UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI SCIENZE E TECNOLOGIE

**Dipartimento di Chimica**
**Scuola di Dottorato in Chimica**
**Ciclo XXX**
**Anno academico: 2017/2018**

**PhD thesis**

# Synthesis of selective C-type lectin antagonist glycomimetics

**LAURA MEDVE**
Matr. Nr. R11166

Tutor: Prof. Anna Bernardi
Coordinator: Prof. Emanuela Licandro

IMMUNOSHAPE
Marie Skłodowska-Curie Innovative Training Network
H2020-MSCA-ITN-2014-ETN-642870

# TABLE OF CONTENTS

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **AAL** | ***Aleuria aurantia* lectin** |
| **ADIBO** | **azadibenzocyclooctyne** |
| **ADMET** | **adsorption-distribution-metabolism-excretion-toxicity** |
| **ADME** | **adsorption-distribution-metabolism-excretion** |
| **AIDS** | **acquired immunodeficiency syndrome** |
| **AIEC** | **adherent invasive *Escherichia coli*** |
| **APC** | **antigen presenting cell** |
| **ASGPR** | **asialoglycoprotein receptor** |
| **AUC** | **analytical ultracentrifugation** |
| **AZBCN** | **azidopropanol-clicked BCN-amine** |
| **BCN** | **bicyclononyne** |
| **BDCA2** | **blood dendritic-cell antigen-2** |
| **BSA** | **bovine serum albumin** |
| ***C. (albicans)*** | ***Candida (albicans)*** |
| **CD** | **cluster of differentiation** |
| **cDNA** | **complementary deoxyribonucleic acid** |
| **CFG** | **Consortium for Functional Glycomics** |
| **CLEC1** | **C-type lectin-like receptor 1** |
| **CLR** | **C-type lectin receptor** |
| **CNS** | **central nervous system** |
| **ConA** | ***Concanavalin A*** |
| **CRD** | **carbohydrate recognition domain** |
| **CTLD** | **C-type lectin-like domain** |
| **CTLF** | **C-type lectin-like fold** |
| **CuAAC** | **copper(I)-catalyzed alkyne-azide cycloadditions** |
| **Cy3** | **cyanine 3** |
| **DAMP** | **damage-associated molecular patterns** |
| **DBU** | **1,8-Diazabicyclo[5.4.0]undec-7-ene** |
| **DC** | **dendritic cell** |
| **DCAR** | **Dendritic cell immunoactivating receptor** |
| **DCAL** | **Dendritic-cell-associated C-type lectin** |
| **DCIR** | **dendritic cell immunoreceptor** |
| **DCM** | **dichloromethane** |
| **DC-SIGN** | **dendritic cell-specific intercellular adhesion molecule-3 grabbing nonintegrin** |
| **DC-SIGNR** | **DC-SIGN related molecule** |
| **Dectin-2** | ***dendritic cell*-associated C-type lectin-2** |
| **DHB** | **dihydroxybenzoic acid** |
| **DIFO** | **difluorinated cyclooctyne** |

| | |
|---|---|
| **DIPEA** | **N,N-Diisopropylethylamine** |
| **DLS** | **dynamic light scattering** |
| **DMF** | **Dimethylformamide** |
| **DMSO** | **Dimethyl sulfoxide** |
| **DNA** | **deoxyribonucleic acid** |
| **DNGR-1** | **dendritic cell receptor-1** |
| **DOL** | **degree of labeling** |
| **ECD** | **extracellular domain** |
| **ECFP4** | **extended connectivity fingerprint 4** |
| **EDC** | **N-(3-dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride** |
| **EDTA** | **Ethylenediaminetetraacetic acid** |
| **ELLA** | **enzyme linked lectin assay** |
| **EPN** | **Glu-Pro-Asn tri-aminoacidic motif** |
| **ERAD** | **endoplasmic reticulum-associated protein degradation** |
| **ESI** | **electrospray ionisation** |
| **Fas** | **first apoptosis signal** |
| **Fc** | **fragment crystallizable** |
| **FCFP4** | **functional-class fingerprint 4** |
| **FcRγ** | **Fc receptor γ-subunit** |
| **FITC** | **fluorescein isothiocyanate cadaverine** |
| **FMO** | **frontier molecular orbitals** |
| **GAG** | **glucosaminoglycan** |
| **GBP** | **glycan binding protein** |
| **GMI-1070** | **Rivipansel** |
| **GMP** | **good manufacturing practice** |
| **GNP** | **gold nanoparticle** |
| **GPI** | **glycosylphosphatidylinositol** |
| **HA** | **hemagglutinin** |
| **HCV** | **hepatitis C virus** |
| **HDM** | **house dust mite** |
| **HGP** | **Human Genome Project** |
| **HIV** | **Human immunodeficiency virus** |
| **HOMO** | **highest occupied molecular orbital** |
| **HRMS** | **high-resolution mass spectrometry** |
| **HSQC NMR** | **Heteronuclear single quantum correlation spectroscopy** |
| **ICAM** | **intercellular adhesion molecules** |
| **IFN-γ** | **interferon-γ** |
| **IgG1** | **immunoglobulin G-1** |
| **IL** | **interleukin** |
| **IRF** | **interferon regulatory factor** |
| **ITAM** | **immunoreceptor tyrosine-based activation motif** |
| **ITC** | **Isothermal titration microcalorimetry** |
| **ITO** | **indium tin oxide** |

| | |
|---|---|
| **LAM** | **lipoarabinomannan** |
| **$Le^X$, $Le^Y$, $Le^A$, $Le^B$, $sLe^X$** | **Lewis-type and sialylated Lewis-type antigens** |
| **LCA** | ***Lens culinaris* agglutinin** |
| **LC** | **Langerhans cell** |
| **LFA** | **lymphocyte function-associated antigen** |
| **LMWH** | **low molecular weight heparin** |
| **LNFP-III** | **Lacto-*N*-fucopentaose III** |
| **LSECtin** | **lymph node sinusoidal endothelial cell C-type lectin** |
| **L-SIGN** | **DC-SIGNR (DC-SIGN-related receptor)** |
| **LUMO** | **lowest unoccupied molecular orbital** |
| **Lys313** | **lysine residue** |
| **MAG** | **myelin-associated glycoprotein** |
| **MALDI** | **matrix assisted laser desorption ionization** |
| **MBL** | **mannose-binding lectin (MBP)** |
| **MBP** | **mannose/mannan-binding protein (MBL)** |
| **MCL** | **macrophage C-type lectin** |
| **mCPBA** | ***meta*-chloroperbenzoic acid** |
| **MD** | **molecular dynamics** |
| **mDCIR2** | **mouse dendritic cell inhibitory receptor 2** |
| **MeCN** | **acetonitrile** |
| **MeOH** | **methanol** |
| **MHC** | **major histocompatibility complex** |
| **MGL** | **macrophage galactose lectin** |
| **Mincle** | **macrophage-inducible C-type lectin** |
| **MM** | **molecular mechanics** |
| **MR** | **mannose receptor** |
| **MRI** | **magnetic resonance imaging** |
| **mRNA** | **messenger ribonucleic acid** |
| **MS** | **mass spectrometry** |
| ***M. (tuberculosis)*** | **Mycobacterium (tuberculosis)** |
| **MW** | **molecular weight** |
| **NDP52** | **Nuclear Dot Protein 52 kDa** |
| **NGL** | **neoglycolipid** |
| **NGP** | **neo-glycoproteins** |
| **NHS** | ***N*-hydroxysuccinimide** |
| **NK** | **natural killer cell** |
| **NMR** | **nuclear magnetic resonance** |
| **NOD-like receptor, NLR** | **nucleotide-binding oligomerization domain-like receptor** |
| ***P. aeruginosa*** | **Pseudomonas aeruginosa** |
| **PA-IL – LecA** | ***Pseudomonas aeruginosa* lectin A** |
| **PA-IIL – LecB** | ***Pseudomonas aeruginosa* lectin B** |
| **PAINS** | **pan-assay interference compounds** |
| **PAMP** | **pathogen-associated molecular patterns** |

| | |
|---|---|
| **PBS** | **phosphate buffer saline** |
| **PCR** | **Polymerase chain reaction** |
| **PD** | **pharmacodynamic** |
| **PDB** | **protein data bank** |
| **PEG** | **poly(ethylene glycol)** |
| **PIM** | **phosphatidylinositol-mannoside** |
| **PK** | **pharmacokinetic** |
| **PMT** | **photomultiplier tube** |
| **PNA** | ***Peanut agglutinin* lectin** |
| **PNP** | ***para*-nitrophenyl ester** |
| **PRR** | **pattern-recognition receptors** |
| **PSA** | ***Pisum sativum* agglutinin** |
| **QPD** | **Gln-Pro-Asp tri-aminoacidic motif** |
| **RCA120** | ***Ricinus communis* agglutinin 120** |
| **RDA** | **reporter displacement assays** |
| **RegIIIa** | **hepatointestinal pancreatic/pancreatitis associated protein, HIP or PAP** |
| **RFU** | **relative fluorescence unit** |
| **rhMBL** | **recombinant human** |
| **RIG-I-like receptors** | **retinoic acid-inducible gene I receptors** |
| **RNA** | **ribonucleic acid** |
| **ROS** | **reactive oxigen species** |
| **RT** | **room temperature** |
| **RT-PCR** | **reverse transcription-polymerase chain reaction** |
| **RU** | **resonance or response units** |
| ***S. (mansoni)*** | ***Schistosoma (mansoni)*** |
| **SAR** | **structure and activity relationships** |
| **SARS-CoV** | **Severe Acute Respiratory Syndrome Coronavirus** |
| **SDS-PAGE** | **Sodiumdodecyl sulfate polyacrylamide gel electrophoresis** |
| **SEC** | **Size Exclusion Chromatography** |
| **SIV-1** | **Simian Immunodeficiency Virus-1** |
| **SLE** | **systemic lupus erythematosus/supported liquid extraction** |
| **SLS** | **static light-scattering** |
| **SNA** | ***Sambucus nigra* agglutinin** |
| **SPAAC** | **strain-promoted alkyne-azide cycloaddition** |
| **SPR** | **surface plasmon resonance** |
| **SRCL** | **scavenger receptor C-type lectin** |
| **STD** | **saturation-transfer difference** |
| **TAMP** | **tumor-associated molecular patterns** |
| **TBTA** | **(tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine)** |
| **TCA** | **trichloroacetoimidate** |
| **TCR** | **T-cell receptors** |
| **TEA** | **triethylamine** |
| **TEM** | **transmission electronmicroscopy** |

| | |
|---|---|
| **TFA** | **trifuoroacetic acid** |
| **THF** | **tetrahydrofuran** |
| **TLC** | **thin-layer chromatography** |
| **TLR** | **Toll-like receptor** |
| **TMSOTf** | **Trimethylsilyl trifluoromethanesulfonate** |
| **TOF** | **time-of-flight** |
| **TNF-α** | **tumour necrosis factor-α** |
| **trNOESY** | **transferred nuclear Overhauser-effect spectroscopy** |
| **Trp182** | **tryptophan** |
| **TS** | **transition state** |
| **UPEC** | **uropathogenic *Escherichia coli*** |
| **WGA** | **wheat germ agglutinin** |
| **WND** | **Trp-Asn-Asp** |
| **$\Theta_{SPR}$** | **surface plasmon resonance angle** |

# OVERVIEW

This thesis discusses a glycomimetic approach to selectively target C-type lectin receptors.

In the **Introduction**, the function, biological importance and characteristics of C-type lectin receptors are explained. Several aspects of interactions with natural and synthetic ligands are discussed, in order to introduce the Reader to the challenges of C-type lectin targeting. Closely related to this, a general summary of carbohydrate ligands is offered, with particular emphasis on their pharmaceutical relevance. Later, the advantages of glycomimetic ligands are highlighted, with a detailed overview of the molecules developed in our group.

**Part I** is divided into two chapters. **Chapter I** describes the design and synthesis of a mannose- and fucose-based glycomimetic ligand library.
Earlier results indicated that modification of a mannosylated glycomimetic scaffold can increase affinity towards C-type lectin receptors, and, at the same time, generate selectivity against others. Therefore, a glycomimetic library was built on this scaffold, with the purpose of screening various C-type lectin receptors. The ligands carry mannose or fucose as a natural sugar residue, and they are diversified with amide substituents, selected with the help of chemoinformatic tools. We anticipated that primary screenings of the library with C-type lectins will be able to identify high affinity and possibly selective ligands for further elaboration.

The microarray-based screening with different C-type lectins is discussed in **Chapter II**.
During their synthesis, ligands were equipped with a linker moiety to allow their presentation on the microarray chips, or later, in multivalent systems. Microarrays are powerful tools to study lectin-carbohydrate interactions, as their size allows hundreds of simultaneous assays, requiring only minute amounts of the protein. The first part of this work focused on the optimization of printing conditions and validation of the screening setup with lectins of known carbohydrate specificity. Once optimal conditions were established, the library was screened against various human C-type lectin receptors, available through the Immunoshape Network. Promising results were obtained whose confirmation and elaboration is underway.

**Part II** consists of a single chapter (**Chapter III**) that describes the design and synthesis of high affinity glycomimetic antagonists of the dendritic cell-specific intercellular adhesion molecule-3 grabbing nonintegrin (DC-SIGN).
This receptor is of particular interest among C-type lectins, as it promotes numerous viral (HIV, HCV, Ebola, Dengue), bacterial (*Mycobacterium tuberculosis*) and parasital (*Schistosoma mansoni*) infections. Based on the X-ray crystal structure of a glycomimetic ligand/DC-SIGN complex, a virtual fragment screening identified a new binding pocket in the vicinity of the ligand. This pocket can be reached with modifications of the glycomimetic molecules, and fragment-like structures could engage in favourable interactions at this site. These additional contacts could significantly increase the overall affinity of glycomimetic ligands towards DC-SIGN, therefore we set off to explore interactions around the binding site. The results obtained open up new perspectives for glycomimetic ligand design and validate previous models.

To conclude, the thesis finishes with a brief summary of results and an outlook on future directions in this research.

# 1. INTRODUCTION – C-TYPE LECTIN RECEPTORS AND THEIR CARBOHYDRATE LIGANDS

The immune system is undoubtedly one of the most marvellous machineries ever crafted by Nature to protect the living organism. A healthy immune system, just like an impregnable fortress, shields its owner from malignant external invaders by comprising an entire apparatus of complex recognizing, signalling and executing mechanisms. These are the fundamental steps of discrimination between self and non-self structures and the latters' neutralization or destruction – the core tasks in host defence against pathogens. The immune system is capable of more than fighting against foreign species, though. It can memorize and remember previously encountered harmful structures to trigger an adequate response more rapidly on a following confrontation; and it is responsible for the detection and elimination of potentially dangerous abnormal self-cells. Because of the closely interacting partners, and biological processes relying on each other, even the smallest dent in the walls of this fort can start a cascade of malfunctioning and expose the individual to infections and diseases.
When a pathogen – organism capable of causing infection – enters the body, the evolutionarily older *innate immune system* represents the first line of defence. This immediate response is non-specific: it recognizes primarily the molecular patterns present on viruses, bacteria, fungi and parasites – the ancient miniature-sized threats for higher level organisms – and destroys them. The *adaptive immune system* developed upon the cellular and humoral elements of the innate system to recognize and learn specific structures and to adapt the immune response, thus enhancing the efficiency and storing the information for a recurring encounter. During the first exposure to a new pathogen, the adaptive response is delayed, but during this period a more specific response is set in motion and an immunological memory is acquired, which will facilitate a faster, improved reaction to a similar structure upon a next meeting.
The two levels of immune defence are tightly linked, complementing each other. The first cells to confront antigens, the antigen presenting cells (APCs), belong to the innate system, but they also set off the adaptive pathway by presenting the antigens (or their fragments) to T-cells - the main mediators in the cellular immunity - and to B-lymphocytes - involved in the humoral system.

## 1.1 THE CELL-MEDIATED IMMUNE RESPONSE

Non-self, damaged-self or altered-self associated structures within the organism are detected by sentinels of the immune system, the antigen presenting cells. APCs internalize and process the antigens - either by completely demolishing them, or by chopping them into smaller bits – and present these antigenic fragments to T-lymphocytes. T-cells, however, do not become activated by „free" antigens, they only recognize antigens that are bound to the major histocompatibility complex (MHC) – a group of cell-surface proteins that mediate the immune signalling. The two major classes of MHC molecules relevant for the display of antigens are MHC class I and MHC class II. The first group can be found on the cell-surface of almost every nucleated cell of the body and the antigens exhibited in the complex are typically endogenous peptides from the degradation of cytosolic proteins (*endogenous pathway*). In contrast, MHC class II expression is a characteristic of specialized APCs, who are in charge of displaying degraded extracellular proteins to T-cells (*exogenous pathway*), in addition to intracellular peptides.

T-cells express on their surface antigen-specific T-cell receptors (TCR), which can interact with antigen-MHC complexes of both classes, depending on the presence of additional co-stimulatory glycoproteins on the cell surface. The antigen-MHC class I complex on the surface of the cell membrane is exclusively recognized by cytotoxic T-cells, also known as $CD8^+$ cells. These white blood cells can immediately kill the signalling tumoral, infected or otherwise damaged cells. MHC class II-bound antigens are presented to another subgroup of T-cells, the T-helper cells (also referred to as $CD4^+$ cells), which - upon activation - initiate distinct processes crucial in the adaptive immune defence. Activated T-cells first proliferate, then differentiate into effector cells or memory cells, and simultaneously release various cytokines and chemokines which regulate, stimulate or suppress the immune signalling.[1] (**Figure 1**)

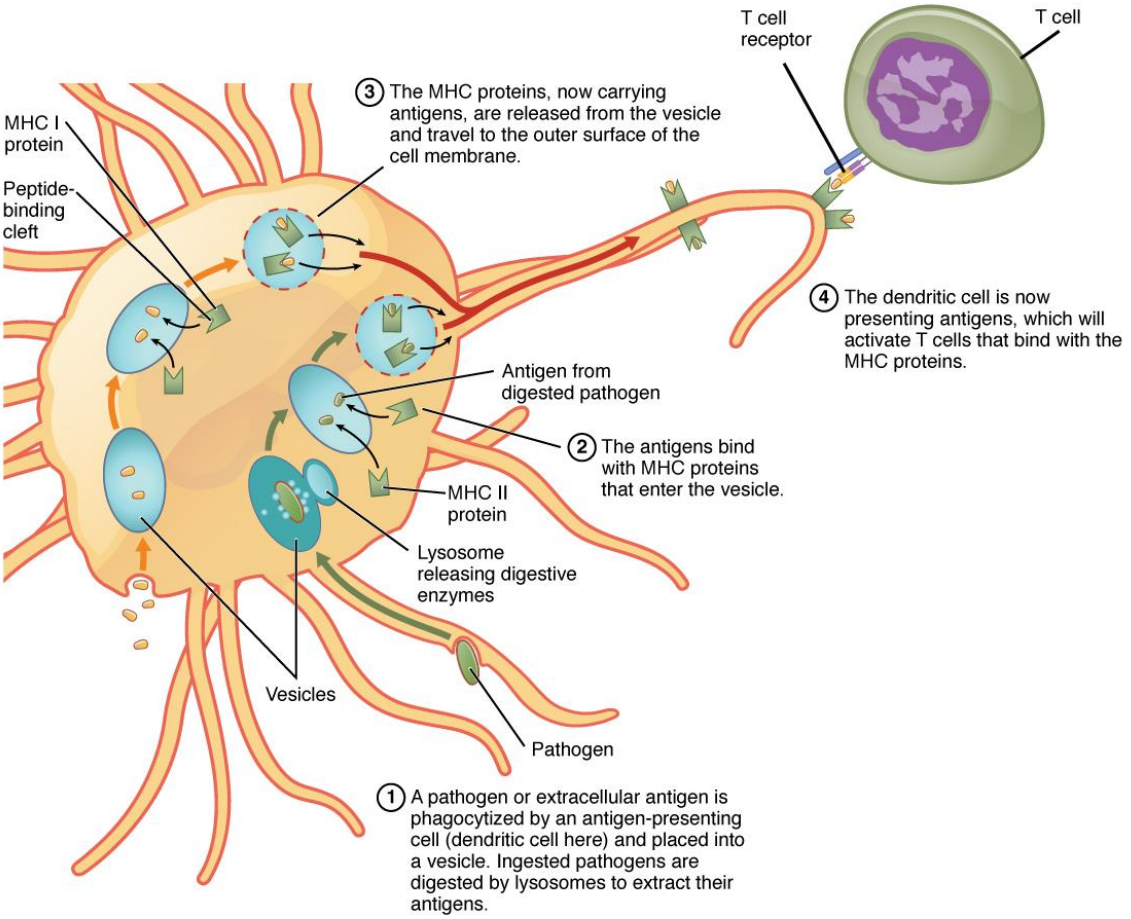



***Figure 1.*** *Schematic illustration of pathogen internalisation, processing and presentation by APCs. APCs encounter antigens from damaged, apoptic or infected cells ready to be presented to T-cells; or pathogens that professional APCs first proteolyze to different degrees of degradation in the phagolysosome. The antigens are then loaded onto MHC class I or class II molecules (depending on the cell type) and are expressed on the cell surface. Various subtypes of T-cells are able to recognize the antigen-MHC complex by their specific T-cell receptors and use additional glycoproteins and co-stimulation for the interaction. Modified from: Anatomy and Physiology. OpenStax CNX, 2016.*[2]

APCs can be classified into two groups. Non-professional or amateur APCs can serve as occasional antigen-presenters, but have otherwise different major functions. Essentially all nucleated cells in animals express MHC class I molecules coupled with β-2-microglobulin, to be able to activate cytotoxic T-cells. When intracellular proteins are degraded by the proteasome, the smaller peptide chains are loaded onto MHC class I molecules in the endoplasmic reticulum and the complexes are inserted into the cell membrane.
The antigens originating from the normal protein turnover of the cell will be tolerated by T-cells, but unfamiliar peptides will lead to the activation of cytotoxic T-cells for cell-destruction. When the antigen is presented on MHC class I molecules, the complex is recognized by the specialized T-cell receptor on the surface of the cytotoxic T-cell membrane. However, for active signalling another co-receptor on the cell-surface of cytotoxic T-cells is also essential. CD8 (cluster of differentiation 8) glycoproteins support the MHC class I-antigen-TCR interaction by binding with their extracellular IgV-like domain to the $\alpha_3$ portion of the MHC class I constant region (**figure 2.A**) – cytotoxic T-cells are therefore also called $CD8^+$ T-cells. This collective interaction sends off a primary signal (signal-1) to the T-cells which alone is not sufficient to activate their cytotoxic function.
Signal-2 is a confirmation that the T-cells encountered a genuine foreign antigen, and do not induce auto-immune response against host antigens. CD28 is a co-receptor on the cell surface of naïve T-cells that interacts with B7.1 (CD80) and B7.2 (CD86) ligands – proteins, whose expression on APCs is upregulated after the detection of antigens. In lack of any of the two interactions, T-cells are not activated.
When the two signals together stimulate cytotoxic T-cells, anti-tumoral, anti-viral or anti-microbial cytokines are secreted, most importantly TNF-α (tumour necrosis factor-α) – an "acute phase reaction" inducer with pro-inflammatory and apoptotic effects; and IFN-γ (interferon-γ) – that activates macrophages and promotes MHC class II expression. At the same time, cytotoxic granules are released from the $CD8^+$ T-cells, which contain perforin and granzymes. The close proximity of the two cells and the tight immune synapse ensure that the cytotoxic species are targeting only the signalling cell and not healthy ones nearby. Lastly, an interaction of the Fas(first apoptosis signal)-receptor (expressed by the target cell) and the Fas-ligand (found on the cytotoxic T-cells) activates caspases - a set of protease enzymes which initiate programmed cell death, eventually killing the abnormal cell. (**Figure 2.B**) Cytotoxic T-cells are able to perform several rounds of this recognition-elimination cycle during their lifespan.

