UMEÅ UNIVERSITET

Development and Evaluation of Tools to Explore Posttranslational HexNAc-Tyrosine and Mucin-Type *O*-Glycosylation

Sandra Behren

Doctoral Thesis, Department of Chemistry
Umeå University, 2021

Responsible publisher under Swedish law: Dean of the Faculty of Science and Technology

Dissertation for degree of Doctor of Philosophy
ISBN: 978-91-7855-646-5 (print)
ISBN: 978-91-7855-647-2 (digital)
Printed by: VMC-KBC Umeå
Electronic version available at http://umu.diva-portal.org/
Umeå, Sweden 2021

The only impossible journey is the one you never begin.

-Tony Robbins

Table of Contents

Abstract

Glycosylation is the most abundant form of post-translational modifications (PTMs). Recently, *O*-glycosylation attracted much attention in the glycoproteomic field due to its association with various diseases, such as pathogenic infections and cancer. However, glycoproteomic analysis of *O*-linked glycosylation is highly challenging due its structural diversity and complexity. New and efficient methods need to be developed to get a better understanding of the biological functions of *O*-glycans. In the presented thesis, glycopeptide microarrays were used as tools to explore the role of mucin type *O*-glycosylation in cancer, bacterial adhesion processes and galectin recognition on a molecular level, and to get insights into a new group of tyrosine *O*-glycosylation. A better understanding of these carbohydrate-protein interactions on a molecular level would facilitate the development of glycomimetic inhibitors to fight bacterial infections or block glycan binding proteins involved in cancer progression, or improve the design of novel carbohydrate-based cancer vaccines.

In the first part of this work, tools were developed to elucidate the role of a new group of PTMs, where *N*-acetylhexosamine (HexNAc = α-GalNAc, α- or β-GlcNAc) was found to modify the hydroxyl group of tyrosine. Synthetic glycopeptides carrying this new modification and glycopeptide microarray libraries were prepared to evaluate the abilities of plant lectins (carbohydrate-binding proteins) to detect HexNAc-*O*-Tyr modifications. These lectins are commonly used in glycoproteomic work flows to detect and enrich glycopeptides and -proteins. Additionally, HexNAc-*O*-Tyr-specific rabbit antibodies were raised and immunologically analyzed by enzyme-linked immunosorbent assays, western blot and microarray binding studies.

In the second part of the presented thesis, synthetic mucin glycopeptide microarray libraries were prepared and employed to explore carbohydrate-protein interactions of galectins, bacterial lectins and tumor specific antibodies. Mucin glycoproteins are part of the mucus barrier that protects the host against invading pathogens. However, bacteria and viruses have co-evolved with the human host and developed strategies to promote virulence, for example by adhering to glycans on the host cell-surface. To combat bacterial infections, their virulence and pathogenicity must be understood on a molecular level. In this work, mucin

glycopeptides were enzymatically modified with different fucose motifs and used to determine the fine binding specificities of fucose-recognizing lectins LecB from *Pseudomonas aeruginosa* and the *Clostridium difficile* toxin A. Furthermore, a synthesis strategy was developed to generate simplified mucin core glycopeptides that can be used as scaffolds to enzymatically generate LacdiNAc modified glycopeptides. They can be used in microarray binding studies to evaluate the glycan binding preferences of various proteins, including the *Helicobacter pylori* lectin LabA and human galectins, which play roles in cancer development and progression. Aberrant glycosylation of mucin glycoproteins has been associated with various types of cancer. Tumor specific carbohydrate antigens on mucins represent attractive antigenic targets for the development of effective anti-cancer vaccines. In this work, antibodies induced by tumor-associated MUC1 glycopeptide-bacteriophage Qβ vaccine conjugates were immunologically analyzed using MUC1 glycopeptide microarray libraries.

List of Abbreviations

4PL	four-parameter logistic regression
aa	amino acid
Ac	acetyl
AC_2O	acetic anhydride
ATP5B	ATP synthase subunit beta
Boc	*tert*-butyloxycarbonyl
BSA	bovine serum albumin
Bu	butyl
c	concentration
C(1-4)T1/2	core (1-4) type-1 or -2
C1GalT	β1,3-galactosy transferase
C2GnT-1	β1,6-acetylglucosaminyl transferase
C2GnT-2	β1,6-acetylglucosaminyl transferase
C3GnT	β1,3-*N*-acetylglucosaminyl transferase
calc.	calculated
cat.	catalytic
CBPs	carbohydrate binding proteins
CF	cystic fibrosis
CFTR	cystic fibrosis transmembrane conductance regulator
^{C}Hex	cyclohexane
CID	collision induced dissociation
COPD	chronic obstructive pulmonary diseases
COSY	correlated spectroscopy
CRD	carbohydrate recognition domain
d	duplet or day
DCC	*N*,*N*'-dicyclohexyl carbodiimide
DCM	dichloromethane
dd	duplet of duplet
DIC	*N*,*N*'-diisopropyl carbodiimide
DIPEA	diisopropylethylamine
DMAP	4-(dimethylamino) pyridine
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DMTST	dimethyldithiosulfonium triflate
ECD	electron-capture dissociation
ECM1	Extracellular matrix protein 1
ELISA	enzyme-linked immunosorbent assay
Eq/equiv	equivalents
ER	endoplasmic reticulum
ESI	electrospray ionization
Et	ethyl
ETD	electron-transfer dissociation
Ex	excitation
FA	formic acid
Fmoc	*N*-(9H-fluoren-9-yl)-methoxycarbonyl
Fuc	L-fucose
FUT	fucosyltransferases
Gal	D-galactose or galectin
GalNAc	*N*-acetyl-D-galactosamine
Glc	D-glucose
GlcNAc	*N*-acetyl-D-glucosamine
grad	Gradient
GSL II	*Griffonia simplicifolia* lectin II
GST	glutathione-S-transferase
HATU	1-[Bis(dimethylamino) methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HBP	hexosamine biosynthetic pathway
HBTU	*O*-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyl unronium hexafluoro phosphate
HBTU	2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HCD	high-energy collision dissociation
HexNAc	*N*-acetylhexosamine
hGal	human galectins
HMBC	heteronuclear multiple bond correlation
HOAT	1-hydroxy-7-azabenzotriazole
HOBt	1-hydroxybenzotriazole

HPLC	high performance liquid chromatography
HR	high resolution
HRP	horseradish peroxidase
HSQC	heteronuclear single quantum coherence
HOBt	1-hydroxybenzotriazole
Hz	Hertz
Ig	immunoglobulin
*i*Pr	isopropyl
J	coupling constant
K_D	dissociation constant
Lac	lactose
LacdiNAc	GalNAc-*β*-1,4-GlcNAc
LacNAc	*N*-acetyllactosamine
Le	Lewis
LG	leaving group
LWAC	lectin weak affinity chromatography
m	multiplet
M	molarity or mega
mAb	monoclonal antibody
MALDI	matrix assisted laser desorption ionization
Man	D-mannose
mbar	millibar
Me	methyl
mRNA	messenger ribonucleic acid
MS	molecular sieves or mass spectrometry
MUC	mucins
NBS	*N*-bromosuccinimide
NCE	normalized collision energy
Neu5Ac	*N*-acetylneuraminic acid
Neu5Gc	*N*-glycolylneuraminic acid
NHS	*N*-hydroxysuccinimide
NIS	*N*-iodosuccinimide
NMP	*N*-methylpyrrolidone
NMR	nuclear magnetic resonance
NUCB1	Nucleobindin-1
NUCB2	Nucleobindin-2
OGA	*O*-GlcNAcase
OGT	*O*-linked *N*-acetylglucosamine transferase
OTf	trifluoromethanesulfonate
p	*para*
PA-IIL	homotetrameric lectin LecB
PaTox[G]	*Photorhabdus asymbiotica* protein toxin
Pbf	2,2,4,6,7-pentamethyldihydro benzofuran-5-sulfonyl
PBS/ PBST	phosphate buffered saline/PBS+Tween-20
PEG	polyethylene glycol
PG	protecting groups
Ph	phenyl
PMP	*para*-methoxy phenyl
ppGalNAcT	polypeptide *N*-acetylgalactosamine transferase
ppm	parts per million
PRAP1	proline-rich acidic protein-1
PTMs	post-translational modifications
p-TsOH	*para*-toluenesulfonic acid
q	quartet
quart	quaternary
R_f	retention factor
RNA	ribonucleic acid
RP	reverse phase
R_t	retention time
s	singlet
SEA	sea urchin sperm protein, enterokinase and agrin
Sia	sialic acid
SLe	sialyl-Lewis
SPPS	solid phase peptide synthesis
ß1,3GalT	β1,3-Gal-transferase
ß1,3GlcNAc T	β1,3-GlcNAc-transferase
ß1,4GalT	β1,4-Gal-transferase
ST	sialyl-Thomsen-Friedenreich antigen
ST3Gal-I to -V	α2,3-sialyltransferases
ST6GalNAc	α2,6-*N*-acetylgalactosamine sialyltransferases
ST_N	sialyl-Thomsen Nouveau antigen
Surf.	surface
t	triplet
T/TF	Thomsen-Friedenreich antigen
TACAs	tumor-associated carbohydrate antigens
TBS	*tert*-butyldimethylsilyl
TcdA	*Clostridium difficile* toxin A
TEG	triethylene glycol

tert	tertiary	Troc	2,2,2-trichloroethoxy carbonyl
Tf	triflate	UDP	uridine diphosphate
TFA	trifluoroacetic acid	UV	ultraviolet
TfOH	trifluoromethanesulfonic acid	VDAC1	voltage-dependent anion-selective channel protein 1
THF	tetrahydrofuran	VLPs	virus-like particles
TIPS	triisopropylsilane	VNTR	variable number of tandem repeats
TLC	thin layer chromatography	VVA	*Vicia villosa* lectin
TMS	tetramethylsilyl	WGA	*Wheat germ agglutinin*
T_N	Thomsen-Nouveau antigen	α	specific optical rotation
TOCSY	total correlation spectroscopy	δ	chemical shift
TOF	time of flight	λ	wavelength
TR	Tandem repeats		

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

List of Publications

This thesis is based on the following papers, referred to in the text by their roman numners.

I **S. Behren**‡, M. Shorlemer‡, G. Schmidt, K. Aktories, U. Westerlind. Antibodies directed against GalNAc- and GlcNAc-*O*-Tyrosine posttranslational modifications - a new tool for glycoproteomic detection. Paper under review.

II **S. Behren**, M. Schorlemer, Y. Xiao, R.J. Woods, F. Marcelo, U. Westerlind. Deciphering the Molecular Recognition of GalNAc- and GlcNAc-*O*-Tyrosine Glycopeptides by Plant Lectins. Advanced manuscript.

III **S. Behren**‡, J. Yu‡, C. Pett, M. Schorlemer, V. Heine, T. Fischöder, L. Elling, U. Westerlind. Bacteria Lectin Recognition Towards Fucose Binding Motifs Highlights the Impact of Presenting Mucin Core Glycopeptides. Preprint. doi:10.33774/chemrxiv-2021-79qhk

IV **S. Behren**, T. Funder, L. Elling, U. Westerlind. Synthesis of Simplified Mucin Cores 1-4 MUC1 and MUC5AC Glylcopeptides for Enzymatic Modification with LacdiNAc Motifs. Manuscript.

V **S. Behren**, M. Schorlemer, J. Jiménez-Barbero, U. Westerlind. Binding Specificities of Human Galectins toward Mucin Glycopeptide Libraries. Manuscript.

VI X. Wu, C. McKay, C. Pett, J. Yu, M. Schorlemer, S. Ramadan, S. Lang, **S. Behren**, U. Westerlind, M. G. Finn, X. Huang, Synthesis and Immunological Evaluation of Disaccharide Bearing MUC-1 Glycopeptide Conjugates with Virus-like Particles. *ACS Chem Biol* **2019**, *14*, 10, 2176-2184; doi:10.1021/acschembio.9b00381.

VII X. Wu, H. McFall-Boegeman, Z. Rashidijahanabad, K. Liu, C. Pett, J. Yu, M. Schorlemer, S. Ramadan, **S. Behren**, U.Westerlind, X. Huang. Synthesis and immunological evaluation of the unnatural β-linked mucin-1 Thomsen-Friedenreich conjugate. *Org. Biomol. Chem.* **2021**, *19*, 2448-2455; doi:10.1039/D1OB00007A.

‡ Authors contibuted equally. Shared 1st author.
All papers have been printed with permission from the publishers.

Author contributions

Paper I: Microarray fabrication, Synthesis of antigen peptide vaccine conjugates, immunological analysis of antibodies by ELISA and microarray assays, purification of antibodies by affinity chromatography, western blot analysis, data analysis, major writing.

Paper II: Microarray fabrication, microarray binding studies with VVA, WGA, GSL II and β-GlcNAc-specific antibody, data analysis, major writing.

Paper III: Fucosylation of glycopeptides, microarray fabrication, microarray binding studies with LecB and TcdA, data analysis, major writing.

Paper IV: Synthesis of simplified mucin core threonine building blocks, glycopeptide synthesis, major writing.

Paper V: Microarray fabrication, flurescent labeling of human galectins, microarray binding studies with labeled galectins, data analysis, major writing.

Paper VI: Microarray fabrication, microarray binding studies with raised antibodies, data analysis.

Paper VII: Microarray fabrication, microarray binding studies with raised antibodies, data analysis.

Papers by the author, but not included in this thesis

S. Behren, U. Westerlind. Glycopeptides and -Mimetics to Detect, Monitor and Inhibit Bacterial and Viral Infections: Recent Advances and Perspectives. *Molecules* **2019**, *24*, 1004; doi:10.3390/molecules2406100.

S. Jiang, T. Wang, S. Behren, U. Westerlind, K. Gawlitza, J. L. Persson, K. Rurack. Sialyl-Tn antigen-imprinted Dual Fluorescent Core-shell Nanoparticles for Selective Labeling of Cancer Cells. Manuscript under review.

C. Pett, S. Behren, G. Larson, J. Nilsson, U. Westerlind. Universal assignment of Sialic Acid Isomers by LC/MS-Based Glycoproteomics. Manuscript.

Enkel sammanfattning på svenska

Glykosylering är en grupp av posttranslationella modifieringar som är vanligt förekommande på proteiner. Kolhydrater som är länkade till proteiner via aminosyrorna; serin, treonin eller tyrosin brukar betecknas som O-glykosylerade proteiner. Dessa modifieringar spelar en viktig roll i många sjukdomsprocesser, bland annat inom cancer och infektionsbiologi. En utmaning med att både identifiera och studera O-glykosylerade proteiner och deras biologiska funktioner beror på att de bundna kolhydraterna ofta består av komplexa strukturer. Diversitet på proteinerna uppnås dessutom genom att en specifik aminosyra bindningsposition kan modifieras med olika typer av kolhydratstrukturer samt genom variationer mellan vilka möjliga aminosyra glykosylerings ”sites” på det specifika proteinet som blir modifierade. Det finns ett behov av att utveckla effektiva verktyg för att detektera och identifiera dessa protein-modifieringar som är en viktig del i cellernas sätt att kommunicera med varandra och sin omgivning. För att förstå denna kommunikation så finns det även ett behov av att kartlägga interaktionspartners (kolhydratbindande proteiner) som känner igen specifika kolhydrater på O-glykosylerade proteiner.

I denna avhandling har ett bibliotek av syntetiska mucin *O*-glykopeptider tagits fram som modellstrukturer för användning i microarray-bindningsstudier av kolhydratbindande proteiner (lektiner) som är involverade i cancer och bakterieadhesions processer. Bland annat studerades Galectiner som genom interaktioner med kolhydrater på tumörceller kan möjliggöra spridning av metastaser och avstängning av viktiga immunceller. Mucin glykopeptid biblioteket nyttjades även i bindningsstudier av lektiner från bakterierna *Pseudomonas aeruginosa* och *Clostridium difficile. P. aeruginosa* infekterar framförallt patienter med kronisk inflammation i luftvägarna medan *C. difficile* orsakar infektion i mag-tarm kanalen hos immunförsvagade patienter. Bättre förståelse för struktur-funktion samband i dessa bindningsinteraktioner är av vikt för att utveckla framtida ”glykomimetika” som exempelvis kan blockera bakterieadhesion till våra celler och därmed förhindra infektioner från specifika bakterier. Det syntetiska mucin *O*-glykopeptid-biblioteket användes även i mikroarray-bindningsstudier för utvärdering av korsreaktivitet och selektivitet hos antikroppar som bildats efter

immunisering med potentiella syntetiska cancervaccin. Att utvärdera bindningsspecificitet av dessa antikroppar är av vikt för att utveckla vaccin som är selektivt riktade mot tumörceller utan att det samtidigt utvecklas ett immunologiskt minne riktat mot friska celler.

I detta arbete har glykopeptider dessutom syntetiserats för att ta fram samt utvärdera kemiska verktyg som möjliggör studier av en nyligen upptäckt grupp av posttranslationella proteinmodifieringar, tyrosin ”*O*-HexNAcylering”. Däribland generades antikroppar som specifikt känner igen *N*-acetylhexosamin strukturer (HexNAc = α-GalNAc, α- or β-GlcNAc) på tyrosin. Dessa antikroppar är nu tillgängliga för att användas som intressanta detektionsverktyg vid identifiering av HexNAc-tyrosin O-glykosylering på olika proteiner. Därutöver utvärderades bindningsspecificitet av olika lektiner som skulle kunna vara intressanta för anrikning av HexNAc-tyrosin modifierade peptider genom lektin affinitets-kromatografi. På detta sätt ökas koncentrationen av HexNAc tyrosin modifierade peptider i proverna och därmed möjliggörs effektiv strukturanalys genom detektion med masspektrometri. Med tillgång till dessa nya verktyg så kan biologiska funktioner av HexNAc-tyrosin O-glykosylering börja studeras.

1 Introduction

Early on, natural compounds with the general formula $C_x(H_2O)_n$ were referred to as *carbohydrates*, a term that was derived from “hydrates of carbon.” For a long time, carbohydrates such as the polysaccharides starch and glycogen were thought of as sources of fuel and energy storage for living organisms. Also, cellulose, which consists of β-1,4-glycosidic linked glucose monomers, is the major structural component of plant cell walls. Later on it was discovered that all cells in the human body are covered by the glycocalyx, a gel-like layer consisting of free glycans or glycans attached to lipids or proteins, thus generating the glycoconjugates glycolipids, proteoglycans and glycoproteins. Human glycans are build-up from only ten monosaccharides: D-glucose (Glc), D-*N*-acetylglucosamine (GlcNAc), D-galactose (Gal), D-*N*-acetylgalactosamine (GalNAc), D-mannose (Man,) L-fucose (Fuc), D-*N*-acetylneuraminic acid (Neu5Ac), D-xylose (Xyl), D-glucuronic acid (GlcA) and L-iduronic acid (IdoA) (**Figure 1**).

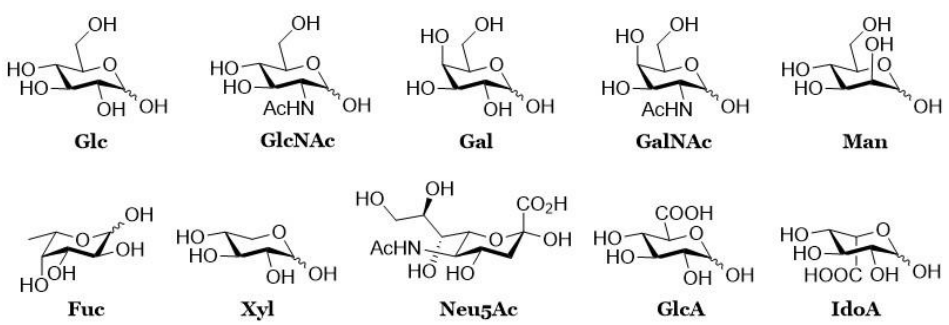


Figure 1. The ten human carbohydrate building blocks.

These glycans are structurally very complex and diverse. This complexity and diversity derives from the many possible ways in which monosaccharides can be linked together to form more intricate structures. Additionally, each glycosidic linkage connecting two carbohydrate moieties can be formed in two possible stereoisomers, termed α or β, at the anomeric carbon of one sugar, thus generating a new stereogenic center. In addition, the various hydroxyl groups of one monosaccharide permit the generation of several possible regioisomers. These isomers exhibit different three-dimensional structures and biological activities.

Finally, a monosaccharide can serve as a branching point by being involved in more than two glycosidic linkages. The structural glycan diversity can be further increased by acylation, sulfation, and phosphorylation.

In the late 19th century, Emil Fischer, who was awarded the Nobel Prize in 1902, laid the basis for understanding the organic chemistry of carbohydrates with his pioneering work on the classification of monosaccharide structures.[1] He established a system for the nomenclature and configurational assignment of carbohydrates.[2-3] Additionally, Fischer contributed greatly to the field of carbohydrate chemistry by exploring the chirality of sugars. He showed that both enantiomers of a given carbohydrate rotate the plane of polarized light with the same magnitude, but in opposite directions. In his work, he determined the absolute configuration of glucose, galactose, fructose, mannose, xylose and arabinose, and categorized the structures using a two-dimensional formula, which is today known as the *Fischer projection*, to relate the configurations of these chiral molecules. Furthermore, he postulated that the D- and L-symbols should be assigned depending on the spatial orientation of the carbohydrate substituents and not on the direction of the compound's optical rotation.

With the discovery of glycan diversity and complexity the roles that carbohydrates and their conjugates play in various complex biological processes became a field of interest. These processes include, for example, cell-cell interactions, molecular recognition, signal transduction, cell growth and proliferation, immune response and inflammation, as well as viral and bacterial infections.[4] The study of complex carbohydrates and their conjugates is hindered by their inherent structural diversity, complexity, low abundancy, and macro- and micro-heterogeneiety. While macro-heterogeneity refers to the site occupancy of glycosylation, micro-heterogeneity relates to the variations in glycan structure at a specific site. Both forms of heterogeneity can strongly impact the biochemical and physical protein properties. To elucidate complex glycan structures, classical chemical methods such as melting point and optical rotation analysis were shown to be inadequate. Progress in this area of carbohydrate research became possible only in the 1960s after NMR-spectroscopy and chromatographic methods were developed. Usually, combinations of tools and methods are required to study complex glycosylation. Nowadays, these tools and methods also include chemical

modification or cleavage, radioactive labeling, the use of enzymes such as endoglycosidases and exoglycosidases, lectins (which are carbohydrate-binding proteins), antibodies, cloning of glycosyltransferases as well as the genetic manipulation of glycosylation in living cells and organisms.

In recent years, highly improved mass spectrometry techniques and novel glycoinformatic tools have led to rapid progress in glycoproteomic studies.[5-7] In order to enable the analysis of intact glycopeptides, different fragmentation techniques including electron-transfer dissociation (ETD), electron-capture dissociation (ECD), collision-induced dissociation (CID) and high-energy collision dissociation (HCD) have been developed. The advantage of intact glycopeptide analysis is that information on both the glycosylation sites and glycan structures are obtained.

One approach to overcome the lack of homogenous natural glycan samples is the chemical synthesis of structurally well-defined glycans. These can then be used to develop new glycoproteomic techniques by allowing definitive structure assignment. Additionally, they can be applied to evaluate the interaction of carbohydrates with carbohydrate-binding proteins and thus contribute to the understanding of many biological processes. The glycosylation reaction is a central reaction in carbohydrate synthesis. In nature, this reaction is repeatedly executed by a variety of glycosyltransferases to yield complex glycans. However, the chemical formation of the glycosidic linkage is more complicated. The first described glycosylation reaction was performed by Arthur Michael in 1879.[8] Also, Wilhelm Koenigs and Eduard Knorr discovered the Ag_2CO_3-promoted glycosylation reaction of acetobromoglucose.[9] This reaction is known today as the *Koenigs-Knorr* reaction. Another pioneering approach on the glycosylation was carried out by Emil Fischer.[10] He performed the reaction under harsh acidic conditions using an excess of the glycosyl acceptor. Since then, carbohydrate chemistry has evolved into a broad research area. Carbohydrate chemists have developed increasingly refined strategies to tackle two fundamental problems of carbohydrate synthesis: i) Protecting groups to selectively mask hydroxyl and amine groups. They are also used to control the regioselectivity of reactions (see Chapter 1.3.1.2). ii) The stereoselective formation of glycosidic bonds. To solve this problem, methods for stereoselective glycosylation reactions were established to connect particular sugars (see Chapter 1.3.1.1).

One pioneer in the field of carbohydrate chemistry was the Canadian Raymond Lemieux.[11] In the 1950s, he reported the first chemical synthesis of sucrose and discovered the anomeric effect, which explains the preference of large electronegative substituents at the anomeric center for the more hindered axial position. Later on, Lemieux identified the endo- and exo- as well as the reverse anomeric effects.[12-13] Furthermore, he introduced ^{1}H and ^{13}C NMR spectroscopy to the field of carbohydrate structure analysis. In the 1970s, Lemieux set a milestone by developing the halide-ion-catalyzed glycosylation and by synthesizing the Lewis a blood group trisaccharide as well as the blood group A type-2 tetrasaccharide. In the 1980s, the introduction of better leaving groups enabled the synthesis of many biologically relevant oligosaccharides.[14] Since then, new leaving groups, activation methods, glycosylation conditions and strategies have been developed to synthesize more complex glycans. In 2001, new synthesis approaches to efficiently and rapidly generate oligosaccharides such as automated solid-phase synthesis and programmed one-pot synthesis have been established.[15-16] Another major innovation was the application of carbohydrate microarrays which has become an important tool to elucidate the roles glycans play in the biological system.[17] This microarray-based technology is extensively used to analyze carbohydrate interactions with protein receptors, antibodies, RNA, bacteria and viruses (Chapters 3 and 4).

1.1 *O*-glycosylation in eukaryotes – It's biosynthesis and biological functions

It has been predicted that more than 50 % of all human proteins are co- or post-translationally modified by mono- or oligosaccharides.[18] Protein glycosylation is one of the most abundant and diverse forms of post-translational modifications (PTMs) and is divided into two main classes: Whereas *N*-glycans are linked to the amine function of the amino acid asparagine (Asn or N) through an amide bond, *O*-glycans are attached to the hydroxyl groups of serine (Ser or S), threonine (Thr or T), or tyrosine (Tyr or Y). The protein *O*-glycosylation is a structurally more diverse group of modifications and are more challenging to study compared to *N*-glycosylation. The availability of specific endoglycosidases further facilitates glycoproteomic analysis of *N*-glycosylation over protein *O*-glycosylation. A number of different *O*-linked glycosylations have been identified, including *O*-mannosylation, *O*-fucosylation and *O*-galactosylation on hydroxyproline. The "mucin-type" (*O*-*N*-acetylgalactosamine-type or *O*-GalNAc-type) and the *O*-GlcNAcylation (*O*-*N*-acetylglucosamine- or *O*-GlcNAc-type) belong to the most common types of *O*-glycosylation.

1.1.1 *O*-GlcNAcylation – A nutrient sensor

Since its discovery in the 1980s, *O*-GlcNAcylation has been found to play important roles in various cellular functions by modifying nuclear, mitochondrial and cytosolic proteins.[19] *O*-GlcNAcylation is usually not further elongated to generate more complex carbohydrate structures. The *O*-GlcNAc residue is transferred from uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc) to Ser or Thr residues on the protein backbone under release of UDP by one single human enzyme, the *O*-linked *N*-acetylglucosamine transferase (OGT).[20] These glycosylation sites also overlap with phosphorylation sites. Consequently, *O*-GlcNAcylation competes with phosphorylation to occupy the same site or to sterically hinder adjacent positions.[21] The *O*-GlcNAc residue can be removed again by the enzyme *O*-GlcNAcase (OGA).

As a result, this PTM is controlled by the availability of its substrate UDP-GlcNAc as well as the activities of both OGT and OGA. The concentration of UDP-GlcNAc in cells depends on its biosynthesis by the hexosamine biosynthetic pathway (HBP). The HBP integrates the flux

through many metabolic pathways, including the glucose, amino acid, nucleotide, and fatty acid metabolisms that are linked to nutrient intake.[22-23] Consequently, *O*-GlcNAc cycling is considered as an intracellular nutrient sensor. For example, glucose deprivation leads to reduced intracellular UDP-GlcNAc levels and thereby decreased levels of protein *O*-GlcNAcylation. As a result, OGT mRNA transcription is being upregulated and the OGT modifies the glycogen synthase in turn, which ultimately results in ~60% glycogen synthase activity.[24] As mentioned above, *O*-GlcNAc also has an extensive cross-talk with phosphorylation.[23] It not only inhibits protein *O*-phosphorylation, but *O*-GlcNAcylation and phosphorylation also regulate each other's enzymes that catalyze the cycling of these PTMs. Thereby, *O*-GlcNAc can regulate signaling, mitochondrial activity, and cytoskeletal functions. Additionally, *O*-GlcNAcylation is involved in the regulation of many other biological processes including transcription, epigenetics, protein expression and stability.[25] For example, *O*-GlcNAcylation regulates transcription as it modifies the RNA polymerase II and the basal transcription complex.[26-27] Furthermore, it modulates the activities of many transcription factors such as STAT5 (signal transducer and activator of transcription 5) and NF-κB (nuclear factor- κB).[28-29] *O*-GlcNAc also plays a role in protein expression and stability. Many ribosomal proteins as well as associated translational factors are *O*-GlcNAcylated.[30] Also, *O*-GlcNAcylation increases the half-life of proteins by, for example, blocking ATPase activity and thereby reducing proteasome-catalyzed degradation.[31] In addition, *O*-GlcNAcylation has been implicated to mediate epigenetics by modifying histones (H2A, H2B, H3, and H4) and some epigenetic regulators.[32] Additionally, *O*-GlcNAc can modulate phosphorylation, acetylation, ubiquitination and methylation of histones.[33]

Aberrant *O*-GlcNAcylation has been associated with many diseases including neurodegenerative diseases and cancer. For example, it has been reported that *O*-GlcNAcylation is increased in cancers, contributing to the proliferation and growth of tumor cells since it integrates the nutrient flow with the metabolic pathways. It also regulates many proteins involved in cancer initiation and proliferation. *O*-GlcNAcylation also plays crucial roles in neurodegenerative diseases such as Alzheimer’s disease. The amyloid precursor protein is *O*-GlcNAcylated, and *O*-GlcNAcylation of the Tau protein decreases the phosphorylation and thus the cytotoxicity of Tau.[34-35]

1.1.2 Mucin-type *O*-glycosylation – A complex post-translational modification

The mucin-type *O*-glycosylation is contrary to *O*-GlcNAcylation the structurally most diverse form of protein *O*-glycosylation. There are three distinct regions that are recognized in *O*-glycans: the innermost carbohydrate residues constituting the core region, elongation with type-1 or type-2 chains and terminal epitopes. The initial step of mucin-type glycosylation is the addition of GalNAc to the acceptor amino acids Ser or Thr which is catalyzed by a large family of polypeptide GalNAc-transferases (ppGalNAc-Ts) (**Figure 2**A).[36] This structure is also known as the T_N-antigen (Thomsen-Nouveau antigen). Recently, GalNAc was found to also modify the hydroxyl group of Tyr. However, only a few glycoproteins carrying this new posttranslational modification are reported until now and the enzymes responsible for coupling of GalNAc to Tyr are unknown. The T_N-antigen can also be sialylated by a family of α2,6-*N*-acetylgalactosamine sialyltransferases (ST6GalNAc-I, -II and -IV) to form the corresponding sialyl T_N-antigen (ST_N).[37-38] Usually, the T_N-antigen is elongated at the C3 and/or C6 position, thus forming so-called core structures. At least eight different core types, of which cores 1-4 are more common than cores 5-8, have been found in mammalian glycoproteins.

The cores are synthesized in the Golgi apparatus by the stepwise transfer of carbohydrate residues from sugar nucleotide donors to acceptor substrates by the sequential action of glycosyltransferases. The basic core 1 motif, also named T- or Thomson-Friedenreich-antigen, is generated by the ubiquitously expressed β1,3-galactosyltransferase (C1GalT) that catalyzes the addition of galactose (Gal) in a β1,3-linkage to the GalNAc residue.[39-40] The core 1 can be sialylated at the Gal residue by the α2,3-sialyltransferase ST3Gal-I, thus forming the respective sialyl T-antigen (ST).[37-38] The core 2 structure is generated by addition of GlcNAc in a β1,6-linkage to the GalNAc residue of core 1 in a glycosylation catalyzed by a β1,6-acetylglucosaminyltransferase (C2GnT-1).[41] The core 3 structure is produced by addition of GlcNAc in a β1,3-linkage to the T_N-antigen by a β1,3-*N*-acetylglucosaminyltransferase (C3GnT).[42] This reaction competes with the core 1 formation. The core 4 is formed by addition of GlcNAc to the GalNAc residue of core 3 in a β1,6-linkage by another member of the β1,6-acetylglucosaminyltransferase (C2GnT-2) family.[43]

Subsequently, the *O*-glycan core structures can be elongated with type-1 or -2 *N*-acetyllactosamine (LacNAc) chains by alternating addition of GlcNAc and Gal residues via either β1,3-GlcNAc-transferase (β1,3GlcNAcT), or β1,3- or β1,4-Gal-transferase (β1,3GalT, β1,4GalT). The core structures can be either linear or branched, leading to the formation of i or I antigens, respectively.

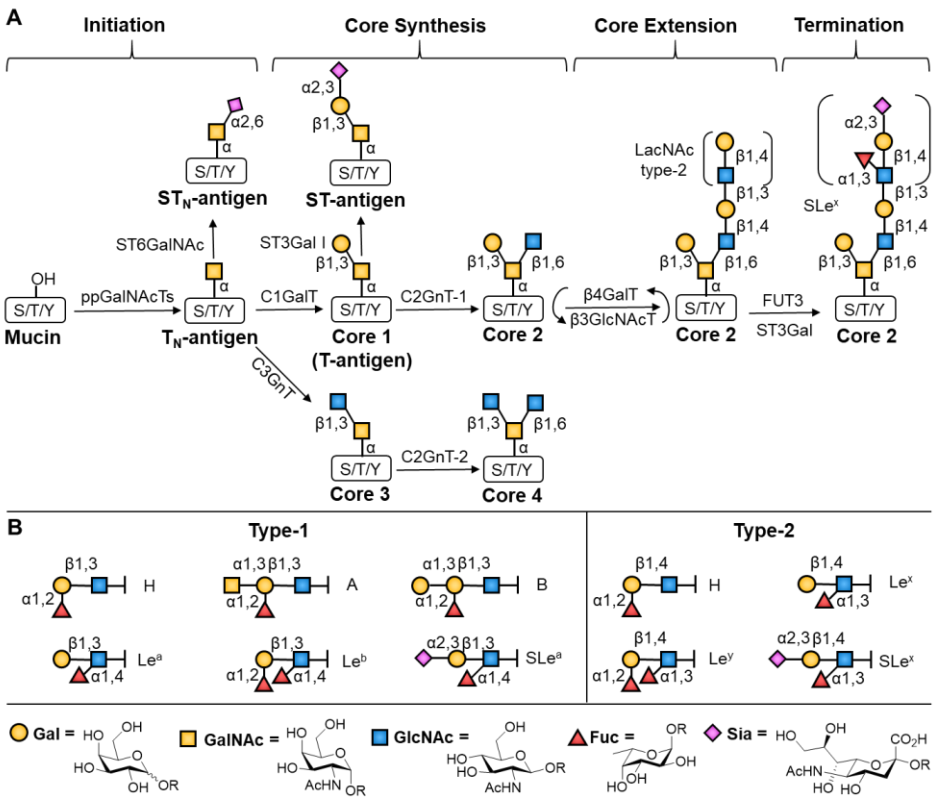


Figure 2. A) Schematic representation of pathways involved in *O*-glycan core formation, sequential elongation and termination; B) Terminal type-1 and -2 histo blood group antigens (H, A and B) and Lewis (Le^x, Le^a, Le^y, Le^b and the corresponding SLe^x and SLe^a) antigen determinants of mucin O-glycans.

Finally, the glycan chains can be terminated by the four sugar moieties fucose (Fuc), Gal, GalNAc and/or sialic acid (Sia), often in α–anomeric configuration, or sulfation. This way, the histo-blood group antigens such as A, B, H, or Lewis antigens such as Lewis a (Le^a), Lewis b (Le^b), Lewis x (Le^x) and Lewis y (Le^y) as well as the corresponding sialyl-Lewis antigens are formed (**Figure 2**B). The exact glycan structures depend among other things on the glycosyltransferases expressed in the cells. As a result, the terminal carbohydrate residues on the glycans are heterogeneous and vary within tissue types. This structural complexity and diversity allows the host to cope with various pathogens. These carbohydrates are added by different glycosyltransferase families.

In case of sialylation, human mucins usually contain the sialic acid *N*-acetylneuraminic acid (Neu5Ac) only, which can be *O*-acetylated on carbons 7, 8 and 9.[44] Human cells have lost the ability to synthesize the sialic acid *N*-glycolylneuraminic acid (Neu5Gc) due to the deletion of gene coding for the enzyme CMP-Neu5Ac hydroxylase.[45] However, Neu5Gc-containing glycoconjugates have been found in different human tissues in low amounts due to its dietary incorporation (.g. red meat and animal milk).[46] Additionally, Neu5Gc has been found on human cancer cells.[47] A family of α2,3-sialyltransferases (ST3Gal-I to –VI) attaches sialic acids (Sia) to terminal galactose moieties.[37-38] The α2,6-sialyltransferases ST6Gal-I and -II catalyze the transfer of Sia to the terminal Gal residue normally of type-2 disaccharides.[37-38] This modification is mainly found on *N*-glycans, but also on *O*-glycans to a lesser extent.[48] The three GalNAc α2,6-sialyltransferases ST6GalNAc-I, -II, and -IV catalyze, as described above, the transfer of Sia to the proximal GalNAc residue of *O*-glycans. Since sialylated glycans are not recognized by many other glycosyltransferases, further chain elongation is hindered.

An exception thereof is the α1,3/4-fucosyltransferases (FUT3) that transfers a fucose unit to either the 3- or the 4-position of the GlcNAc residue of type-1 and type-2 LacNAc disaccharides.[49] This way, the Lewis antigens Le^x and Le^a as well as their corresponding sialyl-Lewis x and sialyl-Lewis a determinants are formed. FUT3 can also act after a α1,2-fucosyltransferase to form Le^b or Le^y. The α1,2-fucosylation of the terminal Gal residue is performed by only two enzymes, FUT1 and FUT2.[50-51] The acceptor specificity of these enzymes is slightly different: While FUT1 prefers the type-2 LacNAc motif, FUT2 preferably modifies Galβ1,3GalNAc glycans.[52-53] Both enzymes show similar activity toward

type-1 LacNAc glycans. FUT1 and FUT2 are responsible for the expression of H-antigens and play consequently important roles in the formation of ABO blood group antigens.

The glycans can also be terminated by sulfation and the sulfate is added by two sulfotransferase families: The GST-family facilitates 6-*O*-sulfation on the 6 position of Gal, GalNAc or GlcNAc.[54] Sulfation of glycans at the 3 position of Gal residues is facilitated by the Gal3ST family.[55]

Whereas sulfates and sialic acid residues on Gal or GlcNAc moieties impart negative charges to mucin glycoproteins, fucose introduces hydrophobicity. Because of their characteristics, these terminal carbohydrates contribute to the physical and/or biological properties of mucins. Therefore, alterations of terminal glycosylation of mucins in diseases can alter the physical properties of mucins and thereby the rheological mucus properties.

1.2 The mucin glycoprotein family

Mucins (MUC) are highly *O*-glycosylated proteins (carbohydrate content 50-90 wt%) ubiquitously found on the epithelial cell surface.[56] They are a major constituent of the mucus layer, an aqueous gel consisting of water, ions, lipids, proteins and mucins. The mucus layer is a dynamic defensive barrier that contributes to the innate immune system that protects the epithelial tissues of the gastrointestinal and respiratory tract, and the ductal surfaces of breast, pancreas and kidney tissue from physical and chemical stress, toxins and invading pathogens.[57] Mucins have immunomodulatory roles in the mucus layer.[58-61]

To date, 21 mucin genes have been identified in humans (HUGO Gene Nomenclature Committee, www.genenames.org/cgi-bin/genefamilies/set/648). Human mucin genes exhibit Variable Number of Tandem Repeats (VNTR) loci, which are chromosomal regions where a short nucleotide sequence motif is repeated a variable number of times. These VNTR regions encode Tandem Repeat (TR) mucin peptide sequences that are rich in proline, threonine and serine (PTS) and form a scaffold for the attachment of *O*-linked glycans. The resulting dense glycan packing along the peptide backbone is responsible for the mucin filament structure and consequently for their functions.

Mucins can be divided into three subfamilies: membrane-bound/transmembrane (MUC1[62], MUC4[63] and MUC16[64]), secreted (gel-forming) (MUC2[65], MUC5AC[66], MUC5B[67] and MUC6[68]) and soluble (non-gel-forming) (e.g. MUC7, MUC8, MUC9, MUC20) mucins (**Figure 3**). Secreted mucins are major components of the mucus layer. They are stored in secretory granules and can be rapidly released within seconds to minutes in response to secretagogues.[69-70] In contrast, membrane-bound mucins are integrated into the cell membrane and are extended further from the glycocalyx than most cell surface receptors due to their rod-like conformation. This way, they can interfere with the adhesion of pathogens to the cell surface.

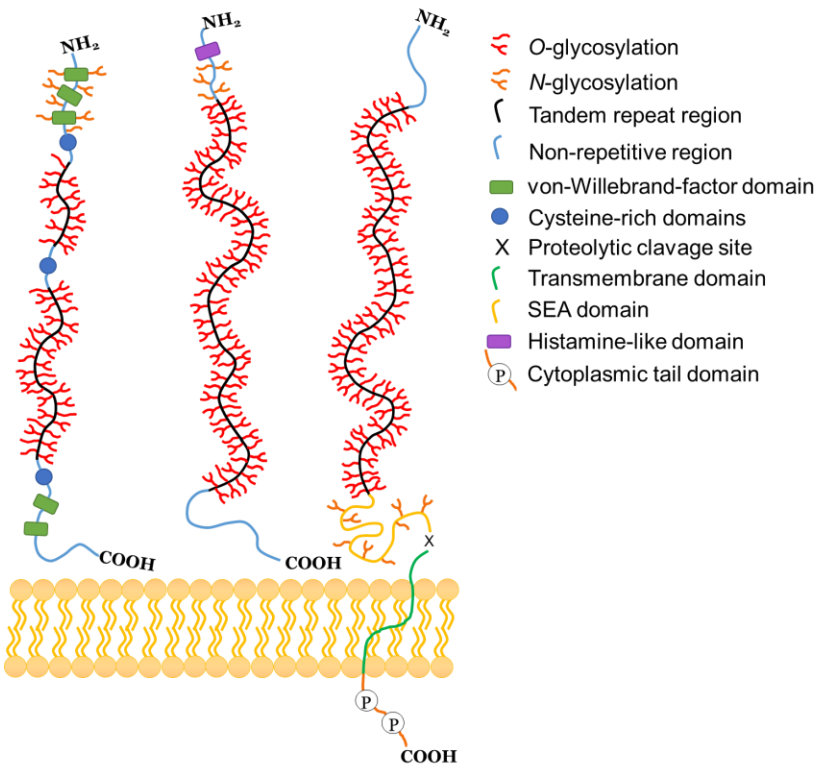


Figure 3. Schematic representation of the structures from secreted-gel-forming, secreted-non-gel forming and membrane-bound mucins (from left to right).

Mucins are mainly produced by goblet cells and mucous cells in surface epithelium. The mucin biosynthesis entails the transcription of a MUC gene to encode a MUC mRNA that is subsequently translated as apomucin in the rough endoplasmic reticulum (ER), which is finally posttranslationally glycosylated by various glycosyltransferases. The *O*-glycosylation is a stepwise process that starts in the *cis*-Golgi or in an intermediate compartment between ER and Golgi, and continues in the different compartments of the Golgi apparatus and the *trans*-Golgi network.

The MUC1 glycoprotein is a membrane-bound mucin that is ubiquitously found on epithelial cells. It contains an extracellular, an SEA (sea urchin sperm protein, enterokinase and agrin domain), a transmembrane domain and a *C*-terminal cytoplasmic tail. The mature MUC1 forms a stable heterodimeric complex that is generated by auto-catalytically cleavage of the precursor protein at the SEA domain during posttranslational processing[71]. The *C*-terminal subunit contains the cytoplasmic domain, the hydrophobic transmembrane domain and a short extracellular sequence. The larger extracellular *N*-terminal subunit contains the densely *O*-glycosylated VNTR domain and exhibits a rod-like structure that protrudes 200-500 nm into the extracellular space and surpasses the glycocalyx thickness (~10 nm).[72] The VNTR domain consist of a 20 amino acid repeat sequence (PAPGSTAPPAHGVTSAPDTR), which is repeated 20-125 times, and the total VNTR number depends on the individual genetic polymorphism.[73-74] Each MUC1 TR contains five potential glycosylation sites (3xT, 2xS) and is rich in proline, which distorts the secondary structure to its linear structure.[75]

The MUC5B glycoprotein is a secreted and major airway mucin. Due to amino acid deletions and insertions, its repeat region is non-tandem and degenerate and only 22 out of 55 possible repeats are present, making MUC5B a unique mucin.[76] It contains seven cysteine-rich regions, which are involved in the dimerization and multimerization of this glycoprotein by formation of disulfide bridges. In MUC5B, the VNTR region consists of 29 amino acids repeat sequence (ATGSTATPSSTPGTTHTPPVLTTTRTTPT).[9,10]

Muc5AC is - next to MUC5B - a major airway mucin. Additionally, it is one of the main secreted mucins in the stomach. MUC5AC is a large oligomeric mucin that is secreted by the surface epithelial cells.[66, 77] Its

tandem repeat domain is interrupted several times by a 130 amino acid cysteine-rich peptide sequence.[66] The MUC5AC TR sequence consist of eight amino acids (TTSTTSAP) and the individual repeats are interrupted by other amino acids. The amino acid sequence GTTPSPVPTTSTTSAP derived from the MUC5AC TR has often been applied in GalNAc-transferase assays.[78-82] It has nine potential glycosylation sites: six threonine and three serine residues.

1.2.1 Mucins in diseases

Alterations in terminal and core mucin glycosylation, which potentially alter the physical properties of mucins and thus the rheological mucus properties (viscosity, elasticity), is strongly associated with disease, such as diagnosis and prognosis of cancer, and pathogenic infections of, for example, the respiratory or gastrointestinal tracts.

1.2.1.1 Mucins in cancer

Membrane-bound mucins regulate growth factors, inflammatory signaling, transcription, apoptosis, differentiation, metastatic behavior and protection from the immune system in cancer cells.[83] For example, specific carbohydrate motifs can control growth factor signaling that is crucial for cancer development.[84] Additionally, changes in *O*-glycosylation, including aberrant and truncated glycosylation, cause loss of apical cell polarization, and alteration of adhesion and anti-adhesion effects.[85-86] They also impact tumor progression and metastasis by influencing cell recognition, trafficking and downstream signaling as well as cell-cell- and cell-matrix-interactions.

In carcinomas, mucins are the main carrier of aberrant and truncated glycosylation leading to the formation of TACAs (tumor associated carbohydrate antigens) (**Figure 4**).[87] TACAS are formed by, for example, abnormal fucosylation that increases the expression of Lewis structures such as Le^x and Le^a, and the expression of T- and T_N-antigens and sulfated glycans.[83] Additionally, the SLe^x and SLe^a determinants on glycans are often overexpressed and correlate with tumor progression such as metastasis and cancer cell invasion. Increased expression levels of sialyltransferases lead to premature sialylation and the formation of truncated glycan such as the Sialyl-T_N- and Sialyl-T-antigens.[85, 88] The expression of these antigens in human tumors has been associated with

poor prognosis. The LacdiNAc determinant has also been associated with various cancer types (see Chapter 4.2). These determinants have been identified cancer cell lines using mass spectrometry or by specific antibodies.

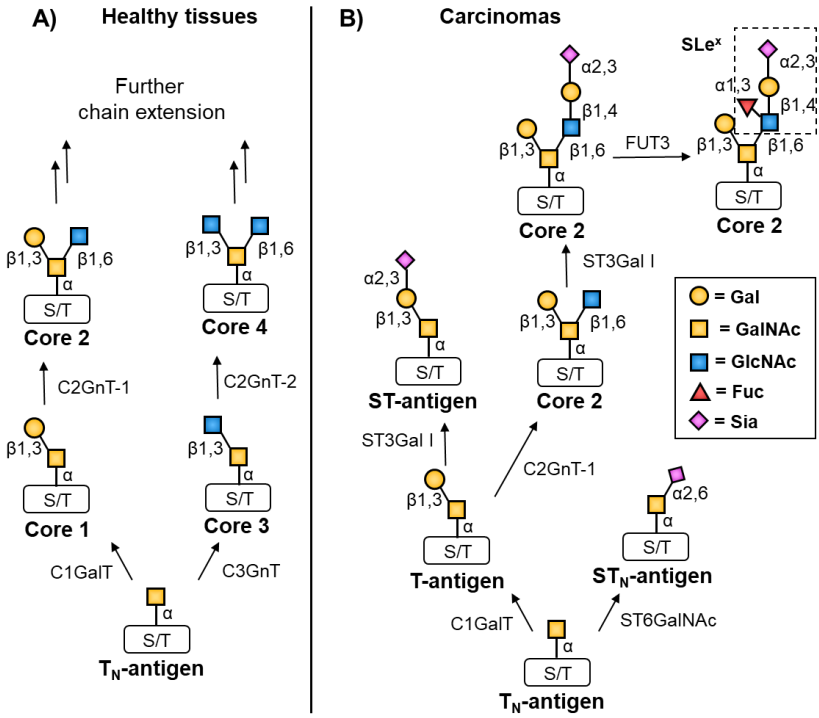


Figure 4. A) Core structures of mucins expressed by healthy tissue. B) Changes that occur in cancer result in the expression of tumor-associated T_N-, ST_N-, T-, ST- and SLe^x-antigens on mucins.

TACAs are associated with different cancer types such as breast, gastric, colorectal, pancreatic and lung cancer. The glycan composition of cancer cells can change as they evolve through different stages of disease. There are various mechanisms that lead to the changes in mucin-type *O*-glycosylation observed in cancer.[83] For example, mutations in the chaperone Cosmc responsible for the correct folding of the core 1 enzyme, C1GalT, result in the formation of T_N- or ST_N-antigens. Furthermore, ppGalNAc-Ts can be relocated to the endoplasmic reticulum upon stimulation of the proto-oncogene *Src*. This results in an increased T_N-antigen density, which blocks the core 1 C1GalT and core 2 C2GnT glycosyltransferases, which in turn leads to increased expressions of T_N-,

ST_N-, T- and ST-antigens. The most important mechanism is the alteration of the expression levels of various glycosyltransferases that install mucin core structures and/or terminal modifications.[89] For example, the formation of TACAs on mucins also can be attributed to downregulation of glycosyltransferases such as core 2 C2GnT, and premature sialylation by increased sialyltransferase expression.[85, 88] These changes are diverse and differ depending on the different cancer types and tissues. The altered glycosylation patterns occurring in cancer are promising targets to design and develop novel cancer diagnostic and therapy strategies. For example, TACAs are used in many approaches as cancer biomarkers: Overexpressed tumor-derived truncated mucins are secreted or shed into the extracellular space that surrounds the cancer cells. As a result, they can be detected in blood samples of cancer patients, and are thus diagnostic for cancer.[90]

Additionally, antibody-based therapies have been able to target TACA cancer markers, or tumor-associated antigens that are overexpressed on the surface of cancer cells. Many cancer antigens have already been used for therapeutic and diagnostic approaches.[91] Due to its high expression levels in various tumors, MUC1 presents one of the most important tumor markers which makes it a promising target for antibody-based therapies.[92] Furthermore, MUC1 is a target for the design of cancer vaccines that prevent cancer progression and metastasis. These vaccines include subunit, glycopeptide, DNA, viral vected, and dendritic cell vaccines.[93]

1.2.1.2 Mucins in airway diseases

Airway mucins are the major constituents of the mucus layer and contribute to the mucocilliary defense system protecting the respiratory tract against environmental toxins and pathogens. The different carbohydrate determinants on the mucins also function as ligands to various pathogens such as viruses and bacteria. In healthy individuals, these pathogens are often cleared by the mucus. Acute threats to the respiratory tract and specific secretagogues such as environmental toxins[94], bacterial-derived products[95], or infectious pathogens trigger lung inflammatory/immune response mediators.[96-97] Certain mediators can upregulate mucin gene expression which leads to airway mucin overproduction. Additionally, some mediators function as secretagogues and induce mucin hypersecretion by initiating a secretory cascade

resulting in the release of mucins stored in secretory granules from surface goblet and/or glandular secretory cells within minutes. These released mucins and their glycans protect the epithelial airway cells by entrapping particles and pathogens which are then removed by mucocilliary clearance.[98] Usually, this mucin overproduction reverts to baseline level after a couple of days in response to anti-inflammatory mediators and mechanisms.[99] However, patients with chronic airway diseases including asthma, chronic obstructive pulmonary diseases (COPD), or cystic fibrosis (CF) develop chronic mucin overproduction. In these diseases, airway remodeling such as goblet cell hyperplasia (elevated expression of Goblet cells in the airways), or glandular hyperplasia (increase in cell number) and hypertrophy (increase in cell size) can be caused by airway remodeling due to specific inflammatory/immune response mediators.[96] These processes lead to increased baseline levels of airway mucin production. This chronic mucin overproduction causes an aberrant flow of the mucus, which contributes to the formation of mucus plugs and ultimately airway obstruction, and therefore to the high morbidity and mortality associated with these airway diseases (**Figure 5**). Acute attacks, also called exacerbations, result in pulmonary obstruction and thus to an advancement in the disease stage, which ultimately leads to death.

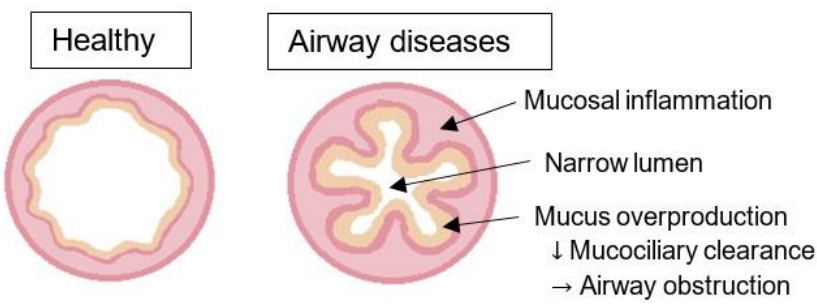


Figure 5. Schematic illustration of healthy airways and airways of patients suffering from airway diseases.

In these patient groups, changes in terminal mucin glycosylation such as altered levels of sialylation, sulfation and fucosylation have been identified.[100] In CF, for example, the airway mucins show an increase in fucosylation and sulfation and a decrease in sialylation.[101-103] These altered glycosylation patterns influence the biophysical and biological

properties of airway mucins which results in non-optimal transport properties of the mucus. Consequently, an environment is created in which pathogens can flourish and contribute to the inflammation. Usually, the majority of potential pathogenic infections is prevented by mucus clearance, but bacteria and viruses have co-evolved with the human host and developed strategies to promote immune escape and virulence.[104-105] For example, pathogenic bacteria adhere to mucin carbohydrate ligands on the host cell surface. To prevent immune cell recognition, they can also manipulate the glycan structures of the host by using specific glycosidases and as a result trigger inflammation, promote biofilm formation or build-up their own glycan shield.[106-107] Furthermore, mucin glycans are also targets of bacterial protein toxins that promote cell adhesion to enable intracellular protein toxin delivery.[104]

Even if the general symptoms of the airway diseases are similar, the origins are different. The knowledge about the fundamental cause of asthma is limited, but it is hypothesized that genetic predisposition and exposure to inhaled substances that trigger allergy are major factors.

COPD is a complex of diseases that is characterized by chronic airway obstruction and is associated with chronic bronchitis (mucus hypersecretion with goblet cell and submucosal gland hyperplasia) or emphysema (destruction of airway parenchyma). The small airways of patients with chronic bronchitis are chronically inflamed with increased levels cytokines such as interleukin-6, -1β, -8, tumor necrosis factor-α (TNF-α).[108] These cytokines are involved in cascades that activate tissue remodeling and/or MUC gene regulation.

CF is characterized by mutations in the gene that cystic fibrosis transmembrane conductance regulator (CFTR) protein, which is a adenosine monophosphate-regulated chloride ion channel.[109-110] These mutations lead to malfunction or even absence of this surface protein. Malfunction of CFTR causes electrolyte imbalances over the epithelial surface, leading to depletion of airway surface liquid water.[111] Together with the mucin overproduction, the electrolyte imbalance results in highly viscous mucous in the lungs. Cystic fibrosis is also characterized by persistent bacterial infections, for example by *Pseudomonas aeruginosa* (Chapter 4.1), which is confined to the airway lumen.[112-113]

1.3 Synthesis of carbohydrates, glycosylated amino acids and glycopeptides

1.3.1 Common methods for carbohydrate synthesis

In nature, the anomeric linkage between the carbohydrate units is important since often only one anomer is biological active. Therefore, stereoselective formation of glycosyl linkages is required during glycosylation reactions. The stereoselective outcome is influenced by several factors such as leaving groups (LG), protecting groups (PG), the solvent system, catalysts/promoters and temperature. Another requirement to synthesize complex glycans from basic building blocks is a unified protecting group strategy. Protecting groups temporarily block functional groups (e.g. hydroxyl groups, amines) that would otherwise also react and lead to the formation of side products. Many protecting groups also have a significant influence on the reactivity of glycosyl donors and acceptors and on the stereoselective outcome of the glycosylation reaction.

