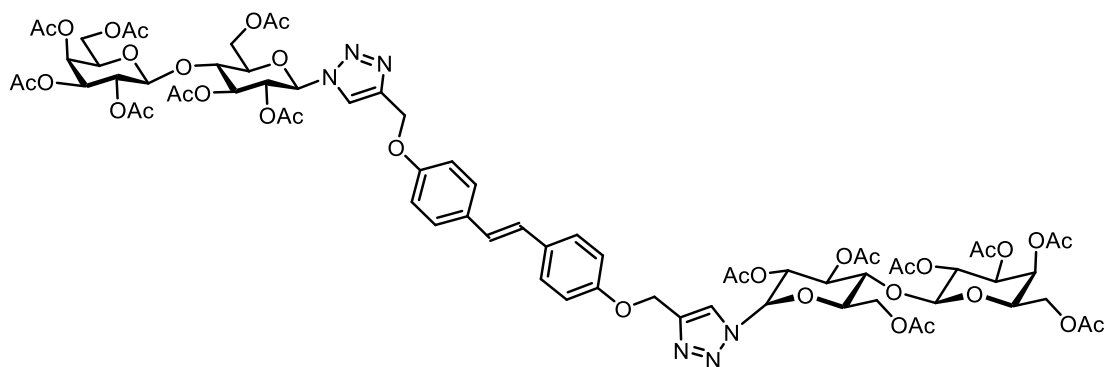
¹H NMR (500 MHz, DMSO-*d*₆) δ = δ 8.18 (s, 4H, triazole), 7.74 (d, *J* = 7.1 Hz, 4H, NH), 6.87 (d, *J* = 8.4 Hz, 8H, aromatic H), 6.80 (d, *J* = 8.8 Hz, 8H, aromatic H), 5.50 (d, *J* = 5.3 Hz, 4H, H-1), 5.03 (s, 8H, CH₂), 4.67 (br s, OH), 4.58 (dt, *J* = 13.8, 6.8 Hz, 4H, CH₂), 4.51 (dt, *J* = 13.9, 7.0 Hz, 4H, CH₂), 4.17 (ddd, *J* = 11.8, 7.0, 5.2 Hz, 4H, H-2), 3.91 (t, *J* = 6.1 Hz, 4H, H-5), 3.70 (br d, *J* = 3.0 Hz, 4H, H-4), 3.54 (d, *J* = 6.0 Hz, 8H, H-6a, H-6b), 3.48

(dd, $J = 11.3, 3.0$ Hz, 4H, H-3), 3.02 (dt, $J = 13.9, 7.0$ Hz, 4H, CH₂), 2.93 (dt, $J = 13.8, 6.9$ Hz, 4H, CH₂), 1.79 (s, 12H, NHAc).

¹³C NMR (126 MHz DMSO-*d*₆) $\delta = 170.2$ (C), 156.9 (C, aromatic), 142.8 (C, triazole), 138.5 (C), 136.9 (C, aromatic), 132.5 (CH, aromatic), 125.2 (CH, triazole), 114.3 (CH, aromatic), 84.8 (CH, C-1), 72.7 (CH, C-5), 68.6 (CH, C-4), 67.9 (CH, C-3), 61.5 (CH₂), 61.4 (CH₂, C-6), 50.5 (CH, C-2), 49.8 (CH₂), 30.3 (CH₂), 23.1 (CH₃, NHAc).

ES-HRMS calcd for C₇₈H₁₀₀N₁₆O₂₄S₄Na 1795.5877, found m/z 1795.5870 [M+Na]⁺

(E)-1,2-Bis[4-(1-(β -D-lactosepentaacetate)-1,2,3-triazol-4-ylmethoxy)-phenyl]ethane (AG125)⁵



To a solution of AG089 (107.7 mg, 0.42 mmol) in degassed THF-H₂O (1:1, 6mL) were added lactose azide (561 mg, 0.92 mmol), sodium ascorbate (49 mg, 0.25 mmol) and Cu₂SO₄·5H₂O (62 mg, 0.25 mmol). The reaction mixture was stirred under inert atmosphere in MW at 50°C, 120 W for 30 min. Tetrahydrofuran was then removed under diminished pressure followed by the dilution of the solution with CH₂Cl₂. This mixture was then washed with water. The aqueous layer was re-extracted with a further portion of CH₂Cl₂. The combined organic layers were dried over Na₂SO₄ and the solvent was removed under diminished pressure. Chromatography of the crude material (95:5 CH₂Cl₂-MeOH) gave the desired compound (600 mg, 89 %) as a colourless solid.

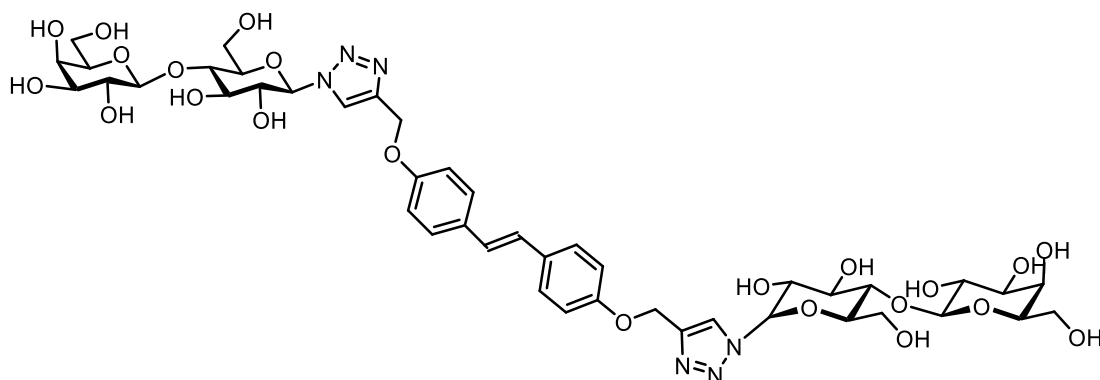
¹H NMR (500 MHz, chloroform-d) δ 7.78 (s, 2H, triazole H), 7.40 (d, J = 8.8 Hz, 4H, aromatic H), 6.94 (d, J = 8.8 Hz, 4H, aromatic H), 6.91 (s, 2H, alkene H), 5.83 (d, J = 9.1 Hz, 2H, H-1), 5.39 (overlapping signals, 4H, H-2 and), 5.35 (d, J = 3.4 Hz, 2H), 5.20 (s, 4H, CH₂), 5.12 (dd, J = 10.4, 7.9 Hz, 2H, H-2'), 4.96 (dd, 10.4, 3.4 Hz, 2H, H-3'), 4.52 (d, J = 7.9 Hz, 2H, H-1'), 4.47 (d, J = 11.0 Hz, 2H, H-6), 4.18 – 4.06 (overlapping signals, 6H, H-6 and H-6' protons), 3.98 – 3.83 (m, 6H, H-4, H-5, H-5'), 2.15 (s, 6H, AcOMe), 2.09 (s, 6H, AcOMe), 2.06 (s, 6H, OMe), 2.05 (s, 6H, AcOMe), 2.04 (s, 6H, AcOMe), 1.96 (s, 6H, AcOMe), 1.82 (s, 6H, AcOMe).

¹³C NMR (126 MHz, chloroform-d) δ = 170.3 (C), 170.2 (C), 170.1 (C), 170.0 (C), 169.4 (C), 169.1 (C), 169.0 (C), 157.5 (C, aromatic), 144.8 (C, triazole), 131.0 (C, aromatic),

127.5 (CH, aromatic), 126.3 (CH, alkene), 121.1 (CH, triazole), 115.0 (CH, aromatic), 101.1 (CH, C-1'), 85.5(CH, C-1), 75.9 (CH, C-4), 75.6 (CH,C-5), 72.6 (CH,C-3), 70.9 (CH,C-3'), 70.8 (CH,C-2), 70.5 (CH,C-2'), 69.0, 66.6 (CH,C-4'), 61.9 (CH₂), 61.7(CH₂,C-6), 60.8 (CH₂,C-6'), 20.8 (CH₃, AcOMe), 20.7(CH₃, AcOMe), 20.6 (CH₃, AcOMe), 20.6 (CH₃, AcOMe), 20.6 (CH₃, AcOMe), 20.5(CH₃, AcOMe), 20.1(CH₃, AcOMe).

ES-HRMS calcd for C₇₂H₈₆N₆NaO₃₆ 1633.4981, found m/z 1633.4985 [M+Na]⁺

(E)-1,2-Bis[4-(1-(β-D-lactose)-1,2,3-triazol-4-ylmethoxy)-phenyl]ethane
(AG127)⁵



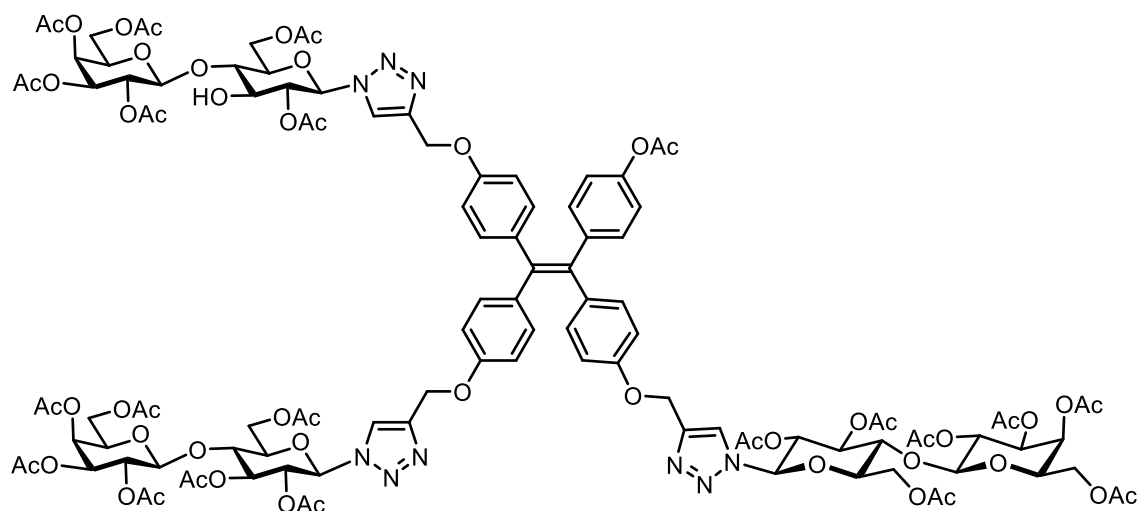
To a 0°C cooled suspension of AG125 (400 mg, 0,25 mmol) in dry MeOH (10 mL), freshly prepared 1M NaOMe in MeOH was added until the solution reached pH 10. The reaction mixture was allowed to warm up at room temperature and stirred overnight. Glacial acetic acid was added to neutralise (pH=7) the solution. [Note: it is possible use Amberlite IR-120 H+ instead of AcOH, but filtration is extremely hard due to the low solubility of AG127]. The solvent was removed under diminished pressure. Reverse phase column (3 volumes of water were flushed through the column, to ensure salt removal, followed by a 6:4 MeCN-H₂O) gave the title compound as white solid (250 mg, 96%). [Note: dissolve the compound, in a minimal amount of water with few drops of MeCN. If the solid crashes out, add one drop of AcOH and slightly heat up].

¹H NMR (500 MHz, DMSO-d₆) δ = 8.46 (s, 2H, triazole H), 7.52 (d, J = 8.8 Hz, 4H, aromatic), 7.11 – 7.00 (overlapping signals, 6H, aromatic H, alkene H), 5.67 (d, J = 9.3 Hz, 2H, H-1), 5.17 (s, 4H, CH₂), 4.26 (d, J = 7.9 Hz, 2H, H-1'), 3.87 (t, J = 9.3 Hz, 2H, H-2), 3.78 (d, J = 10.6 Hz, 2H, H-6), 3.70 – 3.44 (overlapping signals, 16H), 3.42 – 3.29 (overlapping signals, 4H).

¹³C NMR (126 MHz, DMSO-d₆) δ= 157.9 (C), 143.1 (C), 130.8 (C), 127.9 (CH, aromatic), 126.4 (CH, aromatic), 124.4 (CH, triazole), 115.3 (CH, aromatic), 104.2 (CH, C-1'), 87.4 (CH, C-1), 80.2 (CH), 78.2 (CH), 76.0 (CH), 75.6 (CH), 73.7 (CH), 72.2 (CH), 71.0 (CH), 68.6 (CH), 61.4 (CH₂), 60.9 (CH₂), 60.5 (CH₂).

ES-HRMS calcd for C₄₄H₅₇N₆O₂₂ 1021.3526, found m/z 1021.3566 [M-H]⁻

4-(1,2,2-tris[4-(1- β -D-Lactopyranosepentaacetate)-1,2,3-triazol-phenyl]vinyl]phenol (AG133)



To a solution of AG120 (50 mg, 0.094 mmol) in degassed THF-H₂O (6:4, 25mL) were added lactose azide (176 mg, 0.28 mmol), sodium ascorbate (16 mg, 0.08 mmol) and Cu₂SO₄ · 5H₂O (19 mg, 0.08 mmol). The reaction mixture was stirred under inert atmosphere at 50°C for 14h. Tetrahydrofuran was then removed under diminished pressure followed by the dilution of the solution with CH₂Cl₂. This mixture was then washed with water. The aqueous layer was re-extracted with a further portion of CH₂Cl₂. The combined organic layers were dried over Na₂SO₄ and the solvent was removed under diminished pressure. Chromatography of the crude material (98:2→93:3 CH₂Cl₂-MeOH) gave the title compound (186 mg, 78 %) as a colourless solid.

¹H NMR (500 MHz, chloroform-d) δ = 7.79 (s, 2H, triazole H), 7.78 (s, 1H, triazole H), 6.99 (d, J = 8.6 Hz, 2H, aromatic H), 6.95 – 6.90 (overlapped signals, 5H, aromatic H), 6.83 (d, J = 8.6 Hz, 2H, aromatic H), 6.74 – 6.69 (overlapped signals, 5H, aromatic H), 5.80 - 5.87 (overlapped signals, 3H, H-1), 5.39 – 5.41 (overlapped signals, 6 H, H-2 and H-3), 5.36 (d, J = 3.4 Hz, 3H, H-4'), 5.15 – 5.09 (overlapped signals, 8H, CH₂ and H-2'), 4.98 (dt, J = 10.5, 2.9 Hz, 3H, H-3'), 4.55 – 4.51 (overlapped signals, 3H, H-1'), 4.50 – 4.45 (overlapped signals, 3H, H-6), 4.18 – 4.06 (overlapped signals, 9H, H-6 and H-6'), 3.98 – 3.83 (overlapped signals, 9H, H-4, H-5, H-5'), 2.25 (s, 3H, AcOMe), 2.16 (s, 9H, AcOMe), 2.11 – 2.09 (overlapped signals, 9H, AcOMe), 2.08 – 2.04 (overlapped signals, 27H, AcOMe), 1.97 (s, 6H, AcOMe), 1.87 – 1.84 (overlapped signals, 9H, AcOMe).

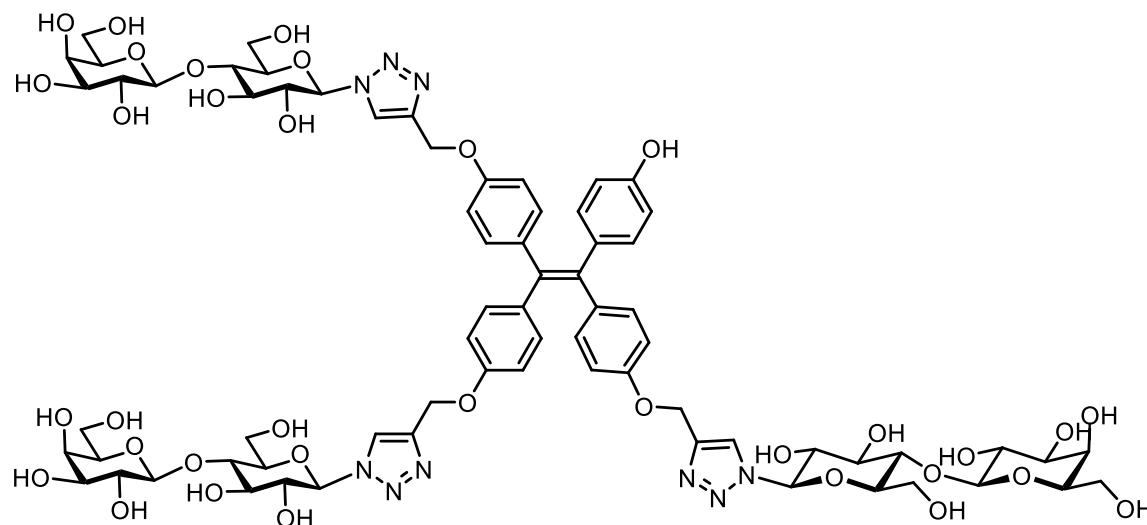
Chapter 6

^{13}C NMR (125 MHz, chloroform- d) δ =170.3 (C), 170.2 (C), 170.1 (C), 170.0 (C), 169.4 (C), 169.1 (C), 169.0 (C), 156.6 (C, aromatic) , 144.8 (C, triazole), 132.6 (CH, aromatic), 121.1 (CH, triazole), 113.9 (CH, aromatic), 110.0, 101.1 (CH, C-1'), 85.5 (CH, C-1), 75.9 (CH, C-4), 75.6 (CH,C-5), 72.6 (CH,C-3), 70.9 (CH,C-3'), 70.8 (CH,C-2), 70.5 (CH,C-2'), 69.0 (CH,C-2'), 66.6 (CH,C-4'), 61.8 (CH₂, C-6), 61.8 (CH₂) 60.8 (CH₂,C-6'), 20.8 (CH₃, AcOMe), 20.7 (CH₃, AcOMe), 20.6 (CH₃, AcOMe), 20.6 (CH₃, AcOMe), 20.6 (CH₃, AcOMe), 20.5 (CH₃, AcOMe), 20.2 (CH₃, AcOMe).

ES-HRMS calcd for C₁₁₅H₁₃₃N₉NaO₅₆ 2558.7734, found m/z 2558.7732 [M+Na]⁺

4-(1,2,2-tris[4-(1-β-D-Lactopyranosyl)-1,2,3-triazol-phenyl]vinyl)phenol

(AG133-OH/ AG146)



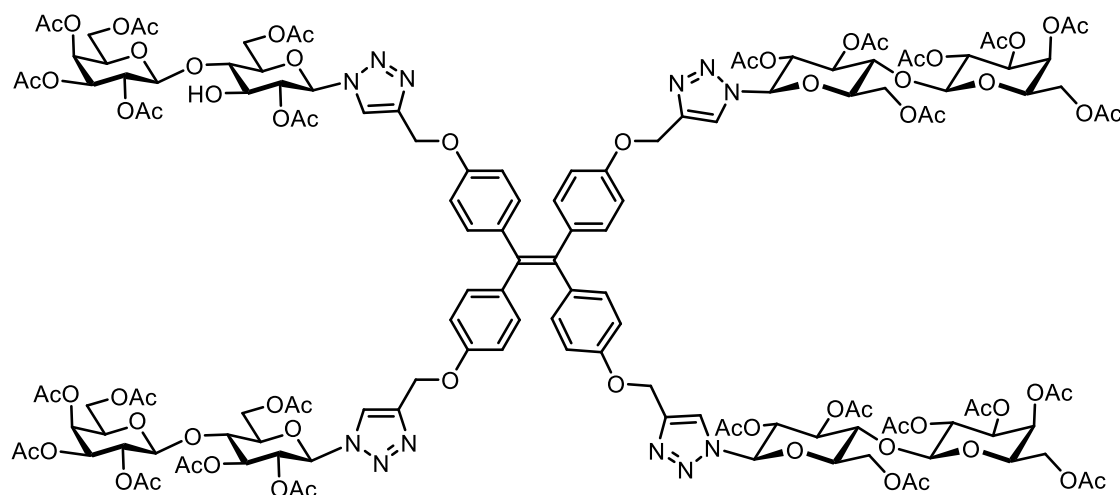
To a 0°C cooled suspension of AG133 (186 mg, 0,07 mmol) in dry MeOH (6 mL), freshly prepared 1M NaOMe in MeOH was added until the solution reached pH 10. The reaction mixture was allowed to warm up at rt and stirred overnight. Glacial acetic acid was added to neutralise (pH=7) the solution. [Note: it is possible to use Amberlite IR- 120 H⁺ instead of AcOH,]. The solvent was removed under diminished pressure. Reverse phase column (3 volumes of water were flushed through the column, to ensure salt removal, followed by a 6:4 MeCN-H₂O) gave the title compound as a white solid (106 mg, 94%). [NOTE: dissolve the compound, in a minimal amount of water with few drops of MeCN. If the solid crashes out, add one drop of AcOH and slightly heat up].

¹H NMR (500 MHz, DMSO-d₆) δ 8.43 – 8.37 (overlapped signals, 3H, triazole), 6.91 – 6.84 (overlapped signals, 6H, aromatic H), 6.79 - 6.82 (overlapped signals, 6H, aromatic H), 6.74 (d, J = 8.4 Hz, 2H, aromatic H), 6.51 (d, J = 8.5 Hz, 2H, aromatic H), 5.64 (d, J = 9.3 Hz, 3H, H-1), 5.57 (br s, 3H, OH), 5.10 (br s, 3H, OH), 5.04 (s, 6H, CH₂), 4.87 (br s, 3H, OH), 4.78 (br s, 1H, OH), 4.65 (br s, 3H, OH), 4.51 (br s, 2H, OH), 4.24 (d, J = 7.2 Hz, 3H, H-1'), 3.89 – 3.81 (overlapped signals, 3H, H-2), 3.76 (d, J = 11.6 Hz, 3H, H-6), 3.66 – 3.43 (overlapped signals, 27H), 3.37 – 3.31 (overlapped signals, 10H).

¹³C NMR (125 MHz, DMSO-d₆) δ =156.8 (C), 143.0 (C, triazole), 137.6 (C)132.5 (CH, aromatic), 124.4 (CH, triazole), 114.2 (CH, aromatic), 104.2 (CH, C-1'), 87.4 (CH, C-1), 80.2 (CH), 78.2 (CH), 76.0 (CH), 75.6 (CH), 73.7 (CH), 72.2 (CH, C-2), 71.0 (CH), 68.6 (CH), 60.9(CH₂), 60.5 (CH₂).

ES-HRMS calcd for C₇₁H₈₈N₉O₃₄ 1610.5434, found m/z 1610.5440 [M-H]⁻

1,1,2,2-Tetrakis[4-(1-(β -D-lactosepentaacetate)-1,2,3-triazol-4-ylmethoxyphenyl)]ethane (AG140)³



To a solution of AG088 (75 mg, 0.14 mmol) in degassed THF-H₂O (2:1, 15 mL) were added lactose azide (500 mg, 0.92 mmol), sodium ascorbate (25 mg, 0.13 mmol) and CuSO₄ 5H₂O (32.5 mg, 0.13 mmol). The reaction mixture was stirred under inert atmosphere at 50°C for 14h. Tetrahydrofuran was then removed under diminished pressure followed by the dilution of the solution with CH₂Cl₂. This mixture was then washed with water. The aqueous layer was re-extracted with a further portion of CH₂Cl₂. The combined organic layers were dried over Na₂SO₄ and the solvent was removed under diminished pressure. Chromatography of the crude material (49:1 → 42:1 CH₂Cl₂-MeOH) gave the title compound (344 mg, 72 %) as a colourless solid.

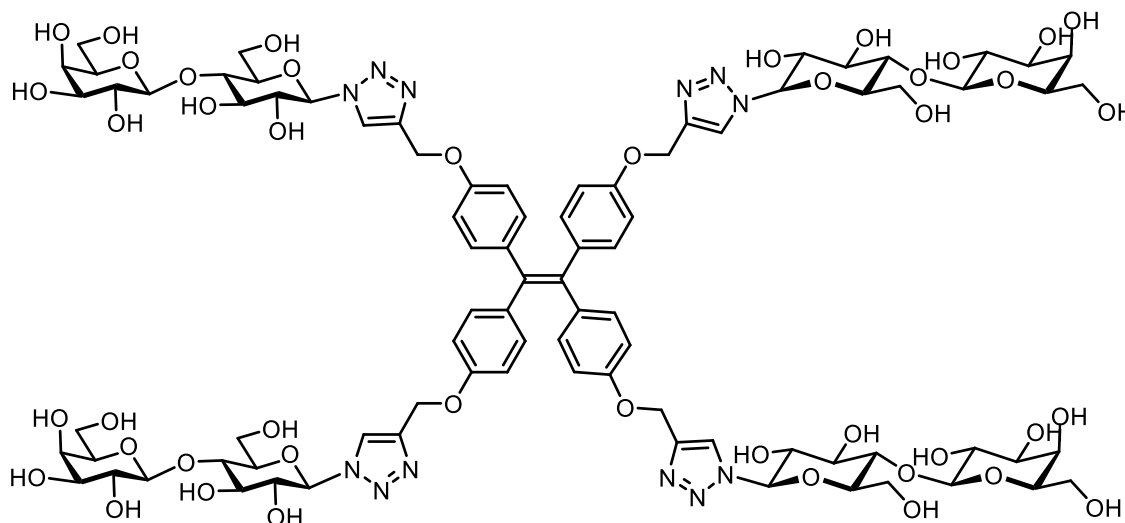
¹H NMR (500 MHz, chloroform-*d*) δ = 7.79 (s, 4H, triazole), 6.93 (d, *J* = 8.7 Hz, 8H, aromatic), 6.71 (d, *J* = 8.8 Hz, 8H, aromatic), 5.85 (d, *J* = 9.1 Hz, 4H, H-1), 5.39 - 5.42 (overlapping signals, 8H, H-2 and H-3), 5.37 (d, *J* = 3.4 Hz, 4H, H-4'), 5.16 - 5.10 (m, 4H, H-2'), 5.12 (s, 8H, CH₂), 4.98 (dd, *J* = 10.4, 3.5 Hz, 4H, H-3'), 4.54 (d, *J* = 7.9 Hz, 4H, H-1'), 4.48 (d, *J* = 10.9 Hz, 4H, H-6), 4.19 - 4.05 (overlapping signals, 12H, 6H, H-6 and H-6'), 3.90-3.98 (overlapping signals, 12H, H-4, H-5, H-5'), 2.16 (s, 12H, AcOMe), 2.10 (s, 12H, AcOMe), 2.07 (s, 12H, AcOMe), 2.06 (s, 12H, AcOMe), 1.97 (s, 12H, AcOMe), 1.85 (s, 12H, AcOMe).