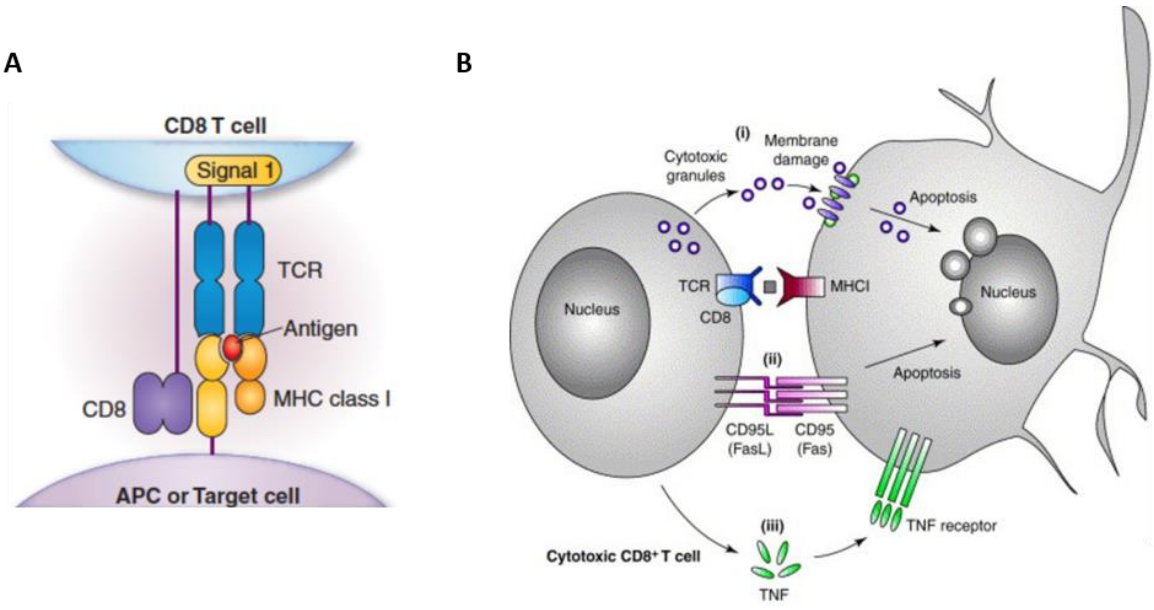



***Figure 2.*** *A. Antigen presented on MHC class I protein complex (with β-2-microglobulin) to a cytotoxic T-cell. CD8 glycoprotein, expressed on the CD8⁺ T-cell surface strengthens the interaction by connecting with the constant region of the MHC molecule.*

*B. CD8⁺ T-cell activation by APC signalling and initiation of apoptosis. Antigen presented by MHC class I molecules is recognized by a specific T-cell receptor, and after a verifying CD28-mediated co-stimulation (not shown in the figure), the cytotoxic T-cell begins its effector function. Fas-FasL binding activates caspases. TNF-a and cytotoxic granules are released towards the tumoral, infected or damaged cell. Modified from: Neumann et al.*[3]

As opposed to amateur APCs, professional APCs are fully committed to their role in the immune system and this class of immune cells include monocytes and macrophages, dendritic cells (DCs) and B-cells. Similarly to other cell types in the phagocyte family, mast cells and neutrophils exhibit professional antigen-presenting capacity like macrophages and DCs, but only after activation.

Professional APCs ingest the pathogen by phagocytosis, fluid phase pinocytosis or receptor-mediated endocytosis and then degrade it to different levels by various mechanisms. Macrophages, for example, engulf the detected invaders in phagosomes, which vacuoles later mature by joining lysosomes – and create an acidic organelle, where the pathogen proteins are degraded by hydrolysis. This powerful proteolysis is facilitated not only by the low pH, but also by the presence of protease enzymes, so that the peptide chains are destroyed to the constituent amino acid level.

On the contrary, in DCs - although they also have phagocytic activity until their maturation and express a wide range of proteases - the degradation of internalized material is significantly less efficient than in macrophages. The remaining larger structures, however, preserve the antigenic information of the original pathogen, making DCs, under normal circumstances, the most potent subset of phagocytes to stimulate naïve T- and indirectly B-cells - forming the link between the innate and adaptive immune response.

DCs originate from the bone marrow, but then spread all throughout the peripheries, like the skin, oral-, nasal-, pulmonary-, genital- and gastrointestinal mucosa - practically every lymphoid and most of the non-lymphoid tissue. DCs, just like macrophages, display on their surface numerous phagocytic receptors, such as Toll-like receptors (TLR), scavenger receptors, retinoic acid-inducible gene I receptors (RIG-I-like receptors), nucleotide-binding oligomerization domain-like receptors (NOD-like receptors, NLRs), and C-type lectin receptors (CLRs) to capture pathogens, infected cells or apoptosing cells. These receptors belong to the pattern-recognition receptors (PRR) and are exceptionally good at detecting unknown molecular fingerprints among healthy cells. They play the key role in distinguishing between healthy and non-self- (pathogen-associated molecular patterns - PAMP), damaged-self- (damage-associated molecular patterns - DAMP) or

altered-self (tumor-associated molecular patterns - TAMP)-related structures in the organism. In addition to their antigen-receptors, DCs also present several adhesion molecules which act as co-stimulatory factors during T-cell activation and facilitate DC migration within the lymphatics and their homing after they accomplished T-cell activation in the lymph nodes. These molecules include intercellular adhesion molecules (ICAM) 1, 2 and 3; members of the TNF family; the previously mentioned CD80 and CD86 for signal-2 initiation and integrin proteins, like LFA-1 and LFA-3 (lymphocyte function-associated antigen 1 and 3). Apoptotic cell phagocytosis is restricted to the immature phase of DC lifespan and induces the expression of specialized receptors, like $\alpha_v\beta_5$ integrin and CD36.

When surface PRRs capture unfamiliar structures originating from pathogens, or when DCs receive inflammatory signals, their maturation begins. DCs have a set of sophisticated control mechanisms to lower the phagosomal activity and acidification level in their endo- or phagolysosomes. Unlike macrophages, they digest proteins only partially - to fragments of some 10 amino acids, which antigenic fractures contain the information for the specific immune response and immunological memory of the adaptive immune system. The epitopic peptide antigens are then conjugated to MHC class II molecules in the phagosomal or endosomal compartments and are transported to the DC cell surface via the endoplasmic reticulum. Mature DCs lose their phagocytic capability and migrate to secondary lymphoid organs, such as lymph nodes, tonsils or the spleen for T- and B-lymphocyte activation. At the same time, other DCs are attracted to the site of pathogen entry.

In the secondary lymphoid tissue, the antigens on DCs are presented in conjunction with the MHC class II molecules to T-helper cells – who will later orchestrate an entirety of coordinated mechanisms in the immune response. The initial interaction is formed between the antigen-MHC complex's $\alpha 1/\beta 1$ domain and the specific T-cell receptor. In this case, the co-receptor strengthening the bond is the glycoprotein CD4, expressed on T-helper cells, thus these T-lymphocytes are also called $CD4^+$ T-cells. CD4 binds with its D1 region to the $\beta 2$ region of the MHC class II molecule. (**Figure 3.A**) This alone triggers only a set of intracellular pathways though; for the complete activation and proliferation a synergistic effect of signal-1 and the aforementioned signal-2 must occur. Without signal-2, T-helper lymphocytes become anergic, and lose their capacity to induce immune response until their apoptosis. After the first time of receiving both signals simultaneously, signal-1 will be enough in the future to activate antigen-experienced T-cells, and this is valid also after their differentiation into T-memory cells. In other words, during the next encounter with the same antigen, T-cell receptor detection of the antigen-MHC molecule complex will activate T-effector cells more rapidly than the first time.

The two signals lead to T-helper cell proliferation, although different receptor signals from the APC produce different T-helper cell subsets. This is achieved by the release of different cytokines both by the DC and the T-helper cell itself, and one of the major initiators of T-helper cell proliferation is IL-2 (interleukin-2). Numerous other cytokines and chemokines take part in the communication and steer the proliferation in certain directions. Eventually, the activated T-helper cells differentiate into various subsets of T-helper cells, T-effector cells and T-memory cells; stimulate macrophages, natural killer cells (NKs), neutrophils, eosinophils and basophils; and instruct B-cells to produce specific antibodies. (**Figure 3.B**)

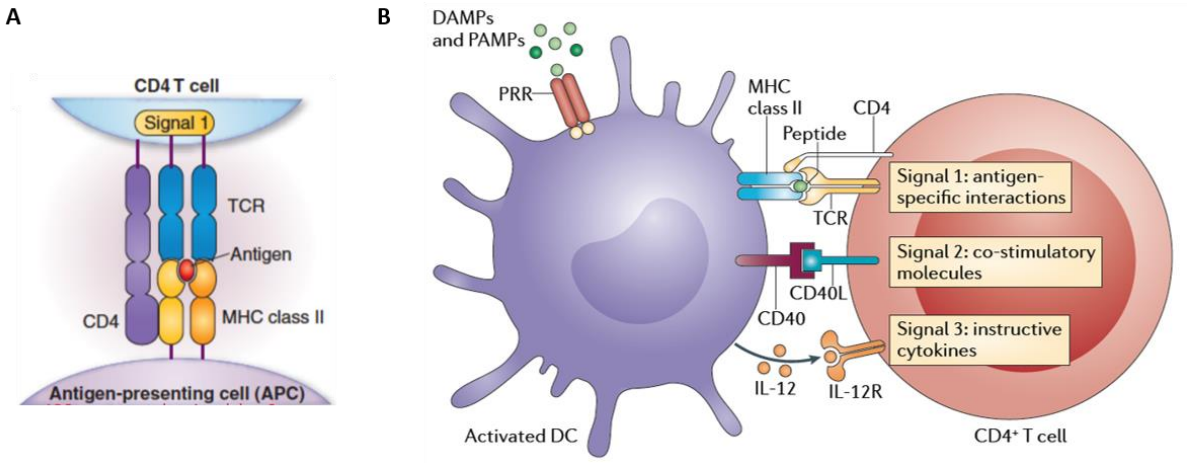



***Figure 3.*** *A. Antigen presented on MHC class II protein complex to a T-helper cell. The specific T-cell receptor binds to the α1/β1 domain and CD4 binds to the β2 region of the MHC class II molecule*

*B. CD4⁺ T-cell activation by APC signalling. Signal-1, formed between the T-cell receptor and MHC-antigen complex initiates endocytic pathways within the T-lymphocyte, but this alone is not enough to activate T-cells. Additional co-stimulation is required (CD28 – CD80/CD86 interaction) for T-helper cell proliferation. As a result, various cytokines and chemokines are secreted, and part of the T-helper cells differentiate into T-effector and T-memory cells. Modified from Kambayashi et al.*[4]

## 1.2 CARBOHYDRATES ENCODE BIOLOGICAL INFORMATION

Glycans are elemental constituents of every plant and animal tissue; ubiquitously present in living systems in the form of oligo- and polysaccharides, or in conjugation with other biomolecules as glycoconjugates. They exhibit enormous structural diversity, mainly due to the fact that individual carbohydrates are multifaceted building blocks.

Monosaccharides are inherently different from nucleic- and amino acids with regards to constructing biopolymers. The latter monomers are bound to form linear macromolecular constructs, as they can only be linked by the two ends of a limited number of internally variable elementary units. The biodiversity of proteins is determined by the residues appended to the $C_\alpha$ of each amino acid, and more advanced structures are formed by chain folding and peptide arrangement into refined complexes. Similarly, DNA and RNA consist of repetitive units of only a few nucleotides; and yet, the sequences contain information of the entire genome - sufficient encoded instruction to synthesize the full set of vital proteins. Carbohydrate oligomerization, on the other hand, can be achieved by the formation of glycosidic linkages on more than two functional groups (hydroxyl groups) of the same monosaccharide - these bonds then determine various regio- and stereoisomers, such as α and β anomers. Additionally, the most common modifications of natural glycans for altered bioactivity are acetylation, methylation, sulfatation, phosphorylation and epimerization of different functional groups.

The multiple positions available for functionalization on the same monomer allow chain branching, which is a singularity of oligo- and polysaccharides among other biopolymers. More interestingly, asymmetrically branched chains can be distinguished from linear or symmetrical multiantennary glycans exhibiting the same saccharide epitopes. This distinct feature was decisively demonstrated by the modulated recognition between systematically branched oligosaccharides and GBPs (glycan binding proteins). In the studies of Wang *et al.*[5] complex *N*-glycans were chemoenzymatically modified into asymmetric multiantennary structures and their interactions with lectins and influenza-virus agglutinins were analyzed on microarray chips. The minimal epitopic elements were recognized differently when presented on linear, symmetrically- or asymmetrically-branched oligosaccharides. This indicates that the glycan environment around specific carbohydrate ligands and the capacity to form branched chains affect the biological information that oligo- and polysaccharides carry.

Carbohydrates are abundantly employed by biological organisms to store energy; maintain tissue structure and integrity; facilitate transport mechanisms; or to improve the efficacy of specific processes, such as the folding of polypeptides synthesized in the endoplasmic reticulum. In addition, hydrophilic saccharide moiety decorations are used to improve the solubility of other biomolecules; and to protect peptides in glycoproteins from the effects of protease enzymes and antibodies. Furthermore, glycans are crucial in numerous recognition and signalling events, determining their selectivity and effectiveness. Consequently, they are capable of initiating differing pathways, for example during fertilization,[6] embryogenesis,[7] disease development,[8] or in the immune response.[9]

The information encrypted by sugars is often referred to as 'glycomics'[10] (analogously to genomics and proteomics), but this discipline is still far from comprehensive understanding. The study, analysis and quantification of interactions between individual carbohydrates and proteins pose a serious challenge, as monovalent glycans have generally very low affinities towards GBPs. Most of the naturally occurring mono- and oligosaccharides typically show dissociation constants within the mM range towards their receptors,[11] while *in vivo* events triggered by glycans require a stronger binding interaction between proteins and their ligands. To be able to observe and assess the functions of carbohydrates in living systems though, one has to look beyond individual sugars.

To overcome this problem, glycans are presented to their receptors in multivalent ensembles of oligosaccharides and glycoconjugates, and so, the accumulated affinities compensate for the individually

insufficient interaction, setting a threshold for initiating biological events. The increased activity of these constructs is known as the "glycoside cluster effect",[11b, 12] which was clearly demonstrated during numerous valence-corrected binding assays,[13] both *in vivo* and *in vitro*.[14] Multivalent carbohydrate structures are widespread within organisms, both in soluble form (e.g. in blood, the cytosol or subcellular organelles) or as constitutional units of the cell membrane surface – and one of the best examples of multivalent glycan display in biological environments is the glycocalyx.

The three-dimensional matrix of glycans and glycoconjugates coating the eukaryotic and bacterial cell surface – the glycocalyx – is a major platform of communication between the cell and external molecules, pathogens and other cells. Various types of glycoconjugates are embedded into the lipid bilayer, with their carbohydrate segments exposed to the extracellular media. The glycan patterns, the density and composition of the glycocalyx may be indicative of the cell's developmental, physiological or pathological status.[15] It undergoes dynamic changes during cell differentiation or degeneration, but pathogenic modifications or other malignant transformations also leave their distinct marks on the coating of cells. For example, abnormal embryonic[16] and tumoral cells[17] show recognizable characteristics in their glycocalyx, which can be associated with atypical or uncontrolled cell growth - these cells are identified by detection of non-self or altered-self antigens and patterns on their surface.

The glycocalyx is a densely occupied nano-layer on the external face of the cell membrane: its typical dimensions range from a few nanometers (e.g. in adhesion-related molecules) to some micrometers (for example in carotid arteries).[18] It is built mainly from glycoconjugates of various kinds, where glycans are covalently bound to other biomolecules, such as proteins (glycoproteins and proteoglycans) and lipids (glycolipids). (**Figure 4**)

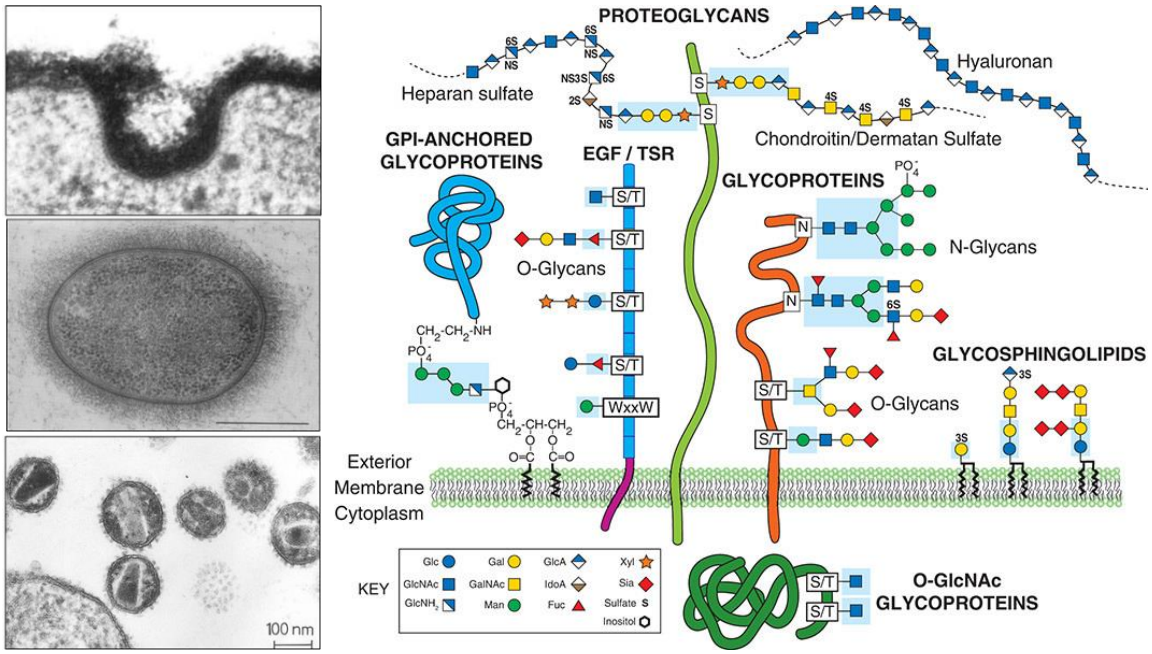



***Figure 4.*** *Left top: Electron micrograph of fibroblast glycocalyx, stained with ruthenium red (120 000x). Modified from: Martinez-Palomo et al.*[19]

*Left middle: Klebsiella pneumonia glycocalyx visualized by substitution fixation with osmium tetraoxide. Modified from: Amako et al.*[20]

*Left bottom: Human immunodeficiency virus (HIV-1) particles observed by transmission electron microscopy. Glycoprotein spikes on the virion detected after 1% tannic acid treatment of virus-producing H9 cells. Modified from: Gelderblom et al.*[21]

*Right: Major classes of glycoconjugates. In glycoproteins and proteoglycans, oligo- and polysaccharides are covalently bound to cell-membrane integrated peptide chains. Glycolipids, such as glycoglycerolipids and glycosphingolipids are amphiphilic conjugates of fatty acid chains embedded in the lipid bilayer and glycans are exposed to the extracellular side of the membrane. GPI-cores anchor glycoproteins to the cell-surface by tying their C-terminus to a lipid via a glycosylphosphatidylinositol-linker. Modified from: Essentials of Glycobiology, 3rd Edition. Cold Spring Harbor, 2015-2017.*[22]

The structure of glycoproteins depends on both the conformation of the polypeptide backbone and the distribution and organization of glycans. *N*- and *O*-glycosylations and ethanolamine-phosphate linkages are among the typical functionalization methods that bind oligosaccharides to peptides.[23] Saccharide-adorned proteins have altered physicochemical and biological properties from the native peptide – this strategy often improves the stability, solubility or specificities of the protein. Examples of glycopeptides can be found among enzymes, antibodies, hormones, cytokines and receptor proteins - some of which were mentioned in the previous chapter. For the proper functioning of glycoproteins, it is extremely important that the peptides are appropriately glycosylated during the post-translational phase.[24]

Proteoglycans are constructs of a core protein covalently attached to one or more large glucosaminoglycan (GAG) chains. Because of the presence of numerous sulfate and uronic acid groups, these polysaccharides are of polyanionic character, and they are built-up from small disaccharide blocks. Depending on the repetitive unit, the major classes of proteoglycans are: hyaluronic acid, chondroitin sulfate, dermatan sulfate, heparin, heparin sulfate and keratin sulfate. These molecules are commonly found in connective tissue where they maintain the overall organization, integrity and porosity of the tissue matrix.[25]

The amphiphilic nature of glycolipids grants these glycoconjugates a mobility to laterally diffuse within the cell membrane.[26] The lipophilic lipid chains are integrated into the bilayer and connected to glycans through a

phosphate group. The freedom to assemble allows the organization of two-dimensional multivalent arrays on the membrane surface, therefore these molecules can actively enhance their avidity towards GBPs.[27]

Another class of glycoconjugates, the glycosylphosphatidylinositol (GPI) anchors tether the C-terminus of proteins on the extracellular side of the membranes to fatty acids integrated into the lipid membrane during post-translational modification.[28]

In the past decades, numerous tools have been provided by scientists to observe or analyse the interactions of native glycans; or to exceed their pharmacokinetic (PK) and pharmacodynamic (PD) behaviour by artificial constructs.[29] Synthetic glycoconjugates are nano-scale materials whose architecture is designed to imitate or emulate the structure of the native molecules. Glycoclusters,[30] glycodendrimers,[31] neo-glycoproteins (NGPs)[32] and recombinant glycopeptides,[33] for example, typically resemble smaller, globular glycoproteins with a relatively-high density of saccharides; whereas fabricated glycopolymers[34] simulate long, extended, polyvalent glycoproteins and proteoglycans. Different carbohydrate-based structures can be organized into higher-level multivalent constructs by fine-tuned carrier-scaffolds, like cyclodextrins,[35] calixarenes,[36] rotaxanes,[37] nanofibers,[38] macrocyclic peptide scaffolds[39] and glycosylated nanoparticles[40] – offering help for targeting, delivery and bioimaging. It has also been demonstrated that glycan interactions can be investigated in the context of the complex glycocalyx environment. Mimetics of mucin (a type of densely glycosylated glycoprotein) with well-defined glycosylation patterns and dimensions were functionalized with lipids and then passively inserted into the cell membrane to model and evaluate properties of the glycocalyx.[41] Later the same research group achieved a precise, bio-orthogonal method to synthesize membrane-protein chimeras[42] – taking a step towards engineering the glycocalyx.

The biological importance of the glycocalyx spans from being a physical barrier between the cell and its environment; through structural and modulatory tasks; to engaging receptors on neighbour cells or in the extracellular media during specific recognition, cellular adhesion and signalling events. Such biological processes involve proteins like enzymes, glycan-specific antibodies, glycosaminoglycan-binding proteins, and another wide class of GBPs, the lectins.