1.3.1.1 Common glycosylation methods

The glycosidic bond can exist in two different anomeric forms that are termed α– and β–anomeric bonds. To form a specific anomer it is necessary to control the stereoselectivity of the glycosylation reaction and the reactivity at the anomeric center. The reactivity of a carbohydrate building block strongly depends on its configuration and substituents. During the glycosylation reaction, a glycosyl donor with a leaving group is activated by a promoter/catalyst. After leaving group departure, the formed intermediate oxocarbenium ion can then be attacked by a glycosyl acceptor with a free hydroxyl group (**Figure 6**).

Figure 6. General acid-catalyzed glycosylation reaction; LG = leaving group, PG = protecting group, E = electrophilic promoter/catalyst, ROH = acceptor.

To date, various glycosylation methods using different glycosyl donors and promoters/catalysts have been reported. Common donors are the glycosyl halides, thioglycosides, trichloroacetimidates or 4-pentenyl glycosides (**Figure 7**). In this thesis, thioglycosides have been employed as glycosyl donors to generate the simplified mucin core structures (see 4.2) since they are easy to prepare and stable under many reaction conditions. Thioglycosides are usually activated with soft electrophiles that generate iodonium-ions such as NIS/TfOH, or sulfonium-ions, for example dimethyldithiosulfonium triflate (DMTST).

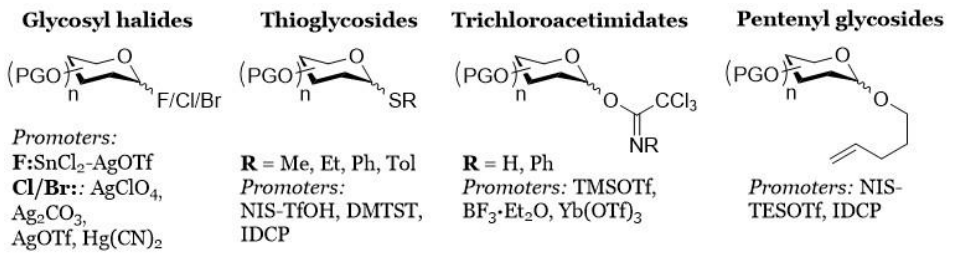


Figure 7. Common leaving groups and corresponding promoters for chemical glycosylations; PG = protecting group.

1.3.1.2 Protecting group chemistry

Every monosaccharide exhibits multiple stereocenters, hydroxyl groups and in some cases also amine or carboxylate functions.[114-115] As a result, carbohydrate synthesis requires complex protecting group manipulations that usually involve multistep synthesis protocols to discriminate between these functionalities. Even though the primary function of protecting groups is to mask a certain functionality on the carbohydrate ring, protecting groups in carbohydrate chemistry have more tasks. They can participate in reactions directly or indirectly and thus strongly influence the overall reactivity of a monosaccharide building block and the stereochemical outcome.[114-115]

Successful synthesis strategies depend on the use of permanent and temporary protecting groups. Permanent groups such as acetyl, benzoyl groups and benzyl ethers must be stable under all reaction conditions applied in the multistep synthesis. Usually, they are removed in a single reaction step at the end of the synthesis and thus have to be cleavable under the same reaction conditions. Temporary protecting groups include esters such as acetyl, chloroacetyl and benzoyl groups, ethers such as

methoxybenzyl, benzyl, trityl, silyl and allyl ethers, or acetals that simultaneously protect vicinal hydroxyl groups. Their usage strongly depend on the nature of the permanent protecting groups and the structural complexity of the target compound.

As a result, the use of temporary and permanent protecting groups have to be 'orthogonal' to each other.[116] The term 'orthogonal' in protecting group chemistry was introduced by Merrifield in 1977.[117] The advantage of orthogonal protecting groups is that they can be selectively removed one at a time without affecting other groups. Efficient synthesis of complex oligosaccharides requires a uniform synthesis strategy to avoid multiple protecting group manipulation steps during and also at the end of the synthesis. Due to their structural complexity, often a number of temporary protecting groups have to be employed during the course of synthesis. Because the carbohydrate hydroxyl functions display similar chemical reactivities, the preparation of complex glycans requires their selective protection. In recent years major progress in the field of protecting group chemistry has been made to regioselectivity protect functional groups of carbohydrates.[118-119]

The overall reactivity of carbohydrate building blocks is strongly influenced by the stereo-electronic properties of arming or disarming protecting groups.[120-121] The principle of 'armed' and 'disarmed' protecting groups was first introduced by Fraise-Reid in 1988.[122-123] Disarming protecting groups are electron-withdrawing groups including esters, amides or acyl groups. They increase the nucleophilicity of the leaving group and thereby deactivate glycosyl donors since their electron-withdrawing properties destabilize the oxocarbenium ion upon leaving group activation. If the electron withdrawing group is located at the 2 position of the donor, this effect is especially strong. On the other hand, armed protecting groups such as ethers activate carbohydrate building blocks.

Protecting groups such as esters, carbamates or amides installed at the C-2 position can give anchimeric assistance (neighboring group participation). This participating group can stabilize the oxonium ion which is transiently formed after leaving group departure by forming the more stable 1,2-dioxocarbenium ion (Figure 8). This dioxocarbenium ion can then be stereoselectively opened in a *cis*- (A) or *trans*-fashion (B) leading to the formation of an α- or β–glycosidic linkage. Due to the directing influence of participating neighboring groups, the 1,2-*trans*-glycoside can

be stereoselectively formed (C). If the nucleophilic attack of the glycosyl acceptor occurs instead at the dioxocarbenium ring, an orthoester will be generated (D). Non-participating neighboring groups such as ethers and azides lead to the formation of both α- and β-glycosides. The synthesis of 1,2-*cis*-glycosides requires the employment of non-participating groups. However, it is challenging to perform the glycosylation reaction with full stereoselectivity. Even though the formation of the α-anomer is favored due the anomeric effect, diastereoselective mixtures of α- and β-glycosides are commonly obtained.[124] The 'anomeric effect' was introduced by Lemieux in 1958 and originally described the tendency of electronegative substituents of a cyclohexly ring to prefer the sterically less favored axial position.[12]

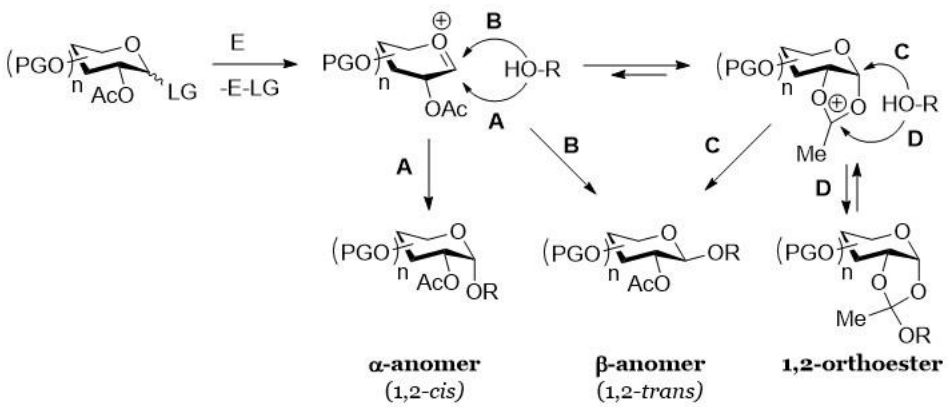


Figure 8. Possible reaction pathways in the glycosylation reaction of donors exhibiting a participating neighboring group on C-2. LG = leaving group, PG = protecting group, E = electrophilic promoter/catalyst, ROH = acceptor.

1.3.2 Mucin Glycopeptide synthesis

Due to the micro- and macroheterogeneiety of glycans and their low abundancy on proteins, it is challenging to obtain homogenous glycopeptides from biological samples. Chemical synthesis is a reliable method to generate structurally well-defined glycopeptides. To synthesize glycopeptides, three different strategies can be applied: The first approach involves the direct attachment of the glycan to the full-length target peptide. Therefore, suitably protected peptides and glycans are synthesized in a convergent fashion, which are subsequently condensed to form glycopeptides. This method is sometimes applied to synthesize *N*-glycopeptides by forming an amide bond between the glycan and aspartate. Preparation of *O*-glycopeptides requires the formation of an *O*-glycosidic bond, and challenges of stereochemical control caused by a structurally more complex peptide acceptor would not make this strategy feasible.

The second strategy to synthesize glycopeptides is chemoenzymatic elongation of simple glycopeptides. This approach employs the incorporation of simple glycosyl amino acid building blocks in SPPS and further glycan elongation by enzymatic modification. It can be applied for the synthesis of both *N*- and *O*-glycopeptides. In case of *N*-glycans, the initial, asparagine-linked β-GlcNAc can be extended *en bloc* with pre-synthesized complex oligosaccharides using endoglycosidases such as Endo M (*Mucor hiemalis*) and Endo A (*Arthrobacter protophormiae*). In contrast, *O*-glycans are assembled by stepwise enzymatic elongation using diverse glycosyl transferases.

The third and most often applied approach to generate glycopeptides is the stepwise glycopeptide assembly on solid-support using conveniently protected glycosylated amino acid building blocks. Solid phase peptide synthesis (SPPS) is an attractive method due to its fast peptide assembly, the possibility of automation and the reduction of chromatographic purification steps. The initial SPPS procedure reported by *Merrifield* in 1963 involved an acid-labile *tert*-butyloxycarbonyl (Boc) group as *N*-terminal protecting group during peptide assembly and aqueous hydrogen fluoride to release peptide from the solid support. This strategy is not feasible for glycopeptide synthesis since *O*-glycosidic bonds are sensitive to strong acids.

A solution for this problem presents the fluoren-9-ylmethoxycarbonyl (Fmoc)-SPPS strategy which was also applied in this work. Here, the *N*-α-

amino group is protected with the base-labile Fmoc group that can be removed by treatment with piperidine. This deprotection method is not basic enough to induce racemization of the amino acids, or β-elimination of the carbohydrate from Ser or Thr. In general, Fmoc-SPPS starts with an amino acid that is preloaded onto a solid-support such as TentaGel® R TRT resins via a cleavable linker (**Figure 9**). The TentaGel resin consist of a trityl linker is coupled to a low cross-linked polystyrene matrix via polyethylene glycol (PEG) chains. This linker is stable under the conditions used for peptide assembly, but can be cleaved under acidic conditions using trifluoroacetic acid (TFA). The Fmoc-SPPS proceeds then from the *C*- to the *N*-terminus by stepwise assembly of the peptide sequence.

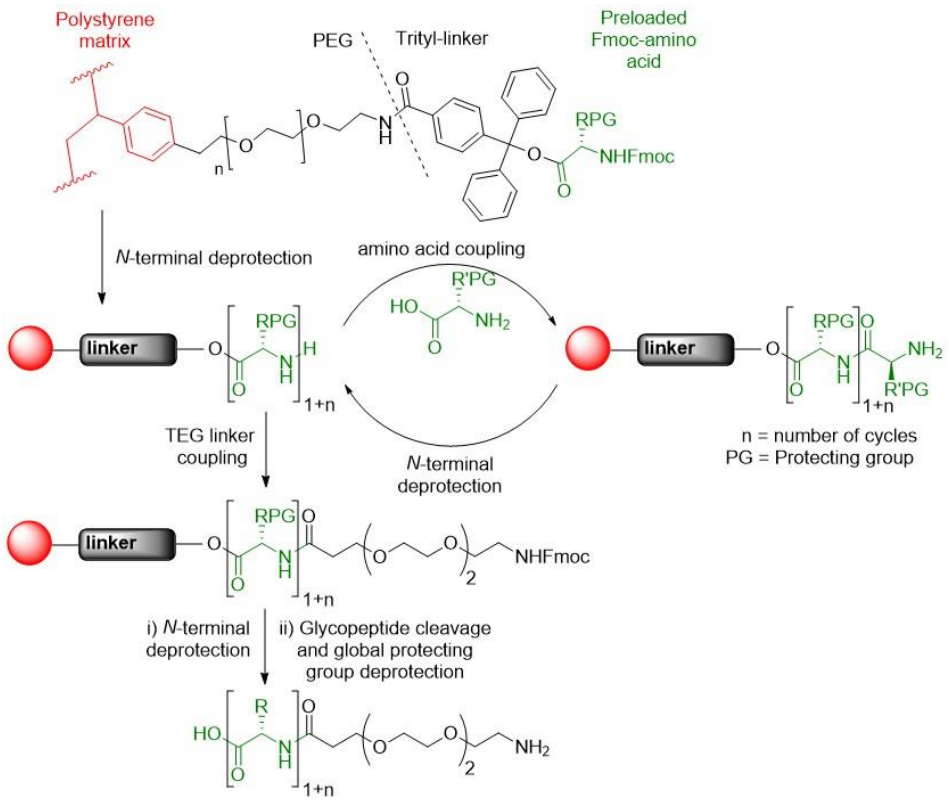


Figure 9. Principle of glycopeptide SPPS.

Non-glycosylated amino acids are usually coupled using a combination of 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and 1-hydroxybenzotriazole (HOBt). These agents give high coupling yields while preventing amino acid racemization. Because glycosylated amino acids are incorporated in lower excess (1.5 equiv) and in a smaller reaction volume, a more reactive system consisting of the coupling reagents 1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU) and 1-hydroxy-7-azabenzotriazole (HOAT) is used. Other typical coupling reagents are for example carbodiimides such as *N*,*N*'-dicyclohexylcarbodiimide (DCC), *N*,*N*'-diisopropylcarbodiimde (DIC) or *N*-ethyl-*N*'-(dimethyaminopropyl)-carbodiimide-hydrochloride (EDC·HCl), or phosphonium salts, including benzotriazol-1-yloxytris(pyrrolidino)phosphonium hexafluorophoshat (PyBOP).[125] After full sequence assembly, the glycopeptides are cleaved from the solid-support with simultaneous global amino acid side-chain deprotection using trifluoroacetic acid (TFA). In the process, carbocations that are generated during the release of the side-chain protection groups can cause side-reactions due to their uncontrolled addition. In this work, triisopropylsilane (TIPS) was used as a scavenger to prevent these side-reactions. Finally, the glycans on the peptide backbone are globally deacetylated using either a mild sodium methoxide/methanol system according to Zemplén, or a sodium hydroxide/water/methanol system.[126] Alternatively, the acetyl groups can be cleaved using hydrazine.[127] Strong basic reaction conditions may result in side reactions that are related to the stability of the glycosidic bond that connects the αGalNAc residue to the threonine or serine amino acid. These side reactions include deprotonation of the Cα-hydrogen by the base, or β-elimination of the glycan (Figure 10). While the deprotonation leads to epimerization of the stereogenic centers on Cα of the amino acids, which results in the formation of an equilibrium of D- and L-amino acids via an enolate intermediate, the β-elimination generates the α,β-unsaturated Ser or Thr alkene via an E1cB mechanism. However, these side reactions progress slowly compared with the glycan deacetylation, most probably due to a protective effect caused by the deprotonation of the peptide amide bond nitrogen. The third method is especially suitable to synthesize *O*-glycopeptides and has been applied in this thesis to produce simplified mucin core structures which were subsequently used in a semi-synthetic-

chemoenzymatic procedure to generate LacdiNAc modified glycopeptides (Chapter 4.2).

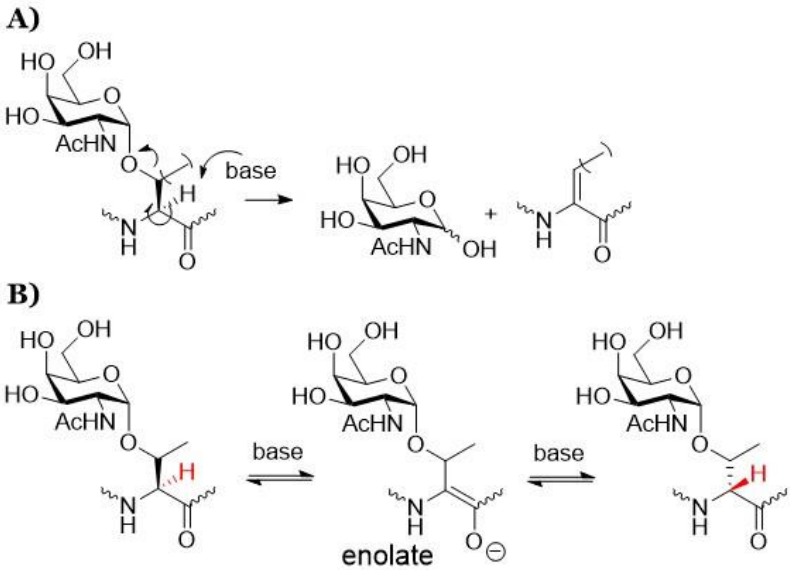


Figure 10. A) β-Elimination of glycans at high pH; B) Base-catalyzed epimerization of glycosylated amino acids.

1.3.3 Enzymatic modification

Synthetic carbohydrate chemistry requires multiple and selective protection and deprotection strategies to generate complex glycan structures, thus limiting feasibility and economic viability. Enzymatic glycosylation represents an important method to complement synthetic techniques. In all living organisms there are enzymes that catalyze specific reactions between carbohydrates. Enzymes that hydrolyze glycosidic bonds are termed glycosidases. In contrast, enzymes, which attach carbohydrates to an acceptor substrate using suitable sugar donors, are called glycosyltransferases. The most common donor sugars are nucleoside diphosphate sugars such as UDP-Gal or GDP-Man. Additionally, nucleoside monophosphate sugar donors (e.g. CMP-Neu5Ac), or lipid phosphates (e.g. dolichol phosphate oligosaccharides) can be used.[128] Acceptor substrates are usually other sugars, but lipids, peptides, glycopeptides, proteins, glycoproteins, nucleic acids, antibiotics, or another small molecules can also be modified.

Glycosyltransferases are either isolated from natural sources or recombinantly expressed. They are highly regioselective and stereospecific with respect to the glycosidic bond that is formed. Their major advantage is that they incorporate unprotected carbohydrate precursors thus avoiding tedious chemical protecting group chemistry. In the enzymatic glycan biosynthesis, a donor sugar is activated in the first step. Then, the activated carbohydrate moiety is transferred to an acceptor sugar. As a result, eight different nucleotide mono- or diphosphates (UDP-Glc, UDP-GlcNAc, UDP-Gal, UDP-GalNAc, GDP-Man, GDP-Fuc, UDP-GlcA, and CMP-NeuNAc) can be used as monosaccharide donors.

In contrast to the long-standing hypothesis that one enzyme is specific for one substrate or catalyzes one reaction, it is known today that many carbohydrate-processing enzymes exhibit a certain versatility in donor and/or acceptor substrate recognition.[129-131] Glycosyltransferases can tolerate modifications to the acceptor as long as specific structural requirements such as appropriate stereochemistry or availability of the reactive hydroxyl group are met.[132-133] A major limitation of enzyme-catalyzed glycosylations is that the glycosyltransferase can be inhibited by nucleoside diphosphates generated during the reaction. One strategy that is also applied in this thesis is to prevent this enzymatic inhibition by the addition of phosphatase to the reaction.[134] The phosphatase degrades nucleoside diphosphates by removing the phosphate group.

In the context of this work, glycosyltransferases were used to further diversify our mucin core glycopeptide library (glycopeptide synthesis was performed by Dr. Yu Jin, Dr. Christian Pett. and Dr. Manuel Schorlemer). These glycopeptides were enzymatically modified with poly-*N*-acetyllactosamine (poly-LacNAc), Neu5Ac, Neu5Gc, deaminated neuraminic acid and fucose at different linkages.

In this thesis, fucosyl transferases have been applied (Chapter 4.1) to selectively modify glycopeptides with Lewis a, x, b and y determinants. Additionally, a chemo-enzymatic approach to generate LacdiNAc-modified glycopeptides was developed as mentioned above (Chapter 4.2).

1.4 Carbohydrate microarrays

Over the years, glycan-protein interactions have been studied using different forms of arrays where glycans or glycoconjugates were attached to for example resins[135] or plates[136]. Based on DNA-microarray techniques, which were developed in the early 90s for large-scale DNA mapping and sequencing[137], and for transcript-level analyses[138]; several independent research groups reported the generation of carbohydrate microarrays in 2002. While two studies published the robotic printing of non-covalently bound samples on arrays,[139-140] other groups focused on covalent-immobilization of glycan samples onto slides.[141-143] The microarray technique allows high-throughput and parallel analysis of various glycan-protein interactions using only small amounts of samples in a miniaturized format. Furthermore, the ligand presentation is an important factor for carbohydrate recognition. Because monovalent carbohydrate-protein interactions are typically weak, with equilibrium dissociation constants (K_D) in the high micro- to millimolar range, many carbohydrate-binding proteins contain several binding sites or assemble into oligomers with multiple binding sites to allow simultaneous multiple protein-carbohydrate interactions to form a multivalent complex. The formation of such a complex can drastically enhance the overall affinity, which is called avidity, of these interactions, and can also have a considerable effect on the selectivity of the recognition event. As a result, the carbohydrate ligand presentation such as density, spacing and orientation of the ligands as well as the linker length and flexibility must match the geometry of the protein binding sites, and can thus have a major impact on recognition. One strategy to control the ligand presentation is to vary the glycan density by altering the average glycan spacing.[144-145] However, these approaches do not allow a precise control of the ligand presentation at a molecular level. To address this issue, alternative strategies have been developed. For example, multivalent glycoconjugates such as native glycoproteins,[146] neoglycoproteins,[147-149] neoglycopeptides,[150] glyco clusters[151], glycopolymers,[152] glycodendrimers[153-154] and glycopeptides[155] have been printed on microarrays. Because the peptide backbone also contributes to the recognition, glycopeptide microarrays have been applied to study the impact of the amino acid sequence, and of different glycosylation sites on glycan recognition.

1.4.1 Glycan and glycopeptide microarrays - applications

Since their introduction in 2002, the applications of glycan and glycoconjugate microarrays have been swiftly expanded. In particular, they have been used to analyze the binding properties of carbohydrate binding proteins (CBPs) such as lectins (e.g. plant lectins, C-type lectins, siglecs and galectins), antibodies, cells, and viral and bacterial proteins.[141, 156-163] Additionally, carbohydrate microarrays have been applied to identify new CBPs and to profile the substrate specificity of glycan processing enzymes.

The peptide backbone is often recognized by CBPs as part of the glycan binding motif, and the particular glycosylation sites can strongly impact protein binding. Glycopeptide microarrays mimic the natural glycan presentation on the cell surface and are useful tools to study, in addition to the binding specificities, the impact of glycosylation site placement and peptide sequence on CBP binding (Figure 11). For example, glycopeptide microarrays can be used in immunodiagnostics to map antibody recognition profiles (Chapters 3 and 4.4).[164-166] Glycopeptide microarrays can also be used to study enzyme substrate specificities and enzyme activities, or to determine enzymatic glycosylation sites.[167] They can also be applied to profile lectin specificities (Chapter 3) and to explore carbohydrate-pathogen interactions (Chapter 4.1).

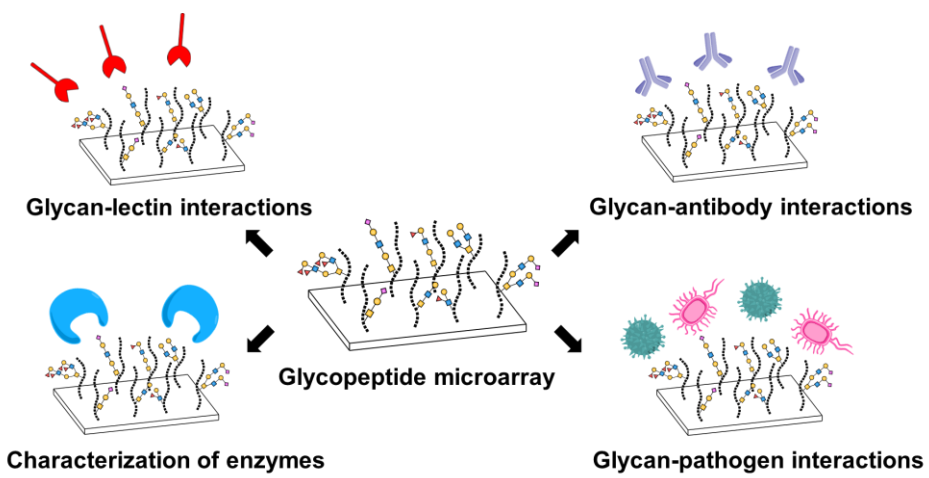


Figure 11. Applications of glycopeptide microarrays.

2 Motivation

Among a variety of post-translational modifications, glycosylation is the most abundant form, with more than 50 % of all mammalian proteins being co- or post-translationally modified by mono- or oligosaccharides. Because of their structural diversity and complexity, carbohydrates and glycoconjugates play key roles in many biological processes. Recently, *O*-glycosylation has drawn much attention in the glycoproteomic field due to its association with various diseases, such as pathogenic infections and cancer. The “mucin-type” (*O*-GalNAc-type) and the *O*-GlcNAcylation belong to the most common types of *O*-glycosylation. Glycoproteomic analysis of *O*-glycans is highly challenging due their structural diversity and complexity, and nearby and dense glycosylation sites make proteolytic cleavage and site specific analysis demanding. In contrast to *N*-glycan proteome analysis, endoglycosidases are still not available for efficient *O*-glycan release. As a result, efficient methods need to be developed to explore and analyze *O*-glycosylation to better understand its biological functions, and the roles it plays in disease. Advancements in microarray technology during the last decade have enabled the high-throughput evaluation of carbohydrate-protein interactions using only small sample amounts. However, biological samples are heterogenous regarding site-specific glycosylation and stoichiometric carbohydrate occupancy, which makes their analysis challenging.

In this work, structurally well-defined synthetic mucin glycopeptides were prepared and used to develop a glycopeptide microarray platform mimicking natural glycan presentation. These glycopeptide microarrays were used to evaluate the binding specificities of carbohydrate-binding proteins, including plant lectins, which are commonly used in affinity-based enrichment methods of glycans, bacterial lectins and human galectins, on a molecular level. Additionally, glycopeptide microarrays were used to evaluate the abilities of induced specific antibodies to detect a new group of tyrosine *O*-glycosylation, and for epitope mapping of cancer-specific serum antibodies generated from administration of synthetic vaccines. A better understanding of these carbohydrate-protein interactions on a molecular level could advance the development of carbohydrate-based inhibitors to block proteins that are involved in cancer development and progession such as galectins, or to fight bacterial

infections. Additionally, efficient carbohydrate-based cancer vaccines could be designed to elicit strong immune responses in patients.

In the first part of this thesis, tools to enable analytical and functional studies of a new group of tyrosine glycosylation were developed and evaluated (Chapter 3). In the last decade, several studies reported that *N*-acetylhexosamine (HexNAc = α-GalNAc, α- or β-GlcNAc) does not only modify serine and threonine, but also tyrosine residues. In 2011, this new type of protein *O*-glycosylation was first discovered on Tyr-681 of amyloid-β in the frame of a glycoproteomic study on CSF from Alzheimer's patients. Even though HexNAc-*O*-Tyr modifications were identified on several glycoproteins, it was not possible to conclude if the identified Tyr modified peptides were of the *O*-GlcNAc- or the mucin-type due to lack of efficient methods to assign gylcoforms and glycosidic linkages. The fact that Tyr-*O*-HexNAcylation exists in addition to Ser and Thr modifications raises questions about their biosynthesis, expression and biological functions, which are unexplored to a large extent. This highlights the need to develop novel tools to specifically detect, identify and enrich *O*-GalNAc- and *O*-GlcNAc-Tyr modifications, and to evaluate the ability of lectin-based enrichment methods to detect these new modifications. The aims of this project were to evaluate the ability of plant lectins commonly used in lectin weak affinity chromatography in glycoproteomic work flows, and to raise and evaluate HexNAc-*O*-Tyr-specifc antibodies generated from immunization of rabbits with synthetic vaccines to detect and identify HexNAc-*O*-Tyr modifications on glycopeptides and -proteins.

In the second part of this thesis, glycopeptide microarrays were used to evaluate binding specificities of carbohydrate-binding proteins, including bacterial lectins and human galectins, and to map epitopes of cancer-specific serum antibodies generated from administration of synthetic vaccines to mice. During infection, pathogenic bacteria cause virulence by adhering to glycans on membrane-bound mucins via bacterial lectins. To fight bacterial infections, good knowledge of the fine binding spficficities of these lectins is essential to design efficient glycomimetic inhibitors. One aim of this project was, to employ glycopeptide microarrays to explore the roles of different terminal fucose motifs, and of ligand presentation on different glycosylation sites of the peptide backbone in bacterial lectin recognition events of the *Pseudomonas aeruginosa* lectin LecB and the *Clostridium difficile* toxin A (Chapter 4.1). Aberrant glycosylation of membrane-bound mucins is also

associated with various cancer types. In carcinomas, mucins are overexpressed and the glycan structures attached to the mucins are altered due to changes in the expression levels of various glycosyltransferases, which leads to the formation of tumor associated carbohydrate antigens. The terminal LacdiNAc motif is expressed on *N*- and *O*-glycans and has been associated in addition to a possible role in *Helicobacter pylori* infection with many cancer types. However, the exact roles that LacdiNAc *O*-glycosylation plays in these diseases are not clear. Another aim of this project was to develop a synthesis strategy to prepare simplified mucin core glycopeptides that could be further enzymatically modified to generate LacdiNAc modified glycopeptides (Chapter 4.2). The glycopepides could be used in microarray binding studies to evaluate the glycan binding preferences of various proteins, including the *H. pylori* lectin LabA and human galectins, which play roles in cancer development and progression. Galectins are thus interesting targets for the development of therapeutic cancer treatments such as carbohydrate-based inhibitors that block glycan-galectin interactions. An additional aim of this project was to evaluate the fine binding specifcities of different human galectins using glycopeptide microarrays (Chapter 4.3). A better understanding of how galectins interact with host glycans on a molecular level could be useful in the generation of efficient galectin inhibitors, and may improve our comprehension of the roles that individual galectins play in their respective biological processes. Since cancer is a leading cause of death worldwide, there is need for novel cancer therapeutics. Because MUC1 is overexpressed in many cancers, it is an attractive antigenic target for the development of effective anti-cancer vaccines. The aim of this project was to use glycopeptide microarrays for epitope mapping of cancer-specific serum antibodies generated from immunization of mice with synthetic vaccines in order to evaluate the applicability of these vaccines in cancer treatment (Chapter 4.4).

More detailed motivations are given in each chapter of this thesis.

3 Project 1: Development and Evaluation of Tools to Explore HexNAc-*O*-Tyrosine Glycosylation (Papers I and II)

3.1 Introduction - HexNAc-*O*-Tyr: a new PTM

3.1.1 HexNAc-*O*-Tyr modification on glycoproteins

In the last decade, several studies reported on a new group of PTMs, where *N*-acetylhexosamine (HexNAc = α-GalNAc, α- or β-GlcNAc) residues are attached to the hydroxyl group of the amino acid tyrosine (HexNAc-*O*-Tyr). In 2011, this type of protein *O*-glycosylation was first identified to modify Tyr-681 in the amyloid-β peptide DAEFRHDSGYEVHHQK in the frame of a glycoproteomic study on CSF from Alzheimer's patients.[168] Further glycoproteomic studies led to the discovery of HexNAc-*O*-Tyr on several other proteins including Nucleobindin-1 (NUCB1), Nucleobindin-2 (NUCB2), the Extracellular matrix protein 1 (ECM1), CD44 and the proline-rich acidic protein-1 (PRAP1).[169-170] Additionally, glycoproteomic studies of isolated mitochondrial proteins led to the identification of HexNAc-*O*-Tyr on the ATP synthase subunit beta (ATP5B), the voltage-dependent anion-selective channel protein 1 (VDAC1) and aspartate aminotransferase (ASAT).[171] Another study showed that the *Photorhabdus asymbiotica* protein toxin ($PaTox^G$) was found to modify a tyrosine residue in the switch II region of host GTPases such as RhoA, Rac and Cdc42 with α-GlcNAc.[172-173] In most of these studies it was not possible to conclude if the identified Tyr modified peptides were of the *O*-GlcNAc- or the mucin-type and the glycoforms were assigned based on general knowledge about carbohydrate biosynthesis, tissue-specific expression of glycosyltransferases, or lectin binding specificities. Since serine and threonine residues are usually considered to be the expected amino acid acceptors on potential glycosylation sites, tyrosine was consequently not previously considered as a potential acceptor amino acid. The fact that Tyr-*O*-HexNAcylation exists in addition to the common Ser and Thr modifications raises questions about their biosynthesis, expression and biological functions, which are unexplored to a large extend so far.

3.1.2 Plant lectins in bioanalytical applications

Lectins are carbohydrate binding proteins that recognize and bind to specific carbohydrate motifs via their carbohydrate recognition domains (CRDs), and both mammalian and plant lectins are commonly used as tools to study glycosylation due to their ability to discriminate between different glycan structures. Consequently, lectins are frequently used in glycoproteomic work flows to detect, identify and/or enrich glycoproteins or tryptic glycopeptides.[171, 174-176]

In the Chapter 3.1.1 mentioned studies, the HexNAcylation types were often defined according to the binding specificities of the respective lectins used for lectin weak affinity chromatography (LWAC): For example, the *Vicia villosa* lectin (VVA), which specifically recognizes GalNAc,[177] and *Wheat germ agglutinin* (WGA), which has been reported to recognize GlcNAc, GalNAc and also NeuNAc modifications.[178-180] Because VVA recognizes both α- and β-GalNAc, a proper assignment of the linkages was not possible. Additionally, since WGA recognizes both GlcNAc and GalNAc residues, it could not be deduced weather the identified tyrosine HexNAcylations were of the *O*-GlcNAc- or the mucin-type. The lack of knowledge about HexNAc-*O*-Tyr lectin recognition, and of efficient methods to determine the exact structures and linkages of the attached carbohydrate moieties makes the study of *O*-HexNAc on tyrosine extremely challenging.

In this work, the plant lectins WGA and VVA were used to explore lectin recognition of tyrosine HexNAcylations. Additionally, the ability of *Griffonia simplicifolia* lectin II (GSL II) to recognize GlcNAc-*O*-Tyr was evaluated. GSL II binds to the non-reducing terminal GlcNAc residues and is usually used to profile glycans,[181-182].

3.2 Motivation

Although HexNAc-*O*-Tyr modifications were identified on several glycoproteins, our knowledge about their biological roles, biosynthesis and expression is limited. Usually, serine and threonine are the expected glycosylation sites. Therefore, tyrosine residues were previously not considered as potential glycosylation sites, and the existence of this new group of PTMs raises questions about their biological functions, which are unexplored so far. A recent study reported that several plant lectins used in LWAC detected α- and β-GalNAc-*O*-Tyr. However, these lectins also recognize *O*-GalNAc on Ser and Thr. The lack of efficient methods to specifically detect and enrich these modifications, and to efficiently discriminate between the attached carbohydrate residues hinders the study of *O*-HexNAcylation on tyrosine. This highlights the need to develop novel tools to specifically detect, identify and enrich *O*-GalNAc- and *O*-GlcNAc-Tyr modifications, and to evaluate already established glycoproteomic methods such as lectin weak affinity chromatography regarding their applicability to detect these new modifications.

The aims of this project were to evaluate the ability of i) plant lectins commonly used in LWAC in glycoproteomic work flows to detect and identify HexNAc-*O*-Tyr modifications, and ii) *O*-GalNAc- and *O*-GlcNAc-Tyr-specific antibodies generated from immunization of rabbit with synthetic vaccines for selective detection and identification of HexNAc-*O*-Tyr modifications on glycopeptides and glycoproteins in order to gain new insights into the biological roles of Tyr *O*-HexNAcylations.

To enable these studies, an extensive and structurally well-defined library of mono- and bivalent synthetic α-GalNAc, α- and β-GlcNAc-O-Tyr, -Ser and -Thr glycopeptides was prepared and subsequently immobilized on NHS-modified microarray slides. The glycopeptide microarrays were then applied to evaluate potential recognition of tyrosine HexNAcylation by the plant lectins VVA, WGA and GSL II, which are commonly used to detect, identify and/or enrich glycopeptides in glycoproteomic studies. Therefore, lectin binding profiles and avidities toward glycopeptides presenting different HexNAc isoforms on Ser, Thr or Tyr were determined and apparent surface K_D values were calculated. Additionally, the impact of bivalent ligand presentation on lectin binding was explored.

Furthermore, the glycopeptide microarray library was used to evaluate the binding specificities of HexNAc-specific polyclonal antibodies generated to selectively detect peptides and proteins carrying this new

PTM. To induce specific antibodies against this tyrosine modification, rabbits were immunized with antigen glycopeptide-CRM conjugates designed to exhibit a high α-GalNAc- or β-GlcNAc-*O*-Tyr density. The obtained rabbit sera were immunologically analyzed and evaluated by enzyme-linked immunosorbent assay (ELISA). Additionally, the obtained HexNAc-*O*-Tyr specific antibodies were probed to selectively detect this modification on protein level. Therefore, the rabbit antisera were affinity purified against GalNAc- and GlcNAc-*O*-Tyr antigen peptides, re-evaluated by ELISA and microarray assays and finally used to specifically detect RhoA modified with α-GlcNAc-*O*-Tyr on a Western blot.

3.3 Results and discussion

3.3.1 Generation of the HexNAc-*O*-Ser/Thr/Tyr glycopeptide library

A peptide library consisting of structurally well-defined synthetic nonglycosylated, monovalent and divalent glycosylated peptide sequences carrying α-GalNAc, α- or β-GlcNAc on serine, threonine or tyrosine was prepared by Dr. Manuel Schorlemer and the author to evaluate the binding preferences of the plant lectins WGA, VVA and GSL II, as well as of HexNAc-*O*-Tyr-specific rabbit antibodies (Figure 12). Therefore, glycosylated amino acid building blocks were assembled according to our reported synthesis procedures using common acceptor amino acids and corresponding glycosyl donors.[183-184] The obtained glycosylated amino acids were subsequently incorporated into Fmoc-SPPS to prepare a GalNAc/GlcNAc-Tyr/Thr/Ser library containing 86 (glyco)peptides (Appendix Table S1). This library included the synthetic peptide sequences shown in Figure 12, which were previously identified by LWAC and LC/MS as tryptic glycopeptide fragments that were modified with HexNAc-*O*-Tyr. Additionally, glycopeptides identified with nearby Ser or Thr sites were prepared including glycan isomer analogs modified at specific, or nearby potential glycosylation sites. Furthermore, MUC1 tandem repeat sequence analogs were synthesized by replacing known Ser/Thr sites with HexNAc-*O*-Tyr.

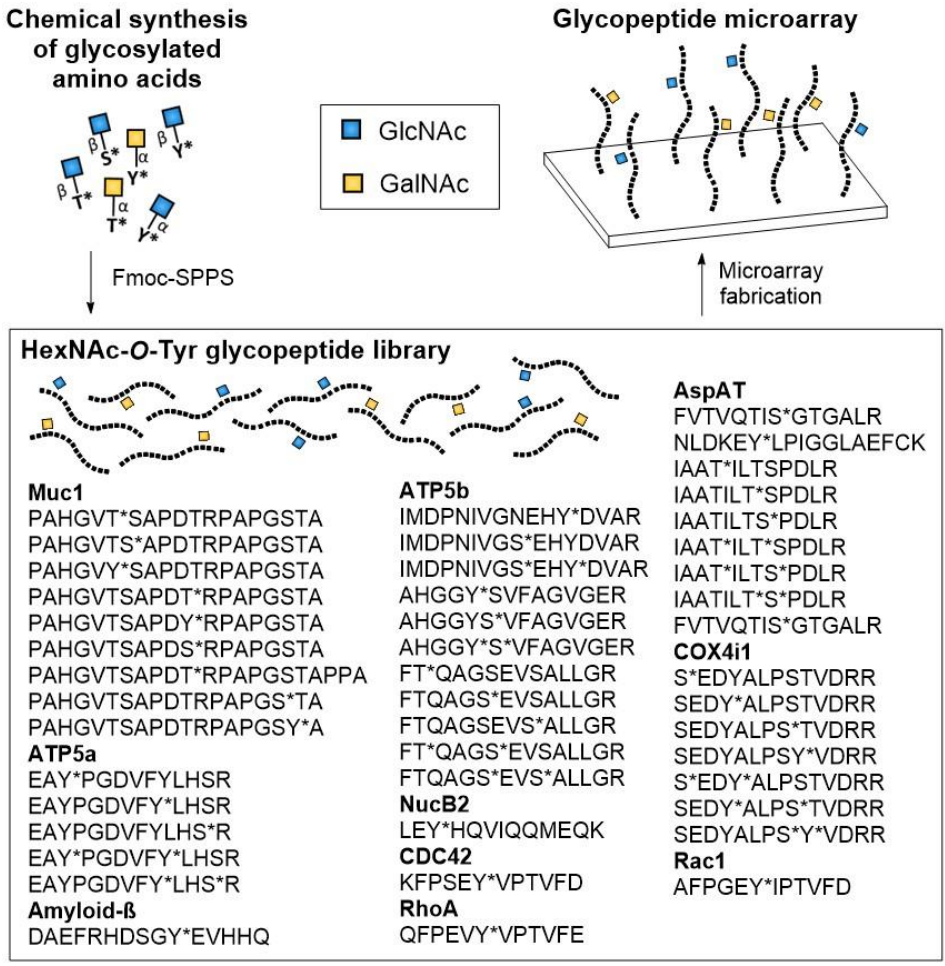


Figure 12. Synthesis of glycosylated amino acids, glycopeptides and a glycopeptide microarray library.

3.3.2 Evaluation of plant lectins using glycopeptide microarrays

The glycopeptide library was then used to evaluate the abilities of the plant lectins VVA, WGA and GSL II to detect and identify HexNAc-*O*-Tyr modifications at the glycopeptide level. The binding preferences of these plant lectins regarding HexNAcylation on Ser and Thr residues are well known, but besides one study that used VVA to detect a few α-GalNAc-*O*-Tyr tripeptides, their binding specificities toward HexNAc-*O*-Tyr modifications have not been explored. Therefore, the microarrays were incubated with a dilution series of the respective biotinylated lectins (VVA: 0.025 – 400 µg/mL, 15 concentrations), GSL II: 0.05 – 12.5 µg/mL, 9 nine concentrations, WGA: 0.003 – 400 µg/mL, 15 concentrations) and, subsequently, with Cy5-streptavidin for fluorescent detection. The lectin binding preferences toward the different glycopeptide epitopes as well as the surface dissociation constants (Surf. K_D) were determined. Curve fitting was carried out by non-linear regression using the saturation binding − specific binding with Hill Slope equation in GraphPad Prism 8. Surprisingly, WGA showed a different binding pattern for some of the glycopeptides that indicated a secondary binding event that might be caused by different carbohydrate binding modes of the two distinct WGA binding sites types.[185] The respective sample data were fitted using the saturation binding − specific binding for two sites in GraphPad Prism 8.

VVA, WGA and GSL II showed strong binding to HexNAc-O-Tyr modified glycopeptides and none of the unglycosylated peptides were recognized by any lectin as shown in Figure 13 and Figure 14. Microarray analysis showed that VVA specifically bound to α-GalNAc on both Thr and Tyr in a low nanomolar range with a preference for Tyr- over Thr-GalNAcylation (Figure 13, Figure 14, Table 1). As expected, α- and β-GlcNAc modifications on either Ser, Thr or Tyr were not recognized. Additionally, VVA binding to the MUC1 glycopeptides **P6-P20** was influenced by the different glycosylation sites indicating that the amino acid sequence also plays a role in lectin recognition.

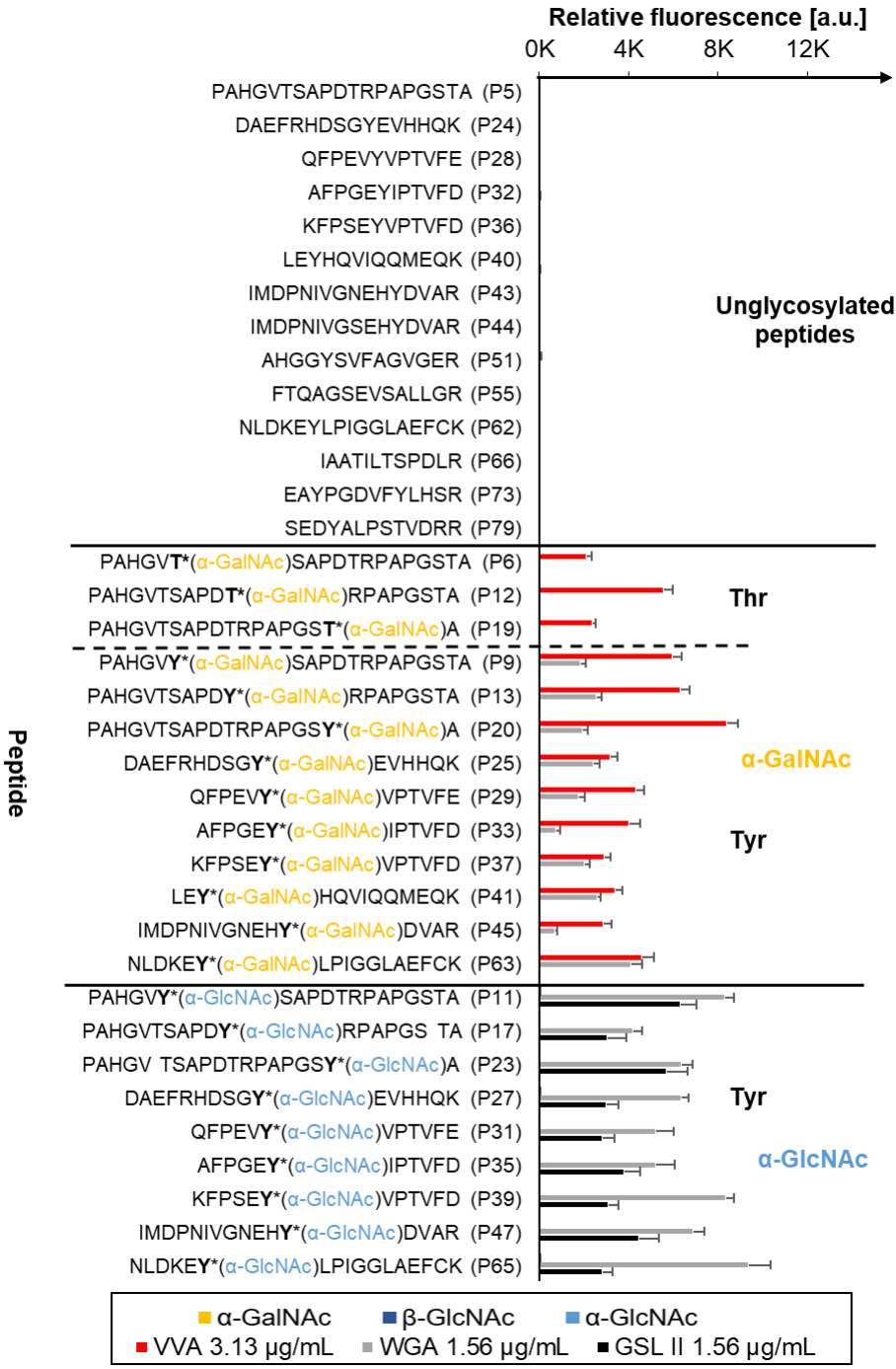


Figure 13. Binding of VVA, WGA and GSL II toward unglycosylated, α-GalNAc-Thr/Tyr and α-GlcNAc-*O*-Tyr glycopeptides. [a.u.] = arbitrary units.

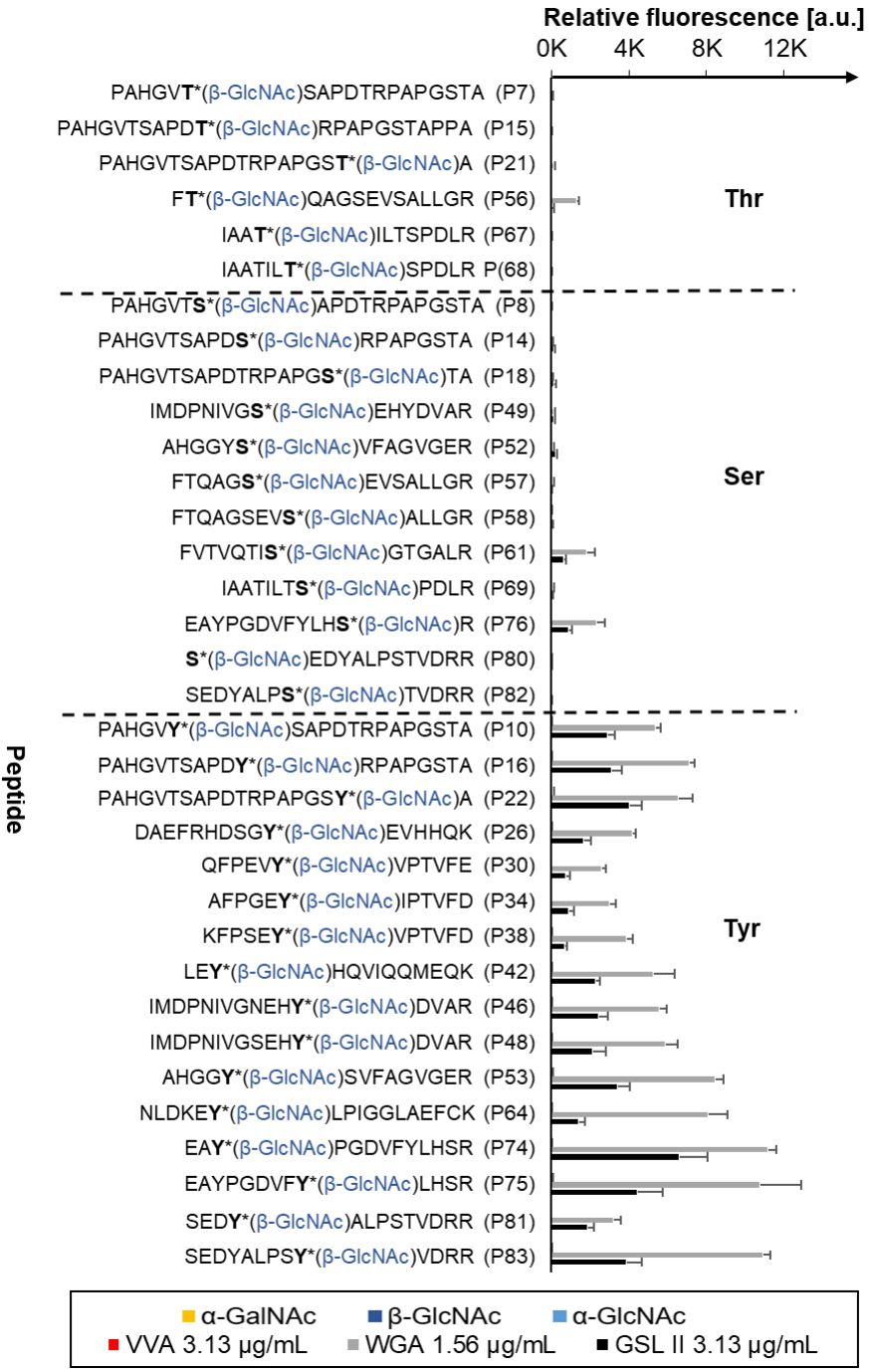


Figure 14. Binding of VVA, WGA and GSL II toward β-GlcNAc-*O*-Thr/Ser/Tyr glycopeptides. [a.u.] = arbitrary units.

Table 1. Surface dissociation constants (Surf. $_{KD}$ values, Mean ± SEM in nM) for microarray binding study of VVA. T/Y* = α-GalNAc.

ID	Glycopeptide sequence	K_D ± SEM [nM]
P6	PAHGVT*SAPDTRPAPGSTA	33.5 ± 3.0
P9	PAHGVY*SAPDTRPAPGSTA	20.1 ± 1.3
P12	PAHGVTSAPDT*RPAPGSTA	27.0 ± 1.7
P13	PAHGVTSAPDY*RPAPGSTA	20.4 ± 1.4
P19	PAHGVTSAPDTRPAPGST*A	31.8 ± 5.1
P20	PAHGVTSAPDTRPAPGSY*A	15.1 ± 1.0
P25	DAEFRHDSGY*EVHHQK	33.9 ± 2.7
P29	QFPEVY*VPTVFE	33.6 ± 1.5
P33	AFPGEY*IPTVFD	33.5 ± 19.5
P37	KFPSEY*VPTVFD	41.2 ± 2.4
P41	LEY*HQVIQQMEQK	29.3 ± 1.5
P45	IMDPNIVGNEHY*DVAR	30.0 ± 2.1
P63	NLDKEY*LPIGGLAEFCK	36.4 ± 1.8

WGA recognized, in line with its reported binding specificities, all carbohydrate isoforms (Figure 13, Figure 14). It exhibited a strong preference for HexNAc modifications on Tyr over the respective Ser and Thr analogs. These findings can be explained by the establishment of additional π-π- and CH-π interactions between the aromatic system of the tyrosine residue and the hydrophobic WGA binding pocket. Furthermore, α-GlcNAc-*O*-Tyr glycopeptides were better binders than the corresponding β-GlcNAc- or α-GalNAc-*O*-Tyr analogs with α-GalNAc structures being the weakest binders (Table 2). Additionally, WGA recognition of GlcNAcylation on Thr, Ser and Tyr residues was strongly influenced by their particular glycosylation sites. These findings imply that WGA binding affinity also depends on the peptide sequence. While HexNAc-*O*-Tyr glycopeptides were recognized by WGA without exception, surprisingly not all α-GalNAc-Thr and β-GlcNAc-Ser/Thr glycopeptides were detected by WGA. The WGA-β-GlcNAc-Tyr interactions were less dependent on the peptide sequence, which is likely due to the further extension of the β-GlcNAc-Tyr residue from the peptide backbone compared with Ser or Thr binding sites. Consequently, β-GlcNAc-Tyr modifications are better accessible for the WGA binding pocket. Interestingly, many binding curves of the α- and β-GlcNAc glycopeptides showed a secondary binding event, which might be caused by different multivalent binding modes of the lectin. Further binding studies should be performed to explore this in more detail.

Table 2. Surface dissociation constants (Surf. $_{KD}$ values, Mean ± SEM in nM) for microarray binding of WGA. **T/Y*** = α-GalNAc**,** **Y*** = α-GlcNAc, **S/T/Y*** = β-GlcNAc.