¹³C NMR (126 MHz, chloroform-*d*) δ = 170.3 (C), 170.2 (C), 170.1 (C), 170.0 (C), 169.4 (C), 169.1 (C), 169.0 (C), 156.6 (C), 144.8 (C), 138.5 (C), 132.6 (CH, aromatic), 113.9 (CH, triazole), 110.0 (CH, aromatic), 101.1 (CH, C-1'), 85.5 (CH, C-1), 75.9 (CH, C-4),

75.6 (CH,C-5), 72.6 (CH,C-5), 70.9 (CH,C-3'), 70.5 (CH,C-2), 69.0 (CH,C-2') , 66.6 (CH,C-4'), 61.8 (CH₂, C-6), 61.8 (CH₂), 60.8 (CH₂, C-6'), 20.8 (CH₃, AcOMe), 20.7 (CH₃, AcOMe), 20.6 (CH₃, AcOMe), 20.6 (CH₃, AcOMe), 20.6 (CH₃, AcOMe), 20.5, 20.2 (CH₃, AcOMe).

ES-HRMS calcd for C₁₄₀H₁₆₇N₁₂O₇₁ 3151.9826, found m/z 3151.9820 [M+H]⁺

1,1,2,2-Tetrakis[4-(1-β-D-Lactopyranosyl)-1,2,3-triazol-4-ylmethoxy-phenyl]ethene (AG140-OH/AG14)³



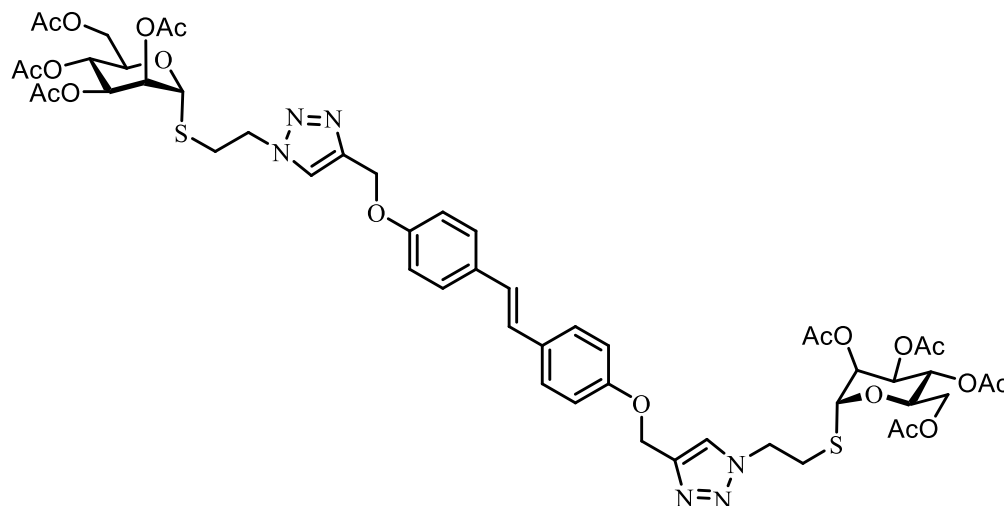
To a 0°C cooled suspension of AG140 (344 mg, 0.1 mmol) in dry MeOH (10 mL), freshly prepared 1M NaOMe in MeOH was added until the solution reached pH 10. The reaction mixture was allowed to warm up at room temperature and stirred overnight. Glacial acetic acid was added to neutralise (pH=7) the solution. [Note: it is possible to use Amberlite IR-120 H⁺ instead of AcOH,]. The solvent was removed under diminished pressure. Reverse phase column (3 volumes of water were flushed through the column, to ensure salt removal, followed by a 6:4 MeCN-H₂O) gave the title compound as a white solid (187.4 mg, 93 %). [Note: dissolve the compound, in a minimal amount of water with few drops of MeCN. If the solid crashes out, add one drop of AcOH and slightly heat up].

¹H NMR (500 MHz, DMSO-d₆) δ = 8.43 (s, 4H, triazole), 6.90 (d, J = 8.5 Hz, 8H, aromatic H), 6.84 (d, J = 8.6 Hz, 8H, aromatic H), 5.66 (d, J = 9.3 Hz, 4H, H-1), 5.65 – 5.61 (overlapping signals, 4H, OH), 5.20 (br s, 4H, OH), 5.07 (s, 8H, CH₂), 4.95 (br s, 4H, OH), 4.92 (br s, 2H, OH), 4.75 – 4.65 (overlapping signals, 8H, OH), 4.59 (s, 4H, OH), 4.26 (d, J = 7.1 Hz, 4H, H-1'), 3.87 (appntm, 4H, H-2), 3.78 (m, 4H), 3.70 – 3.45 (overlapping signals, 30H), 3.40 – 3.33 (overlapping signals, 8H).

¹³C NMR (125 MHz, DMSO-d₆) δ = 171.1 (C), 156.9 (C), 143.0 (C), 137.0 (C), 132.5 (CH, aromatic), 124.4 (CH, triazole), 114.3 (CH, aromatic), 104.2 (CH, C-1'), 87.4 (CH, C-1), 80.2 (CH), 78.2 (CH), 76.1 (CH), 75.6 (CH), 73.7 (CH), 72.2 (CH, C-2), 71.0 (CH), 68.6 (CH₂), 60.9 (CH₂), 60.5 (CH₂).

ES-HRMS calcd for C₈₆H₁₁₁N₁₂O₄₄ 2015.6817, found m/z 2015.6818 [M-H]⁻

(*E*)-1,2-Bis[4-(1-(α -D-mannosepentaacetate)-1,2,3-triazol-4-ylmethoxy)-phenyl]ethane (AG137)



To a solution of AG089 (75 mg, 0.26 mmol) in degassed THF-H₂O (1:1, 14mL) were added AG132 (279 mg, 0.64 mmol), sodium ascorbate (35 mg, 0.17 mmol) and CuSO₄ · 5H₂O (43 mg, 0.17 mmol). The reaction mixture was stirred under inert atmosphere at 50°C for 14h. Tetrahydrofuran was then removed under diminished pressure followed by the dilution of the solution with CH₂Cl₂. This mixture was then washed with water. The aqueous layer was re-extracted with a further portion of CH₂Cl₂. The combined organic layers were dried over Na₂SO₄ and the solvent was removed under diminished pressure. Chromatography of the crude material (49:1 → 32:1 CH₂Cl₂-MeOH) gave the title compound (209 mg, 70 %) as a colourless solid.

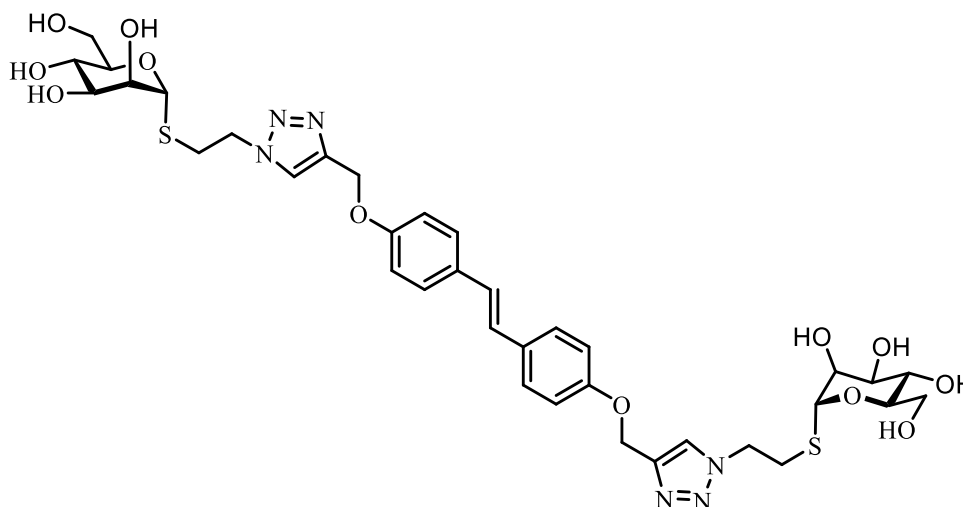
¹H NMR (500 MHz, chloroform-*d*) δ = 7.69 (s, 2H, triazole), 7.42 (d, *J* = 8.8 Hz, 4H, aromatic H), 6.97 (d, *J* = 8.7 Hz, 4H, aromatic H), 6.92 (s, 2H, alkene), 5.31 (dd, *J* = 3.4, 1.5 Hz, 2H, H-2), 5.29 (overlapped signals, 2H, H-4), 5.25 (d, *J* = 1.5 Hz, 2H, H-1), 5.24 (s, 4H, CH₂), 5.20 (dd, *J* = 10.0, 3.3 Hz, 2H, H-3), 4.67 – 4.54 (m, 4H, CH₂), 4.33 – 4.27 (overlapped signals, 4H, H-5, H-6a), 4.16 – 4.09 (m, 2H, H-6b), 3.16 (dq, *J* = 29.8, 7.1 Hz, 4H, CH₂), 2.17 (s, 6H, AcOMe), 2.05 (s, 6H, AcOMe), 2.05 (s, 6H, AcOMe), 1.99 (s, 6H, AcOMe).

Chapter 6

^{13}C NMR (125 MHz, chloroform-*d*) δ = 195.6 (C, triazole), 169.9 (C), 169.9 (C), 157.6 (C, aromatic), 154.2 (C), 144.4 (C), 131.0 (C, aromatic), 127.5 (CH, aromatic), 126.3 (CH, alkene), 123.0 (CH, triazole), 115.0 (CH, aromatic), 82.9 (CH, H-1), 70.7 (CH, C-2), 69.5 (CH, C-5), 69.1 (CH, C-4), 66.1, (CH, C-3), 62.4 (CH₂, C-6), 62.1 (CH₂), 49.7 (CH₂), 31.6 (CH₂), 20.9 (CH₃, AcOMe), 20.7 (CH₃, AcOMe), 20.7 (CH₃, AcOMe), 20.6 (CH₃, AcOMe).

ES-HRMS calcd for C₅₂H₆₂N₆NaO₂₀S₂ 1177.3358, found m/z 1177.3363 [M+Na]⁺

**(E)-1,2-Bis[4-(1-(α -D-mannose)-1,2,3-triazol-4-ylmethoxy)-phenyl]ethane
(AG139)**



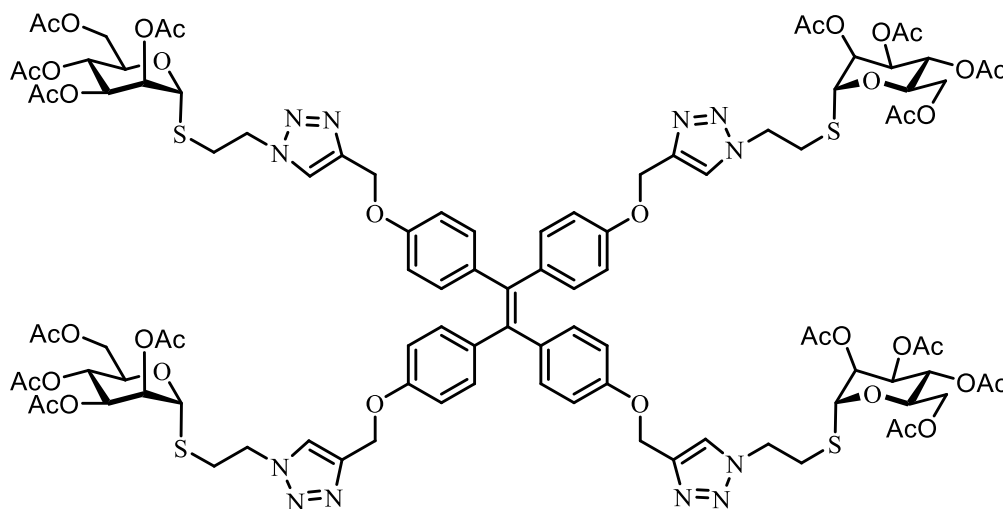
To a 0°C cooled suspension of AG137 (209 mg, 0.18 mmol) in dry MeOH (22 mL), NaOMe in MeOH was added until the solution reached pH 10. The reaction mixture was allowed to warm up at room temperature and stirred overnight. Glacial acetic acid was added to neutralise (pH=7) the solution. [Note: it is possible to use Amberlite IR- 120 H⁺ instead of AcOH,]. The solvent was removed under diminished pressure. Reverse phase column (3 volumes of water were flushed through the column, to ensure salt removal, followed by a 6:4 MeCN-H₂O) gave the title compound as a white solid (141 mg, 95%). [Note: dissolve the compound, in a minimal amount of water with few drops of MeCN. If the solid crashes out, add one drop of AcOH and slightly heat up].

¹H NMR (500 MHz, DMSO-*d*₆) δ = 8.26 (s, 2H, triazole), 7.49 (d, *J* = 8.9 Hz, 4H, aromatic H), 7.03 (s, 2H, allylic), 7.02 (d, *J* = 8.9 Hz, 4H, aromatic H), 5.23 (s, 2H, H-1), 5.14 (s, 4H, CH₂), 5.00 (d, *J* = 4.4 Hz, 2H, OH), 4.83 (d, *J* = 5.3 Hz, 2H, OH), 4.71 (d, *J* = 5.4 Hz, 2H, OH), 4.68 – 4.51 (overlapped signals, 4H, CH₂), 3.72 (ddd, *J* = 11.7, 6.0, 2.1 Hz, 2H, CH₂), 3.69 – 3.61 (overlapped signals, 4H), 3.50 – 3.42 (m, 2H, CH₂), 3.39- 3.395 (overlapped signals, 4H), 3.08 (td, *J* = 6.9, 2.5 Hz, 4H, CH₂).

¹³C NMR (125 MHz, DMSO-*d*₆) δ = 157.9 (C), 142.9 (C), 130.8 (C), 127.9 (CH), 126.4 (CH), 125.2 (CH), 115.4 (CH), 110.0, 85.8 (CH, C-1), 75.3 (CH), 72.1 (CH), 71.9 (CH), 67.7 (CH), 61.6 (2 X CH₂), 49.7 (CH₂), 31.2 (CH₂).

ES-HRMS calcd for C₃₆H₄₅N₆O₁₂S₂ 817.2537, found *m/z* 817.2535 [M-H]⁻

1,1,2,2-Tetrakis[4-(1-thio- α - D-mannopyranosyl-pentaacetate)-1,2,3-triazol-4-ylmethoxy-phenyl]ethene (AG134)



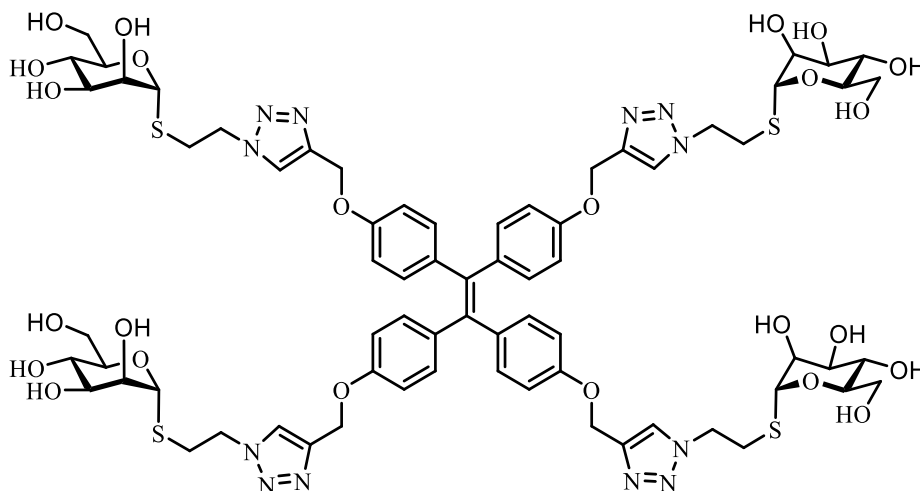
To a solution of AG088 (85 mg, 0.15 mmol) in degassed THF-H₂O (1:1, 15mL) were added AG132 (267 mg, 0.62 mmol), sodium ascorbate (36 mg, 0.18 mmol) and CuSO₄ · 5H₂O (45 mg, 0.18 mmol). The reaction mixture was stirred under inert atmosphere at 50°C for 14h. Tetrahydrofuran was then removed under diminished pressure followed by the dilution of the solution with CH₂Cl₂. This mixture was then washed with water. The aqueous layer was re-extracted with a further portion of CH₂Cl₂. The combined organic layers were dried over Na₂SO₄ and the solvent was removed under diminished pressure. Chromatography of the crude material (98:2→93:3 CH₂Cl₂-MeOH) gave the title compound (160 mg, 78 %) as a colourless solid.

¹H NMR (500 MHz, chloroform-*d*) δ = 7.68 (s, 4H, triazole), 6.93 (d, *J* = 8.7 Hz, 8H, aromatic H), 6.72 (d, *J* = 8.8 Hz, 8H, aromatic H), 5.31 (dd, *J* = 3.4, 1.4 Hz, 4H, H-3), 5.29 (br s, 4H, H-5), 5.26 (br s, 4H H-1), 5.19 (dd, *J* = 10.0, 3.4 Hz, 4H, H-4), 5.14 (s, 8H, CH₂), 4.61 (dq, *J* = 17.2, 7.1 Hz, 8H, CH₂), 4.31 -4.26 (overlapped signals, 8H, H-2, H-6), 4.16 – 4.07 (m, 4H, CH₂, H-6'), 3.17 (overlapped signals, 8H, CH₂), 2.16 (s, 12H, AcOMe), 2.05 (s, 12H, AcOMe), 2.04 (s, 12H, AcOMe), 1.99 (s, 12H, AcOMe).

¹³C NMR (125 MHz, chloroform-*d*) δ = 170.5 (C), 169.9 (C), 169.7 (C), 169.6 (C), 156.6 (C), 144.3 (C), 137.2 (C), 132.6 (CH, aromatic), 123.0 (CH, triazole), 113.9 (CH, aromatic), 82.9 (CH, C-1), 70.7 (CH, C-3), 69.5 (CH, C-2), 69.1 (CH, C-5), 66.1 (CH, C-4), 62.4 (CH₂, C-6), 61.8 (CH₂), 49.6 (CH₂), 31.5 (CH₂), 20.8 (CH₃, AcOMe), 20.7 (CH₃, AcOMe), 20.7 (CH₃, AcOMe), 20.6 (CH₃, AcOMe).

ES-HRMS calcd for C₁₀₂H₁₂₀N₁₂NaO₄₀S₄ 2303.6505, found *m/z* 2303.6499[M+Na]⁺

1,1,2,2-Tetrakis[4-(1-thio- α -D-mannopyranosyl)-1,2,3-triazol-4-ylmethoxy-phenyl]ethene (AG138)



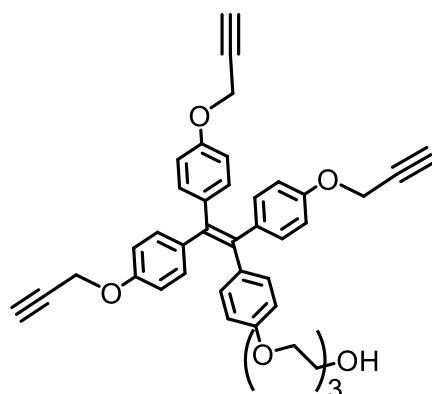
To a 0°C cooled suspension of AG134 (88 mg, 0.038 mmol) in dry MeOH (8 mL), freshly prepared 1M NaOMe in MeOH was added until the solution reached pH 10. The reaction mixture was allowed to warm up at rt and stirred overnight. Glacial acetic acid was added to neutralise (pH=7) the solution. [Note: it is possible to use Amberlite IR- 120 H⁺ instead of AcOH,]. The solvent was removed under diminished pressure. Reverse phase column (3 volumes of water were flushed through the column, to ensure salt removal, followed by a 6:4 MeCN-H₂O) gave the title compound as a white solid (60 mg, 98%). [NOTE: dissolve the compound, in a minimal amount of water with few drops of MeCN. If the solid crashes out, add one drop of AcOH and slightly heat up].

¹H NMR (500 MHz, DMSO-*d*₆) δ = 8.25 (s, 4H, triazole), 6.89 (d, *J* = 8.7 Hz, 8H, aromatic H), 6.82 (d, *J* = 8.8 Hz, 8H, aromatic H), 5.24 (s, 4H, H-1), 5.05 (s, 8H, CH₂), 4.88 (brs, 4H, OH), 4.79 (brs, 4H, OH), 4.69 – 4.49 (overlapping signals, 8H, CH₂), 3.73 (dd, *J* = 11.3, 4.3 Hz, 4H, H-5), 3.68 (s, 4H, H- Man), 3.68 – 3.62 (m, 4H, H-2), 3.47 (dt, *J* = 11.6, 5.9 Hz, 4H, CH₂), 3.41-3.37 (overlapping signals, 8H, H-Man, H- Man), 3.10-3.01 (overlapping signals, 8H, CH₂).

¹³C NMR (125 MHz, DMSO-*d*₆) δ = 156.9 (C), 142.9 (C), 138.5 (C), 136.9 (C), 132.5 (CH, aromatic), 125.2 (CH, triazole), 114.3 (CH, aromatic), 85.8 (CH, C-1), 75.3 (CH), 72.1 (CH), 71.9 (CH), 67.7 (CH), 61.6 (CH₂), 61.4 (CH₂), 49.6 (CH₂), 31.1 (CH₂).

ES-HRMS calcd for C₇₀H₈₇N₁₂O₂₄S₄ 1607.4839, found *m/z* 1607.4844 [M-H]⁻

2-(2-(2-(4-(1,2,2-tris(4-(prop-2-yn-1-yloxy)phenyl)vinyl)phenoxy)ethoxy)ethoxy)ethan-1-ol (AG142)



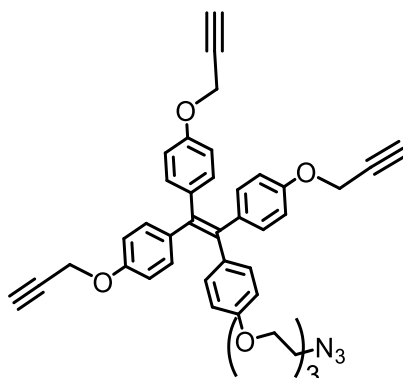
To a solution of AG141 (1 g, 1.95mmol) in dry DMF (50 mL) were added 2-[2-(2-chloroethoxy)ethoxy]ethanol (697 μ L, 4.89 mmol), K_2CO_3 (2.02 g, 14062 mmol) and potassium iodide (324 mg, 1.95 mmol). The reaction mixture was stirred at 100 °C for 16 h then diluted with DCM and saturated NH_4Cl solution. The organic phase was washed with water and brine. The organic phase was dried over Na_2SO_4 and the solvent removed under reduced pressure. Column chromatography (1:1 cyclohexaneF-EtOAc \rightarrow 1:4 cyclohexane -EtOAc) yielded the titled compound (1.1 g, 95%).

1H NMR (500 MHz, Chloroform-d) δ 6.97 – 6.87 (8H, aromatic H), 6.70 (d, J = 8.7 Hz, 6H, aromatic H), 6.64 (d, J = 8.1 Hz, 2H, aromatic H), 4.61 (s, 6H, CH_2), 4.06 (t, J = 4.3 Hz, 2H, CH_2), 3.82 (t, J = 4.3 Hz, 2H, CH_2), 3.75 – 3.66 (m, 6H, CH_2), 3.65 – 3.55 (m, 2H, CH_2), 2.54 – 2.47 (overlapping signals, 2H, alkyne), 2.42 – 2.36 (m, 1H, alkyne).

^{13}C NMR (125 MHz, Chloroform-d) δ 157.1 (C), 156.1 (C), 138.8 (C), 138.5 (C), 137.6 (C), 137.6 (C), 136.9 (C), 132.6 (CH, aromatic), 114.1 (CH, aromatic), 113.9 (CH, aromatic), 78.7 (C), 75.6 (CH, alkyne), 75.5 (CH, alkyne), 72.6 (CH_2), 70.9 (CH_2), 70.5 (CH_2), 69.9 (CH_2), 67.3 (CH_2), 61.9 (CH_2), 55.9 (CH_2).

ES-HRMS calcd for $C_{41}H_{37}O_7$ 641.2539, found m/z 641.2534 [M-H] $^-$

4,4',4''-(2-(4-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)phenyl)ethene-1,1,2-triyl)tris((prop-2-yn-1-yloxy)benzene) (AG185/AG149)



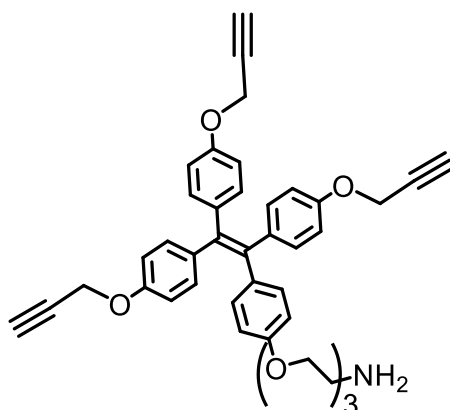
To a 0°C cooled solution of AG142 (600 mg, 0.93 mmol, 1 eq.) in THF (10 mL), PPh₃ (783 mg, 3 mmol, 3.2 eq.) and DIAD (603 mg, 3 mmol, 3.2 eq.) were added. The reaction mixture was stirred at 0°C until complete reaction of the reagents. (*Note*: Check by TLC 7:3 C-EtOAc, it takes ca 1h. In case you just see PPh₃ add DIAD and *vice versa*). DPPA (847 mg, 3.1 mmol, 3.3 eq.) is then added and the reaction is allowed to warm up at room temperature and stirred for 2 h. The reaction is then diluted with EtOAc (100 mL) and washed with H₂O (3x 15 mL). The organic phase was dried over Na₂SO₄ and the solvent removed under reduced pressure. Column chromatography (6:4 cyclohexane-EtOAc) yielded the titled compound (500 mg, 80%).