## 1.3 About the versatile nature of carbohydrate-lectin interactions

Lectins are carbohydrate-specific proteins that vary greatly in size, structure and functions. One discrete modular element is common, however - lectins bind their sugar partners primarily through a carbohydrate recognition domain (CRD).

Lectins are found in animals both in soluble forms (in serum or in the extracellular matrix) and as integral membrane proteins. The mannose-binding lectin[43] (MBL; also referred to as mannose/mannan-binding protein, MBP) is one example of soluble lectins in the circulation. It is an important receptor in the immune system that recognizes terminal mannose and *N*-acetylglucosamine moieties expressed on the surface of pathogens, such as *Candida albicans* (fungal), HIV and influenza (viral), *Salmonella* and *Streptococci* (bacterial) and *Leishmania* (parasital pathogen). MBL also triggers the lectin pathway of the complement system which can eventually kill the pathogen.

The expression of integral membrane lectins mirrors the presentation of their cell-surface glycan partners: the CRDs are oriented towards the extracellular or luminal side of the membrane, linked by a neck to the transmembrane segment, and reaching into the cytoplasm with their N-terminus region of various lengths. The primary interaction between lectins and sugars is formed between the CRD and a carbohydrate unit of the lectin's specificity, while other domains on the lectin peptide mediate different functions in response to the recognition.

In principle, the CRD structure determines the main carbohydrate specificity of these proteins, but the recognition event is largely dependent on the presentation of glycans.[11b] In addition to the essential monosaccharide composition, the stereochemistry of different functional groups, the epitope-density, the three dimensional arrangement, as well as the supramolecular environment (e.g. glycoconjugates of the glycocalyx) affects the affinity of glycans towards many lectins. Consequently, many classes of lectins belong to the pattern recognition receptors (PRRs).

Apart from the main CRD, numerous lectins display extended or secondary binding sites that can accommodate additional glycan or aglycon motifs. These sites increase of affinity towards certain ligands, but can also determine selectivity among various carbohydrates. They may also provide an explanation for the discriminative recognition of identical epitopes on asymmetrically branched N-glycans in contrast to the linear or symmetrical multiantennary oligosaccharides (see **1.2**). The diverse ligand specificity of lectins is further broadened by multiple ligand-binding modes - some examples will be discussed in the following section.

It had been demonstrated that some lectins show major structural changes upon binding to carbohydrates, in order to tune their fine selectivity. Sialic acid-binding immunoglobulin-type lectin 7 (Siglec-7), for instance, features a peptide loop crucial for ligand-specificity in siglec-fold proteins, that undergoes significant conformational changes during co-crystallization with a synthetic oligosaccharide similar to α(2,8)-disialylated GT1b ganglioside.[44]

The low affinity of individual glycans towards proteins was already mentioned in **1.2** (from the carbohydrate point of view), and lectin-sugar interactions are no exception to this phenomenon. The dissociation constants ($K_D$) of natural glycans and lectins lie typically within the $10^{-3}$M - $10^{-6}$M range,[11b, 45] and this can be attributed to several distinct causes. These receptors usually feature relatively shallow binding sites, rather than deep, well-defined pockets, which could tightly accommodate ligands. There are only a few examples of lectins which grab their glycan partners in hollow cavities. One of these rare cases are the M-type lectins, where a deep cleft is found at the end of a barrel-shaped structure.[46] FimH, a bacterial lectin expressed by *Escherichia coli* promotes bacterial adherence to host cells – in this adhesin two tyrosines and a lysine residue form the so-called "tyrosine-gate" that leans over the more restricted CRD.[47]

The flat binding sites of lectins allow a high solvent access to the receptor-ligand complex, thus giving room for disruptive hydrogen-bonding and formation of salt-bridges under physiological conditions. At the same time, while glycosidic bonds are flexible in solution, the loss of this flexibility upon binding a single conformation results in disadvantageous entropy loss, that is compensated by beneficial enthalpy gain from newly formed linkages, such as H-bonds.[11a, 45]

To enhance affinity between carbohydrate ligands and lectins, natural glycans and glycoconjugates are organized into multivalent structures of various types, where recognition motifs are presented in close proximity to each other (glyco-cluster effect, see **1.2**). However, this multivalent appearance is also reflected by the lectin partners, who often form oligomers – usually from dimers to hexamers. Oligomerization is a type of hemophilic protein-protein interaction, as it involves two subunits of the same lectin; however lectins are capable of forming heterophilic protein-protein complexes as well. Galectin-8, a galactose specific lectin was shown to attract Nuclear Dot Protein 52 kDa (NDP52) for the activation of anti-bacterial autophagy against *Salmonella* proliferation. This is achieved by binding NDP52 on the convex side of the galectin-8 CRD, which allows simultaneous interaction with carbohydrates on the concave side of the domain.[46]

When several binding sites of a lectin engage glycans, the interaction can be described by two generally accepted models. The face-to-face model explains the binding of glycans and glycoconjugates exhibiting clustered carbohydrate epitopes, in which case the affinity is greatly increased. The trimeric asialoglycoprotein receptor, for example, has an approximately $10^6$-fold affinity enhancement when it binds to a trivalent glycan structure displaying terminal LacNAc motifs, in comparison with individual LacNAc.[48] The bind-and-slide (or internal diffusion) model[49] is used for globular and linear glycoproteins, and synthetic multivalent glycomaterials. In this scenario, one binding site of a lectin engages a single glycan-epitope at a time, but dynamically dissociates and rebinds to adjacent epitopes, this way increasing the persistence time (and observed avidity) of the complex. Further binding sites on the multivalent lectin - which do not participate in the primary glycan binding - are capable of cross-linking other glycoconjugates.

The outcome of these mechanisms leads to different biological results. Oligomeric lectins can bind carbohydrates on the same multivalent structure in a 1:1 complex (lectin inhibition); bridge two or more glycoconjugates on the same cell's surface (e.g. clustering of membrane-glycoproteins); or alternatively, cross-link glycans expressed on different cells.[50] The ability to form agglutinates from neighbouring cells is a characteristic of lectins, and this technique is often used for their detection and characterization.[51]

Indeed, the first plant lectin, ricin, was discovered in 1888 when castor bean (*Ricinus communis*) seed extracts were reported to agglutinate erythrocytes, hence the name 'hemagglutination'. Numerous other plant lectins were observed and characterized since then and although their function is not entirely understood, they are thought to have effects on plant growth, tissue development and defence reactions.[52] They are present in bulbs, fruits and seeds where these lectins play key roles in protecting the plant from herbivores, pests and microbial pathogens.[53]

Legume lectins (also known as L-type lectins) are the most studied plant lectins thanks to the high access to soluble lectins in the extracts of legume seeds. Structurally, the monomers consist almost always of antiparallel β-sheets which are connected by loops and β-bends - α-helices are rarely found in these proteins. Typically, 3 antiparallel sheets – a flat back-sheet, a concave front-sheet and a linking top-sheet – build up a dome (the "lectin fold"). Glycan-binding is mediated by one $Ca^{2+}$ ion and one transition metal ion (usually $Mn^{2+}$) which are based in the vicinity of the primary CRD - on the top of the front sheet, in the apex of the fold.[54]

The carbohydrate specificity of legume lectins varies significantly, however, in spite of their almost superimposable crystal structures and sharing the major part of their amino acid composition. This is the result of their variable tertiary structure, or more specifically, the size of one of the four main loops. Furthermore, additional binding sites, as well as aqueous interactions and post-translational modifications of the lectins can affect the recognition of carbohydrates.[55] As a consequence, some of these lectins bind mannose or glucose

with high specificity (e.g. *Concanavalin A*)[56] while there are others with exclusive affinity towards sialylated oligosaccharides (e.g. *Maackia amurensis*,[57] *Sambucus nigra*).[58] Legume lectins are more often than not organized into oligomers; regardless to being soluble or membrane-bound proteins, they usually construct tetramers or dimers.

This property – in addition to increased avidity when binding carbohydrates – stands behind the mitogenetic activity of legume lectins, as they are capable of stimulating DNA synthesis by quiescent cells.[55] Moreover, they can mimic natural protein ligands, like cross-linking growth factor receptors through their glycans, and this provides a useful application in cell growth control studies.[59] Consumption of undenaturated lectins originating from uncooked beans intensely stimulates intestinal epithelial cell surface receptors and lead to severe toxic effects in animals – this likely belongs to the plant defence activity.

Researchers have since long exploited the abundantly available plant lectins as biological tools. Immobilized lectins are ideal for the purification of oligosaccharides and glycopeptides in affinity chromatography. Lectins of well-characterized sugar-specificity are often used to detect certain structures on glycoproteins, or to determine the carbohydrate composition of various structures.[52a, 59] As some lectins show selectivity for blood group antigens and are excellent hemagglutinating agents, these proteins may be used for blood typing.[60]

Viruses and bacteria purposefully use lectins to adhere to membrane surfaces or to invade host cells. Influenza virus displays two proteins as "spikes" on the outer surface of the virion capsid to help infecting host cells.[61] The first protein – the trimeric viral hemagglutinin (HA) facilitates entering the cell membrane by binding to sialic acid residues on the glycoproteins of the glycocalyx. After replication of the virus, the other protein, neuraminidase removes sialic acid from the cell-surface glycans to enable release of the new virus particles. This way, the newly produced virions will not be retained by the same cell and are free to spread within the organism. Similarly, bacteria, like *E. coli* expresses lectins for binding to carbohydrate structures on host cells in the urinary or gastrointestinal tract, without being washed away by extracellular fluids. Two types of adhesins, FimH and PapG are presented on the tips of pili and fimbriae – short projections on the bacterial surface.[62]

Lectins are essential means of communication and transportation between cells in multicellular organisms and between invading pathogens and host tissues. However, some lectins are directly involved in intracellular glycoprotein trafficking. For these purposes, the glycoside motifs on the glycoproteins serve as a "handles" which lectins grab in order to move these glycoproteins or to read out information. Calnexin and calreticulin play a crucial role in the quality control of protein folding in the endoplasmic reticulum.[63] Incorrectly folded peptides are retained and presented to folding factors until the right structure is obtained, before they proceed in the secretory system. If the protein folding is beyond repair, another class of lectins, the M-type lectins are responsible to submit the faulty peptide to endoplasmic reticulum-associated protein degradation (ERAD).[64] The glycoprotein transport is facilitated by the L-type lectin subfamily (see legume lectins).[65] To avoid the release of destructive enzymes from the cells and to direct the freshly synthesized hydrolases from the *trans*-Golgi towards lysososmes, the enzymes are tagged with mannose-6-phosphate groups which are recognized by P-type lectins.[66] During infection, inflammation and tissue damage, oligomannose *N*-glycans are released into the circulation and are cleared out by the mannose receptor (MR), found on macrophages and the sinusoidal cells of the liver.[67]

Other lectins are involved in adhesion processes and contribute to cell migration. Selectins are cell-adhesion glycoproteins expressed on endothelial cells, lymphocytes and platelets, with a single chain integrated into the membrane. During leukocyte homing to peripheral lymph nodes, adhesive interactions are formed between the selectins on the blood vessel lining and the glycans expressed by the leukocytes. The strength of these interactions depends on the density and clustering of selectins and glycan moieties, but it is generally weak, and thus, the leukocytes keep rolling along the vessels until binding to an integrin enables their migration through the endothelium.[68]

Often intracellular signalling is also induced or mediated by lectin receptors. Siglecs, a class of the I-type lectin superfamily are able to initiate signals which will then activate or suppress immunological responses.[69] On the other hand, the B-cell associated CD22 molecule, a member of the Siglec family, is necessary to avoid the overactivation of the defence mechanism in the humoral immune system, thus decreasing the risk of developing auto-immune reactions.[70] In another example, when the macrophage-inducible C-type lectin (Mincle) - a CLR expressed by macrophages - binds to pathogenic glycans on the cell wall of *Mycobacterium tuberculosis*, a signalling pathway is induced and by the end, the macrophage starts secreting pro-inflammatory cytokines.[71]
Occasionally, lectins can take part in bactericidal pore-formation. RegIIIa (hepatointestinal pancreatic/pancreatitis associated protein, also known as HIP or PAP) attaches to intestinal bacteria through binding to its peptidoglycan backbone. Consequently, the lectin forms hexameric pores, inducing uncontrolled ion-efflux which causes lysis of Gram-positive bacteria.[72]
Lectins are fundamental elements of the cell-mediated immune response - from the step of recognizing molecular patterns on antigens and pathogens (PRRs); through the antigen uptake and internalization; to the antigen-presenting of APCs and signal transmission. At least three families of lectins serve as PRRs. They are grouped based on the architecture of their structural domain and the minimal carbohydrate epitope required for recognition. Siglecs recognize sialic acid residues,[73] while galectins possess a specificity for galactose containing sugars - typically derivatives of lactose and *N*-acetyllactosamine.[74] The largest and most diverse family of lectins, the C-type lectin receptors, are involved in all of the abovementioned steps during the cellular immune response.

## 1.4 THE STRUCTURE AND FUNCTIONS OF C-TYPE LECTIN RECEPTORS

### 1.4.1 CLRs display a highly conserved CRD structure

C-type lectin receptors (CLRs) display at least one independently folding, modular domain defined by conserved residues – the C-type lectin-like domain (CTLD).[75] The overall fold of the CTLD is determined by the architecture of a hydrophobic core held together by disulfide bonds formed between Cys residues. In addition to the CTLD, CLRs typically require one of more $Ca^{2+}$ ions to bind sugars (hence the letter 'C' in their name) – the metal is ligated by various sets of conserved residues and they collectively represent primary or secondary sugar-binding sites.

Not every protein that presents the CTLD is $Ca^{2+}$-dependent, or binds saccharides at all, thus there is a distinction between CLRs and proteins featuring the CTLD, although $Ca^{2+}$ mediated carbohydrate recognition is the most common function of vertebrate CTLDs. Taking an even broader view, there are structural similarities between C-type lectin CRDs and other protein domains without significant sequence analogy – domain structures of this type are commonly described as C-type lectin-like folds (CTLF).

The crystal structure of numerous CTLD containing protein (with or without bound ligands) revealed a common compact, globular structure of about 115-130 amino acid residues, adorned with a characteristic double-loop (loop-in-a-loop). This is stabilized by two highly conserved disulfide bonds and other conserved hydrophobic and polar interactions. (**Figure 5**)

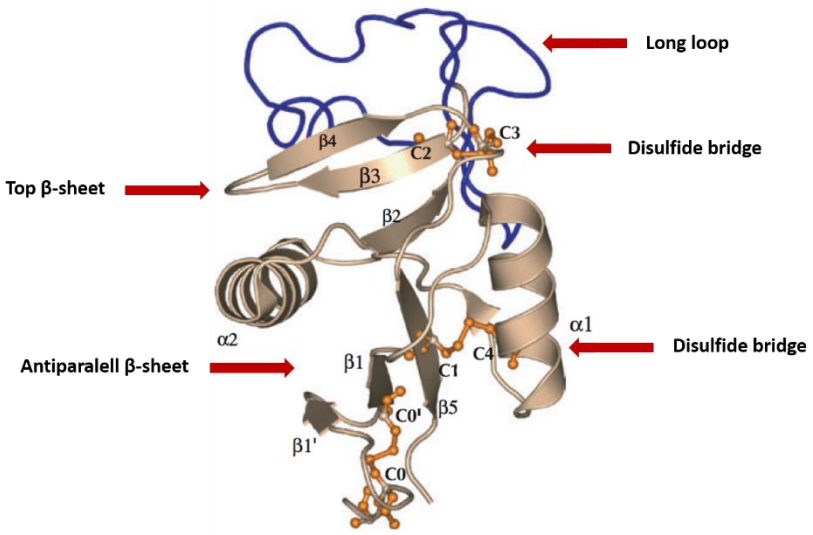



***Figure 5.*** *Double-loop structure of the CTLD. The domain is defined by an overall disordered loop, with an antiparallel β-sheet located at the bottom of the domain (β1 and β5 strands) and a second β-sheet on the top, and two additional α-helices (α1 and α2). The CRD defining long loop enters and exits the domain at the same location, close to the β2, β3 and β4 strands. The domain is stabilized by 4 Cys residues which link β5 with α1 and β3 with β5. Modified from: Zelensky et al.*[75b]

The overall domain is defined by a disordered loop whose N- and C-terminal β-strands (β1 and β5) assemble into an antiparallel β-sheet under the domain. In CLRs, the presence of a secondary loop (the long loop) defines the $Ca^{2+}$-dependent CRD, however this loop contains only a few conserved structures – it is both evolutionarily and structurally flexible. The variable primary and secondary structure of the CRD-defining loop might be a result of the enormous evolutionary pressure on CLRs, as they have to adjust to the rapid transformation of

microbes. In some bacterial CTLDs, this whole loop can be entirely missing and in a few cases, this loop is involved in domain-swapping dimerization.[76]

This long loop enters and exits the core domain at one location, close to a second β-sheet which is formed by the β2, β3 and β4 strands. Two α-helices (α1 and α2) adjoin the main domain and the whole CTLD is stabilized by four highly conserved cysteine residues – from one side they link the main domain loop (β5) with α1 via a disulphide bridge, and on the top of the domain, they fix the long loop entrance and exit site by connecting β3 with β5. There is also a highly conserved WIGL motif located on the β2 strand which is used as a landmark in sequence analysis.

Typically, there are up to four $Ca^{2+}$ ions present on each CRD, from which usually the second serves as the sugar-binding site. Some CLRs, however, lack the canonical $Ca^{2+}$-mediated sugar binding site completely – for example dectin-1 and layilin. Eventually, a few highly characteristic motifs around the main $Ca^{2+}$ determine the carbohydrate-specificity of the CLR itself. In the majority of these lectins, the primary CRD is formed around a $Ca^{2+}$ centre which is stabilized by carboxylic and amide moieties either in the peptide backbone or in the side chains – these include Asn, Asp, Glu and Gln residues. The first two amino acids are present in a highly conserved WND triplet located on the β4 strand. While Asn and Asp are directly involved in the metal coordination, Trp is a highly conserved follower of these residues and serves as a key point to detect the motif in the sequence. Additionally, there are two variants of specificity-defining group of residues present in the majority of CLRs, the EPN and QPD triplets. These amino acid residues create an intricate network of interactions with the $Ca^{2+}$ ion as well as H-bonding with each other and the carbohydrate ligand. During ligand binding, an eightfold coordination of the metal ion is accomplished by the WND and EPN/QPD residues and two hydroxyl groups from a monosaccharide.

In both the EPN and QPD triplets, the internal Pro has a conserved *cis* conformation which maintains the backbone conformation and positions the two adjacent carbonyl groups in the appropriate positions for $Ca^{2+}$ coordination. The Glu-Asn (EPN) and Gln-Asp (QPD) residues are directly coordinating the $Ca^{2+}$ and form further H-bonds with the carbohydrate ligands – whose nature is specifically determined by these triplets. During ligand binding, two vicinal hydroxyl groups of specific carbohydrates form a complex with the amino acid residues in the CRD. In EPN containing CLRs, two equatorial hydroxyl groups in positions 3 and 4 of the sugar ring are ligated by the $Ca^{2+}$, and therefore these lectins recognize primarily mannose, glucose and *N*-acetylglucosamine related structures, as well as fucose-containing saccharides, which use OH groups in position 2 and 3 for coordination with the $Ca^{2+}$. The QPD sequence determines a preference for galactose and *N*-acetylgalactosamine, where the hydroxyl groups in position 3 and 4 have an equatorial-axial geometry.

The proof for this sugar-specificity determination was provided for the first time by Iobst *et al.*,[77] who replaced the originally Man/Glc defining EPN sequence in murine MBL with a QPD triplet. The obtained QPD mutant showed dramatically increased affinity towards Gal-type structures coupled with a loss of affinity towards Man, although the naturally QPD-containing CRDs discriminate between the two carbohydrates more efficiently. In the same model, crystallographic analysis did not reveal any significant structural difference in the geometry of the $Ca^{2+}$ binding site, except for the fact that swapping the H-bond donor and acceptor across the carbohydrate binding plane changes the formerly asymmetrical (Man-type) pattern of H-bonding to symmetrical (Gal-type). This leads to the assumption that other H-bonding networks are formed between the saccharide and the amino acid residues, and these effectively determine the ligand specificity, orientation and conformation. This is supported by the observation that in many cases CLRs bind sugars atypical for their pre-determined preference - EPN motif containing CRDs occasionally recognize galactosides (e.g. CEL-IV,[78] a CLR isolated from sea cucumber *Cucumaria echinata*) and binding sites displaying the QPD sequence accept mannosylated structures (e.g. *Tc*-CTL-1,[79] a CLR secreted by the parasitic nematode *Toxocara canis*). (**Figure 6**)

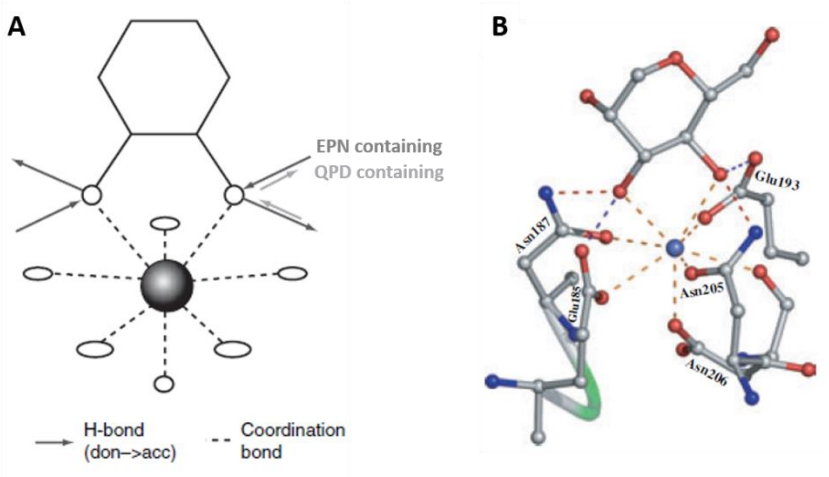



***Figure 6.*** *Generalized presentation of the primary $Ca^{2+}$ binding site in CLRs. The residues contributing to sugar-binding are organized around a $Ca^{2+}$ core, and typically comprise Asn, Asp, Glu and Gln. Upon ligation of a monosaccharide, an eightfold coordination and H-bond network is created around the metal. The carbohydrate residue uses its 3 and 4 hydroxyl groups (occasionally 2 and 3) to directly coordinate the $Ca^{2+}$. EPN motifs in the CRD sequence grant room for two equatorial OH groups in these positions, leading to the recognition of Man and Glc-type monosaccharides (or Fuc, in the case of OH at positions 2 and 3). The QPD triplet allows the ligation of Gal-type carbohydrates where 3-OH is directed equatorially and 4-OH is in an axial position. Modified from: Zelensky et al.*[75b]

