



# **LEctPROFILE kits : towards quality control and new potential applications**

Vittoria Federica Vena

## **► To cite this version:**

Vittoria Federica Vena. LEctPROFILE kits : towards quality control and new potential applications. Biochemistry, Molecular Biology. Université d'Orléans, 2022. English. NNT : 2022ORLE1060 . tel-04310082

**HAL Id: tel-04310082**

**<https://theses.hal.science/tel-04310082>**

Submitted on 27 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# UNIVERSITÉ D'ORLÉANS

*ÉCOLE DOCTORALE SANTE, SCIENCES BIOLOGIQUES ET CHIMIE DU VIVANT*

**ICOA UMR7311 - GLYcoDiag**

## THÈSE présentée par : **Vittoria Federica VENA**

soutenue le : 13/12/2022

pour obtenir le grade de : **Docteur de l'Université d'Orléans**

Discipline/ Spécialité : Biologie - Biochimie

**LectPROFILE kits: towards quality control and new potential applications**

Kits LECTPROFILE : vers un contrôle de qualité et de nouvelles applications potentielles

**THÈSE dirigée par :**

**Ludovic Landemarre** Directeur GLYcoDiag

**RAPPORTEURS :**

**Annabelle Varrot** Directrice de recherche, CERMAV- CNRS, Grenoble

**Ján Tkac** Directeur de recherche, Slovak Academy of Sciences, Bratislava

**JURY :**

**Chrystel Lopin-Bon** Professeure, ICOA, **Présidente du jury**

**Annabelle Varrot** Directrice de recherche, CERMAV- CNRS

**Ján Tkac** Directeur de recherche, Slovak Academy of Sciences

**Julie Bouckaert** Chargée de recherche, CNRS, Lille

**Richard Daniellou** Professeur, AgroParisTech

**Ludovic Landemarre** Directeur GLYcoDiag – Directeur de Thèse

## Acknowledgments

*First of all, I sincerely thank Dr. Annabelle Varrot and Dr. Jan Tkac for accepting to review my manuscript, as well as Prof. Chrystel Lopin-Bon and Dr. Julie Buckaert for accepting to be part of my thesis defence committee.*

*My gratitude to Dr. Ludovic Landemarre for giving me opportunity to be part of his GLYcoDiag team and to be part of the synBIOcarb project. Thank you for guiding me through my research but also for showing me how is it to work in a scientific company and for finding always the time to answer my questions.*

*A special thanks goes to Dr. Benoît Roubinet for his support, advices, patience, and for sharing his experience as an organic chemist. It was a pleasure working with you.*

*Thanks to my former voisin de bureau and friend Yasmina Traore, I am really happy to have gotten to know you (and petite Cassandre!!) during these three years and to have shared a lot of snacks.*

*Thank to Dr. Mateja Senicar, who is a great colleague and an amazing human being. I miss already our lunch break in the sun (or not) and all the adventures, conversations, laughs (or not), we have shared. I am sure there will be many more. Thank you for supporting me and sharing your knowledge.*

*Thanks to Prof. Richard Daniellou and the amazing Glycobiochimie team Dr. Pierre Lafite, Arnaud Paris, Damien Bretagne, Nastassja Burrini. You have truly helped me during the course of my thesis and you've always been available to answer my questions and to help me solve my doubts.*

*I would love to thank all synBIOcarb PIs and PhD students: unfortunately we did not have many occasions to get together in person, but when we did it was really great. Thanks to Dr. Simona Notova and Juvissan Aguedo with whom I collaborated and as well a thank you to Dr. Anne Imberty, who welcomed me in her laboratory, and Dr. Jan Tcak that agreed for the collaboration with Juvi.*

*PhD is a great life experience but in order to be awesome, you have to have the right people around you. I was lucky enough to be surrounded by very good friends who turned into a family and with whom we had the greatest fun in the petite Orléans:*

*The itaño! gang: mi tocaya e moglie Fede (e Gateau ovviamente) grazie per essere stata al mio fianco, come ci insegnano, nella gioia e nel dolore. Fe, Moni, Cla, Elenita: grazie per la compagnia, la spesa, le chiacchierate, le mangiate (special mention to carbonara), viaggi,*

*bevute, passeggiate, risate, feste. Siete stati come una "home away from home" e un grande supporto nei momenti di sconforto.*

*El grupo de los latinos: Hectorcito y los colombianos, Fabiancito, Alfredo y los mexicanos. Gracias por la música (no solo reguetón jaja), las fiestas, el baile, el chisme y el desorden que han traído en nuestra vida. Orléans es un lugar más feliz gracias a ustedes.*

*Los hijos de Orléans (or should I say "Los hijos de les Enfants Terribles"): you all know who you are. I will never forget Rue du Grenier à Sel. A thank you to my dear friend Mahdi, who is always there for me.*

*The ADSO crew: it was a honour to be vice-president and I know I left the organization in very good hands. Thanks Pierre-Yves for the good memories, after all, we met for the first time at ADSO!*

*Simo, perché dice te l'avevo detto ma poi alla fine mi aiuta sempre.*

*Federico, che mi ha accolta a Manchester quando avevo bisogno e ancora adesso riesce sempre ad essere presente al momento giusto.*

*Giulia e Carme con cui abbiamo scattato le migliori foto da trentenni a Parigi: queste sono solo le ultime di una lunga serie di momenti passati insieme e spero le prime di altri che verranno.*

*Bice, che è venuta fino ad Orléans con me per assicurarsi che sopravvivessi.*

*Vale, che ovunque io sia riesce sempre a trovare qualche giorno per venire a trovarmi.*

*Paola, Roberto e Gianluca che sono la mia famiglia della porta accanto.*

*Zia Pina e Zio Franco che da molti km di distanza riescono sempre ad esserci col loro affetto.*

*Mia madre e mio padre a cui è piaciuta molto Orléans, mi dispiace adesso vi toccherà venire fino a Lille. Grazie di supportare sempre le mie scelte vagabonde.*

*Mia sorella Berry, che, insieme a Sirius, è la mia persona preferita.*

*E infine Andrea: non ci sono abbastanza parole per ringraziarti, in tutte le lingue che conosciamo. Grazie per i km fatti e per avere sempre creduto in me e soprattutto in noi.*



## Abbreviations

ADCP	Antibody-dependent cellular phagocytosis
Ag	Antigen
Ab	Antibody
$\alpha 1,3$ GT	$\alpha 1,3$ galactosyl-transferase
ADCC	Antibody-Dependent Cell-mediated Cytotoxicity
APC	Antigen-Presenting Cells
BC50	half of the maximum ligand concentration
Bmax	maximum number of lectin binding sites
BSA	Bovine Serum Albumin
CDC	complement-dependent cytotoxicity
CE	Capillary Electrophoresis
CHO	Chinese Hamster Ovary Cells
CMAH	Cytidine monophosphate-N-acetylneuraminic acid hydroxylase
CMP	Cytidine monophosphate
CL	Constant region of the light chain
CH1,2,3	Constant region of the heavy chain, domain 1, 2, 3
CQA	Critical Quality Attribute
CRD	Carbohydrate-Recognition Domain
CTD	carbohydrate-deficient Transferrin
CuAAC	copper-catalysed azide-alkyne cycloaddition
EDC	1-ethyl-3-(3- (dimethylamino)propyl)carbodiimide
EGFR	Epidermal Growth Factor Receptor

ESI/MS	ElectroSpray Ionisation/ Mass Spectrometry
Fab	fragment antigen-binding
Fc	fragment crystallizable
FDA	Food and Drug Administration
FCS	Fetal Calf Serum
FcγR	Fcγ receptor
Galα1-3Gal	Galili antigen, xenotransplantation epitope
GCTA1	α1,3galactosyl-transferase gene
HA	Hemagglutinin
HD-antibodies/HD-Abs	Hanganutziu Deicher- antibodies
HPLC	High Performance Liquid Chromatography
MAAI	<i>Maackia amurensis leukagglutinin</i>
MOA	<i>Marasmius oreades agglutinin</i>
MOAβT	<i>Marasmius oreades β trefoil</i>
mAbs	monoclonal antibodies
HOBt	1-hydroxy- benzotriazole
IC50	half-maximum inhibitory concentration
IgG	Immunoglobulin
ITC	Isothermal Titration Calorimetry
ITN	Innovative Training Network
IVIG	Intravenous Immunoglobulin Therapy
MALDI-TOF/ MS	Matrix Assisted Laser Desorption ionisation Ionisation- Time Of Flight/ Mass Spectrometry
MOA	<i>Marasmius oreades agglutinin</i>
MOAβT	<i>Marasmius oreades agglutinin</i> beta trefoil
Neu5Ac	acetyl neuraminic acid

NANA	acetyl neuraminic acid
Neu5Gc	glycolyl neuraminic acid
NeoGalili	Neoglycoprotein Galili
NMR	Nuclear magnetic resonance
NeoGP	Neoglycoprotein
NeoF	Neoglycoprotein Fucose
NeoM	Neoglycoprotein Mannose
NeoL	Neoglycoprotein Lactose
NeoNeuAc	Neoglycoprotein NeuAc
NeoNeuGc	Neoglycoprotein NeuGc
Neu $\alpha$ 2-3	$\alpha$ 2-3 neuraminidase
NeuT	Total neuraminidase
RIP	Relative Inhibitory Potency
PCa	Prostate cancer
PNGase F	Peptide -N-Glycosidase F
PSA	Prostate Specific Antigen
QbT	Quality by Testing
QbD	Quality by Design
SA	sialidase
Sia	Sialic acid
SPR	Surface Plasmon Resonance
ST	sialyltransferase
TAA	Tumour Associated antigen
VL	variable region of the light chain

VH	variable region of the heavy chain

## Table of figures

### Chapter 1 – Introduction and thesis objective

Fig. 1. 1 synBIOcarb network. On the left, highlighted in yellow all countries where the beneficiaries and partners are located. On the right, the logos of all beneficiaries and partners involved in the network.	2
Fig. 1. 2 synBIOcarb work packages	3
Fig. 1. 3 Thesis workflow.	5
Fig. 1. 4 The non-template glycosylation drastically increases biological functional diversity, adapted from [5].	9
Fig. 1. 5 Classification of lectin based on their CRDs and their presence across the kingdoms of life adapted from [4]	10
Fig. 1. 6 Lectin array design, adapted from [13].	12
Fig. 1. 7 Glycan array principle, adapted from [20]	14
Fig. 1. 8 SPR principle adapted from <a href="https://bit.ly/2Es3k2s">https://bit.ly/2Es3k2s</a> .	15
Fig. 1. 9 ITC principle adapted from <a href="https://bit.ly/2rP1OAL">https://bit.ly/2rP1OAL</a>	15
Fig. 1. 10 Principle of HPLC, adapted from <a href="https://bit.ly/3UWsiMH">https://bit.ly/3UWsiMH</a>	16
Fig. 1. 11 Schematic representation of CE system, adapted from <a href="https://bit.ly/3SHSR64">https://bit.ly/3SHSR64</a>	17
Fig. 1. 12 Derivatization of sialic acid linkages. Adapted from the poster "Rapid and high-throughput methods for discrimination of sialic acids linkages in glycoproteins" J. Aguedo and F. Vena.	18
Fig. 1. 13 MALDI-TOF/MS for the discrimination of $\alpha$ 2,3 and $\alpha$ 2,6 sialic acid linkages workflow. Adapted from the poster "Rapid and high-throughput methods for discrimination of sialic acids linkages in glycoproteins" J. Aguedo and F. Vena.	18
Fig. 1. 14 A 2020 best-selling drugs, adapted from [25]. B Biotherapeutics global revenues from 2005 to 2006 adapted from [26].	19
Fig. 1. 15 Different glycosylation according to the kind of expression system.	23
Fig. 1. 16 A GLYcoPROFILE in direct binding. B GLYcoPROFILE in inhibition.	27

### Chapter 2 – Neoglycoproteins

Fig. 2. 1 A structure of the neoRSLVI in complex with aMethylFucose (spheres), all binding sites result occupied, adapted from [51]. B Cartoon representation of a Janus lectin, able to cross-link biological surfaces thanks to its double specificity for fucosylated glycans on one side and sialylated ones on the other side, adapted from [52].	31
Fig. 2. 2 Mechanisms of interaction of multivalent partners, adapted from [8]	32
Fig. 2. 3 Types of multivalent synthetic glycoconjugates, adapted from [8]	33
Fig. 2. 4 A CuAAC via 1,3-dipolar cycloaddition of azides and alkynes reaction resulting in 1,2,3-triazoles, adapted from [63]. B CuAAC reaction mechanism (adapted from, <a href="http://bitly.ws/uSiZ">http://bitly.ws/uSiZ</a> )	36
Fig. 2. 5 A Reaction between the primary amines of the BSA and PNHS ester, followed by the click-chemistry reaction (CuAAC) between BSA-alkyne and the glycan functionalized with an azide group. B Reaction between the primary amines of the BSA and the NHS, followed by the click-chemistry reaction (CuAAC) between BSA-azide and the glycan functionalized with an alkyne group.	38
Fig. 2. 6 SDS-PAGE of BSA functionalization with alkyne and azide groups. The protein samples were analysed under denaturing conditions on 14% polyacrylamide gel. Line 1: Protein marker, Line 2: BSA, Line 3: BSA-alkyne, Line 4: BSA-azide.	39
Fig. 2. 7 A MALD-TOF/MS spectrum of BSA B MALD-TOF/MS spectrum of BSA-alkyne C MALD-TOF spectrum of BSA-azide.	40
Fig. 2. 8 CuAAC reaction between BSA-azide and Gal $\alpha$ 1-3Gal-N-Acetyl-Propargyl.	40

Fig. 2. 9 SDS-PAGE of NeoGalili synthesis. The protein samples were analysed under denaturing conditions on 14% polyacrylamide gel. Line 1: Protein marker, Line 2: BSA, Line 3: BSA-alkyne, Line 4: BSA-azide, Line5: NeoGalili neoglycoprotein. _____	41
Fig. 2. 10 MALD-TOF spectrum of NeoGalili neoglycoprotein. _____	41
Fig. 2. 11 GLYcoPROFILE in direct binding of the NeoGalili neoglycoprotein labelled. The neoglycoprotein at three different concentrations was added to the plate coated with six different lectins. The data shown are the average of experiments performed in triplicate. _____	42
Fig. 2. 12 CuAAC reaction between BSA-alkyne and 6'-sialylactose-N3 or 3'-sialylactose-N3. _____	43
Fig. 2. 13 SDS-PAGE of Neo6'SL and Neo3'SL synthesis. The protein samples were analysed under denaturing conditions on 14% polyacrylamide gel. Line 1: Protein marker, Line 2: BSA-alkyne, Line 3: Neo6'SL, Line 4: Neo3'SL. _____	43
Fig. 2. 14 <b>A</b> MALD-TOF spectrum of Neo6'SL <b>B</b> MALD-TOF spectrum of Neo3'SL. _____	44
Fig. 2. 15 <b>A</b> GLYcoPROFILE in direct binding of the Neo6'SL neoglycoprotein labelled. The neoglycoprotein at three different concentrations was added to the plate coated with four different lectins. <b>B</b> GLYcoPROFILE in direct binding of the Neo3'SL neoglycoprotein labelled. The neoglycoprotein at three different concentrations was added to the plate coated with four different lectins. The data shown are the average of experiments performed in triplicate. _____	44
Fig. 2. 16 CuAAC reaction between BSA-azide and propargylNeu5Ac or propargylNeu5Gc. _____	45
Fig. 2. 17 SDS-PAGE of NeoNuAc and NeoNeuGc synthesis. The protein samples were analysed under denaturing conditions on 14% polyacrylamide gel. Line 1: Protein marker, Line 2: BSA, Line 3: BSA-azide, Line 4: NeoNuAc, Line 4: NeoNeuGc. _____	46
Fig. 2. 18 <b>A</b> MALD-TOF spectrum of NeoNeuAc <b>B</b> MALD-TOF spectrum of NeoNeuGc. _____	46
Fig. 2. 19 <b>A</b> GLYcoPROFILE in direct binding of the NeoNeuAc neoglycoprotein labelled. The neoglycoprotein at three different concentrations was added to the plate coated with seven different lectins. <b>B</b> GLYcoPROFILE in direct binding of the NeoNeuGc neoglycoprotein labelled. The neoglycoprotein at three different concentrations was added to the plate coated with seven different lectins. The data shown are the average of experiments performed in triplicate. _____	47

### Chapter 3 – The Galili antigen

Fig. 3. 1 <b>A</b> , the Galili antigen trisaccharide. <b>B</b> , the reaction catalysed by the $\alpha$ 1,3GT, adapted from [67]. _____	59
Fig. 3. 2 <b>A</b> , Cetuximab structure. It contains two glycosylation sites in the Asn 299 on the Fc portion and two at Asn88 in the Fab portion, adapted from [71]. <b>B</b> , the cetuximab mechanism: it binds to the EGFR and prevents its dimerization, blocking all its pro-tumour functions and promoting apoptosis, adapted from [70]. _____	59
Fig. 3. 3 Schematic representation of the distribution of hypersensitivity reactions to Cetuximab and red meat in the southeast region of North America and its connection to the increasing deer population carrying ticks and tick bites, adapted from [72]]. _____	61
Fig. 3. 4 Galsafe engineered pigs used for xenotrasplantation, adapted from [77]. _____	62
Fig. 3. 5 Schematic representation of the viral vaccine mechanism of action, adapted from [79]. _____	63
Fig. 3. 6 Representation of the $\alpha$ -Gal glycolipids and how they are recognized by anti-gal Abs when injected in tumour tissues, adapted from [66]. _____	64
Fig. 3. 7 MOA $\beta$ -trefoil portion crystal structure in complex with the Galili antigen. This figure was created with Pymol, adapted from the crystal structure of Grahn et al, ([85] PDB: 2IHO). _____	65
Fig. 3. 8 Crystal structure of GSLI-B <sub>4</sub> in complex with the xenograft antigen. PDB code 1HQL, adapted from [86]. _____	66
Fig. 3. 9 <b>A</b> SDS-PAGE analysis of MOA $\beta$ T The protein sample was analyzed under denaturing conditions on 14% polyacrylamide gel. <b>Line 1</b> : Protein marker, <b>Line 2</b> : MOA $\beta$ T (17,2 kDa). <b>B</b> GLYcoPROFILE in direct binding of MOA $\beta$ T interacting with labelled Porcine Thyroglobulin, Bovine Thyroglobulin, Invertase, NeoGalili and Neo $\beta$ Gal, NeoF. _____	67

Fig. 3. 10 GLYcoPROFILE in direct binding of GSLI-B <sub>4</sub> interacting with labelled Porcine Thyroglobulin, Bovine Thyroglobulin, Invertase, NeoGalili, NeoL, NeoF.	68
Fig. 3. 11 <b>A</b> Comparison of the binding of MOAβT at decreasing quantity/well to the NeoGalili neoglycoprotein at a fixed concentration (1.15 μM) on covalent plate (black) and adsorption plate (pink). The two experiments can be compared because the signal in both cases is measured in absorbance. <b>B</b> Binding of MOAβT at decreasing quantity/well to the NeoGalili neoglycoprotein at a fixed concentration (1.15 μM) on adsorption plate, this time the interaction is reported as fluorescence intensity.	69
Fig. 3. 12 <b>A</b> MOAβT decreasing quantity/well (4 μg/well to 0.12 μg/well) plotted against the absorbance at 450 nm resulting from the lectin interaction with the NeoGalili neoglycoprotein at a fixed concentration (0.3 μM) on covalent plate. <b>B</b> MOAβT decreasing quantity/well (1 μg/well to 0.1 μg/well) plotted against the absorbance at 450 nm resulting from the lectin interaction with the NeoGalili neoglycoprotein at a fixed concentration (0.3 μM) on covalent plate.	70
Fig. 3. 13 Comparison of MOAβT binding to NeoGalili neoglycoprotein when MOAβT is at a fixed quantity/well of either 1 μg/well or 0.5 μg/well with NeoGalili added in serial dilutions.	71
Fig. 3. 14 Direct binding curve of MOAβT immobilized on covalent plate at the fixed quantity of 1 μg/well with NeoGalili neoglycoprotein applied to the plate at 1.15 μM and serially diluted.	72
Fig. 3. 15 Inhibition experiment performed with MOAβT at 1 μg/well interacting with labelled NeoGalili at the fixed concentration of 0.5 μg/mL and with the non-labelled inhibitors applied in serial dilution, from a starting concentration of Porcine Thyroglobulin=12μM, NeoGalili=12 μM, Melibiose=50 mM.	73
Fig. 3. 16 Direct binding curve of MOAβT coated on adsorption plate at the fixed quantity of 1 μg/well with NeoGalili neoglycoprotein applied to the plate at 2.3 μM and serially diluted.	74
Fig. 3. 17 Direct binding curve of MOA coated on adsorption plate at the fixed quantity of 1 μg/well with NeoGalili neoglycoprotein applied to the plate at 2.3 μM and serially diluted.	75
Fig. 3. 18 Direct binding curve of GSLI-B <sub>4</sub> coated on adsorption plate at 1μg/well interacting with NeoGalili neoglycoprotein applied to the plate at 2.3 μM and serially diluted.	76
Fig. 3. 19 Inhibition experiment performed with GSLI-B <sub>4</sub> interacting with labelled NeoGalili at the fixed concentration of 35 μg/mL and with the non-labelled inhibitors applied in serial dilution, from a starting concentration of Porcine Thyroglobulin=12μM, NeoGalili=12 μM, Melibiose=100 mM, α-galactose=100 mM. The IC <sub>50</sub> for α-galactose and Melibiose could not be calculated because of the saturation of the signal.	77
Fig. 3. 20 <b>A</b> Infliximab GLYcoPROFILE in direct binding with a panel of lectins. <b>B</b> Rituximab GLYcoPROFILE in direct binding with a panel of lectins.	79
Fig. 3. 21 Cetuximab GLYcoPROFILE in direct binding with a panel of lectins.	80
Fig. 3. 22 MOAβT standard curve.	82
Fig. 3. 23 GSLI-B <sub>4</sub> standard curve.	83
Fig. 3. 24 HPyL standard curve.	83
Fig. 3. 25 SNA standard curve.	84
Fig. 3. 26 MAAI standard curve.	84
Fig. 3. 27 PSA standard curve.	85

## Chapter 4 – Sialic acid

Fig. 4. 1 Structure of sialic acid [90].	111
Fig. 4. 2 Asialoglycoprotein receptors. ASGP receptors in the liver do not recognize desialylated proteins, targeting them for uptake and degradation [91]	112
Fig. 4. 3 Roles of sialic acids. They can act as ligands for intrinsic receptors as well as extrinsic receptors[92].	112
Fig. 4. 4 N-Acetyl neuraminic acid. The red circle highlights the acetyl functional group on C <sub>5</sub>	113
Fig. 4. 5 Neu5Ac acid metabolism in mammalian cells [93].	114
Fig. 4. 6 Human Transferrin glycoforms [95]	115
Fig. 4. 7 Sialome organization [4].	116

Fig. 4. 8 IgG structure and its glycosylation. The bottom domain is the Fc and the top one is the Fab. V <sub>L</sub> = variable region of the light chain, C <sub>L</sub> = constant region of the light chain, V <sub>H</sub> = variable region of the heavy chain, C <sub>H1,2,3</sub> = Constant region of the heavy chain, domain 1, 2, 3, Ag= antigen [109].	119
Fig. 4. 9 Different glycan structures that can be found on IgG. G= galactose, F= fucose, S= sialic acid [105].	120
Fig. 4. 10 <b>A</b> N-Glycolyl neuraminic acid biosynthesis. <b>B</b> taxa are found in blue, not found in purple and purple and blue where only certain species have lost the capacity to make it [114]	122
Fig. 4. 11 MAA-I in complex with sialyllactose, PDB code 1DBN [64]	125
Fig. 4. 12 SNA in complex with N-Acetylgalactosamine, PDB code 3CA3 [65].	125
Fig. 4. 13 structure of VP1. PDB code 4POQ [115].	126
Fig. 4. 14 3D structure of WGA [116].	126
Fig. 4. 15 <b>A</b> SDS PAGE <b>analysis of MAAI</b> . The protein sample was analysed under denaturing conditions on 14% polyacrylamide gel. Line 1: Protein marker, Line 2: MAAI (35 kDa) <b>B</b> GLYcoPROFILE in direct binding of MAAI interacting with labelled Neo6'SL, Neo3'SL, NeoMa and NeoF at 40 µg/mL.	128
Fig. 4. 16 <b>A</b> SDS PAGE analysis of SNA. The protein sample was analysed under denaturing conditions on 14% polyacrylamide gel. Line 1: Protein marker, Line 2: SNA (33 kDa) <b>B</b> GLYcoPROFILE in direct binding of SNA interacting with labelled Neo6'SL, Neo3'SL and NeoMannose at 40 µg/mL.	128
Fig. 4. 17 GLYcoPROFILE in direct binding representing the interaction between the lectin SNA and the labelled neoglycoprotein Neo6'SL at a starting concentration of 0.6 µM. The curve is obtained from four independent replicates. All measurements were performed in duplicates.	130
Fig. 4. 18 GLYcoPROFILE in direct binding representing the interaction between the lectin MAAI and the labelled neoglycoprotein Neo3'SL at a starting concentration of 0.6 µM. The curve is obtained from four independent replicates. All measurements were performed in duplicates.	130
Fig. 4. 19 Inhibition curves of SNA with the inhibitors Neo6'SL, Fetuin, Transferrin at the starting concentration of 10 µM and 6'SL at the starting concentration of 5mM. The labelled tracer Neo6'SL is at the fixed concentration of 20 µg/mL. The curve is obtained from three independent replicates. All measurements were performed in duplicates.	132
Fig. 4. 20 Inhibition curves of SNA with the inhibitors Neo3'SL and Fetuin at the starting concentration of 10 µM and 3'SL at the starting concentration of 5mM. The labelled tracer Neo3'SL is at the fixed concentration of 20 µg/mL. The curve is obtained from three independent replicates. All measurements were performed in duplicates.	132
Fig. 4. 21 SNA and MAAI standard curves obtained with labelled Neo6'SL and Neo3'SL.	135
Fig. 4. 22 Interaction of MAAI and SNA specific lectins with labelled Fetuin at three different concentrations. The intensity of interaction of each lectin was studied before and after Fetuin digestion by total neuraminidase (NeuT) and α2-3 neuraminidase (Neuα2-3). Two independent assays (each with n = 3 replicates) were carried out.	136
Fig. 4. 23 MALDI-TOF/MS. <b>A</b> Possible sialylated structures present on Fetuin. <b>B</b> Relative abundance of α2-6 and α2-6+α2-3- sialic acid linkages present in sialylated glycans.	137
Fig. 4. 24 Interaction of MAAI and SNA specific lectins with labelled transferrin at three different concentrations. The intensity of interaction of each lectin was studied before and after transferrin digestion by total neuraminidase (NeuT) and α2-3 neuraminidase (Neuα2-3). Two independent assays (each with n = 3 replicates) were carried out.	138
Fig. 4. 25 Interaction of MAAI and SNA specific lectins with labelled transferrin native and denatured at three different concentrations. Two independent assays (each with n = 3 replicates) were carried out.	139
Fig. 4. 26 MALDI-TOF/MS <b>A</b> Possible sialylated structures present on Transferrin. <b>B</b> Relative abundance of α2-6 and α2-6+α2-3- sialic acid linkages present in sialylated glycans.	140
Fig. 4.27 <b>A</b> Interaction of MAAI and SNA specific lectins with labelled porcine thyroglobulin at three different concentrations. <b>B</b> Interaction of MAAI and SNA specific lectins with labelled bovine thyroglobulin at three different concentrations. The intensity of interaction of each lectin was studied before and after thyroglobulin digestion by total neuraminidase (NeuT) and α2-3 neuraminidase (Neuα2-3). For each protein two independent assays (each with n = 3 replicates) were carried out.	141
Fig. 4. 28 MALDI-TOF/MS <b>A</b> Possible sialylated structures present on porcine thyroglobulin. <b>B</b> Relative abundance of α2-6 and α2-6+α2-3- sialic acid linkages present in sialylated glycans.	142



Fig. 4. 29 <b>A</b> Possible sialylated structures present on bovine thyroglobulin. <b>B</b> Relative abundance of $\alpha$ 2-6 and $\alpha$ 2-6+ $\alpha$ 2-3-sialic acid linkages present in sialylated glycans.	143
Fig. 4. 30 Intracellular reaction of the non-fluorescent CFDA-SE cells-permeant molecule which diffuses into the cells and it is converted into the fluorescent CFSE by intracellular esterases.	145
Fig. 4. 31 <b>A</b> CHO standard curve with linear regression equation. <b>B</b> GLYcoPROFILE of CHO k1 adherent cells. The data are mean values of three replicates. The wells "without lectins" correspond to the empty wells where only cells are deposited. The percentage of bound cells is calculated based on the standard curve, considering that $2 \times 10^5$ cells/well are deposited.	146
Fig. 4. 32 <b>A</b> . GLYcoPROFILE of CHO supernatant at three different concentrations on a panel of lectins. The supernatant is labelled at a concentration of 1 mg/mL and then serially diluted to 0.25 mg/mL <b>B</b> . GLYcoPROFILE of CHO medium at three different concentrations on a panel of lectins. The medium is labelled at a concentration of 1 mg/mL and then serially diluted to 0.25 mg/mL <b>C</b> Illustrative bar chart of the subtraction of cell supernatant GLYcoPROFILE from which the values of fluorescence intensity of the cell medium GLYcoPROFILE are removed. Below 500 fluorescence units the signal is considered background noise. The subtraction was made only for the highest concentration of 1 mg/mL.	147
Fig. 4. 33 <b>A</b> B16 standard curve with linear regression equation. <b>B</b> GLYcoPROFILE of B16 melanoma cells. The data are mean values of three replicates. The wells "without lectins" correspond to the empty wells where only cells are deposited. The percentage of bound cells is calculated based on the standard curve, considering that $2 \times 10^5$ cells/well are deposited.	148
Fig. 4. 35 SDS-PAGE on 14% polyacrylamide gel of IgG after elution from affinity column with protein A. Line 1: Protein marker Line 2 sample 1 Line 2 sample 2 Line 3 sample 4 Line 5 sample 4.	149
Fig. 4. 36 GLYcoPROFILE® of purified IgG from Sample 1, 2, 3 and 4 performed in direct binding with the lectins SNA and MAAI.	150
Fig. 4. 37 <b>A</b> SDS PAGE analysis of HPyL. The protein sample was analysed under denaturing conditions on 14% polyacrylamide gel. <b>Line 1</b> : Protein marker, <b>Line 2</b> : HPyL (35 kDa). <b>B</b> GLYcoPROFILE in direct binding of HPyL interacting with labelled NeNeuGc, NeoNeuAc, and NeoFucose at $2.3 \mu\text{M}$ .	151
Fig. 4. 38 GLYcoPROFILE in direct binding of WGA interacting with labelled NeoNeuAc, NeoNeuGc and NeoFucose at $0.6 \mu\text{M}$ .	152
Fig. 4. 39 <b>A</b> GLYcoPROFILE in direct binding representing the interaction between the lectin WGA and the labelled neoglycoprotein NeoNeu5Ac at a starting concentration of $2 \mu\text{M}$ . The curve is obtained from three independent replicates. All measurements were performed in duplicates. <b>B</b> GLYcoPROFILE in direct binding representing the interaction between the lectin HPyL and the labelled neoglycoprotein NeoNeu5Gc at a starting concentration of $2 \mu\text{M}$ . The curve is obtained from three independent replicates. All measurements were performed in duplicates	153
Fig. 4. 40 Inhibition curves of WGA with the inhibitors NeNeuAc and Transferrin at the starting concentration of $10 \mu\text{M}$ . The labelled tracer NeoNeuGc is at the fixed concentration of $50 \mu\text{g/mL}$ . All measurements were performed in duplicates.	153

## Table of tables

Table 3. 1 MOA $\beta$ T characteristics.	66
Table 3. 2 BC50 and Bmax calculated for MOA $\beta$ T based on the experimental curves shown in Fig 3.13.	71
Table 3. 3 Characteristics of mAbs.	79
Table 3. 4 Calculation of Cetuximab equivalents of neoglycoprotein based on the standard curves in Annex I.	80
Table 4. 1 IC50 and RIP values calculated for all inhibitors tested with the lectins SNA and MAAI. The calculation is based on the inhibition curves in Fig.4.16 and 4.17.	131
Table 4. 2 Characteristics of glycoproteins considered in the study.	134

Table 4. 3	Equivalents of neoglycoprotein and ratio $\alpha 2,3/ \alpha 2,6$ on Fetuin, calculated at the maximum concentration based on the standard curve Fig. 4.18.	137
Table 4. 4	Summary of glycosidic linkages relative abundances and ratio $\alpha 2,6/ \alpha 2,3$ on Fetuin.	138
Table 4. 5	Equivalents of neoglycoprotein and ratio $\alpha 2,3/ \alpha 2,6$ on Transferrin, calculated at the maximum concentration based on the standard curve Fig. 4.18.	139
Table 4. 6	Comparison of equivalents of neoglycoprotein and ratio $\alpha 2,3/ \alpha 2,6$ on Transferrin native and denatured, calculated at the maximum concentration based on the standard curve Fig. 4.18.	140
Table 4. 7	Summary of glycosidic linkages relative abundances and ratio $\alpha 2,6/ \alpha 2,3$ on Transferrin.	141
Table 4. 8	Equivalents of neoglycoprotein and ratio $\alpha 2,3/ \alpha 2,6$ on Transferrin, calculated at the maximum concentration based on the standard curve Fig. 4.18.	142
Table 4. 9	Summary of glycosidic linkages relative abundances and ratio $\alpha 2,6/ \alpha 2,3$ on both Porcine and Bovine Thyroglobulin.	143
Table 4. 10	Sample 1, 2, 3, 4 obtained by EFS and purified on Protein A. The concentration of purified IgG was estimated by Bradford Protein Assay.	149
Table 4. 11	Quantity of $\alpha 2,3$ sialic acid and $\alpha 2,6$ sialic acid expressed in equivalents of neoglycoproteins. The calculation is made only for the highest IgG experimental concentration.	150
Table 4. 12	List of all lectins used through this thesis and corresponding specificities.	158

## Table of Contents

<b>Chapter 1- Introduction and thesis objective</b>	1
1.1 Project description and thesis structure	2
1.1.1 Description du projet et structure de la thèse	6
1.2 Introduction	9
1.3 Lectins	10
1.4 Techniques to study glycobiological interactions	12
1.4.1 Lectin arrays	12
1.4.2 Glycan arrays	13
1.4.3 SPR	14
1.4.4 ITC	15
1.5 Other techniques to study glycans	16
1.5.1 HPLC	16
1.5.2 Capillary electrophoresis	16
1.5.3 Mass spectrometry	17
1.6 Biotherapeutic glycoproteins	19
1.6.1 History	19
1.6.2 Glycosylation in biotherapeutics	20
1.6.3 Production systems	20
1.6.4 Biotherapeutics regulation	24
1.7 Research aim	26
1.7.1 Objectif de la recherche	27
<b>Chapter 2 – Neoglycoproteins</b>	29
Abstract	30
Résumé	30
2.1 Multivalency: a powerful tool	31
2.2 Different types of multivalent synthetic glycoconjugates	33
2.3 Neoglycoproteins	34
2.3.1 Applications	34
2.3.2 Synthesis	35
2.3.3 Synthesis by click chemistry	36
2.4 Results and discussion	37
2.4.1 Synthesis of BSA-alkyne and BSA-azide	38
2.4.2 Synthesis of NeoGalili neoglycoprotein	40
2.4.3 Synthesis of Neo6'SL and Neo3'SL neoglycoproteins	42
2.4.4 Synthesis of NeoNeuAc and NeoNeuGc neoglycoproteins	45

Conclusion.....	47
<b>Paper 1- Homo- and Heterovalent Neoglycoproteins as Ligands for Bacterial Lectins .....</b>	<b>49</b>
<b>Chapter 3 – The Galili antigen.....</b>	<b>56</b>
Abstract .....	57
Résumé .....	57
3.1 The Galili antigen .....	58
3.2 Applications Galili antigen LEctPROFILE kit .....	59
3.2.1 In biotherapeutics – the Cetuximab case study .....	59
3.2.2 In organ transplantation .....	61
3.2.3 In vaccines and other clinical settings.....	63
3.3 The <i>Marasmius oreades</i> agglutinin .....	65
3.4 GSLI-B <sub>4</sub> .....	66
Results and discussion .....	66
3.5 MOAβT Purification .....	66
3.6 GSLI-B <sub>4</sub> Functionality .....	67
3.7 Parameters of the Galili antigen LEctPROFILE kit .....	68
3.7.1 Determination of MOAβT quantity/well .....	68
3.7.2 Direct binding in absorbance on covalent plate .....	71
3.7.3 Inhibition in absorbance on covalent plate.....	72
3.7.4 MOAβT studies on adsorption plate with fluorescence detection.....	73
3.7.5 Direct binding in fluorescence on adsorption plate .....	74
3.7.6 Comparison between MOAβT and rMOA .....	74
3.7.7 GSLI-B <sub>4</sub> direct binding in fluorescence on adsorption plate .....	75
3.7.8 GSLI-B <sub>4</sub> inhibition in fluorescence on adsorption plate.....	76
3.8 Monoclonal antibodies: a proof of concept .....	77
Conclusion.....	80
Annex I – standard curves .....	82
<b>Paper 2 - Extending Janus lectins architecture: characterization and application to protocells .....</b>	<b>85</b>
<b>Chapter 4 – Sialic acid .....</b>	<b>108</b>
Abstract .....	109
Résumé .....	110
4.1 Biochemical and functional overview .....	111
4.2 N-acetyl neuraminic acid.....	113
4.2.1 Neu5Ac in forensic sciences .....	114
4.2.2 Neu5Ac and specific linkages .....	115
4.3 Applications of α2-3/ α2-6 LEctPROFILE kit.....	117
4.3.1 In pathogens recognition and in cancer .....	117
4.3.2 In biotherapeutics.....	118
4.3.3 In antibodies .....	119

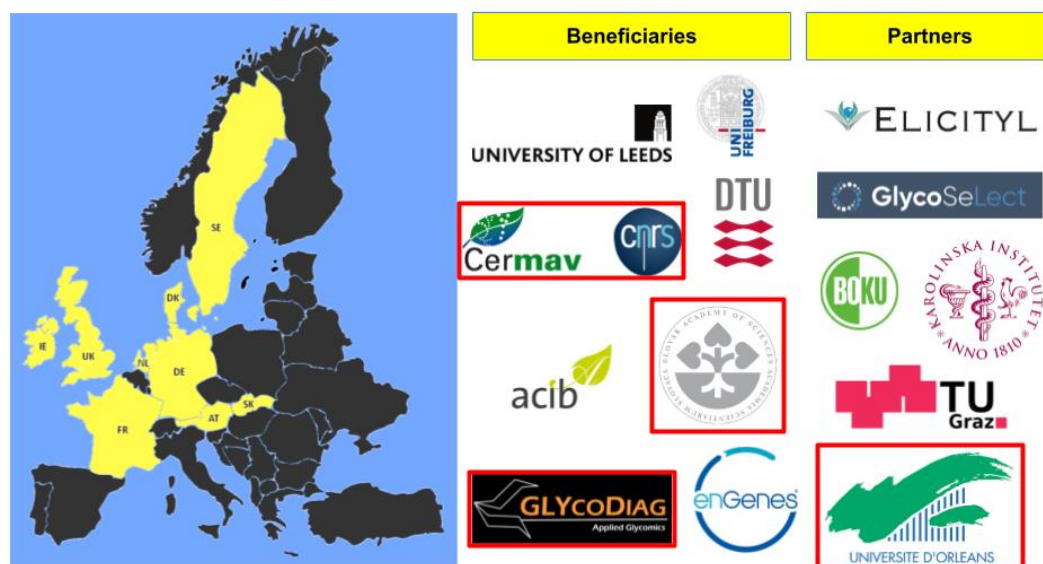
4.4 N-glycolyl neuraminic acid .....	121
4.5 Applications of Neu5Ac/Neu5Gc LEctPROFILE kit .....	122
4.5.1 In cancer .....	123
4.5.2 In biotherapeutics .....	124
4.5 Lectins considered in this study .....	124
4.5.1 MAAI .....	124
4.5.2 SNA .....	125
4.5.3 HPyL .....	126
4.5.4 WGA .....	126
4.5.5 SSL .....	127
Results and discussion .....	127
4.6 MAAI Purification .....	127
4.7 SNA functionality .....	128
4.8 Parameters of $\alpha$ 2-3/ $\alpha$ 2-6 LEctPROFILE kit .....	129
4.8.1 Direct binding .....	129
4.8.2 Inhibition .....	130
4.9 Validation of the kit: a comparison study .....	133
4.9.1 GLYcoPROFILE® and MALDI-TOF-MS of reference glycoproteins .....	133
4.9.2 Fetuin .....	135
4.9.3 Transferrin .....	138
4.9.4 Porcine and Bovine Thyroglobulin .....	141
4.10 GLYcoPROFILE® of cells .....	144
4.10.1 Chinese Hamster Ovary Cells .....	144
4.10.2 Melanoma Cells .....	147
4.11 GLYcoPROFILE® of serum IgG .....	148
4.12 Neu5Ac/Neu5Gc LEctPROFILE kit .....	151
4.12.1 HPyL Purification .....	151
4.12.2 WGA functionality .....	151
4.13 Parameters of Neu5Ac/Neu5G LEctPROFILE kit .....	152
4.13.1 Direct binding .....	152
4.13.2 Inhibition .....	153
Conclusion .....	154
Annex II .....	155
<b>Paper 3 - Polymers of a transition-state sialyl cation strongly Inhibit Bacterial Sialidases .....</b>	<b>159</b>
<b>Chapter 5 – Materials and methods .....</b>	<b>167</b>
5. Materials and methods .....	168
5.1 Neoglycoprotein synthesis .....	168
5.1.1 Mass spectrometry .....	168
5.1.2 Functionalization of BSA with an alkyne or azido group .....	168
5.1.3 Synthesis of neoglycoprotein by click chemistry .....	168

5.2. Biotinylation of neoglycoproteins .....	169
5.3 Determination of protein concentration with Bradford Protein Assay .....	169
5.4 Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE electrophoresis) ..	169
5.5 GLYcoPROFILE® .....	169
5.6 Mammalian cells culture and GLYcoPROFILE® of cells .....	171
5.6.1 CHO K1 .....	171
5.6.2 B16 .....	172
5.7 Enzymatic digestion .....	172
5.8 Serum IgG Purification .....	172
<b>Chapter 6 – Conclusive remarks</b> .....	174
Final conclusion and perspectives .....	175
Conclusion finale et perspectives .....	177
List of communications .....	180
Bibliography .....	182

## **Chapter 1- Introduction and thesis objective**

## 1.1 Project description and thesis structure

Synthetic Glycobiology is a rapidly growing field aiming at engineering and redesigning glycans, glycan-active enzymes and glycan-binding proteins for fundamental research, biomedicine and biotechnology. My PhD thesis has been carried out in the framework of the synBIOcarb (<https://synbiocarb.science/>) Innovative Training Network (ITN) part of the European Union's Horizon 2020 research and innovation programme, under the Marie Skłodowska Curie grant agreement no. 814029. SynBIOcarb started in 2019 and it is a consortium of European academic and industrial beneficiaries where 15 PhD students, called Early Stage Researchers (ESR), have been trained. In the network there are other partners, with the purpose to contribute to the consortium, giving the students the adequate support required and the opportunity to carry out secondments. In the map in *Fig. 1.1*, are highlighted the European countries that take part in the synBIOcarb ITN and a list of the beneficiaries and partners of the network; in red are highlighted our main collaborators. SynBIOcarb is a multidisciplinary network and each one of the participant has high expertise either in chemistry, biochemistry, molecular biology and protein engineering, structural biology, cell biology, bioanalytical tools. The final common goal is the development and exploitation of synthetic glycobiology for diagnostics and targeted drug delivery through the building of collaborations.



*Fig. 1. 1* synBIOcarb network. On the left, highlighted in yellow all countries where the beneficiaries and partners are located. On the right, the logos of all beneficiaries and partners involved in the network.



SynBIOcarb is organized into 4 work packages, schematically presented in *Fig. 1.2* . My PhD thesis is part of the **workpackage 3** and has taken place at the company GLYcoDiag with the Université d'Orléans as partner institution. GLYcoDiag was funded by Dr. Ludovic Landemarre in 2005 and it is a company specialized in glycoanalysis; the first one to launch lectin array research, application and services in France. In my PhD project, the already established lectin array-based method called GLYcoPROFILE® technology [1](described in Paragraph 1.4.1) has been tailored to serve the aim of the thesis that is **building a lectin kit** for the detection of the **major undesirable glycan structures** found mainly on biotherapeutic glycoproteins. The goal is to allow real-time monitoring and quality control **of these biotherapeutics, but also other potential applications (vaccines quality control, diagnostics)** .

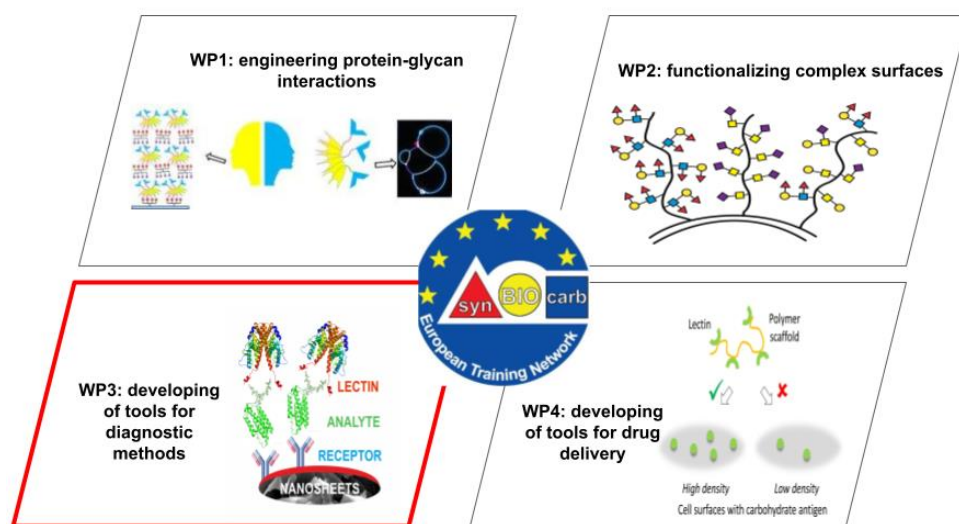


Fig. 1. 2 synBIOcarb work packages

The research has been mostly focused on two glycans: the **Galili antigen** and the **sialic acid**. Given the wide range of opportunities offered by this field, during the last three years my project has broadened. It has become interesting for us to imagine new areas other than the quality control, in which a rapid lectin kit could turn out to be a huge help in the detection of target glycans.

To help the reader navigate the thesis, here it is briefly described its structure:

**Chapter 1** is a general introduction on the subject, especially focused on the study of glycans and lectin-glycan interactions and the importance of rapid quality control of biotherapeutics.

Chapter 2, 3 and 4 contain the main results and are designed for the reader to be read independently: each one has its own abstract, introduction, results and discussion and conclusion.

**Chapter 2** is the prelude to Chapter 3 and 4 because it contains the results regarding the neoglycoprotein synthesis which, together with lectins, are the main tools for the development of the kits. At the end of this chapter, it is attached the paper in collaboration with Dr. Goyard and Prof. Renaudet of University of Grenoble-Alpes: “Homo- and Heterovalent Neoglycoproteins as Ligands for Bacterial Lectins”.

**Chapter 3** is entirely dedicated to the Galili antigen. At the end of this chapter, it is attached the paper in collaboration with Dr. Simona Notova of Cermav in Grenoble: “Extending Janus lectins architecture: characterization and application to protocells”.

**Chapter 4** is entirely dedicated to sialic acid. At the end of this chapter, it is attached the paper in collaboration with Dr. Gouin of University of Nantes: “Polymers of a transition-state sialylation strongly Inhibit Bacterial Sialidases”.

**Chapter 5** is dedicated to materials and methods and it is in common for all experiments because the techniques used are the same for all the three results chapters.

**Chapter 6** is a short summary of what are the thesis main accomplishments and the future perspectives.

The experimental work can be divided into four big chunks that put together have led to reach the final objective of building the lectin kits. The workflow is represented in *Fig. 1.3*.

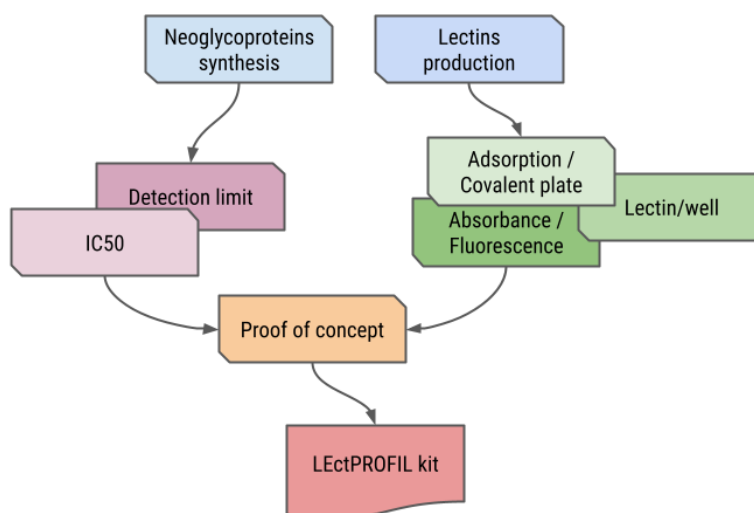
The four steps are:

1. Purification of highly specific lectins. This part of the work intended to target above all the lectins that are specific towards the glycans of interest. The lectins are either recombinant or from natural sources. Concerning recombinant lectins, the first one that was purified is MOA $\beta$ T and this was done in collaboration with Dr. Simona Notova of the synBIOcarb network, during my secondment at the Cermav-CNRS in Grenoble in January 2020. The other recombinant one is HPyL specific for Neu5Gc, purified at GLYcoDiag. Regarding the natural lectins, the lectin MAAI was purified by me, and the lectin SNA has been purified at GLYcoDiag before my arrival.

2. Synthesis of neoglycoproteins. Neoglycoproteins are a fundamental part of the construction of the kit, because they are used as standard probes to set the reference parameters for the kits use.

**3.** Study of the kits' parameters: choice of the quantity lectin/well, choice of the type of plate, if adsorption or covalent, the method of detection if in absorbance or fluorescence, the detection limit, half-maximum inhibitory concentration (IC50).

**4.** Proof of concept, meaning testing the kit with relevant molecules, one part of this work was done in collaboration with Juvissa Aguedo , another PhD student of the synBIOcarb network.



*Fig. 1. 3 Thesis workflow.*

Being my research inserted in the context of a company, the next step after the proof of concept is the redaction of all the documents required for the kits to be suitable for applications(user guide, specifications).

The good specifications and the precise definition of all the kits parameters are of fundamental importance at a research level to ensure stability and reliability to the potential users.

### 1.1.1 Description du projet et structure de la thèse

La glycobiologie synthétique est un domaine en plein essor visant l'ingénierie et la reconception des glycanes, des enzymes actives sur les glycanes et des protéines de liaison aux glycanes pour la recherche fondamentale, la biomédecine et la biotechnologie. Ma thèse de doctorat a été réalisée dans le cadre du réseau de formation innovant (ITN) synBIOcarb (<https://synbiocarb.science/>) faisant partie du programme de recherche et d'innovation Horizon 2020 de l'Union européenne, sous la convention de subvention Marie Skłodowska Curie n° 814029. SynBIOcarb a commencé en 2019 et c'est un consortium de bénéficiaires académiques et industriels européens où 15 doctorants, appelés Early Stage Researchers (ESR), ont été formés. Dans le réseau, il y a d'autres partenaires, dans le but de contribuer au consortium, en donnant aux étudiants le soutien adéquat requis et la possibilité d'effectuer des détachements. Sur la carte de la *Fig. 1.1*, sont mis en évidence les pays européens qui participent au RIT synBIOcarb et une liste des bénéficiaires et des partenaires du réseau ; en rouge sont mis en évidence nos principaux collaborateurs. SynBIOcarb est un réseau multidisciplinaire et chacun des participants possède une grande expertise en chimie, biochimie, biologie moléculaire et ingénierie des protéines, biologie structurale, biologie cellulaire, outils bioanalytiques. L'objectif final commun est le développement et l'exploitation de la glycobiologie synthétique pour le diagnostic et l'administration ciblée de médicaments grâce à la mise en place de collaborations.

SynBIOcarb est organisé en 4 workpackages, présentés schématiquement dans la *Fig. 1.2*. Ma thèse fait partie du workpackage 3 et s'est déroulée au sein de la société GLYcoDiag avec l'Université d'Orléans comme institution partenaire. GLYcoDiag a été fondée par le Dr. Ludovic Landemarre en 2005 et est une société spécialisée dans la glycoanalyse ; la première à lancer la recherche, l'application et les services de lectin array en France. Dans mon projet de thèse, la méthode déjà établie basée sur les réseaux de lectines appelée technologie GLYcoPROFILE® [1] (décrite dans le paragraphe 1.4.1) a été adaptée pour servir l'objectif de la thèse qui est de construire un kit de lectines pour la détection des principales structures glycaniques indésirables trouvées principalement sur les glycoprotéines biothérapeutiques. L'objectif est de permettre le suivi et le contrôle de qualité en temps réel de ces produits biothérapeutiques, mais aussi d'autres applications potentielles (contrôle de qualité des vaccins, diagnostics).

La recherche a été principalement axée sur deux glycanes : l'antigène Galili et l'acide sialique. Étant donné le large éventail de possibilités offertes par ce domaine, mon projet s'est élargi au cours des trois dernières années. Il est devenu intéressant pour nous d'imaginer de nouveaux domaines autres que le contrôle de qualité, dans lesquels un kit de lectine rapide pourrait s'avérer d'une grande aide pour la détection des glycanes cibles.

Pour aider le lecteur à s'orienter dans cette thèse, nous décrivons ici brièvement sa structure:

Le **chapitre 1** est une introduction générale sur le sujet, particulièrement axée sur l'étude des glycanes et des interactions lectine-glycane et sur l'importance du contrôle rapide de la qualité des produits biothérapeutiques.

Les chapitres 2, 3 et 4 contiennent les principaux résultats et sont conçus pour être lus indépendamment les uns des autres: chacun a son propre résumé, son introduction, ses résultats, sa discussion et sa conclusion.

Le **chapitre 2** est le prélude aux chapitres 3 et 4 car il contient les résultats concernant la synthèse des néoglycoprotéines qui, avec les lectines, sont les principaux outils pour le développement des kits. À la fin de ce chapitre est joint le document rédigé en collaboration avec le Dr Goyard et le professeur Renaudet de l'Université de Grenoble-Alpes : "Homo- et Heteroalent Neoglycoproteins as Ligands for Bacterial Lectins".

Le **chapitre 3** est entièrement consacré à l'antigène Galili. À la fin de ce chapitre est joint l'article rédigé en collaboration avec le Dr Simona Notova du Cermav à Grenoble : "Extending Janus lectins architecture : characterization and application to protocells".

Le **chapitre 4** est entièrement consacré à l'acide sialique. À la fin de ce chapitre est joint l'article rédigé en collaboration avec le Dr Gouin de l'Université de Nantes: " Polymers of a transition-state sialyl cation strongly Inhibit Bacterial Sialidases ".

Le **chapitre 5** est consacré aux matériaux et méthodes et il est commun à toutes les expériences car les techniques utilisées sont les mêmes pour les trois chapitres de résultats.

Le **chapitre 6** est un bref résumé des principales réalisations de la thèse et des perspectives d'avenir.

Le travail expérimental peut être divisé en quatre grandes parties qui, mises ensemble, ont permis d'atteindre l'objectif final de construire les kits de lectines. Le flux de travail est représenté dans la Fig. 1.3.

Les quatre étapes sont :

1. Purification de lectines hautement spécifiques. Cette partie du travail avait pour but de cibler avant tout les lectines qui sont spécifiques aux glycanes d'intérêt. Les lectines sont soit recombinantes, soit de source naturelle. Concernant les lectines recombinantes, la première qui a été purifiée est MOA $\beta$ T et cela a été fait en collaboration avec le Dr Simona Notova du réseau synBIOcarb, lors de mon détachement au Cermav-CNRS de Grenoble en janvier

2020. L'autre recombinant est le HPyL spécifique du Neu5Gc, purifié chez GLYcoDiag. Concernant les lectines naturelles, la lectine MAAI a été purifiée par mes soins, et la lectine SNA a été purifiée au GLYcoDiag avant mon arrivée.

2. Synthèse des néoglycoprotéines. Les néoglycoprotéines sont une partie fondamentale de la construction du kit, car elles sont utilisées comme sondes standards pour établir les paramètres de référence pour l'utilisation des kits.

3. Etude des paramètres des kits : choix de la quantité de lectine/puits, choix du type de plaque, si adsorption ou covalente, la méthode de détection si en absorbance ou fluorescence, la limite de détection, la concentration inhibitrice à 50 % (IC50).

Ma recherche étant insérée dans le contexte d'une entreprise, l'étape suivante après la preuve de concept est la rédaction de tous les documents requis pour que les kits soient adaptés aux applications (guide d'utilisation, spécifications).

Les bonnes spécifications et la définition précise de tous les paramètres des kits sont d'une importance fondamentale au niveau de la recherche pour assurer la stabilité et la fiabilité aux utilisateurs potentiels.

## 1.2 Introduction

The evolution of genome sequencing, from the 1970 onwards, has opened up a completely new world in front of the human eyes: the possibility to understand the information hidden inside our cells, the genetic code. The central dogma of molecular biology, DNA copied into mRNA translated into proteins, finds its strength in the fact that DNA replication and translation are template-driven, meaning that proteins are primary genes products. Considering that the human genome contains only 25000 genes, which is not much more than the one of *Drosophila melanogaster* (20000 genes), the hypothesis that came out was that human complexity lies in post-translational modifications of proteins [2]. One of them, and the most common one, is protein glycosylation. Hence, as it turned out, genome decoding does not explain all biological functions that our cells perform because it does not include a whole class of macromolecules: carbohydrates, also called glycans, or simply sugar chains. Glycans are not only energy-storage molecules, but they participate in many biological processes [3]. The **glycome** is defined as the entire pool of glycans, free or attached to other macromolecules, such as proteins or lipids, that is synthesized by our cells “under specific conditions of time, space, and environment” [4]. As the definition says, the glycome can change according to many factors and its synthesis does not depend on directly from the DNA sequence. It depends instead on enzymes, encoded by DNA, like glycosidases and glycosyltransferases that can attach carbohydrates in different order, organization, in a linear or branched configuration. The fact that it is not possible to predict the glycomes structure increases its complexity and makes its decodification very challenging.

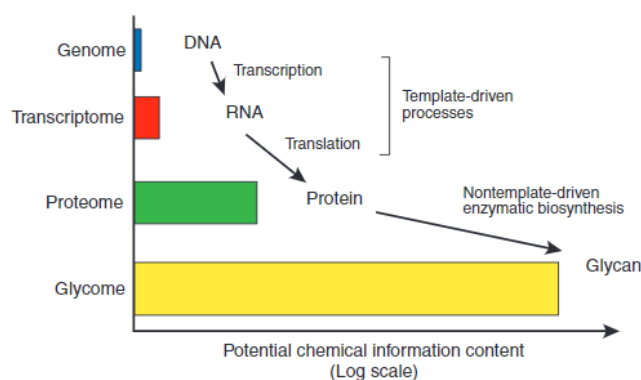


Fig. 1. 4 The non-template glycosylation drastically increases biological functional diversity, adapted from [5].

One of the keys to unlocking the so-called “glycocode” and deciphering the information that glycans come from natural inspiration from glycobiological interactions. Indeed, the main mechanisms driving plants and animals lives result from glycobiological interactions, that involve glycans specific recognition by glycan recognition molecules such as **lectins**.

### 1.3 Lectins

The name “lectin” originates from the Latin word “legere” that means “to select”. Lectins in fact are proteins with at least one non-catalytic domain, and of non-immune origin, which are able to specifically recognize, or select, and to reversibly bind carbohydrates. Lectins fall into the group of glycan-binding proteins, in which also sulphated glycosaminoglycan-binding proteins are placed. In 1919, the first lectin discovered was the *Canavalia ensiformis* or ConA or jack bean agglutinin, named from the leguminous plant it comes from and from the fact that it is able to clump red-blood cells. Years later, Landsteiner confirmed the idea that the agglutination was due to the interaction of the haemagglutinin with sugars in the blood, saying that “the actions of plant haemagglutinins resemble antibody reactions in all essentials” [6]. Nowadays we know that lectins/glycans interaction, unlike antibodies/antigens ( $K_d$   $10^{-8}$ - $10^{-12}$  M), have a relatively low affinity ranging from millimolar for monosaccharides ligand to micromolar for oligosaccharides [7]. Lectins can be either a single carbohydrate-recognition domain (CRD) or an entire protein (containing more domains other than the CRD, e.g. proteolytic domains) and are ubiquitous, meaning they are present in all living organisms. They can be classified according to their monosaccharide specificity but more recently they are classified based on their CRD structure (Fig. 1.5) [4].

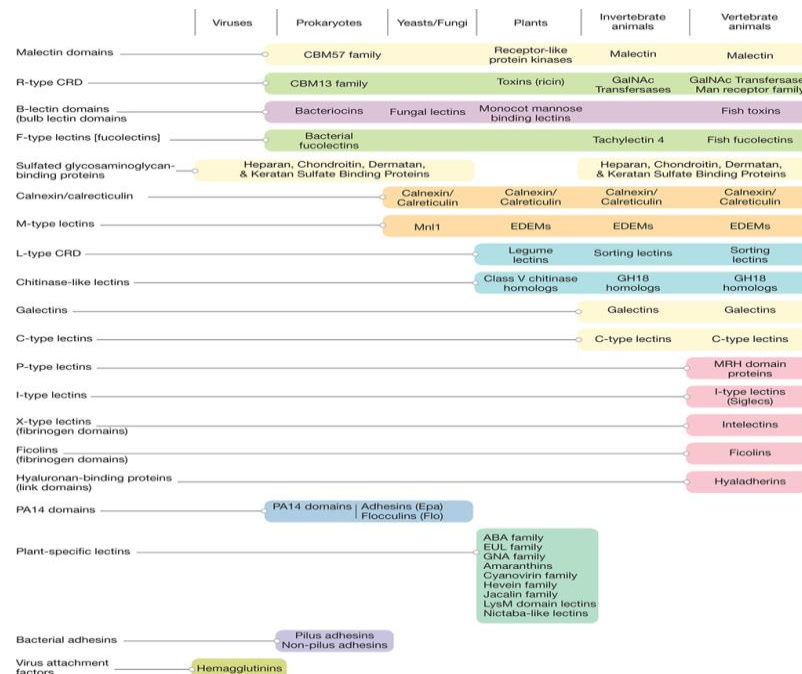


Fig. 1. 5 Classification of lectin based on their CRDs and their presence across the kingdoms of life adapted from [4]

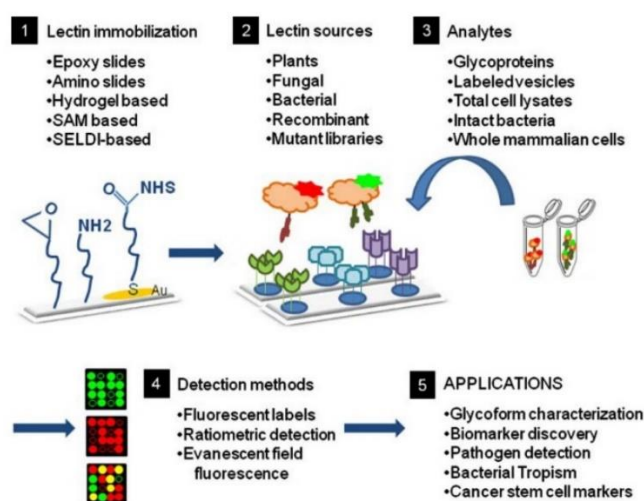


On one hand, many lectins are involved in immune responses, on the other hand, pathogens also often use their own lectins to attach to host cells during infection. Lectins also function in communication between cells in multicellular organisms, resulting in movement of molecules, cells, and information [4]. Lectins interact with their ligands mostly through weak interactions, mainly the formation of a hydrogen-bond network or, in the case of some positively or negatively charged carbohydrates (e.g. sialic acid) or modified carbohydrates (e.g. sulphated sugars), through electrostatic interactions [8]. Moreover, hydrophobic interactions ensure the interaction between the hydrophobic component of the glycan structure and the aromatic residues present in the CRD. Metal cations, such as  $\text{Ca}^{2+}$  and  $\text{Mn}^{2+}$ , can also be involved in the recognition as with  $\text{Ca}^{2+}$  for C-type animal lectins. The reasons of the low affinity for the glycan–lectin interactions may be due to several variables: to the shallow binding pockets on the surface of the protein, to the limited changes in the conformations of the binding site after the ligand binding, the restricted binding surface available for the ligand, the inability of the ligand to exhibit multiple binding modes [3]. If the **affinity is the strength of the interaction that occurs between a single protein and a single ligand**, in the case of lectins it would be better to speak about avidity. The avidity compensates for the rather low affinity and it can be defined as the **sum of all the strengths of all the multiple simultaneous interactions between the lectin and the interacting glycan**, leading to an affinity enhancement [9]. The importance of multivalent interactions in generating high affinity was demonstrated by Ashwell and co-workers, who showed that desialylated serum glycoproteins exposing several Gal residues could increase of 10000-fold the affinity towards rabbit hepatic lectin [10]. So, for lectins to be able to cross-link glycan-containing structures, from which their agglutination ability come, they either are organized as oligomers of CRDs, either contain multiple binding sites on a single polypeptide chain, or again they are constituted by the association of polypeptides containing CRDs.

## 1.4 Techniques to study glycobiological interactions

### 1.4.1 Lectin arrays

Since their discovery, lectins have been used for all sorts of applications, spacing from lectin affinity chromatography, blood typing, histology, flow cytometry and blotting [11]. With the advent of glycomics, they have been applied for lectin arrays, providing a fast, high-throughput screening platform that allows to monitor at the same time several lectin-glycan interactions. Lectin arrays were first introduced in 2005 [12] and the basic format consists of **lectins immobilized on a surface** that could be a plate or a glass slide, through **adsorption, covalent or biochemical interaction**. Methods for lectin immobilization could be carbene insertion, biotin–avidin bridge, coupling through primary amine on lysine of lectin with N-hydroxysuccinimide (NHS)-derivates esters, exploitation of self-assembled monolayers of thiols on gold-coated surfaces, and 3D hydrogel surfaces. None of this methods guarantees the complete control of the optimal lectin orientation [13]. The interacting sample is usually found in solution, and can be of various nature, from glycoproteins to cells, to vesicles (*Fig. 1.6*).



*Fig. 1. 6 Lectin array design, adapted from [13].*

The detection of the interaction can be achieved in three different ways: **direct mode**, **indirect mode** or **competitive mode**. In the direct mode the binding partner is labelled usually with a fluorophore, and the fluorescence is detected after the binding [12]. In the indirect mode, the

partner is not labelled but it can be recognized by a fluorescent secondary antibody or the initial incubation with a biotinylated antibody is followed by incubation with fluorescent-labelled streptavidin. In this case, there is one step more before the fluorescence detection. In the competitive mode the binding partner could be either labelled or not labelled. If it is not labelled, another well-characterized glycan partner is labelled and the two molecules compete for the lectins on the array. In the GLYcoPROFILE® in inhibition, the interaction is measured as percentage of inhibition of the binding of the labelled molecule to the lectin (Paragraph 1.7). If the binding partner is labelled, the competitor molecule is labelled as well but with a different dye. One example is the dual colour measurement where both samples compete for the lectin immobilized on the array, and there is different ratio of two dyes after the incubation time due to the different affinity of each binding partner for the lectin [14]. One drawback of lectin array are the washing steps that need to be performed to remove the unbound molecules. These steps are time-consuming and could be detrimental in the case of very weak lectin-glycan interactions. Hirabayashi *et al.* [15] found a solution to this problem through evanescent-field fluorescence detection for weak interactions between lectins and carbohydrates in a liquid phase. This method does not require the washing step after a probing reaction. Another method, reverse phase dot blot lectin array is formed by immobilizing the samples containing glycoproteins on a surface, which is then analysed by addition of biotinylated lectins over the chip surface [16]. Chen *et al.* [17] designed a sandwich format of antibody microarray that capture multiple glycoproteins followed by detection with lectins. Lectin arrays find their application particularly in the following of the progression of pathologies such as cancer: Qiu *et al.* [18], showed the identification of potential plasma biomarkers for colorectal cancer, and Bertok *et al.* [19], showed the use of serum samples from healthy individuals, benign prostate hyperplasia patients, and prostate cancer patients.

#### 1.4.2 Glycan arrays

Glycan arrays are another high-throughput screening technique to study lectin-glycan interactions but in this case, a variety of oligosaccharides, glycolipids, or glycoproteins are bound to solid supports and used to probe the carbohydrate-binding properties of proteins or cells. They can be printed in many different ways: chemically or enzymatically, without or with different types of linkers, but also the density of glycans on the surface can be adjusted according to the needs. Since they allow the screening of hundreds or thousands of samples at the same time, they are a valuable method for biomarker discovery, especially in the initial searching phase [20]. Using an O-glycopeptide microarray, Wandall *et al.* demonstrated the presence of higher levels of antibodies against O-glycosylated mucin 1 in patients with breast,

ovarian, and prostate cancer compared to that of healthy controls [21]. Although these glycan arrays provide information about carbohydrate-interacting proteins, they do not allow us to directly examine changes in glycosylation.

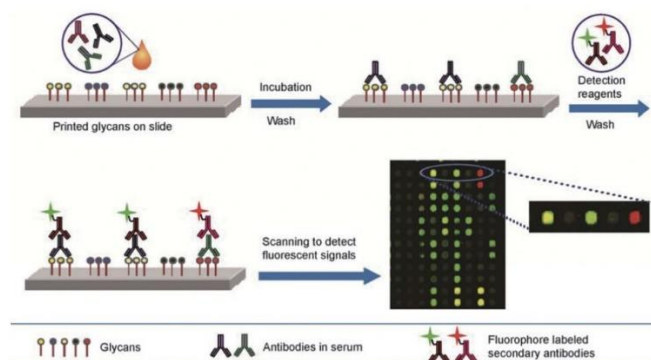


Fig. 1. 7 Glycan array principle, adapted from [20]

### 1.4.3 SPR

Surface plasmon resonance (SPR) is a label-free bioanalytical method which allows real-time monitoring of protein-ligand interactions. The ligand is immobilized on a sensor chip of carboxymethylated dextran covalently attached to a gold surface and the analyte is injected in continuous flow over it. The analyte of interest is prepared in solution in different concentrations to flow over the ligand. The increase in mass associated with a binding event causes a proportional increase in the refractive index ( $\delta\theta$ ) of refracted light. The variation of resonance angle is quantified as resonance units (RU). Measurement of binding of different concentrations of ligand gives signal dependent on time, which is called sensorgram. It is possible to determine the kinetic of the binding: during the injection of an analyte, the binding response increase is due to the formation of analyte–ligand complexes at the surface and the sensorgram is dominated by the association phase. After a certain time of injection, a steady state is reached, in which binding and dissociating molecules are in equilibrium. The decrease in response after the end of the analyte injection is due to the dissociation of the complexes, defining the dissociation phase (*Fig 1.8*). Among the studies about glycan-protein interactions with SPR, the glycan binding preferences of the hemagglutinin derived from influenza A viruses were explored in order to understand their potential ability for interspecies transmission [22].

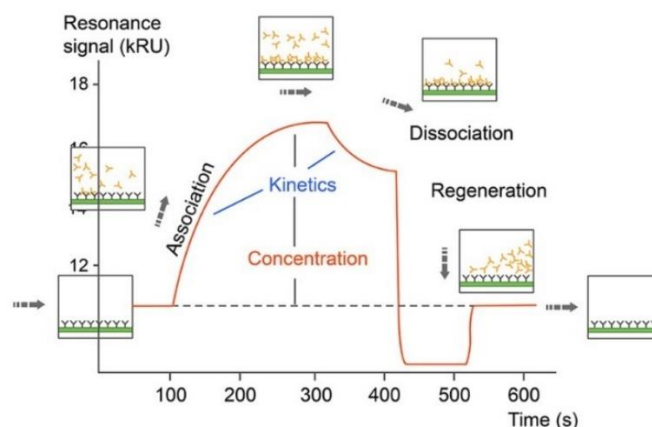


Fig. 1. 8 SPR principle adapted from <https://bit.ly/2Es3k2s>.

#### 1.4.4 ITC

Isothermal Titration Calorimetry (ITC) is a label-free quantification technique that works by directly measuring the heat that is either released or absorbed during a biomolecular binding event. By measuring the heat transfer during the binding between a glycan and its receptor, it is possible to calculate the binding stoichiometry ( $n$ ), binding constants ( $K_d$ ), and thermodynamics parameters (variations of enthalpy ( $\Delta H$ ) and entropy ( $\Delta S$ )). The reference cell and the sample cell are set to the desired experimental temperature. The ligand is loaded into a syringe which sits in a very accurate injection device. The injection device is inserted into the sample cell containing the protein of interest. If there is a binding of the ligand to the protein, heat changes of a few millionths of a degree Celsius are detected and measured (Fig. 1.9).

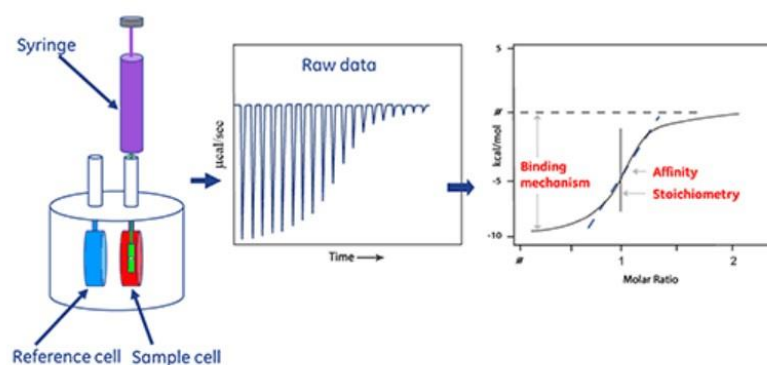
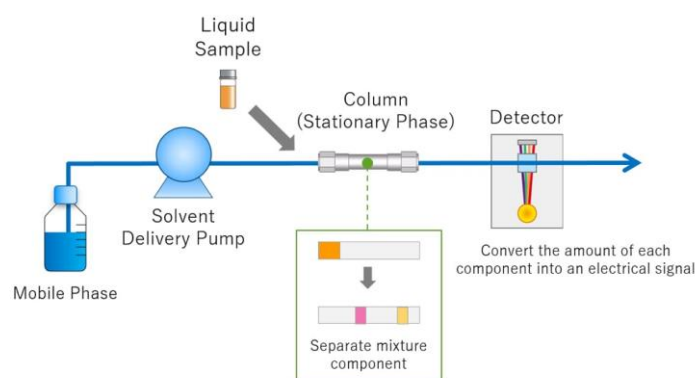


Fig. 1. 9 ITC principle adapted from <https://bit.ly/2rP1OAL>

## 1.5 Other techniques to study glycans

### 1.5.1 HPLC

High Performance Liquid Chromatography is a technique that separates compounds dissolved in a liquid sample and allows qualitative and quantitative analysis of what components and how much of each component are contained in the sample. The solvent used to separate components in a liquid sample for HPLC analysis is called the mobile phase. The mobile phase is delivered to a separation column, otherwise known as the stationary phase, and then to the detector at a stable flow rate controlled by the solvent delivery pump. The compounds separated in the column are detected by a detector downstream of the column and each compound is identified and quantified (*Fig. 1.10*).



*Fig. 1. 10* Principle of HPLC, adapted from <https://bit.ly/3UWsiMH>

### 1.5.2 Capillary electrophoresis

Capillary Electrophoresis (CE) is another micro-separative technique that separates ions based on their electrophoretic mobility with the use of an applied voltage. It follows the basic principle of all electrophoretic separation methods: different particles with different effective charges and sizes migrate with different velocities. A separation compartment is a narrow capillary usually made of fused silica, filled with an electrolytic solution (*Fig. 1.11*). CE has been shown to be particularly effective for the resolution of protein glycoforms that can be difficult using conventional methods, among which the studies upon Carbohydrate-deficient Transferrin [23].

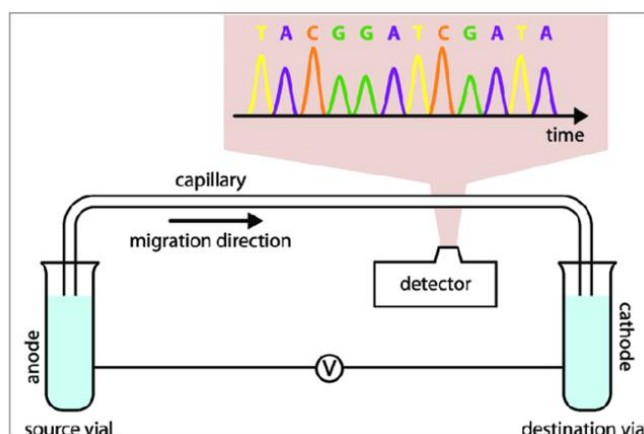


Fig. 1. 11 Schematic representation of CE system, adapted from <https://bit.ly/3SHSR64>

### 1.5.3 Mass spectrometry

Mass spectrometry (MS) is a sensitive technique used to detect, identify and quantitate molecules based on their mass-to-charge ( $m/z$ ) ratio. Mass spectrometers consist of an ion source that converts molecules into ions in the gas phase, a mass analyser that separates ionised analytes on the basis of the  $m/z$  ratio, and a detector that records the number of ions for each  $m/z$  value. Soft ionisation methods, suitable for the analysis of biological macromolecules, include: Matrix Assisted Laser Desorption ionisation Ionisation (MALDI) and ElectroSpray Ionisation (ESI). The mass analyser is fundamental to MS technology; one of the most used is the time-of-flight mass analyser (TOF). During the course of the thesis, we have performed experiments together with Juvisan Aguedo, a synBIOcarb PhD student at the Slovak Academy of Sciences, in order to compare the LEctPROFILE kit and the MALDI-TOF/MS for the discrimination of  $\alpha$ 2-3 and  $\alpha$ 2-6 sialic acid linkages (Paragraph 4.4.6). The analysis performed by Juvisan Aguedo is MALDI-TOF/MS with linkage-specific derivatization of N-glycans. The first step is the glycan release from the sample using PNGase F (Peptide - N-Glycosidase F). Afterwards, the glycans are purified using hydrophilic interaction liquid chromatography solid phase extraction. Then, the glycans are derivatized using 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide (EDC) and 1-hydroxy- benzotriazole (HOBt) as activators. EDC is meant for initial carboxylic acid activation and HOBt for catalysing the subsequent conversion to an ester. In Fig.1.12 a, it is shown the reaction of  $\alpha$ 2-6 sialic acid that undergoes ethyl esterification and has a mass shift of +28 Da, while in Fig.1.12 b,  $\alpha$ 2-3 sialic acid undergoes a lactonization, causing a mass shift of 618 Da [24].