¹H NMR (500 MHz, Chloroform-d) δ 6.96 – 6.88 (m, 8H, aromatic H), 6.70 (d, J = 7.9 Hz, 6H, aromatic H), 6.64 (d, J = 8.7 Hz, 2H, aromatic H), 4.61 (d, J = 2.2 Hz, 6H, CH₂), 4.06 (t, J = 5.2 Hz, 2H, CH₂), 3.83 (t, J = 5.0 Hz, 2H, CH₂), 3.75 – 3.71 (m, 2H, CH₂), 3.70 – 3.65 (overlapping signals, 4H, CH₂), 3.38 (t, J = 5.0 Hz, 2H, CH₂), 2.53 – 2.48 (overlapping signals, 3H, alkyne);

¹³C NMR (125 MHz, Chloroform-d) δ 157.2 (C), 156.1 (C), 138.9 (C), 138.5 (C), 137.7 (C), 136.9 (C), 132.6 (CH, aromatic), 114.1 (CH, aromatic), 113.9 (CH, aromatic), 78.8 (C), 75.6 (CH, alkyne), 75.5 (CH, alkyne), 71.0 (CH₂), 70.9 (CH₂), 70.2 (CH₂), 70.0 (CH₂), 67.3 (CH₂), 55.9 (CH₂), 50.8 (CH₂);

ES-HRMS calcd for C₄₁H₃₇N₃O₆Na 690.2580, found m/z 690.2584 [M+Na]⁺

2-(2-(2-(4-(1,2,2-tris(4-(prop-2-yn-1-yloxy)phenyl)vinyl)phenoxy)ethoxy)ethoxy)ethan-1-amine (AG150)



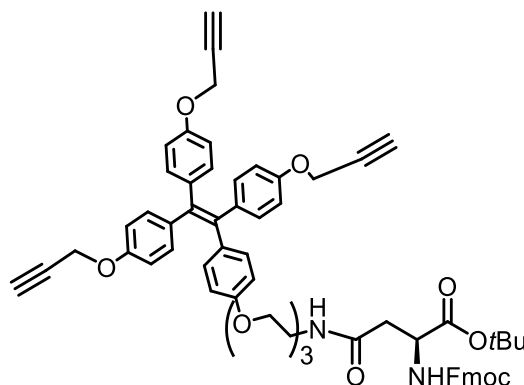
To a solution of AG149/AG185 (500 mg, 0.75mmol) in THF (5 mL), PBU_3 (303 mg, 1.5 mmol) was added followed by water (500 μL). The reaction was stirred at room temperature under nitrogen atmosphere for 2 h. Upon completion, the reaction mixture was quenched with Et_3N and filtered through Celite. The solvent was removed in vacuo. Column chromatography (95:05 \rightarrow 85:13:02 CHCl_3 -MeOH- NH_4OH) yielded the titled compound (427 mg, 85%).

^1H NMR (500 MHz, Chloroform-d) δ 6.97 – 6.86 (overlapping signals, 8H, aromatic H), 6.70 (d, J = 8.6 Hz, 6H, aromatic H), 6.66 – 6.60 (overlapping signals, 2H, aromatic H), 4.72 – 4.52 (overlapping signals, 6H, CH_2), 4.11 – 3.99 (m, 2H, CH_2), 3.85 – 3.77 (m, 2H, CH_2), 3.74 – 3.69 (m, 2H, CH_2), 3.68 – 3.61 (m, 2H, CH_2), 3.52 (t, J = 5.2 Hz, 2H, CH_2), 2.87 (t, J = 5.1 Hz, 2H, CH_2), 2.61 – 2.42 (overlapping signals, 3H, alkyne);

^{13}C NMR (125 MHz, Chloroform-d) δ 157.1(C), 155.9 (C), 138.7 (C), 138.3 (C), 137.5 (C), 136.7 (C), 132.5 (CH, aromatic), 114.0 (CH, aromatic), 113.7 (CH, aromatic), 78.6 (C), 75.4 (CH, alkyne), 75.4 (CH, alkyne), 73.2(CH_2), 70.8 (CH_2), 70.3 (CH_2), 69.8 (CH_2), 67.2 (CH_2), 41.7 (CH_2);

ES-HRMS calcd for $\text{C}_{41}\text{H}_{40}\text{NO}_6$ 642.2777, found m/z 642.2780 $[\text{M}+\text{H}]^+$

tert-butyl N2-(((9H-fluoren-9-yl)methoxy)carbonyl)-N4-(2-(2-(2-(4-(1,2,2-tris(4-(prop-2-yn-1-yloxy)phenyl)vinyl)phenoxy)ethoxy)ethoxy)ethyl)-L-asparaginate (AG151)



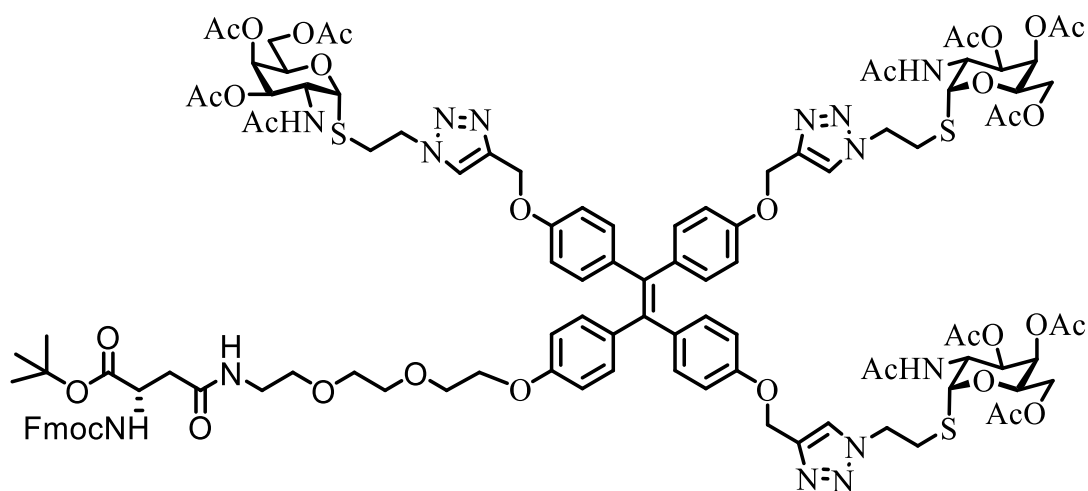
To a solution of AG150 (363 mg, 0.57 mmol) in dry DCM (20 mL), Fmoc-L-Asp *Ot*-Bu (234 mg, 0.57 mmol) and EEDQ (155 mg, 0.63 mmol) were added. The reaction mixture was stirred at 30 °C for 16 h. The solvent was removed in vacuo. Column chromatography (8:2 → 1:1 cyclohexane-EtOAc) yielded the titled compound (524 mg, 89%).

¹H NMR (500 MHz, Chloroform-*d*) δ 7.75 (d, *J* = 7.6 Hz, 2H, aromatic Fmoc), 7.58 (d, *J* = 7.3 Hz, 2H, aromatic Fmoc), 7.39 (t, *J* = 7.5 Hz, 2H, aromatic Fmoc), 7.30 (t, *J* = 7.4 Hz, 2H, aromatic Fmoc), 6.96 – 6.85 (overlapping signals, 8H, aromatic H), 6.79 (*brs*, 1H, NH), 6.74 – 6.65 (overlapping signals, 6H, aromatic H), 6.61 (d, *J* = 8.7 Hz, 2H, aromatic H), 5.93 (d, *J* = 8.1 Hz, 1H, NH), 4.67 – 4.56 (overlapping signals, 6H, CH₂), 4.51 (s, 1H, H-Asp), 4.41 (t, *J* = 6.1 Hz, 2H, CH-Fmoc), 4.25 – 4.18 (m, 1H, CH-Fmoc), 4.01 (t, *J* = 4.7 Hz, 2H, CH₂), 3.77 (t, *J* = 4.7 Hz, 2H, CH₂), 3.66 (d, *J* = 5.1 Hz, 2H, CH₂), 3.61 (d, *J* = 4.2 Hz, 2H, CH₂), 3.54 (d, *J* = 4.9 Hz, 2H, CH₂), 3.49 – 3.39 (m, 2H, CH₂), 2.94 – 2.82 (m, 1H, CH₂-Asp), 2.60 (dd, *J* = 16.7, 6.2 Hz, 1H, CH₂-Asp), 2.47-2.51 (overlapping signals, 3H, alkyne), 1.44 (s, 9H, *t*-butyl).

¹³C NMR (125 MHz, Chloroform-*d*) δ 157.0 (C), 155.9 (C), 155.9 (C), 143.7 (C), 141.3 (C), 138.7 (C), 138.3 (C), 137.5 (C), 136.7 (C), 132.5 (CH, aromatic), 127.8 (CH, aromatic), 127.1 (CH, aromatic), 125.0 (CH, aromatic), 120.0 (CH, aromatic), 114.0 (CH, aromatic), 113.7 (CH, aromatic), 78.6 (C, alkyne), 75.4 (CH, alkyne), 75.4 (CH, alkyne), 70.7 (CH₂), 70.4 (CH₂), 69.7 (CH₂), 69.6 (CH₂), 67.1 (CH₂), 55.7 (CH₂), 51.3 (CH, Fmoc HSQC), 47.1 (CH, Fmoc), 39.5 (CH₂), 28.0 (CH₃).

ES-HRMS calcd for C₆₄H₆₃N₂O₁₁ 1035.4432, found *m/z* 1035.4441 [M+H]⁺

tert-butyl N2-(((9H-fluoren-9-yl)methoxy)carbonyl)-N4-(2 1-(ethyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-1-thio- α -D galactopyranosyl)phenyl)vinyl)phenoxy)ethoxy)ethoxy)ethyl)-L-asparaginate (AG152)



To a solution of AG151 (374 mg, 0.36 mmol) in degassed THF/H₂O (50 mL, 1:1), AG123 (548 mg, 1.27 mmol), sodium ascorbate (62 mg, 0.31 mmol) and CuSO₄•5H₂O (74 mg, 0.3 mmol) were added respectively. The reaction mixture was stirred at 35 °C for 16 h, then the THF was removed in vacuo and the water phase was extracted with DCM (3X 25 mL). Column chromatography (98:2 → 95:05 DCM-MeOH) yielded the titled compound (631 mg, 78%).

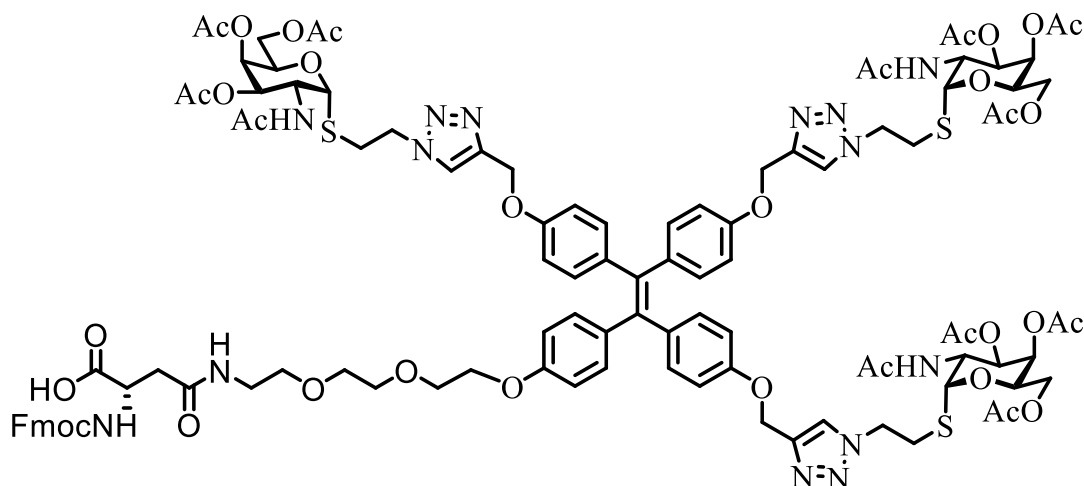
¹H NMR (500 MHz, Chloroform-*d*) δ 7.74 (d, $J = 7.5$ Hz, 2H, aromatic Fmoc), 7.66 (s, 3H, triazole), 7.59 (dd, $J = 7.2, 2.5$ Hz, 2H, aromatic Fmoc), 7.38 (t, $J = 7.4$ Hz, 2H, aromatic Fmoc), 7.28 (t, $J = 7.3$ Hz, 2H, aromatic Fmoc), 6.96 - 6.85 (overlapped signals, 8H, aromatic H), 6.68-6.72 (overlapped signals, 6H, aromatic H), 6.63 (d, $J = 7.2$, 2H, aromatic H), 6.35 (brs, 1H, NH), 6.11 (d, $J = 8.1$ Hz, 1H, NH), 6.01 (d, $J = 8.0$ Hz, 1H, NHAc), 5.97 (d, $J = 8.2$ Hz, 1H, NHAc), 5.84 (d, $J = 7.9$ Hz, 1H, NHAc), 5.70 - 5.63 (overlapped signals, 3H, H-1), 5.36-5.38 (overlapping signals, 3H, H-4), 5.13 (s, 2H, CH₂), 5.12 (s, 2H, CH₂), 5.11 (s, 2H, CH₂), 5.00 (dd, $J = 11.8, 3.1$ Hz, 3H, H-3), 4.77 - 4.68 (overlapping signals, 3H, H-2), 4.65 - 4.54 (overlapping signals, 6H, CH₂), 4.51 (q, $J = 7.8, 6.5$ Hz, 3H, H-5), 4.44 (dt, $J = 8.4, 4.6$ Hz, 1H, CH Asp), 4.41 - 4.36 (m, 1H, CH₂ Fmoc), 4.33 - 4.25 (m, 1H, CH₂ Fmoc), 4.21 (t, $J = 7.1$ Hz, 1H, CH Fmoc), 4.17 - 4.07 (overlapping signals, 6H, H-6), 4.05 - 4.02 (m, 2H, CH₂ TEG), 3.80 - 3.76 (m, 2H, CH₂ TEG), 3.64 (d, $J = 4.5$ Hz, 2H, CH₂ TEG), 3.59 (d, $J = 4.5$ Hz, 2H, CH₂ TEG), 3.53 (t, $J = 5.1$ Hz, 2H, CH₂ TEG), 3.42 (t, $J = 5.1$ Hz, 2H, CH₂ TEG), 3.21 - 3.11 (m, 3H, CH₂), 3.10 - 2.97 (m, 3H, H), 2.83

(dd, $J = 15.5, 4.9$ Hz, 1H, CH₂ Asp), 2.66 (dd, $J = 15.7, 4.2$ Hz, 1H, CH₂ Asp), 2.16 (s, 9H, NHAc), 1.99 (s, 9H, OAc), 1.97 (s, 9H, OAc), 1.96 (s, 9H, OAc), 1.45 (s, 9H, *t*-Bu).

¹³C NMR (126 MHz, cdCl₃) δ 171.0 (C), 170.6 (C), 170.6 (C), 170.5 (C), 170.4 (C), 170.4 (C), 170.2 (C), 170.1 (C), 156.9 (C), 156.6 (C), 156.5 (C), 144.3 (C), 143.9 (C), 141.2 (C), 138.6 (C), 138.4 (C), 137.3 (C), 137.2 (C), 136.8 (C), 132.5 (CH, aromatic), 127.7 (CH, Fmoc aromatic), 127.1 (CH, Fmoc aromatic), 125.2 (CH, Fmoc aromatic), 123.1 (CH, Fmoc aromatic), 119.9 (CH, triazole), 113.9 (CH, aromatic), 113.9 (CH, aromatic), 113.7 (CH, aromatic), 85.1 (CH, C-1), 82.2 (C), 70.6 (CH₂, TEG), 70.3 (CH₂, TEG), 69.7 (CH₂, TEG), 68.2 (CH, C-3), 67.6 (CH, C-5), 67.1 (CH, C-4), 62.0 (CH₂, C-6), 61.9 (CH₂, C-6), 51.5 (CH, Asp), 49.6 (CH₂), 48.5 (CH, C-2), 47.1 (CH, Fmoc), 39.3 (CH₂, TEG), 37.9 (CH₂ Asp), 31.1 (CH₂), 27.9 (CH₃, *t*-Bu), 23.2 (CH₃, NHAc), 20.7 (CH₃, OAc), 20.7 (CH₃, OAc), 20.6 (CH₃, OAc).

ES-HRMS calcd for C₁₁₂H₁₃₄N₁₄NaO₃₅S₃ 2353.8196, found m/z 2353.8201 [M+Na]⁺

N2-(((9H-fluoren-9-yl)methoxy)carbonyl)-N4-(2 1-(ethyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-1-thio- α -D galactopyranosyl)phenyl)vinyl)phenoxy)ethoxy)ethoxy)ethyl)-L-asparaginate (AG153)



A solution of AG152 (1.2 g, 0.53 mmol) in TFA/DCM/H₂O (48:48:4, 10 mL) was allowed to stir at room temperature for 2h. The solvent was removed in vacuo and an azeotropic mixture with toluene removed under vacuo several times. Column chromatography (98:2:0.1 → 90:10:0.1 DCM-MeOH-AcOH) yielded the titled compound (1.15 g, 95%).

¹H NMR (600 MHz, chloroform-d) δ 7.82 – 7.75 (m, 2H, aromatic Fmoc), 7.76 – 7.70 (overlapping signals, 3H, triazole), 7.58 (d, J = 6.4 Hz, 2H, aromatic Fmoc), 7.38 (t, J = 7.4 Hz, 2H, aromatic Fmoc), 7.28 (t, J = 7.3 Hz, 2H, aromatic Fmoc), 7.26 – 7.22 (m, 1H, aromatic Fmoc), 7.17 (d, J = 7.1 Hz, 1H, aromatic Fmoc), 6.97 – 6.85 (overlapping signals, 8H, aromatic H), 6.65-6.69 (overlapping signals, 6H, aromatic H), 6.61 (d, J = 8.4 Hz, 2H, aromatic protons), 6.54 (br s, 1H, NH), 6.20 (br s, 1H, NH), 5.70 (d, J = 4.8 Hz, 3H, H-1), 5.39 (s, 3H, H-4), 5.21 – 5.07 (overlapping signals, 6H, CH₂), 5.01 (d, J = 11.7 Hz, 3H, H-3), 4.71 (dt, J = 11.8, 7.4 Hz, 3H, H-2), 4.66 – 4.56 (overlapping signals, 6H, CH₂), 4.55 – 4.45 (overlapping signals, 4H, CH Asp and H-5), 4.39 – 4.28 (m, 2H, CH₂ Fmoc), 4.22 – 4.17 (m, 2H, CH Fmoc), 4.17 – 4.08 (overlapping signals, 6H, H-6), 4.08 – 4.01 (m, 2H, CH₂ TEG), 3.86 – 3.79 (m, 2H, CH₂ TEG), 3.72 – 3.67 (m, 2H, CH₂ TEG), 3.66 – 3.58 (m, 2H, CH₂ TEG), 3.57 – 3.47 (m, 2H, CH₂ TEG), 3.39 – 3.30 (m, 2H, CH₂ TEG), 3.23 – 3.14 (overlapping signals, 3H, CH₂), 3.14 – 3.04 (overlapping signals, 3H, CH₂), 2.86 – 2.78 (m, 1H, CH₂ Asp), 2.83 - 2.79 (m, 1H, CH₂ Asp), 2.17 (d, J = 2.0 Hz, 9H, NHAc), 2.08 – 1.96 (overlapping signals, 27H, OAc).

^{13}C NMR (125 MHz, chloroform- d) δ 173.2 (C), 171.2 (C), 170.8 (C), 170.4 (C), 159.2 (C), 159.0 (C), 156.7 (C), 156.2 (C), 143.6 (C), 141.2 (C), 138.7 (C), 138.5 (C), 137.4 (C), 132.6 (CH, aromatic), 129.0 (CH, Fmoc aromatic), 128.2 (CH, Fmoc aromatic), 127.8 (CH, Fmoc aromatic), 127.1 (CH, Fmoc aromatic), 125.3 (CH, Fmoc aromatic), 125.1 (CH, Fmoc aromatic), 120.0 (CH, triazole), 113.8 (CH, aromatic), 113.7 (CH, aromatic), 84.6 (CH,C-1), 70.7 (CH₂, TEG), 70.4 (CH₂, TEG), 70.0 (CH₂, TEG), 69.7 (CH₂, TEG), 69.0 (CH₂, TEG), 68.0 (CH, C-3), 67.6 (CH, C-5), 67.3 (CH₂, TEG), 66.9 (CH, C-4), 61.9 (CH₂, C-6), 61.0 (CH₂,) , 50.9 (CH, Asp), 50.2 (CH₂), 49.0 (CH, C-2), 46.9 (CH, Fmoc), 39.8 (CH₂, TEG), 37.8 (CH₂, Asp), 30.9 (CH₂), 22.6 (CH₃, NHAc), 22.6 (CH₃, NHAc), 22.5 (CH₃, NHAc), 20.6 (CH₃, OAc), 20.6 (CH₃, OAc), 20.6 (CH₃, OAc), 20.6 (CH₃, OAc).

ES-HRMS calcd for C₁₀₈H₁₂₅N₁₄O₃₅S₃ 2273.7594, found m/z 2273.7592 [M-H]⁻

$[\alpha]_{\text{D}}^{20} + 59$ (c 0.13, CHCl₃).

6.3. Peptide synthesis and characterisation

6.3.1. General

SPPS of peptides and glycopeptides was carried out on a Syro I peptide synthesizer by Multisynth GmbH, Witten. Solid phase resins were purchased from Rapp Polymere GmbH, Tübingen. Protected amino acid building blocks for Fmoc-SPPS (Novabiochem®) were purchased from Merck KGaA, Darmstadt and Merck Schuchardt OHG, Hohenbrunn. N,N-dimethylformamide, N-methylpyrrolidone, trifluoroacetic acid and piperidine were purchased from Biosolve Chimie SARL, Dieuze, France. Coupling reagents HBTU and HATU (Novabiochem®) were purchased from Merck KGaA, Darmstadt. HOBt hydrate was purchased from Sigma-Aldrich GmbH, Steinheim and recrystallized from absolute ethanol and dried at reduced pressure. HOAt was purchased from GL Biochem, Shanghai, China.

RP-HPLC: Analytic RP-HPLC was performed on a Dionex U-3000 (Thermo Scientific) system (DR-3600 six channel degasser, LPG-3x00 pump, TCC-3100 column compartment, DAD-3000 UV/VIS diode array detector, WPS-3000 autosampler). A Luna C18(2) (3 μm , 100Å, 150 x 2.0 mm) column from Phenomenex was applied for analytical HPLC.

Preparative RP-HPLC was performed on a Dionex U-3000 system (HPG-3200P, VWD-3400 UV/VIS detector, AFC-3000 sampler). A Luna C18(2) (10 μm , 100Å, 250 x 21.2 mm) column from Phenomenex was applied for preparative HPLC. The flow rate was set to 20 mL/min.

(Glyco-)Peptides were detected at $\lambda = 214$ nm. The chromatograms were analyzed with Dionex Chromeleon (version 6.80DU10a Build 2826(171948)).

6.3.2. General protocol for automated glycopeptide-solid phase peptide synthesis

Automated solid phase synthesis standard protocol: The syntheses were performed with Syro I automated peptide synthesizer by MultisynTech GmbH, Witten, on a 2 mL x 24 reactor plate. The synthesis scale is reported in the experimental section for each single peptide. Stock solution of the relevant Fmoc-aa-OH 0.5M in DMF (Fmoc-Phe-OH 0.5 M in NMP), HBTU and HOBt 0.45 M in DMF, DIPEA 2M in NMP and 40% piperidine in DMF were preloaded accordingly to the required synthesis. Preloaded TentaGel-Fmoc-aa-Trt was used, if not otherwise specified. The resin was manually swelled with 500 μ L of DCM for 30 min and then washed with DMF for 5 times (500 μ L, 40s vortex, 20s vac). The Fmoc-protecting group was initially cleaved using 40% piperidine in DMF (46.2 μ L per 1 μ mol batch size; 2 x 3 min + 1 x 9 min; 15s vortex, 45 s break), then the resin was washed with DMF (46.2 μ L per 1 μ mol batch size; 6 x 1 min, 15 s vortex; 45 s break). In automated reaction cycles the corresponding Fmoc-aa-OH (8 equiv.), HBTU (7.6 equiv.), HOBt (8 equiv.) and DIPEA (16 equiv.) were added in the reactor and shaken by vortex (40 min, 15 s vortex; 2.75 min break). A reaction cycle was concluded by Fmoc-deprotection before the next cycle was carried out. Glycosylated amino acids, spacer and MGL/ligand with the adopted coupling reagents were added manually into the synthesis reactors. The coupling condition are reported in the experimental section for each single peptide.

6.3.3. Glycosylated amino acids manual coupling protocol

To the solution of glycosylated amino acids (T* and S*) in DMF (1.5 eq in 200 μ L) were added HATU (1.4 eq), HOAt (1.4 eq) and DIPEA (2.8 eq). The solution was quickly added to the reactor and shaken by vortex (15 s vortex; 2.45 min break) at room temperature for 8h. After the reaction, the resin was washed with DMF for 5 times (500 μ L, 40s vortex, 20s vac). The Fmoc-group was cleaved by triple addition of 500 μ L each of 20 vol% piperidine in DMF (3 x 3 x 9 min; 15s vortex, 45 s break) and the resin washed with DMF (5 x 1 min 6 x 500 μ L, 15 s vortex; 45 s break).

6.3.4. Spacer manual coupling protocol

To the solution of 3-(2-(2-(2-aminoethoxy)ethoxy)ethoxy)propanoic acid in DMF (3 eq in 200 μ L) were added HBTU (3 eq) , HOBT (3 eq) and DIPEA (6 eq) . The solution was quickly added to the reactor and shaken by vortex (15 s vortex; 2.45 min break) at room temperature for 3h. After the reaction, the resin was washed with DMF for 5 times (500 μ L, 40s vortex, 20s vac). The the Fmoc-group was cleaved by triple addition of 500 μ L each of 40 vol% piperidine in DMF (3 x 3 x 9 min; 15s vortex, 45 s break) and the resin washed with DMF (5 x 1 min 6 x 500 μ L, 15 s vortex; 45 s break).

6.3.5. Biotinated Lysine manual coupling protocol

To the solution of Lys (Biotin) in NMP (3 eq in 200 μ L) were added HBTU (3 eq), HOBT (3 eq) and DIPEA (6 eq). The solution was quickly added to the reactor and shaken by vortex (15 s vortex; 2.45 min break) at room temperature for 3h. After the reaction, the resin was washed with DMF for 5 times (500 μ L, 40s vortex, 20s vac). The the Fmoc-group was cleaved by triple addition of 500 μ L each of 40 vol% piperidine in DMF (3 x 3 x 9 min; 15s vortex, 45 s break) and the resin washed with DMF (5 x 1 min 6 x 500 μ L, 15 s vortex; 45 s break).

6.3.6. AG153 manual coupling protocol

To the solution of AG153 in DMF (2 eq. in 200 μ L) were added HATU (1.8 eq.), HOAt (1.8 eq.) and DIPEA (3.6 eq.). The solution was quickly added to the reactor and shaken by vortex (15 s vortex; 2.45 min break) at room temperature for 8h. After the reaction, the resin was washed with DMF for 5 times (500 μ L, 40s vortex, 20s vac). The the Fmoc-group was cleaved by triple addition of 500 μ L each of 20 vol% piperidine in DMF (3 x 3 x 9 min; 15s vortex, 45 s break) and the resin washed with DMF (5 x 1 min 6 x 500 μ L, 15 s vortex; 45 s break).