ID	Glycopeptide sequence	K_{D1} ± SEM [nM]	K_{D2} ± SEM [nM]
P10	PAHGV**Y***SAPDTRPAPGSTA	96.9 ± 13.7	0.50 ± 0.09
P11	PAHGV**Y***SAPDTRPAPGSTA	59.4 ± 12.0	0.60 ± 0.27
P13	PAHGVTSAPD**Y***RPAPGSTA	111.8 ± 7.6	
P16	PAHGVTSAPD**Y***RPAPGSTA	137.1 ± 36.2	0.75 ± 0.08
P17	PAHGVTSAPD**Y***RPAPGS TA	75.5 ± 12.4	0.10 ± 0.46
P20	PAHGVTSAPDTRPAPGS**Y***A	180.7 ± 13.5	
P22	PAHGVTSAPDTRPAPGS**Y***A	62.9 ± 12.1	0.62 ± 0.10
P23	PAHGV TSAPDTRPAPGS **Y***A	66.3 ± 9.2	0.49 ± 0.37
P25	DAEFRHDSG**Y***EVHHQK	64.7 ± 3.7	
P26	DAEFRHDSG**Y***EVHHQK	88.0 ± 23.0	0.42 ± 0.08
P27	DAEFRHDSG**Y***EVHHQK	77.6 ± 16.1	0.56 ± 0.10
P30	QFPEV**Y***VPTVFE	197.7 ± 18.7	
P31	QFPEV**Y***VPTVFE	126.3 ± 8.6	
P34	AFPGE**Y***IPTVFD	87.5 ± 11.8	
P35	AFPGE**Y***IPTVFD	68.7 ± 6.4	
P38	KFPSE**Y***VPTVFD	117.0 ± 11.3	0.41 ± 0.06
P39	KFPSE**Y***VPTVFD	103.6 ±12.3	0.73 ± 0.13
P41	LE**Y***HQVIQQMEQK	64.3 ± 4.0	
P42	LE**Y***HQVIQQMEQK	78.1 ± 32.7	0.64 ± 0.11
P46	IMDPNIVGNEH**Y***DVAR	80.2 ± 8.2	0.50 ± 0.13
P47	IMDPNIVGNEH**Y***DVAR	63.1 ± 6.9	0.57 ± 0.41
P48	IMDPNIVGSEH**Y***DVAR	74.6 ± 8.7	0.48 ± 0.14
P50	IMDPNIVG**S***EH**Y***DVAR	110.7 ± 19.1	0.74 ± 0.17
P53	AHGG**Y***SVFAGVGER	125.0 ± 20.5	0.45 ± 0.05
P54	AHGG**Y*S***VFAGVGER	63.5 ± 6.2	0.25 ± 0.08
P59	F**T***QAG**S***EVSALLGR	73.1 ± 4.3	
P61	FVTVQTI**S***GTGALR	173.7 ± 23.0	
P63	NLDKE**Y***LPIGGLAEFCK	71.6 ± 3.8	
P64	NLDKE**Y***LPIGGLAEFCK	152.3 ± 58.9	0.60 ± 0.10
P65	NLDKE**Y***LPIGGLAEFCK	148.3 ± 35.5	0.88 ± 0.16
P70	IAA**T***IL**T***SPDLR	228.9 ± 11.28	
P71	IAA**T***ILT**S***PDLR	143.5 ± 11.1	
P72	IAATIL**T*S***PDLR	426.6 ± 42.5	
P74	EA**Y***PGDVFYLHSR	109.3 ± 25.8	0.65 ± 0.07
P75	EAYPGDVF**Y***LHSR	477.9 ± 2389.5	0.40 ± 0.14
P76	EAYPGDVFYLH**S***R	126.3 ± 11.2	
P77	EA**Y***PGDVF**Y***LHSR	1.1 ± 0.1	
P78	EAYPGDVF**Y***LH**S***R	1.6 ± 0.1	
P81	SED**Y***ALPSTVDRR	62.7 ± 3.9	
P83	SEDYALPS**Y***VDRR	76.5 ± 38.2	0.78 ± 0.16
P84	**S***ED**Y***ALPSTVDRR	110.7 ± 9.3	0.54 ± 0.12
P85	SED**Y***ALP**S***TVDRR	1.3 ± 0.1	
P86	SEDYALP**S*Y***VDRR	633.5 ± 265.4	1.21 ± 0.20

The binding avidities of WGA for different bivalent glycopeptides were evaluated and found to be frequently enhanced compared with the avidities observed for the corresponding monovalent analogs (Figure 15, Table 2). Again, a secondary binding event could be observed for WGA binding to a few of the bivalent glycopeptides. Depending on the fitted curve shapes of the fit for the two binding sites, it was assumed that K_{D1} contributed to overall avidity to a greater extent than K_{D2}. Consequently, K_{D1} was used to compare WGA binding to mono- and bivalent glycopeptides. Microarray analysis and Surf. K_D value determination showed that WGA binding was often enhanced for bivalent peptides. For example, the bivalent peptide **P78** EAYPGDVF**Y***LH**S***R (Surf. K_D = 1.6 nM) showed a higher binding affinity in comparison to the respective monovalent peptides **P75** EAYPGDVF**Y***LHSR (Surf. K_D = 478 nM) and **P76** EAYPGDVFYLH**S***R (Surf. K_D = 126 nM). However, the exact placement of HexNAc-glycosylation sites strongly influenced lectin binding and adjacent glycosylation sites could sterically hinder lectin binding leading to a derease in the overall avidity. For example, WGA binding to the bivalent glycopeptide **P72** IAATIL**T***__S__*PDLR (Surf. K_D = 427 nM) was dramatically increased compared to the recognition of the corresponding monovalent glycopeptides **P68** IAATIL**T***SPDLR and **P69** IAATILT**S***PDLR. However, the adjacent placement of the Thr and Ser glycosylation sites of **P72** sterically hindered lectin binding and the peptide showed decreased binding avidities compared with the other bivalent analogs **P70** IAA**T***IL**T***SPDLR (Surf. K_D = 229 nM) and **P71** IAA**T***ILT**S***PDLR (Surf. K_D = 144 nM). The same phenomenon was also observed for the bivalent glycopeptide **P86** SEDYALP**S***__Y__*VDRR (Surf. K_D = 634 nM). Interestingly, WGA showed enhanced binding to the bivalent peptide **P54** AHGG**Y***__S__*VFAGVGER (Surf. K_D = 64 nM) in comparison with the monovalent peptides **P52** AHGGY**S***VFAGVGER and **P53** AHGG**Y***SVFAGVGER (Surf. K_D = 229 nM), even though the glycosylation sites were situated next to each other. While WGA binding to bivalent glycopeptides containing only Ser and Thr glycosylation sites was generally increased in comparison to the affinities observed for the corresponding monovalent derivatives, WGA recognition of bivalent glycopeptides carrying a glycan on a Tyr residue depended on the exact placement of both glycosylations sites in relation to each other. For example, bivalent ligand presentation on peptides **P85** SED**Y***ALP**S***TVDRR (Surf. K_D = 1.3 nM) and **P86** SEDYALP**S***__Y__*VDRR

(Surf. K_D = 634 nM) either enhanced or decreased WGA binding compared with the monovalent peptides **P80** SED**Y***ALPSTVDRR (Surf. K_D = 1.3 nM) and **P83** SEDYALPS**Y***VDRR (Surf. K_D = 77 nM). These findings indicate that the exact placement of the glycosylation sites in the amino acid sequence strongly impacts WGA binding and that appropriate spacing between the two glycosylation is crucial for optimal peptide backbone ligand presentation and lectin recognition.

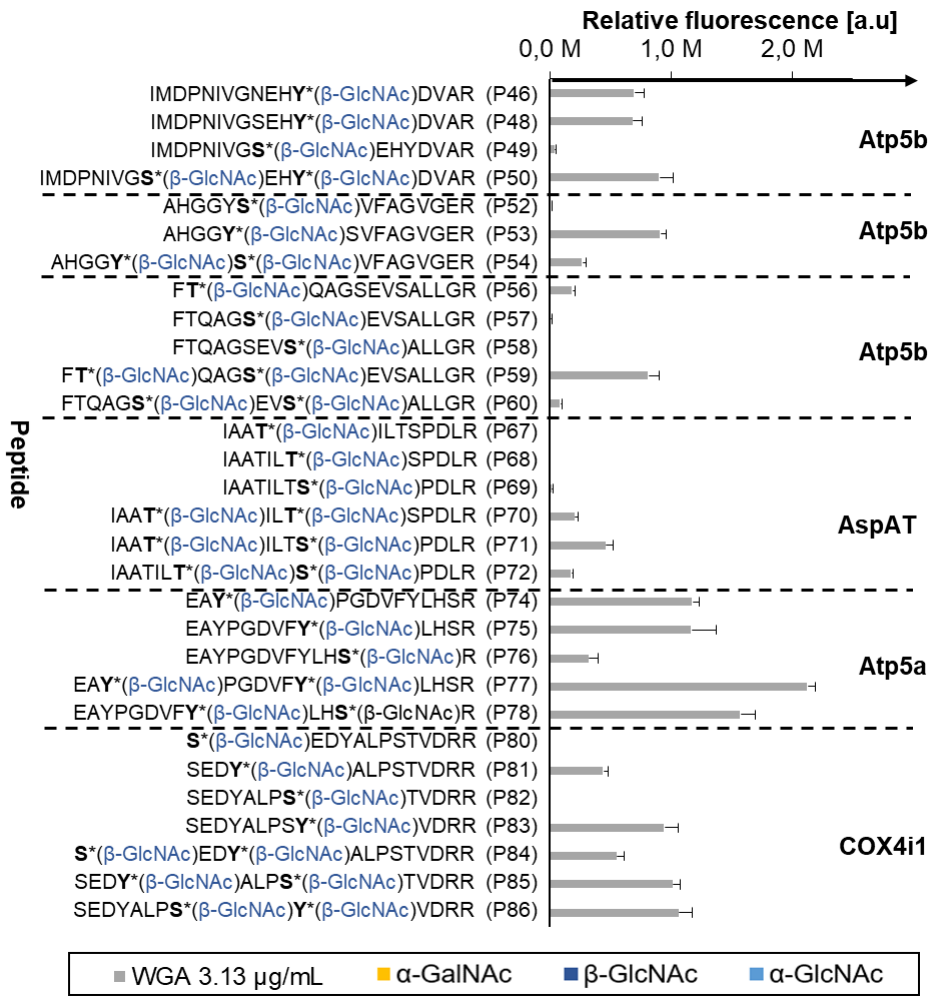


Figure 15. Contribution of bivalent ligand presentation to WGA glycopeptide recognition. [a.u.] = arbitrary units.

As expected, GSL II recognized both α-GlcNAc and β-GlcNAc glycopeptides, but none of the α-GalNAc analogs (Figure 13, Figure 14). Generally, no binding preference to either of the GlcNAc anomers could be observed, although, the lectin preferably bound to *O*-GlcNAcylated Tyr over Thr/Ser. GSL II bound to GlcNAc-O-Tyr modified glycopeptides in a low nanomolar range (Table 3). As already observed for WGA, GSL II detected all peptides with O-GlcNAcylated tyrosine, but many of the GlcNAc-O-Thr/Ser peptides were weakly recognized, or not recognized at all, and no Surf. K_D values could be determined for these samples. Additionally, no secondary binding events were observed for GSL II binding to any of the glycopeptides, indicating that the modes of binding are different for the two plant lectins.

The impact of bivalent ligand presentation on the binding strength was also evaluated for GSL II (Figure 16). As observed for WGA recognition, the GSL II binding depended on the respective glycosylation sites and their spacing. However, the differences in binding avidities for the different bi- and monovalent glycopeptides were not as dramatic as for WGA since GSL II bound to all *O*-GlcNAc peptides in a low nanomolar range (Surf. K_D = 2.9 - 24.6 nM). For example, the bivalent peptide **P54** AHGG**Y***-**S***VFAGVGER (Surf. K_D = 10.3 nM) was again a better binder than the corresponding monovalent peptide **P52** AHGGY**S***VFAGVGER. However, its affinity was slightly lower in contrast to the monovalent tyrosine peptide **P53** AHGG**Y***SVFAGVGER (Surf. K_D = 6.8 nM) because of sterical hindrance cause by the neighboring placement of the two glycosylation sites. In contrast to the WGA data, GSL II showed slightly enhanced binding to the bivalent peptide **P86** SEDYALP**S*Y***VDRR (Surf. K_D = 7.0 nM) in comparison to the other bivalent peptides **P84** **S***ED**Y***ALPSTVDRR (Surf. K_D = 14.2 nM) and **P85** SED**Y***ALP**S***TVDRR (Surf. K_D = 13.3 nM) as well as to the respective monovalent peptide **P83** SEDYALPS**Y***VDRR (Surf. K_D = 9.9 nM). Also, an additional Ser glycosylation site in peptides containing a glycosylated Tyr residue did not strongly affect GSL II binding. For example, GSL II showed similar recognition for the bivalent peptides **P77** EA**Y***PGDVF**Y***LHSR (Surf. K_D = 2.9 nM) and **P78** EAYPGDVF**Y***LH**S***R (Surf. K_D = 6.5 nM), and for the respective monovalent peptides **P74** EA**Y***PGDVFYLHSR (Surf. K_D = 4.4 nM) and **P75** EAYPGDVF**Y***LHSR (Surf. K_D = 6.4 nM), indicating that the Tyr glycosylation site plays a more important role in lectin binding events than the Ser site.

These results indicate that the exact placement and spacing of the glycosylation sites in the amino acid sequence, and the respective acceptor amino acid influence GSL II. The impact on the binding strength is not as severe as observed for WGA binding.

Table 3. Surface dissociation constants (Surf. $_{KD}$ values, Mean ± SEM in nM) for microarray binding study of GSL II.

ID	Glycopeptide sequence	K_D ± SEM [nM]
P10	PAHGV**Y*****SAPDTRPAPGSTA	11.3 ± 2.
P11	PAHGVY*SAPDTRPAPGSTA	7.3 ± 0.8
P16	PAHGVTSAPD**Y***RPAPGSTA	13.2 ± 2.7
P17	PAHGVTSAPDY*RPAPGS TA	9.7 ± 1.4
P22	PAHGVTSAPDTRPAPGS**Y***A	7.1 ± 0.6
P23	PAHGV TSAPDTRPAPGSY*A	4.6 ± 0.5
P26	DAEFRHDSG**Y***EVHHQK	5.1 ± 0.5
P27	DAEFRHDSGY*EVHHQK	3.9 ± 0.3
P31	QFPEVY*VPTVFE	9.8 ± 1.7
P34	AFPGE**Y***IPTVFD	7.5 ± 0.5
P35	AFPGEY*IPTVFD	19.4 ± 3.9
P38	KFPSE**Y***VPTVFD	8.0 ± 0.6
P39	KFPSEY*VPTVFD	14.6 ± 3.1
P42	LE**Y***HQVIQQMEQK	8.2 ± 0.8
P46	IMDPNIVGNEH**Y***DVAR	11.6 ± 2.1
P47	IMDPNIVGNEHY*DVAR	10.7 ± 1.5
P48	IMDPNIVGSEH**Y***DVAR	10.9 ± 1.4
P50	IMDPNIVG**S***EH**Y***DVAR	12.4 ± 1.5
P53	AHGG**Y***SVFAGVGER	6.8 ± 0.9
P54	AHGG**Y*S***VFAGVGER	10.3 ± 2.0
P64	NLDKE**Y***LPIGGLAEFCK	12.0 ± 1.7
P65	NLDKEY*LPIGGLAEFCK	24.6 ± 3.6
P74	EA**Y***PGDVFYLHSR	4.4 ± 0.6
P75	EAYPGDVF**Y***LHSR	6.4 ± 1.1
P76	EAYPGDVFYLH**S***R	15.2 ± 3.8
P77	EA**Y***PGDVF**Y***LHSR	2.9 ± 0.2
P78	EAYPGDVF**Y***LH**S***R	6.5 ± 1.0
P81	SED**Y***ALPSTVDRR	11.8 ± 1.0
P83	SEDYALPS**Y***VDRR	9.9 ± 1.5
P84	**S***ED**Y***ALPSTVDRR	14.2 ± 1.5
P85	SED**Y***ALP**S***TVDRR	13.3 ± 1.5
P86	SEDYALP**S*Y***VDRR	7.0 ± 0.8

Y* = α-GlcNAc, S/T/Y* = β-GlcNAc

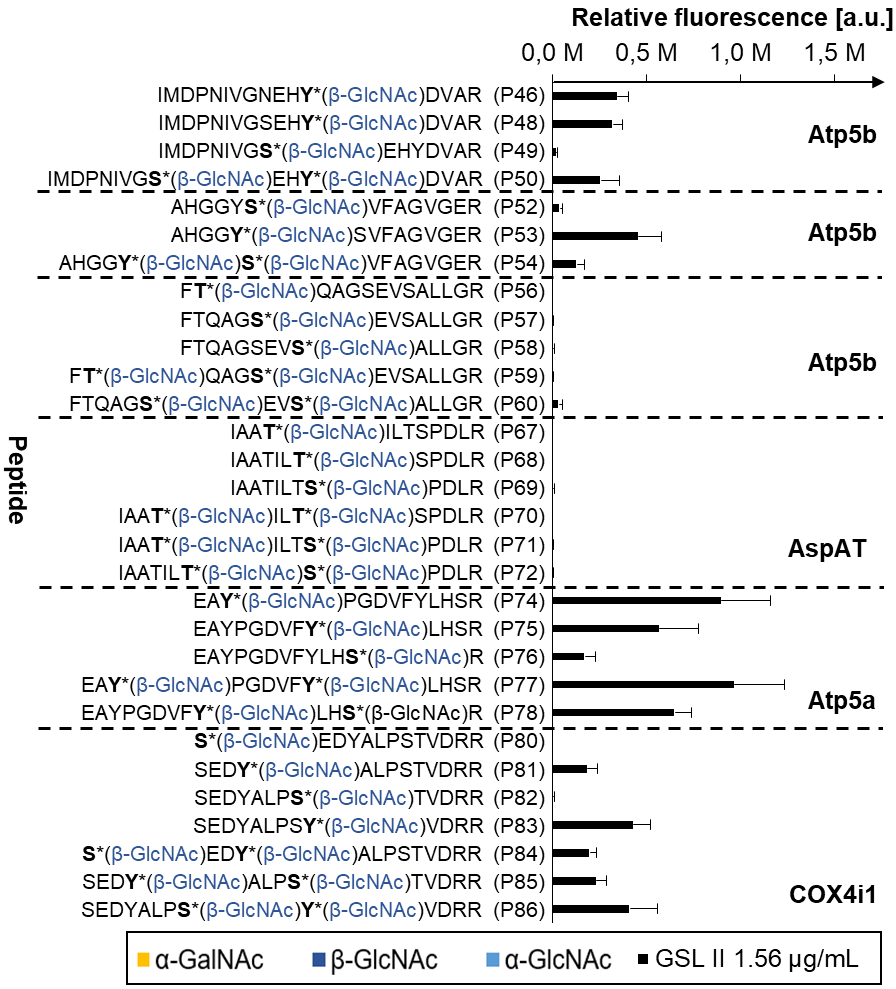


Figure 16. Contribution of bivalent ligand presentation to GSL II glycopeptide recognition. [a.u.] = arbitrary units.

In conclusion, microarray binding analyses of the plant lectins VVA, WGA and GSL II indicate that HexNAc-*O*-Tyr glycopeptides can be efficiently detected and captured from biological samples in bioanalytical applications by all three lectins depending on their respective binding specificities. Nonetheless, quantitative enrichment could be dependent on the respective acceptor amino acid the HexNAc-ligand is attached to, the placement, spacing and steric accessibility of the glycosylation sites, especially when more than one is present on the same peptide backbone. However, these lectins also detect HexNAcylation on Ser and Thr

glycosylation sites as well. Consequently, new tools to specifically detect, identify and enrich this new PTM need to be developed to gain insights into the biological roles and biosynthesis of this new tyrosine modification. In the next part of this project, HexNAc-*O*-Tyr-specific polyclonal antibodies were generated to specifically detect this novel modification.

3.3.3 Synthesis of HexNAc-*O*-Tyr antigen peptide-CRM vaccine conjugates

To generate HexNAc-*O*-Tyr specific antibodies, α-GalNAc-*O*-Tyr and β-GlcNAc-*O*-Tyr antigen glycopeptide-CRM vaccine conjugates were generated by coupling antigen glycopeptides to the carrier protein CRM^{127} via a non-immunogenic linker. Mice were then immunized with the vaccines and the obtained antisera were immunologically analyzed by ELISA and microarray assays. The glycopeptide antigens were coupled to BSA to generate the corresponding BSA conjugates for ELISA antibody endpoint titer determination.

To construct antigen glycopeptide-CRM vaccine conjugates **CRM-1** and **-2**, glycopeptides **P2** and **P3**, which serve as B-cell epitopes and consist of the short peptide sequence GYYA that is glycosylated on Tyr with α-GalNAc or β-GlcNAc, respectively, were first coupled to diethyl squarate, followed by conjugation to CRM^{197}, which was used as carrier protein to elicit strong immune responses (Figure 17). For ELISA antibody endpoint titer determination, glycopeptide antigens **P2**, **P3** and **P4** were coupled to BSA to prepare the corresponding BSA conjugates **BSA-1** to **-3**. The average loadings of the protein conjugates **CRM-1** (10 mol/mol), **CRM-2** (10 mol/mol), **BSA-1** (19 mol/mol), **BSA-2** (17 mol/mol) and **BSA-3** (24 mol/mol) were determined by MALDI-TOF mass spectrometry.

To generate HexNAc-*O*-Tyr specific antibodies, two rabbits each were immunized with α-GalNAc-Tyr **CRM-1** (rabbit **R1** and **R2**) and β-GlcNAc-Tyr **CRM-2** (rabbit **R3** and **R4**), and the obtained rabbit antisera were immunologically analized by ELISA and microarray binding experiments.

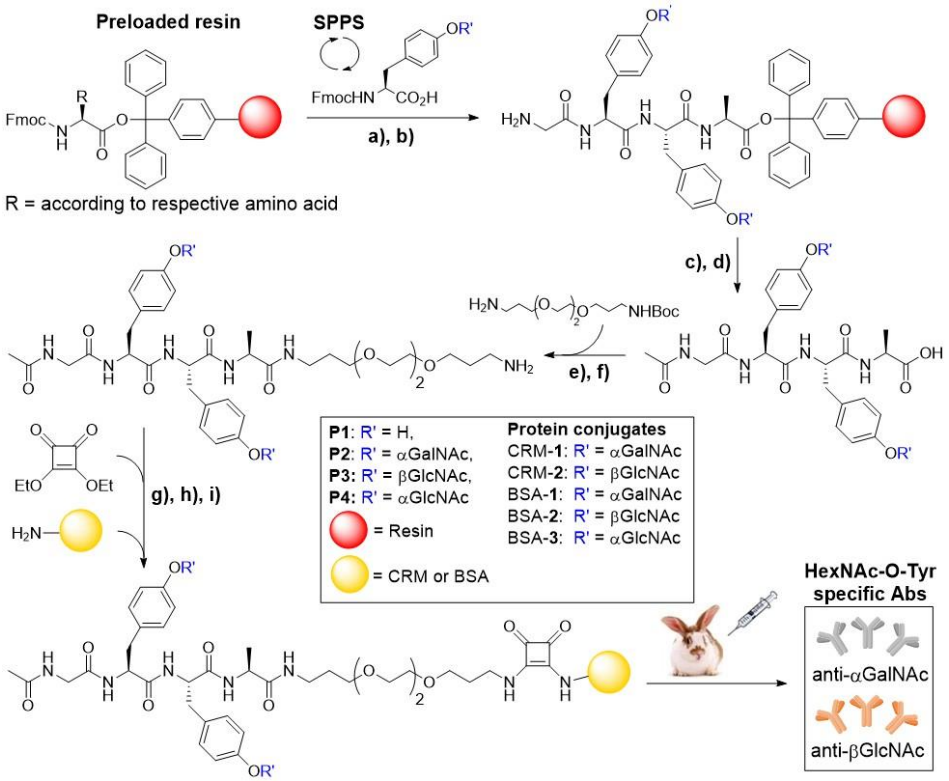


Figure 17. Synthesis of glycopeptide glycopeptide vaccine and BSA conjugates. a) Fmoc removal: 20 % piperidine in DMF; b) Amino acid coupling: Fmoc-Xaa-OH (8.0 equiv), HOBt (7.6 equiv), HBTU (7.6 equiv), DIPEA (16 equiv) in DMF, 40 min; Fmoc-Tyr(HexNAc)-OH, (1.5 equiv), HATU (1.4 equiv), HOAt (1.4 equiv), DIPEA (3.0 equiv) in DMF, 8 h; c) Capping: acetic anhydride, HOBt, DIPEA, DMF, 2 h; d) Release from resin: TFA/TIPS/H_2O (95:5:5), 2 h; e) Spacer coupling: *N*-Boc-4,7,10-trioxa-1,13-tridecanediamine, HATU, HOAt, DIPEA, DMF, 4 h; f) Boc-removal: DCM/TFA (3:1), 4 h; g) Deacetylation: 0.2 M NaOH/MeOH (pH 10.0), 18 h; h) 3,4-diethoxy-3-cyclobutene-1,2-dione, sat. Na_2CO_3 (pH 8.0), EtOH/H_2O (1:1), RT, 3.5 h; i) 75 mM Na_2HPO_4 (pH 9.0-9.5), RT, 3 d.

3.3.4 Evaluation of HexNAc-*O*-Tyr specific antibodies by ELISA and microarray binding experiments

For ELISAs, the rabbit sera were added to wells coated with **BSA-1** (α-GalNAc-*O*-Tyr), **BSA-2** (β-GlcNAc-O-Tyr), or BSA to determine the cutoff value of **BSA-1** and **BSA-2** at 2-fold dilutions (1:8000 to 1:16 384 000), and subsequently probed with a secondary antibody, followed by incubation with streptavidin-conjugated horseradish peroxidase (HRP) for colorimetric detection. The endpoint titer is defined as the reciprocal of the highest dilution of a serum that gives a positive signal above the absorbance cutoff value.[186] All endpoint titers were determined by non-linear regression using the non-linear four-parameter logistic regression (4PL) in GraphPad Prism 8. The obtained endpoint titers indicated high concentrations of specific antibodies and were determined to be approximately 2 024 000 and 4 096 000 for rabbits **R1** and **R2**; and 4 096 000 and 2 048 000 for rabbits **R3** and **R4**, respectively (Figure 18).

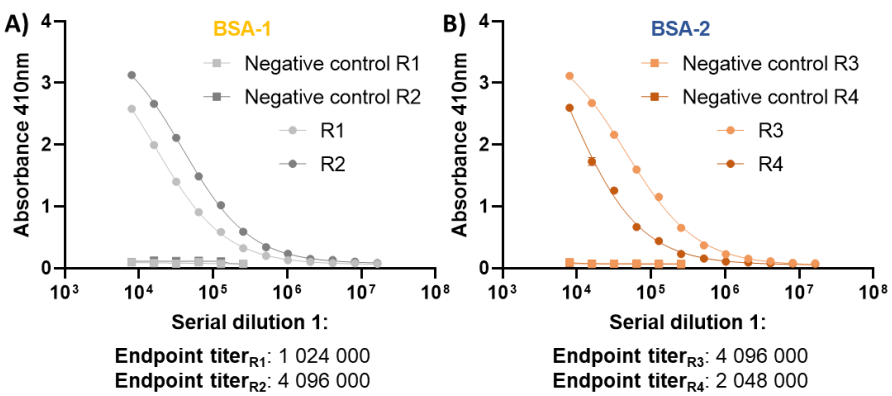


Figure 18. Endpoint titer determination using ELISA assays of A) anti-α-GalNAc-*O*-Tyr rabbit sera **R1** and **R2**, ELISA plate coated with **BSA-1** (α-GalNAc conjugate); B) anti-β-GlcNAc-*O*-Tyr rabbit sera **R3** and **R4,** ELISA plate coated with **BSA-2** (β-GlcNAc conjugate).

Next, the polyclonal antibodies were evaluated using the glycopeptide microarray library in order to determine if they displayed specificity for α-GalNAc, α- or β-GlcNAc-*O*-Tyr glycopeptides over the corresponding Ser and Thr modifications. The microarrays were incubated with dilution series of the anti-α-GalNAc-*O*-Tyr antisera **R1**, **R2** and the anti-β-GlcNAc-*O*-Tyr antisera **R3**, **R4** to adjust for variations in antibody titers. The rabbit antibodies were detected by a secondary biotinylated goat anti-rabbit IgG

and subsequently with a Streptavidin-Cy5 conjugate for fluorescent detection. The microarray binding studies showed that the rabbit sera exhibited different binding patterns toward the glycopeptides. These differences depended on the respective glycosylation site and also on the different GalNAc or GlcNAc isomers modifying the same glycosylation site. While all rabbit antisera showed high avidities for GlcNAc- and GalNAc-*O*-Tyr glycopeptides (Figure 19), the respective Ser and Thr analogs were usually not recognized (Figure 20). Rabbit antisera **R2** and **R3** bound weakly to β-GlcNAc-*O*-Ser only at high serum concentrations. Additionally, rabbit sera **R1, R3**, and **R4** exhibited weak avidities for the unglycosylated antigen peptide **P1**. Surprisingly, all rabbit antibodies showed cross-reactivity between the different HexNAc-*O*-Tyr isomers, implying that they tolerate minor conformational differences of closely related monosaccharides. Among the different glycoforms modifying tyrosine, all antisera preferably bound to α-GlcNAc-*O*-Tyr glycopeptides over the corresponding α-GalNAc and β-GlcNAc-*O*-Tyr analogs. Thereby, it was irrelevant if the antibodies were directed against the α-GalNAc or β-GlcNAc antigen peptides **P2** or **P3**.

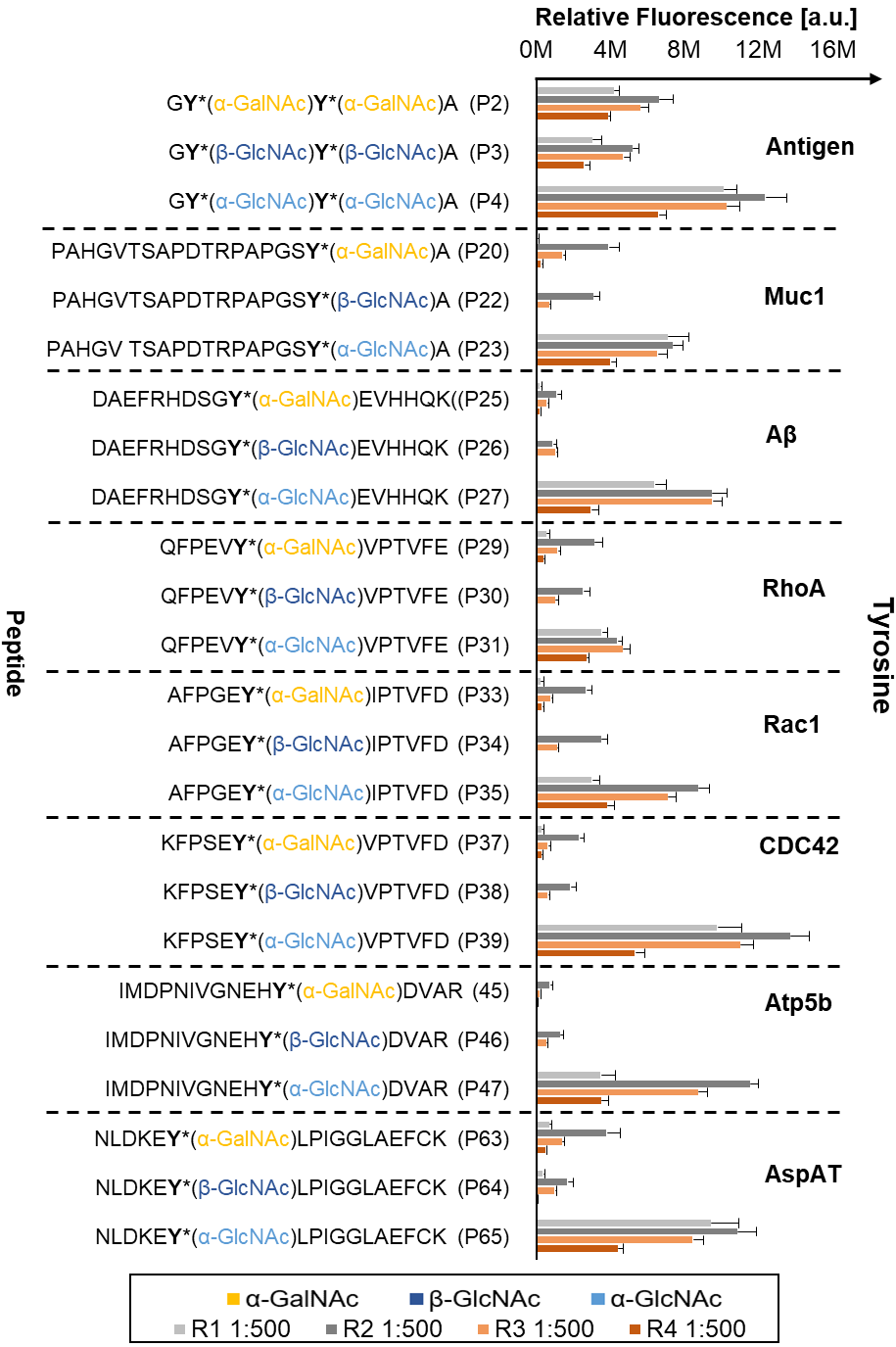


Figure 19. Binding of rabbit sera **R1**, **R2**, **R3** and **R4** at dilution 1:500 toward α-GalNAc-, β-GlcNAc and α-GlcNAc-*O*-Tyr glycopeptides. [a.u.] = arbitrary units.

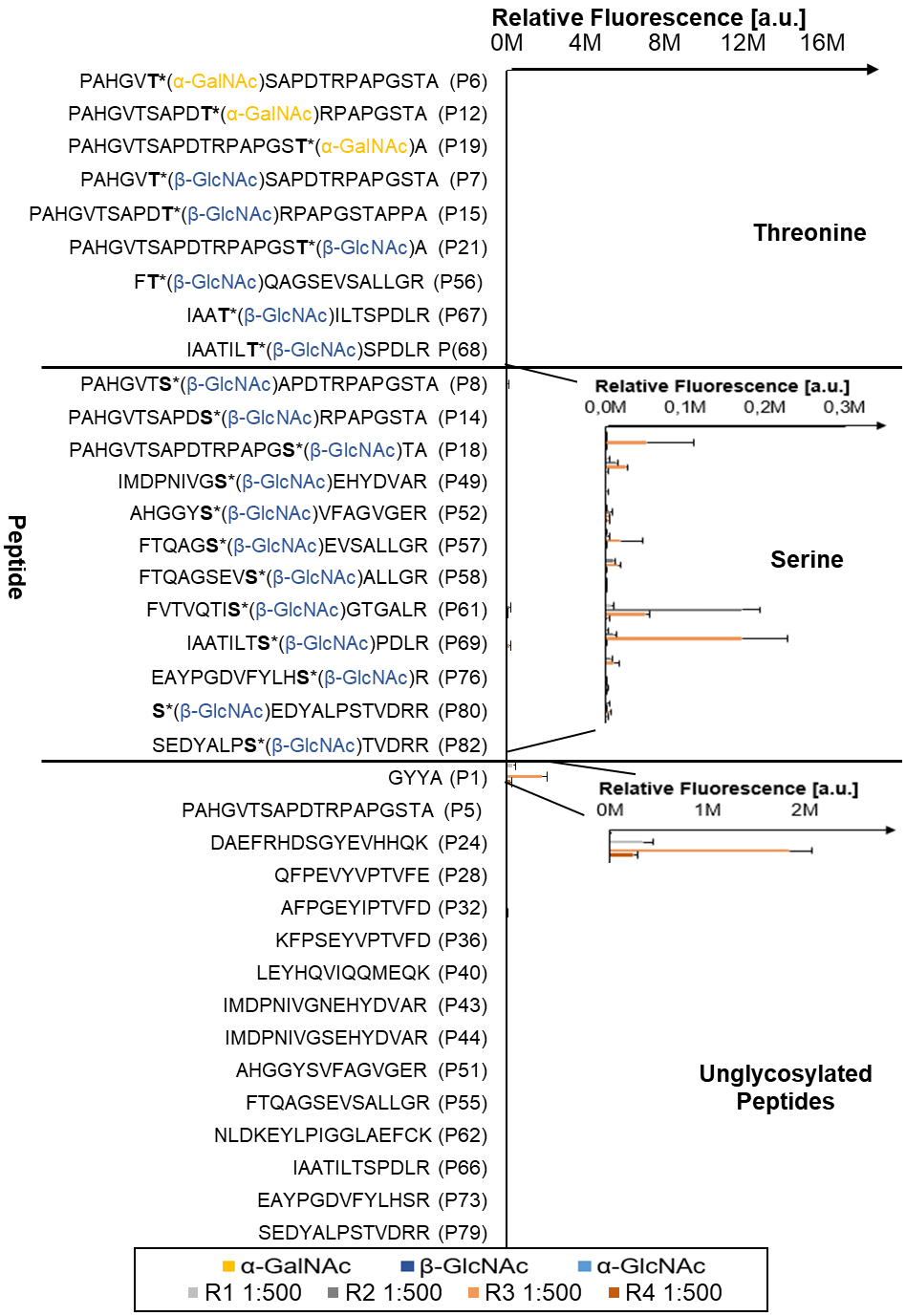


Figure 20. Binding of rabbit sera **R1**, **R2**, **R3** and **R4** at dilution 1:500 toward unglycosylated as well as α-GalNAc- and β-GlcNAc-*O*-Thr/Ser peptides.

To specifically detect HexNAc-*O*-Tyr modified proteins by western blot analysis, monospecific α-GalNAc-, α-GlcNAc- and β-GlcNAc-*O*-Tyr antibodies were obtained from rabbit serum **R2** by affinity purification against the antigen glycopeptides **P2** (α-GalNAc-*O*-Tyr)**, P3** (β-GlcNAc-*O*-Tyr) and **P4** (α-GlcNAc-*O*-Tyr). Rabbit serum **R2** was chosen for this experiment due to its high endpoint titer and affinities toward α-GalNAc and α/β-GlcNAc-*O*-Tyr glycopeptides. Additionally, it was the only serum that did not recognize the unglycosylated antigen peptide **P1**. Antigen peptides **P2**, **P3** and **P4** were immobilized on HiTrap NHS-Activated HP affinity columns and purified polyclonal antibodies **AP-1**, **AP-2** and **AP-3** were obtained from rabbit serum **R2** by affinity enrichment, respectively (Figure 21). SDS-PAGE analysis of the purified rabbit antibodies **AP-1**, **AP-2** and **AP-3** showed a single band indicative of high purity.

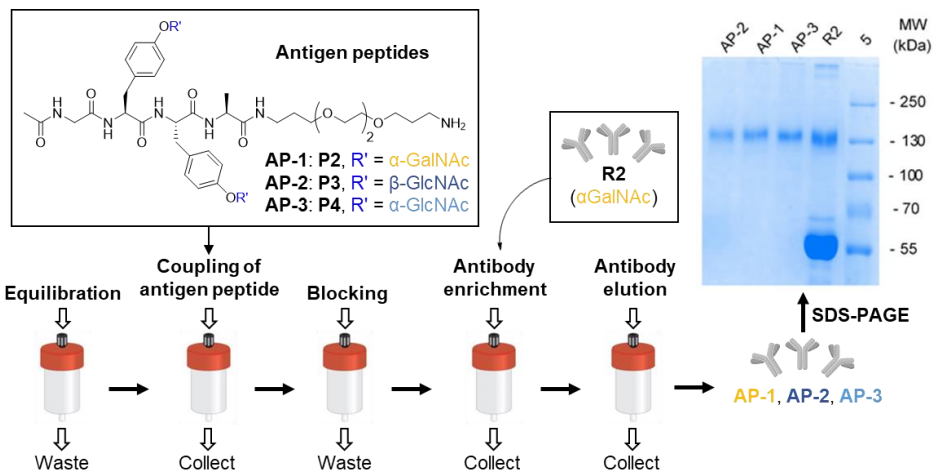


Figure 21. Work flow of affinity purification of polyclonal rabbit antibodies and purity verification by SDS-PAGE.

The endpoint titers of the affinity-purified antibodies **AP-1**, **AP-2** and **AP-3** were then determined by ELISA. ELISA analysis showed, that all purified antibodies **AP-1**, **AP-2** and **AP-3** exhibited good immunoreactivity with the antigen conjugates **BSA-1** (α-GalNAc-O-Tyr), **BSA-2** (β-GlcNAc-O-Tyr) and **BSA-3** (α-GlcNAc-O-Tyr). The endpoint titers are shown in Figure 22 and were determined to be in a low ng/mL range. These results indicate that high concentrations of specific high-affinity antibodies could be enriched during the affinity purification experiments.

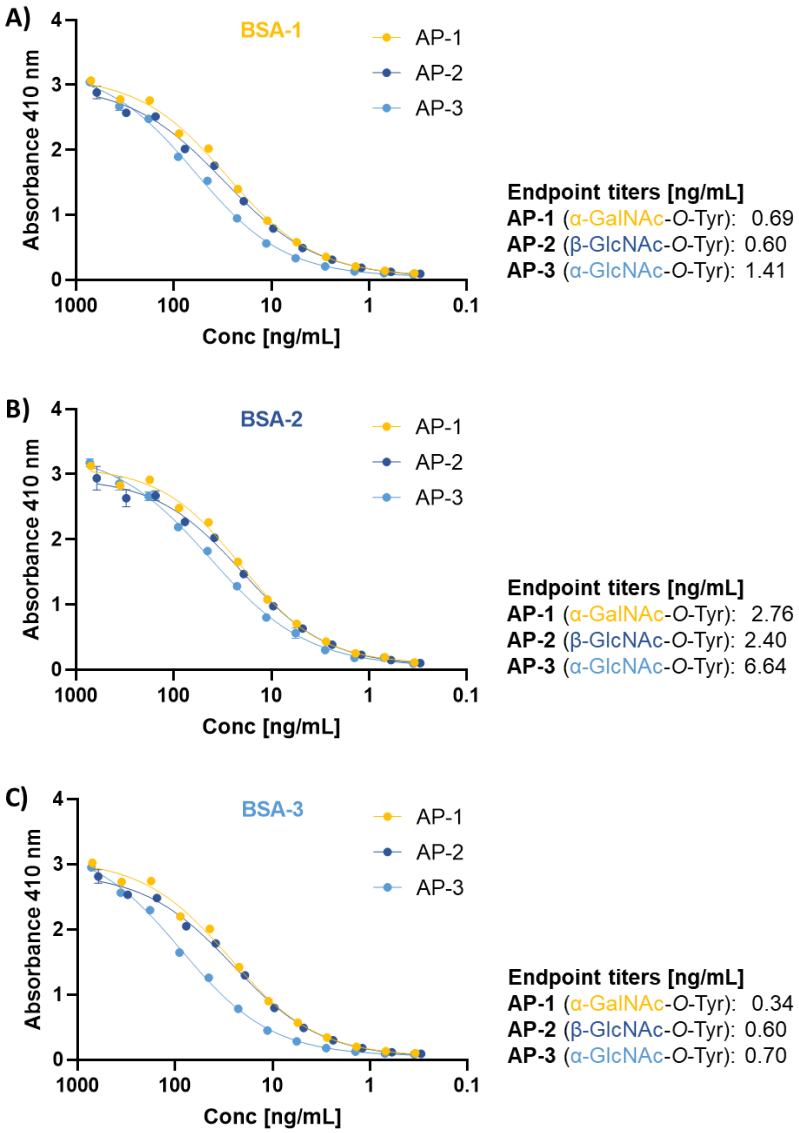


Figure 22. Endpoint titer determination of affinity-purified antibodies **AP-1**, **AP-2** and **AP-3** using ELISA assays. The ELISA plates were coated with A) antigen peptide BSA conjugate **BSA-1** (α-GalNAc-*O*-Tyr); B) antigen peptide BSA conjugate **BSA-2** (β-GlcNAc-*O*-Tyr); C) antigen peptide BSA conjugate **BSA-3** (α-GlcNAc-*O*-Tyr).

Finally, the binding specificity of the purified antibodies **AP-1**, **AP-2** and **AP-3** were evaluated using glycopeptide microarray binding experiments. The glycopeptide microarrays were incubated with a dilution series (ranging from 1:50 to 1:64 000) of the respective affinity purified rabbit antibodies, followed by detection with a secondary biotinylated goat anti-rabbit antibody and Cy5-labeled streptavidin for fluorescent detection. Microarray analysis showed that none of the purified antibodies showed cross-reactivity toward either the unglycosylated peptides or glycosylated Ser and Thr peptides (Figure 23). Interestingly, the affinity purifications did not dramatically improve the antibody selectivity towards the corresponding HexNAc-*O*-Tyr glycoforms coupled to the affinity column, and **AP-1**, **AP-2** and **AP-3** displayed similar binding preferences for the different GalNAc and GlcNAc isomers on tyrosine (Figure 24). Consequently, the higher affinity that **AP-1**, **AP-2** and **AP-3** displayed toward α-GlcNAc- over the α-GalNAc- and β-GlcNAc-*O*-Tyr glycopeptides remained.

3.3.5 Ability of a O-β-GlcNAc mAb to recognize O-β-GlcNAc on tyrosine

In order to determine if the affinity-purified antibodies **AP-1**, **AP-2** and **AP-3** could be applied to specifically detect Tyr-*O*-HexNAcylation using common antibody concentrations, the commercial *O*-β-GlcNAc-specific monoclonal mouse antibody CTD 110.6 was evaluated as a comparison. This antibody is usually applied to detect of *O*-β-GlcNAcylation on Ser and Thr residues, and its binding specificities toward *O*-β-GlcNAc-Tyr modifications has not been explored so far. The mAb binding specificities toward *O*-β-GlcNAc-Ser, -Thr and -Tyr glycopeptides were determined by a microarray binding study, and were compared with those of the purified Tyr specific rabbit antibodies **AP-1**, **AP-2** and **AP-3**. The glycopeptide microarray library was incubated with a dilution series (ranging from 1:50 to 1:2 000) of the monoclonal β-O-GlcNAc specific mouse antibody with subsequent detection using a secondary Cy5-conjugated anti-mouse antibody. Microarray analysis showed that while the *O*-β-GlcNAc mAb was selective for *O*-β-GlcNAc modified peptides, it surprisingly did not recognize all glycopeptides that were *O*-β-GlcNAcylated on Ser and Thr (Figure 23). However, CTD 110.6 bound to all β-GlcNAc-O-Tyr glycopeptides and exhibited even higher binding affinities to peptides GlcNAcylated on Tyr compared to the respective Ser and Thr analogs. These results indicate that hydrophobic interactions with the Tyr residue can increase the binding of the monoclonal antibody drastically. Furthermore, the affinity purified antibodies **AP-1-3** exhibited binding intensities that were comparable with the monoclonal *O*-β-GlcNAc-specific antibody at similar concentrations. Consequently, the HexNAc-*O*-Tyr-specific antibodies **AP-1**, **AP-2** and **AP-3** can be used for bioanalytical experiments in concentrations that are similar to the mAb application concentration.

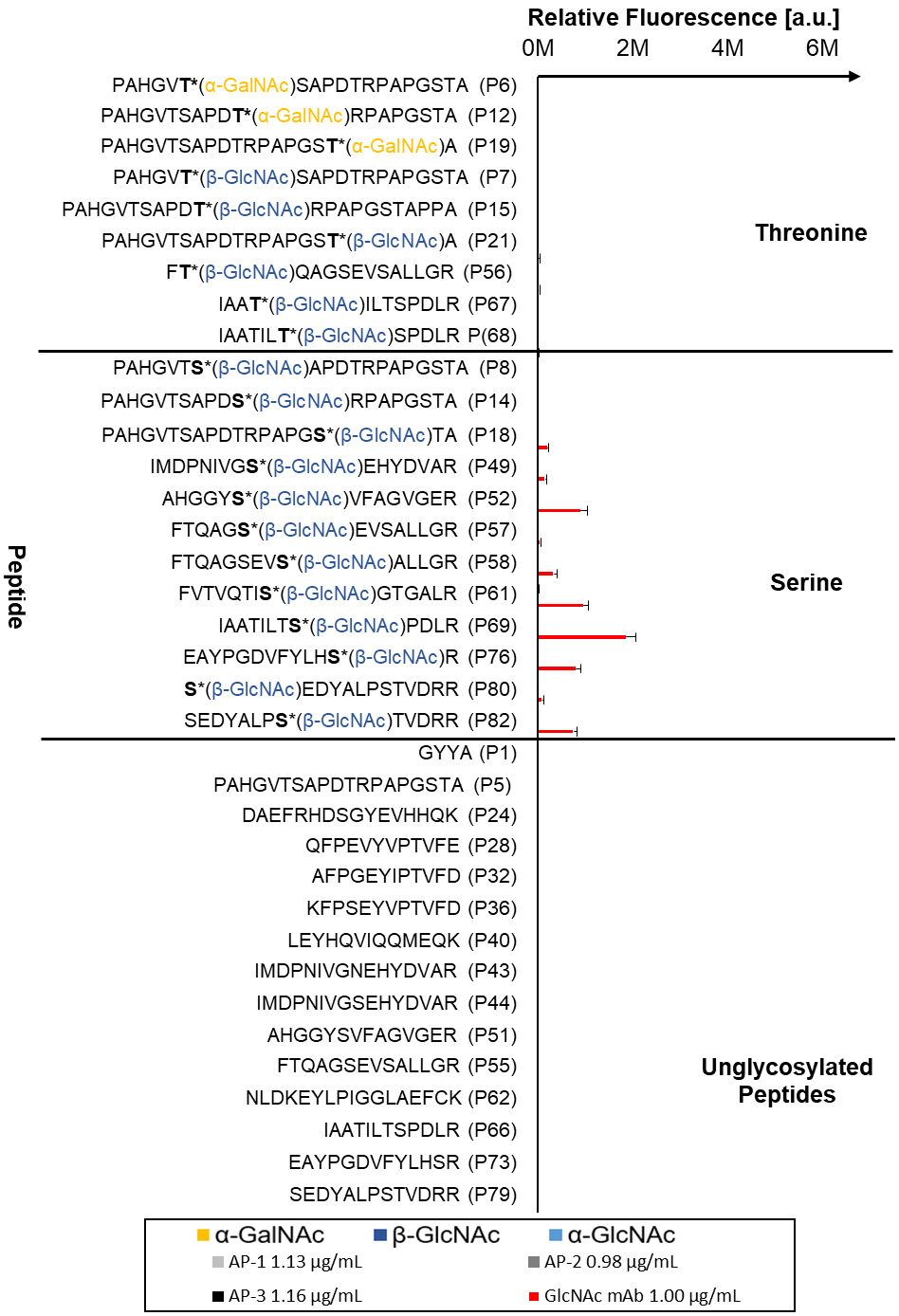


Figure 23. Binding of affinity purified rabbit antibodies **AP-1**, **AP-2**, **AP-3** and anti-*O*-β-GlcNAc mAb toward α-GalNAc-*O*-Thr or β-GlcNAc-*O*-Thr/Ser glycopeptides.

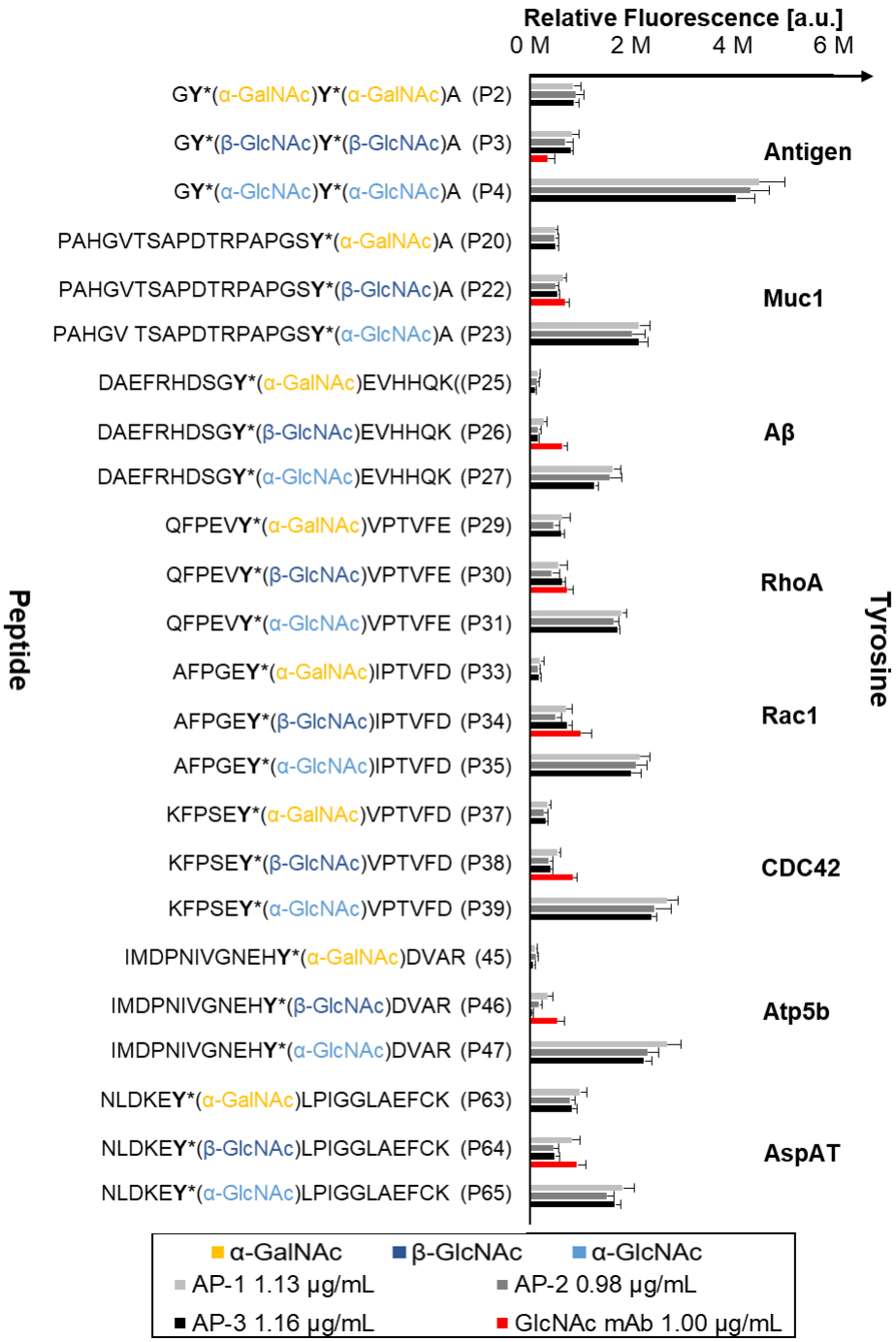


Figure 24. Binding of affinity purified rabbit antibodies **AP-1**, **AP-2**, **AP-3** and anti-*O*-β-GlcNAc toward α-GalNAc-, α- or β-GlcNAc-*O*-Tyr glycopeptides.

3.3.6 Specific detection of RhoA modified with α-GlcNAc-*O*-Tyr

The affinity-purified antibodies **AP-1**, **AP-2** and **AP-3** were finally applied to specifically detect the GTPase RhoA that was enzymatically α-GlcNAcylated on Tyr. Western blot analysis was used to test if the affinity-purified antibodies can detect this modification. RhoA was enzymatically glycosylated with α-GlcNAc using UDP-GlcNAc and the glycosyltransferase domain of the *Photorhabdus asymbiotica* protein toxin ($PaTox^{G}$). After SDS gel electrophoresis of α-GlcNAcylated RhoA, as well as unmodified RhoA as a negative control, the proteins were transferred to a polyvinylidene fluoride membrane. The purified rabbit antibodies **AP-1**, **AP-2** and **AP-3** were used to probe both proteins. For fluorescent detection, the membranes were incubated with a secondary Alexa Fluor 488 labeled donkey anti-rabbit antibody. Whereas HexNAc-*O*-Tyr specific rabbit antibodies did not recognize the unmodified RhoA, the α-GlcNAcylated RhoA was successfully detected (Figure 25). This shows that rabbit antibodies **AP-1**, **AP-2** and **AP-3** can be used to detect α-GlcNAc-O-Tyr-modified proteins.

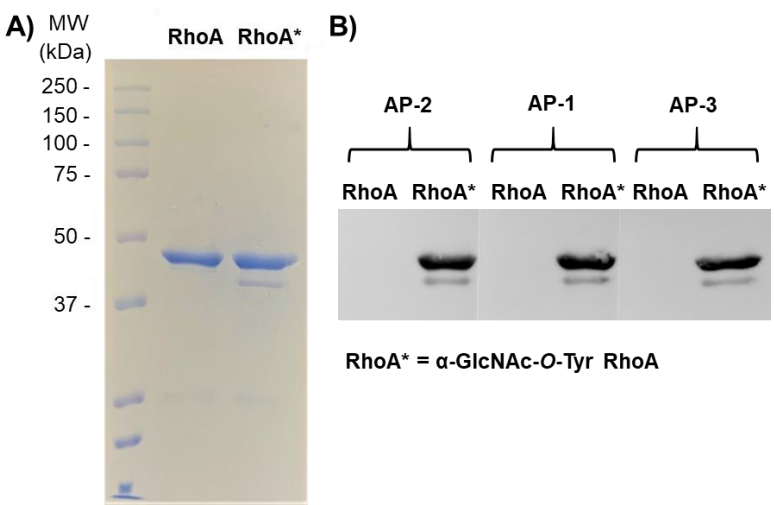


Figure 25. A) Coomassie staining of RhoA (MW = 48620 Da, lane 2) and α-GlcNAc-*O*-Tyr modified RhoA* (lane 3); B) Western blot of RhoA (lanes 1, 3 and 5) and α-GlcNAc-RhoA* (lanes 2, 4 and 64) probed with HexNAc-*O*-Tyr specific antibodies **AP-1** (lanes 3 and 4), **AP-2** (lanes 1 and 2) and **AP-3** (lanes 5 and 6). The lower band indicates partial breakdown of RhoA.

3.4 Summary and conclusion

In summary, a library consisting of 86 synthetic unglycosylated and α-GalNAc-, β-GlcNAc- and α-GlcNAc-*O*-Tyr/Ser/Thr glycopeptides was generated and subsequently printed on microarray slides. The glycopeptide microarray library was then used to evaluate the ability of the plant lectins VVA, WGA and GSL II to detect and identify tyrosine *O*-HexNAcylation on a peptide level. All tested lectins showed similar or lower binding affinities to HexNAcylated Ser and Thr residues compared to the corresponding HexNAc-*O*-Tyr analogs. Furthermore, none of the unglycosylated peptides were recognized by either lectin. These findings support the hypothesis that lectins specifically bind to the carbohydrate residue in a site specific and presentation dependent mode. This can be influenced not only by the respective acceptor amino acid, but also the amino acid sequence of the peptide backbone. Additionally, the spacing and steric accessibility of multivalent ligands are crucial factors for lectin binding. Since the evaluated plant lectins are not specific for HexNAcylation on Tyr, but also recognize HexNAc modifications on Ser and Thr residues, new tools to specifically detect, identify and enrich HexNAc-*O*-Tyr needed to be developed. Therefore, highly specific anti-α-GalNAc-, β-GlcNAc- and α-GlcNAc-*O*-Tyr antibodies were generated by immunizing rabbits with synthetic β-GlcNAc- and α-GalNAc-*O*-Tyr antigen peptide-CRM vaccines. The binding specificities of the obtained sera were evaluated by ELISA and microarray binding assays, and as a result, GalNAc- and GlcNAc-O-Tyr antigen peptide-BSA conjugates were generated. ELISA as well as microarray binding analysis showed that the raised antibodies exhibited strong binding to all HexNAc isoforms with a preference for α-GlcNAc-O-Tyr glycopeptides. As desired, low or no binding affinities could be observed to unglycosylated or HexNAcylated Ser or Thr glycopeptides. To evaluate the abilities of the rabbit antibodies to specifically detect HexNAcylation on tyrosine in more complex samples, a rabbit serum with a high titer and high affinities avidities for all HexNAc-*O*-Tyr glycopeptide isomers was purified by affinity enrichment against antigen peptides. The purified monospecific antibodies were applied in a western blot analysis to specifically detect the host GTPase RhoA that was enzymatically modified with α-GlcNAc on tyrosine using the bacterial toxin PaTox^{G}. While unmodified RhoA was not recognized by the affinity-purified rabbit antibodies, α-GlcNAc-*O*-Tyr-RhoA was successfully detected. Consequently, HexNAc-*O*-Tyr specific polyclonal

rabbit antibodies were generated that are able to detect α-GlcNAc-*O*-Tyr modifications on protein level. These results indicate that the specific antibodies can also be applied to detect HexNAc-*O*-Tyr glycosylation in biological samples. Additionally, the *O*-β-GlcNAc specific monoclonal mouse antibody (CTD 110.6), which is commonly used to detect Ser- and Thr-*O*-GlcNAcylation, was also found to detect *O*-β-GlcNAc on tyrosine.
In conclusion, the affinity enriched HexNAc-*O*-Tyr specific rabbit antibodies as well as the *O*-β-GlcNAc-specific mAb CTD 110.6 and the plant lectins, VVA, WGA and GSL II, exhibit complementing affinities and could be combined to enable selective detection, identification and enrichment of α-GalNAc-, β-GlcNAc- and α-GlcNAc-O-Tyr modified proteins and tryptic glycopeptides. As a result of this work, new tools are now available to detect and enrich proteins carrying this new group of PTMs, and to explore the glycobiology behind tyrosine *O*-HexNAcylation.