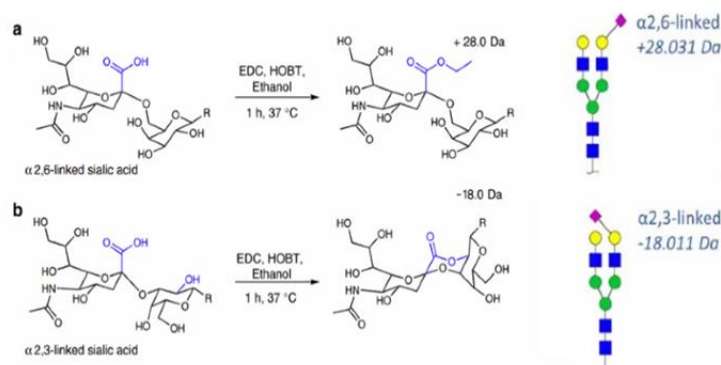


Fig. 1. 12 Derivatization of sialic acid linkages. Adapted from the poster “Rapid and high-throughput methods for discrimination of sialic acids linkages in glycoproteins” J. Aguedo and F. Vena.

Then the sample are spotted for MALDI-TOF/MS analysis (Fig1.13 ).

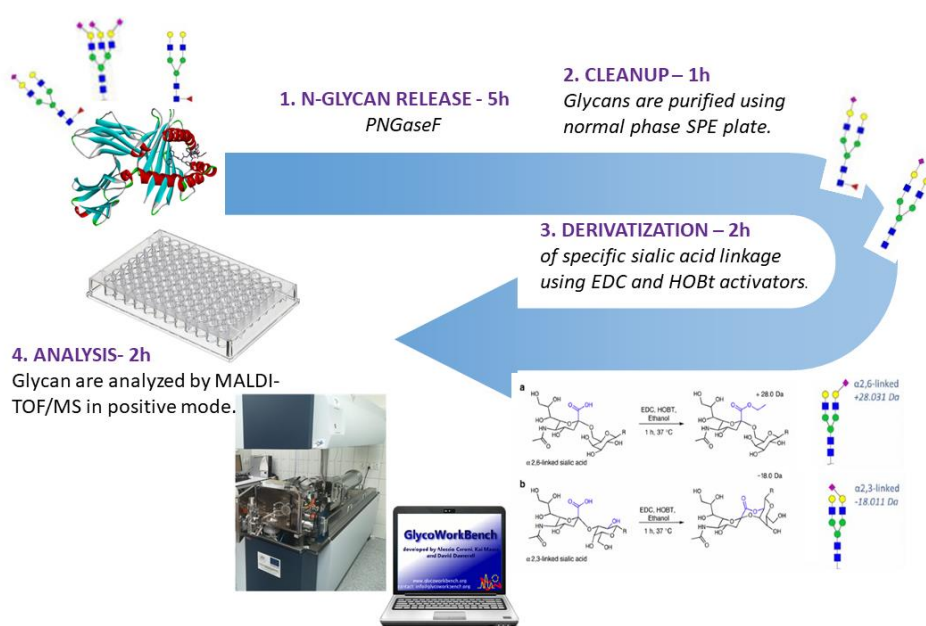


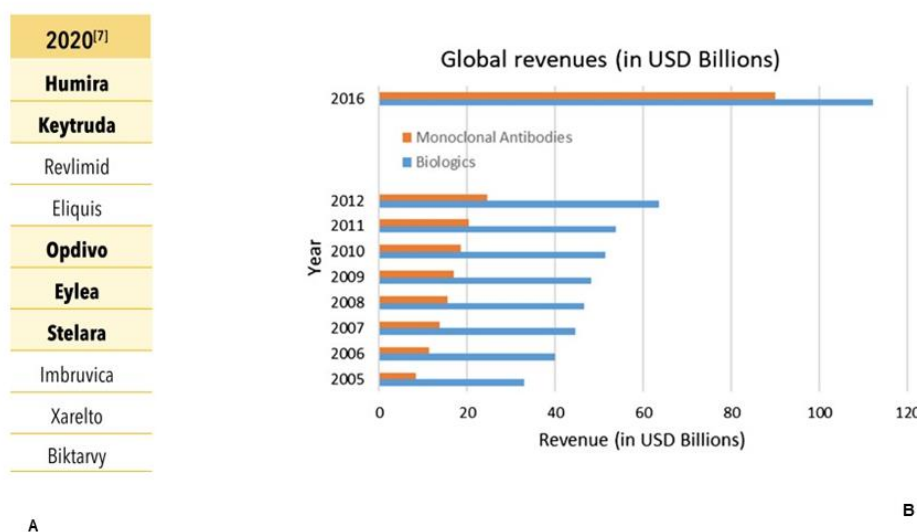
Fig. 1. 13 MALDI-TOF/MS for the discrimination of  $\alpha 2$ -3 and  $\alpha 2$ -6 sialic acid linkages workflow. Adapted from the poster “Rapid and high-throughput methods for discrimination of sialic acids linkages in glycoproteins” J. Aguedo and F. Vena.



## 1.6 Biotherapeutic glycoproteins

### 1.6.1 History

The first drug was developed by Joseph Buchner in 1828 it's Sialicin, used for pain relief, known nowadays as Aspirin. It is the first of a series of drugs called chemical molecules or small molecules, chemically synthesized organic compounds of known structure. 150 years later, in 1986, the first glycosylated biotherapeutic drug, Orthoclone, was marketed. It is a monoclonal antibody (mAb) that functions as immunosuppressant. Biotherapeutics, biologics or biological drugs are medical products produced from living organisms such as humans, animals, plants, microorganism and/or biotechnology methods. They can be fusion proteins, growth factors, cytokines, therapeutic enzymes, and hormones. They are produced in complex and highly controlled manufacturing processes [25]. Of the 2020 top global best-selling drugs, five of them are biological drugs and they are heavily glycosylated (highlighted in *Fig 1.14A*). From the list in *Fig 1.14 A*, it can be seen as well that of the five biotherapeutics, three are mAbs, Humira (Adanulinumab), Keytruda ( Pembrolizumab) and Opdivo (Nivolumab ).



*Fig. 1. 14 A* 2020 best-selling drugs, adapted from [25]. *B* Biotherapeutics global revenues from 2005 to 2006 adapted from [26].

Monoclonal antibodies are the biggest class of biotherapeutics on the market, accounting for more than 80 billion dollars in 2016 (*Fig. 1.14 B*). Their history dates back to the 1960s with the work of César Milstein and Georges Köhler that developed the hybridoma technology

fusing rodent lymphocytes and myeloma cells. They are used for the treatment of a variety of disorders, including cancer, autoimmune diseases, cardiovascular disease, infections [26].

### 1.6.2 Glycosylation in biotherapeutics

Glycosylation is a fundamental post-translational modification, essential for protein folding, stability and molecular recognition, and it goes without saying that has also a strong impact on biotherapeutics. Glycosylation of a biotherapeutic protein is critical for product solubility, stability, pharmacokinetics and pharmacodynamics (PK/PD), bioactivity, and safety [27]. Existing research recognises the critical role played by glycosylation for the structure and stability of biotherapeutics *in vivo*. It is the case of Erythropoietin, a hormone containing three N-linked and one O-linked glycans, whose role is to stimulate red blood cells production. It was demonstrated that glycans play a role in the stabilization of erythropoietin in denaturing conditions compared to the non-glycosylated protein. Sialylation protects erythropoietin from clearance by the liver *in vivo*, because the sialic acid capping prevents the exposure of galactose residues that are recognized by the Asialoglycoprotein receptors (Paragraph 4.1) [28]. The crucial role of glycosylation is very well-studied also for mAbs: their functionality it is not only linked to the antigen binding domain (Fab) in their variable region, but also to the effector function mediated by the antibody Fc fragment (fragment crystallizable). The interaction of Fc with the Fc $\gamma$ -receptors (Fc $\gamma$ R) expressed on immune cell surfaces or the molecules of the complement triggers the ADCC (Antibody-Dependent Cell-mediated Cytotoxicity) and CDC (Complement-Dependent Cytotoxicity). The mechanism is not yet completely elucidated but it seems that sialylation has an anti-inflammatory effect, because it dramatically decreases mAb affinity for the Fc $\gamma$ -receptor, inhibiting ADCC [29]. On the other side, the lack core fucosylation of IgG N-glycans, in which fucose is added  $\alpha$ 1,6 to the protein-adjacent GlcNAc of N-glycans, enhances ADCC via increased Fc $\gamma$ R binding and signalling [30].

### 1.6.3 Production systems

Various production systems exist for the manufacturing of biotherapeutics. Indeed, recombinant DNA and hybridoma technology have a lot improved the production of customized recombinant proteins. Animal and human cells are the main platforms for the manufacturing, even though non-human mammalian cells are the most used at industrial scale [31]. This trend is mostly driven by the increased attention towards the biotherapeutics glycosylation state: non-human mammalian cells give to biotherapeutics a human-like glycosylation, with the advantage that are much easier to manipulate than human cells. Human-like glycosylation makes them safer, in terms of being less immunogenic, and more

efficient, with longer half-life in the blood. There are different types of cells currently used for the manufacturing:

- **Non-human mammalian cells**

Chinese hamster ovary cells or **CHO**: together with murine myeloma cells they are mammalian host gold-standard for the production mAbs and Fc-fusion proteins [32]. They account for the 70% of recombinant biopharmaceuticals production. The advantages are that they are less sensitive to contaminations by human viruses, making them safe for commercialization, they grow very fast in suspension and they support many gene-expressing systems, giving the possibility of upscaling the reaction in the bioreactors and increase the production. Although CHO cells possess many advantages for glycoprotein productions, they are unable to produce some types of human glycosylation, such as  $\alpha$ -2,6-sialylation and  $\alpha$ -1,3/4-fucosylation. Moreover, CHO cells produce glycans for humans that do not occur in human cells and are for this reason potentially immunogenic for patients: N-glycolylneuraminic acid (Neu5Gc) and Gal- $\alpha$ 1,3-Gal ( $\alpha$ -gal), even though these occurring at very low levels, <2% and <0.2% respectively [33].

**Murine myeloma cells**: NS0 and SP2/0 cells are used to produce mAbs. They exhibit small amounts of glycoforms, with additional  $\alpha$ -1,3-Gal and N-glycolylneuraminic acid in considerably high levels, that result to be immunogenic in humans [32]. The case of Cetuximab is explanatory of the danger of this kind of residues (Chapter 3, Paragraph 3.2.1).

**BHK**: Baby Hamster kidney fibroblasts are mostly been used for the production of veterinary viral vaccines. Only two marketed recombinant glycoproteins are currently manufactured in these cells, Factor VIIa and Factor VIII, which are clotting factors [34].

- **Human cell lines**

Human embryonic kidney cells or **HEK 293 and HT-1080** of fibrosarcoma origin are the most used human cell line to produce research-proteins for many years. Human cell-based expression systems produce recombinant proteins with glycosylation more similar to their natural counterpart [34].

**PER.C6 cell line** rFVIIIFc; Biogen, Cambridge, MA was created from human embryonic retinal cells, immortalized via transfection with the adenovirus E1 gene. This system was originally developed for the production of human adenovirus vectors for use in vaccine development and gene therapy. An advantage of PER.C6 is their ability to produce a high level of protein when used in the production of human IgG (Immunoglobulin) [35].

- **Hybridoma cells** are the second most frequently used hosts for recombinant glycoprotein production [36].
- **Transgenic animals:** in transgenic goat milk it is produced human anti-thrombin a, approved in Europe and the US in 2006 and 2009, respectively. The glycosylation profile changes significantly depending on the transgenic host used to express the proteins (goat, rabbit, chicken) [34].
- **Insect cells** infected with the viral vector baculovirus are mostly used for the development of virus-like particles and, as a consequence, vaccines. However, they lack of either galactosyltransferase and/or sialyltransferase activity, and even though they produce N-glycan precursors, these result in either high mannose or paucimannose residues that do not develop further into terminal galactose and/or sialic acid residues [35].
- **Plants** expression systems are characterized for the absence of sialylation and the presence of potentially immunogenic  $\alpha$ 1,3-fucose and  $\beta$ 1,2-xylose residues in plant-derived glycoproteins glycoforms. There are currently a few plant-derived therapeutics approved in Europe for topical use in human and others are in clinical trials, including glucocerebrosidase manufactured in carrot cells [34].
- **Yeast** high mannose-type N-glycans contain from five to nine mannose residues and are found in yeasts [32] which may confer a short half-life and render proteins less efficacious and even immunogenic in humans
- **Bacterial expression systems** are convenient for the rapid cell growth and high yields. However, they are not able to glycosylate protein because they don't possess the required enzymes and cellular compartments. They are therefore generally used for production of non-glycosylated proteins, including some mAbs, hormones, cytokines and enzymes. Another problem is that mammalian proteins produced in bacteria often are not soluble, resulting in the formation of inclusion bodies and thus they need to be extracted from inclusion bodies and refolded [35]. The Swiss company GlycoVaxyn, now part of GSK, has designed a bioconjugation platform which enables the *in vivo* synthesis of protein-polysaccharide complexes in *E. coli*, via recombinant DNA technology.

All these systems have their advantages and disadvantages, and above all, they present very different glycosylation pathways (*Fig 1.15*).

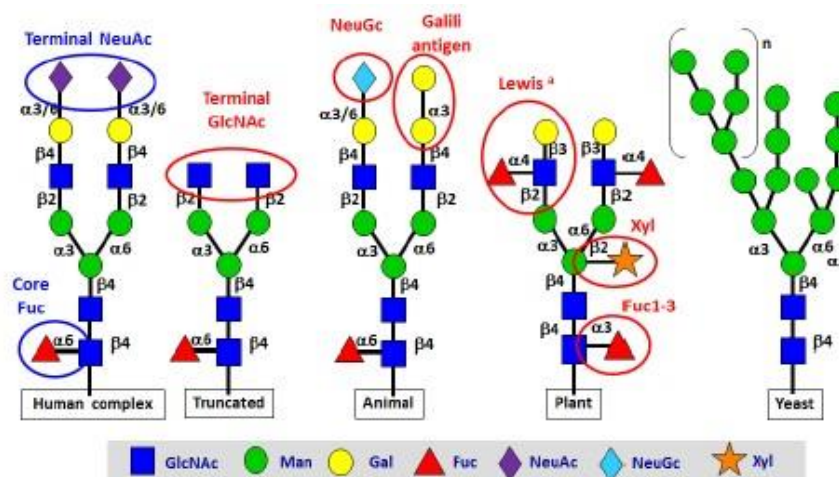


Fig. 1. 15 Different glycosylation according to the kind of expression system.

Hence, the choice of the production system influence the glycosylation pathway of the biotherapeutics. Being glycosylation non-template driven, the risk of glycan heterogeneity is high and it is connected not only to the production system, but also to the culture conditions. Microheterogeneity is defined as the variation in the structure of the attached glycans, while macroheterogeneity is the variability of the glycosylation sites and glycans number [37].

Genetic engineering is one way to introduce the required type of glycosylation. CHO cells were transfected with the gene for  $\alpha$ 2-6-sialyltransferase in order to increase the sialylation of various biologics, making them more human-like [38]. Another notable example of glycoengineering is human erythropoietin on which two additional N-linked glycosylation sites were introduced, allowing for a threefold increase in half-life and higher potency with a reduced frequency of administration in patients [39]. In recent years, there have also been developments in the use of transgenic insect cells, engineered to mimic mammalian cell glycosylation of expressed proteins [40].

Process optimization is the other approach to introduce the wanted glycan profile on biotherapeutics: media composition (glycosylation substrate feeding or media supplements) and process optimization (pH, temperature, and osmolality) can be employed to increase sialic acid content [34]. One method that has been investigated is supplementation of the media with 25 mM glucose that enabled production of chimeric mAbs exhibiting greater homogeneity of the protein glycoforms [41] to mitigate the toxic effects of ammonia generated by cellular respiration, the media can be supplemented with proline, threonine, and glycine resulting in increased sialic acid content of recombinantly expressed tissue plasminogen activator, used to treat stroke patients [42]. Metals such as iron have been added to cell culture media and

were all shown to alter glycosylation levels of diverse proteins such as interferon- $\gamma$  in CHO [43].

#### 1.6.4 Biotherapeutics regulation

Having discussed the importance of glycosylation at a molecular level, now its impact on the marketing of biotherapeutics will be explored. Throughout the stages of drug development, all the data regarding the production process and its safety and efficacy are reviewed by regulatory agencies [36]. Glycosylation is considered a **critical quality attribute (CQA)** that according to ICH Q8 [44] is a physical, chemical, biological, or microbiological property or characteristic that should be within an appropriate limit, range, or distribution to ensure the desired product quality. **ICH** is the International Council of Harmonization of Technical Requirements for Pharmaceuticals for Human Use, born in 1990, whose aim is to achieve an amity worldwide on the regulations to be applied in order to have safer, more effective and high quality medicines. Regulatory standards for biotechnology-derived pharmaceuticals are generally harmonized among the European Union, Japan and United States. The ICH Q6B [45] is the one taking care of glycosylation, in the specific section 2.1.1 “Physicochemical properties”. It is stated that the composition, physical properties and structure of a product have to be determined. It is admitted a degree of heterogeneity but this must be consistent with the one of the batches used in preclinical and clinical studies and it doesn't have to be detrimental for the safety and efficacy of the product.

Therefore, the degree of heterogeneity and the glycosylation profile of biotherapeutics has to be clearly defined during the research and clinical trials and has to be respected during the upscale of the process and the following production phases. Any alteration from the established glycosylation pattern has to be evaluated. The case of the company Genzyme is an example of the importance of the glycosylation profile of biotherapeutics. The company developed the drug Myozyme for the treatment of the Pompe disease, a glycogen storage disorder, made in relatively small 160-liter batches in the United States, that was only commercially available to children and infants. The biotechnology company in 2009 had asked the U.S. Food and Drug Administration for permission to make Myozyme in 2000-liter batches, in order to ease supply restraints and make the product available to adult patients, but the FDA delayed the approval of the medicine on large scale production for a difference in the glycans composition, causing a loss in the revenue of the company. The FDA determined that since the drug coming out of the 2000-liter batches was chemically slightly different than Myozyme, it needed to be approved as a separate medicine under a separate name, now marketed as Lumizyme [46].

The glycosylation control becomes even more important when it comes to biosimilars. A biosimilar is a biologic which is highly similar to the original medicine already approved, in terms of safety, purity and efficacy [25]. It is the equivalent of a generic drug for small molecules and the first biosimilar that has been put on the market was infliximab biosimilar in 2013 [31]. The expiring of the patents in the next 20 years will contribute to the rise of biosimilars [25]. Glycosylation is critical in the development of biosimilars because it has to be in accordance with the one of the original medicine, therefore quality control will play a vital role in the European biologics market.

Production of drugs involves a highly-controlled environment constrained by strict regulations and standards, and compared to chemical synthesis of small molecules, bioproduction of therapeutics includes many steps due to the complexity and the biological origins of these drugs. As such, cost of goods is higher for biologics than for small molecules [31]. The quality control is one of these steps to ensure that the end product has the desired properties. The widespread approach for drug quality control testing and approval is the so-called **quality by testing** (QbT). In QbT at the end of each batch, the product is tested for compliance with the desired quality. The QbT approach uncouples product end quality from the manufacturing process because doesn't require the knowledge that put in relationship the bioprocess inputs with the product quality outputs. On the contrary, the more recent idea is that quality must be built into the product. This is the main goal of the other approach, the **quality by design** (QbD) that means to build quality at every stage of process development [36]. This concept was developed by Dr. Joseph M. Juran who believed that most quality problems are related to the way a product is designed in the first place. The QbD consists of the identification of CQA, product and process design, control for each step of the manufacturing process and process continuous improvement [47].

The current methods of quality control of biotherapeutics glycosylation make difficult to have an in process quality control as required by the QbD approach. The tools most commonly used for glycan characterization include HPLC and CE, followed MS. MALDI-TOF/MS analysis gives a reasonably good glycoform profile rapidly enough, although getting actual percentage results takes days. NMR imaging can also be used for glycans; but the drawback is that it requires reasonably pure glycan structure in relatively large quantity. When looking at the glycoproteins, it is important to determine glycosylation site occupancy as well as their overall intact protein profiles. For example, CE with sodium-dodecyl sulphate gel (CE-SDS) is readily used to determine the degree of occupancy at Asn 297 site for an IgG1 but for more complex glycoproteins, currently available methods are not yet optimal for quantitative purposes. Using peptide mapping, it can be seen whether a site is occupied, but that is difficult to quantify even using liquid chromatography-mass spectroscopy (LC-MS) [48].

## **1.7 Research aim**

It would be ideal to have rapid tool allowing to examine more samples in-process with the aim of controlling and potentially optimising conditions in real time. The ideal tool would be affordable for the manufacturers, cost-effective, fast and easy-to use. The techniques described above give highly detailed carbohydrate structures but they present different drawbacks. First of all is the “human factor”: they require knowledge, experience and training to be carried out correctly and to have reliable and reproducible results. Not all scientist have the skills to carry out experiments using a mass spectrometer. In addition those machines are very expensive and not all laboratories have one. Another factor is the accuracy: if carbohydrates need to be released for the analysis, it is very unlikely that they will be released at 100%, it is usually around 85%. Especially when it comes to O-linked glycans, because the methods of releasing are not as efficient as for N-linked glycans [48]. Another important point to consider is that the end product is a glycoprotein, so it would be in the best interest to look at the entire protein with the glycans still on it. Moreover, these techniques tend to be time-consuming. The development of LEctPROFILE kits is meant to provide a convenient and complementary methodology to satisfy the requirements of the QbD process. First of all, many lectins show promiscuous binding: one of the goal for the development of the kits is the purification of highly-specific lectins for the recognition of unwanted glycan structures on biotherapeutics. It can be carried out by non-specialized researcher: it is not needed a printing machine to spot the lectins, because it is in a 96-well plate format. Everything can be done manually with the basic tools that every laboratory has and without a specific training for the users. There is no need to remove glycans but the glycoprotein can stay in its natural condition. The glycan-containing sample has to be labelled prior to the analysis only if the direct-binding mode is chosen (*Fig. 1.16 A*). To avoid this additional step and above all to avoid adding unwanted variability due to the labelling efficiency, a secondary antibody can be used. Even better, the assay can be performed in a inhibition mode(*Fig. 1.16 B*): the sample competes with a previously labelled neoglycoprotein that has the same specificity of the sample for the lectin immobilized on the plate. Another advantage is that is a high-throughput analysis, and the plate can be designed according to the user’s needs, customizing the plate according to the number of samples. It is also a very fast method that in 2 hours give the desired results, using for the detection a plate reader that is found in the majority of lab facilities. Results analysis is also taken into consideration during the development: the user will only need to insert the raw data and the output will be given thanks to a predesigned data sheet.



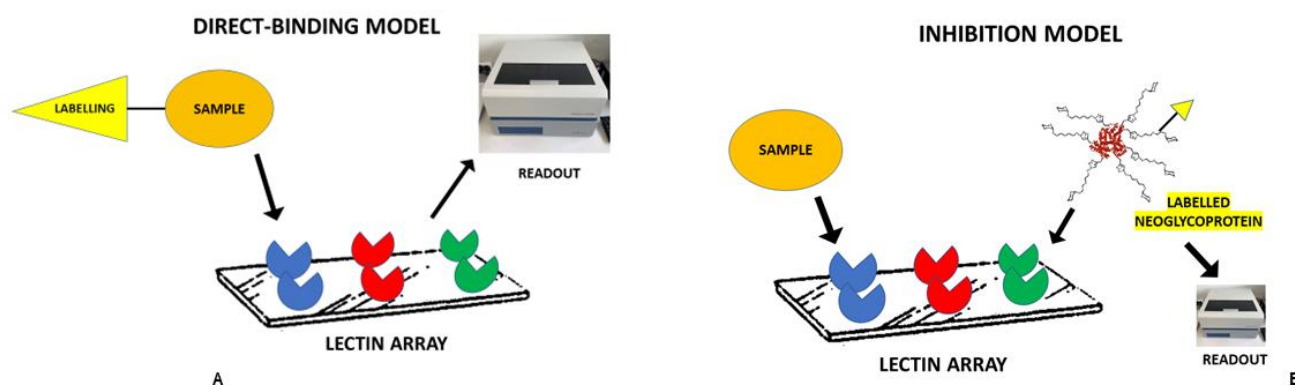


Fig. 1. 16 A GLYcoPROFILE in direct binding. B GLYcoPROFILE in inhibition.

### 1.7.1 Objectif de la recherche

L'idéal serait de disposer d'un outil rapide permettant d'examiner un plus grand nombre d'échantillons en cours de processus dans le but de contrôler et éventuellement d'optimiser les conditions en temps réel. L'outil idéal serait abordable pour les fabricants, rentable, rapide et facile à utiliser. Les techniques décrites ci-dessus donnent des structures d'hydrates de carbone très détaillées, mais elles présentent différents inconvénients. Le premier est le "facteur humain" : elles nécessitent des connaissances, de l'expérience et une formation pour être réalisées correctement et obtenir des résultats fiables et reproductibles. Tous les scientifiques n'ont pas les compétences nécessaires pour réaliser des expériences à l'aide d'un spectromètre de masse. En outre, ces machines sont très coûteuses et tous les laboratoires n'en possèdent pas. Un autre facteur est la précision : si les hydrates de carbone doivent être libérés pour l'analyse, il est très peu probable qu'ils soient libérés à 100 %, ils le sont généralement à environ 85 %. Surtout lorsqu'il s'agit de glycanes O-liés, car les méthodes de libération ne sont pas aussi efficaces que pour les glycanes N-liés [48]. Un autre point important à prendre en compte est que le produit final est une glycoprotéine, il serait donc dans le meilleur intérêt d'examiner la protéine entière avec les glycanes encore présents. En outre, ces techniques ont tendance à prendre beaucoup de temps. Le développement des kits LEctPROFILE a pour but de fournir une méthodologie pratique et complémentaire pour répondre aux exigences du processus QbD. Tout d'abord, de nombreuses lectines présentent une liaison promiscuous : l'un des objectifs du développement des kits est la purification de lectines hautement spécifiques pour la reconnaissance des structures glycaniques indésirables sur les produits biopharmaceutiques. Il peut être réalisé par un chercheur non spécialisé : il n'est pas nécessaire d'avoir une machine à imprimer pour repérer les lectines, car il s'agit d'un format de plaque à 96 puits. Tout peut être fait manuellement avec les outils de base dont dispose tout laboratoire et sans formation spécifique pour les utilisateurs. Il n'est

pas nécessaire d'éliminer les glycanes mais la glycoprotéine peut rester dans son état naturel. L'échantillon contenant des glycanes doit être marqué avant l'analyse uniquement si le mode de liaison directe est choisi (*Fig. 1.16 A*). Pour éviter cette étape supplémentaire et surtout pour éviter d'ajouter une variabilité indésirable due à l'efficacité du marquage, un anticorps secondaire peut être utilisé. Mieux encore, le test peut être réalisé en mode inhibition (*Fig. 1.16 B*) : l'échantillon entre en compétition avec une néoglycoprotéine préalablement marquée qui a la même spécificité que l'échantillon pour la lectine immobilisée sur la plaque. Un autre avantage est qu'il s'agit d'une analyse à haut débit, et que la plaque peut être conçue selon les besoins de l'utilisateur, en personnalisant la plaque en fonction du nombre d'échantillons. C'est également une méthode très rapide qui permet d'obtenir les résultats souhaités en 2 heures, en utilisant pour la détection un lecteur de plaques que l'on trouve dans la majorité des laboratoires. L'analyse des résultats est également prise en compte lors du développement : l'utilisateur n'aura qu'à insérer les données brutes et la sortie sera donnée grâce à une feuille de données prédéfinie.

## **Chapter 2 – Neoglycoproteins**

## **Abstract**

Neoglycoproteins are multivalent macromolecules based on a protein backbone on which glycans are linked. They are used for the study of protein-glycan interactions with interesting potential applications for human health [49] (imaging, pharmacology). We want to exploit the neoglycoproteins multivalency and high affinity to the target lectins (MAAI, SNA, MOA $\beta$ T, GSLIB4, HPyL, WGA) in order to define the characteristics of the LEctPROFILE kits that we intend to develop. To do so, we have established a protocol for batch production of five main neoglycoproteins of interest, and we have demonstrated their purity, activity and specificity towards the target lectins. This work puts the basis for the following development of the LEctPROFILE kits. At the end of the chapter, it is also included a paper in collaboration with Dr. David Goyard and Professor Olivier Renaudet of the University of Grenoble-Alpes, in which we have explored multimeric glycodendrimers and the synthesis of multispecific neoglycoproteins.

## **Résumé**

Les néoglycoprotéines sont des macromolécules multivalentes. La synthèse de ces molécules est large et variées. Elles sont utilisées notamment pour l'étude des interactions protéines-glycanes avec des applications potentielles intéressantes pour la santé humaine. Nous souhaitons exploiter la multivalence et la haute affinité des néoglycoprotéines pour définir les caractéristiques des kits LEctPROFILE que nous comptons développer. Pour ce faire, nous avons établi un protocole de production de cinq néoglycoprotéines d'intérêt, et nous avons démontré leur pureté, leur activité et leur spécificité vis-à-vis des lectines cibles. Ce travail participe à mettre en place les bases du développement ultérieur des kits LEctPROFILE qui seront abordés dans les différents chapitre de cette thèse. En fin de chapitre est également inclus un article en collaboration avec le Dr David Goyard et le Professeur Olivier Reanudet de l'Université de Grenoble-Alpes, dans lequel nous avons exploré l'utilisant de glycodendrimères multivalents ainsi que la synthèse et rôle de néoglycoprotéines multispécifiques comme ligands des lectines bactériennes.

## 2.1 Multivalency: a powerful tool

As introduced in Chapter 1, Paragraph 1.1, lectins are involved in many biological functions such as immunity, cell adhesion, and trafficking, thanks to their ability to cross-link glycans through multivalent interactions. A growing body of research has been directed into replicating lectin-glycan interactions to understand better their mechanism and to develop new drugs and new detection systems. **Multivalent interactions** are defined as “specific simultaneous associations of multiple ligands (or epitopes) present on a molecular construct or biological surface, that binds to multiple receptors expressed on a complementary entity” [50]. Therefore, given that it is known that multivalency occurs in nature, the field of synthetic glycobiology has been studying new ways to design new molecules with increased multivalency. One approach can be lectins engineering: recent examples are **Neolelectins** like NeoRSLVI, consisting of a monomeric lectin with a  $\beta$ -propeller architecture and six active binding sites (Fig. 2.1A) [51]. Another example are **Janus lectins**, the first one of which is a bispecific lectin with two faces, one recognizing fucosylated glycoconjugates and the other one sialylated ones, with avidities in the nanomolar range thanks to the multivalent presentation of the binding sites (Fig. 2.1B) [52].

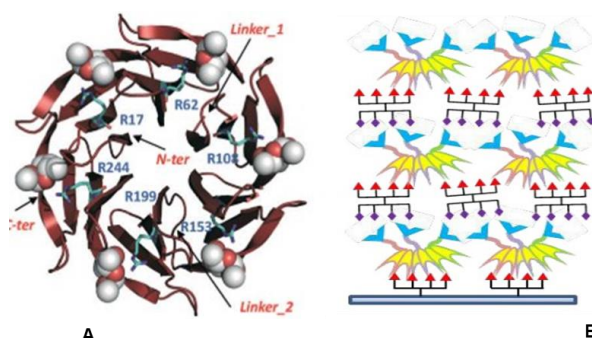


Fig. 2. 1 **A** structure of the neoRSLVI in complex with  $\alpha$ Methylfucose (spheres), all binding sites result occupied, adapted from [51]. **B** Cartoon representation of a Janus lectin, able to cross-link biological surfaces thanks to its double specificity for fucosylated glycans on one side and sialylated ones on the other side, adapted from [52].

Another approach that allows to study protein-glycan interactions is from the glycan's point of view. In nature, the affinity is increased when glycans are presented linked to other structures, such as glycolipids or glycoproteins. In Fig. 2.2, are represented the four different mechanisms of interaction of multivalent partners [8] [53]:

**a) The chelate binding.** The  $n$  binding sites of a homomultimeric lectin can interact with  $m$  ligands from a multi-glycosylated molecule on the cell surface.

**b) The receptor clustering.** It happens when monovalent lectins or ligands are anchored to the cell membrane. For example, in the presence of a multivalent ligand, receptors can diffuse through the lipid bilayer and become clustered by the multivalent ligand.

**c) The subsite binding.** Some proteins in addition to the first binding site possess a second one that can be occupied by a multivalent glycan with different affinity compared to the first binding site.

**d) Statistical reassociation.** A monovalent lectin could improve the affinity interacting with multivalent glycans present in close proximity of the binding site and in high concentration.

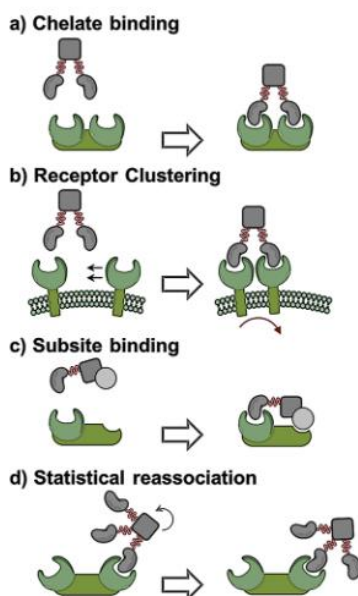


Fig. 2. 2 Mechanisms of interaction of multivalent partners, adapted from[8]

The **glycoside cluster effect** occurs when multivalency is achieved on both sides, the one of the protein and the one of the glycan [9]. However, multivalent glycans in nature are structurally complex and thus not suitable to study physiological mechanisms. Therefore, another way to increase the avidity and consequently the affinity, is to chemically synthesise multivalent glycans choosing the type of scaffold, playing with their size, the density of ligand, the orientation and the flexibility. The general idea is to have a multivalent glycoconjugate that has one glycan specificity displayed in a way that results in a strong multivalency.

## 2.2 Different types of multivalent synthetic glycoconjugates

Different types of multivalent glycoconjugates have been designed with different valencies and topologies (Fig 2.3).

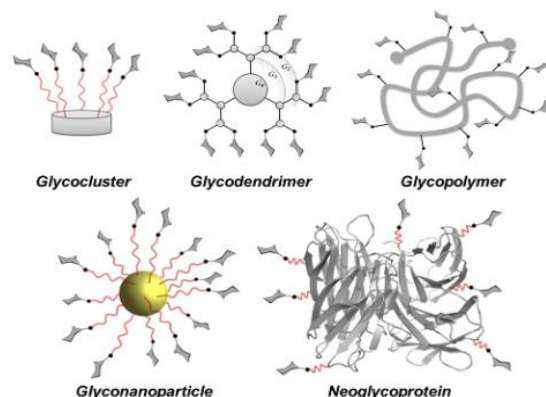


Fig. 2. 3 Types of multivalent synthetic glycoconjugates, adapted from [8]

- **Glycoclusters** usually contain a limited number of linked glycoconjugates of which the orientation and position are synthetically controlled [8].
- **Glycodendrimers** are macromolecules completely built around a core unit and completely branched throughout their structures. Also in this case, the valency and the terminal functional groups are well-controlled [3].
- **Glycopolymers** are wholly synthetic multivalent or polyvalent cluster glycosides [3]. For example, the polymer polyacrylates were used to incorporate 5-N-acetyl neuraminic acid in order to reveal the maximal inhibition of the influenza virus receptor-binding activity [54].
- **Glyconanoparticles** are made of magnetic materials tethered with specific glycans. The most used are made of iron oxide and gold for their nontoxic and biodegradable properties [8]. Reynolds *et al.* [55] showed that gold nanoparticles exhibiting Gal ligands in a multivalent fashion had a 3000-fold increase in the affinity of the lectin PA-IL, from *Pseudomonas aeruginosa*, over the monovalent ligand.

- **Neoglycoproteins** defined as a glycoproteins produced by the covalent attachment of mono-, oligosaccharides, or dendrimers to proteins that may or may not already contain naturally occurring, covalently bound carbohydrate [56]. The main difference with a glycoprotein is that of course, the linkage with the glycans is not natural. In addition the occupancy and glycan density are controlled, as well as the homogeneity of glycosylation.

## 2.3 Neoglycoproteins

### 2.3.1 Applications

Extensive research has shown that neoglycoproteins are suitable for many practical applications other than the study of biological interactions. Applications for imaging include human serum albumin modified with  $^{99m}\text{Tc}$ -galactose for radiolabelling. It is exploited in order to estimate the hepatocyte mass and function by imaging the expression of the asialoglycoprotein receptor in the liver [49].

Neoglycoproteins have been also used successfully as drug carriers in order to selectively release drugs inside cells. Methotrexate, which acts as a metabolic inhibitor within the cells, is a chemotherapeutic drug for the treatment of leukemia. It was synthesized a fucosyl serum albumin conjugate to methotrexate neoglycoprotein which was more cytotoxic than free methotrexate, indicating that methotrexate bound to the neoglycoprotein is more actively internalized than the free drug [49].

Pathogenic infections are often mediated by an initial glycan-protein interaction, as it happens for example, for the recognition of sialic acid on human cells by the influenza hemagglutinin. Since neoglycoproteins can present multiple glycans epitopes, their potential immunogenic nature have been used to design vaccines against viruses, bacteria, parasites. One case is the one of the gram negative bacterium *Neisseria meningitidis* responsible of the epidemic meningitis in Sub-Saharan Africa. Human Serum Albumin conjugates of synthetic fragments of bacterium capsular polysaccharide of this bacterium were synthesized and it was showed that even small as monosaccharide fragments was recognized by polyclonal Meningococcus A antiserum [57].



### 2.3.2 Synthesis

The first method for preparing neoglycoproteins was developed in 1929 by Goebel and Avery, who attached a variety of mono-and oligo-saccharides to ovalbumin and horse serum-globulin [58].

A number of criteria must be considered in developing the protocol for the chemical synthesis of neoglycoproteins:

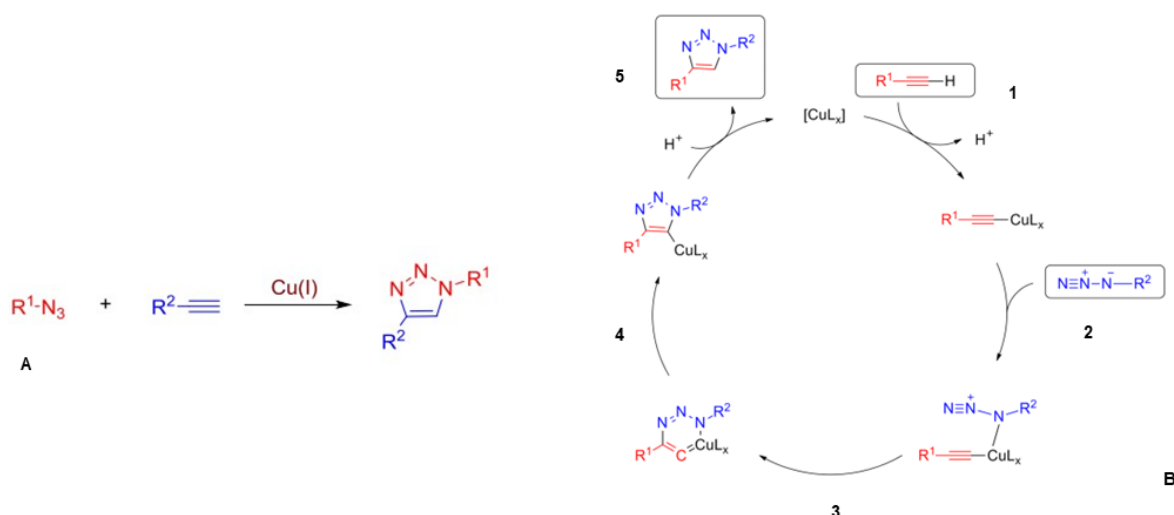
- Reproducibility is one of the main factors, therefore the conditions of preparation of glycans need to be stable and applicable to a great variety of glycans. It is preferable if the reaction is rapid and works under mild conditions. The modification should be specific, preferably for one functional group. and stable to physiological conditions.
- The protein chosen as a scaffold, it should maintain its native structure and its biological activity, again underlining the importance of keeping mild conditions for the reaction. If the modified protein needs to be used *in vivo*, the linkage should be resistant to metabolic and enzymic degradation. The modified protein should be easy to purify from the reaction mixture, and the modification must be analysed and quantified.
- The linkage between carbohydrate and protein should not introduce long-chain hydrocarbons, aromatic rings, or ionic species, as such groups as these may interfere with the activity of the protein, or interact non-specifically in biological systems [56].

There are many ways to synthesise neoglycoproteins, for example the conjugation to cysteine with terminal thiol, conjugation to aspartic or glutamic acid through their carboxylic group [1]. The conjugation to lysine with terminal primary amine is one of the most used. The most common way to do this is through reductive amination. In this type of reaction, the protein is solubilized in borate buffer at pH around 8 or 9 with a large excess of glycans. The reducing end of the target glycan forms a Schiff's base and react with the lysins  $\epsilon$ -amino group on the protein. Since Schiff's base is not a stable product, by adding cyanoborohydride ( $\text{NaBH}_3\text{CN}$ ) in solution, the equilibrium of the reaction is pushed towards the formation of the product [59]. Another more versatile strategy, is to use a linker for the modification of lysins on the protein scaffold. The linker can be of different length and it can present different active groups: isothiocyanate, 2-imino-2-methoxyethyl -1-thioglycoside, carbodiimide-activated carboxylic acid. All these groups are activated in order to react with primary amine, while, on the other side, the glycan moiety remains intact [60].

In this thesis it is explored only the synthesis of neoglycoprotein by click-chemistry after modification of the lysine amino acid with a propargyl or azido linker.

## 2.3.3 Synthesis by click chemistry

The **CuAAC click-chemistry reaction** (copper-catalysed azide-alkyne cycloaddition), was discovered in 2002 by two independent groups [61] [62], and, since then, it has been improved and now it has become an excellent tool in organic synthesis. This reaction occurs between two molecular entities that are functionalized by azido and alkyne groups, giving as product the 1,4-disubstituted-1,2,3-triazole molecules (*Fig.2.4 A*). It is an advantageous compound because it is chemically inert against oxidation, reduction and hydrolysis under acidic or basic conditions. The advantage of the addition of catalytic amount of copper is that the reaction goes much faster, taking few hours, and can be performed at room temperature. Additionally, this reaction is very convenient because it does not require any products separation since it has a very high regioselectivity being the 1,4- disubstituted isomer the only product formed (*Fig. 2.4, A, B*).



*Fig. 2. 4 A* CuAAC via 1,3-dipolar cycloaddition of azides and alkynes reaction resulting in 1,2,3-triazoles, adapted from [63]. *B* CuAAC reaction mechanism (adapted from, <http://bitly.ws/uSiZ>)

It can be conducted in **aqueous conditions** or in mixtures of water and organic solvents such as dimethyl sulfoxide (**DMSO**). In order for the reaction to work efficiently it is better to use a mixture of **Cu(II)** like CuSO<sub>4</sub> because Cu(I) salts oxidize in aqueous environment. The addition of sodium ascorbate produces Cu(I) in situ. **Ascorbic acid** in water oxidize to dehydroascorbate and reduce Cu(II) to Cu(I). As Cu(I) is unstable in aqueous solvents the

stabilizing ligand Tris((1-benzyl-4-triazolyl)methyl)amine (**TBTA**) is used to improve the reaction outcome.

## 2.4 Results and discussion

In the recent years, the synthesis of neoglycoproteins has been a focal point for GLYcoDiag's research and development. The previous Glycodiag's PhD student, Dr. Blanka Didak, mainly worked on neoglycoproteins as strong multivalent ligands for C-type lectins (especially DC-SIGN and Langerin), by preparing a large variety of multimeric mannosylated neoglycoproteins and neoglycoclusters. During the course of my thesis, the neoglycoproteins synthesis was optimised with the purpose of having the right probes for the development of the LEctPROFILE kits. The neoglycoproteins are, in this case, addressed towards an application; they work as model macromolecules that, thanks to their high avidity and strong specificity for the target lectin, help to design the assay in order to have all the information for the analysis of an unknown sample. Thus, in this section are presented the results that are the prerequisites for the implementation of the Galili antigen LEctPROFILE kit, the  $\alpha$ 2-6/ $\alpha$ 2-3 LEctPROFILE kit and the Neu5Ac/Neu5Gc LEctPROFILE kit.

For the synthesis of neoglycoproteins used for this thesis, the scaffold used is the protein Bovine Serum Albumin (**BSA**). It is a protein of 66.5 kDa and consists of three homologous domains, conferring a heart-shaped molecule. It is soluble in neutral and alkaline medium and it is big enough to allow a heavy sugar substitution [49]. It is cheap, enabling to produce large batches of neoglycoproteins and it is not glycosylated so there is no interference in the glycans of interest-lectin interactions.

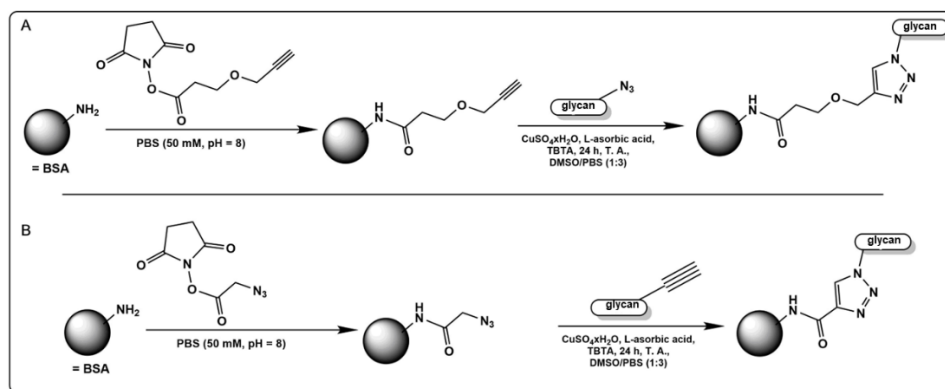
The neoglycoproteins synthesis is carried out following two main steps:

1. The first step is the functionalization of BSA with Propargyl-N-hydroxysuccinimidyl ester (PNHS) obtaining BSA-alkyne (*Fig. 2.5 A*). In fact, PNHS react with primary amines ( $-NH_2$ ) of lysine. This step was optimized by Dr. Blanka Didak who obtained a ratio of 35-40 alkyne/BSA, which is very good, considering that on BSA there are around 60 lysins, leaving 20 lysins for additional modifications such as labelling with biotin.

The "reverse" protocol to obtain BSA-azide was optimized during the course of my thesis thanks to Dr. Benoît Roubinet. This functionalization is carried out again via the primary amine groups ( $-NH_2$ ) of BSA reacting with the N-hydroxysuccinimide ester (NHS) of the reagent 2-azidoacetic acid NHS ester (*Fig. 2.5 B*). In this case, a ratio of

20-25 azide groups/BSA is expected. The reaction product is purified on Sephadex G-25 and the ratio is verified by MALDI-TOF/MS. This protocols are optimized to obtain around 100 mg of functionalized protein.

- The second step is the actual CuAAC reaction (*Fig. 2.5 A, B – right part*) between the BSA-alkyne or BSA-azide prepared previously and the alkyne or azide group of a simple or complex glycan. The number of glycans conjugated on BSA is determined by MALDI-TOF/MS analysis. Further steps for characterization of these glycoconjugates are SDS-PAGE, Bradford for protein quantification. After neoglycoprotein biotinylation, their specificity is evaluated through GLYcoPROFILE in direct binding. This protocol is optimized to obtain around 5 mg of neoglycoprotein.

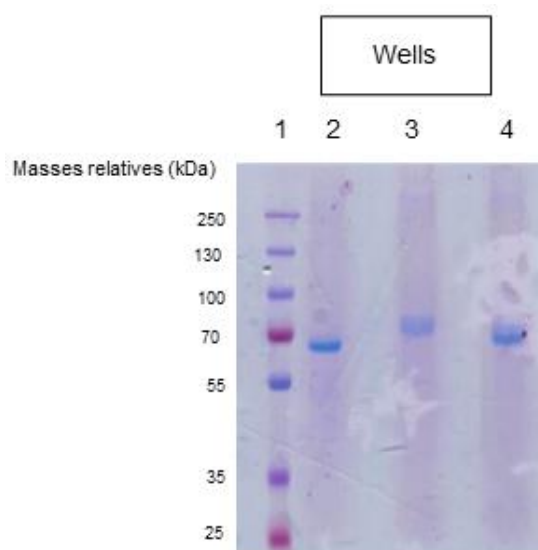


*Fig. 2. 5* **A** Reaction between the primary amines of the BSA and PNHS ester, followed by the click-chemistry reaction (CuAAC) between BSA-alkyne and the glycan functionalized with an azide group. **B** Reaction between the primary amines of the BSA and the NHS, followed by the click-chemistry reaction (CuAAC) between BSA-azide and the glycan functionalized with an alkyne group.

#### 2.4.1 Synthesis of BSA-alkyne and BSA-azide

We performed the synthesis of BSA containing alkyne groups and BSA containing azide groups, in order to have the right scaffolds to build the desired neoglycoproteins for the LEctPROFILE kits. For the synthesis of BSA-alkyne, the chemical reaction happens between the primary amines  $\text{NH}_2$  of BSA lysins and PNHS (*Fig. 2.5 A*). For the synthesis of BSA-azide, the chemical reaction happens between the primary amines  $\text{NH}_2$  of BSA lysins and NHS (*Fig. 2.5 B*). *Fig 2.6* shows the SDS-PAGE that was carried out at the end of the reactions, after the products purification by sephadex G-25. The most interesting aspect of this gel is that the band at line 3 representing the BSA-alkyne, and the band at line 4 representing the BSA-azide, clearly show an increase in the molecular weight compared the simple BSA at line 2.

This is the first confirmation that the reaction worked. The MALD-TOF spectra in *Fig 2.7 A, B* and *C* are the final proof of the success of the synthesis. On the top, the spectrum of BSA confirms the molecular weight of around 66 kDa of the pure protein. On the bottom left, in the spectrum of the BSA-alkyne we can see a single peak indicating the purity of the protein and the increased molecular weight of 70674 Da compared to the BSA 66 kDa. On the right, in the spectrum of the BSA-azide we can see a single peak indicating the purity of the protein and the increased molecular weight of 68206 Da. The molecular weight calculated in the MALD-TOF spectra for the two proteins is also in accordance with what we see on the SDS-PAGE (*Fig 2.6*). The calculated number of alkyne/BSA calculated from the MALD-TOF analysis is 37 and the calculated number of azide/BSA is 21, as expected.



*Fig. 2. 6* SDS-PAGE of BSA functionalization with alkyne and azide groups. The protein samples were analysed under denaturing conditions on 14% polyacrylamide gel. Line 1: Protein marker, Line 2: BSA, Line 3: BSA-alkyne, Line 4: BSA-azide.

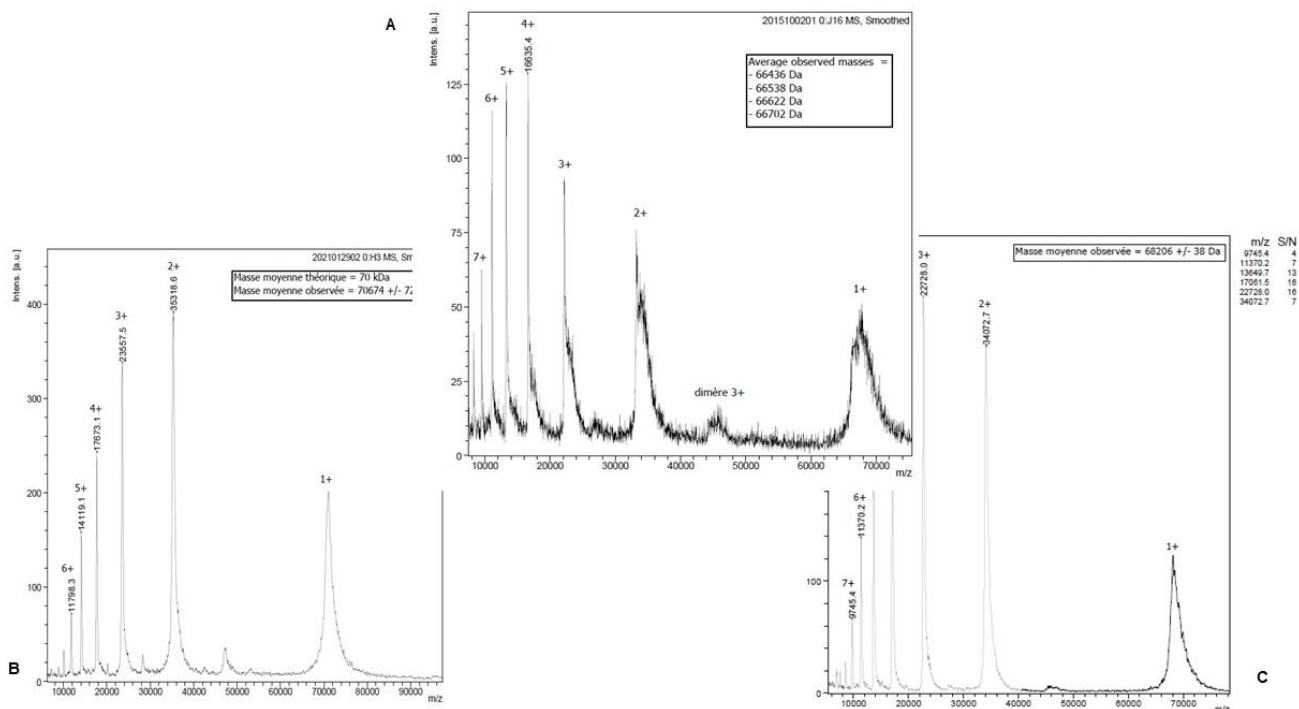


Fig. 2. 7 **A** MALD-TOF/MS spectrum of BSA **B** MALD-TOF/MS spectrum of BSA-alkyne **C** MALD-TOF spectrum of BSA-azide.

## 2.4.2 Synthesis of NeoGalili neoglycoprotein

We performed the present synthesis of the NeoGalili neoglycoprotein to have a highly specific probe for the development of the Galili antigen LEctPROFILE kit (Chapter 3). The reaction, represented in Fig. 2.8, takes place between the BSA-azide and the oligosaccharide Gal $\alpha$ 1-3Gal-N-Acetyl-Propargyl to give as product the BSA conjugated with Gal $\alpha$ 1-3Gal, also called NeoGalili neoglycoprotein.

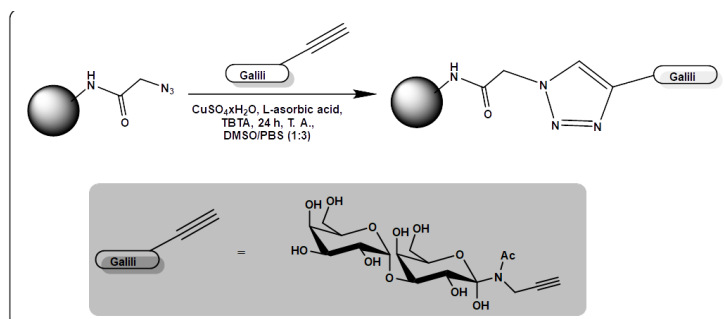


Fig. 2. 8 CuAAC reaction between BSA-azide and Gal $\alpha$ 1-3Gal-N-Acetyl-Propargyl.

After purification of the reaction product, the SDS-PAGE was performed. The gel shown in Fig 2.9 confirmed the conjugation of the Gal $\alpha$ 1-3Gal sugar moieties to the BSA was successful. In fact the NeoGalili neoglycoprotein loaded on line 5 had a molecular weight higher than the expected band of 70 kDa for the sole BSA-azide. The presence of the sugar moieties was further confirmed after PAS staining for glycoproteins, which resulted in a stronger violet colour in line 5 compared to the other bands.

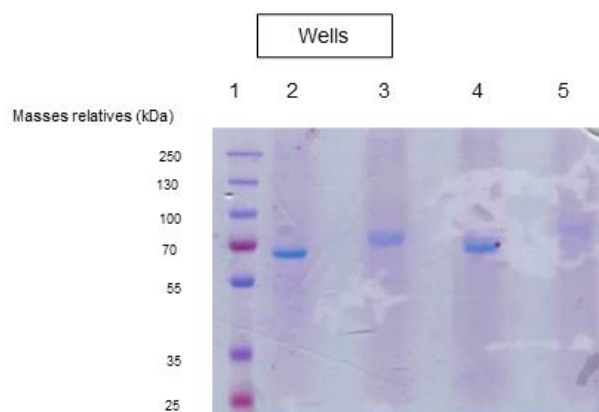


Fig. 2. 9 SDS-PAGE of NeoGalili synthesis. The protein samples were analysed under denaturing conditions on 14% polyacrylamide gel. Line 1: Protein marker, Line 2: BSA, Line 3: BSA-alkyne, Line 4: BSA-azide, Line5: NeoGalili neoglycoprotein.

An increase of the molecular weight to 74464 Da, in complete accordance with the SDS-PAGE, was further confirmed by MALDI-TOF/MS analysis (Fig 2.10), where the purity of the neoglycoprotein is also confirmed by the presence of a single peak. The number of sugar moieties/BSA calculated is 16.

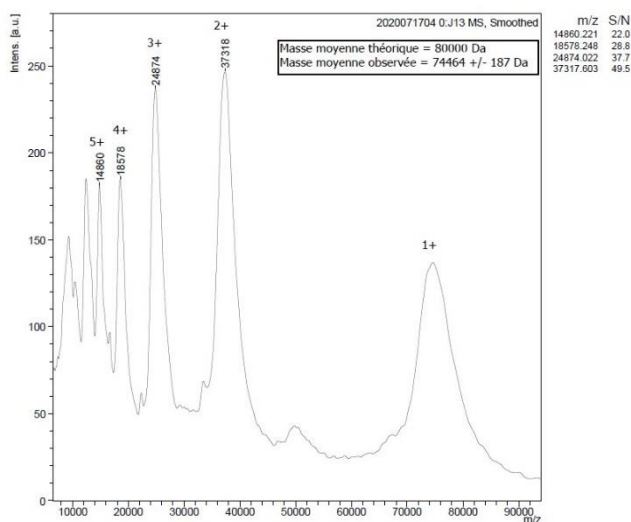
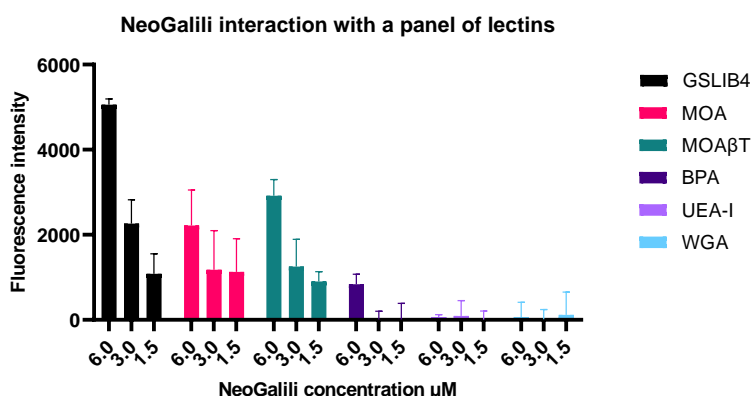


Fig. 2. 10 MALDI-TOF spectrum of NeoGalili neoglycoprotein.

In the final part of the characterization of the NeoGalili, we labelled the neoglycoprotein and we tested it through the GLYcoPROFILE in direct binding, to see the interaction with a panel of lectins (refer to Annex II, Chapter 4); as shown in the chart in *Fig 2.11*. The interaction was tested at three different concentrations of NeoGalili. The NeoGalili showed a strong interaction (fluorescence intensity = 5055) with the lectins GSLIB4, MOA and MOA $\beta$ T which are highly specific to the Galili antigen, confirming the selectivity of the lectins for this neoglycoprotein. The choice of BPA is to see if there is a cross-reaction of the neoglycoprotein with lectins specific for  $\beta$ -Gal. As expected, the NeoGalili showed no interactions (fluorescence intensity < 0) with UEA-I and WGA, the first one specific for fucose and the second one for Neu5Ac and GlcNac. From the chart, it can be seen that by far the highest NeoGalili interaction is with the lectin GSLIB4 and after with MOA $\beta$ T and the full MOA.



*Fig. 2. 11* GLYcoPROFILE in direct binding of the NeoGalili neoglycoprotein labelled. The neoglycoprotein at three different concentrations was added to the plate coated with six different lectins. The data shown are the average of experiments performed in triplicate.

Overall, these results indicate that the synthesis via click chemistry of the NeoGalili neoglycoprotein was successful and therefore this neoglycoprotein makes an optimal probe for the development of the Galili antigen LEctPROFILE kit.

#### 2.4.3 Synthesis of Neo6'SL and Neo3'SL neoglycoproteins

In order to synthesize the neoglycoproteins Neo6'SL and Neo3'SL, needed as probes for the presence of  $\alpha$ 2-3/  $\alpha$ 2-6 linkages within the LEctPROFILE kit, we used the BSA functionalized with the alkyne group to react with the azide group on the thrisaccharide 6'-sialylactose- $N_3$  and 3'-sialylactose- $N_3$  (*Fig. 2.12*).



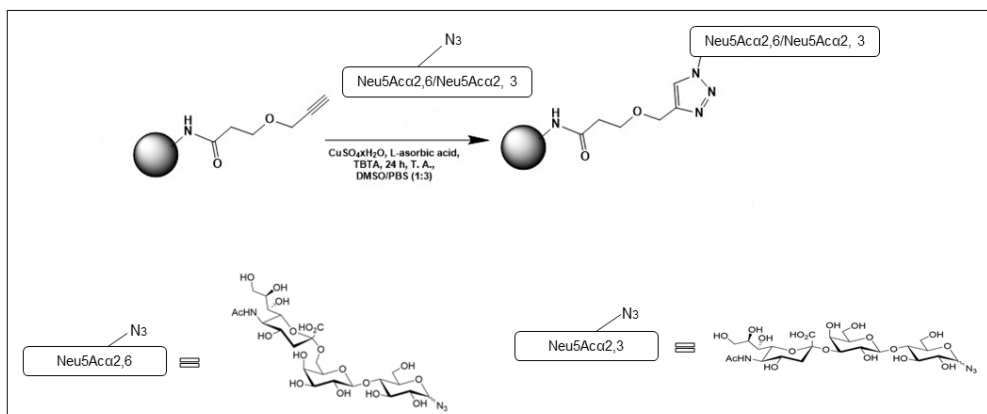


Fig. 2. 12 CuAAC reaction between BSA-alkyne and 6'-sialylactose-N3 or 3'-sialylactose-N3.

After purification of the reaction product, the SDS-PAGE presented in Fig 2. 13 confirmed that the reaction was successful, with a clear molecular weight shift of the Neo3'SL and Neo6'SL towards 70 kDa.

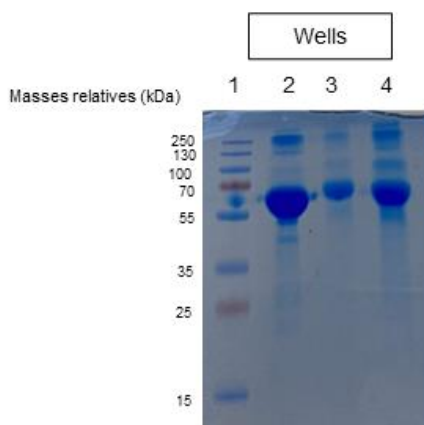


Fig. 2. 13 SDS-PAGE of Neo6'SL and Neo3'SL synthesis. The protein samples were analysed under denaturing conditions on 14% polyacrylamide gel. Line 1: Protein marker, Line 2: BSA-alkyne, Line 3: Neo6'SL, Line 4: Neo3'SL.

The result was further confirmed by the MALDI-TOF/MS spectra in Fig 2.14 A and B. The Neo6'SL spectrum (Fig. 2.14 A) shows a single peak of the pure neoglycoprotein at a molecular weight of 72185 Da, resulting from the Neo2,6 linked to 3 sugar moieties to the BSA-alkyne. In Fig. 2.14 B it is shown the spectrum of Neo3'SL, with the single peak at 72875 Da, meaning that we have added 4 sugar moieties to the BSA-alkyne. The number of glycans/BSA was lower compared to the Galili neoglycoprotein and may result from the nature of sialic acid linkages which tend have a low stability.

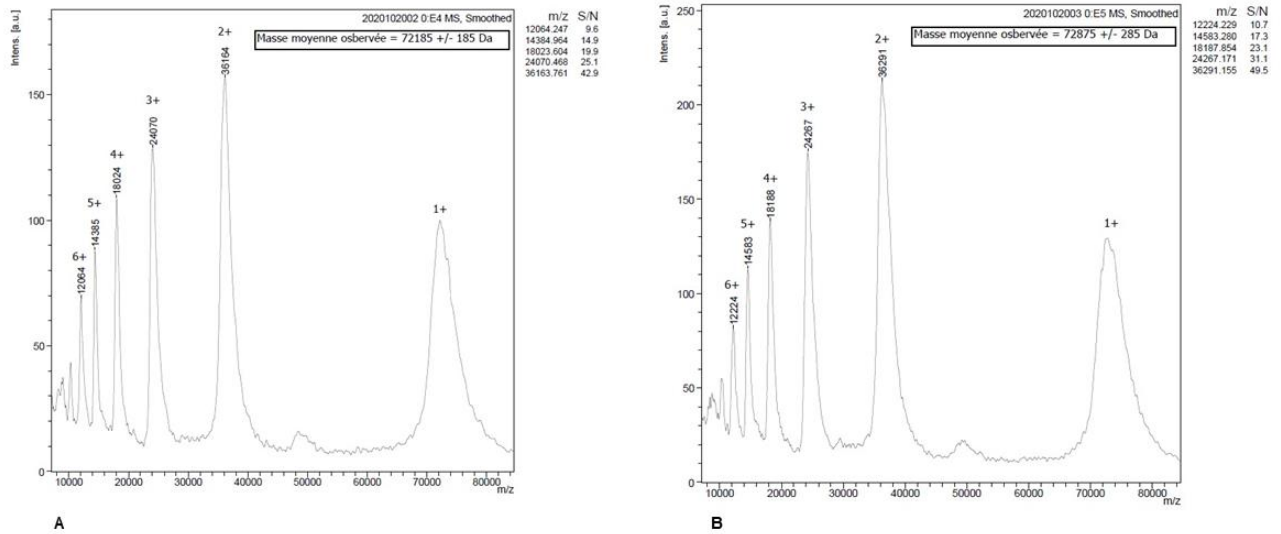


Fig. 2. 14 **A** MALD-TOF spectrum of Neo6'SL **B** MALD-TOF spectrum of Neo3'SL.

The results of the GLYcoPROFILE (Fig. 2.15A) indicate that Neo6'SL, tested at three different concentrations on a panel of lectins, is highly specific for SNA, while it does not show interaction with MAAI, PNA specific for lactose and PSA for mannose (Refer to Annex II, Chapter 4). Neo3'SL was also tested at three different concentrations on the same lectins as Neo6'SL and, in Fig.2.15 B it can be seen a strong specificity for MAAI, proving the reverse specificity of the two lectins SNA and MAAI that is central for the development of the  $\alpha$ 2-3/  $\alpha$ 2-6 LEctPROFILE kit. Neo3'SL, similarly to Neo6'SL, does not interact with PNA and PSA, which were used as negative control due to their specificity respectively for lactose and PSA for mannose.

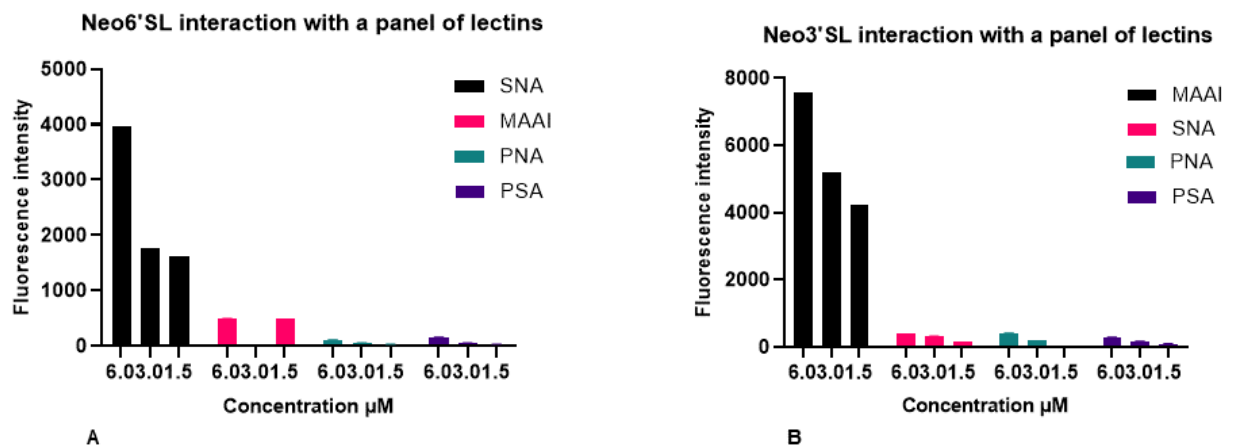


Fig. 2. 15 **A** GLYcoPROFILE in direct binding of the Neo6'SL neoglycoprotein labelled. The neoglycoprotein at three different concentrations was added to the plate coated with four different lectins. **B** GLYcoPROFILE in direct binding of the Neo3'SL neoglycoprotein labelled. The neoglycoprotein at three different concentrations was

added to the plate coated with four different lectins. The data shown are the average of experiments performed in triplicate.

#### 2.4.4 Synthesis of NeoNeuAc and NeoNeuGc neoglycoproteins

Lastly, the glycoprotein NeoNeuAc and NeoNeuGc, meant for the Neu5Ac/Neu5Gc LEctPROFILE kit, were synthesized from BSA-alkyne and propargyl Neu5Ac or propargyl Neu5Gc (Fig 2.16).

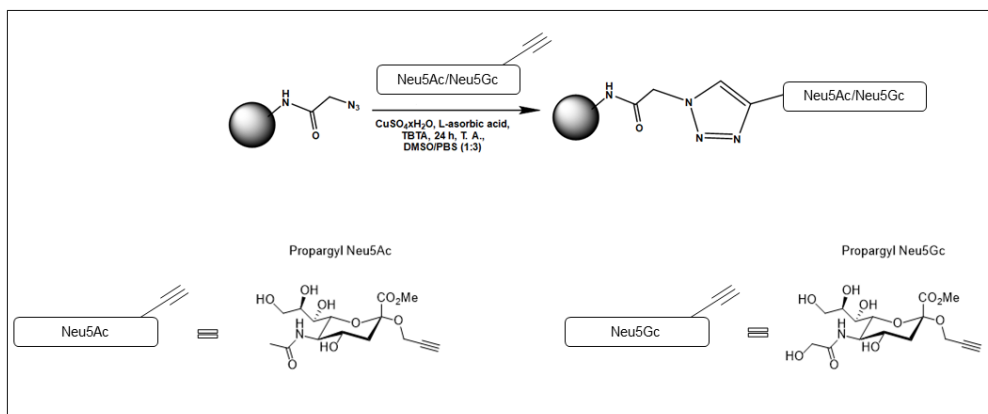


Fig. 2. 16 CuAAC reaction between BSA-azide and propargylNeu5Ac or propargylNeu5Gc.

The SDS-PAGE in Fig 2.17 shows an evident increase in molecular weight from BSA in line 2, to BSA-azide in line 3, to NeoNeuAc and NeoNeuGc in line 4 and 5. The increase is once again confirmed by the MALD-TOF spectra: in Fig. 2.18 A, the NeoNeuAc spectrum shows a single peak with a molecular weight of 69392 Da, and in Fig 2.18 B the NeoNeuGc spectrum shows a single peak with a molecular weight of 73437 Da. The Neu5Ac moieties/BSA calculated are 4 and the Neu5Gc/BSA moieties calculated are 13.

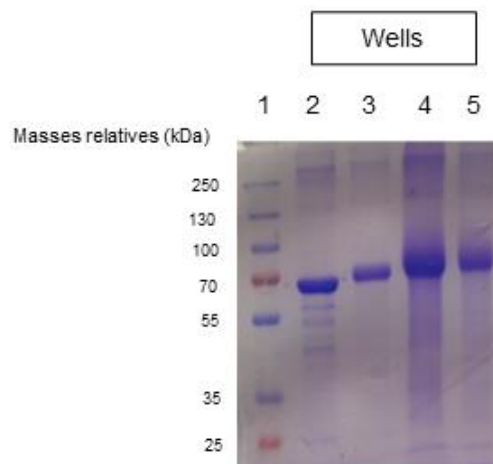


Fig. 2. 17 SDS-PAGE of NeoNuAc and NeoNeuGc synthesis. The protein samples were analysed under denaturing conditions on 14% polyacrylamide gel. Line 1: Protein marker, Line 2: BSA, Line 3: BSA-azide, Line 4: NeoNeuAc, Line 5: NeoNeuGc.

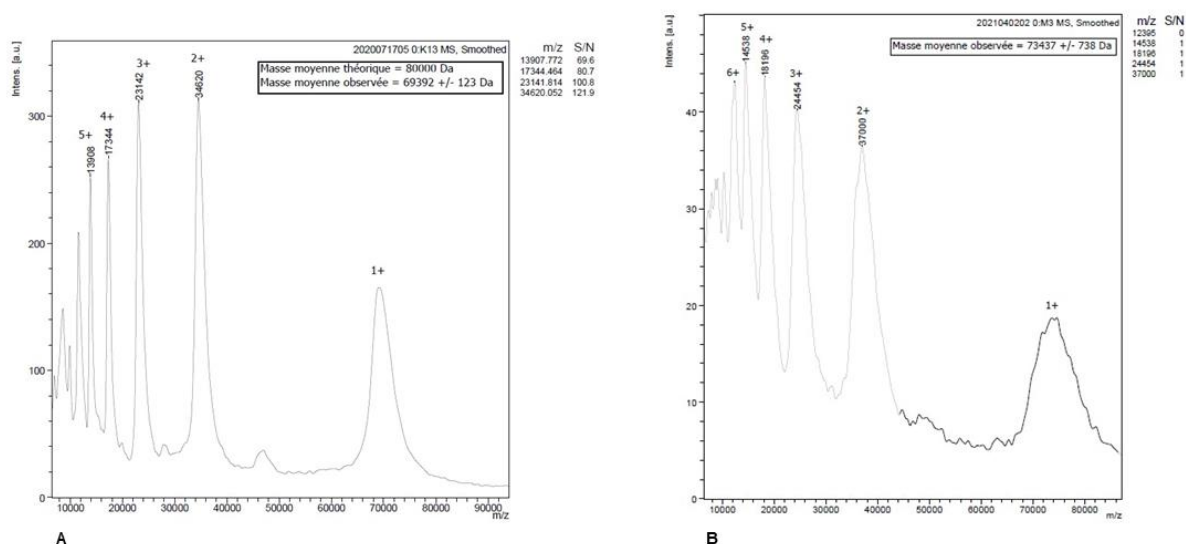
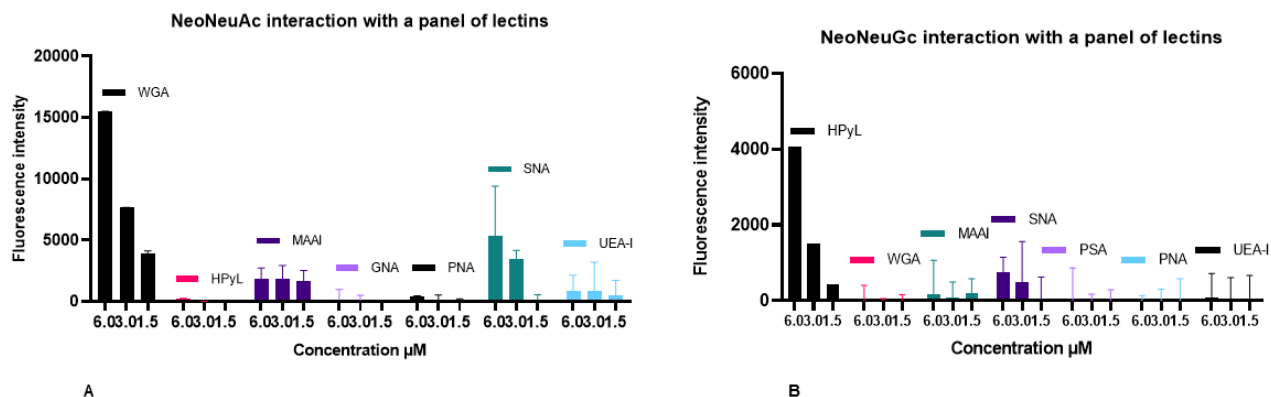


Fig. 2. 18 **A** MALD-TOF spectrum of NeoNeuAc **B** MALD-TOF spectrum of NeoNeuGc.