6.3.7. Release of the peptides from the solid phase resin protocol

After peptide synthesis the resin was washed with dichloromethane, isopropanol and diethylether (5 x 500 μ L each; 5 x 1 min; 15 s vortex; 45 s break). The resin was dried in an airstream for 30 min and then transferred from the synthesis reactor into a 2 mL syringe clogged by a frit. The glycopeptides were cleaved from the resin by three additions of TFA/TIPS/H₂O 15:0.9:0.9 (1 x 120 min + 2 x 10 min; 15 s vortex; 2.45 min break). The combined filtrates were co-evaporated with toluene. C18-Carttige filtration protocol for glycosylated peptides: glycosylate peptides were dissolved in water (5 mL) and loaded to prepacked (MeOH 5mL x 5, H₂O 5mL x 5) C18-column (Waters Sep-Pak® Vac 6cc (1 g) . The column was eluted with H₂O (5 mL x 5) and with CH₃CN/H₂O 70:30 (5 mL x 7). Fractions containing acetonitrile were combined, the organic solvent removed in vacuo and the aqueous residue was lyophilized to give the crude glycopeptide product.

6.3.8. Removal of the carbohydrate acetyl protecting groups protocol

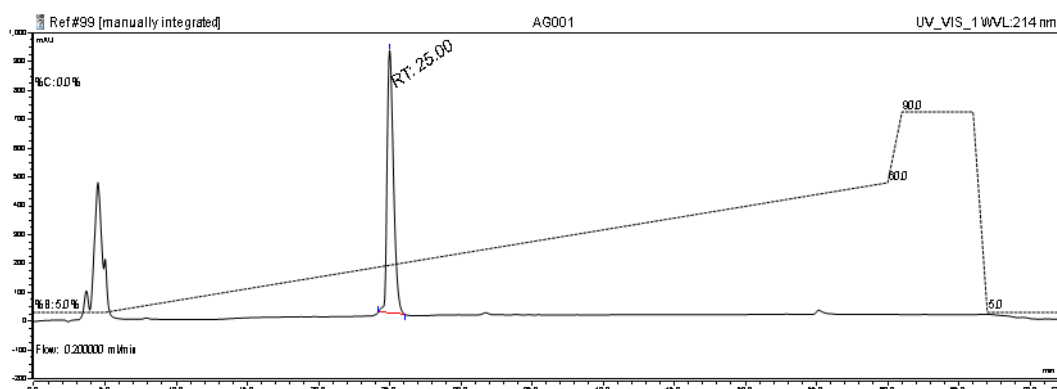
The crude glycopeptides were dissolved in 5-10 mL methanol and were treated with 1% sodium methoxide in methanol until a pH of 9.5 was reached (wet pH paper). The reaction was stirred at room temperature and it was followed by analytical HPLC control and MALDI-TOF (16-24 h).

The reaction mixture was neutralized with acetic acid and the solvents removed in vacuo. The crude glycopeptide was purified by preparative HPLC and lyophilized from water to give the product as a colorless lyophilizate foam.

6.3.9. Peptides characterisation

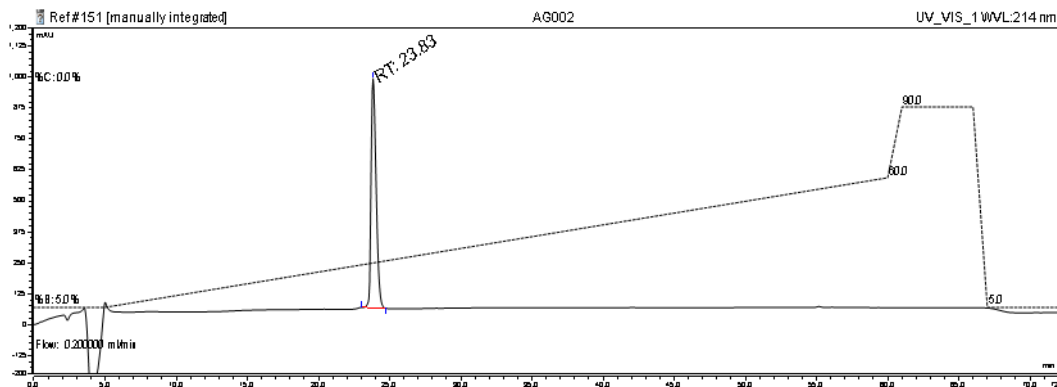
Peptide AG001, Sequence: NH₂-HGVTSAPDTRPAPGSTAPPA

Yield: 54% (12.7 mg, 6.73 μmol). Analytical HPLC Rt= 25.00 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 20.40 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95)→(45:55), 5-45 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XL™ hybride FT), *m/z*: 943.9734 ([M+2H]²⁺, calc. 943.9719), 629.6506 ([M+3H]³⁺, calc. 629.6505).



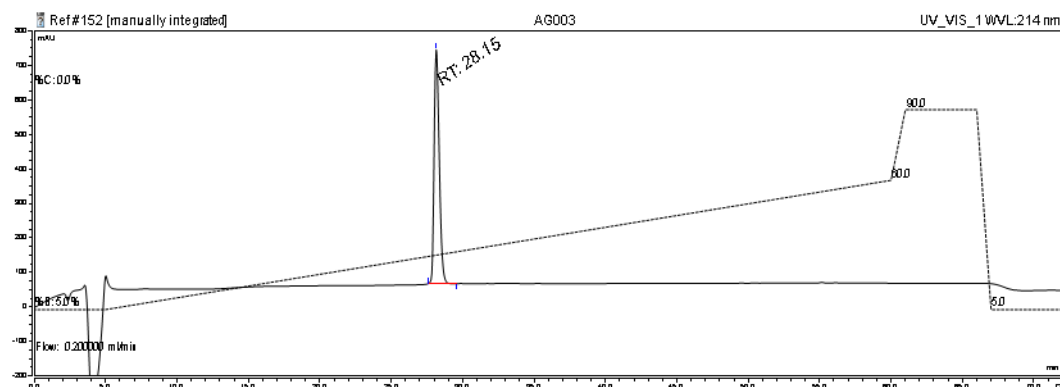
Peptide AG002, Sequence: NH₂-PAHGVTSAPDTRPAPGSTA

Yield: 55% (12.23 mg, 6.83 μmol). Analytical HPLC Rt= 23.83 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 19.31 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95)→(30:70), 5-30 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XL™ hybride FT), *m/z*: 895.4458 ([M+2H]²⁺, calc. 895.4456), 597.2993 ([M+3H]³⁺, calc. 597.2996).

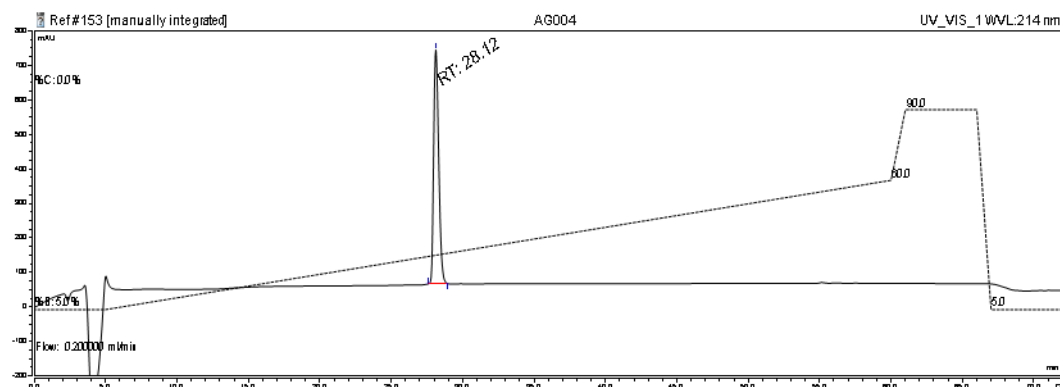


Peptide AG003, Sequence: NH₂-ASTGHATPLPVTD

Yield: 78% (12.42 mg, 9.81 μ mol). Analytical HPLC Rt= 28.15 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μ m, Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 23.77 min (GLSciences Inc, InertSustain C18, 5 μ m, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95) \rightarrow (30:70), 5-30 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1266.6333 ([M+H]⁺, calc. 1266.6330), 633.8199 ([M+2H]²⁺, calc. 633.8204).

Peptide AG004, Sequence: NH₂-GHATPLPVTD

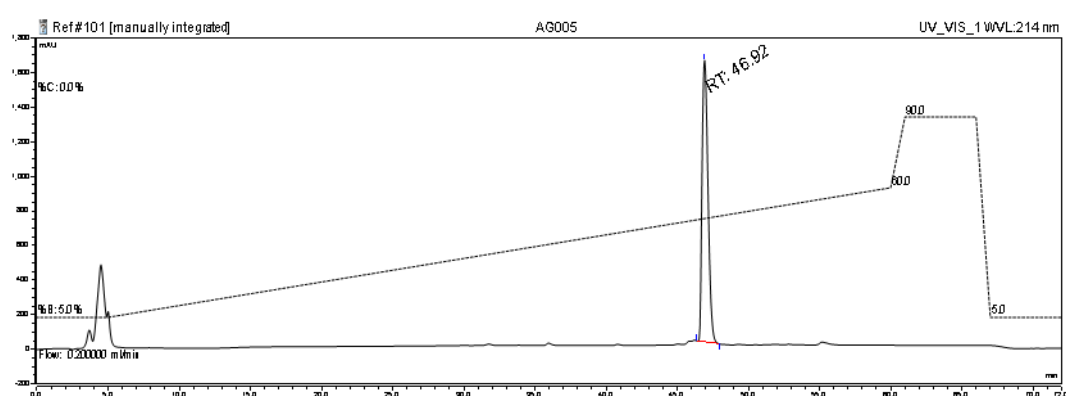
Yield: 80% (10.05 mg, 9.98 μ mol). Analytical HPLC Rt= 28.12 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μ m, Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 23.83 min (GLSciences Inc, InertSustain C18, 5 μ m, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95) \rightarrow (30:70), 5-30 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1007.5172 ([M+H]⁺, calc. 1007.5162), 504.2614 ([M+2H]²⁺, calc. 504.2620).



Chapter 6

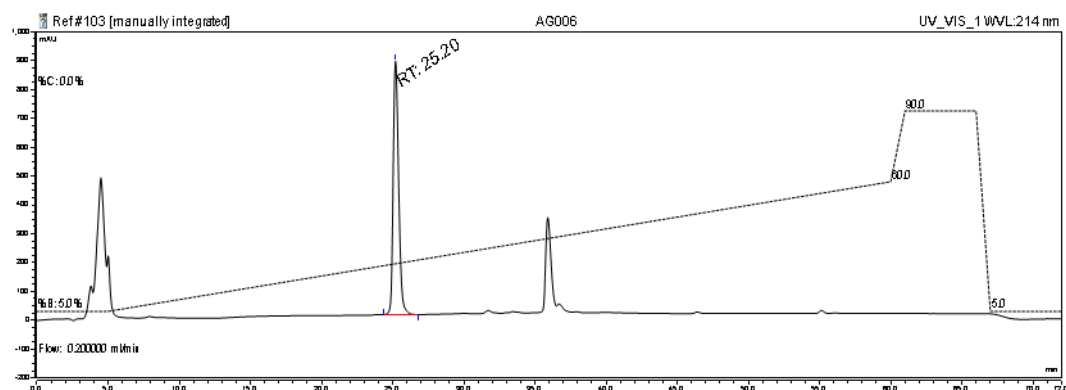
Peptide AG005, Sequence: NH₂-FNNFTVSFWLRVPKVSASHLE

Yield: 53% (16.31 mg, 6.58 μmol). Analytical HPLC Rt= 46.92 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 41.75 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95)→(45:55), 5-45 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1240.1531 ([M+2H]²⁺, calc. 1240.1523), 827.1040 ([M+3H]³⁺, calc. 827.1041), 620.5796 ([M+4H]⁴⁺, calc. 620.5800).



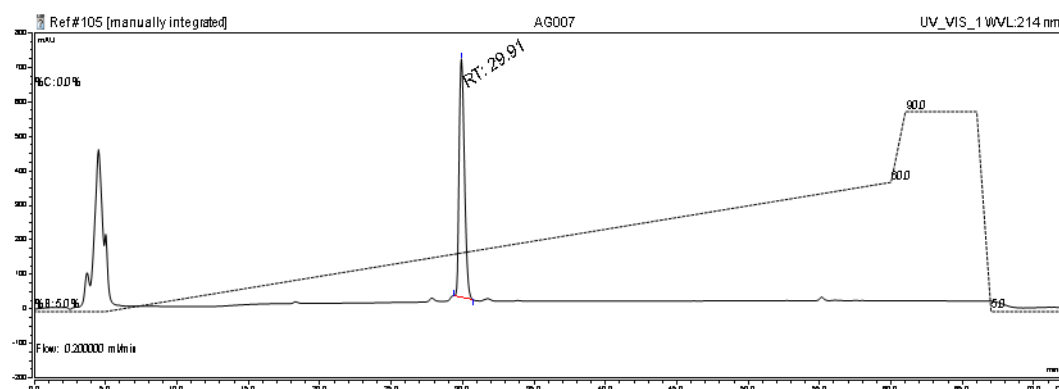
Peptide AG006, Sequence: NH₂-AAGPTPSAAGPPVAS

Yield: 72% (11.18 mg, 8.94 μmol). Analytical HPLC Rt= 25.20 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 21.37 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95)→(30:70), 5-30 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1250.6395 ([M+H]⁺, calc. 1250.6381), 625.8229 ([M+2H]²⁺, calc. 625.8230).

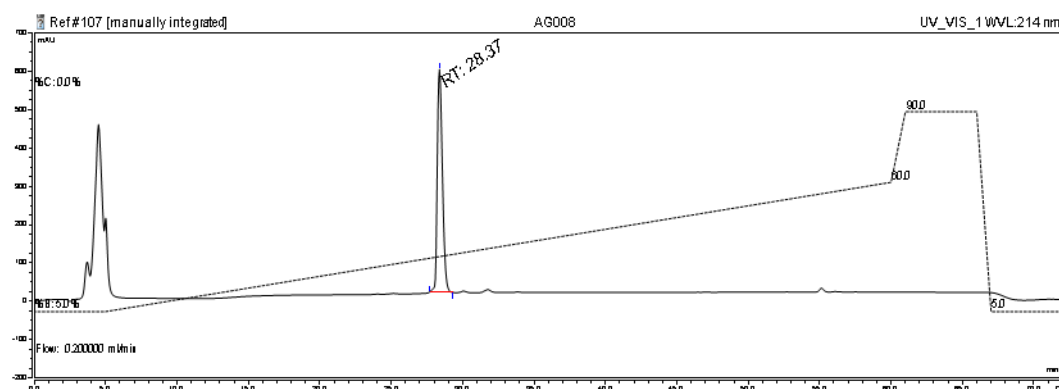


Peptide AG007, Sequence: NH₂-INTTADEKDPTNPFR

Yield: 84% (18.13 mg, 10.55 μmol). Analytical HPLC Rt= 29.91 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 25.36 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95)→(37:63), 5-37min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 859.9218 ([M+2H]²⁺, calc. 859.9214), 573.6165 ([M+3H]³⁺, calc. 573.6168).

Peptide AG008, Sequence: NH₂-Spacer-VVNSTTGPEHLR

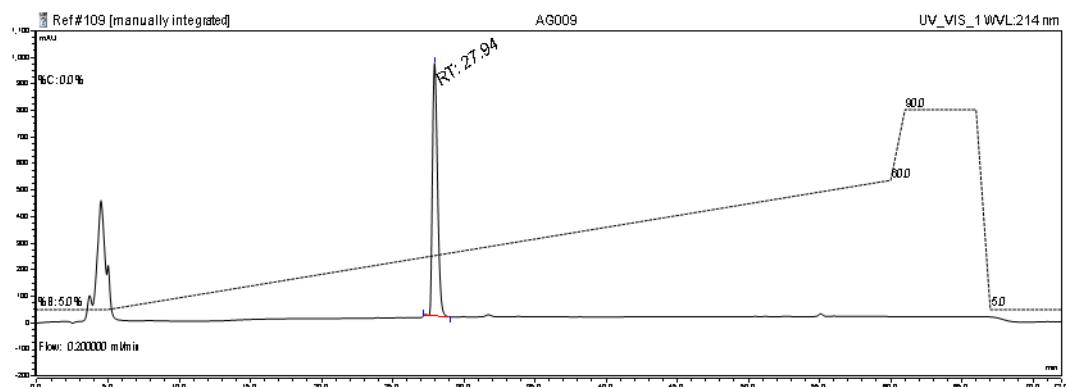
Yield: 91% (17.83 mg, 11.36 μmol). Analytical HPLC Rt= 28.73 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 23.77 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95)→(37:63), 5-37min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 785.4154 ([M+2H]²⁺, calc. 785.4157), 523.9460 ([M+3H]³⁺, calc. 523.9464).



Chapter 6

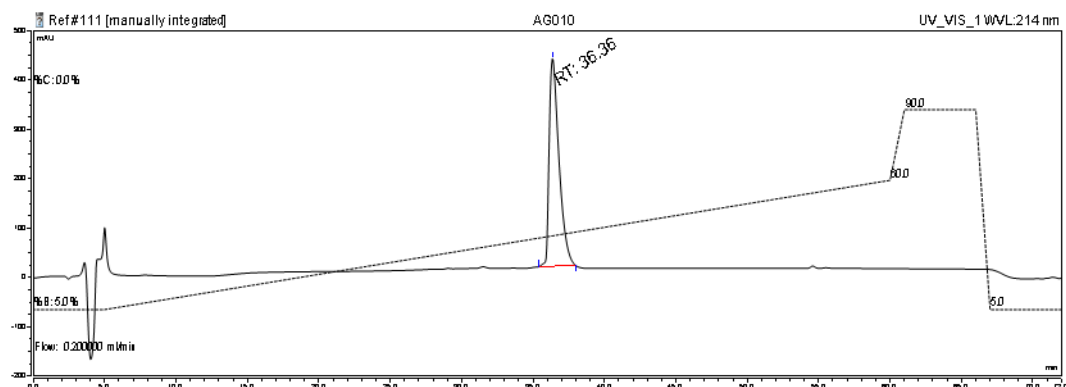
Peptide AG009, Sequence: NH₂-Spacer-WVSNKTEGR

Yield: 91% (14.5 mg, 11.33 μmol). Analytical HPLC Rt= 27.94 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 23.53 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95)→(37:63), 5-37min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 640.3358 ([M+2H]²⁺, calc. 640.3358), 427.2262 ([M+3H]³⁺, calc. 427.2264).



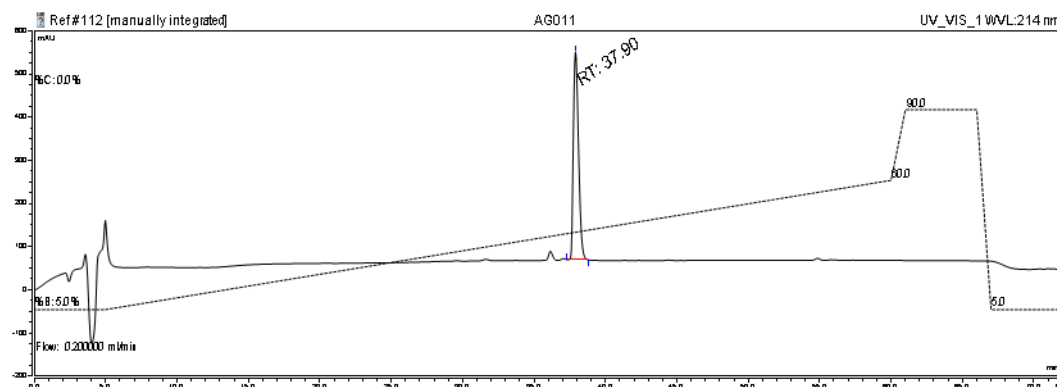
Peptide AG010, Sequence: NH₂-Spacer-NLTALPPDLPK

Yield: 41% (7.12 mg, 5.15 μmol). Analytical HPLC Rt= 36.36 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 32.21 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95)→(37:63), 5-37min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1381.7952 ([M+H]⁺, calc. 1381.7942), 691.4011 ([M+2H]²⁺, calc. 691.4010).



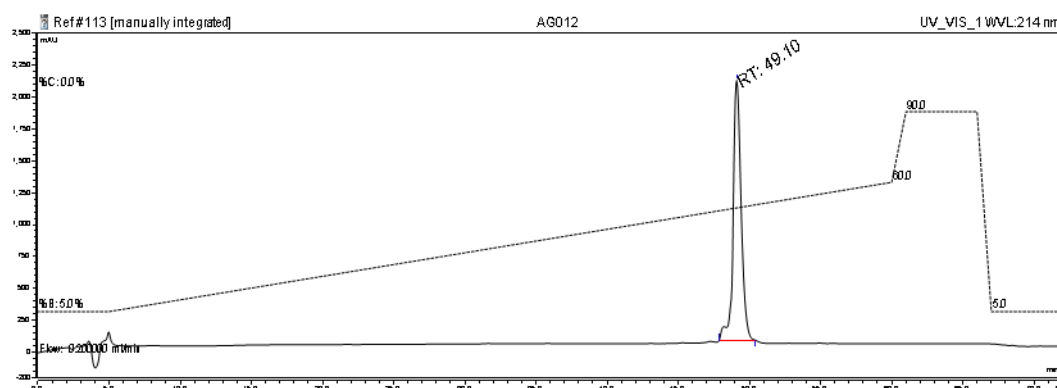
Peptide AG011 Sequence: NH₂-Spacer-LQNLTLPTNASIK

Yield: 41% (8.34 mg, 5.16 μ mol). Analytical HPLC Rt= 37.90 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μ m, Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 33.77 min (GLSciences Inc, InertSustain C18, 5 μ m, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95) \rightarrow (37:63), 5-37min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1615.9275 ([M+H]⁺, calc. 1615.9270), 808.4675 ([M+2H]²⁺, calc. 808.4674).



Peptide AG012 Sequence: Biotin-FNNFTVSFWLRVPKVSASHLE-Spacer-HGVTSAPDTRPAPGSTAPPA

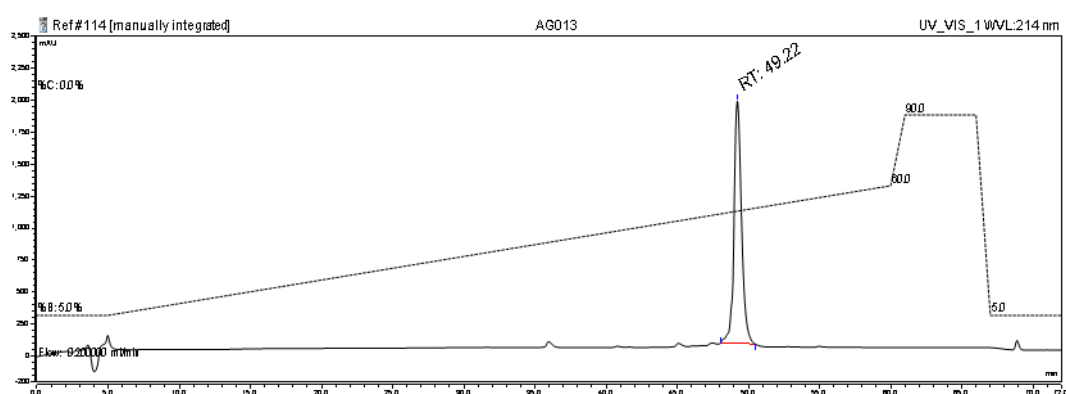
Yield: 10% (6.21 mg, 1.30 μ mol). Analytical HPLC Rt= 49.10 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μ m, Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 29.45 min (GLSciences Inc, InertSustain C18, 5 μ m, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80) \rightarrow (50:50), 5-35min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1195.1089 ([M+4H]⁴⁺, calc. 1195.1086), 956.2894 ([M+5H]⁵⁺, calc. 956.2885).



Chapter 6

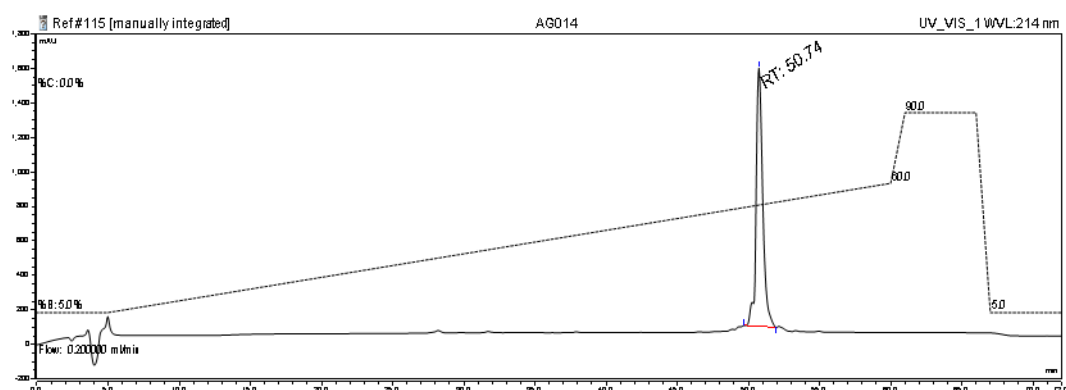
Peptide AG013 Sequence: Biotin-FNNFTVSFWLRVPKVSASHLE-Spacer-PAHGVTSAPDTRPAGSTA

Yield: 4% (2.15 mg, 0.46 μmol). Analytical HPLC R_t = 49.22 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC R_t = 29.20 min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80) \rightarrow (50:50), 5-35min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1170.8459 ($[\text{M}+4\text{H}]^{4+}$, calc. 1170.8454), 936.8781 ($[\text{M}+5\text{H}]^{5+}$, calc. 936.8779), 780.8991 ($[\text{M}+6\text{H}]^{6+}$, calc. 780.8996).



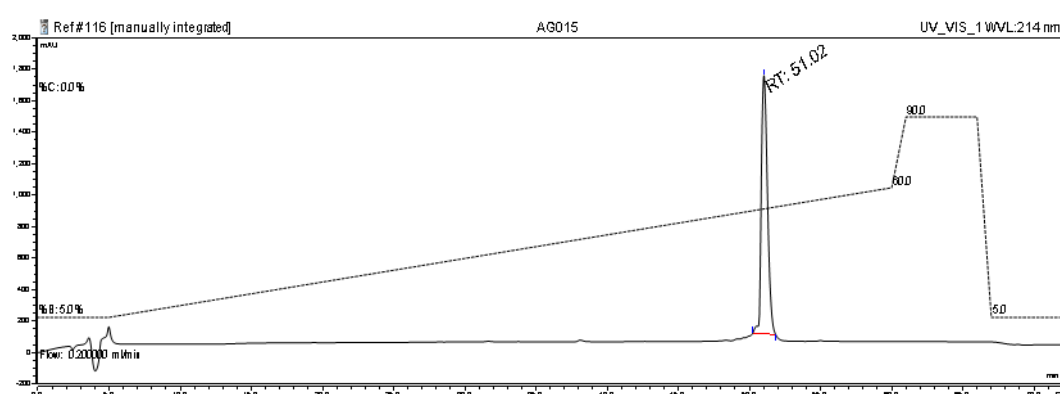
Peptide AG014, Sequence: Biotin-FNNFTVSFWLRVPKVSASHLE-Spacer-ASTGHATPLPVTD

Yield: 41% (21.16 mg, 5.09 μmol). Analytical HPLC R_t = 50.74 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC R_t = 31.01 min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80) \rightarrow (50:50), 5-35min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1386.3753 ($[\text{M}+3\text{H}]^{3+}$, calc. 1386.3745), 1040.0337 ($[\text{M}+4\text{H}]^{4+}$, calc. 1040.0329), 832.2273 ($[\text{M}+5\text{H}]^{5+}$, calc. 832.2278).