The primary monosaccharide interactions within the $Ca^{2+}$-binding site are often accompanied by several additional interactions which may fundamentally deviate empirical observations from predictions based on the peptide sequence. For example, while langerin - a CLR expressed on the surface of Langerhans cells - binds mannose structures at the primary sugar-binding site in accordance with its EPN motif, it can also accommodate sulphated galactose in the same pocket.[80] This non-optimal $Ca^{2+}$ ligation was later attributed to the formation of salt-bridges between the sulfate-group and two positively charged Lys residues - this charge-charge interaction compensates for the disfavourable coordination geometry of the galactose hydroxyl groups. It was similarly difficult to explain, why the blood dendritic-cell antigen-2 (BDCA2) recognizes selectively galactose-terminated biantennary *N*-linked glycans on microarrays when it has a high specificity towards Man-related sugars and displays the EPN motif at the $Ca^{2+}$ site.[81] Competition assays then determined the binding epitope as a trisaccharide Galβ1-3/4GlcNAcβ1-2Man and further crystallographic and mutagenesis studies revealed that a mannose residue is indeed ligated at the conserved $Ca^{2+}$ binding site, however a wide, shallow groove adjacent to the primary binding site forms extra interactions with the galactose unit. Likewise, extended or secondary binding pockets near the primary epitope-binding site offer room for extra ligand moieties in the case of mouse dendritic cell inhibitory receptor 2 (mDCIR2) and a biantennary, complex *N*-glycan, bearing bisecting GlcNAc.[82] X-ray crystallography demonstrated that mDCIR2 accommodates one arm of the multiantennary glycan near the primary $Ca^{2+}$ sugar binding site. On the other hand, the dendritic cell specific intercellular adhesion molecule-3 grabbing nonintegrin (DC-SIGN) binds simultaneously two arms of a similar GlcNAc terminated branched glycan.[83] Although the main interactions are formed between a core mannose and the primary sugar-binding site, the two branches on the nearby CRD surface maximize the affinity between the lectin and glycan. DC-SIGN is well-known to bind numerous highly mannosylated oligosaccharides through accessory binding sites – analogously to the previous example – by accommodating the saccharide arms adjacent to the mannose residue ligated in the primary binding site. (**Figure 7**)

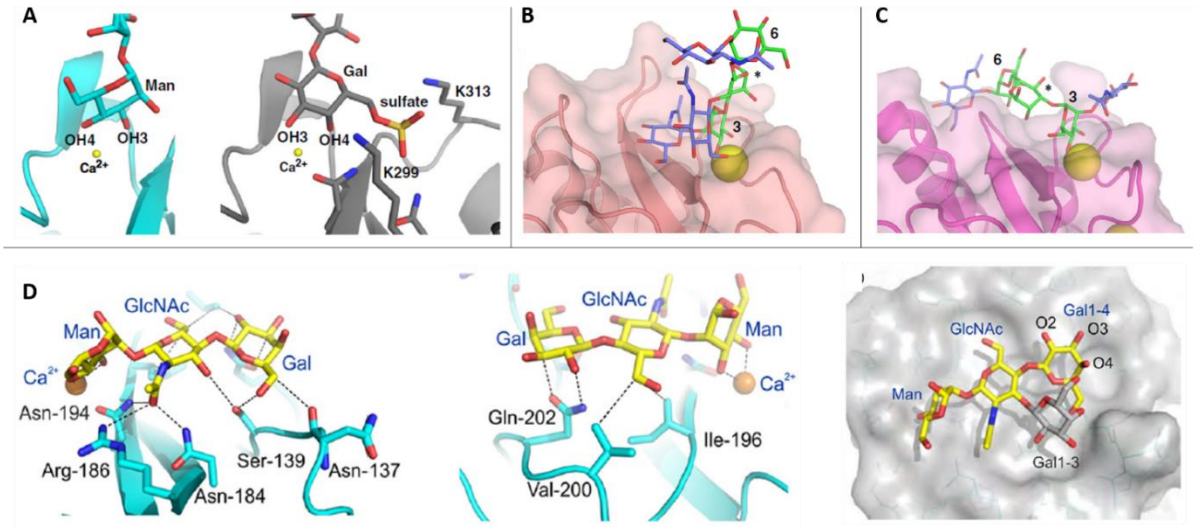



***Figure 7.*** *A. Non-optimal sugar-binding in the CRD. Left - Binding of Man in the CRD of langerin, as anticipated by the presence of the EPN triplet (langerin-Man5 complex; PDB – Protein Data Bank - code: 3P5D). Right – non-optimal recognition of sulphated galactose by forming salt-bridges between the sulfate group and two Lys residues (langerin-6$SO_4$-Galβ1-4GlcNAc complex; PDB code: 3P5I). Modified from: Nagae et al.*[84]

*B, C, D. Extended and adjacent secondary sugar-binding sites. B. Mouse DCIR2 binds to different arms of bisected GlcNAc-terminated N-glycans (mDCIR2-branched N-glycan complex, PDB code: 3VYK). Modified from: Nagae et al.*[84]

*C. DC-SIGN binds simultaneously two branches of a biantennary oligosaccharide, as a result of interactions between the glycan and secondary binding sites close to the primary $Ca^{2+}$ site (DC-SIGN-branched GlcNAc-terminated N-glycan complex, PDB code: 1K9I) Modified from: Nagae et al.*[84]

*D. BDCA2 recognizes Gal-terminated biantennary N-glycans via accommodating Gal residues in a shallow groove adjacent to the primary $Ca^{2+}$ binding site (BDCA2-Galβ1-3/4GlcNAcβ1-2Man complex, PDB code: 4ZET). Modified from: Jégouzo et al.*[81]

Even though high-resolution crystallography provides an intimate insight to the spatial arrangements of sugar recognition, it may not always reveal every aspect of the interactions. In addition to exhibiting adjacent binding sites close to the primary one, DC-SIGN was confirmed to utilize multiple binding modes to recognize highly mannosylated oligosaccharides.[85] In particular, $Man_6$ co-crystallized with DC-SIGN was discovered to have two binding modes (3:1 ratio), and the resolution of the binding epitope Manα1-2Man revealed that the two orientations differ in ligand coordination by a 180° rotation about the C3-C4 bond bisector (this two-fold rotation axis is a recurring geometrical feature of multiple-mode binding). Coordinates could be assigned for the major complex, but the minor one lacked the definition of a few ligand atoms. The less populated conformation was resolved by saturation-transfer difference nuclear magnetic resonance (STD-NMR) studies,[86] and thus, both ligand orientations are now well-characterized. It is important to stress that the presence of the Manα1-2Man motif significantly improves DC-SIGN's affinity towards oligosaccharides due to both the additional contacts and the entropic enhancement of binding through different conformations.

Another example of multiple binding modes in the same $Ca^{2+}$ binding site involves the macrophage galactose lectin (MGL). The CLR binds Gal and GalNAc with different affinity which can be traced back to the dual recognition mode of the two sugar rings. STD-NMR experiments complemented with molecular dynamics (MD) simulations unveiled opposite trends in the major interaction patterns for the two ligands; as well as further H-bonds and CH-π interactions between the NHAc group and amino acid residues in the peptide backbone.[87] These observations offer an explanation to the affinity contrast between the two similar ligands and highlight the restrictions of relying only on one structural-analysis technique. (**Figure 8**)

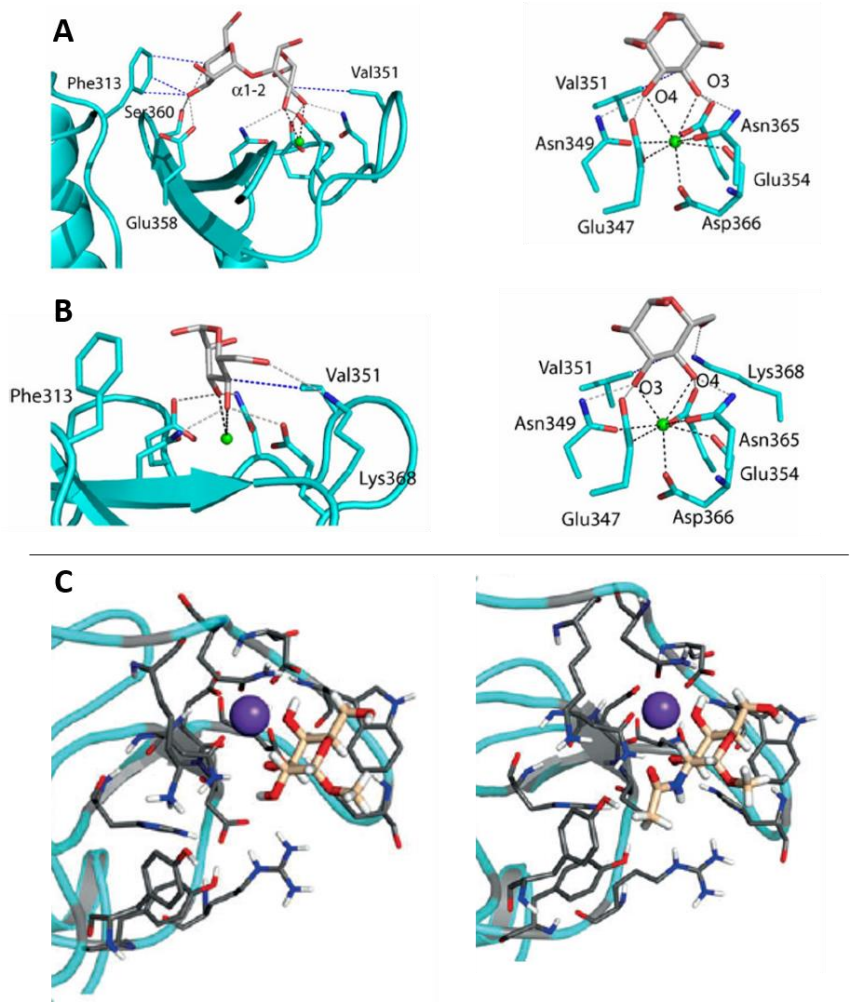



***Figure 8.*** *Multiple sugar binding mode in the primary $Ca^{2+}$ binding site.*
*A and B. Two orientations observed in the crystals of DC-SIGN-$Man_6$ complex (3 : 1 ratio for major : minor, PDB code: 2IT5). Co-crystallization with the epitopic Manα1-2Man disaccharide (PDB code: 2IT6) reveals the conformation of the major orientation (A), most importantly H-bonding with Phe313 and Ser360. In the minor orientation (B) only one Man residue is visible, rotated by a 180° axis bisecting the C3-C4 bond, forming the typical coordination interactions and H-bonding. In Man6, a second Manα1-2Man unit at the non-reducing end interacts with Val351, but this interaction was not confirmed by crystallographic data about Manα1-2Man. Further computational predictions and their confirmation by STD-NMR solved the minor binding mode of the disaccharide. Modified from: Feinberg et al.*[85]
*C. Dual binding-mode of Gal and GalNAc in the CRD of MGL. Selected molecular dynamic (MD) simulation conformations representing the two orientations (supported by STD-NMR) data show a rotation of the galactopyranose ring of about 180°. This orientation allows the formation of additional CH-π interactions between the NHAc group of GalNAc and the peptide residues, significantly enhancing the affinity of GalNAc towards MGL. Modified from: Marcelo et al.*[87]

More remote secondary binding sites are capable of enhancing affinity towards extended ligands, such as larger oligosaccharides, glycoconjugates and additional aglycon motifs appended to the glycan core. For instance, the mouse scavenger receptor C-type lectin (SRCL) $Ca^{2+}$ site coordinates the 3 and 4 hydroxyl groups of Gal in the $Lewis^X$ blood group antigen (Galβ1-4[Fucα1-3]GlcNAcβ), as predicted by the QPD motif in its sequence. Furthermore, the relatively apolar "B" face of Gal is oriented towards a tryptophan sidechain, which allows packing of this side of the sugar with the aromatic functionality, while the Fuc moiety interacts with other residues (e.g. Lys), reaching out from the CRD surface.[88] Although the human CRD has not been co-crystallized with the ligand, probably similar interactions account for the restricted specificity of the two ligands

towards $Le^X$ and $Le^A$. In another example, Mincle coordinates one Glc of a trehalose (Glcα1-α1Glc) headgroup of a glycolipid, while a hydrophobic groove in the CRD accommodates a long acyl chain on the 6-OH of the sugar,[71] significantly improving the affinity between the lectin and the glycoconjugate.

Occasionally, the ligand may form additional interactions not only around the CRD, but with the neck region as well. Langerin, on one hand, binds small heparin oligosaccharides (containing 1-8 trisaccharide units) at the primary $Ca^{2+}$ binding site – a common key structural element is used for coordination. On the other hand, heparin oligosaccharides larger than 6 kDa use a hexasaccharide motif that fits into a positively charged groove formed by the base of two CRDs and the neck domain – this binding takes place in a $Ca^{2+}$-independent manner.[89] (**Figure 9**)

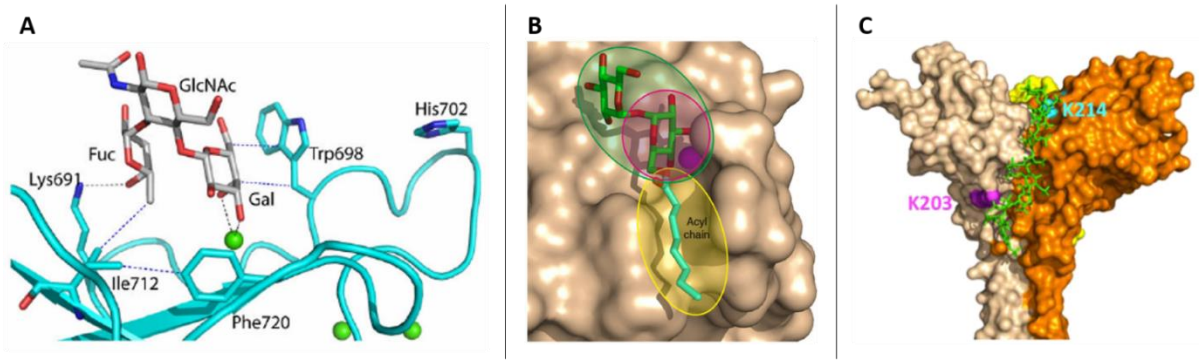



***Figure 9.*** *Secondary binding sites and further additional interactions between CLRs and ligands. A. Mouse SRCL (PDB code: 2OX8) binding $Le^X$ by coordinating two vicinal hydroxyl groups of Gal (QPD triplet in the peptide); packing the apolar face of galactopyranose with a nearby Trp and forming additional H-bonds between the Fuc and a Lys residue. Modified from: Feinberg et al.*[88b]

*B. A hydrophobic groove in Mincle is capable to accommodate apolar chains of glycolipids, while the minimal epitope monosaccharide is ligated at the primary sugar-binding site (cow Mincle-trehalose monobutyrate complex, PDB code: 4ZRV). Modified from: Taylor et al.*[90]

*C. More remote additional interaction between langerin and heparin. Smaller heparin trisaccharide motifs are bound at the primary $Ca^{2+}$ binding site. Heparin polysaccharides larger than 6 kDa, however, fit a hexasaccharide unit into a positively charged region located at the junction of the neck and two CRDs. The 3D model of langerin extracellular domain (ECD) and 6 kDa heparin was built based on molecular modelling simulations supported by STD-NMR and trNOESY experiments, using the crystal structure of langerin complexed with laminaritriose (PDB code: 3P5H). Modified from: Muñoz-García et al.*[89b]

Many CLR-ligand complexes have been structurally explored and the results have successfully explained unexpected or vague ligand recognition observations, but further factors are yet to be revealed. These examples were meant to pinpoint that unique binding modes; extended and secondary recognition sites; and additional interactions with saccharide or aglycon moieties powerfully expand the diversity among CLRs. The lectins will consequently possess finely tuned saccharide-specificity – a requirement to professionally orchestrate a wide range of biological events.

The sugar binding activity of a CRD can be strictly dependent on the $Ca^{2+}$ concentration of the surrounding media. SIGNR1 (a mouse homologue of DC-SIGN) showed important differences in binding sialylated glycans on IgG when crystallized at low $Ca^{2+}$ levels compared to the conventional recognition modes, because the secondary adjacent $Ca^{2+}$ site was absent from its structure.[91] This raises the question that if the metal ions stabilize the overall architecture of the CRD, but are missing from the crystal, the obtained image might not represent the binding geometry in physiological environment. In Mincle, a similarly absent accessory $Ca^{2+}$ binding site causes re-arrangement at the primary coordination site, which does not allow sugar-binding at lower metal concentrations.[92] Only higher $Ca^{2+}$ levels enable the canonical sugar ligation at the primary site,

when the secondary site is occupied by a metal ion. As biological environments vary in their $Ca^{2+}$ concentrations, as well as in their acidity, ligand binding and release occurs in different compartments. Endocytic CLRs tend to sharply lose their $Ca^{2+}$ binding and with it, their sugar-binding activity at lower pH levels in the endosome, resulting in the discharge of their saccharide ligands.

### 1.4.2 Classification of C-type lectins based on the CRD arrangement and oligomerization

The C-type lectin CRDs display different evolutionary patterns that is portrayed by several conserved receptors across the animal kingdom, or the absence of certain lectins in some species. Furthermore, a few relatively recent divergences and duplications were observed in the case of some receptors.[90] Close orthologues of cell adhesion molecules (such as selectins), MR and SRCL were identified in different species, testifying the conservation of endogenous receptors. One of the most striking examples for the absence of orthology is demonstrated by the DC-SIGN family. Humans express two closely related genes for DC-SIGN and its homologue, DC-SIGNR (DC-SIGN related molecule); while the mouse genome encodes eight SIGN-R lectins – significantly different from the human receptors. The structural and functional differences among these proteins limit the applicability of mouse models.
Drickamer *et al.* defined 14 families of CTLD-containing proteins,[93] which system was later expanded into 17 different classes by Zelensky and Gready.[75b] (**Figure 10**) These subdivisions of the protein superfamily are based either on phylogeny or on the overall domain architecture and sequence alignment, rather than the oligomerization patterns. The largest and most extensively studied families of CTLD-containing proteins are the NK-receptors (e.g. NKR-P1), collectins (e.g. MBL), selectins (L-/E-/P-selectins), and the myeloid C-type lectins (for example DC-SIGN) – from which the last three groups belong to the more restricted family of C-type lectins.

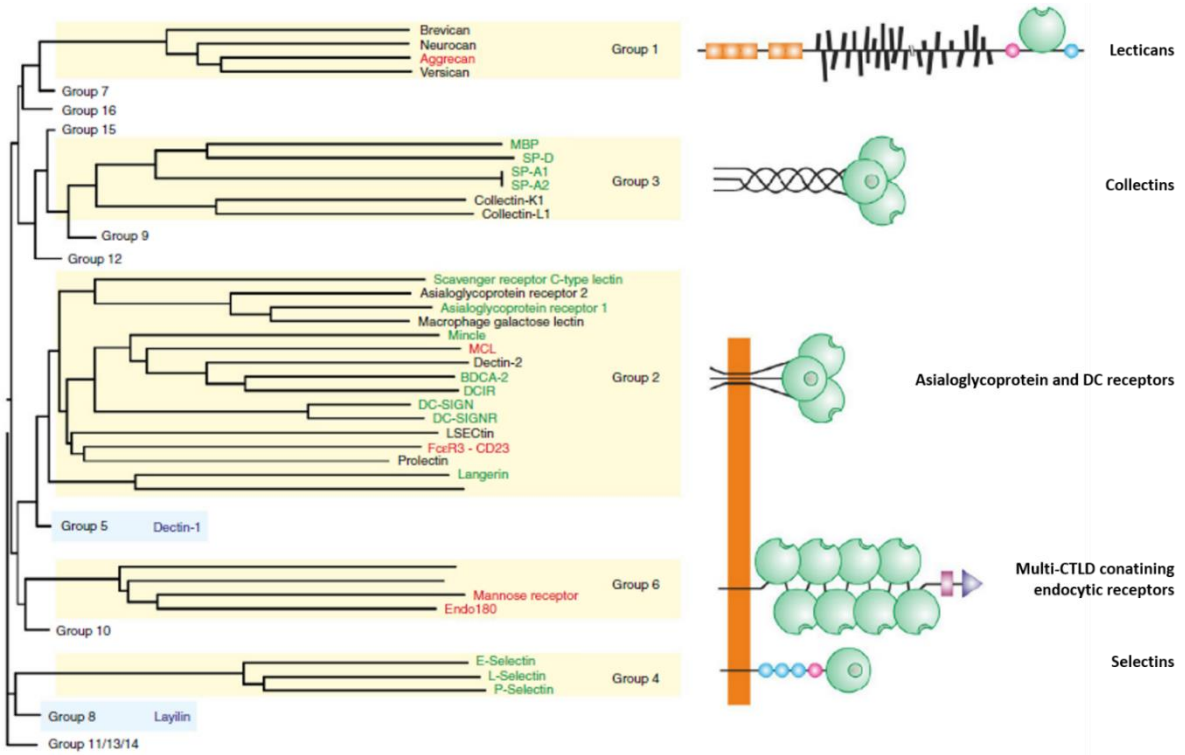



***Figure 10.*** *Major subclasses of the C-type lectin receptors and their schematic representation with respect to the CTLDs, neck regions and transmembrane domains (if applicable). Modified from: Drickamer et al.*[94]

As discussed earlier, oligomerization is one of the strategies to enhance the rather poor affinity between sugars and CLRs. C-type lectin CRDs are no exception and are often found in clusters in the cell-membrane or in soluble forms. The spatial arrangement of the multiple CRDs is diverse: in a few cases, multiple CRDs are decorating a single neck (e.g. MR),[95] while in other cases, lectin polypeptides bearing only one CRD assemble into homo-oligomers (e.g. DC-SIGN,[96] langerin,[97] MBL[98]), occasionally hetero-oligomers (e.g. macrophage C-type lectin - MCL - and dectin-2),[99] to increase avidity. Some CLRs, like MBL, present their CRDs in a fixed, moderately rigid geometrical arrangement, whereas other oligomers are flexibly linked (e.g. DC-SIGN).

The architecture of the assemblies can be of crucial importance however, and may deeply affect the biological role of the CLR. The trimeric langerin undergoes significant structural alterations upon the replacement of one residue at the interface between two CRDs.[100] This sole mutation perturbs the correct presentation of the CRDs and destabilizes the global oligomerization architecture, consequently causing abnormalities in the membrane structure of Birbeck granules – cytoplasmic organelles in Langerhans cells essential in the degradation of HIV-1.[101]

In a biological environment, CLR homo-oligomers can organize into higher order multimeric structures. DC-SIGN is known to form well-defined, discrete microdomains of 1-2 tetramers on the surface of DCs,[102] which domains facilitate binding and internalization of virus particles (e.g. HIV-1), although no rearrangement of the domains was observed during this process.

### 1.4.3 Immunological functions and signalling of CLRs

The rich diversity of CLRs in terms of distribution in tissues, structure and specificity strongly implies their wide range of biological functions, which often require delicate coordination and selectivity. The central role of myeloid cells (such as monocytes, macrophages, neutrophils and DCs) in anti-microbial host-defense is primarily mediated by the CLRs expressed on their surface. These receptors are involved in pathogen recognition and binding to the cell walls of viruses, bacteria, fungi and parasites by detecting non-self molecular patterns (PRRs, see previous chapters), and are directly or indirectly involved in their internalization and degradation (endocytosis and phagocytosis), followed by the further antigen-processing and presentation. The myeloid CLRs modulate the immune response by inducing the proliferation and differentiation of various $T_H$ cell subsets based on the specific CLR and pathogen. They dampen or enhance certain immunological reactions by regulating natural killer functions or activating the complement system,[103] as well as by triggering alternative signalling pathways in order to direct gene expression or platelet activation. CLRs also help myeloid cells in cell adhesion, for example during their migration and antigen-presentation to T-cells – some examples will be detailed at specific CLR discussions.