NeoNeuAc and NeoNeuGc show reverse specificity towards WGA and HPyL, as shown in the GLYcoPROFILE (Fig. 2.19 A and B). NeoNeuAc does not significantly interact with SNA or MAAI, in fact, according to literature the monosaccharide shows interaction with these two lectins only in the millimolar range [64] [65]. Neither of the two neoglycoproteins shows interaction with the mannose-specific lectins PSA and GNA, the lactose-specific lectin PNA and the fucose-specific lectin UEA-I (refer to Annex II, Chapter 4), underlining once again the selectivity of the interaction lectin-neoglycoprotein.



**Fig. 2. 19** **A** GLYcoPROFILE in direct binding of the NeoNeuAc neoglycoprotein labelled. The neoglycoprotein at three different concentrations was added to the plate coated with seven different lectins. **B** GLYcoPROFILE in direct binding of the NeoNeuGc neoglycoprotein labelled. The neoglycoprotein at three different concentrations was added to the plate coated with seven different lectins. The data shown are the average of experiments performed in triplicate.

## Conclusion

In this Chapter, it the synthesis of five neoglycoproteins of interest, by copper-catalysed azide-alkyne cycloaddition, is described. The neoglycoproteins produced will serve as standard probes for testing the performances of the three LEctPROFILE. The protocol developed for the synthesis of these molecules resulted in good yield, of 5 mg/synthesis, and high reliability batch to batch, key parameters to have reliable reagents for the kits development. The neoglycoprotein were also characterised through SDS-PAGE, MALDI-TOF/MS analysis and GLYcoPROFILE, allowing to confirm a high level of purity, quantify the ratio of sugar/BSA and study their specificity to each neoglycoprotein. Overall, these results build a solid ground for the next steps in the development of the kits.

Dans ce chapitre, il est décrit la synthèse de cinq néoglycoprotéines d'intérêt dans le cadre de cette thèse. Ces synthèses ont été réalisées par cycloaddition azide-alkyne catalysée par le cuivre ou « chimie click ». L'importance de ce travail réside dans le fait que toutes les néoglycoprotéines produites sont ensuite utilisées comme sondes standards pour le développement de trois kits LEctPROFILE. Les résultats de cette étude indiquent qu'avec ce protocole nous avons un bon rendement 5 mg/synthèse avec une excellente reproductibilité de lot à lot. Ce dernier paramètre est d'autant plus important pour assurer la totale reproductibilité des kits qui en sont issus et réduire la variabilité des résultats. Nous avons développé également une caractérisation complète de ces néoglycoprotéines, qui nous permet de maîtriser le degré de pureté le ratio glycanes/BSA et la fonctionnalité de chaque

néoglycoprotéine. Globalement, ces résultats constituent une base solide pour les prochaines étapes du développement des kits.

# Paper 1- Homo- and Heterovalent Neoglycoproteins as Ligands for Bacterial Lectins

ChemPlusChem

Research Article  
doi.org/10.1002/cplu.202100481

www.chempluschem.org



## Homo- and Heterovalent Neoglycoproteins as Ligands for Bacterial Lectins

David Goyard,<sup>[a]</sup> Benoît Roubinet,<sup>[b]</sup> Federica Vena,<sup>[b]</sup> Ludovic Landemarre,<sup>[b]</sup> and Olivier Renaudet<sup>\*[a]</sup>

Click chemistry gives access to unlimited set of multivalent glycoconjugates to explore carbohydrate-protein interactions and discover high affinity ligands. In this study, we have created supramolecular systems based on a carrier protein that was grafted by Cu(I)-catalyzed azide-alkyne cycloaddition with tetravalent glycodendrons presenting  $\alpha$ Gal,  $\beta$ Gal and/or  $\alpha$ Fuc. Binding studies of the homo- (4a–c) and heterovalent (5) neoglycoproteins (neoGPs) with the LecA and LecB lectins from *P. aeruginosa* has first confirmed the interest of the multivalent

presentation of glycodendrons by the carrier protein (IC<sub>50</sub> up to 2.8 nM). Moreover, these studies have shown that the heterovalent display of glycans (5) allows the interaction with both lectins (IC<sub>50</sub> of 10 nM) despite the presence of unspecific moieties, and even with similar efficiency for LecB. These results demonstrate the potential of multivalent and multispecific neoGPs as a promising strategy to fight against resistant pathogens.

### Introduction

In the past decades, glycosylated compounds based on multivalent synthetic scaffolds revealed essential tools for studying carbohydrate-protein (lectin) interactions and modulating biological events.<sup>[1,2,3]</sup> To design high affinity molecules, computational methods and structural data of the target protein allow the choice of sugar units, valency and linking arm, however predicting the influence of the scaffold itself remains difficult. Besides being easy to synthesize, it should ideally be chosen to be both soluble and stable *in vivo*, to bring the sugar units at the closest proximity of the protein binding pockets and to have low flexibility to limit entropic cost of the interaction. If simple scaffolds fulfil these criteria, the development of chemical ligation has significantly extended the geometrical limits of multivalent systems to optimize their properties.<sup>[4,5,6]</sup> By combining scaffolds having either identical or different characteristics, a diversity of glycosylated architectures are now accessible and some of them are currently investigated as pathogen inhibitors, immunomodulators, delivery or theranostic tools.<sup>[7,8,9,10,11]</sup>

In the present study, we report the conjugation of glycodendrons to a carrier protein using a convergent approach based on the Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC),<sup>[12]</sup> thus providing a new series of neoglycoproteins (neoGP) with clusters of glycans distributed at different

positions (Figure 1). Given the dense and expanded expression of glycoconjugates at the surface of the cell membrane, we reasoned that the combination of synthetic glycodendrons within a protein carrier could easily provide a diversity of relevant neoglycoprotein-based reagents for studying and/or modulating carbohydrate-protein interactions. In particular, the utilization of glycans involved in bacterial infections might have the potential to inhibit the infection process and to provide alternative to antibiotic treatments against resistant pathogens.<sup>[13]</sup> To this aim, we used Bovine Serum Albumin (BSA), a naturally unglycosylated protein as the carrier model. BSA was first functionalized with alkyne linkers, then tetravalent  $\alpha$ - and  $\beta$ -galactosylated ( $\alpha$ Gal and  $\beta$ Gal) or  $\alpha$ -fucosylated ( $\alpha$ Fuc) dendrons functionalized with an azido group were conjugated to provide homovalent neoGPs. The resulting structures may thus bind selectively to the LecA and LecB lectins of *Pseudomonas aeruginosa*.<sup>[14,15]</sup> These two tetrameric calcium-dependent lectins are involved in the bacterial adhesion to the host cells, the biofilm formation and the lethal effect for immunocompromised patients, thus making them ideal biological targets for this study. In addition, recent studies have highlighted the importance of heterovalent structures as high affinity and selective ligands for carbohydrate-binding proteins, often leading to original properties.<sup>[16,17,18,19,20,21,22]</sup> To this aim and to extend the interest of the present approach, another neoGP combining randomized grafting of the  $\alpha$ Gal and  $\alpha$ Fuc dendrons was prepared to provide a heterovalent structure. The binding ability of these homo- and heterovalent neoGPs was next evaluated with these two bacterial lectins.

[a] Dr. D. Goyard, Prof. O. Renaudet  
Univ. Grenoble Alpes, CNRS, DCM UMR 5250  
38000 Grenoble (France)  
E-mail: olivier.renaudet@univ-grenoble-alpes.fr

[b] Dr. B. Roubinet, F. Vena, Dr. L. Landemarre  
GLYcoDiag  
2 Rue du cristal, 45100 Orléans (France)

Supporting information for this article is available on the WWW under  
<https://doi.org/10.1002/cplu.202100481>

This article is part of a Special Collection celebrating the 10th Anniversary of ChemPlusChem.



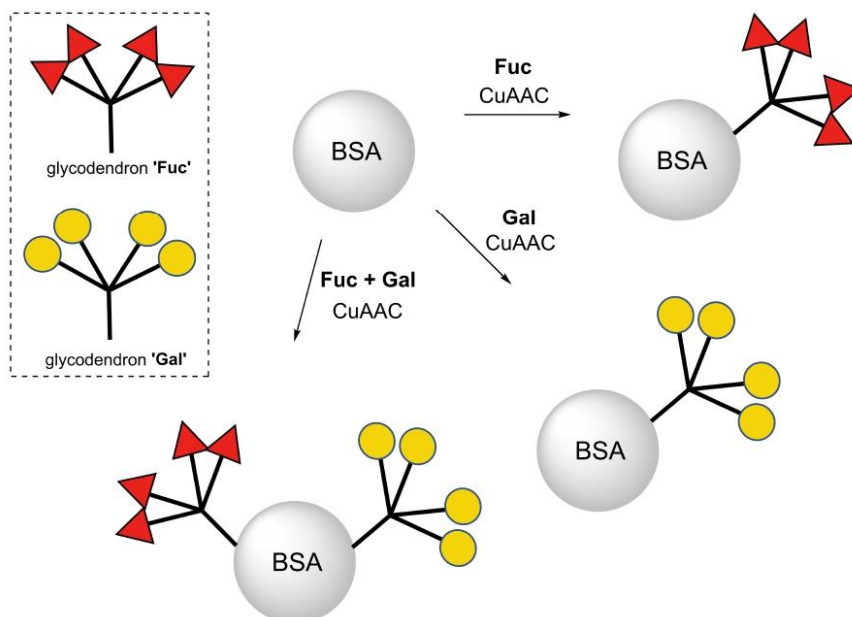


Figure 1. Assembly of homo- and heterovalent neoGPs from BSA and glycodendrons 'Fuc' and 'Gal' by CuAAC.

## Results and Discussion

### Synthesis of neoGPs

We previously reported several series of fully synthetic multivalent structures decorated with  $\beta$ Gal or  $\alpha$ Fuc.<sup>[23,24]</sup> Binding studies revealed that their affinity with LecA and LecB was closely dependent on either the valency or the ligand geometry. For example, only fucosylated structures of higher valency (*i.e.* 16 copies of  $\alpha$ Fuc) were found as excellent ligands for LecB with a significant impact of the sugar display on the affinity.<sup>[23]</sup> By contrast, tetravalent galactosylated structures have shown nanomolar affinity for LecA, while the increase of valency to sixteen  $\beta$ Gal had only a limited effect on the binding.<sup>[24]</sup> Here we first synthesized tetravalent dendrons displaying both glycans to be further conjugated to BSA as the carrier protein. A polylysine structure functionalized with azido groups (**1**) was synthesized by solid-phase peptide synthesis using the Fmoc/tBu strategy (Scheme 1).

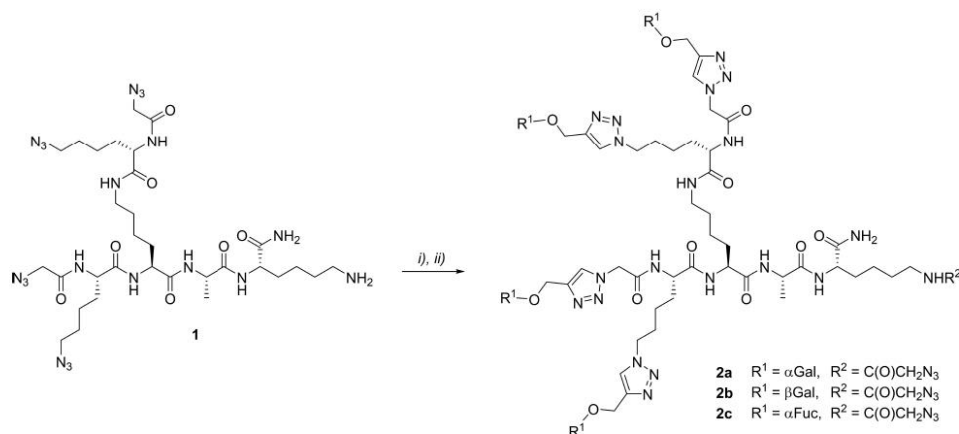
Propargylated  $\alpha$ Fuc and  $\beta$ Gal<sup>[24,25]</sup> were next conjugated by CuAAC under standard condition using copper sulfate, tris(3-hydroxypropyl-triazolylmethyl)amine (THPTA) and sodium ascorbate in a mixture of DMF and PBS (pH 7.4). After 2 hours, semi-preparative RP-HPLC purification provided tetravalent compounds that were subsequently functionalized on the free side chain of the C-terminal lysine with azido *N*-hydroxy-succinimidyl ester of azidoacetic acid (Figure S1–9).

The neoGPs were synthesized using a similar protocol. BSA was functionalized at the lysine side chain with a commercial propargyl linker by amide coupling (Scheme 2).

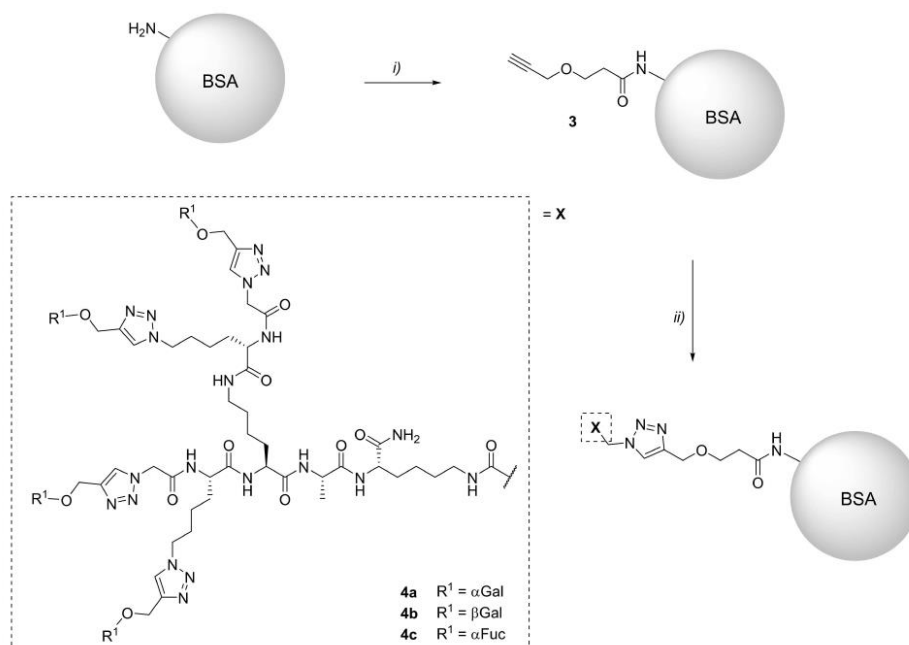
After 5 h in PBS, the crude mixture was purified by size-exclusion chromatography on Sephadex G25 and MALDI-TOF analysis of BSA-alkyne **3** indicated that among the 60 lysines present in BSA, approximately 33 were functionalized with an alkyne group (Figure S10). Finally, glycodendrons **2a–c** (25 equiv. over the 33 reactive groups present in **3**) were conjugated to **3** by CuAAC, with tris(benzyltriazolylmethyl)amine (TBTA), L-sodium ascorbate and CuSO<sub>4</sub> in PBS similarly to previous reports.<sup>[26,27,28]</sup> After 24 h at room temperature, the resulting homovalent neoGPs **4a–c** were purified on Sephadex G25. The purity was confirmed by SDS-PAGE and MALDI-TOF experiments enabled to estimate that between 12 to 19 tetravalent glycodendrons were coupled to BSA, which corresponds to 49–75 glycans per BSA and molecular weights from 92 to 105 kDa for homovalent neoGPs **4a–c** (Table 1, Figure S11 and S14). Of note, when the coupling reaction was performed with an excess of the fucosylated glycodendron **2c** over BSA-alkyne **3** (*i.e.* 50 equiv.), the quantity of  $\alpha$ Fuc only increased to 73 (versus 53 using 25 equiv.), thus suggesting that the remaining alkyne groups are less accessible in the protein to be fully functionalized under these experimental conditions (Figure S12).

The value presented in the table are rounded values; [a] MW of neoGPs **4a–c** was calculated on the basis of the number of eq. of glycodendrons **2a–c**: MW calc. = MW of BSA-alkyne **3** +





**Scheme 1.** Synthesis of the polylysine-based glycodendrons **2a–c**. Reagents and conditions: i) Propargyl glycoside,  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , THPTA, sodium ascorbate, DMF/PBS (pH 7.5, 1:1), r.t., 2 h; ii) *N*-hydroxy-succinimidyl ester of azidoacetic acid, DIPEA, DMF, r.t., 2 h.



**Scheme 2.** Synthesis of homovalent neoGPs **4a–c**. Reagents and conditions: i) 3-(2-Propyn-1-yloxy)propanoic acid 2,5-dioxo-1-pyrrolidinyl ester, PBS (pH 8.0), r.t., 5 h; ii) **2a–c**,  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , TBTA, sodium ascorbate, PBS (pH 7.5), r.t., 24 h.

25 × MW of **2a–c**; [b] MW of neoGP**4c** was calculated on the basis of the number of eq. of glycodendron **2c**, considering that BSA-alkyne **3** displays a maximum of 33 alkyne groups: MW calc. = MW of BSA-alkyne **3** + 33 × MW of **2c**; [c] MW of neoGP**5**

was calculated on the basis of the number of eq. of mixed glycodendrons **2a–c**: MW calc. = MW of BSA-alkyne **3** + 25 × (50% MW of **2a** + 50% MW of **2c**); [d] nDendron/BSA = (MW found for **4a–c** – MW of BSA-alkyne **3**)/MW of **2a–c**; [e] n

**Table 1.** MALDI-TOF analysis of the functionalization of BSA-alkyne 3 with glycodendrons 2a–c by CuAAC.

Compound	MW calc. [Da]	MW found [Da]	nDendron/BSA	nGlycan/BSA
neoGP 4a (αGal, 25 eq.)	114578 <sup>[d]</sup>	103371	19 <sup>[d]</sup>	75 <sup>[d]</sup>
neoGP 4b (βGal, 25 eq.)	114578 <sup>[d]</sup>	91894	12 <sup>[d]</sup>	49 <sup>[d]</sup>
neoGP 4c (αFuc, 25 eq.)	112978 <sup>[d]</sup>	93322	13 <sup>[d]</sup>	53 <sup>[d]</sup>
neoGP 4c (αFuc, 50 eq.)	126664 <sup>[b]</sup>	101321	18 <sup>[d]</sup>	73 <sup>[d]</sup>
neoGP 5 (αFuc/αGal, 50/50) <sup>[f]</sup>	113778 <sup>[c]</sup>	103303	19 <sup>[d]</sup>	76 <sup>[d]</sup>

Glycan/BSA = n Dendron/BSA × 4; [f] The ratio of 50% of αGal and 50% of αFuc residues was arbitrarily set to determine the number of dendrons and glycans per BSA and nGlycan/BSA.

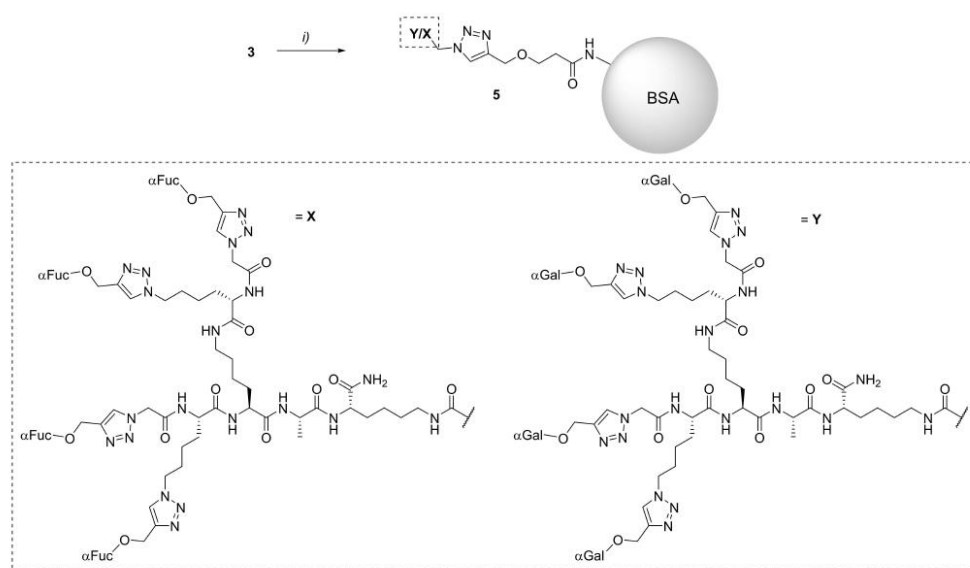
In addition, with respect to the role of LecA and LecB in the infectious and lethal processes mediated by *P. aeruginosa*, heteroclusters combining βGal and αFuc could simultaneously bind these two lectins to strengthen inhibitory effect as demonstrated in previous studies based on heterovalent peptide glycodendrimers,<sup>[29]</sup> pillar[5]arene-containing [2]rotaxane<sup>[30]</sup> or hexadecavalent cyclopeptide-based glycodendrimers.<sup>[24]</sup> To this aim, we performed the CuAAC reaction between BSA-alkyne 3 and a mixture of 2a and 2c to afford the heterovalent neoGP 5 with randomized distribution

of glycodendrons exposed by BSA (Scheme 3, Table 1, Figure S13 and Figure S15).

### Direct binding assays

The binding ability of the homovalent neoGPs 4a–c was studied with LecA and LecB using the GLYcoPROFILE® technology (GLYcoDiag, France).<sup>[26,27,28,31]</sup> NeoGPs 4a–c were first labelled with biotin using the protocol described previously then were incubated in microtitration plates coated with lectins. Streptavidin labelled with 5-(4,6-dichlorotriazinyl)-aminofluorescein (DTAF) was finally incubated in each well and the interaction was directly detected by reading the fluorescence intensity (FI) provided by the neoGP bound to the lectin. While no interaction was observed with the unspecific Wheat Germ Agglutinin (WGA), this experiment indicated that the galactosylated 4a–b and fucosylated 4c neoGPs were able to bind LecA and LecB, showing a FI proportional to the neoGP concentration (Figure S16–18). In addition, at lower concentration, we observed higher FI for 4a than for 4b which could be attributed to the higher valency in 4a rather than a clear preference of LecA for the alpha anomer of galactose. Altogether, these results confirm the benefit of the multivalent presentation of glycans.

Finally, we evaluated the interaction of the heterovalent neoGP 5, which statistically displays identical content of αGal and αFuc units, with LecA and LecB (See Figure S19–20). As expected, the fluorescent signal observed with LecA is higher



**Scheme 3.** Synthesis of heterovalent neoGP 5 with randomized distribution of glycodendrons. Reagents and conditions: i) 2a (12.5 eq.) and 2c (12.5 eq.), CuSO<sub>4</sub>·5H<sub>2</sub>O, TBTA, sodium ascorbate, PBS (pH 7.5), r.t., 24 h.

when BSA only displays  $\alpha$ Gal due to the higher valency (75 in **4a** vs 38 in **5**). Conversely, when the grafting ratio of  $\alpha$ Fuc increases (**4c**) compared to **5**, the fluorescent signal observed with LecB is of higher intensity. More interestingly, while containing half less specific glycan for a given lectin, the heterovalent neoGP **5** shows slightly lower binding ability with LecA and LecB than the parent homovalent neoGPs **4a–c**. This result suggests that the heterovalent display of glycans in BSA allows the simultaneous interaction with both lectins without significant loss of specificity and efficiency.

### Inhibition assays

To confirm these observations, we next performed inhibition assays with LecA and LecB using biotinylated neoGPs as tracers (respectively neoGP **4a** for LecA and neoGP **4c** for LecB).<sup>[31,26,27,28]</sup> The neoGPs **4a–c** and **5**, the two commercial neoglycoproteins NeoGa and NeoF functionalized with monovalent  $\alpha$ Gal or  $\alpha$ Fuc (grafting ratio of 13 and 14 units, respectively), and the corresponding monosaccharides have been tested as inhibitors (Table 2).

In this assay, lectin is immobilized in microplate wells and simultaneously incubated with a solution containing the biotinylated tracer and various concentration of the inhibitors. After a final incubation with streptavidin-DTAF, the intensity of fluorescence enables to determine the inhibition potency of the tested compounds (Figure S21–24). With galactosylated derivatives and LecA, we observed a modest  $IC_{50}$  value (247 nM) for the monovalent NeoGa, which promotes a 53-fold improvement compared to galactose (13  $\mu$ M). Additionally, both homovalent galactosylated neoGPs **4a–b** revealed  $IC_{50}$  in the nanomolar range with LecA. The best neoGP **4b** displaying  $\beta$ Gal even reaches a relative potency (rp) of 4700 over free galactose, i.e. almost 100-fold improvement when reported to the number of sugar units (rp/n). While containing 12  $\beta$ -galactosylated dendrons (i.e. 49  $\beta$ Gal), this compound indeed shows a rp/n value 3.6 times better than **4a** having higher valency (i.e. 75  $\alpha$ Gal). Unlike what was suggested by the direct binding assays, this result seems to indicate that  $\beta$ Gal units allow stronger inhibitory

effect than  $\alpha$ Gal. However, when alpha galactosylated ligands are compared, the heterovalent neoGP **5**, which contains 38 copies of  $\alpha$ Gal, was found slightly less efficient than **4a** (75  $\alpha$ Gal) while showing comparable rp/n (34 for **5** vs. 26 for **4a**). This highlights that the inhibition potency is closely related to the number of  $\alpha$ Gal and more importantly, that the presence of fucose units does not interfere with the inhibition process.

With LecB, the neoF with 14 sugar units showed a 200-fold better inhibition over fucose (rp/n=14), whereas neoGP **4c** displaying 53 glycans showed a remarkable improvement factor of 4200 times over fucose to inhibit the interaction, which corresponds to a rp/n of 79. Additionally, with the heterovalent neoGP **5** displaying only 38  $\alpha$ Fuc (combined with 38  $\alpha$ Gal units), the rp value reached similar efficiency albeit with lower rp/n. This unexpected result suggests that the presence of  $\alpha$ Gal in proximity of  $\alpha$ Fuc in **5** could induce a favourable orientation of the fucose units to inhibit the interaction of the tracer with LecB. This observation demonstrates the importance of spatial presentation of glycans rather than the valency for inhibition as already observed in other studies.<sup>[3]</sup>

### Conclusion

In this paper, we report the synthesis of multivalent neoglycoproteins by click chemistry. The functionalization of lysine residues of BSA with alkyne linkers provided approximately 33 anchoring groups to the protein that were subsequently conjugated with tetravalent glycodendrons presenting  $\alpha$ Gal,  $\beta$ Gal and/or  $\alpha$ Fuc by CuAAC. Homo- and heterovalent neoGPs (**4a–c** and **5**) functionalized with up to 19 randomly distributed dendrons were thus prepared and their binding abilities were studied with the LecA and LecB lectins from *Pseudomonas aeruginosa* using the GLYcoPROFILE® technology. In a direct assay, we first showed the potential of the homovalent neoGPs **4a–c** for recognition with the specific lectin and we observed that the presence of randomized fucosylated and galactosylated glycodendrons in BSA (**5**) allows the interaction with both LecA and LecB without any loss of efficiency. Moreover, inhibition assays have highlighted that the presence of  $\beta$ Gal (**4b**) ensures a stronger inhibition for LecA, whereas  $\alpha$ Gal (**4a**) at higher valency is required to promote similar effect. Unexpectedly, the inhibitory potency of the heterovalent neoGP **5** with LecB proved similar to the fucosylated neoGP (**4c**), thus suggesting that a more favorable orientation of the fucose units in **5**. Overall the simplicity of the synthetic strategy and the promising biological potency of carrier protein grafted with glycodendrimers opens the route towards the development of new therapeutic agents against resistant pathogens. In particular, due to their structural feature, such neoGP could efficiently mimic the cell surface and prevent the infection of host cells from pathogens. A large diversity of neoGPs is currently developed to explore this perspective.

**Table 2.** Inhibition of the interaction of biotinylated neoGPs to LecA and LecB-coated plates with homo- and heterovalent neoGPs **4a–c** and **5**.

Compound	n <sup>[a]</sup>	IC <sub>50</sub> [ $\mu$ M]	rp <sup>[b]</sup>	rp/n <sup>[c]</sup>
Gal (LecA)	1	13.2 $\pm$ 1.2	1	1
NeoGa <sup>[d]</sup> (LecA)	13	0.247 $\pm$ 0.025	53	4
neoGP <b>4a</b> (LecA)	75	0.0069 $\pm$ 0.0009	1913	26
neoGP <b>4b</b> (LecA)	49	0.0028 $\pm$ 0.0007	4714	96
neoGP <b>5</b> (LecA)	76 (38 $\alpha$ Gal)	0.010 $\pm$ 0.0012	1320	34
Fuc (LecB)	1	42.0 $\pm$ 2.9	1	1
NeoF <sup>[e]</sup> (LecB)	14	0.208 $\pm$ 0.017	201	14
neoGP <b>4c</b> (LecB)	53	0.010 $\pm$ 0.0013	4200	79
neoGP <b>5</b> (LecB)	76 (38 $\alpha$ Fuc)	0.010 $\pm$ 0.0014	4200	110

[a] Number of sugar units; [b] Relative potency "rp" = IC<sub>50</sub>(monosaccharide)/IC<sub>50</sub>(neoGP); [c] Relative potency/sugar "rp/n"; [d] Neoglycoprotein displaying  $\alpha$ Gal commercialized by GLYcoDiag, France; [e] Neoglycoprotein displaying  $\alpha$ Fuc commercialized by GLYcoDiag, France.

## Experimental section

### General methods

All chemical reagents were purchased from Aldrich (Saint Quentin Fallavier, France) or Acros (Noisy-Le-Grand, France) and were used without further purification. Commercial neoglycoproteins NeoGa and NeoF (*i.e.* BSA functionalized with  $\alpha$ Gal and  $\alpha$ Fuc, respectively), LecA and LecB 96-well plates (LectPROFILES plates) were obtained from GLYcoDiag (France). All protected amino acids and Fmoc-Gly-Sasrin® resin was obtained from Advanced ChemTech Europe (Brussels, Belgium), Bachem Biochimie SARL (Voisins-Les-Bretonneux, France) and France Biochem S.A. (Meudon, France). The EZ-link sulfo-NHS-LC-biotin was purchased from ThermoFisher Scientific (Ref: 21335) and the number of conjugated biotin quantified by the HABA/avidine method (ThermoFisher Scientific\_UserGuide: HABA). For glycopeptides, analytical RP-HPLC was performed on a Waters alliance 2695 separation module, equipped with a Waters 2489 UV/visible detector. Analyses were carried out at 1.23 mL min<sup>-1</sup> (Waters XBridge Shield RP18 3.5  $\mu$ m, C<sub>18</sub>, 100  $\times$  4.6 mm) with UV monitoring at 214 nm and 250 nm using a linear A-B gradient (buffer A: 0.09% CF<sub>3</sub>CO<sub>2</sub>H in water; buffer B: 0.09% CF<sub>3</sub>CO<sub>2</sub>H in 90% acetonitrile). Preparative HPLC was performed on Waters equipment consisting of a Waters 2545 controller and a Waters 2487 dual absorbance detector. Purifications were carried out at 22 mL min<sup>-1</sup> (VP 250  $\times$  21 mm nucleosil 100-7 C<sub>18</sub>) with UV monitoring at 214 nm and 250 nm using a linear A-B gradient. <sup>1</sup>H NMR spectra were recorded on Bruker Avance III 500 MHz spectrometers and chemical shifts ( $\delta$ ) were reported in parts per million (ppm). Spectra were referenced to the residual proton solvent peaks relative to the signal of D<sub>2</sub>O (4.79 ppm). All mass spectrometry characterizations were performed at Mass Spectrometry facility, PCN-ICMG, Grenoble. ESI high resolution mass spectra of glycopeptides were measured on a LTQ Orbitrap XL spectrometer from Thermo Scientific. For the neoGPs, Sephadex-G25 was obtained from GE Healthcare. Chemicals and solvents were used as received from commercial sources without further purification. Fluorescence was recorded with a Fluostar Optima spectrometer (BMG labtech, Offenburg, Germany). The MALDI-TOF analysis were carried out with an UltrafleXtreme from Bruker.

### General procedure for the preparation of glycodendrons by CuAAC (2a–c)

Propargyl glycoside (4.4 eq.) and azide-functionalized scaffold 1 (1 eq.) were dissolved in 1 mL of a 1:1 mixture of DMF and PBS buffer (pH 7.5). A solution of CuSO<sub>4</sub>·5H<sub>2</sub>O (0.5 eq.) and THPTA (1 eq.) in PBS was added to a solution of sodium ascorbate (3 eq.) in PBS. This mixture was added to the solution containing the azide and alkyne which was degassed with argon and stirred at r.t. for 2 hours after which RP-HPLC showed completion of the reaction. Chelex® resin was then added to the reaction mixture which was stirred for 45 minutes. The resin was filtered off, rinsed with water and the filtrate was concentrated under reduced pressure. The crude product was used directly in the next step. The resulting tetravalent compound 2 (1 eq.) was dissolved in dry DMF (1 mL), DIPEA was added to reach pH ~9–10 (c.a. 20  $\mu$ L) then *N*-hydroxy-succinimidyl ester of azidoacetic acid (1.5 eq.) was added. The reaction mixture was stirred at r.t. for 1 hour after which RP-HPLC showed completion of the reaction. The mixture was diluted with water (3 mL) and purified by semi-preparative RP-HPLC. Fractions containing the product were combined and lyophilized. Glycodendron 2a: 82% yield; HRMS (ESI<sup>+</sup>) *m/z*: calc. for C<sub>69</sub>H<sub>109</sub>N<sub>23</sub>O<sub>32</sub> [M-2H]<sup>2+</sup>: 885.8810, found 885.8808; RP-HPLC: R<sub>t</sub> = 5.09 min (C<sub>18</sub>,  $\lambda$  = 214 nm, 0–30% B in 15 min). Glycodendron 2b: 84% yield; HRMS (ESI<sup>+</sup>) *m/z*: calc. for C<sub>69</sub>H<sub>113</sub>N<sub>23</sub>O<sub>32</sub> [M+2H]<sup>2+</sup>: 887.8956, found 887.8971; RP-

HPLC: R<sub>t</sub> = 6.65 min (C<sub>18</sub>,  $\lambda$  = 214 nm, 0–30% B in 15 min). Glycodendron 2c: 74% yield; HRMS (ESI<sup>+</sup>) *m/z*: calc. for C<sub>69</sub>H<sub>113</sub>N<sub>23</sub>O<sub>28</sub> [M+2H]<sup>2+</sup>: 855.9058, found 855.9074; RP-HPLC: R<sub>t</sub> = 9.12 min (C<sub>18</sub>,  $\lambda$  = 214 nm, 0–30% B in 15 min).

### Functionalization of BSA with an alkyne chain (3)

A solution of 3-(2-Propyn-1-yloxy)propanoic acid 2,5-dioxo-1-pyrrolidinyl ester (2 mg, 60 eq.) was added in a solution of BSA (10 mg) in PBS and the reaction mixture was stirred for 5 h at room temperature. BSA-alkyne (3) was purified on a Sephadex G25 and its purity was controlled by SDS-PAGE (Figure S14–15). 90% yield. The number of alkyne group per BSA was determined by MALDI-TOF (Figure S10).

### Conjugation of BSA-alkyne 3 with glycodendrons 2a–c by CuAAC

TBTA (5 mg/mL), L-ascorbic acid (3 mg/mL), CuSO<sub>4</sub>·5H<sub>2</sub>O (3 mg/mL) and glycodendron 2a, 2b or 2c (10 mg/mL, 25 eq. or 50 eq.) was added to a solution of BSA-alkyne 3 (2 mg/mL) in PBS. The solution was stirred 24 h at room temperature then the crude mixture was purified by size-exclusion chromatography on a Sephadex G25 gel. The purity of each neoglycoproteins was controlled by SDS-PAGE (Figure S14–15) and the ratio of glycodendrimers by BSA was determined by MALDI-TOF (Figure S11–13).

### Biotinylated NeoGPs

The labelling of NeoGPs was carried out by amide coupling with the EZ-link sulfo-NHS-LC-biotin (10 equivalents) in PBS at pH 7.2 for 30 min at room temperature. The excess of biotin reagent was removed by dialysis process. The number of biotin coupled on the neoglycoprotein is determined according to the standard HABA/avidine quantification method.

### Direct binding assays

The assays were performed according GlycoDiag's protocol already described.<sup>[26,27,28,31]</sup> For direct binding mode, each neoglycoproteins were preliminary labeled with biotin according to standard protocol already described.<sup>[28]</sup> Briefly, the different compounds (range of concentrations) prepared in PBS supplemented with 1 mM CaCl<sub>2</sub> and 0.5 mM MgCl<sub>2</sub> were deposited in each well of lectin (50  $\mu$ L each) in triplicate and incubated for two hours at room temperature. After washing with PBS buffer, the streptavidin-DTAF conjugate was added (50  $\mu$ L) and incubated for 30 min. The plate was then washed again with PBS. Finally, 100  $\mu$ L of PBS were added for reading the plate using a fluorescence reader ( $\lambda_{\text{ex}}$  = 485 nm,  $\lambda_{\text{em}}$  = 530 nm). The intensity of the signal was directly correlated with the ability of the compound to be recognized by the lectin.

### Inhibition assays

The interaction profiles of each compound were determined through an indirect method based on the inhibition by the compound of the interaction between a specific couple lectin-glycan (a neoglycoprotein labeled with biotin and used as a tracer). Briefly, a mix of tracer (fixed concentration) and the corresponding compounds (range of concentrations) prepared in PBS supplemented with 1 mM CaCl<sub>2</sub> and 0.5 mM MgCl<sub>2</sub> is deposited in each well (50  $\mu$ L each) in triplicates and incubated two hours at room temperature. After washing with PBS buffer, the conjugate streptavidin-DTAF was added (50  $\mu$ L) and incubated 30 min more.

The plate was washed again with PBS. Finally, 100  $\mu$ L of PBS was added for the readout of fluorescent plate performed with a fluorescence reader ( $\lambda_{\text{ex}}=485$  nm,  $\lambda_{\text{em}}=530$  nm). The signal intensity was inversely correlated with the capacity of the compound to be recognized by the lectin and expressed as inhibition percentage with comparison with the corresponding tracer alone.

### Acknowledgments

This work was supported by CNRS, Université Grenoble Alpes, ICMG FR 2607, the French ANR projects Glyco@Alps (ANR-15-IDEX-02), LectArray (ANR-19-CE18-0019-03), Labex ARCANE and CBH-EUR-GS (ANR-17-EURE-0003). O.R. acknowledges the European Research Council Consolidator Grant "LEGO" (647938) for D.G. and the Proof of Concept Grant "THERA-LEGO" (963862). The European Union's Horizon 2020 research and innovation program (Marie Skłodowska-Curie grant agreement No. 814029 ITN SynBIOcarb) is acknowledged for F.V.

### Conflict of Interest

The authors declare no conflict of interest.

**Keywords:** bacterial lectin · click chemistry · glycodendrimer · multivalency · neoglycoprotein

- [1] O. Renaudet, R. Roy, *Chem. Soc. Rev.* **2013**, 42, 4515–4517.
- [2] Y. Kim, J. Y. Hyun, I. Shin, *Chem. Soc. Rev.* **2021**, 50, 10567–10593.
- [3] S. Cecioni, A. Imbert, S. Vidal, *Chem. Rev.* **2015**, 115, 525–561.
- [4] E. M. Sletten, C. R. Bertozzi, *Angew. Chem. Int. Ed.* **2009**, 48, 6974–6998; *Angew. Chem.* **2009**, 121, 7108–7133.
- [5] K. Villadsen, M. C. Martos-Maldonado, K. J. Jensen, M. B. Thygesen, *ChemBioChem* **2017**, 18, 574–612.
- [6] W. Tang, M. L. Becker, *Chem. Soc. Rev.* **2014**, 43, 7013–7039.
- [7] A. Bernardi, J. Jiménez-Barbero, A. Casnati, C. De Castro, T. Darbre, F. Fieschi, J. Finne, H. Funken, K.-E. Jaeger, M. Lahmann, T. K. Lindhorst, M. Marradi, P. Messner, A. Molinaro, P. V. Murphy, C. Nativi, S. Oscarson, S. Penadés, F. Peri, R. J. Pieters, O. Renaudet, J.-L. Reymond, B. Richichi, J. Rojo, F. Sansone, C. Schäffer, W. B. Turnbull, T. Velasco-Torrijos, S. Vidal, S. Vincent, T. Wennekes, H. Zuilhof, A. Imbert, *Chem. Soc. Rev.* **2013**, 42, 4709–4727.
- [8] L. Mousavifar, R. Roy, *Molecules* **2021**, 26, 2428.
- [9] C. Pifferi, A. Ruiz-De-Angulo, D. Goyard, C. Tiertant, N. Sacristán, D. Barriales, N. Berthet, J. Anguita, O. Renaudet, A. Fernández-Tejada, *Chem. Sci.* **2020**, 11, 4488–4498.
- [10] R. Sharma, K. Naresh, Y. M. Chabre, R. Rej, N. K. Saadeh, R. Roy, *Polym. Chem.* **2014**, 5, 4321–4331.
- [11] B. Todaro, S. Achilli, B. Liet, E. Laigre, C. Tiertant, D. Goyard, N. Berthet, O. Renaudet, *Biomater. Sci.* **2021**, 9, 4076–4085.
- [12] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, 41, 2596–2599; *Angew. Chem.* **2002**, 114, 2708–2711.
- [13] A. Imbert, A. Varrot, *Curr. Opin. Struct. Biol.* **2008**, 18, 567–576.
- [14] A. Imbert, M. Wimmerová, E. P. Mitchell, N. Gilboa-Garber, *Microbes Infect.* **2004**, 6, 221–228.
- [15] E. Mitchell, C. Houles, D. Sudakevitz, M. Wimmerová, C. Gautier, S. Pérez, A. M. Wu, N. Gilboa-Garber, A. Imbert, *Nat. Struct. Biol.* **2002**, 9, 918–921.
- [16] J. L. Jiménez Blanco, C. Ortiz Mellet, J. M. García Fernández, *Chem. Soc. Rev.* **2013**, 42, 4518–4531.
- [17] C. Müller, G. Despras, T. K. Lindhorst, *Chem. Soc. Rev.* **2016**, 45, 3275–3302.
- [18] M. González-Cuesta, C. Ortiz Mellet, J. M. García Fernández, *Chem. Commun.* **2020**, 56, 5207–5222.
- [19] M. I. García-Moreno, F. Ortega-Caballero, R. Rísquez-Cuadro, C. Ortiz Mellet, J. M. García Fernández, *Chem. Eur. J.* **2017**, 23, 6295–6304.
- [20] R. S. Bagul, M. Hosseini, T. C. Shiao, N. K. Saadeh, R. Roy, *Polym. Chem.* **2017**, 8, 5354–5366.
- [21] C. Ortiz Mellet, J.-F. Nierengarten, J. M. García Fernández, *J. Mater. Chem. B* **2017**, 5, 6428–6436.
- [22] J. P. Ribeiro, S. Villringer, D. Goyard, L. Coche-Guerente, M. Höferlin, O. Renaudet, W. Römer, A. Imbert, *Chem. Sci.* **2018**, 9, 7634–7641.
- [23] N. Berthet, B. Thomas, I. Bossu, E. Dufour, E. Gillon, J. Garcia, N. Spinelli, A. Imbert, P. Dumy, O. Renaudet, *Bioconjugate Chem.* **2013**, 24, 1598–1611.
- [24] D. Goyard, B. Thomas, E. Gillon, A. Imbert, O. Renaudet, *Front. Chem.* **2019**, 7, 666.
- [25] E. Fernandez-Megia, J. Correa, I. Rodríguez-Meizoso, R. Riguera, *Macromolecules* **2006**, 39, 2113–2120.
- [26] C. Assailly, C. Bridot, A. Saumonneau, P. Lottin, B. Roubinet, E. Krammer, F. François, F. Vena, L. Landemarre, D. Alvarez Dorta, D. Deniaud, C. Grandjean, C. Tellier, S. Pascual, V. Montembault, L. Fontaine, F. Daligault, J. Bouckaert, S. G. Guin, *Chem. A Eur. J.* **2021**, 27, 3142–3150.
- [27] M. Cauwel, A. Sivignon, C. Bridot, M. C. Nongbe, D. Deniaud, B. Roubinet, L. Landemarre, F.-X. Felpin, J. Bouckaert, N. Barnich, S. G. Guin, *Chem. Commun.* **2019**, 55, 10158–10161.
- [28] Y. Brissonnet, C. Assailly, A. Saumonneau, J. Bouckaert, M. Maillason, C. Petitot, B. Roubinet, B. Didak, L. Landemarre, C. Bridot, R. Blossey, D. Deniaud, X. Yan, J. Bernard, C. Tellier, C. Grandjean, F. Daligault, S. G. Guin, *Chem. Eur. J.* **2019**, 25, 2358–2365.
- [29] G. Michaud, R. Visini, M. Bergmann, G. Salerno, R. Bosco, E. Gillon, B. Richichi, C. Nativi, A. Imbert, A. Stocker, T. Darbre, J.-L. Reymond, *Chem. Sci.* **2016**, 7, 166–182.
- [30] S. P. Vincent, K. Buffet, I. Nierengarten, A. Imbert, J.-F. Nierengarten, *Chem. Eur. J.* **2016**, 22, 88–92.
- [31] L. Landemarre, E. Duverger, **2013**, pp. 221–226.

Manuscript received: October 26, 2021

Revised manuscript received: December 3, 2021

Accepted manuscript online: December 17, 2021

## **Chapter 3 – The Galili antigen**

### Abstract

The Gal $\alpha$ 1-3Gal epitope, also called Galili antigen, is found in mammals except from humans, apes and Old World Monkeys that lost its encoding gene (GGTA1) for the enzyme  $\alpha$ 1,3-galactosyltransferase ( $\alpha$ 1,3GT) during evolution. As a consequence, 1% of the circulating antibodies in humans are anti-Gal antibodies. This is an obstacle for the production of biotherapeutics in non-human mammalian cell lines, such as murine myeloma cells. The case of the Cetuximab monoclonal antibody, which is produced in these type of cells, was a striking case of hypersensitive reaction among patients due to the Galili antigen. For the same reason, anti-Gal antibodies are an obstacle for the transplantation of organs from pig into humans. To date, there is no a fast, high-throughput screening able to detect the presence of the Galili antigen on biotherapeutics and transplanted tissues and organs. We want to address this gap by building a lectin array highly specific and easy to use, in order to detect this antigen on biotherapeutics and biological samples. In this chapter, a review about the Galili antigen and an overview on the potential application of the Galili antigen LEctPROFILE kit are described. The results about the study of the kit are presented as well as a proof of concept using the above-cited mAbs Cetuximab and two other therapeutic mAbs. The Galili-antigen specific lectins MOA $\beta$ T and GSLIB4 are carefully characterized and studied for the development of the kit. At the end of the chapter is enclosed the paper “Extending Janus lectins architecture: characterization and application to protocells” by Notova *et al.*, in which the purification and characterization of MOA $\beta$ T is described.

### Résumé

L'épitope Gal $\alpha$ 1-3Gal, également appelé antigène Galili, est exprimé chez tous les mammifères à l'exception des humains, des singes et des singes de l'Ancien Monde qui ont perdu le gène codant (GGTA1) pour l'enzyme  $\alpha$ 1,3-galactosyltransférase ( $\alpha$ 1,3GT) au cours de l'évolution. En conséquence, 1% des anticorps circulants chez l'homme sont des anticorps anti-Galili. Cette caractéristique crée un obstacle pour la production de biothérapeutiques dans des lignées cellulaires de mammifères non humains, telles que les cellules de myélome murin. Le cas de l'anticorps monoclonal Cetuximab, qui est produit dans ce type de cellules, a été un cas frappant de réaction d'hypersensibilité chez les patients dû à l'antigène Galili. Pour la même raison, les anticorps anti-Gal sont un obstacle à la transplantation d'organes de porc chez l'homme. Il n'existe pas à ce jour de criblage rapide à haut débit capable de détecter la présence de l'antigène Galili sur les biothérapies et les tissus et organes destinés à être greffés. Nous voulons combler cette lacune en développant une puce à lectine hautement spécifique et simple d'utilisation, afin de détecter cet antigène sur des échantillons biothérapeutiques et biologiques. Dans ce chapitre, un examen de l'antigène Galili et un aperçu des applications potentielles du kit LectPROFILE d'antigène Galili sont décrits. Les résultats de l'étude du kit sont présentés ainsi qu'une preuve de concept utilisant le

mAbs cetuximab cité ci-dessus et deux autres mAbs thérapeutiques. Les lectines spécifiques de l'antigène Galili MOA $\beta$ T et GSLIB4 ont été particulièrement caractérisées et étudiées pour le développement du kit. L'article « Extending Janus lectins architecture: characterization and application to protocells » de Notova *et al.*, dans lequel la purification et la caractérisation de MOA $\beta$ T est décrite, est joint dans son intégralité, à la fin du chapitre.

#### 3.1 The Galili antigen

The enzyme  $\alpha$ 1,3-galactosyltransferase ( **$\alpha$ 1,3GT**), encoded by the gene **GGTA1**, synthesizes the  $\alpha$ 1,3-galactose epitopes (Gal $\alpha$ 1-3Gal $\beta$ 1-4GlcNAc-R) (*Fig. 3.1A*), which can be found at the external part of glycoproteins and glycolipids.  $\alpha$ 1,3GT uses N-acetyllactosamine residues as the sugar acceptor and UDP-Gal as the sugar donor (*Fig3. 1B*). This reaction takes place in the Golgi apparatus of non-primate mammals, prosimians and New World monkeys (South American monkeys). This enzyme is not found in humans, apes and Old World monkeys (African monkeys) that in contrast produce antibodies (Abs) against the **Gal $\alpha$ 1-3Gal $\beta$ 1-4GlcNAc-R** epitope, to which was given the name of **Galili antigen** from the scientist who discovered it [66]. It can also be called  $\alpha$ -Gal epitope or xenotransplantation epitope for reasons that will be explained further on in this chapter (Paragraph 3.2.2). The gene GGTA1 was inactivated in ancestral Old World monkeys and apes around 20–28 million years ago. One hypothesis on the disappearance is that it was due to an epidemic caused by some infectious agents spread in the Old World and expressing the Galili antigen. The pathogen exerted a selective pressure for these lineages to develop anti-Gal antibodies and to suppress the GGTA1 gene in order to, respectively, get rid of the Galili structure to survive the pathogen and to avoid autoimmune reactions. The other hypothesis is that a pathogen in the Old World was using the Galili antigen to enter the cells and it caused a selective pressure for primates to inactivate the GGTA1 gene. Consequently, the Galili structure started to be recognized as an antigen and antibodies against it started to be produced [67]. Now, it is calculated that there are around 1% of circulating anti-Gal IgG in immunocompetent individuals, produced in response to intestinal bacteria, many of which express this carbohydrate antigen [68]. Antibodies can bind either only the disaccharide but also the pentasaccharide as long as it contains the Gal $\alpha$ 1-3Gal epitope [69].



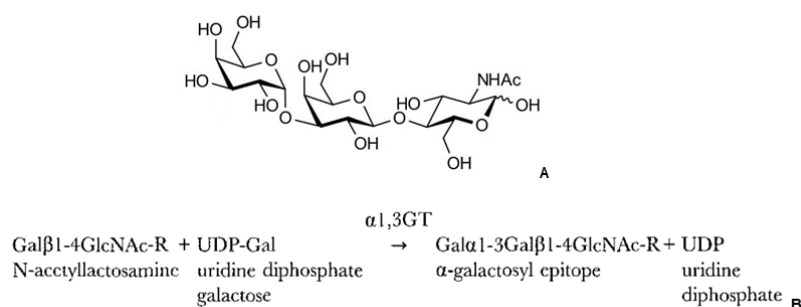


Fig. 3. 1 **A**, the Galili antigen trisaccharide. **B**, the reaction catalysed by the  $\alpha 1,3GT$ , adapted from [67].

### 3.2 Applications Galili antigen LEctPROFILE kit

#### 3.2.1 In biotherapeutics – the Cetuximab case study

**Cetuximab**, also known with the commercial name Erbitux, is a monoclonal antibody directed against the extracellular domain of the Epidermal Growth Factor Receptor (EGFR), preventing its dimerization and thus disrupting the induction of cell proliferation, invasion, apoptosis, angiogenesis, metastasis (Fig. 3. 2 B) [70]. It is a **medication** used for the treatment of metastatic colorectal cancer and squamous-cell carcinoma of the head and neck. This **chimeric mouse-human IgG** is produced in murine myeloma cells, and it contains two glycosylation sites on the conserved Fc portion and one per Fab portion (Fig. 3. 2 A) [71]. For the description of the IgG structure refer to Chapter 4, Paragraph 4.3.3.

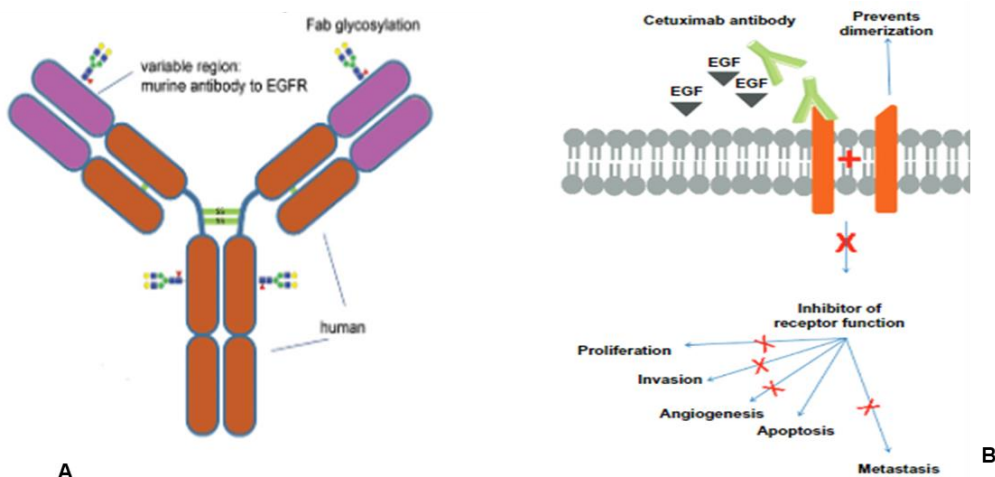


Fig. 3. 2 **A**, Cetuximab structure. It contains two glycosylation sites in the Asn 299 on the Fc portion and two at Asn88 in the Fab portion, adapted from [71]. **B**, the cetuximab mechanism: it binds to the EGFR and prevents its dimerization, blocking all its pro-tumour functions and promoting apoptosis, adapted from [70].

**Hypersensitivity** is defined as the allergic immune reactions, stimulated by soluble antigens, that can have different degrees of intensity and, in the case of anaphylaxis, it could be fatal. In the early

2000s, some patients started to show immediate onset of anaphylaxis after the administration of intravenous Cetuximab. Further studies pointed in the direction of **anti-Gal IgE Abs** and found that individuals with anti-Gal IgE  $\geq 0.35$  IU/mL were more likely to develop a reaction against Cetuximab. As seen in Paragraph 3.1, all humans have IgG antibodies specific for the disaccharide Gal $\alpha$ 1-3Gal and Cetuximab is produced in murine myeloma cells which can synthesise this glycan motif. Of the total  $\alpha$ -gal in Cetuximab, most of it is located on the Fab domain [72]. Anti-Gal-producing B cells occasionally undergo isotypes switch to produce IgE antibodies, which cause allergic reactions by binding to  $\alpha$ -gal epitopes [73].

The investigation of this phenomena started considering the **geographical area** where the hypersensitivity was more common and it turned out it was in the South-eastern regions of the United States, mainly in North Carolina, Arkansas, Missouri, Virginia, and Tennessee. In the same regions there were some cases of people, having spent much time outdoors, developing gastrointestinal symptoms, skin rash and urticaria around 3 to 6 hours after red meat consumption. After careful analysis, it was stated that also in this case the allergy was triggered by anti-Gal IgE Abs.

What was left to be understood was the reason why so many cases were concentrated in the same area, what was the link between Cetuximab anaphylaxis and **red meat allergy** and what is the mechanism behind the isotype switch from IgG to IgE. Most patients of the very same regions with red meat hypersensitivity reported having recently been bitten by the **Lone Star tick** (*Amblyomma americanum*). Isotype switch to anti-Gal IgE was observed in several individuals following a bite from this tick. In Europe and Australia, the isotype switch was connected to different tick species than in the United States. In Europe, *Ixodes ricinus* was implicated, while in Australia the relevant tick was *Ixodes holocyclus*. There are three theories to explain the isotope switch: the response could be caused by some component of tick saliva, by some residual glycans from a previous blood meal or by another organism commensal of the tick [72].

The big appearance in North America was due to the increase of deer's population on the territory since these animals are a major carrier of lone star ticks. In turn, the increasing deer population was linked to several reasons, among which the enactment of leash laws for dogs and a decrease in the number of hunters. This phenomenon got the name of  **$\alpha$ -gal syndrome** [72] (Fig. 3. 3). Therefore, the conclusion was that the Galili contained in the Cetuximab administered to those patients was enough to trigger a strong immune response in individuals having already a significant anti-Gal IgE titre. Thus, the Galili antigen LEctPROFILE kit could be used as a screening method to detect as early as possible (research stage) the level of Galili antigen in biotherapeutics and make sure that it does not overcome the threshold established by the Regulatory Agencies (Chapter 1, Paragraph 1.6.4).

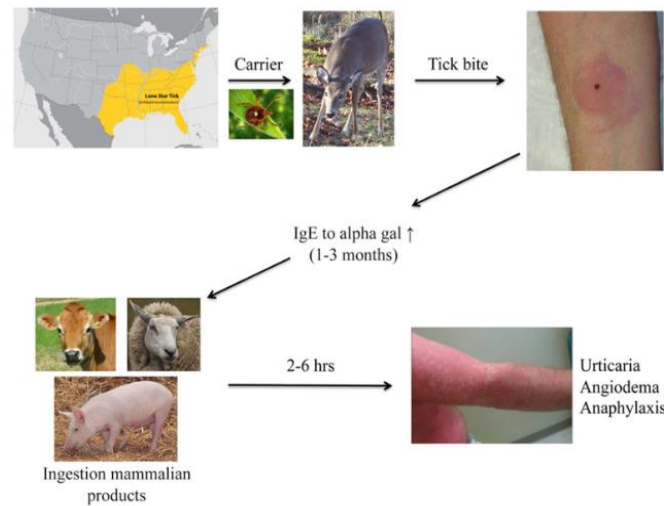


Fig. 3. 3 Schematic representation of the distribution of hypersensitivity reactions to Cetuximab and red meat in the southeast region of North America and its connection to the increasing deer population carrying ticks and tick bites, adapted from [72]].

### 3.2.2 In organ transplantation

Transplantation is the implant of a living tissue or an organ from a donor into a recipient. It mostly consists of **allotransplantation**, when an organ is transplanted from a person, living or deceased depending on the organ donated, to another one. Since 1980, it has been a revolutionary medical practice but it still presents many setbacks which are really hard to overcome. First of all, patients must take immunosuppressive drugs that do not always prevent rejection. Second, transplantation is a treatment, not a cure, and the implanted organs last a determined period of time (on average around 10 years [74]) after which they should be replaced. And here pops up the third and main problem: the shortage of organs coming from compatible donors. According to the US Health Resources and Services Administration, currently, there are more than 100000 people waiting for an organ and every nine minutes a person is added to the list. To solve this critical issue, it has been thought that organs coming from other mammals could be used: pig organs have attracted attention for their size which is really similar to humans. Mammal organs, that are not humans, used for transplantation, are defined as “**xenograft**” from which the process takes the name “**xenotransplantation**”.

There is anyway an **immunologic barrier** between pigs and humans and one of the immunogenic epitopes is the Galili antigen that is produced in pigs and for which humans have circulating antibodies (Paragraph 3.1). The Galili antigen for this reason is also called **xenotransplantation epitope** and it is to be expected that transplantation of pig organs into humans results in the binding of anti-Gal Abs to the Galili antigen on the xenograft cells with their consequent destruction. Immunosuppressive drugs are not sufficient to avoid the organ rejection, unless used at a dose that

completely inhibits the immune system, preventing the recipient from being able to respond to eventual infections [73].

In the early 2022, the first xenograft transplant was performed at the University of Maryland School of Medicine (Baltimore, MD, USA), on the patient David Bennett, a 57-year-old man with advanced heart failure and a type of arrhythmia called ventricular fibrillation, who was declared ineligible to receive a human heart. The genetically modified pig heart was provided by Revivicor, a regenerative medicine company based in Blacksburg, Virginia (USA). They are known for the production of the so-called “**Galsafe pigs**” that are  $\alpha 1,3\text{GT}$ -deficient [75]. Other than the Galili antigen as carbohydrate knockouts, the gene edits included the Sda blood group antigen, and N-glycolylneuraminic acid. Seven more genetic modifications were added: 6 human genes were inserted that are two anti-inflammatory genes, two genes that promote normal blood coagulation and prevent blood vessel damage, and two other regulatory proteins that help downsize the antibody response plus a modification for a growth hormone to control the size of the heart and reduce the chance the organ would grow once implanted [76].

Unfortunately, after two months from the transplant, the patient died from a porcine virus. Of course, the Galili antigen is not the only source of rejection or death, nevertheless, the fast screening of knock-out pigs could help the detection of the unwanted glycan and add a small brick in the improvement of these organs. The LEctPROFILE kit could be an interesting option for detection of the Galili antigen in knock-out animals.

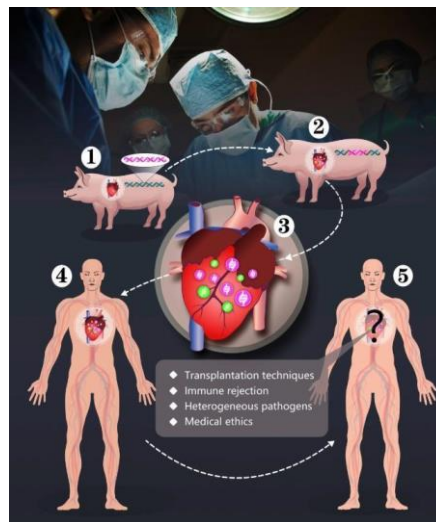


Fig. 3. 4 Galsafe engineered pigs used for xenotraspantation, adapted from [77].

### 3.2.3 In vaccines and other clinical settings

Since humans have natural anti-Gal antibodies and the Galili antigen is quite easy to chemically synthesise, it could be potentially applied to increase the **immunogenicity of viral vaccines**. The immunogenicity of influenza vaccine is not optimal, leading to the main problem of the **low uptake of the vaccine** by antigen presenting cells (APC). This happens because the quantity of hemagglutinin A (HA), the viral envelope glycoprotein included in the annual vaccines, is very low (15 µg HA of each vaccinating strain) [78]. Moreover, the APCs at the vaccination site uptake the vaccine by random endocytosis because they do not recognize any glycoprotein marker. The mechanism of action of viral vaccines is shown in *Fig. 3.5*: the viral antigen is recognized by the APCs that could be macrophages, dendritic cells and Langerhans cells of the skin. Dendritic cells further process the viral glyco/protein into peptides and present them by major histocompatibility complex (MHC) activating the cell-mediated immune response. MHCII presents the antigen to CD4+T cells. CD4+T cells differentiate into CD4+T helper cells that help activate CD8+T cells that release granzyme, perforins and interferon  $\gamma$ , and B cells that produce antibodies. Eventually, memory T cells and memory B cells are produced [79]. If the viral antigen injected with the vaccine is not a sufficient amount, it will not be enough to activate a good adaptive immune response and the efficacy of the vaccine will be lowered.

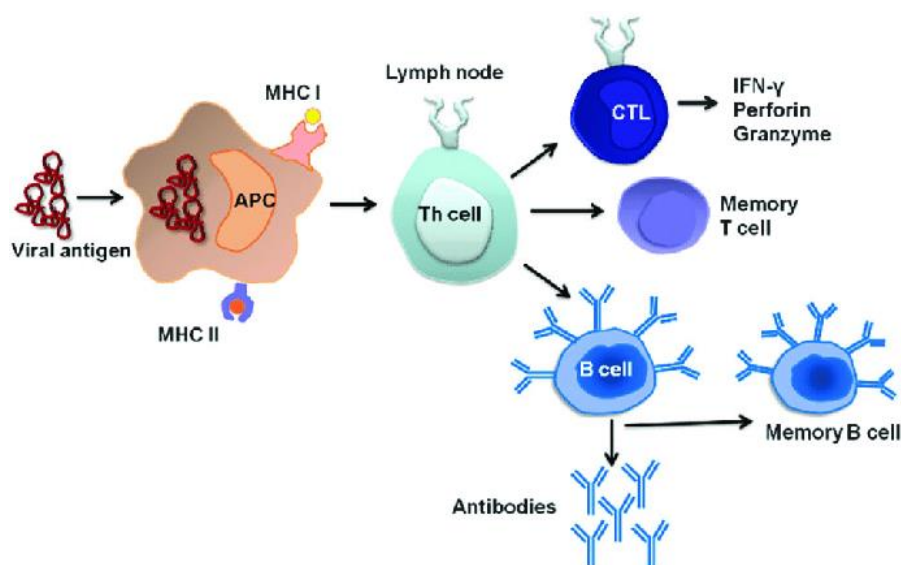


Fig. 3. 5 Schematic representation of the viral vaccine mechanism of action, adapted from [79].

The HA of influenza has 6-8 N-linked glycans [80] whereas major glycoprotein in the envelope of HIV is the gp120 which has 24 N-linked glycosylation sites. Studies were carried out to see how the immunogenicity of vaccines against these pathologies could be improved by carrying alpha-gal epitopes on their multiple carbohydrate chains. Flu virus is usually produced in embryonated chicken eggs. Chickens lack  $\alpha$ -gal epitopes, and since Sia is removed from viral glycoproteins by viral

neuraminidase, the carbohydrate chains on HA terminate with N-Acetylglucosamine (Gal $\beta$ 1–4GlcNAc-R). Thus,  $\alpha$ -gal epitopes could be added to all carbohydrate chains by recombinant  $\alpha$ 1,3GT.  $\alpha$ 1,3GT knockout mice immunised with inactivated flu virus vaccine expressing  $\alpha$ -gal epitopes presented higher anti-flu virus antibody and T cell response in comparison to knockout mice immunised with the original inactivated flu virus lacking  $\alpha$ -gal epitopes [81]. The gp120 carbohydrate chains, produced in CHO cells, have the terminal structure Sia-Gal $\beta$ 1–4GlcNAc-R. They were converted to  $\alpha$ -gal epitopes by incubation of the glycoprotein with a mixture of neuraminidases for removal of Sia plus  $\alpha$ 1,3GT and UDP-Gal. Immunizations of knockout mice with the gp120 expressing  $\alpha$ -gal epitopes resulted in higher anti-gp120 antibody response and a similar higher T cell response, in comparison to the immune response in knockout mice receiving similar immunizations with gp120 lacking  $\alpha$ -gal epitopes. This principle could be applied to different types of vaccines including the ones for cancer [82].

Since anti-Gal Abs are the most abundant natural antibodies in humans, there are other clinical settings in which the Galili antigen/anti-Gal interaction could be potentially exploited : one example is the injection into patients with solid tumours of glycolipids with carbohydrate chains capped with Galili epitopes (Fig. 3.6). They would act as Tumor Associated Antigen (TAA) recruiting APC to the site and starting the immune response against cancer cells. In clinical studies, patients with advanced-stage cancer survived more than expected after the injection and no patients developed clinical signs of toxicity or autoimmunity [83].

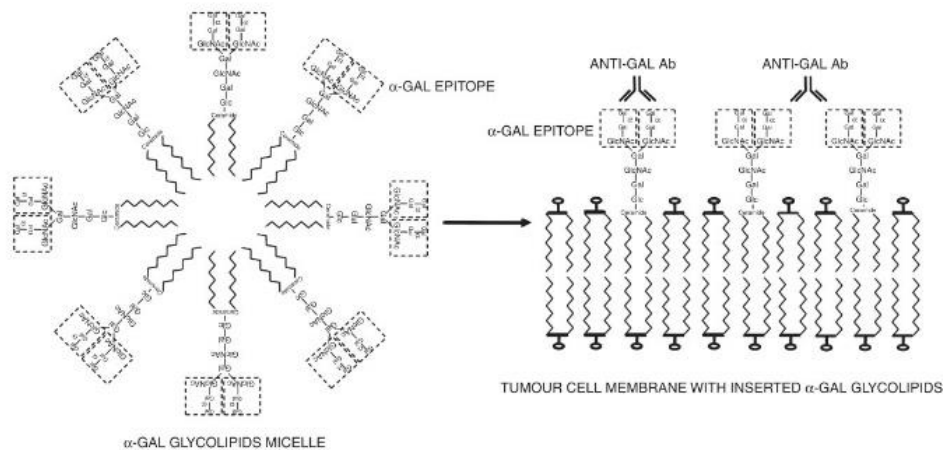


Fig. 3. 6 Representation of the  $\alpha$ -Gal glycolipids and how they are recognized by anti-gal Abs when injected in tumour tissues, adapted from [66].

The same principle can be applied also for accelerating the healing of wounds, burns, regeneration of ischaemic myocardium and of injured nerves;  $\alpha$ -gal nanoparticles applied to the injury site will bind anti-Gal and may induce rapid macrophage migration and activation for the secretion of vascular endothelial growth factor but further studies need to be conducted on mammalian animal models

[66]. Taking these studies into account, the Galili antigen LEctPROFILE kit could be a fast screening tool for the presence and relative quantification of Galili antigen during all phases of the vaccine and  $\alpha$ -gal glycolipids and nanoparticles production.

### 3.3 The *Marasmius oreades* agglutinin

MOA (*Marasmius Oreades agglutinin*) is a lectin from the edible fairy ring mushroom. MOA is specific for oligosaccharides with terminal non-reducing Gal $\alpha$ 1-3Gal linked galactose, while disaccharides linked in  $\alpha$ 1,2,  $\alpha$ 1,4, and  $\alpha$ 1,6 show no binding. It agglutinates group B blood cells which display the branched trisaccharide epitope Gal $\alpha$ (1,3)-[Fuc $\alpha$ (1,2)]Gal $\beta$ -epitopes but no binding is detected with blood group A or H oligosaccharides [82]. It also binds to linear oligosaccharides such as Gal $\alpha$ 1-3Gal1-4Glc on glycosphingolipid, isoGb3 and the Galili antigen Gal $\alpha$ 1-3Gal1-4GlcNAc [12]. The overall protein sequence is around 290 amino acids. The entire structure consists of a N-terminal  $\beta$ -trefoil domain and a C-terminal proteolytic domain [84]. Three carbohydrate-binding sites are located on the trefoil domain but they have different degrees of affinities for the ligand, with the  $\alpha$  and  $\beta$  sites appearing to have an higher affinity than the  $\gamma$  (Fig. 3. 7). The proteolytic domain doesn't bear binding sites but is important for protein homodimerization [85].

In our case, only the N-terminal  $\beta$  trefoil domain (156 amino acids) was cloned, expressed in *E. coli* and purified by affinity chromatography to perform the experiments described in this thesis. We named it **MOA $\beta$ T**, which means MOA beta trefoil.

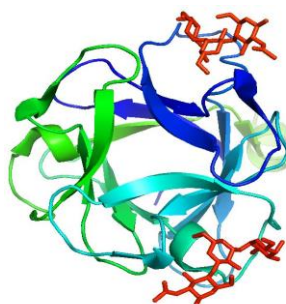


Fig. 3. 7 MOA  $\beta$ -trefoil portion crystal structure in complex with the Galili antigen. This figure was created with Pymol, adapted from the crystal structure of Grahn *et al*, ([85] PDB: 2IHO).

Summary of characteristics	
Number of amino acids	293 whole structure/ 156 construct only $\beta$ -trefoil
Molecular weight $\beta$ -trefoil	17244,01 Da
Theoretical pI	5.14

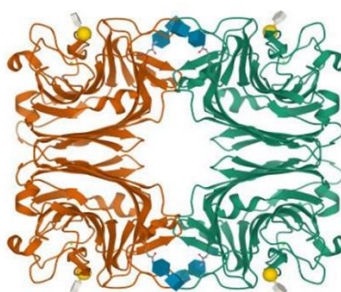


Extinction coeff	52940
------------------	-------

Table 3. 1 MOA  $\beta$ T characteristics.

### 3.4 GSLI-B<sub>4</sub>

*Griffonia simplicifolia lectin-1* is a carbohydrate-binding protein present in the seeds of the African leguminous shrub. It is a mixture of five tetrameric isolectins that vary in their content of A and B subunits. It is highly specific for  $\alpha$ -Galactose and the xenotransplantation antigen was often studied by using this lectin in the course of the years [69]. We have decided to include it in the Galili antigen LEctPROFILE kit study because it is instructive to compare it with MOA $\beta$ T. Structural data suggest that the GSLI-B<sub>4</sub> isolectin has a very restricted binding site that does not appear to easily accommodate glycans beyond the monosaccharide unit [86]. This was verified by the X-ray crystallographic structure of the B4 isolectin complexed with Gal $\alpha$ 1-3Gal in which it is shown that only the non-reducing galactosyl group makes contact with the lectin (*Fig. 3.8*) [87]. Moreover, Kirkeby and Moe, using an enzyme-linked lectin assay, showed that galactosyl disaccharides linked in  $\alpha$ 1-2,  $\alpha$ 1-3, and  $\alpha$ 1-4-galactobiosyl groups as well as the linear Galili antigen, linked to human serum albumin were very similar in their binding affinity to GSLI-B<sub>4</sub> [88].



*Fig. 3. 8* Crystal structure of GSLI-B<sub>4</sub> in complex with the xenograft antigen. PDB code 1HQL, adapted from [86].

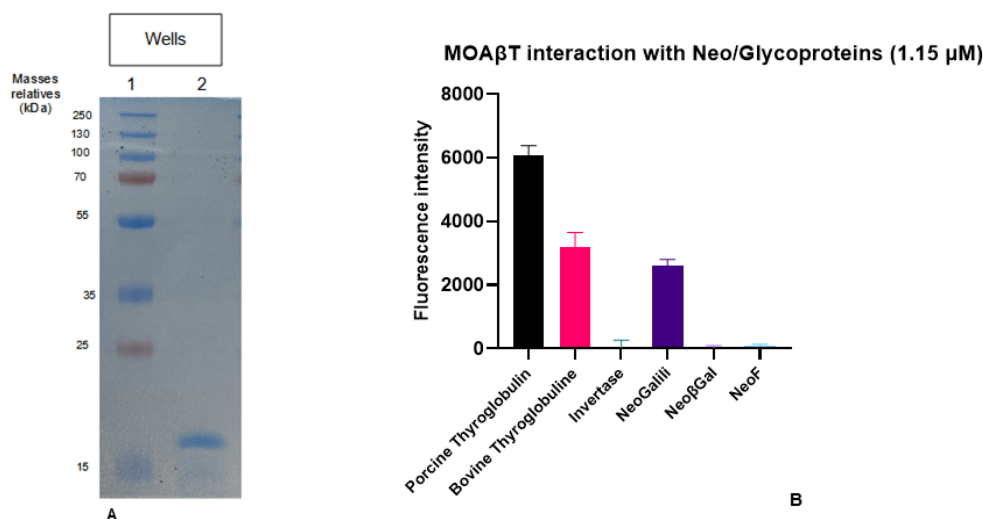
## Results and discussion

### 3.5 MOA $\beta$ T Purification

The  $\beta$ -trefoil domain was cloned and purified during my secondment at the CERMAV (Centre de Recherches sur les Macromolécules Végétales) in Grenoble, in January 2020. The purification protocol and lectin characterization are described in the paper at the end of this chapter (Notova et al., 2022). At GLYcoDiag, we optimized the purification protocol for industrial scale purification. The purification procedure is property of GLYcoDiag and for this reason it will not be described in detail



in this thesis. All lectins manufactured at GLYcoDiag are carefully quantified through the Bradford protein assay: the yield of MOA $\beta$ T reached for the industrial batch production is of around 15 mg of protein/ culture litre. We checked the lectin purity through SDS-PAGE: in *Fig. 3.9A* it is shown the lectin at the expected molecular weight of around 17 kDa. No impurities were detected. We tested also the lectin activity and selectivity for the target sugar through GLYcoPROFILE in direct binding. The bar chart in *Fig. 3.9B* shows the selectivity of the protein for the thyroglobulin glycoproteins known to bear the Galili antigen, and for the NeoGalili neoglycoprotein.



*Fig. 3. 9 A* SDS-PAGE analysis of MOA $\beta$ T The protein sample was analyzed under denaturing conditions on 14% polyacrylamide gel. **Line 1:** Protein marker, **Line 2:** MOA $\beta$ T (17,2 kDa). **B** GLYcoPROFILE in direct binding of MOA $\beta$ T interacting with labelled Porcine Thyroglobulin, Bovine Thyroglobulin, Invertase, NeoGalili and Neo $\beta$ Gal, Neof.

### 3.6 GSLI-B<sub>4</sub> Functionality

This lectin was not produced at GLYcoDiag but it was purchased from Vector. We checked the functionality of the lectin and its selectivity through GLYcoPROFILE in direct binding. The bar chart in *Fig. 3.10*, shows that the specificity of GSLI-B<sub>4</sub> for NeoGalili is higher than thyroglobulin, in accordance to literature: Kirkeby and Moe obtained higher affinity when thyroglobulin was denatured, proving that the accessibility of the galactosyl residues needs to be taken into account [88].