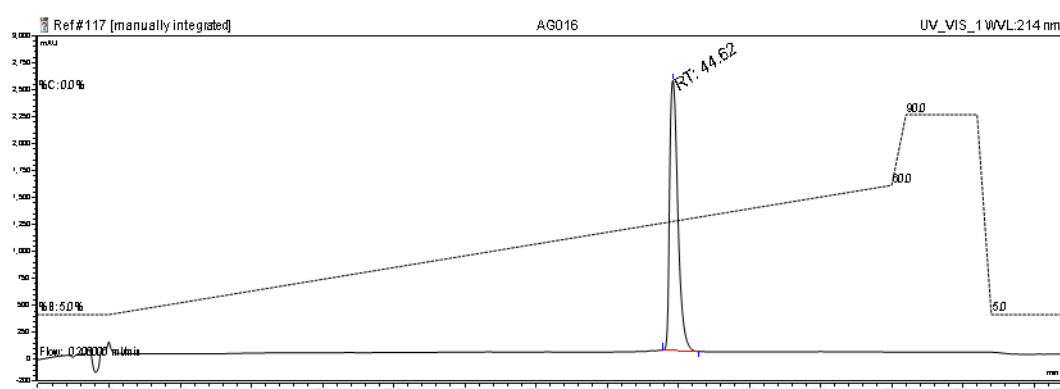
Peptide AG015, Sequence: Biotin-FNNFTVSFWLRVPKVSASHLE-Spacer-GHATPLPVTD

Yield: 44% (21.33 mg, 5.47 μmol). Analytical HPLC R_t = 51.02 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC R_t = 31.22 min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80) \rightarrow (50:50), 5-35min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1300.0028 ($[\text{M}+3\text{H}]^{3+}$, calc. 1300.0023), 975.2542 ($[\text{M}+4\text{H}]^{4+}$, calc. 975.2537), 780.4041 ($[\text{M}+5\text{H}]^{5+}$, calc. 780.4045).



Peptide AG016 Sequence: NH_2 -FNNFTVSFWLRVPKVSASHLE-Spacer-HGVTSAPDTRPAPGSTAPPA

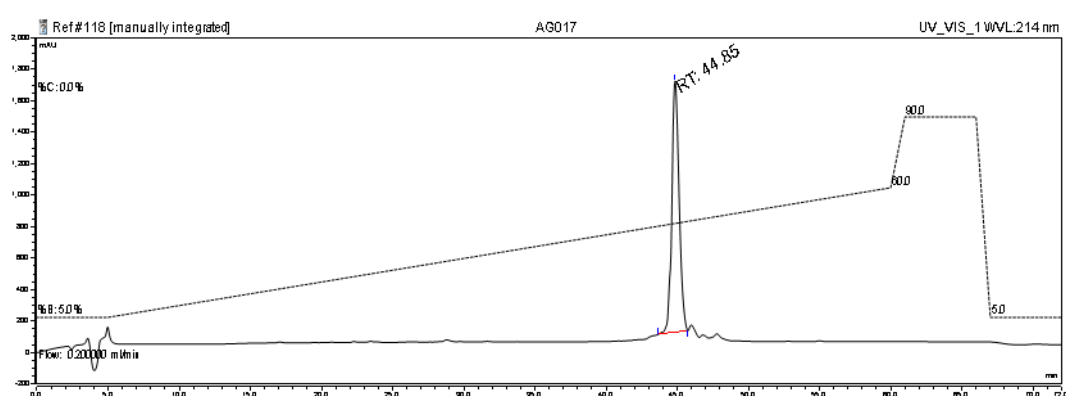
Yield: 37% (20.95 mg, 4.61 μmol). Analytical HPLC R_t = 44.62 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC R_t = 24.89 min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80) \rightarrow (45:55), 5-30min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1517.7835 ($[\text{M}+3\text{H}]^{3+}$, calc. 1517.7830), 1138.5894 ($[\text{M}+4\text{H}]^{4+}$, calc. 1138.5892), 911.0730 ($[\text{M}+5\text{H}]^{5+}$, calc. 911.0729), 759.3951 ($[\text{M}+6\text{H}]^{6+}$, calc. 759.3954).



Chapter 6

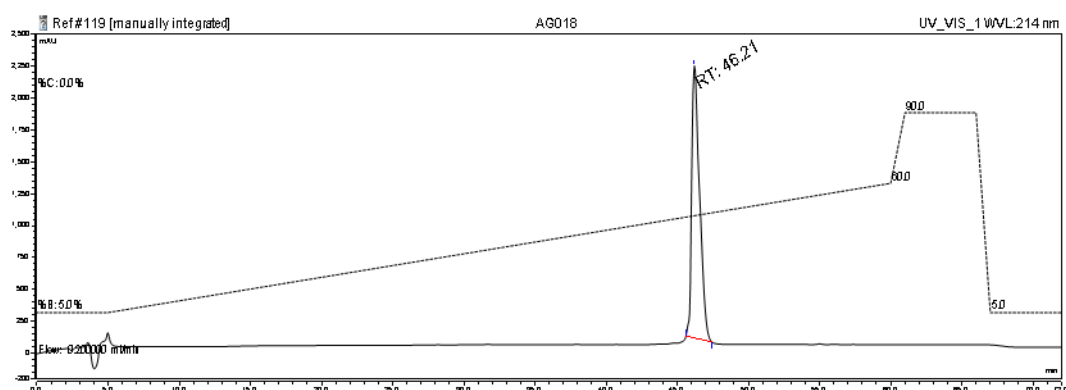
Peptide AG017 Sequence: NH₂-FNNFTVSFWLRVPKVSASHLE-Spacer-PAHGVTSAPDTRPAGSTA

Yield: 10% (5.51 mg, 1.24 μmol). Analytical HPLC Rt= 44.85 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 24.93 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80)→(45:55), 5-30min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1485.4328 ([M+3H]³⁺, calc. 1485.4321), 1114.3263 ([M+4H]⁴⁺, calc. 1114.3260), 891.6624 ([M+5H]⁵⁺, calc. 891.6624), 743.2195 ([M+6H]⁶⁺, calc. 743.2200).



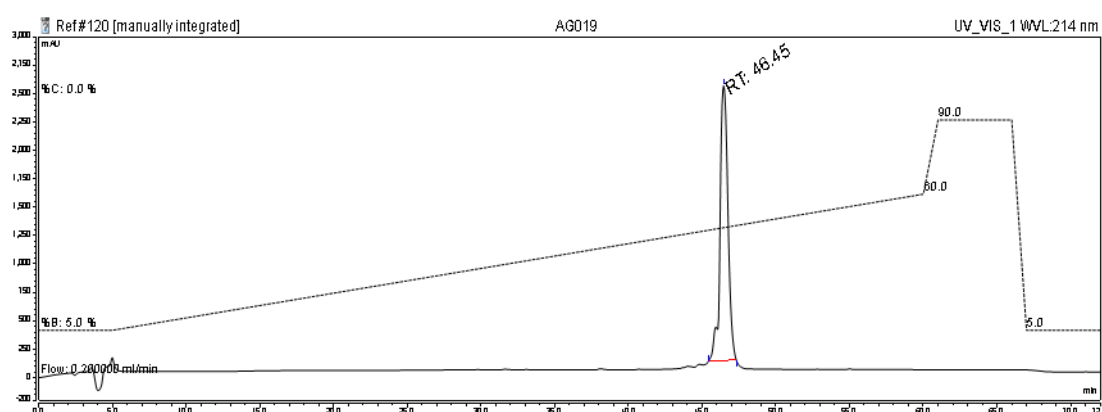
Peptide AG018 Sequence: NH₂-FNNFTVSFWLRVPKVSASHLE-Spacer-ASTGHATPLPVTD

Yield: 48% (23.8 mg, 6.06 μmol). Analytical HPLC Rt= 46.21 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 26.03 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80)→(45:55), 5-30min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1311.0159 ([M+3H]³⁺, calc. 1311.0153), 983.5140 ([M+4H]⁴⁺, calc. 983.5135), 787.0119 ([M+5H]⁵⁺, calc. 787.0123).



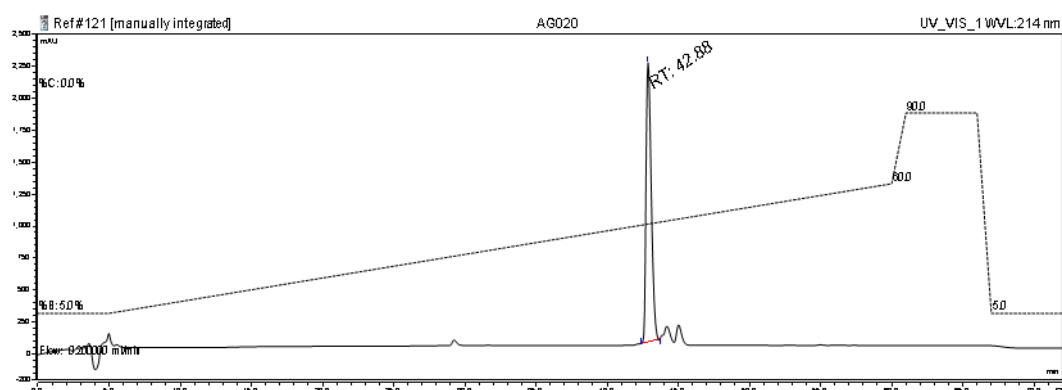
Peptide AG019, Sequence: NH₂-FNNFTVSFWLRVPKVSASHLE-Spacer-GHATPLPVTD

Yield: 41% (18.61 mg, 5.07 μmol). Analytical HPLC Rt= 46.45 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 26.37 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80)→(60:40), 5-45min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XL™ hybride FT), *m/z*: 1224.6431 ([M+3H]³⁺, calc. 1224.6431), 918.7342 ([M+4H]⁴⁺, calc. 918.7343), 735.1883 ([M+5H]⁵⁺, calc. 735.1890).



Peptide AG020, Sequence: NH₂-MGL-GHATPLPVTD

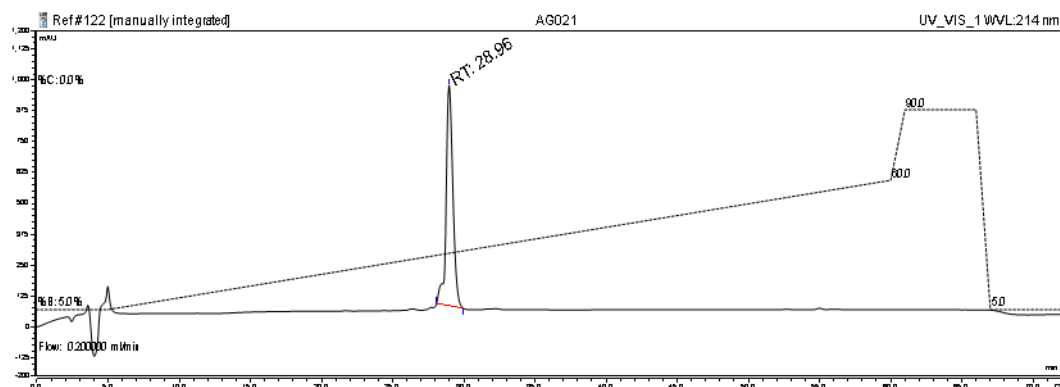
Yield: 21% (6.95 mg, 2.61 μmol). Analytical HPLC Rt= 42.88 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 33.95 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (10:90)→(55:45), 5-50 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XL™ hybride FT), *m/z*: 1333.0601 ([M+2H]²⁺, calc. 1333.0604), 889.0425 ([M+3H]³⁺, calc. 889.0429).



Chapter 6

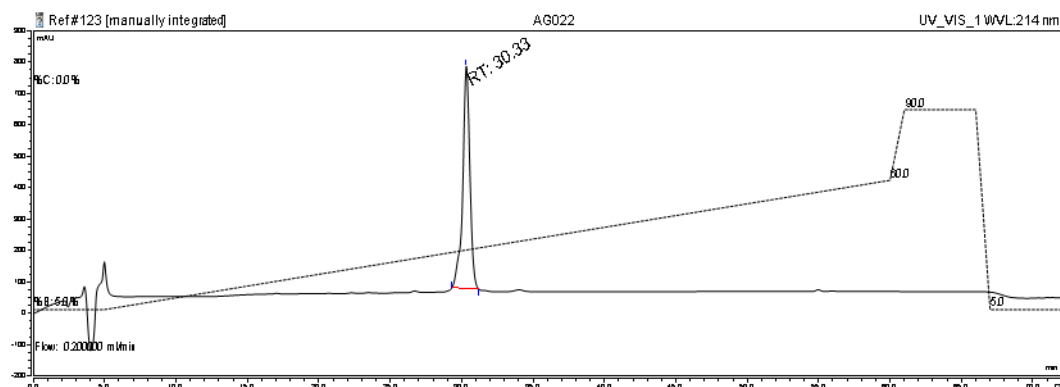
Peptide AG021 Sequence: Biotin-HGVTSAPDTRPAPGSTAPPA

Yield: 50% (13.19 mg, 6.24 μmol). Analytical HPLC R_t = 28.96 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC R_t = 24.81 min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95) \rightarrow (45:55), 5-45 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1057.5126 ($[\text{M}+2\text{H}]^{2+}$, calc. 1057.5124), 705.3436 ($[\text{M}+3\text{H}]^{3+}$, calc. 705.3442).



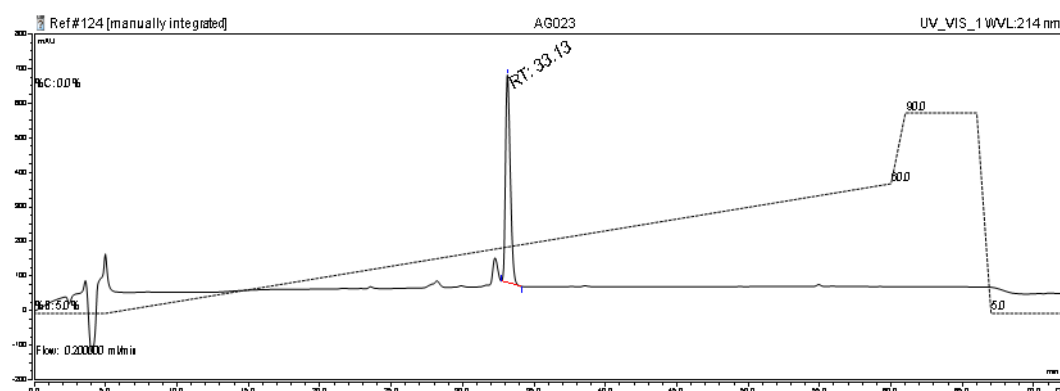
Peptide AG022, Sequence: Biotin-PAHGVTSAPDTRPAPGSTA

Yield: 54% (13.63 mg, 6.76 μmol). Analytical HPLC R_t = 30.33 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC R_t = 26.15 min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95) \rightarrow (30:70), 5-30 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1008.9862 ($[\text{M}+2\text{H}]^{2+}$, calc. 1008.9860), 672.9925 ($[\text{M}+3\text{H}]^{3+}$, calc. 672.9933).



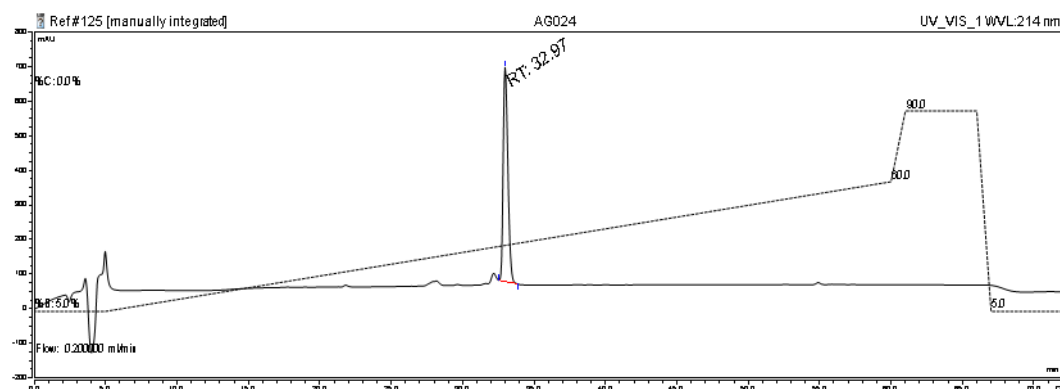
Peptide AG023, Sequence: Biotin-ASTGHATPLPVTD

Yield: 55% (10.31 mg, 6.91 μmol). Analytical HPLC R_t = 33.13 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC R_t = 29.69 min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95) \rightarrow (35:65), 5-35 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1492.7115 ($[\text{M}+\text{H}]^+$, calc. 1492.7106), 746.8590 ($[\text{M}+2\text{H}]^{2+}$, calc. 746.8592).



Peptide AG024, Sequence: Biotin-GHATPLPVTD

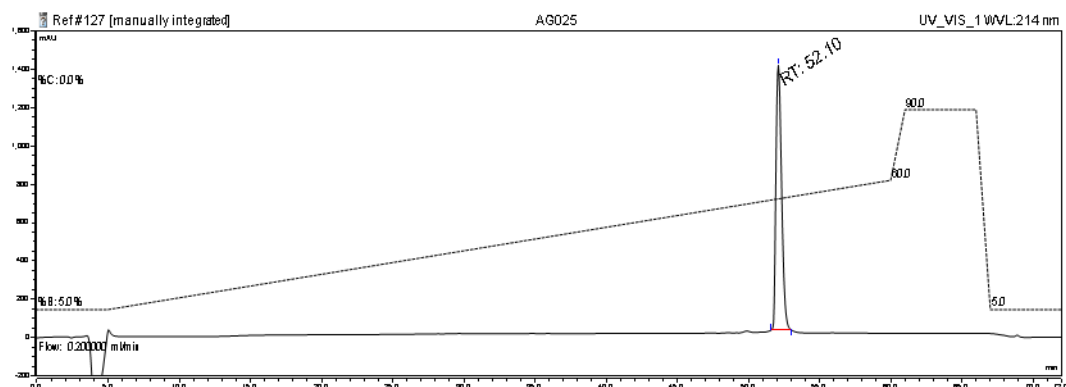
Yield: 70% (10.72 mg, 8.69 μmol). Analytical HPLC R_t = 32.97 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC R_t = 29.57 min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95) \rightarrow (35:65), 5-35 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1233.5945 ($[\text{M}+\text{H}]^+$, calc. 1233.5938), 617.3005 ($[\text{M}+2\text{H}]^{2+}$, calc. 617.3008).



Chapter 6

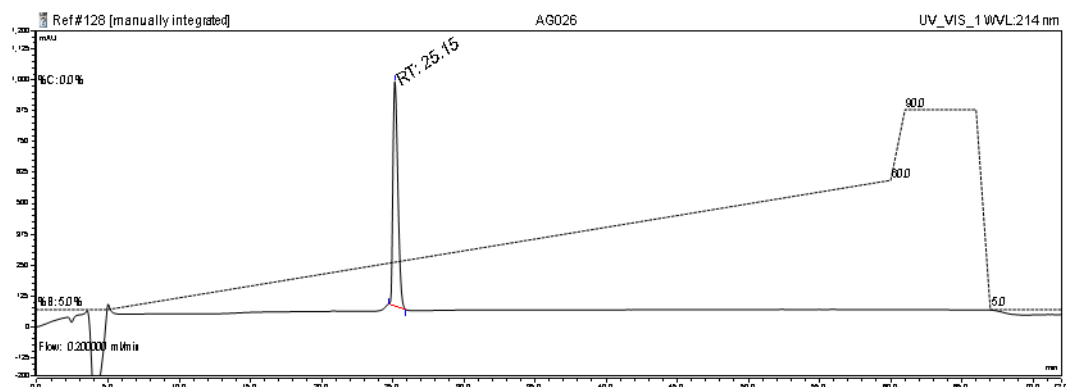
Peptide AG025, Sequence: Biotin-FNNFTVSFWLRVLPKVSASHLE

Yield: 39% (13.26 mg, 4.90 μmol). Analytical HPLC $R_t= 52.10$ min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC $R_t= 32.57$ min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80) \rightarrow (50:50), 5-35min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1353.1920 ($[\text{M}+2\text{H}]^{2+}$, calc. 1353.1911), 902.4638 ($[\text{M}+3\text{H}]^{3+}$, calc. 902.4633).



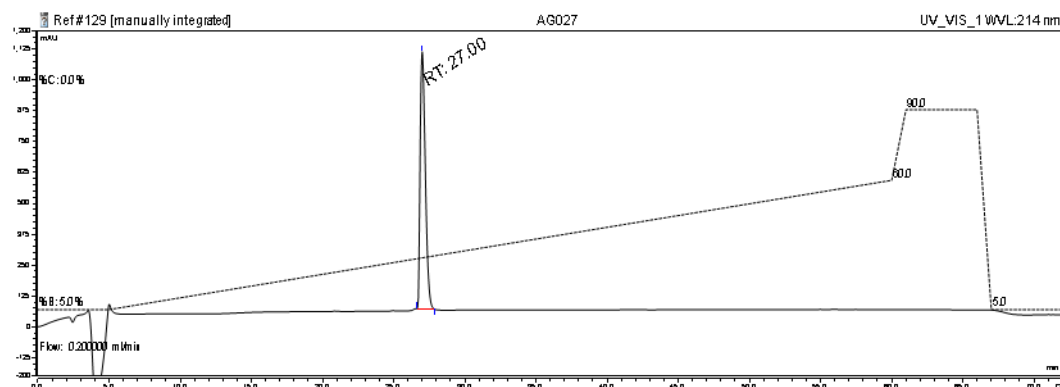
Peptide AG026, Sequence: AcNH-HGVTSAPDTRPAPGSTAPKA

Yield: 56% (13.73 mg, 7.01 μmol). Analytical HPLC $R_t= 25.15$ min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC $R_t= 20.83$ min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95) \rightarrow (30:70), 5-30 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 981.0006 ($[\text{M}+2\text{H}]^{2+}$, calc. 981.0000), 654.3353 ($[\text{M}+3\text{H}]^{3+}$, calc. 654.3359).



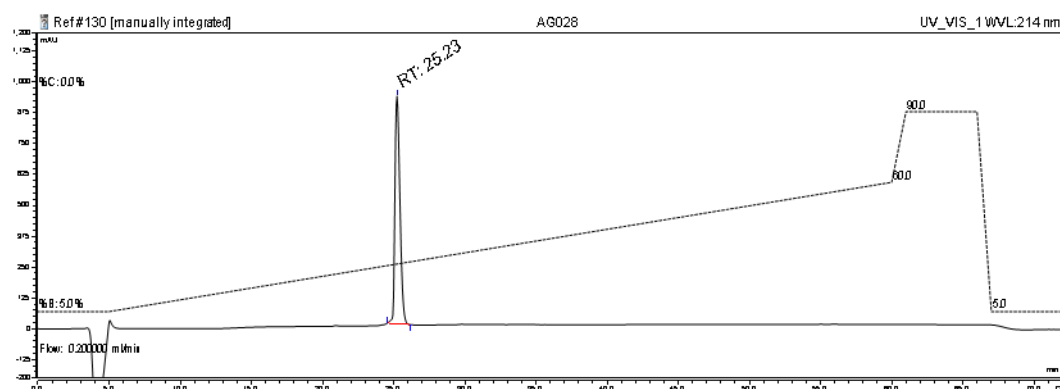
Peptide AG027, Sequence: AcNH-Spacer-HGVTSAPDTRPAPGSTAPKA

Yield: 30% (8.2 mg, 3.79 μmol). Analytical HPLC R_t = 27.00 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC R_t = 22.49 min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95) \rightarrow (30:70), 5-30 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1082.5595 ($[\text{M}+2\text{H}]^{2+}$, calc. 1082.5579), 722.0411 ($[\text{M}+3\text{H}]^{3+}$, calc. 722.4012).



Peptide AG028, Sequence: AcNH-HGVTSAPDTRPAPGSTAPKA

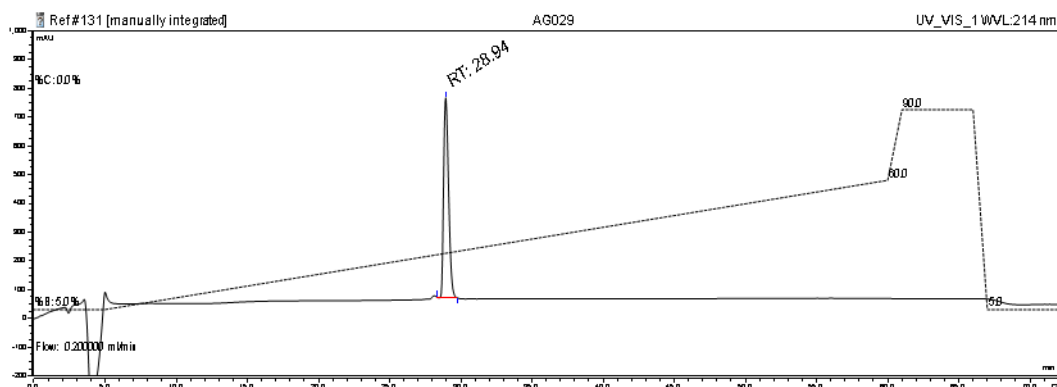
Yield: 43% (10.44 mg, 5.33 μmol). Analytical HPLC R_t = 25.23 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC R_t = 20.86 min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95) \rightarrow (30:70), 5-30 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 981.0014 ($[\text{M}+2\text{H}]^{2+}$, calc. 981.0000), 654.3358 ($[\text{M}+3\text{H}]^{3+}$, calc. 654.3359).



Chapter 6

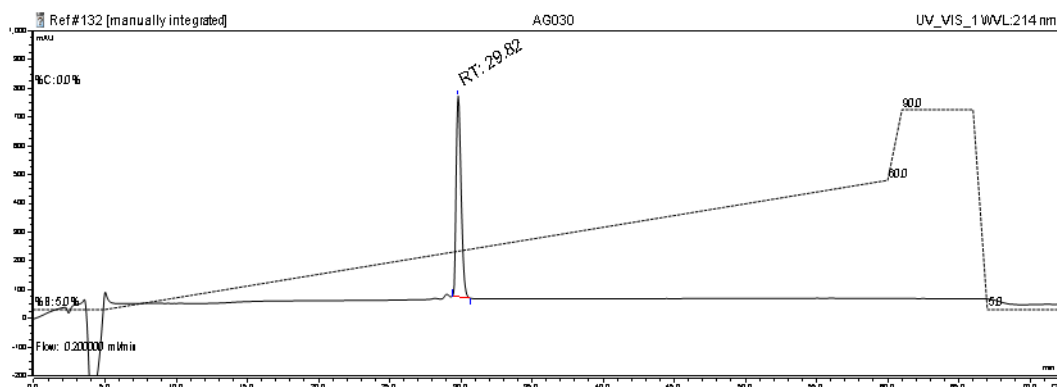
Peptide AG029, Sequence: AcNH-GHATPLPVTDKA

Yield: 80% (12.85 mg, 9.96 μmol). Analytical HPLC R_t = 28.94 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC R_t = 24.43 min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95) \rightarrow (30:70), 5-30 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1248.6602 ($[\text{M}+\text{H}]^+$, calc. 1248.6588), 624.8331 ($[\text{M}+2\text{H}]^{2+}$, calc. 624.8333).