Not only do CLRs initiate immune responses, but some of them play a role in maintaining immune homeostasis. Dectin-1, for instance, has a major role in gastrointestinal homeostasis due to its capability to sense mycobiota in the guts, because it supresses fungal-mediated inflammation after mucosal damage.[104] Furthermore, dectin-1 polymorphism was detected in the intestinal tissue of patients with severe ulcerative colitis.[105]

Abnormal CLR expression was demonstrated in different autoimmune responses. mDCIR deficient mice are more likely to develop autoimmune diseases because of the excessive expansion of DCs whose growth would normally be downregulated by this CLR.[106] In humans, the gene encoding DCIR is associated with rheumatoid arthritis.[107] Patients with systemic lupus erythematosus (SLE) showed significantly reduced levels of BDCA2, expressed by plasmacytoid DCs, which is required for limiting the TLR-induced type I IFN secretion.[108] IFN-α and β are therefore excessively produced, and their concentration in circulation increases – this might bear pathogenic significance in SLE.

CLRs were confirmed to initiate or alter allergic responses. Dectin-2 facilitates the development of allergy to house dust mite (HDM) allergens,[109] as it induces the production of cysteinyl leukotriene and IL-33, leading to airway inflammation and Th2 immunity. The macrophage CLR Mincle, upon detecting glycolipids released after cell-damage, initiates skin allergy and consequently promotes the development of experimental autoimmune uveoretinitis.[110]

In addition to the previously mentioned pathological conditions, C-type lectins exhibit immense therapeutic potential to detect and target dead and tumoral cells.[17, 111] Roughly half of the human proteins carry glycan appendages as glycosylation is the most common post- and co-translational transformation modification. Upon malignant transformations or tumor progression however, major alterations in the expression levels of glycosyltransferases in the Golgi-apparatus cause aberrant glycosylation patterns on the tumoral cell population. One of the most characteristic glycan alteration is the increased size and branching of N-glycans, while another prominent sign is the upregulation of *N*-acetylneuraminic acid and fucose in the cell-surface glycoconjugates as a result of the increased sialyltransferase and fucosyltransferase activity.[112] Overexpression of Lewis-type and sialylated Lewis-type antigens ($Le^X$, $Le^Y$, $sLe^X$, $sLe^A$) is reported to occur in carcinomas,[113] and excessively fucosylated cell-surface glycoproteins are observed in ovarian,[114] prostate[115] and colorectal cancers,[116] for example. These warning structural marks of malignant cells and, similarly, the characteristic antigens on the surface of necrotic cells can be recognized and effectively distinguished from the normal glycosylation patterns. CLEC9A (also referred to as DNGR-1) is reported to detect dead cell antigens, promote

their phagocytosis by DCs and cross-present the epitopes to $CD8^+$ T-cells.[117] Macrophages recruit neutrophils to damaged tissue by expressing various chemokines and TNF as a consequence of Mincle recognizing a necrotic cell antigen released from dead cells, for example after ischemic stroke[118] and traumatic brain injuries.[119]

Adhesion molecules, such as DC-SIGN, DC-SIGNR and LSECtin (lymph node sinusoidal endothelial cell C-type lectin) have been affiliated with cancer occurrence, progression and metastasis. DC-SIGN binds to tumor-associated carcinoembryonic antigens containing aberrantly glycosylated forms of $Le^A$ and $Le^B$ on cell-surface oligosaccharides that causes attenuated function of DC-mediated antitumor-immunity and leads to colon cancer progression and subsequent metastasis by immune evasion.[120] Similarly, LSECtin present on B16 melanoma cells hinders the antitumor immune response by the suppression of tumor-specific T-cell response.[121] On the other hand, some CLRs, like MR, LSECtin and the selectins also promote metastasis by contributing to the circulation of cancerous cells in the blood or lymphatic system, their adhesion to endothelial cells and the eventual migration through the endothelium.[122]

Occasionally, CLRs are directly contributing to anti-tumoral responses, as it was reported in the case of the DC and macrophage associated dectin-1.[123] The lectin binds to tumor surface N-glycans and triggers a cascade of signaling events, including the nuclear translocation of interferon regulatory factor 5 (IRF5) and specific gene expression, in order to promote NK-mediated tumor-cell killing via hemophilic interactions.

Two engaging possible applications of the selective recognition of tumor-antigens are the application of DCs in cancer diagnostics[124] and the exploitation of their signalling functions in vaccine development where traditional vaccination fails (e.g. infectious diseases like AIDS and tuberculosis, or inflammatory diseases, such as autoimmune diseases and asthma). Partially successful attempts were made to target CD205 (or DEC-205) where mouse models indicated that the method can stimulate $CD4^+$ T-lymphocytes, but the approach could not be transferred efficiently into nonhuman primates and humans.[125] *Ex vivo* generated DC vaccination therapies combined with delivering tumor antigens to the antigen-presenting interfaces were proven to elicit directed immune responses,[126] and recently a set of metastatic melanoma immunotherapy clinical trials were based on primary myeloid DCs in the natural circulation.[127] These studies demonstrated feasible and safe DC vaccination that improved progression-free survival rates of patients with virtually no side effects. According to the authors, the production of the myeloid DC vaccines is fast and highly controlled, and the effects include multifunctional, antigen-specific T-cell responses. It still remains open, however, whether the laborious and costly development of specific DC-subset targeting vaccination could compete with the protein-based or particle-based therapies (TLR adjuvants).[128] DC vaccines could have an advantage at targeting receptors that are engaged in certain, desired immune responses - like antigen loading on one of the two MHC molecules versus the other, or the initiation or redirection of specific signalling pathways.

For adequate immune response, the proper differentiation of $T_H$ cells is crucial, and for this, DCs need to convey specific and detailed information about pathogens. The PAMPs on the cell-walls of bacteria, viruses, fungi and parasites are akin to fingerprints which is only recognized by a particular set of PRRs. Among PRRs, CLRs are directly or indirectly (through the modulation of TLR signalling) involved in the initiation of specialized gene expression pathways in response to certain pathogens. As mentioned in an earlier chapter, upon presentation of antigens in conjugation with MHC molecules, $CD4^+$ cells will proliferate and differentiate into distinct $T_H$ cell subsets, based on the invading pathogen and the PRR induced signalling. $T_H1$ cells express IFN$\gamma$ for the activation of macrophages, and these phagocytes can effectively clear the microbes. $T_H2$ cells secrete a number of interleukins, such as IL-4, IL-5, IL-13, in order to trigger the humoral immune response (mediated by antibodies and the complement system). Finally $T_H17$ cells express IL-17 for the stimulation of phagocytes which eliminate fungi and bacteria from the extracellular media.[129] Therapeutic attempts have been aiming to direct T cell differentiation towards these specialized subsets of $T_H$ cells in order to combat

tumor-induced immunosuppression, viral, bacterial and fungal infections as well as to redirect allergic responses caused by $T_H2$ cells towards favourable $T_H1$ responses.[130]

Several aspects of the signalling pathways still remain among the less understood functions of CLRs. Some lectins of the CLR superfamily, such as Mincle, dectin-2, BDCA2 and CLEC5A induce signalling by assembling with immunoreceptor tyrosine-based activation motif (ITAM)-containing adaptor molecules, like Fc receptor γ-chain (FcRγ) and DAP12. Other C-lectins, for example dectin-1, DC-SIGN, DCIR and MICL (myeloid C-type lectin-like receptor) trigger a cascade of signaling steps by activating various protein kinases and phosphatases which then engage directly or indirectly with their cytoplasmic domains. Although numerous CLRs, like DC-SIGN, BDCA2, DCIR and MICL are capable of initiating signaling pathways that affect the TLR induced gene-expression at the transcriptional or post-transcriptional level, they require signaling by other PRRs to effectively trigger the gene-expression. At the same time, dectin-1, dectin-2 and Mincle can initiate gene expression after carbohydrate ligand binding, regardless to other PRRs. In all the pathways listed here, a central mediator of inducible gene expression is the transcription factor nuclear factor-κB (NF-κB) and other supplementary transcription factors.[131]

CLRs and TLRs can and need to engage in coordinated crosstalk to promote specific gene expression.[132] When DC-SIGN binds mannose-containing pathogens, it modulates the immune response activated by TLR4, but this "conversation" is dependent on the preceding activation of NK-κB by TLR3, TLR4 or TLR5. Upon ligand recognition by DC-SIGN, the serine/threonine protein kinase (RAF1) becomes activated by a complex cascade of events, and by the end, the p65 subunit of NF-κB undergoes phosphorylation. This phosphorylation allows binding of histone acetyltransferase CREB-binding protein and p300 to p65, and the consequent acetylation of p65 on several Lys residues which enhances its DNA binding affinity. Therefore the nuclear activity of p65 and the transcriptional rate increases, and accordingly, the gene expression rate increases as well. Accumulating evidence indicates however, that other factors still remain hidden in the DC-SIGN triggered signaling pathway, as fucose-structures initiate a pathway independent from RAF1 activation.

On the contrary, BDCA2 signalling[133] is facilitated by the non-covalent association with the ITAM-containing signalling adaptor molecule FcRγ, followed by its phosphorylation and recruitment of spleen tyrosine kinase (Syk). This induces the assembly of a signaling complex, comprising caspase recruitment domain family member 9 (CARD9), B-cell lymphoma 10 (BCL-10) and mucosa associated lymphoid tissue lymphoma translocation gene 1 (MALT1). This complex normally participates in the activation of NK-κB or expression of cytokines, but BDCA2 recognition of saccharide ligands results in calcium mobilization via the Syk-dependent activation of another signaling complex. When the high-avidity ligation of ITAM-associated receptors triggers transient calcium signalling and this is coupled with the activation of NK-κB and mitogen-activated protein kinase pathways, a collective signalling with TLRs start pro-inflammatory cytokine production.

As opposed to BDCA2, DCIR and MICL present immunoreceptor tyrosine-based inhibitory motifs (ITIMs) in their cytoplasmic tails and sH2-domain containing protein tyrosine phosphatase 1 (sHP1) and sHP2 are recruited by these motifs upon ligand binding.[134] None of the two lectins are capable of inducing immunological responses alone, they rather alter the signalling pathways initiated by other PRRs. When antibodies bind to DCIR, the CLR-antibody complex is internalized and transferred into endosomal compartments to join TLR8 and TLR9. Upon activation of the CLR, the ITIM motif in the tail becomes phosphorylated and the phosphatases gather around the cytoplasmic tail. The following pathway is yet to be understood, but eventually results in the suppression of IL-12 and TNF production by myeloid DCs (mediated by TLR8) and IFNα and TNF secretion by plasmacytoid DCs (induced by TLR9). In other words, the presence of ITIM motifs in the cytoplasmic tails leads to the inhibition of cytokine response induced by other PRRs. (**Figure 11**)

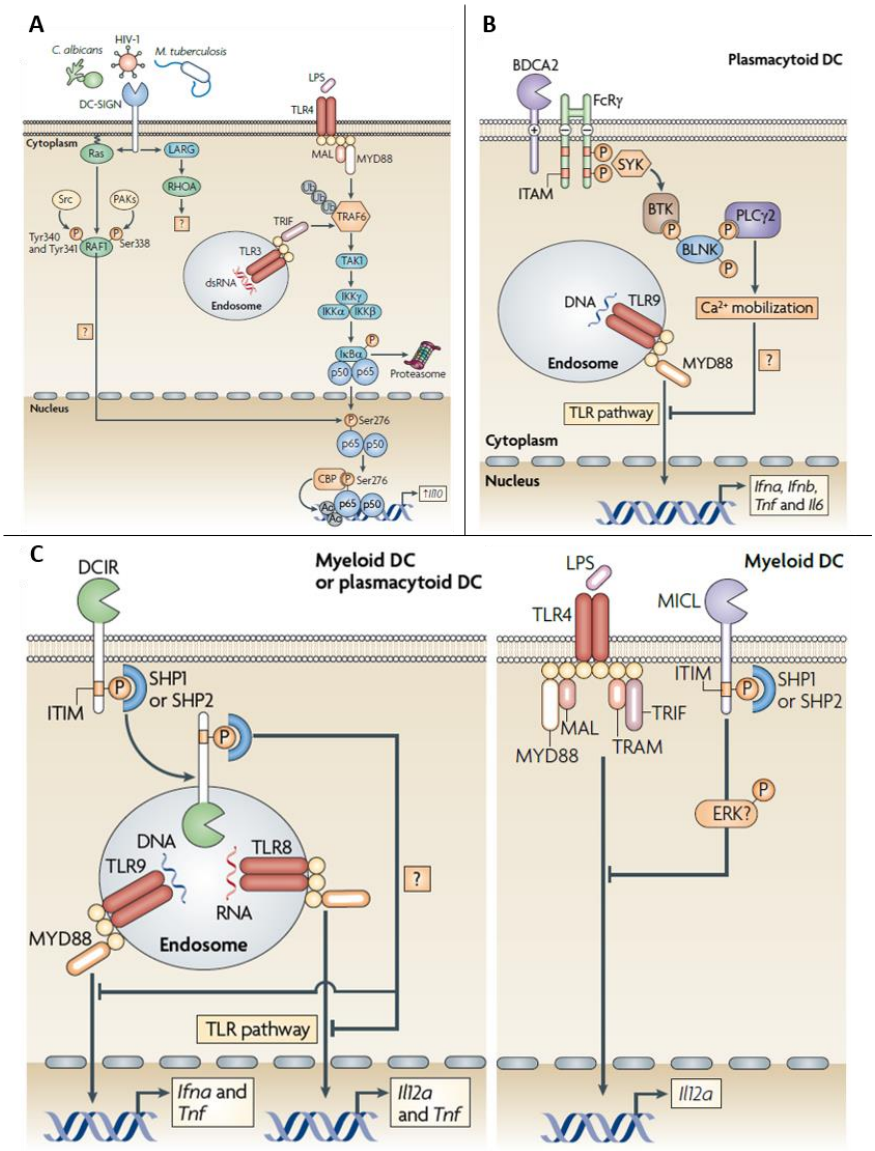



***Figure 11.*** *TLR dependent signalling pathways induced by various CLRs. A. DC-SIGN engages in crosstalk with various TLRs through the RAF1-mediated modulation of the TLR-induced NF-κB activation; and by the end of a complex cascade, the gene expression rate increases.*

*B. The cytoplasmic tail of BDCA2 associates with the ITAM-containing signalling adaptor molecule FcRγ, which in turn recruits Syk-kinase to activate the CARD9-BCL-10-MALT1 complex to initiate NK-κB induced cytokine expression. As a result of joint signalling, pro-inflammatory cytokines are secreted.*

*C. DCIR and MICL contain ITIM motifs in their cytoplasmic tails which associate with sHP1 and sHP2 phosphatases. A collective signalling with TLRs is necessary to suppress cytokine expression induced by other PRRs. Modified from: Geijtenbeek et al.*[131]

Dectin-1 is a particular CLR, as it can initiate gene expression without the help of other PRRs or adaptor molecules.[135] The lectin recognizes β-1,3-glucans on fungal pathogens, and following their binding, induces NF-κB activation. Dectin-1 contains the YXXL motif in its cytoplasmic domain and Syk associates with the phosphorylated group of residues. Similarly to BDCA2, the CARD9-BCL-10-MALT1 complex assembly activates NF-κB – this is the canonical NF-κB signalling pathway. However, dectin-1 also employs the RAF1 mediated pathway (independent of Syk-signalling, but involving crosstalk) and the two routes interlink at the

NF-κB stage. The downstream signalling steps also induce a non-canonical pathway unique of dectin-1, as well as the NLRP3 inflammasome.[136] The non-canonical pathway is presumably a major checkpoint during the induction of specific T-cell responses and together with the canonical NF-κB pathways, they fine-tune the induced cytokine responses by cross-regulating each other. In summary, dectin-1 possesses a specialized, delicate control system over the cytokine gene transcription profile.

Dectin-2 and Mincle both associate in the cytoplasm with the ITAM-bearing signalling adaptor molecule FcRγ and induce its phosphorylation upon ligand binding.[137] The recruitment of Syk kinase is followed by a number of undefined steps, supposedly analogous to the other FcRγ-paired receptors, but eventually the Syk-CARD9-BCL-10-MALT1 complex activates NF-κB and thus induces gene transcription. These two receptors are different from BDCA2 however, as they are able to trigger pro-inflammatory cytokine-production (e.g. TNF, IL-6) independently from TLRs, and because they are present on several myeloid-derived APCs, whereas BDCA2 is expressed solely by plasmacytoid DCs. (**Figure 12**)

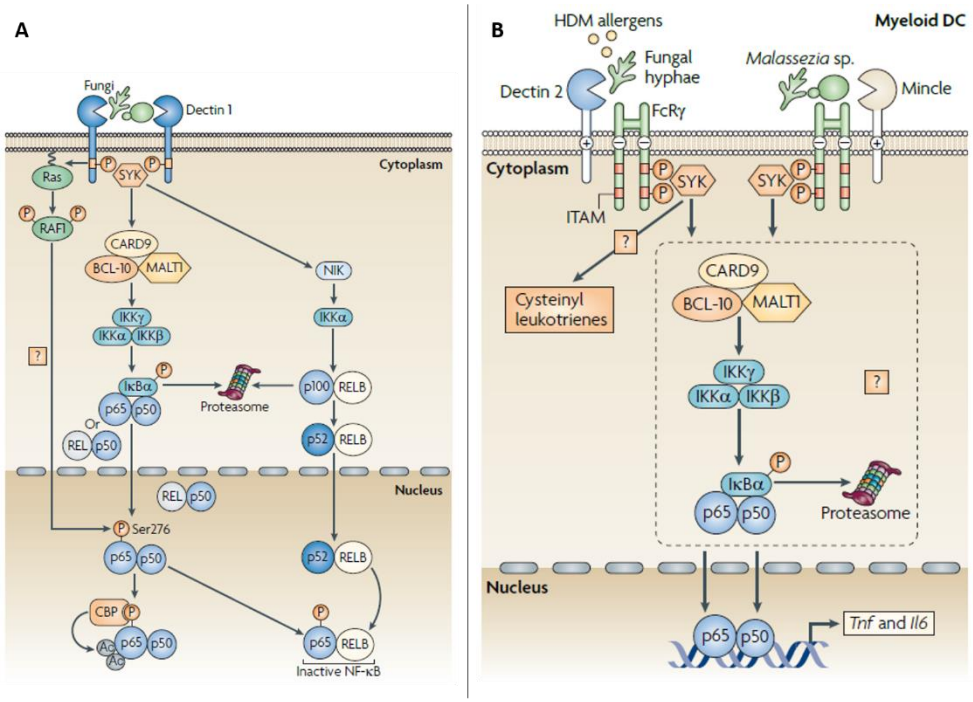



***Figure 12.*** *Independent signalling pathways induced by various CLRs. A. Dectin-1 has a unique signalling pathway, comprising RAF1-, Syk- and CARD9-BCL-10-MALT1 complex-mediated NK-κB activation together with a non-canonical signalling pathway. This delicate system is capable of activating cytokine gene transcription without additional PRR signalling or assembling with adaptor molecules.*

*B. Dectin-2 and Mincle associate with the ITAM containing FcRγ adaptor molecule to recruit Syk-kinase. The CARD9-BCL-10-MALT1 complex activates the NK-κB via unknown pathways to produce pro-inflammatory cytokines – this gene expression induction is not dependent on signals from other PRRs. Modified from: Geijtenbeek et al.*[131]

The abovementioned signalling pathways rarely ever operate alone – they are in constant communication and coordination to collectively steer the T-cell differentiation patterns. The subsequent immunological response relies on the combined signalling pathways induced by the different PRRs present on DCs. The hypothesis that manipulations of the receptor and signalling crosstalk will result in controlled $T_H$ cell differentiation is one of the pillars of CLR targeting therapeutic approaches.

## 1.5 CARBOHYDRATE-BASED DRUGS AND VARIOUS APPROACHES IN LECTIN-TARGETING

### 1.5.1 Carbohydrate-based pharmaceuticals

The extensive involvement of CLRs in biological events renders this group of receptors an attractive target for therapeutic strategies. The main lines of CLR-based research focuses on the prevention of microbial adhesion to the host cells and prevention of the subsequent colonization and invasion of the body; suppression of inflammatory reactions and the modulation of specific immune responses. Phagocytic CLRs, for example, are outstanding targets to promote or dampen certain APC functions, such as DC and macrophage mediated endocytosis and to direct immunological signalling. So far, the majority of lectin-targeted investigations (not limited to C-type lectins) focused on extracellular receptors, like siglecs, galectins, selectins; on bacterial and viral lectins; and on APC receptors, like DC-SIGN.

Selective targeting of CLRs has enormous potential in the cell-specific delivery of drugs, vaccine agents or genes. For instance, asialoglycoprotein receptor (ASGPR) is used to transfer genes exclusively into parenchymal hepatocytes, as the receptor is expressed solely on the surface of these liver cells. This application was first described when coupling artificial tetra-antennary galactose oligosaccharides (selective ligands of ASGPR) to polylysine-plasmid DNA complexes lead to the transfer of the luciferase gene into murine and human hepatocyte cell lines.[138] As this and other cases illustrate, successful attempts to harness CLR-mediated processes for medical purposes usually rely on glycan-protein interactions, while there are only a few examples of non-carbohydrate based CLR ligands with a therapeutic relevance. Before rationally realizing any medical strategies, one needs to look at the already existing saccharide structures with a biomedical application and be cautious with certain boundaries associated with natural glycans to successfully overcome the difficulties.