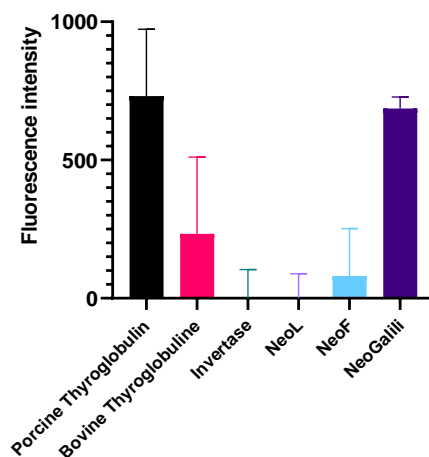
GSLIB4 interaction with Neo/Glycoproteins (1.15  $\mu$ M)

Fig. 3. 10 GLYcoPROFILE in direct binding of GSLI-B<sub>4</sub> interacting with labelled Porcine Thyroglobulin, Bovine Thyroglobulin, Invertase, NeoGalili, NeoL, NeoF.

### 3.7 Parameters of the Galili antigen LEctPROFILE kit

To optimize the kit for comprehensive glycans analysis, a number of parameters were studied. First of all, since the lectin MOA $\beta$ T was new in GLYcoDiag's catalogue of lectins, the optimal quantity of lectin/well was evaluated. Moreover, the type of detection, meaning the measurement of the interaction lectin-glycans either in fluorescence or absorbance, was studied. Once completed these two steps, the neoglycoprotein NeoGalili was chosen as a model probe for the characterization of the sensitivity of the kit and studies upon the IC<sub>50</sub> determination. Except from the choice of the type of plate and the quantity of lectin/well, the same process was applied to the lectin GSLI-B<sub>4</sub>. Furthermore, a set of glycoproteins and neoglycoproteins was tested on the kit to prove the specificity of the two lectins and for further studies on the IC<sub>50</sub>. To validate the assay for potential applications, therapeutic monoclonal antibodies were tested on the kit.

#### 3.7.1 Determination of MOA $\beta$ T quantity/well

In order to establish the optimum detection signal, the determination of the right quantity of lectin/well is a step of paramount importance. At the same time, we chose the most suitable plate for the by testing the lectin on adsorption plates and on covalent plates. The lectin was immobilized on covalent a adsorption plate at decreasing concentrations starting from 4  $\mu$ g/well. To test the optimal interaction only based on the lectin quantity/well, not dependent from the neoglycoprotein concentration, the NeoGalili was added in each well at the same concentration of 1.15  $\mu$ M. The detection of the signal was measured in absorbance units for the covalent plate, and in absorbance and fluorescence units for the adsorption plates (Fig. 3.11 A, B).

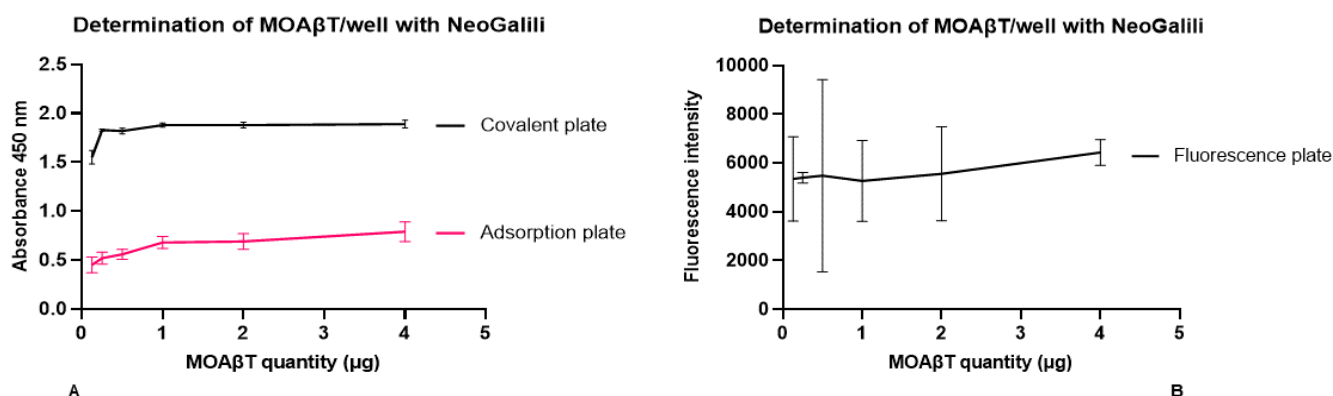
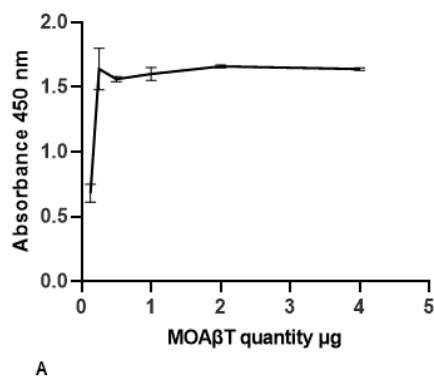


Fig. 3. 11 **A** Comparison of the binding of MOAβT at decreasing quantity/well to the NeoGalili neoglycoprotein at a fixed concentration (1.15 μM) on covalent plate (black) and adsorption plate (pink). The two experiments can be compared because the signal in both cases is measured in absorbance. **B** Binding of MOAβT at decreasing quantity/well to the NeoGalili neoglycoprotein at a fixed concentration (1.15 μM) on adsorption plate, this time the interaction is reported as fluorescence intensity.

As it can be seen in Fig 3.11 A, in both the adsorption and covalent plate the Absorbance at 450 nm stays at a constant value until the quantity of lectin decreases to 0.5 μg. The main difference between the covalent and the adsorption plate is that the covalent plate seems to show a more intense signal (Abs 400 nm > 1.5 versus Abs 400 nm = 0.6) and therefore probably a better sensitivity. In Fig 3.11 B, we see that the intensity of the signal in fluorescence on adsorption plate it's strong enough to think of a good interaction between MOAβT and NeoGalili; there is a slight decrease of the intensity following the decrease of the quantity lectin/well that however are not directly proportional. This makes us think of an experimental error more than a real difference connected to the quantity of lectin/well which is confirmed by the repetition of the experiment that gave the same results (data not shown) and by the higher standard deviation reported at lower concentrations (e.g. at 0.5 μg, the standard deviation = 1000 fluorescence units). Therefore, the quantity chosen in any condition is 1 μg/well because it is the right quantity to achieve a good signal while more would give the same result with a higher waste of lectin and less would compromise the quality of the signal. It needs to be pointed out that we chose to follow up with experiments on the covalent plate because of the strong signal that might have resulted in a very sensitive LectPROFILE kit.

To confirm that the right quantity was 1 μg/well, and show that the high intensity of the signal on the covalent plate was not dependent on the NeoGalili high concentration, we performed an experiment with MOAβT at decrescent concentrations, and NeoGalili at the lower concentration of 0.3 μM. We can see in Fig 3.12 A, that again, no matter we decreased the NeoGalili concentration, the absorbance remains high and at a plateau up until around 0,5 μg/well, with a big drop at lower quantities of lectin. To see this event magnified, we decided to keep the same concentration of NeoGalili but to apply it on decreasing concentrations of lectin going from 1 μg/well to 0.1 μg/well. Fig 3.12 B clearly shows a drop of the signal directly proportional to the lectin quantity.

Determination of MOAβT/well on covalent plate with NeoGalili



Determination of MOAβT/well on covalent plate with NeoGalili

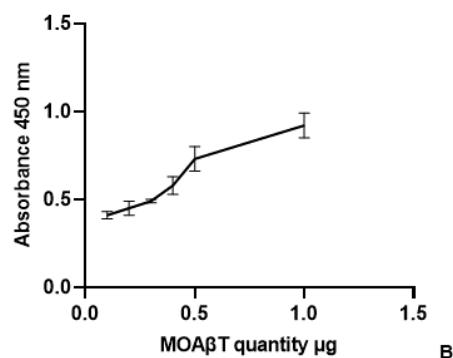


Fig. 3. 12 **A** MOAβT decreasing quantity/well (4 μg/well to 0.12 μg/well) plotted against the absorbance at 450 nm resulting from the lectin interaction with the NeoGalili neoglycoprotein at a fixed concentration (0.3 μM) on covalent plate. **B** MOAβT decreasing quantity/well (1 μg/well to 0.1 μg/well) plotted against the absorbance at 450 nm resulting from the lectin interaction with the NeoGalili neoglycoprotein at a fixed concentration (0.3 μM) on covalent plate.

The last proof of the fact that the optimal quantity of lectin is 1 μg/well was obtained making two experimental curves in direct binding of NeoGalili interaction with MOAβT at 1 μg/well and 0.5 μg/well. This time, the lectin was immobilized at a fixed quantity/well and the neoglycoprotein was added in serial dilutions starting from 0.85 μM. The results of the direct binding of NeoGalili neoglycoprotein on different quantities of MOAβT/well are shown in Fig. 3.13: the data obtained fit saturation binding curves obtained using a non-linear regression model from which the **BC50** and **Bmax** were calculated. The BC50 represents half of the maximum ligand concentration at which there is interaction with the lectin. This value can also be called  $K_d$  and reflects a surface concentration of the binding sites because the area in the microtiter plate is constant. The Bmax is the plateau value of adsorption and reflects the maximum number of the lectin binding sites [69]. These values are shown in the Table 3.2. The  $K_d$  is 10 times lower when the lectin is coated at 1 μg/well, meaning an higher sensitivity. For both curves the number of binding sites calculated are around 2, in accordance with the crystallographic studies that reveal three binding sites/monomer, each one with different ligand occupancy and with ELLA studies that calculates 4 binding sites using the full MOA including the dimerization domain [85] [69].

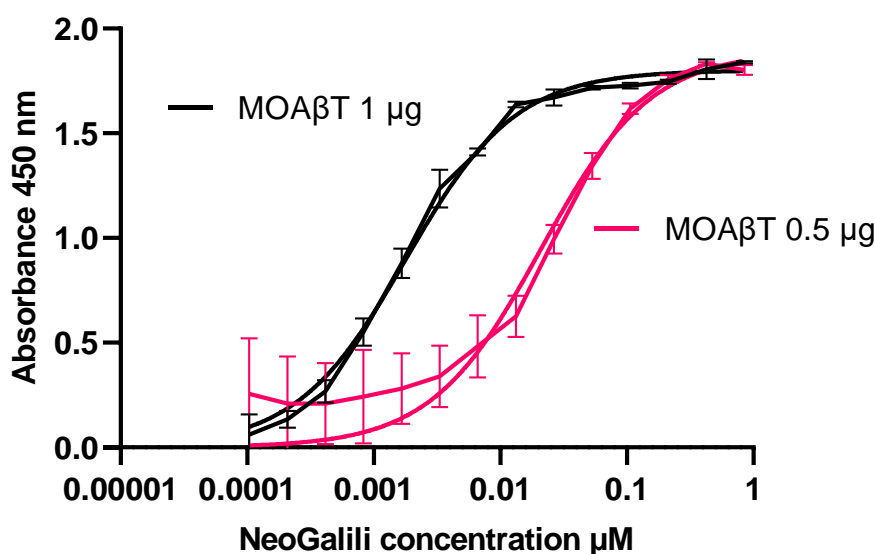
Comparison MOA $\beta$ T binding curves to NeoGalili

Fig. 3. 13 Comparison of MOA $\beta$ T binding to NeoGalili neoglycoprotein when MOA $\beta$ T is at a fixed quantity/well of either 1  $\mu$ g/well or 0.5  $\mu$ g/well with NeoGalili added in serial dilutions.

MOA $\beta$ T 0.5 $\mu$ g/well		MOA $\beta$ T 1 $\mu$ g/well	
BC <sub>50</sub>	B <sub>max</sub>	BC <sub>50</sub>	B <sub>max</sub>
20 nM	1.9	2 nM	1.8

Table 3. 2 BC<sub>50</sub> and B<sub>max</sub> calculated for MOA $\beta$ T based on the experimental curves shown in Fig 3.13.

## 3.7.2 Direct binding in absorbance on covalent plate

To continue the development of the Galili antigen LEctPROFILE kit, it was chosen the covalent immobilization of MOA $\beta$ T in the microtiter plate at 1  $\mu$ g/well. More experiments in direct binding were performed to confirm the value for the BC<sub>50</sub> and to calculate the **detection limit** of the kit, meaning the absorbance units below which the interaction between the lectin and the neoglycoprotein is considered non-specific. The data shown in Fig 3.14 represent a direct binding curve made with MOA $\beta$ T immobilized at 1  $\mu$ g/well followed by the application of the labelled NeoGalili at 1.15  $\mu$ M and serially diluted. Again, the value calculate for the BC<sub>50</sub> with the non-linear regression model it is confirmed to be in the nanomolar range even though in this case it is ten times higher (40 nM) than the one calculated for the previous experiment (2 nM) at the same quantity of lectin/well (1  $\mu$ g/mL, Fig 3.13). The detection limit of the kit was settled at 0.1 absorbance units considering that the significative value of interaction of the wells lectin-PBS is above 0.1 absorbance units. We can see from the experimental curve that also the detection limit is very low and it can be included in the range going from 0.001-0.01  $\mu$ M.

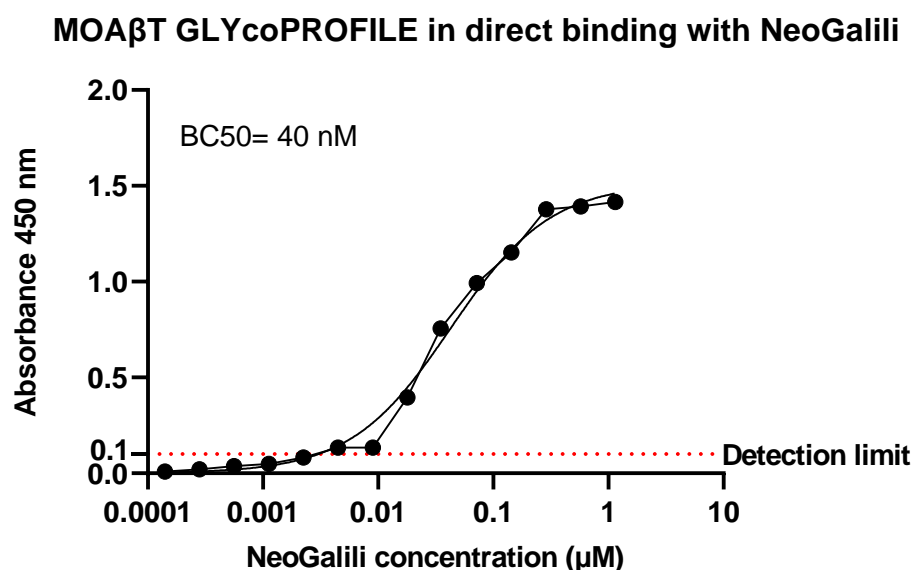
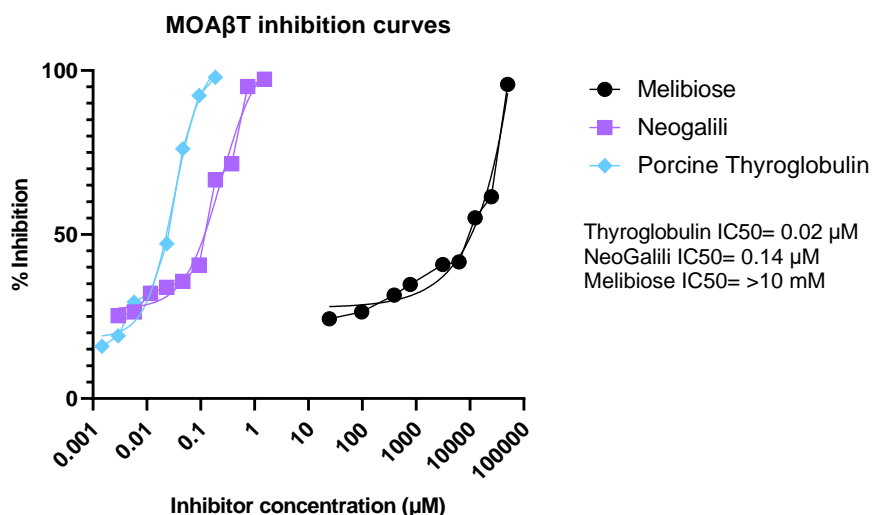


Fig. 3. 14 Direct binding curve of MOA $\beta$ T immobilized on covalent plate at the fixed quantity of 1  $\mu$ g/well with NeoGalili neoglycoprotein applied to the plate at 1.15  $\mu$ M and serially diluted.

### 3.7.3 Inhibition in absorbance on covalent plate

The inhibition experiments are designed to understand the ability of a compound to inhibit the binding between MOA $\beta$ T and the NeoGalili neoglycoprotein. The inhibition assay proves to be particularly advantageous because the compounds of interest do not need to be labelled, saving time and reagents, and they can directly be put in relation with the already characterised neoglycoprotein. The NeoGalili is well-characterized and it is used as a reference inhibitor so that, using the Galili antigen LECTPROFILE kit, a relative quantity of Galili antigen for an unknown compound can be estimated in relation to the NeoGalili. In Fig 3.15 the labelled NeoGalili functions as a tracer applied to the lectin plate at a fixed concentration of 0.5  $\mu$ g/mL. The compounds of choice, named inhibitors, compete with the tracer for the binding to the lectin. The data fit a non-linear regression model from which the respective IC<sub>50</sub> were calculated. As model compounds, we used the glycoproteins porcine thyroglobulin and the disaccharide melibiose (Gal $\alpha$ 1-6Glc). They were used respectively at a starting concentration of 12  $\mu$ M for the NeoGalili and thyroglobulin and at 50 mM for the melibiose and then serially diluted. It is evident the affinity of the porcine thyroglobulin for the lectin is higher than the NeoGalili affinity for the lectin, in fact the IC<sub>50</sub> of Porcine Thyroglobulin is almost ten times less than the one of the NeoGalili neoglycoprotein. This can be due either to the avidity effect caused by a bigger quantity of NeoGalili on the thyroglobulin or a better arrangement of the sugar moieties on this glycoprotein. Hence, MOA $\beta$ T has a strong selectivity for the Gal $\alpha$ 1-3Gal linkage and it is clear that melibiose has a poorly inhibitory effect on MOA $\beta$ T unless it is used at very high concentration (100 mM). No inhibition is obtained with the galactose monosaccharide

(not shown). The low affinity of MOA $\beta$ T for the monosaccharide has a structural reason: the binding sites are found on the protein surface and are extended and shallow, and this is the main feature that make this lectin more suitable for the study on the Galili antigen than the lectin GSLI-B<sub>4</sub> (Paragraph 3.6).



*Fig. 3. 15* Inhibition experiment performed with MOA $\beta$ T at 1  $\mu$ g/well interacting with labelled NeoGalili at the fixed concentration of 0.5  $\mu$ g/mL and with the non-labelled inhibitors applied in serial dilution, from a starting concentration of Porcine Thyroglobulin=12 $\mu$ M, NeoGalili=12  $\mu$ M, Melibiose=50 mM.

### 3.7.4 MOA $\beta$ T studies on adsorption plate with fluorescence detection

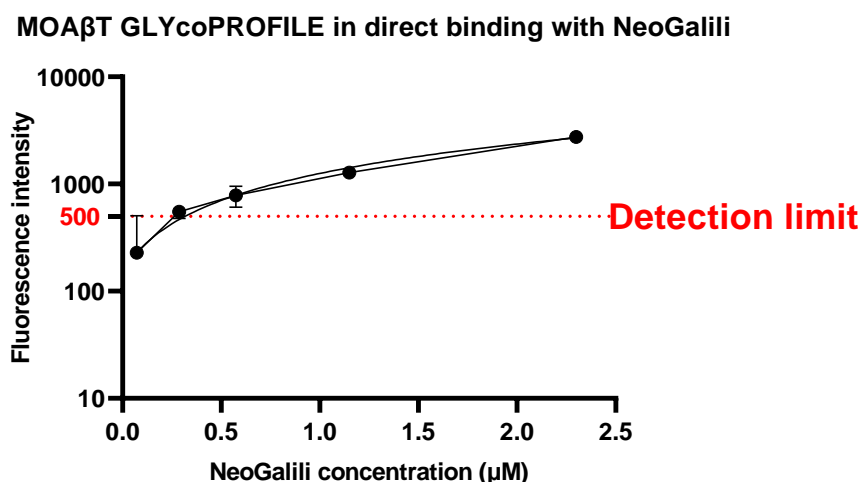
At the same time, further studies were carried on using the adsorption plate in fluorescence. The choice to continue also in this direction was due to different reasons:

1. The other lectin considered for this kit, GSLI-B<sub>4</sub>, gives good results in fluorescence and it was interesting to see the comparison with MOA $\beta$ T.
2. There are lectins that cross-react with the extravidin-peroxidase used in the absorbance assay, so in the case the use of the kit needs to be extended to those lectins, fluorescence would be the only possible choice.
3. The use of extravidin-peroxidase adds one step more to the experiment protocol, which does not take up much time when it comes only to one test but thinking it on an industrial scale is for sure time-consuming.
4. The use of the substrate for extravidin-peroxidase makes the kit development more expensive.
5. The covalent binding of the lectin to the plate cannot be easily controlled.

The lectin MOA $\beta$ T was compared to the full MOA and to GSLI-B<sub>4</sub> in order to have a clear overview on the potential use of these lectins for the kit.

### 3.7.5 Direct binding in fluorescence on adsorption plate

We performed the GLYcoPROFILE in direct binding of MOA $\beta$ T on adsorption plate in fluorescence in order to define the kits characteristics for this specific type of plate and this method of detection. As shown *Fig. 3.16 A*, knowing that the fluorescence detection is slightly less sensitive than the absorbance, the starting concentration of the NeoGalili neoglycoprotein was set at a double concentration of 2.3  $\mu$ M if compared to the experiment in *Fig. 3.14*. Since in this case we do not see saturation in the binding of MOA $\beta$ T to the NeoGalili neoglycoprotein, a simple linear regression analysis was used to calculate the BC50. [69].



*Fig. 3. 16* Direct binding curve of MOA $\beta$ T coated on adsorption plate at the fixed quantity of 1  $\mu$ g/well with NeoGalili neoglycoprotein applied to the plate at 2.3  $\mu$ M and serially diluted.

The BC50 calculated for MOA $\beta$ T interaction with NeoGalili neoglycoprotein is in the micromolar range (1 $\mu$ M) while the one calculated in absorbance on covalent plate is in the nanomolar range (40 nM). A different sensitivity was expected because the signal has a higher amplification in absorbance thanks to the use of the colorimetric reaction. The detection limit of the kit, set at 500 fluorescence units, for this reason is affected too, being around 0.3  $\mu$ M, which would not be suited for low-concentration samples. In order to see if this is due to the fact that MOA $\beta$ T has less avidity for the neoglycoprotein because of the lack of the dimerization domain, experiments in direct binding with the full MOA were performed.

### 3.7.6 Comparison between MOA $\beta$ T and rMOA

The GLYcoPROFILE in direct binding of the full MOA was performed in order to compare it with the one of MOA $\beta$ T (*Fig. 3.16*). In the course of the thesis I have attempted to purify the natural MOA directly from the fairy ring mushrooms but with poor results, so we performed the experiments with the commercial lectin. In order for the experiments to be comparable, the same parameters as



in the experiment in Fig. 3.16 were kept for the curve shown in Fig. 3.17. No matter the higher signal (max fluorescence intensity around 10000 fluorescence units), also in this case the BC50 is in the millimolar range (1.2  $\mu\text{M}$ ) and it is difficult to calculate the detection limit for 500 fluorescence units because below 0.5  $\mu\text{M}$  the signal drops. Overall, the lack of dimerization domain does not seem to affect the binding performance of MOA $\beta$ T, as it is also highlighted in the paper at the end of this chapter (Notova et al.,).

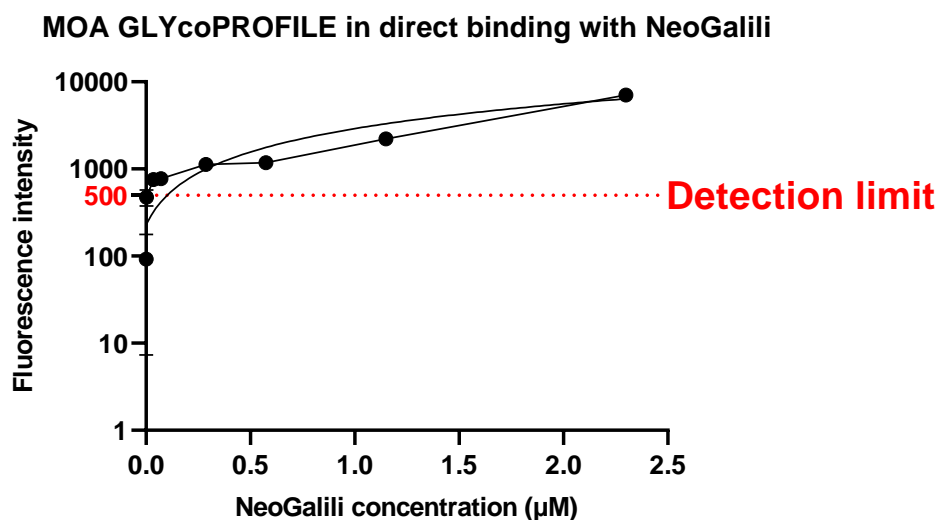


Fig. 3. 17 Direct binding curve of MOA coated on adsorption plate at the fixed quantity of 1  $\mu\text{g}/\text{well}$  with NeoGalili neoglycoprotein applied to the plate at 2.3  $\mu\text{M}$  and serially diluted.

### 3.7.7 GSLI-B<sub>4</sub> direct binding in fluorescence on adsorption plate

The same experiment in direct binding that were performed for MOA $\beta$ T and full were carried out for GSLI-B<sub>4</sub>. The use of the adsorption plate in fluorescence was chosen due to previous studies and for the advantages of this detection method over the absorbance one, already described in Paragraph 3.7.4. As shown in Fig. 3.18, GSLI-B<sub>4</sub> presents a very good interaction with the NeoGalili reflected in the high fluorescence intensity measured (>10000 fluorescence units). Again, in this case we do not see saturation in the binding of GSLI-B<sub>4</sub> to the NeoGalili neoglycoprotein, hence a simple linear regression was used to estimate the BC50. The BC50 is in the micromolar range (0.9  $\mu\text{M}$ ) as for MOA and MOA $\beta$ T but the detection limit is lower, around 0.1  $\mu\text{M}$ , showing a better sensitivity compared to the MOA $\beta$ T in fluorescence on adsorption plate, and to full MOA.

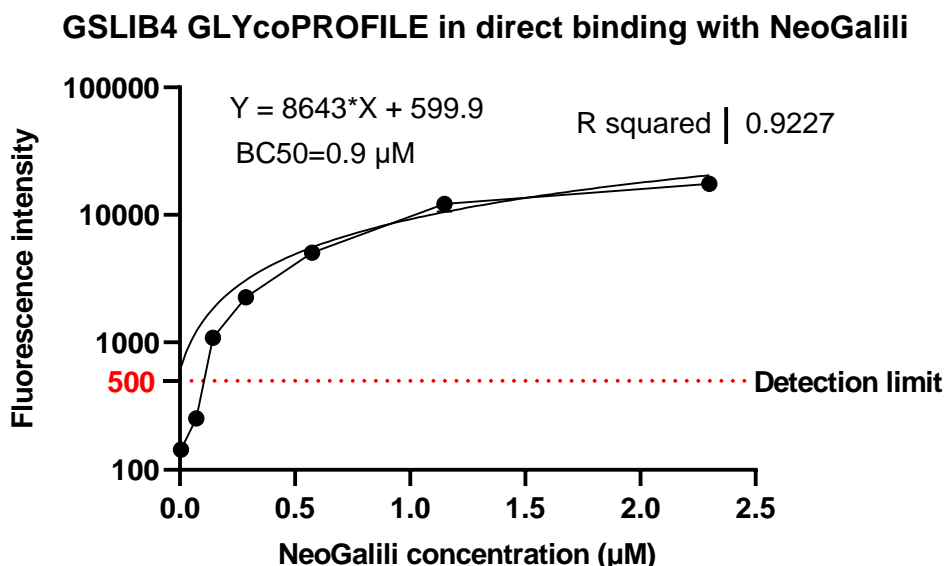
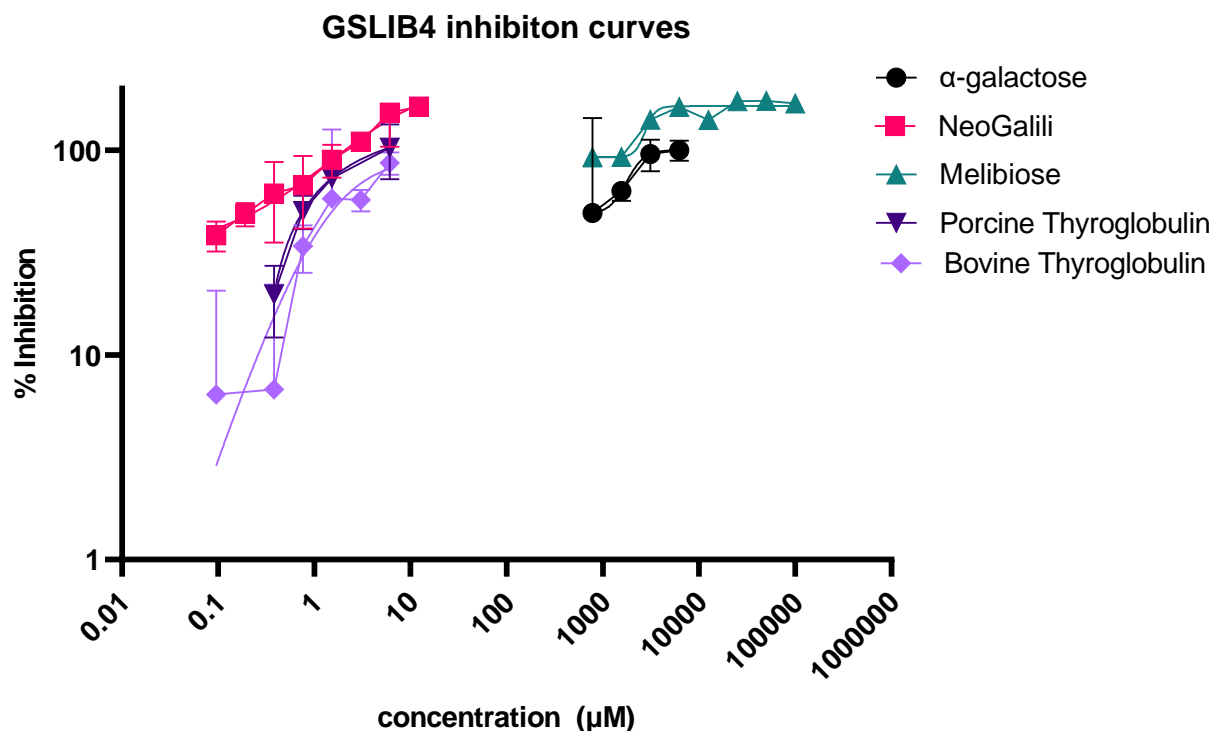


Fig. 3. 18 Direct binding curve of GSLI-B<sub>4</sub> coated on adsorption plate at 1 µg/well interacting with NeoGalili neoglycoprotein applied to the plate at 2.3 µM and serially diluted.

### 3.7.8 GSLI-B<sub>4</sub> inhibition in fluorescence on adsorption plate

The inhibition experiments were designed to understand the ability of a compound to inhibit the binding between GSLI-B<sub>4</sub> and the NeoGalili neoglycoprotein and it follows the same principle explained for MOAβT in Paragraph 3.7.3, this time in fluorescence on adsorption plate. In the experiment in Fig. 3.19 the labelled NeoGalili functions as a tracer applied to the lectin plate at the fixed concentration of 35 µg/mL. In this case, as model compounds for inhibition, we used the glycoproteins porcine and bovine thyroglobulin, the NeoGalili neoglycoprotein, the disaccharide Melibiose (Galα1-6Glc) and the monosaccharide α-galactose. They were used respectively at a starting concentration of 12 µM for the NeoGalili and thyroglobulin and at 100 mM for melibiose and galactose and then serially diluted. The IC<sub>50</sub> could not be calculated for galactose and melibiose because of the signal saturation. It is observed saturation also with NeoGalili, so in order to have a more precise calculation of the IC<sub>50</sub>, the experiment need to be repeated. Even though we performed the inhibition assay on two different plates and with two different detection methods for GSLI-B<sub>4</sub> and MOAβT, it is evident that the first lectin, as described literature [87] [69], shows a far less selective binding, interacting not only with the monosaccharide but also being inhibited to saturation by the melibiose that barely inhibits MOAβT. The explanation is once again in the binding site structure: for GSLI-B<sub>4</sub> it is a deep pocket, less large and less prone to accommodate large glycans.



NeoGalili IC<sub>50</sub>= 0.7  $\mu$ M  
 Porcine Thyroglobulin IC<sub>50</sub>= 1.19  $\mu$ M  
 Bovine Thyroglobulin IC<sub>50</sub>= 1.28  $\mu$ M

*Fig. 3. 19* Inhibition experiment performed with GSLI-B4 interacting with labelled NeoGalili at the fixed concentration of 35  $\mu$ g/mL and with the non-labelled inhibitors applied in serial dilution, from a starting concentration of Porcine Thyroglobulin=12 $\mu$ M, NeoGalili=12  $\mu$ M, Melibiose=100 mM,  $\alpha$ -galactose= 100 mM. The IC<sub>50</sub> for  $\alpha$ -galactose and Melibiose could not be calculated because of the saturation of the signal.

### 3.8 Monoclonal antibodies: a proof of concept

In the Paragraph 3.2.1 we have seen how the presence of the wrong sugar moiety, the Galili antigen on Cetuximab, could affect patient's safety. To validate the kit we have performed the GLYcoPROFILE® of Cetuximab and two other mAbs, Infliximab and Rituximab. Their characteristics are summarized in the *Table 3.3*. These mAbs were chosen because they are extensively used worldwide for diseases treatment, and their glycosylation is well documented [71] [88]. In Fig. 3.20 and 3.21, the panel of chosen lectins is based on the known glycan structures that can be found on the mAbs. A list of all lectins mentioned throughout this thesis and their specificity is enclosed in Annex 2 at the end of Chapter 4. The experiment was performed with native mAbs and digested with total neuraminidase (NeuT) and total neuraminidase+ $\alpha$ -galactosidase, as negative control. The availability of the sample was limited, therefore the concentration of mAbs used for the experiment

could not exceed the 40 µg/mL. This is reflected in the Infliximab GLYcoPROFILE in Fig. 3.20A where the signal for all lectins is below the detection limit so it cannot be considered as significant. The same happens for the Rituximab in Fig. 3.20B for which the most abundant glycan species are known to be G0F, G1F, G2F but the only glycan moiety that is detected is the core fucose interacting with the lectin PSA. For the Cetuximab in Fig. 3.21, despite the low quantity of sample available, we can see a GLYcoPROFILE consistent with published data obtained with LC/MS [71]. One hypothesis for this difference between the three mAbs, that goes beyond the sample concentration, is that the mAbs Infliximab and Rituximab only contain glycans on the Fc domain while the Cetuximab presents the N-glycans also on the Fab region (Fig 3.2). Those are more exposed so it is possible that they are more accessible to the interaction with lectins, whereas the Fc glycans are positioned inside a pocket formed by the two CH2 domain of the antibody (Fig. 4.7, Chapter 4) so they are difficult to be detected. In order to have a quantification of the glycan moieties, we have transformed the fluorescence intensity into neoglycoprotein equivalents (Table 3.4). It means that for each lectin included in this experiment a direct binding curve was performed with a corresponding neoglycoprotein known to have high affinity for the lectin. For each curve a linear regression equation was calculated (Annex I) and from each y values corresponding to the fluorescence intensity of the native antibody interaction with the lectin, a x value was calculated that corresponds to the concentration of the glycan moiety on the mAb, equivalent to the concentration of neoglycoprotein.

Name	Commercial name	Molecular weight	Mechanism	Clinical use	Production	Type of glycans based on literature
<b>Cetuximab</b>	Erbixux	152 kDa including carbohydrates	EGFR inhibitor (Paragraph 3.2.1)	Treatment of metastatic colorectal cancer and head and neck cancer	Murine myeloma cells	G0F, G1F, G2, high mannose, hybrid structures, Low levels of G2Fgal1 (Asn299 Fc domain), three glycan species with NGNA (Asn88 Fd domain) [71].
<b>Infliximab</b>	Remicade	149 kDa	Inhibition of the pro-inflammatory cascade signalling via binding to	Chron's disease, Rheumatoid Arthritis, Ankylosing Spondylitis,	Mouse myeloma cells (SP2/0 cells)	M3G0F, M5, G1, G1F, M6, G2, G2F [88]

			TNF- $\alpha$ with high affinity	Psoriatic arthritis		
Rituximab	Rituxan	145 kDa	Binds to protein CD20, on the surface of immune system B cells triggering cell death	Treatment of patients with stage III-IV follicular lymphoma who are chemoresistant or are in their second or subsequent relapse after chemotherapy	Chinese Hamster Ovary cells	G0, G1F, G2F, low levels of high mannose and NANA sialic acid are detected as well [71]

Table 3. 3 Characteristics of mAbs.

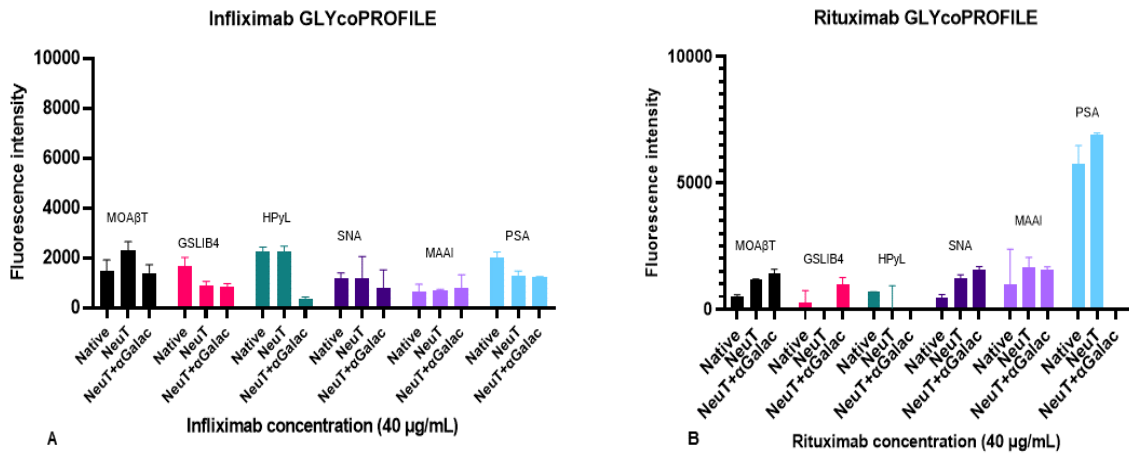


Fig. 3. 20 **A** Infliximab GLYcoPROFILE in direct binding with a panel of lectins. **B** Rituximab GLYcoPROFILE in direct binding with a panel of lectins.

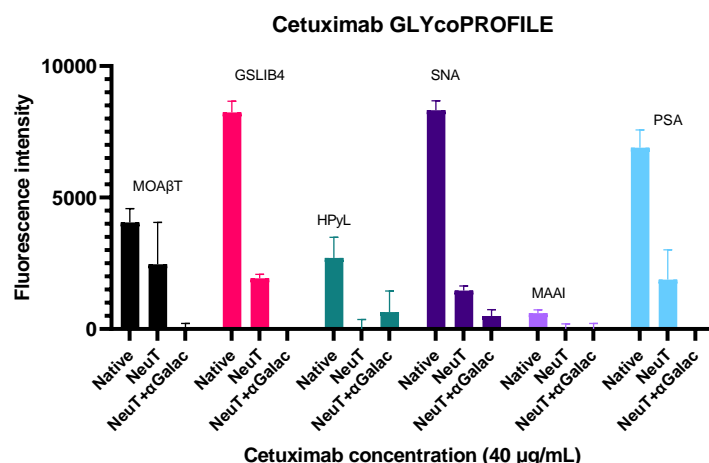


Fig. 3. 21 Cetuximab GLYcoPROFILE in direct binding with a panel of lectins.

	Lectins					
	MOAβT	GSLIB4	HPyL	SNA	MAAI	PSA
Cetuximab	Neoglycoprotein equivalents					
	7.69	0.84	2.2	1.95	0.08	4.89

Table 3. 4 Calculation of Cetuximab equivalents of neoglycoprotein based on the standard curves in Annex I.

## Conclusion

In this Chapter, we provided a complete characterization of the kit in relation with the lectins GSLI-B<sub>4</sub> and MOAβT. The fact that GSL1B4 shows affinity not only for the Galili antigen but also for different α-galactosylated carbohydrates, monosaccharides and disaccharides in particular, could be seen as a drawback in using the lectin as a reliable marker for the xenotransplantation antigen. On another point of view, used jointly with MOAβT that has a strong selectivity for the Galili antigen, could provide information on the presence of species of galactosyl-oligosaccharides/disaccharides/monosaccharides in the sample. Therefore, these two lectins combined are complementary tools. The properties of the β-trefoil domain of MOA lectin were compared with the full MOA. Both proteins showed the same affinity towards NeoGalili confirming that there is no loss of activity due to the lack of the dimerization domain. What is to notice is that MOAβT shows less sensitivity when used in fluorescence on adsorption plate. Therefore, the LEctPROFILE kit can be used in fluorescence on adsorption plate when the two lectins are combined and when other lectins are added to the screening as shown for the tested mAbs (Paragraph 3.8).

When the screening needs to be specifically designed for the high-sensitive detection of the Galili antigen, only MOA $\beta$ T covalently immobilized on the plate is sufficient. Not only this screening can be performed in direct binding labelling the sample of interest or using secondary antibodies, it can also be performed in inhibition using the tested compounds in competition with the tracer NeoGalili, expressing the results as equivalents of neoglycoprotein or comparing the IC<sub>50</sub> to the ones of the reference compounds.

Dans ce chapitre, nous avons fourni une caractérisation complète du kit en relation avec les lectines GSLI-B4 et MOA $\beta$ T. Le fait que GSL1B4 présente une affinité non seulement pour l'antigène Galili mais aussi pour différents glucides  $\alpha$ -galactosylés, monosaccharides et disaccharides en particulier, pourrait être considéré comme un inconvénient à l'utilisation de la lectine comme marqueur fiable de l'antigène de la xénotransplantation. D'un autre point de vue, utilisée conjointement avec la MOA $\beta$ T qui a une forte sélectivité pour l'antigène Galili, pourrait fournir des informations sur la présence d'espèces de galactosyl-oligosaccharides/disaccharides/monosaccharides dans l'échantillon. Par conséquent, ces deux lectines combinées sont des outils complémentaires. Les propriétés du domaine -trefoil de la lectine MOA ont été comparées à celles de la MOA complète. Les deux protéines ont montré la même affinité envers le NeoGalili, confirmant qu'il n'y a pas de perte d'activité due à l'absence du domaine de dimérisation. Ce qui est à noter, c'est que la MOA $\beta$ T montre moins de sensibilité lorsqu'elle est utilisée en fluorescence sur plaque d'adsorption. Par conséquent, le kit LEctPROFILE peut être utilisé en fluorescence sur plaque d'adsorption lorsque les deux lectines sont combinées et lorsque d'autres lectines sont ajoutées au criblage, comme cela a été montré pour les AcM testés (paragraphe 3.8). Lorsque le criblage doit être spécifiquement conçu pour la détection très sensible de l'antigène Galili, seul le MOA $\beta$ T immobilisé de manière covalente sur la plaque est suffisant. Non seulement ce criblage peut être réalisé en liaison directe en marquant l'échantillon d'intérêt ou en utilisant des anticorps secondaires, mais il peut également être réalisé en inhibition en utilisant les composés testés en compétition avec le traceur NeoGalili, en exprimant les résultats en équivalents de néoglycoprotéine ou en comparant les IC<sub>50</sub> à ceux des composés de référence.

Annex I – standard curves

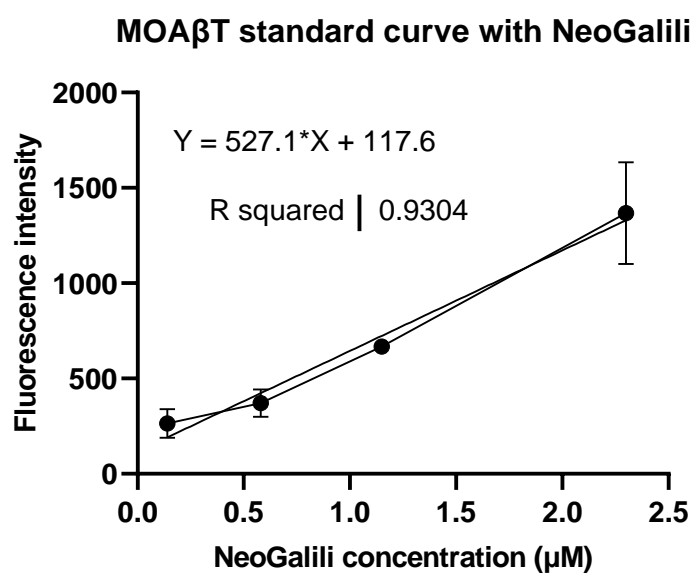


Fig. 3. 22 MOAβT standard curve.



GSLIB4 standard curve with NeoGalili

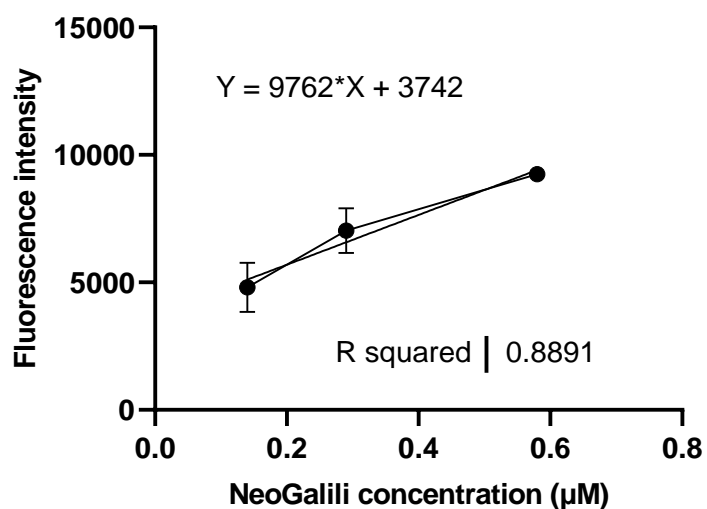


Fig. 3. 23 GS LI-B<sub>4</sub> standard curve.

HPyL standard curve with NeoNeuGc

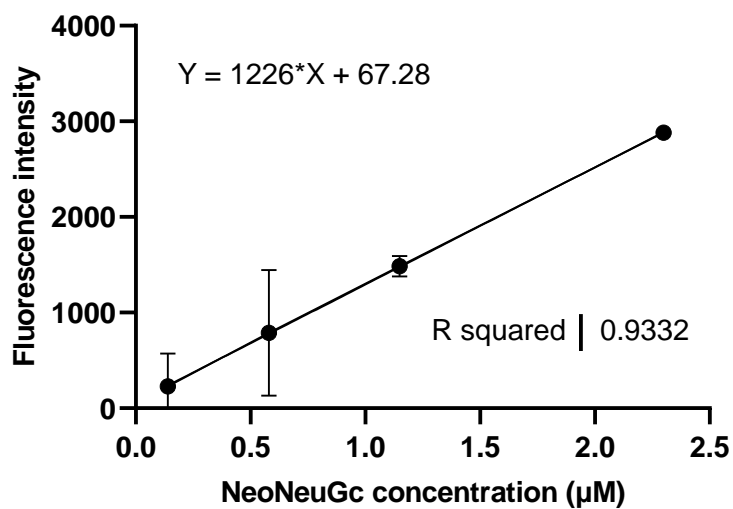


Fig. 3. 24 HPyL standard curve.

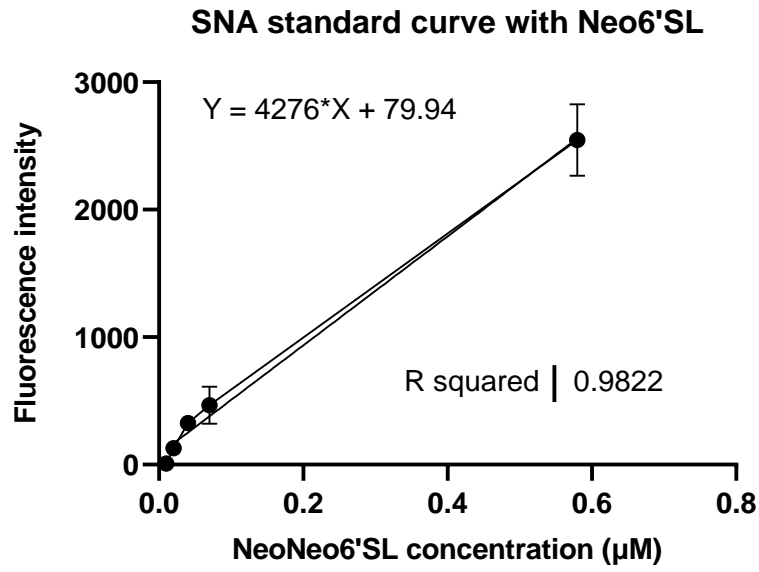


Fig. 3. 25 SNA standard curve.

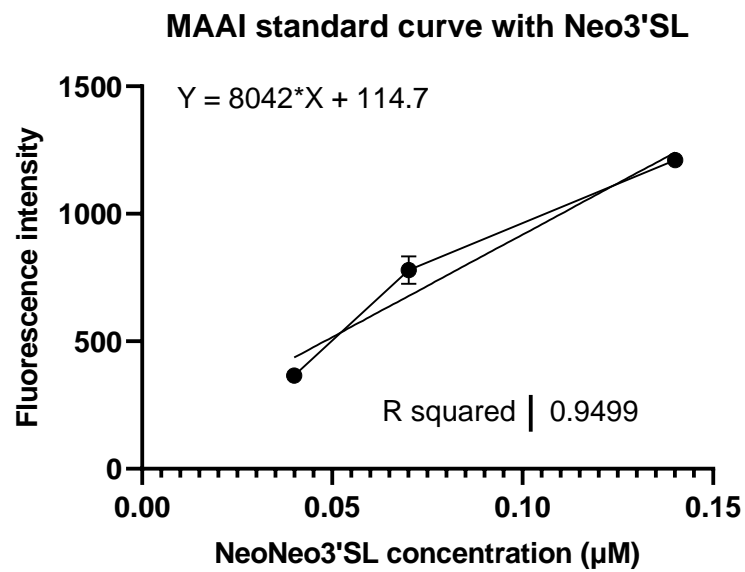


Fig. 3. 26 MAAI standard curve.

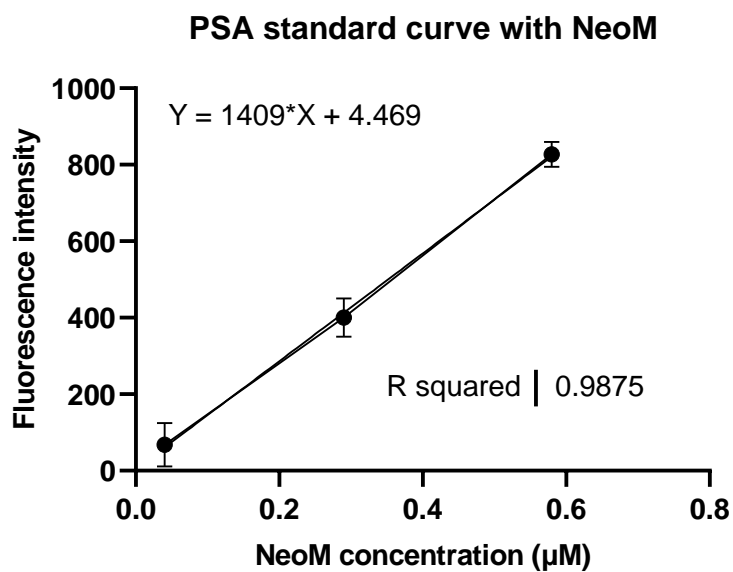


Fig. 3. 27 PSA standard curve.

## Paper 2 - Extending Janus lectins architecture: characterization and application to protocells

This paper is currently on BioRxiv and it was submitted to Computational and StructuralBiotechnology Journal.

### Extending Janus lectins architecture: characterization and application to protocells

Simona Notova <sup>a</sup>, Lina Siukstaite <sup>b,c</sup>, Francesca Rosato <sup>b,c</sup>, Federica Vena <sup>d</sup>, Aymeric Audfray <sup>e</sup>, Nicolai Bovin <sup>f</sup>, Ludovic Landemarre <sup>d</sup>, Winfried Römer <sup>b,c,g</sup>, Anne Imberty <sup>a\*</sup>

<sup>a</sup> Université Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France

<sup>b</sup> Faculty of Biology, University of Freiburg, 79104 Freiburg, Germany

<sup>c</sup> Signalling Research Centers BIOS and CIBSS, University of Freiburg, 79104 Freiburg, Germany

<sup>d</sup> GLYcoDiag, 2 rue du Cristal, 45100 Orléans, France

<sup>e</sup> Malvern Panalytical SAS, 30 rue du docteur Levy, 69200 Vénissieux

<sup>f</sup> Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Science, Moscow 117997, Russian Federation

<sup>g</sup> Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg, 79104 Freiburg, Germany

\* To whom correspondence should be addressed. Anne Imberty (anne.imberty@cermav.cnrs.fr)

### Abstract

Synthetic biology is a rapidly growing field with applications in biotechnology and biomedicine. Through various approaches, remarkable achievements, such as cell and tissue engineering, have been already accomplished. In synthetic glycobiology, the engineering of glycan binding proteins is being developed for producing tools with precise topology and specificity. We developed the concept of chimeric lectins, i.e., Janus lectin, with increased valency, and additional specificity. The novel engineered lectin, assembled as a fusion protein between the  $\beta$ -propeller domain from *Ralstonia solanacearum* and the  $\beta$ -trefoil domain from fungus *Marasmius oreades*, is specific for fucose and  $\alpha$ -galactose and its unique protein architecture allows to bind these ligands simultaneously. The protein activity was tested with glycosylated giant unilamellar vesicles, resulting in the formation of proto-tissue-like structures through cross-linking of such protocells. The synthetic protein binds to H1299 lung epithelial cancer cells by its two domains. The biophysical properties of this new construct were compared with the two already existing Janus lectins, RSL-CBM40 and RSL-CBM77<sub>RL</sub>. Denaturation profiles of the proteins indicate that the fold of each has a significant role in protein stability and should be considered during protein engineering.

**Keywords:** synthetic biology, protein engineering, carbohydrate-binding protein, Galili epitope, lectins

## Introduction

Due to their ability to crosslink various cells, such as red blood cells, lectins are originally known as agglutinins. They are generally multivalent and their selective interaction with glycoconjugates found many applications in the field of biotechnology and biomedicine as glycan-profiling tools (Sharon and Lis 2004). Therefore, fine-tuning the specificity or the valency of lectins is a promising approach for obtaining novel tools. Synthetic biology strategies, such as the engineering of protein architecture, bring novel possibilities for lectin applications (Fettis et al. 2019; Irumagawa et al. 2022; Oh et al. 2022; Ramberg et al. 2021; Ross et al. 2019). Lectin engineering is a novel domain of synthetic glycobiology (Terada et al. 2016; Ward et al. 2021) and it can be performed at different levels, from glycan specificity to supramolecular architecture.

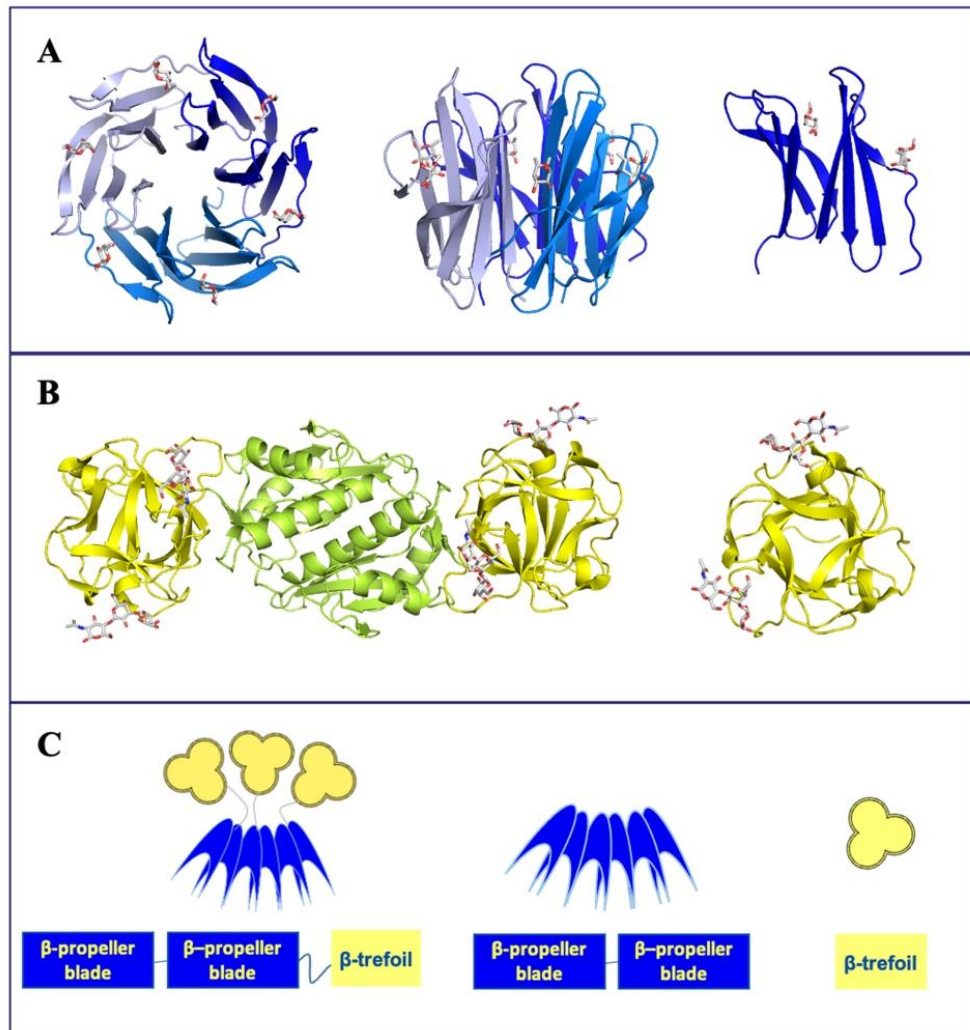
The trimeric oligomerization of *Ralstonia solanacearum* lectin (RSL) in a  $\beta$ -propeller shape (Fig. 1A) was recently used as a scaffold for building Janus lectins (Notova et al. 2022b; Ribeiro et al. 2018) with two faces and different specificities. The two Janus lectins, RSL-CBM40 and RSL-CBM77<sub>Rf</sub>, were obtained by fusion of the monomeric RSL with two carbohydrate-binding modules (CBMs) with different specificity but rather similar shape, consisting of a single  $\beta$ -sandwich domain. This resulted in the creation of synthetic bispecific chimeras able to establish the interaction with fucose on one side and sialic acid (RSL-CBM40) or homogalacturonan (RSL-CBM77<sub>Rf</sub>), respectively, on the other side.

Other protein domains with different specificity, but also topology diverging from the  $\beta$ -sandwich CBM, could be considered for building novel Janus lectins from the trimeric RSL scaffold.  $\beta$ -trefoils are robust domains that contain an internal tandem forming a three-lobed architecture with almost no secondary structures (Murzin et al. 1992). They display a large variety of functions, and many show carbohydrate-binding activity. They are considered as lectins due to their multivalence, but also as CBMs (e.g., CBM13) since they sometimes occur as domains associated with glyco-active enzymes or toxins (Notova et al. 2020; Notova et al. 2022a). While their shape is well conserved, their sequences do not show strong similarity, except for the presence of hydrophobic residues in the core and a QXW repeat found in most sequences (Hazes 1996). A large number of crystal structures of  $\beta$ -trefoil lectins are available (212 from 63 different proteins), and based on their sequences they have been classified into 12 classes in the UniLectin3D database (Bonnardel et al. 2019). Datamining of genomes based on the lobe sequence signature of each class resulted in the prediction of thousands of putative  $\beta$ -trefoil lectins spanning all kingdoms of life (Notova et al. 2022a). The threefold symmetry results in the presence of three carbohydrate-binding sites referred to as  $\alpha$ ,  $\beta$ , and  $\gamma$  although variations in amino acid sequences sometimes result in only one or two active binding sites.

Among the known structures of  $\beta$ -trefoil lectin, the agglutinin from the fairy ring mushroom *Marasmius oreades* (MOA) presents an interesting strict specificity for  $\alpha$ Gal containing oligosaccharides. The lectin rose attention already in the second half of the 20<sup>th</sup> century due to its high selectivity and affinity to blood group B oligosaccharide whereas very poor binding was detected to blood group A or H oligosaccharides (Estola and Elo 1952; Winter et al. 2002).

MOA is specific for oligosaccharides with terminal non-reducing  $\alpha$ 1-3 linked galactose, while galactosides linked in  $\alpha$ 1-2,  $\alpha$ 1-4, and  $\alpha$ 1-6 showed no binding (Winter et al. 2002). Due to its preference for the  $\alpha$ Gal1-3Gal disaccharide, MOA binds efficiently to blood group B (Gal $\alpha$ 1-3[Fuc $\alpha$ 1-2]Gal) but also to linear oligosaccharides such as Gal $\alpha$ 1-3Gal $\beta$ 1-4Glc on glycosphingolipid isoGb3 and Gal $\alpha$ 1-3Gal $\beta$ 1-4GlcNAc (Galili epitope) on glycoproteins (Macher and Galili 2008). This latter epitope is present on the cell surface of most mammals, but not in humans, apes, and Old World monkeys due to the deactivation of the  $\alpha$ 3-galactosyltransferase during evolution (Galili et al. 1988). Humans possess preformed antibodies directed against Gal $\alpha$ 1-3Gal, which are responsible for hyperacute rejection of animal (mainly porcine) organs in xenotransplantation attempts (Cooper et al. 1994). They are also responsible for a strong immune response against biodrugs such as therapeutical antibodies if produced in rodent cell culture with active  $\alpha$ 1-3galactosyltransferase (Chung et al. 2008). Highly selective lectins are therefore needed for the identification of such epitopes (Kruger et al. 2002; Tateno and Goldstein 2004; Winter et al. 2002).





**Figure 1: Schematic representations of Janus lectin RSL-MOA and crystal structures of its individual protein components.** A) Left and middle: Top and side view of the crystal structure of trimeric RSL (blue cartoon) in complex with  $\alpha$ -methyl fucoside (sticks) with in total of six fucose binding sites. Right: structure of RSL monomer (PDB code 2BT9). B) Left: Cartoon representation of the crystal structure of dimeric MOA,  $\beta$ -trefoil domain (yellow) in complex with  $\text{Gal}\alpha 1\text{-}3\text{Gal}\beta 1\text{-}4\text{GlcNAc}$  and dimerization domain (green). Right:  $\beta$ -trefoil domain of MOA with different orientations (PDB code 2IHO). C) Schematic representation of Janus lectin RSL-MOA, RSL, MOA $\beta$ T ( $\beta$ -trefoil domain of MOA), and peptide domains.

The crystal structure of MOA revealed that the lectin assembles as a homodimer with the monomer composed of two distinct domains, adopting  $\beta$ -trefoil fold at the N-terminus and  $\alpha/\beta$  fold at the C-terminus (Fig. 1B). The C-terminal domain serves as a dimerization interface and retains a proteolytic function (Cordara et al. 2016; Juillot et al. 2016). The lectin  $\beta$ -trefoil

domain has three-fold symmetry, and the conserved motif (Gln-X-Trp)<sub>3</sub> is involved in the hydrophobic core of the structure. Co-crystals of MOA with Galα1-3Galβ1-4GlcNAc revealed that each binding site has different ligand occupancy, emphasizing the fact that slight differences in amino acids might affect the binding (Fig. 1B) (Grahn et al. 2007).

In the present study, we designed a Janus lectin as a fusion protein of monomeric RSL at the N-terminus and the MOA β-trefoil domain at the C-terminus (Fig. 1C). We produced, and characterized the Janus lectin RSL-MOA with double specificity toward fucosylated and α-galactosylated glycans. The ability to bind these epitopes was tested with H1299 lung epithelial cancer cells and giant unilamellar vesicles. We also compared the biophysical behavior of this novel Janus-lectin with RSL-CBM40 (Ribeiro et al. 2018) and RSL-CBM77<sub>RF</sub> (Notova et al. 2022b). Additionally, we have engineered the β-trefoil domain of MOA (MOAβT) and compared its activity with RSL-MOA.

## Results

### *Design and production of the β-trefoil domain of MOA (MOAβT)*

MOA occurs naturally as a dimer, assembled by a strong association between the two C-terminal domains. Monomeric MOA composed only of the β-trefoil domain would be of interest for biotechnology applications. The gene of the MOA β-trefoil domain was defined as the 156 N-terminal AAs of MOA full sequence and was amplified by PCR and subcloned into the vector pET TEV as a fusion with an N-terminal His-tag sequence. The resulting protein was named MOAβT and was recombinantly produced in soluble form in the bacterium *E. coli* BL21(DE3). The protein was purified by immobilized metal ion chromatography followed by His-tag cleavage by TEV (tobacco etch virus) protease. The resulting protein has an estimated molecular weight of 17.2 kDa (Supplementary Fig. 1A). TEV cleavage followed by additional immobilized metal ion chromatography reduced the sample contaminants and the protein purity was verified by 12% SDS PAGE electrophoresis (Supplementary Fig. 1A).

### *Design and production of Janus lectin RSL-MOA*

Janus lectin RSL-MOA was designed as a gene fusion of monomeric RSL (N-terminus) and β-trefoil domain of lectin MOA (C-terminus). The fusion of the two protein domains by the eight AAs linker PNGELLSS results in the protein sequence of 255 amino acids. The gene sequence was optimized for bacterial expression, and the synthesized gene was subcloned into the plasmid pET25b+. The protein was produced in soluble form in the bacterium *E. coli* KRX. The protein was subsequently purified by affinity chromatography on an agarose-mannose column due to the interaction between the RSL domain and mannose residues. Protein analysis by 12% SDS PAGE electrophoresis showed that Janus lectin RSL-MOA has an apparent size of 83 kDa, which corresponds to a trimer. RSL oligomerization appears to be resistant to denaturation conditions. A smaller amount of dimers and monomers were also visible, probably because of partial oligomer denaturation (Supplementary Fig. 1B).

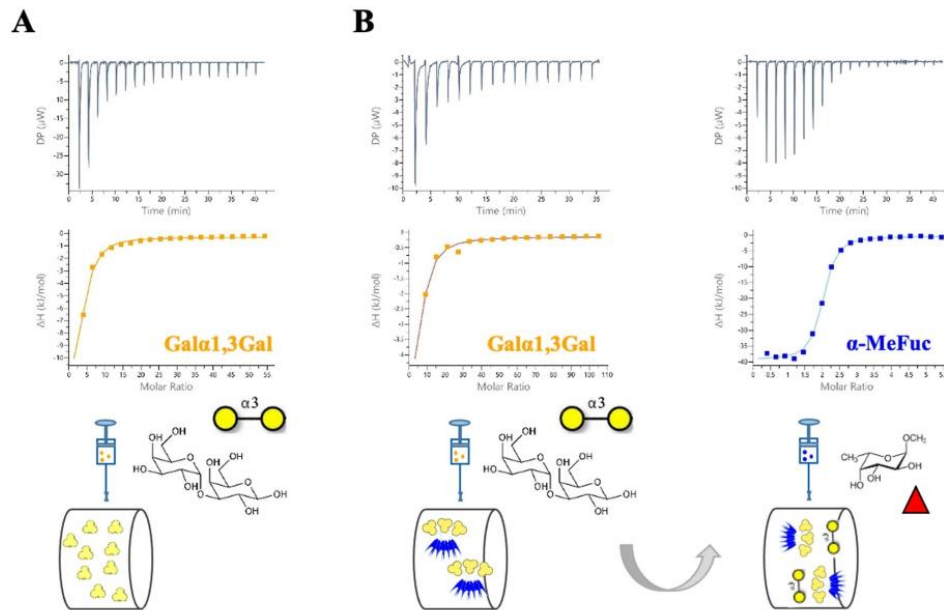


#### *Biophysical characterization of MOA $\beta$ T and RSL-MOA by isothermal titration calorimetry (ITC)*

MOA $\beta$ T affinity towards Gal $\alpha$ 1-3Gal disaccharide, the terminal disaccharide of Galili epitope, was assayed by titration microcalorimetry. Large exothermic peaks were obtained at the beginning of the titration. The affinity was not strong enough to obtain a sigmoidal curve, and therefore, as recommended in such cases (Turnbull and Daranas 2003), the stoichiometry (N) was fixed, using a value of N = 2 since, as previously shown by Grahm et al., (2007), MOA binding sites have different affinities, and we expected at least two of them to be active. A dissociation constant (K<sub>d</sub>) of 150  $\mu$ M was obtained, which confirms the functionality of the isolated  $\beta$ -trefoil domain, and is in good agreement with the K<sub>d</sub> measured for the whole MOA protein, i.e., 182  $\mu$ M (Winter et al. 2002) (Fig. 2A).

RSL-MOA was designed to possess six binding sites for fucose and up to nine binding sites for  $\alpha$ -galactose on the opposite faces. The functionality of both binding interfaces was tested by various biophysical approaches. To this end, we designed an ITC experiment with a consecutive injection of both ligands. First,  $\alpha$ Gal1-3Gal was titrated into the cell containing RSL-MOA. Subsequently, the cell content from the first experiment (complex RSL-MOA/Gal $\alpha$ 1-3Gal) was titrated by  $\alpha$ -methyl fucoside ( $\alpha$ -MeFuc) (Fig. 2B).

The thermogram obtained by titrating Gal $\alpha$ 1-3Gal in RSL-MOA solution was very similar to the one obtained for MOA $\beta$ T. The dissociation constant (K<sub>d</sub> = 226  $\mu$ M) was comparable to the K<sub>d</sub> for the isolated  $\beta$ -trefoil domain (above) and for the whole MOA (Winter et al. 2002). The subsequent titration by  $\alpha$ -MeFuc resulted in a sigmoid shape, due to stronger affinity, with a measured stoichiometry of N= 2, corresponding to the presence of two fucose binding sites per RSL monomer. The K<sub>d</sub> was measured to be 0.4  $\mu$ M, in excellent agreement with the previously measured affinity (0.7  $\mu$ M) for RSL (Kostlanová et al. 2005). This experiment proved that both parts of RSL-MOA are functional and able to bind their ligands at the same time.



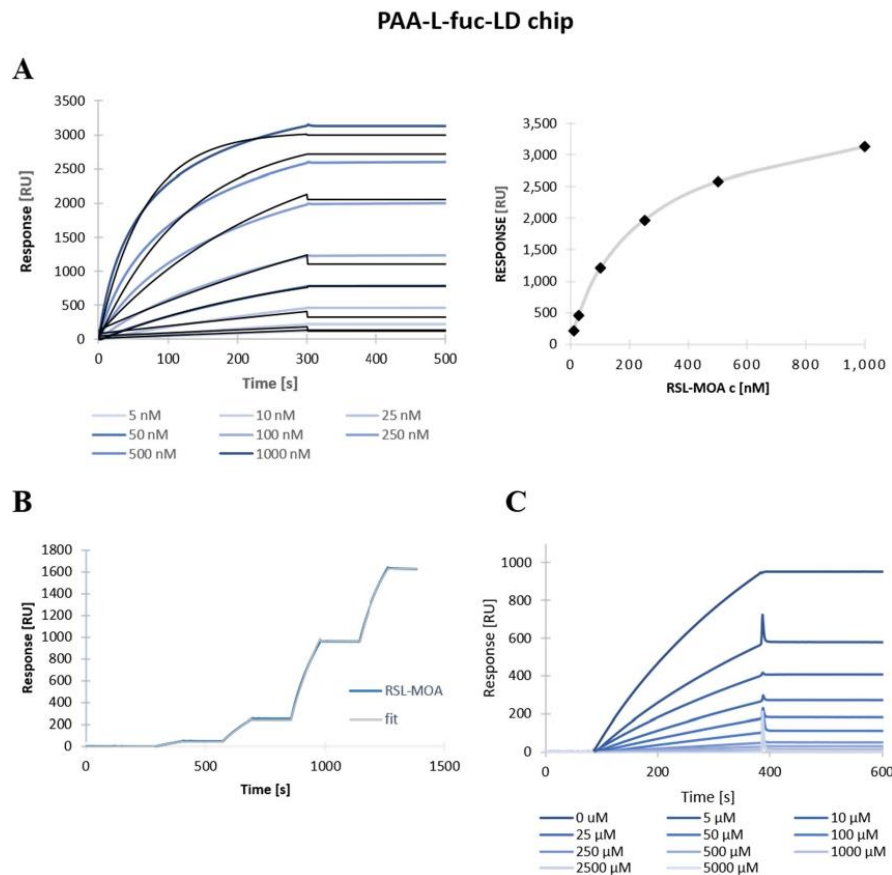
**Figure 2: ITC thermographs of MOAβT and Janus lectin RSL-MOA.** A) MOAβT binds Galα1-3Gal with a similar affinity as parental protein MOA. B) RSL-MOA can simultaneously bind two binding partners, i.e., Galα1-3Gal (yellow circles) and α-MeFuc (red triangle) confirming the activity of both functional domains.

#### Biophysical characterization of RSL-MOA by surface plasmon resonance

To evaluate the effect of multivalency and therefore measure avidity instead of affinity, surface plasmon resonance (SPR) was used. Streptavidin CM5 chips were functionalized by different ligands, i.e., biotinylated PAA-α-fucose, PAA-α-galactose, and PAA-β-galactose.

No binding of RSL-MOA could be observed with α-galactose chip, even with high density surface (200 μg/mL biotinylated PAA-α-Gal), probably because of its very low affinity for the monosaccharide α-galactose ( $K_D = 8$  mM) (Winter et al. 2002). On the opposite, RSL-MOA bound efficiently to fucosylated chips. In order to avoid mass transfer (de Mol and Fischer 2010), low density (LD) CM5 chip was prepared with mixture of PAA-α-fucose/PAA-β-galactose in the ratio 1:9 (200 μg/mL) and RSL-MOA was injected in increasing concentrations (5 to 1000 nM). The regeneration step was performed between each injection using 1 M fucose solution. As shown in Figure 3A, a dose-dependent response was obtained, with a steep association phase and very weak dissociation event. The steady-state analysis showed that even at the highest protein concentration (1000 nM) the chip surface was not saturated by RSL-MOA, i.e., the plateau phase was not reached (Fig. 3A). Nonetheless, the fitting of the kinetics

constant could be performed and values for affinity ( $K_D = 4.9 \times 10^{-10}$  M) and kinetics ( $k_{on}$  and  $k_{off}$ ) were estimated (Table 1).



**Figure 3: SPR sensorgrams of Janus lectin RSL-MOA on CM5-PAA- $\alpha$ -fuc LD chip.** A) Titration experiment of different concentrations of RSL-MOA (blue) and fitting curves (1:1 binding fit) (black) (left) and steady-state analysis of RSL-MOA titration (right). B) Single-cycle kinetics of RSL-MOA with concentrations of 10, 50, 250, and 500 nM. C) Inhibition of 50 nM RSL-MOA by various concentration of fucose.

As an alternative, a single-cycle kinetic experiment was performed. RSL-MOA was injected at four concentrations (10 to 500 nM) (Fig. 3B). The advantage of this method is that no regeneration step is needed. The estimated  $K_D$  of  $5.9 \times 10^{-12}$  M was two orders lower if compared to the titration experiment; however, such a variation is acceptable in binding events characterized by avidity. In order to evaluate the capacity of monosaccharides to inhibit the multivalent binding and compete with protein binding to the chip surface, 50 nM RSL-MOA



was pre-incubated with various concentrations of fucose (5 to 5000  $\mu\text{M}$ ). As shown in Figure 3C, the complete inhibition was achieved in the presence of high concentrations of fucose (2500 and 5000  $\mu\text{M}$ ), and a  $\text{IC}_{50}$  value of 7.8  $\mu\text{M}$  was obtained.

**Table 1: SPR statistics from titration and single-cycle kinetics experiments.** The experiments were carried out on LD PAA- $\alpha$ -Fuc CM5 chips with a flow of 10  $\mu\text{L}/\text{min}$ .