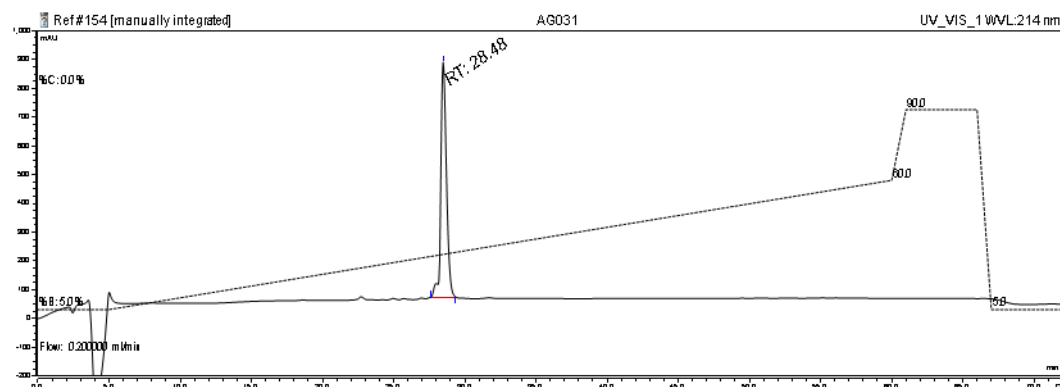
Peptide AG030, Sequence: AcNH-ASTGHATPLPVKA

Yield: 79% (12.74 mg, 9.87 μmol). Analytical HPLC R_t = 29.82 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC R_t = 25.24 min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (5:95) \rightarrow (30:70), 5-30 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1291.7027 ($[\text{M}+\text{H}]^+$, calc. 1291.7010), 646.3543 ($[\text{M}+2\text{H}]^{2+}$, calc. 646.3544).



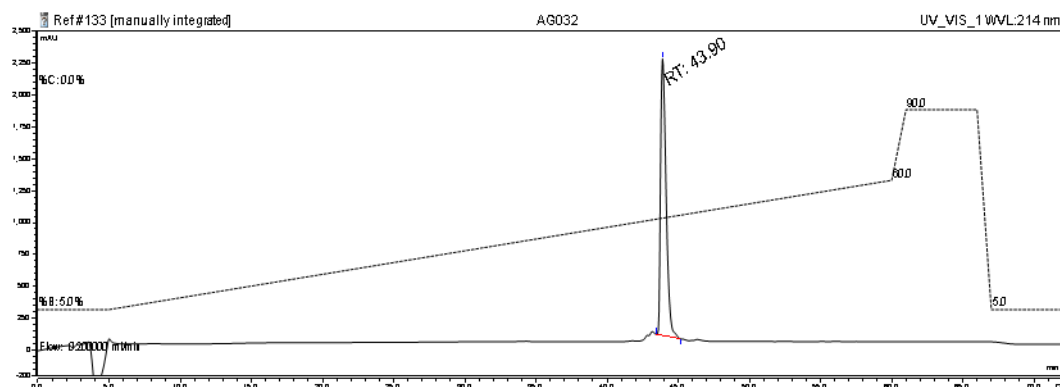
Peptide AG031, Sequence: NH₂-HGVT SAPDTRPAPGSTAPK(Biotin)A

Yield: 37% (10.01 mg, 4.67 μmol). Analytical HPLC Rt= 28.48 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 19.41 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (10:90)→(30:70), 5-25 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1073.0341 ([M+2H]²⁺, calc. 1073.0335), 715.6909 ([M+3H]³⁺, calc. 715.6916).



Peptide AG032, Sequence: NH₂-MGL-GHATPLPVTDK(Biotin)A

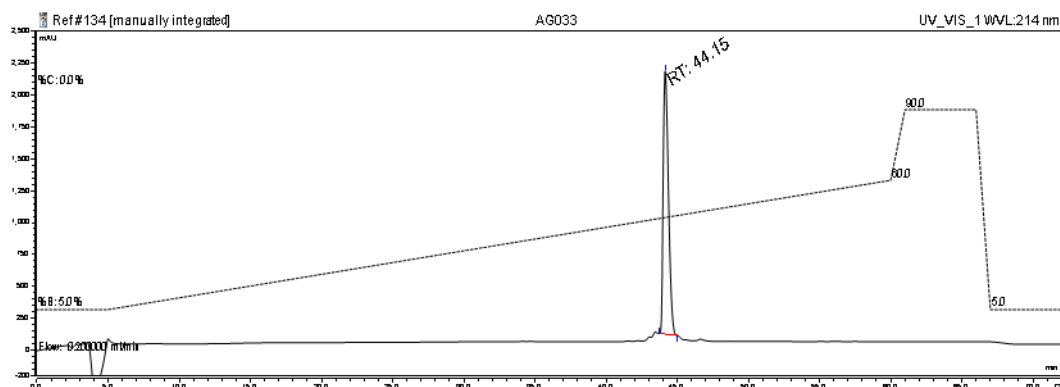
Yield: 44% (10.09 mg, 3.27 μmol). Analytical HPLC Rt= 43.90 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 34.74 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (10:90)→(55:45), 5-50 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1545.6665 ([M+2H]²⁺, calc. 1545.6653), 1030.7806 ([M+3H]³⁺, calc. 1030.7794).



Chapter 6

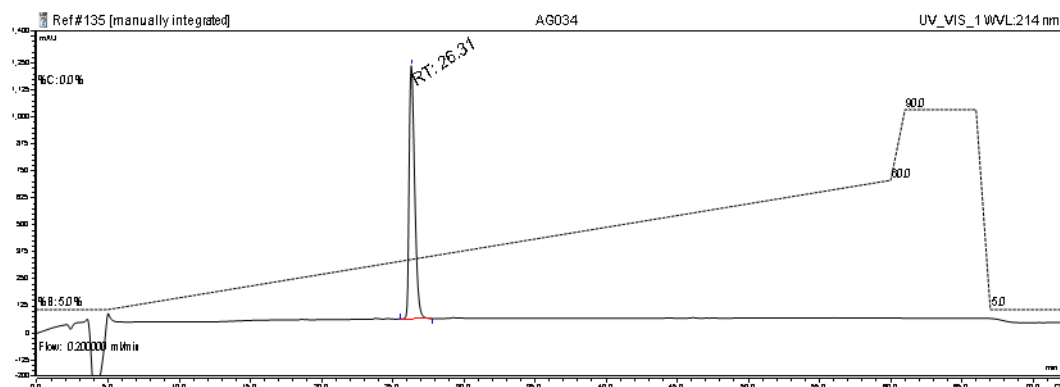
Peptide AG033, Sequence: NH₂-MGL-ASTGHATPLPVK(Biotin)A

Yield: 52% (12.33 mg, 3.94 μmol). Analytical HPLC Rt= 44.15 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 34.89 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (10:90)→(55:45), 5-50 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1567.1872 ([M+2H]²⁺, calc. 1567.1864), 1045.1275 ([M+3H]³⁺, calc. 1045.1268).



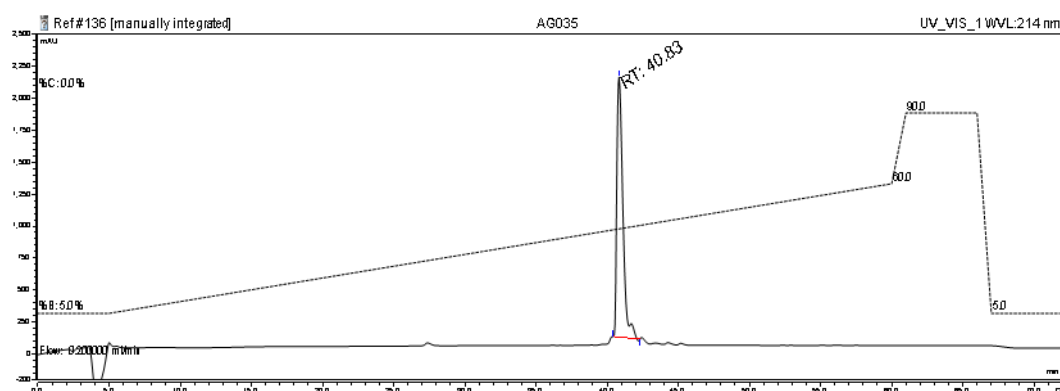
Peptide AG034, Sequence: NH₂-HGVTSAPDT*RPAPGS*T*APK(Biotin)A, *=GalNAc

Yield: 38% (13.1 mg, 4.76 μmol). Analytical HPLC Rt= 26.31 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 33.83 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (10:90)→(35:65), 5-105 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1377.6538 ([M+2H]²⁺, calc. 1377.6526), 918.7718 ([M+3H]³⁺, calc. 918.7710).



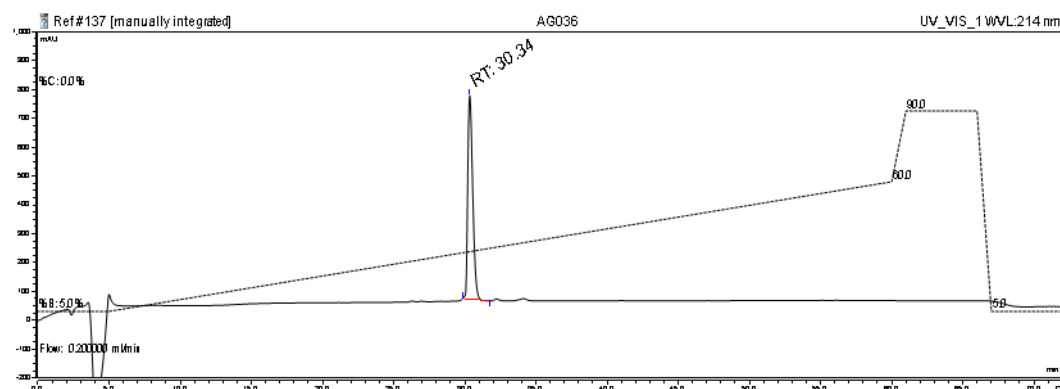
Peptide AG035, Sequence: NH₂-MGL-HGVTSAPDT*RPAPGS*T*APK(Biotin)A,
 *=GalNAc

Yield: 20% (6.62 mg, 1.50 μmol). Analytical HPLC Rt= 40.83 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 47.15 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80)→(55:45), 5-145min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XL™ hybride FT), *m/z*: 1471.3036 ([M+3H]³⁺, calc. 1471.3033), 1103.7301 ([M+4H]⁴⁺, calc. 1103.7294).



Peptide AG036, Sequence: NH₂-GHAT*PLPVT*DK(Biotin)A, *=GalNAc

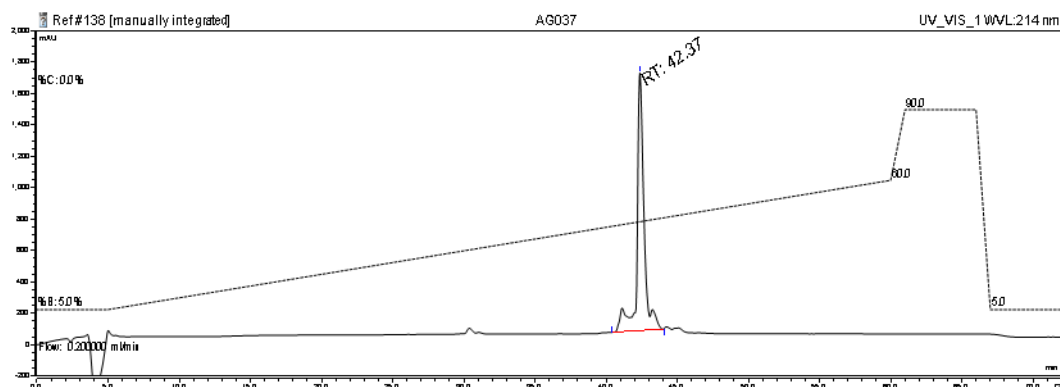
Yield: 76% (17.56 mg, 9.55 μmol). Analytical HPLC Rt= 30.34 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 46.44 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (10:90)→(35:65), 5-105 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XL™ hybride FT), *m/z*: 1838.8870 ([M+H]⁺, calc. 1838.8846), 919.9475 ([M+2H]²⁺, calc. 919.9462).



Chapter 6

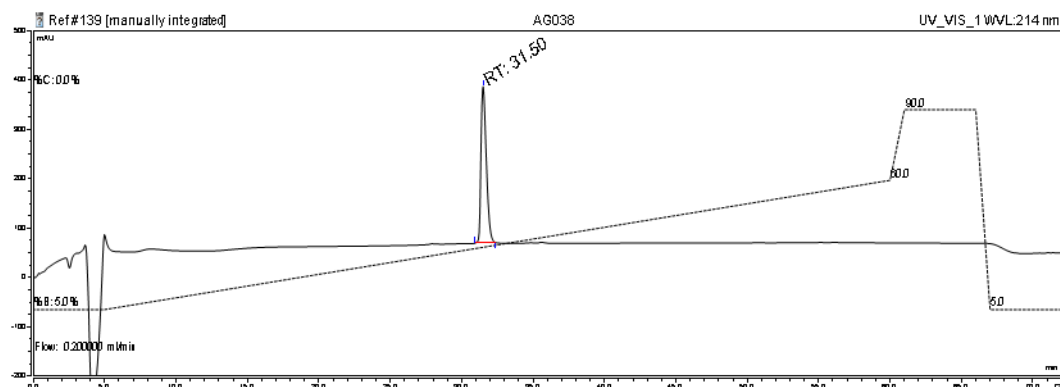
Peptide AG037, Sequence: NH₂-MGL-GHAT*PLPVT*DK(Biotin)A, *=GalNAc

Yield: 35% (9.1 mg, 2.60 μmol). Analytical HPLC Rt= 42.37 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 52.19 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (10:90)→(35:65), 5-105 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1749.2481 ([M+2H]²⁺, calc. 1749.2463), 1166.5014 ([M+3H]³⁺, calc. 1166.5001).



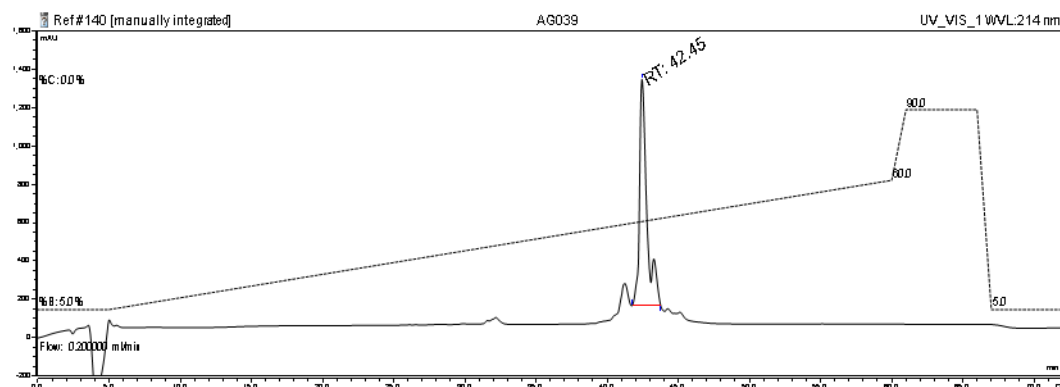
Peptide AG038, Sequence: NH₂-AS*T*GHAT*PLPVK(Biotin)A, *=GalNAc

Yield: 23% (5.94 mg, 2.85 μmol). Analytical HPLC Rt= 31.50 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 29.18 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (15:85)→(35:65), 5-85 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 695.6739 ([M+3H]³⁺, calc. 695.6739).



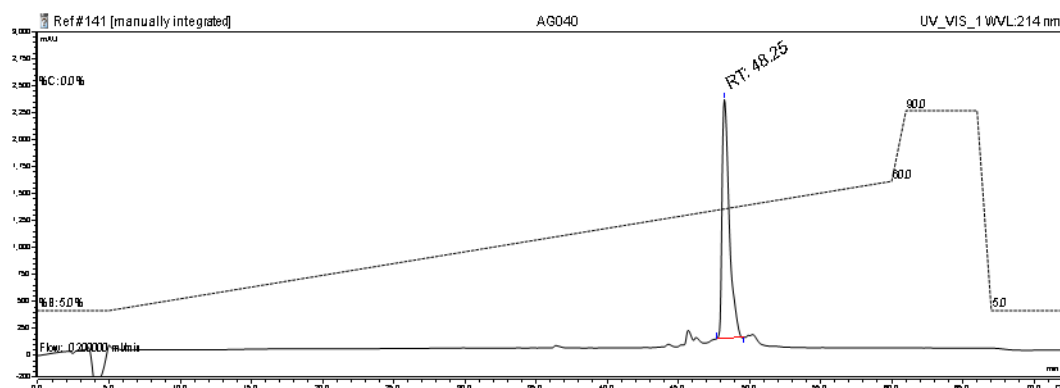
Peptide AG039, Sequence: NH₂-MGL-AS*T*GHAT*PLPVK(Biotin)A, *=GalNAc

Yield: 18% (4.94 mg, 1.32 μmol). Analytical HPLC Rt= 42.45 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 40.81 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80)→(50:50), 5-85min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XL™ hybride FT), *m/z*: 1871.8066 ([M+2H]²⁺, calc. 1871.8054).



Peptide AG040, Sequence: NH₂-MGL-FNNFTVSFWLRVPKVSASHLE-Spacer-GHAT*PLPVT*DK(Biotin)A, *=GalNAc

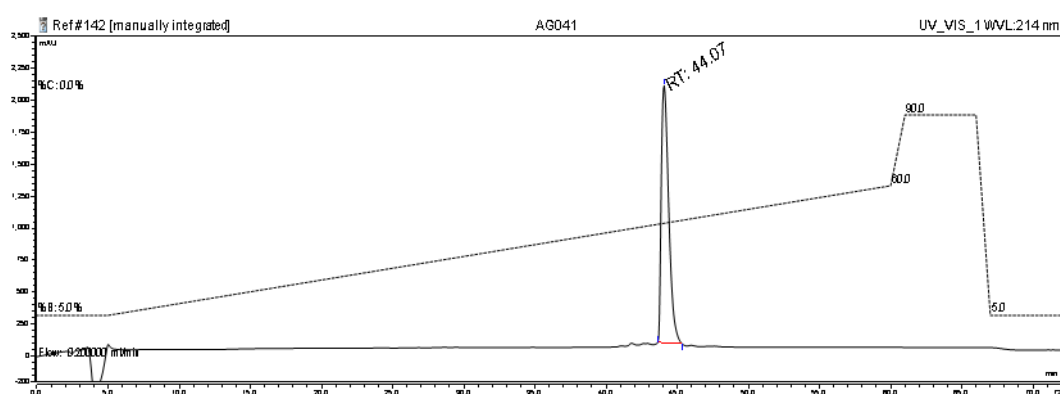
Yield: 5% (2.37 mg, 0.39 μmol). Analytical HPLC Rt= 48.25 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 76.37 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80)→(55:45), 5-145min, wavelength=214 nm);



Chapter 6

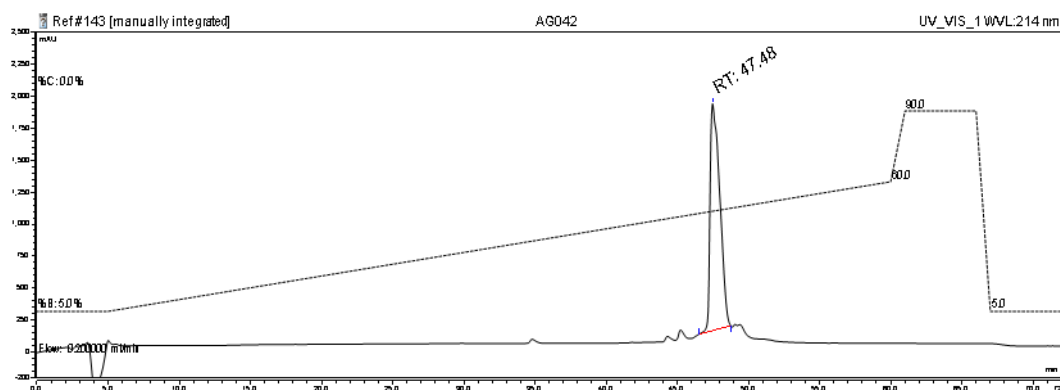
Peptide AG041, Sequence: NH₂-FNNFTVSFWLRVPKVSASHLE-Spacer-HGVTSAPDT*RPAPGS*T*APPA, *=GalNAc

Yield: 6% (3.59 mg, 0.70 μmol). Analytical HPLC Rt= 44.07 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 59.80 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80)→(50:50), 5-125min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XL™ hybride FT), *m/z*: 1290.9000 ([M+4H]⁴⁺, calc. 1290.8988), 1032.9216 ([M+5H]⁵⁺, calc. 1032.9206), 860.9352 ([M+6H]⁶⁺, calc. 860.9351).



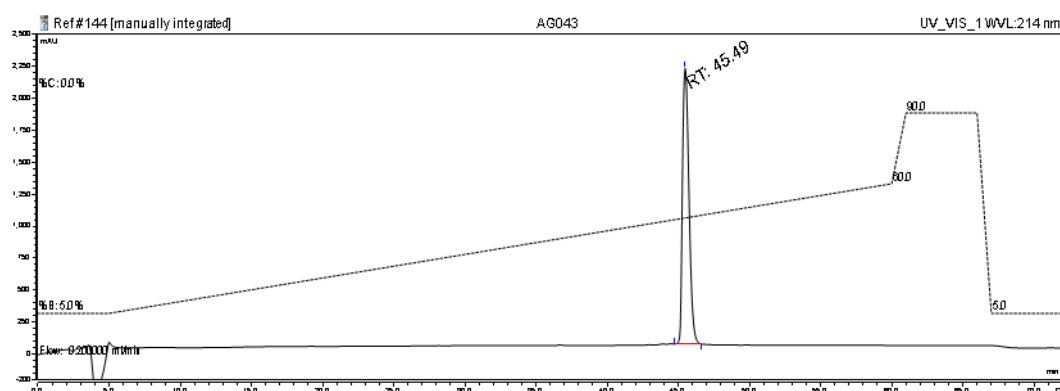
Peptide AG042, Sequence: NH₂-MGL-FNNFTVSFWLRVPKVSASHLE-Spacer-HGVTSAPDT*RPAPGS*T*APPA, *=GalNAc

Yield: 1% (1.01 mg, 0.15 μmol). Analytical HPLC Rt= 47.48 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 74.47 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80)→(55:45), 5-145min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XL™ hybride FT), *m/z*: 1364.4409 ([M+5H]⁵⁺, calc. 1364.4399), 1137.2017 ([M+6H]⁶⁺, calc. 1137.2013).



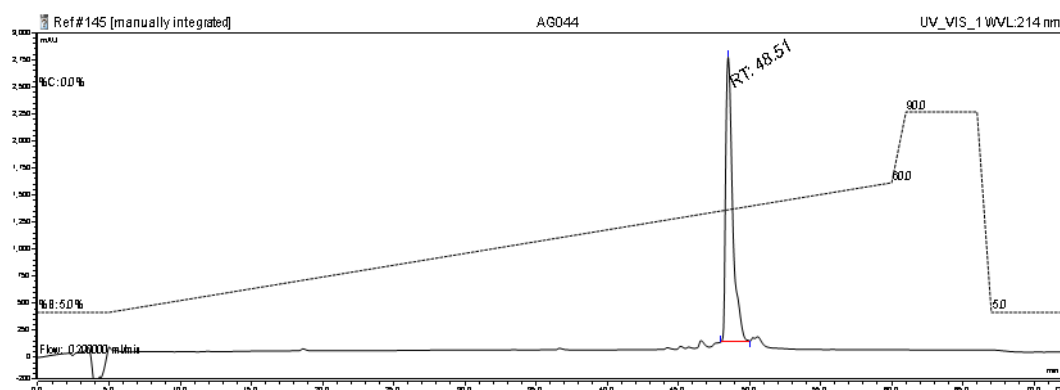
Peptide AG043, Sequence: NH₂-FNNFTVSFWLRVPKVSASHLE-Spacer-GHAT*PLPV
T*D, *=GalNAc

Yield: 12% (5.98 mg, 1.47 μmol). Analytical HPLC Rt= 45.49 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 48.56 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80)→(50:50), 5-85min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1020.2757 ([M+4H]⁴⁺, calc. 1020.2740), 816.4211 ([M+5H]⁵⁺, calc. 816.4207).



Peptide AG044, Sequence: NH₂-MGL-FNNFTVSFWLRVPKVSASHLE-Spacer-
GHAT*PLPVT*D, *=GalNAc

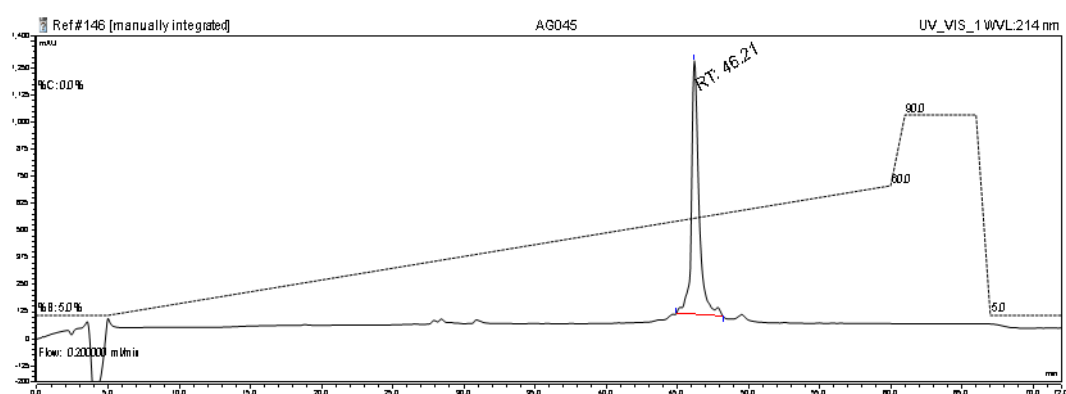
Yield: 4% (2.42 mg, 0.42 μmol). Analytical HPLC Rt= 48.51 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 77.21 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80)→(55:45), 5-145min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1912.5634 ([M+3H]³⁺, calc. 1912.5616), 1147.9405 ([M+5H]⁵⁺, calc. 1147.9401).