4 Project 2 – Tools to explore mucin-type glycosylation (Papers III – VII)

4.1 Bacterial lectin recognition of fucosylated mucin glycopeptides (Paper III)

Mucin glycoproteins are major components of the mucosal protective barrier, which protects the epithelial tissues from many bacterial and virus infections. However, bacteria and viruses have co-evolved with the human host and developed strategies to promote immune escape and virulence. For example, pathogenic bacteria cause virulence by adhering to specific glycan motifs on membrane-bound mucins on the host cell-surface. These carbohydrate-protein interactions can also promote biofilm formation, protein toxin delivery, or trigger inflammation.

Terminal *O*-fucosylated glycan epitopes on mucin glycoproteins are key ligands for many bacterial and viral lectins. A family of ten fucosyltransferases (FUT1–7 and FUT9–11) is responsible for the addition of a fucose residue to either a terminal galactose of glycan chains in an α-1,2-linkage, or to a subterminal GlcNAc in an α-1,3- or α-1,4-linkage.[187] This way, the blood group A-, B- and H-antigens, and Lewis epitopes are generated. Fucosylated glycans also play important roles in airway diseases such as cystic fibrosis or the chronic obstructive pulmonary disease. Individuals suffering from these diseases show, in addition to a high mucin secretion, an increased expression of fucosylated glycans on lung mucins. Thus, they are more susceptible to bacterial pathogens such as *Pseudomonas aeruginosa* or *Clostridium difficile,* which express fucose-binding adhesins.

P. aeruginosa is a Gram-negative opportunistic bacterium that colonizes various human tissues and organs and leads to severe and chronic infection of the respiratory and/or urinary tracts, eyes as well as skin.[188-189] So far, several *P. aeruginosa* adhesins and lectins have been identified, which mediate the bacterial adhesion process. For example, the small soluble homotetrameric lectin LecB (PA-IIL) specifically recognizes and binds to L-fucosides with a preference for α- over β-L-fucosides.[190-192] This lectin is a key virulence factor in *P. aeruginosa* infections and is involved in bacterial biofilm formation.[193] LecB also mediates the attachment to the host cell during infection by interacting with fucosylated glycans on the epithelial cell surface. Additionally, it has

been shown to have a cytotoxic effect on epithelial airway cells and to arrest ciliary beating of human airway epithelium.[194-195] To combat *P. aeruginosa* infection, multivalent glycoconjugates and glycomimetics including glycopeptides, -dendrimers or –polymers that inhibit LecB binding have been developed.[196]

C.difficile is a Gram-positive opportunistic bacterium that often causes recurrent mild to severe gastrointestinal disorders.[197] The *C.difficile* toxin A (TcdA) is is a primary virulence factor for *C. difficile* infection.[198] This multi-domain protein possesses four functionally distinct domains: The carbohydrate recognition domain, which interacts with host cell carbohydrate structures to initiate toxin internalization, the translocation domain, which translocates the catalytic glucosyltransferase domain into the cell cytosol, the autoprotease domain, which is required for proteolytic cleavage of the toxin, and the glycosyltransferase domain.[198] After TcdA-binding to glycan ligands of intestinal epithelial cells and toxin delivery into the cell,[199-201] TcdA monoglucosylates and thereby inactivates critical host GTPases such as the RhoA protein family, Rac and Cdc42, leading to inflammation, tissue damage, and ultimately cell death.[202-203] Therapeutics to treat *C. difficile* infections are limited and often employ strong antibiotics. To tackle this problem, glycoconjugates have been developed to interfere with TcdA binding and thereby inhibit TcdA-mediated cell toxicity.[201, 204] TcdA has in previous studies been reported to recognize the Galili epitope, but also fucosylated Le^y, Le^x, Sialyl-Le^x and sulfo-Le^x glycans.[199-201]

A better understanding of the bacterial adhesion processes of *P. aeruginosa* and *C. difficile* on a molecular level could advance the development of new glycoconjugates and -mimetics for anti-biofilm and anti-adhesion therapies, as the fine binding specificities of these lectins depend on glycan epitope presentation by the underlying core structures, the peptide backbone and the specific glycosylation sites.

4.1.1 Motivation

The presentation of terminal carbohydrate motifs on the underlying mucin core structures, the exact placement of glycosylation sites on the peptide backbone, as well as the peptide amino acid sequence are potentially essential for the fine binding specificities and biological function of bacterial lectins. However, these factors are often not considered in studies of lectin-glycan interactions even if the glycan orientation and structural rigidity may define possible limitations for ligand recognition.

In this work, we employed fucosylated structures displayed on glycan cores of mucin 1 (MUC1) and mucin 5B (MUC5B) tandem repeat sequences to explore the roles of different terminal fucose motifs, and of ligand presentation on different glycosylation sites of the peptide backbone in bacterial lectin recognition events of the *P. aeruginosa* lectin LecB and the *C. difficile* toxin A (TcdA). For this study, a library of 63 synthetic α1,2-, α1,3- and α1,4-fucosylated mono- and bivalent MUC1 and MUC5B glycopeptides was generated and immobilized on microarrays. The glycopeptides carried mucin core-1 to core-4, as well as LacNAc elongated core structures on distinct mucin peptide tandem repeat glycosylation sites that were enzymatically modified with Lewis x (Le^x), Lewis a (Le^a), Lewis y (Le^y), Lewis b (Le^b) and H-type motifs (Figure 26, Figure 27). The obtained fucose glycopeptide microarray library was used to elucidate the fine binding specificities of the *P.aeruginosa* lectin LecB and the *C. difficile* toxin A (TcdA).

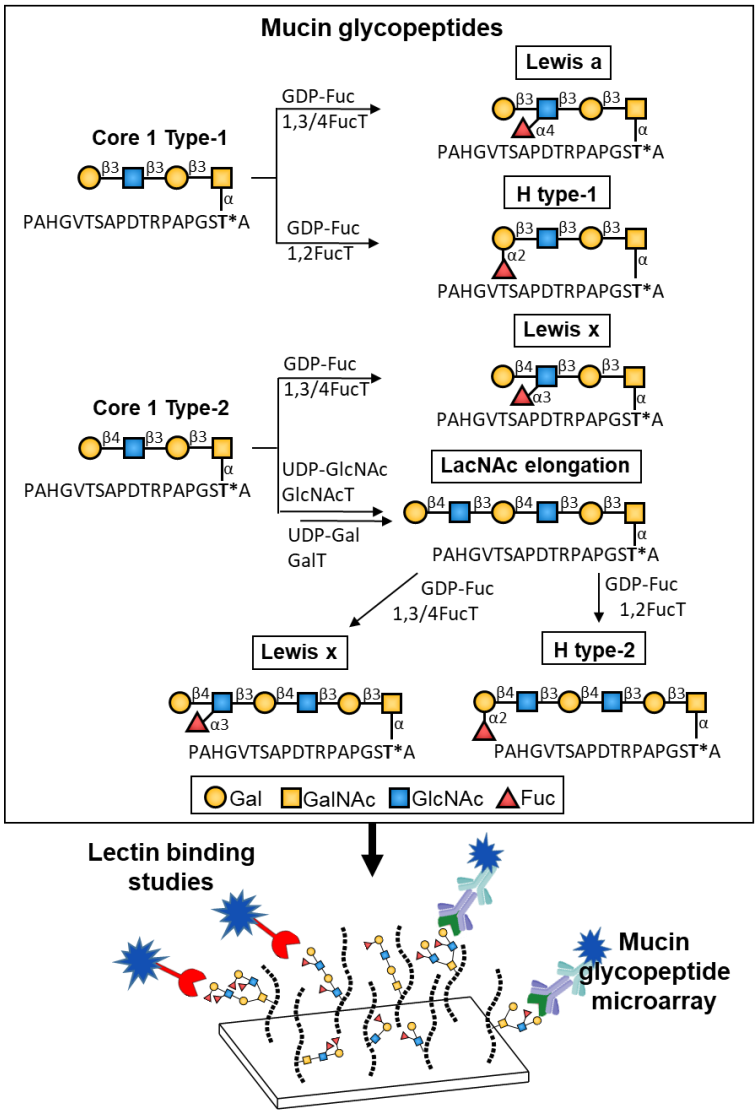
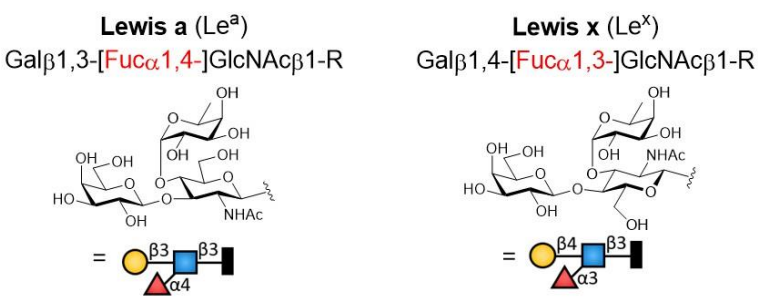


Figure 26. Example of enzymatic LacNAc elongation and fucosylations made to generate a fucosyl glycopeptide microarray library for evaluation of binding specificities of the *P. aeruginosa* lectin LecB and the *C. difficile* toxin A.

4.1.2 Results and Discussion

4.1.2.1 Preparation of a fucosylated mucin glycopeptide library

To study the lectin-glycan interactions of LecB and TcdA, LacNAc (type-1 and type-2) elongated type-1 and type-2 mucin core 1 to core 4 glycosylated threonine building blocks were prepared and incorporated into the MUC1 and MUC5B tandem repeat sequences, PAHGVTSAPDT*RPAPGST*A and AT*PSST*PGT*THTP (T* = possible glycosylation sites), by Fmoc-SPPS. The glycosylated amino acid building blocks and the mucin core glycopeptides were synthesized by Dr. Manuel Schorlemer and Dr. Christian Pett. The glycopeptides were enzymatically modified with LacNAc and/or fucosylated with Le^a, Le^x and H-type as well as bi-fucosylated Le^b and Le^y motifs (**Figure 27**). Selected glycopeptides were extended with additional LacNAc using the *Helicobacter pylori* β-1,3-*O*-*N*-acetylglucosaminyltransferase (β3GlcNAcT) and a fusion protein of human β-1,4-*O*-galactosyltransferase (His_6-Propeptide-catβ4GalT-1, β4GalT). Subsequently, the different fucose motifs were generated using *Helicobacter pylori* α1,3/4-*O*-fucosyltransferase and/or *H. mustelae* α1,2-*O*-fucosyltransferase. In the frame of this project, a fucosylated mucin glycopeptide library, which was previously prepared by Dr. Jin Yu, was further extended with additional Le^a, Le^x and H-type glycopeptides, as well as with glycopeptides exhibiting Le^b and Le^y motifs.



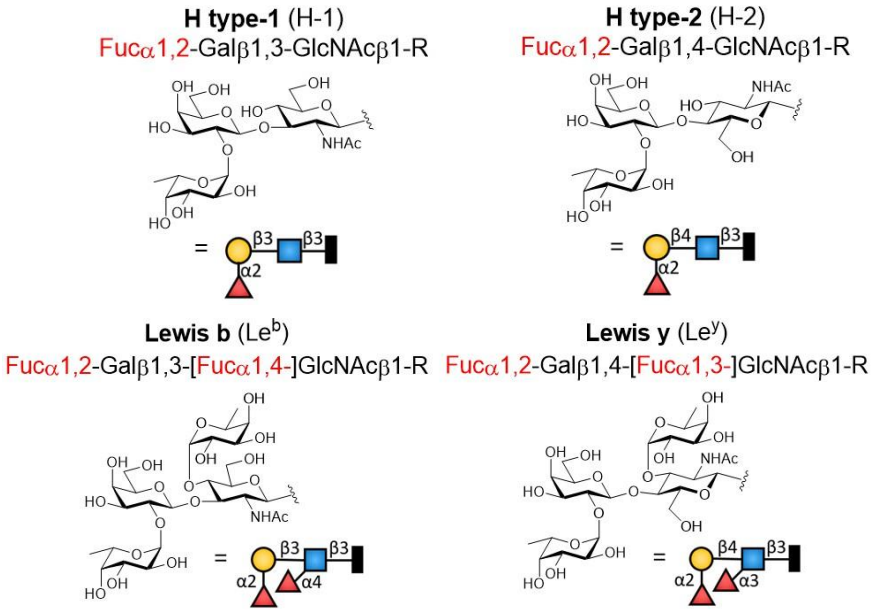


Figure 27. Schematic representation of the prepared fucosylated terminal glycan antigens.

In order to generate Le^b and Le^y motifs, the order of the applied fucosyltransferases was crucial for the synthesis outcome (Figure 28). In an initial attempt, selected glycopeptides were first α1,3/4-fucosylated, followed by modification with α1,2-fucose. However, the desired bi-fucosylated products could not obtained in satisfying yields, since α1,3/4-fucosylated glycans were found to be poor α1,2-*O*-fucosyltransferase-substrates. Consequently, the order of the applied fucosyltransferases was reversed so the α1,2- was followed by the α1,3/4-fucosylation. This adjustment resulted in the successful preparation of the desired Le^b and Le^y mucin glycopeptides and the fucosylated mucin glycopeptide library was printed on NHS-activated microarrays. This extensive library of *O*-fucosyl MUC1 and MUC5B glycopeptides with structurally well-defined, and closely related core structures provided a unique platform to explore fine specificities of the fucose-binding proteins LecB and TcdA.

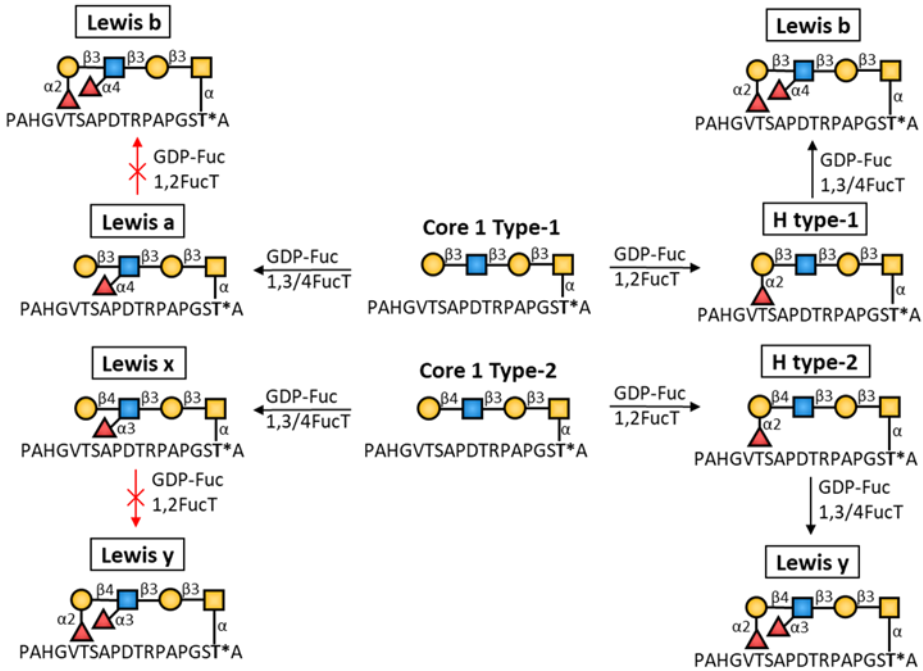


Figure 28. Enzymatic synthesis of Lewis b and y modified glycopeptides.

4.1.2.2 Recognition of fucosylated glycopeptides by LecB from *Pseudomonas aeruginosa*

The fucosylated glycopeptide microarray library was used to determine the binding preferences of LecB. Microarrays were incubated with a dilution series of LecB-biotin (31 nM – 16 µM) and subsequently incubated with Cy5-labeled streptavidin for fluorescent detection. Additionally, apparent surface dissociation constants (Surf. K_D) for LecB binding toward the fucosylated MUC1 and MUC5B glycopeptides were calculated. Curve fitting was carried out by non-linear regression using the saturation binding − specific binding with Hill Slope equation in GraphPad Prism 8. From here on, the core 1 to core 4 type-1 and type-2 mucin core structures will be referred to as C(1-4)T1/2 in the text and figures, for example $C2_{Tet}T1$ stands for core 2 tetrasaccharide type-1, and $C2_{Hex}T2$ for core 2 hexasaccharide type-2 etc.. Microarray analysis of LecB showed that the fine binding specificities strongly depended on the particular fucose motif, the underlying core structures, LacNAc-extension as well as on the presenting peptide backbone and the distinct glycosylation sites. LecB exhibited a broad selectivity toward all fucosylated MUC1 and MUC5B glycopeptides and bound to them in a high nanomolar to low micromolar range (Surf. $K_{D,Muc1}$ = 0.16 – 2.97 µM and Surf. $K_{D,Muc5B}$ = 0.39 – 2.91 µM).

The Le^a and H-type-2 fucosylated MUC1 glycopeptides were better binders than the respective H-type-1 and Le^x fucosylated glycopeptides with Le^x glycans being the weakest binders (Figure 29, Table 4). The observed binding pattern was consistent with binding data from previous glycan recognition studies of LecB.[205-207] The higher avidities for α1,4-fucosylated glycopeptides are based on the additional hydrogen bond to the protein backbone the Le^a antigen can form due to the favorable steric location of the GlcNAc *O*-6 position. On the other hand, if the GlcNAc *N*-acetyl group of the Le^x antigen is located in the same position it would lead to sterical hindrance. Consequently, the Le^x glycan must adapt into a less favorable conformation upon binding to LecB.[208-209] Additionally, LecB bound to various fucose antigens presented on different core structures with different affinities (Figure 29, Table 4). The H-antigen peptides were recognized with an increasing affinity in the order core 3 ≤ core 1 < core 4 ≈ core 2 tetrasaccharide ≤ core 2 hexasaccharide. Whereas LecB showed comparable binding to the linear H-antigen core 1 and 3 glycans, the branched core 2 and 4 structures with both arms

carrying one fucose residue each showed an increased binding affinities. The terminal carbohydrate epitopes on the branches are oriented in opposite directions, thus presenting a spatial arrangement that may favor intra- or intermolecular multivalent lectin binding.

LecB bound to the α1,4-fucosylated glycopeptides in a pattern similar to the one observed for the corresponding type-2 H glycans. The Le^a-antigens on the different cores were recognized with the increasing affinity order core 3 < core 1 ≤ core 2 tetrasaccharide < core 2 hexasaccharide. Because the Le^a modified core 2 tetrasaccharide glycopeptides carry only one fucose residue, their binding affinities are similar to the values for corresponding the core 1 analogs. In contrast, the fucosylated core 2 hexasaccharide contains an additional α1,4-fucosylated LacNAc unit which can participate in multivalent interactions that lead to an overall higher avidity. Therefore, it was a better binder than the respective tetrasaccharide derivative.

Glycopeptides modified with the Le^x antigen were bound by LecB in the following increasing affinity order: core 3 < core 2 tetrasaccharide < core 1 tetrasaccharide < core 4 < core 2 hexasaccharide. The branched core 2 hexasaccharide and the core 4 glycopeptides contain an additional fucose residue each that can participate in multivalent binding events. In comparison to the core 1, core 2 tetrasaccharide and core 3 that carry only one fucose unit each, the hexasaccharide derivative showed enhanced binding. The additional arm on the core 2 tetrasaccharide may further lead to steric hindrance of the bound Le^x conformation, thus decreasing the binding avidity compared with the linear core 1 and 3 structures.

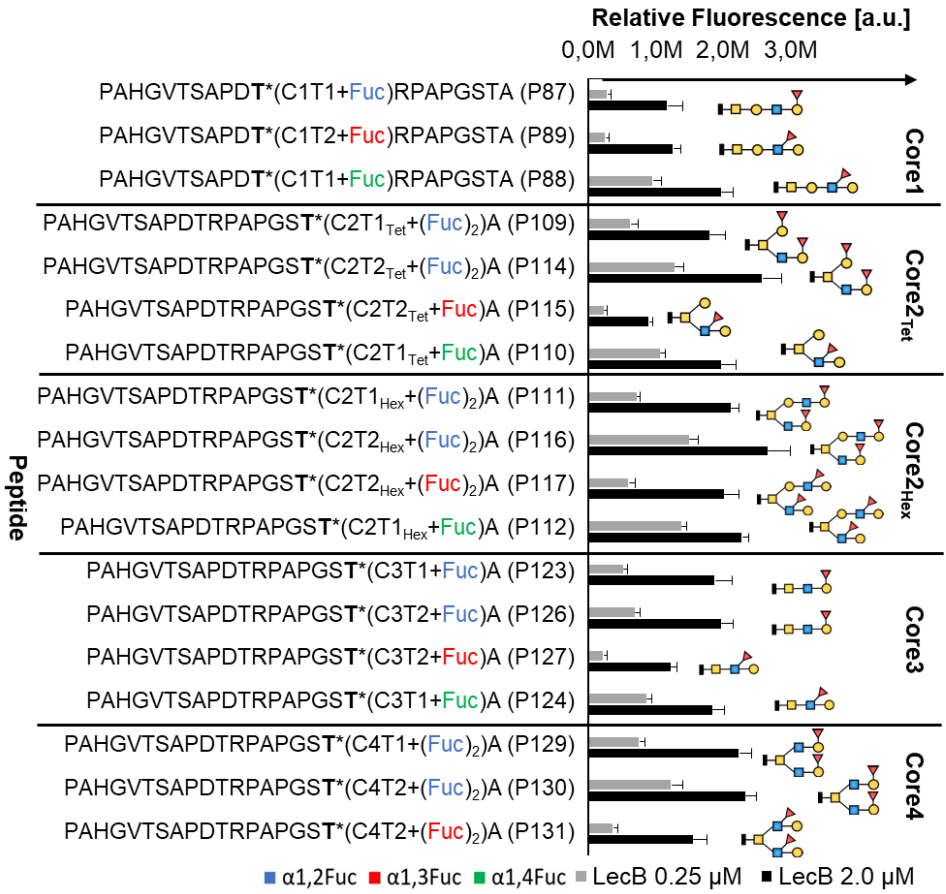


Figure 29. LecB binding toward Le^a- Le^x and H-type-1 and -2 modified MUC1 core structures. [a.u.] = arbitrary units.

Table 4. Surf. K_D values for LecB binding toward Le^a- Le^x and H type-1 and -2 modified MUC1 core structures.

ID	Glycopeptide sequence	K_D ± SEM [µM]
P87	PAHGVTSAPD**T***(C1T1+α1,2Fuc)RPAPGSTA	1.10 ± 0.10
P88	PAHGVTSAPD**T***(C1T1+α1,4Fuc)RPAPGSTA	0.36 ± 0.03
P89	PAHGVTSAPD**T***(C1T2+α1,3Fuc)RPAPGSTA	1.19 ± 0.09
P101	PAHGVTSAPDTRPAPGS**T***(C1T1+α1,2Fuc)A	0.64 ± 0.03
P104	PAHGVTSAPDTRPAPGS**T***(C1T2+α1,2Fuc)A	0.48 ± 0.03
P109	PAHGVTSAPDTRPAPGS**T***(C2T1$_{Tet}$+(α1,2Fuc)$_2$)A	0.54 ± 0.03
P114	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Tet}$+(α1,2Fuc)$_2$)A	0.37 ± 0.03
P115	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Tet}$+α1,3Fuc)A	1.72 ± 0.19
P110	PAHGVTSAPDTRPAPGS**T***(C2T1$_{Tet}$+α1,4Fuc)A	0.25 ± 0.02
P111	PAHGVTSAPDTRPAPGS**T***(C2T1$_{Hex}$+(α1,2Fuc)$_2$)A	0.53 ± 0.03
P116	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Hex}$+(α1,2Fuc)$_2$)A	0.28 ± 0.03
P117	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Hex}$+(α1,3Fuc)$_2$)A	0.91 ± 0.06
P112	PAHGVTSAPDTRPAPGS**T***(C2T1$_{Hex}$+(α1,4Fuc)$_2$)A	0.19 ± 0.01
P123	PAHGVTSAPDTRPAPGS**T***(C3T1+α1,2Fuc)A	0.82 ± 0.05
P126	PAHGVTSAPDTRPAPGS**T***(C3T2+α1,2Fuc)A	0.60 ± 0.04
P127	PAHGVTSAPDTRPAPGS**T***(C3T2+α1,3Fuc)A	1.76 ± 0.15
P124	PAHGVTSAPDTRPAPGS**T***(C3T1+α1,4Fuc)A	0.31 ± 0.02
P129	PAHGVTSAPDTRPAPGS**T***(C4T1+(α1,2Fuc)$_2$)A	0.55 ± 0.03
P130	PAHGVTSAPDTRPAPGS**T***(C4T2+(α1,2Fuc)$_2$)A	0.37 ± 0.03
P131	PAHGVTSAPDTRPAPGS**T***(C4T2+(α1,3-Fuc)$_2$)A	1.17 ± 0.09

In agreement with previous studies, LecB showed higher binding avidities for H-antigen type-2 glycopeptides than for the type-1 analogs (Figure 30, Table 4).[205, 207] As described above, H-antigen peptides were recognized with the increasing affinity order: core 3 ≤ core 1 < core 4 ≈ core 2 tetrasaccharide ≤ core 2 hexasaccharide. Thereby, Surf. K_D values for LecB recognition of type-1 α1,2-fucosylated core structures were determined to be 0.82 µM, 0.64 µM, 0.55 µM, 0.54 µM and 0.53 µM, respectively. In contrast, LecB showed lower Surf. K_D values for the corresponding type-2 α1,2-fucosylated type-2 cores with 0.60 µM , 0.48 µM, 0.37 µM, 0.37 µM and 0.28 µM, respectively.

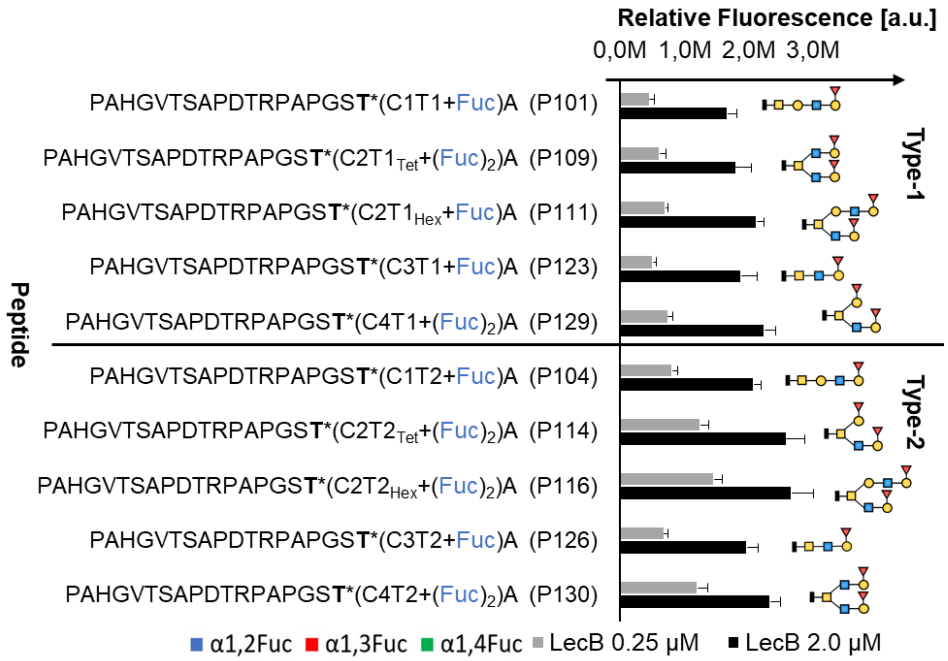


Figure 30. Comparison of LecB-binding affinities toward type-1 and type-2 H-antigen MUC1 glycopeptides. [a.u.] = arbitrary units.

Microarray analysis of LecB binding preferences toward fucosylated glycopeptides showed that peptides glycosylated in the GSTA region were better binders than the respective derivatives glycosylated in the PDTR region (Figure 31, Table 5). These findings indicate that the exact glycosylation sites play important roles for lectin binding.

Additionally, the impact of bivalent ligand presentation on the MUC1 peptide backbone on LecB recognition was evaluated. Generally, the lectin showed higher avidities for bivalent rather than for monovalent MUC1 glycopeptides (Figure 31, Table 5). For example, the bivalent H-antigen glycopeptide **P132** (Surf. K_D = 0.50 µM) was a better binder than the corresponding monovalent PDTR **P87** (Surf. K_D = 1.10 µM) and GSTA **P101** (Surf. K_D = 0.64 µM) peptides. The same LecB recognition pattern was also observed for the mono- and bivalent Le^a and Le^x glycopeptides.

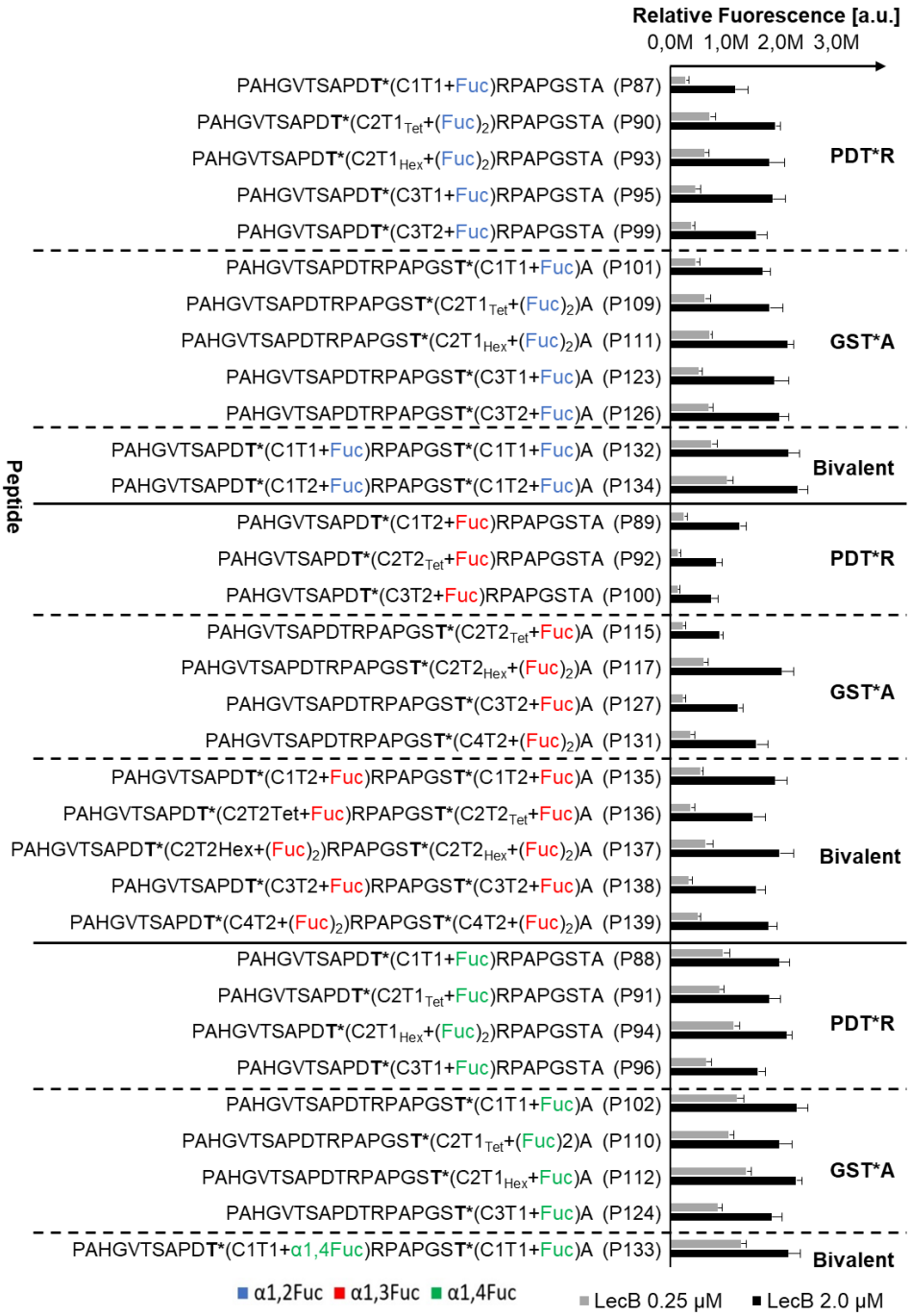


Figure 31. Comparison of LecB-binding affinities toward mono- and bi-fucosylated MUC1 glycopeptides. [a.u.] = arbitrary units.

Table 5. Surf. K_D values for LecB binding toward mono- and bi-fucosylated MUC1 core structures.

ID	Glycopeptide sequence	K_D ± SEM [µM]
P87	PAHGVTSAPD**T*(C1T1+Fuc)RPAPGSTA	1.10 ± 0.10
P90	PAHGVTSAPD**T*(C2T1$_{Tet}$+(Fuc)$_2$)RPAPGSTA	0.55 ± 0.04
P93	PAHGVTSAPD**T*(C2T1$_{Hex}$+(Fuc)$_2$)RPAPGSTA	0.63 ± 0.05
P95	PAHGVTSAPD**T*(C3T1+Fuc)RPAPGSTA	1.02 ± 0.07
P99	PAHGVTSAPD**T*(C3T2+Fuc)RPAPGSTA	1.17 ± 0.09
P101	PAHGVTSAPDTRPAPGS**T*(C1T1+Fuc)A	0.64 ± 0.03
P104	PAHGVTSAPDTRPAPGS**T*(C1T2+Fuc)A	0.48 ± 0.03
P109	PAHGVTSAPDTRPAPGS**T*(C2T1$_{Tet}$+(Fuc)$_2$)A	0.54 ± 0.03
P111	PAHGVTSAPDTRPAPGS**T*(C2T1$_{Hex}$+(Fuc)$_2$)A	0.53 ± 0.03
P123	PAHGVTSAPDTRPAPGS**T*(C3T1+Fuc)A	0.82 ± 0.05
P126	PAHGVTSAPDTRPAPGS**T*(C3T2+Fuc)A	0.60 ± 0.04
P132	PAHGVTSAPD**T*(C1T1+Fuc)RPAPGS**T*(C1T1+Fuc)A	0.50 ± 0.03
P134	PAHGVTSAPD**T*(C1T2+Fuc)RPAPGS**T*(C1T2+Fuc)A	0.35 ± 0.02
P89	PAHGVTSAPD**T*(C1T2+Fuc)RPAPGSTA	1.19 ± 0.09
P92	PAHGVTSAPD**T*(C2T2$_{Tet}$+Fuc)RPAPGSTA	2.13 ± 0.23
P115	PAHGVTSAPDTRPAPGS**T*(C2T2$_{Tet}$+Fuc)A	1.72 ± 0.19
P117	PAHGVTSAPDTRPAPGS**T*(C2T2$_{Hex}$+(Fuc)$_2$)A	0.91 ± 0.06
P127	PAHGVTSAPDTRPAPGS**T*(C3T2+Fuc)A	1.76 ± 0.15
P131	PAHGVTSAPDTRPAPGS**T*(C4T2+(Fuc)$_2$)A	1.17 ± 0.09
P135	PAHGVTSAPD**T*(C1T2+Fuc)RPAPGS**T*(C1T2+Fuc)A	0.79 ± 0.05
P136	PAHGVTSAPD**T*(C2T2$_{Tet}$+Fuc)RPAPGS**T*(C2T2$_{Tet}$+Fuc)A	1.12 ± 0.11
P137	PAHGVTSAPD**T*(C2T2$_{Hex}$+(Fuc)$_2$)RPAPGS**T*(C2T2$_{Hex}$+(Fuc)$_2$)A	0.54 ± 0.04
P138	PAHGVTSAPD**T*(C3T2+Fuc)RPAPGS**T*(C3T2+Fuc)A	1.26 ± 0.08
P139	PAHGVTSAPD**T*(C4T2+(Fuc)$_2$)RPAPGS**T*(C4T2+(Fuc)$_2$)A	0.75 ± 0.04
P88	PAHGVTSAPD**T*(C1T1+Fuc)RPAPGSTA	0.36 ± 0.03
P91	PAHGVTSAPD**T*(C2T1$_{Tet}$+Fuc)RPAPGSTA	0.31 ± 0.02
P94	PAHGVTSAPD**T*(C2T1$_{Hex}$+(Fuc)$_2$)RPAPGSTA	0.24 ± 0.02
P96	PAHGVTSAPD**T*(C3T1+Fuc)RPAPGSTA	0.42 ± 0.03
P102	PAHGVTSAPDTRPAPGS**T*(C1T1+Fuc)A	0.26 ± 0.02
P110	PAHGVTSAPDTRPAPGS**T*(C2T1$_{Tet}$+Fuc)A	0.25 ± 0.02
P112	PAHGVTSAPDTRPAPGS**T*(C2T1$_{Hex}$+(Fuc)$_2$)A	0.19 ± 0.01
P124	PAHGVTSAPDTRPAPGS**T*(C3T1+Fuc)A	0.31 ± 0.02
P133	PAHGVTSAPD**T*(C1T1+Fuc)RPAPGS**T*(C1T1+Fuc)A	0.16 ± 0.01

Fuc = α1,2Fuc, Fuc = α1,3Fuc, Fuc = α1,4Fuc

Next, LecB binding toward a selection of fucosylated LacNAc elongated MUC1 glycopeptides was evaluated (Figure 32, Table 6). Previous studies reported that the chain length of oligosaccharides carrying Le^a and Le^x motifs can enhance the binding affinity of LecB.[206] In line with these findings, LacNAc elongation of different Le^a and Le^x modified mucin cores was found to increase LecB-binding up to 3.3-fold in comparison with the corresponding non-extended analogs. These results indicate that the additional fucose residue on the LacNAc extension contributes to avidity enhancement by allowing bidentate or multivalent binding interactions with LecB. Furthermore, LacNAc extended H-antigen peptides were better binders than the respective Le^x glycopeptides. For example, the LacNAc elongated H type glycopeptides **P106** (Surf. K_D = 0.32 µM) and **P97** (Surf. K_D = 0.54 µM) were better binders than the respective α1,3-fucosylated analogs **P108** (Surf. K_D = 0.90 µM) and **P98** (Surf. K_D = 0.91 µM).

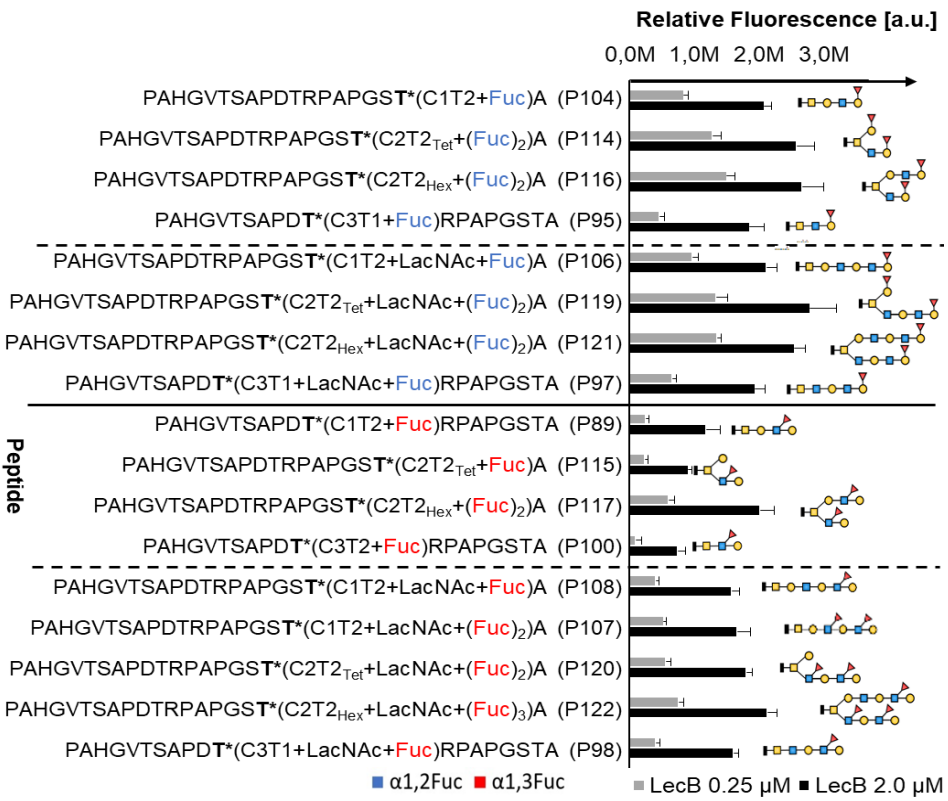


Figure 32. Influence of LacNAc elongation on LecB-binding towards H-type and Le^x modified MUC1 glycopeptides. [a.u.] = arbitrary units.

In case of α1,2-fucosylation, the avidity was not increased by LacNAc elongation of core 2 glycopeptides **P119** (Surf. K_D = 0.34 µM) and **P121** (Surf. K_D = 0.29 µM) compared to their non-elongated analogs **P114** (Surf. K_D = 0.37 µM) and **P116** (Surf. K_D = 0.28 µM). This is likely due to the fact that only terminal galactose residues of glycan chains are α1,2-fucosylated and thus the added LacNAc unit cannot contribute with an additional fucose residue to the glycan-lectin binding interaction. Nevertheless, LecB binding toward LacNAc extended core 1 and core 3 structures was slightly enhanced. These findings indicate that LacNAc extension of the core 1 and core 3 chains leads to an optimized ligand presentation for the LecB binding pockets. However, these results stand in contrast to a previous study on *N*-glycans, where LecB showed an enhanced affinity for a di-LacNAc bi-antennary H type-2 glycan in comparison with the mono-LacNAc analog, which was caused by a favored sterical fit towards the LecB binding pockets.[210] Since mucin type *O*-glycans are structurally very different from the mannose containing *N*-glycans, it can be concluded that the spatial orientations of the mucin core branches in relation to the LecB binding pockets promote different binding modes. As a result, different LecB binding preferences for *O*-glycans can be observed.

Table 6. Surf. K_D values for LecB binding toward LacNAc elongated and non-extended fucosylated MUC1 core structures.

ID	Glycopeptide sequence	K_D ± SEM [µM]
P104	PAHGVTSAPDTRPAPGS**T*** (C1T2+Fuc)A	0.48 ± 0.03
P114	PAHGVTSAPDTRPAPGS**T*** (C2T2$_{Tet}$+(Fuc)$_2$)A	0.37 ± 0.03
P116	PAHGVTSAPDTRPAPGS**T*** (C2T2$_{Hex}$+(Fuc)$_2$)A	0.28 ± 0.03
P95	PAHGVTSAPD**T*** (C3T1+Fuc)RPAPGSTA	1.02 ± 0.07
P106	PAHGVTSAPDTRPAPGS**T*** (C1T2+LacNAc+Fuc)A	0.32 ± 0.02
P119	PAHGVTSAPDTRPAPGS**T*** (C2T2$_{Tet}$+LacNAc+(Fuc)$_2$)A	0.34 ± 0.03
P121	PAHGVTSAPDTRPAPGS**T*** (C2T2$_{Hex}$+LacNAc+(Fuc)$_2$)A	0.29 ± 0.02
P97	PAHGVTSAPD**T*** (C3T1+LacNAc+Fuc)RPAPGSTA	0.54 ± 0.03
P89	PAHGVTSAPD**T*** (C1T2+Fuc)RPAPGSTA	1.19 ± 0.09
P115	PAHGVTSAPDTRPAPGS**T*** (C2T2$_{Tet}$+Fuc)A	1.72 ± 0.19
P117	PAHGVTSAPDTRPAPGS**T*** (C2T2$_{Hex}$+(Fuc)$_2$)A	0.91 ± 0.06
P100	PAHGVTSAPD**T*** (C3T2+Fuc)RPAPGSTA	2.97 ± 0.32
P108	PAHGVTSAPDTRPAPGS**T*** (C1T2+LacNAc+Fuc)A	0.90 ± 0.05
P107	PAHGVTSAPDTRPAPGS**T*** (C1T2+LacNAc+(Fuc)$_2$)A	0.79 ± 0.06
P120	PAHGVTSAPDTRPAPGS**T*** (C2T2$_{Tet}$+LacNAc+(Fuc)$_2$)A	0.76 ± 0.04
P122	PAHGVTSAPDTRPAPGS**T*** (C2T2$_{Hex}$+LacNAc+(Fuc)$_3$)A	0.49 ± 0.03
P98	PAHGVTSAPD**T***(C3T1+LacNAc+Fuc)RPAPGSTA	0.91 ± 0.05

Fuc = α1,2Fuc, Fuc = α1,3Fuc

Furthermore, the influence of double fucosylated Le^b and Le^y core structures on LecB binding was explored. The higher fucose content in Le^b antigens did not enhance the LecB binding affinity due to avidity effects (Figure 33, Table 7).[205, 207] In fact, these glycopeptides were even weaker binders than their corresponding H-antigen and Le^a modified glycopeptide analogs. These results suggest that the additional fucose residue on the same core structure may sterically hinder LecB binding. In contrast, Le^y antigen peptides were better LecB binders than the respective Le^x glycans, but showed decreased avidities compared with the H-type glycopeptides.

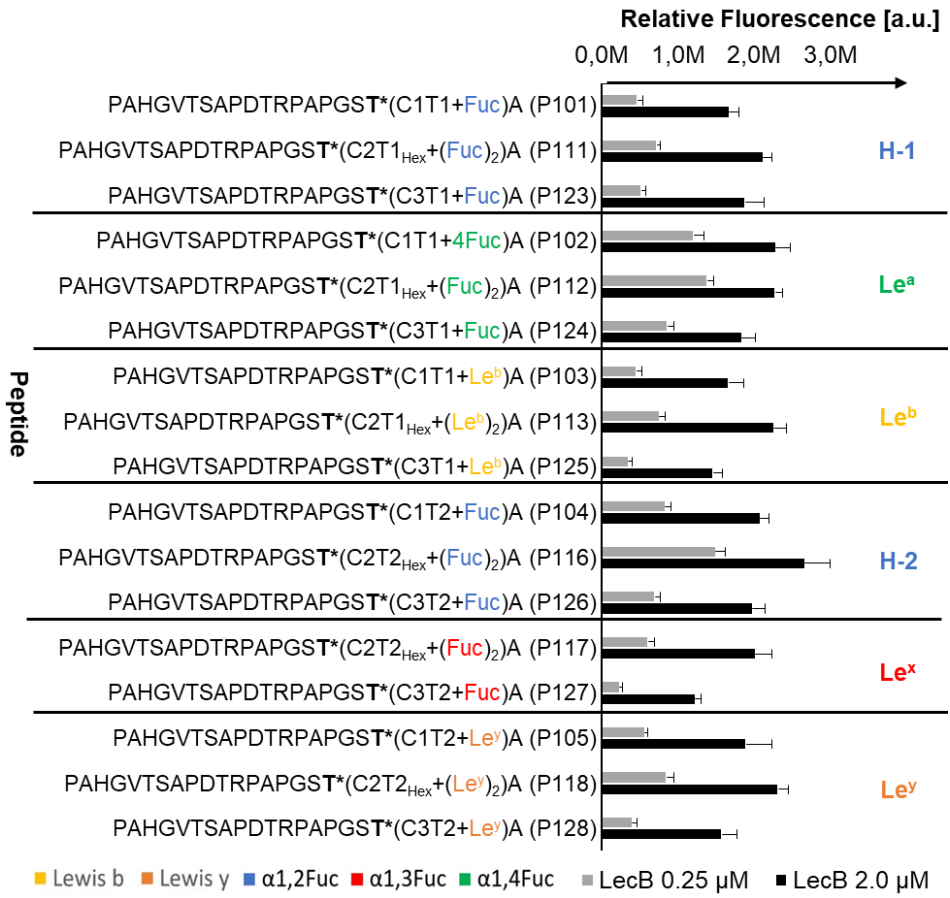


Figure 33. Comparison of LecB binding between double-fucosylated Le^b and Le^y antigens on MUC1 glycopeptides, and their respective mono-fucosylated derivatives.

Table 7. Surf. K_D values for LecB binding toward double-fucosylated Le^b and Le^y antigens on MUC1 glycopeptides, and their respective mono-fucosylated derivatives.

ID	Glycopeptide sequence	K_D ± SEM [μM]
P101	PAHGVTSAPDTRPAPGST*(C1T1+α1,2Fuc)A	0.64 ± 0.03
P111	PAHGVTSAPDTRPAPGST*(C2T1$_{Hex}$+(α1,2Fuc)$_2$)A	0.53 ± 0.03
P123	PAHGVTSAPDTRPAPGST*(C3T1+α1,2Fuc)A	0.82 ± 0.05
P102	PAHGVTSAPDTRPAPGST*(C1T1+α1,4Fuc)A	0.26 ± 0.02
P112	PAHGVTSAPDTRPAPGST*(C2T1$_{Hex}$+(α1,4Fuc)$_2$)A	0.19 ± 0.01
P124	PAHGVTSAPDTRPAPGST*(C3T1+α1,4Fuc)A	0.31 ± 0.02
P103	PAHGVTSAPDTRPAPGST*(C1T1+Le^b)A	0.82 ± 0.05
P113	PAHGVTSAPDTRPAPGST*(C2T1$_{Hex}$+(Le^b)$_2$)A	0.54± 0.03
P125	PAHGVTSAPDTRPAPGST*(C3T1+Le^b)A	1.06 ± 0.06
P104	PAHGVTSAPDTRPAPGST*(C1T2+α1,2Fuc)A	0.48 ± 0.03
P116	PAHGVTSAPDTRPAPGST*(C2T2$_{Hex}$+(α1,2Fuc)$_2$)A	0.28 ± 0.03
P126	PAHGVTSAPDTRPAPGST*(C3T2+α1,2Fuc)A	0.60 ± 0.04
P117	PAHGVTSAPDTRPAPGST*(C2T2$_{Hex}$+(α1,3Fuc)$_2$)A	0.91 ± 0.06
P127	PAHGVTSAPDTRPAPGST*(C3T2+α1,3Fuc)A	1.76 ± 0.15
P105	PAHGVTSAPDTRPAPGST*(C1T2+Le^y)A	0.72 ± 0.05
P118	PAHGVTSAPDTRPAPGST*(C2T2$_{Hex}$+(Le^y)$_2$)A	0.51 ± 0.03
P128	PAHGVTSAPDTRPAPGST*(C3T2+Le^y)A	1.18 ± 0.08

Finally, the fine binding specificities of LecB toward selected MUC5B glycopeptides were determined. The observed binding patterns were in line with the obtained MUC1 data (Figure 34, Table 8). For example, the H-type glycopeptides **P140** (Surf. K_D = 0.74 µM) and **P142** (Surf. K_D = 0.39 µM) were better binders than the corresponding Le^x analogy **P141** (Surf. K_D = 2.19 µM) and **P143** (Surf. K_D = 1.45 µM). Thereby, LacNAc elongation on these Muc5B glycopeptides also enhanced LecB binding. As observed for the MUC1 glycopeptides, Le^x antigen glycopeptides were recognized in the increasing affinity order: core 3 < core 1 < core 2 hexasaccharide. However, bivalent glycan presentation on the MUC5B peptide backbone did not have a major impact on LecB-binding. Unexpectedly, the monovalent α1,3-fucosylated glycopeptide **P141** was even a slightly better binder than both corresponding bivalent peptides; **P146** (Surf. K_D = 2.32 µM) and **P149** (Surf. K_D = 2.91 µM). Additionally, the exact placement of the second threonine glycosylation site influenced LecB binding. Here, the avidity toward all bivalent peptides was increased when the spatial distance between the glycosylation sites was bigger. These findings indicate that the ligand presentation on the MUC5B peptide backbone as well as on distinct glycosylation sites is important for LecB binding, as well as for binding strength enhancement through bidentate and multivalent binding interactions.

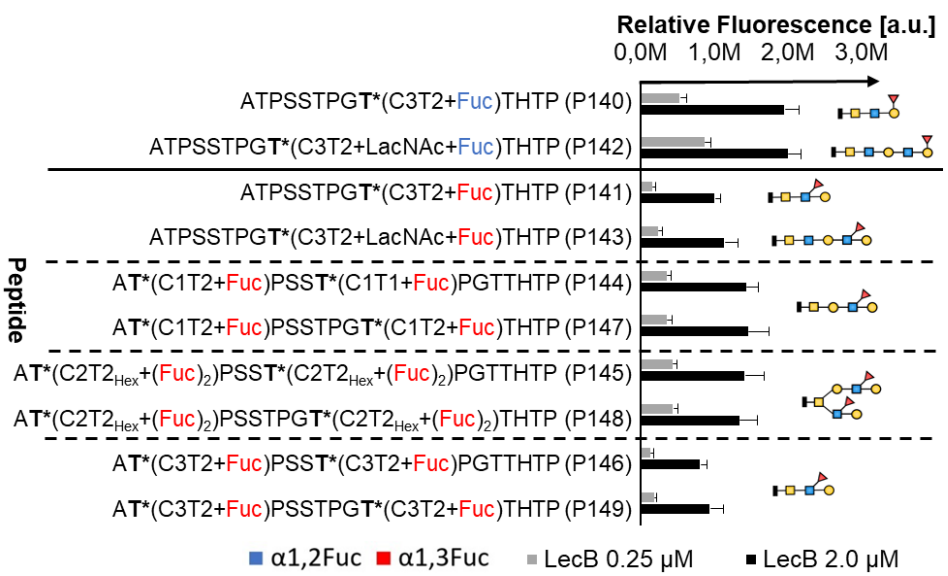


Figure 34. LecB-binding toward different H type and Le^x MUC5B core structures.

Table 8. Surf. K_D values for LecB binding toward different H type and Le^x MUC5B core structures.

ID	Glycopeptide sequence	K_D ± SEM [µM]
P140	ATPSSTPG**T***(C3T2+Fuc)THTP	0.74 ± 0.05
P142	ATPSSTPG**T***(C3T2+LacNAc+Fuc)THTP	0.39 ± 0.02
P141	ATPSSTPG**T***(C3T2+Fuc)THTP	2.19 ± 0.24
P143	ATPSSTPG**T***(C3T2+LacNAc+Fuc)THTP	1.45 ± 0.13
P144	A**T***(C1T2+Fuc)PSS**T***(C1T1+Fuc)PGTTHTP	0.95 ± 0.06
P147	A**T***(C1T2+Fuc)PSSTPG**T***(C1T2+Fuc)THTP	0.96 ± 0.08
P145	A**T***(C2T2$_{Hex}$+(Fuc)$_2$)PSS**T***(C2T2$_{Hex}$+(Fuc)$_2$)PGTTHTP	0.79 ± 0.07
P148	A**T***(C2T2$_{Hex}$+(Fuc)$_2$)PSSTPG**T***(C2T2$_{Hex}$+(Fuc)$_2$)THTP	0.87 ± 0.08
P146	A**T***(C3T2+Fuc)PSS**T***(C3T2+Fuc)PGTTHTP	2.32 ± 0.31
P149	A**T***(C3T2+Fuc)PSSTPG**T***(C3T2+Fuc)THTP	2.91 ± 0.44

Fuc = α1,2Fuc, Fuc = α1,3Fuc

In conclusion, the microarray binding study showed that the fine binding specificities of LecB toward fucosylated MUC1 and MUC5B glycopeptides strongly depended on the different fucose motifs, the ligand presentation on the underlying core structures, LacNAc extension as well as on the particular glycosylation sites on the respective peptide backbones. The microarray analysis showed that LecB exhibited a broad selectivity toward all fucosylated MUC1 and MUC5B peptides, as it bound to the fucose antigens presented on varying core structures with different affinities. In line with previous studies, H-antigen type-2 glycopeptides were better binders than the type-1 analogs.[205, 207] Additionally, the placement of the glycosylation sites was shown to impact lectin binding. For example, higher binding avidities were observed for MUC1 peptides glycosylated in the GSTA region than for the respective derivatives glycosylated in the PDTR region. Whereas bivalent ligand presentation on MUC1 peptides generally increased the avidity, no major impact on lectin binding was observed for bivalent MUC5B glycopeptides. Furthermore, LacNAc extension of MUC1 and MUC5B glycans also generally enhanced LecB binding, with the increase in magnitude depending on the underlying core structure. Lastly, LecB binding toward double fucosylated mucin core structures showed that Le^b antigen glycopeptides were weaker LecB binders than the corresponding H-antigen and Le^a analogs, and Le^y core

structures were better binders than the respective Le^x glycans. However, Le^y antigens showed weaker binding compared with the H type glycopeptides.