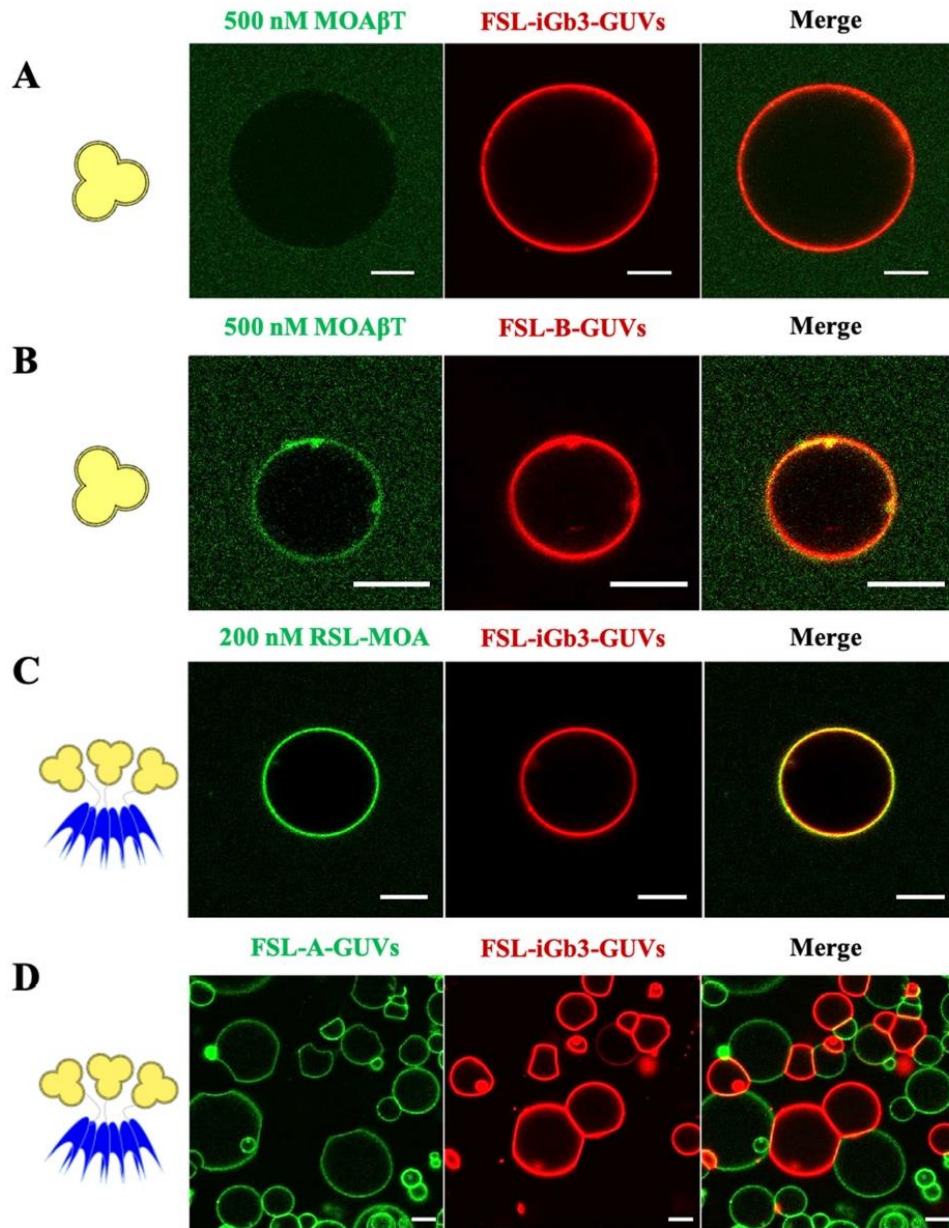
Experimental condition	$k_{\text{on}}$ (1/Ms)	$k_{\text{off}}$ (1/s)	$K_D$ (M)	$R_{\text{MAX}}$ (RU)	$\text{Chi}^2$
Chip CM5 PAA- $\alpha$ -fuc LD Classical kinetics	$4.42 \cdot 10^{-2}$	$2.08 \cdot 10^{-7}$	$4.9 \cdot 10^{-10}$	3030	8.23
Chip CM5 PAA- $\alpha$ -fuc LD Single cycle kinetics	$1.84 \cdot 10^4$	$1.09 \cdot 10^{-7}$	$5.9 \cdot 10^{-12}$	1960	40.4

The ability of RSL-MOA to bind the glycan-decorated surface of SPR chips was confirmed by several experimental setups. The variations in the affinity constant just illustrate the difficulty to quantify binding when avidity is the major event. These results are in agreement with previous ones evaluating RSL avidity for the fucose chip (Arnaud et al. 2013). This avidity constant is at least 1000-fold higher than the affinity measured for fucose monosaccharide in solution, confirming the very strong cluster effect resulting from the topology of the RSL  $\beta$ -propeller with the presentation of six binding sites on the same face, binding very efficiently to fucose presented in a multivalent manner on a surface.

*MOA $\beta$ T binds, while RSL-MOA binds and crosslinks differentially glyco-decorated giant unilamellar vesicles*

The binding properties of MOA $\beta$ T and RSL-MOA were tested with glyco-decorated giant unilamellar vesicles (GUVs). The GUVs were fluorescently labelled with Atto647N- or Atto488-DOPE lipid and contained functional spacer lipid (FSL) with different terminal saccharides attached, i.e. FSL-isoglobotriose, FSL-A, and FSL-B carrying Gal $\alpha$ 1-3Gal $\beta$ 1-4Glc, blood group A and blood group B terminal trisaccharides, respectively (Figure 4 A-D). These glyco-decorated GUVs were incubated with MOA $\beta$ T and RSL-MOA lectins, which were both fluorescently labeled with same Atto488 fluorophore (Figure 4 A-D).

MOA $\beta$ T-Atto488 (500 nM, green) almost did not bind to FSL-iGb3-containing GUVs (red) (Fig. 4A). While, on the contrary, MOA $\beta$ T-Atto488 bound FSL-B-containing GUVs and induced the tubule-like structures (Fig. 4B) indicating its ability to bind glycan structures on the membranes.



**Figure 4: The binding properties of MOA $\beta$ T and RSL-MOA to glyco-decorated GUVs.** A) 500 nM MOA $\beta$ T-Atto488 (green) shows almost no binding to the FSL-iGb3-GUVs (red, labelled with fluorescent lipid Atto647N-DOPE). B) 500 nM MOA $\beta$ T-Atto488 (green) binds to FSL-B-GUVs (red, labelled with fluorescent lipid Atto647N-DOPE). C) 200 nM RSL-MOA-Atto488 (green) binds to FSL-iGb3-GUVs (red, labelled with fluorescent lipid Atto647N-DOPE). D) Unlabeled 200 nM RSL-MOA

*crosslinks two different populations GUVs FSL-iGb3-GUVs (red, labelled with fluorescent lipid Atto647N-DOPE) and FSL-A-GUVs (green, labelled with fluorescent lipid Atto488-DOPE) and crosslinks them. The GUVs were composed of DOPC, cholesterol, glycolipid of choice, and membrane dye to the proportion of 64.7:30:5:0.3 mol%, respectively. Scale bars are 10  $\mu$ m.*

The Janus lectin RSL-MOA-Atto488 (green) binds efficiently to the surface of FSL-iGb3-GUVs (red), even at a concentration of 200 nM (Fig. 4C). The difference is striking when compared to the absence of binding of MOA $\beta$ T-Atto488 (500 nM) (Fig. 4A). The super-multivalency resulting from the presence of three  $\beta$ -trefoil domains, so in total nine possible  $\alpha$ Gal binding sites, makes a very strong difference in terms of ability to bind to the glycosylated surface.

The capacity of RSL-MOA to crosslink GUVs carrying different oligosaccharides was tested using two populations of fluorescently labeled vesicles, FSL-iGb3-GUVs (labelled with Atto647N-DOPE, red) and FSL-A-GUVs (labelled with Atto488-DOPE, green). When the unlabeled RSL-MOA was incubated with the liposomes, cross-linking is observed between red and green GUVs, as well as, cross-linking between GUVs of the same color. The cross-linking between same population GUVs is possibly due to lectin topology and multivalency,  $\beta$ -trefoil of MOA and  $\beta$ -propeller of RSL. Herein, due to the double specificity of lectin RSL-MOA, we confirmed that multivalent and two-site-oriented topology has clear potential in the cross-linking of glyco-decorated liposomes and proto-tissues (Fig. 4D).

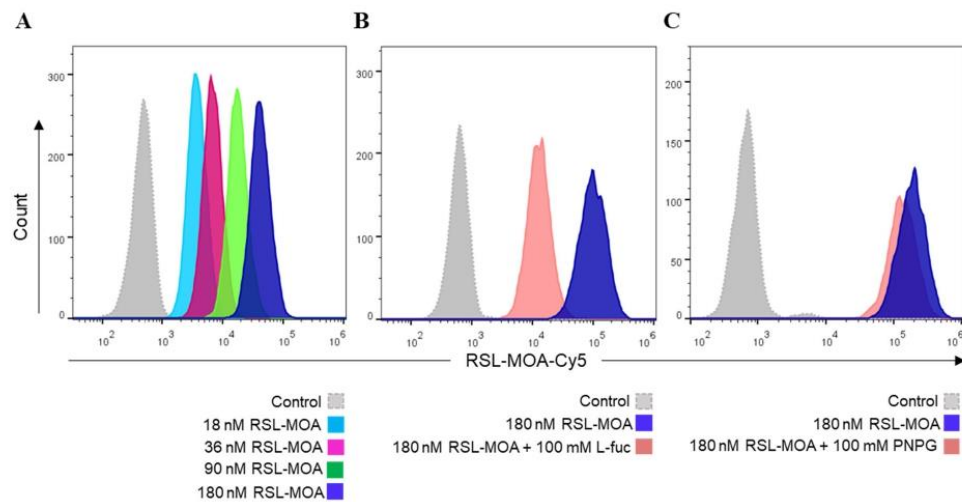
#### *RSL-MOA binds to human epithelial cells*

The ability of RSL-MOA to recognize and bind to glycans on the surface of human cells was investigated by flow cytometry. RSL-MOA was fluorescently labeled with a Cy5 dye (RSL-MOA-Cy5) and incubated with the non-small cell lung cancer cell line H1299, for 30 minutes at 4°C. As depicted in the histograms of fluorescence intensity (Fig. 5A), cells treated with increasing concentrations of RSL-MOA-Cy5 (0.07 – 0.7  $\mu$ M) showed binding of the lectin to the cell surface in a dose-dependent manner. In order to confirm lectin specificity and glycan-driven binding to receptors at the plasma membrane of H1299 cells, a number of inhibition assays was performed. RSL-MOA-Cy5 was pre-incubated with 100 mM fucose or 100 mM synthetic analogue of  $\alpha$ -Gal epitope, p-nitrophenyl- $\alpha$ -D-galactopyranoside (PNPG), respectively. The high concentrations of ligands were chosen due to previous observations with lectin RSL, for which a complete inhibition of binding to fucosylated receptors on the surface of H1299 cells was achieved only in the presence of 100 mM fucose (Siukstaite et al. 2021).

The presence of 100 mM fucose significantly decreased RSL-MOA-Cy5 binding to treated cells by approx. 50% compared to unblocked RSL-MOA-Cy5 (Fig. 5B), but did not abolish it entirely. Increasing the concentration up to 500 mM of fucose did not result in a further decrease in protein binding. The interaction of the engineered lectin with the cell surface was therefore not completely inhibited upon saturation of the fucose-binding sites, suggesting that the MOA domain partially compensates for the binding of Janus lectin to H1299 cells. On the



other hand, 100 mM PNPG had almost no effect on RSL-MOA-Cy5 binding to treated cells, as indicated by the histograms in Fig. 5C. Fluorescence of cells incubated with RSL-MOA-Cy5 or RSL-MOA-Cy5 pre-treated with PNPG exhibit a similar intensity, as illustrated by the overlapping histograms (Fig. 5C). Reasonably, inhibition of protein binding was not achieved due to the low affinity of MOA toward monosaccharides (Winter et al. 2002). These results suggest that both RSL and MOA domains of Janus lectin are able to bind to glycans present on H1299 cancer cells.

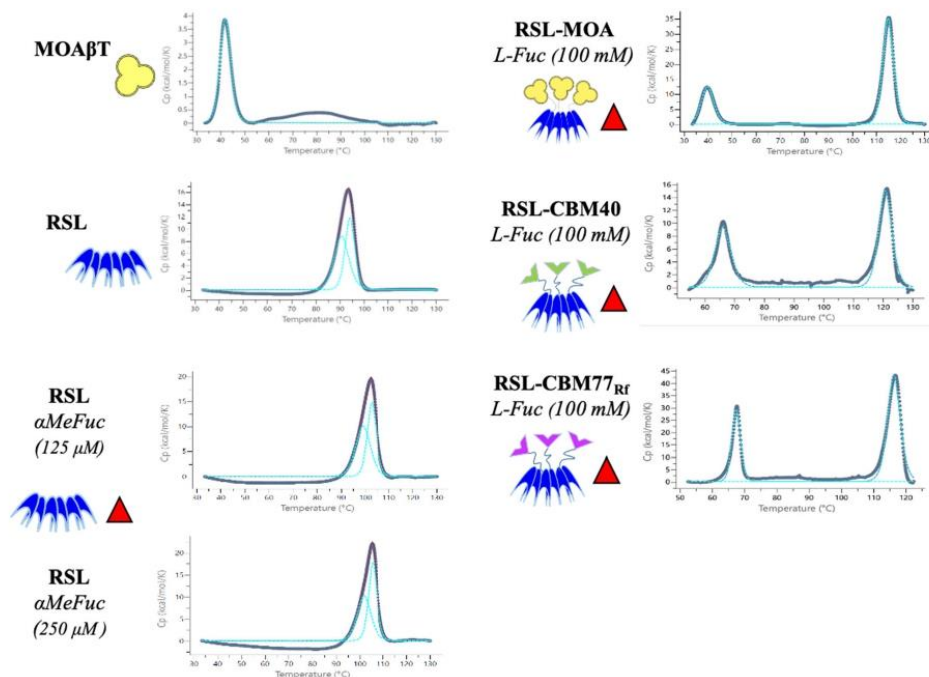


**Figure 5: RSL-MOA shows dose-dependent binding to H1299 cells.** A) Representative histogram plot of gated living H1299 cells pre-incubated with fluorescently labeled RSL-MOA-Cy5. Histograms of fluorescence intensity reveal a dose-dependent trend in binding according to used lectin concentrations. B) Representative histograms of fluorescence intensity of H1299 treated with RSL-MOA-Cy5 pre-incubated with 100 mM fucose. Inhibition of the RSL domain with high concentrations of fucose determines a shift in fluorescence intensity towards lower values (pink histogram), suggesting a reduced binding of lectin to H1299 cells. C) Representative histograms of fluorescence intensity of H1299 treated with RSL-MOA-Cy5 pre-incubated with 100 mM PNPG shows almost no reduction in lectin binding to cell surface. The control experiments correspond to samples of H1299 cells without any addition of lectin sample. The number of cells within the live population (y-axis) is plotted against the fluorescence intensity of RSL-MOA-Cy5 (x-axis).

#### Comparison of thermal stability of Janus lectins and their distinct domains

Janus lectins are created by assembling protein domains with different biophysical properties. It is therefore of interest to evaluate the effect of the fusion of these modules, and also to compare the biophysical properties of Janus lectins.

The thermal stability of RSL-MOA and its constituting domains was measured with differential scanning calorimetry (DSC). Thermograms of MOA $\beta$ T and RSL are displayed in Figure 6. These two modules have very different thermal stability with a denaturation/midpoint temperature ( $T_m$ ) of 42°C for the MOA $\beta$ T and 92°C (with two  $T_m$  at 90.5°C and 93.8°C) for the  $\beta$ -propeller RSL. It could be that MOA $\beta$ T is easily denatured as it represents only a domain of the whole original protein. The very strong thermal resistance of RSL is related to the compact and robust architecture of the  $\beta$ -propeller, and the two slightly different  $T_m$ s extracted from the curve may correspond to the dissociation of the trimer followed by the unfolding of each monomer. As expected, adding  $\alpha$ -MeFuc as a ligand to the protein solution increased the  $T_m$  of RSL to values of approximately 100°C and 103°C, for sugar concentrations of 125 and 250  $\mu$ M, respectively.



**Figure 6: Thermal stability analysis of MOA $\beta$ T, RSL, and Janus lectins RSL-MOA, RSL-CBM40, and RSL-CBM77<sub>Rf</sub> with or without ligands. The figure displays experimental conditions, fitted data, and schematic representation of proteins and ligands.**

The thermal unfolding of RSL-MOA and other Janus lectins was carried out in the presence of a high concentration of fucose since the protein was tested after the purification procedure. RSL-MOA displayed two events of denaturation at very different temperatures, i.e., 40°C and 115°C. From the results obtained on the separated domains, it is clear that the lower  $T_m$  corresponds to the unfolding of the MOA  $\beta$ -trefoil and the second event, with higher  $T_m$ , corresponds to the denaturation of RSL. The  $T_m$  of the MOA moiety is not different when



isolated or linked to RSL. The  $T_m$  of the RSL moiety is higher than for the  $\beta$ -propeller alone, which is probably due to the high concentration of fucose in the medium.

**Table 2: Thermal signatures of denaturation profiles of MOA $\beta$ T, RSL, and Janus lectins. The scan rate was fixed at 200 °C/min.**

Protein	Ligand	Protein (g/L & $\mu$ M)	$T_{m1}$ (°C)	$T_{m2}$ (°C)
MOA $\beta$ T	-	0.5 (29)	41.8	-
	-	0.25 (25)	90.5	93.8
RSL	125 $\mu$ M $\alpha$ MeFuc	0.25 (25)	98.9	102.5
	250 $\mu$ M $\alpha$ MeFuc	0.25 (25)	101.6	105.3
RSL-MOA	100 mM Fuc	0.25 (9)	39.5	114.5
RSL-CBM77 <sub>Rf</sub>	100 mM Fuc	0.5 (18)	67.4	116.6
RSL-CBM40	100 mM Fuc	0.5 (16)	65.5	121.1

Other Janus lectins were also assayed by DSC for comparison. The proteins RSL-CBM40 and RSL-CBM77<sub>Rf</sub> were obtained as described previously (Notova et al. 2022b; Ribeiro et al. 2018) and the buffer was supplemented by fucose as in the case of RSL-MOA. As already experienced with RSL-MOA, two separate events of denaturation were observed (Fig. 6) (Table 2). The high-temperature  $T_m$ , corresponding to the denaturation of RSL, is observed, with a value as high as 121°C for RSL-CBM40. CBM77<sub>Rf</sub> and CBM40 have a similar  $\beta$ -sandwich fold, and they present rather similar melting temperatures of 67°C and 65°C, respectively. This value is more than 25 °C higher than the one observed for MOA, indicating a clear difference in thermal stability between  $\beta$ -sandwich and  $\beta$ -trefoil.

## Discussion and Conclusion

In this work, we extended the concept of Janus lectin and therefore developed a universal strategy for increasing lectin valency and introducing an additional specificity (Notova et al. 2022b; Ribeiro et al. 2018). Janus lectins are engineered as fusion chimeras and due to the presence of the lectin RSL, these synthetic proteins assemble as trimers, resulting in the multiplication of lectin binding sites and thus expected higher affinity toward ligands. The first two Janus lectins, RSL-CBM40 (Ribeiro et al. 2018; Siukstaite et al. 2021) and RSL-CBM77<sub>Rf</sub> (Notova et al. 2022b) were designed as fusion chimeras of a lectin and CBM domain with a possible application as drug carriers or protein crosslinker in plant cell wall engineering,

respectively. However, here we propose an alternative where the Janus lectin RSL-MOA is composed of two individual lectin domains,  $\beta$ -trefoil from lectin MOA and  $\beta$ -propeller of RSL.

The biophysical properties of RSL-MOA were compared with MOA $\beta$ T, the engineered  $\beta$ -trefoil domain of MOA lectin. Both proteins showed the same affinity toward ligands in solution (ITC experiments) confirming their activity. However, their behavior was different when tested with glyco-decorated liposomes (GUV experiments). We observed that MOA $\beta$ T showed almost no binding to FSL-iGb3-GUVs while RSL-MOA interacts with such vesicles and even crosslinks them. This is probably conditioned by the super-multivalency of engineered RSL-MOA that presents nine binding sites, instead of only three in MOA $\beta$ T. On the other hand, MOA $\beta$ T displayed active binding with blood group B GUVs (FSL-B-containing GUVs) confirming that the lectin domain is able to bind glycoconjugates on the membrane surface. RSL-MOA also has the ability to bind two different GUVs and crosslink them. This confirms that the engineered topology does not affect protein binding and such a lectin is of interest for the development of a highly selective tool for the construction of proto-tissues.

The SPR analysis of RSL-MOA revealed the strong dose-dependent binding of the  $\beta$ -propeller RSL toward low-density PAA- $\alpha$ -Fuc CM5 chips. Even though almost no dissociation event was observed, the fitting procedure allowed for the estimation of binding constants. Depending of the experimental design, some variations are observed in affinity and kinetics, due to the strong avidity of the system, generating complexity. Nonetheless, RSL-MOA showed at least 1000 higher affinity for surface-presented oligosaccharides compared to the solution state, supporting the fact that the addition of MOA does not affect its binding properties.

The ability of RSL-MOA to recognize glycans exposed at the surface of H1299 cells was analyzed by flow cytometry, demonstrating a strong dose-dependent binding. H1299 cells are characterized by the presence of highly fucosylated glycoconjugates, among others, on their surface (Jia et al. 2018). Furthermore, the effect of saturation of RSL-MOA sugar-binding sites with 100 mM fucose or 100 mM PNPG, was checked and resulted in a decrease in the interaction between lectin and cells. We believe that the presence of the galactose-specific MOA domain partially compensates the inhibition of fucose-dependent interaction between the lectin and the cell surface. Our hypothesis is supported by previous observations with the lectin RSL, which showed similar binding properties to H1299 cells that could be fully inhibited by 100 mM fucose. On the other hand, 100 mM PNPG did not affect RSL-MOA binding, suggesting that the synthetic analogue of  $\alpha$ -galactose (PNPG) did not bind MOA with sufficient affinity, as already observed previously with  $\alpha$ -galactose monosaccharide (Winter et al. 2002) and confirmed by SPR observations in the present work.

The thermal stabilities of three Janus lectins, i.e., RSL-MOA, RSL-CBM40, and RSL-CBM77<sub>RE</sub>, were compared by DSC. During protein denaturation, two events of unfolding are observed for all Janus lectins implying the fact that each protein domain has a different stability. Structurally similar CBMs ( $\beta$ -sandwich fold) share almost the same  $T_m$ , while  $\beta$ -trefoil is much more thermally unstable. On the other hand, the  $\beta$ -propeller RSL showed extremely high

T<sub>m</sub>, whereas the addition of the ligand has a tremendous effect on its stability. Our findings, therefore, suggest that the structural fold of the individual domains should be taken into account when designing novel Janus lectins if stability is to be considered.

Until now, three Janus lectin with different specificities and architecture have been engineered. Therefore, we are confident that with this strategy, the novel bispecific lectins with improved valency can be obtained and find their applications in numerous fields of biotechnology and biomedicine.

## Methods

### Gene design and cloning

#### MOAβT

The gene for the β-trefoil domain of MOA was obtained by polymerase chain reaction where the plasmid pET-25b+RSL-MOA was used as a template. The primers, TTACATATGAGCTTACGTCGTGGC (Forward) and ATTACTCGAGTTACATGCGGTTGAAGTACC (Reverse) were designed to align to the full sequence of the β-trefoil domain of MOA and were ordered from Eurofins Genomics (Ebersberg, Germany). The restriction enzyme sites of NdeI and XhoI were added at 5' and 3' ends, respectively. Subsequently, the gene *moaβt* and plasmid pET-TEV vector (Houben et al. 2007) were digested by RE NdeI and XhoI and ligated resulting in the pET-TEV-MOAβT vector. After transformation by heat shock in the *E. coli* DH5α strain, a colony screening was performed, and the positive plasmids were amplified and controlled by sequencing.

#### RSL-MOA

The original amino acid sequence of the lectin MOA from *Marasmius oreades* was obtained from the PDB database. The gene *rsl-moa* was designed as a fusion chimera with RSL at the N-terminus and the β-trefoil domain of MOA at the C-terminus via the linker PNGELLSS. The gene was ordered from Eurofins Genomics (Ebersberg, Germany) after codon optimization for the expression in the bacteria *Escherichia coli*. The restriction enzyme sites of NdeI and XhoI were added at 5' and 3' ends, respectively. The synthesized gene was delivered in plasmid pEX-A2-RSL-MOA. Subsequently, plasmid pEX-A2-RSL-MOA and the pET-25b+ were digested by NdeI and XhoI restriction enzymes to ligate *rsl-moa* in pET-25b+. After transformation by heat shock in the *E. coli* DH5α strain, a colony screening was performed, and the positive plasmids were amplified and controlled by sequencing.

### Protein expression

*E. coli* BL21(DE3) cells were transformed by heat shock with pET-TEV-MOAβT plasmid prior pre-culture in Luria Broth (LB) media with 25 µg/mL kanamycin at 37°C under agitation at 180 rpm overnight. The next day, 10 mL of preculture was used to inoculate 1 L LB medium with 25 µg/mL kanamycin at 37°C and agitation at 180 rpm. When the culture reached OD<sub>600nm</sub>



of 0.6 - 0.8, the protein expression was induced by adding 0.1 mM isopropyl  $\beta$ -D-thiogalactoside (IPTG), and the cells were cultured at 16°C for 20 hours.

*E. coli* KRX (Promega) cells were transformed by heat shock with the pET-25b+-RSL-MOA plasmid and pre-cultured in LB media substituted with 50  $\mu$ g/mL ampicillin at 37°C under agitation at 180 rpm overnight. The following day, 10 mL of preculture was used to inoculate 1 L LB medium with 50  $\mu$ g/mL ampicillin at 37°C and agitation at 180 rpm. When reached an OD<sub>600nm</sub> of 0.6 - 0.8, the protein expression was induced by adding 1% L-rhamnose, and the cells were cultured at 16°C for 20 hours.

The cells were harvested by centrifugation at  $14000 \times g$  for 20 min at 4°C and the cell paste was resuspended in 20 mM Tris/HCl pH 7.5, 100 mM NaCl (Buffer A), and lysed by a pressure cell disruptor (Constant Cell Disruption System) with a pressure of 1.9 kBar. The lysate was centrifuged at  $24\,000 \times g$  for 30 min at 4°C and filtered on a 0.45  $\mu$ m syringe filter prior to loading on an affinity column.

### Protein purification

#### MOA $\beta$ T

The cell lysate was loaded on 1 mL HisTrap column (Cytiva) pre-equilibrated with Buffer A. The column was washed with Buffer A to remove all contaminants and unbound proteins. The MOA was eluted by Buffer A in steps during which the concentration of imidazole was increased from 25 mM to 500 mM. The fractions were analyzed by 12% SDS PAGE and those containing MOA $\beta$ T were collected and deprived of imidazole by dialysis in Buffer A. The N-terminal His-tag was removed by TEV cleavage with the ratio 1:50 mg of TEV:protein in the presence of 0.5 mM EDTA and 1 mM TCEP over night at 19°C. After, the protein mixture was repurified on 1 mL HisTrap column (Cytiva) and the pure protein was concentrated by Pall centrifugal device with MWCO 3 kDa and stored at 4°C.

#### RSL-MOA

The cell lysate was loaded on 10 mL D-mannose-agarose resin (Merck) pre-equilibrated with Buffer A. The column was washed with Buffer A to remove all contaminants and unbound proteins and the flow-through was collected. RSL-MOA was eluted by Buffer A with the addition of 100 mM D-mannose or 100 mM L-fucose in one step. Due to not sufficient binding capacity of the column, the flow-through was reloaded on the column several times and the protein was eluted as described previously. The fractions were analyzed by 12% SDS PAGE and those containing RSL-MOA were collected and dialyzed against Buffer A. The protein was concentrated by Pall centrifugal device with MWCO 30 kDa and the pure protein fractions were pooled, concentrated, and stored at 4°C.

### Isothermal Titration Calorimetry (ITC)

ITC experiments were performed with MicroCalITC200 (Malvern Panalytical). Experiments were carried out at  $25^\circ\text{C} \pm 0.1^\circ\text{C}$ . Protein and ligand samples were prepared in Buffer A. The

ITC cell contained proteins in a concentration range from 0.05 mM to 0.2 mM. The syringe contained the ligand solutions in a concentration from 50  $\mu$ M to 10 mM. 2  $\mu$ L of ligands solutions were injected into the sample cell at intervals of 120 s while stirring at 750 rpm. Integrated heat effects were analyzed by nonlinear regression using one site binding model (MicroCal PEAQ-ITC Analysis software). The experimental data were fitted to a theoretical curve, which gave the dissociation constant (Kd) and the enthalpy of binding ( $\Delta$ H).

### Surface plasmon resonance (SPR)

The SPR experiments were performed using a Biacore X100 biosensor instrument (GE Healthcare) at 25°C. Biotinylated polyacrylamide-attached (PAA) sugars, such as PAA- $\alpha$ -fucose, PAA- $\alpha$ -galactose, and PAA- $\beta$ -galactose (Lectinity) were immobilized on CM5 chips (GE Healthcare) pre-coated with streptavidin, as previously described (Ribeiro et al., 2016). In the sample cell, the PAA-sugars were immobilized either as a mixture for low-density chips (1:9 of sugar of interest:non-interacting sugar) or as an individual PAA-sugar. The immobilization levels of chip CM5 LD PAA- $\alpha$ -fucose:PAA- $\beta$ -galactose (1:9) were FC1: streptavidin 3694 RU, PAA- $\beta$ -galactose 49 RU, FC2: streptavidin 3541 RU, PAA- $\alpha$ -fucose:PAA- $\beta$ -galactose 596 RU. The solutions were prepared at a concentration 200  $\mu$ g/mL in 10 mM HEPES buffer pH 7.5 with 100 mM NaCl and 0.05% Tween 20 (Buffer S). The reference cell was prepared in the same way as described for the sample cell, however, this time only non-interacting sugars were used. All the experiments were carried out in Buffer S. For the titration experiments, different concentrations of protein were injected onto the chip with the flow of 10 or 30  $\mu$ L/min, and the protocol included steps: association time 300 s, dissociation time 300 s, 2x regeneration step for 180 s by 1 M fucose. The single-cycle kinetics experiment was carried out with the flow of 10  $\mu$ L/min with an association and dissociation time of 120 s, while the regeneration was done only at the end of the run. For the inhibition assay, RSL-MOA with the concentration of 50 nM was pre-incubated with various concentrations of L-fucose for a minimum of 1 h at 4°C and subsequently injected onto the fucose chip. The used protocol was identical to the one used for titration experiments. The data analysis was performed by BIAevaluation software.

### Protein labeling

RSL-MOA and MOA $\beta$ T were dissolved at 1 mg/mL in Dulbecco's phosphate-buffered saline (PBS) and stored at 4°C prior to usage. For fluorescent labeling, NHS-ester conjugated Atto488 (Thermo Fisher) or Cy5 (GE Healthcare) were used. Fluorescent dyes were dissolved at a final concentration of 10 mg/mL in water-free DMSO (Carl Roth GmbH & Co), aliquoted, and stored at -20 °C before usage according to the manufacturer's protocol. For the labeling reaction, 100  $\mu$ L of lectin (1 mg/mL) was supplemented with 10  $\mu$ L of a 1 M NaHCO<sub>3</sub> (pH 9.0) solution. Hereby, the molar ratio between dye and lectin was 5:1 for RSL-MOA (cell assays) and 2:1 for RSL-MOA and MOA $\beta$ T for GUV assays. The labeling mixture was incubated at 4°C for 90 min, and uncoupled dyes were separated using Zeba Spin<sup>TM</sup> desalting columns (7 kDa MWCO, 0.5 mL, Thermo Fischer). Labeled lectins were stored at 4°C, protected from light.



### **Composition and preparation of giant unilamellar vesicles (GUVs)**

GUVs were composed of 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), cholesterol (both AvantiPolar Lipids, United States), Atto647N 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE; Sigma Aldrich), Atto488 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE; Sigma Aldrich) and one of the following glycolipids at a molar ratio of 64.7:30:0.3:5. The glycolipids are FSL-A (Function-Spacer-Lipid with blood group A trisaccharide) (SigmaAldrich), FSL-B (Function-Spacer-Lipid with blood group B trisaccharide) (Sigma Aldrich) or FSL-isoGb3 (Function-Spacer-Lipid with iso-globotriaosyl saccharide).

GUVs were prepared by the electroformation method as earlier described (Madl et al., 2016). Briefly, lipids dissolved in chloroform of a total concentration of 0.5 mg/mL were spread on indium tin oxid-covered (ITO) glass slides and dried in a vacuum for at least one hour or overnight. Two ITO slides were assembled to create a chamber filled with sucrose solution adapted to the osmolality of the imaging buffer of choice, either HBSS (for live-cell imaging) or PBS (for GUV-only imaging). Then, an alternating electrical field with a field strength of 1 V/mm was implemented for 2.5 hours at RT. Later we observed the GUVs in chambers manually built as described (Madl et al., 2016).

### **Imaging of RSL-MOA and MOA $\beta$ T binding to GUVs**

Samples containing GUVs and lectins were imaged using a confocal fluorescence microscope (Nikon Eclipse Ti-E inverted microscope equipped with Nikon A1R confocal laser scanning system, 60x oil immersion objective, NA = 1.49, and four laser lines: 405 nm, 488 nm, 561 nm, and 640 nm). Image acquisition and processing were made using the software NIS-Elements (version 4.5, Nikon) and open-source Fiji software (<https://imagej.net/software/fiji/>).

### **Cell culture**

The human lung epithelial cell line H1299 (American Type Culture Collection, CRL-5803) was cultured in Roswell Park Memorial Institute (RPMI) medium supplemented with 10% fetal calf serum (FCS) and 4 mM L-glutamine at 37°C and 5% CO<sub>2</sub>, under sterile conditions. Cells were cultivated in standard TC-dishes 100 (Sarstedt AG & Co. KG) until 90% confluence, detached with trypsin (0.05% trypsin-EDTA solution; Sigma-Aldrich Chemie GmbH) and re-seeded for a subculture or for experiments. For experiments, cells were incubated with different concentrations of RSL-MOA for indicated time points.

### **Flow Cytometry Analysis**

H1299 cells were detached with 2 mL of 1.5 mM EDTA in PBS, and  $1 \times 10^5$  cells were counted and transferred to a U-bottom 96 well plate (Sarstedt AG & Co. KG). To quantify protein binding to cell surface receptors, cells were incubated with different concentrations of fluorescently labeled RSL-MOA-Cy5 lectin for 30 min at 4°C and protected from light compared to PBS-treated cells as a negative control. For the saturation of glycan-binding sites, 180 nM RSL-MOA-Cy5 was preincubated with 100 mM soluble L-fucose or with 100 mM 4-

Nitrophenyl  $\alpha$ -D-galactopyranoside (PNPG; Sigma-Aldrich, Chemie GmbH), for 30 min at RT and in the absence of light. At the end of pre-incubation, the solution was diluted 100 times and added to cells for 30 min at 4°C, in the dark. Subsequently, cells were centrifuged at 1600  $\times$  g for 3 min at 4°C and washed twice with FACS buffer (PBS supplemented with 3% FCS v/v). After the last washing step, the cells were resuspended with FACS buffer and transferred to FACS tubes (Kisker Biotech GmbH Co. KG) on ice and protected from light. The fluorescence intensity of treated cells was monitored at FACS Gallios (Beckman Coulter Inc.) and further analyzed using FlowJo V.10.5.3.

### Differential scanning calorimetry (DSC)

The DSC experiments were carried out in Micro-Cal PEAQ DSC instrument (Malvern Panalytical). The protein samples were prepared in Buffer A, which could be substituted with the addition of a ligand. The sample cell contained the protein solutions in the concentration range of 9-29 mM. The reference cell contained the same buffer as present in the sample but without protein. The increase in temperature was measured from 20-130°C with a scan rate of 200°C/min. The data were analyzed by Micro-Cal PEAQ DSC software with a non-two-state and progress baseline method. fitting model.

**Acknowledgment:** This research was funded by the European Union Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement synBIOcarb (No. 814029). AI acknowledges support from Glyco@Alps (ANR-15-IDEX-02) and Labex Arcane/CBH- EUR-GS (ANR-17-EURE-0003). Moreover, WR acknowledges support by the Ministry for Science, Research and Arts of the State of Baden-Württemberg (Az: 33-7532.20), and by the Freiburg Institute for Advanced Studies (FRIAS). This publication is partially based upon work from COST Action CA18103 (INNOGLY), supported by COST (European Cooperation in Science and Technology).

### Bibliography

- Arnaud J, Claudinon J, Tröndle K, Trovaslet M, Larson G, Thomas A, Varrot A, Römer W, Imberty A, Audfray A. 2013. Reduction of lectin valency drastically changes glycolipid dynamics in membranes but not surface avidity. *ACS Chem. Biol.*, 8:1918-1924, 10.1021/cb400254b
- Bonnardel F, Mariethoz J, Salentin S, Robin X, Schroeder M, Pérez S, Lisacek F, Imberty A. 2019. UniLectin3D, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands. *Nucleic Acids Res.*, 47:D1236–D1244, 10.1093/nar/gky832
- Chung CH, Mirakhur B, Chan E, Le QT, Berlin J, Morse M, Murphy BA, Satinover SM, Hosen J, Mauro D, *et al.* 2008. Cetuximab-induced anaphylaxis and IgE specific for galactose- $\alpha$ -1,3-galactose. *N. Engl. J. Med.*, 358:1109-1117, 10.1056/NEJMoa074943
- Cooper DK, Koren E, Oriol R. 1994. Oligosaccharides and discordant xenotransplantation. *Immunol. Rev.*, 141:31-58, 10.1111/j.1600-065x.1994.tb00871.x



- Cordara G, Van Eerde A, Grahn EM, Winter HC, Goldstein IJ, Krenzel U. 2016. An unusual member of the papain superfamily: Mapping the catalytic cleft of the *Marasmius oreades* agglutinin (MOA) with a caspase inhibitor. *PLoS ONE*, 10.1371/journal.pone.0149407
- de Mol NJ, Fischer MJ. 2010. Surface plasmon resonance: a general introduction. *Methods Mol. Biol.*, 627:1-14, 10.1007/978-1-60761-670-2\_1
- Estola E, Elo J. 1952. Occurrence of an exceedingly weak A blood group property in a family. *Ann. Med. Exp. Biol. Fenn.*, 30:79-87,
- Fettis MM, Farhadi SA, Hudalla GA. 2019. A chimeric, multivalent assembly of galectin-1 and galectin-3 with enhanced extracellular activity. *Biomater Sci*, 7:1852-1862, 10.1039/c8bm01631c
- Galili U, Shohet SB, Kobrin E, Stults CL, Macher BA. 1988. Man, apes, and Old World monkeys differ from other mammals in the expression of alpha-galactosyl epitopes on nucleated cells. *J. Biol. Chem.*, 263:17755-17762,
- Grahn E, Askarieh G, Holmner A, Tateno H, Winter HC, Goldstein IJ, Krenzel U. 2007. Crystal structure of the *Marasmius oreades* mushroom lectin in complex with a xenotransplantation epitope. *J. Mol. Biol.*, 369:710-721, 10.1016/j.jmb.2007.03.016
- Hazes B. 1996. The (QxW)<sub>3</sub> domain: a flexible lectin scaffold. *Protein Sci.*, 5:1490-1501, 10.1002/pro.5560050805
- Houben K, Marion D, Tarbouriech N, Ruigrok RW, Blanchard L. 2007. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. *J. Virol.*, 81:6807-6816, 10.1128/JVI.00338-07
- Irumagawa S, Hiemori K, Saito S, Tateno H, Arai R. 2022. Self-Assembling Lectin Nano-Block Oligomers Enhance Binding Avidity to Glycans. *Int J Mol Sci*, 23, 10.3390/ijms23020676
- Jia L, Zhang J, Ma T, Guo Y, Yu Y, Cui J. 2018. The Function of Fucosylation in Progression of Lung Cancer. *Front Oncol*, 8:565, 10.3389/fonc.2018.00565
- Juillot S, Cott C, Madl J, Claudinon J, Van Der Velden NSJ, Künzler M, Thuenauer R, Römer W. 2016. Uptake of *Marasmius oreades* agglutinin disrupts integrin-dependent cell adhesion. *Biochim. Biophys. Acta*, 1860:392–401, 10.1016/j.bbagen.2015.11.002
- Kostlanová N, Mitchell EP, Lortat-Jacob H, Oscarson S, Lahmann M, Gilboa-Garber N, Chambat G, Wimmerová M, Imberty A. 2005. The fucose-binding lectin from *Ralstonia solanacearum*: a new type of b-propeller architecture formed by oligomerisation and interacting with fucoside, fucosyllactose and plant xyloglucan. *J. Biol. Chem.*, 280:27839-27849, 10.1074/jbc.M505184200
- Kruger RP, Winter HC, Simonson-Leff N, Stuckey JA, Goldstein IJ, Dixon JE. 2002. Cloning, expression, and characterization of the Galalpha 1,3Gal high affinity lectin from the mushroom *Marasmius oreades*. *J. Biol. Chem.*, 277:15002-15005, 10.1074/jbc.M200165200
- Macher BA, Galili U. 2008. The Galalpha1,3Galbeta1,4GlcNAc-R (alpha-Gal) epitope: a carbohydrate of unique evolution and clinical relevance. *Biochim. Biophys. Acta*, 1780:75-88, 10.1016/j.bbagen.2007.11.003



- Murzin AG, Lesk AM, Chothia C. 1992. beta-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1 beta and 1 alpha and fibroblast growth factors. *J. Mol. Biol.*, 223:531-543, 10.1016/0022-2836(92)90668-a
- Notova S, Bonnardel F, Lisacek F, Varrot A, Imberty A. 2020. Structure and engineering of tandem repeat lectins. *Curr. Opin. Struct. Biol.*, 62:39-47, 10.1016/j.sbi.2019.11.006
- Notova S, Bonnardel F, Rosato F, Siukstaite L, Schwaiger J, Bovin N, Varrot A, Römer W, Lisacek F, Imberty aA. 2022a. A pore-forming  $\beta$ -trefoil lectin with specificity for the tumor-related glycosphingolipid Gb3. *BioRxiv*, 10.1101/2022.02.10.479907
- Notova S, Cannac N, Rabagliati L, Touzard M, Mante J, Navon Y, Coche-Guérente L, Lerouxel O, Heux L, Imberty A. 2022b. Building artificial plant cell wall on lipid bilayer by assembling polysaccharides and engineered proteins. *BioRxiv*, 10.1101/2022.07.25.501355
- Oh YJ, Dent MW, Freels AR, Zhou Q, Lebrilla CB, Merchant ML, Matoba N. 2022. Antitumor activity of a lectibody targeting cancer-associated high-mannose glycans. *Mol Ther*, 30:1523-1535, 10.1016/j.ymthe.2022.01.030
- Ramberg KO, Guagnini F, Engilberge S, Wronska MA, Rennie ML, Perez J, Crowley PB. 2021. Segregated Protein-Cucurbit[7]uril Crystalline Architectures via Modulatory Peptide Tectons. *Chemistry*, 27:14619-14627, 10.1002/chem.202103025
- Ribeiro JP, Villringer S, Goyard D, Coche-Guerente L, Höferlin M, Renaudet O, Römer W, Imberty A. 2018. Tailor-made Janus lectin with dual avidity assembles glycoconjugate multilayers and crosslinks protocells. *Chem. Sci.*, 9:7634 -7641 10.1039/C8SC02730G
- Ross JF, Wildsmith GC, Johnson M, Hurdiss DL, Hollingsworth K, Thompson RF, Mosayebi M, Trinh CH, Paci E, Pearson AR, et al. 2019. Directed Assembly of Homopentameric Cholera Toxin B-Subunit Proteins into Higher-Order Structures Using Coiled-Coil Appendages. *J. Am. Chem. Soc.*, 141:5211-5219, 10.1021/jacs.8b11480
- Sharon N, Lis H. 2004. History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology*, 14:53R-62R, 10.1093/glycob/cwh122
- Siukstaite L, Rosato F, Mitrovic A, Müller PF, Kraus K, Notova S, Imberty A, Römer W. 2021. The two sweet sides of Janus lectin drive crosslinking of liposomes to cancer cells and material uptake *Toxins*, 13:792, 10.3390/toxins13110792
- Tateno H, Goldstein IJ. 2004. Partial identification of carbohydrate-binding sites of a Gal $\alpha$ 1,3Gal $\beta$ 1,4GlcNAc-specific lectin from the mushroom *Marasmius oreades* by site-directed mutagenesis. *Arch. Biochem. Biophys.*, 427:101-109, 10.1016/j.abb.2004.04.013
- Terada D, Kawai F, Noguchi H, Unzai S, Hasan I, Fujii Y, Park SY, Ozeki Y, Tame JRH. 2016. Crystal structure of MytiLec, a galactose-binding lectin from the mussel *Mytilus galloprovincialis* with cytotoxicity against certain cancer cell types. *Sci. Rep.*, 6:28344, 10.1038/srep28344
- Turnbull WB, Daranas AH. 2003. On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J. Am. Chem. Soc.*, 125:14859-14866, 10.1021/ja036166s
- Ward EM, Kizer ME, Imperiali B. 2021. Strategies and Tactics for the Development of Selective Glycan-Binding Proteins. *ACS Chem. Biol.*, 16:1795-1813, 10.1021/acscchembio.0c00880

## **Chapter 4 – Sialic acid**

**Abstract**

Sialic acids are a family with a wide structural diversity, found mostly at the outer end of the glycan chains. They can be divided in two big subgroups, the N-acetyl neuraminic acid (Neu5Ac) and N-glycolyl neuraminic acid (Neu5Gc). While the latter is not found in humans, the former is of paramount importance in humans. In fact it can be linked to the glycan chain with  $\alpha$ 2-3 or  $\alpha$ 2-6 linkage, and thanks to its location and variety of structures it is in charge of mediating many physiological and pathological functions like cell-cell communication, immune response, intra membrane transport, bacterial and viral infections.

Alteration of sialylation could lead to pathological states and it is well-known as a hallmark of cancer. Moreover, the changing of the ratio  $\alpha$ 2-3/ $\alpha$ 2-6 linkage is involved in pathogens tropism, in anti-inflammatory activity of antibodies, and in recombinant biotherapeutic glycoproteins are different depending on the system of production.

In this chapter, a brief review of sialic acid biochemistry and role in biological systems is provided with a major focus on the  $\alpha$ 2-3/ $\alpha$ 2-6 linkage and the significance of their recognition by specific lectins. Here, a LEctPROFILE kit able to discriminate between the two linkages is developed, providing a rapid semi-quantificative tool with industrial and research and development applications. A comparison with MALDI-TOF-MS results is showed to prove the reliability of the kit. The sialylation of IgGs from healthy patients and patients with pathologies is explored, and CHO and melanoma cells are also tested on the kit. At the end of the chapter, preliminary results about the development of Neu5Ac/Neu5Gc LEctPROFILE kit are reported.

**Résumé**

Les acides sialiques sont une famille de sucres avec une grande diversité structurelle, que l'on trouve principalement à l'extrémité externe des chaînes de glycanes. Ils peuvent être divisés en deux grands sous-groupes, l'acide N-acétyl neuraminique (Neu5Ac) et l'acide N-glycolyl neuraminique (Neu5Gc), ce dernier n'étant pas exprimé chez l'homme. Le premier, au contraire, est d'une importance primordiale chez l'homme, il peut être lié à la chaîne glycanique avec une liaison  $\alpha$ 2-3 ou  $\alpha$ 2-6, et grâce à sa localisation et à la variété de ses structures, il est impliqué dans la médiation de nombreux aspects physiologiques tels que la communication cellule-cellule, la réponse immunitaire, le transport intramembranaire, les infections bactériennes et virales.

L'altération de la sialylation pourrait conduire à des états pathologiques et elle est bien connue comme une caractéristique du cancer. De plus, la modification du rapport de liaison  $\alpha$ 2-3/ $\alpha$ 2-6 est impliquée dans le tropisme des agents pathogènes, dans l'activité pro- ou anti-inflammatoire des anticorps et présente un impact dans les fonctions effectrices des glycoprotéines recombinantes puisque la sialylation est différente selon le système de production.

Dans ce chapitre, après un bref résumé de la biochimie des acides sialiques et de leur rôle dans les systèmes biologiques, l'accent est ensuite placé sur la liaison  $\alpha$ 2-3/ $\alpha$ 2-6 et l'importance de leur reconnaissance par des lectines spécifiques. Nous verrons le développement d'un kit LEctPROFILE destiné à discriminer les liaisons  $\alpha$ 2-3 et  $\alpha$ 2-6 afin de fournir une semi-quantification de chacune et d'avoir un outil rapide susceptible de répondre à des applications potentielles industrielles et de recherche et développement. Une comparaison avec les résultats MALDI-TOF-MS prouve la fiabilité du kit. La sialylation des IgG de patients sains et de patients atteints de pathologies est explorée ; les surnageants de culture des cellules CHO et de mélanome sont également testés. A la fin du chapitre, les résultats préliminaires du développement du kit LEctPROFILE Neu5Ac/Neu5Gc sont décrits.

## 4.1 Biochemical and functional overview

The family of **nine-carbon backbone acid-amino sugars** can be called **Sialic acid** (Sia) or **Neuraminic acid** (Fig. 4.1). Both names are retained because they were given separately by two different scientists around the same period. Gunnar Blix isolated it from salivary mucins in 1936 and named his substance “sialic acid” after the Greek word for saliva (σίαλον). Ernst Klenk isolated it from brain glycolipids in 1941 and named it “neuraminic acid” from neurons in the brain [4]. In nature, around 50 derivatives of sialic acid are known, due to the fact that its **amino group** is **substituted** either by an acetyl or glycolyl residue, and the hydroxyl groups may be methylated or esterified with acetyl, lactyl, sulfate or phosphate group (Fig. 4.1). Non-substituted sialic acid does not occur in nature [89].

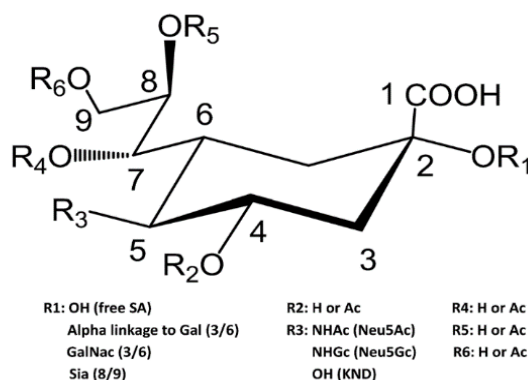
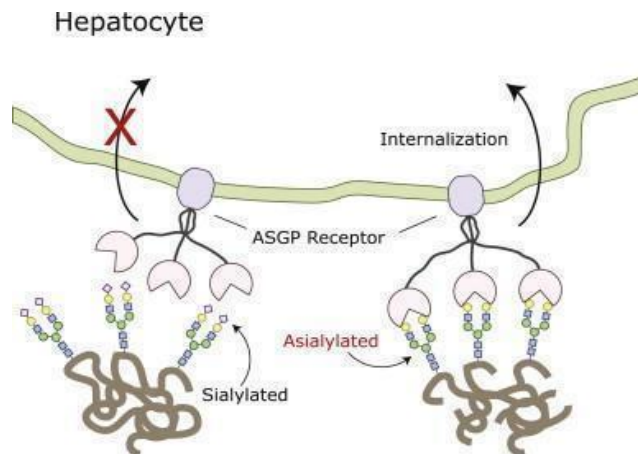


Fig. 4. 1 Structure of sialic acid, adapted from [90].

Sias are abundant on various glycoproteins and glycolipids and are mainly found at **the outermost non-reducing end of glycan chains**, linked to other sugars such as galactose with  **$\alpha$ 2–3- or  $\alpha$ 2–6-glycosidic bonds** or through poly-sialic acid repeats in  $\alpha$ 2–8-linkage. They are found on all types of cells and on secreted proteins, in vertebrates and higher invertebrates and the sum of all sialoglycans on a cell is named “sialome”. Their location, ubiquitous distribution, and peculiar structural features give sialic acids the role of mediators of many biological functions that are different depending on the cell type they are located on. In this thesis, more attention will be given to the functions connected to the objective of the study, which is the building of kits discriminating the  $\alpha$ 2–3 from the  $\alpha$ 2–6 linkage and Neu5Ac from Neu5Gc. They will be described more in detail later on in the corresponding paragraphs.

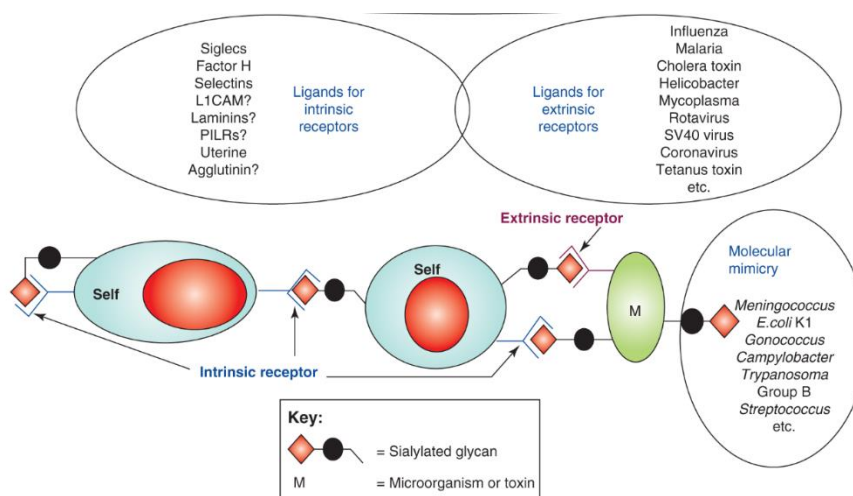
Because Sias are negatively charged, they play a role in the **transport** of positively charged compounds via  $\text{Ca}^{2+}$  channels. They are involved in **conformational changes** of glycoproteins which are important for their organization on cell membranes, and they contribute to the **resistance**

of serum glycoproteins **to proteolytic degradation** [90]. In fact, when initially sialylated molecules age, they progressively lose their sialic acid moieties. Asialoglycoproteins are recognized by galactose-specific lectins on hepatocytes, called asialoglycoprotein receptors (ASGPR) that bind them, resulting in their uptake and degradation (*Fig. 4.2*) [91].



*Fig. 4. 2* Asialoglycoprotein receptors. ASGP receptors in the liver do not recognize desialylated proteins, targeting them for uptake and degradation, adapted from [91]

They are ligands for **intrinsic receptors** such as Siglecs, factor H, selectins, all of them allowing cell-cell communication in multicellular organisms. On the other hand, they can act as ligands for **extrinsic receptors** that allow the interaction between host and pathogens like viral hemagglutinins, bacterial adhesins, bacterial toxins, and parasite-binding proteins (*Fig. 4.3*).



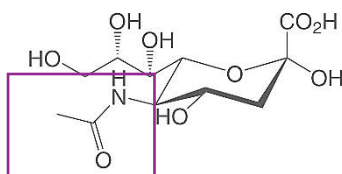
*Fig. 4. 3* Roles of sialic acids. They can act as ligands for intrinsic receptors as well as extrinsic receptors, adapted from [92].

This is the main reason why Sias and their receptors give many opportunities for **therapeutic development**. An example of drugs based on Sia metabolism are Relenza and Tamiflu, competitive inhibitors of influenza A and B sialidases, another one is Rivipansel, a Sia-mimetic drug that inhibits selectins and has shown promising results in clinical trials. Pathogens can also display the so-called “molecular mimicry”, which means that they express Sias to evade the host immune response (*Fig. 4.3*) [92].

Sialylation often results altered in presence of diseases. This is again connected to the fact that sialylation is involved in the regulation of many intrinsic receptors. Selectins recognize malignant tumour cells causing their interaction with platelets, leukocytes and endothelium, facilitating the motility and metastasis. Tumour Sias also bind to factor H to limit the complement activation and modulate cell attachment to the matrix, promoting invasion and metastasis [4].

#### 4.2 N-acetyl neuraminic acid

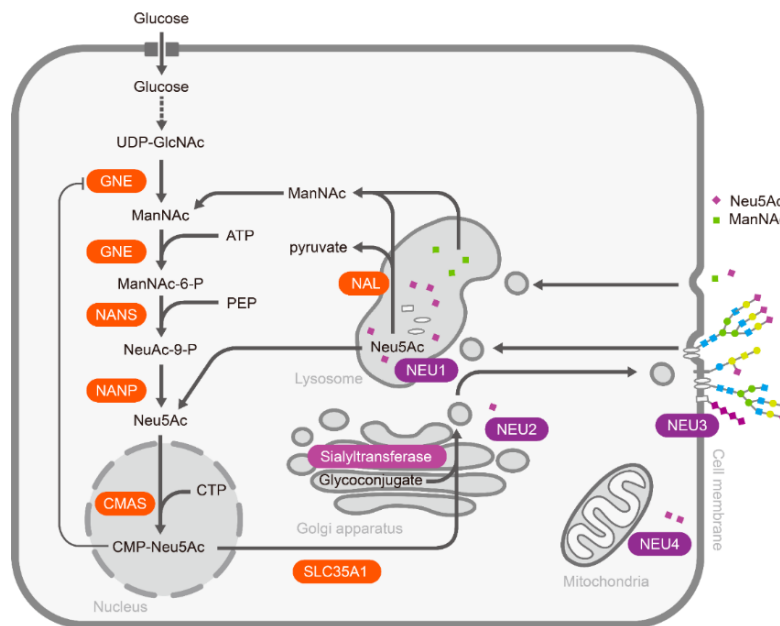
**N-acetyl neuraminic acid** (Neu5Ac or NANA) is the most common sialic acid derivative found in mammals and it occurs when the C5 amino group is substituted with an **acetyl group** (*Fig. 4.4*).



*Fig. 4. 4* N-Acetyl neuraminic acid. The red circle highlights the acetyl functional group on C<sub>5</sub>

As shown in the *Fig. 4.5*, in mammalian cells the Neu5Ac biosynthesis starts in the cytosol with UDP-GlcNAc converted into ManNAc-6-P by the bifunctional enzyme UDP-GlcNAc 2-epimerase/ManNAc-6-kinase coded by the gene Gne. Reaction of ManNAc-6-P with phosphoenolpyruvate (PEP) by Neu5Ac 9-phosphate synthase (NANS) results in Neu5Ac-9-P, which is then dephosphorylated by a specific phosphatase to release free Neu5Ac in the cytoplasm. Neu5Ac is transferred to the nucleus and activated by the cytosine 5'-monophosphate N-acetylneuraminic acid synthetase (CMAS) to form CMP-Neu5Ac. After activation, CMP-Neu5Ac is transferred to the glycoconjugates in the Golgi apparatus by a family of linkage-specific sialyltransferases. Eventually, sialylated glycoconjugates are transported to the cell membrane or

packaged for secretion (*Fig. 4.5*) [93]. To summarise, Neu5Ac is biosynthesized in the cytosol, converted to activated CMP-Neu5Ac in the nucleus, transferred to lipid and protein glycans in the Golgi and eventually transported to the cell surface or secreted [94].



*Fig. 4. 5* Neu5Ac acid metabolism in mammalian cells, adapted from [93].

#### 4.2.1 Neu5Ac in forensic sciences

Human Transferrin is the iron transport serum glycoprotein and it consists of a polypeptide chain with two binding sites for iron and two N-linked glycan chains. It can exist in different glycoforms, related to the number of Sia antennae (n-antennary) that it contains (*Fig. 4.6*), which can go from zero to four. The major human transferrin glycoform in physiological conditions consists of four sialic acid residues (tetrasialo) and it is about 80% of the total. [95]. It is known that a regular consumption of 50–80 g of alcohol per day for at least one week is associated with an increase of and typically of asialo- transferrin (no glycosylation) and disialo-transferrin. This increase has been adopted in clinical and forensic medicine as a biomarker of alcohol abuse, with the name of carbohydrate deficient Transferrin (CDT). Its measurement is little affected by other conditions and it is apparently independent from liver comorbidities [96].



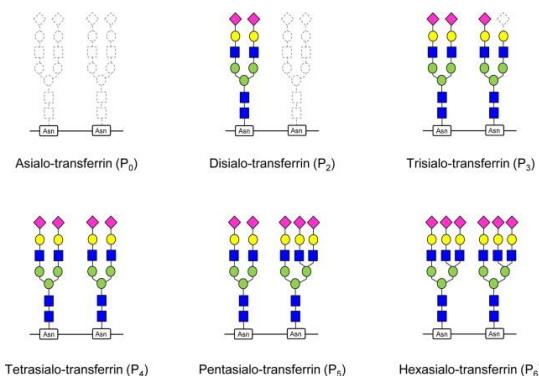


Fig. 4. 6 Human Transferrin glycoforms, adapted from [95]

In this thesis, transferrin was used as a model glycoprotein for the development of the  $\alpha$ 2-3/ $\alpha$ 2-6 LEctPROFILE kit and the NeuAc/NeuGc LEctPROFILE kit .

#### 4.2.2 Neu5Ac and specific linkages

As mentioned in paragraph 4.1, Sias in vertebrates occurs usually at the terminal end of the glycan chain studied (N-, O- glycans and glycolipids) and they only extend further if linked to another Sia. The “sialome” is compared to a forest, where the class of glycans (glycoproteins, glycolipids) are the trees, the sugars at the core and below the Sias are the branches, the sialic acid linkages are the stems and the Sias structures are the leaves and flowers (*Fig. 4. 7*) [4]. Focusing on the stems or linkages, diversity in sialic acid presentation is generated by different glycosidic bonds from the 2-carbon to the underlying sugar chain. The most common sialyl-terminals are  $\text{S}\alpha$ 2-3Galb-(2-3) and  $\text{S}\alpha$ 2-6Galb1-(2-6).

Sialic acids in  **$\alpha$ 2-3 linkage with galactose** are generated by six different  $\alpha$ 2-3 sialyltransferases called ST3GAL1, ST3GAL2, ST3GAL3, ST3GAL4, ST3GAL5 and ST3GAL6. In mice, ST3GAL1 is most abundant in the spleen, liver, bone marrow, thymus, and salivary glands. ST3GAL2 is most abundant in the brain , where  $\alpha$ 2-3-sialylated glycolipids are common. ST3GAL5 is expressed in the brain, skeletal muscle, adrenals, and liver, and St3gal6 is most expressed in testes.

ST3GAL1 generates  $\text{Sia}\alpha$ 2-3Gal $\beta$ 1-3GalNAc $\alpha$ -Ser/Thr that is important for the viability of peripheral CD8+ T cells. Mice lacking ST3GAL1 show decreased cytotoxic T-cell responses with an increase in the apoptotic death of naïve CD8+ T cells [4].

**Sialic acid linked in  $\alpha$ 2-6 with galactose** is less common than  $\alpha$ 2-3 sialic acid in tissue glycoproteins but more common in plasma glycoproteins. It is mainly generated by two  $\alpha$ 2-6 sialyltransferases, ST6GAL1 and ST6GAL2. In mice, ST6GAL1 is expressed in hepatocytes where it is responsible for  $\alpha$ 2-6 sialylation of serum glycoproteins and in lymphocytes where it is found on the glycoproteins of the antigen receptor complex. Mice lacking ST6GAL1 show diminished antibody

responses to T-lymphocyte-dependent and -independent antigens. Changes in sialic acid linkages also affect cell-mediated immunity. Sias are known to bind Siglecs (Sialic acid-binding immunoglobulin-type lectins) that are proteins located on the surface of immune cells, responsible for their adhesion and activation of the immune response. The Siglec CD22 on B cells mediates the adhesion specifically through the recognition of sialic acids in  $\alpha$ 2-6 linkage. The extracellular domain of CD22 on B lymphocytes specifically recognizes Sia $\alpha$ 2-6Gal $\beta$ 1-4GlcNAc- and in the absence of  $\alpha$ 2-6 sialic acid on glycans, CD22 shows increased clustering with the B-cell receptor (BCR) leading to the down-regulation of its signal. Mice lacking ST6GAL1 also show reduced B-lymphocyte proliferation response and reduced B-cell surface IgM levels [97].

ST6GAL2 expression is mainly restricted to the embryonic and adult brain but its functions are currently unknown [4].

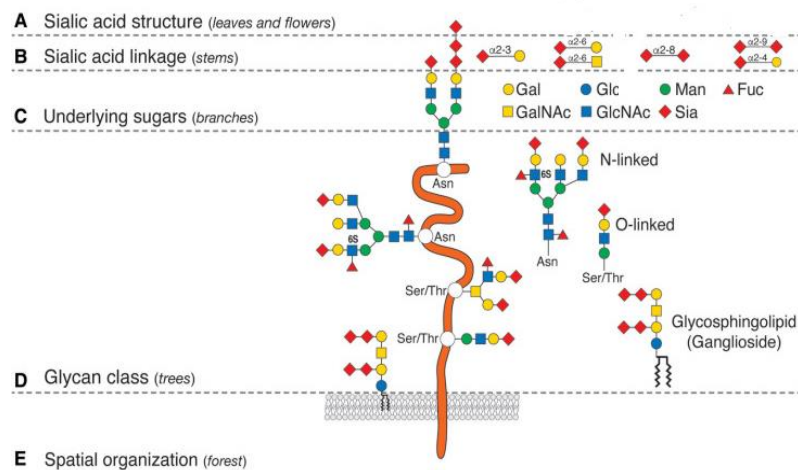


Fig. 4. 7 Sialome organization, adapted from [4].

$\alpha$ 2-3 and  $\alpha$ 2-6 sialyllactose are present in a variety of tissues and are connected to many biological functions so it is clear that modifications in their expression leads to various consequences. The development of the  $\alpha$ 2-3/  $\alpha$ 2-6 LEctPROFILE kit is intended to have a rapid glycoprofile of the samples of interest and to detect the relative ratio/quantity of  $\alpha$ 2-3/  $\alpha$ 2-6 sialyllactose that they contain. Below are reported examples in which the ratio/determination quantification of  $\alpha$ 2-3/  $\alpha$ 2-6 sialyllactose can be of paramount importance.

### 4.3 Applications of $\alpha$ 2-3/ $\alpha$ 2-6 LEctPROFILE kit

#### 4.3.1 In pathogens recognition and in cancer

Many pathogens have evolved receptors for Sias in order to enter and infect human cells (Fig. 4.3). The viral hemagglutinin of influenza infects cells binding to cell-surface Sia and fusing with the host cell membrane. The virion is released inside the cells, where it exploits the host machine to replicate and assemble new virions that eventually are released from the host cell surface through the action of a viral neuraminidase. The hemagglutinin directs species and tissue tropism: human influenza virus for example binds preferentially to sialic acid residues in  $\alpha$ 2,6 linkage, whereas **avian influenza** recognizes the  $\alpha$ 2,3-linkage of sialic acids found in the intestinal epithelium of birds, where it resides [98]. The possible jump from birds to humans has drawn interest for the dangers that it could bring, as it could cause an epidemic outburst. This could happen if the hemagglutinin component of the virus underwent specific mutations, switching to preferentially binding  $\alpha$ 2-6-linked sialic acids, which are enriched on the airway epithelium of humans. It occurred in 1918 when the influenza pandemic, very well-known as Spanish flu, killed at least 20 million people. What is believed is that there was no direct jump from birds to humans, but most likely pigs acted as a 'mixing vessel' as they have both  $\alpha$ 2-3- and  $\alpha$ 2-6-linked sialic acids on the epithelium.

Glycans bearing  $\alpha$ 2-3 sialic acid residues have also been implicated in bacterial pathogenesis. *Helicobacter pylori* is a human and primate-specific pathogen found in the gastric mucus layer or attached to the gastric epithelium. Infection by this bacterium, which affects about half the world population, results in chronic active gastritis and is a risk factor for the development of peptic ulcer disease, gastric adenocarcinoma, and gastric lymphoma. The wild-type bacteria binds to N-acetyllactosamine-based gangliosides with terminal  $\alpha$ 2-3-linked Neu5Ac, while gangliosides with terminal Neu5Gc, Neu5Ac $\alpha$ 2-6, or Neu5Ac $\alpha$ 2-8, Neu5Ac $\alpha$ 2-3 are not recognized [99].

The study of  $\alpha$ 2-6- and  $\alpha$ 2-3-sialylation is a promising biomarker for cancer diagnosis since increasing levels of  $\alpha$ 2-3 linkage point towards tumour malignancy. **Prostate cancer (PCa)** is one example. The current golden standard method for PCa screening is the blood quantification of the Prostate Specific Antigen (**PSA**), a glycoprotein produced by normal, as well as malignant, cells of the prostate gland. The higher a man's PSA level, the more likely it is that he has PCa (PSA > 10 ng/mL). However, this kind of test is not accurate, in fact men with levels of PSA that are in a range considered normal or intermediate were found to have PCa. Moreover, high levels of PSA could also be connected to benign prostate hypertrophy due to excessive exercise or linked to urinary tract infection or assumption of some medications [100]. Only about 25% of people who undergo prostate biopsy due to an elevated PSA level are diagnosed with PCa. Therefore, a more accurate test is needed. The levels of  $\alpha$ 2-3-linked sialic acids on PSA could improve PCa diagnosis since it illustrated great potential in discriminating malignant from benign conditions. During the years there have been different studies in this direction: electrochemical label-free glycoprofiling showed that *Maackia amurensis agglutinin* recognizing  $\alpha$ 2-3-terminal sialic acids binding to serum PSA was significantly

higher for PCa samples than for healthy controls [101]. In another study, the evaluation of  $\alpha$ 2-3-sialylated PSA in serum with a lectin immunoaffinity column revealed that  $\alpha$ 2-3-sialic acid on PSA exhibited high performance in discriminating between high-risk prostate cancer patients and the benign prostate hyperplasia individuals [102].

#### 4.3.2 In biotherapeutics

It has been estimated that in 2018 **Chinese hamster ovary cells (CHO)** had been used for the production of 84% of FDA-approved **biotherapeutics** [103]. They are, in fact, advantageous for many reasons. They are less susceptible to infections by human viruses, making them safer for commercial purposes. They are easy to manipulate since they duplicate quickly both in adhesion and in suspension, and for this reason are suitable for large-scale industrial suspension culture. Moreover, the biotherapeutics produced by CHO have a human-like glycosylation, thus making the products more compatible with the human hosts. Despite all these upsides, CHO cells are not able to produce some types of human glycosylation such as  $\alpha$ 2,6-sialylation, as they lack the ST6 sialyltransferase gene [104]. As described in paragraph 4.1, sialylation helps to keep glycoproteins in the bloodstream, preventing the recognition by receptors on liver hepatocytes. Increased protein sialylation may decrease the frequency of injection or the amount of therapeutic protein used in a single dose. Moreover, the fact that both linkages are present in humans' glycoproteins has an impact on their activity inside the body (Paragraph 4.2.2). Hence, many strategies have been put in place to increase sialylation and recreate a glycosylation profile beneficial for the host that could also generate savings for the industry. One of these approaches is the supplementation of CHO culture media with different precursors of the sialic acid pathway to improve the sialylation of interferon- $\gamma$  [94]. Another strategy is the transient expression of sialylation enzymes into CHO cells, as in the case of ST6Gal1 to increase the sialylation of Trastuzumab F243A mutant antibody [105].

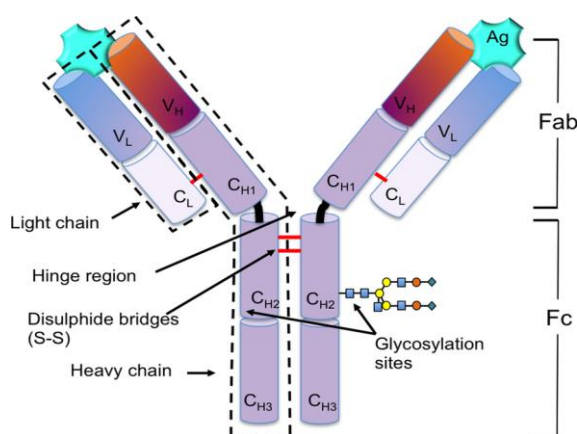
Finally, other expression systems have been exploited, among which human cell lines HEK293 cells and the HT-1080, respectively from human embryo kidney [106] and fibrosarcoma origin [107]. Only few biotherapeutics produced in these cell lines have been approved so far. Another one, the *Nicotiana benthamiana* plant expression system was genetically modified on a combination of stably transformed plants with transient expression modules. They are intended to carry the human-type sialylation pathway and to generate mono- and disialylated structures with  $\alpha$ 2-3 and  $\alpha$ 2-6 linkage [108].

For all the cases mentioned above, the  $\alpha$ 2,3/ $\alpha$ 2,6 LEctPROFILE kit could be a tool to detect the right clones in the early phases of production of these engineered strains, accelerating drug development and production, reducing manufacturing costs and improving drug efficiency.

## 4.3.3 In antibodies

**Immunoglobulins or IgGs** are the most common type of antibodies (Ab) found in blood circulation, representing 75% of serum antibodies in humans. The four IgGs subclasses, IgG1, 2, 3 and 4 are named based on their abundance in serum. IgGs are large globular proteins of around 150 KDa. They are part of the humoral immunity and they control infections of body tissues binding to viruses, bacteria and fungi.

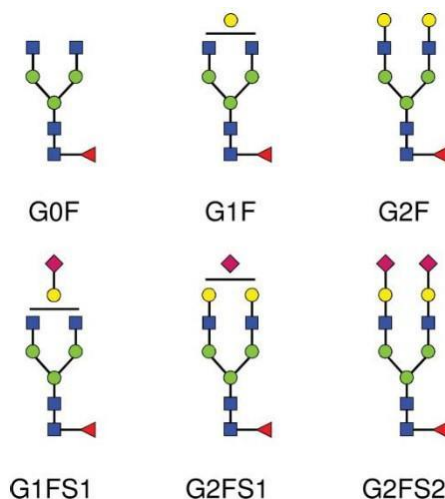
They are represented with the classical Y shape, where the bottom part is made of the **Fc** domain, and the upper part is the **Fab** domain. The Fab consists of the light chain variable and constant domains (VL and CL) and the first two domains of the heavy chain, one variable and one constant (VH and CH1). The Fc portion consists of two constant domains, CH2 and CH3. The hinge region is a short sequence of the heavy chains linking the Fab region to the Fc region (*Fig. 4.8*). The two heavy chains are linked to each other and to a light chain each by disulfide bonds [109] (*Fig. 4.8*). The variable domains on the Fab are the ones responsible for the recognition of the antigen. The Fc domain coordinates the pathogen killing through antibody's effector functions such as the antibody-dependent cell-mediated cytotoxicity (ADCC), antibody-dependent cellular phagocytosis (ADCP) and complement-dependent cytotoxicity (CDC). The effector functions can be triggered by the Fc interaction with the FcγR inducing the release of cytotoxic granules resulting in killing of infected cells or triggering the activation of the complement cascade or the activation of phagocytic immune cells.



*Fig. 4. 8* IgG structure and its glycosylation. The bottom domain is the Fc and the top one is the Fab. V<sub>L</sub>= variable region of the light chain, C<sub>L</sub>= constant region of the light chain, V<sub>H</sub>= variable region of the heavy chain, C<sub>H1,2,3</sub>= Constant region of the heavy chain, domain 1, 2, 3, Ag= antigen, adapted from [109].

At the position of the highly conserved **Asparagine 297 (N297)** on each CH<sub>2</sub> domain there are two glycosylation sites mediating interchain protein–protein interactions. While N-glycans are normally exposed at the surface of the proteins, the **Fc N-glycans lie within a pocket formed by the two**

**CH2 domains of the antibody** where they interact with internal amino acid residues through hydrogen and CH-p bonds. As a consequence of this embedment, the Fc N-glycans are mostly limited to di-antennary complex types with partial galactosylation and low sialylation. Fc-sialylation in circulating human IgGs is generally believed to be mainly of  $\alpha 2-6$  type [94]. The most common glycan on circulating human IgGs is a core fucosylated complex structure with one galactose (G1F), followed by fucosylated complex glycans with 0 and 2 galactoses (G0F and G2F). There can be glycans also in the Fab regions but they are not conserved (*Fig. 4.9*).



*Fig. 4. 9* Different glycan structures that can be found on IgG. G= galactose, F= fucose, S= sialic acid [105] .

Studies over the past 30 years have provided important information regarding how glycosylation affects Fc-mediated effector functions. Antibody glycosylation was found to be changing in numerous diseases; one that was particularly studied is the autoimmune rheumatoid arthritis, in which are present immune complexes consisting only of immunoglobulin (most commonly IgG), implying that the IgG is both the 'antibody' and the 'antigen'. Autoantigenic reactivity was localised to the constant-region domains of IgG where there is a decrease of the level of outer-arm galactosylation of the complex N-linked glycans. As a consequence, there is a decrease of Neu5Ac $\alpha$ (2-6)Gal segments and an increase in IgG molecules bearing oligosaccharides terminating in N-acetylglucosamine. Immune complexes formed through G0 IgG can bind and activate mannan-binding lectin and a pseudo-complement cascade [110]. Amongst patients with the most severe rheumatoid arthritis, articular amelioration occurred during pregnancy, linked to a six-fold reduction in the actual concentration of IgG lacking galactose, with consequent remission of symptoms during gestation [111].

Studies on intravenous immunoglobulin therapy (IVIG) for rheumatoid arthritis show that the **anti-inflammatory activity** of purified IgG from pooled human plasma can be attributed to the **presence of terminal sialic acid**. In fact, removal of the terminal sialic acid of IVIG, in animal models, abrogates the anti-inflammatory activity. On the contrary, enrichment of the sialylated fraction of IVIG

enhances this activity. This study was conducted only on the Fc portion, and it was the first time when it was demonstrated the independence of the Fc from the antigen-binding domain. Through sequential mass spectrometric based glycan analysis and treatment with either  $\alpha$ 2-3 or  $\alpha$ 2-3/  $\alpha$  2-6 sialidase, it was determined **that the linkage type present on IVIG is the 2-6 linkage** to the penultimate galactose. No  $\alpha$ 2-3 linkages were detected on intact IVIG and no attenuation of the anti-inflammatory activity was observed following  $\alpha$ 2-3 SA treatment [112]. In another paper by the same author [113] it is theorized that the addition of sialic acid moieties has structural implications and consequently the effector function of the IgG is affected. In fact, the glycans at N-297 keep the Fc region in an open conformation to promote the binding to the Fc $\gamma$ R. The addition of sialic acid would reduce this binding and drive the IgG to bind to the lectin receptors SIGN-R1 or DC-SIGN thereby attenuating autoantibody-initiated inflammation. However, this mechanism of sialylated IgG/Fc to bind DC-SIGN was later discussed in other studied and not confirmed [30] and it remains under debate. Whether  $\alpha$ 2-6-sialylated IgG/Fc actively inhibits inflammation remains an interesting topic and still there is much to investigate. The effect probably depends upon the context of the disease and metabolic factors.