Chapter 6

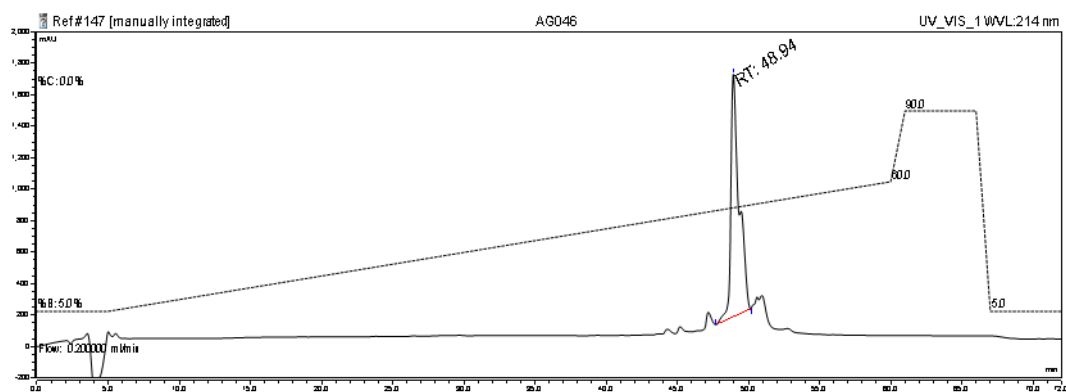
Peptide AG045, Sequence: $\text{NH}_2\text{-FNNFTVSFWLRVPKVSASHLE-Spacer-AS*T*GHAT*PLPV, *=GalNAc}$

Yield: 3% (1.8 mg, 0.42 μmol). Analytical HPLC $R_t = 46.21$ min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC $R_t = 64.73$ min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80) \rightarrow (55:45), 5-145min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1442.0713 ($[\text{M}+3\text{H}]^{3+}$, calc. 1442.0698), 1081.8056 ($[\text{M}+4\text{H}]^{4+}$, calc. 1081.8043), 856.6455 ($[\text{M}+5\text{H}]^{5+}$, calc. 865.6450).



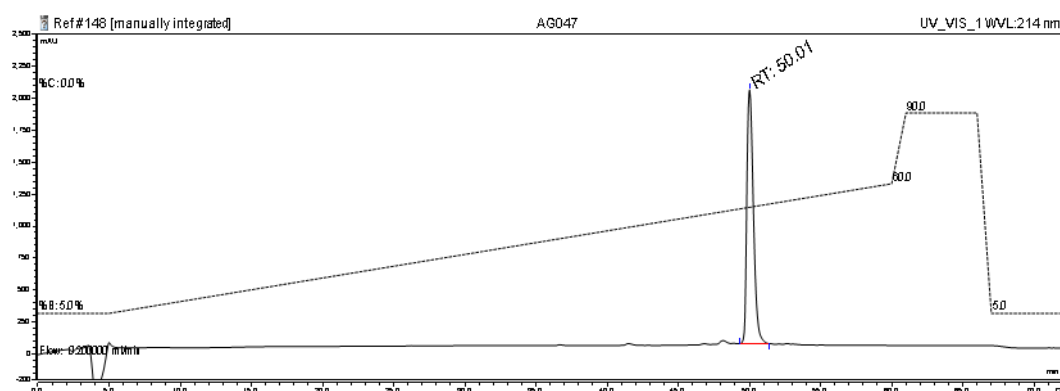
Peptide AG046, Sequence: $\text{NH}_2\text{-MGL-FNNFTVSFWLRVPKVSASHLE-Spacer-AS*T*GHAT*PLPV, *=GalNAc}$

Yield: 1% (0.71 mg, 0.12 μmol). Analytical HPLC $R_t = 48.94$ min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC $R_t = 79.32$ min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80) \rightarrow (55:45), 5-145min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1496.2040 ($[\text{M}+4\text{H}]^{4+}$, calc. 1496.2036), 1197.1649 ($[\text{M}+5\text{H}]^{5+}$, calc. 1197.1644).



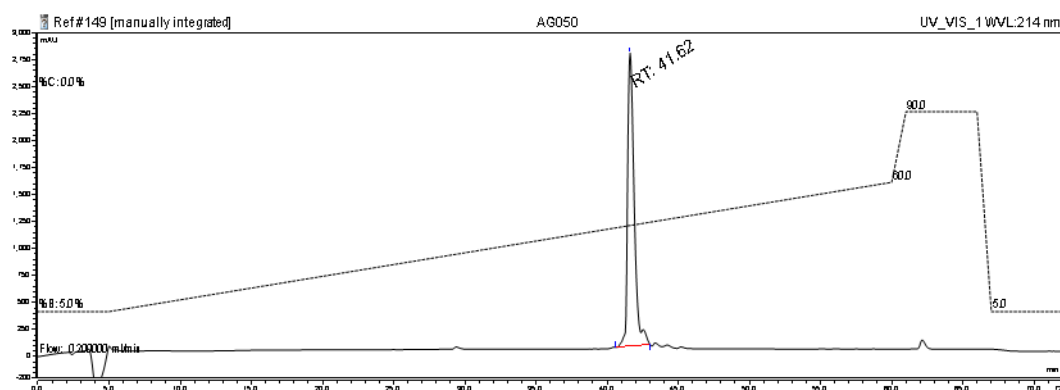
Peptide AG047, Sequence: Biotin-FNNFTVSFWLRVPKVSASHLE-Spacer-GHAT*PLPVT*D, *=GalNAc

Yield: 18% (9.64 mg, 2.24 μmol). Analytical HPLC R_t = 50.01 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC R_t = 60.05 min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (25:75) \rightarrow (55:45), 5-125min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1435.3903 ($[\text{M}+3\text{H}]^{3+}$, calc. 1435.3885), 1076.7944 ($[\text{M}+4\text{H}]^{4+}$, calc. 1076.7934).



Peptide AG050, Sequence: NH_2 -MGL-HGVTSAPDTRPAPGSTAPK(Biotin)A

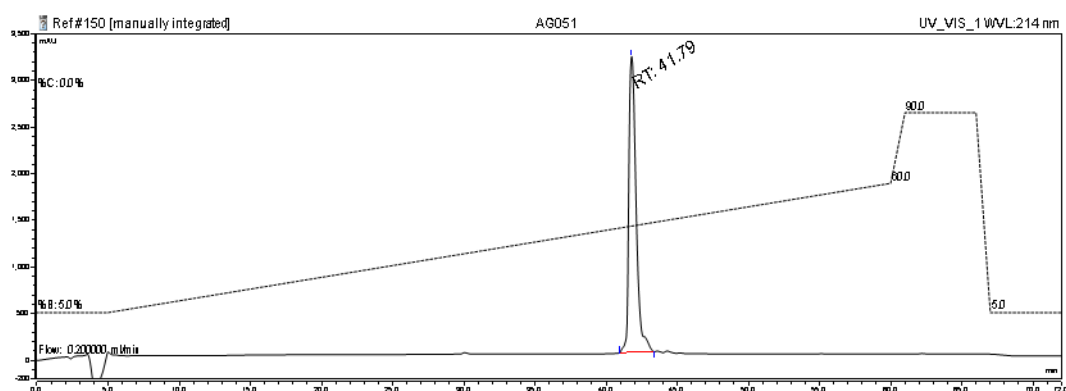
Yield: 29% (8.33 mg, 2.19 μmol). Analytical HPLC R_t = 41.62 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm , Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC R_t = 51.23 min (GLSciences Inc, InertSustain C18, 5 μm , 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (20:80) \rightarrow (55:45), 5-145min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), m/z : 1901.9331 ($[\text{M}+2\text{H}]^{2+}$, calc. 1901.8319), 1268.2244 ($[\text{M}+3\text{H}]^{3+}$, calc. 1268.2239).



Chapter 6

Peptide AG051, Sequence: NH₂-MGL-Spacer-HGVTSAPDTRPAPGSTAPK(Biotin)A

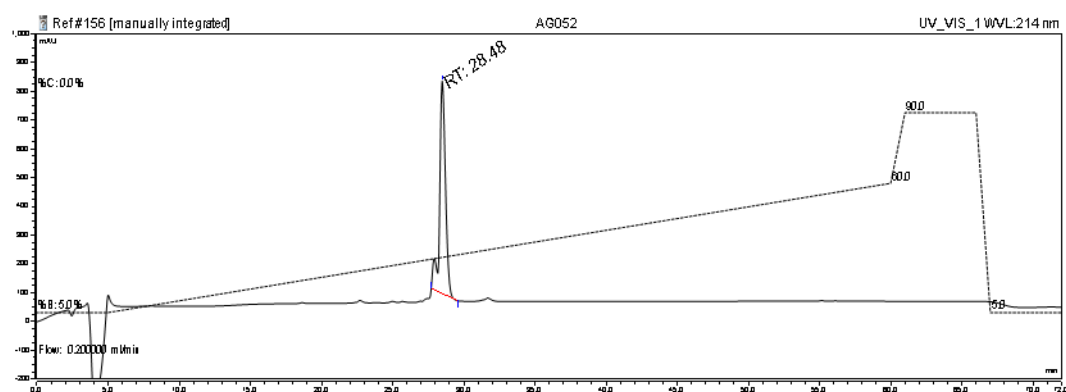
Yield: 31% (9.36 mg, 2.34 μmol). Analytical HPLC Rt= 41.79 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 31.63 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (25:75)→(60:44), 5-145min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1335.9294 ([M+3H]³⁺, calc. 1335.9291), 1002.1995 ([M+4H]⁴⁺, calc. 1002.1988).



Peptide AG052, Sequence: NH₂-HGVTSAPDTRPAPGSTAPK(Biotin)A

Yield: 48% (7.77 mg, 3.62 μmol). Analytical HPLC Rt= 28.48 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(60:40), 5-60 min, wavelength=214 nm). Semi-preparative HPLC Rt= 19.46 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (15:85)→(35:65), 5-85 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 695.6739 ([M+3H]³⁺, calc. 695.6739).

HR-ESI-MS (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1073.0344 ([M+2H]²⁺, calc. 1073.0335), 715.6912 ([M+3H]³⁺, calc. 715.6916).



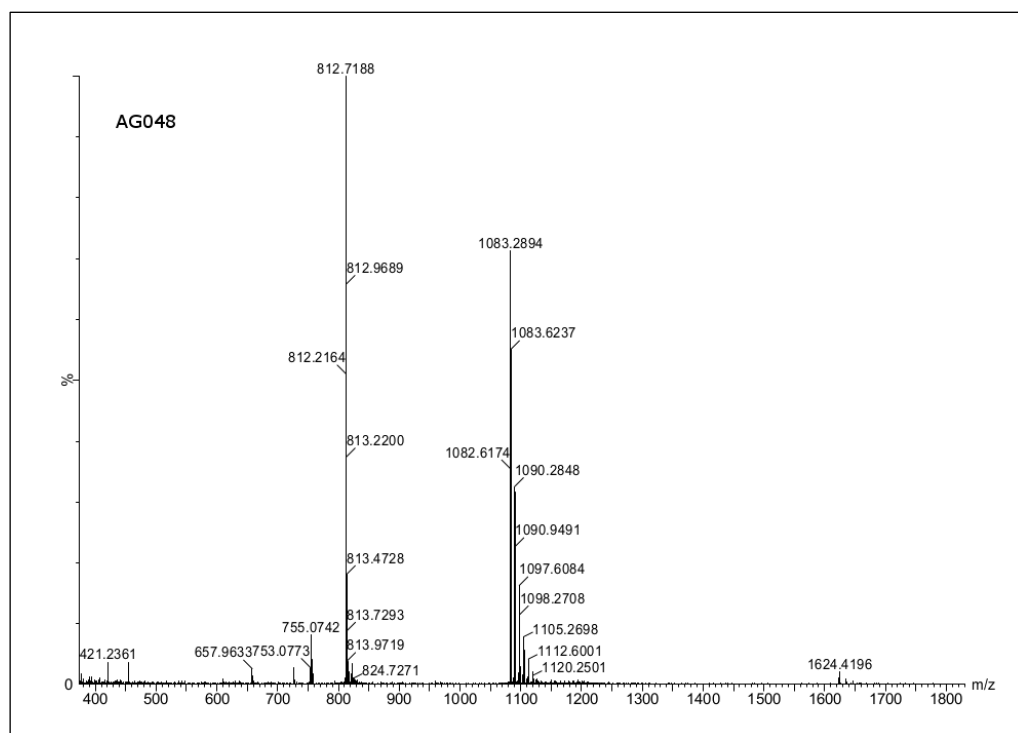
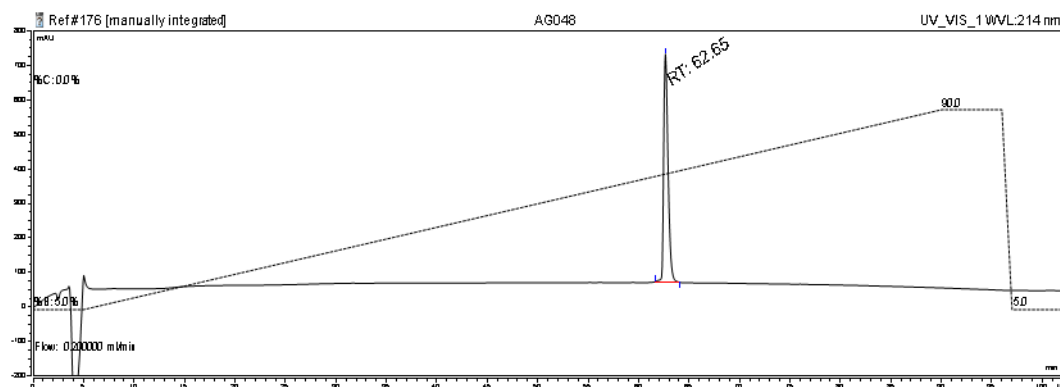
6.3.10. Coiled-coil peptide synthesis and characterisation

The peptides were synthesized by stepwise solid-phase peptide synthesis using the Fmoc strategy and starting with Rink amide MBHA resin low loading (100-200 mesh) resin (For peptides AG048 12.5 μmol scale/peptide and for peptides AG053 7.5 μmol scale/peptide). The coupling of the glycosylation site was performed using 2 eq of the protected MGL amino acid building block AG153, which was pre-activated manually using 1.8 eq HATU/HOAt and 3.6 eq of DIPEA, followed by Fmoc deprotection. The standard Fmoc amino acids were coupled automatically on a MultisynTech peptide synthesizer using 8 eq of the amino acid and the HBTU/HOBt reagents. Fmoc deprotection was done according to standard conditions, 20% piperidine in DMF. Peptides AG048 and AG053 were capped with an *N*-terminal acetyl group using a solution of 0.015 M HOBt, 0.5 M Ac₂O, 0.125 M DIPEA in NMP on *N*-terminus for 45 min. The obtained peptides were then released from the resin, and all acid sensitive side-chain protecting groups were simultaneously removed using TFA/TIPS/H₂O 15:0.9:0.9 in 2.5 h followed by solvent concentration, lyophilization and purification using a C-18 cartridge (1g of C-18 material, Waters). For saccharide deprotection, the *O*-acetyl groups were cleaved by transesterification in methanol using catalytic amounts of NaOMe at pH 9-9.5 for 16-24 h (deprotection followed by analytical HPLC) to yield peptides AG048 and AG053, which were purified by preparative HPLC (Phenomenex Luna C18 (2), 21.20x250mm).

Chapter 6

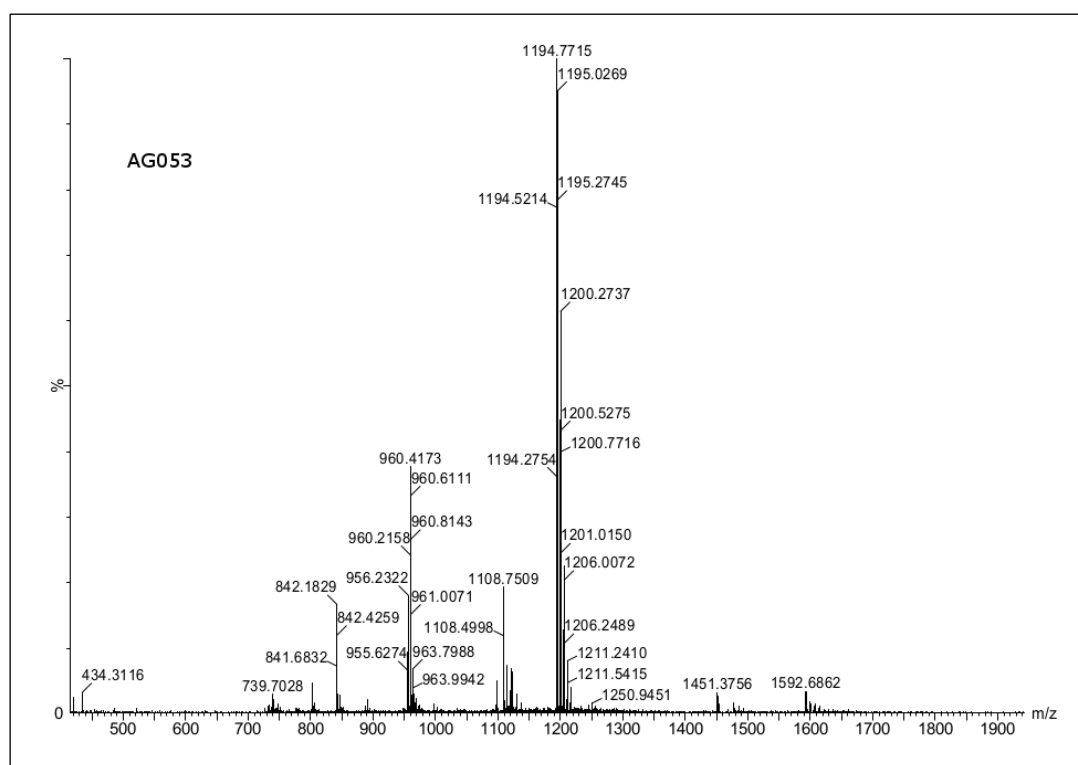
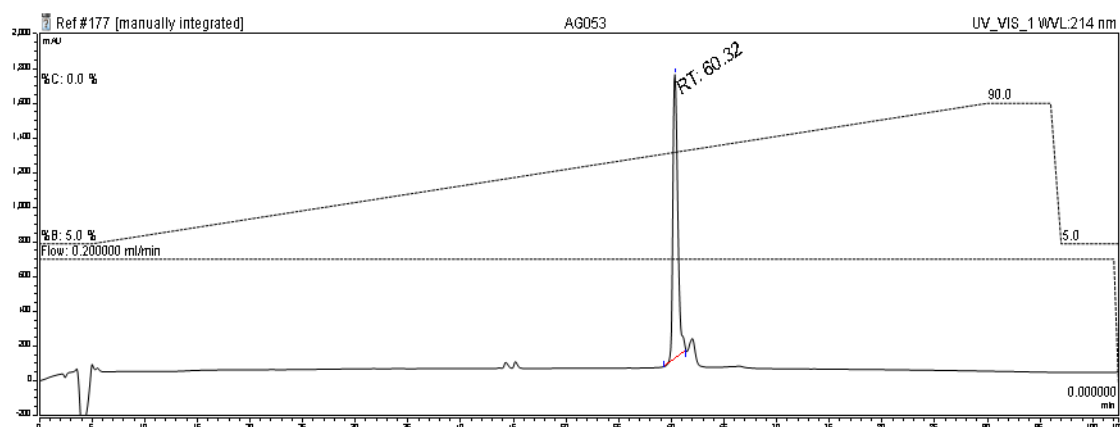
Peptide AG048 Sequence: AcNH-GIAAIEQKIAAIEQKIAAIEQKIAAIEWKG-NH₂

Yield: 54% (22.06 mg, 6.79 μmol). Analytical HPLC Rt= 62.65 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μm, Grad: eluent A/B + 0.1% TFA (5:95)→(90:10), 5-90 min, wavelength=214 nm). Semi-preparative HPLC Rt= 32.84 min (GLSciences Inc, InertSustain C18, 5 μm, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (10:90)→(90:10), 5-50 min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XL™ hybride FT), *m/z*: 1082.9607 ([M+3H]³⁺, calc. 1082.9605).



Peptide AG053, Sequence: AcNH-GIAAIEQKIAAIEWKIAAIEQKIAAIE-MGL-KG-NH₂

Yield: 18% (6.33 mg, 1.33 μ mol). Analytical HPLC Rt= 60.32 min (Phenomenex Luna C18 (2), 2.0 x 150 mm, 3 μ m, Grad: eluent A/B + 0.1% TFA (5:95) \rightarrow (90:10), 5-90 min, wavelength=214 nm). Semi-preparative HPLC Rt= 39.01 min (GLSciences Inc, InertSustain C18, 5 μ m, 6.0x250mm, Grad: eluent A/B + 0.1 % TFA (40:60) \rightarrow (60:44), 5-85min, wavelength=214 nm); *HR-ESI-MS* (Thermo LTQ Orbitrap XLTM hybride FT), *m/z*: 1194.8564 ([M+4H]⁴⁺, calc. 1194.8569), 956.0875 ([M+5H]⁵⁺, calc. 956.0871).



6.4. Protein: experimental procedure and characterisation

6.4.1. Plasmid preparation hMGL cloning

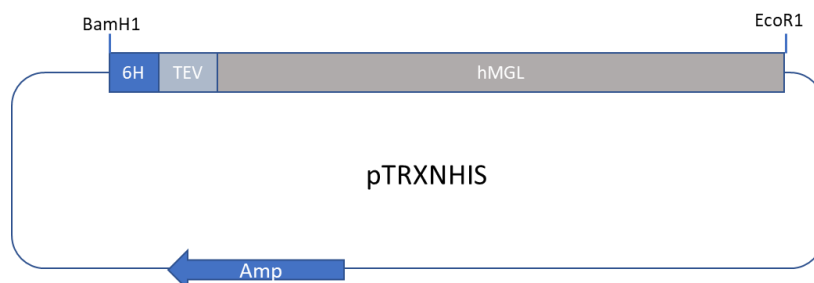


Figure 1. Simplified map of hMGL plasmids

For expression of the hMGL constructs a 6-His tag system consisting of plasmid pTRXNHIS was chosen. The plasmid encodes a 6-His tag followed by a TEV protease recognition epitope (E-N-L-Y-F-Q---G), unique cloning site BamH1, EcoR1 and stop codon. Synthetic clones corresponding to amino acids: - hMGL carbohydrate recognition domain, 181-309 (library code17-39); - hMGL full extracellular fragment isoform3, 81-309 (library code17-41); -hMGL extracellular fragment isoform3 short tail,145-309 (library code17-50), -hMGL extracellular fragment isoform3 long tail, 93-309 (library code17-49), -hMGL coiled-coil section, 73-179 (library code17-51); and -hMGL full extracellular fragment isoform3 were purchased from IDT (Coraville, IA) company as GBlocks and was amplified by polymerase chain reaction (PCR) using appropriate primers. PCR protocol: with initial denaturation 98°C for 2 min and then 30 cycles at 98°C for 10 sec, 60 °C for 15 sec ,72 °C for 30 sec and at 72°C for 5 min for the final extension. The DNA generated by polymerase chain reaction was double digested with BamH1-HF and EcoR1-HF and inserted between the BamH1-HF and EcoR1-HF sites of expression vector pTRXNHIS. The plasmid was transformed into DH5 α cells line. Colonies were picked, miniprep and the presence of the insert was verified by restriction digestion, agarose gel electrophoresis and overlapping DNA sequencing using T7 forward and T7 terminator primers.

6.4.2. Protein expression and purification

All the proteins were expressed and purified following the procedure reported below: 100 ng of plasmid DNA was used to transform C41(DE3) cells using standard procedures and colonies were grown overnight on LB agarose plates containing 100 μ g/ml ampicillin. A single colony was used to inoculate a 10 mL overnight culture containing 100 μ g/ml

ampicillin at 37°C. This was transferred into 1 L of LB Miller Broth containing 100 µg/ml of ampicillin. The cells were grown at 37°C until $OD_{600} = 0.6$ and protein expression was induced at 37°C for 4h by adding a final concentration of 1mM IPTG to give the desired protein as inclusion body. The cells were harvested and washed with PBS buffer.

To remove soluble contaminating *e. coli* proteins, the inclusion body were washed with PBS and purified alternating wash/sonication steps, figure 2, (25 mL of buffer for 1L culture, buffer 1: 20mM Tris pH 8.0, 0.5% Triton-100, 5 mM EDTA pH 8, 25 mM b-ME; buffer 2: 20mM Tris pH 8.0, 1 % Triton-100, 5 mM EDTA pH 8, 25 mM b-ME, 500 mM NaCl; buffer 3: 20mM Tris pH 8.0, 5 mM EDTA pH 8, 25 mM b-ME, 1 M NaCl). The protein was solubilized with 25 mL of 2 M NH_4OH at room temperature stirring the solution for 1h.

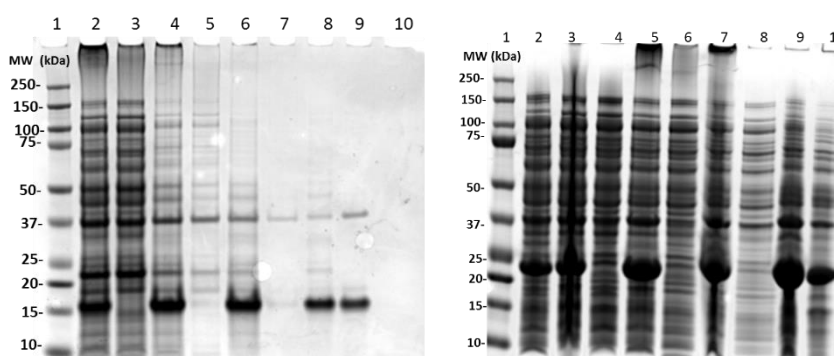


Figure 2. On the left SDS-PAGE of inclusion body purification, 6His hMGL (181-309) 16.9 kDa. Lane 1 Bio-rad Precision Plus Protein kaleidoscope marker (10 µL), lane 2 pellet suspended in Buffer#1, lane 3 supernatant #1, lane 4 pellet suspended in Buffer#2, lane 5 supernatant #2, lane 6 pellet suspended in Buffer#3, lane 7 supernatant #3, lane 8 pellet suspended in 2 M NH_4OH , lane 9 protein in 2 M NH_4OH . Where not otherwise indicated the sample, concentration is: 7 µL from 25 mL + 18 µL SDS-page buffer: 950 µL + 50 µL 14.3 M b-ME.