4.1.2.3 Recognition of fucosylated glycopeptides by the *Clostridium difficile* toxin A

To determine the fine binding specificities of TcdA, the fucosylated MUC1 and MUC5B glycopeptide library was incubated with a dilution series of TcdA (27 nM – 3.5 µM), followed by detection using the mouse anti-TcdA mAb TGC-2 and an anti-mouse Cy5-labeled antibody for fluorescence readout. Additionally, apparent surface dissociation constants (Surf. K_D) for LecB binding toward the fucosylated MUC1 and MUC5B glycopeptides were calculated. Curve fitting was carried out by non-linear regression using the saturation binding - specific binding with Hill Slope equation in GraphPad Prism 8.

This lectin was previously reported to recognize the Galili epitope, but also fucosylated Le^y, Le^x, Sialyl-Le^x and sulfo-Le^x glycans.[199-201] Microarray analysis showed that the *C. difficile* toxin A selectively recognized α1,3-fucosylated Muc1 and Muc5B core structures. TcdA bound to Le^x and Le^y antigen glycopeptides in the high nanomolar to low micromolar range (Surf. K_D = 0.28 - 2.46µM). Additionally, the fine TcdA binding specificities depended on the underlying core structures, LacNAc-extension as well as on the respective glycosylation sites. H type, Le^a and Le^b glycopeptides were not recognized.

MUC1 peptides glycosylated in the PDTR and GSTA with different Le^x modified core structures were recognized by TcdA in the increasing order C2T2 tetrasaccharide < C1T2 < C3T2, and C2T2 tetrasaccharide < C4T2 < C2T2 hexasaccharide ≤ C3T2 (Figure 35,Table 9). These results suggest, that even though branched structures exhibit two Le^x units they are less favored than the linear core 1 and core 3 glycans presenting only one Le^x residue. Additionally, the TcdA showed preferred binding toward peptides **P115** (Surf. K_D = 2.01 µM) and **127** (Surf. K_D = 0.93 µM) that were modified in the GSTA region over the respective peptides **P92** (Surf. K_D = 2.46 µM) and (Surf. K_D = 2.15 µM) **P100** that were glycosylated in the PDTR region. As observed in the LecB binding study, bivalent glycopeptides showed higher binding avidities compared with the respective monovalent glycopeptides. These findings can again be related to the multivalent binding effect.

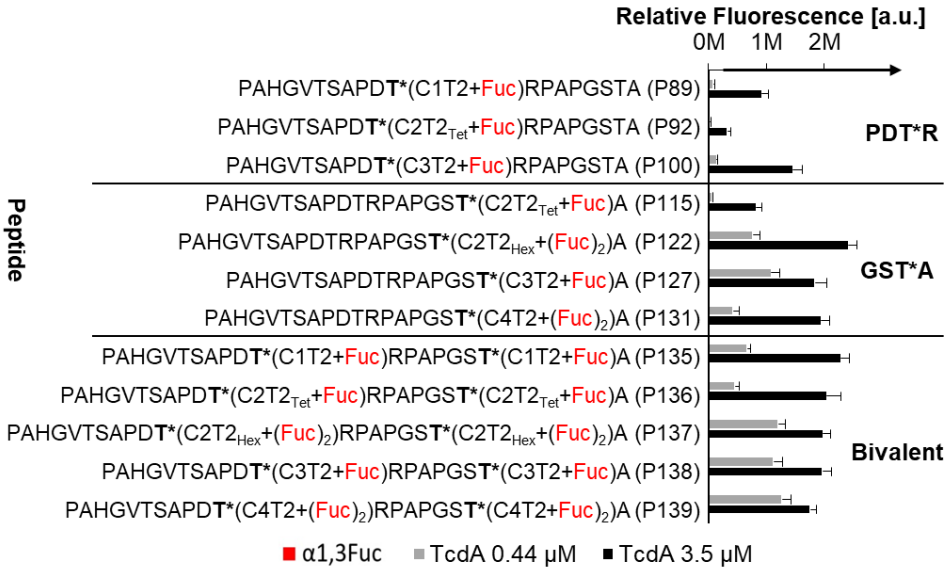


Figure 35. Comparison of TcdA binding affinities toward mono- and bi-fucosylated MUC1 glycopeptides. [a.u.] = arbitrary units.

Table 9. Surf. K_D values for TcdA binding toward mono- and bi-fucosylated MUC1 Le^x modified core structures.

ID	Glycopeptide sequence	K_D ± SEM [µM]
P89	PAHGVTSAPD**T***(C1T2+Fuc)RPAPGSTA	1.94 ± 0.15
P92	PAHGVTSAPD**T***(C2T2$_{Tet}$+Fuc)RPAPGSTA	2.46 ± 0.19
P100	PAHGVTSAPD**T***(C3T2+Fuc)RPAPGSTA	2.15 ± 0.25
P115	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Tet}$+Fuc)A	2.01 ± 0.15
P122	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Hex}$+LacNAc+(Fuc)$_3$)A	1.16 ± 0.12
P127	PAHGVTSAPDTRPAPGS**T***(C3T2+Fuc)A	0.93 ± 0.08
P131	PAHGVTSAPDTRPAPGS**T***(C4T2+(Fuc)$_2$)A	0.4 ± 0.04
P135	PAHGVTSAPD**T***(C1T2+Fuc)RPAPGS**T***(C1T2+Fuc)A	1.28 ± 0.16
P136	PAHGVTSAPD**T***(C2T2$_{Tet}$+Fuc)RPAPGS**T***(C2T2$_{Tet}$+Fuc)A	1.24 ± 0.14
P137	PAHGVTSAPD**T***(C2T2$_{Hex}$+(Fuc)$_2$)RPAPGS**T*** (C2T2$_{Hex}$+(Fuc)$_2$)A	0.28 ± 0.03
P138	PAHGVTSAPD**T***(C3T2+Fuc)RPAPGS**T***(C3T2+Fuc)A	0.51 ± 0.05
P139	PAHGVTSAPD**T***(C4T2+(Fuc)$_2$)RPAPGS**T***(C4T2+(Fuc)$_2$)A	0.23 ± 0.03

Fuc = α1,3Fuc

Next, the influence of LacNAc extension of the MUC1 core structures on TcdA recognition was explored. Surprisingly, LacNAc elongation decreased TcdA binding dramatically (Figure 36, Table 10). Even though the LacNAc elongated glycopeptides **P120** and **P122** (Surf. K_D = 1.16 µM) present an additional fucose unit for multivalent binding, the corresponding shorter peptides **P115** (Surf. K_D = 2.01 µM) and **P117** (Surf. K_D = 0.79 µM), respectively, might have a better sterical fit to the TcdA binding sites. Also, the more rigid shorter structures might lead to a benifical entropy effect.

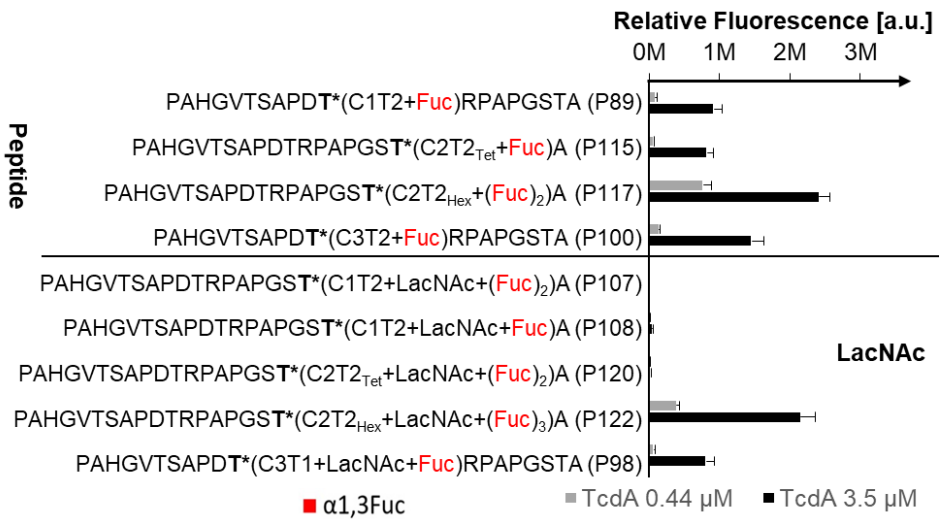


Figure 36. Influence of LacNAc elongation on TcdA binding towards Le^x modified MUC1 glycopeptides. [a.u.] = arbitrary units.

Table 10. Surf. K_D values for TcdA binding toward LacNAc elongated Le^x modified MUC1 glycopeptides.

ID	Glycopeptide sequence	K_D ± SEM [µM]
P89	PAHGVTSAPD**T***(C1T2+Fuc)RPAPGSTA	1.94 ± 0.15
P115	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Tet}$+Fuc)A	2.01 ± 0.15
P117	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Hex}$+(Fuc)$_2$)A	0.79 ± 0.07
P100	PAHGVTSAPD**T***(C3T2+Fuc)RPAPGSTA	2.15 ± 0.25
P122	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Hex}$+LacNAc+(Fuc)$_3$)A	1.16 ± 0.12
P98	PAHGVTSAPD**T***(C3T1+LacNAc+Fuc)RPAPGSTA	1.98 ± 0.21

Fuc = α1,3Fuc

Bivalent fucose presentation on the same glycan structure did not enhance TcdA binding avidities (Figure 37, Table 11). Here, TcdA showed similar binding to the branched Le^y modified peptide **P118** (Surf. K_D = 0.75 µM) as to the corresponding Le^x derivative **P117** (Surf. K_D = 0.79 µM). The linear Le^y glycopeptide **P128** (Surf. K_D = 1.42 µM) was an even weaker binder than the respective Le^x analog **P127** (Surf. K_D = 0.93 µM). These findings suggest that the 1,2-fucosylation on the terminal Gal residue either does not add an stabilizing effect by interaction with TcdA, or does not participate in the binding interaction with TcdA.

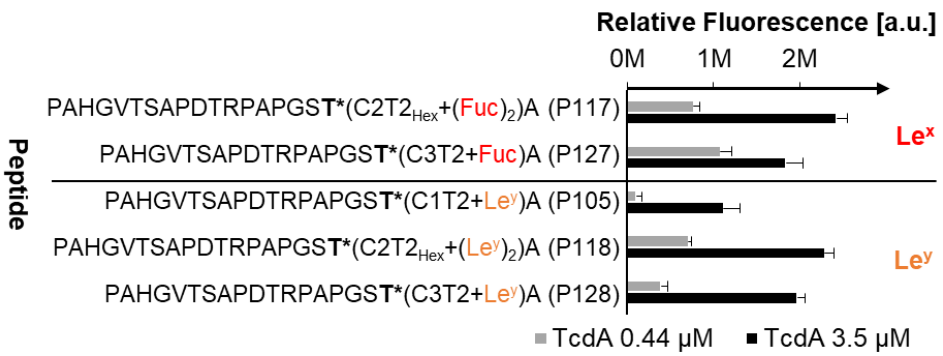


Figure 37. Comparison of TcdA binding toward double-fucosylated MUC1 Le^y antigens, and their respective mono-fucosylated Le^x modified derivatives. [a.u.] = arbitrary units.

Table 11. Surf. K_D values for TcdA binding toward double-fucosylated MUC1 Le^y antigens, and their respective mono-fucosylated Le^x modified derivatives.

ID	Glycopeptide sequence	K_D ± SEM [µM]
P117	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Hex}$+(Fuc)$_2$)A	0.79 ± 0.07
P127	PAHGVTSAPDTRPAPGS**T***(C3T2+Fuc)A	0.93 ± 0.08
P105	PAHGVTSAPDTRPAPGS**T***(C1T2+Le^y)A	2.15 ± 0.15
P118	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Hex}$+(Le^y)$_2$)A	0.75 ± 0.05
P128	PAHGVTSAPDTRPAPGS**T***(C3T2+Le^y)A	1.42 ± 0.12

Fuc = α1,3Fuc

Finally, the fine binding specificities of TcdA toward fucosylated MUC5B glycopeptides were determined. As observed for MUC1 glycopeptides, TcdA did not recognize H type MUC5B glycopeptides and specifically bound to α1,3-fucosylated glycans (Figure 38, Figure 12). In line with the MUC1 data, LacNAc elongation decreased TcdA binding for glycopeptide **P143** (Surf. K_D = 2.08 µM) compared with the respective shorter analog **P141** (Surf. K_D = 1.31 µM). Additionally, TcdA showed enhanced binding to the bivalent Le^x peptides **P146** (Surf. K_D = 0.63 µM) and **P149** (Surf. K_D = 0.59 µM) compared with the monovalent glycopeptide **P141**. The placement of the second glycosylation site also only had a minor impact on lectin binding. Here, the avidities for peptides **P147** (Surf. K_D = 1.30 µM) and **P148** (Surf. K_D = 0.52 µM), which had with a bigger distance between the glycosylation sites, were slightly decreased compared with the binding strengths for glycopeptides **P144** (Surf. K_D = 1.19 µM) and **P145** (Surf. K_D = 0.39 µM)**.** In contrast to the MUC1 data, core specificity for bivalent Le^x glycopeptides was observed in the increasing avidity order C1T2 < C3T2 < C2T2Hex.

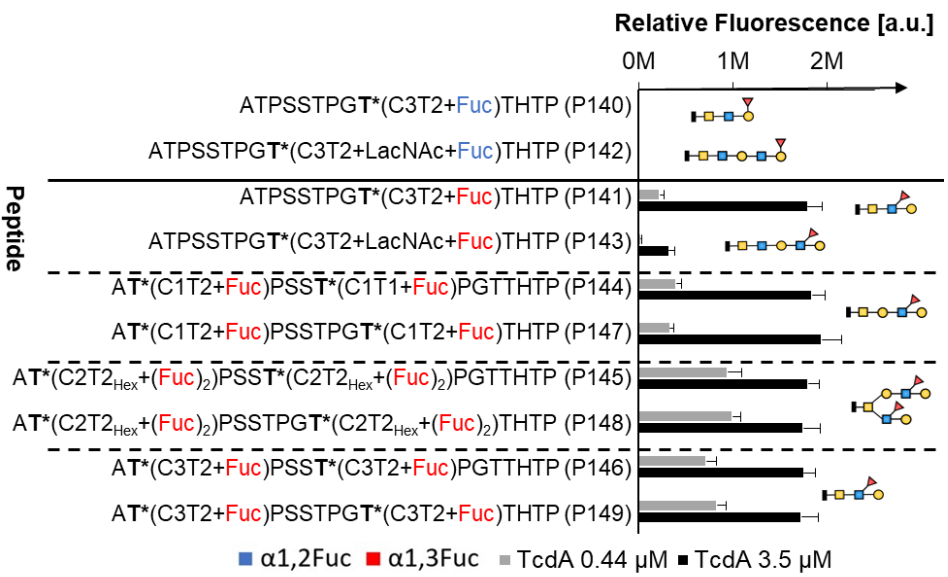


Figure 38. TcdA binding toward different H type and Le^x MUC5B core structures. [a.u.] = arbitrary units.

Table 12. Surf. K_D values for TcdA binding toward different Le^x MUC5B core structures.

ID	Glycopeptide sequence	K_D ± SEM [µM]
P141	ATPSSTPG**T*****(C3T2+Fuc)THTP	1.31 ± 0.09
P143	ATPSSTPG**T*****(C3T2+LacNAc+Fuc)THTP	2.08 ± 0.19
P144	A**T*****(C1T2+Fuc)PSS**T*****(C1T1+Fuc)PGTTHTP	1.19 ± 0.14
P147	A**T*****(C1T2+Fuc)PSSTPG**T*****(C1T2+Fuc)THTP	1.30 ± 0.10
P145	A**T*****(C2T2$_{Hex}$+(Fuc)$_2$)PSS**T*****(C2T2$_{Hex}$+(Fuc)$_2$)PGTTHTP	0.39 ± 0.04
P148	A**T*****(C2T2$_{Hex}$+(Fuc)$_2$)PSSTPG**T*****(C2T2$_{Hex}$+(Fuc)$_2$)THTP	0.52 ± 0.06
P146	A**T*****(C3T2+Fuc)PSS**T*****(C3T2+Fuc)PGTTHTP	0.63 ± 0.04
P149	A**T*****(C3T2+Fuc)PSSTPG**T*****(C3T2+Fuc)THTP	0.59 ± 0.05

Fuc = α1,3Fuc

In conclusion, the microarray binding study showed that the fine binding specificities of TcdA toward fucosylated MUC1 and MUC5B glycopeptides strongly depended on the ligand presentation on the underlying core structures, LacNAc extension as well as on the particular glycosylation sites on the respective peptide backbones. TcdA selectively recognized α1,3-fucosylated MUC1 and MUC5B peptides and bound to the Le^x antigens presented on varying core structures with different affinities. Additionally, the placement of the glycosylation sites impacted lectin binding. TcdA also showed preferred binding toward MUC1 peptides glycosylated in the GSTA region over the respective PDTR-modified peptides, and the spacing of MUC5B glycosylation sites slightly influenced the binding strength. While bivalent ligand presentation on MUC1 and MUC5B peptides generally increased the overall avidity, LacNAc elongation on different core structures decreased TcdA recognition. Lastly, TcdA showed similar binding strengths toward double fucosylated Le^y MUC1 core structures as well as to the respective monofucosylated Le^x glycopeptides.

4.1.3 Conclusion

In summary, a library of fucosylated mucin core 1-4 MUC1 and MUC5B tandem repeat glycopeptides was generated to study the fine binding specificities of the fucose-recognizing bacterial lectins LecB from *P. aeruginosa* and toxin A from *C. difficile.* Selected glycopeptides were extended with additional LacNAc using the *H. pylori* β-1,3-*O*-*N*-acetylglucosaminyltransferase (β3GlcNAcT)and a fusion protein of human β-1,4-*O*-galactosyltransferase (His_6-Propeptide-catβ4GalT-1, β4GalT). Subsequently, the different fucose motifs, including the Le^a, Le^x and H-type as well as bi-fucosylated Le^b and Le^y antigens, were enzymatically coupled using *H. pylori* α1,3/4-*O*-fucosyltransferase and/or *H. mustelae* α1,2-*O*-fucosyltransferase. The order of the applied fucosyltransferases was also crucial to prepare the Le^b and Le^y determinants. The obtained fucosylated mucin glycopeptide library was printed on NHS-activated microarrays, which was then applied to determine the binding preferences of LecB and TcdA. Whereas TcdA exclusively bound to α1,3-fucosylated Muc1 and Muc5B core structures, LecB exhibited a broader selectivity towards all fucosylated glycopeptides. Additionally, both lectins exhibited unique fine specificities that strongly depended on the different fucose motifs, presenting peptide backbone, underlying core structures, LacNAc-extension as well as placement of the glycosylation sites on the MUC1 and MUC5B glycopeptides.

These findings highlight the importance of the evaluated structural glycopeptide properties in lectin binding interactions since they can define the glycan orientation, structural rigidity or possible limitations for ligand recognition, which may be essential for the biological functions of these lectins. Based on the gained knowledge about the fine binding specificities of LecB and TcdA, novel glycoconjugates and -mimetics for anti-biofilm and anti-adhesion therapies to fight *P. aeruginosa* and *C. difficile* infections could be developed.

4.2 Synthesis of simplified core 1 to core 4 MUC1 and MUC5AC glycopeptides as scaffolds for enzymatic modifications (Paper IV)

4.2.1 LacdiNAc in cancer and bacterial lectin interactions

The terminal GalNAcβ-1,4-GlcNAc (LacdiNAc) motif is expressed on the epithelial surface of many organs including the salivary glands, heart, stomach, small intestines and colon where it was found to modify glycan chains of both *N*- and *O*-glycans (Figure 39).[211-212]

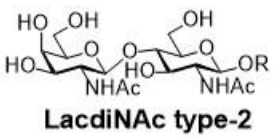


Figure 39. Structure of the LacdiNAc type-2 motif.

LacdiNAc is predominantly found on *N*-glycans and several mammalian *N*-glycoproteins are known to carry this modification.[213-216] Additionally, LacdiNAc was found to modify several *O*-glycoproteins such as the murine zona pellucida glycoprotein 3 responsible for initial sperm-egg binding,[217] the bovine pro-opiomelanocortin that is a precursor for corticotropin and endorphin,[218] as well as the extracellular matrix related glycoproteins ECM1, AMACO, nidogen-1, α-dystroglycan and neurofascin.[219] The presence of LacdiNAc epitopes in the gastric mucosa[220] and on gastric *O*-glycans was recently reported where LacdiNAc modified core 2 and 3 *O*-glycans.[221-222] The LacdiNAc determinants on *N*- and *O*- glycans can be further modified by fucosylation, sialylation, and sulfation.[214, 223] Additionally, phosphorylation of the LacdiNAc GlcNAc residue [GalNAcβ-1,4-(phosphor)GlcNAc] has been found on *O*-glycans of recombinant and endogenous extracellular matrix/matrix-related proteins in both humans and bovines.[219]

Despite its relatively low abundance in mammalian glycoproteins, LacdiNAc and poly-LacdiNAc has been reported in various types of cancer, and can be used as an efficient diagnostic marker for human breast, pancreatic and ovarian cancer.[212] For example, the LacdiNAc expression level in *N*-glycans is decreased in breast cancer.[224] In contrast, its expression level is increased in prostate,[225] ovarian,[226] and pancreatic cancers[227]. These findings suggest that the up- or

downregulation of this carbohydrate motif depends on the tumor type. Even though, LacdiNAc is usually found on *N*-glycans of human cancer cells, a recent study reported that the β-4-N-acetylgalactosaminyltransferases β4GalNAcT3[228] and β4GalNAcT4,[229] which are responsible for the biosynthesis of LacdiNAc and are expressed in a tissue-specific manner, also transfer GalNAc to the β-1,6-linked residue of core 2 *O*-glycans.[230] Because abberant changes in the expression levels of these glycosyltransferases have been associated with various types of cancer, it might be possible that the LacdiNAc motif expressed on *O*-glycans also plays a role in human cancer. However, the exact functions of the LacdiNAc group on *O*-glycans together with its fucosylated, sialylated and/or sulfated forms have to be elucidated by future studies.

It was recently reported that LacdiNAc might be involved in the adhesion process of the bacterium *Helicobacter pylori. H. pylori* is a spiral-shaped Gram-negative bacterium that colonizes the human gastric mucosa and is associated with chronic gastritis and other severe gastroduodenal diseases such as peptic and gastric ulcers, gastric cancer and mucosa-associated lymphoid tissue lymphoma.[231-233] By infecting approximately 50% of the world's population, it is the most common bacterial infection worldwide.[234] *H. pylori* adherence to specific carbohydrate mucin epitopes on mucous epithelial cells and the mucus layer that covers the gastric epithelium is mediated by different adhesins including BabA and SabA.[235-236] Additionally, the LacdiNAc-binding adhesin LabA was reported to bind to LacdiNAc motifs on gastric Muc5Ac mucins.[211] However, more recent studies indicate that LacdiNAc might not be the physiological LabA ligand.[237-238] Microarray binding studies using LacdiNAc modified mucin glycopeptides could give an insight as to whether this disaccharide is a ligand for LabA.

4.2.2 Motivation

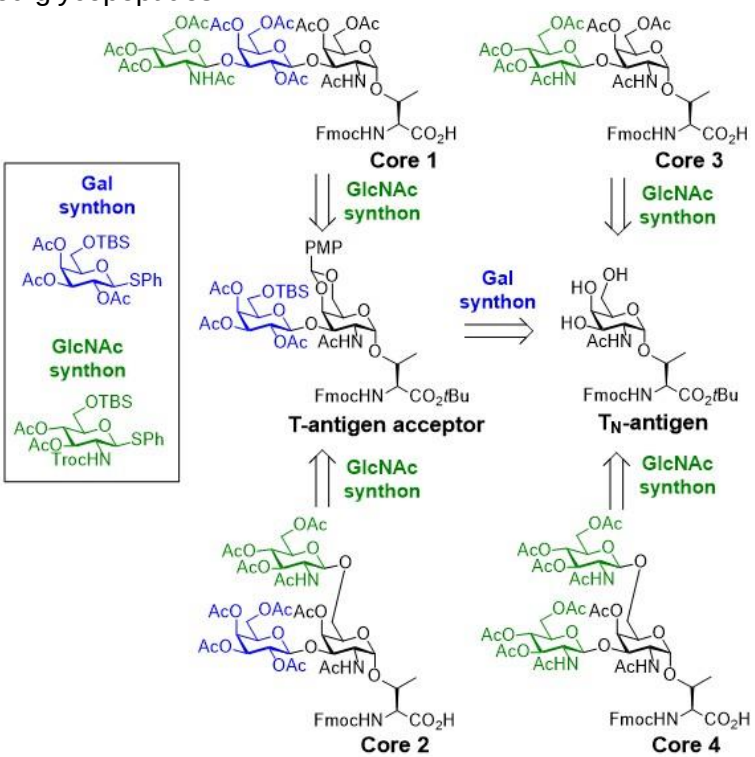
In previous work, LacdiNAc was found to modify terminal positions of both *N*- and *O*-glycan chains. However, the exact roles that LacdiNAc *O*-glycosylation plays in cancer, *Helicobacter pylori* adhesion or other LacdiNAc-protein interactions are not clear. In order to characterize interactions between glycans exhibiting the LacdiNAc motif and different LacdiNAc-binding proteins on a molecular level, I aimed to expand the previously prepared synthetic mucin glycopeptide library with LacdiNAc-modified mucin core structures. To achieve this goal, a unified synthesis strategy was designed to prepare MUC1 and mucin 5AC (MUC5AC) glycopeptides carrying simplified mucin core structures. The strategy employed the generation of a GlcNAc building block that was used to selectively extend GalNAc-Thr (T_N-antigen) or Gal-GalNAc-Thr (T-antigen) amino acids to generate the respective simplified cores 1 to 4. In the glycosylation reactions, orthogonal protecting groups that direct the reactivity and stereoselectivity of the reaction were applied. The obtained core-glycosylated Fmoc-protected threonine building blocks were then incorporated into Muc1 and Muc5AC tandem repeat sequences by Fmoc-SPPS to generate the desired simplified mucin core glycopeptides.

The obtained GlcNAc-elongated mucin core peptides could then be applied as scaffolds in enzymatic reactions to generate the desired type-2 LacdiNAc motifs. Finally, the obtained LacdiNAc-modified glycopeptide library could be used to prepare glycopeptide microarrays that can be applied in binding studies of the *H. pylori* adhesin LabA and other LacdiNAc-recognizing proteins. Using another enzymatic approach, the prepared simplified mucin core glycopeptides could also be interesting to generate a library consisting of sialylated, sulfated and/or fucosylated LacNAc or LacdiNAc modified glycopeptides.

4.2.3 Results and discussion

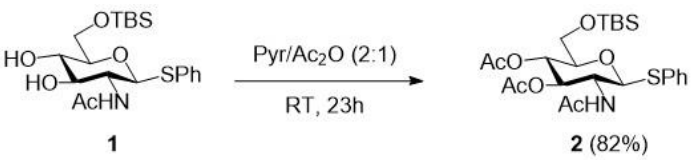
4.2.3.1 Synthesis of simplified mucin core threonine building blocks

To synthesize MUC1 and MUC5AC glycopeptides that are glycosylated on threonine glycosylation sites with simplified mucin core 1 to core 4 structures, Fmoc-protected GlcNAc-extended GalNAc-(T_N-antigen) and Gal-GalNAc-(T-antigen)-threonine building blocks were synthesized as simplified cores 1 to 4 (**Scheme 1**). The synthesis was achieved using orthogonal protecting groups that direct reactivity and stereoselectivity of the glycosylation reactions. The obtained Fmoc-protected amino acid building blocks were then incorporated into MUC1 and MUC5AC tandem repeat sequences by Fmoc-SPPS to generate the corresponding simplified glycopeptides.



Scheme 1. A unified synthesis strategy to prepare GlcNAc-extended core 1–4 threonine building blocks using common precursors. *t*Bu = *tert*-butyl, Troc = trichloroethoxycarbonyl, TBS = *tert*-butyl dimethylsilyl, PMP = *p*-methoxyphenyl.

The GlcNAc-elongated core 1-4 threonine amino acids were synthesized in preparative amounts in a few steps from common precursor building blocks using a convergent synthesis protocol. Therefore, the *N*-trichloroethoxycarbonyl (Troc) protected β-D-Glc thioglycoside **2** was designed as glycosyl donor to elongate previously reported T_N- or T-acceptor amino acids to form the respective core structures. A *tert*-butyldimethylsilyl (TBS) protecting group was introduced into the 6-position of glycosyl donor **2** to increase its reactivity, and an *N*-Troc group into the 2-position as neighboring participating β-directing group. Thioglycoside **2** was synthesized starting from the known Glc-*N*Troc-protected thioglycoside **1** in 82% yield by acetylation of the free hydroxyl functions using pyridine/acetic anhydride (Ac_2O) (Figure 40).



1 **2** (82%)

Figure 40. Synthesis of the β-D-Glc thioglycoside glycosyl donor **2**.

The obtained glycosyl donor **2** was then applied in glycosylation reactions to generate the desired GlcNAc-elongated Fmoc-protected core 1-4 threonine building blocks. The same synthesis strategy was applied to generate all simplified core amino acids: Glycosylation reactions were performed at 0 °C using the promoter system *N*-iodosuccinimide (NIS)/ trifluoromethanesulfonic acid (TfOH). Then, TBS and *p*-methoxyphenyl groups were removed by treatment with 80 % aqueous acetic acid, followed by acetylation of the hydroxyl groups using pyridine/acetic anhydride and DMAP in catalytic amounts. The *N*-Troc group was then converted to an acetamide under reductive conditions using Zn/Ac_2O with simultaneous acetylation of the free amine. Finally, the *C*-terminal tBu protecting group was removed by treatment with trifluoroacetic acid (TFA) in dichloromethane using anisole as cation scavenger.

By applying this synthesis strategy, the simplified GlcNAc core 1 trisaccharide building block **7** was prepared starting from the thioglycoside donor **2** and the known T_N-antigen acceptor 3 (Figure 41). In a glycosylation reaction, donor **2** was coupled to acceptor **3** to yield sufficient amounts of the core 1 trisaccharide **4** that was formed with complete regio- and stereoselectivity. The TBS and *p*-methoxyphenyl

groups of the trisaccharide **4** were removed, followed by acetylation to obtain compound **5**. Reductive elimination of the *N*-Troc group with simoultaneous acetylation of the amine function gave the *t*Bu-protected compound **6**. In a final step, the *t*Bu ester was cleaved to give the Fmoc-protected core 1 trisaccharide threonine building block **7** in a total yield of 20 % over 4 steps starting from the common T-antigen precursor **3**.

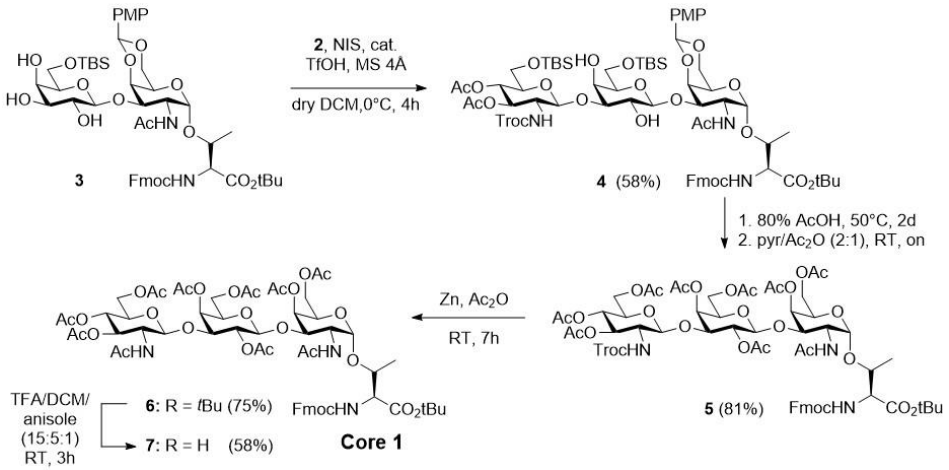


Figure 41. Synthesis of the Fmoc-protected simplified core 1 threonine building block **7**.

To prepare the core 2 trisaccharide **13**, the *p*-methoxyphenyl protecting group of the reported T-antigen **8** was first hydrogenolytically cleaved using $Pd(OH)_2$ to make position-6 accessible for glycosylation (Figure 42). The obtained the T-antigen acceptor **9** was then glycosylated with thioglycoside donor **2** and the trisaccharide **10** was formed with the desired stereo- and regioselectivity. Then, the TBS group of **10** was removed, followed by an acetylation step to give compound **11.** The reductive elimination of the *N*-Troc group was followed by another acetylation step to generate compound **12.** Finally, the *t*Bu group was removed and the desired Fmoc-protected core 2 trisaccharide **13** was obtained in a total yield of 39 % over 4 steps starting from the T-antigen precursor **9**.

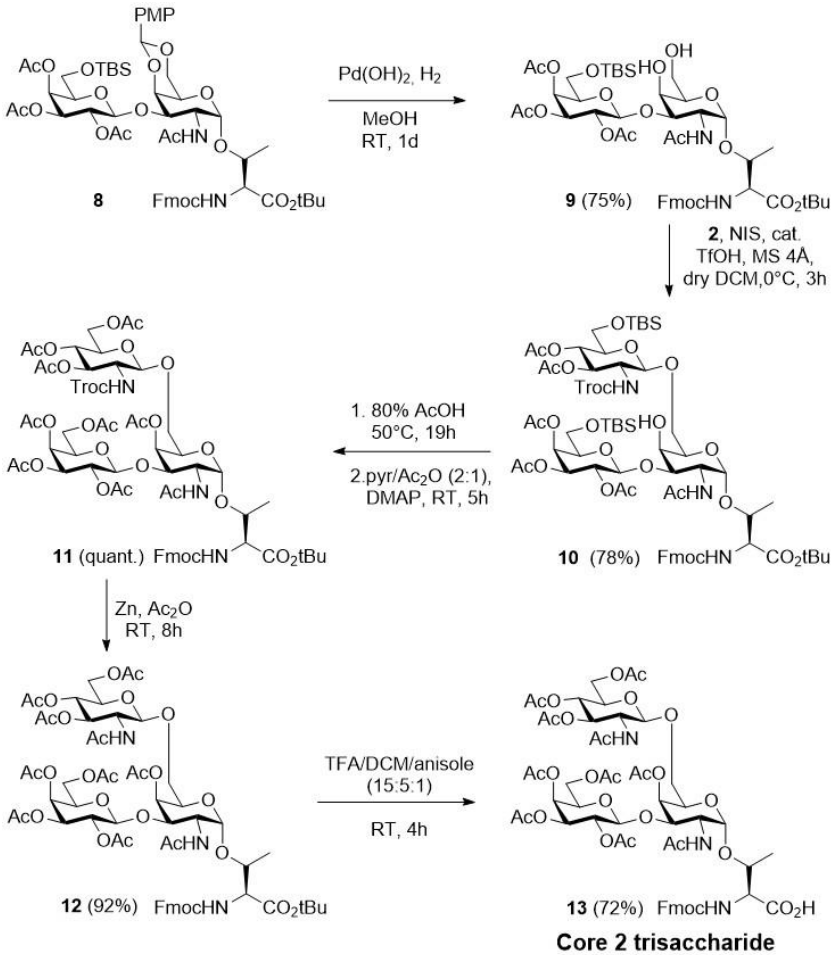


Figure 42. Synthesis of the Fmoc-protected simplified core 2 threonine building block **13**.

The GlcNAc elongated core 3 Fmoc-SPPS building block **18** was synthesized by glycosylation of the of the common T_N-antigen acceptor **14** with the thioglycoside donor **2** (Figure 43). The coupling reaction gave a mixture of the desired disaccharide coupling product **15** and a byproduct **16** modified with a phenyl-sulfenyl group at the *N*-Troc amine as verified by NMR and MS. This side reaction has been observed before and a possible reaction mechanism was proposed.[183] In accordance with the above-described synthesis protocol, the TBS and *p*-methoxyphenyl groups of disaccharides **15** and **16** were cleaved and the free hydroxyl groups were acetylated to yield compounds **17** and **18**, respectively. The *N*-Troc and *N*-Troc-SPh groups of **17** and **18** were subsequently

converted to the corresponding acetamide by reductive elimination with simultaneous acetylation of the generated free amine, leading to the formation of compound **19**. In the final step, the *t*Bu ester was removed to give the desired core 3 disaccharide Fmoc-protected threonine building block **20** in a total yield of 41 % over 4 steps starting from the common T_N-antigen precursor **14**.

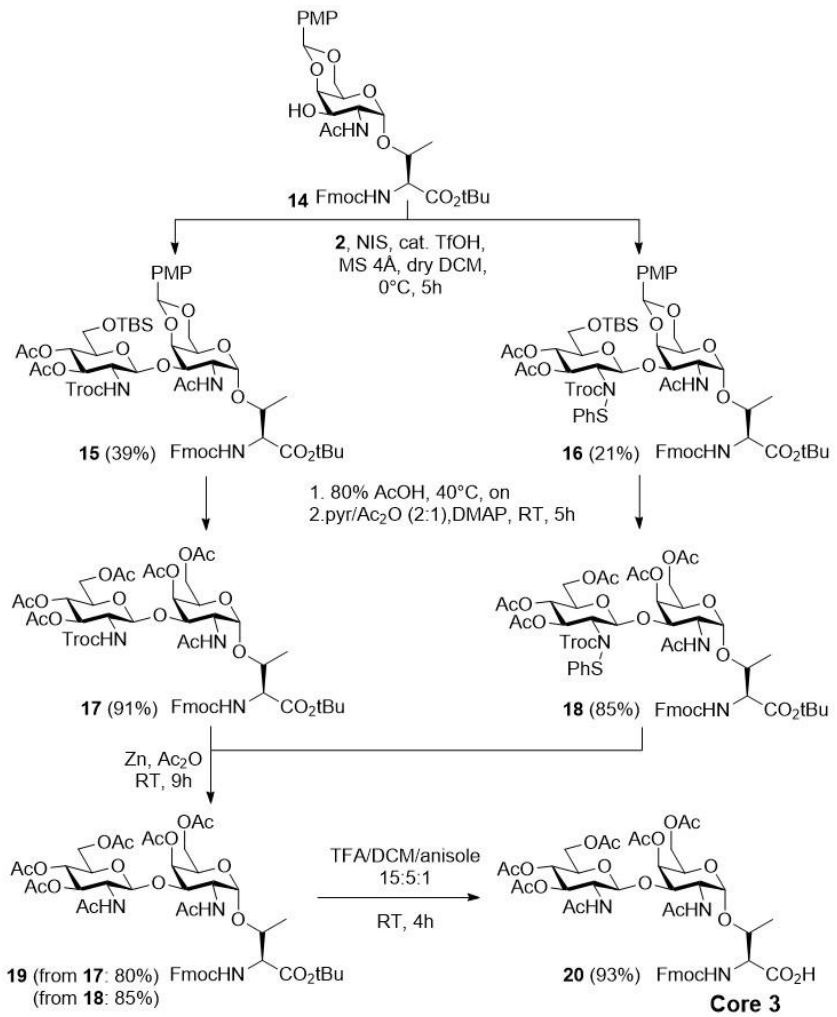


Figure 43. Synthesis of the Fmoc-protected simplified core 3 threonine building block **20**.

The GlcNAc elongated core 4 trisaccharide building block **27** was synthesized by double glycosylation of the common deacetylated T_N-acceptor building block **21** using thioglycoside donor **2** (Figure 44). As

already observed for the core 3 glycosylation reaction, the *N*-Troc-SPh adduct **23** was formed in addition to the desired product **22**. The TBS groups of **22** and **23** were removed and the free hydroxyl groups were acetylated to yield compounds **24** and **25**, respectively. Reductive elimination of the *N*-Troc and *N*-Troc-SPh groups of **24** and **25** with simultaneous acetylation gave compound **26**. In the last step, the *t*Bu ester was cleaved to obtain the desired Fmoc-protected core 4 trisaccharide threonine building block **27** in an overall yield of 35 % over 4 steps from the common T_N-antigen precursor **21**.

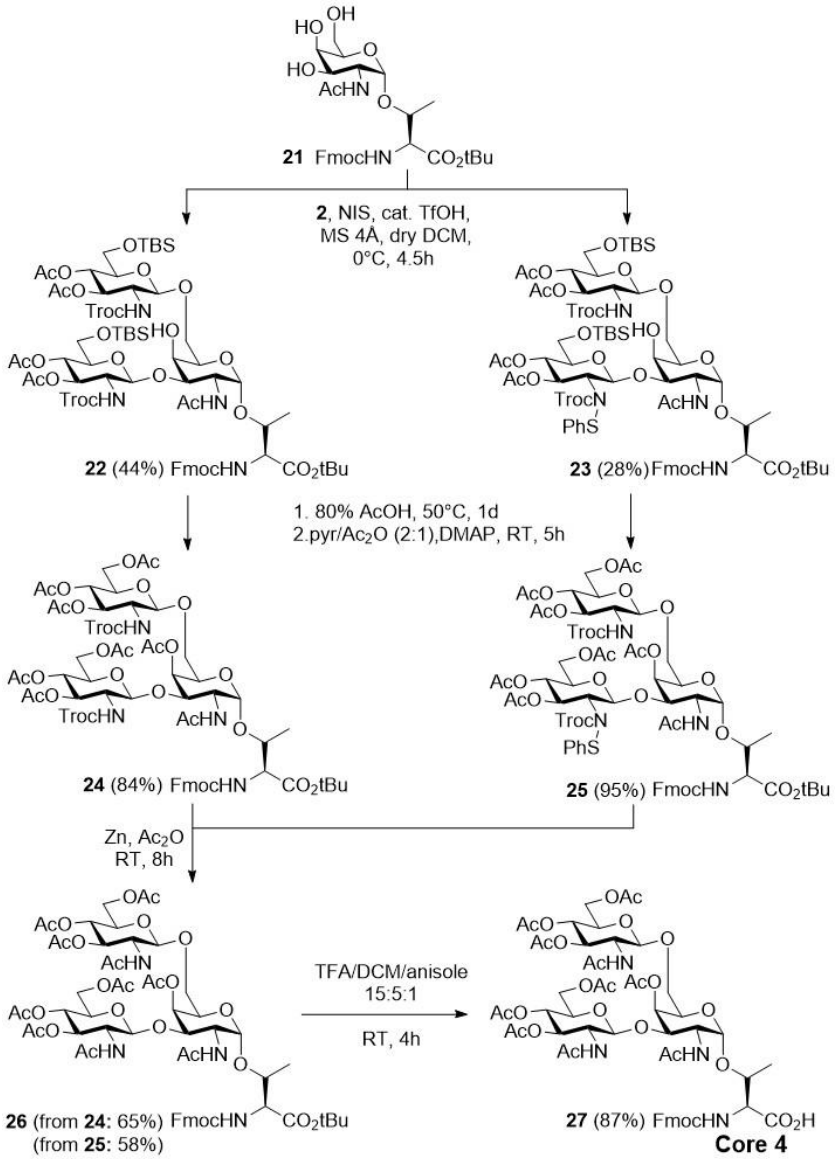


Figure 44. Synthesis of the Fmoc-protected simplified core 4 threonine building block **27**.

4.2.3.2 Synthesis MUC1 and MUC5AC peptides carrying simplified mucin cores

The obtained GlcNAc extended core 1-4 Fmoc-SPPS threonine building blocks **7**, **13**, **20** and **27** were then incorporated into of MUC1 and MUC5AC tandem repeat peptides by Fmmoc-SPPS (Figure 45).[183-184] This way, a library of MUC1 PAHGVT*SAPDT*RPAPGST*A and MUC5AC GT*T*PSPVPT*TST*T*SA (T* = possible glycosylation site) glycopeptides modified with the simplified core structures at different glycosylation sites was generated. The stepwise peptide assembly was carried out by coupling of standard Fmoc-amino acids (8 equiv) using HBTU/HOBt[239] according to the Fmoc-SPPS protocol for glycopeptide synthesis.[183] The glycosylated Fmoc-threonine building blocks (1.5 equiv) were coupled using the more reactive HATU/HOAt[240-241] with extended reaction times. The glycosylated amino acids were pre-activated in a smaller volume of solvent and added manually to the resin. The two following standard amino acids were double-coupled. After full mucin peptide assembly, a triethylenglycol spacer was coupled to the *N*-terminus for subsequent immobilization on NHS-activated microarray slides. Afterwards, the glycopeptides were cleaved from the solid support using TFA/triisopropylsilane (TIPS)/H_2O. After a desalting step on a C-18 cartridge, the glycan *O*-acetyl groups were removed using catalytic amounts of NaOMe in methanol at pH 9.5. Finally, the deacetylated glycopeptides were purified by preparative HPLC to obtain glycopeptides **P150**-**P178**.

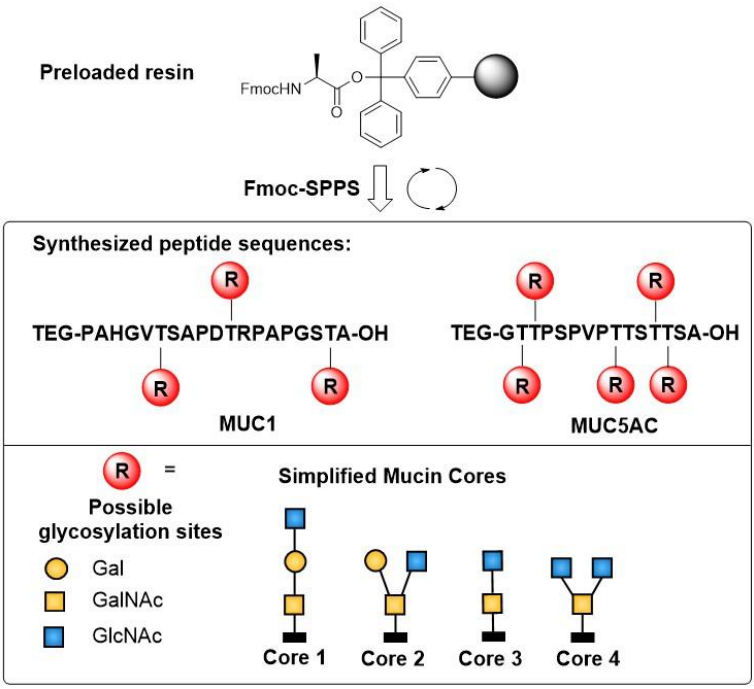


Figure 45. Fmoc-SPPS of MUC1 and MUC5AC peptides carrying simplified mucin core structures.

4.2.4 Conclusion

In summary, an efficient methodology was developed to synthesize simplified mucin core MUC1 and MUC5AC glycopeptides. Based on a convergent synthesis strategy, simplified core 1 to 4 Fmoc-protected building blocks **7**, **13**, **20** and **27** were synthesized in a few steps from common acceptor amino acids and glycosyl donors including the GlcNAc thioglycoside donor **2** for core elongation. The newly synthesized building blocks were then incorporated into MUC1 and MUC5AC peptides that were mono-glycosylated on different threonine residues with the GlcNAc-elongated core structures.

The obtained simplified mucin core glycopeptides represent useful scaffolds for further enzymatic modifications and elongation using different enzymatic approaches to further diversify and expand the mucin glycopeptide library. A library of type-2 LacdiNAc modified MUC1 and MUC5AC glycopeptides could be generated to evaluate the fine binding specificities of galectin-3, which was reported to also recognize this

carbohydrate motif.[242-243] Therefore, the obtained glycopeptides could be further enzymatically elongated using the human *N*-acetylgalactosaminyltransferase variant β4GalTY284L and printed on microarrays (Figure 46). Additionally, this LacdiNAc glycopeptide library could be used to evaluate binding specificities of other LacdiNAc-recognizing proteins such as the *H. pylori* adhesin LabA.

Furthermore, the simplified mucin core glycopeptides could act as substrates for other enzymes. For example, the sulfotransferase CHST2[244] requires the non-elongated GlcNAc residue on the core structure as substrate. The obtained GlcNAc-6-SO_3^- modified glycopeptides can then be elongated using Gal or GalNAc transferases to generate the corresponding 6-sulfated LacNAc or LacdiNAc motifs (Figure 46). Additionally, the LacdiNAc, and sulfo LacNAc and LacdiNAc glycopeptides could be further modified by fucosylation or sialylation to build up an extensive glycopeptide library. The glycopeptide library could then be widely applied in microarray binding studies to explore the binding specificities of galectins, siglecs and a variety of bacterial lectins, as well as to study binding interactions with the glycan-binding modules of mucinases (enzymes capaple of degrading mucins).

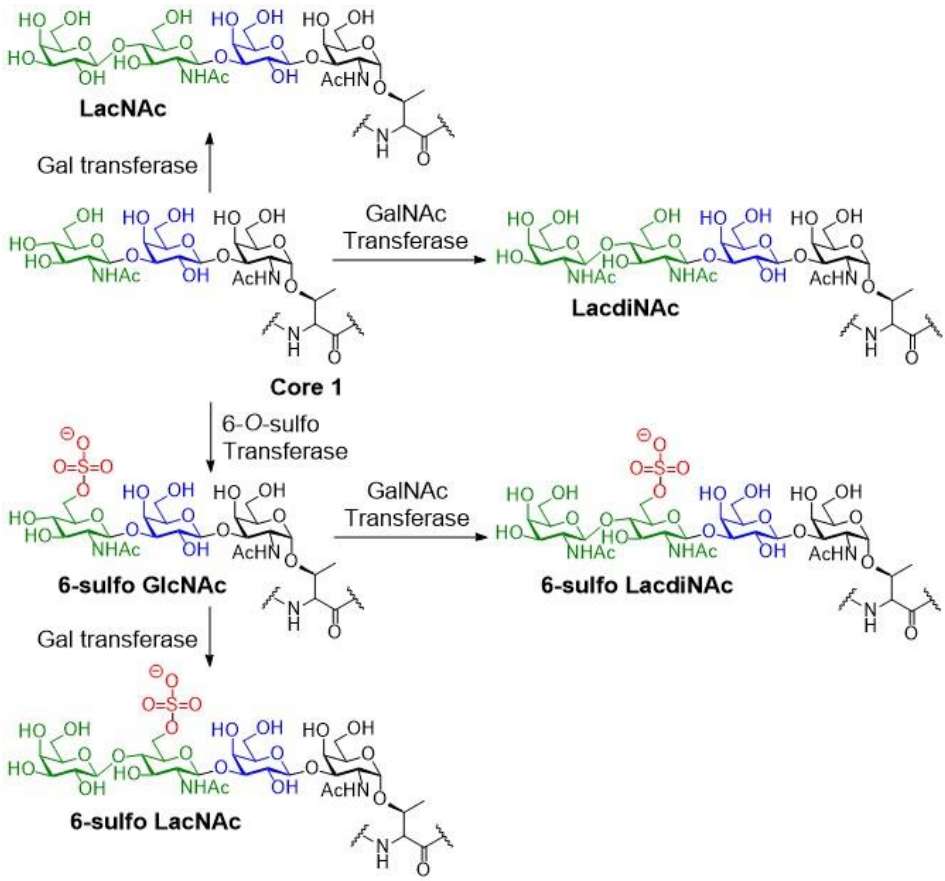


Figure 46. Possible enzymatic modifications of the GlcNAc-elongated core 1.

4.3 Galectin recognition of MUC1 glycopeptides (Paper V)

4.3.1 Galectins – the galactose recognizing proteins

Galectins (Gal) are a family of soluble proteins with conserved carbohydrate recognition domains that typically recognize β-galactosides.[245] So far, 15 mammalian galectins have been identified that are divided based on their structural organization of CRDs into proto, tandem repeat and chimera types (Figure 47).[246] The proto type galectins have one CRD and can form non-covalent dimers. The tandem repeat type galectins contain a *C*- and *N*-terminal recognition domain that are connected via a short linker peptide. Galectin-3 is the only chimera type galectin and has two distinct domains, the *C*-terminal CRD and an *N*-terminal collagen-like domain, which enables Gal-3 to oligomerize to pentamers. Galectins do not only differ in their structural organization of CRDs and oligomerization, but also show differences in their cellular location and tissue expression.[247-248] While some galectins are distributed widely in different cell and tissue types, others are more selectively expressed.

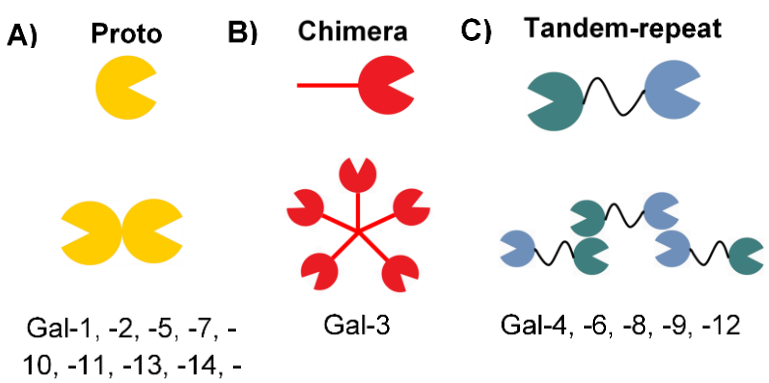


Figure 47. Schematic representation of the three groups of galectins. A) Proto-type galectins contain one CRD (top) and dimerize (bottom). B) The chimera type Gal-3 (top) oligomerizes to pentamers upon binding glycan ligands (bottom). C) Tandem-repeat type galectins contain two covalently linked CRDs (top) and can oligomerize via their *N*- and *C*-termini (bottom).

Galectins are involved in many biological processes, many of which are directly linked with immunity and disease.[249-250] Extracellularly, galectins can interact via their CRDs with cell-surface glycans for example on immune cells and can have an impact on many processes including cytokine and mediator production, cell adhesion, apoptosis, and chemo attraction.[251-253] Galectins also control cell-cell and cell-matrix interactions and adhesion processes.[254] Additionally, they modulate receptor functions such as clustering and endocytosis by forming lattices with cell-surface glycoprotein receptors, thus triggering a cascade of transmembrane signaling events.[255] Intracellularly, galectins participate in signaling pathways and also modulate biologic responses including cell differentiation, cell migration and apoptosis.[254]

To understand different biological functions of the galectin family members, it is necessary to elucidate their fine carbohydrate-binding specificities and to identify their endogenous receptor glycans. Galectins do not bind to specific individual ligands, but each recognizes a set glycan motifs.[256] The minimal units recognized by galectins is the LacNAc disaccharide (Gal-β-1,4-GlcNAc), which is found on *N*- and *O*-glycans, and the T-antigen (Gal-β-1,3-GalNAc).[257] Structurally, galectins require 4-OH and 6-OH groups of the galactose residue, and the 3-OH group of the penultimate GlcNAc for carbohydrate recognition, and substitution on these positions usually reduces or completely abolishes binding. Significant differences in glycan-binding preferences of the individual galectins have been reported. These structural variations include *N*-glycan branching, LacNAc extension, and terminal glycan modification including sulfation, sialylation or fucosylation.[257-259] While galectins-1, -2, -3, -7 and -9 showed increased affinity for branched bi-, tri-, and tetra-antennary *N*-glycans, the CRDs of galectin-4 and -8 did not recognize these *N*-glycans.[259] Furthermore, galectins-2, -4*N*, -7 and -9*C* preferably bind to type-1 LacNAc containing *N*-glycans, and Gal-1, -3, -8*N* and –9*N* favor type-2 LacNAc.[259] Additionally, galectins, including Gal-2, -3, -4*N*, -8*C*, -9*N* and -9*C*, which have the Glu-water-Arg-water motif were shown to recognize the endocyclic oxygen O5 and the *N*-acetyl group of the penultimate GalNAc residue of the T-antigen.[260] As a result, Gal-1 and -8*N,* which don't contain the Glu-water-Arg-water motif, do not bind the T-antigen. The α-2,3-Sialyl-T-antigen was reported to be recognized by Gal-2, -3, -4*N*, -9*N* and 9*C*.[247, 259] While, the ST-antigen was not bound by the *C*-terminal CRD of galectin-8, Gal-8*N* showed a strongly enhanced

affinity for this epitope.[259, 261-262] While α-2,3-sialylation strongly enhances Gal-8*N* binding, it decreases recognition by other galectins.[259, 263] Previous studies showed that α-2,6-sialylation inhibits binding of galectins to *N*-glycans and thereby acts as a negative regulator of galectin-dependent cell responses.[257, 264] However, binding of Gal-3 to α2,6-sialylated poly-LacNAc could be observed in some studies, even though the binding reduced compared to Gal-3 recognition of unsialylated or α2,3-sialylated poly-LacNAc.[265-266] Galectin-3 might bind laterally to internal LacNAc units within the extended LacNAc chain, which weakens the inhibitory effect of the α-2,6-linked sialic acid residue on the terminal LacNAc and ultimately leads to an increase in affinity.[265, 267] Additionally, galectins generally don't recognize Lewis x, Lewis a and Lewis b epitopes with the exception of the *C*-terminal CRD of galectin-8.[257, 259, 268] Gal-8C was shown to exhibit a high specificity for glycans that contain blood group A and B glycans.[263]

Since galectins are involved in a variety of glycan-dependent processes, they are potential therapeutic targets for inflammatory disease and cancer treatment.[249-250]. Galectins-1, -2, -3 -4, -8 and -9 are, for example, greatly increased in various cancer types.[269-270] Galectin-3 is one of the most studied galectins and plays important roles in tumor cell transformation, migration, invasion and metastasis.[271-272] As a result, Gal-3 might be used as a potential biomarker for cancer diagnosis, and represents an attractive target for cancer treatment.[273] For example, the small molecule galectin-3 inhibitors GB0139 and GB1211 from Galecto are in clinical trials, and the drug GMI-1757 from GlycoMimetics is in pre-clinical trials. An improved understanding of how galectin-3 and all other galectins bind to glycans on a molecular level could also advance the development of glycan-based inhibitors for specific galectin members.