Glycan and glycan-derived drugs typically rely on the formation or disruption of carbohydrate-protein interactions. The most common goals behind this principle are the modulation of enzyme activity during disease-related glycan synthesis and processing; and the interference with cellular recognition and signalling events through the directed alteration of ligand-receptor interactions. Glycan-structures are therefore good candidates in the treatment and diagnostics of numerous inflammatory, immunological, oncological, neurodegenerative and infectious diseases. In particular, poly- and oligosaccharides are widespread anti-inflammatory, anticoagulant and antithrombotic agents. The marketed poly- and oligosaccharide drugs are usually derived from heparin, obtained from animal tissue, or synthetic low molecular weight heparin (LMWH).[139] Small-molecule carbohydrate and glycomimetic drugs, on the other hand, offer clinical potential, for example, as antibacterial, antiviral and anti-adhesive agents.[140] Glycosylated proteins, comprise adaptive immunity-related agents, such as engineered antibodies,[141] or erythropoietin glycoforms that are used, for example, in the treatment of anaemia.[142] Pharmacopoeia statistics from the US, Europe, Japan and China showed that from ca. 100 marketed and under-development carbohydrate-drugs (98 registered in the US in 2015) about 77% consisted of polysaccharide structures (or derivatives), and 11.5-11.5% of oligo- and monosaccharides (and derivatives).[143]

Carbohydrate-based drugs are found among blood- and cardiovascular therapeutics, like anticoagulants - one example from the LMWH family is Fondaparinux, a synthetic coagulation-cascade enzyme inhibitor.[144] Topiramate is an anticonvulsive fructopyranose derivative used for epilepsy.[145] The drug lacks the detailed description of its action mechanism, although it presumably blocks the action potential that would otherwise be repeatedly stimulated by sustained depolarization of neurons. Acarbose, a drug in the treatment of diabetes mellitus type 2, is an α-glycosidase inhibitor oligosaccharide that blocks the release of glucose from large

polysaccharides.[146] The monovalent iminosugar Miglitol, although structurally quite different, functions as an anti-diabetic drug, based on the same principle.[147] Another relatively simple-structured iminosugar - Miglustat - functions as a competitive and reversible inhibitor of glucosylceramide synthase – an activity which is exploited against lysosomal storage disorders, in mild and moderate type I Gaucher's disease.[148] Antiviral pharmaceuticals also involve examples of sugar-based drugs, like the neuraminidase inhibitors Zanamivir and Oseltamivir in the treatment of Influenza A and B virus.[149] These small molecule ligands bind to the sialic acid-removing enzyme expressed on the viral surface and prohibit the escape of the virus from the single infected cell, preventing its spreading in the tissue (see **1.3**). On the margins of carbohydrate therapeutics, there are some saccharide-based drugs which act based on physicochemical principles - such examples include Icodextrin, a colloidal osmotic agent in peritoneal dialysis.[150]

### 1.5.2 The poor drug-like properties of natural glycans and limitations of carbohydrate drug-design

As indicated by the contradiction between the vast number of medically important functions of glycans and, in spite of this, the limited number of approved carbohydrate drugs, native sugars exhibit suboptimal drug-like properties that hinder their broad therapeutic application. The high number of hydroxyl groups and additional charged functional groups, such as sulfate- and carboxylate-functionalities, lead to high polarity, and thus, a low tissue-permeability of unmodified carbohydrates. This pharmacokinetic burden poses a problem for orally administered drugs, as they are not absorbed into enterocytes in the small intestine by passive transport, thus significantly reducing the active ingredient's bioavailability. Carbohydrate structures that have reached the circulatory system have short plasma half-life because they are subjected to degradation by plasma glycosidases and liver microsomes, and they are rapidly cleared from the blood by the organic cation and anion transport systems found in the liver and kidneys.[140b]

As discussed earlier in details, monovalent carbohydrates display insufficient affinities towards their protein receptors, well-below therapeutic potential, but the synthesis of larger oligo- and polysaccharides is a cumbersome procedure. The synthetic routes almost always comprise numerous steps with low atom efficiency, therefore a reasonable alternative is the use natural saccharide sources. Indeed, LMWHs are obtained by heparin depolymerization by different methods, and many polysaccharide carrier scaffolds are based on modified starch from plants. The control over these processes does not always reach the pharmaceutical standards of certain national authorities though, and so, approvals for drugs from natural carbohydrate sources are often regionally restricted. Poorly controlled purification or derivatization from these saccharide sources can result in mixed composition, vaguely understood mechanism of action and challenging analysis and quality control. Advances in the recent years, however, opened new horizons in controlled oligo- and polysaccharide synthesis. The growing exploration and application of chemoenzymatic reactions,[151] as well as optimized automated carbohydrate-synthesis[152] will undoubtedly accelerate the integration of these techniques into the large-scale production of biologically relevant glycomolecules.

Laborious synthesis is not the sole reason standing behind the scarcity of carbohydrate drugs, however. The glyco-code is encrypted in an intricate structural complexity and heterogeneity that is close to impossible to decipher in the absence of specialized, accurate and sensitive analytical tools. The thorough understanding and quantification of structure and activity relationships (SAR) is an inevitable step during modern rational drug design. At the same time, the accurate observation and measurement of glycan-target interactions is required to identify hit and lead structures and to further develop their properties into drug candidates. The hitherto hindered progress of glycoscience for medical applications has greatly benefited from the technical progress of

the past decades in the fields of NMR spectroscopy, X-ray crystallography, molecular modelling and microarrays.

During binding event studies, two major aspects of carbohydrate-protein binary complexes need to be characterized. On one hand, identification of the functionalities and binding sites that contribute to the interaction will offer later a valuable guidance to improve the chemical composition of glyco-drugs for higher affinity. On the other hand, the three-dimensional organization of pharmacophores ensure the required spatial arrangement of these groups, in order to maximize contact between the ligand and receptor. Presumably the most detailed insight into the molecular basis of glycan-recognition is provided by X-ray crystallographic structures of receptor or receptor-ligand complexes. For many therapeutically relevant lectins, the CRD structure has been described and can facilitate rational ligand-design, but in lack of an accessible X-ray structure, or an appropriate homology model, NMR serves as a useful alternative to examine the binding event. Two techniques - the transferred nuclear Overhauser-effect spectroscopy (trNOESY) and saturation transfer difference (STD-NMR) – are particularly helpful to assess carbohydrate-binding specificities.[153] The former serves as a tool to observe the bound ligand-conformation in the CRD, while the second provides useful information to map the glycan-epitopes and protein sites which are in close contact during binding. Even if NMR data cannot fully replace structural characterization by X-ray crystallography, the two analytical tools complement each other and offer deeper SAR understanding. When none of the two techniques provide sufficient information to start rational design of therapeutically desired ligands, molecular mechanics (MM) and molecular dynamics (MD) are used to open up new perspectives in the structure of ligand-receptor complexes. An additional NMR technique consists of recording the protein NMR in the absence or in the presence of the (sugar) ligand. Using labelled protein allows the identification of signals that are perturbed in the presence of the ligand and therefore are the most likely to interact with it.[154]

New high affinity or selective carbohydrate lead structures can be identified by *in silico* virtual fragment screenings and powerful and facile high-throughput methods, such as the screening of libraries of small molecules in automated assays, NMR-guided fragment-based approaches and glycan microarrays. However, in order to target specific proteins by the means of rational drug design, the drug-like properties of previously identified leads need to be improved and adjusted according to the specific requirements. The pivotal criteria in the development of ideal carbohydrate-therapeutics are the enhanced affinity and selectivity of ligands towards the target receptors; the satisfactory PK/PD/ADMET (absorption, distribution, metabolism, excretion and toxicity) properties for appropriate bioavailability and the straightforward and well-controlled synthetic strategy. All of these requirements point in the direction of two main strategic approaches for the development of novel glyco-drugs – the rational design of optimized glycomimetic ligands, and the construction of multimeric systems to improve affinity through increased avidity.[30a, 140b, 155]

To accomplish these goals, one needs to address and purposely eliminate the above discussed obstacles lying in the way of glycan-based therapeutics during carbohydrate drug-design.[140b] Based on the binding characteristics of a natural ligand in the CRD, a synthetic carbohydrate can be designed to present the crucial epitopes in a pre-organized manner to mimic the required contacts with different sites at the protein surface. Additionally though, in the case of well-defined binding pockets, further elements can also be included and oriented towards the protein surface to form additional contacts, and thus, provide enthalpic contributions. Optimally, these newly included functionalities would increase the otherwise low affinity between the glycomimetic ligand and receptor, and ideally present a very unique binding architecture for the given CRD, which will not be bound by other proteins. In other words, selectivity among receptors can be achieved by accommodating specific epitopes in one CRD which would either not interact with any site in other CRDs (no enhanced affinity) or block binding with others, as a result of a spatial clash (decreased affinity). Once the pivotal points in the binding network are identified, the attention can be turned to improving the drug-likeness of the molecule.

Bioavailability can be improved or adjusted by optimizing the molecular mass, polarity and number of H-bond donors and acceptors within the molecule. These properties can be tuned by elimination of functional groups which do not contribute to binding interactions or the conformation and by the bioisosteric replacement of motifs which cannot be omitted. If absorption by passive transport is inherently impaired, rational design allows the exploitation of active transport systems for example in the intestine, liver, kidney or the brain; by appending specific functional groups that promote the crossing of these barriers. If the glycomimetic contains charged-groups, transporter assays specific for the ion should be performed from early on during the development.
Similarly, regular testing of the metabolic stability throughout chemical modifications of the lead structure helps to elongate the serum half-life of the drug. Enzymatic resistance can be enhanced, for example, by incorporating structural motifs that prohibit the recognition of the molecule by glycosidase and glycosyltransferase enzymes. Occasionally, smartly designed drug release can profit from the activity of enzymes, allowing the oral administration of a prodrug, which will be degraded into an active metabolite with finely tuned ADME profile.
Aglycon motifs in glycomimetic structures, as well as carrier scaffolds and nano-cores in multivalent glycoconstructs potentially pose the problem of toxicity. This might be one of the major pitfalls, for example, in the case of gold nanoparticles (GNPs), and needs to be carefully evaluated. Additionally, similarly to every drug-discovery process, minimizing the costs of development (and possible production) is a basic requirement that one needs to consider.
Oseltamivir - a competitive inhibitor of the neuraminidase enzyme on the Influenza virus surface - is a benchmark among glycomimetic drugs.[156] The molecule, structurally similar to sialic acid, was developed from a carbohydrate lead by the omission and transformation of polar and enzymatically sensitive groups. It is orally available as an ester, and about 80% of the administered prodrug is released as the active metabolite. (**Figure 13**)

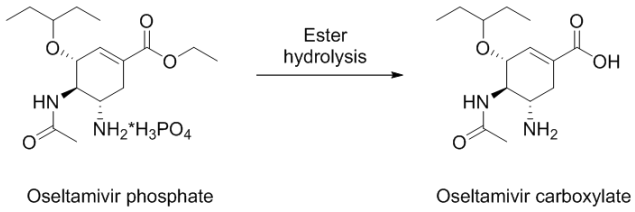



***Figure 13.*** *Oseltamivir drug release from prodrug. The orally administered ethylester is hydrolysed to yield the bioactive carboxylic acid.*

CLR-, and in general lectin-targeting has always been challenging, as even if the binding site is maximally occupied by the carbohydrate ligand, the flat and extended CRD is an easily accessible region in the protein (for further details on CRD structures, see **1.4.1**). Therefore, similarly to clustered, multimeric natural glycoconjugates, artificial constructs can also present several copies of the epitopic saccharide unit. The affinity of multivalent glycoconstructs towards CLRs is highly dependent on the spatial arrangement and rigidity of the carrier scaffold, as well as the length of linkers which tether the monovalent units to the core. Although in the case of densely functionalized multivalent systems, there is not always precise control over the presentation characteristics of the monovalent ligands (GNPs, glycopolymers, NGPs, larger glycodendrimers); well-defined scaffolds can be tailored for certain CLRs. The appropriate spacing of ligands can be tuned by rigid spacers in cases when the spatial presentation and the distance between adjacent binding sites is known (e.g. from X-ray crystallography).[157]

## 1.5.3 Some examples of successful lectin-targeting

### 1.5.3.1 The broader class of mammalian lectins

1.5.3.1.1 Siglecs

Examples of successful lectin targeting can be found in the Siglec family - co-receptors that play a role in leukocyte signalling during adaptive and innate immune responses. One member of this lectin class, the myelin-associated glycoprotein (MAG) inhibits neurite outgrowth after injuries of the central nervous system (CNS).[158] Small molecule antagonists of MAG, displaying nanomolar affinities and optimal stability in the spinal cord fluid, are encouraging results to pursue the glycomimetic strategy in the search for siglec inhibitors.

1.5.3.1.2 Selectins

Several examples of promising antagonists have been developed against the three members of the selectin-family. E-selectin is primarily expressed on endothelial cells upon stimulation by inflammatory cytokines; P-selectin is present on activated platelets, and intracellularly stored supplies of the lectin can be rapidly recruited to the platelet surface if necessary, while L-selectin plays a role in homing and trafficking of leukocytes. Targeting selectins has an attractive therapeutic potential in the treatment of diseases where cell-adhesion, escape from the blood stream and migration of leukocytes has been implied in the pathologic progress. All three selectins recognize $sLe^A$ and $sLe^X$-type cancer-related antigens and therefore they have been associated with the false recognition of tumor cells as migrating leukocytes; their extravasation from the bloodstream and metastasis to other organs.[159] Selectins are also involved in acute and chronic inflammation processes, and this has been exploited to develop glyconanoparticles for the *in vivo*, pre-symptomatic visualization of brain-diseases.[160] Dextran-coated nanoparticles with high Fe-content were functionalized with glycans through an amine-linker, and the obtained glyconanoparticles were efficiently detected as cerebral endothelial markers of acute inflammation by magnetic resonance imaging (MRI).[161] Rivipansel (GMI-1070), a pan-selectin glycomimetic antagonist reduces soluble E-selectin levels in the bloodstream.[162] The drug-candidate promotes accelerated recovery from vaso-occlusive crisis in patients with sickle-cell disease and has successfully completed phase I and II clinical trials. Another glycomimetic inhibitor of E-selectin has been developed for potential anti-inflammatory applications. A $sLe^X$ mimic was linked with the best ligands identified for the second binding-site by NMR-guided fragment-based approach.[163] $K_D$ values of the optimized structures raged within 30-89 nM and the ligand-receptor complex half-life reached minutes, as opposed to the typical value of a few seconds.

### 1.5.3.2 Anti-bacterial agents

Probably one of the most interesting perspectives in lectin targeting lies within the development of anti-adhesive agents against microbial infections.[140a] Bacteria in the enteric, oral and respiratory mucosa benefits from adherence to host tissue because it increases the chances of host invasion and colonization, as well as unperturbed biofilm-formation. Pathogens gain improved resistance against natural clearing mechanisms of the host, like destruction by immune-factors, bacteriolytic enzymes and they are less likely to be eradicated by antibiotics. At the same time, they gain better access to nutrients, and can more easily deliver their toxins. The alarming emergence of antibiotic-resistant bacterial strains and the possibility of intervention at an early stage of pathogen invasion turned the focus towards anti-adhesive therapies in the treatment of many infectious diseases. Anti-adhesive agents which inhibit lectins involved in pathogen-host cell interaction can prevent the adhesion of microbes to the host, and their spreading within the body, and they are effective in combatting

biofilm-formation. One of their major expected advantages, though, is that they lack bactericidal effect, and thus they do not contribute to the propagation of resistant strains, like the excessive use of antibiotic drugs. Pathogens typically adhere to host tissue by two methods: either by host cell PRRs recognizing carbohydrate structures present on the invader's surface, or by expressing adhesin proteins themselves, which will then bind to host-cell surface glycans. Bacteria often express several types of adhesins, therefore a potent anti-adhesive drug must inhibit multiple lectins of this kind.

1.5.3.2.1 FimH - *E.coli*

FimH has been linked to uropathogenic *E. coli* (UPEC),[164] as well as with adherent invasive E. coli (AIEC), isolated from the small intestine of patients with Crohn's disease.[165] UPEC causes urinary tract infection by migrating from the rectal microbiota through the urethra until it reaches the bladder. Infections are usually treated with oral antibiotics, however, and therefore recurrent cases hold the threat of drug resistance. Studies of the CRD of FimH adhesin over 200 strains showed an invariant structure: a deep sugar-binding pocket with a specificity for mannose, where the entrance is narrowed by a conserved group of residues, the so-called tyrosine-gate (Tyr48, Tyr137, Ile52, see **1.5.3**).[47, 166] Therefore, rational glycomimetic development began to form additional interactions with the hydrophilic sidechains found in this region.[167] In the synthetic ligands, a Man monosaccharide element creates a H-bonding network in the pocket with its 2, 3, 4 and 6 hydroxyl groups, and an α-linked hydrophobic aglycon motif that interacts with the gate. Alkyl chains (e.g. *n*-butyl) were found to significantly increase the affinity of monovalent ligands, but the *n*-heptyl substituted glycomimetic dimerized two CRDs *in vitro* which causes a likely overestimation of the observed binding affinity. Nanomolar antagonists were developed by incorporating aromatic aglycon motifs in the anomeric position because these groups can form favourable π-π interactions with the Tyr residues.[168] The most potent inhibitors belong to the class of biphenyl derivatives. During a PK/PD profile optimization process,[169] 3'-chloro-4'-(α-D-mannopyranosyloxy)biphenyl-4-carbonitrile emerged as the best drug candidate – an orally available molecule that is absorbed in the intestine, renally excreted and has a finely-tuned residence time in mouse bladder to reduce the bacterial load 1000-fold.

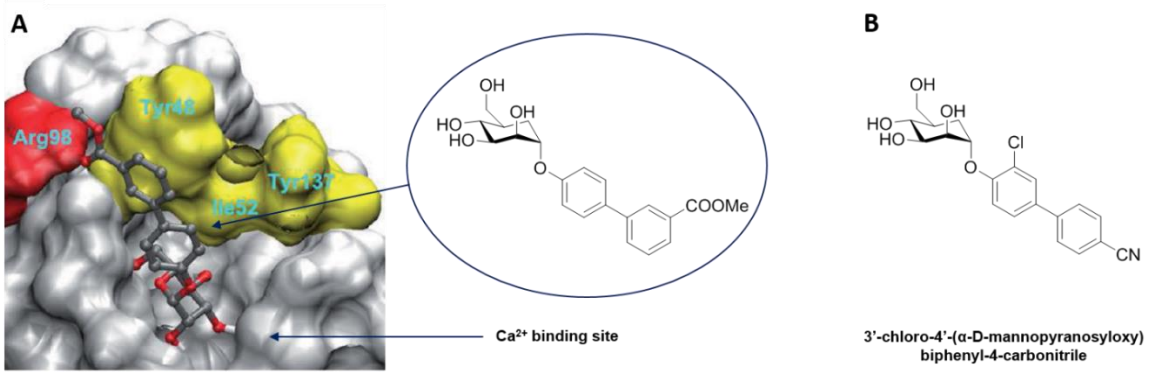



***Figure 14.*** *FimH antagonists and binding mode in the CRD. A. X-ray crystal structure of a biphenyl-group containing a mannose-based lead, complexed with FimH CRD (PDB code: 3MCY). The aryl groups pack against the tyrosine gate above the principal binding site, creating π-π interactions with the lipophilic residues in this region. B. Optimized structure with enhanced PK/PD properties. Modified from: Pang et al.*[168]

1.5.3.2.2 PA-IL and PA-IIL (LecA and LecB) – *Pseudomonas aeruginosa*

*Pseudomonas aeruginosa* is a Gram-negative bacterium present in healthy flora that becomes a severe threat for hospitalized, immunocompromised or otherwise critically ill patients and for patients with cystic fibrosis. The traditional treatment involves antibiotics; although the bacterium is particularly resistant, mainly as a result of its capability to form biofilms.[170] Bacterial cells which are shielded from external conditions by this protective clustering are less likely to be accessed by host-defense mediators and 1000-times more resistant against antibiotic treatment.[171] Attempts were made to combat the biofilm formation of *P. aeruginosa* by autoinduction of quorum sensing triggered by analogues of the native mediator.[172] Quorum sensing is a coordinated mechanism of bacteria to adjust gene-expression levels to the local population density. The expression of two virulence factor lectins – PA-IL (LecA) and PA-IIL (LecB) – are controlled by this machinery.[173] These tetrameric lectins are originally soluble intracellular proteins, but upon their release, they promote adhesion of the bacteria to host cells and biofilm formation by cross-linking glycans expressed on bacterial cells.[174] LecA recognizes terminal α-linked Gal saccharides and requires one $Ca^{2+}$ ion for binding,[175] while LecB has an exceptionally high affinity - 1.6 μM - towards L-Fuc (in lectin-monosaccharide relation) and uses two $Ca^{2+}$ ions.[176] Bacterial adherence can be reversed in the case of both LecA and LecB by the addition of D-Gal (for LecA) and D-Man (for LecB). These monosaccharides are included as a supplementary agent during combination therapies with Tobramycin – an antibiotic drug particularly effective against *P. aeruginosa*, that prohibits mRNA translation into protein synthesis.[177] In spite of the $Ca^{2+}$ requirement, the lectins do not share sequence or folding characteristics with CLRs, but LecA displays a highly similar Gal-binding mode to animal lectins, like the QPDWG mutant rat-MBL and the tunicate Gal-specific CLR, TC14, isolated from *Polyandrocarpa misakiens*. Interestingly, LecA binds Gal 10-fold more strongly than animal CLRs, which can be attributed to an additional interaction with 6-OH.[175] LecB, on the other hand, in addition to the unusually low $IC_{50}$ value for monosaccharides, has a binding affinity towards $Le^A$ in the nanomolar range (210 nM). In an attempt to develop simpler glycan molecules, the trisaccharide structure was simplified to a Fucα(1-4)GlcNAc disaccharide, but the higher flexibility of the molecule resulted in an increased entropy loss that was only slightly compensated by newly formed H-bonds.[178] The breakthrough in LecB glycomimetic ligand development was provided by the Titz group, after screening 4 series of α-mannoside glycomimetics in competitive binding assays on ELLA (enzyme linked lectin assay) plates.[179] The highest-affinity binders were identified as 6-modified α-mannosides, particularly the cinnamide and sulfonamide derivatives. (**Figure 15**) X-ray images were not able to provide sufficient information about the complex structure, but further biochemical assays, NMR studies, MD simulations and bacterial adhesion tests showed two distinct binding modes where the Man core was accommodated the binding-cleft, while the aromatic rings got in close contact with the protein surface. Additionally, in the case of the sulfonamide derivative, a further H-bond was formed between the NH and Asp96. The most promising ligands were transferred into fucosylated glycomimetics, decorated with cinnamide and sulfonamide moieties, deliberately aiming for increased drug-likeness.[180] By designing highly potent monovalent ligands, the authors intended to circumvent two major issues of LecB inhibitors: the probability of stabilizing the biofilm by bridging several tetrameric LecB proteins by larger oligosaccharides; and the restricted use of multivalent constructs by topical application, instead of oral administration. The most potent ligands compete with the affinity of $Le^A$ – for example, a tiophene sulfonamide decorated fucoside reached a $K_D$ value of 290 nM. These derivatives have improved pharmacokinetic and pharmacodynamic behavior – low MW, oral availability, high concentrations in plasma and urine and receptor-ligand half-lives of minutes. Moreover, they show selectivity towards LecB over host lectins, low toxicity; high metabolic stability *in vitro* and superior binding kinetics and thermodynamics. Most importantly, unlike the natural ligands, the monovalent glycomimetic ligands were able to efficiently inhibit biofilm formation by *P. aeruginosa*.