It appears to be a certain variability in Fc glycans across individuals. At the Croatian company Genos, under the leadership of Gordon Luc, studies were carried out through the hydrophilic interaction liquid chromatography (HILIC) method, performed on a large scale of patients affected by different pathologies. The biggest differences in glycosylation were found at different stages of life and among individuals coming from different parts of the world; the HILIC method, does not takes into account the linkage of sialic acid though [30]. Therefore, insight in the sialylation pattern through the  $\alpha$ 2-3/ $\alpha$ 2-6 LEctPROFILE kit may provide valuable clues on how sialylation affects IgGs effector functions in different pathologies and individuals.

#### 4.4 N-glycolyl neuraminic acid

It is a glycoform of Sia occurring when there is the addition of a single oxygen atom to the N-acetyl group of Neu5Ac at the C5 amino group, resulting in the glycolyl group (*Fig. 4.10 B*).

It is found in most mammals, with exceptions for taxa that have lost the capacity to synthesize Neu5Gc (*Fig 4.10*) including New World primates, Mustelidae, pinnipeds, Procyonidae, hedgehog, bats, spermwhale, and white-tailed deer and humans [114].

The only known biosynthetic pathway generating Neu5Gc is the conversion of CMP-N-acetylneuraminic acid into CMP-Neu5Gc, which is catalysed by the CMP-Neu5Ac hydroxylase enzyme encoded by CMAH gene. This reaction happens in the cytosol and it is irreversible. Cmah catalyses the transfer of an oxygen atom to CMP-Neu5Ac, generating CMP-Neu5Gc. CMP-Neu5Gc can then be transported into the Golgi apparatus in the same manner as CMP-Neu5Ac (*Fig. 4.10 A*) to add Neu5Gc to newly synthesised glycoconjugates. These two nucleotides seem to be used interchangeably by the Golgi CMP-Sia transporter and by the mammalian sialyltransferases, which

transfers Sia residues to cell surfaces. Moreover, Neu5Ac or Neu5Gc released from glycoconjugates during lysosomal degradation processes can also be exported back into the cytosolic compartment by a specific transporter. Again, there appears to be no major difference in their conversion by CMP-Sia synthase [115].

Although found in mammals, humans are incapable of synthesising Neu5Gc due to an inactivating single-exon deletion mutation in the CMAH gene [116]. The reason why this mutation occurred in the first place it's still a matter of discussion. Many pathogens use Sias to bind and infect, and some specifically target Neu5Gc so one explanation is the loss of Neu5Gc to survive a lethal infectious pathogen. Another possibility that does not exclude the first one, it is called the “cryptic female choice” which is the selective fertility of Neu5Gc-deficient females with Neu5Gc- deficient males, leading to positive selection of this genotype [114].

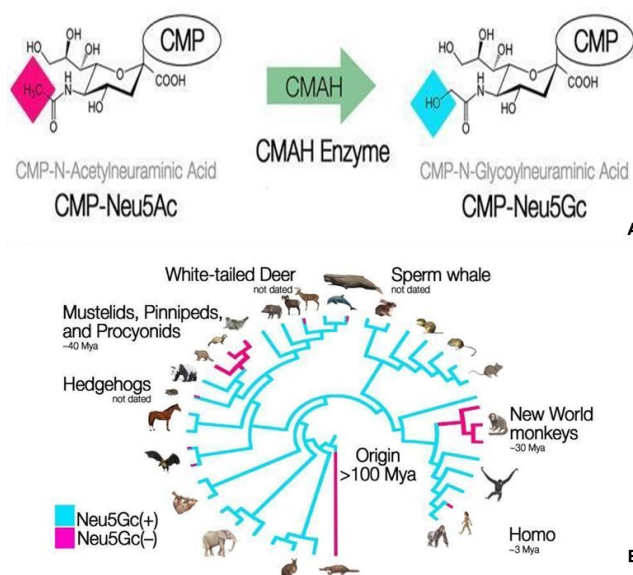


Fig. 4. 10 **A** N-Glycolyl neuraminic acid biosynthesis. **B** taxa are found in blue, not found in purple and purple and blue where only certain species have lost the capacity to make it, adapted from [117]

#### 4.5 Applications of Neu5Ac/Neu5Gc LEctPROFILE kit

Although humans cannot produce Neu5Gc, not even via alternative pathways, it is detected in some tissues and since meat and dairy products are a big part of the human diet, it was hypothesized the existence of a metabolic incorporation of Neu5Gc from mammalian foods. Human cell cultures were shown to take up Neu5Gc from the foetal calf serum contained in the culture medium and insert it into cell-surface glycoconjugates, demonstrating the existence of an intrinsic biochemical pathway compatible with Neu5Gc [115].

Over the years, many groups have taken up the challenge to elucidate the mechanism of this accumulation. Tangvoranuntaku *et al.* have proposed that the food chewed in the mouth is embodied



in salivary mucin, but this has resulted to be poor and without any connection to the findings in endothelial, epithelia, cancers, or foetuses [118]. Afterwards, Banda *et al.* proposed a mechanism to elucidate this accumulation using a Neu5Gc-deficient mouse with a human-like defect in Cmah. They followed the ingested free Neu5Gc by fluorescent tagging of free sialic acids with 1,2-diamino-4,5-methylenedioxybenzene dihydrochloride (DMB) for HPLC, and the Neu5Gc-glycoproteins by Neu5Gc IgY. The former showed rapid absorption into the circulation and urinary excretion, the latter was absorbed into the small intestinal wall, appeared in circulation at a steady-state level for several hours, and was incorporated into multiple peripheral tissue glycoproteins and glycolipids. This proved that Neu5Gc can be metabolically incorporated from food [116]. This is the first example where a non-human dietary molecule becomes incorporated onto human cell surfaces, also being epidemiologically linked to the increased cancer risk [119]. Hence, Neu5Gc presence has consequences on health and it can be specifically linked with development of anti-Neu5Gc antibodies and diseases.

#### 4.5.1 In cancer

It is described that certain types of cancer, including colon carcinomas, retinoblastomas, breast cancers, and melanomas, show an increase of Neu5Gc and Neu5Gc antibodies [119]. The first to describe the presence of anti-Neu5Gc antibodies in humans were Hanganutziu and Deicher, from which now they are called HD-antibodies. They observed a “serum sickness” derived from the development of these Abs after administration of therapeutic animal antiserum to humans. Later, it was demonstrated that they are directed against the Neu5Gc antigen contained in gangliosides. They were later detected in some cancer-patients' human serum, without the patients having ever received animal serum, and afterwards it had been shown that also healthy individuals can have high levels of anti-Neu5Gc antibodies [119]. The frequency of cancer patients possessing HD-Abs is 80% of hepatocellular carcinomas and 25–28% in colon and gastric cancer cases. In general, HD-antibodies from patients with melanoma and hepatic cancers are identified as IgG and IgM [120]. The appearance of HD-Abs in tumours has been connected to the chronic inflammatory state of the cancer tissue. Inflammation is the biological response of body tissues after an injury and consists in the recruitment of tissue mast cells and the migration of leukocytes from the blood vessels to the interested site. The idea would be that Neu5Gc is a potential link between diet and human cancer. Red meat and milk-derived products triggers the accumulation of Neu5Gc with consequential “xenosialitis” an inflammation due to reaction against a xeno-sialic acid that is now part of “self” and HD-Abs production. The inflammation causes the recruitment of innate immunity growth factors, cytokines, and angiogenic agents and the Neu5Gc the recruitment of antibodies to the inflammation site. This causes a chronic situation in which the immune response is not sufficient to shut down the inflammation, but instead stimulates tumour growth and survival [121]. The inflammation status favours angiogenesis and cell extravasation, originating the metastatic tumours.

This hypothesis matches with the decreased cancer risk associated with veganism and the rarity of carcinomas in the chimpanzees who normally express Neu5Gc and thus do not make antibodies against it. In any case, an increase in the Neu5Gc titre correlated well with the chance to develop cancer and analysis of tissues through the NeuAc/NeuGc LEctPROFILE kit could be of diagnostic significance to give an early indication of the inflammatory status.

Xenosialitis is connected also to the formation of atherosclerotic plaques. Studies have shown that Neu5Gc can be detected in the endothelium and the sub-endothelium of the atherosclerotic plaques. In vitro experiments with human endothelial cells fed with Neu5Gc and subsequently exposed to serum containing anti-Neu5Gc glycan antibodies showed the presence of IgG and complement deposition which increase endothelial activation, cytokine production, all events associated with early atherogenesis. These effects were inhibited by Neu5Gc- $\alpha$ -methyl glycoside, a specific competitor to anti-Neu5Gc antibodies [114].

#### 4.5.2 In biotherapeutics

The use of systems for the expression of high levels of protein, such as bacterial, yeast and insect cells for production of biotherapeutic glycoproteins, are limited to small proteins without extensive post-translational modification. Besides CHO, another widely-used expression system is mouse myeloma cells. The drawback of these cells is that they have very high levels of Neu5Gc and patients receiving products coming from them might enhance their own intrinsic anti-Neu5Gc antibody production.

A more serious possibility is the formation of antigen–antibody complexes, which could secondarily trigger an immune reaction against the biotherapeutic agent. Some limitations of mammalian cell culture systems have been overcome by the expression of recombinant proteins in transgenic animals such as human recombinant antithrombin (rhAT) into the milk of transgenic goats. The study of Zhou *et al.* compared the carbohydrate structures in rhAT from several transgenic goat lines and clones. rhAT carbohydrates differ from human plasma AT with increasing levels of Neu5Ac and this carbohydrate heterogeneity was found also in different transgenic goat lines [122].

For this reason it is of paramount importance to profile the glycosylation in the early stages of production, in order to select the right clones from the beginning of the process and maintain a glycosylation homogeneity.

### 4.5 Lectins considered in this study

#### 4.5.1 MAAI

The first lectin considered for the  $\alpha$ 2-3/ $\alpha$ 2-6 LEctPROFILE kit is the sialic-acid binding lectin *Maackia amurensis leukagglutinin*. It is a L-type lectin extracted from the seeds of the legume tree *Maackia amurensis*, also referred commercially as MAL or MAA-I. For simplicity we will refer to it as MAAI. It is made of two subunits and in solution it assembles in the two monomers of the “canonical

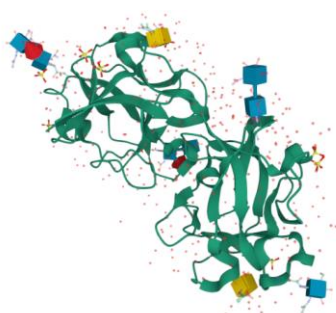
legume lectin dimer” containing one carbohydrate binding domain each, as shown in *Fig. 4.11*. The molecular weight of the dimer is around 74 kDa with one monomer weighing around 35 and the other around 34 kDa. It is inhibited by low concentration of  $\alpha$ 2-3-sialyllactose (NeuAc2-3Gal $\beta$ 1-4Glc), but it is not inhibited by either 2-6-sialyllactose or free NeuAc [64]. Studies show that the 50% of inhibition is reached at a concentration of 0.5 mM of the trisaccharide  $\alpha$ 2-3-sialyllactose and lactose was found to inhibit at 5 mM [123].



*Fig. 4. 11* MAA-I in complex with sialyllactose, PDB code 1DBN, adapted from [64]

#### 4.5.2 SNA

The other lectin in the  $\alpha$ 2-3/ $\alpha$ 2-6 LEctPROFILE kit is the *Sambucus Nigra* agglutinin (SNA) extracted from black elderberry bark. It is a type-II ribosome-inactivating protein made of two linked  $\beta$ -trefoil domains, containing two carbohydrates recognition domains on opposite faces at both ends of the protein (*Fig.4.12*). Although some oligomeric species are present, the monomeric species predominates in solution [65]. Neu5Ac alone is not inhibitory for this lectin. On the contrary, a series of oligosaccharides containing terminal Neu5Ac-Gal linkages, especially those containing the  $\alpha$ 2-6-linkage, are highly inhibitory: N-acetyllactosamine gives 50% of inhibition at 8.5 mM, lactose at 12 mM [124].



*Fig. 4. 12* SNA in complex with N-Acetylgalactosamine, PDB code 3CA3, adapted from [65].

### 4.5.3 HPyL

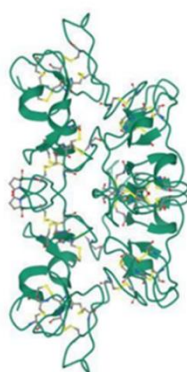
Human polyomavirus 9 (HPyV9) was identified in 2011 from the serum and urine of asymptomatic kidney transplant patients. VP1 is the major polyomavirus capsid protein and it interacts with glycans terminating with Neu5Gc. The protein adopts a jelly roll fold and assembles into a homo-pentamer around a central cavity (*Fig. 4. 13*) [125]. The molecular weight of the monomer is around 30 kDa and it contains one carbohydrate recognition domain per monomer.



*Fig. 4. 13* structure of VP1. PDB code 4POQ [125].

### 4.5.4 WGA

*Weat germ agglutinin* WGA was isolated from *Triticum vulgaris*. The lectin has two subunits and a molecular weight of 36 kDa. WGA selectively binds to N-Acetyl glucosamine (GlcNAc) and to N-acetylneuraminic acid (sialic acid) residues of glycoproteins and glycolipids. It has a molecular weight of around 36 kDa and it is a dimer containing two carbohydrate-binding sites [126].



*Fig. 4. 14* 3D structure of WGA, adapted from [126] .

#### 4.5.5 SSL

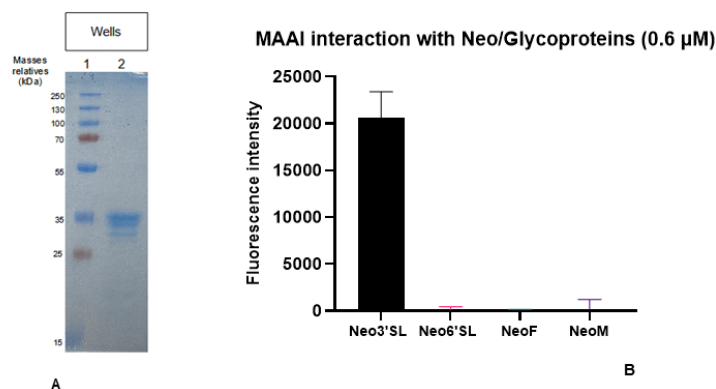
*Scylla serrata* lectin takes its name from the homonymous marine crab, and it's a C-type lectin specific for Neu5Gc. Crustaceans cannot synthesise sialic acids, but most probably they make sialic acid binding lectins as a defence against bacteria sialoconjugates. The molecular weight of pure SSL is about 55 kDa and under reducing conditions it resolves into two subunits of 30 kDa and 25 kDa. NeuGc inhibits hemagglutination activity of the purified lectin at a concentration of 0.6 mM [127]. To date, there are no crystal structures of this protein and the number of binding sites is unknown.

The purification of this lectin was carried out starting from the natural source, the crab *Scylla serrata*. The yield of the purification was too low to perform any experiment and it was not also possible to detect a clear band on the SDS-PAGE, even though signs of lectin activity were detected in the extract.

### Results and discussion

#### 4.6 MAAI Purification

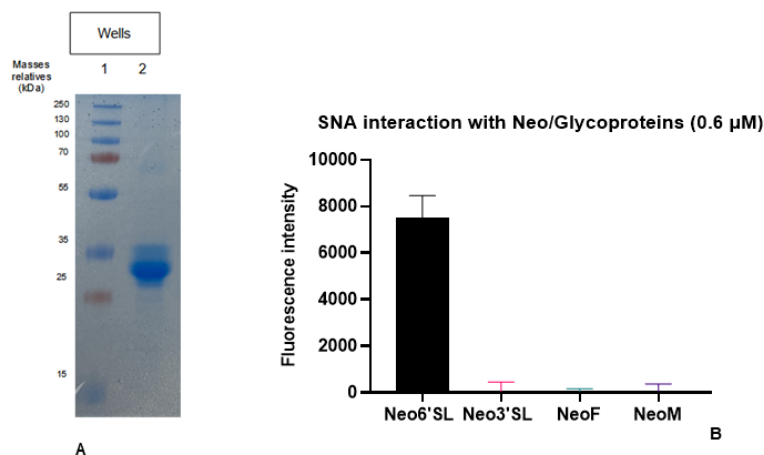
At GLYcoDiag we established purification protocol for the industrial production of MAAI based on Cummings [128], starting from the seeds of the leguminous tree *Maackia amurensis*. The purification procedure is property of GLYcoDiag and for this reason it will not be described in detail in this thesis. The protocol was successful in the purification of around 10 mg/lectin per 100 g of seeds. The GLYcoPROFILE in direct binding in *Fig 4.15B* shows the lectin is active and interacts activity  $\alpha$ 2-3-sialyllactose that is highlighted in the interaction with the neoglycoprotein Neo3'SL. The lectin is highly selective for the  $\alpha$ 2-3 Sia linkage and does not interact with the neoglycoproteins Neo6'SL, NeoM, NeoF. Moreover, it resulted in no impurities in the end product, as shown in the line 2 of the SDS-PAGE (*Fig 4.15A*).



**Fig. 4. 15 A** SDS PAGE **analysis of MAAI**. The protein sample was analysed under denaturing conditions on 14% polyacrylamide gel. Line 1: Protein marker, Line 2: MAAI (35 kDa) **B** GLYcoPROFILE in direct binding of MAAI interacting with labelled Neo6'SL, Neo3'SL, NeoMa and NeoF at 40  $\mu$ g/mL.

#### 4.7 SNA functionality

SNA purification was also carried out at GLYcoDiag: the protocol was developed before I joined the laboratory. Again, the purity of the lectin was assessed by SDS-PAGE. In *Fig. 4.16 A*, we can see in the line 2 of the gel the pure lectin. The activity and selectivity were tested by GLYcoPROFILE in direct binding: in *Fig. 4.16* the lectin shows interaction with Neo6'SL and no interaction with Neo3'SL, NeoF and NeoL, confirming the reverse specificity compared to MAAI.



**Fig. 4. 16 A** SDS PAGE analysis of SNA. The protein sample was analysed under denaturing conditions on 14% polyacrylamide gel. Line 1: Protein marker, Line 2: SNA (33 kDa) **B** GLYcoPROFILE in direct binding of SNA interacting with labelled Neo6'SL, Neo3'SL and NeoMannose at 40  $\mu$ g/mL.

#### 4.8 Parameters of $\alpha$ 2-3/ $\alpha$ 2-6 LEctPROFILE kit

The kit was optimized to discriminate the  $\alpha$ 2-3/ $\alpha$ 2-6 glycosidic linkages. In this specific case, the choice of the type of plate and the method of detection were immediately directed to the fluorescence detection on adsorption plate, based on previous studies conducted at GLYcoDiag before my arrival (data not shown). Based on previous studies performed at GLYcoDiag (data not shown), the quantity of lectin/well chosen for SNA is 2.5  $\mu$ g/well and for MAAI is 1  $\mu$ g/well.

##### 4.8.1 Direct binding

We performed the GLYcoPROFILE in direct experiments to test the specificity of the lectins MAAI and SNA and their selectivity towards the corresponding Sia linkages and to test the sensitivity of the kit, determining the limit of detection. The determination of the detection limit is useful to establish the minimum concentration of unknown sample necessary to have a robust interaction with the lectin on the plate. The neoglycoproteins Neo3'SL and Neo6'SL were chosen as a model probes. In *Fig. 4.17*, it is represented the direct binding curve of the lectin SNA interaction with Neo6'SL at starting concentration of 0.6  $\mu$ M and then serially diluted on the plate. The data obtained fit a non-linear regression model, from which the BC50 was calculated. The BC50, as for the Galili antigen LEctPROFILE kit, represents half of the maximum ligand concentration at which there is interaction with the lectin and for SNA interaction with Neo3'SL resulted to be 0.2  $\mu$ M. Then, we set the detection limit at 500 fluorescence units, below which, the signal can be considered as background noise. The concentration of Neo6'SL giving a signal below 500 fluorescence units in the chart in *Fig. 4.17* is 0.01  $\mu$ M.

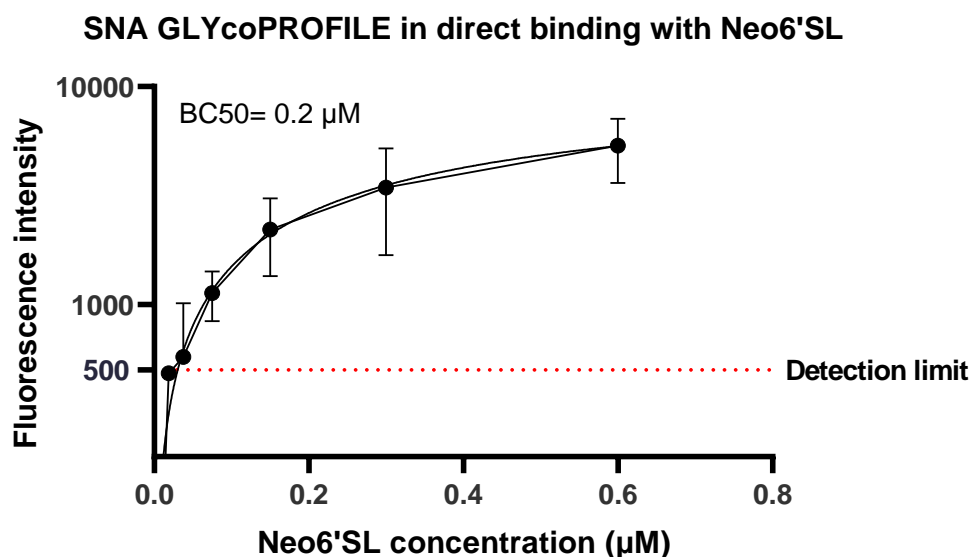


Fig. 4. 17 GLYcoPROFILE in direct binding representing the interaction between the lectin SNA and the labelled neoglycoprotein Neo6'SL at a starting concentration of 0.6  $\mu\text{M}$ . The curve is obtained from four independent replicates. All measurements were performed in duplicates.

The direct binding curve was performed as well for MAAI interacting with Neo3'SL at a starting concentration of 0.6  $\mu\text{M}$  and then serially diluted on the plate. In Fig 4.18 is shown the chart obtained: the data fit a non-linear regression model and the calculated BC50 is also in this case 0.2  $\mu\text{M}$ . The detection limit is below 0.01  $\mu\text{M}$ , concentration of Neo3'SL.

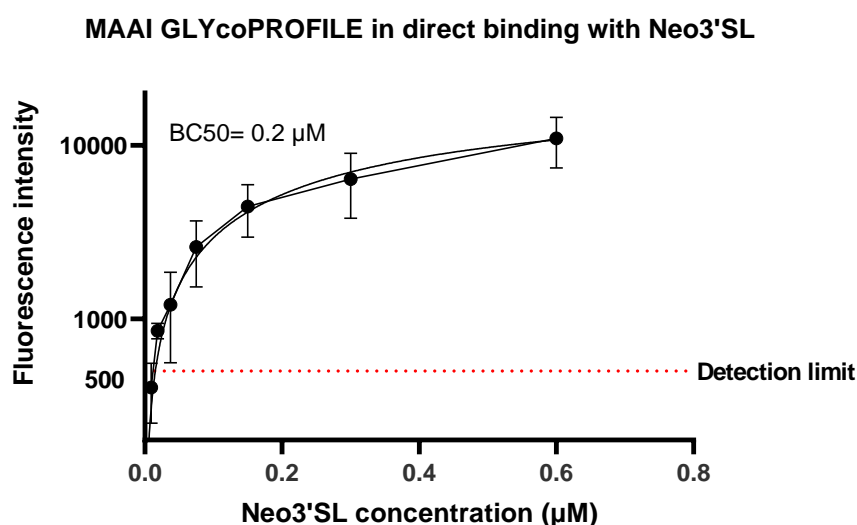


Fig. 4. 18 GLYcoPROFILE in direct binding representing the interaction between the lectin MAAI and the labelled neoglycoprotein Neo3'SL at a starting concentration of 0.6  $\mu\text{M}$ . The curve is obtained from four independent replicates. All measurements were performed in duplicates.

What emerges from this direct binding experiment is that the sensitivity of the kit is very good and it could allow to detect Sia linkages at concentrations up to 0.01  $\mu\text{M}$ .

#### 4.8.2 Inhibition

The inhibition experiments are designed to understand the ability of a compound to compete for the binding between the lectin and the specific neoglycoprotein. The labelled neoglycoproteins are the “tracer” applied to the lectin plate at a fixed concentration. The inhibitors are the non-labelled compounds applied to the plate in serial dilutions. For each tested inhibitor it was calculated the IC50 (half-maximal inhibitory concentration), which is the half concentration of inhibitor necessary to displace the interaction between the lectin and the tracer. After the determination of the IC50, the Relative Inhibitory Potency (RIP) was calculated (Table 4.1) for each inhibitor. This was done with



the purpose of standardizing all values and to have reference values to which a target unknown sample can be compared to. For the RIP, the obtained IC50 values were normalized to the IC50 value of nonconjugated trisaccharide 6'SL for SNA and 3'SL for MAAI. Thus, stronger binding to SNA or MAAI results in a lower IC50 value and a higher RIP. RIP was calculated according to the formula:

$$RIP = \frac{IC50(\text{simple trisaccharide})}{IC50(\text{inhibitor})}$$

The GLYcoPROFILE in inhibition is carried out using as tracer the labelled neoglycoproteins Neo6'SL and Neo3'SL, respectively, for SNA and MAAI. The choice of the concentration was determined according to the direct binding curve: the concentration that gives the 50% of interaction between the lectin and the neoglycoprotein (BC50) is chosen as the concentration of tracer for the inhibition experiment and for both lectins it was fixed at 20 µg/mL (*Fig. 4.17* and *4.18*). In *Fig. 4.19* are represented the SNA inhibition curves with Neo6'SL non labelled, transferrin, fetuin and the trisaccharide α2-6-sialyllactose (6'SL). The glycoproteins fetuin and transferrin were chosen because they are known to contain sialic acid linked both in α2-3 and α2-6 [129]. They were used respectively at a starting concentration of 10 µM for Neo6'SL, Fetuin and Transferrin and at 5 mM for 6'SL. In the experiment was also included simple Neu5Ac at 50 mM but it showed no inhibition either with SNA, nor with MAAI (data not shown), in accordance with literature [124] [123]. For MAAI inhibition curves, *Fig. 4.20*, were chosen Neo3'SL non labelled, Fetuin and the trisaccharide α2-3-sialyllactose (3'SL) but not transferrin because this glycoprotein doesn't show any interaction with the lectin in direct binding (*Fig. 4.24*). They were used respectively at a starting concentration of 10 µM for Neo3'SL and Fetuin and at 5 mM for 'SL.

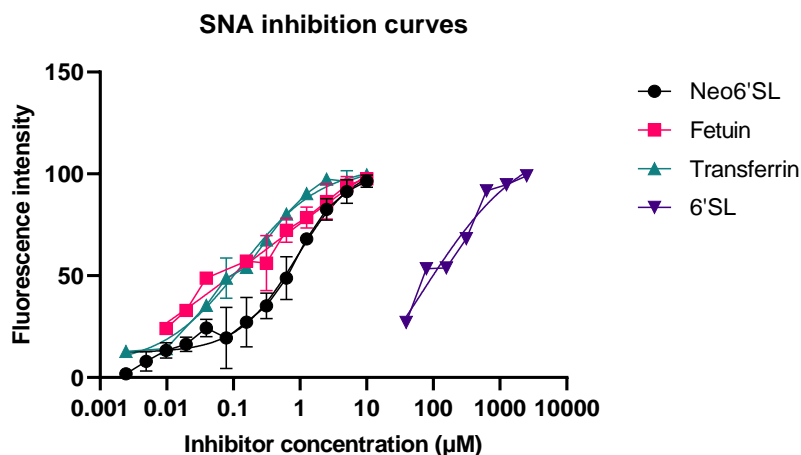


Fig. 4. 19 Inhibition curves of SNA with the inhibitors Neo6'SL, Fetuin, Transferrin at the starting concentration of 10  $\mu\text{M}$  and 6'SL at the starting concentration of 5mM. The labelled tracer Neo6'SL is at the fixed concentration of 20  $\mu\text{g/mL}$ . The curve is obtained from three independent replicates. All measurements were performed in duplicates.

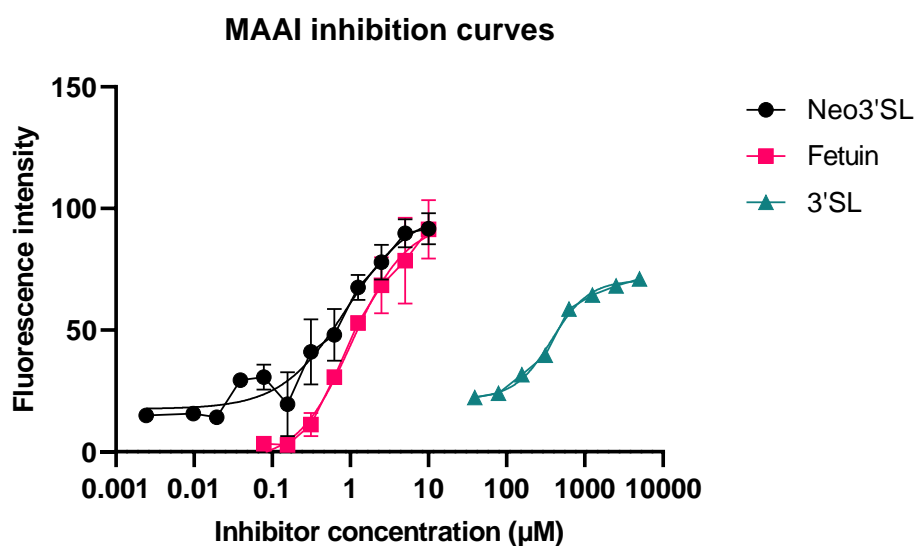


Fig. 4. 20 Inhibition curves of SNA with the inhibitors Neo3'SL and Fetuin at the starting concentration of 10  $\mu\text{M}$  and 3'SL at the starting concentration of 5mM. The labelled tracer Neo3'SL is at the fixed concentration of 20  $\mu\text{g/mL}$ . The curve is obtained from three independent replicates. All measurements were performed in duplicates.

SNA			MAAI		
Compound	IC50	RIP	Compound	IC50 ( $\mu\text{M}$ )	RIP
6'SL	200	1	3'SL	500	1
Neo6'SL	0.6	333	Neo3'SL	0.6	833
Fetuin	0.15	1333	Fetuin	1.12	446

Transferrin	0.15	1333
-------------	------	------

Table 4. 1 IC50 and RIP values calculated for all inhibitors tested with the lectins SNA and MAAI. The calculation is based on the inhibition curves in Fig.4.16 and 4.17.

The values of IC50 and RIP obtained for the two lectins are reported in Table 4.1. The IC50 values obtained for the two lectins with the respective trisaccharide are in accordance with previous studies [124] [123]. It is interesting to notice that the RIP for fetuin interacting with SNA is higher than for fetuin interacting with MAAI, that could mean this glycoprotein presents more  $\alpha$ 2-6 linkages than  $\alpha$ 2-3 which would be in accordance with literature[129].

#### 4.9 Validation of the kit: a comparison study

Taking into account the applications reviewed at the beginning of this chapter, it is of great interest the quantification of  $\alpha$ 2-3- and  $\alpha$ 2-6- sialyllactose in a rapid and simple way. The goal of this study is to do it through two relevant and complementary methods, the MALDI-TOF-MS and the GLYcoPROFILE®. The purpose of the collaboration with Juvissan Aguedo of the Slovak Academy of Sciences was to calculate the ratio of  $\alpha$ 2-6- and  $\alpha$ 2-3- glycosidic linkages contained in reference commercial glycoproteins through the two techniques and compare the results obtained.

##### 4.9.1 GLYcoPROFILE® and MALDI-TOF-MS of reference glycoproteins

We used four reference native glycoproteins listed in the Table 4.2. The choice of these particular glycoproteins is due to their interesting applications, for example transferrin, which has gained attention in the forensic field (Paragraph 4.2.1), and to the fact that their glycosylation was previously described. However, in literature is missing the exact quantification of the Sia linkages on these glycoproteins, a part from Fetuin for which the percentage of  $\alpha$ 2-3- and  $\alpha$ 2-6- sialyllactose has been described [28], hence, for this reason, it can be considered as a control.

Name	Supplier	Characteristics	Glycosylation according to literature
<b>Fetuin from foetal calf serum</b>  48.4 kDa	BioRad 4430-2204	Blood protein produced in the liver and secreted into the bloodstream. It is a binding protein mediating the transport and availability of a wide variety of cargo substances in the bloodstream.	3N- an 3 O-linked (mucine-type)glycans [130]  Complex glycan containing NeuAc (around 60 % $\alpha$ 2-6/40 % $\alpha$ 2-3) [129].

Name	Supplier	Characteristics	Glycosylation according to literature
<b>Human Transferrin</b> 76 kDa	Sigma T3309	It is the main iron transporter serum glycoprotein. It consists of a polypeptide chain with two binding sites for iron.	2 N-linked complex glycans containing NeuAc (a2,3/a2,6) <a href="https://glyconnect.expasy.org/browser/proteins?f=transferrin">https://glyconnect.expasy.org/browser/proteins?f=transferrin</a>
<b>Bovine Thyroglobulin</b> 670 kDa	Sigma T1001	Bovine thyroglobulin, is a precursor of the iodated thyroid hormones thyroxine (T4) and triiodothyronine (T3)	13 N-glycosylation sites complex, hybrid, high mannose [31]. In bovine thyroglobulin, on the contrary to the human one, no O-glycosylation occurs
<b>Porcine Thyroglobulin</b> 670kDa	Sigma T1126	Porcine thyroglobulin, is a precursor of the iodated thyroid hormones thyroxine (T4) and triiodothyronine (T3)	High mannose/complex 3 N-linked sites containing Galili, a2,6/a2,3 NeuAc, O-linked containing NeuAc/NeuGc [131].

Table 4. 2 Characteristics of glycoproteins considered in the study.

The GLYcoPROFILE experiments were carried out in direct binding. As negative control, digestion with  $\alpha$ 2-3 neuraminidase (Neu $\alpha$ 2-3) and total neuraminidase (NeuT) was carried out in order to confirm that the results were in good agreement with the expected specificities of the lectins. Neuraminidase is the common name for the Acetyl-neuraminyl hydrolase also known as Sialidase.  $\alpha$ 2-3 Neuraminidase (Neu $\alpha$ 2-3) is a highly specific exoglycosidase that catalyses the hydrolysis of  $\alpha$ 2-3 linked N-acetylneuraminic acid residues from glycoproteins and oligosaccharides. The  $\alpha$ 2-3,6,8 Neuraminidase (NeuT) catalyses the hydrolysis of  $\alpha$ 2-3,  $\alpha$ 2-6, and  $\alpha$ 2-8 linked N-acetylneuraminic acid residues from glycoproteins and oligosaccharides. What we expected to see is a drop of fluorescence intensity for the lectins SNA and MAAI interaction with samples treated with NeuT and a drop only in the fluorescence intensity interaction of MAAI with the samples treated with Neu $\alpha$ 2-3.

The ratio  $\alpha$ 2-3/  $\alpha$ 2-6 was calculated based on the equivalents of neoglycoprotein determined for each glycoprotein, for the highest concentration tested. To do so, we performed a direct binding curve for SNA and MAAI with the corresponding neoglycoproteins Neo6'SL and Neo3'SL. For each curve a linear regression equation was calculated (*Fig. 4.21*) and from each y values corresponding

to the fluorescence intensity of the native glycoprotein interaction with the lectin, a  $x$  value was calculated that corresponds to the equivalent concentration in neoglycoprotein. We have decided to give the fluorescence a quantification in equivalents of neoglycoprotein because the neoglycoproteins are an optimized tool, representing the Sia linkages and they function as a standard of which we know precisely the concentration.

The MALDI-TOF/MS with derivatization of sialic acid linkages was performed at the Slovak Academy of Sciences in Bratislava by another PhD student of the synBIOcarb network, Juvissan Aguedo (Description of the method: Chapter 1, Paragraph 1.5.3).

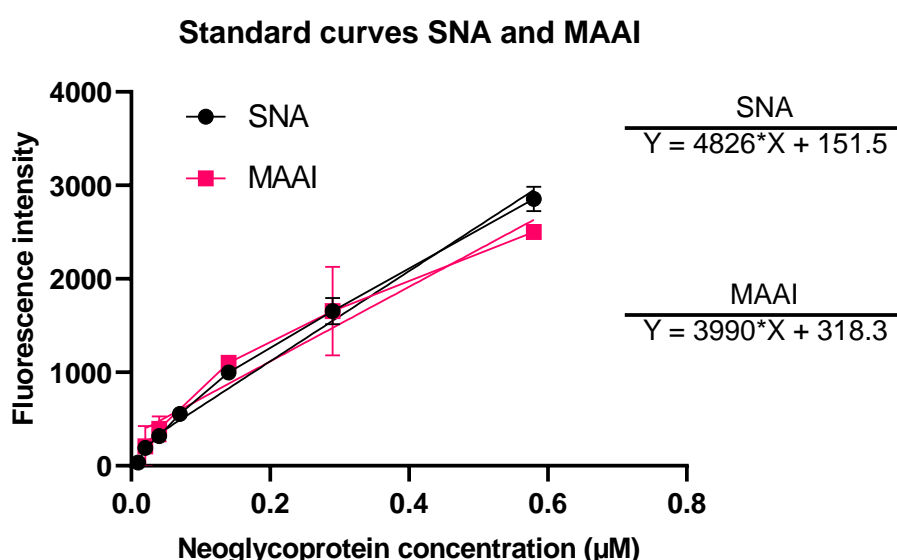


Fig. 4. 21 SNA and MAAI standard curves obtained with labelled Neo6'SL and Neo3'SL.

#### 4.9.2 Fetuin

The use of bovine fetuin is widespread as a model glycoprotein for the study of the biosynthesis, structure, functions and lately also for the carbohydrate characterization. N-linked and O-linked oligosaccharide on Bovine Fetuin have been extensively described, and to date it is known that the N-linked account for the 80% of Fetuin oligosaccharide while the O-linked for the 20% [132].

Sialic acid linkage isomers contained into fetuin were recently studied by Guttman and Lee by Ion Mobility Spectrometry in addition to conventional glycoproteomic platforms with the main purpose of differentiating sialic acid linkages on both N- and O-linked glycopeptides [129]. They state that with their technique, the overall ratios of sialic acid linkage isomers observed on Fetuin is 38.4% for  $\alpha$ 2-3 and 59.4% for  $\alpha$ 2-6, with a 2.2% attributed to sialylation at the GlcNAc. These results are consistent

with previous quantitative measurements of  $\alpha 2\text{-}6\text{:}\alpha 2\text{-}3$  sialylation ratio of Fetuin that range from 62:38 up to 49:51. The oscillation of the percentage could be possibly due to differences in the methods used for glycan isolation, the sample preparations and the data collection.

Observing the GLYcoPROFILE in Fig. 4.22, the experiment gives coherent and reliable results. In fact, as expected, a decrease in the signal for the interaction SNA/NeuT+Fet and MAAI/NeuT+Fet. The digestion of the glycoprotein with  $\alpha 2\text{-}3$  neuraminidase worked too: being MAAI the lectin that recognize the  $\alpha 2\text{-}3$  linkage, there is a decrease only in the interaction MAAI/Neu $\alpha 2\text{-}3$ +Fetuin but not for SNA/Neu $\alpha 2\text{-}3$ +Fetuin. The ratio  $\alpha 2\text{-}6\text{:}\alpha 2\text{-}3$  sialylation, obtained as equivalents of neoglycoprotein, is 56:44: this observation is in line with the previously cited studies [129]. To confirm the level of  $\alpha 2\text{-}3$ - or  $\alpha 2\text{-}6$ -terminally sialylated structures, the background level of interaction obtained after neuT digestion was subtracted from the results obtained for the interaction of each lectin with the native glycoprotein at the maximum concentration, following the method of Damiani *et al* [133]. The calculated ratio  $\alpha 2\text{-}6\text{:}\alpha 2\text{-}3$ , which is 54:46, is in agreement with literature and the calculation based on the neoglycoprotein equivalents.

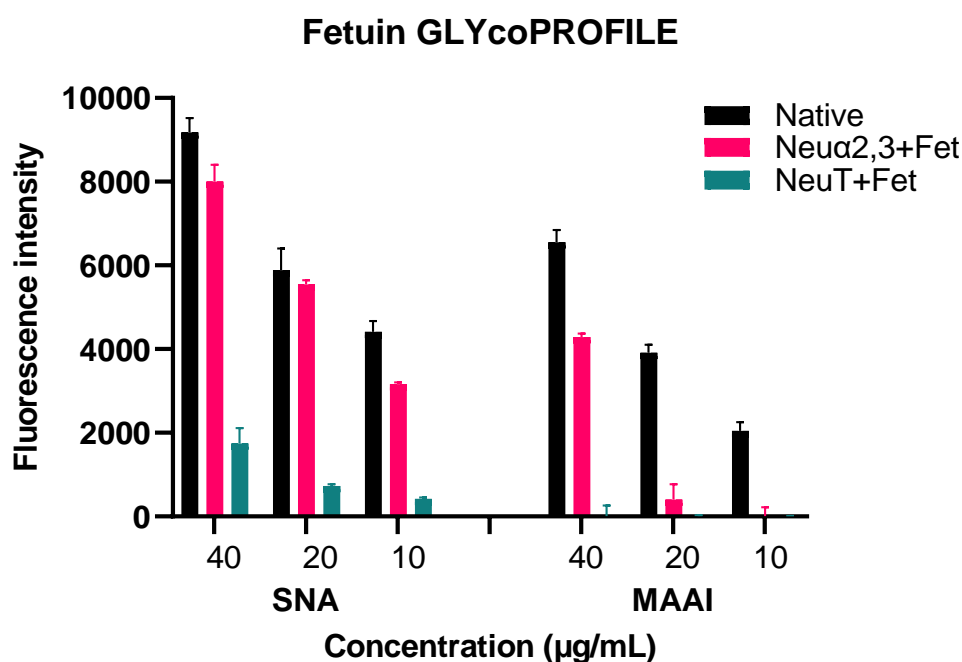


Fig. 4. 22 Interaction of MAAI and SNA specific lectins with labelled Fetuin at three different concentrations. The intensity of interaction of each lectin was studied before and after Fetuin digestion by total neuraminidase (NeuT) and  $\alpha 2\text{-}3$  neuraminidase (Neu $\alpha 2\text{-}3$ ). Two independent assays (each with  $n = 3$  replicates) were carried out.

Concentration ( $\mu\text{g/mL}$ )	Equivalent of neoglycoprotein ( $\mu\text{g/mL}$ )		% $\alpha 2\text{-}6\text{:}\alpha 2\text{-}3$
	SNA	MAAI	54:46

40	1.9	1.6	
----	-----	-----	--

Table 4. 3 Equivalents of neoglycoprotein and ratio  $\alpha$ 2-6: $\alpha$ 2-3 on fetuin, calculated at the maximum concentration based on the standard curve Fig. 4.21.

However, the results reported for fetuin by MALDI-TOF-MS are not in agreement with the GLYcoPROFILE neither with literature. According to MS data, the most abundant sialylated glycan in Fetuin is a non-core fucosylated bi-antennary di-sialylated glycan with full  $\alpha$ 2-6-sialylation (Fig. 4.23 A). The difference of ratio  $\alpha$ 2-6: $\alpha$ 2-3 (Table 4.4) is too drastic (97:3 versus 56:44) to attribute the diverging results to the techniques used so an hypothesis might be that the lack of  $\alpha$ 2-3 sialylation is due to the fact that only the fetuin N-glycans were analysed in this MALDI-TOF-MS experiments, in fact, the sample preparation includes only the digestion with PNGase F.

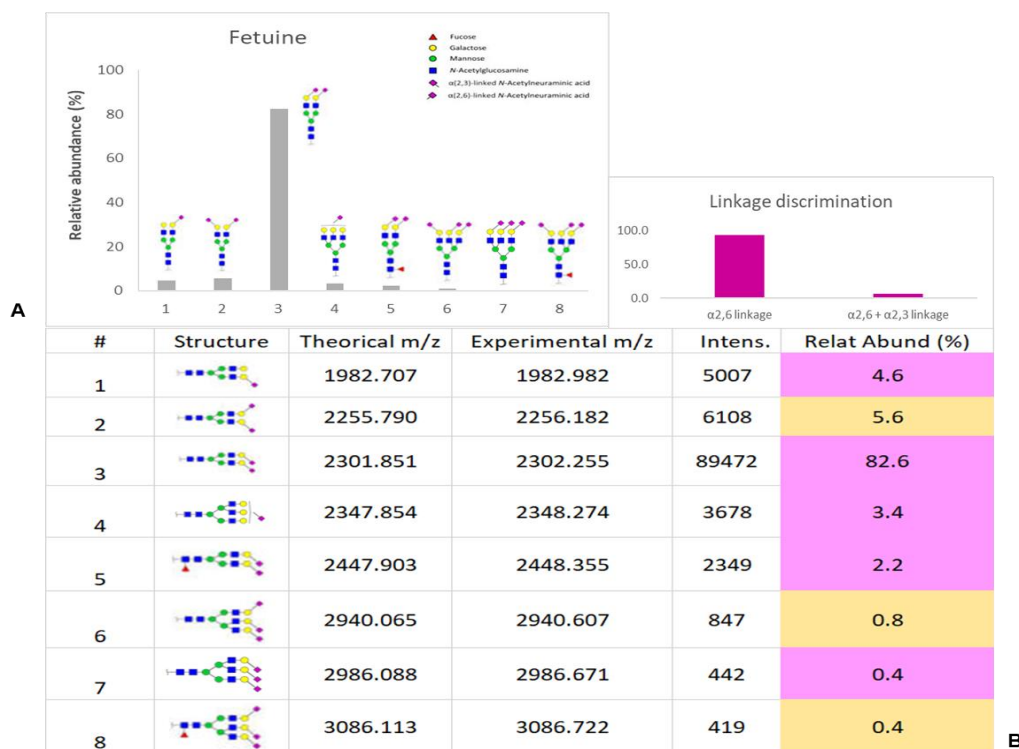


Fig. 4. 23 MALDI-TOF/MS. **A** Possible sialylated structures present on Fetuin. **B** Relative abundance of  $\alpha$ 2-6 and  $\alpha$ 2-6+ $\alpha$ 2-3- sialic acid linkages present in sialylated glycans.

Linkage	Relative abundance %	Ratio $\alpha$ 2-6: $\alpha$ 2-3
$\alpha$ 2-6 sialyllactose	96.79	



$\alpha$ 2-3 sialyllactose	3.3	97:3
$\alpha$ 2-6 + $\alpha$ 2-3 linkage	6.8	

Table 4. 4. Summary of glycosidic linkages relative abundances and ratio  $\alpha$ 2-6: $\alpha$ 2-3 on calculated according to MALDI-TOF/MS with linkages derivatization.

#### 4.9.3 Transferrin

The studies that focus on human transferrin sialylation have as their main object the discrimination of transferrin glycoforms, classified by the number of sialic acids branches found on the two N-glycosylation sites of this glycoprotein. We wanted to have a more complete view on this matter, searching for the type of Sia linkages found on human transferrin and the ratio  $\alpha$ 2-6: $\alpha$ 2-3. The GLYcoPROFILE in Fig. 4.24 implies that Transferrin only contains the  $\alpha$ 2-6 linkage, given that the only interaction that we see is with the lectin SNA. This is not in accordance with the information found on Glyconnect by Expasy, neither with the MALDI-TOF-MS results and could result from a limited accessibility of the  $\alpha$ 2-3 linkage by the lectin.

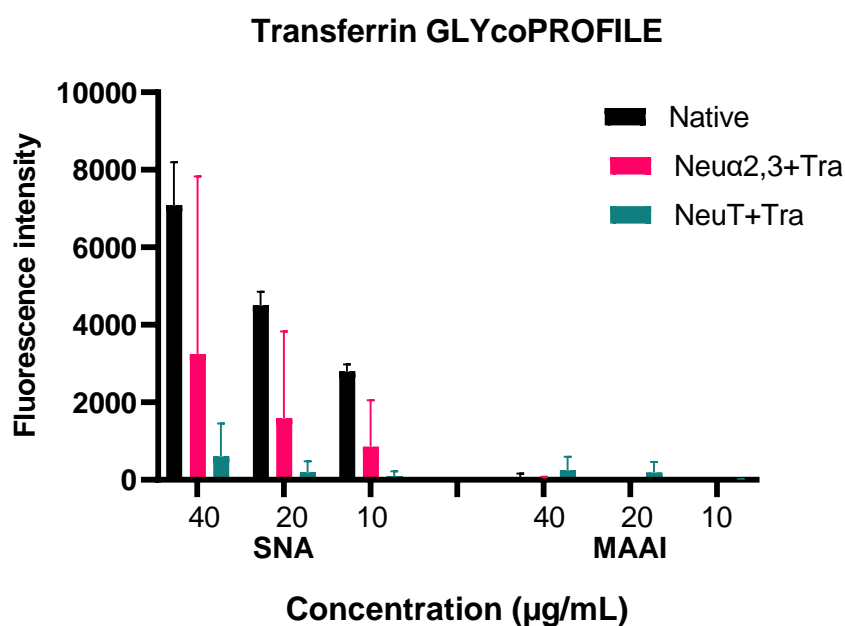


Fig. 4. 24 Interaction of MAAI and SNA specific lectins with labelled transferrin at three different concentrations. The intensity of interaction of each lectin was studied before and after transferrin digestion by total neuraminidase (NeuT) and  $\alpha$ 2-3 neuraminidase (Neu $\alpha$ 2-3). Two independent assays (each with n = 3 replicates) were carried out.

Concentration ( $\mu\text{g/mL}$ )	Equivalent of neoglycoprotein ( $\mu\text{g/mL}$ )		Ratio $\alpha 2\text{-}6\text{:}\alpha 2\text{-}3$
	SNA	MAAI	
40	1.46	0,016	99:1

Table 4. 5 Equivalents of neoglycoprotein and ratio  $\alpha 2\text{-}6\text{:}\alpha 2\text{-}3$  on transferrin, calculated at the maximum concentration based on the standard curve Fig. 4.21.

To know if this hypothesis of the accessibility was correct, we performed a GLYcoPROFILE of transferrin both in native and in denaturing conditions. As shown in Fig.4.25, the fluorescence units value of MAAI incubated with denatured transferrin are 10 folds higher than values obtained after lectin incubation with native transferrin. For both lectins we can see an increase of fluorescence intensity when transferrin is in denaturing conditions, but, the difference is that for SNA the increment is proportional to the fluorescence values of the native transferrin at the three different concentrations, while for MAAI the native transferrin gives a signal close to zero, proving that there is a problem of accessibility. The ratio  $\alpha 2\text{-}6\text{:}\alpha 2\text{-}3$  calculated on the neoglycoprotein equivalents goes from 99:1 for native transferrin to 56:44 for the denatured (Table 4.6).

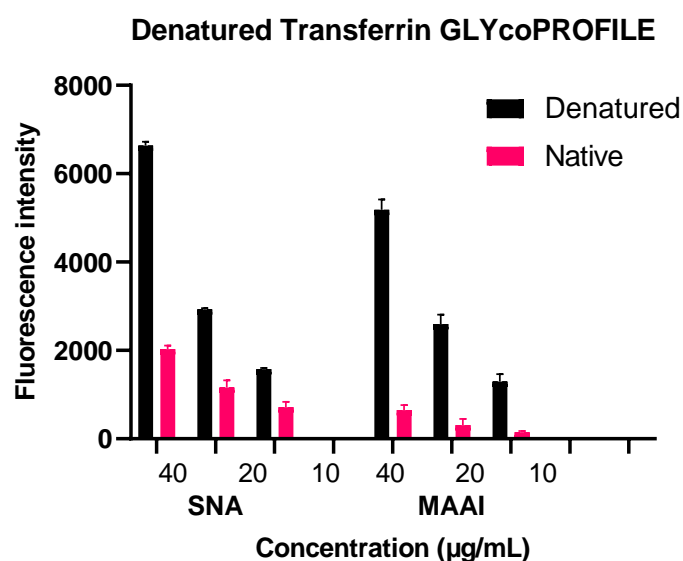


Fig. 4. 25 Interaction of MAAI and SNA specific lectins with labelled transferrin native and denatured at three different concentrations. Two independent assays (each with  $n = 3$  replicates) were carried out.

	Concentratio	Equivalent of neoglycoprotein	Ratio $\alpha 2\text{-}3\text{:}\alpha 2\text{-}6$
--	--------------	----------------------------------	--

Glycoprotein	n (µg/mL)	(µg/mL)		
		SNA	MAAI	
Transferrin native	40	0.42	0.016	99:1
Transferrin denatured	40	1.4	1.1	56:44

Table 4. 6 Comparison of equivalents of neoglycoprotein and ratio  $\alpha$ 2-6: $\alpha$ 2-3 on transferrin native and denatured, calculated at the maximum concentration based on the standard curve Fig. 4.21.

According to the MALDI-TOF MS, the most abundant species is a non-core fucosylated tri-antennary tri-sialylated glycan with two  $\alpha$ 2-6-linkages and one  $\alpha$ 2-3-linkage (Fig 4.26). The ratio  $\alpha$ 2-6: $\alpha$ 2-3 (Table 4.7) is comparable to the one obtained in GLYcoPROFILE, being 65:35, in this case the small difference can be attributed to the assay.

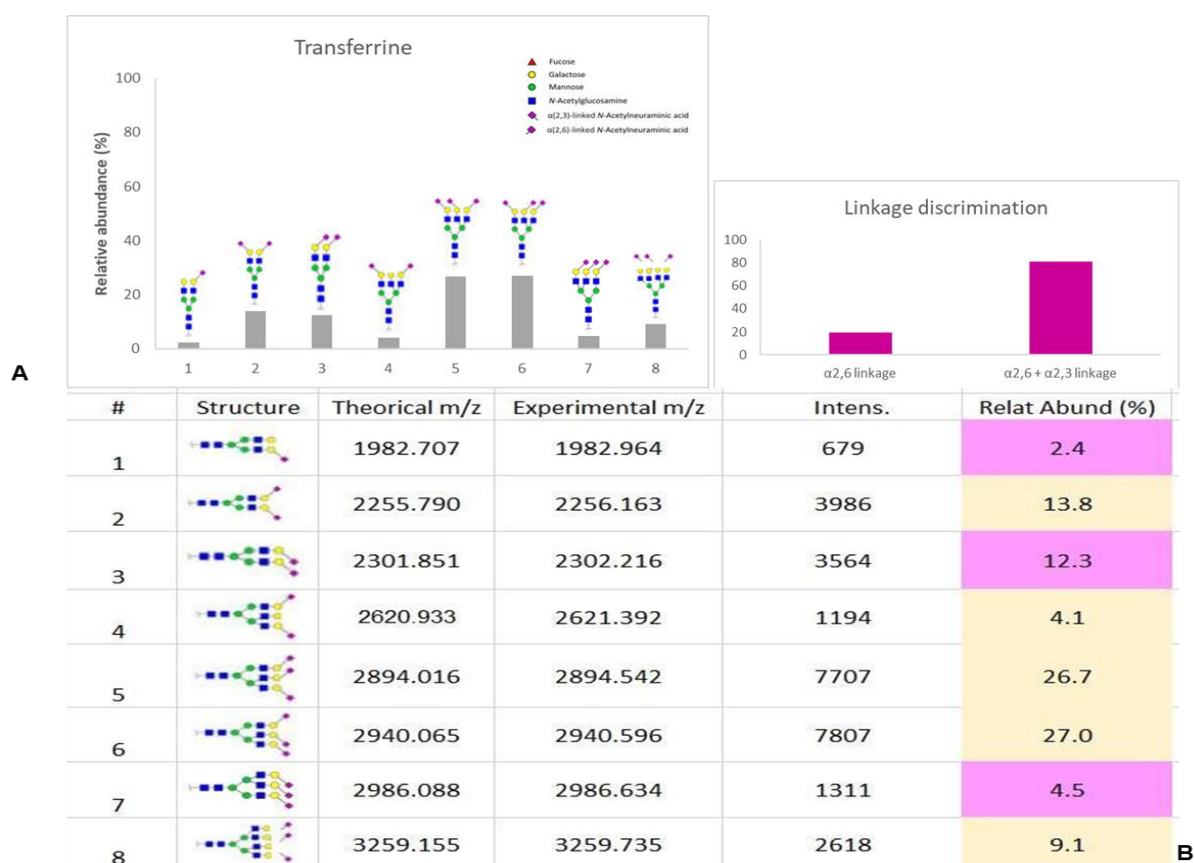


Fig. 4. 26 MALDI-TOF/MS **A** Possible sialylated structures present on Transferrin. **B** Relative abundance of  $\alpha$ 2-6 and  $\alpha$ 2-6+ $\alpha$ 2-3- sialic acid linkages present in sialylated glycans.

Linkage	Relative abundance %	Ratio $\alpha$ 2-6: $\alpha$ 2-3
$\alpha$ 2-6 sialyllactose	56.6	65:35
$\alpha$ 2-3 sialyllactose	29.83	
$\alpha$ 2-6 + $\alpha$ 2-3 linkage	80.7	

Table 4. 7 Summary of glycosidic linkages relative abundances and ratio  $\alpha$ 2-6: $\alpha$ 2-3 on transferrin calculated according to MALDI-TOF/MS with linkages derivatization.

#### 4.9.4 Porcine and Bovine Thyroglobulin

The GLYcoPROFILE for both the Thyroglobulin is shown in Fig.4.27 A and B. In the case of both Porcine and Bovine Thyroglobulin there is more variability in the two experiments considering the error bars. The digestion with NeuT and Neu $\alpha$ 2-3 neuraminidase did not seem to completely work, probably because of the big size of these glycoproteins. Overall, we see a ratio  $\alpha$ 2-3: $\alpha$ 2-6 around 55:45 for Porcine Thyroglobulin and 52:48 for Bovine Thyroglobulin.

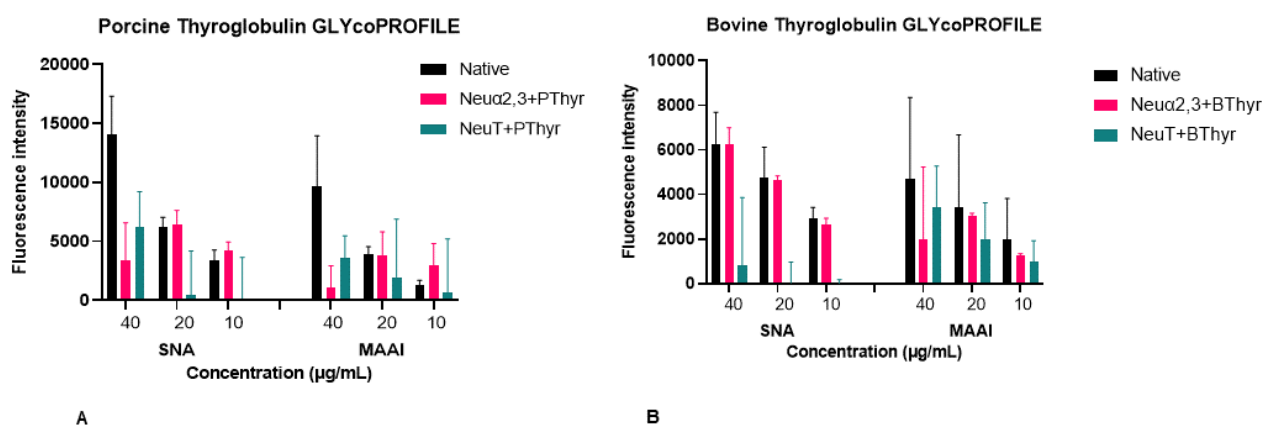


Fig. 4.27 **A** Interaction of MAAI and SNA specific lectins with labelled porcine thyroglobulin at three different concentrations. **B** Interaction of MAAI and SNA specific lectins with labelled bovine thyroglobulin at three different concentrations. The intensity of interaction of each lectin was studied before and after thyroglobulin digestion by total neuraminidase (NeuT) and  $\alpha$ 2-3 neuraminidase (Neu $\alpha$ 2-3). For each protein two independent assays (each with  $n = 3$  replicates) were carried out.

Glycoprotein	Concentration (μg/mL)	Equivalent of neoglycoprotein (μg/mL)		Ratio $\alpha$ 2-6: $\alpha$ 2-3
		SNA	MAAI	

Porcine Thyroglobulin	40	2.9	2.4	55:45
Bovine Thyroglobulin	40	1.3	1.2	52:48

Table 4. 8 Equivalents of neoglycoprotein and ratio  $\alpha$ 2-6: $\alpha$ 2-3 on both thyroglobulins, calculated at the maximum concentration based on the standard curve Fig. 4.21.

According to the MALDI-TOF-MS, Bovine Thyroglobulin presents a fucosylated bi-antennary di-sialylated glycan with fully  $\alpha$ 2-6-sialylation and Porcine Thyroglobulin presents a fucosylated bi-antennary mono-sialylated glycan with one  $\alpha$ 2-6-sialylation (Fig. 4.28). Only  $\alpha$ 2-6 glycosidic linkage is reported (Table 4.9) as opposite as the GLYcoPROFILE in in which also  $\alpha$ 2-3 is detected.

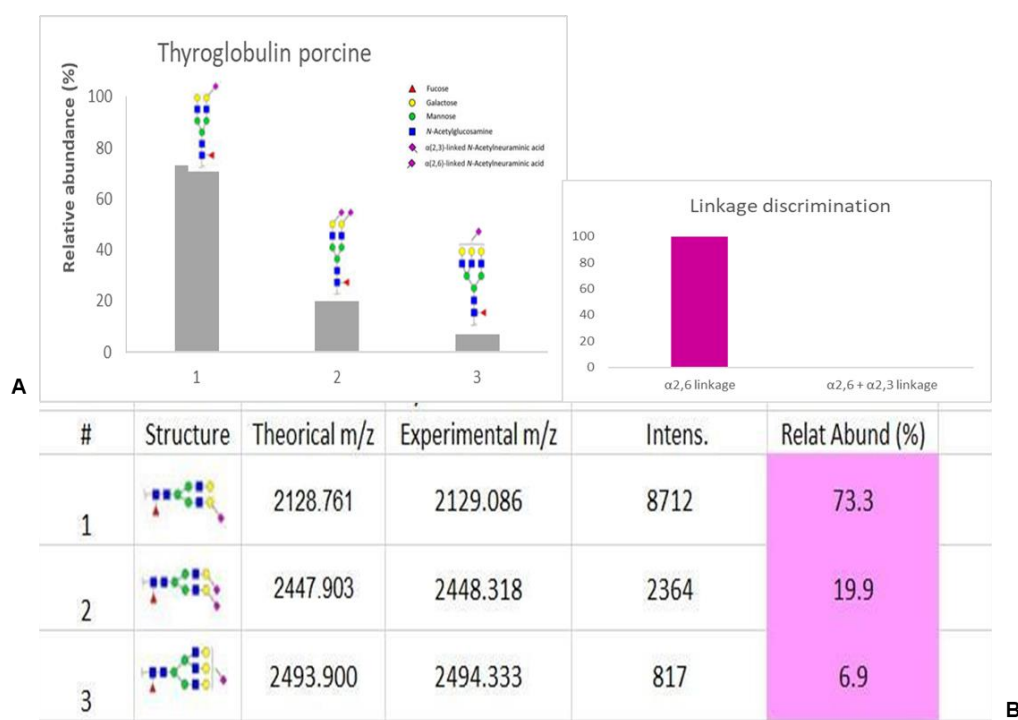


Fig. 4. 28 MALDI-TOF/MS **A** Possible sialylated structures present on porcine thyroglobulin. **B** Relative abundance of  $\alpha$ 2-6 and  $\alpha$ 2-6+ $\alpha$ 2-3- sialic acid linkages present in sialylated glycans.

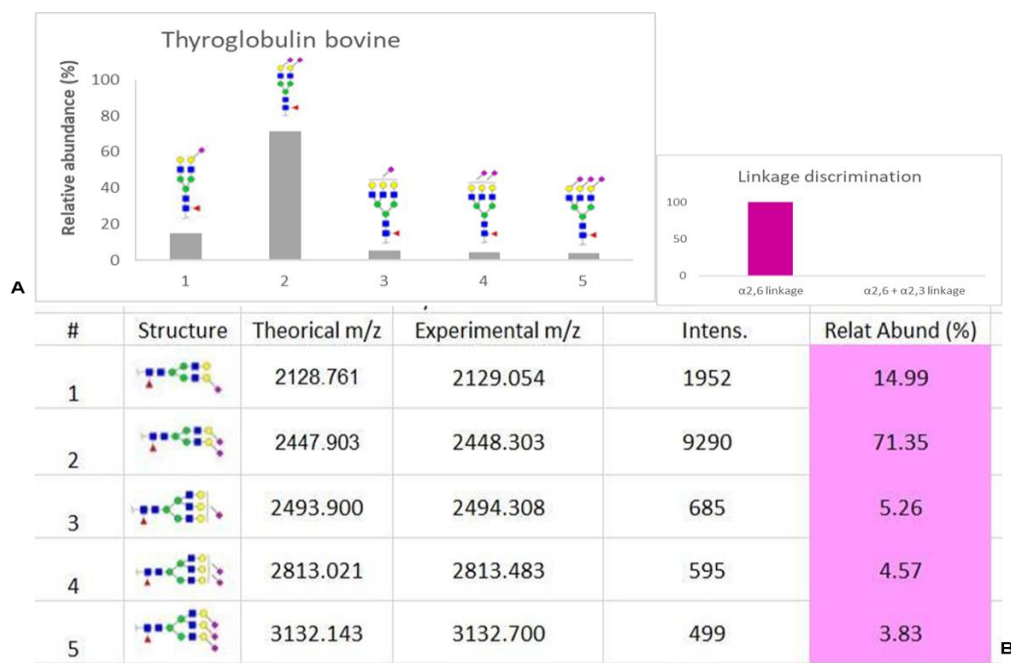


Fig. 4. 29 **A** Possible sialylated structures present on bovine thyroglobulin. **B** Relative abundance of  $\alpha$ 2-6 and  $\alpha$ 2-6+ $\alpha$ 2-3-sialic acid linkages present in sialylated glycans.

Linkage	Relative abundance %
$\alpha$ 2-6 sialyllactose	100
$\alpha$ 2-3 sialyllactose	0
$\alpha$ 2-6 + $\alpha$ 2-3 linkage	0

Table 4. 9 Summary of glycosidic linkages relative abundances of both porcine and bovine thyroglobulin calculated according to MALDI-TOF/MS with linkages derivatization.

The two methods agree on the fact that the  $\alpha$ 2-6 linkage is the one always in major quantity; sialylated glycans containing  $\alpha$ 2-6-linkages are more abundant in fetuin, porcine and bovine thyroglobulin whereas glycans with both linkages  $\alpha$ 2-6- and  $\alpha$ 2-3- were found in lower levels, except in transferrin, which presents higher abundance of glycans containing  $\alpha$ 2-6- and  $\alpha$ 2-3- sialylation. However, the two methods are not completely in accordance because with the GLYcoPROFILE we saw the presence of  $\alpha$ 2-3 linkages also in fetuin and in both thyroglobulins. In the case of transferrin, we see that the two methods give a very similar ratio  $\alpha$ 2-6: $\alpha$ 2-3, but, in the case of the GLYcoPROFILE we needed to perform the denaturation of transferrin to gain accessibility to the  $\alpha$ 2-3 linkages: Further analyses need to be performed to clarify why there is such a big difference in the detection of the  $\alpha$ 2-3 linkages. One idea could be to introduce in the study another glycoprotein well characterized and of a lower molecular weight than the thyroglobulins, which can be for example the

alpha(1)-Acid glycoprotein (AGP), with a molecular weight of 41-43 kDa and heavily glycosylated (45%), which is also already described [129].

This work wants to show the first step towards a complementary and high throughput method for discrimination of  $\alpha$ 2-6- and  $\alpha$ 2-3-sialylation to be used in biomarker discovery and in quality control of biotherapeutics proteins. The advantage of MALDI-MS method for derivatized glycans is the possibility to get additional structural information regarding the exact position of sialic acid linkage and when both linkages  $\alpha$ 2-6- and  $\alpha$ 2-3- are present on the glycan. The advantage of the GLYcoPROFILE is to have a very fast analysis because the results are obtained in a couple of hours and moreover the glycans don't need to be removed from the glycoproteins so we can obtain additional information on their accessibility.

#### 4.10 GLYcoPROFILE® of cells

##### 4.10.1 Chinese Hamster Ovary Cells

After the validation of the kit with commercial glycoproteins, the goal was to have the proof of concept of its exploitability for cell lines, in the perspective of potential research and development applications, such as following the production of a recombinant biotherapeutic glycoprotein in cell lines. The goal of the experiments described here is to highlight the glycosylation and, in particular, the  $\alpha$ 2-3/ $\alpha$ 2-6 sialylation of **glycoproteins** expressed at the **surface** of CHO cells or **released** in the culture medium. The brief history of CHO cell line is that in 1957, a parental CHO cell line was initiated from a biopsy of an adult, female Chinese hamster from which the CHO cell line was then derived as a subclone. For this study, we have carried out the GLYcoPROFILE [134] in direct binding of CHO k1 adherent cells, their supernatant and their culture medium. On the same day of the GLYcoPROFILE, cells are harvested and labelled with CFSE (Carboxyfluorescein succinimidyl ester). It is a fluorescent molecule resulting from CFDA-SE (carboxyfluorescein diacetate succinimidyl ester), a cell permeant molecule which diffuse into the cells and it is cleaved by intracellular esterases (*Fig 4.30*). The CFSE binds indiscriminately with intracellular free amines to generate covalent fluorescent dye-protein conjugates resulting in live cells with an intracellular fluorescent label.



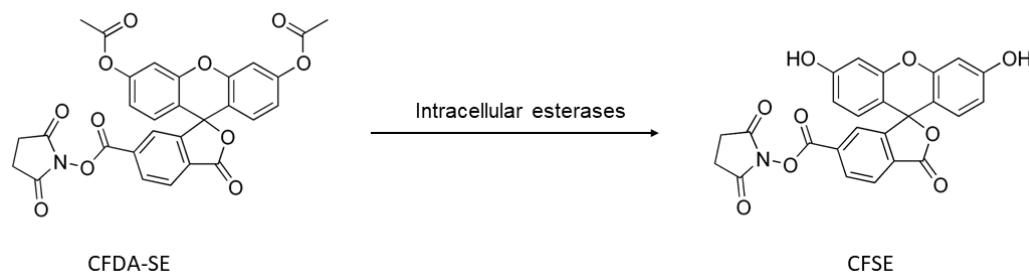


Fig. 4. 30 Intracellular reaction of the non-fluorescent CFDA-SE cells-permeant molecule which diffuses into the cells and it is converted into the fluorescent CFSE by intracellular esterases.

Generally, for cells GLYcoPROFILE, the results are not expressed in terms of fluorescence intensity but rather as percentage of bound cells, which is more comparable between experiments, while the fluorescence can move according to the quality and the concentration of cell during the labelling. The percentage is obtained thanks to a reference linear regression equation calculated from the standard curve in Fig 4.31A.

For the experiment in Fig. 4.31B, the panel of lectins was chosen based on previous results (data not shown) and on literature (for a detailed description of all lectins used in this thesis refer to Annex I). According to literature, both CHO grown in monolayer (adherent) and in suspension have a similar range of glycan structures. High mannose structures, complex N-glycans (mixture of bi-, tri-, tetra-antennary structures) fucosylated at the core and bearing LacNAc extensions, capped with sialic acid residues. Sia is predominant in the Neu5Ac form, the NeuGc in small quantities can be attributed to the presence of foetal bovine serum in the culture medium and to the fact that CHO can synthesise a small amount of CMP-NeuGc. Not a significant number of hybrid structures was observed [135]. The results are in agreement with what it was expected in terms of glycan structures: the lectins ConA, GNA and PSA highlight the presence of high-mannose and core-fucosylated structures, WFA and DSA respectively GalNAc and GlcNAc, AIA could underline the presence of Lac residues while PHA-L PHA-E the presence of complex N-glycans. It is striking the interaction with MAAI and the absence of signal for SNA, linked to the well-known lack of ST6 and thus  $\alpha$ 2-6-sialylation in CHO cells.

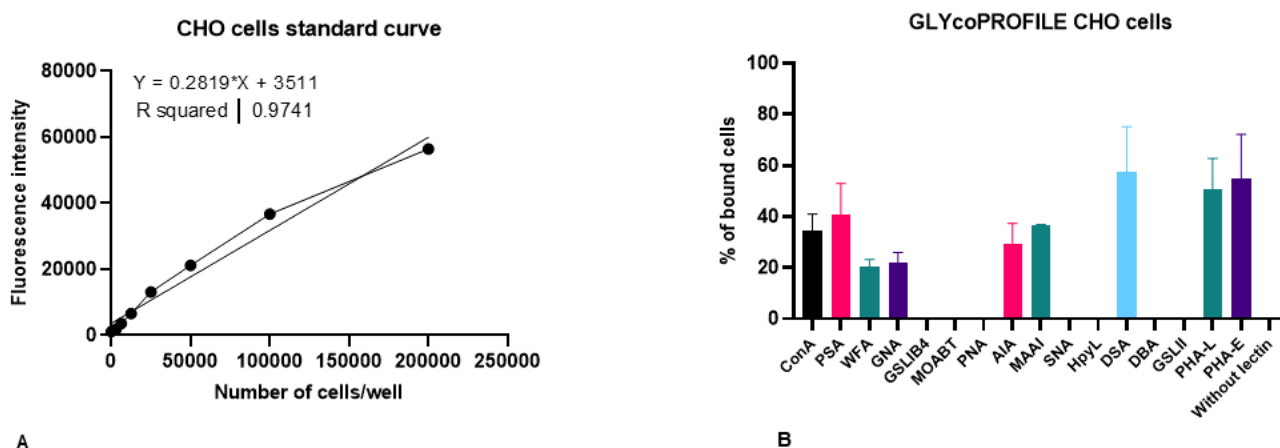
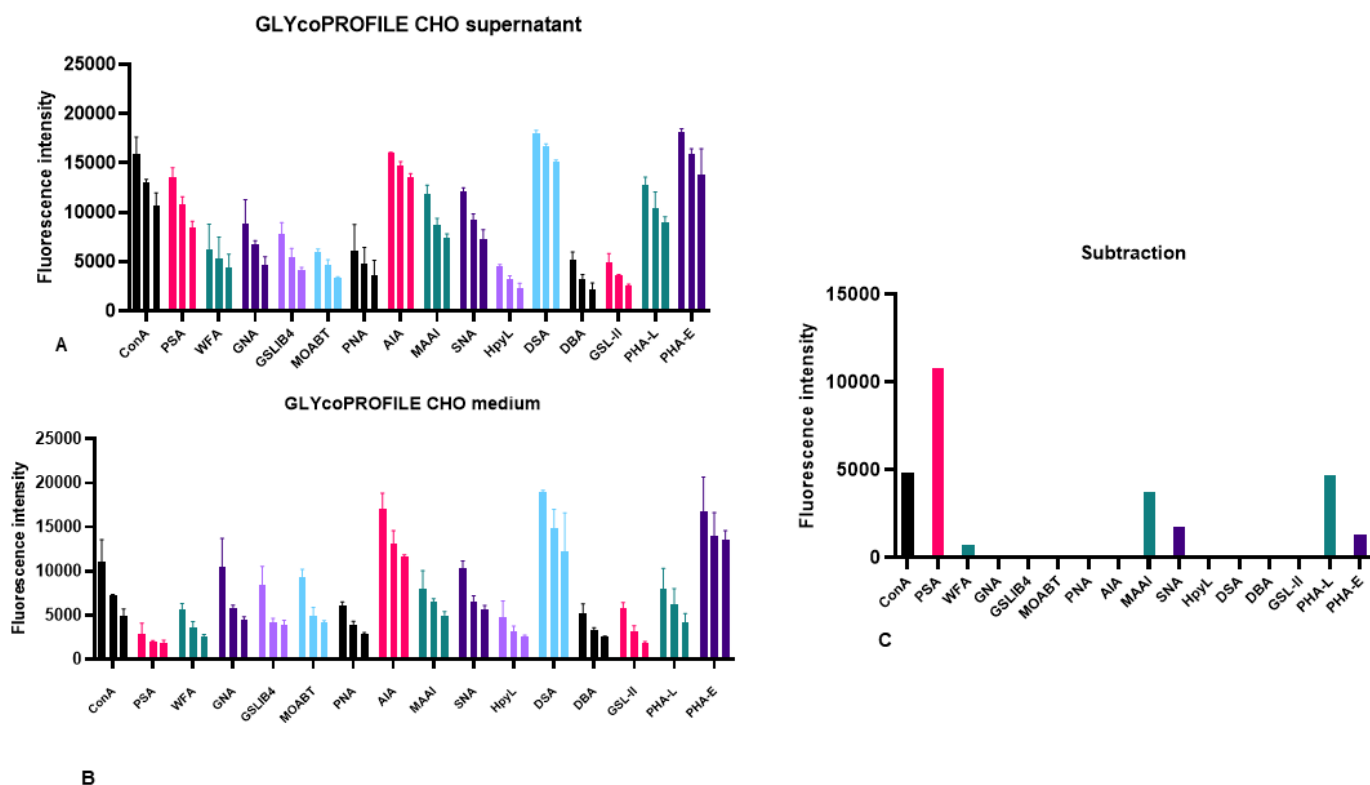


Fig. 4. 31 **A** CHO standard curve with linear regression equation. **B** GLYcoPROFILE of CHO k1 adherent cells. The data are mean values of three replicates. The wells “without lectins” correspond to the empty wells where only cells are deposited. The percentage of bound cells is calculated based on the standard curve, considering that  $2 \times 10^5$  cells/well are deposited.