4.3.2 Motivation

Galectin binding to mucins and mucin *O*-glycans appears to be important in cancer progression, however, the exact roles these interactions play are not well understood. For example, galectin-3 interacts with MUC1 and promotes EGFR dimerization and activation in epithelial cancer cells, which might play a role in EGFR-associated tumorigenesis and cancer progression.[274] Additionally, galectin-3 was reported to modulate the expression of MUC2 in human colon cancer cells, nevertheless, the specific regulatory mechanisms are unknown.[275] The parts that mucins play in interactions with other circulating galectins that are overexpressed in cancer patients are also not well explored.[269]

Since galectins play key roles many biological processes that are associated with inflammatory diseases and cancer, they are interesting targets for the development of therapeutic agents to combat these diseases.[249-250] A better understanding of how galectins interact with mucin *O*-glycans on a molecular level is not only essential to generate efficient galectin inhibitors, but also to elucidate the roles mucins play in galectin dependent cancer processes. So far, glycan arrays have often been used to explore the binding specificities of individual galectins. However, information about galectin binding to *O*-glycans and O-glycopeptides is limited.

In this work, the binding specificities of human galectins toward different core MUC1 glycopeptides were determined by microarray analysis. Additionally, the influence of modifications on *O*-glycan mucin core structures such as LacNAc extension, fucosylation and sialylation on galectin binding was explored. Therefore, selected human galectins were fluorescently labeled for detection and subsequently screened against MUC1 glycopeptide microarray libraries. The *N*- and *C*-terminal carbohydrate recognition domains of the tandem repeat-type galectins 4 and 8 were individually evaluated regarding their distinct binding specificities. Common Fmoc-protected mucin-core glycosylated threonine building blocks were synthesized and incorporated into MUC1 glycopeptides. Selected glycopeptides were further enzymatically modified with LacNAc, sialylation or fucosylation. The obtained glycopeptides were printed on microarrays and the microarray libraries were used to elucidate the binding specificities of human galectins-1, -3, -4, -7 and -8.

4.3.3 Results and discussion

4.3.3.1 Fluorescent labeling of human galectins

In order to explore the fine binding specificities of human galectins (hGal) toward galactose containing *O*-glycans, hGal-1, -3 CRD, the *N*- and *C*-terminal CRDs of Gal-4, hGal-7, the full length Gal-8, as well as its *N*- and *C*-terminal CRDs were fluorescently labeled using Alexa Fluor 488-NHS (Table 13). Therefore, all galectins were coupled to NHS-Alexa Fluor 488 under basic conditions in the presence of lactose, which blocked attachment of the fluorescent dye to or near the galectin CRDs (Figure 48). The human galectins were kindly provided by the Jiménez-Barbero group.

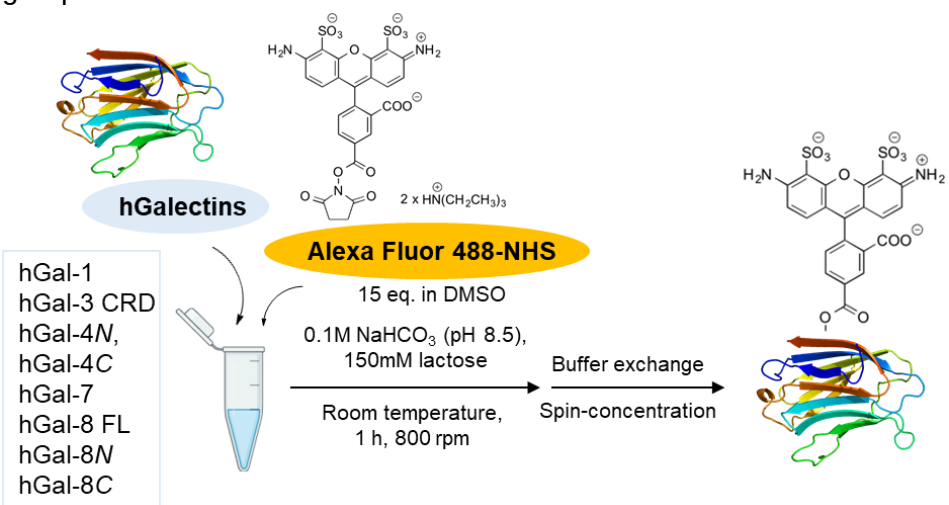


Figure 48. Alexa Fluor 488 labeling of human galectins.

Table 13. Amino acid sequences of human galectins.

Human galectin	Amino acid sequence
hGal-1	MACGLVASNLNLKPGECLRVRGEVAPDAKSFVLNLGKDSNNLCLH FNPRFNAHGDANTIVCNSKDGGAWGTEQREAVFPFQPGSVAEVCI TFDQANLTVKLPDGYEFKFPNRLNLEAINYMAADGDFKIKCVAFD
hGal-3 CRD	MLIVPYNLPLPGGVVPRMLITILGTVKPNANRIALDFQRGNDVAFHF NPRFNERNRRVIVCNTKLACDINACNWGREERQSVFPFESGKPFKI QVLVEPDHFKVAVNDAHLLQYNHRVKKLNEISKLGISGDIDLTSASY TMI
hGal-*4N*	EGDIHMAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASE HMKRFFVNFVVGQDPGSDVAFHFNPRFDGWDKVVFNTLQGGKW GSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHRLPL QMVTHLQVDGDLQLQSINFIGTLVPRGSMAISDPNSSSVDKLAAAL EHHHHHH LVPRGS = Thrombin cleavage site HHHHHH = His_5-tag
hGal-4*C*	MLPTMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAIN FKVGSSGDIALHINPRMGNGTVVRNSLLNGSWGSEEKKITHNPFGP GQFFDLSIRCGLDRFKVYANGQHLFDFAHRLSAFQRVDTLEIQGDV TLSYVQIGTLVPRGSMAISDPNSSSVDKLAAALEHHHHHH LVPRGS = thrombin cleavage site HHHHHH = His_5-tag
hGal-7	MSNVPHKSSLPEGIRPGTVLRIRGLVPPNASRFHVNLLCGEEQGSD AALHFNPRLDTSEVVFNSKEQGSWGREERGPGVPFQRGQPFEVLI IASDDGFKAVVGDAQYHHFRHRLPLARVRLVEVGGDVQLDSVRIF
hGal-8 FL	MMLSLNNLQNIIYNPVIPFVGTIPDQLDPGTLIVIRGHVPSDADRFQV DLQNGSSMKPRADVAFHFNPRFKRAGCIVCNTLINEKWGREEITYD TPFKREKSFEIVIMVLKDKFQVAVNGKHTLLYGHRIGPEKIDTLGIYG KVNIHSIGFSFSSDLQSTQASSLELTEISRENVPKSGTPQLRLPFAA RLNTPMGPGRTVVVKGEVNANAKSFNVDLLAGKSKDIALHLNPRLN IKAFVRNSFLQESWGEEERNITSFPFSPGMYFEMIIYCDVREFKVAV NGVHSLEYKHRFKELSSIDTLEINGDIHLLEVRSW
hGal-8*N*	MMLSLNNLQNIIYNPVIPFVGTIPDQLDPGTLIVIRGHVPSDADRFQV DLQNGSSMKPRADVAFHFNPRFKRAGCIVCNTLINEKWGREEITYD TPFKREKSFEIVIMVLKDKFQVAVNGKHTLLYGHRIGPEKIDTLGIYG KVNIHSIGFSFSSGSLVPRGSLEHHHHHH GS = Linker between hGal-8*N* and thrombin cleavage site LVPRGS= thrombin cleavage site HHHHHH = His_5-tag
hGal-8*C*	HMRLPFAARLNTPMGPGRTVVVKGEVNANAKSFNVDLLAGKSKDI ALHLNPRLNIKAFVRNSFLQESWGEEERNITSFPFSPGMYFEMIIYC DVREFKVAVNGVHSLEYKHRFKELSSIDTLEINGDIHLLEVRSWGSL VPRGSLEHHHHHH GS = Linker between hGal-8*C* and thrombin cleavage site LVPRGS= thrombin cleavage site HHHHHH = His_5-tag

4.3.3.2 Galectin recognition of mucin core glycopeptides

The fluorescently Alexa Fluor 488 labeled human galectins were then applied in microarray-based binding studies to explore their binding preferences toward different unmodified, LacNAc extended, fucosylated and sialylated MUC1 glycopeptides. Therefore, threonine building blocks carrying type-1 and type-2 mucin core structures 1 to 4 were synthesized and incorporated into the MUC1 tandem repeat sequence PAHGV**T***SAPD**T***RPAPGS**T***A (**T*** = possible glycosylation site) using our reported Fmoc-SPPS protocol for glycopeptide synthesis by Dr. Christian Pett and Dr. Manuel Schorlemer.[183-184] This way, a MUC1 glycopeptide library was prepared containing mono- bi- and trivalent peptides to assess the impact of various mucin antigens and cores, as well as of ligand presentation on different glycosylation sites on galectin binding. The obtained MUC1 glycopeptide library was then either printed on NHS-activated hydrogel slides (Table 14), or selected glycopeptides were further modified with LacNAc and/or α-1,3-, α-1,4-fucose, α-2,3- or α-2,6-sialic acid by Dr. Christian Pett, or Dr. Jin Yu, followed by immobilization on microarrays (Table 15, Table 16). Enzymatic LacNAc elongation was performed using the *H. pylori* β-1,3-*O*-*N*-acetylglucosaminyltransferase (β3GlcNAcT) and a fusion protein of human β-1,4-*O*-galactosyltransferase (His_6-Propeptide-catβ4GalT-1, β4GalT). Lewis a and Lewis x motifs were generated using the *H. pylori* α-1,3/4-*O*-fucosyltransferase. Additionally, glycopeptides were α2,3-sialylated by either the α-2,3-*O*-sialyltransferase PmST1 from *Pasteurella multocida,*[276] or Rat2,3-OST[277]. α-2,6-sialylation of mucin core structures was carried out using the α-2,6-*O*-sialyltransferase Pd2,6ST from *Photobacterium damsela*.[278]

Table 14. MUC1 core glycopeptide library.

ID	Glycopeptide sequence
P6	PAHGV**T*** (T_N)SAPDTRPAPGSTA
P12	PAHGVTSAPD**T*** (T_N)RPAPGSTA
P19	PAHGVTSAPDTRPAPGS**T*** (T_N)A
P179	PAHGV**T*** (T_N)SAPD**T*** (T_N)RPAPGSTA
P180	PAHGV**T*** (T_N)SAPDTRPAPGS**T*** (T_N)A
P181	PAHGVTSAPD**T*** (T_N)RPAPGS**T*** (T_N)A
P182	PAHGV**T*** (T_N)SAPD**T*** (T_N)RPAPGS**T*** (T_N)A
P183	PAHGV**T*** (T)SAPDTRPAPGSTA
P184	PAHGVTSAPD**T*** (T)RPAPGSTA
P185	PAHGVTSAPDTRPAPGS**T*** (T)A
P186	PAHGV**T*** (T)SAPD**T*** (T)RPAPGSTA
P187	PAHGV**T*** (T)SAPDTRPAPGS**T*** (T)A
P188	PAHGVTSAPD**T*** (T)RPAPGS**T*** (T)A
P189	PAHGV**T*** (T)SAPD**T*** (T)RPAPGS**T*** (T)A
P190	PAHGV**T*** (C1T1)SAPDTRPAPGSTA
P191	PAHGVTSAPD**T*** (C1T1)RPAPGSTA
P192	PAHGVTSAPDTRPAPGS**T*** (C1T1)A
P193	PAHGV**T*** (C1T1)SAPD**T*** (C1T1)RPAPGSTA
P194	PAHGV**T*** (C1T1)SAPDTRPAPGS**T*** (C1T1)A
P195	PAHGVTSAPD**T*** (C1T1)RPAPGS**T*** (C1T1)A
P196	PAHGV**T*** (C1T1)SAPD**T*** (C1T1)RPAPGS**T*** (C1T1)A
P197	PAHGV**T*** (C1T2)SAPDTRPAPGSTA
P198	PAHGVTSAPD**T*** (C1T2)RPAPGSTA
P199	PAHGVTSAPDTRPAPGS**T*** (C1T2)A
P200	PAHGV**T*** (C1T2)SAPD**T*** (C1T2)RPAPGSTA
P201	PAHGV**T*** (C1T2)SAPDTRPAPGS**T*** (C1T2)A
P202	PAHGVTSAPD**T*** (C1T2)RPAPGS**T*** (C1T2)A
P203	PAHGV**T*** (C1T2)SAPD**T*** (C1T2)RPAPGS**T*** (C1T2)A
P204	PAHGV**T*** (C3T1)SAPDTRPAPGSTA
P205	PAHGVTSAPD**T*** (C3T1)RPAPGSTA
P206	PAHGVTSAPDTRPAPGS**T*** (C3T1)A
P207	PAHGV**T*** (C3T1)SAPD**T*** (C3T1)RPAPGSTA
P208	PAHGV**T*** (C3T1)SAPDTRPAPGS**T*** (C3T1)A
P209	PAHGVTSAPD**T*** (C3T1)RPAPGS**T*** (C3T1)A
P210	PAHGV**T*** (C3T1)SAPD**T*** (C3T1)RPAPGS**T*** (C3T1)A
P211	PAHGV**T*** (C3T2)SAPDTRPAPGSTA
P212	PAHGVTSAPD**T*** (C3T2)RPAPGSTA
P213	PAHGVTSAPDTRPAPGS**T*** (C3T2)A
P214	PAHGV**T*** (C3T2)SAPD**T*** (C3T2)RPAPGSTA
P215	PAHGV**T*** (C3T2)SAPDTRPAPGS**T*** (C3T2)A
P216	PAHGVTSAPD**T*** (C3T2)RPAPGS**T*** (C3T2)A
P217	PAHGV**T*** (C3T2)SAPD**T*** (C3T2)RPAPGS**T*** (C3T2)A
P218	PAHGV**T*** ($C2T1_{Tet}$)SAPDTRPAPGSTA
P219	PAHGVTSAPD**T*** ($C2T1_{Tet}$)RPAPGSTA
P220	PAHGVTSAPDTRPAPGS**T*** ($C2T1_{Tet}$)A

P221	PAHGV**T***(C2T1$_{Tet}$)SAPD**T***(C2T1$_{Tet}$)RPAPGSTA
P222	PAHGV**T***(C2T1$_{Tet}$)SAPDTRPAPGS**T***(C2T1$_{Tet}$)A
P223	PAHGVTSAPD**T***(C2T1$_{Tet}$)RPAPGS**T***(C2T1$_{Tet}$)A
P224	PAHGV**T***(C2T1$_{Tet}$)SAPD**T***(C2T1$_{Tet}$)RPAPGS**T***(C2T1$_{Tet}$)A
P225	PAHGV**T***(C2T1$_{Hex}$)SAPDTRPAPGSTA
P226	PAHGVTSAPD**T***(C2T1$_{Hex}$)RPAPGSTA
P227	PAHGVTSAPDTRPAPGS**T***(C2T1$_{Hex}$)A
P228	PAHGV**T***(C2T1$_{Hex}$)SAPD**T***(C2T1$_{Hex}$)RPAPGSTA
P229	PAHGV**T***(C2T1$_{Hex}$)SAPDTRPAPGS**T***(C2T1$_{Hex}$)A
P230	PAHGVTSAPD**T***(C2T1$_{Hex}$)RPAPGS**T***(C2T1$_{Hex}$)A
P231	PAHGV**T***(C2T1$_{Hex}$)SAPD**T***(C2T1$_{Hex}$)RPAPGS**T***(C2T1$_{Hex}$)A
P232	PAHGV**T***(C2T2$_{Tet}$)SAPDTRPAPGSTA
P233	PAHGVTSAPD**T***(C2T2$_{Tet}$)RPAPGSTA
P234	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Tet}$)A
P235	PAHGV**T***(C2T2$_{Tet}$)SAPD**T***(C2T2$_{Tet}$)RPAPGSTA
P236	PAHGV**T***(C2T2$_{Tet}$)SAPDTRPAPGS**T***(C2T2$_{Tet}$)A
P237	PAHGVTSAPD**T***(C2T2$_{Tet}$)RPAPGS**T***(C2T2$_{Tet}$)A
P238	PAHGV**T***(C2T2$_{Tet}$)SAPD**T***(C2T2$_{Tet}$)RPAPGS**T***(C2T2$_{Tet}$)A
P239	PAHGV**T***(C2T2$_{Hex}$)SAPDTRPAPGSTA
P240	PAHGVTSAPD**T***(C2T2$_{Hex}$)RPAPGSTA
P241	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Hex}$)A
P242	PAHGV**T***(C2T2$_{Hex}$)SAPD**T***(C2T2$_{Hex}$)RPAPGSTA
P243	PAHGV**T***(C2T2$_{Hex}$)SAPDTRPAPGS**T***(C2T2$_{Hex}$)A
P244	PAHGVTSAPD**T***(C2T2$_{Hex}$)RPAPGS**T***(C2T2$_{Hex}$)A
P245	PAHGV**T***(C2T2$_{Hex}$)SAPD**T***(C2T2$_{Hex}$)RPAPGS**T***(C2T2$_{Hex}$)A
P246	PAHGV**T***(C4T1)SAPDTRPAPGSTA
P247	PAHGVTSAPD**T***(C4T1)RPAPGSTA
P248	PAHGVTSAPDTRPAPGS**T***(C4T1)A
P249	PAHGV**T***(C4T1)SAPD**T***(C4T1)RPAPGSTA
P250	PAHGV**T***(C4T1)SAPDTRPAPGS**T***(C4T1)A
P251	PAHGVTSAPD**T***(C4T1)RPAPGS**T***(C4T1)A
P252	PAHGV**T***(C4T1)SAPD**T***(C4T1)RPAPGS**T***(C4T1)A
P253	PAHGV**T***(C4T2)SAPDTRPAPGSTA
P254	PAHGVTSAPD**T***(C4T2)RPAPGSTA
P255	PAHGVTSAPDTRPAPGS**T***(C4T2)A
P256	PAHGV**T***(C4T2)SAPD**T***(C4T2)RPAPGSTA
P257	PAHGV**T***(C4T2)SAPDTRPAPGS**T***(C4T2)A
P258	PAHGVTSAPD**T***(C4T2)RPAPGS**T***(C4T2)A
P259	PAHGV**T***(C4T2)SAPD**T***(C4T2)RPAPGS**T***(C4T2)A

First, the binding speficifities of all human galectins toward unmodified type-1 and type-2 MUC1 core structures 1 to 4 were evaluated (Figure 50, Figure 51). Since Gal is the minimal recognition motif of galectins, T_N-antigens were as expected not or only weakly recognized. In accordance with previous studies,[259-260] hGal-3 and hGal-8*C* bound to T-antigens and hGal-8*N,* which does not have the Glu-water-Arg-water motif, did not recognize T-antigens on MUC1. However, no binding for hGal-4*N* was detected. Unexpectedly, hGal-1 showed weak affinities for T-antigens, even though it does not contain the Glu-water-Arg-water motif. Generally, galectins displayed only weak binding to T-antigens on MUC1, which might be explained by the fact, that multivalent binding and clustering of T-antigens on mucins strongly impact galectin recognition. Additionally, the placement of the particular glycosylation sites on the peptide backbone impacted galectin binding indicating that the underlying peptide backbone plays a role in galectin recognition events. Here, distinct binding profiles to mono- and bivalent glycopeptides could be observed for each galectin. Monovalent glycopeptides were often recognized in the decreasing affinity order GSTA > PDTR > GVTS, and bivalent derivatives as follows PDTR/GSTA > GVTS/GSTA > GVTS/PDTR. The bi- and trivalent glycopeptides can participate in multiple simultaneous binding interactions leading to an overall enhanced galectin binding strength. Thereby, trivalent ligand presentation on the same peptide backbone showed the highest binding affinity. Furthermore, the human galectins showed divergent binding preferences for type-1 and type 2 glycans on the different mucin core structures.

For example, the proto type galectins hGal-1 and hGal-7exhibited different binding profiles for the mucin core structures. While hGal-1 showed a certain preference for type-2 glycans over the corresponding type-1 analogs, hGal-7 showed a strong preference for core type-1 glycopeptides with the core 3 and core 4 type-2 structures hardly recognized. Besides the preferences for either type-1 or type-2 LacNAc units, both galectins showed stronger binding to the branched core 2 hexasaccharide and core 4 derivatives over the corresponding linear core 1 and core 3 structures. The arms of the branched core structures are oriented in opposite directions, thus presenting a spatial arrangement that may favor intermolecular multivalent galectin binding, and increasing the overall affinity. Core 2 hexasaccharides, which are presenting two LacNAc units, were better binders than the respective core 2

tetrasaccharide glycopeptides. These findings indicate that extra LacNAc elongation on the 3-arm, as found on core 4 and core 2_{Hex} structures, is important for galectin binding. The exact inner core structure (core-1-based vs. core-3 based glycans) seems not to be important for hGal-1 and hGal-7 binding with only a minor preference for the core 1 over the core 3 glycans.

Both galectins showed the same binding pattern for monovalent glycopeptides, which were recognized in the decreasing affinity order GSTA > PDTR > GVTS. However, whereas hGal-7 bound to bivalent peptides as follows PDTR/GSTA ≥ GVTS/GSTA > GVTS/PDTR, no distinct pattern was observed in case of hGal-1.

The chimera type galectin-3 CRD did not show a preference for either type-1 or type-2 core structures. Here, galectin-3 recognition depended on the respective core structure and/or glycosylation site(s). As observed for the proto type galectins hGal-1 and -7, branched core 4 and core 2_{Hex} structures were better binders than the unbranched core 1 and 3 analogs. These results suggest, that the additional arm on the core structure enhances the overall binding through intra- and intermolecular multivalent interactions of galectin-3 pentamers. However, the additional LacNAc unit on the core 2 hexasaccharide had no major impact in comparison to the binding strength of galectin-3 toward the core 2 tetrasaccharides. hGal-3 CRD did not show a specific binding pattern for the different mucin core structures. In line with the data obtained for the proto type galectins, this chimera type galectin recognized monovalent glycopeptides in the decreasing affinity order GSTA > PDTR > GVTS, and bivalent analogs as follows PDTR/GSTA > GVTS/GSTA > GVTS/PDTR. Galectin-3 binding was enhanced by multivalent ligand presentation and the trivalent glycopeptides showed were better binders than the bi- and monovalent peptide analogs with the monovalent glycopeptides being the weakest binders.

The tandem repeat galectins hGal-4 and hGal-8 contain *C*- and *N*-terminal carbohydrate recognition domains. Binding studies with human galectins showed that the binding profiles of tandem-repeat type galectins are essentially equivalent to the sum the *N*- and *C*-terminal recognition domains that show different binding specificities.[259] In this work, differences in binding between the *N*- and *C*-terminal CRDs of hGal-4 (hGal-4*N* and hGal-4*C*) and hGal-8 were analyzed. Additionally, the full

length hGal-8 was evaluated and its binding pattern was compared with those of the individual CRDs.

Surprisingly, similar recognition patterns for both hGal-4 recognition sites were observed. Type-1 core structures were better binders than the corresponding type-2 derivatives. However, both binding domains did not bind to type-2 core 3 and core 4 glycopeptides, and showed only weak binding to core 3 type-1 glycans. These findings indicate that the inner core structure is highly important for hGal-4 binding and that the Gal-GalNAc-based core structures are better binders than glycans having an inner core consisting of GlcNAc-GalNAc-based glycans. Structure analysis using X-ray crystallography with lactose or LacNAc ligands showed that the structures of human galectin-2, -3, -7, -9, and -10 possess similar tertiary structures and that amino acids crucial for glycan binding in the S4, S5 and S6 β-sheets are well conserved among galectins-1 to -9 (Figure 49).[261] While the Gal-β-1,3-GalNAc and additional LacNAc units of core 1 and core 2 glycans can together occupy the conserved and extended galectin binding pockets, core 3 and core 4 glycans do not have an extended LacNAc structure and can only participate with the core GlcNAc residue as part of a LacNAc unit in galectin binding. This explains the preference of hGal-4 for core 1 and core 2 glycopeptides.

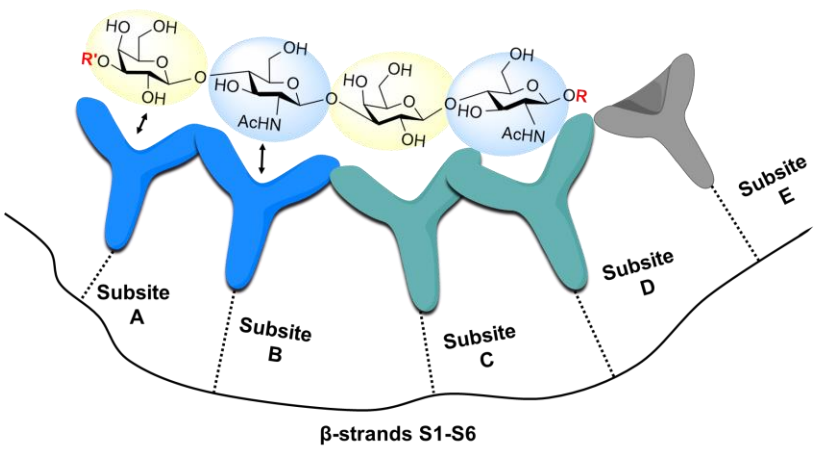


Figure 49. Schematic representation of a poly-LacNAc type-2 glycan bound in the CRD subsites A-D. The green subsites represent the first, and the blue subsites the second binding pocket. R = non-reducing; R' = reducing end.

Consistent with the observations for the proto galectins, hGal-4*N* and -*C* showed higher affinities for core 2 hexasaccharide over the tetrasaccharide glycopeptides suggesting that LacNAc extension on the 3-arm is important for galectin-4 binding. No distinct differences in binding pattern for either of the hGal-4 CRDs to the core structures was observed. However, the placement of the glycosylation site(s) strongly impacted hGal-4*N* and –*C* binding and monovalent glycopeptides were generally recognized in the decreasing affinity order: GSTA > PDTR > GVTS; and bivalent glycopeptides as follows: PDTR/GSTA > GVTS/GSTA > GVTS/PDTR.

In case of galectin-8, the full length protein, the *N*-terminal and the *C*-terminal recognition domains were compared. Distinct binding differences between the full length hGal-8 and its *N*- and *C*-terminal CRDs were observed. While core 3, $C2T2_{Tet}$ and C4T2 structures were not recognized by the full length protein, the C1T2 glycopeptides were good binders and better recognized than the corresponding C1T1 structures. The $C2T2_{Hex}$ was recognized very well, while the C2T1 and C4T1 where weak binders. In contrast to the full length protein, the hGal-8 *N*-terminal recognition domain also bound to $C2T2_{Tet}$, and multivalent C3T1 and C4T2 glycopeptides. Besides weak binding to C3T2 and C4T2, all other type-1 and type-2 core structures were well recognized by hGal-4*N.* A preference for branched core 2 glycans over linear core 1 was also observed. In contrast, hGal-*8C* exhibited a binding pattern very similar to hGal-8 FL and recognized C3T1 and $C2T2_{Tet}$ structures only weakly. In agreement with data obtain for the other evaluated galectins, mono- and divalent glycopeptides which were recognized by the full length galectin-8, as well as hGal-*8N* and -*C* in the decreasing affinity order: GSTA > PDTR > GVTS, and PDTR/GSTA ≥ GVTS/GSTA ≥ GVTS/PDTR.

In summary, the individual human galectins showed no or weak binding to T- and T_N-antigens on MUC1 glycopeptides. Different recognition patterns for type-1 and -2 core structures and variations in recognition toward the core structures were observed. In some cases, core 1 and core 2 glycans were better recognized than core 3 and core 4 glycopeptides due to the additional Gal residue in core 1 and core 2 structures that can be bound by the galectin binding pockets. Discriminations among linear and branched core structures and preferences for ore hexasaccharide (presenting two LacNAc units) over core 2 tetrasaccharide glycopeptides were observed in some cases.

Generally, galectin binding was enhanced by multivalent ligand presentation on the same peptide backbone and was strongest for trivalent glycopeptides, followed by the corresponding di- and monovalent derivatives. Additionally, the placement of the glycosylation site in the amino acid sequence influenced galectin binding. Monovalent glycan structures were mostly recognized in the in the decreasing affinity order: GSTA > PDTR > GVTS, and bivalent glycopeptides as follows PDTR/GSTA ≥ GVTS/GSTA ≥ GVTS/PDTR. These findings indicate that the exact glycan structures and glycosylation sites are potentially essential for the fine binding specificities and biological function of human galectins.

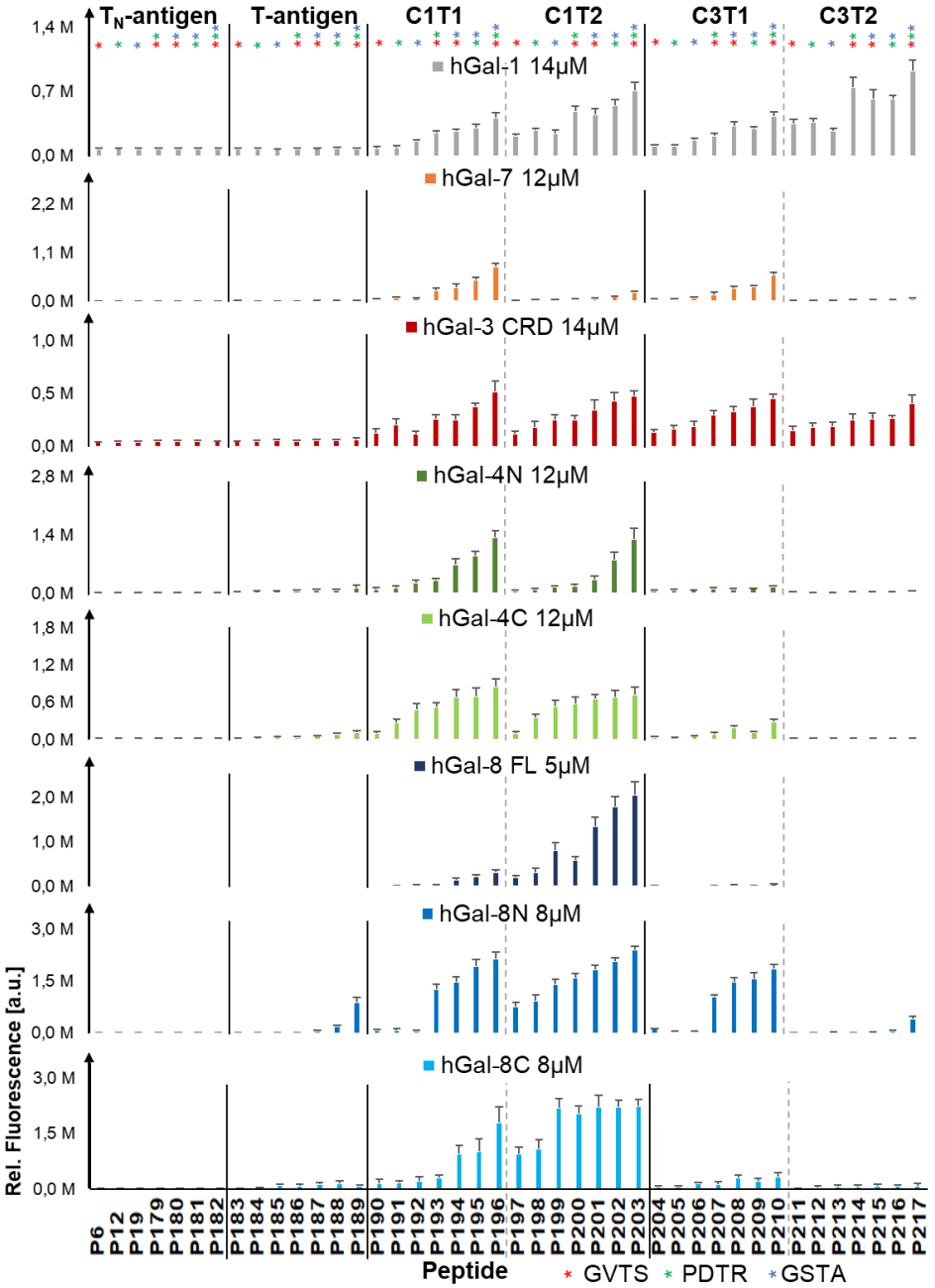


Figure 50. Binding of human galectins toward T-, T_N-antigen, core 1 and core 3 structures on mono-, bi- and trivalent MUC1 glycopeptides. [a.u.] = arbitrary units.

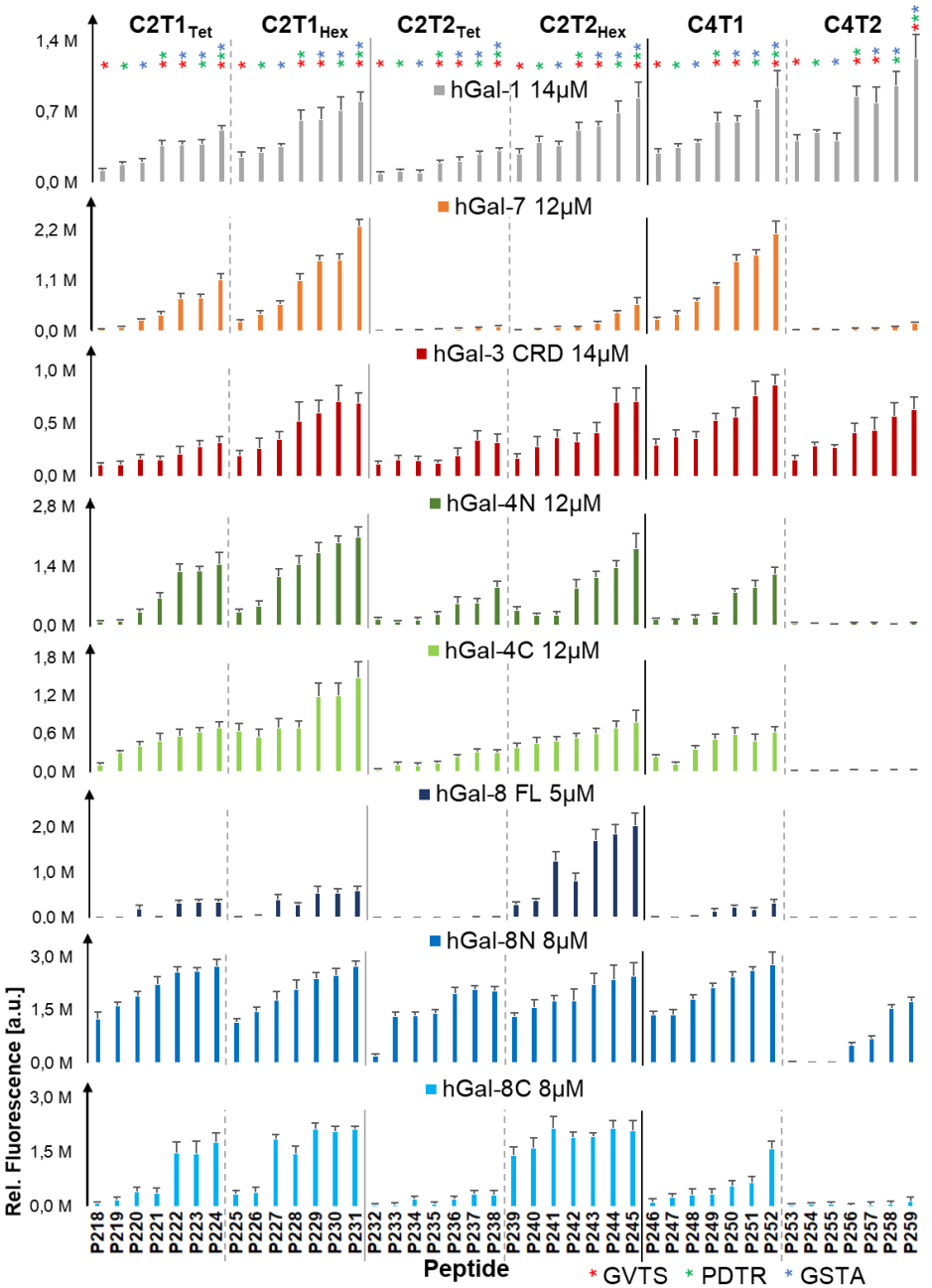


Figure 51. Binding of human galectins toward core 2 and core 4 structures on mono-, bi- and trivalent MUC1 glycopeptides. [a.u.] = arbitrary units.

Previous studies reported, that glycan modifications such as LacNAc elongation, sialylation and fucosylation strongly influence galectin binding.[257-259] Since the 4-OH and 6-OH groups of the galactose residue of LacNAc, and the 3-OH group of the penultimate GlcNAc are essential for galectin recognition, α-2,6-sialylation and fucosylation of the GlcNAc residue inhibit galectin binding.[257, 268] In contrast, α-2,3-sialylation was shown to not have a major impact on galectin recognition and can even slightly enhance the binding affinity.

In order to explore the impact of *O*-glycan modification on galectin binding, selected glycopeptides were enzymatically modified with LacNAc and/or α-1,3-, α-1,4-fucose, α-2,3- or α-2,6-sialic acid by Dr. Christian Pett, or Dr. Jin Yu as described above (Table 15). In agreement with previous studies, microarray analysis showed that LacNAc extended glycopeptides were generally better binders than the corresponding unelongated glycopeptide analogs (Figure 52).[259] Additionally, α2,3- and α-2,6-sialylation and α-1,3- and α-1,4-fucosylation often decreased or abolished galectin binding.

The proto-type galectin hGal-1 showed stronger binding affinities for LacNAc elongated mucin core structures. α-1,3- and α-1,4-fucosylations were found to generally decreased hGal-1 recognition and the addition of another fucose residue on the same core strucuture completely inhibited hGal-1 binding. While α-2,3-sialylation of glycans only lead to a slight reduction or equal binding compared to the non-sialylated structures in hGal-1 recognition, α-2,6-sialylation strongly decreased or even abolished hGal-1 binding. Additionally, LacNAc elongation of sialylated glycopeptides did not enhance galectin-1 binding due to the presence of a LacNAc unit with accessible 4- and 6-OH groups, but further decreased binding.

The other evaluated proto-type galectin hGal-7 displayed a similar recognition pattern as hGal-1. It also recognized LacNAc elongated glycans with increased affinity, and fucosylation, as well as sialylation reduced or completely inhibited hGal-7 binding. However, LacNAc elongation of sialylated glycans slightly increased hGal-7 recognition.

The chimera-type galectin hGal-3 CRD showed a similar recognition pattern as the proto-type galectins and bound to LacNAc elongated core structures with increased affinity in comparison with the respective shorter glycans. Additionally, hGal-3 CRD binding to both α-1,3- and α-1,4-

fucosylated, and α-2,3- and α-2,6-sialylated glycopeptides was strongly reduced or inhibited.

The *N*- and *C*-terminal CRDs of the tandem repeat galectin hGal-4 exhibited different binding patterns for the different glycan modifications. While hGal-4*N* showed enhanced binding for LacNAc elongated core structures, the impact of LacNAc extension on hGal-4*C* recognition depended on the particular core structure, and the binding was, for example, decreased for LacNAc elongated C1T2 and $C2T2_{Hex}$ glycopeptides. Both recognition domains displayed reduced or inhibition of binding for fucosylated sialylated glycans with exception of the α1,4-fucosylated $C2T1_{Hex}$ glycopeptide where hGal-4*C* showed increased recognition.

In accordance with previous observations, the full length tandem repeat-type galectin hGal-8 displayed binding preferences for LacNAc extended, fucosylated and sialylated glycopeptides that were a combination of the binding patterns of its individual *N*- and *C*-terminal CRDs. While hGal-8*N* showed increased affinity for LacNAc elongated core 3 and core 2 tetrasaccharide structures and decreased affinities for core 1 and core 2 hexasaccharide glycopeptides compared with the binding to the unelongated derivatives, hGal-8*C* exhibited increased recognition of LacNAc elongated core 1 and core 3 glycans and decreased binding for core 2 hexasaccharide in comparison with the shorter glycopeptides. In contrast, LacNAc elongation generally enhanced binding of the full length galectin-8. The two hGal-8 CRDs also displayed distinct binding patterns for fucosylated and sialylated glycopeptides. The *N*-terminal recognition domain showed a decreased binding for α-1,3-and α-1,4-fucosylated branched core 2 tetra- and hexasaccharide glycopeptides, and increased binding for the α-1,4-fucosylated type-1 core 3 and α-1,3-fucosylated type-2 core 1 glycopeptides. However, hGal-8*N* recognition was inhibited by LacNAc elongation of the α-1,3-fucosylated core 1 peptides. In contrast, reduced binding of hGal-8*C* was observed to fucosylated glycans with exception of the α-1,4-fucosylated type-1 core 1 and 2 hexasaccharide glycopeptides which displayed an enhanced binding affinity.

Again, hGal-8 FL exhibited a combined recognition pattern of the individual CRDs and the full length galectin-8 bound to the α-1,4-fucosylated type-1 core 1 and 2 hexasaccharide glycans with increased affinity, but showed decreased binding to the α-1,4-fucosylated type-1

core 3, and α-1,3-fucosylated type-2 core 1 and core 2 tetra- and hexasaccharide structures. Additionally, the binding pattern of hGal-8 FL for sialylated glycopeptides was similar to the one of hGal-8*C* and sialylation generally decreased the binding strength. Here, α-2,6-sialylation showed a stronger inhibitory effect on hGal-8 FL and -8*C* than α-2,3-sialylation. On the other hand, hGal-8*N* binding was strongly decreased or blocked by α-2,6-sialylation and its affinities for α-2,3-sialylated glycopeptides was, in line with previous observations, mostly enhanced.[259, 261-262] These findings highlight that the *N*- and *C*-terminal recognition domains of the tandem repeat-type galectin hGal-8 may fulfill different functions in the galectin-glycan interactions.

In summary, the fine specificities off all evaluated galectins toward LacNAc elongation, fucosylation and sialylation on MUC1 glycopeptides strongly depended on the particular mucin core structures and on the linkage of the added fucose and sialic acid residues. LacNAc elongation often lead to an incrase in galectin binding.

The impact of α-2,3- and α-2,6-sialylation on mucin core structures on galectin was further explored. Selected MUC1 glycopeptides were α-2,3- and α-2,6-sialylated by Dr. Christian Pett as decribed above, and the obtained glycopeptide library was printed on microarrays (Table 16). Microarray analysis showed that α-2,6-sialylation reduced binding of all evaluated galectins. The α-2,3-Sialyl-T-antigen was reported to be recognized by Gal-2, -3, -4*N*, -9*N* and 9*C*.[247, 259] In agreement with these data, α-2,3-ST_N-antigens were bound by hGal-3, hGal-4*N* and hGal-8*N,* but not by hGal-4*C*, -7, and hGal-8*C*. Additionally, equal or reduced affinities for α-2,3-sialylated mucin core structures was observed with exception of hGal-8*N*, which exhibited increased affinities for α-2,3-sialylated glycopeptides (Figure 53). However, this specific binding pattern was not fully reflected in the binding preferences of the full length galectin-8. All other galectins displayed a preference for α-2,3-sialylated over the α-2,6-sialylated mucin core glycopeptides. Data obtained for hGal-7 binding to sialylated glycopeptides were inconclusive.

Table 15. Peptide list of unmodified, fucosylated and sialylated MUC1 glycopeptides.

ID	Glycopeptide sequence
P191	PAHGVTSAPDT*(C1T1)RPAPGSTA
P88	PAHGVTSAPDT*(C1T1+α1,4Fuc)RPAPGSTA
P198	PAHGVTSAPDT*(C1T2)RPAPGSTA
P199	PAHGVTSAPDTRPAPGST*(C1T2)AP
P260	PAHGVTSAPDT*(C1T2+LacNAc)RPAPGSTA
P261	PAHGVTSAPDTRPAPGST*(C1T2+LacNAc)A
P89	PAHGVTSAPDT*(C1T2+α1,3Fuc)RPAPGSTA
P108	PAHGVTSAPDTRPAPGST*(C1T2+LacNAc+α1,3Fuc)A
P107	PAHGVTSAPDTRPAPGST*(C1T2+LacNAc+(α1,3Fuc)$_2$)A
P262	PAHGVTSAPDTRPAPGST*(C1T2+LacNAc+α2,3Sia)A
P263	PAHGVTSAPDT*(C1T2+α2,6Sia)RPAPGSTA
P264	PAHGVTSAPDTRPAPGST*(C1T2+LacNAc+α2,6Sia)A
P265	PAHGVTSAPDTRPAPGST*(C1T2+LacNAc+(α2,6Sia)$_2$)A
P204	PAHGVTSAPDT*(C3T1)RPAPGSTA
P276	PAHGVTSAPDT*(C3T1+LacNAc)RPAPGSTA
P98	PAHGVTSAPDT*(C3T1+LacNAc+α1,3Fuc)RPAPGSTA
P96	PAHGVTSAPDT*(C3T1+α1,4Fuc)RPAPGST(P96)
P266	PAHGVTSAPDT*(C3T1+LacNAc+α2,3Sia)RPAPGSTA
P267	PAHGVTSAPDT*(C3T1+α2,6Sia)RPAPGSTA
P268	PAHGVTSAPDT*(C3T1+LacNAc+α2,6Sia)RPAPGSTA
P219	PAHGVTSAPDT*(C2T1$_{Tet}$)RPAPGSTA
P91	PAHGVTSAPDT*(C2T1$_{Tet}$+α1,4Fuc)RPAPGSTA
P226	PAHGVTSAPDT*(C2T1$_{Hex}$)RPAPGSTA
P94	PAHGVTSAPDT*(C2T1$_{Hex}$+)α1,4Fuc)$_2$)RPAPGSTA
P220	PAHGVTSAPDTRPAPGST*(C2T2$_{Tet}$)A
P269	PAHGVTSAPDTRPAPGST*(C2T2$_{Tet}$+LacNAc)A
P92	PAHGVTSAPDT*(C2T2$_{Tet}$+α1,3Fuc)RPAPGSTA
P120	PAHGVTSAPDTRPAPGST*(C2T2$_{Tet}$+LacNAc+(α1,3Fuc)$_2$)A
P270	PAHGVTSAPDTRPAPGST*(C2T2$_{Tet}$+LacNAc+α2,3Sia)A
P271	PAHGVTSAPDTRPAPGST*(C2T2$_{Tet}$+LacNAc+(α2,3Sia)$_2$)A
P272	PAHGVTSAPDTRPAPGST*(C2T2$_{Tet}$+LacNAc+α2,6Sia)A
P273	PAHGVTSAPDTRPAPGST*(C2T2$_{Tet}$+LacNAc+(α2,6Sia)$_2$)A
P241	PAHGVTSAPDTRPAPGST*(C2T2$_{Hex}$)A
P274	PAHGVTSAPDTRPAPGST*(C2T2$_{Hex}$+(LacNAc)$_2$)A
P117	PAHGVTSAPDTRPAPGST*(C2T2$_{Hex}$+(α1,3Fuc)$_2$)A
P122	PAHGVTSAPDTRPAPGST*(C2T2$_{Hex}$+LacNAc+(α1,3Fuc)$_3$)A
P275	PAHGVTSAPDTRPAPGST*(C2T2$_{Hex}$+LacNAc+(α2,3Sia)$_2$)A
P276	PAHGVTSAPDTRPAPGST*(C2T2$_{Hex}$+LacNAc+(α2,6Sia)$_4$)A

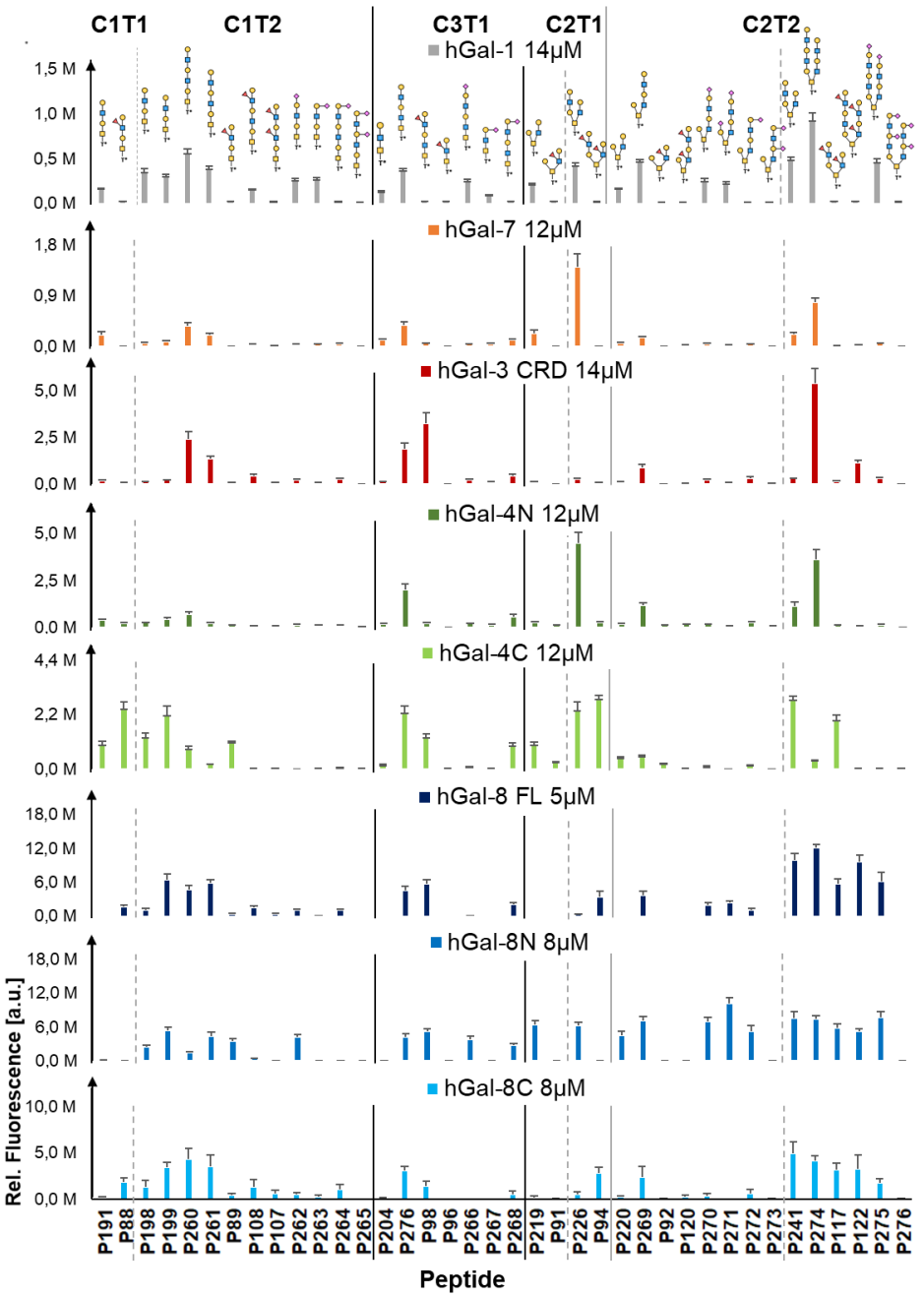


Figure 52. Binding of human galectin-8*C* toward unmodified, fucosylated and sialylated MUC1 glycopeptides.

Table 16. Part I. Peptide list of unmodified, α-2,3- and α-2,6-sialylated MUC1 glycopeptides.

ID	Glycopeptide sequence
P183	PAHGV**T*** (T)SAPDTRPAPGSTA
P184	PAHGVTSAPD**T*** (T)RPAPGSTA
P185	PAHGVTSAPDTRPAPGS**T*** (T)A
P186	PAHGV**T*** (T)SAPD**T*** (T)RPAPGSTA
P187	PAHGV**T*** (T)SAPDTRPAPGS**T*** (T)A
P188	PAHGVTSAPD**T*** (T)RPAPGS**T*** (T)A
P189	PAHGV**T*** (T)SAPD**T*** (T)RPAPGS**T*** (T)A
P191	PAHGVTSAPD**T*** (C1T1)RPAPGSTA
P192	PAHGVTSAPDTRPAPGS**T*** (C1T1)A
P197	PAHGV**T*** (C1T2)SAPDTRPAPGSTA
P198	PAHGVTSAPD**T*** (C1T2)RPAPGSTA
P199	PAHGVTSAPDTRPAPGS**T*** (C1T2)A
P203	PAHGV**T*** (C1T2)SAPD**T*** (C1T2)RPAPGS**T*** (C1T2)A
P233	PAHGVTSAPD**T*** ($C2T2_{Tet}$)RPAPGSTA
P227	PAHGVTSAPDTRPAPGS**T*** ($C2T1_{Hex}$)A
P240	PAHGVTSAPD**T*** ($C2T2_{Hex}$)RPAPGSTA
P241	PAHGVTSAPDTRPAPGS**T*** ($C2T2_{Hex}$)A
P211	PAHGV**T*** (C3T2)SAPDTRPAPGSTA
P212	PAHGVTSAPD**T*** (C3T2)RPAPGSTA
P213	PAHGVTSAPDTRPAPGS**T*** (C3T2)A
P217	PAHGV**T*** (C3T2)SAPD**T*** (C3T2)RPAPGS**T*** (C3T2)A

Table 17. Part II. Peptide list of unmodified, α-2,3- and α-2,6-sialylated MUC1 glycopeptides.