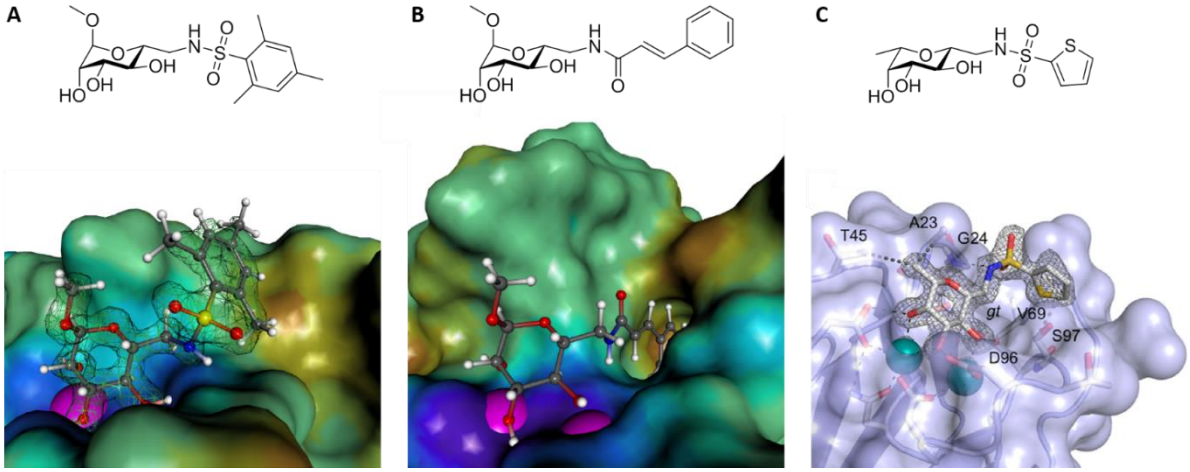



***Figure 15.*** *LecB inhibitors and their crystal structure in the CRD. A and B: 6-OMe-Fuc-based sulphonamide and cinnamide inhibitors and their binding mode in the LecB CRD (PDB codes: 3ZDV, 5A3O respectively). C: Optimized structure of a sulphonamide derivative ($K_D$: 290 μM) with improved selectivity and pharmacokinetic properties, and its binding mode in the CRD (PDB code: 5MAY). Modified from Hauck et al. and Sommer et al.*[179-180]

As mentioned earlier, lectin-targeted anti-adhesive therapeutic approaches can involve either the adhesins expressed by pathogens, or the receptors found in the host – primarily APC-surface receptors. Dendritic cell-induced signalling pathways and immune responses are greatly influenced by the type of encountered pathogen and the nature of the PRR (TLRs and CLRs) that captured the microbe. Various DC subsets express different CLRs, such as DC-SIGN, DC-SIGNR, Langerin, MR, DEC205, BDCA-type lectins, dectin-type lectins, DCIR, DCAR, DCAL-type lectins, CLEC1, MGL - the majority of these work as endocytic receptors. MR and DEC205 display 8-10 CRDs on a single neck and are type I transmembrane proteins, while the rest belong to the type II transmembrane receptors. Arguably, the most important CLR involved in first-line host-defense is DC-SIGN, a dendritic-cell expressed PRR that is among the firsts to encounter pathogens within the body. Precisely DC-SIGN's versatile role in the immune response, and its interlaced functions with other what makes DC-SIGN an attractive target for pathogens to escape defense mechanisms and infect the host.

# 1.6 DC-SIGN TARGETING IN THE CENTRE OF ANTI-MICROBIAL AND IMMUNOMODULATING STRATEGIES

## 1.6.1 The structure and primary functions of DC-SIGN

Humans express two closely related DC-SIGN homologues, DC-SIGN (CD209) and DC-SIGNR (L-SIGN, DC-SIGN related protein, CD209L, CD299), whereas the mouse genome encodes 8 homologues (SIGNR1-8), although SIGNR-6 is likely a pseudogene.[181] The mouse homologue family, however, has undergone such a significant divergence that none of the SIGNR receptors are real orthologues of DC-SIGN – their functions, signalling pathways, cell and tissue distribution patterns and ligand-binding characteristics are only partially overlapping with the human homologue.[182] *In vivo* studies of DC-SIGN are therefore restricted to SIGNR knockout mice and transgene introduction of human DC-SIGN – the majority of information about the biological role of the human receptor originates from *in vitro* investigations.

DC-SIGN is primarily expressed by immature DCs in the peripheries and mucosa, like the skin, lungs, rectum, cervix and placenta, but it is also found on the surface of mature DCs that migrated to lymphoid organs. DC-SIGN is not exclusively associated with DCs, though, as the receptor is present on other APCs in immunoprivileged tissues, such as macrophages in the dermis and intestinal mucosa, tumor-associated inflammatory M2 macrophages, lymph node sinus endothelial cells and Hofbauer cells in the placenta.[183] The expression of DC-SIGN is stimulated by IL-4, a cytokine that is produced by $T_H2$ cells and additionally promotes naïve T helper cell differentiation into $T_H2$ cells in a positive feedback-loop.[184] DC-SIGN is often co-expressed with other CLRs (e.g. MR), but studies until now indicate that this receptor recognizes a broader repertoire of pathogen-associated ligands than other DC-related PRRs with a mannose-specificity.[185]

The CRD of DC-SIGN exhibits the Glu-Pro-Asn motif around the primary $Ca^{2+}$ binding site,[83] and accordingly, the lectin binds Man and Fuc-containing stuctures: internal mannose trisaccharides (Manα1-3Manα1-6Man) in branched glycans; terminal mannose disaccharides (Manα1-2Man) - for example in the *Mycobacterium tuberculosis* bacterial cell-wall component lipoarabinomannan (ManLAM); the four Lewis-type antigens – $Le^X$, $Le^Y$, $Le^A$, $Le^B$, and $sLa^A$ but not $sLe^X$.[96, 186] Notably, DC-SIGN has a generally higher affinity towards fucosylated natural ligands than towards mannosylated ones, and it is the main receptor for $Le^X$ on DCs,[187] while mannosylated ligands are captured by several other CLRs expressed on the surface of APCs. The CRD structure-ligand binding relationship will be discussed in more details later on.

A DC-SIGN monomer consists of a globular CRD oriented towards the extracellular site of the cell-membrane; a neck domain that includes 7 complete and one incomplete 23-amino acid tandem repeat units; the transmembrane region and a short cytoplasmic tail. Characteristic recycling and internalization motifs – Leu-Leu, triacidic Glu-Glu-Glu cluster and an incomplete ITAM - are found in the tail region, which direct the captured pathogens towards lysosomes and MHC class II containing late endosomes for processing and presentation to T-cells.[185a]

*In vivo*, DC-SIGN organizes into coiled coil oligomers by autonomous tetramerization of the neck domains in a highly coordinated process that presumably skips the dimerization state. As these regions are not fully helical (ca. 40% α-helical structure) and there is only little interaction between the individual CRDs, the four clustered monomers are flexibly linked.[96] This loose, 4-stranded bundle differs from several other CLR oligomers that contain fully helical necks and are therefore more intimately associated. At least 6 repeat units are needed for oligomerization in the SIGN family of CLRs.[188] (**Figure 16**)

This allows multimeric interactions between the extracellular domain (ECD) and pathogens leading to increased affinity through enhanced avidity, as well as a more efficient screening of the environment for ligands.

The CRDs are pointed ca. 32 nm away from the cell surface and the individual CRDs are presented approximately 5 nm away from each other.[189] The spatially constrained arrangement of the CRDs suggests that merely the multimeric presentation of ligands is not sufficient to enhance affinity, their geometry can improve or reduce avidity effects. Neck polymorphism has been described on the genetic level (5 alleles),[190] and the resulting variable neck lengths affect susceptibility to pathogens, such as HIV-1. Menon *et al.* reported a certain degree of flexibility upon ligand binding,[191] as established by surface force measurements between DC-SIGN presenting lipid bilayers and $Man_9$ functionalized neoglycolipids (NGLs). Ligand engagement was accompanied by a "contraction" in the CRD length – from 33 nm (unligated) to 27 nm (ligated form) – and this conformational change presumably facilitates the engagement of all CRDs when approaching pathogens that exhibit numerous copies of ligands.

Once the receptor-ligand complex is formed, the neck plays a central role in the internalization of pathogens for further processing in endosomes and lysosomes of decreasing acidity. The oligomerization state of the domain depends on the pH and ionic strength of these compartments, and in the studies of Tabarani *et al.*,[189] DC-SIGN ECD tetramer dissociation into monomers was observed around pH ~ 5, with an approximate transition midpoint at pH ~ 5.9. It was previously shown *in vivo* that DC-SIGN releases its ligands in the acidic luminal environments around pH ~ 6, and the demonstrated acidity-dependent change in the oligomerization state, the extended receptor conformation and the CRD organization and spatial presentation may provide the molecular-level explanation for this phenomenon.

In addition to tetramerization, DC-SIGN expression on the cell surface is organized also on a second level. The lectin oligomers assemble into well-defined membranous microdomains, and these sites are located in the so-called lipid rafts – a region in the cell membrane with high concentrations of cholesterol and glycosphingolipids.[192] Importantly, microdomain formation by the receptor enhances virus binding and uptake (e.g. HIV-1),[102] compared to randomly localized DC-SIGN. Tetrameric presentation and organization into microdomains allow the formation of multiple interactions between the CRDs and PAMPs – and a facilitated uptake of pathogens that differ in size. As it was discussed in earlier sections, multivalent ligands (both natural and artificial) benefit from this synergistic affinity enhancement of individual interactions, and avidity can be purposefully exploited to design inhibitors of DC-SIGN.

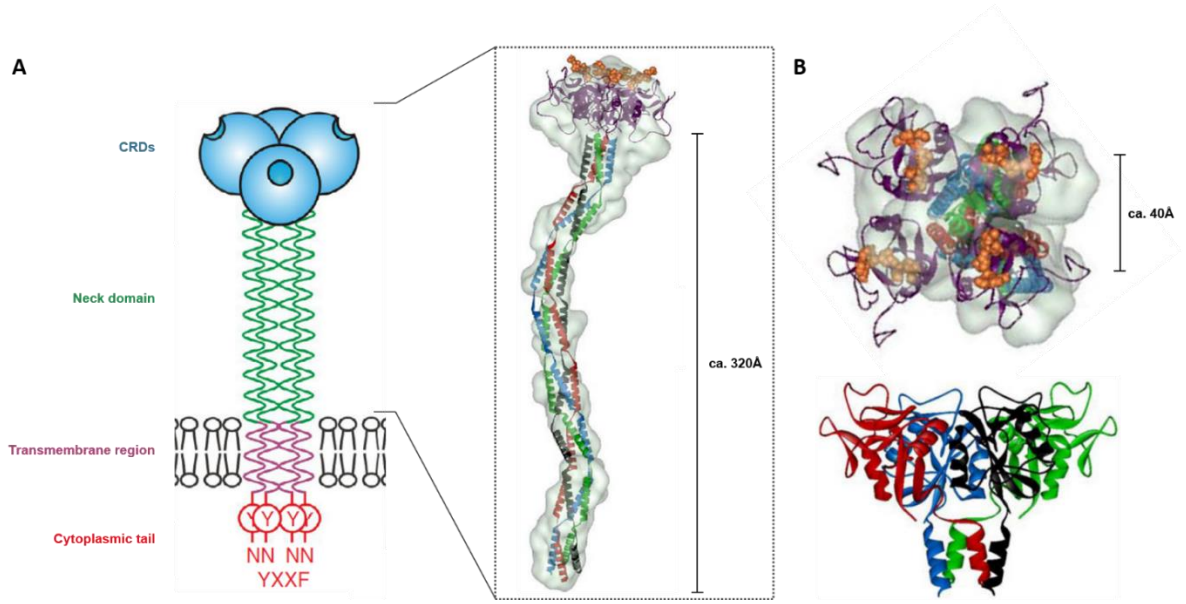



***Figure 16.*** *The structure of DC-SIGN tetramer and CRD. A. Schematic image of the type II transmembrane receptor. Four CRDs are projected towards the extracellular surface through a long, coiled-coil neck that contains seven and a half tandem repeats. The transmembrane domain connects the extracellular domain (overall conformation represented on the right, based on a GASBOR model) to a short cytoplasmic domain that contains di-leucine (LL) and triacidic (EEE) clusters, as well as an incomplete ITAM-motif.*

*B. CRD tetramers in top- and side-view. Four copies of $Man_3GlcNAc_2$ are ligated at the four sugar-binding sites (based on the same GASBOR model). Two neck repeats are included in the bottom image. Modified from: Drickamer et al. and Tabarani et al.*[94, 189]

The primary role of DC-SIGN is pattern recognition to capture microbes encountered by DCs, followed by pathogen internalization and submission to classical endolysosomal degradation and MHC class II presentation to T-lymphocytes. Additionally, DC-SIGN is able to modulate signalling pathways induced by other PRRs, such as TLRs (see the earlier discussion of DC-SIGN-TLR crosstalk), but it cannot initiate DC maturation and cytokine secretion alone.[132] DC-SIGN recognizes molecular patterns on pathogens, as well as self-glycoproteins, such as intercellular adhesion molecules (ICAM). This capability endows DC-SIGN with an important function as cell-adhesion mediator during DC migration and antigen presentation. More specifically, DC-SIGN binding to heavily glycosylated ICAM-2 and 3 drives important cell-cell adhesion processes, regulated by the cell and tissue specific glycosylation patterns. DC-SIGN uses ICAM-2 present on endothelial cells as a "handle" during DC migration from peripheries and in the blood to lymph nodes.[193] This CLR-ligand promoted trafficking is similar to the selectin-mediated rolling of lymphocytes along the blood vessel linings. DC-SIGN-ICAM-3 interactions, on the other hand, are crucial in DC-T-cell signal transmission, because the transient adhesion gives time to T-lymphocyte TCRs to find and recognize MHC-antigen complexes on the DC surface, and the created immune synapse is further stabilized by the adhesive receptor. The importance of DC-SIGN's contribution to this immunological communication was illustrated by inhibiting the CLR with mABs – an obstruction that abrogated DC-T-lymphocyte clustering and prevented resting T-cell proliferation.[21] ICAM-3 is expressed by endothelial cells and neutrophils as well, therefore DC-SIGN-ICAM-3 binding facilitates attachment to these cells too. Other self-glycoproteins recognized by DC-SIGN include macrophage receptor-1, carcinoembryonic antigen, butyrophilin, the Fc (fragment crystallizable) region of immunoglobulins and milk bile-salt-stimulated lipase.[182]

An intriguing study from Geijtenbeek and co-workers proposes that the receptor is unique among EPN motif-containing CLRs in the sense that it can distinguish and react differently to Man and Fuc structures.[194] In

healthy higher-order mammals, mannosylated saccharide conjugates are present almost exclusively in the ER, during the maturation of glycoproteins (post-translational modification) and on the surface of invading microbes (viral envelope, bacterial cell-walls). According to the reported observations, if DC-SIGN detects Man-structures, the TLR induced DC response is modulated towards pro-inflammatory reactions (e.g. chemokine expression) to eliminate the pathogen or the damaged cell which has exposed these – normally intracellular – elements. On the contrary, fucosylated saccharides are expressed by a large number of healthy cells, and these antigens require tolerance by the immune system. In this case, DC-SIGN contributes to immune-homeostasis by steering the TLR induced pathways towards anti-inflammatory signalling. Notably, a few pathogens learned how to subvert this mechanism, as this is seen in the case of some *Helicobacter pylori* mutants that express $Le^Y$ antigens and accordingly, exhibit improved pathogenicity.
The suggested differential recognition of mannose and fucose and the opposite signalling pathways induced by the two epitopes are not the only characteristics of DC-SIGN that opportunistic pathogens hijack. Microbes exploit DC-SIGN to infect DCs and furthermore, to shift the naïve T-cell differentiation towards $T_H1$ or $T_H2$ cells – towards a response that favours the pathogens' persistence against eradication. Pathogens, like *Schistosoma mansoni* and *Helicobacter pylori*, expressing fucosylated structures exert pressure on the $T_H1/T_H2$ equilibrium towards adaptive $T_H2$ responses, mediated by atypical NK-κB family member Bcl3 activation. Indeed, a common feature among these microbes is that they cause long(lifetime)-lasting, chronic infections.[187, 195] Because of the IL-4 dependent expression of DC-SIGN, high levels of the receptor expression are related to $T_H2$ cell polarization.

### 1.6.2 DC-SIGN helps the uptake of pathogens and promotes infections

Interestingly, DC-SIGN was discovered only "after" the identification of one of its pathogen ligands, as during its first description in 2000 by Geijtenbeek *et al.*,[196] the receptor was found to be identical to a previously cloned HIV-1 envelope binding CLR from placenta, reported in 1992 by Curtis *et al.*[197] Since then, numerous pathogenic ligands were associated with DC-SIGN. Viruses are bound primarily through their exposed glycoproteins in the viral envelope – these include HIV-1 (DC-SIGN ligand: gp120),[196] Ebola virus (GP1),[198] Dengue virus (E glycoprotein),[199] Hepatitis C (E1, E2),[200] Cytomegalovirus,[201] SIV-1 (Simian Immunodeficiency Virus-1),[202] SARS-CoV (Severe Acute Respiratory Syndrome Coronavirus; S glycoprotein),[203] Lassa virus (GP1),[204] West Nile virus (E glycoprotein).[205] Bacteria, such as *Helicobacter pylori*,[195] *Klebsiella pneumonia*,[185a] *Mycobacterium tuberculosis*[206] and *Neisseria meningitides*[207] are recognized by bacterial cell-wall constituents, like ManLAM and phosphatidylinositol-mannosides (PIMs).[208] DC-SIGN captures also parasite-specific ligands, such as Man-capped lipophosphoglycans and $Le^X$-type structures - for example, from *Leishmania* (*Leishmania pifanoi*)[209] and *Schistosoma* (*Schistosoma mansoni* soluble egg antigen; and fungal epitopes of *Candida albicans*[210] and *Aspergillus fumigatus*.[211] DC-SIGN has been associated furthermore with allergic reactions through the capture of house dust mite and dog-allergens.[212] Detailed carbohydrate profiling of the lectin allows the prediction of pathogen recognition.[213]
The envelope glycoproteins of various viruses contain diverse *N*-linked oligosaccharides whose differential expression leads to discriminative recognition by DC-SIGN, suggesting selectivity in pathogen binding. For example, vesicular stomatitis virus doesn't bind to DC-SIGN, in spite of its heavily glycosylated viral capsule.[214] In other viruses, for instance HIV-1, the role of the high affinity DC-SIGN ligand gp120 is not restricted to recognition by the CRD as a PAMP, but the glycoprotein plays a role in the viral uptake as well.[215]

1.6.2.1 DC-SIGN helps HIV *cis*- and *trans*-infections

HIV-1 is the most common and pathogenic strain of HIV that causes HIV infection and leads, over time, to acquired immunodeficiency syndrome (AIDS). This is due to the fact that upon entering the host, the viral particles are able to escape and impair the eradicating immune response by infecting vital cells of the immune system, such as dendritic cells, macrophages and $CD4^+$ T helper cells. Infected $CD4^+$ T helper cells are then killed by several mechanisms, either directly by the virus itself, or by a number of virus-induced immunological, self-targetting processes (for example by $CD8^+$ cytotoxic T-lymphocytes). The drastic decrease in $CD4^+$ T-cell count results in abolishment of the cell-mediated immune system, and the host becomes susceptible to infections that would otherwise be no challenge to a healthy immune system.

The main route of virus transmission from host to host is during sexual intercourse, but HIV-positive mothers can transmit the virus (as free viral particles and infected immune cells) during childbirth - through the exposure to blood and vaginal fluid, or during lactation via breastmilk (mother-to-child transmissions). During sexual transmission, dendritic cells in the mucosal genital tissue are among the first host cells to encounter the virus. DCs capture HIV-1 at the site of entry, and then migrate to lymphoid organs, effectively transporting the virus to T helper cells. Here, HIV-1 is presented and transmitted to $CD4^+$ lymphocytes, or – in higher concentrations – to other DCs, and so, the immune cell infection process begins.

HIV-1 uses several pathways to enter DCs, replicate in these cells and infect other immune mediators. On the surface of DCs, both TLRs and CLRs encounter HIV-1, although the receptors recognize different PAMPs on the viral envelope. DC-SIGN binds with high affinity the mannosylated residues of gp120 (nM range for the entire ECD, mM in the case of individual CRDs),[216] but the glycoprotein is found at relatively low density on the viral capsule as trimers, altogether in 8-10 copies. The receptor-gp120 interaction triggers several mechanisms that protect the virus from intracellular degradation, promotes its transmission to T-cells and other DCs, and initiates viral replication. Even though the mechanisms are not entirely understood on the molecular level, gp120 takes part in all the crucial steps from capture, through internalization and replication, to presentation to other cells. Once HIV-1 entered DCs, the virus is oriented into non-lysosomal acidic intracellular compartments, but the receptor-bound virus retains its potency to infect other cells, presumably because gp120 maintains its native conformation that protects the virus from destruction.[185a] DC-SIGN-HIV-1 complexes "hide" close to the DC membrane and are relatively stable until they are recycled to the cell surface, to present the virus particles to other immune cells. The protective role of DC-SIGN was indicated in studies when deletion of the receptor's cytoplasmic tail or neutralization of the pH in the intracellular organelles - where the DC-SIGN-HIV-1 complex resides within DCs – significantly reduced viral uptake and *trans*-infection to T-cells.[215] Binding to HIV-1 subverts DC-SIGN's conventional signalling routes and modulates TLR induced adaptive responses.[132] In particular, TLR-8 is able to stimulate NF-κB transcription factor upon activation by viral ssRNA, but for successful, full length transcription and thus virus replication and dissemination, DC-SIGN-HIV-1 stimulated pTEF-b transcription-elongation factor is also necessary.

Interestingly, for an enhanced HIV-1 *trans*-infection to T-lymphocytes, DC-SIGN-mediated internalization and intact storage of the viral particles is not required - yet, the lectin increases the rate of transmission to T-cells, even if it loses its internalization capacity.[217] HIV-1 uptake and replication in DCs is initially mediated by receptors, like CD4 and chemokine receptors, such as CCR5 and CXCR4, but gp120 binding by the – in this context - co-receptor DC-SIGN enhances the interactions between CD4 and HIV-1, and it is absolutely required for successful production of new virus particles. During replication, viral uptake by CD4 and similar receptors is followed by fusion with the cell membrane, viral uncoating, reverse transcription of the virus ssRNA and eventually the integration of HIV-1 in the host genome.

When infected DCs get in close contact with T-lymphocytes in lymph nodes (*trans*-infection), or immature DCs (*cis*-infection), receptor-bound HIV-1 recycles back to the DC surface and is presented to T-cells. DC-

SIGN plays a stabilizing role in the formation of these infectious synapses, and again, facilitates virus transmission to T-cells.[218]

Although some aspects of DC-SIGN's involvement in all the above mentioned interlocking processes still remains unexplored, it seems that the CLR is an unavoidable, key-mediator of HIV-1 infection and dissemination. This is greatly supported by DC-SIGN inhibition experiments – blocking the receptor (e.g. with the natural ligand mannan or CLR-specific antibodies) prevents binding to viral gp120 and as a result, abrogates DC infection and virus transmission to T-lymphocytes. (**Figure 17**).