The GLYcoPROFILE® of CHO supernatant was performed to see if there is an important secretion of glycoproteins containing glycan structures of interest that should be taken into consideration. To have a complete view on the cell supernatant glycosylation, the cell medium GLYcoPROFILE® has also been performed and subtracted to the cell supernatant GLYcoPROFILE® at the maximum concentration. This is done because cell medium is enriched with foetal bovine serum which contains glycoproteins like fetuin and it needs to be removed from the overall estimation of glycan structures contained on glycoproteins secreted by CHO cells. Before the GLYcoPROFILE, the proteins contained into the cells medium and supernatant are quantified through the Bradford protein assay and then, they are labelled with biotin. The fluorescent intensities of the lectins obtained from the medium GLYcoPROFILE in Fig 4.32A are subtracted from the fluorescent intensities of the lectins obtained from the supernatant GLYcoPROFILE in Fig 4.32B. The subtraction in Fig 4.32C shows an increase in Mannose/core Fucose, complex glycans and sialic acid linked in  $\alpha$ 2-3  $\alpha$ 2-6, meaning that this glycan motifs come only from the culture medium hence they are not a result of the glycoprotein secretion by the CHO cells into the supernatant.



**Fig. 4.32** **A.** GLYcoPROFILE of CHO supernatant at three different concentrations on a panel of lectins. The supernatant is labelled at a concentration of 1 mg/mL and then serially diluted to 0.25 mg/mL **B.** GLYcoPROFILE of CHO medium at three different concentrations on a panel of lectins. The medium is labelled at a concentration of 1 mg/mL and then serially diluted to 0.25 mg/mL **C** Illustrative bar chart of the subtraction of cell supernatant GLYcoPROFILE from which the values of fluorescence intensity of the cell medium GLYcoPROFILE are removed. Below 500 fluorescence units the signal is considered background noise. The subtraction was made only for the highest concentration of 1 mg/mL.

#### 4.10.2 Melanoma Cells

**B16 melanoma** is a **murine tumour cell line** used for research as a **model for human skin cancers**, especially for the study of metastasis and solid tumour formation. They were discovered and maintained in the Jackson Laboratories in Maine in 1954 when a tumour developed naturally behind the ear of a C57BL/6 mouse. Glycosylation, and in particular sialylation, plays a key role in the development and progression of cancer. Therefore, we have decided to establish a reference glycosylation profile for this cell line in order to give an additional tool to researchers in the investigation of malignant human skin cancer and a possible starting point for comparison with human melanoma cells lines and cells derived from patients. For this experiment, the panel of lectins was chosen based on previous results (AnnexII).

The results shown in *Fig. 4.33* are obtained with the same procedure described for CHO cells, with adaptation to melanoma cells. Their similarity to humans' glycosylation can be seen in the presence of both  $\alpha 2-3$  and  $\alpha 2-6$  sialyllactose, in the absence of the Galili epitope evident in the interaction with both MAAI and SNA and no interaction with MOA $\beta$ T and GSLIB4. The presence of mannose and

core fucose glycan structures is highlighted by ConA and PSA while DSA, GSLII, WGA, BC2LA indicate the presence of GlcNAc oligomers and terminal GlcNAc. PHA-L and PHA-E underline the presence of complex glycans.

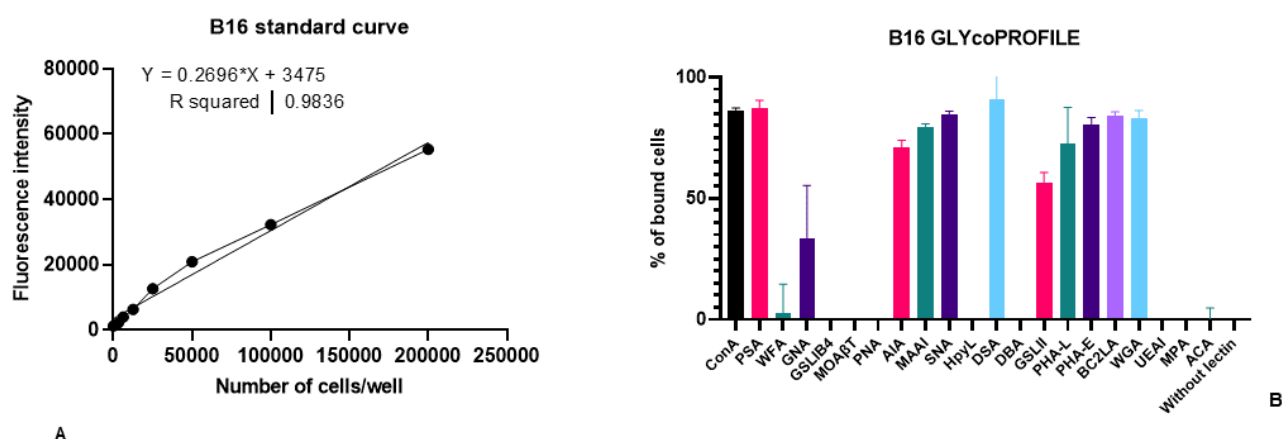


Fig. 4.33 **A** B16 standard curve with linear regression equation. **B** GLYcoPROFILE of B16 melanoma cells. The data are mean values of three replicates. The wells “without lectins” correspond to the empty wells where only cells are deposited. The percentage of bound cells is calculated based on the standard curve, considering that  $2 \times 10^5$  cells/well are deposited.

The cells GLYcoPROFILE is a proof of the fact that this methodology is very convenient, because it gives reliable results in a rapid way. The application for quality control of cell lines during the manufacturing of biotherapeutic glycoproteins could provide fast and important information in order to monitor the biotherapeutics and to keep their glycosylation in compliance with the regulation agencies' requirements.

#### 4.11 GLYcoPROFILE® of serum IgG

As seen in Paragraph 4.3.3, despite sialylation is one of the main determinants of antibodies effector function, our knowledge about the changes of Sia linkages has been hindered by the structural variability and the distribution of this carbohydrate among the world population. For this reason, we have carried out experiments on samples coming from the serum of healthy donors, and sick donors. The goal was to calculate the ratio  $\alpha 2,6:\alpha 2,3$  sialyllactose of the samples and try to identify the main differences in the sialic acid glycosidic linkages that occur in healthy and sick individuals. The serum samples 1,2 are from healthy patients while 3,4 are one from a case of malaria and the other from a case of sepsis. They were purified on Protein A to obtain the IgGs. The results of Bradford protein assay are presented in Table 4.10.

Sample	Reference	Concentration purified IgG (mg/mL)	Inflammatory patient status (based on C-reactive protein)
1	6220523799	0,8	Healthy
2	62220523780	1	Healthy
3	812853798	0.025	Inflammation
4	811501141	0.1	Inflammation

Table 4. 10 Sample 1, 2, 3, 4 obtained by EFS and purified on Protein A. The concentration of purified IgG was estimated by Bradford Protein Assay.

According to the SDS-PAGE in Fig. 4.35, the IgGs coming from there four samples after elution are relatively pure due to the presence of two main bands corresponding to the heavy and light chains, respectively at around 55 kDa and 25 kDa.

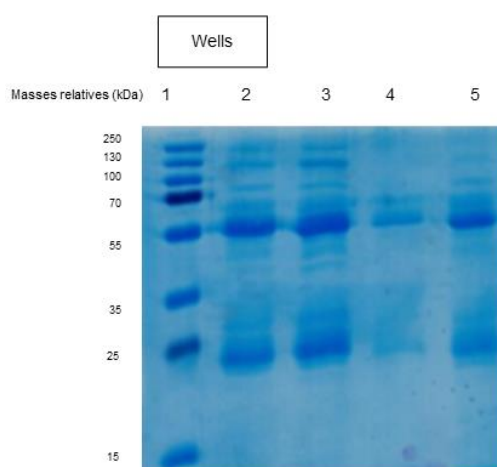
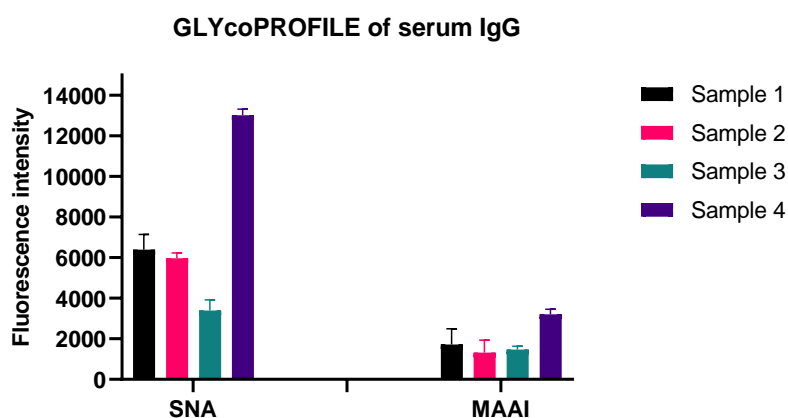


Fig. 4. 34 SDS-PAGE on 14% polyacrylamide gel of IgG after elution from affinity column with protein A. Line 1: Protein marker Line 2 sample 1 Line 2 sample 2 Line 3 sample 4 Line 5 sample 4.

In the GLYcoPROFILE in direct binding in *Fig. 4.36*, samples 1,2,3 seem to contain IgG with only  $\alpha$ 2-6 linked sialic acid, while sample 4 seem to contain a much bigger quantity of  $\alpha$ 2-6 linked sialic acid and a slightly more significant quantity of  $\alpha$ 2-3 linked sialic compared to the others.



*Fig. 4. 35* GLYcoPROFILE® of purified IgG from Sample 1, 2, 3 and 4 performed in direct binding with the lectins SNA and MAAI.

In *Table 4.11* the quantity of  $\alpha$ 2-3 sialyllactose  $\alpha$ 2-6 sialyllactose sialic acid is expressed in equivalents of neoglycoprotein, and the ratio  $\alpha$ 2-6: $\alpha$ 2-3 is calculated. Calculations are based on the standard curve (*Fig.4.21*).

Sample	Equivalent of neoglycoprotein ( $\mu$ g/mL)		Ratio $\alpha$ 2-6: $\alpha$ 2-3
	SNA	MAAI	
1	1.32	0.43	76:24
2	1.24	0.33	78:22
3	0.70	0.37	65:35
4	2.7	0.8	77:23

*Table 4. 11* Quantity of  $\alpha$ 2-3 sialic acid and  $\alpha$ 2-6 sialic acid expressed in equivalents of negolycoproteins. The calculation is made only for the highest IgG experimental concentration.

From this data it is difficult to understand whether the significant glycosylation change on the IgG of sample 4 is due to the disease, especially because we do not have data for this patient in healthy

conditions and the purified IgGs of the sample 3 and 4 are at a very low concentration (Table 4.10) that might affect the detection of IgG sialic acid linkages. However, the healthy individuals show a very similar profile, letting us think that the changes could be attributed to the disease.

## 4.12 Neu5Ac/Neu5Gc LEctPROFILE kit

### 4.12.1 HPyL Purification

The lectin from HPyV9 was expressed and purified at GLYcoDiag following the protocol described by Khan *et al.* [125]. The purification protocol for industrial scale purification was optimised at GLYcoDiag and it is property of GLYcoDiag, for this reason it will not be described in detail in this thesis. In Fig. 4.37A, the SDS-PAGE shows the pure lectin at the expected molecular weight of 17 kDa. No impurities were detected. The bar chart in Fig. 4.37 B shows the selectivity of the protein for the NeoNeuGc neoglycoprotein but not for the NeoNeuAc and NeoF underlying the selectivity of the lectin for glycolyl neuraminic acid.

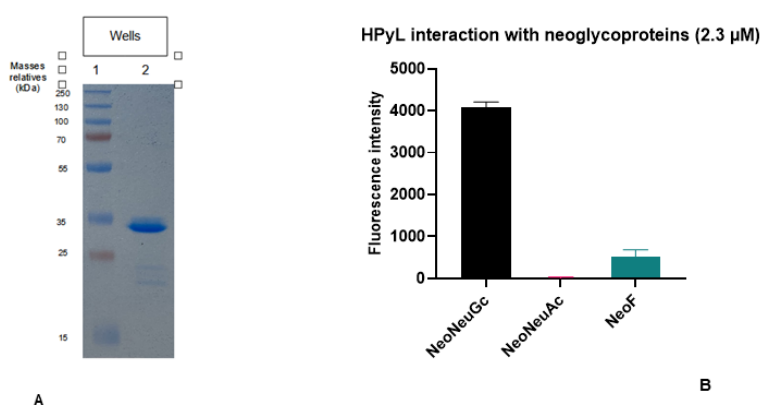


Fig. 4. 36 **A** SDS PAGE analysis of HPyL. The protein sample was analysed under denaturing conditions on 14% polyacrylamide gel. **Line 1**: Protein marker, **Line 2**: HPyL (35 kDa). **B** GLYcoPROFILE in direct binding of HPyL interacting with labelled NeNeuGc, NeoNeuAc, and NeoFucose at 2.3 μM.

### 4.12.2 WGA functionality

WGA was not produced at GLYcoDiag but it was supplied by Medicago. In Fig. 4.38 it is shown the GLYcoPROFILE in direct binding that we performed to ensure the selectivity of the lectin for NeoNeuAc. The lectin interacts with NeoNeuAc but it does not interact with NeoNeuGc or NeoF confirming the reverse specificity compared to HPyL.

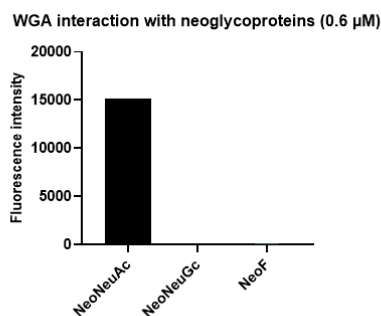


Fig. 4. 37 GLYcoPROFILE in direct binding of WGA interacting with labelled NeoNeuAc, NeoNeuGc and NeoFucose at 0.6 μM.

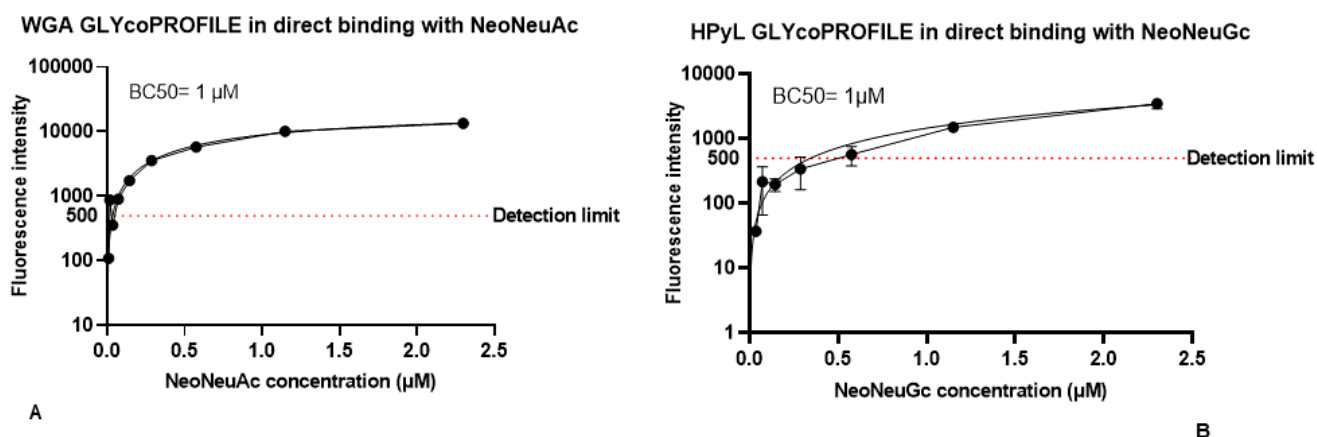
#### 4.13 Parameters of Neu5Ac/Neu5G LEctPROFILE kit

The study of this kit was finalized at the discrimination of the Neu5Ac from the Neu5Gc. The type of plate and the method of detection were immediately directed to the fluorescence detection on adsorption plate, because WGA in absorbance cross-react with the peroxidase, therefore, since the goal was to have a unique kit containing HPyL and WGA, fluorescence was chosen. Based on previous studies performed at GLYcoDiag, the quantity of lectin/well chosen for WGA and HpyL is 2.5 μg/well.

##### 4.13.1 Direct binding

The GLYcoPROFILE in direct binding was used to test the specificity of the lectins WGA and HpyL towards the corresponding neoglycoproteins and to test the sensitivity of the kit, determining the limit of detection. The neoglycoproteins NeoNeuAc and NeoNeuGc were chosen as a model probes. The preliminary results in Fig. 4.39 A and B show the direct binding curve of WGA with NeoNeuAc and HPyL with NeoNeuGc. The data obtained fit the non-linear regression model from which the BC50 is calculated, in both cases it is around 1 μM. The limit of detection was set at 500 fluorescence units. For WGA, the limit is around 0.03 μM. For HPyL the sensitivity is low, being the detection limit ten times higher than the one of WGA, around 0.3 μM.



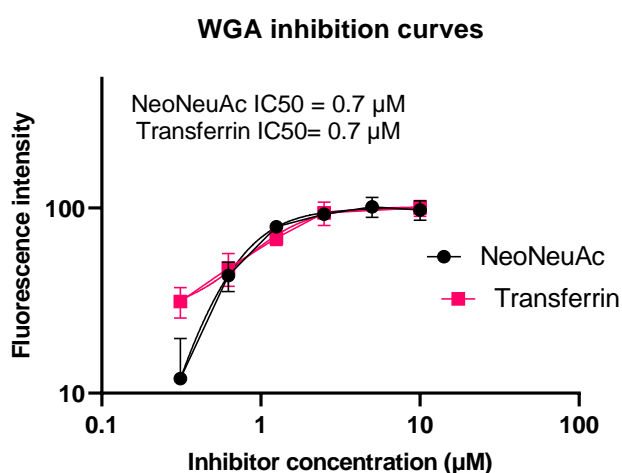


*Fig. 4. 38* **A** GLYcoPROFILE in direct binding representing the interaction between the lectin WGA and the labelled neoglycoprotein NeoNeu5Ac at a starting concentration of 2  $\mu\text{M}$ . The curve is obtained from three independent replicates. All measurements were performed in duplicates. **B** GLYcoPROFILE in direct binding representing the interaction between the lectin HPyL and the labelled neoglycoprotein NeoNeu5Gc at a starting concentration of 2  $\mu\text{M}$ . The curve is obtained from three independent replicates. All measurements were performed in duplicates

#### 4.13.2 Inhibition

The GLYcoPROFILE in inhibition was carried out only for WGA with a set of glycoproteins and neoglycoproteins for further studies on the IC<sub>50</sub>. The results of shown in *Fig. 4.40* are preliminary data of the GLYcoPROFILE in inhibition of the lectin WGA. It was carried out using as tracer the labelled neoglycoproteins NeoNeuAc at the concentration of 50  $\mu\text{g/mL}$ , and as inhibitors, NeoNeuAc and human transferrin, at a starting concentration of 10  $\mu\text{M}$ . The IC<sub>50</sub> calculated for the two inhibitors are the same, 0.7  $\mu\text{M}$ .

In the future, further experiment need to be performed to have a complete view of this kit.



*Fig. 4. 39* Inhibition curves of WGA with the inhibitors NeNeuAc and Transferrin at the starting concentration of 10  $\mu\text{M}$ . The labelled tracer NeoNeuGc is at the fixed concentration of 50  $\mu\text{g/mL}$ . All measurements were performed in duplicates.

## Conclusion

The main objective of this chapter was to build a sialic acid-targeted LEctPROFILE kits and more specifically, one for the recognition of Neu5Ac and Neu5Gc and the other for the discrimination of  $\alpha$ 2-3/ $\alpha$ 2-6 linkages. The first one mentioned only underwent preliminary studies, thus future investigations are required to establish all the required parameters and to verify the feasibility of the assay.

On the other side, the development of the  $\alpha$ 2-3/ $\alpha$ 2-6 LEctPROFILE kit was efficient and we have obtained a complete view. Depending of the mode in which the assay is performed, results can be expressed differently. In the direct binding mode, the quantity of the glycosidic  $\alpha$ 2-3 and  $\alpha$ 2-6 linkages can be expressed as equivalents of neoglycoprotein or as ratio  $\alpha$ 2-3/ $\alpha$ 2-6 whereas in the inhibition mode the results can be expressed as RIP. These characteristics together determine the value of the sample and allow to evaluate unknown compounds. The advantage of this methodology is the fact that the glycans do not need to be detached from the sample and, in inhibition, the sample doesn't even need to be labelled.

Such a tool can elucidate not only  $\alpha$ 2-3/ $\alpha$ 2-6 linkages, but including more lectins in the assay, other glycan structures can be elucidated with minimal sample requirements and without additional sample preparation steps. One example is the GLYcoPROFILE of cells, that can be applied either for the detection of glycan moieties in the in-process quality control of biotherapeutics, either in cancer research for biomarker discovery.

Moreover, we have tested IgG from human serum. It would be of great interest to further these studies and test IgG from individuals with pathologies like Rheumatoid arthritis at different stages or in order to monitor the glycosylation changes of the disease. In addition, a bigger pool of patients would be needed in order to identify a repetitive pattern. Along with this it would be also interesting to test transferrin coming from alcohol abusive patients to see the sialylation differences with healthy individuals. All this information could allow the screening of a large number of samples in a short time.

L'objectif principal de ce chapitre était de construire des kits LEctPROFILE ciblés sur les acides sialiques et plus spécifiquement, un pour la reconnaissance de Neu5Ac et Neu5Gc et l'autre pour la discrimination des liaisons  $\alpha$ 2-3/ $\alpha$ 2-6. Le premier mentionné n'a fait l'objet que d'études préliminaires, et des recherches futures sont donc nécessaires pour établir tous les paramètres requis et vérifier la faisabilité du test.

D'autre part, le développement du kit  $\alpha 2$ -3/ $\alpha 2$ -6 LEctPROFILE a été efficace et nous avons obtenu une vue complète. Selon le mode de réalisation du test, les résultats peuvent être exprimés différemment. Dans le mode de liaison directe, la quantité des liaisons glycosidiques  $\alpha 2$ -3 et  $\alpha 2$ -6 peut être exprimée en équivalents de néoglycoprotéine ou en rapport  $\alpha 2$ -3/ $\alpha 2$ -6 alors que dans le mode d'inhibition, les résultats peuvent être exprimés en RIP. Ces caractéristiques déterminent ensemble la valeur de l'échantillon et permettent d'évaluer des composés inconnus. L'avantage de cette méthodologie est le fait que les glycanes n'ont pas besoin d'être détachés de l'échantillon et, en inhibition, l'échantillon n'a même pas besoin d'être étiqueté.

Un tel outil peut élucider non seulement les liaisons  $\alpha 2$ -3/ $\alpha 2$ -6, mais en incluant plus de lectines dans le test, d'autres structures de glycanes peuvent être élucidées avec des exigences minimales en matière d'échantillons et sans étapes supplémentaires de préparation des échantillons. Un exemple est le GLYcoPROFILE de cellules, qui peut être appliqué soit pour la détection de parties de glycanes dans le contrôle de qualité en cours de fabrication de produits biothérapeutiques, soit dans la recherche sur le cancer pour la découverte de biomarqueurs.

En outre, nous avons testé des IgG provenant de sérum humain. Il serait très intéressant de poursuivre ces études et de tester les IgG de personnes atteintes de pathologies telles que la polyarthrite rhumatoïde à différents stades ou afin de surveiller les changements de glycosylation de la maladie. En outre, un plus grand nombre de patients serait nécessaire afin d'identifier un modèle répétitif. Parallèlement, il serait également intéressant de tester la transferrine provenant de patients alcooliques pour voir les différences de sialylation avec les individus sains. Toutes ces informations pourraient permettre le dépistage d'un grand nombre d'échantillons en peu de temps.

## Annex II

Short Name	Common name	Glycans structures specificity
<b>AIA</b>	<i>Artocarpus intergrifolia</i>	Gal $\alpha$ 1-6 or Gal $\beta$ 1-3GalNAc (T-antigen)>> lactose
<b>Con A</b>	<i>Canavalia ensiformis</i>	Man > Glc ; branched mannoses

<b>GNA</b>	<i>Galanthus nivalis</i>	Terminal mannoses. Mana1-3Man ; $\alpha$ 2-macroglobulin ; mannopentaose
<b>PHA E</b>	<i>Phaseolus vulgaris</i>	Gal $\beta$ 1-4GlcNAc $\beta$ 1-2Man, the bisecting GlcNAc $\beta$ 1-4Man is essential
<b>PHA L</b>	Phaseolus vulgaris	Gal $\beta$ 1-4GlcNAc $\beta$ 1-6Man of branched structures of N-glycans, Gal $\beta$ 1-4GlcNAc $\beta$ 1-2Man
<b>PNA</b>	Arachis hypogaea	Lactose, T- antigen
<b>PSA</b>	Pisum sativum	Man > Glc ; enhanced by Fuca1-6 on the core GlcNAc-Asn N-glycopeptides, IgM1A mouse
<b>SNA</b>	Sambucus nigra	Neu5Ac $\alpha$ 2-6Gal
<b>MAAI</b>	Maackia amurensis	Neu5Ac $\alpha$ 2-3Ga
<b>HPyL</b>	Human Polyomavirus 9 VP1	Neu5Gc
	Wisteria floribunda	GalNAc ; GalNAc $\alpha$ 6Gal > GalNAc $\alpha$ 3GalNAc

<b>WFA</b>		(Forssman antigen) > GalNAc
<b>GSLIB4</b>	Griffonia simplicifolia isoB4	$\alpha$ -Gal
<b>MOA<math>\beta</math>T</b>	Marasmius oreades agglutinin ( $\beta$ -Trefoil part)	Gal- $\alpha$ -1,3-Gal
<b>DSA</b>	Datura stramonium	GlcNAc $\beta$ 4GlcNAc oligomers, Galb4GlcNAc
<b>DBA</b>	Dolichos biflorus	Terminal GalNAc, GalNAca3GalNAc (Forssman), blood group A trisaccharide
<b>GSLII</b>	Griffonia simplicifolia	Terminal GlcNAc
<b>BC2LA</b>	<i>Burkholderia cenocepacia lectin A</i>	Man $\alpha$ 1-2, Man $\alpha$ 1-3, Man $\alpha$ 1-6, dimanoside
<b>WGA</b>	<i>Triticum vulgare</i>	GlcNAc; GlcNAc $\beta$ 4 oligomers, core of Asn linked oligasacchide; Neu5Ac
<b>UEAI</b>	<i>Ulex europeus</i>	Fucose

<b>MPA</b>	<i>Maclura pomifera</i>	Galb3GalNAc (T antigen), Gala6Glc (melibiose)
<b>ACA</b>	<i>Amaranthus caudatus</i>	Galb3GalNAca-O-R (T-antigen)

Table 4. 12 List of all lectins used through this thesis and corresponding specificities.

## Paper 3 - Polymers of a transition-state sialyl cation strongly Inhibit Bacterial Sialidases

### Polymers of a transition-state sialyl cation strongly inhibit bacterial sialidases.

Coralie Assailly,<sup>[a]</sup> Clarisse Bridot,<sup>[b]</sup> Amélie Saumonneau,<sup>[c]</sup> Paul Lottin,<sup>[d]</sup> Benoit Roubinet,<sup>[e]</sup> Eva-Maria Krammer,<sup>[b]</sup> Francesca François,<sup>[d]</sup> Frederica Vena,<sup>[e]</sup> Ludovic Landemarre,<sup>[e]</sup> Dimitri Alvarez Dorta,<sup>[a]</sup> David Deniaud,<sup>[a]</sup> Cyrille Grandjean,<sup>[c]</sup> Charles Tellier,<sup>[c]</sup> Sagrario Pascual,<sup>[d]</sup> Véronique Montembault,<sup>[d]</sup> Laurent Fontaine,<sup>[d]</sup> Franck Daligault,<sup>[c]</sup> Julie Bouckaert,<sup>[b]</sup> Sébastien G. Gouin<sup>\*[a]</sup>

- [a] C. Assailly, Dr D. Alvarez Dorta, Dr D. Deniaud, S.G. Gouin  
Université de Nantes, CNRS, CEISAM UMR, 6230  
F-44000, Nantes, France
- [b] Dr. C. Bridot, Dr. E.-M. Krammer, Dr. J. Bouckaert  
Unité de Glycobiologie Structurale et Fonctionnelle (UGSF), UMR8576 CNRS,  
Université de Lille, Lille 59000, France
- [c] Dr. A. Saumonneau, Dr C. Grandjean, Dr C. Tellier, Dr F. Daligault  
Université de Nantes, UFIP, UMR CNRS 6286, UFR des Sciences et des Techniques,  
2 rue de la Houssinière, BP 92208, 44322 Nantes Cedex 3, France
- [d] P. Lottin, F. François, Dr. S. Pascual, Dr. V. Montembault, Dr L. Fontaine  
Unité de Glycobiologie Structurale et Fonctionnelle (UGSF), UMR8576 CNRS,  
Université de Lille 1, Lille 59000, France
- [e] Dr. B. Roubinet, F. Vena, Dr. L. Landemarre  
Glycodiag, Bâtiment Physique-Chimie,  
Rue de Chartres, BP6759, 45067 Orléans cedex 2, France

Supporting information for this article is available.

**Abstract:** Bacterial sialidases (SA) are validated drug targets expressed by common human pathogens such as *Streptococcus pneumoniae*, *Vibrio cholerae* or *Clostridium perfringens*. Non-covalent inhibitors of bacterial SA capable of reaching the submicromolar level are rarely reported. We developed multi- and polyvalent compounds based on the transition state sialyl cation 2-deoxy-2,3-didehydro-N-acetylneuraminic (DANA). Poly-DANA inhibits the catalytic activity of SA from *S. pneumoniae* (NanA) and the symbiotic microorganism *B. thetaiotaomicron* (BtSA) at the picomolar and low nanomolar levels when expressed in moles of molecules and of DANA, respectively. Each DANA grafted to the polymer surpasses the inhibitory potential of the monovalent analogue by more than four orders of magnitude, which represents the highest multivalent effect reported so far for an enzyme inhibition. The synergistic interaction was shown to operate exclusively in the catalytic domain, offering interesting perspectives for the multivalent inhibition of other SA families lacking such as a carbohydrate-binding domain, such as viral, parasitic or human SA.

### Introduction

Sialidases (SA) are a family of exoglycosidases hydrolysing sialic acid, a negatively charged  $\alpha$ -keto monosaccharide with nine carbon backbone atoms displayed at the outermost end of glycans, glycoproteins or glycolipids. SA are widely present in animal species, and unbalanced human SA activity plays a crucial role in diabetes, inflammation, cancers and cardiovascular and lysosomal storage diseases.<sup>[1]</sup> SA are also common virulent factors of pathogenic microorganisms, enabling them to develop biofilms, to feed on host sialic acid, or to infect cells by unmasking binding ligands. SA are therefore particularly relevant targets for the development of anti-infective therapies against viruses, bacteria, fungi and parasites. Considering the rapid development of antimicrobial resistance to prescribed antibiotics or antifungals, there is an urgent need to develop therapeutic approaches

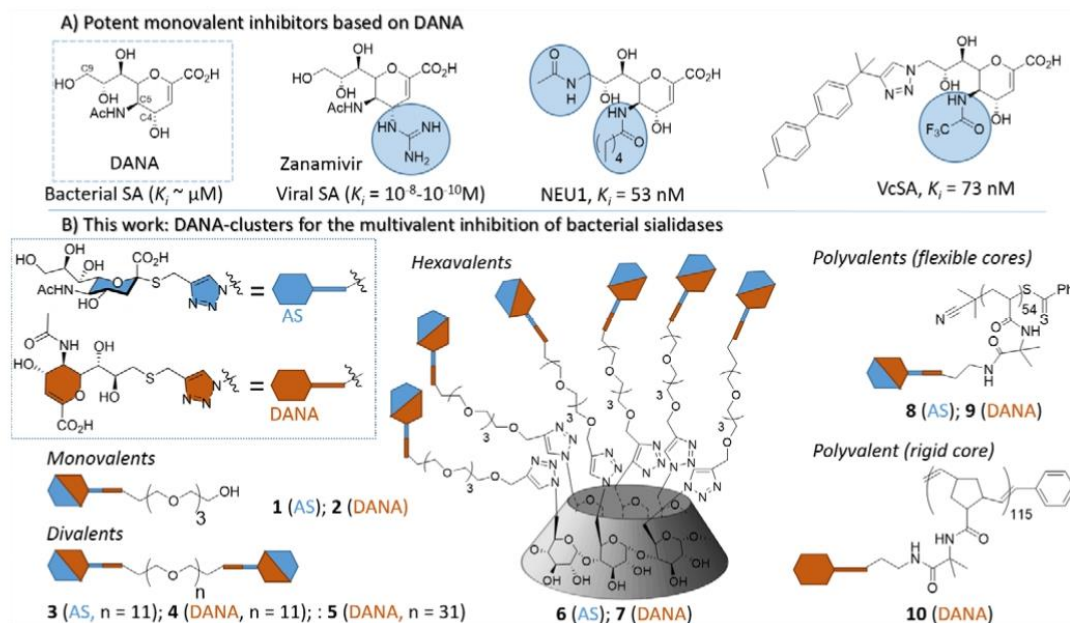
targeting such new virulent factors. Active research in this domain was rewarded by the approval of anti-influenza drugs such as oseltamivir (Tamiflu) and zanamivir (Relenza) blocking viral SA and preventing virion release and spread from the infected cells. There has been less success with bacterial, fungal or parasitic SAs, and there are few reports of inhibitory activities for bacterial SA in the nanomolar range.<sup>[2,3]</sup> The most promising inhibitors of bacterial and human SA (NEU1-4)<sup>[4]</sup> are based on 2-deoxy-2,3-didehydro-N-acetylneuraminic (DANA), a transition-state analogue of the sialyl cation after SA cleavage (Figure 1A). DANA modifications with adequate pharmacophores at the C-5 or C-9 positions were shown to improve affinity and selectivity for specific hNEU isoenzymes.<sup>[1,4-8]</sup> The group of Chen developed triazole-linked peptide in C-9 position of DANA with micromolar and selective inhibition of *V. cholerae* SA.<sup>[9]</sup> Virulent factors interfering with host sialoglycans may also be efficiently targeted or inhibited with synthetic glycodusters bearing several copies of the sialic acid epitope. These multivalent inhibitors proved particularly effective for lowering the binding ability of human adenovirus or influenza virus strains, by blocking trimeric sialic acid binding proteins at the surface of the viral capsides.<sup>[10-14]</sup> The viral neuraminidase, a homotetrameric enzyme allowing virions release from cell surfaces after sialic acid cleavages, was also efficiently inhibited by the same approach. Oligo- and poly-mers of the drug zanamivir (ZV), showed impressive *in vivo* protection of infected mice, higher than the parent ZV.<sup>[15-17]</sup> Polymeric ZV are however not suited for bacterial SA inhibition. First SA catalytic domains do not arrange in multimers at the bacterial surface, and second the ZV ligand displays only a weak binding affinity for these enzymes due to different active site architectures.<sup>[18]</sup> Bacterial SAs from the GH33 family share a six-bladed  $\beta$ -propeller catalytic domain (CAT), often flanked with a lectinic domain or carbohydrate-binding module (CBM). The role of these CBM is not clearly understood, but they were proposed to significantly improve the enzymatic catalytic efficiency for polysialylated

surfaces.<sup>[19]</sup> Inspired by these results, we recently designed non-hydrolysable polyvalent thiosialosides to bind to the CAT and CBM domains simultaneously. These sialoclusters displayed submicromolar binding inhibition capacity and inhibitory activity for the pathogenic SA from *Vibrio cholerae* (VcSA) and *Streptococcus pneumoniae* (NanA).<sup>[20]</sup> Each clustered thiosialoside showed enhanced inhibitory capacity up to three orders of magnitude compared with relevant monovalent references.

Surprisingly, these synergistic inhibitory effects were also observed for NanA catalytic domains alone (CAT), truncated from CBM. More potent CAT binder than the millimolar affinity thiosialosides may therefore provide higher levels of inhibition. These results also suggested that CAT domains of SA are sensitive to multivalent inhibitory effects as previously shown with multivalent iminosugars and specific glycosidases.<sup>[21–24]</sup> In this work, we select the DANA ligand to develop the first class of multi- and polyvalent sialyl cation transition-state inhibitors of bacterial SA (Figure 1B). The DANA-clusters inhibition potentials were assessed towards SA from *Bacterioides thetaiotaomicron* (BtSA), and *Streptococcus pneumoniae* (NanA) both belonging to the GH33 family. These enzymes were produced both in their full length (CAT+CBM) and as CAT domains alone (for NanA).

The SA from *B. thetaiotaomicron*, a symbiotic commensal microbe of the intestinal tract, has a unique CBM at the N-terminal end of the CAT that, shares low sequence similarity (6 to 10%) to other CBM present in the CAZY database. A specificity of this SA is that the CAT and CBM domains are very close to each other (17Å) compared with the same two SA domains of other pathogenic bacteria.<sup>[25]</sup> To date, little attention has been paid to the inhibition of SA from commensal bacteria despite their potential interest as an obesity-regulating target.<sup>[25]</sup>

*S. pneumoniae* remains a major cause of respiratory infections such as pneumonia, sinusitis or otitis, which can progress to forms of purulent meningitis. The *S. pneumoniae* SA NanA, is expressed by 100% of strains and is an established virulence factor contributing to infection and virulence.<sup>[26]</sup> Here, we designed synthetic DANA clusters with an unprecedented level of inhibitory capacity for GH 33 family SA enzymes. We studied the inhibitory capacity and binding interaction of the compounds on the full length BtSA and NanA and on the truncated NanA catalytic domain (NanA-CAT). Structural information on ligand binding was obtained from a NanA-CAT co-crystal structure, molecular dynamic simulation on DANA-NanA complexes and dynamic light scattering (DLS) experiments.



**Figure 1.** A) The most potent SA inhibitors previously reported were obtained by making structural modifications (shown in light blue) at C-4, C-5, and C-9 positions of DANA, a pan-inhibitor of bacterial SA with  $K_i$  in the  $\mu\text{M}$  range. These modifications were rewarded by the launch of zanamivir (Relenza®),<sup>[27]</sup> an oral drug to treat influenza, and by the identification of potent and selective inhibitors of human neuraminidase (NEU1)<sup>[31]</sup> and *Vibrio cholerae* sialidase (VcSA).<sup>[3]</sup> B) In this work we propose an alternative strategy to potentially inhibit bacterial SAs. DANA clusters were designed as multivalent transition-state inhibitors of bacterial SAs. The efficiency of the mono- and multivalent DANA derivatives **2**, **4**, **5**, **7**, **9** and **10** (in brown) was compared with mono- and multivalent analogues **1**, **3**, **6**, **8** bearing non-hydrolysable thiosialoside substrates (in blue).



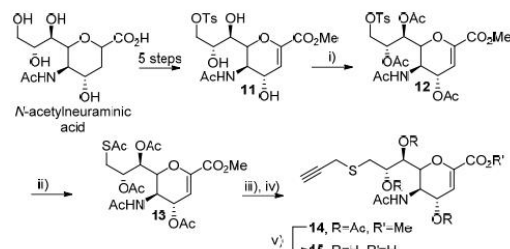
## Results and Discussion

Previously described thiosialosides **1** and **3**<sup>[20]</sup> and new compounds **6** and **8** were first designed to compare the potential of the multivalent DANA with these substrate analogues.

Three types of DANA clusters were selected to provide potential multivalent inhibitory effects on bacterial SAs (Figure 1B). These compounds bear chemical groups attached to DANA structurally identical to the monovalent reference **1**, which is necessary to calculate potential synergistic multivalent effects while excluding different binding signatures due to specific interactions in proximity to the binding site. The divalent DANA **4** and **5** with short and long spacer arm lengths, respectively, are potentially able to interact in the primary and a putative secondary binding site at the protein surface. Different levels of binding efficiency are expected depending on the spacer arm length. We previously estimated from polymer theory that the selected spacer arms for **4** and **5**, bearing 10 and 30 PEG units, attain average distances of 22 and 33 Å, respectively.<sup>[20]</sup> As previously observed with lectins, this value should match the distance between two protein binding sites for an optimal fit.<sup>[28]</sup> The  $\alpha$ -cyclodextrin (CD) scaffold of compound **7** was selected for a planar orientation of six DANA ligands. CD can be easily derivatized at the primary rim, a feature largely exploited for the design of glycoclusters.<sup>[29–31]</sup> These size-controlled DANA-clusters were completed by two different types of DANA polymers **9** and **10** obtained by RAFT polymerization and ROMP, respectively. Glycopolymers designed for lectin inhibition have already been shown to provide large affinity enhancements for their targets due to their high density sugar contents, large distance spanning, and high adaptability.<sup>[32,33]</sup> Members of this panel of compounds display unique topologies and architectures resulting from their different scaffolds. This structural diversity should cover the requirement for expected multivalent interaction through chelate, aggregative or statistical binding.

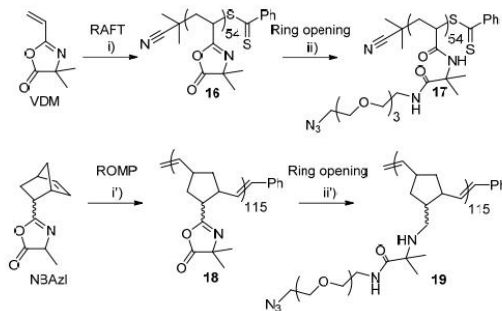
The synthesis was started by the chemical functionalization of DANA by an alkyne group, to subsequently graft the ligand on azido-armed EG, cyclodextrin or polymeric scaffolds by conventional copper-catalysed azide-alkyne cycloaddition (CuAAC). DANA was modified at C-9, a position previously shown to accept a wide range of chemical group (Figure 1A), and not buried in SA binding sites.<sup>[34]</sup>

First, we adapted previously described procedures for the five-step synthesis of **11** (supp. mat). Briefly, protected *N*-acetylneuraminic acid was chlorinated in the anomeric position with acetyl chloride,<sup>[35,36]</sup> and treated with DBU to form the intracyclic double bond. Acetate deprotection of the hydroxyl groups was followed by selective C-9 activation with *p*-toluene sulfonyl chloride in pyridine. Unstable compound **11**<sup>[37]</sup> was readily protected with acetates in pyridine to form **12**. The tosyl group was substituted with potassium thioacetate in DMF to give **13** with a moderate yield of 56%. Diverse bases were envisaged for the selective thiol deprotection followed by an *in situ* alkylation with propargyl-bromide. The best reaction yields were obtained with an equimolar mixture of sodium methyl thiolate and potassium carbonate. After flash chromatography purification, **14** was deprotected with lithium hydroxide to form the expected compound **15** quantitatively.



**Scheme 1.** Synthesis of alkynyl-DANA **10**. i)  $\text{Ac}_2\text{O}$ , pyr,  $0^\circ\text{C} \rightarrow \text{rt}$ , 1d, quant, ii) KSAc, DMF,  $60^\circ\text{C}$ , 80 min, 56%, iii) NaSMe,  $\text{K}_2\text{CO}_3$ , MeOH,  $0^\circ\text{C}$ , 30 min, then propargyl bromide,  $0^\circ\text{C} \rightarrow \text{rt}$ , 2h, workup, , evaporation, iv), pyr,  $\text{Ac}_2\text{O}$ , rt, 12h 90% over 3 steps, v) LiOH, MeOH, rt, 24h, quant.

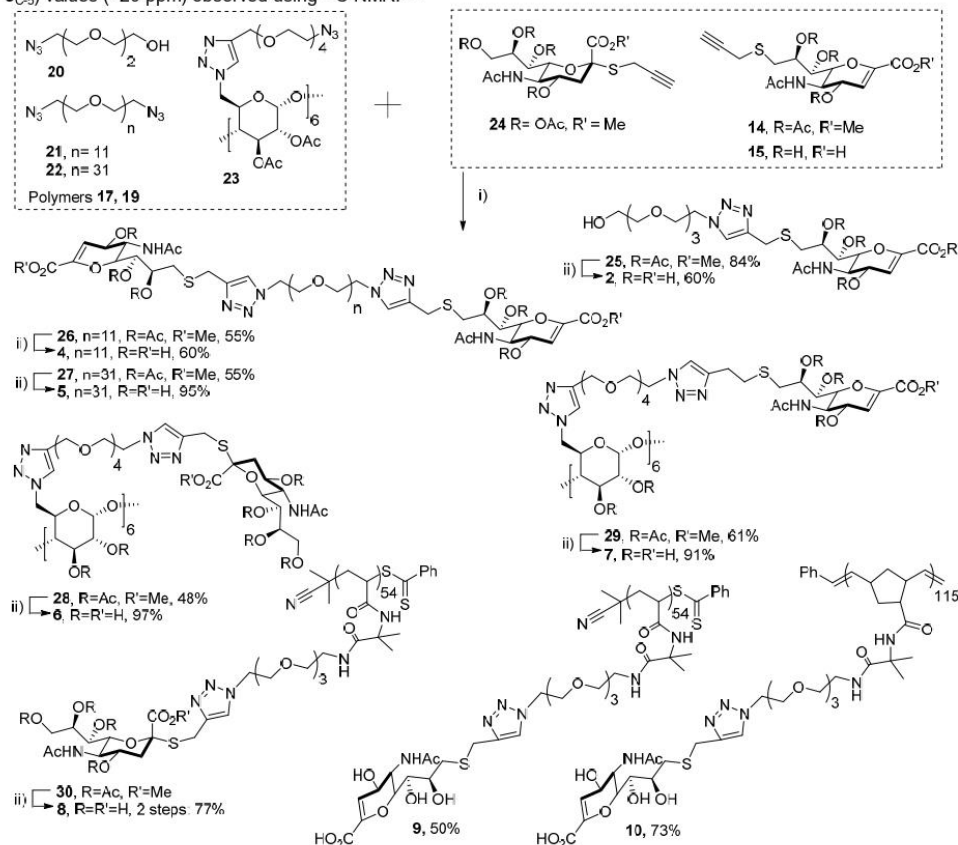
The synthesis of mono-azido-EG<sub>3</sub> **20**, di-azido-EG<sub>11</sub>, **21**, **22** and peracetylated azido-EG<sub>3</sub>-CD **23** (Scheme 3) was described in previous articles.<sup>[29,31]</sup> The two polyvalent azides **17** and **19** were obtained as depicted in Scheme 2. 2-Vinyl-4,4-dimethyl-5-oxazolone (VDM)<sup>[38,39]</sup> was polymerized by reversible addition-fragmentation chain transfer (RAFT) polymerization to form the poly(VDM)<sub>54</sub> **16** with 35% conversion (the conversion of the monomer was deliberately limited in order to access well-defined polymers<sup>[38,39]</sup>) after precipitation in hexane ( $\overline{M}_{n,\text{NMR}} = 7800 \text{ g.mol}^{-1}$ ,  $\overline{DP}_{n,\text{NMR}} = 54$ ,  $M_{n,\text{SEC}}^{\text{THF}} = 8300 \text{ g.mol}^{-1}$  and  $\overline{D} = 1.14$ ). Polymer **18** was prepared by ring-opening metathesis polymerization (ROMP) of norbornenyl azlactone (NBAzl) using ruthenium-based Grubbs' third generation catalyst leading to a poly(NBAzl)<sub>115</sub> of  $\overline{M}_{n,\text{NMR}} = 23700 \text{ g.mol}^{-1}$ ,  $\overline{DP}_{n,\text{NMR}} = 115$ ,  $M_{n,\text{SEC}}^{\text{THF}} = 25300 \text{ g.mol}^{-1}$ ,  $\overline{D} = 1.12$ .<sup>[40,41]</sup> The azlactone moieties of polymers **16** and **18** were easily reacted with one equivalent of amine **12** to form the corresponding azido-polymers **17** and **19**, respectively, as demonstrated by  $^1\text{H}$  NMR spectroscopy and SEC analysis ( $\overline{M}_{n,\text{SEC}}^{\text{THF}} = 16300 \text{ g.mol}^{-1}$ ,  $\overline{D}_{\text{SEC}} = 1.31$  and  $\overline{M}_{n,\text{SEC}}^{\text{DMF}} = 51100 \text{ g.mol}^{-1}$ ,  $\overline{D} = 1.16$  for **17** and **19**, respectively).



**Scheme 2.** Synthesis of azido-polymers **17** and **19** from VDM and NBAzl. i) 2-cyanoprop-2-ylbenzodithioate, AIBN  $60^\circ\text{C}$ , 24h, 35% ii) 11-Azido-3,6,9-trioxundecan-1-amine, THF,  $40^\circ\text{C}$ , 26h, quant.

The azido-functionalized scaffolds **20–23** and azido polymers **17**, **19** were engaged in a copper-catalysed azide alkyne cycloaddition with acetyl-protected or unprotected DANA

derivatives **14**, **15**, and thiosialoside **24**. Azide conversion was followed by  $^1\text{H}$  NMR and FT-IR and exclusive formation of the 1,4-triazole regioisomers **25–29** was demonstrated by the large  $\Delta(\delta_{\text{C-4}} - \delta_{\text{C-5}})$  values ( $>20$  ppm) observed using  $^{13}\text{C}$  NMR.<sup>[42]</sup>



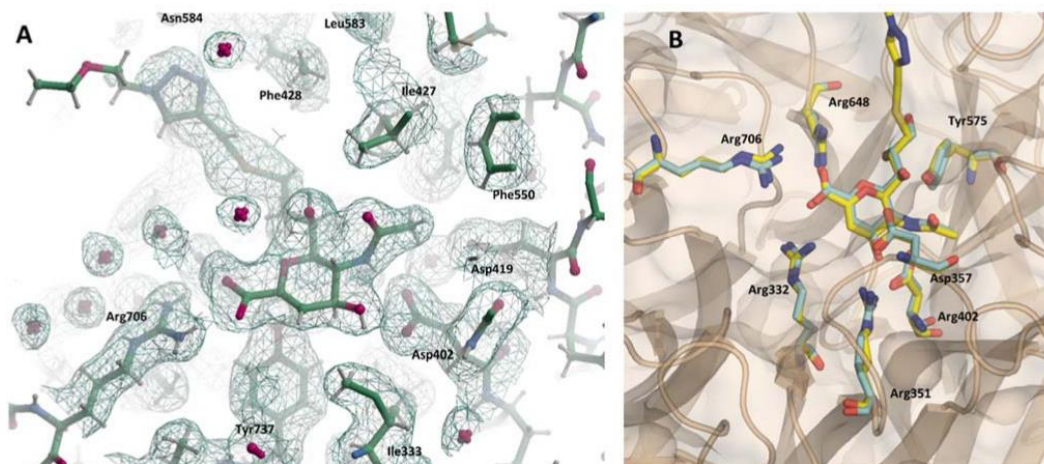
**Scheme 3.** Chemical synthesis of compounds **1–9**.

Once the chemical synthesis has been achieved, we checked that the introduction of a triazole moiety and EG linker at the C-9 position of DANA did not impair ligand-binding ability in the SA catalytic domains. The NanA-CAT was produced, purified,<sup>[20]</sup> and co-crystallized either with DANA or compound **2**, using the sitting-drop (vapor diffusion) method. Good diffraction-quality data for DANA-NanA-CAT and **2**-NanA-CAT were collected and resolved at 2.7 Å and 1.9 Å, respectively, using the previously described NanA-CAT structure from PDB entry 2YA5 as a model for molecular replacement.<sup>[18]</sup> Ligand **2** was shown to interact in the catalytic pocket of NanA-CAT and adopts a similar orientation in NanA-CAT as with DANA (Figure 2) or the influenza drug zanamivir.<sup>[18]</sup> Three arginines make four strong salt bridges (3.0 Å) with the carboxylic acid oxygens of DANA, of which Arg706 forms a bifurcated interaction to orient the inhibitor in the pocket (Figure 2A). Arg332 and Arg648 each make a single salt bridge with an oxygen of the carboxylic group, symmetrically on each side of the ligand. The second face of their guanidinium groups is charge-

compensated in salt bridges with Glu753 and Glu632, respectively.

The similar orientation of the DANA in DANA-NanA-CAT and **2**-NanA-CAT indicates that chemical modification of C9 on the glycerol tail of DANA does not impair binding (Figure 2B). A small difference is seen between these crystal structures in that a fourth arginine, Arg351, contributes a salt bridge (3.0 Å) with O4 of ligand **2**, whereas in the complex with DANA this bond remains a long-distance interaction (3.7–3.8 Å). The other changes are situated at the exit of the pocket, where Tyr680 and Asp587 change their hydrogen bond pattern and orientation, due to the modification of O9 to a sulfur atom and linkage to the triazole group.

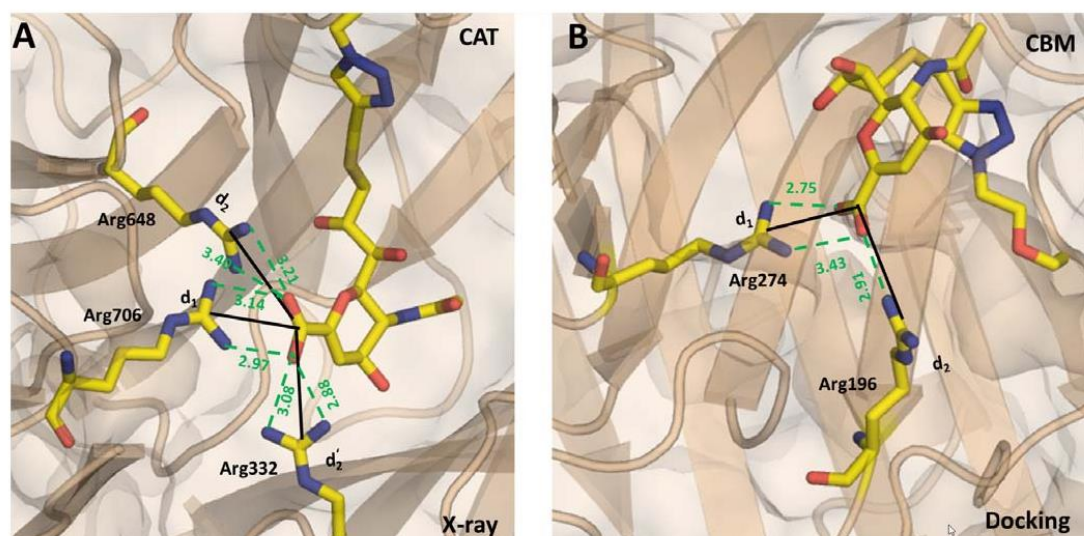
Molecular dynamics (MD) simulations were performed on the crystallized NanA-CAT complex with monovalent **2'** (**2** with one PEG linker instead of four, Figure S25) and on docking results of monovalent **2'** to NanA-CBM (Figure 3).



**Figure 2.** Binding of monovalent **2** in NanA-CAT. (A) Electron density of **2** in the binding site of the catalytic domain of NanA-CAT (PDB entry code: 7A5X). (B) Superposition, using secondary structure matching (SSM)<sup>431</sup> of monovalent **2** (yellow) and DANA (blue) bound in the NanA-CAT binding site showing a similar orientation of the functional groups and amino-acid ligands in the catalytic site. The figures have been prepared (A) using COOT<sup>44</sup> and (B) PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.)

To estimate the relative binding preferences of **2'** for the two protein domains, we used two quantities:  $N_{end}$ , which represents the number of simulations (out of four= $N_{tot}$ ) in which the compound remains bound at the end of the 30 ns simulations and  $T$ , which represents the time percentage of the simulations in which the ligand is found in the active site.

Compound **2'** binds almost exclusively to the NanA-CAT domain ( $N_{end}/N_{tot}=4/4$  and  $T=100\%$ ) and in contrast to the NanA-CBM simulations ( $N_{end}/N_{tot}=0/4$  and  $T=0.02\%$ ) does not leave the binding site.



**Figure 3.** Binding criteria used in the MD simulations. Compound **2'** was defined as bound when salt bridges with binding site arginine residues were kept. To measure that in the (A) NanA-CAT the distance with Arg706 ( $d_1$ ) had to be kept below 6 Å and in addition at least one of the two other saltbridges  $d_2$  and  $d_2'$  had to be below 6 Å at the same time; and B) for NanA-CBM two distances ( $d_1$  and  $d_2$ ) had to concurrently be below 6 Å. The distance  $d_1/d_2/d_2'$  is defined as between the carbon atom of the guanidinium arginine sidechain group of corresponding residues and the carbon atom of the C2 positioned carboxylate group of the sugar moiety of compound **2'**. The saltbridge length are shown in green and given in Å.

As the specific binding of **2** in the SA catalytic domain is established, the inhibitory capacity of compounds **1-10** was assessed against NanA-CAT, NanA and BtSA. The full procedure to produce and purify these recombinant proteins











(whole or truncated) fused with an affinity tag is described in the supplementary material. The inhibitory activity of compounds **1-10** was studied with the substrate **2'**-(4-methylumbelliferyl)- $\alpha$ -D-*N*-acetylneuraminic acid (4-MU-



NANA), releasing the fluorescent 4-methylumbelliferone ( $\lambda_{\text{ex}} = 360$  nm and  $\lambda_{\text{em}} = 455$  nm) after enzymatic sialic acid hydrolysis. Molecular concentration values required to inhibit 50% of the enzymatic activity of NanA-CAT, NanA and BtSA measured in moles of molecules ( $\text{IC}_{50}^{\text{Mol}}$ ) are listed in Table 1. As polymers **8-10** possess a large average number of DANA ligands (54 or 115), their  $\text{IC}_{50}^{\text{Mol}}$  value may be considered to give an inflated inhibitory potency result. We therefore also expressed the inhibitory capacity of polymers **8-10** in mol of DANA (valency-corrected  $\text{IC}_{50}^{\text{Mol}} \times V = \text{IC}_{50}^{\text{Vc}}$ ) instead of molecules. Monovalent compound **2** was used as a relevant reference to calculate potential synergistic inhibitory effect due to multivalency. The enhancement factor for each clustered DANA is provided here by the valency-corrected relative inhibitory potency ( $\text{RIP}^{\text{Vc}} = \text{IC}_{50}(\mathbf{2}) / V \times \text{IC}_{50}^{\text{Mol}}$ ). Monovalent DANA **2** showed significantly lower  $\text{IC}_{50}$  values (3-10 fold) for NanA, NanA-CAT and BtSA compared with the ones measured for thiosialoside **1**, (either measured in this

study, Table 1 or previously reported).<sup>[20]</sup> These results are in accordance with previously observed micromolar  $\text{IC}_{50}$  of DANA towards NanA,<sup>[45]</sup> and confirm the superiority of this transition-state inhibitor compared with sialoside substrate analogues. Compared with **2**, divalent DANA **4** and **5** showed only a moderately improved inhibitory capacity for NanA and NanA-CAT but a strong multivalent effect was observed on BtSA. Compound **5** is a strong inhibitor ( $\text{IC}_{50} = 46$  nM) for this enzyme also showing high selectivity towards NanA ( $\text{IC}_{50} = 44$   $\mu\text{M}$ ). Each of the two DANA groups of **5** are about 900-fold more potent than the monovalent reference **2**. Further increasing the DANA valency with hexavalent CD **7** led to significantly higher inhibitory potency for NanA ( $\text{IC}_{50} = 0.29$   $\mu\text{M}$ ,  $\text{RIP}^{\text{Vc}} = 59$ ) and very strong inhibition of BtSA ( $\text{IC}_{50} = 2.4$  nM,  $\text{RIP}^{\text{Vc}} = 5763$ ). When the divalent and CD scaffolds are decorated with thiosialosides (cpds **3** and **6**), much less potent analogues were obtained with  $\text{IC}_{50}$ : in the micromolar range at best.

**Table 1.** Inhibitory activity of compounds **1-10**. Molecular and valency-corrected inhibitory concentration values ( $\text{IC}_{50}^{\text{Mol}}$  and  $\text{IC}_{50}^{\text{Vc}}$ , respectively) are given in micromolar units for new compounds **2, 4-10** and previously reported compounds **1**<sup>[20]</sup> and **3**<sup>[20]</sup>. The valency-corrected relative inhibitory potencies ( $\text{RIP}^{\text{Vc}}$ ) showed the enhancement factor of each clustered DANA compared with **2**.

Cpds		NanA		NanA-CAT		BtSA	
Cpds	Valency	$\text{IC}_{50}^{\text{Mol}}$ ( $\text{IC}_{50}^{\text{Vc}}$ )	$\text{RIP}^{\text{Vc}}$	$\text{IC}_{50}$ ( $\text{IC}_{50}^{\text{Vc}}$ )	$\text{RIP}^{\text{Vc}}$	$\text{IC}_{50}$ ( $\mu\text{M}$ )	$\text{RIP}^{\text{Vc}}$
	1	1042 $\pm$ 20 <sup>[a]</sup>	0.1	426 $\pm$ 53	0.2	271 $\pm$ 9	0.3
	1	103 $\pm$ 5	1	97 $\pm$ 11	1	83 $\pm$ 11	1
	2	206 $\pm$ 23 <sup>[a]</sup>	0.25	258 $\pm$ 2	0.2	>1000	<0.1
	2	5 $\pm$ 1	10	77 $\pm$ 7	0.6	0.81 $\pm$ 0.04	51
	2	44 $\pm$ 5	1	39 $\pm$ 2	1.2	0.046 $\pm$ 0.002	902
	6	>1000	<0.1	921 $\pm$ 53	0.05	1.3 $\pm$ 0.2	11
	6	0.29 $\pm$ 0.02	59	2 $\pm$ 0.2	8	0.0024 $\pm$ 0.0002	5763
	54	>20 (>1000)	<0.1	0.199 $\pm$ 0.052 (8 $\pm$ 2)	12	6 $\pm$ 1 (327 $\pm$ 1)	<0.3
	54	0.00014 $\pm$ 8.10 <sup>-8</sup> (0.0077 $\pm$ 4.10 <sup>-7</sup> )	13 376	0.00055 $\pm$ 3.10 <sup>-8</sup> (0.030 $\pm$ 2.10 <sup>-6</sup> )	3 222	0.00048 $\pm$ 1.10 <sup>-8</sup> (0.026 $\pm$ 7.10 <sup>-7</sup> )	3192
	115	0.000032 $\pm$ 9.10 <sup>-9</sup> (0.0037 $\pm$ 1.10 <sup>-7</sup> )	27 837	0.000690 $\pm$ 8.10 <sup>-8</sup> (0.078 $\pm$ 10.10 <sup>-6</sup> )	1 243	0.000067 $\pm$ 2.10 <sup>-9</sup> (0.0077 $\pm$ 2.10 <sup>-7</sup> )	10 779

[a] Values previously described values in ref <sup>[20]</sup>

The flexible and rigid DANA polymers **9** and **10** proved highly potent with unprecedented levels of inhibitory capacity for the targets. The polymers reached picomolar and low nanomolar levels in the inhibition for NanA, NanA-CAT when expressed in moles of molecules or DANA groups, respectively. NanA-CAT was more potently inhibited by the flexible poly-DANA **9** ( $\text{IC}_{50}^{\text{Mol}} = 550$  pM,  $\text{IC}_{50}^{\text{Vc}} = 30$  nM) while lower inhibitory values were obtained with poly-DANA **10** and NanA ( $\text{IC}_{50}^{\text{Mol}} = 32$  pM,  $\text{IC}_{50}^{\text{Vc}} = 3.7$  nM) or BtSA ( $\text{IC}_{50}^{\text{Mol}} = 67$  pM,  $\text{IC}_{50}^{\text{Vc}} = 7.7$  nM). For these two polymers, each DANA ligand surpassed the inhibitory capacity of the monovalent reference **2** by three ( $\text{RIP}^{\text{Vc}}$  **10**-BtSA = 1243, Table 1) to more than four orders of magnitudes

( $\text{RIP}^{\text{Vc}}$  **10**-NanA = 27 837). We also determined the inhibitory constants ( $K_i$ ) and valency-corrected inhibitory constants ( $K_i^{\text{Vc}}$ ) for the DANA ligands attached to the two best inhibitors for NanA and BtSA (**9**, **10** and **7**, **10**, respectively). As the compounds are highly potent inhibitors, DANA concentration used during the assay was not significantly higher than [SA] and we had to apply the Morrison equation for tight-binding inhibitors.<sup>[46,47]</sup> Indeed, equations of the Henri-Michaelis-Menten type are only valid if the enzyme concentration is much lower than the inhibitor concentration. We also assumed a competitive binding mode for the multivalent DANA as suggested from crystallography and MD simulations showing

DANA interaction in the catalytic binding pocket. The low  $K_i^{VC}$  values measured for the best inhibitors **9** ( $K_i^{VC} = 17.4 \pm 1.7$  nM) and **10** ( $K_i^{VC} = 16.2 \pm 1.9$  nM) for NanA and, **7** ( $K_i = 0.56 \pm 0.23$  nM) and **10** ( $K_i^{VC} = 10.7 \pm 1.6$  nM) for BtSA, confirmed the strong potency and synergistic effects offered by these multivalent-DANA.

Next, we evaluated whether the strong inhibition of the enzymatic activity provided by the multivalent DANA was correlated with a strong binding affinity for the targets. More precisely, we assessed whether the compounds could efficiently prevent NanA and NanA-CAT binding to a labelled multi-sialylated protein mimicking natural sialyl glycans at cell surfaces. BtSA was not evaluated in this assay due to its much lower binding affinity for the labelled protein. NanA constructs were coated onto a microplate surface and labelled bovine serum albumin (BSA) was coated with non-hydrolysable thiosialosides.<sup>[20]</sup> This system, adapted from "lectin profiling" technology, allowed determination of the half maximal inhibitory concentrations of binding values  $IC_{50}$  and to calculate the valency-corrected relative inhibitory potency of binding  $RIPb^{VC}$  (Table 2).

The DANA derivatives showed  $IC_{50}^{Mol}$  one to two-order of magnitude lower against NanA-CAT than against NanA, meaning that the compounds are more efficient in blocking the catalytic site alone compared to the full-length enzyme bearing the CBM (sialoside-binding domain). NanA-CBM display a higher intrinsic affinity for sialylated surfaces than do NanA-CAT<sup>[20]</sup> and the lower binding capacity of the compounds for NanA compared with NanA-CAT, suggest that the DANA ligands interact in the CAT domain only, as observed in our structural and simulation data.

The binding inhibition potential observed with compounds **1-10** shows significant homologies with their inhibitory capacity (except for **7** on NanA) because (i) the DANA derivatives are significantly more potent than their sialoside derivatives, (ii) the low-valency compounds **2-7** show a moderate level of synergistic effects, if any, and (iii) the DANA polymers **9-10** are highly potent inhibitors of NanA binding. These compounds displayed picomolar  $IC_{50}^{Mol}$  and low nanomolar  $IC_{50}^{VC}$  inhibition level. **10** was the most potent polymer with  $IC_{50}^{Mol} = 31$  pM and  $IC_{50}^{VC} = 3.6$  nM for NanA-CAT. **9-10** also display outstanding multivalent effects for NanA with  $RIPb^{VC}$  values of 17242 and 38461 (Table 2), respectively. Thus, the DANA polymers combine an outstanding potential for blocking the enzymatic hydrolysis activities and binding capacities of NanA. Both effects are also observed on the truncated domain NanA-CAT, showing that DANA binding occurs at catalytic domains as suggested by X-ray structural data and molecular dynamic simulations (Figures 2, 3).











Dynamic light scattering (DLS) experiments were performed by means of titration to polymer **10**, carrying 115 DANA molecules. DLS analysis showed that upon titration of the sialidase proteins to polymer **10**, the hydrodynamic radius ( $R_h$ ) for the protein peak doubles in size for those BtSA and NanA proteins (Figures S17-S24). Thus, the multivalent DANA **10** has the potency to cluster SA catalytic domains in a similar fashion to that previously reported between GalNAc binding lectins and a submaxillary mucin polymer bearing ~2300  $\alpha$ -GalNAc residues. The lectins displayed a much higher affinity for the polymeric GalNAc, explained by a reduced off-rate of the complex due to a lectin diffusion mechanism along the

GalNAc chain.<sup>[48]</sup> The high synergistic effects observed here with multivalents **9** and **10** are likely due to a similar "bind and jump" mechanism where the catalytic domain diffuses from one DANA ligand to another.

## Conclusion

In conclusion, we developed polymers of the sialyl cation transition-state analogue DANA with unprecedented level of inhibition for a bacterial SA of the GH 33 family. The DANA clusters provided valency-corrected multivalent effects exceeding four orders of magnitude on NanA or BtSA. As far as we know, such a level of synergistic inhibition has never been achieved with previously described multivalent enzyme inhibitors. The improved inhibition and affinity being also observed on the truncated catalytic domain NanA-CAT alone, we forecast that this strategy may also be successful in blocking other classes of biologically relevant bacterial, viral or human SA lacking a CBM domain.

**Table 2.** Binding inhibition of compounds **1-10** in  $\mu$ M.

Cpds	Val	NanA		NanA-CAT	
		$IC_{50}^{Mol}$ ( $IC_{50}^{VC}$ )	$RIPb^{VC}$	$IC_{50}^{Mol}$ ( $IC_{50}^{VC}$ )	$RIPb^{VC}$
	1	8420 $\pm$ 100 <sup>[a]</sup>	0.06	— <sup>[b]</sup>	—
	1	>500	1	2.4 $\pm$ 0.8	1
	2	880 $\pm$ 200 <sup>[a]</sup>	<0.28	— <sup>[b]</sup>	—
	2	160 $\pm$ 20	>1.5	0.40 $\pm$ 0.05	3
	2	310 $\pm$ 50	>0.8	0.20 $\pm$ 0.02	6
	6	60 $\pm$ 5	>1.4	4.5 $\pm$ 2.0	0.09
	6	>250	>0.3	0.9 $\pm$ 0.1	0.4
	54	0.055 (3.0 $\pm$ 1.5)	>166	0.014 (0.8 $\pm$ 0.1)	3
	54	0.00054 (0.029 $\pm$ 0.004)	>17242	0.000068 (0.0037 $\pm$ 0.0013)	648
	115	0.00011 (0.013 $\pm$ 0.002)	>38461	0.000031 (0.0036 $\pm$ 0.0012)	667

## Acknowledgements

This work was carried out with financial support from the *Centre National de la Recherche Scientifique (CNRS)*, the *Ministère de l'Enseignement Supérieur et de la Recherche* in France and the National Agency for Research (ANR project HICARE 17-CE07-028-01). We acknowledge SOLEIL for providing synchrotron radiation facilities and would like to thank Leonard Chavas for his assistance in using beamline Proxima1.

**Keywords:** Sialidases • Multivalency • Inhibitors • Enzymes •

- [1] T. Guo, R. Héon-Roberts, C. Zou, R. Zheng, A. V. Pshezhetsky, C. W. Cairo, *J. Med. Chem.* **2018**, *61*, 11261–11279.
- [2] W. Li, A. Santra, H. Yu, T. J. Slack, M. M. Muthana, D. Shi, Y. Liu, X. Chen, *J. Org. Chem.* **2019**, *84*, 6697–6708.
- [3] Hinou Hiroshi, Miyoshi Risho, Takasu Yasuaki, Kai Hirokazu, Kuroguchi Masaki, Arioka Shingo, Gao Xiao-Dong, Miura Nobuaki, Fujitani Naoki, Omoto Shinya, Yoshinaga Tomokazu, Fujiwara Tamio, Noshi Takeshi, Togame Hiroko, Takemoto Hiroshi, Nishimura Shin-Ichiro, *Chem. – Asian J.* **2011**, *6*, 1048–1056.
- [4] T. Guo, P. Dätwyler, E. Demina, M. R. Richards, P. Ge, C. Zou, R. Zheng, A. Fougerat, A. V. Pshezhetsky, B. Ernst, C. W. Cairo, *J. Med. Chem.* **2018**, *61*, 1990–2008.
- [5] C. W. Cairo, *MedChemComm* **2014**, *5*, 1067–1074.
- [6] A. Albohy, Y. Zhang, V. Smutova, A. V. Pshezhetsky, C. W. Cairo, *ACS Med. Chem. Lett.* **2013**, *4*, 532–537.
- [7] C. D. Hunter, N. Khanna, M. R. Richards, R. Rezaei Darestani, C. Zou, J. S. Klassen, C. W. Cairo, *ACS Chem. Biol.* **2018**, *13*, 922–932.
- [8] S. Magesh, S. Moriya, T. Suzuki, T. Miyagi, H. Ishida, M. Kiso, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 532–537.
- [9] T. J. Slack, W. Li, D. Shi, J. B. McArthur, G. Zhao, Y. Li, A. Xiao, Z. Khedri, H. Yu, Y. Liu, X. Chen, *Med. Chem.* **2018**, *26*, 5751–5757.
- [10] P. Kiran, S. Bhatia, D. Lauster, S. Aleksic, C. Fleck, N. Peric, W. Maison, S. Liese, B. G. Keller, A. Herrmann, R. Haag, *Chem. – Eur. J.* **2018**, *24*, 19373–19385.
- [11] S. Spjut, W. Qian, J. Bauer, R. Storm, L. Frängsmyr, T. Stehle, N. Arnberg, M. Elofsson, *Angew. Chem. Int. Ed.* **2011**, *50*, 6519–6521.
- [12] R. Caraballo, M. Saleeb, J. Bauer, A. M. Liaci, N. Chandra, R. J. Storm, L. Frängsmyr, W. Qian, T. Stehle, N. Arnberg, M. Elofsson, *Org. Biomol. Chem.* **2015**, *13*, 9194–9205.
- [13] D. Lauster, S. Klenk, K. Ludwig, S. Nojumi, S. Behren, L. Adam, M. Stadtmüller, S. Saenger, S. Zimmer, K. Hönzke, L. Yao, U. Hoffmann, M. Bardua, A. Hamann, M. Witzentrath, L. E. Sander, T. Wolff, A. C. Hocke, S. Hippenstiel, S. De Carlo, J. Neudecker, K. Osterrieder, N. Budisa, R. R. Netz, C. Böttcher, S. Liese, A. Herrmann, C. P. R. Hackenberger, *Nat. Nanotechnol.* **2020**, DOI 10.1038/s41565-020-0660-2.
- [14] M. Mammen, G. Dahmann, G. M. Whitesides, *J. Med. Chem.* **1995**, *38*, 4179–4190.
- [15] L. Fu, Y. Bi, Y. Wu, S. Zhang, J. Qi, Y. Li, X. Lu, Z. Zhang, X. Lv, J. Yan, G. F. Gao, X. Li, *J. Med. Chem.* **2016**, *59*, 6303–6312.
- [16] A. K. Weight, J. Haldar, L. Álvarez de Cienfuegos, L. V. Gubareva, T. M. Tumpey, J. Chen, A. M. Klibanov, *J. Pharm. Sci.* **2011**, *100*, 831–835.
- [17] T. Honda, S. Yoshida, M. Arai, T. Masuda, M. Yamashita, *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1929–1932.
- [18] H. Gut, G. Xu, G. L. Taylor, M. A. Walsh, *J. Mol. Biol.* **2011**, *409*, 496–503.
- [19] S. Thobhani, B. Ember, A. Siriwardena, G.-J. Boons, *J. Am. Chem. Soc.* **2003**, *125*, 7154–7155.
- [20] Y. Brissonnet, C. Assailly, A. Saumonneau, J. Bouckaert, M. Maillason, C. Petitot, B. Roubinet, B. Didak, L. Landemarre, C. Bridot, R. Blossey, D. Deniaud, X. Yan, J. Bernard, C. Tellier, C. Grandjean, F. Daligault, S. G. Gouin, *Chem. – Eur. J.* **2019**, *25*, 2358–2365.
- [21] S. G. Gouin, *Chem. – Eur. J.* **2014**, *20*, 11616–11628.
- [22] P. Compain, A. Bodlemer, *ChemBioChem* **2014**, *15*, 1239–1251.
- [23] N. Kanfar, E. Bartolami, R. Zelli, A. Marra, J.-Y. Winum, S. Ulrich, P. Dumy, *Org. Biomol. Chem.* **2015**, DOI 10.1039/C5OB01405K.
- [24] C. O. Mellet, J.-F. Nierengarten, J. M. G. Fernández, *J. Mater. Chem. B* **2017**, DOI 10.1039/C7TB00860K.
- [25] K.-H. Park, M.-G. Kim, H.-J. Ahn, D.-H. Lee, J.-H. Kim, Y.-W. Kim, E.-J. Woo, *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2013**, *1834*, 1510–1519.
- [26] D. Parker, G. Soong, P. Planet, J. Brower, A. J. Ratner, A. Prince, *Infect. Immun.* **2009**, *77*, 3722–3730.
- [27] M. von Itzstein, W.-Y. Wu, G. B. Kok, M. S. Pegg, J. C. Dyason, B. Jin, T. Van Phan, M. L. Smythe, H. F. White, S. W. Oliver, P. M. Colman, J. N. Varghese, D. M. Ryan, J. M. Woods, R. C. Bethell, V. J. Hotham, J. M. Cameron, C. R. Penn, *Nature* **1993**, *363*, 418–423.
- [28] V. Wittmann, R. J. Pieters, *Chem. Soc. Rev.* **2013**, *42*, 4492–4503.
- [29] J. Bouckaert, Z. Li, C. Xavier, M. Almant, V. Caveliers, T. Lahoutte, S. D. Weeks, J. Kovensky, S. G. Gouin, *Chem. – Eur. J.* **2013**, *19*, 7847–7855.
- [30] C. Decroocq, D. Rodriguez-Lucena, V. Russo, T. Mena Barragán, C. Ortiz Mellet, P. Compain, *Chem. – Eur. J.* **2011**, *17*, 13825–13831.
- [31] V. Lehot, Y. Brissonnet, C. Dussouy, S. Brument, A. Cabanettes, E. Gillon, D. Deniaud, A. Varrot, P. Le Pape, S. G. Gouin, *Chem. – Eur. J.* **2018**, *24*, 19243–19249.
- [32] C. R. Becer, M. I. Gibson, J. Geng, R. Ilyas, R. Wallis, D. A. Mitchell, D. M. Haddleton, *J. Am. Chem. Soc.* **2012**, *132*, 15130–15132.
- [33] S. Brument, C. Cheneau, Y. Brissonnet, D. Deniaud, F. Halary, S. G. Gouin, *Org. Biomol. Chem.* **2017**, *15*, 7660–7671.
- [34] Y.-S. Hsiao, D. Parker, A. J. Ratner, A. Prince, L. Tong, *Biochem. Biophys. Res. Commun.* **2009**, *380*, 467–471.
- [35] R. Kuhn, P. Lutz, D. L. Macdonald, *Chem. Ber.* **1966**, *99*, 611–617.
- [36] S. Wolf, S. Warnecke, J. Ehrst, F. Freiburger, R. Gerardy-Schahn, C. Meier, *ChemBioChem* **n.d.**, *13*, 2605–2615.
- [37] T. G. Warner, *Biochem. Biophys. Res. Commun.* **1987**, *148*, 1323–1329.
- [38] M. E. Levere, H. T. Ho, S. Pascual, L. Fontaine, *Polym. Chem.* **2011**, *2*, 2878–2887.
- [39] S. Pascual, T. Blin, P. J. Saikia, M. Thomas, P. Gosselin, L. Fontaine, *J. Polym. Sci. Part Polym. Chem.* **2010**, *48*, 5053–5062.
- [40] V. Lapinte, J.-C. Brosse, L. Fontaine, *Macromol. Chem. Phys.* **2004**, *205*, 824–833.
- [41] V. Lapinte, L. Fontaine, V. Montembault, I. Campistron, D. Reyx, *J. Mol. Catal. Chem.* **2002**, *190*, 117–129.
- [42] N. A. Rodios, *J. Heterocycl. Chem.* **1984**, *21*, 1169–1173.
- [43] E. Krissinel, K. Henrick, *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2256–2268.
- [44] P. Emsley, K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126–2132.
- [45] Z. Xu, S. von Grafenstein, E. Walther, J. E. Fuchs, K. R. Liedl, A. Sauerbrei, M. Schmidtke, *Sci. Rep.* **2016**, *6*, DOI 10.1038/srep25169.
- [46] D. J. Murphy, *Anal. Biochem.* **2004**, *327*, 61–67.
- [47] R. A. Copeland, *Enzymes*, John Wiley & Sons, Inc., New York, USA, **2000**.
- [48] T. K. Dam, T. A. Gerken, B. S. Cavada, K. S. Nascimento, T. R. Moura, C. F. Brewer, *J. Biol. Chem.* **2007**, *282*, 28256–28263.