ID	Glycopeptide sequence
P277	PAHGV**T***(ST)SAPDTRPAPGSTA
P278	PAHGVTSAPD**T***(ST)RPAPGSTA
P279	PAHGVTSAPDTRPAPGS**T***(ST)A
P280	PAHGV**T***(ST)SAPD**T***(ST)RPAPGSTA
P281	PAHGV**T***(ST)SAPDTRPAPGS**T***(ST)A
P282	PAHGVTSAPD**T***(ST)RPAPGS**T***(ST)A
P283	PAHGV**T***(ST)SAPD**T***(ST)RPAPGS**T***(ST)A
P284	PAHGVTSAPD**T***(C1T1+α2,3Sia)RPAPGSTA
P285	PAHGVTSAPDTRPAPGS**T***(C1T1+α2,3Sia)A
P286	PAHGV**T***(C1T2+α2,3Sia)SAPDTRPAPGSTA
P287	PAHGVTSAPD**T***(C1T2+α2,3Sia)RPAPGSTA
P288	PAHGVTSAPDTRPAPGS**T***(C1T2+α2,3Sia)A
P289	PAHGV**T***(C1T2+Sia)SAPD**T***(C1T2+Sia)RPAPGS**T***(C1T2+Sia)A
P290	PAHGVTSAPD**T***(C2T2$_{Tet}$+α2,3Sia)RPAPGSTA
P291	PAHGVTSAPD**T***(C2T2$_{Tet}$+α2,3Sia)RPAPGSTA
P292	PAHGVTSAPD**T***(C2T2$_{Hex}$+α2,3Sia)RPAPGSTA
P293	PAHGV**T***(C3T2+α2,3Sia)SAPDTRPAPGSTA
P294	PAHGVTSAPDTRPAPGS**T***(C3T2+α2,3Sia)A
P295	PAHGVTSAPD**T***(C3T2+α2,3Sia)RPAPGSTA
P296	PAHGVTSAPDTRPAPGS**T***(C1T1+α2,6Sia)A
P297	PAHGVTSAPDTRPAPGS**T***(C1T1+(α2,6Sia)$_2$)A
P273	PAHGVTSAPD**T***(C1T2+α2,6Sia)RPAPGSTA
P298	PAHGVTSAPDTRPAPGS**T***(C1T2+α2,6Sia)A
P299	PAHGVTSAPDTRPAPGS**T***(C1T2+(α2,6Sia)$_2$)A
P300	PAHGVTSAPDTRPAPGS**T***(C2T1$_{Hex}$+α2,6Sia)A
P301	PAHGVTSAPD**T***(C2T2$_{Hex}$+α2,6Sia)RPAPGSTA
P302	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Hex}$+α2,6Sia)A
P303	PAHGVTSAPDTRPAPGS**T***(C2T2$_{Hex}$+(α2,6Sia)$_2$)A
P304	PAHGV**T***(C3T2+α2,6Sia)SAPDTRPAPGSTA
P305	PAHGVTSAPD**T***(C3T2+α2,6Sia)RPAPGSTA
P306	PAHGVTSAPDTRPAPGS**T***(C3T2+α2,6Sia)A
P307	PAHGV**T***(C3T2+α2,6Sia)SAPD**T***RPAPGS**T***(C3T2+α2,6Sia)A

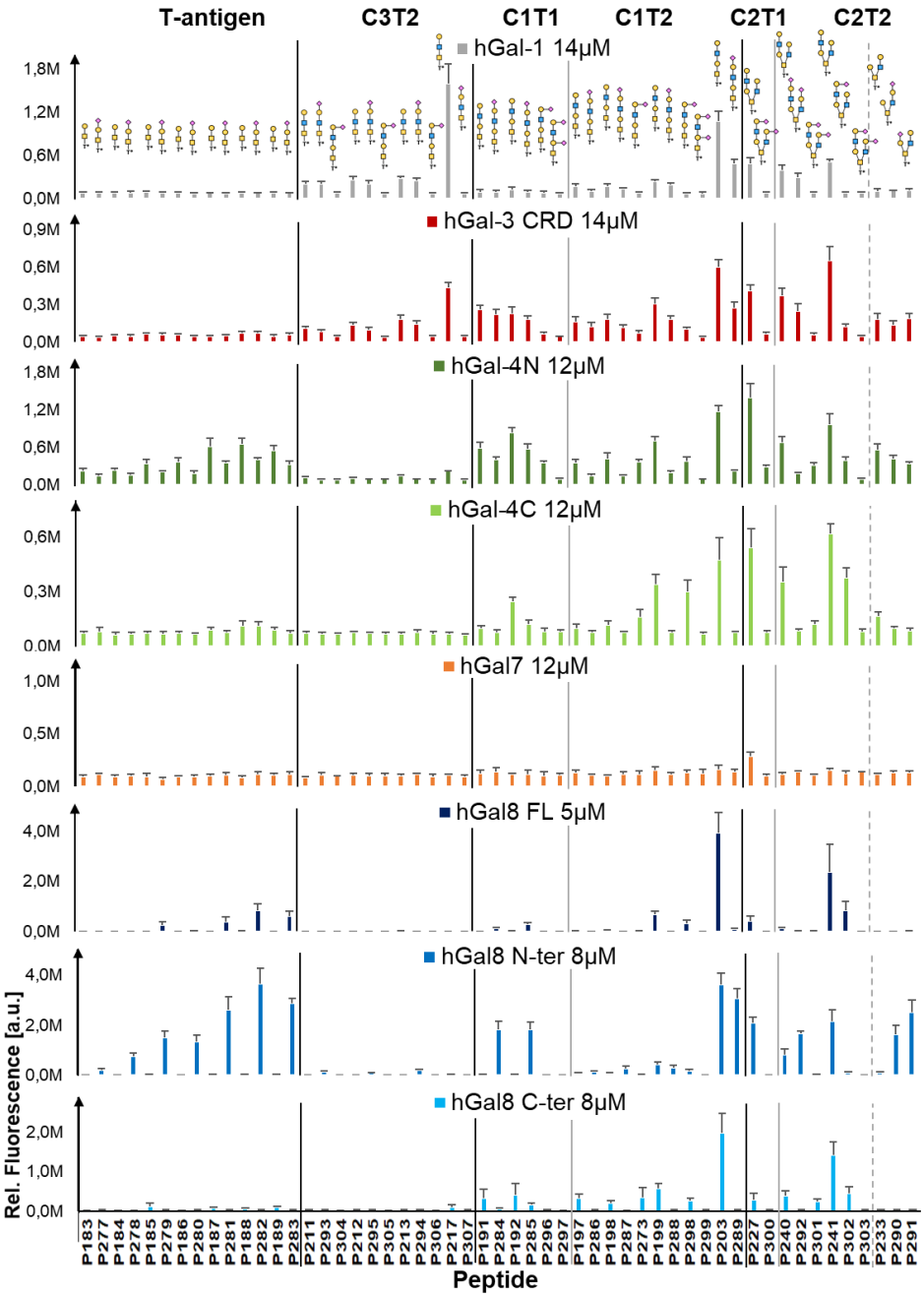


Figure 53. Binding of human galectin-8*C* toward unmodified, α-2,3- and α-2,6 sialylated MUC1 glycopeptides. [a.u.] = arbitrary units.

4.3.4 Conclusion

In summary, the binding specificities of human galectins-1, -3 CRD, -4*N*, -4*C*, -7, -8 FL, -8*N* and -8*C* were determined using MUC1 glycopeptide libraries. Additionally, the influence of LacNAc extension, fucosylation and sialylation of *O*-glycan mucin core structures on galectin binding was explored. Therefore, all human galectins were fluorescently labeled with Alexa Fluor 488 for detection. MUC1 glycopeptides bearing different mucin core structures were synthesized by Fmoc-SPPS and selected glycopeptides were further enzymatically elongated with LacNAc, and/or modified with α-2,3- and α-2,6-sialylation, or α-1,3- and α-1,4-fucosylation. The obtained glycopeptides were immobilized on microarrays and the microarray libraries were used to elucidate the fine binding specificities of the selected human galectins. Microarray analysis showed, that T- and T_N-antigens were not or only weakly bound by the galectins. All galectins showed distinct binding preferences for the different mucin core structures and also for type-1 and type 2 glycans on the cores. The exact placement of the glycosylation site(s) on the peptide backbone had a strong impact on galectin recognition. Monovalent glycopeptides were generally bound in the decreasing affinity order: GSTA > PDTR > GVTS, and bivalent derivatives as follows: PDTR/GSTA > GVTS/GSTA > GVTS/PDTR. These results suggest that the peptide backbone plays a secondary role in galectin recognition events. Additionally, bi- and trivalent ligand presentation increased galectin affinities due to the multivalent cluster effect, and all galectins showed preferred binding to tri- over di- and monovalent glycopeptides with the peptide analogs displaying only one ligand being the weakest binders. In line with previous studies that reported that LacNAc elongation, sialylation and fucosylation of glycans strongly influenced galectin binding, LacNAc extension generally enhanced galectin binding, and α-2,6-sialylation and α-1,3- and α1,4-fucosylation mostly decreased binding.[257-259] The reduced affinities result from modification and thus blocking of the Gal 6-OH group and the GlcNAc 3-OH group, which are essential for galectin recognition. [257, 268] However, α2,3-sialylation was shown to often decrease galectin recognition and only the hGal-8*N* exhibited increased affinities for α2,3-sialylated glycan structures.

The determined binding specificities of selected proto,- chimera and tandem repeat-type galectins for unmodified, LacNAc extended, fucosylated and sialylated mucin core glycopeptides are important to

better understand of how galectins interact with *O*-glycans on cell surfaces and consequently to get insights of the specific roles the individual galectins play in their respective biological processes. Many of these processes are also associated with inflammatory diseases and cancer.[249-250] Therefore, galectins are interesting targets to development novel strategies to fight these diseases. Good knowledge of the distinct binding preferences of each galectin could help to design galectin-specific carbohydrate-based inhibitors.

4.4 Immunological evaluation of antibodies induced by tumor-associated MUC1 glycopeptide-bacteriophage Qβ vaccine conjugates (Papers VI and VII)

4.4.1 MUC1 in cancer vaccines

Since cancer is a leading cause of death worldwide, there is an urgent need for novel and better cancer therapeutics.[279] One approach in therapeutic cancer treatment is immunotherapy including vaccination.[280-281] MUC1 is ubiquitously found on epithelial cell surfaces and is overexpressed in many cancers, including breast, lung, pancreatic, colon, prostate, and ovarian cancer.[83, 282] As a result, MUC1 represents an attractive antigenic target for the development of effective anti-cancer vaccines.[93, 283] Earlier MUC1-based vaccines typically employed the unmodified MUC1 peptide as the antigen.[284] However, due to the natural tolerance of the immune system towards endogenous structures, MUC1 glycopeptide vaccines are only weakly immunogenic, thus making it challenging to elicit a strong MUC1 based immune response.[284-285]
As a result, new strategies to elicit strong, tumor-specific antibody responses need to be developed. One approach involves the introduction of tumor-associated carbohydrate antigens (TACAs) into the MUC1 vaccine such as Sialyl-T_N-,T_N-, T- and Sialyl-T-antigens.[164-165, 286] On tumor cells, the formation of TACAs on the MUC1 peptide tandem repeats can be attributed to MUC1 overexpression, downregulation of the core 2 β-1,6-*N*-acetylglucosaminyltransferase-1 (C2GnT-1), mutation of the Cosmc-gene, which is essential for T-synthase activity, and premature sialylation by increased sialyltransferases expression.[85, 88, 287-288] As a result, T_N-,Sialyl-T_N-, T- and Sialyl-T-antigens are dominant over branched and elongated core 2 structures.

The induction of humoral immune responses directed against TACAs represents a valuable asset for tumor immunotherapy. A synthetic antitumor vaccine has to meet strict requirements. It needs to elicit a strong tumor-specific immune response, has to overcome the natural immune tolerance, and should lead to immunological memory. Several methods to increase the antigenicity of mucin glycopeptides have been reported. For example, MUC1 glycopeptide B-cell epitopes have been conjugated to different immune stimulants such as carrier proteins, including the keyhole limpet hemocyanin, Tetanus toxoid and CRM, T-cell epitope peptides, or other immune stimulating adjuvants such as

lipopeptide-derived Toll-like receptor-2 ligands.[289-292] Human MUC1 transgenic mice are capable of mimicking MUC1 immunotolerance in humans and have been shown to produce higher levels of specific anti-MUC1 antibodies when immunized with MUC1-T_N glycopeptide vaccine in comparison to vaccines carrying the unglycosylated MUC1 peptide.[165, 286]

The advantage of using glycopeptides instead of the unmodified MUC1 sequences for vaccine design is that the generated antibodies recognize structures and conformations found on MUC1 at the surface of tumor cells.[164] The immunogenicity such structures is greatly enhanced, probably due to carbohydrate-induced favorable conformational changes of the mucin peptide backbone.[293] For example, glycosylation of MUC1 with ST_N- and T_N-antigens in the PDTR and GSTA regions strongly influences the MUC1 peptide backbone conformation, which in turn influences antibody recognition.[294] Additionally, the glycosylation site in the antigen peptide is important to elicit a strong and specific immune response. A study on monoclonal antibodies showed that while almost all monoclonal antibodies bound to the PDTR motif, much fewer antibodies bound to the GSTA or GVTS motifs on the peptide backbone, implying that the PDTR motif is immune dominant in mice immunized with MUC1 vaccines.[295] Due to the limited availability of glycopeptide samples for bioassays such as ELISA, surface plasmon resonance, or microarrays, evaluation of the detailed specificity of the raised antibodies is often not addressed. This is problematic since the immunological memory could potentially produce antibodies with cross-reactivity to epitopes exposed on healthy cells.

4.4.2 Motivation

MUC1 is an attractive antigenic target for anticancer vaccines. Tumor associated MUC1 exhibits various TACAs including the T-, T_N- and ST_N-antigens. Whereas MUC1 peptides and MUC1-T_N glycopeptides have most often been used in tumor vaccines, MUC1 glycopeptides bearing T- or ST_N-antigens have been less explored in tumor models.[164, 289, 291, 296] KLH and TTox are immunogenic carrier proteins that have been conjugated with TACAs to induce strong immune responses.[289, 297-298] Additionally, various non-protein carriers have been tested to induce strong anti-TACA immune responses, including dendrimers,[299] gold nanoparticles,[300-301] and virus-like particles (VLPs) such as the bacteriophage Qβ.[302] VLPs are attractive carriers that are highly immunogenic due to their size that promotes vaccine uptake by antigen-presenting cells, repetitive structure promoting B-cell recognition, and ability to cross-link B-cell receptors. This study showed that Qβ elicited high levels of T_N-antigen selective antibodies.[302]

In this work, T- or ST_N-antigen peptide-Qβ conjugates were prepared as potential anticancer vaccines to induce T- or ST_N-antigen specific antibodies in mice. MUC1 glycopeptides SAPDT*RPAP (T* = glycosylation site) carrying the T-, or ST_N-epitopes were synthesized and subsequently conjugated to the carrier bacteriophage Qβ. The instability of the α-*O*-glycosidic linkage between the GalNAc residue and the peptide backbone of the native MUC1 glycopeptide to glycosidases presents a potential drawback for vaccine design. As a result, a MUC1 glycopeptide SAPDT*RPAP (T* = glycosylation site) bearing a T-antigen was synthesized, where the GalNAc residue was coupled to threonine in an unnatural β-glycosidic linkage, to overcome this hurdle. The obtained β-T-antigen peptide was then conjugated to the virus-like particle Qβ. To evaluate the translational potential of these vaccine constructs as cancer vaccines, MUC1.Tg mice were immunized with Qβ-MUC1-α-T, Qβ-MUC1-β-T and Qβ-MUC1-ST_N. All conjugates induced high levels of IgG antibodies in clinically relevant human MUC1 transgenic mice. To evaluate the ability of the raised polyclonal antibodies to recognize different MUC1 glycoforms upon vaccine stimulation, as well as their specificity, synthetic mucin glycopeptide libraries were used in microarray binding studies.

4.4.3 Results and discussion

4.4.3.1 Synthesis of Qβ-MUC1 conjugates carrying TACAs and antibody induction (Papers VI and VII)

To obtain tumor antigen specific mouse antibodies, MUC1 TACA antigen glycopeptides **P308** and **P309**, were prepared by incorporating synthetic Fmoc-protected T_N-, ST_N-, α-T- and β-T-antigen threonine building blocks into the short MUC1 sequence SAPDT*RPAP (T* = glycosylation site) by Fmoc-SPPS. Antigen glycopeptides **P308** and **P309** were subsequently conjugated to Qβ virus-like particles to generate the Qβ-MUC1 vaccine conjugates **28** and **29** (Figure 54). Additionally, the antigen peptide **P310** carrying the non-natural β-T-antigen was synthesized, which was shown to exhibit an increased stability toward glycosidase cleavage. The peptide **P310** was then conjugated to bacteriophage Qβ to generate the Qβ-MUC1 vaccine conjugate **30**. The MUC1 glycopeptides **P308** - **P310** were then coupled to bacteriophage Qβ and the obtained conjugates Qβ-MUC1 **28** -**30** were used to induce anti-TACA antibodies in MUC1.Tg mice. High levels of IgG antibodies were induced and the binding specificities of the induced antibodies towards mucin core MUC1 glycopeptides were evaluated using microarray binding assays. The syntheses of the Fmoc-protected T_N-, ST_N-, α-T- and β-T-antigen threonine building blocks, glycopeptides **P308** to **P310**, and Qβ-MUC1vaccine conjugates **28** -**30** were performed by the Huang group.

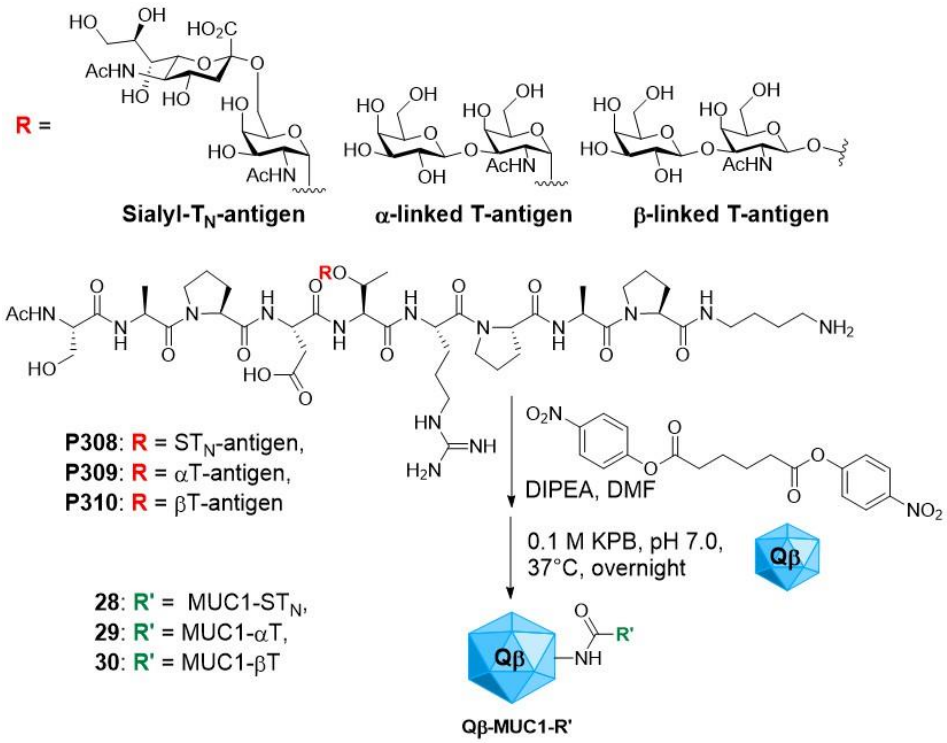


Figure 54. Synthesis of MUC1 glycopeptides **P308** - **P310** exhibiting the T_N-, ST_N-, αT- and βT-antigens, respectively, and their conjugation to bacteriophage Qβ.

4.4.3.2 Immunological evaluation of anti-αT-MUC1 mouse antibodies by microarray assay

In order to evaluate the ability of the Qβ-MUC1 vaccine conjugate **29** to induce strong and specific immune responses in vivo, and its possible applicability in cancer treatment, the binding specificities of four α-T-MUC1 mouse antisera toward different core MUC1 glycopeptides were determined (list of peptides see Chapter 4.3.3 Table 14).

Microarray analysis showed that all α-T-MUC1 antibodies exhibited a broad selectivity toward all MUC1 glycopeptides with slightly different binding patterns. Peptides glycosylated in the PDTR region were the best binders (Figure 55, Figure 56). Interestingly, no increased binding to peptide **188** containing the antigen-glycopeptide motif PDT*(α-T-antigen)R was observed. Bivalent glycopeptides containing the PDTR glycosylation site were recognized with a similar strength as the corresponding monoglycosylated PDTR glycopeptide analogs. In contrast, binding to bivalent peptides glycosylated in the GVTS and GSTA regions was decreased. These results imply that the induced antibodies exhibited a certain selectivity for the glycosylation site of the antigen peptide in the Qβ-MUC1-αT conjugate **29**. Furthermore, no cross-reactivities could be observed for MUC5B glycopeptides or the glycoproteins fetuin from fetal bovine serum, poly(LacNAc)$_3$-BSA-neoglycoprotein, human ICAM-1, human transferrin, porcine stomach mucin and bovine submaxillary mucin. These findings indicate that the generated antibodies were specific for glycosylated MUC1 rather than the glycan epitope only.

Because tumor-associated carbohydrate antigens on MUC1 at the tumor cell surface are presented with a certain glycan microheterogeneity, antibody cross-reactivity to related tumor-associated glycan structures might be desired for efficient cancer treatment.[303-304] However, these antibodies could potentially cross-react with core epitopes on healthy cells. Since a synthetic antitumor vaccine should lead to immunological memory, this cross-reactivity might cause severe autoimmune diseases. As a result, the impact of cross-reactive antisera on healthy cells needs to be further explored. It might be more beneficial for cancer treatment if the induced antibodies were more specific to the carbohydrate antigen structure presented on the MUC1 glycopeptide vaccine conjugate.

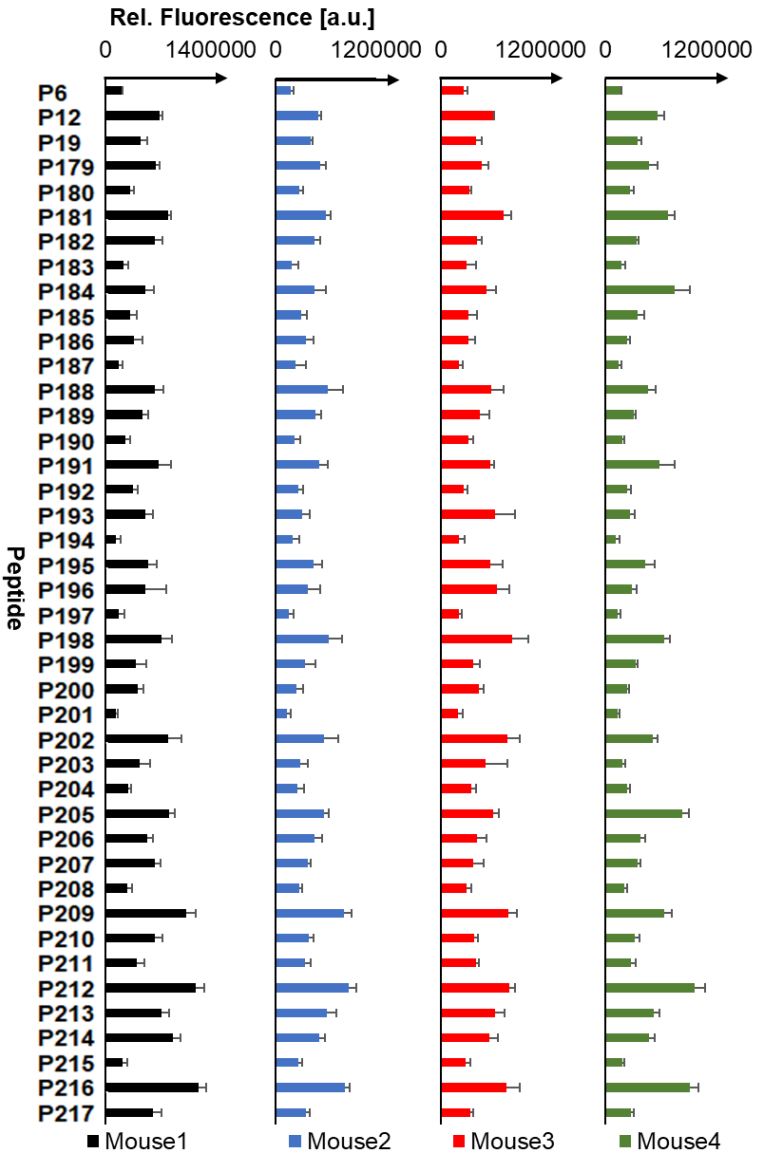


Figure 55. Binding of anti-αT-MUC1 antibodies toward T-, T_N-antigen, core 1 and core 3 structures on mono-, bi- and trivalent MUC1 glycopeptides. All sera were tested at 1:25 dilutions. [a.u.] = arbitrary units.

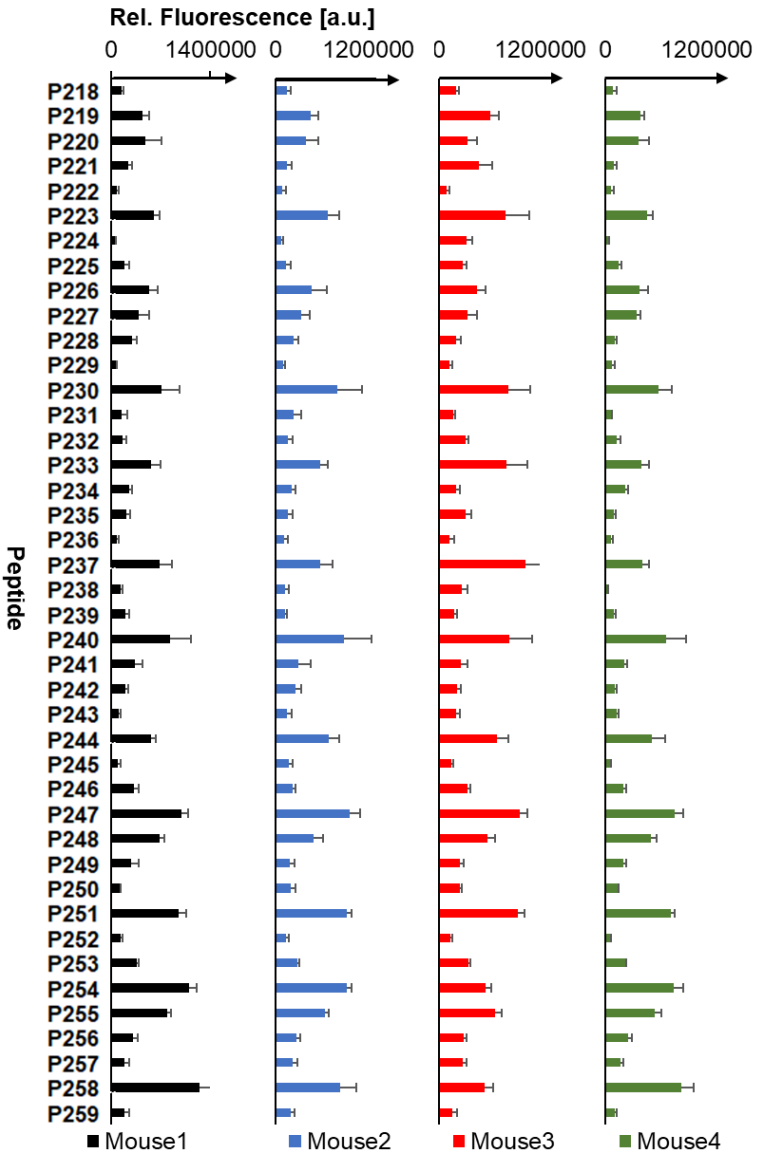


Figure 56. Binding of anti-αT-MUC1 antibodies toward core 2 and core 4 structures on mono-, bi- and trivalent MUC1 glycopeptides. All sera were tested at 1:25 dilutions. [a.u.] = arbitrary units.

4.4.3.3 Immunological evaluation of anti-βT-MUC1 mouse antibodies by microarray assay

The binding specificities of five β-T-MUC1 mouse antisera toward different core MUC1 glycopeptides were explored and compared with the data for the α-T-MUC1 antibodies in order to evaluate the ability of the Qβ-MUC1 vaccine conjugate **30** to induce strong and specific immune responses in vivo, and if the antibodies directed against the unnatural β-linkage showed differences in binding, or an improved selectivity for the carbohydrate antigen and the glycosylation site in comparison with the antibodies directed against the α-T-antigen (list of peptides see Chapter 4.3.3 Table 14). In line with the α-T-MUC1 antibody data, anti-β-T-MUC1 antibodies exhibited a broad selectivity toward all MUC1 glycopeptides (Figure 57, Figure 58). Again, slight differences in the binding patterns of the different mouse sera were observed. Unexpectedly, mono- and bivalent peptides glycosylated in the GSTA region were slightly stronger recognized than the respective derivatives glycosylated in the PDTR region. The anti-βT-MUC1 mouse antibodies also did not bind to MUC5B glycopeptides or glycoproteins including the α1-acid glycoprotein, fetuin from fetal bovine serum, BSA, albumin from human serum, poly(LacNAc)$_3$-BSA-neoglycoprotein, human ICAM-I, human transferrin, mucin from porcine stomach and mucin from bovine submaxillary. These results confirmed the specificity of the induced anti-βT-MUC1 antibodies toward MUC1 glycopeptides indicating that both the glycan and the MUC1 peptide backbone are important for antibody recognition.

In summary, the antibodies generated from immunization with the unnatural β-T-MUC-Qβ vaccine **30** displayed a similar binding pattern as the antibodies derived from the α-T-MUC1-Qβ vaccine **29**. No improved selectivity for neither the glycan epitope, nor the glycosylation site could be observed. The increased glycosidase cleavage stability of vaccine conjugate **30** is useful for the design of novel cancer vaccines.

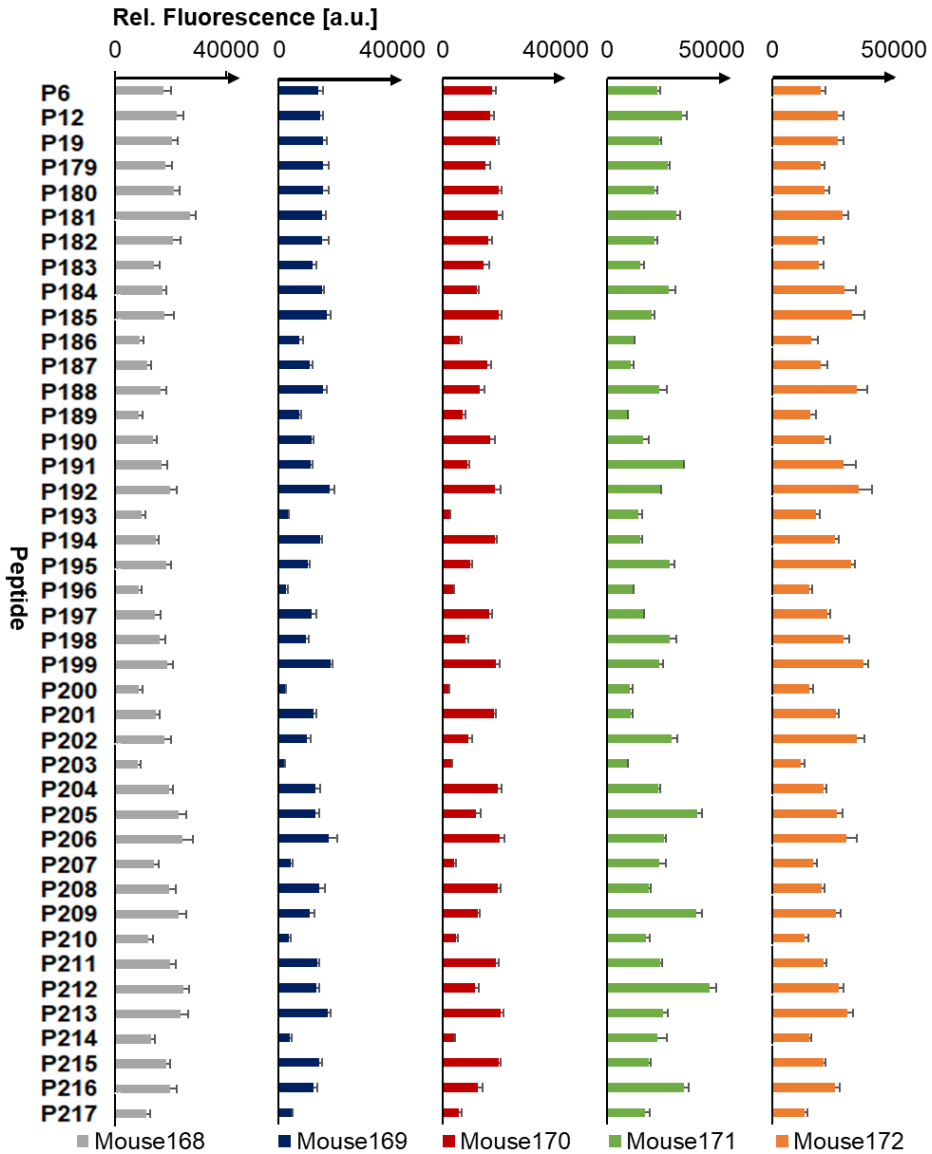


Figure 57. Binding of anti-βT-MUC1 antibodies toward T-, T_N-antigen, core 1 and core 3 structures on mono-, bi- and trivalent MUC1 glycopeptides. All sera were tested at 1:20 dilutions. [a.u.] = arbitrary units.

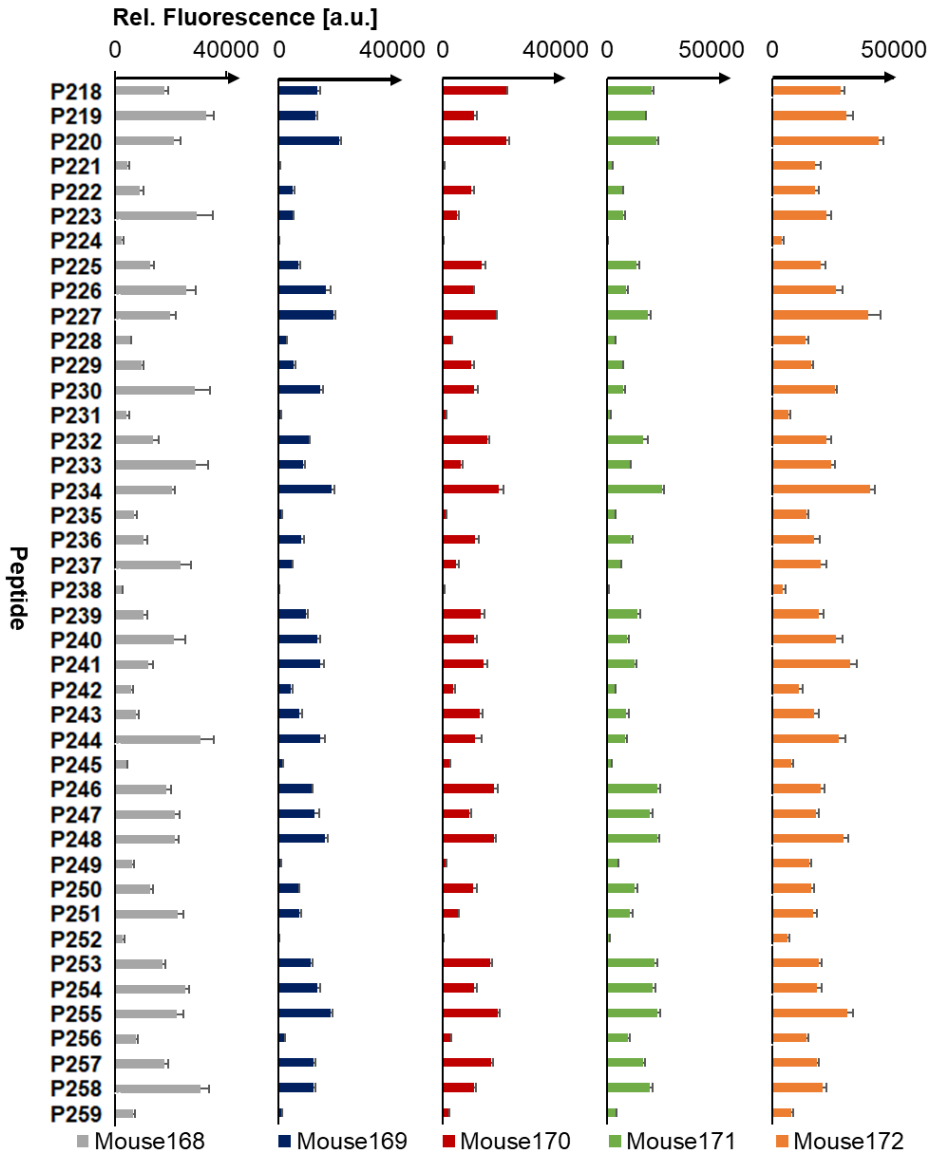


Figure 58. Binding of anti-βT-MUC1 antibodies toward core 2 and core 4 structures on mono-, bi- and trivalent MUC1 glycopeptides. All sera were tested at 1:20 dilutions. [a.u.] = arbitrary units.

4.4.3.4 Immunological evaluation of anti-ST_N-MUC1 mouse antibodies by microarray assay

To evaluate the ability of the Qβ-MUC1 vaccine conjugate **28** to induce strong and specific immune responses in vivo, and its possible applicability in cancer treatment, the binding specificities of five ST_N-MUC1 antisera from mice immunized with the Qβ-MUC1-ST_N conjugate **28** toward different core MUC1 glycopeptides were determined (list of peptides see Chapter 4.3.3 Table 14). Additionally, their specificities for the carbohydrate epitope were compared with the specificities of the antibodies induced by the vaccine conjugates **29** and **30** that carry the T-antigen.

Microarray analysis showed that all ST_N-MUC1 mouse antibodies bound strongly to MUC1 glycopeptides carrying the T_N-antigen on different glycosylation sites (Figure 59, Figure 60). Strong recognition of mono- and bivalent T_N-antigen peptides glycosylated in the PDTR region was observed indicating that the mouse antibodies exhibited a high selectivity for the glycosylation site of the antigen peptide in the Qβ-MUC1-ST_N conjugate **28** that was used to immunize the mice.

The increase in binding for T_N-glycosylation in the PDTR region in combination with glycosylation in the GVTS and/or GSTA domains can be attributed to glycan clustering effects in antibody recognition that are often overlooked, even though MUC1 antibodies that preferably bind to clustered T_N-antigen in the GSTA region have been reported.[305] Additionally, mouse sera **6** to **9** also showed weak interactions with other core MUC1 glycopeptides with varying recognition patterns. No cross-reactivities were observed for MUC5B glycopeptides or the glycoproteins fetuin from fetal bovine serum, poly(LacNAc)$_3$-BSA-neoglycoprotein, human ICAM-1, human transferrin, porcine stomach mucin and bovine submaxillary mucin. Consequently, the generated ST_N-MUC1 antibodies were specific for glycosylated MUC1 rather than the glycan epitope only.

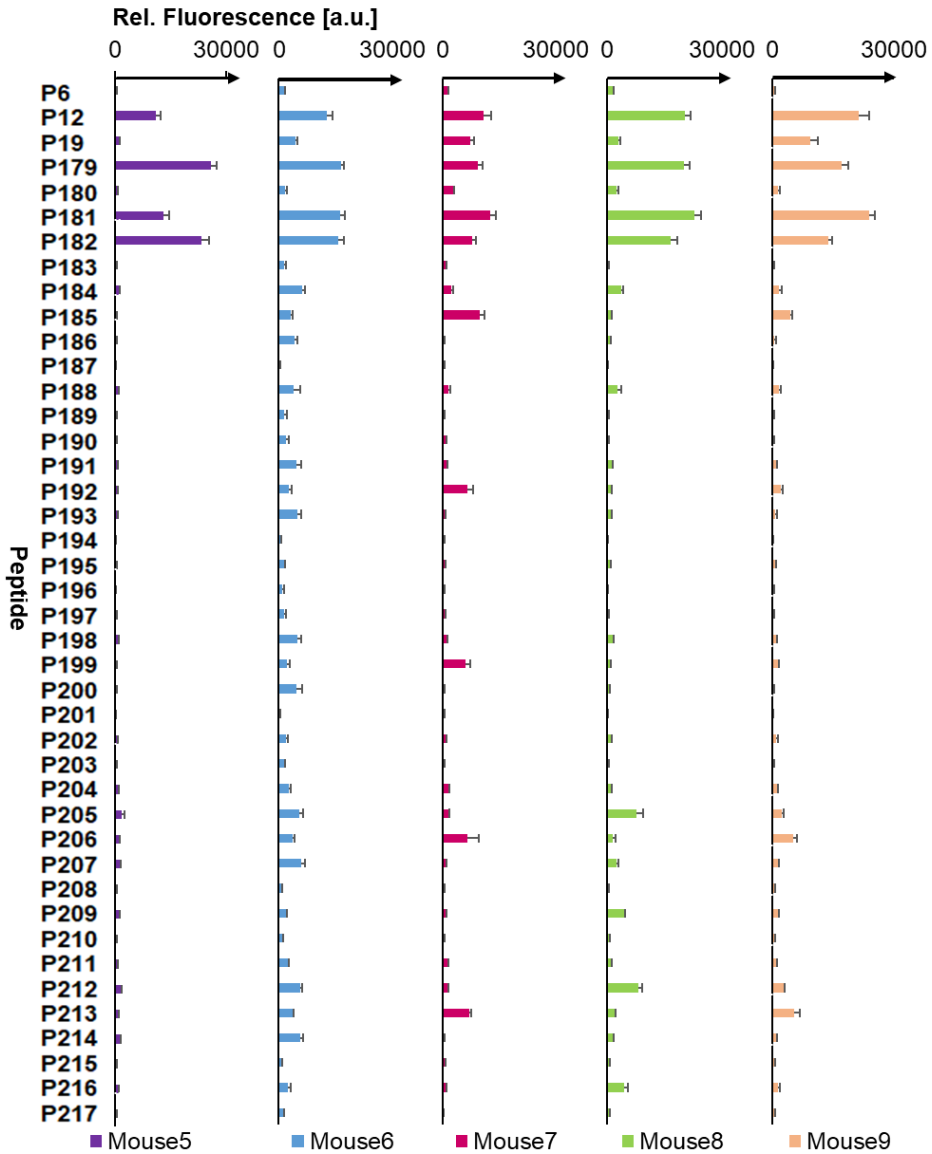


Figure 59. Binding of anti-ST_N-MUC1 antibodies toward T-, T_N-antigen, core 1 and core 3 structures on mono-, bi- and trivalent MUC1 glycopeptides. All sera were tested at 1:100 dilutions. [a.u.] = arbitrary units.

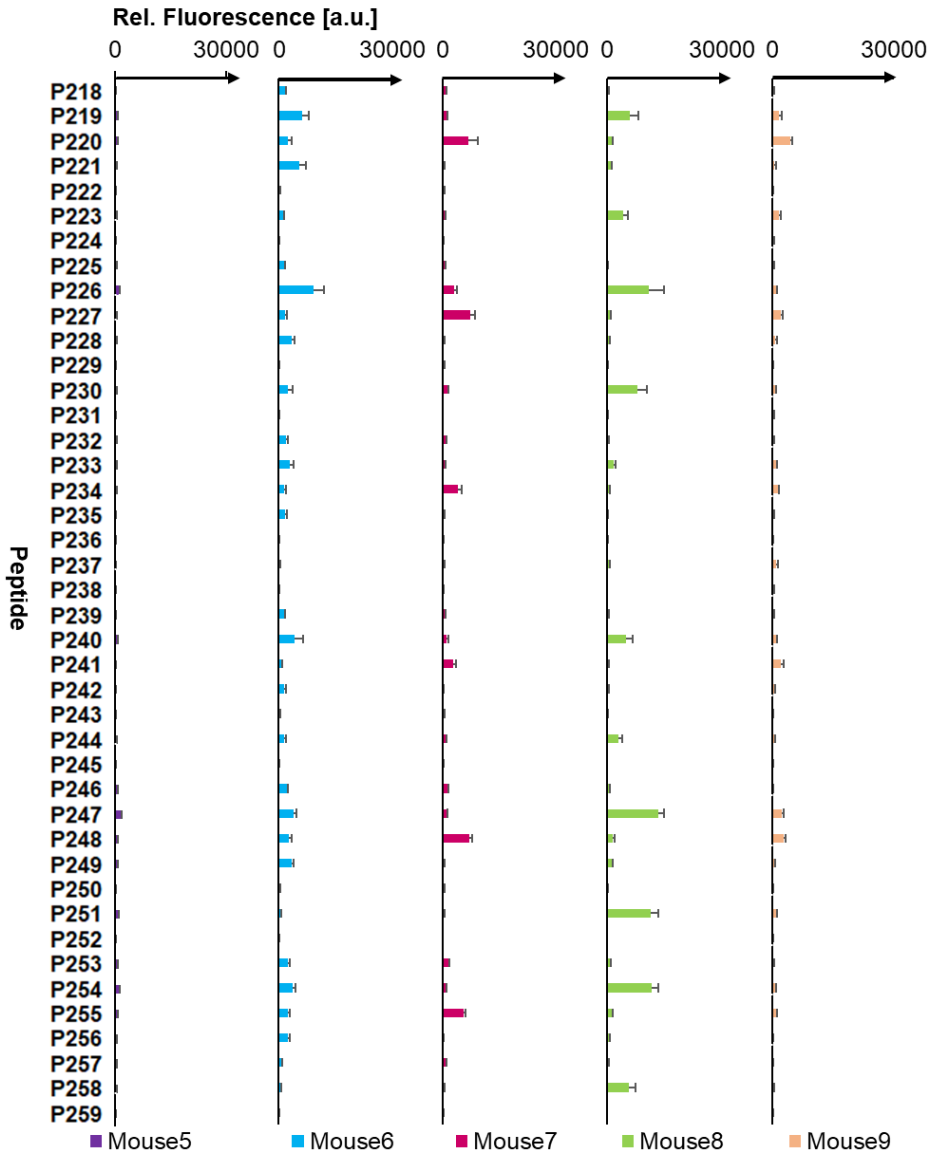


Figure 60. Binding of anti-ST_N-MUC1 antibodies toward core 2 and core 4 structures on mono-, bi- and trivalent MUC1 glycopeptides. All sera were tested at 1:100 dilutions. [a.u.] = arbitrary units.

Finally, the binding-specificities of the ST_N-MUC1 mouse sera antisera were evaluated using selected T_N-, ST_N-, T- and ST-antigen MUC1 glycopeptides (Table 18). As expected, the peptide **P312** PAHGVTSAPDT*(ST_N)RPAPGSTA containing the antigen peptide of the Qβ-MUC1-ST_N conjugate **28** was the best binder (Figure 61). Surprisingly, sialylation of the T_N- and T-antigen peptides decreased antibody binding compared with the corresponding un-sialylated MUC1 glycopeptide derivatives.

In summary, the antibodies generated from immunization with the ST_N-MUC1-Qβ vaccine **28** were highly selective for MUC1 glycopeptides glycosylated with the T_N-antigen and for the ST_N-modified peptide **P312** that encompasses the antigen peptide **P308**. As a result, these antibodies showed a higher specificity for the carbohydrate antigen epitope than the antibodies induced by the T-MUC1-Qβ vaccines **29** and **30**. Because the anti-ST_N-MUC1 only exhibited weak cross-reactivity for other mucin core structures and might not cross-react with mucins on healthy cells, ST_N-MUC1-based vaccines might be more useful for therapeutic cancer treatment than T-MUC1-based vaccines.

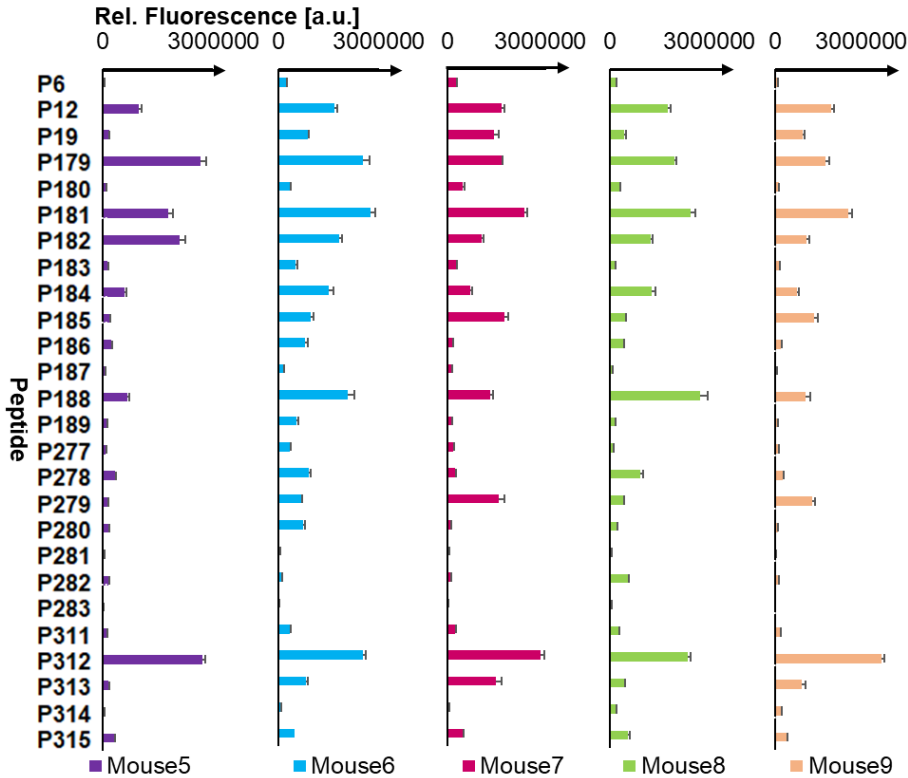


Figure 61. Binding of anti-ST_N-MUC1 antibodies toward T_N-, T-, ST_N- and ST-antigen MUC1 glycopeptides. All sera were tested at 1:100 dilutions. [a.u.] = arbitrary units.

Table 18. List of T_N-, T-, ST_N- and ST-antigen MUC1 glycopeptides.

ID	Glycopeptide sequence
P6	PAHGV**T*** (T_N)SAPDTRPAPGSTAPPA
P12	PAHGVTSAPD**T*** (T_N)RPAPGSTAPPA
P19	PAHGVTSAPDTRPAPGS**T*** (T_N)APPA
P179	PAHGV**T*** (T_N)SAPD**T*** (T_N)RPAPGSTAPPA
P180	PAHGV**T*** (T_N)SAPDTRPAPGS**T*** (T_N)APPA
P181	PAHGVTSAPD**T*** (T_N)RPAPGS**T*** (T_N)A PPA
P182	PAHGV**T*** (T_N)SAPD**T*** (T_N)RPAPGS**T*** (T_N)APPA
P183	PAHGV**T*** (T)SAPDTRPAPGSTAPPA
P184	PAHGVTSAPD**T*** (T)RPAPGSTA PPA
P185	PAHGVTSAPDTRPAPGS**T*** (T)A PPA
P186	PAHGV**T*** (T)SAPD**T*** (T)RPAPGSTAPPA
P187	PAHGV**T*** (T)SAPDTRPAPGS**T*** (T)APPA
P188	PAHGVTSAPD**T*** (T)RPAPGS**T*** (T)APPA
P189	PAHGV**T*** (T)SAPD**T*** (T)RPAPGS**T*** (T)APPA
P277	PAHGV**T*** (ST)SAPDTRPAPGSTAPPA
P278	PAHGVTSAPD**T*** (ST)RPAPGSTA PPA
P279	PAHGVTSAPDTRPAPGS**T*** (ST)A PPA
P280	PAHGV**T*** (ST)SAPD**T*** (ST)RPAPGSTAPPA
P281	PAHGV**T*** (ST)SAPDTRPAPGS**T*** (ST)APPA
P282	PAHGVTSAPD**T*** (ST)RPAPGS**T*** (ST)APPA
P283	PAHGV**T*** (ST)SAPD**T*** (ST)RPAPGS**T***(ST)APPA
P311	PAHGV**T*** (ST_N)SAPDTRPAPGSTAPPA
P312	PAHGVTSAPD**T*** (ST_N)RPAPGSTAPPA
P313	PAHGVTSAPDTRPAPGS**T*** (ST_N)APPA
P314	PAHGV**T*** (ST_N)SAPDTRPAPGS**T*** (ST_N)APPA
P315	PAHGV**T*** (ST_N)SAPDTRPAPGS**T*** (T_N)APPAHGV**T*** (ST_N) SAPDTRPAPGS**T*** (T_N)APPA

4.4.4 Conclusion

In summary, MUC1-ST_N and –α-T glycopeptide bacteriophage Qβ conjugates **28** and **29** were prepared. Additionally, the Qβ-β-T-MUC1 vaccine **30** was generated that exhibited a higher stability toward glycosidase cleavage than the natural α-linked T-antigen analog. The resulting conjugates were used to immunize MUC1.Tg mice and elicited high levels of anti-MUC1 IgG antibodies. The induced antibodies were immunologically analyzed using microarray analysis and were shown to be specific for MUC1. Mouse sera induced by the α- and β-T-MUC1-Qβ vaccine conjugates **29** and **30** bound to a broad range of MUC1 glycoforms. While peptides glycosylated in the PDTR and/or GSTA regions were found to be the best, binding to bivalent peptides glycosylated in the GVTS and GSTA regions was decreased. These results suggest that the induced antibodies exhibited a certain selectivity for the glycosylation site of the antigen peptides in the vaccine conjugates **29** and **30**. Mouse sera induced by the ST_N-MUC1-Qβ vaccine conjugate **28** displayed a higher specificity than the antibodies obtained from immunization with vaccines **29** and **30** and were highly selective for MUC1 peptides glycosylated with the TN-antigen, as well as the antigen peptide SAPDT*(ST_N)RPAP **P308**. It could also been shown that a simple change of the stereochemistry of the glycosidic bond coupling the glycan to the peptide part could increase the stability against glycosidase cleavage, and represents an efficient strategy for anti-cancer vaccine epitope design.

Usually, the development of cancer vaccines, which elicit antibodies that are highly specific to the carbohydrate antigen structure presented on the vaccine conjugate, are desired for efficient cancer treatment. However, tumor-associated carbohydrate antigens on MUC1 at the tumor cell surface are presented with a certain carbohydrate microheterogeneity. As a result, antibody cross-reactivity to other tumor-associated glycan structures might be beneficial for effective cancer treatment.[303-304] Nevertheless, these antibodies could potentially cross-react with mucin core structures on healthy cells. Because a synthetic antitumor vaccine usually creates an immunological memory, this cross-reactivity might lead to severe autoimmune diseases. Consequently, the impact of antibody cross-reactivity with healthy cells needs to be further explored. Glycopeptide microarrays are valuable tools for epitope mapping of antibodies induced by glycopeptide vaccines and to discover possible cross-reactivities with glycan structures that are also found on

healthy cells. By using glycopeptide microarray binding studies, the most efficient glycopeptide epitope structures could be identified and used to design new and efficient synthetic vaccines that induce strong immune responses with optimal tumor selectivity.

5 Final conclusions and relevance

The aims of this thesis was to prepare structurally well-defined synthetic mucin glycopeptides to build-up a glycopeptide microarray platform that were used as tools to evaluate the binding specificities of carbohydrate-binding proteins, including plant lectins, bacterial lectins and human galectins and antibodies on a molecular level. Additionally, glycopeptide microarrays were used to evaluate the abilities of induced specific antibodies to detect a new group of tyrosine *O*-glycosylation. The advantage of using glycopeptides for the preparation of microarrays is that glycopeptide microarrays mimic the natural glycan presentation on the cell surface. Consequently, they and are useful tools to study, in addition to the binding specificities, the impact of glycosylation site placement and the peptide sequence on carbohydrate-binding proteins.

Chapter 3 Development and Evaluation of Tools to Explore HexNAc-*O*-Tyrosine Glycosylation

HexNAc-*O*-Tyr-CRM vaccines were prepared to raise specific rabbit antibodies against this new group of modifications. Immunological evaluation by ELISA and glycopeptide microarray assays showed that high levels of HexNAc-*O*-Tyr-specific antibodies were induced. The antibodies could detect α-GlcNAc-*O*-Tyr modified RhoA by western blot analysis. Additionally, the abilities of lectin-based enrichment methods, as well as of a β-GlcNAc specific monoclonal antibody to detect HexNAc-*O*-Tyr modifications were evaluated. The lectins and antibody bound tightly to this PTM. As a result of this work, new tools are now available to detect and enrich proteins carrying this new group of PTMs, and to explore the glycobiology behind tyrosine *O*-HexNAcylation.

Since HexNAc-*O*-Tyr specific rabbit antibodies, *O*-β-GlcNAc-specific mAb CTD 110.6 and the plant lectins, VVA, WGA and GSL II, exhibit complementing affinities it would be interesting to use a combined approach of these tools in a glycoproteomic study to enable selective detection, identification and enrichment of α-GalNAc-, β-GlcNAc- and α-GlcNAc-O-Tyr modified proteins and tryptic glycopeptides.

Chapter 4 Tools to explore mucin-type glycosylation

Glycopeptide microarrays were used as tools to evaluate binding specificities of bacterial lectins LecB from *Pseudomonas aeruginosa* and the *Clostridium difficile* toxin A, and human galectins. They were also used to map epitopes of cancer-specific serum antibodies generated from administration of synthetic vaccines to mice. To enable the study of these protein-glycan interactions, structurally well-defined glycopeptides were prepared to build-up a glycopeptide microarray platform. This platform contained mucin core glycopeptides, as well as fucosylated and sialylated glycan structures.

Glycopeptides can easily be enzymatically modified and the mucin core glycopeptide library was further extended with Le^x, Le^a, Le^y and Le^b motifs. A fucosylated glycopeptide library was used to study the fine binding specificities of the fucose-recognizing *P. aeruginosa* lectin LecB and the *C. difficile* toxin A. Both proteins exhibited unique fine specificities that strongly depended on the different fucose motifs, presenting peptide backbone, underlying core structures, LacNAc-extension as well as placement of the glycosylation sites.

This glycopeptide library could furthermore be used to study other fucose-binding bacterial and viral lectins in order to get a better understanding of the individual pathogenic adhesion processes. Additionally, information about fine binding specificities are useful for the development of efficient glycomimetic inhibitors to fight pathogenic infections.

To be able to study a broad variety of carbohydrate-protein interactions, we aim to further extend and diversify our glycopeptide library. In this context, simplified mucin core threonine building blocks were synthesized and incorporated into mucin glycopeptides.

The next step would be to employ the simplified mucin glycopeptides in enzymatic reactions to easily access LacdiNAc, sulfo-LacdiNAc and sulfo-LacNAc modified glycan structures. Additional enzymatic fucosylation and sialylation of the obtained glycopeptides would further diversify the glycopeptide library. This platform could then be used to study proteins such as galectins and LabA from *H. pylori* that possibly recognize LacdiNAc, sulfo-LacdiNAc and/or sulfo-LacNAc structures.

Unmodified, LacNAc elongated and/or α-1,3- and α-1,4-fucosylated, and/or α-2,3- and α-2,6-sialylated mucin glycopeptide microarrays were used to study galectins binding interactions. All galectins showed distinct binding preferences for the different mucin cores, type-1 and type 2 glycans, sialyation and fucosylation on the core structures. The exact placement of the glycosylation site(s) on the peptide backbone had a strong impact on galectin recognition. Because galectins play important roles, knowledge about their fine binding specificities could advance the development of therapeutic cancer treatments such as carbohydrate-based inhibitors, and could also improve our comprehension of the roles the individual galectins play in their respective biological processes.

Further studies need to be performed to explore the binding preferences of galectins. For example, the newly extended fucosylated glycopeptide library that was created to study LecB and TcdA could be used to also explore galectin recognition to α-1,2-fucosylated, and Le^y and Le^b modified glycopeptides. Additionally, the LacdiNAc, sulfo-LacdiNAc and sulfo-LacNAc modified glycopeptide library could be used to explore galectin binding. Galectin binding studies with different sample concentrations would also be interesting to obtain a more complete picture of galectin binding interactions.