On the right: SDS-PAGE of inclusion body purification, 6His hMGL (81-309) 23.9 kDa. Lane 1 Bio-rad Precision Plus Protein kaleidoscope marker (10 µL), lane 2 uninduced 10 mL pellet sonicated in 1 mL PBS, lane 3 pellet suspended in Buffer#1, lane 4 supernatant #1, lane 5 pellet suspended in Buffer#2, lane 6 supernatant #2, lane 7 pellet suspended in Buffer#3, lane 8 supernatant #3, lane 9 pellet suspended in 2 M NH_4OH , lane 10 protein in 2 M NH_4OH . Where not otherwise indicated the sample, concentration is: 7 µL from 25 mL + 18 µL SDS-page buffer: 950 µL + 50 µL 14.3 M b-ME.

The soluble protein was slowly diluted 10 x with the refolding buffer, filtered and refolded by dialysis against the same buffer (5L X 2) (refolding buffer: 20 mM Tris pH 7.5, 500 mM NaCl, 25 mM $CaCl_2$, 1 mM reduced glutathione, 0.1 mM oxidized glutathione, refolding occurs in 8-16 h. The addition of 400 mM lysine allows refolding in 4 h).

The refolded protein was filtered and purified by affinity chromatography on cComplete His-Tag purification resin from Roche. The soluble protein was incubated with the resin at 4°C for 5 h, washed with 20 mM Tris pH 7.5, 500 mM NaCl and 5 mM imidazole and

eluted with 20 mM Tris pH 7.5, 150 Mm NaCl and 200 mM imidazole, figure 3.

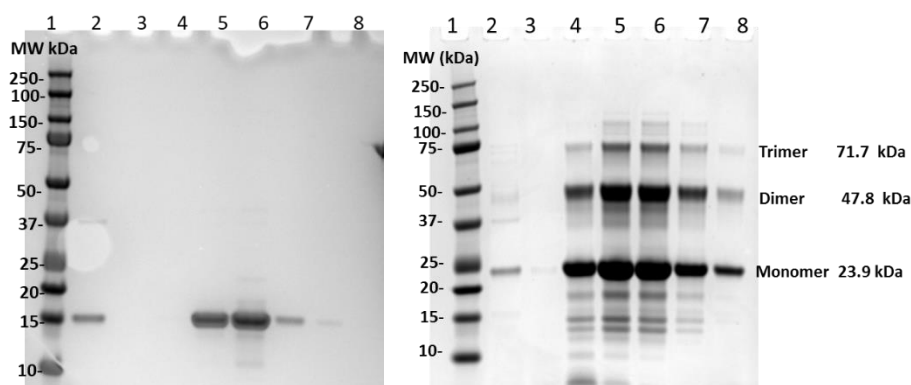


Figure 3. On the left: 6His hMGL CRD (181-309) 16.9 kDa. Lane 1 Bio-rad Precision Plus Protein kaleidoscope marker (10 µL), lane 1 sample before purification, lane 3 flow through, lane 4 wash step, lane 5 elution test tube #1, lane 6 elution test tube #2, lane 7 elution test tube #3, lane 8 elution test tube #4.

On the right: SDS-PAGE of Ni-NTA purification, 6His hMGL (81-309) monomer 23.9 kDa, dimer 47.8 kDa, trimer 71.7 kDa. Lane 1 Bio-rad Precision Plus Protein kaleidoscope marker (10 µL), lane 2 flow through, lane 3 wash step, lane 4 elution test tube #1, lane 5 elution test tube #2, lane 6 elution test tube #3, lane 7 elution test tube #4, lane 8 elution test tube #4.

The protein was concentrated, and the buffer exchanged with low salt ion exchange buffer then the His tag was removed treating the protein with TEV at 4°C for 16 h.

The proteins were purified with ion exchange chromatography, low salt buffer: 20 mM Tris pH 7.5, 40 mM galactose, 2 mM CaCl₂; high salt buffer: 20 mM Tris pH 7.5, 40 mM Galactose, 2 mM CaCl₂, 1M NaCl, and with size exclusion chromatography, size exclusion buffer (SEC buffer): 20 mM Tris pH 7.5, 40 mM Galactose, 2 mM CaCl₂, 150 mM NaCl following the protocol reported in table 1.

Table 3 Ion exchange and size exclusion purification protocols.

Protein	Ion exchange	Size exclusion
6 His hMGL CRD, 181-309	HiPrep QFF 16/10 column, 1 mL/min, 0-40min: 0 % low salt buffer, 40-120 min: 0→50% high salt buffer, 120-170 wash step 50→100 % high salt buffer. RT = 85 min.	Supradex 75 column, 0.5 mL/min, 100% SEC buffer. RT = 182 min.
hMGL CRD, 181-309	HiPrep QFF 16/10 column, 1 mL/min, 0-40min: 0 % low salt buffer, 40-120 min: 0→50% high salt buffer, 120-170 wash step 50→100 % high salt buffer RT = 78 min.	Supradex 75 column, 0.5 mL/min, 100% SEC buffer. RT = 173 min.
hMGL isoform3 81-309	HiPrep QFF 16/10 column, 1 mL/min, 0-40min: 0 % low salt buffer, 40-125 min: 0→50% high salt buffer, 125-140 wash step 50→100 % high salt buffer.	Supradex 200 column, 0.5 mL/min, 100% SEC buffer

RT = 114 min.

RT =95 min

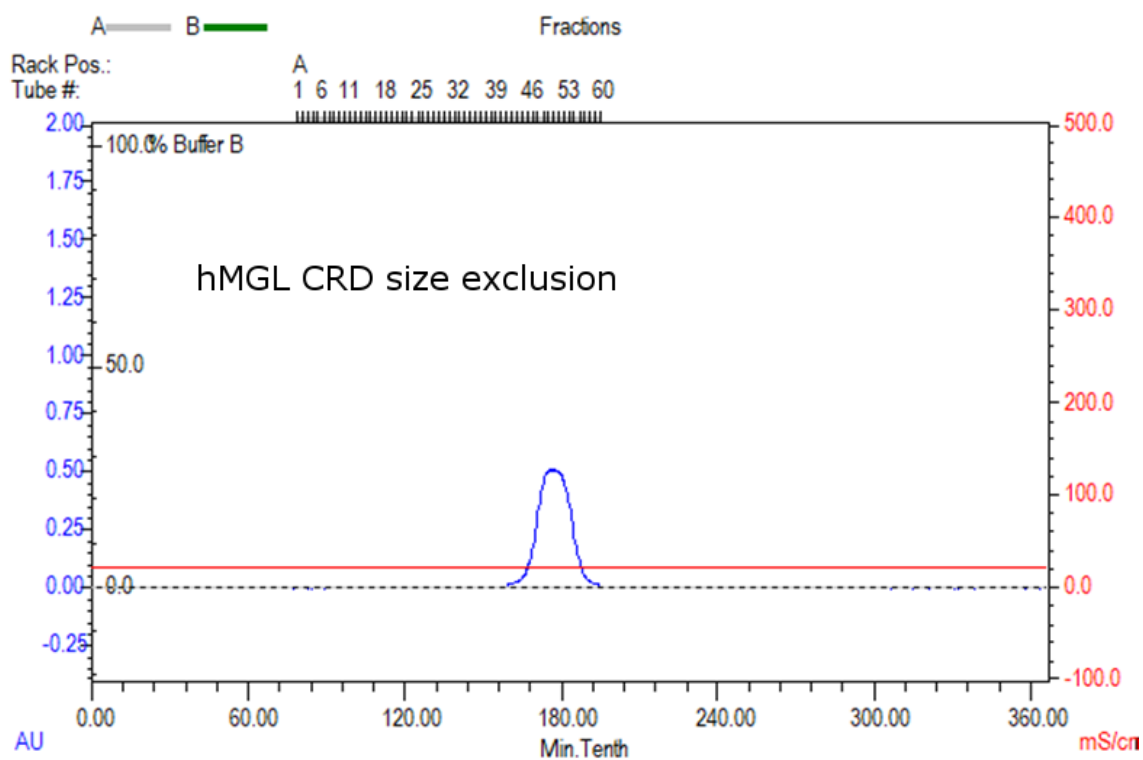
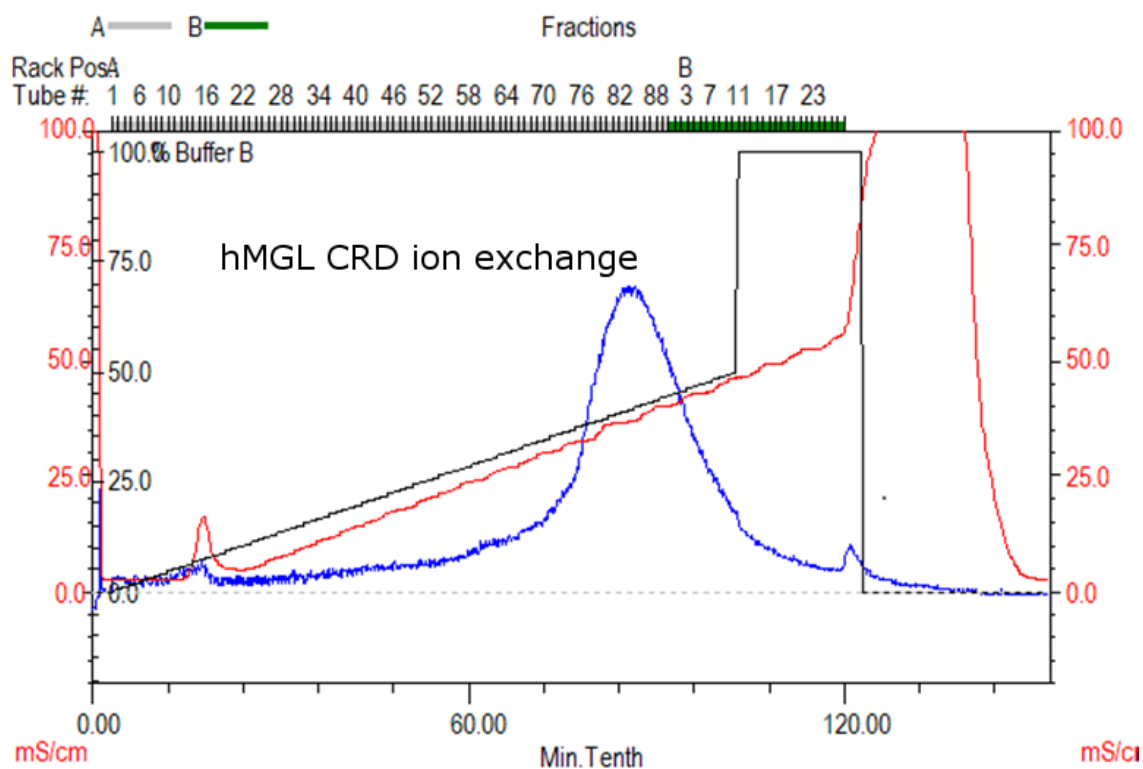
hMGL	full	HiPrep QFF 16/10 column, 1 mL/min, 0-40min: 0 % low salt buffer, 40-125 min: 0→50% high salt buffer,125-140 wash step 50→100 % high salt buffer.	Supradex 200 column, 0.5 mL/min, 100% SEC buffer
------	------	--	--

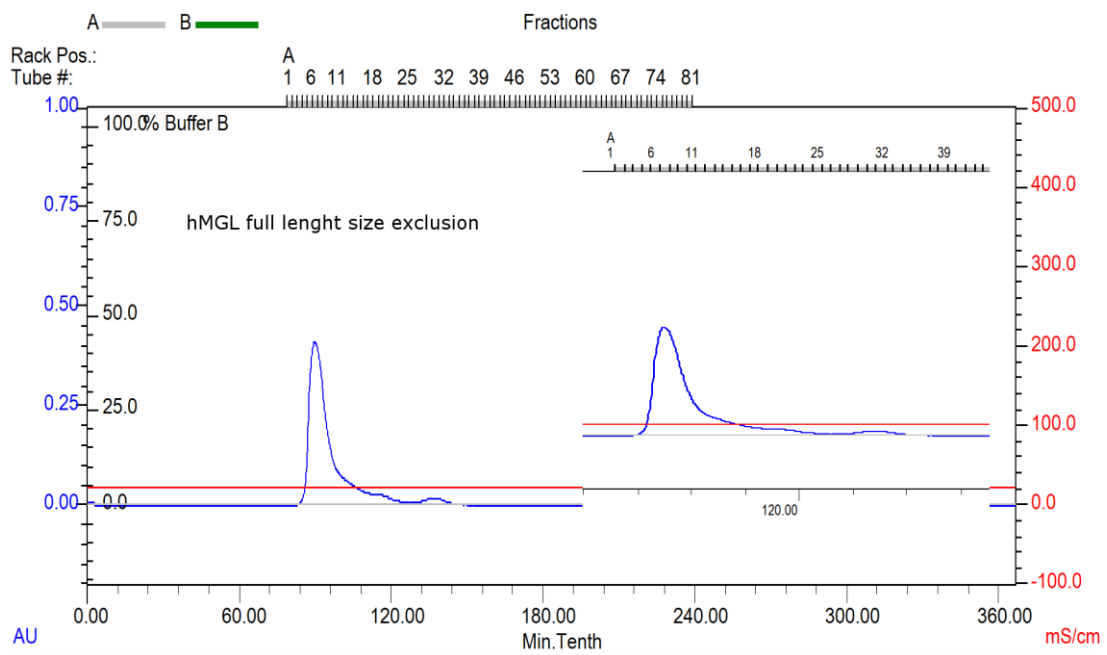
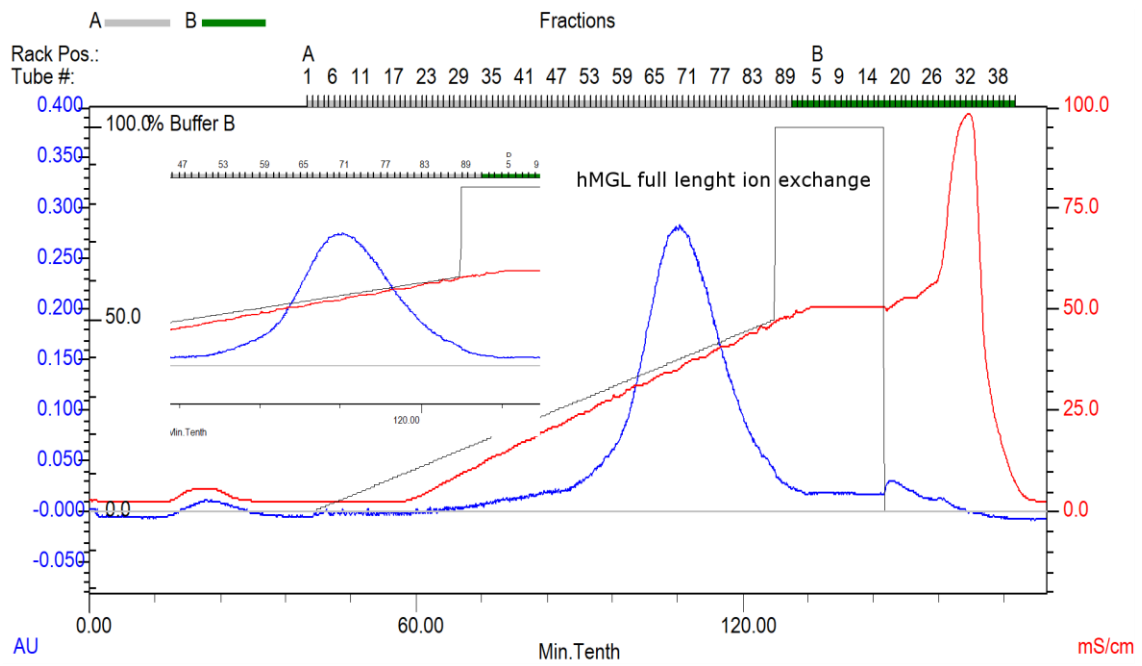
RT = 117 min.

RT =90 min

The fractions corresponding to hMGL were collected, concentrated and applied to polyacrilamide column washed with 20 mM Tris pH 7.5, 2 mM CaCl₂, 150 mM NaCl.

Chapter 6





6.4.3. hMGL CRD crystallisation

hMGL CRD, 181-309

Refolded and purified hMGL 181-309 was concentrated up to 23 mg/mL in buffer containing 20 mM Tris pH 7.5, 2 mM CaCl₂, 150 mM NaCl. Crystals were grown as rod cluster, in space group P2₁2₁2₁ at 18 °C by sitting drop vapour diffusion from drops containing 1 µL of protein and 1 µL of reservoir solution containing 20% (w/v) PEG 3000 and 100 mM sodium citrate tribasic/hydrochloric acid pH 5.5. Crystals were transferred stepwise in reservoir solutions containing ethylene glycol 5-20 % as cryoprotectant and flash frozen in liquid nitrogen for data collection.

Co-crystallisation hMGL CRD, 181-309 with GalNAc/ methyl 2-(acetylamino)-2-deoxy-1-thio-alpha-D-galactopyranose / Ser-T_N antigen

Refolded and purified hMGL 181-309 was concentrated up to 67 mg/mL (23 mg/mL for T_N Antigen) in buffer containing 20 mM Tris pH 7.5, 2 mM CaCl₂, 150 mM NaCl. GalNAc/ methyl 2-(acetylamino)-2-deoxy-1-thio-alpha-D-galactopyranose / Ser-T_N antigen was added in 10 molar fold excess and incubated at room temperature for 1h to obtain a protein concentration of 60 mg/mL (20 mg/mL for Ser-T_N antigen). Crystals were grown as needle cluster, in space group P3₁2 at 18 °C by sitting drop vapour diffusion from drops containing 1 µL of protein-ligand complex and 1 µL of reservoir solution containing 100 mM Tris pH 6-7, 200 mM MgCl₂, NaCl 2M-2.8M. Crystals were transferred stepwise in reservoir solutions containing ethylene glycol 5-15% as cryoprotectant and flash frozen in liquid nitrogen for data collection.

6.4.4. Data collection and crystal structure analysis

Diffraction data were recorded using an Pilatus 3 6M detector at beamline ESRF, ID-30B. The data were collected at ESRF (Grenoble, France). In all cases, data were collected at 100 K. A total of 360 frames per crystal were recorded with an oscillation range of 0.5° and an exposure time of 0.5 or 1 s per frame. X-ray wavelengths and data collection statistics are listed in table 2, chapter 5. For all data sets, data processing and scaling was carried out using HKL2000.¹⁴

All structures were solved by molecular replacement using the program Phaser¹⁵ with the protein coordinates of ASGPR (PDB:1DV8¹⁶) as the starting model. Ligands, ions and water molecules were only included in the model when unbiased positive difference electron density was clearly visible. All the structures were refined using the program Refmac5¹⁷ with 5% of unique reflections excluded from refinement calculations for cross validation purposes.

Ligand restraints were generated using the CCP4i monomer sketcher¹⁸, Coot¹⁹ and the PRODRG server.²⁰ Electron density map were inspected and model building were done using Coot. Geometry was validated, and Ramachandran plots were calculated.

Distances were measured using Coot. The structure analysis and the images were produced using UCSF Chimera.²¹

6.5. Isothermal titration calorimetry methodology

Binding constants of the glycans with hMGL CRD and hMGL full-length were measured by isothermal titration calorimetry (ITC) using a MicroCal VP-ITC MicroCalorimeter (MicroCal Inc.). hMGL was dialyzed against a buffer containing 20 mM Tris pH 7.5, 2 mM CaCl₂, 150 mM NaCl. The lyophilised glycoclusters were dissolved in the same buffer. The experiments were performed at 25 °C. The sample cell was loaded with hMGL solution, and the protein was titrated with the sugar solution as typical for forward ITC analysis. Protein concentration was kept low (7-20 μM) to avoid precipitation and it was determined by measuring the absorbance at 280 nm, after centrifuging the sample at high speed for 10 min. We used molar concentrations of multivalent carbohydrate analogues, not their epitope equivalence. A total of 33 automatic injections of 8 μl each were carried out with spacing delays of 320 seconds between each injection. The heats of dilution, determined by the titration of the sugar into the same buffer alone, were subtracted from the raw titration data before data analysis. The integrated heats derived from the injection series were fitted to a single site model by the Origin 7 software (OriginLab), which yielded the molar enthalpy of the association and the association constant. The stoichiometry was also estimated in the analyses.

6.6. Bibliography

- (1) Kaltner, H.; Manning, J. C.; García Caballero, G.; Di Salvo, C.; Gabba, A.; Romero-Hernández, L. L.; Knospe, C.; Wu, D.; Daly, H. C.; O'Shea, D. F.; et al. Revealing Biomedically Relevant Cell and Lectin Type-Dependent Structure-Activity Profiles for Glycoclusters by Using Tissue Sections as an Assay Platform. *RSC Adv.* **2018**, *8* (50), 28716–28735. <https://doi.org/10.1039/c8ra05382k>.
- (2) Cagnoni, A. J.; Varela, O.; Gouin, S. G.; Kovensky, J.; Uhrig, M. L. Synthesis of Multivalent Glycoclusters from 1-Thio- β -D-Galactose and Their Inhibitory Activity against the β -Galactosidase from *E. Coli*. *J. Org. Chem.* **2011**, *76* (9), 3064–3077. <https://doi.org/10.1021/jo102421e>.
- (3) Hu, X. M.; Chen, Q.; Wang, J. X.; Cheng, Q. Y.; Yan, C. G.; Cao, J.; He, Y. J.; Han, B. H. Tetraphenylethylene-Based Glycoconjugate as a Fluorescence “Turn-on” Sensor for Cholera Toxin. *Chem. - An Asian J.* **2011**, *6* (9), 2376–2381. <https://doi.org/10.1002/asia.201100141>.
- (4) Andrus, M. B.; Mendenhall, K. G.; Meredith, E. L.; Soma Sekhar, B. B. V. Glycolate Aldol Reactions with Boron Enolates of Bis-4-Methoxyphenyl Dioxanone. *Tetrahedron Lett.* **2002**, *43* (10), 1789–1792. [https://doi.org/10.1016/S0040-4039\(02\)00152-1](https://doi.org/10.1016/S0040-4039(02)00152-1).
- (5) Kutzner, T. J.; Gabba, A.; FitzGerald, F. G.; Shilova, N. V; García Caballero, G.; Ludwig, A.-K.; Manning, J. C.; Knospe, C.; Kaltner, H.; Sinowatz, F.; et al. How Altering the Modular Architecture Affects Aspects of Lectin Activity: Case Study on Human Galectin-1. *Glycobiology* **2019**, *29* (8), 593–607. <https://doi.org/10.1093/glycob/cwz034>.
- (6) Wu, J.; Sun, S.; Feng, X.; Shi, J.; Hu, X. Y.; Wang, L. Controllable Aggregation-Induced Emission Based on a Tetraphenylethylene- Functionalized Pillar[5]Arene via Host-Guest Recognition. *Chem. Commun.* **2014**, *50* (65), 9122–9125. <https://doi.org/10.1039/c4cc03127j>.
- (7) Kaltner, H.; Manning, J. C.; García Caballero, G.; Di Salvo, C.; Gabba, A.; Romero-Hernández, L. L.; Knospe, C.; Wu, D.; Daly, H. C.; O'Shea, D. F.; et al. Revealing Biomedically Relevant Cell and Lectin Type-Dependent Structure-Activity Profiles for Glycoclusters by Using Tissue Sections as an Assay Platform. *RSC Adv.* **2018**, *8* (50), 28716–28735. <https://doi.org/10.1039/c8ra05382k>.
- (8) Dowlut, M.; Hall, D. G.; Hindsgaul, O. Investigation of Nonspecific Effects of Different Dyes in the Screening of Labeled Carbohydrates against Immobilized Proteins. *J. Org. Chem.* **2005**, *70* (24), 9809–9813. <https://doi.org/10.1021/jo051503w>.
- (9) Knapp, S.; Myers, D. S. α -GlcNAc Thioconjugates. *J. Org. Chem.* **2001**, *66* (10), 3636–3638. <https://doi.org/10.1021/jo010088e>.
- (10) André, S.; O'Sullivan, S.; Koller, C.; Murphy, P. V; Gabius, H. J. Bi- to Tetravalent Glycoclusters Presenting GlcNAc/GalNAc as Inhibitors: From Plant Agglutinins to Human Macrophage Galactose-Type Lectin (CD301) and Galectins. *Org. Biomol. Chem.* **2015**, *13* (14), 4190–4203. <https://doi.org/10.1039/c5ob00048c>.
- (11) Gelas, J.; Horton, D. Kinetic Acetonation of D-Mannose: Preparation of 4,6-Mono- and 2,3,4,6-Di-O-Isopropylidene-d-Mannopyranose. *Carbohydr. Res.* **1978**, *67* (2), 371–387. [https://doi.org/10.1016/S0008-6215\(00\)84126-9](https://doi.org/10.1016/S0008-6215(00)84126-9).
- (12) Ernst, B.; Winkler, T. Summary: The Anomeric Hydroxyl Group of Various Furanose Andpyranose Hemiacetals Can Be Replaced by A. *Tetrahedron Lett.* **1989**, No. 23, 3081–3084.
- (13) Haque, M. B.; Roberts, B. P.; Tocher, D. A. Enantioselective Radical-Chain Hydrosilylation of Alkenes Using Homochiral Thiols as Polarity-Reversal Catalysts. *J. Chem. Soc. - Perkin Trans. 1* **1998**, No. 17, 2881–2889. <https://doi.org/10.1039/a803280g>.
- (14) Otwinowski, Z.; Minor, W. Processing of X-Ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* **1997**, *276* (January 1993), 307–326. [https://doi.org/10.1016/S0076-6879\(97\)76066-X](https://doi.org/10.1016/S0076-6879(97)76066-X).
- (15) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (4), 658–674. <https://doi.org/10.1107/S0021889807021206>.
- (16) Sanhueza, C. A.; Baksh, M. M.; Thuma, B.; Roy, M. D.; Dutta, S.; Préville, C.; Chrnyk, B. A.; Beaumont, K.; Dullea, R.; Ammirati, M.; et al. Efficient Liver Targeting by Polyvalent Display of a

- Compact Ligand for the Asialoglycoprotein Receptor. *J. Am. Chem. Soc.* **2017**, *139* (9), 3528–3536. <https://doi.org/10.1021/jacs.6b12964>.
- (17) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (4), 355–367. <https://doi.org/10.1107/S0907444911001314>.
- (18) Potterton, E.; Briggs, P.; Turkenburg, M.; Dodson, E. A Graphical User Interface to the CCP4 Program Suite. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **2003**, *59* (7), 1131–1137. <https://doi.org/10.1107/S0907444903008126>.
- (19) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (12 I), 2126–2132. <https://doi.org/10.1107/S0907444904019158>.
- (20) Schüttelkopf, A. W.; Van Aalten, D. M. F. PRODRG: A Tool for High-Throughput Crystallography of Protein-Ligand Complexes. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (8), 1355–1363. <https://doi.org/10.1107/S0907444904011679>.
- (21) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. <https://doi.org/10.1002/jcc.20084>.