## **Chapter 5 – Materials and methods**



## 5. Materials and methods

The ultrapure water, purified through the Aqua Service Distribution (ASD) water purification system, was used for all solutions and buffer preparation. The buffer were filtered at 0.2  $\mu\text{m}$  (Millipore filters) using a filtration system and those used for mammalian cells culture were sterilised by autoclaving (autoclave Vapour line, VWR) prior to use. When not otherwise specified, PBS (Phosphate buffer saline) pH 7.4 was supplemented with cations  $\text{CaCl}_2$  1M, and  $\text{MgCl}_2$  0.5 M.

### 5.1 Neoglycoprotein synthesis

3-(2-Propyn-1-yloxy)propanoic acid 2,5-dioxo-1-pyrrolidiny ester (Propargyl-NHS ester), (2,5-Dioxopyrrolidin-1-yl) 2-azidoacetate, 1-[(Azidoacetyl)oxy]pyrrolidine-2,5-dione (Azido-NHS ester), L-ascorbic acid,  $\text{CuSO}_4$ , and 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) were obtained from Sigma Aldrich. Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA) was obtained from TCI. Oligosaccharides 6'-sialylactose- $\text{N}_3$ , 3'-sialylactose- $\text{N}_3$ , propargylNeu5Ac and propargylNeu5Gc were obtained from Professor Olivier Renaudet laboratory, University of Grenoble-Alpes and Gal $\alpha$ 1-3Gal-N-Acetyl-Propargyl was obtained from Elicityl. Bovine serum albumin (BSA) was purchased from H2B/Seqens. Sephadex used for the purification of neoglycoproteins was obtained from Cytiva.

#### 5.1.1 Mass spectrometry

(MALDI-TOF/MS) analysis was recorded on an UltrafleXtreme from Bruker by The Facility for Mass Spectrometry and Proteomics at the Centre for Molecular Biophysics (CBM) in Orléans.

#### 5.1.2 Functionalization of BSA with an alkyne or azido group

A solution of Propargyl-NHS ester or Azido-NHS ester (60 equivalents) was added to a solution of BSA in PBS (50 mM, pH 8). The reaction mixture was stirred for 5 h at room temperature. BSA-alkyne or BSA-azide was purified on a Sephadex G25 and its purity was controlled by SDS-PAGE. The protein concentration was determined by Bradford. The number of alkyne or azido group per BSA was determined by MALDI-TOF/MS.

#### 5.1.3 Synthesis of neoglycoprotein by click chemistry

Functionalized oligosaccharide (60 equivalents) was added to a solution of BSA-alkyne or BSA-azide (5 mg/mL) in PBS 10 mM pH 7.4. Then, TBTA (5 mg/mL in DMSO, 35 equivalents), L-ascorbic acid (3 mg/mL in PBS 10 mM, pH 7.4, 21 equivalents) and  $\text{CuSO}_4$  (3 mg/mL in PBS 10 mM, pH 7.4, 10.5 equivalents) were successively added to the reaction mixture. The solution was stirred 24 h at room temperature then the crude mixture was purified by size-exclusion chromatography on a Sephadex G25 gel. The purity of each neoglycoproteins was controlled by Bradford, SDS-PAGE and the number of glycans/BSA was determined by MALDI-TOF/MS. The



functionality of the neoglycoproteins was checked by GLYcoPROFILE in direct binding after labelling with biotin.

### 5.2. Biotinylation of neoglycoproteins

The labelling of NeoGPs was carried out by amide coupling with the EZ-link sulfo-NHS-LC-biotin (10 equivalents) in PBS 10 mM at pH 7.4 for 30 min at room temperature. The excess of biotin reagent was removed by dialysis process. The number of biotin coupled on the neoglycoprotein was determined according to the standard HABA (4'-hydroxyazobenzene-2-carboxylic acid)/ avidin quantification method. This value is related to the biotin amount of the sample. The same protocol was applied to the biotinylation of cells culture medium and cell culture supernatant.

### 5.3 Determination of protein concentration with Bradford Protein Assay

The protein concentration is determined by Bradford Protein Assay. This test is based on the on a colour change of Coomassie Blue reagent (from BioRad) when it reacts with proteins. It gives a blue colour, which has a maximum of absorbance between 465 and 595 nm. The colour intensity is directly proportional to the protein concentration. Briefly, the test was performed in a flat bottom microtiter plate. The samples were diluted in PBS directly on the microplates in serial dilutions. The standard curve was prepared in triplicate, using a BSA solution prepared at 100 µg/ml in PBS. 50 µL of Coomassie Blue reagent is added to the plate, immediately followed with measurement of absorbance at 595 nm with a microplate reader. The concentration of sample was calculated by reporting the absorbance value from the linear equation of the standard curve.

### 5.4 Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE electrophoresis)

Proteins were controlled for their purity by SDS-PAGE electrophoresis using stacking gel (4%, w/v) and running gel (14%, w/v) in Tris/glycine buffer. The samples were diluted (4:1) in Laemmli sample buffer, boiled at 96 °C for 5 min and loaded into the wells. Electrophoresis was conducted using Bio-Rad Mini-PROTEAN® Electrophoresis System at constant voltage (200 V). All samples sizes were compared to Thermo Scientific PageRuler Plus Pre-stained Protein Ladder. Protein were fixed and visualized on the gel by incubating for approximately 30 min in a staining solution of Coomassie Brilliant Blue G250 (Sigma). The staining was removed from the gel by heating in water until bands appeared visible or after staining with a MeOH/AcOH solution for 1 h. In order to reveal the glycoproteins present in a sample, the glycoproteins band can be revealed by using the Schiff reagent PAS (Periodic Acid Schiff stain, Sigma) protocol. Scan records of the gels were taken.

### 5.5 GLYcoPROFILE®

The covalent and adsorption 96-microtiter plates for GLYcoPROFILEs study were purchased from Biomat. Streptavidin-DTAF and extravidine-peroxydase solution, the revealing solution used in GLYcoPROFILEs were respectively obtained from Jackson Immuno Research and Sigma Aldrich. The OPD tablet was purchased from Sigma Aldrich. All absorbance and fluorescence measurements

were recorded on a FluoStar Omega (BMG LabTech). The full recombinant MOA was purchased from Sobioda, French distributor for the Japanese lectins Fujifilm WAKO. GSLI-B<sub>4</sub> was purchased from Vector. WGA was purchased from Medicago.

GLYcoDiag's GLYcoPROFILE is a lectin-array based technology. Lectins were immobilized overnight at 4°C on a 96-well microtiter plate. The quantity of each lectin was determined empirically in order to get the optimum binding. For covalent plates, surfaces are functionalized with carboxylic groups that can react covalently through amidification with the free amino groups of lectins using the EDC activating reagent. After lectins incubation overnight, the plates were washed with PBS/Tween 20 0,05% and saturated with PBS/BSA 0.5%.

GLYcoPROFILE in direct binding: after 30 minutes incubation time with PBS/BSA 0.5%, the plates were washed with PBS. Then, the labelled glycoprotein/neoglycoprotein or the monoclonal antibodies were applied to the plate in serial dilutions and they were left for 30 minutes at room temperature. After this 30 min of incubation, the plates were washed with PBS, and streptavidin-DTAF for fluorescent plates or ExtrAvidin-peroxidase for absorbance plate were added. They were incubated 30 minutes at room temperature, sheltered from light. In the case of monoclonal antibodies, before adding the streptavidin-DTAF, a biotinylated secondary antibody was added to the plate and it was incubated for 30 minutes at room temperature. For the fluorescent plates, after washing to remove the excess of streptavidin-DTAF solution, PBS was added in the plates and the fluorescence was read at  $\lambda_{exc} = 485\text{nm}$  and  $\lambda_{em} = 520\text{nm}$  with the help of a fluorescent reader. The fluorescence intensity is directly correlated to the concentration of biotinylated neoglycoprotein that interacts with the lectin. The fluorescence background of the lectin was corrected by removing the fluorescence intensity obtained of the lectin with only PBS and without any glycoconjugate. Moreover, the data were corrected by the gain of the plate measured by the plate reader. In the absorbance test, the plates were washed to remove the excess of the extrAvidin-peroxidase. Then, OPD peroxidase substrate was added and incubated in the dark for 15 minutes more. The reaction was stopped with 1M HCl. The plate was read at 450nm with the absorbance plate reader. The absorbance is directed correlated with the concentration of biotinylated neoglycoprotein that interacts with the lectin. The absorbance signal was corrected for the background by removing the absorbance value obtained with only PBS and without any glycoconjugate. Both for fluorescence and for absorbance test, a positive internal control to verify the correct running of the experiment was added.

GLYcoPROFILE in inhibition: the protocol is the same with few differences. After the saturation step with PBS/BSA 0.5% and the washing step, the non-labelled inhibitors were added to the plate in serial dilution while the labelled tracers (a biotinylated neoglycoprotein) were added to the plate at a fixed concentration, determined empirically. The plates were incubated for two hours at room temperature. The fluorescence or absorbance signal then were expressed in % of inhibition.

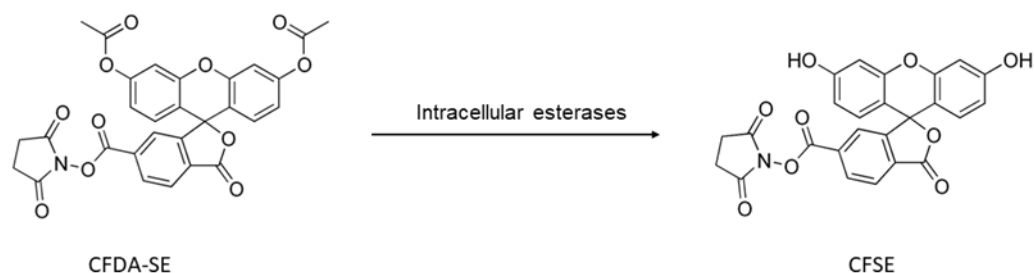
GraphPad Prism software (San Diego, CA, USA) was used for data analysis and graphing. The data presented are mean values of experiments performed in triplicate or duplicate. If the standard deviation did not exceed the 10% of the mean value, error bars are not shown in the graphs.

## 5.6 Mammalian cells culture and GLYcoPROFILE® of cells

All the steps before the labelling of cells and GLYcoPROFILE were performed in sterile conditions under a laminar flow hood. All reagents for culture cells (media and reagents) were purchased from Lonza. CFSE (Carboxyfluoresceinsuccinimidyl ester) for labelling of the cells was purchased from Sigma Aldrich. CHO K1 were obtained from Promogene and B16 were obtained from LGC Standard.

### 5.6.1 CHO K1

Adherent CHO K1 cells were grown in medium RPMI 1640 + L-glutamine 2mM + heat inactivated foetal bovine serum 5% + antibiotics. They were kept in the incubator at 37 °C with 5% of CO<sub>2</sub>. Before reaching the confluence, the cells were harvested using Trypsin/EDTA 1X and after harvesting, cells were washed in PBS pH 7.4 without cations, collected by centrifugation and resuspended in PBS pH 7.4 without cations. The cells were counted using the Trypan blue 0.4% , dye exclusion method. To perform the cells count, the Malassez counting chamber was used. After counting, all cells were diluted to 2x10<sup>6</sup> cells/mL and labelled with CFSE at 40 µM for 15 min at 37 °C under gentle agitation sheltered from the light. CFSE is a fluorescent molecule resulting from CFDA-SE (carboxyfluorescein diacetate succinimidyl ester), a cell permeant molecule which diffuse into the cells and it is cleaved by intracellular esterases (*Fig 5.2*). The CFSE binds indiscriminately with intracellular free amines to generate covalent fluorescent dye-protein conjugates resulting in live cells with an intracellular fluorescent label.



## 5.6.2B16

B16 murine melanoma cells were grown in medium DMEM supplemented with heat inactivated foetal bovine 10%, L-Glutamine 200 mM and mix of antibiotics 1X. They were kept in the incubator at 37°C with 5% of CO<sub>2</sub>. Before reaching the confluence, the cells were harvested using Trypsin/EDTA and after harvesting, cells were washed in PBS pH 7.4 without cations, collected by centrifugation and resuspended in PBS pH 7.4 without cations. The cells were counted using the Trypan blue 0.4% from Sigma. To perform a cell count, the Malassez counting chamber was used. After counting, all cells were diluted to 2x10<sup>6</sup> cells/mL and labelled with CFSE at 40 µM for 15 min at 37 °C under gentle agitation sheltered from the light. After washing with PBS without cations, cells were incubated on the previously prepared LEctPROFILE plate at a concentration of 2x10<sup>5</sup> cells/well. After washing, the fluorescence intensity was measured using the microplate reader. Data analysis was performed as previously described (5.7.1.).

## 5.7 Enzymatic digestion

α2-3 Neuraminidase Sfrom *Streptococcus pneumoniae* (P0743) and α2-3,6,8 Neuraminidase (P0720) from *Clostridium perfringens* were purchased from New England Biolabs. The α-1-(3,6) Galactosidase from green coffee beans was purchased from Sigma.

For the α2-3 Neuraminidase digestion, the glycoproteins or the monoclonal antibodies were incubated with the enzyme in a ratio of 8 units/µg of glycoconjugate. The reaction mix was incubated in a buffer at pH 6 (supplied with the enzyme) at 37°C for 1 hour. After the incubation time, the reaction was stopped on ice and the pH was reset to 7.4 with PBS. The samples were dialyzed against PBS pH 7.4 prior to use in GLYcoPROFILE. For the α2-3,6,8 Neuraminidase digestion, the glycoproteins or the monoclonal antibodies were incubated with the enzyme in a ratio of 2,5 units of enzyme/µg of glycoconjugate. The reaction mix was incubated in a buffer at pH 6 (supplied with the enzyme) at 37°C overnight. After the incubation time, the reaction was stopped and the pH was reset to 7.4 with PBS. The samples were dialyzed against PBS pH 7.4 prior to use in GLYcoPROFILE.

After digestion with α2-3,6,8 Neuraminidase, the monoclonal antibodies were incubated with the α-1-(3,6) Galactosidase in a ratio of 14 units/µg of glycoconjugate. The reaction mix was incubated in PBS pH 7.4 at 37°C for 24 hours and afterwards the reaction was stopped on ice and the pH was reset to 7.4 with PBS. The samples were dialyzed against PBS pH 7.4 prior to use in GLYcoPROFILE.

## 5.8 Serum IgG Purification

The blood samples were provided by Etablissement français du sang (EFS) for healthy donors, and supplied by the department of infectious and tropical diseases of the "Centre hospitalier régional d'Orléans (CHRO)" for sick donors.

The blood was clotted by leaving it undisturbed at room temperature for 15–30 minutes. The clot was removed by centrifugation of the samples at 2,000 x g for 10 minutes at 4°C and filtered through 0.45 µM. The samples were then purified on Protein A after dilution by 3 of sample in the equilibration buffer (0.02M sodium phosphate, pH 7.0). Pure antibodies were eluted with 0.1M glycine-HCl buffer (pH 2.7) and immediately neutralized with a TrisHCl buffer (1M, pH 9.0). The purified antibodies were controlled for their purity by Bradford and SDS-PAGE before their use in GLYcoPROFILE studies. The antibodies were kept at -20°C.

## **Chapter 6 – Conclusive remarks**

### Final conclusion and perspectives

The aim of the synBIOcarb network is to study the glycobiological interactions in order to design targeted drug-delivery systems and diagnostics tools. This ambitious goal requires a high-level of expertise in different fields, brought in the network by all beneficiaries and partners. My role in the network, at the company GLYcoDiag, has been in the part of the development of a new lectin array technology, based on the GLYcoPROFILE platform, for the in-process monitoring of biotherapeutic glycoproteins throughout all stages of the manufacturing process. This goal has taken a new shape in due course, and it has become possible to speculate upon other interesting applications requiring the use of a fast, high-throughput LEctPROFILE kit. This lectin array technology permits a wide variety of compounds to be screened at one time, in contrast to the nature of existing well-established analytical technologies, such as HPLC or CE.

The focus of the research was placed mainly on two glycan motifs: the Galili antigen and the sialic acid. The kits development was approached in several steps. First of all, we purified highly-selective lectins: MOA $\beta$ T, MAAI, HPyL, all of them carefully controlled for their concentration, purity and functionality. The purification of MOA $\beta$ T, performed with Dr. Simona Notova at CERMAV (CNRS, Grenoble) contributed to the production and characterisation of the Janus lectin RSL-MOA, described in the paper “Extending Janus lectins architecture: characterization and application to protocells”, at the end of Chapter 3.

Then, we undertook the synthesis of neoglycoproteins by click-chemistry reaction between the BSA functionalized with an alkyne or azido group and the sugars of interest functionalized with the same groups. We synthesized five neoglycoproteins, NeoGalili, Neo6'SL, Neo3'SL, NeoNeuAc and NeoNeuGc with a good yield and high purity. These neoglycoproteins allowed us to have reference molecules in order for us to define the kits characteristics. In conclusion, the work we published in collaboration with Professor's O. Renaudet's group was an extension of the initial synthesis done for neoglycoprotein. The paper, “Homo- and Heterovalent Neoglycoproteins as Ligands for Bacterial Lectins” (Chapter 2), in which we synthesized neoglycoproteins with clusters of glycans distributed at different position, opens the way for heterofunctionalization of neoglycoproteins.

The development of the Galili antigen LEctPROFILE kit was the following: first, we studied the optimal quantity of lectin MOA $\beta$ T/well to be immobilized on the microtiter plate and the best type of plate. We chose the covalent plate with EDC mediated amination linkage and an absorbance detection because it was the one giving the most robust signal. Then, we tested MOA $\beta$ T also with fluorescence detection: the sensitivity of the kit is lower in fluorescence but it is much more exploitable. The fluorescence detection removes the step of the addition of the substrate for the extravidin-peroxidase that makes the kit cheaper and the experiment faster. In addition, if MOA $\beta$ T is included in a wider kit containing more lectins, the fluorescence is advantageous because some



lectins cross-react with the peroxidase. The lectin GSLIB4 was also tested in fluorescence, having high specificity for the NeoGalili neoglycoprotein and a very good sensitivity. The two lectins MOA $\beta$ T and GSLIB4 can be considered complementary tools on the same kit: MOA $\beta$ T has a strong selectivity for the Galili antigen but it has low affinity for the monosaccharide and disaccharides whereas GSLIB4 has higher affinity for monosaccharides and disaccharides also linked in 1,4 and 1,6.

The development of the  $\alpha$ 2-3/ $\alpha$ 2-6 LEctPROFILE was carried out in fluorescence using the  $\alpha$ 2-3 sialic acid-linkage specific lectin MAAI and  $\alpha$ 2-6 sialic acid-linkage specific SNA. The kit performs well in direct binding mode, where the quantity of the glycosidic  $\alpha$ 2-3 and  $\alpha$ 2-6 linkages can be expressed as equivalents of neoglycoprotein or as ratio  $\alpha$ 2-3/ $\alpha$ 2-6 and in inhibition mode, where the results can be expressed as relative inhibition potency (RIP). Sialic acid was also the focus of another study in collaboration with Dr. S. Gouin *et al.* "Polymers of a transition-state sialyl cation strongly Inhibit Bacterial Sialidases", at the end of Chapter 4 . Here, we evaluated the IC<sub>50</sub> of potential sialidase multivalent inhibitors, thanks to sialic acid néoglycoproteins.

Altogether, considering the work and results presented in this thesis, the objective of the development of kits was met. However, several short-term and long-term perspectives should be considered for the future work. The purification of more lectins that may expand the screening of a wider variety of glycan motifs, such as core fucose, which is another important feature for the functionality of mAbs. Together with this, other than the neoglycoproteins already available at GLYcoDiag, the synthesis of neoglycoproteins containing more complex glycan motifs can be considered. In addition, it would be of great interest to test the kit with a wider pool of macromolecules. For example, more mAbs, IgG from healthy individuals and pathological individuals at different stages in order to monitor the glycosylation changes of the disease. It would be also interesting to use other glycoproteins to go on with the comparative experiments of quantification of the  $\alpha$ 2-3/  $\alpha$ 2-6 linkages with the LEctPROFILE kits and the MALDI-TOF/MS with linkages derivatization. The development of the Neu5Ac/Neu5Gc LEctPROFILE kit; which has been started, needs to be continued.

The ITN synBiocarb offered me the opportunity to build constructive collaborations: I spent one month at the Cermav in Grenoble in January 2020 in Dr. Anne Imberty's group. There, as mentioned before, together with Dr. Simona Notova, we purified the recombinant lectin MOA $\beta$ T that was of great use for my experiments on the LEctPROFILE kit and that is included in the paper at the end of Chapter 3. In addition, I had the chance to compare the kit performances to the MALDI/TOF/MS with derivatization of sialic acid linkages performed by Juvissan Aguedo at the Slovak Academy of Sciences. The synBIOcarb ITN provided me with high-level scientific training in the field of synthetic glycobiology, that helped me to build the knowledge to go ahead during the PhD program. Moreover, the network gave us many training in scientific communication: we had the opportunity to improve

our presenting and writing skills and we got the chance to learn the basics of videomaking and to make a video speaking about our project to a lay audience. The covid epidemic outburst was something completely unexpected and new that had a great impact on people's lives all over the world and naturally, on us as young scientists as well. Unfortunately, our scheduled synBIOcarb biannual meetings were only three and many congresses I was supposed to participate in, were cancelled. No matter this, we were given the training remotely and I experienced to present online and, once travelling was permitted again, I had the chance to go to conferences to present my work (listed below). Last but not least, being in an industrial PhD program gave me the opportunity to look at my research from a more pragmatic and applicable way and gave me the chance to engage in multiple collaborations resulted in publications (listed below).

### Conclusion finale et perspectives

L'objectif du projet Innovative Training Networks (ITN) « synBIOcarb » est d'étudier les interactions glycobiotiques afin de concevoir des systèmes de délivrance de médicaments et des outils de diagnostic ciblés. Cet objectif ambitieux nécessite un haut niveau d'expertise dans différents domaines, qui est constitué par le réseau de tous les bénéficiaires et partenaires de ce projet. Mon rôle dans le réseau, au sein de la société GLYcoDiag, a été de participer au développement d'une nouvelle technologie de lectine array, basée sur la plateforme GLYcoPROFILE, pour le suivi en cours de production des glycoprotéines biothérapeutiques et à toutes les étapes du processus de fabrication. Cet objectif a évolué au cours du temps pour s'adapter également à d'autres applications intéressantes nécessitant l'utilisation d'un kit LECTPROFILE rapide et à haut débit. Cette technologie de réseau de lectines permet de cribler une grande variété de composés en une seule fois, sans nécessiter d'une forte expertise de la technologie, contrairement à la nature des technologies analytiques existantes et bien établies, telles que la spectrométrie de masse, la chromatographie (HPLC) ou l'électrophorèse capillaire (CE).

La recherche s'est principalement concentrée sur deux motifs glycaniques : l'antigène Galili et l'acide sialique. Le développement des kits a été abordé en plusieurs étapes. Tout d'abord, nous avons purifié des lectines hautement sélectives: MOA $\beta$ T, MAAI, HPyL, toutes soigneusement contrôlées pour leur pureté et leur fonctionnalité. La purification de MOA $\beta$ T, réalisée avec le Dr Simona Notova au CERMAV (CNRS, Grenoble), a contribué à la production et à la caractérisation de la « Janus » lectine RSL-MOA, décrite dans l'article "Extending Janus lectins architecture : characterization and application to protocells", à la fin du chapitre 3.

Par la suite, nous avons entrepris la synthèse de néoglycoprotéines par réaction de chimie du click entre la sérum albumine bovine (BSA) fonctionnalisée par des groupements alcyne ou azido et des glycanes d'intérêt fonctionnalisés par les mêmes groupements chimiques. Nous avons ainsi synthétisé cinq néoglycoprotéines, NeoGalili, Neo6'SL, Neo3'SL, NeoNeuAc et NeoNeuGc avec un bon rendement, une grande pureté et selon des protocoles particulièrement reproductibles. Ces

néoglycoprotéines nous ont permis de disposer de molécules de référence afin de définir les caractéristiques des kits que nous allons développer ensuite. En conclusion de cette partie, nous présentons le travail que nous avons publié en collaboration avec le groupe du Professeur O. Renaudet, qui représentait une extension du travail initial mené autour de la synthèse des néoglycoprotéines. L'article "Homo- and Heterovalent Neoglycoproteins as Ligands for Bacterial Lectins" (Chapitre 2), dans lequel nous avons synthétisé des néoglycoprotéines avec des clusters de glycannes distribués sur différentes positions, ouvre la voie à l'hétérofonctionnalisation des néoglycoprotéines.

Le développement du kit LEctPROFILE ciblant l'antigène Galili a été le suivant : tout d'abord, nous avons étudié la quantité optimale de lectine MOA $\beta$ T/puits à immobiliser sur la plaque de microtitration et le meilleur type de plaque. Nous avons choisi initialement une liaison d'amination médiée par l'EDC et une détection par absorbance car c'était celle qui donnait le signal le plus robuste. Ensuite, nous avons également adapté l'utilisation de la lectine MOA $\beta$ T à une détection par fluorescence : la sensibilité du kit est plus faible en fluorescence mais elle est beaucoup plus exploitable. En effet, la détection par fluorescence supprime l'étape de l'ajout du substrat de l'extravidine-peroxydase ce qui rend le kit moins simple et plus rapide d'utilisation. De plus, si la lectine MOA $\beta$ T est incluse dans un kit plus large contenant plus de lectines, la fluorescence est avantageuse car certaines lectines réagissent de manière croisée avec la peroxydase. La lectine GSLIB4 a également été étudiée en fluorescence pour le développement de ce kit. Ayant une haute spécificité pour la néoglycoprotéine NeoGalili et une très bonne sensibilité, les deux lectines MOA $\beta$ T et GSLIB4 peuvent être considérées comme des outils complémentaires sur le même kit : MOA $\beta$ T a une forte sélectivité pour l'antigène Galili mais elle a une faible affinité pour les monosaccharides et disaccharides alors que GSLIB4 a une affinité plus élevée pour les monosaccharides et disaccharides également liés en 1,4 et 1,6.

Le développement du LEctPROFILE  $\alpha$ 2-3/ $\alpha$ 2-6 a été réalisé en fluorescence à l'aide de la lectine MAAI spécifique de la liaison acide sialique  $\alpha$ 2-3 et de la lectine SNA spécifique de la liaison acide sialique  $\alpha$ 2-6. Le kit est performant en mode de liaison directe, où la quantité de liaisons glycosidiques  $\alpha$ 2-3 et  $\alpha$ 2-6 peut être exprimée en équivalents de néoglycoprotéine ou en ratio de liaisons  $\alpha$ 2-3/ $\alpha$ 2-6. En mode d'inhibition/compétition, où les résultats peuvent être exprimés en potentiel d'inhibition relative (RIP), permettant ainsi de comparer aisément l'avidité de nombreux composés glycosylés simultanément. Cette partie de reconnaissance des acides sialiques a également fait l'objet d'une autre étude menée en collaboration avec l'équipe du Dr S. Gouin et basée sur l'évaluation des IC<sub>50</sub> d'inhibiteurs multivalents potentiels de sialidases. Ces résultats ont été publiés dans l'article " Polymers of a transition-state sialyl cation strongly Inhibit Bacterial Sialidases ", ajouté à la fin du chapitre 4.

Dans l'ensemble, au vu des travaux et des résultats présentés dans cette thèse, l'objectif de développement de kits a été atteint. Cependant, plusieurs perspectives à court et à long terme

doivent être envisagées pour les travaux futurs. La purification d'un plus grand nombre de lectines qui pourrait étendre le criblage à une plus grande variété de motifs glycaniques, tels que le « core-fucose », qui est une autre caractéristique importante à contrôler pour la fonction effectrice des anticorps monoclonaux thérapeutiques (mAbs). En plus des néoglycoprotéines déjà disponibles chez GLYcoDiag, la synthèse de néoglycoprotéines contenant des motifs glycaniques plus complexes est aussi à envisager. De plus, il serait très intéressant de tester le kit avec un plus grand nombre de macromolécules. Par exemple, plus de mAbs, IgG provenant d'individus sains et d'individus pathologiques à différents stades inflammatoires afin de suivre les changements de glycosylation de la maladie. Il serait également intéressant d'utiliser d'autres glycoprotéines pour poursuivre les expériences comparatives de quantification des liaisons  $\alpha 2$ -3/  $\alpha 2$ -6 avec les kits LEctPROFILE et le MALDI-TOF/MS avec dérivation des liaisons. Le développement du kit LEctPROFILE Neu5Ac/Neu5Gc, qui a été commencé, doit être également poursuivi.

L'ITN synBiocarb m'a offert l'opportunité de construire des collaborations constructives : J'ai passé un mois au CERMAV de Grenoble en janvier 2020 dans le groupe du Dr Anne Imberty. Là, comme mentionné précédemment, avec le Dr. Simona Notova, nous avons purifié la lectine recombinante MOA $\beta$ T qui a été d'une grande utilité pour mes expériences sur le kit LEctPROFILE « Galili » . En outre, ce travail a fait l'objet d'une publication au travers de la « Janus » lectine MOA $\beta$ T - RSL dans l'article inclus à la fin du chapitre 3. J'ai également eu l'occasion, dans le cadre d'une glycoprotéine modèle, de comparer les performances du kit LEctPROFILE  $\alpha 2$ -3/ $\alpha 2$ -6 aux résultats obtenus par MALDI/TOF/MS avec dérivation des liaisons d'acide sialique, réalisé par Juvissan Aguedo à l'Académie des Sciences de Slovaquie. L'ITN synBIOcarb m'a fourni une formation scientifique de haut niveau dans le domaine de la glycobiologie synthétique, ce qui m'a aidé à acquérir les connaissances nécessaires pour poursuivre mon programme de doctorat. De plus, le volet formation de cet ITN nous a permis de suivre de nombreuses formations en communication scientifique : nous avons eu l'occasion d'améliorer nos compétences en matière de présentation et d'écriture et nous avons eu la chance d'apprendre les bases de la réalisation de vidéos et de faire une vidéo parlant de notre projet à un public profane. Le déclenchement de la pandémie de Covid était quelque chose de complètement inattendu et nouveau qui a eu un grand impact sur la vie des gens dans le monde entier et, naturellement, sur nous aussi en tant que jeunes scientifiques. Malheureusement, les réunions semestrielles de SynBIOcarb que nous avions prévues ont été réduites à trois et de nombreux congrès auxquels je devais participer ont été annulés. Malgré tout, nous avons reçu une formation à distance et j'ai appris à faire des présentations en ligne pour plusieurs réunions et conférences. Une fois que les voyages ont été à nouveau autorisés, j'ai eu la chance d'assister à des conférences pour présenter mes travaux (voir la liste ci-dessous). Enfin, le fait d'être dans un programme de doctorat industriel m'a donné l'occasion de considérer mes recherches d'une manière plus pragmatique et applicable et m'a donné la chance de m'engager dans de multiples collaborations qui ont abouti à des publications (voir ci-dessous).

### List of communications

1. Poster: B. Roubinet, J. Bouckaert, F. Vena, L. Landemarre, Simple method for high throughput screening of FimH ligands, 2019, International Symposium on Glycoconjugates, Milano, 25<sup>th</sup> – 31<sup>st</sup> August.
2. Online Presentation: F. Vena, Lectin array for quality control of recombinant biotherapeutic glycoproteins, 29/04/2021, mi-thèse ICOA.
3. Online Presentation: F. Vena, Lectin array for quality control of recombinant biotherapeutic glycoproteins, 03/06/2021, Group Français des Glycosciences.
4. Poster and Flash Presentation: F. Vena, "Lectin array for quality control of recombinant biotherapeutic glycoproteins" 33<sup>o</sup> Colloque Biotechnocentre, 7th-8th October 2021, Center Parc-Domaine des Hauts de Bruyères, France.
5. Flash Presentation: F. Vena, "Lectin array for quality control of recombinant biotherapeutic glycoproteins" 4th Glycobasque meeting, 11th-12th November 2021, Donostia - San Sebastián, Spain.
6. Video: F. Vena, Diary of a Scientific Researcher: Federica Vena (ESR14) chapter one, New Frontiers Symposium Translational Glycoscience, 18th 19th November 2021, Online event.
7. Presentation: F. Vena, "Lectin array for quality control of recombinant biotherapeutic glycoproteins"  
and  
Poster: J. Aguedo, F. Vena, L. Landemarre, J. Tkac, Rapid and high-throughput methods for discrimination of sialic acid linkages in glycoproteins  
Joint Training Cost Training School and synBIOcarb «Advanced Training in Synthesis and Applications of Multivalent Glyconanomaterials», 29th March 1st April 2022, Prague, Czech Republic.
8. Poster: J. Aguedo, F. Vena, L. Landemarre, J. Tkac, Rapid and high-throughput methods for discrimination of sialic acid linkages in glycoproteins, Group Français des Glycosciences, 30th May-3rd June 2022, Normandy, France.

**Publications**

1. Goyard D., Roubinet B., Vena F, Landemarre, Renaudet O., Homo- and Heterovalent Neoglycoproteins as Ligands for Bacterial Lectins, 2021, ChemPlusChem, 2, 1- 9.
2. Assailly C., Bridot C., Saumonneau A., Lottin P., Roubinet B., Krammer E., François F., Vena F., Landemarre L., Alvarez-Dorta D., Polymers of a transition-state sialyl cation strongly Inhibit Bacterial Sialidases, 2020, Chemistry a European Journal, 27, 3142 – 3150.
3. Simona Notova , Lina Suikstaite b, Francesca Rosato, Federica Vena , Aymeric Audfray, Nicolai Bovin , Ludovic Landemarre , Winfried Römer , Anne Imberty, Extending Janus lectins architecture: characterization and application to protocells”, 2022, BioRxiv. Submitted to Computational and Structural Biotechnology Journal.

## Bibliography

- [1] L. Landemarre and E. Duverger, "Lectin Glycoprofiling of Recombinant Therapeutic Interleukin-7," vol. 988, 2013, doi: 10.1007/978-1-62703-327-5.
- [2] J. C. Venter *et al.*, "The Sequence of the Human Genome," 2001.
- [3] N. Jayaraman, "Multivalent ligand presentation as a central concept to study intricate carbohydrate-protein interactions," *Chem. Soc. Rev.*, vol. 38, no. 12, pp. 3463–3483, Nov. 2009, doi: 10.1039/b815961k.
- [4] A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, G. W. Hart, and J. D. Marth, *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, 2017.
- [5] J. E. Turnbull and R. A. Field, "Emerging glycomics technologies," *Nat. Chem. Biol.*, vol. 3, no. 2, pp. 74–77, 2007, doi: 10.1038/nchembio0207-74.
- [6] N. Sharon, "Lectins: Cell-Agglutinating and Sugar-Specific Proteins provide new tools for studying polysaccharide glycoproteins, and cell surfaces, and for cancer research," 1972.
- [7] O. Sulak, E. Lameignère, M. Wimmerova, and A. Imberty, "Specificity and affinity studies in lectin/carbohydrate interactions," 2009, pp. 357–372.
- [8] S. Cecioni, A. Imberty, and S. Vidal, "Glycomimetics versus multivalent glycoconjugates for the design of high affinity lectin ligands," *Chem. Rev.*, vol. 115, no. 1, pp. 525–561, 2015, doi: 10.1021/cr500303t.
- [9] Y. C. Lee and R. T. Lee, "Carbohydrate-Protein Interactions: Basis of Glycobiology," *Acc. Chem. Res.*, vol. 28, no. 8, pp. 321–327, 1995, doi: 10.1021/ar00056a001.
- [10] P. K. Grewal, *The Ashwell-Morell receptor*, 1st ed., vol. 479, no. C. Elsevier Inc., 2010.
- [11] K. L. Hsu, K. T. Pilobello, and L. K. Mahal, "Analyzing the dynamic bacterial glycome with a lectin microarray approach," *Nat. Chem. Biol.*, vol. 2, no. 3, pp. 153–157, 2006, doi: 10.1038/nchembio767.
- [12] P. Gemeiner *et al.*, "Lectinomics. II. A highway to biomedical/clinical diagnostics," *Biotechnol. Adv.*, vol. 27, no. 1, pp. 1–15, 2009, doi: 10.1016/j.biotechadv.2008.07.003.
- [13] G. Gupta, A. Surolia, and S. G. Sampathkumar, "Lectin microarrays for glycomic analysis," *Omi. A J. Integr. Biol.*, vol. 14, no. 4, pp. 419–436, 2010, doi: 10.1089/omi.2009.0150.
- [14] K. T. Pilobello, D. E. Slawek, and L. K. Mahal, "A ratiometric lectin microarray approach to analysis of the dynamic mammalian glycome," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 28, pp. 11534–11539, 2007, doi: 10.1073/pnas.0704954104.
- [15] A. Kuno *et al.*, "Evanescent-field fluorescence-assisted lectin microarray: A new strategy for glycan profiling," *Nat. Methods*, vol. 2, no. 11, pp. 851–856, 2005, doi: 10.1038/nmeth803.
- [16] S. Angeloni *et al.*, "Glycoprofiling with micro-arrays of glycoconjugates and lectins," *Glycobiology*, vol. 15, no. 1, pp. 31–41, 2005, doi: 10.1093/glycob/cwh143.
- [17] S. Chen *et al.*, "Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays," *Nat. Methods*, vol. 4, no. 5, pp. 437–444, 2007, doi: 10.1038/nmeth1035.
- [18] Y. Qiu *et al.*, "NIH Public Access," vol. 7, no. 4, pp. 1693–1703, 2009, doi: 10.1021/pr700706s.Plasma.
- [19] T. Bertok *et al.*, "Analysis of serum glycome by lectin microarrays for prostate cancer patients - a search for aberrant glycoforms," *Glycoconj. J.*, vol. 37, no. 6, pp. 703–711, 2020, doi: 10.1007/s10719-020-09958-4.
- [20] S. M. Muthana and J. C. Gildersleeve, "Glycan microarrays: Powerful tools for biomarker discovery," *Cancer Biomarkers*, vol. 14, no. 1, pp. 29–41, 2014, doi: 10.3233/CBM-130383.

- [21] H. H. Wandall *et al.*, "Cancer biomarkers defined by autoantibody signatures to aberrant O-glycopeptide epitopes," *Cancer Res.*, vol. 70, no. 4, pp. 1306–1313, 2010, doi: 10.1158/0008-5472.CAN-09-2893.
- [22] E. Suenaga, H. Mizuno, and K. K. R. Penmetcha, "Monitoring influenza hemagglutinin and glycan interactions using surface plasmon resonance," *Biosens. Bioelectron.*, vol. 32, no. 1, pp. 195–201, 2012, doi: 10.1016/j.bios.2011.12.003.
- [23] R. Prasad, R. L. Stout, D. Coffin, and J. Smith, "Analysis of carbohydrate deficient transferrin by capillary zone electrophoresis," *Electrophoresis*, vol. 18, no. 10, pp. 1814–1818, 1997, doi: 10.1002/elps.1150181016.
- [24] K. R. Reiding, D. Blank, D. M. Kuijper, A. M. Deelder, and M. Wuhler, "High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification," *Anal. Chem.*, vol. 86, no. 12, pp. 5784–5793, 2014, doi: 10.1021/ac500335t.
- [25] T. E. G. Community, "Glyco@ALPS - Glyco 2030: A Roadmap for Glycoscience in Europe."
- [26] V. G. Dhara, H. M. Naik, N. I. Majewska, and M. J. Betenbaugh, "Recombinant Antibody Production in CHO and NS0 Cells: Differences and Similarities," *BioDrugs*, 2018, doi: 10.1007/s40259-018-0319-9.
- [27] L. Zhang, S. Luo, and B. Zhang, "The use of lectin microarray for assessing glycosylation of therapeutic proteins," *MAbs*, vol. 8, no. 3, pp. 524–535, 2016, doi: 10.1080/19420862.2016.1149662.
- [28] L. O. Narhi *et al.*, "The effect of carbohydrate on the structure and stability of erythropoietin," *J. Biol. Chem.*, vol. 266, no. 34, pp. 23022–23026, 1991, doi: 10.1016/s0021-9258(18)54457-4.
- [29] C. Raymond, A. Robotham, J. Kelly, E. Lattova, H. Perreault, and Y. Durocher, "Production of Highly Sialylated Monoclonal Antibodies," *Glycosylation*, 2012, doi: 10.5772/51301.
- [30] B. A. Cobb, "The history of IgG glycosylation and where we are now," *Glycobiology*, vol. 30, no. 4, pp. 202–213, 2021, doi: 10.1093/GLYCOB/CWZ065.
- [31] MabDesign, "Immunowatch Edition n°3 - July 2021," *Immunowatch*, vol. 3, no. July, 2021.
- [32] A. Beck, O. Cochet, and T. Wurch, "Glycofi's technology to control the glycosylation of recombinant therapeutic proteins," *Expert Opin. Drug Discov.*, vol. 5, no. 1, pp. 95–111, 2010, doi: 10.1517/17460440903413504.
- [33] M. E. Lalonde and Y. Durocher, "Therapeutic glycoprotein production in mammalian cells," *J. Biotechnol.*, vol. 251, no. April, pp. 128–140, 2017, doi: 10.1016/j.jbiotec.2017.04.028.
- [34] Y. Durocher and M. Butler, "Expression systems for therapeutic glycoprotein production," *Curr. Opin. Biotechnol.*, vol. 20, no. 6, pp. 700–707, 2009, doi: 10.1016/j.copbio.2009.10.008.
- [35] J. Dumont, D. Euwart, B. Mei, S. Estes, and R. Kshirsagar, "Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives," *Crit. Rev. Biotechnol.*, vol. 36, no. 6, pp. 1110–1122, 2016, doi: 10.3109/07388551.2015.1084266.
- [36] I. J. del Val *et al.*, "Application of Quality by Design Paradigm to the Manufacture of Protein Therapeutics," *Glycosylation*, 2012, doi: 10.5772/50261.
- [37] L. Zhang, S. Luo, and B. Zhang, "Glycan analysis of therapeutic glycoproteins," *MAbs*, vol. 8, no. 2, pp. 205–215, 2016, doi: 10.1080/19420862.2015.1117719.
- [38] N. Lin *et al.*, "Chinese hamster ovary (CHO) host cell engineering to increase sialylation of recombinant therapeutic proteins by modulating sialyltransferase expression," *Biotechnol. Prog.*, vol. 31, no. 2, pp. 334–346, 2015, doi: 10.1002/btpr.2038.
- [39] S. Elliott *et al.*, "Control of rHuEPO biological activity: The role of carbohydrate," *Exp. Hematol.*, vol. 32, no. 12, pp. 1146–1155, 2004, doi: 10.1016/j.exphem.2004.08.004.
- [40] T. A. Kost, J. P. Condreay, and D. L. Jarvis, "Baculovirus as versatile vectors for protein expression in



- insect and mammalian cells," *Nat. Biotechnol.*, vol. 23, no. 5, pp. 567–575, 2005, doi: 10.1038/nbt1095.
- [41] B. Liu, M. Spearman, J. Doering, E. Lattová, H. Perreault, and M. Butler, "The availability of glucose to CHO cells affects the intracellular lipid-linked oligosaccharide distribution, site occupancy and the N-glycosylation profile of a monoclonal antibody," *J. Biotechnol.*, vol. 170, no. 1, pp. 17–27, 2014, doi: 10.1016/j.jbiotec.2013.11.007.
- [42] J. M. Roth, "Recombinant Tissue Plasminogen Activator for the Treatment of Acute Ischemic Stroke," *Baylor Univ. Med. Cent. Proc.*, vol. 24, no. 3, pp. 257–259, 2011, doi: 10.1080/08998280.2011.11928729.
- [43] M.-F. Clincke, E. Guedon, F. T. Yen, V. Ogier, and J.-L. Goergen, "Effect of iron sources on the glycosylation macroheterogeneity of human recombinant IFN- $\gamma$  produced by CHO cells during batch processes," *BMC Proc.*, vol. 5, no. S8, p. P114, 2011, doi: 10.1186/1753-6561-5-s8-p114.
- [44] ICH Q8, "EMA/CHMP, 2009, ICH Topic Q 8 (R2) Pharmaceutical Development, Step 5: Note for Guidance on Pharmaceutical Development," *Regul. ICH*, vol. 8, no. June, 2009.
- [45] Ich, "INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN USE ICH HARMONISED TRIPARTITE GUIDELINE SPECIFICATIONS: TEST PROCEDURES AND ACCEPTANCE CRITERIA FOR BIOTECHNOLOGICAL/BIOLOGICAL PRODUCTS Q6B," 1999.
- [46] M. Ratner, "Genzyme's Lumizyme clears bioequivalence hurdles," *Nat. Biotechnol.*, vol. 27, no. 8, pp. 685–685, 2009, doi: 10.1038/nbt0809-685a.
- [47] L. X. Yu *et al.*, "Understanding pharmaceutical quality by design," *AAPS J.*, vol. 16, no. 4, pp. 771–783, 2014, doi: 10.1208/s12248-014-9598-3.
- [48] R. J. Solá and K. Griebenow, "Glycosylation of Therapeutic Proteins," *BioDrugs*, vol. 24, no. 1, pp. 9–21, 2010, doi: 10.2165/11530550-000000000-00000.
- [49] O. Srinivas and M. Monsigny, "3.23 Neoglycoproteins ´.," pp. 477–521, 2007.
- [50] Y. M. Chabre and R. Roy, *Design and Creativity in Synthesis of Multivalent Neoglycoconjugates*, vol. 63, no. 10. 2010.
- [51] J. Arnaud *et al.*, "Membrane deformation by neolectins with engineered glycolipid binding sites," *Angew. Chemie - Int. Ed.*, vol. 53, no. 35, pp. 9267–9270, 2014, doi: 10.1002/anie.201404568.
- [52] J. P. Ribeiro *et al.*, "Tailor-made Janus lectin with dual avidity assembles glycoconjugate multilayers and crosslinks protocells," *Chem. Sci.*, vol. 9, no. 39, pp. 7634–7641, 2018, doi: 10.1039/C8SC02730G.
- [53] L. L. Kiessling, J. E. Gestwicki, and L. E. Strong, "Synthetic multivalent ligands in the exploration of cell-surface interactions," *Curr. Opin. Chem. Biol.*, vol. 4, no. 6, pp. 696–703, 2000, doi: 10.1016/S1367-5931(00)00153-8.
- [54] L. V. Mochalova *et al.*, "Synthetic polymeric inhibitors of influenza virus receptor-binding activity suppress virus replication," *Antiviral Res.*, vol. 23, no. 3–4, pp. 179–190, 1994, doi: 10.1016/0166-3542(94)90016-7.
- [55] M. Reynolds, M. Marradi, A. Imberty, S. Penadés, and S. Pérez, "Multivalent gold glycoclusters: High affinity molecular recognition by bacterial lectin PA-IL," *Chem. - A Eur. J.*, vol. 18, no. 14, pp. 4264–4273, 2012, doi: 10.1002/chem.201102034.
- [56] C. P. Stowell and Y. C. Lee, *Neoglycoproteins the preparation and application of synthetic glycoproteins*, vol. 37, no. C. 1980.
- [57] A. Berkin, B. Coxon, and V. Pozsgay, "Towards a synthetic glycoconjugate vaccine against *Neisseria meningitidis* A," *Chem. - A Eur. J.*, vol. 8, no. 19, pp. 4424–4433, 2002, doi: 10.1002/1521-3765(20021004)8:19<4424::AID-CHEM4424>3.0.CO;2-1.

- [58] B. Y. O. S. W. A. L. D. T. Avery and W. A. L. T. H. E. R. F. Goebel, "I. Anticarbhydrate Antibodies," no. 3, pp. 769–780, 1932.
- [59] R. Roy, E. Katzenellenbogen, and H. J. Jennings, "Improved procedures for the conjugation of oligosaccharides to protein by reductive amination.," *Can. J. Biochem. Cell Biol.*, vol. 62, no. 5, pp. 270–275, 1984, doi: 10.1139/o84-037.
- [60] A. Ogura, A. Kurbangalieva, and K. Tanaka, "Exploring the glycan interaction in vivo: Future prospects of neo-glycoproteins for diagnostics," *Glycobiology*, vol. 26, no. 8, pp. 804–812, 2016, doi: 10.1093/glycob/cww038.
- [61] C. W. Tornøe, C. Christensen, and M. Meldal, "Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides," 2002, doi: 10.1021/jo011148j.
- [62] V. V. Rostovtsev, L. G. Green, V. V. Fokin, and K. B. Sharpless, "A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective 'ligation' of azides and terminal alkynes," *Angew. Chemie - Int. Ed.*, vol. 41, no. 14, pp. 2596–2599, 2002, doi: 10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4.
- [63] E. Haldón, M. C. Nicasio, and P. J. Pérez, "Copper-catalysed azide-alkyne cycloadditions (CuAAC): An update," *Org. Biomol. Chem.*, vol. 13, no. 37, pp. 9528–9550, 2015, doi: 10.1039/c5ob01457c.
- [64] A. Imberty, C. Gautier, J. Lescar, S. Pérez, L. Wyns, and R. Loris, "An unusual carbohydrate binding site revealed by the structures of two Maackia amurensis lectins complexed with sialic acid-containing oligosaccharides," *J. Biol. Chem.*, vol. 275, no. 23, pp. 17541–17548, 2000, doi: 10.1074/jbc.M000560200.
- [65] L. Maveyraud *et al.*, "Structural basis for sugar recognition, including the Tn carcinoma antigen, by the lectin SNA-II from Sambucus nigra," *Proteins Struct. Funct. Bioinforma.*, vol. 75, no. 1, pp. 89–103, Apr. 2009, doi: 10.1002/prot.22222.
- [66] U. Galili, "Anti-Gal: An abundant human natural antibody of multiple pathogeneses and clinical benefits," *Immunology*, vol. 140, no. 1, pp. 1–11, Sep-2013, doi: 10.1111/imm.12110.
- [67] "jhev.1995.1067."
- [68] U. Galili, F. Anaraki, A. Thall, C. Hill-Black, and M. Radic, "One Percent of Human Circulating B Lymphocytes Are Capable of Producing the Natural Anti-Gal Antibody."
- [69] S. Kirkeby, H. C. Winter, and I. J. Goldstein, "Comparison of the binding properties of the mushroom Marasmius oreades lectin and Griffonia simplicifolia I-B4 isolectin to  $\alpha$ galactosyl carbohydrate antigens in the surface phase," *Xenotransplantation*, vol. 11, no. 3, pp. 254–261, May 2004, doi: 10.1111/j.1399-3089.2004.00108.x.
- [70] N. Patil, M. Abba, and H. Allgayer, "Cetuximab and biomarkers in non-small-cell lung carcinoma," *Biologics: Targets and Therapy*, vol. 6, pp. 221–231, 2012, doi: 10.2147/BTT.S24217.
- [71] N. England Biolabs, "Glycoproteomics Technical Guide."
- [72] J. W. Steinke, T. A. E. Platts-Mills, and S. P. Commins, "The alpha-gal story: Lessons learned from connecting the dots," *J. Allergy Clin. Immunol.*, vol. 135, no. 3, pp. 589–596, 2015, doi: 10.1016/j.jaci.2014.12.1947.
- [73] U. Galili, "The  $\alpha$ -gal epitope (Gal $\alpha$ 1-3Gal1-4GlcNAc-R) in xenotransplantation," 2001.
- [74] A. Rana and E. Godfre, "Outcomes in Solid-Organ Transplantation:," vol. 46, no. 1, pp. 75–76, 2019.
- [75] C. J. Phelps *et al.*, "Production of  $\alpha$ 1,3-galactosyltransferase-deficient pigs," *Science (80-. )*, vol. 299, no. 5605, pp. 411–414, Jan. 2003, doi: 10.1126/SCIENCE.1078942.
- [76] B. P. Griffith *et al.*, "Genetically Modified Porcine-to-Human Cardiac Xenotransplantation," *N. Engl. J.*

*Med.*, pp. 35–44, 2022, doi: 10.1056/nejmoa2201422.

- [77] W. Wang, W. He, Y. Ruan, and Q. Geng, "First pig-to-human heart transplantation," *Innov.*, vol. 3, no. 2, p. 100223, 2022, doi: 10.1016/j.xinn.2022.100223.
- [78] B. A. Macher and U. Galili, "The Gal $\alpha$ 1,3Gal $\beta$ 1,4GlcNAc-R ( $\alpha$ -Gal) epitope: A carbohydrate of unique evolution and clinical relevance," *Biochimica et Biophysica Acta - General Subjects*, vol. 1780, no. 2, pp. 75–88, Feb-2008, doi: 10.1016/j.bbagen.2007.11.003.
- [79] V. Sokolova, A. M. Westendorf, J. Buer, K. Überla, and M. Eppele, "The potential of nanoparticles for the immunization against viral infections," *Journal of Materials Chemistry B*, vol. 3, no. 24, Royal Society of Chemistry, pp. 4767–4779, 28-Jun-2015, doi: 10.1039/c5tb00618j.
- [80] P. C. Roberts, W. Garten, and H.-D. Klenk, "Role of Conserved Glycosylation Sites in Maturation and Transport of Influenza A Virus Hemagglutinin," 1993.
- [81] U. M. Abdel-Motal, H. M. Guay, K. Wigglesworth, R. M. Welsh, and U. Galili, "Immunogenicity of Influenza Virus Vaccine Is Increased by Anti-Gal-Mediated Targeting to Antigen-Presenting Cells," *J. Virol.*, vol. 81, no. 17, pp. 9131–9141, Sep. 2007, doi: 10.1128/jvi.00647-07.
- [82] U. Abdel-Motal, S. Wang, S. Lu, K. Wigglesworth, and U. Galili, "Increased Immunogenicity of Human Immunodeficiency Virus gp120 Engineered To Express Gal $\alpha$ 1-3Gal $\beta$ 1-4GlcNAc-R Epitopes," *J. Virol.*, vol. 80, no. 14, pp. 6943–6951, Jul. 2006, doi: 10.1128/jvi.00310-06.
- [83] G. F. Whalen, M. Sullivan, B. Piperdi, W. Wasseff, and U. Galili, "Cancer immunotherapy by intratumoral injection of  $\alpha$ -gal glycolipids," *Anticancer Res.*, vol. 32, no. 9, pp. 3861–3868, 2012.
- [84] S. Juillot *et al.*, "Uptake of Marasmius oreades agglutinin disrupts integrin-dependent cell adhesion," *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1860, no. 2, pp. 392–401, 2016, doi: 10.1016/j.bbagen.2015.11.002.
- [85] E. Grahn *et al.*, "Crystal Structure of the Marasmius Oreades Mushroom Lectin in Complex with a Xenotransplantation Epitope," *J. Mol. Biol.*, vol. 369, no. 3, pp. 710–721, 2007, doi: 10.1016/j.jmb.2007.03.016.
- [86] W. Tempel, L. A. Lipscomb, J. P. Rose, and R. J. Woods, "The xenograft antigen in complex with GS-1-B 4 lectin: crystallization and preliminary X-ray analysis," 2001.
- [87] H. C. Winter, K. Mostafapour, and I. J. Goldstein, "The mushroom Marasmius oreades lectin is a blood group type B agglutinin that recognizes the Gal $\alpha$ 1,3Gal and Gal $\alpha$ 1,3Gal, $\beta$ 1,4GlcNAc porcine xenotransplantation epitopes with high affinity," *J. Biol. Chem.*, vol. 277, no. 17, pp. 14996–15001, 2002, doi: 10.1074/jbc.M200161200.
- [88] C. Lee *et al.*, "mAbs Glycosylation profile and biological activity of Remicade® compared with Flixabi® and Remsima® Glycosylation profile and biological activity of Remicade Ò compared with Flixabi Ò and Remsima Ò," 2017, doi: 10.1080/19420862.2017.1337620.
- [89] R. Schauer, "Sialic acids and their role as biological masks," *Trends Biochem. Sci.*, vol. 10, no. 9, pp. 357–360, Sep. 1985, doi: 10.1016/0968-0004(85)90112-4.
- [90] Z. Z., W. M., and H. S., "Serum sialylation changes in cancer," *Glycoconj. J.*, vol. 35, no. 2, pp. 139–160, 2018.
- [91] E. Maverakis *et al.*, "Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: A critical review," *J. Autoimmun.*, vol. 57, no. January, pp. 1–13, 2015, doi: 10.1016/j.jaut.2014.12.002.
- [92] A. Varki, "Sialic acids in human health and disease," *Trends Mol. Med.*, vol. 14, no. 8, pp. 351–360, 2008, doi: 10.1016/j.molmed.2008.06.002.
- [93] X. Zhou, G. Yang, and F. Guan, "Biological Functions and Analytical Strategies of Sialic Acids in Tumor," *Cells*, vol. 9, no. 2, pp. 1–17, 2020, doi: 10.3390/cells9020273.

- [94] X. Gu and D. I. C. Wang, "Improvement of interferon- $\gamma$  sialylation in Chinese hamster ovary cell culture by feeding of N-acetylmannosamine," *Biotechnol. Bioeng.*, vol. 58, no. 6, pp. 642–648, 1998, doi: 10.1002/(SICI)1097-0290(19980620)58:6<642::AID-BIT10>3.0.CO;2-9.
- [95] N. M. Porpiglia, E. F. De Palo, S. A. Savchuk, S. A. Appolonova, F. Bortolotti, and F. Tagliaro, "A new sample treatment for asialo-Tf determination with capillary electrophoresis: an added value to the analysis of CDT," *Clin. Chim. Acta*, vol. 483, no. March, pp. 256–262, 2018, doi: 10.1016/j.cca.2018.05.019.
- [96] H. Stibler, "Carbohydrate-deficient transferrin in serum: A new marker of potentially harmful alcohol consumption reviewed," *Clin. Chem.*, vol. 37, no. 12, pp. 2029–2037, 1991, doi: 10.1093/clinchem/37.12.2029.
- [97] A. Varki, "Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins," *Nature*, vol. 446, no. 7139, pp. 1023–1029, 2007, doi: 10.1038/nature05816.
- [98] A. Imberty and A. Varrot, "Microbial recognition of human cell surface glycoconjugates," *Curr. Opin. Struct. Biol.*, vol. 18, no. 5, pp. 567–576, 2008, doi: 10.1016/j.sbi.2008.08.001.
- [99] N. Roche, J. Ångström, M. Hurtig, T. Larsson, T. Borén, and S. Teneberg, "Helicobacter pylori and Complex Gangliosides," *Infect. Immun.*, vol. 72, no. 3, pp. 1519–1529, 2004, doi: 10.1128/IAI.72.3.1519-1529.2004.
- [100] T. Bertok *et al.*, "Analysis of serum glycome by lectin microarrays for prostate cancer patients - a search for aberrant glycoforms," *Glycoconj. J.*, 2020, doi: 10.1007/s10719-020-09958-4.
- [101] D. Pihikova *et al.*, "Sweet characterisation of prostate specific antigen using electrochemical lectin-based immunosensor assay and MALDI TOF/TOF analysis: Focus on sialic acid," *Proteomics*, vol. 16, no. 24, pp. 3085–3095, 2016, doi: 10.1002/pmic.201500463.
- [102] E. Llop *et al.*, "Improvement of Prostate Cancer Diagnosis by Detecting PSA Glycosylation-Specific Changes," *Theranostics*, vol. 6, no. 8, pp. 1190–1204, 2016, doi: 10.7150/thno.15226.
- [103] M. Coulet, O. Kepp, and G. Kroemer, "Metabolic Profiling of CHO Cells during the Production of Biotherapeutics," pp. 1–21, 2022.
- [104] M. E. Lalonde and Y. Durocher, "Therapeutic glycoprotein production in mammalian cells," *J. Biotechnol.*, vol. 251, pp. 128–140, 2017, doi: 10.1016/j.jbiotec.2017.04.028.
- [105] C. Raymond, A. Robotham, M. Spearman, M. Butler, J. Kelly, and Y. Durocher, "Production of  $\alpha$ 2,6-sialylated IgG1 in CHO cells," *MAbs*, vol. 7, no. 3, pp. 571–583, 2015, doi: 10.1080/19420862.2015.1029215.
- [106] F. L. Graham, J. Smiley, W. C. Russell, and R. Nairn, "Characteristics of a human cell line transformed by DNA from human adenovirus type 5," *J. Gen. Virol.*, vol. 36, no. 1, pp. 59–72, 1977, doi: 10.1099/0022-1317-36-1-59.
- [107] S. Rasheed, W. A. Nelson-Rees, E. M. Toth, P. Arnstein, and M. B. Gardner, "Characterization of a newly derived human sarcoma cell line (HT-1080)," *Cancer*, vol. 33, no. 4, pp. 1027–1033, 1974, doi: 10.1002/1097-0142(197404)33:4<1027::AID-CNCR2820330419>3.0.CO;2-Z.
- [108] S. Kallolimath *et al.*, "Engineering of complex protein sialylation in plants," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 34, pp. 9498–9503, Aug. 2016, doi: 10.1073/PNAS.1604371113/-/DCSUPPLEMENTAL/PNAS.201604371SI.PDF.
- [109] R. O'Kennedy, C. Murphy, and T. Devine, "Technology advancements in antibody purification," *Antib. Technol. J.*, vol. Volume 6, no. August, pp. 17–32, 2016, doi: 10.2147/anti.s64762.
- [110] R. B. Parekh *et al.*, "Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG."
- [111] G. A. W. Rook *et al.*, "Changes in IgG glycoform levels are associated with remission of arthritis during

- pregnancy," *J. Autoimmun.*, vol. 4, no. 5, pp. 779–794, 1991, doi: 10.1016/0896-8411(91)90173-A.
- [112] R. M. Anthony, F. Nimmerjahn, D. J. Ashline, V. N. Reinhold, J. C. Paulson, and J. V. Ravetch, "A RECOMBINANT IgG Fc THAT RECAPITULATES THE ANTI-INFLAMMATORY ACTIVITY OF IVIG,"
- [113] R. M. Anthony and J. V. Ravetch, "A novel role for the IgG Fc glycan: The anti-inflammatory activity of sialylated IgG Fcs," *J. Clin. Immunol.*, vol. 30, no. SUPPL. 1, pp. 9–14, 2010, doi: 10.1007/s10875-010-9405-6.
- [114] C. Dhar, A. Sasmal, and A. Varki, "From 'Serum Sickness' to 'Xenosialitis': Past, Present, and Future Significance of the Non-human Sialic Acid Neu5Gc," *Front. Immunol.*, vol. 10, no. April, p. 807, 2019, doi: 10.3389/fimmu.2019.00807.
- [115] M. Bardor, D. H. Nguyen, S. Diaz, and A. Varki, "Mechanism of uptake and incorporation of the non-human sialic acid N-glycolylneuraminic acid into human cells," *J. Biol. Chem.*, vol. 280, no. 6, pp. 4228–4237, 2005, doi: 10.1074/jbc.M412040200.
- [116] K. Banda, C. J. Gregg, R. Chow, N. M. Varki, and A. Varki, "Metabolism of vertebrate amino sugars with N-glycolyl groups: Mechanisms underlying gastrointestinal incorporation of the non-human sialic acid xeno-autoantigen N-glycolylneuraminic acid," *J. Biol. Chem.*, vol. 287, no. 34, pp. 28852–28864, 2012, doi: 10.1074/jbc.M112.364182.
- [117] M. O. Altman and P. Gagneux, "Absence of Neu5Gc and Presence of Anti-Neu5Gc Antibodies in Humans—An Evolutionary Perspective," *Front. Immunol.*, vol. 10, no. April, p. 789, 2019, doi: 10.3389/fimmu.2019.00789.
- [118] P. Tangvoranuntakul *et al.*, "Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 21, pp. 12045–12050, 2003, doi: 10.1073/pnas.2131556100.
- [119] M. Hedlund, V. Padler-Karavani, N. M. Varki, and A. Varki, "Evidence for a human-specific mechanism for diet and antibody-mediated inflammation in carcinoma progression," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 48, pp. 18936–18941, 2008, doi: 10.1073/pnas.0803943105.
- [120] Y. N. Malykh, R. Schauer, and L. Shaw, "N-Glycolylneuraminic acid in human tumours," *Biochimie*, vol. 83, no. 7, pp. 623–634, 2001, doi: 10.1016/S0300-9084(01)01303-7.
- [121] R. T. Prehn and L. M. Prehn, "The Autoimmune Nature of Cancer," *Cancer Res.*, vol. 47, no. 4, pp. 927–932, 1987.
- [122] Q. Zhou *et al.*, "Effect of genetic background on glycosylation heterogeneity in human antithrombin produced in the mammary gland of transgenic goats," *J. Biotechnol.*, vol. 117, no. 1, pp. 57–72, 2005, doi: 10.1016/j.jbiotec.2005.01.001.
- [123] R. N. Knibbs, I. J. Goldstein, R. M. Ratcliffe, and N. Shibuya, "Characterization of the carbohydrate binding specificity of the leucoagglutinating lectin from *Maackia amurensis*. Comparison with other sialic acid-specific lectins," *J. Biol. Chem.*, vol. 266, no. 1, pp. 83–88, 1991.
- [124] N. Shibuya, I. J. Goldstein, W. F. Broekaert, M. Nsimba-Lubaki, B. Peeters, and W. J. Peumans, "The elderberry (*Sambucus nigra* L.) bark lectin recognizes the Neu5Ac(alpha 2-6)Gal/GalNAc sequence.," *J. Biol. Chem.*, vol. 262, no. 4, pp. 1596–1601, 1987, doi: 10.1016/s0021-9258(19)75677-4.
- [125] Z. M. Khan *et al.*, "Crystallographic and Glycan Microarray Analysis of Human Polyomavirus 9 VP1 Identifies N-Glycolyl Neuraminic Acid as a Receptor Candidate," *J. Virol.*, vol. 88, no. 11, pp. 6100–6111, 2014, doi: 10.1128/jvi.03455-13.
- [126] C. S. Wright, "2.2 A Resolution Structure Analysis of Two Refined N-acetylneuraminyllactose-Wheat Germ Agglutinin Isolectin Complexes," 1990.
- [127] Mercy and Ravindranath, "Purification and characterization of lectin from *Lathyrus sativus*," *J. Plant Biochem. Biotechnol.*, vol. 5, no. 2, pp. 91–95, 1996, doi: 10.1007/bf03262989.

- [128] W. C. Wang and R. D. Cummings, "The immobilized leucoagglutinin from the seeds of *Maackia amurensis* binds with high affinity to complex-type Asn-linked oligosaccharides containing terminal sialic acid-linked  $\alpha$ -2,3 to penultimate galactose residues," *J. Biol. Chem.*, vol. 263, no. 10, pp. 4576–4585, 1988.
- [129] M. Guttman and K. K. Lee, "Site-Specific Mapping of Sialic Acid Linkage Isomers by Ion Mobility Spectrometry," *Anal. Chem.*, vol. 88, no. 10, pp. 5212–5217, 2016, doi: 10.1021/acs.analchem.6b00265.
- [130] K. M. Dziegielewska *et al.*, "The complete cDNA and amino acid sequence of bovine fetuin. Its homology with  $\alpha$ 2HS glycoprotein and relation to other members of the cystatin superfamily," *J. Biol. Chem.*, vol. 265, no. 8, pp. 4354–4357, 1990, doi: 10.1016/s0021-9258(19)39571-7.
- [131] S. Honda, S. Suzuki, S. Zaiki, and K. Kakehi, "Analysis of N- and O-glycosidically bound sialooligosaccharides in glycoproteins by high-performance liquid chromatography with pulsed amperometric detection," 1990.
- [132] E. D. Green, G. Adelt, J. U. Baenziger, S. Wilson, and H. Van Halbeek, "The asparagine-linked oligosaccharides on bovine fetuin. Structural analysis of N-glycanase-released oligosaccharides by 500-megahertz  $^1\text{H}$  NMR spectroscopy," *J. Biol. Chem.*, vol. 263, no. 34, pp. 18253–18268, 1988, doi: 10.1016/s0021-9258(19)81354-6.
- [133] R. Damiani *et al.*, "Stable expression of a human-like sialylated recombinant thyrotropin in a Chinese hamster ovary cell line expressing  $\alpha$ 2,6-sialyltransferase," *Protein Expr. Purif.*, vol. 67, no. 1, pp. 7–14, Sep. 2009, doi: 10.1016/j.pep.2009.04.005.
- [134] L. Landemarre, P. Cancellieri, and E. Duverger, "Cell surface lectin array: Parameters affecting cell glycan signature," *Glycoconj. J.*, vol. 30, no. 3, pp. 195–203, 2013, doi: 10.1007/s10719-012-9433-y.
- [135] S. J. North *et al.*, "Glycomics profiling of Chinese hamster ovary cell glycosylation mutants reveals N-glycans of a novel size and complexity," *J. Biol. Chem.*, vol. 285, no. 8, pp. 5759–5775, Feb. 2010, doi: 10.1074/jbc.M109.068353.

**Kits LEctPROFILE : vers un contrôle de qualité et de nouvelles applications potentielles**

La glycosylation affecte les propriétés des molécules biothérapeutiques, des médicaments fabriqués dans des systèmes vivants comme les cellules de mammifères, les bactéries ou les levures. La sialylation dans les produits biothérapeutiques peut affecter la demi-vie sérique et accélérer son élimination de l'organisme et peut également modifier les fonctions effectrices des anticorps monoclonaux. La glycosylation des produits biothérapeutiques dépend strictement du système d'expression et du processus de production: l'utilisation d'un système d'expression hôte mammifère non humain est très pratique car elle donne aux médicaments un schéma de glycosylation proche de celui de l'homme, mais elle a ses inconvénients. Les cellules CHO n'expriment que des acides sialiques liés par une liaison  $\alpha$ 2-3 et les cellules murines peuvent ajouter des épitopes glycaniques immunogènes tel que l'antigène Galili et également du Neu5Gc. Les principales techniques utilisées pour l'analyse des glycanes dans la production industrielle, comme la MS et l'HPLC, sont très précises, mais elles peuvent prendre du temps et exiger un haut niveau d'expertise. L'idée principale a été ici de développer un outil complémentaire, basé sur la technologie GLYcoPROFILE®, qui facilite au fabricant industriel le contrôle de la qualité d'une glycoprotéine recombinante thérapeutique au stade le plus précoce possible de la recherche, puis à chaque étape du développement et de la production pour établir un schéma de contrôle des glycosylations tout au long du développement, mais également de la production du futur biomédicament. Parmi les principales caractéristiques des kits LEctPROFILE, il y a la haute spécificité pour les glycanes d'intérêt: c'est un outil qui aide à discriminer les liaisons d'acide sialique et à distinguer les glycanes indésirables tels les motifs Galili et Neu5Gc. Ceci a été rendu possible grâce à la purification des lectines avec une haute sélectivité pour la partie du glycanne cible. En outre, le test doit être rapide, à haut débit et doit permettre à des chercheurs non spécialisés de réaliser l'analyse. Afin de développer ce type de test, nous avons synthétisé des néoglycoprotéines qui fonctionnent comme des sondes spécifiques pour l'étude des performances des kits. Nous avons établi un protocole de production par lots des cinq principales néoglycoprotéines d'intérêt, et nous avons démontré leur pureté, leur activité et leur spécificité vis-à-vis des lectines cibles. Ces molécules standard permettent d'établir des références relatives à la sélectivité des lectines, la limite de détection des kits et l'IC50 de différentes molécules de référence. Au cours de la thèse, l'idée d'application des kits s'est élargie: le même principe d'un screening rapide et facile à utiliser peut être appliqué à d'autres domaines. Par exemple, l'antigène Galili est un obstacle à la transplantation d'organes de porc chez l'homme. À ce jour, il n'existe pas de dépistage rapide et à haut débit capable de détecter la présence de l'antigène Galili sur les tissus et les organes transplantés. Cette lacune peut être comblée en construisant un réseau de lectines afin de détecter cet antigène sur les produits biothérapeutiques et les échantillons biologiques. L'altération de la sialylation peut conduire à des états pathologiques et elle est bien connue comme une caractéristique du cancer. De plus, le changement du rapport  $\alpha$ 2-3/ $\alpha$ 2-6 est impliqué dans le tropisme des agents pathogènes. Un taux important de l'acide sialique NeuGc associé à la présence d'anticorps anti-NeuG est également décrit pour certains types de cancer, y compris les carcinomes du côlon, les rétinoblastomes, les cancers du sein et les mélanomes. Les kits LEctPROFILE axés sur l'acide sialique peuvent permettre une discrimination entre les deux liaisons ou entre Neu5Ac et Neu5Gc afin de fournir une semi-quantification et de disposer d'un outil rapide pour la découverte de biomarqueurs potentiels.

Mots clés : lectines, lectin array, biothérapies, glycobiologie, antigène Galili, acide sialique, néoglycoprotéines

## LEctPROFILE kits: towards quality control and new potential applications

Glycosylation is one of the most important post-translational modifications of proteins: it affects folding, stability, molecular recognition and functionality. It also influences the properties of biotherapeutics, medicinal products manufactured in living systems like mammalian cells, bacteria, yeast. It has been shown that lack of sialylation in biotherapeutics can affect the serum half-life of erythropoietin and accelerate its clearance from the body. It can also modify the monoclonal antibodies effector functions. The presence of sugar moieties like Gal $\alpha$ 1,3Gal (Galili antigen) and glycolyl neuraminic acid (Neu5Gc), can make a medicine unsafe, triggering an immune response from the human's body. Biotherapeutics glycosylation strictly depends on the expression system and on the manufacturing process: the use of non-human mammalian host expression system like Chinese hamster ovary cells and murine myeloma cells is very convenient because it gives the drugs a human-like glycosylation pattern, but it has its downsides. CHO only sialylate in the  $\alpha$ 2,3 linkage affecting the functionality of the biotherapeutic molecules and murine cells can add the immunogenic epitopes Galili antigen and Neu5Gc. When it comes to glycan analysis in industrial production, the main used techniques like mass spectrometry and HPLC are very accurate but they can be time-consuming and require a high-level of knowledge and expertise. The main idea has been to develop a tool that makes easy to the industrial manufacturer the quality control of their biotherapeutics at the earliest possible stage of research, then each stage of the development and production: a lectin array based on the GLYcoDiag's GLYcoPROFILE® technology. Among the main features of the LEctPROFILE kits there is the high-specificity for the glycans of interest: it is a tool that helps to discriminate the sialic acid linkages and to distinguish the unwanted glycans Galili and Neu5Gc. This is achieved thanks to the purification of lectins with high-selectivity for the target glycan moiety. In addition, it has to be fast, high-throughput and it has to allow non-specialized researchers to be able to perform the analysis. In order to develop this kind of assay, we have synthesised neoglycoproteins that work as specific probes for the study of the kits performances. To do so we have established a protocol for batch production of five main neoglycoproteins of interest, and we have demonstrated their purity, activity and specificity towards the target lectins. They allow to understand the lectins selectivity, the detection limit of the kits and the IC50 of different reference molecules. During the course of the thesis, the idea of application of the kits has broadened: the same principle of a fast and easy-to use screening can be applied to other fields than the quality control of biotherapeutics. For example the Galili antigen is an obstacle for the transplantation of organs from pig into humans. To date, there is no a fast, high-throughput screening able to detect their presence on transplanted tissues and organs. This gap can be addressed by building a lectin array in order to detect this antigen on biotherapeutics and biological samples. Another example is the alteration of sialylation that could lead to pathological states and it is well-known as a hallmark of cancer. Moreover, the changing of the ratio  $\alpha$ 2,3/ $\alpha$ 2,6 linkage is involved in pathogens tropism and It is also described that certain types of cancer, including colon carcinomas, retinoblastomas, breast cancers, and melanomas, show an increase of Neu5Gc and Neu5Gc antibodies. The LEctPROFILE kits focused on sialic acid can give a discrimination between the two linkages or between Neu5Ac and Neu5Gc in order to provide a semi-quantification and have a rapid tool for potential biomarker discovery.

Key words: lectins, lectin array, biotherapeutics, glycobiology, Galili antigen, sialic acid, neoglycoproteins



**GLYcoDiag, 2 Rue du Cristal, 40100 Orléans**