The mucin core glycopeptide library was also used for epitope mapping of cancer-specific serum antibodies generated from immunization of mice with synthetic vaccines. Glycopeptide microarrays are useful tools for epitope mapping of antibodies induced by glycopeptide vaccines and to discover possible cross-reactivities with glycan structures that are also found on healthy cells. Glycopeptide microarray binding studies can also help to identify the most efficient glycopeptide epitope structures for the design of efficient synthetic vaccines that induce strong immune responses with optimal tumor selectivity.

6 Acknowledgement

First of all, I would like to express my gratitude to my supervisor Ulrika Westerlind, who gave me the opportunity to work on these projects. Thank you for supervising this thesis, for all the discussions we had, your scientific guidance and the very nice Midsommar this year.
I would like to thank Mikael Elofsson for being my second supervisor, for including me in your group and letting me participate in your seminars.
Many thanks go to my fellow labmates Ganesh, Maruthi, Gergö and Pett for the scientific discussions and all the funny stories. Christian, vielen Dank, dass du mir deine Projekte vererbt hast, für all deine Hilfe in den letzten Jahren und die gemeinsamen Reparaturversuche der LC-MS-Systeme. Christian Hedberg, I really enjoyed our discussions about food, meat and booze a lot!
A big thanks goes to all the people on the chemistry corridor. It was super nice to work with you!

To all my friends in Umeå, thank you for making it worthwile, for the great board game evenings and all the fun time we spend together. Sam, thank you for all the smaller and bigger distractions during the last few weeks, for always cheering me up and supporting me (and for proofreading this thesis of course). Liebe Laura, vielen Dank für all die tollen Stunden auf deinem Balkon und ich hoffe, dass es noch viele mehr werden. Martin, ich werde unsere kleinen großen Fikapausen sehr vermissen.

I also don't want to forget to thank all the people at ISAS in Dormund where I spend the first two years of my PhD. Olli and Svenja, it wouldn't have been the same without you.
Meinen Freunden in Deutschland, vor allem Anna, Katty und Katha, möchte ich ein großes Dankeschön sagen, dass ihr mich nicht vergessen habt.

Mein größter Dank gilt meiner Familie, die mich immer unterstützt und mir das Studium und die Promotion ermöglicht hat. Mein lieber Peter, vielen Dank, dass du fast drei Jahre auf mich gewartet hast, dass du immer an meiner Seite gestanden und mich unterstützt hast (trotz fast 1800 km Luftlinie zwischen uns).

References

[1] F. W. Lichtenthaler, *Angew. Chem. Int. Ed.* **1992**, *31*, 1541-1556.
[2] E. Fischer, *Ber. Dtsch. Chem. Ges.* **1891**, *24*, 1836-1845.
[3] E. Fischer, *Ber. Dtsch. Chem. Ges.* **1891**, *24*, 2683-2687.
[4] A. Varki, *Glycobiology* **2017**, *27*, 3-49.
[5] D. C. Frost, L. Li, in *Adv. Protein Chem. Struct. Biol., Vol. 95* (Ed.: R. Donev), Academic Press, **2014**, pp. 71-123.
[6] S. Suttapitugsakul, F. Sun, R. Wu, *Anal. Chem.* **2020**, *92*, 267-291.
[7] J. L. Abrahams, G. Taherzadeh, G. Jarvas, A. Guttman, Y. Zhou, M. P. Campbell, *Curr. Opin. Struct. Biol.* **2020**, *62*, 56-69.
[8] A. Michael, *Am. Chem. J* **1879**, *1*, 305-312.
[9] W. Koenigs, E. Knorr, *Ber. Dtsch. Chem. Ges.* **1901**, *34*, 957-981.
[10] E. Fischer, *Ber. Dtsch. Chem. Ges.* **1893**, *26*, 2400-2412.
[11] R. Lemieux, *Explorations with sugars : how sweet it was*, Am. Chem. Soc., Washington, DC., **1990**.
[12] R. U. Lemieux, N. J. Chü, *133rd National Meeting of the American Chemical Society, San Francisco, CA. Washington* **1958**, *133*, N31.
[13] R. U. Lemieux, S. Koto, D. Voisin, in *Anomeric Effect, Vol. 87*, Journal of the American Chemical Society, **1979**, pp. 17-29.
[14] R. R. Schmidt, *Angew. Chem. Int. Ed.* **1986**, *25*, 212-235.
[15] O. J. Plante, E. R. Palmacci, P. H. Seeberger, *Science* **2001**, *291*, 1523.
[16] P. Sears, C.-H. Wong, *Science* **2001**, *291*, 2344.
[17] J. Y. Hyun, J. Pai, I. Shin, *Acc. Chem. Res.* **2017**, *50*, 1069-1078.
[18] L. Liu, C. S. Bennett, C.-H. Wong, *Chem. Commun.* **2006**, 21-33.
[19] C. R. Torres, G. W. Hart, *J. Biol. Chem.* **1984**, *259*, 3308-3317.
[20] R. S. Haltiwanger, G. D. Holt, G. W. Hart, *J. Biol. Chem.* **1990**, *265*, 2563-2568.
[21] L. Wells, K. Vosseller, G. W. Hart, *Science* **2001**, *291*, 2376.
[22] N. E. Zachara, G. W. Hart, *Biochimica et Biophysica Acta (BBA) - General Subjects* **2004**, *1673*, 13-28.
[23] G. W. Hart, C. Slawson, G. Ramirez-Correa, O. Lagerlof, *Annu. Rev. Biochem* **2011**, *80*, 825-858.
[24] R. P. Taylor, G. J. Parker, M. W. Hazel, Y. Soesanto, W. Fuller, M. J. Yazzie, D. A. McClain, *J. Biol. Chem.* **2008**, *283*, 6050-6057.
[25] N. Zachara, Y. Akimoto, G. W. Hart, *The O-GlcNAc Modification*, 3rd edition ed., Cold Spring Harbor Laboratory Press, **2017**.
[26] S. Hardivillé, Gerald W. Hart, *Cell Metab.* **2014**, *20*, 208-213.
[27] M. W. Krause, D. C. Love, S. K. Ghosh, P. Wang, S. Yun, T. Fukushige, J. A. Hanover, *Front. Endocrinol. (Lausanne)* **2018**, *9*, 521.
[28] C. Gewinner, G. Hart, N. Zachara, R. Cole, C. Beisenherz-Huss, B. Groner, *J. Biol. Chem.* **2004**, *279*, 3563-3572.
[29] G. W. Hart, *Sci. Signal.* **2013**, *6*, pe26.
[30] T. Ohn, N. Kedersha, T. Hickman, S. Tisdale, P. Anderson, *Nat. Cell Biol.* **2008**, *10*, 1224-1231.
[31] X. Yang, K. Qian, *Nat. Rev. Mol. Cell. Biol.* **2017**, *18*, 452-465.
[32] D. Wu, Y. Cai, J. Jin, *Protein Cell* **2017**, *8*, 713-723.
[33] V. Dehennaut, D. Leprince, T. Lefebvre, *Front. Endocrinol. (Lausanne)* **2014**, *5*, 155-155.
[34] S. A. Yuzwa et al., *Nat. Chem. Biol.* **2012**, *8*, 393-399.
[35] L. S. Griffith, M. Mathes, B. Schmitz, *J. Neurosci. Res.* **1995**, *41*, 270-278.
[36] E. P. Bennett et al., *Glycobiology* **2012**, *22*, 736-756.
[37] A. Harduin-Lepers, R. Mollicone, P. Delannoy, R. Oriol, *Glycobiology* **2005**, *15*, 805-817.

[38] A. Harduin-Lepers, *Glycobiol. Insigths* **2010**, 29-61.
[39] H. Schachter, E. J. McGuire, S. Roseman, *J. Biol. Chem.* **1971**, *246*, 5321-5328.
[40] T. Ju, K. Brewer, A. D'Souza, R. D. Cummings, W. M. Canfield, *J. Biol. Chem.* **2002**, *277*, 178-186.
[41] M. F. Bierhuizen, M. Fukuda, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9326.
[42] T. Iwai et al., *J. Biol. Chem.* **2002**, *277*, 12802-12809.
[43] J.-C. Yeh, E. Ong, M. Fukuda, *J. Biol. Chem.* **1999**, *274*, 3215-3221.
[44] J. T. Hutchins, C. L. Reading, *J. Cell. Biochem.* **1988**, *37*, 37-48.
[45] A. S. Kooner, H. Yu, X. Chen, *Front. Immunol.* **2019**, *10*, 2004.
[46] M. Bardor, D. H. Nguyen, S. Diaz, A. Varki, *J. Biol. Chem.* **2005**, *280*, 4228-4237.
[47] Y. N. Malykh, R. Schauer, L. Shaw, *Biochimie* **2001**, *83*, 623-634.
[48] F. Dall'Olio, *Glycoconj. J.* **2000**, *17*, 669-676.
[49] J. F. Kukowska-Latallo, R. D. Larsen, R. P. Nair, J. B. Lowe, *Genes Dev.* **1990**, *4*, 1288-1303.
[50] R. D. Larsen, L. K. Ernst, R. P. Nair, J. B. Lowe, *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 6674-6678.
[51] R. J. Kelly, S. Rouquier, D. Giorgi, G. G. Lennon, J. B. Lowe, *J. Biol. Chem.* **1995**, *270*, 4640-4649.
[52] P. Kyprianou, A. Betteridge, A. S. Donald, W. M. Watkins, *Glycoconj. J.* **1990**, *7*, 573-588.
[53] A. Sarnesto, T. Köhlin, J. Thurin, M. Blaszczyk-Thurin, *J. Biol. Chem.* **1990**, *265*, 15067-15075.
[54] S. Hemmerich, J. K. Lee, S. Bhakta, A. Bistrup, N. R. Ruddle, S. D. Rosen, *Glycobiology* **2001**, *11*, 75-87.
[55] G. Lamblin et al, *Glycoconj. J.* **2001**, *18*, 661-684.
[56] J. Dekker, J. W. A. Rossen, H. A. Büller, A. W. C. Einerhand, *Trends Biochem. Sci* **2002**, *27*, 126-131.
[57] D. W. Kufe, *Nature reviews. Cancer* **2009**, *9*, 874-885.
[58] M. G. Roy et al., *Nature* **2014**, *505*, 412-416.
[59] K. C. Kim, E. P. Lillehoj, *Am. J. Respir. Cell Mol. Biol.* **2008**, *39*, 644-647.
[60] E. S. Frenkel, K. Ribbeck, *J. Oral Microbiol.* **2015**, *7*, 29759.
[61] M. E. V. Johansson, J. M. H. Larsson, G. C. Hansson, *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 4659.
[62] S. J. Gendler et al., *J. Biol. Chem.* **1990**, *265*, 15286-15293.
[63] N. Porche et al. , *Biochem. Biophys. Res. Commun.* **1991**, *175*, 414-422.
[64] B. W. T. Yin, K. O. Lloyd, *J. Biol. Chem.* **2001**, *276*, 27371-27375.
[65] J. R. Gum, J. W. Hicks, N. W. Toribara, B. Siddiki, Y. S. Kim, *J. Biol. Chem.* **1994**, *269*, 2440-2446.
[66] V. Guyonnet Duperat et al., *Biochem. J* **1995**, *305*, 211-219.
[67] J.-L. Desseyn, V. Guyonnet-Dupérat, N. Porchet, J.-P. Aubert, A. Laine, *J. Biol. Chem.* **1997**, *272*, 3168-3178.
[68] N. W. Toribara et al., *J. Biol. Chem.* **1997**, *272*, 16398-16403.
[69] C. W. Davis, in *Mucus Hypersecretion in Respiratory Disease*, **2002**, pp. 113-131.
[70] P. Verdugo, *Am. Rev. Respir. Dis.* **1991**, *144*, 33-37.
[71] F. Levitin et al., *J. Biol. Chem.* **2005**, *280*, 33374-33386.
[72] J. Hilkens, M. Ligtenberg, H. L. Vos, S. Litvinov, *Trends Biochem. Sci* **1992**, *17*, 359-363.
[73] J. Siddiqui, M. Abe, D. Hayes, E. Shani, E. Yunis, D. Kufe, *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 2320-2323.
[74] S. Gendler, J. Taylor-Papadimitriou, T. Duhig, J. Rothbard, J. Burchell, *J. Biol. Chem.* **1988**, *263*, 12820-12823.
[75] F.-G. Hanisch, S. Müller, *Glycobiology* **2000**, *10*, 439-449.
[76] J. Dufosse et al., *Biochem. J.* **1993**, *293*, 329-337.

[77] C. De Bolós, M. Garrido, F. X. Real, *Gastroenterology* **1995**, *109*, 723-734.
[78] D. Tetaert, C. Richet, J. Gagnon, A. Boersma, P. Degand, *Carbohydr. Res.* **2001**, *333*, 165-171.
[79] D. Tetaert, K. G. Ten Hagen, C. Richet, A. Boersma, J. Gagnon, P. Degand, *Biochem. J.* **2001**, *357*, 313-320.
[80] L. Cheng et al., *FEBS Lett.* **2004**, *566*, 17-24.
[81] T. A. Fritz, J. H. Hurley, L.-B. Trinh, J. Shiloach, L. A. Tabak, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15307.
[82] J. Raman et al., *J. Biol. Chem.* **2008**, *283*, 22942-22951.
[83] J. Taylor-Papadimitriou, J. M. Burchell, Future Medicine Ltd, **2013**, p. 139.
[84] I. G. Ferreira, M. Pucci, G. Venturi, N. Malagolini, M. Chiricolo, F. Dall'Olio, *Int. J. Mol. Sci.* **2018**, *19*, 580.
[85] J. Taylor-Papadimitriou, J. Burchell, D. W. Miles, M. Dalziel, *BBA - Molecular Basis of Disease* **1999**, *1455*, 301-313.
[86] S. E. Baldus, K. Engelmann, F.-G. Hanisch, *Crit. Rev. Clin. Lab. Sci.* **2004**, *41*, 189-231.
[87] B. Monzavi-Karbassi, A. Pashov, T. Kieber-Emmons, *Vaccines* **2013**, *1*, 174-203.
[88] I. Brockhausen, J.-M. Yang, J. Burchell, C. Whitehouse, J. Taylor-Papadimitriou, *Eur. J. Biochem.* **1995**, *233*, 607-617.
[89] R. Gupta, F. Leon, S. Rauth, S. K. Batra, M. P. Ponnusamy, *Cells* **2020**, *9*, 446.
[90] J. L. Magnani, Z. Steplewski, H. Koprowski, V. Ginsburg, *Cancer Res.* **1983**, *43*, 5489.
[91] N. Rodrigues Mantuano, M. Natoli, A. Zippelius, H. Läubli, *Journal for immunotherapy of cancer* **2020**, *8*, e001222.
[92] M. Pourjafar, P. Samadi, M. Saidijam, *Immunotherapy* **2020**, *12*, 1269-1286.
[93] T. Gao, Q. Cen, H. Lei, *Biomed. Pharmacother.* **2020**, *132*, 110888.
[94] J. M. Samet, P. W. Cheng, *Environ. Health Perspect.* **1994**, *102 Suppl 2*, 89-103.
[95] K. B. Adler, D. D. Hendley, G. S. Davis, *Am. J. Pathol.* **1986**, *125*, 501-514.
[96] M. C. Rose, J. A. Voynow, *Physiol. Rev.* **2006**, *86*, 245-278.
[97] S. J. Levine, P. Larivée, C. Logun, C. W. Angus, F. P. Ognibene, J. H. Shelhamer, *Am. J. Respir. Cell Mol. Biol.* **1995**, *12*, 196-204.
[98] M. A. Hollingsworth, B. J. Swanson, *Nat. Rev. Cancer* **2004**, *4*, 45-60.
[99] G. Folkerts, W. W. Busse, F. P. Nijkamp, R. Sorkness, J. E. Gern, *Am. J. Respir. Crit. Care Med.* **1998**, *157*, 1708-1720.
[100] V. Venkatakrishnan, N. H. Packer, M. Thaysen-Andersen, *Expert Rev. Respir. Med.* **2013**, *7*, 553-576.
[101] A. D. Rhim, V. A. Kothari, P. J. Park, A. E. Mulberg, M. C. Glick, T. F. Scanlin, *Glycoconj. J.* **2000**, *17*, 385-391.
[102] M. C. Glick, V. A. Kothari, A. Liu, L. I. Stoykova, T. F. Scanlin, *Biochimie* **2001**, *83*, 743-747.
[103] B. Xia, J. A. Royall, G. Damera, G. P. Sachdev, R. D. Cummings, *Glycobiology* **2005**, *15*, 747-775.
[104] J. Poole, C. J. Day, M. von Itzstein, J. C. Paton, M. P. Jennings, *Nat. Rev. Microbiol.* **2018**, *16*, 440-452.
[105] Y. Watanabe, T. A. Bowden, I. A. Wilson, M. Crispin, *Biochim Biophys Acta Gen Subj* **2019**, *1863*, 1480-1497.
[106] J. Sjögren, M. Collin, *Future Microbiol.* **2014**, *9*, 1039-1051.
[107] E. Roilides et al, *Microbiol. Spectr.* **2015**, *3*, 3.3.22.
[108] K. F. Chung, *Eur. Respir. J.* **2001**, *18*, 50s.
[109] J. M. Rommens, et al., *Science* **1989**, *245*, 1059.
[110] X. Meng, J. Clews, V. Kargas, X. Wang, R. C. Ford, *Cell. Mol. Life Sci.* **2017**, *74*, 23-38.
[111] R. C. Boucher, *J. Intern. Med.* **2007**, *261*, 5-16.
[112] J. R. Govan, V. Deretic, *Microbiol. Rev.* **1996**, *60*, 539-574.

[113] E. E. Smith et al., *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 8487-8492.
[114] D. E. Levy, P. Fügedi, *The organic chemistry of sugars*, CRC Press, **2005**.
[115] J. P. Kamerling, *Comprehensive glycoscience*, Elsevier, **2007**.
[116] K. Ágoston, H. Streicher, P. Fügedi, *Tetrahedron: Asymmetry* **2016**, *27*, 707-728.
[117] G. Barany, R. B. Merrifield, *J. Am. Chem. Soc.* **1977**, *99*, 7363-7365.
[118] J. D. C. Codée, A. Ali, H. S. Overkleeft, G. A. van der Marel, *Comptes Rendus Chimie* **2011**, *14*, 178-193.
[119] B. Ghosh, S. S. Kulkarni, *Chem. Asian J.* **2020**, *15*, 450-462.
[120] J. T. Smoot, P. Pornsuriyasak, A. V. Demchenko, *Angew. Chem. Int. Ed.* **2005**, *44*, 7123-7126.
[121] H. D. Premathilake, A. V. Demchenko, *Top. Curr. Chem.* **2011**, *301*, 189-221.
[122] D. R. Mootoo, P. Konradsson, U. Udodong, B. Fraser-Reid, *J. Am. Chem. Soc.* **1988**, *110*, 5583-5584.
[123] B. Fraser-Reid, Z. Wu, U. E. Udodong, H. Ottosson, *J. Org. Chem.* **1990**, *55*, 6068-6070.
[124] E. Juaristi, G. Cuevas, *Tetrahedron* **1992**, *48*, 5019-5087.
[125] F. Albericio, A. El-Faham, *Org. Process Res. Dev.* **2018**, *22*, 760-772.
[126] G. Zemplén, A. Kunz, *Ber. Dtsch. Chem. Ges.* **1923**, *56*, 1705-1710.
[127] L. A. Marcaurelle, C. R. Bertozzi, *Glycobiology* **2002**, *12*, 69R-77R.
[128] L. L. Lairson, B. Henrissat, G. J. Davies, S. G. Withers, *Annu. Rev. Biochem* **2008**, *77*, 521-555.
[129] A. Hagopian, E. H. Eylar, *Arch. Biochem. Biophys.* **1968**, *128*, 422-433.
[130] M. M. Palcic, O. Hindsgaul, *Glycobiology* **1991**, *1*, 205-209.
[131] L. J. Berliner, R. D. Robinson, *Biochemistry* **1982**, *21*, 6340-6343.
[132] T. L. Lowary, O. Hindsgaul, *Carbohydr. Res.* **1993**, *249*, 163-195.
[133] T. L. Lowary, O. Hindsgaul, *Carbohydr. Res.* **1994**, *251*, 33-67.
[134] C. Unverzagt, H. Kunz, J. C. Paulson, *J. Am. Chem. Soc.* **1990**, *112*, 9308-9309.
[135] R. Liang et al., *Science* **1996**, *274*, 1520.
[136] P. W. Tang, H. C. Gool, M. Hardy, Y. C. Lee, T. Felzi, *Biochem. Biophys. Res. Commun.* **1985**, *132*, 474-480.
[137] C. R. Cantor, A. Mirzabekov, E. Southern, *Genomics* **1992**, *13*, 1378-1383.
[138] M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* **1995**, *270*, 467.
[139] S. Fukui, T. Feizi, C. Galustian, A. M. Lawson, W. Chai, *Nat. Biotechnol.* **2002**, *20*, 1011-1017.
[140] D. Wang, S. Liu, B. J. Trummer, C. Deng, A. Wang, *Nat. Biotechnol.* **2002**, *20*, 275-281.
[141] B. T. Houseman, M. Mrksich, *Chem. Biol.* **2002**, *9*, 443-454.
[142] S. Park, I. Shin, *Angew. Chem. Int. Ed.* **2002**, *41*, 3180-3182.
[143] W. G. T. Willats, S. E. Rasmussen, T. Kristensen, J. D. Mikkelsen, J. P. Knox, *Proteomics* **2002**, *2*, 1666-1671.
[144] P.-H. Liang, S.-K. Wang, C.-H. Wong, *J. Am. Chem. Soc.* **2007**, *129*, 11177-11184.
[145] C. F. Grant, V. Kanda, H. Yu, D. R. Bundle, M. T. McDermott, *Langmuir* **2008**, *24*, 14125-14132.
[146] M. Kilcoyne et al., *Anal. Chem.* **2012**, *84*, 3330-3338.
[147] J. C. Manimala, T. A. Roach, Z. Li, J. C. Gildersleeve, *Angew. Chem. Int. Ed.* **2006**, *45*, 3607-3610.
[148] A. Walz, S. Odenbreit, J. Mahdavi, T. Borén, S. Ruhl, *Glycobiology* **2005**, *15*, 700-708.
[149] O. Oyelaran, Q. Li, D. Farnsworth, J. C. Gildersleeve, *J. Proteome Res.* **2009**, *8*, 3529-3538.
[150] X. Tian, J. Pai, I. Shin, *Chem. Asian J.* **2012**, *7*, 2052-2060.
[151] A. Hoang et al., *Org. Biomol. Chem.* **2017**, *15*, 5135-5139.
[152] K. Godula, C. R. Bertozzi, *J. Am. Chem. Soc.* **2012**, *134*, 15732-15742.

[153] E. Laigre, C. Tiertant, D. Goyard, O. Renaudet, *ACS Omega* **2018**, *3*, 14013-14020.
[154] H. M. Branderhorst, R. Ruijtenbeek, R. M. J. Liskamp, R. J. Pieters, *ChemBioChem* **2008**, *9*, 1836-1844.
[155] U. Westerlind et al., *Angew. Chem. Int. Ed.* **2009**, *48*, 8263-8267.
[156] T. Horlacher, P. H. Seeberger, *OMICS* **2006**, *10*, 490-498.
[157] O. Blixt et al., *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 17033-17038.
[158] J. C. Paulson, O. Blixt, B. E. Collins, *Nat. Chem. Biol.* **2006**, *2*, 238-248.
[159] B. S. Bochner et al., *J. Biol. Chem.* **2005**, *280*, 4307-4312.
[160] T. Horlacher et al., *ChemBioChem* **2010**, *11*, 1563-1573.
[161] O. Blixt, I. Boos, U. Mandel, **2012**.
[162] M. A. Campanero-Rhodes, A. S. Palma, M. Menéndez, D. Solís, *Front. Microbiol.* **2020**, *10*, 2909.
[163] S. Park, I. Shin, *Org. Lett.* **2007**, *9*, 1675-1678.
[164] C. Pett et al., *Chem. Eur* **2017**, *23*, 3875-3884.
[165] X. Wu et al., *J. Am. Chem. Soc.* **2018**, *140*, 16596-16609.
[166] X. Wu et al., *ACS Chem. Biol.* **2019**, *14*, 2176-2184.
[167] S. L. Ivry et al., *Protein Sci.* **2018**, *27*, 584-594.
[168] A. Halim et al., *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 11848.
[169] C. Steentoft et al., *Nat. Methods* **2011**, *8*, 977-982.
[170] S. Y. Vakhrushev et al., *Mol. Cell. Proteom.* **2013**, *12*, 932-944.
[171] J. C. Trinidad, R. Schoepfer, A. L. Burlingame, K. F. Medzihradszky, *Mol. Cell. Proteom.* **2013**, *12*, 3474-3488.
[172] T. Jank et al., *Nat. Struct. Mol. Biol.* **2013**, *20*, 1273-1280.
[173] T. Jank et al., *Nat. Commun* **2015**, *6*, 7807.
[174] J. Yu et al., *Chem. Eur* **2016**, *22*, 1114-1124.
[175] K. Kubota et al., *Anal. Chem.* **2008**, *80*, 3693-3698.
[176] C. Pett et al., *Angew. Chem. Int. Ed.* **2018**, *57*, 9320-9324.
[177] S. E. Tollefsen, R. Kornfeld, *J. Biol. Chem.* **1983**, *258*, 5165-5171.
[178] Y. Nagata, M. M. Burger, *J. Biol. Chem.* **1974**, *249*, 3116-3122.
[179] J. T. Gallagher, A. Morris, T. M. Dexter, *Biochem. J* **1985**, *231*, 115-122.
[180] B. P. Peters, S. Ebisu, I. J. Goldstein, M. Flashner, *Biochemistry* **1979**, *18*, 5505-5511.
[181] P. N. S. Iyer, K. D. Wilkinson, I. J. Goldstein, *Arch. Biochem. Biophys.* **1976**, *177*, 330-333.
[182] S. Nakamura-Tsuruta et al., *J. Biochem.* **2006**, *140*, 285-291.
[183] C. Pett, M. Schorlemer, U. Westerlind, *Chem. Eur* **2013**, *19*, 17001-17010.
[184] C. Pett, U. Westerlind, *Chem. Eur* **2014**, *20*, 7287-7299.
[185] D. Schwefel, C. Maierhofer, J. G. Beck, S. Seeberger, K. Diederichs, H. M. Möller, W. Welte, V. Wittmann, *J. Am. Chem. Soc.* **2010**, *132*, 8704-8719.
[186] A. Frey, J. Di Canzio, D. Zurakowski, *J. Immunol. Methods* **1998**, *221*, 35-41.
[187] M. Schneider, E. Al-Shareffi, R. S. Haltiwanger, *Glycobiology* **2017**, *27*, 601-618.
[188] V. E. Wagner, B. H. Iglewski, *Clin. Rev. Allergy Immunol.* **2008**, *35*, 124-134.
[189] D. Golemi-Kotra, in *xPharm: The Comprehensive Pharmacology Reference* (Eds.: S. J. Enna, D. B. Bylund), Elsevier, New York, **2008**, pp. 1-8.
[190] A. Imberty, M. Wimmerová, E. P. Mitchell, N. Gilboa-Garber, *Microb. Infect.* **2004**, *6*, 221-228.
[191] N. Gilboa-Garber, in *Methods Enzymol., Vol. 83*, Academic Press, **1982**, pp. 378-385.
[192] N. Garber, U. Guempel, N. Gilboa-Garber, R. J. Royle, *FEMS Microbiol. Lett.* **1987**, *48*, 331-334.
[193] D. Tielker et al., *Microbiology* **2005**, *151*, 1313-1323.
[194] E. C. Adam, B. S. Mitchell, D. U. Schumacher, G. Grant, U. Schumacher, *Am. J. Respir. Crit. Care Med.* **1997**, *155*, 2102-2104.

[195] C. Chemani et al., *Infect. Immun.* **2009**, *77*, 2065-2075.
[196] S. Behren, U. Westerlind, *Molecules* **2019**, *24*.
[197] D. A. Leffler, J. T. Lamont, *N. Engl. J. Med.* **2015**, *372*, 1539-1548.
[198] R. N. Pruitt, M. G. Chambers, K. K. S. Ng, M. D. Ohi, D. B. Lacy, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 13467-13472.
[199] K. D. Tucker, T. D. Wilkins, *Infect. Immun.* **1991**, *59*, 73-78.
[200] C.-Y. Yeh, C.-N. Lin, C.-F. Chang, C.-H. Lin, H.-T. Lien, J.-Y. Chen, J.-S. Chia, *Infect. Immun.* **2008**, *76*, 1170-1178.
[201] V. Heine, S. Boesveld, H. Pelantová, V. Křen, C. Trautwein, P. Strnad, L. Elling, *Bioconjugate Chem.* **2019**, *30*, 2373-2383.
[202] K. Aktories, *Nat. Rev. Microbiol.* **2011**, *9*, 487-498.
[203] R. Pruitt, D. B. Lacy, *Front. Cell. Infect. Microbiol.* **2012**, *2*, 28.
[204] M. Cherian Reeja, C. Jin, J. Liu, G. Karlsson Niclas, J. Holgersson, V. B. Young, *Infect. Immun.* **2016**, *84*, 2842-2852.
[205] K. Marotte et al., *ChemMedChem* **2007**, *2*, 1328-1338.
[206] A. M. Wu, J. H. Wu, T. Singh, J. H. Liu, M. S. Tsai, N. Gilboa-Garber, *Biochimie* **2006**, *88*, 1479-1492.
[207] A. M. Boukerb et al., *Front. Microbiol.* **2016**, *7*, 811-811.
[208] S. Perret et al., *The Biochemical journal* **2005**, *389*, 325-332.
[209] E. Mitchell et al., *Nature Structural Biology* **2002**, *9*, 918-921.
[210] R. Sommer et al., *Chemical Science* **2016**, *7*, 4990-5001.
[211] Y. Rossez et al., *J. Infect. Dis.* **2014**, *210*, 1286-1295.
[212] K. Hirano, A. Matsuda, T. Shirai, K. Furukawa, *Biomed Res. Int.* **2014**, *2014*, 981627-981627.
[213] E. D. Green, H. van Halbeek, I. Boime, J. U. Baenziger, *J. Biol. Chem.* **1985**, *260*, 15623-15630.
[214] A. Dell et al., *J. Biol. Chem.* **1995**, *270*, 24116-24126.
[215] S. M. Manzella, S. M. Dharmesh, C. B. Cohick, M. J. Soares, J. U. Baenziger, *J. Biol. Chem.* **1997**, *272*, 4775-4782.
[216] L. V. Hooper, M. C. Beranek, S. M. Manzella, J. U. Baenziger, *J. Biol. Chem.* **1995**, *270*, 5985-5993.
[217] A. Dell et al., *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 15631-15636.
[218] R. A. Siciliano, H. R. Morris, H. P. Bennett, A. Dell, *J. Biol. Chem.* **1994**, *269*, 910-920.
[219] I. Breloy, S. Pacharra, P. Ottis, D. Bonar, A. Grahn, F.-G. Hanisch, *Int. J. Biol. Chem.* **2012**, *287*, 18275-18286.
[220] Y. Ikehara et al., *Glycobiology* **2006**, *16*, 777-785.
[221] D. T. Kenny, E. C. Skoog, S. K. Lindén, W. B. Struwe, P. M. Rudd, N. G. Karlsson, *Glycobiology* **2012**, *22*, 1077-1085.
[222] C. Jin et al., *Mol. Cell. Proteom.* **2017**, *16*, 743-758.
[223] S. B. Yan, Y. B. Chao, H. van Halbeek, *Glycobiology* **1993**, *3*, 597-608.
[224] N. Kitamura, S. Guo, T. Sato, S. Hiraizumi, J. Taka, M. Ikekita, S. Sawada, H. Fujisawa, K. Furukawa, *Int. J. Cancer* **2003**, *105*, 533-541.
[225] R. Peracaula et al., *Glycobiology* **2003**, *13*, 457-470.
[226] E. Machado, S. Kandzia, R. Carilho, P. Altevogt, H. S. Conradt, J. Costa, *Glycobiology* **2011**, *21*, 376-386.
[227] R. Peracaula et al., *Glycobiology* **2003**, *13*, 227-244.
[228] T. Sato et al., *J. Biol. Chem.* **2003**, *278*, 47534-47544.
[229] M. Gotoh et al., *FEBS Lett.* **2004**, *562*, 134-140.
[230] D. Fiete, M. Beranek, J. U. Baenziger, *J. Biol. Chem.* **2012**, *287*, 29204-29212.
[231] D. E. Kirschner, M. J. Blaser, *J. Theor. Biol.* **1995**, *176*, 281-290.
[232] T. L. Cover, M. J. Blaser, *Annu. Rev. Med.* **1992**, *43*, 135-145.
[233] R. H. Hunt, *Scand. J. Gastroenterol. Suppl.* **1996**, *220*, 3-9.
[234] J. K. Y. Hooi et al., *Gastroenterology* **2017**, *153*, 420-429.

[235] D. Ilver et al., *Science* **1998**, *279*, 373.
[236] M. Aspholm-Hurtig et al., *Science* **2004**, *305*, 519.
[237] Y. H. Mthembu et al., *Mol. Omics* **2020**, *16*, 243-257.
[238] V. Paraskevopoulou et al., *Current Research in Structural Biology* **2021**, *3*, 19-29.
[239] V. Dourtoglou, B. Gross, V. Lambropoulou, C. Zioudrou, *Synthesis* **1984**, *1984*, 572-574.
[240] L. A. Carpino, *J. Am. Chem. Soc.* **1993**, *115*, 4397-4398.
[241] L. A. Carpino, A. El-Faham, C. A. Minor, F. Albericio, *J. Chem. Soc., Chem. Commun.* **1994**, 201-203.
[242] T. K. van den Berg et al., *J. Immunol. Res.* **2004**, *173*, 1902.
[243] A. Šimonová et al., *J. Mol. Catal. B: Enzym.* **2014**, *101*, 47-55.
[244] X. Li, T. F. Tedder, *Genomics* **1999**, *55*, 345-347.
[245] A. de Waard, S. Hickman, S. Kornfeld, *J. Biol. Chem.* **1976**, *251*, 7581-7587.
[246] J. Hirabayashi, K.-i. Kasai, *Glycobiology* **1993**, *3*, 297-304.
[247] L. Chiariotti, P. Salvatore, R. Frunzio, C. B. Bruni, *Glycoconj. J.* **2002**, *19*, 441-449.
[248] L. Johannes, R. Jacob, H. Leffler, *J. Cell Sci.* **2018**, *131*.
[249] A. A. Klyosov, P. G. Traber, in *Galectins and Disease Implications for Targeted Therapeutics, Vol. 1115*, American Chemical Society, **2012**, pp. 3-43.
[250] D. Compagno et al., *Biomolecules* **2020**, *10*, 750.
[251] F.-T. Liu, G. A. Rabinovich, *Ann. N.Y. Acad. Sci.* **2010**, *1183*, 158-182.
[252] S. Di Lella et al., *Biochemistry* **2011**, *50*, 7842-7857.
[253] S. Thiemann, L. G. Baum, *Annu. Rev. Immunol.* **2016**, *34*, 243-264.
[254] F.-T. Liu, G. A. Rabinovich, *Nat. Rev. Cancer* **2005**, *5*, 29-41.
[255] G. A. Rabinovich, M. A. Toscano, S. S. Jackson, G. R. Vasta, *Curr. Opin. Struct. Biol.* **2007**, *17*, 513-520.
[256] R.-Y. Yang, G. A. Rabinovich, F.-T. Liu, *Expert Rev. Mol. Med.* **2008**, *10*, e17.
[257] J. Hirabayashi et al., *Biochimica et Biophysica Acta (BBA) - General Subjects* **2002**, *1572*, 232-254.
[258] S. K. Patnaik et al., *Glycobiology* **2006**, *16*, 305-317.
[259] J. Iwaki, J. Hirabayashi, *Trends. Glycosci. Glycotechnol.* **2018**, *30*, SE137-SE153.
[260] C.-F. Bian, Y. Zhang, H. Sun, D.-F. Li, D.-C. Wang, *PLoS One* **2011**, *6*, e25007.
[261] H. Ideo, T. Matsuzaka, T. Nonaka, A. Seko, K. Yamashita, *Int. J. Biol. Chem.* **2011**, *286*, 11346-11355.
[262] S. Carlsson et al., *Glycobiology* **2007**, *17*, 663-676.
[263] D. F. Smith, X. Song, R. D. Cummings, in *Methods Enzymol., Vol. 480* (Ed.: M. Fukuda), Academic Press, **2010**, pp. 417-444.
[264] Y. Zhuo, S. L. Bellis, *Int. J. Biol. Chem.* **2011**, *286*, 5935-5941.
[265] S. R. Stowell et al., *Int. J. Biol. Chem.* **2008**, *283*, 10109-10123.
[266] F. H. M. de Melo et al., *J. Histochem. Cytochem.* **2007**, *55*, 1015-1026.
[267] C. F. Brewer, *Glycoconj. J.* **2002**, *19*, 459-465.
[268] H. Ideo, A. Seko, I. Ishizuka, K. Yamashita, *Glycobiology* **2003**, *13*, 713-723.
[269] H. Barrow et al., *Clin. Cancer. Res.* **2011**, *17*, 7035.
[270] F.-C. Chou, H.-Y. Chen, C.-C. Kuo, H.-K. Sytwu, *Int. J. Mol. Sci.* **2018**, *19*, 430.
[271] L. Song, J.-w. Tang, L. Owusu, M.-Z. Sun, J. Wu, J. Zhang, *Clin. Chim. Acta* **2014**, *431*, 185-191.
[272] M. Farhad, A. S. Rolig, W. L. Redmond, *Oncoimmunology* **2018**, *7*, e1434467-e1434467.
[273] A. Girard, J. L. Magnani, *Trends. Glycosci. Glycotechnol.* **2018**, *30*, SE211-SE220.
[274] T. Piyush, A. R. Chacko, P. Sindrewicz, J. Hilkens, J. M. Rhodes, L.-G. Yu, *Cell Death Differ.* **2017**, *24*, 1937-1947.

[275] S. Song et al., *Gastroenterology* **2005**, *129*, 1581-1591.
[276] H. Yu et al., *J. Am. Chem. Soc.* **2005**, *127*, 17618-17619.
[277] Y. C. Lee et al., *J. Biol. Chem.* **1994**, *269*, 10028-10033.
[278] H. Yu, S. Huang, H. Chokhawala, M. Sun, H. Zheng, X. Chen, *Angew. Chem. Int. Ed.* **2006**, *45*, 3938-3944.
[279] H. Sung et al., *CA Cancer J. Clin.* **2021**, *71*, 209-249.
[280] P. Romero et al., *Sci. Transl. Med.* **2016**, *8*, 334ps339.
[281] S. H. van der Burg, R. Arens, F. Ossendorp, T. van Hall, C. J. M. Melief, *Nat. Rev. Cancer* **2016**, *16*, 219-233.
[282] S. Nath, P. Mukherjee, *Trends Mol. Med.* **2014**, *20*, 332-342.
[283] J. Taylor-Papadimitriou, J. M. Burchell, R. Graham, R. Beatson, *Biochem. Soc. Trans.* **2018**, *46*, 659-668.
[284] M. M. Soares, V. Mehta, O. J. Finn, *J. Immunol. Res.* **2001**, *166*, 6555.
[285] G. A. Rabinovich, Y. van Kooyk, B. A. Cobb, *Ann. N.Y. Acad. Sci.* **2012**, *1253*, 1-15.
[286] S. O. Ryan, M. S. Turner, J. Gariépy, O. J. Finn, *Cancer Res.* **2010**, *70*, 5788.
[287] T. Ju, R. P. Aryal, C. J. Stowell, R. D. Cummings, *J. Cell. Biol.* **2008**, *182*, 531-542.
[288] I. Brockhausen, *Biochimica et Biophysica Acta (BBA) - General Subjects* **1999**, *1473*, 67-95.
[289] N. Gaidzik, U. Westerlind, H. Kunz, *Chem. Soc. Rev.* **2013**, *42*, 4421-4442.
[290] D. M. McDonald, S. N. Byrne, R. J. Payne, *Frontiers in Chemistry* **2015**, *3*.
[291] R. M. Wilson, S. J. Danishefsky, *J. Am. Chem. Soc.* **2013**, *135*, 14462-14472.
[292] U. Westerlind, H. Kunz, *CHIMIA* **2011**, *65*, 30-34.
[293] D. M. Coltart et al., *J. Am. Chem. Soc.* **2002**, *124*, 9833-9844.
[294] D. Zhou, L. Xu, W. Huang, T. Tonn, *Molecules (Basel, Switzerland)* **2018**, *23*, 1326.
[295] M. R. Price et al., *Tumour Biol.* **1998**, *19*, 1-20.
[296] S. Julien et al., *Br. J. Cancer* **2009**, *100*, 1746-1754.
[297] J. Zhu, J. D. Warren, S. J. Danishefsky, *Expert Rev. Vaccines* **2009**, *8*, 1399-1413.
[298] A. Kaiser, N. Gaidzik, U. Westerlind, D. Kowalczyk, A. Hobel, E. Schmitt, H. Kunz, *Angew. Chem. Int. Ed.* **2009**, *48*, 7551-7555.
[299] R. Lo-Man et al., *Cancer Res.* **2004**, *64*, 4987.
[300] R. P. Brinãs et al., Jr., *Bioconjugate Chem.* **2012**, *23*, 1513-1523.
[301] H. Cai et al., *Biorg. Med. Chem.* **2016**, *24*, 1132-1135.
[302] Z. Yin et al., *ACS Chem. Biol.* **2013**, *8*, 1253-1262.
[303] F.-G. Hanisch, T. R. E. Stadie, F. Deutzmann, J. Peter-Katalinic, *Eur. J. Biochem.* **1996**, *236*, 318-327.
[304] S. J. Storr et al., *Glycobiology* **2008**, *18*, 456-462.
[305] A. Borgert et al., *ACS Chem. Biol.* **2012**, *7*, 1031-1039.
[306] H. Yu et al., *Angewandte Chemie International Edition* **2006**, *45*, 3938-3944.

Appendix

General synthesis methods

If not otherwise stated, solvent were purchased in the quality *pro analysis* (p.a.) and used without further purification. Dichloromethane and acetonitrile were dried by reflux over calcium hydride under inert atmosphere and subsequent distillation. Molecular sieve (4 Å, Acros Organics) was activated by heating (>300 °C) for 2 h *in vacuo*. Thin layer chromatography (TCL) was performed using aluminium plates coated with silica (Kieselgel 60 F254, Merck KGaA, Darmstadt) with detection by A) treatment with 5 % sulfuric acid in ethanol and subsequent heat exposure, B) UV light (254 nm). Compounds were purified by flash column chromatography on silica (Silica gel 40 – 63 µm, *VWR*).
High-resolution ESI-spectra were measured with a 6230 TOF LC/MS spectrometer (*Agilent Technologies*) or a *LTQ Orbitrap XL* mass spectrometer (*Thermo Fisher*). NMR spectra were measured on *Avance 400* (Bruker), *Avance 600* (Bruker) and *Ascend 850* (Bruker) 295 K. The reported values for the chemical shifts δ (ppm) were calibrated to the residual proton or carbon resonance signal of the deuterated solvent, relatively correlated to the corresponding tetramethylsilane signal. Signal multiplicity is assigned as follows: S = singlet, d = duplet, t = triplet, q = quartett, m = multiplet, br = broad. Elucidation of the 1H and 13C spectra was performed by gCOSY, TOCSY, gHSQC and gHMBC correlation experiments. Specific rotations ([α]D20) were recorded on a Autopol IV polarimeter (Rudolph Research Analytical).

Fmoc-SPPS protocol of mucin glycopeptides

Glycopeptides were synthesized according to our reported Fmoc-SPPS protocol for glycopeptide synthesis.[183-184] SPPS of was carried out on a *Syro I* peptide synthesizer (*Multisyntech GmbH*). Glycopeptides were synthesized starting from pre-loaded Tentagel R Fmoc-AA-Trt resins (*Rapp Polymere*, Tübingen) on a 13 µmol scale. Coupling of the standard amino acids (8.0 eq, *Novabiochem*®, *Merck KGaA* or *Merck Schuchardt OHG*) was carried out by using HBTU/HOBt/DIPEA (0.95/0.95/2 equiv with respect to amino acid) in DMF for 40 min. The glycosylated amino acid building blocks (1.5 equiv) were pre-activated with HATU/HOAt/DIPEA (0.95/0.95/2 equiv with respect to amino acid) in DMF and coupled manually for 8 h. Fmoc was removed by treatment with 20 % piperidine in DMF. After full glycopeptide assembly, a triethyleneglycol spacer (TEG, 3 equiv) was coupled to the *N*-terminus using HBTU/HOBt/DIPEA (0.95/0.95/2 equiv with respect to spacer) in DMF with 2 h coupling time, followed by Fmoc deprotection. Release of the glycopeptides from the resin with simultaneous removal of the amino acid side-chain protecting groups was performed using TFA/TIPS/H_2O (95:5:5). The crude glycopeptides were desalted on a C-18 cartridge followed by removal of the *O*-acetyl protecting groups on the glycans by treatment with catalytic amounts of NaOMe in MeOH at pH 9.0-9.5 or with 0.2 M NaOH in MeOH/H_2O (pH ~10.0). Finally, the deprotected glycopeptides were purified by preparative HPLC.

General protocols for microarray fabrication and assays

Microarray fabrication

Glycopeptides were printed in a concentration of 50 µM in printing buffer (150 mM NaH_2PO_4/Na_2HPO_4, pH 8.5) in replicates on NHS-activated hydrogel slides (Nexterion® slide H, Schott, Mainz, Germany) using a non-contact piezoelectric spotting device (iONE, M2 automation, Berlin, Germany). Unreacted NHS-groups were blocked at room temperature for 1 h with 25 mM ethanolamine in 100 mM sodium tetraborate buffer (pH 9.0).

Microarray assays

Reagents:

- Incubation-buffer: PBST-buffer (0.2% Tween-20): 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4, 0.2% Tween-20.
- Wash-buffer 1: PBST-buffer (0.05%)
- Wash-buffer 2: PBS-buffer

Incubations of microarray slides were performed at 100 µL/well. The slides were incubated with samples diluted in incubation-buffer for 1 h at room temperature and 70% RH and then washed two times wash buffer 1 and once with wash buffer 2. Subsequently, the slides were incubated with secondary antibodies or Cy5-Streptavidin diluted in incubation-buffer for 1 h at room temperature and 70% RH. After a final washing step, the slides were rinsed with water and spin-dried. Finally, the microarray slides were scanned at wavelegths suitable for the respective fluorophores using a *GenePix 4300A* (Molecular Devices Corporation, Sunnyvale, CA, USA) microarray fluorescence scanner.

Enzymatic glycosylation on mucin glycopeptides

The following enzymes were employed in enzymatic modifications on the mucin glycopeptides:

1. Fucosylation
    - For α-1,3/4-fucosylation on GlcNAc in type-1 and -2 LacNAc containing peptides: α-1,3/4-(*O*)-fucosyltransferase from *Helicobacter pylori*, recombinant (*Escherichia coli*), Chemily, E.C. Number: 2.4.1.65, EN01024.
    - For α-1,2-fucosylation on terminal Gal in type-1 and -2 LacNAc containing peptides: α-1,2-(*O*)-fucosyltransferase from *Helicobacter mustelae*, recombinant (*Escherichia coli*), Chemily, E.C. Number: 2.4.1.67, EN01023
2. For sialylation:
    - For α-2,3-sialylation: PmST1 for α-2,3-sialylation of type-1 and -2 LacNAc containing peptides: α-2,3-(O)-Sialyltransferase 1 from *Pasteurella multocida*, recombinant (*Escherichia coli* BL21(DE3)), *Sigma Aldrich*.[276]
    - For α-2,6-sialylation of LacNAc structures: Pd2,6ST: α-2,6-(*O*)-Sialyltransferase from *Photobacterium damsela*, recombinant (*Escherichia coli* BL21(DE3)), *Sigma Aldrich,* catalogue number: S2076.*[306]*

General procedure for enzymatic α1,3/4-fucosylation of mucin glycopeptides

The MUC1 glycopeptide (1.0 eq) and guanosine 5′-diphospho-β-L-fucose (GDP-Fuc, 2.0 eq per glycosylation site) were dissolved in 50 mM tris(hydroxymethyl)aminomethane hydrochloride buffer (Tris·HCl, pH 8.0, 10 mM $MgCl_2$), followed by addition of α1,3/4-(*O*)-fucosyltransferase and alkaline phosphatase (2.5 U), resulting in a final peptide substrate concentration higher than 8 mM and an enzyme concentration higher than 1 U/mL. The reaction was kept at 30-37°C until full conversion of the peptide substrate was verified by MALDI-TOF or LC-MS. The reaction was neutralized by addition of an equal volume of 40% MeCN and lyophilized. After desalting on C18 cartridges (SPE, *Agilent Spec* 3 mL C18AR 15 mg) and purification by semipreparative HPLC-MS, the pure fucosylated glycopeptides could be obtained.

General procedure for enzymatic sialylation of mucin glycopeptides
The MUC1 glycopeptide (1.0 eq) and citidin-5´-monophospho-*N*-actylneuraminc acid (CMP-Neu5Ac, 2.5 eq per glycosylation site) were dissolved in corresponding reaction buffers (Rat 2,3-OST: 100 mM sodium cacodylate, pH 6.0; PmST1 and Pd2,6ST: 100 mM tris(hydroxymethyl)aminomethane hydrochloride (TRIS·HCl), pH 8.5). Then, sialyltransferase (Rat 2,3-OST: > 0.4 U/mL; PmST1: > 1 U/mL; Pd2,6ST: > 0.4 U/mL) and alkaline phosphatase (2.5 U) were added, resulting in a final peptide substrate concentration higher than 8 mM. The reaction was kept at 30^{o}C until full conversion of the peptide substrate was verified by analytical HPLC-MS. The reaction was neutralized by addition of an equal volume of cold acetonitrile (20% final concentration) and lyophilized. After desalting on C18 cartridges (SPE, *Agilent Spec* 3 mL C18AR 15 mg) and purification by semipreparative HPLC-MS, the pure sialylated glycopeptides could be obtained

Chapter 3

Table S1. List of α-GalNAc-, α-GlcNAc- and β-GlcNAc-*O*-Thr/Ser/Tyr glycopeptides and unglycosylated peptides.

Peptide ID number	Sequence	Glycan
P1	Ac-GYYA-TEG	-
P2	Ac-G**Y*Y***A-TEG	α-GalNAc
P3	Ac-G**Y*Y***A-TEG	β-GlcNAc
P4	Ac-G**Y*Y***A-TEG	α-GlcNAc
P5	TEG-PAHGVTSAPDTRPAPGSTA	-
P6	TEG-PAHGV**T***SAPDTRPAPGSTA	α-GalNAc
P7	TEG-PAHGV**T***SAPDTRPAPGSTA	β-GlcNAc
P8	TEG-PAHGVT**S***APDTRPAPGSTA	β-GlcNAc
P9	TEG-PAHGV**Y***SAPDTRPAPGSTA	α-GalNAc
P10	TEG-PAHGV**Y***SAPDTRPAPGSTA	β-GlcNAc
P11	TEG-PAHGV**Y***SAPDTRPAPGSTA	α-GlcNAc
P12	TEG-PAHGVTSAPD**T***RPAPGSTA	α-GalNAc
P13	TEG-PAHGVTSAPD**Y***RPAPGSTA	α-GalNAc
P14	TEG-PAHGVTSAPD**S***RPAPGSTA	β-GlcNAc
P15	TEG-PAHGVTSAPD**T***RPAPGSTAPPA	β-GlcNAc
P16	TEG-PAHGVTSAPD**Y***RPAPGSTA	β-GlcNAc
P17	TEG-PAHGVTSAPD**Y***RPAPGS TA	α-GlcNAc
P18	TEG-PAHGVTSAPDTRPAPG**S***TA	β-GlcNAc
P19	TEG-PAHGVTSAPDTRPAPGS**T***A	α-GalNAc
P20	TEG-PAHGVTSAPDTRPAPGS**Y***A	α-GalNAc
P21	TEG-PAHGVTSAPDTRPAPGS**T***A	β-GlcNAc
P22	TEG-PAHGVTSAPDTRPAPGS**Y***A	β-GlcNAc
P23	TEG-PAHGVTSAPDTRPAPGS**Y***A	α-GlcNAc
P24	TEG-DAEFRHDSGYEVHHQK	-
P25	TEG-DAEFRHDSG**Y***EVHHQK	α-GalNAc
P26	TEG-DAEFRHDSG**Y***EVHHQK	β-GlcNAc
P27	TEG-DAEFRHDSG**Y***EVHHQK	α-GlcNAc
P28	TEG-QFPEVYVPTVFE	-
P29	TEG-QFPEV**Y***VPTVFE	α-GalNAc
P30	TEG-QFPEV**Y***VPTVFE	β-GlcNAc
P31	TEG-QFPEV**Y***VPTVFE	α-GlcNAc
P32	TEG-AFPGEYIPTVFD	-
P33	TEG-AFPGE**Y***IPTVFD	α-GalNAc
P34	TEG-AFPGE**Y***IPTVFD	β-GlcNAc
P35	TEG-AFPGE**Y***IPTVFD	α-GlcNAc
P36	TEG-KFPSEYVPTVFD	-
P37	TEG-KFPSE**Y***VPTVFD	α-GalNAc
P38	TEG-KFPSE**Y***VPTVFD	β-GlcNAc
P39	TEG-KFPSE**Y***VPTVFD	α-GlcNAc
P40	TEG-LEYHQVIQQMEQK	-
P41	TEG-LE**Y***HQVIQQMEQK	α-GalNAc
P42	TEG-LE**Y***HQVIQQMEQK	β-GlcNAc
P43	TEG-IMDPNIVGSEHYDVAR	-
P44	TEG-IMDPNIVGNEHYDVAR	-

P45	TEG-IMDPNIVGNEHY*DVAR	α-GalNAc
P46	TEG-IMDPNIVGNEHY*DVAR	β-GlcNAc
P47	TEG-IMDPNIVGNEHY*DVAR	α-GlcNAc
P48	TEG-IMDPNIVGSEHY*DVAR	β-GlcNAc
P49	TEG-IMDPNIVGS*EHYDVAR	β-GlcNAc
P50	TEG-IMDPNIVGS*EHY*DVAR	β-GlcNAc/ β-GlcNAc
P51	TEG-AHGGYSVFAGVGER	-
P52	TEG-AHGGYS*VFAGVGER	β-GlcNAc
P53	TEG-AHGGY*SVFAGVGER	β-GlcNAc
P54	TEG-AHGGY*S*VFAGVGER	β-GlcNAc/β-GlcNAc
P55	TEG-FTQAGSEVSALLGR	-
P56	TEG-FT*QAGSEVSALLGR	β-GlcNAc
P57	TEG-FTQAGS*EVSALLGR	β-GlcNAc
P58	TEG-FTQAGSEVS*ALLGR	β-GlcNAc
P59	TEG-FT*QAGS*EVSALLGR	β-GlcNAc/β-GlcNAc
P60	TEG-FTQAGS*EVS*ALLGR	β-GlcNAc/β-GlcNAc
P61	TEG-FVTVQTIS*GTGALR	β-GlcNAc
P62	TEG-NLDKEYLPIGGLAEFCK	-
P63	TEG-NLDKEY*LPIGGLAEFCK	α-GalNAc
P64	TEG-NLDKEY*LPIGGLAEFCK	β-GlcNAc
P65	TEG-NLDKEY*LPIGGLAEFCK	α-GlcNAc
P66	TEG-IAATILTSPDLR	-
P67	TEG-IAAT*ILTSPDLR	β-GlcNAc
P68	TEG-IAATILT*SPDLR	β-GlcNAc
P69	TEG-IAATILTS*PDLR	β-GlcNAc
P70	TEG-IAAT*ILT*SPDLR	β-GlcNAc/β-GlcNAc
P71	TEG-IAAT*ILTS*PDLR	β-GlcNAc/β-GlcNAc
P72	TEG-IAATILT*S*PDLR	β-GlcNAc/β-GlcNAc
P73	TEG-EAYPGDVFYLHSR	-
P74	TEG-EAY*PGDVFYLHSR	β-GlcNAc
P75	TEG-EAYPGDVFY*LHSR	β-GlcNAc
P76	TEG-EAYPGDVFYLHS*R	β-GlcNAc
P77	TEG-EAY*PGDVFY*LHSR	β-GlcNAc/β-GlcNAc
P78	TEG-EAYPGDVFY*LHS*R	β-GlcNAc/β-GlcNAc
P79	TEG-SEDYALPSTVDRR	-
P80	TEG-S*EDYALPSTVDRR	β-GlcNAc
P81	TEG-SEDY*ALPSTVDRR	β-GlcNAc
P82	TEG-SEDYALPS*TVDRR	β-GlcNAc
P83	TEG-SEDYALPSY*VDRR	β-GlcNAc
P84	TEG-S*EDY*ALPSTVDRR	β-GlcNAc/β-GlcNAc
P85	TEG-SEDY*ALPS*TVDRR	β-GlcNAc/β-GlcNAc
P86	TEG-SEDYALPS*Y*VDRR	β-GlcNAc/β-GlcNAc