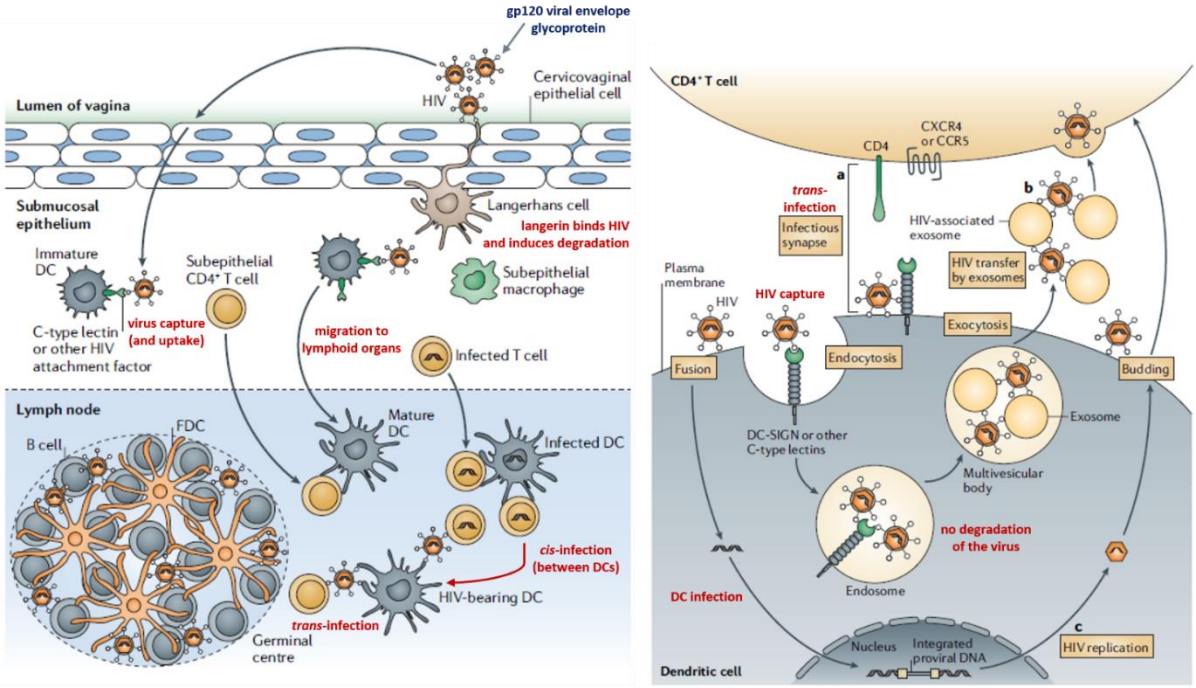



***Figure 17.*** *DC-SIGN mediated HIV-1 capture, internalization and trans/cis-infection. The main routes of virus entry include sexual intercourse, childbirth and lactation. In the genital, rectal, cervical and skin mucosa, the viral particles encounter DCs. DC-SIGN, a PRR expressed on the surface of DCs, recognizes the viral envelope protein gp120, but upon binding, the typical signalling pathways induced by DC-SIGN and TLRs are subverted to favour the viral survival. As a result, after capture and internalization, the virus avoids the classical lysosomal degradation. The unharmed virus is able to replicate in DCs and infect other DCs (cis-infection), and furthermore, to infect CD4+ T-cells in the lymph node (trans-infection). Viral particles are transmitted to T-cells with the help of DC-SIGN. The drastic loss of T-cells leads to the long-term impairment of immune responses, otherwise known as AIDS. Modified from Wu et al.*[219]

1.6.2.2 DC-SIGN mediated uptake of other pathogens

Similarly, DC-SIGN is the target of other viral infections, like Ebola, with analogous, glycoprotein-mediated cellular entry.[198] Ebola virus causes severe hemorrhagic fever with alarmingly high mortality rates during outbreaks. DC-SIGN functions as a co-receptor during Ebola virus uptake by DCs and then both as a *cis*- and *trans*-receptor for the infection of DCs and T-lymphocytes, respectively. The clear demonstration of DC-SIGN's role in Ebola virus cellular entry was provided by viral infection of DC-SIGN and DC-SIGNR expressing Jurkat-cells. These receptor deficient, immortalized human T-lymphocytes are often used to elucidate T-cell signalling and chemokine receptor expression upon viral entry. CLR expression was induced by retroviral vectors in the cells, which are otherwise non-permissive for Ebola virus and viral infection was observed both in th case of DC-SIGN and DC-SIGNR presence on the cell surface. On the contrary, viral infectivity was suppressed by inhibition of the receptors, either by mannosylated ligands or ABs.

Dengue virus is another widely spread pathogen, that normally causes fever, headache, muscle and joint pain, but during a second infection or encounter with a heterologous strain, symptoms may escalate to life threatening-conditions, such as severe bleeding and shock. The single stranded, enveloped RNA virus induces platelet activation (common observation of thrombocytopenia in patients) and infects immature DCs, and Langerhans cells, in a DC-SIGN mediated process.[199] The rate of DC-SIGN expression on the surface of DCs is correlated to Dengue virus infection rates, moreover, antibody-blocking of DC-SIGN strongly diminishes Dengue infections. Similarly to HIV-1, different studies suggest that DC-SIGN enhances, but is not the major mediator of Dengue virus cellular entry, although it can rapidly internalize the pathogen.[220] At the same time, DC-SIGN interactions with mannosylated *N*-glycans on the viral E-glycoprotein increas binding to other receptors on the DC surface, which are later responsible for pathogen uptake. The CLR facilitates *cis*- and *trans*-infections to susceptible cells, but virus dissemination is efficiently blocked by inhibition of DC-SIGN.

The bacterial pathogen *Mycobacterium tuberculosis* is taken up in human monocyte-derived DCs by binding to DC-SIGN.[206] The bacterium is responsible for millions of deaths worldwide each year, but challenges only the impaired immune system, which – unlike the healthy one - allows the development of an active disease. *M. tuberculosis* triggers mainly $T_H1$ responses (IFN-$\gamma$ secretion for microbicidal effect) and stimulates DCs to express co-stimulatory molecules and cytokine expression by TLR-induced pathways.[221] *M. tuberculosis* interacts with various PRRs on the surface of DCs, these include TLR-2 and TLR-4, dectin-1, MR and DC-SIGN.[222]

DC-SIGN binds capping mannose di- and trisaccharide residues in bacterial cell-wall glycoproteins and glycolipids (e.g. ManLAM, phosphatidylinositol-mannosides) in mycobacterial pathogens, like *M. tuberculosis* and *M. bovis*. In the second case, the receptor was shown to mediate bacterium-internalization, and blocking DC-SIGN with antibodies abrogated *M. bovis* infection. *M. tuberculosis*, upon binding to DC-SIGN, is directed to phagosomal compartments in DCs, but instead of directly triggering DC maturation, the DC-SIGN-ManLAM interaction downstream signalling modulates TLR-induced pathways on the transcriptional level, and upregulates the secretion of IL-10, IL-12p35, IL-12p40 and IL-6.[194] The altered immune respons leads to immunosuppressive results that hinders protective immunity against mycobacterial infections, and it can be reversed by antibody-inhibition of DC-SIGN during encounters with *M. tuberculosis*.

The rapidly growing number of newly-discovered DC-SIGN functions gave rise to a number of medically relevant investigations. Targeting DC-SIGN is a plausible method to enhance or suppress DC functions and to steer the immune response in therapeutically desired directions. The highest potential among these attempts lie in the inhibition of the receptor to block pathogen uptake and dissemination (anti-adhesive strategy);[140a] modulation of signalling routes for altered biological outcomes (anti-inflammatory applications);[132] and exploitation of its antigen cross-presentation ability in order to stimulate anti-tumoral $CD8^+$ responses (immunotherapy and vaccine-development).[223]

The wide range of pathogen ligands suggest that DC-SIGN antagonists can effectively ambush and halt numerous microbial infections at the first stage, and that these agents could be used as topical microbicides and virostats. Ideal candidates, however, need to fulfill several criteria. First of all, competitive inhibitors have to exceed the binding affinity of pathogenic ligands towards DC-SIGN (and suppress immune-cell infections *in vivo*), while being selectively recognized by the receptor, in order to avoid unwanted, occasionally opposite physiological effects. A careful evaluation of DC-SIGN-induced or modulated signalling allows the understanding of DC-evoked defensive immune responses and tolerance, but – as it was shown during the previous examples of infectious pathogens, this cannot be predicted with a high degree of certainty at the moment.

The second requirement for selectivity can be easily understood through the example of DC-SIGN and langerin. The HIV-1 surface glycoprotein gp120 is recognized by several Man-specific CLRs, including DC-SIGN (see above), MR and langerin. Langerin is an endocytic receptor expressed by a specialized subset of

DCs, the Langerhans cells (LCs), found in the skin and mucous tissues.[101] While binding to DC-SIGN promotes infection of DCs and viral transmission to T-helper cells, the capture of gp120 by langerin initiates the eradication of HIV-1.[224] The lectin induces formation of specialized, "tennis-racket" shaped cytoplasmic organelles – the Birbeck-granules, and mediates viral internalization into these compartments, where HIV-1 is rapidly degraded. Birbeck-granules display superimposed and zippered membranes and are uniquely present in LCs – they are essential for the eradication of HIV-1 infections that are promoted and spread by DC-SIGN. An optimally designed CLR inhibitor has to be therefore not only efficient against one receptor, but additionally selective among different CLRs.

DC-SIGN antagonists have long been designed and synthesized by several research laboratories, including the Bernardi group, which has been focusing on small-molecule glycomimetics and the tailored design of multivalent constructs.

# 1.7 THE CRD OF DC-SIGN ACCOMMODATES NATURAL AND SYNTHETIC LIGANDS IN MULTIPLE BINDING-MODES

## 1.7.1 The CRD of DC-SIGN determines a specificity for mannosylated and fucosylated ligands

### 1.7.1.1 The structure of the CRD

Up to now, 7 CRD-ligand complexes are available in the PDB and provide an insight to the structure of the DC-SIGN CRD, as well as to the molecular-level interactions between the protein and ligands (1K9I, 1SL4, 1SL5, 2IT5, 2IT6 - with oligosaccharides built from natural monosaccharides and 2XR5, 2XR6 - complexes with glycan-based glycomimetic molecules). Additionally, the structure of Dengue virus, complexed with DC-SIGN CRDs, was obtained by cryogenic electron microscopy and it is accessible in the same database (2B6B).[225] These crystallographic results, coupled with NMR observations and glycan-microarray binding profiles form the core of DC-SIGN inhibitor development. Knowledge about the ligand diversity and binding mechanisms is actively exploited during the development of DC-SIGN specific antagonists, therefore an overview of the main discoveries is provided below, before the discussion of our work on antagonist design, synthesis and evaluation.

The C-terminal CRD exhibits the typical long-form C-type lectin fold (see **1.3** and **1.4.1**) – 12 β-strands, two α-helices and three disulfide bonds organize into a globular domain.[226] (**Figure 18**) One of the two α-helices is unusually long and flanks the domain more closely than in other CLRs, furthermore, the COOH-terminal end of the helix extends the surface formed by the loop linking β-strands 3 and 4. This continuous area offers more room for interactions with extended oligosaccharide ligands. The long loop determining the primary carbohydrate binding-site protrudes from the surface of the domain and accommodates three $Ca^{2+}$ ions, one of which directly coordinates sugars and is therefore designated the primary $Ca^{2+}$ site. The three $Ca^{2+}$ ions also contribute to maintaining the overall conformation of the CRD.[83]

The principal $Ca^{2+}$ is nestled among the Glu347-Asn349-Glu354-Asn365 residues that determine the monosaccharide epitopes that can be ligated here. The binding site defines a flat, shallow hydrophobic "trough" on the protein surface that is bordered, on one side, by Val351.[193]

The peptide sequence contains the EPN motif (Glu347-Pro348-Asn349) and in consistence with this, the lectin has a preference for two equatorial OH-groups in positions 3 and 4 (or 2 and 3). These functionalities simultaneously form coordination bonds with the metal and participate in H-bonding with amino acid residues that also ligate the $Ca^{2+}$. More specifically, the CLR has affinity towards Man- and Fuc-structures. (**Figure 18**)

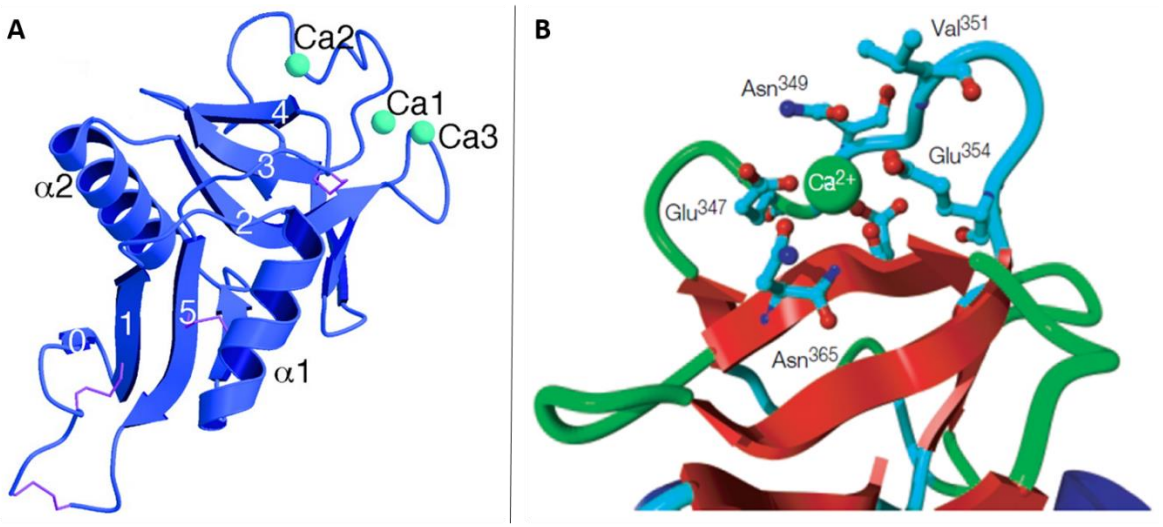



***Figure 18.*** *The structure of DC-SIGN CRD. A. The overall CRD domain. β-strands (numbered 1-5), two α-helices (α1, α2) and disulfide bonds (purple linakges) build a globular domain where a long loop protrudes from the top and accommodates 3 $Ca^{2+}$ ions.*

*B. Zoom on the primary sugar binding site. Amino acid residues directly involved in the ligand coordination (Glu347-Asn349-Glu354-Asn365) line the $Ca^{2+}$ binding site. The sidechains also determine the nature and orientation of the monosaccharide epitopes ligated at this site. Modified from: Feinberg et al.[83] and Geijtenbeek et al.[226]*

## 1.7.2 Natural DC-SIGN ligands assume different binding poses

This prediction was proven early-on in glycan-microarray screenings - numerous oligosaccharide ligands were defined by Guo *et al.* in this robust, high-throughput ligand-binding assay.[227] The poor individual affinity of low-valency carbohydrates can be significantly enhanced by multivalent presentation on the surface of microarray chips – sufficiently to obtain relative binding profiles and identify the most prominent binders. The originally described glycan library consisted of 12 mono-, 33 di-, 43 tri- and 40 larger oligosaccharides that were equipped with a biotin-tag for immobilization on streptavidin-coated wells. The wells were probed with fluorescently labeled DC-SIGN ECD. Highly mannosylated structures (notably $Man_9$ as $Man_9GlcNAc_2Asn$) and branched Fuc-containing ligands bearing a terminal Gal (4 Lewis-type antigens, blood group antigen A and B) stood out among the findings. Likewise, recombinant DC-SIGN-Fc chimera recognized almost exclusively Manα1-2Man terminated oligosaccharides and Fucα1-3GlcNAc and Fucα1-4GlcNAc containing glycans from more than 200 synthetic and natural glycans immobilized on glass-slides by amino functionalities.[228] Secondary incubation with goat anti-human-IgG-Alexa Fluor-488 antibody was used for fluorescent visualization. (**Figure 19**)

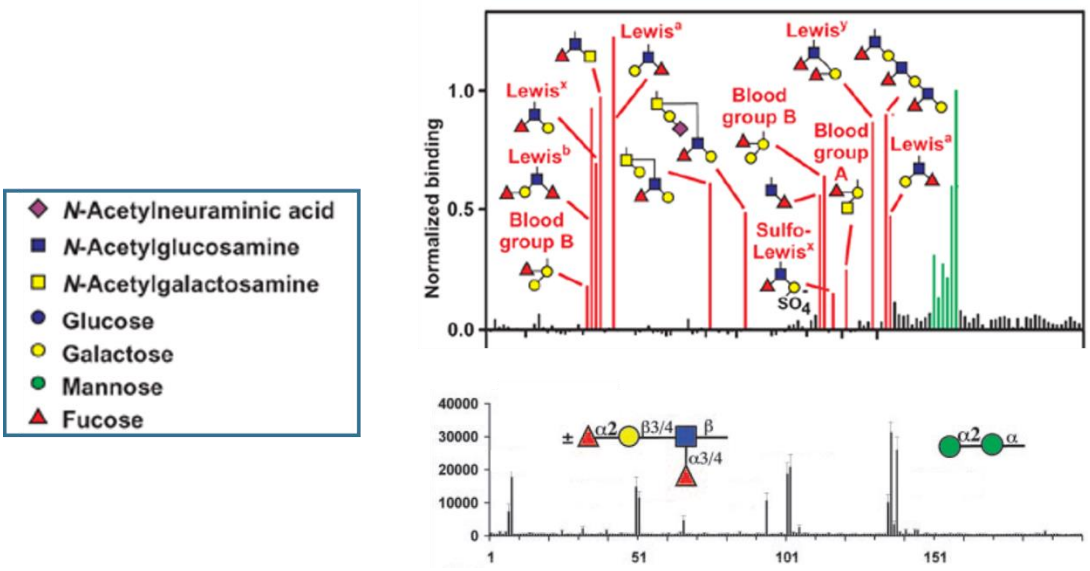



***Figure 19.*** *Top: Array-format screening of ca. 130 natural glycans against DC-SIGN. High mannose and fucosylated structures, like the Lewis and blood group antigens stand out with their affinities towards the CLR. Bottom: More than 200 oligosaccharides show similar results of microarrays against DC-SIGN. Man and Fuc-containing glycans are recognized almost exclusively. Modified from: Guo et al.*[227] *and Blixt et al.*[228]

Microarrays are powerful tools for the determination of relative ligand-binding affinities (more on this topic in Chapter II), but the results can be accurately elucidated solely in the context of a single assay and does not allow the precise estimation of an absolute inhibitory potency. For comparable quantification of affinities, the most commonly used measures are the $IC_{50}$, $K_D$ and $K_i$ values (half maximal inhibitory concentration, dissociation constant and inhibition constant, respectively). Although not interchangeable, the three values are closely related for competitive antagonists and they are suitable to compare ligand potency. With a critical eye, it can be noticed, however, that diverse experimental settings and presentations of the protein and/or ligands; the lack of a standard, widespread reference and the deviation of independent measurements even with an identical setup discourage the usage of precise values. The numbers provided here serve to indicate affinity values compared to references in the same study and assay, as well as the magnitude of binding potency.

Among natural monosaccharides, DC-SIGN displays the highest affinity towards L-Fuc ($K_i$ = 6.7 mM), D-Man ($K_i$ = 13.3 mM) and ManNAc ($K_i$ = 8.7 mM) as determined by radioactivity-based solid phase competition assays with $^{125}$I-$Man_{30}$-BSA (bovine serum albumin) against immobilized protein CRDs.[227] In accordance with the glycan-array screenings, the highly mannosylated $Man_9GlcNAc_2$ was not only the most potent inhibitor on the microarray chips, but the highest affinity binder in solid-phase competition assays as well ($K_i$ = 0.21 mM, where $K_i$(Man) = 7.7 mM).[176] In the fucosylated series, van Leimpt *et al.*[186] performed binding studies with DC-SIGN-Fc and multivalent, biotinylated Fuc-containing neoglycoconjugates immobilized on streptavidin coated plates, followed by fluorescent read-out with goat anti-human Alexa Fluor-488 antibodies. $IC_{50}$ values ranged from low affinities, like 15 µg/ml (L-Fuc), through moderate ($Le^Y$: 5.2 µg/ml; $Le^X$: 5 µg/ml) to the highest potency ligands ($Le^B$: 3.5 µg/ml; $Le^A$: 2.4 µg/ml). During a competition assay between glycomimetic ligands and a mannosylated-BSA surface against DC-SIGN ECD in solution, the $IC_{50}$ values of the reference compounds were measured as following: D-Man (1.8 mM), L-Fuc (1.2 mM) and $Le^X$ (0.8 mM).[229]

The binding affinity of larger, pathogenic glycoproteins has also been determined and compared, although the cell-type used to express DC-SIGN presumably affects the results. HIV-1 gp120 has shown the highest inhibitory potency, compared to ICAM-2 and ICAM-3, regardless to the cell-lines.[193]

The high ligand promiscuity of DC-SIGN is a burden in the way of selective antagonist development, but rational design can profit from a closer look on specific sugar binding interactions in the CRD. As stated

previously, the definitive, trademark ligation between the $Ca^{2+}$ and two hydroxyl groups of a monosaccharide are directly complemented by H-bonding with the peptide backbone and sidechains in the vicinity of the primary binding site. Eventually, the differentiation among specific oligosaccharides is the result of a number of additional contacts around the CRD, and/or preformed conformational states of the ligand and the protein. From the CRD point of view, the binding-site is already preformed to some extent, even without any bound ligand. This, and the crystal structure of the CRD alone suggests that the replacement of water molecules by sugar OH-groups is sufficient to form binding interactions and does not require further rearrangement of the peptide chain.[230] From the ligand point of view, the binding modes of natural high Man-oligosaccharides and Fuc-containing glycans are only partially overlapping, and differences may be explained by the abovementioned two favourable energy contributions to the overall free energy state of the complex.

The majority of oligosaccharides on pathogenic glycoproteins, like gp120 belong to the high Man family and these glycans have been shown to assume various conformations. None of these would match the architecture of the preformed binding site, therefore a conformational entropy penalty accompanies binding events – in particular during the crystallization of ligand-protein complexes. This unfavourable energy debt is compensated by a number of secondary contacts which add up to a significant enthalpy gain that tilts the energy balance of CRD-high Man interactions to the favourable side.

The additional contacts between mannosylated oligosaccharides and DC-SIGN CRD were confirmed by a model built on the X-ray structure of a $Man_5$ mimic pentasaccharide. In the crystallized ligand, the trimannose core terminates in two GlcNAc residues.[83] The CRD of DC-SIGN is unique among CLRs, as it recognizes internal Man-epitopes, rather than terminal ones. In this case, equatorial 3-OH and 4-OH in the α1-3-linked Man are directly coordinated by the $Ca^{2+}$ ion, but these hydroxyl groups also form H-bonds with amino acid residues that ligate the metal. Three carboxylates (Glu347, Glu354, Asp366), two side-chain (Asn349, Asn365) and one backbone (Asp366) carbonyls, and two ligand hydroxyl functionalities create the 8-fold coordination system around the $Ca^{2+}$, with ligation distances ranging between 2.35-2.75Å.[185b] The exocyclic C6 contacts Glu366, while the hydroxyl group on this carbon interacts with Asp367 indirectly, through a water molecule. Man3 and Man4 are supported by Ser360.

The X-ray image unveiled an additional binding motif that bears crucial importance in the recognition of high-Man structures. The oligosaccharide bends over Phe313 to fit into an adjacent groove that is lined by Asn311. This region, and in particular Phe313 are critical for binding high mannose structures, because the aromatic ring presumably slips in between Man1 and Man4 in the $Man_5$ model. Interestingly, Feinberg *et al.* attributed the discrimination between branched and linear mannose-oligosaccharides to the phenylalanine, based on the assumption that the residue could cause disruptive steric clash that may completely prevent the binding of linear structures. This speculation has been challenged by Anderluh *et al.*,[231] based on the fact that linear oligosaccharides were shown to bind in a similar fashion to branched ones, although the affinity of these ligands is lower. Additionally, other ligated complex structures also indicate that the phenylalanine sidechain is flexible and has some degree of freedom to incline in different directions, adjusting to the ligand. The amino acid is capable of forming beneficial interactions as well, like van der Waals contacts, or complementing the binding surface for improved accommodation of the α-linked outer arm Manα1-6Man moiety of a branched N-glycan.[85] In the same glycan, Phe313 abrogates binding to the β-linked core branch exhibiting a similar mannosylated motif – an effective discrimination of analogous epitopes in multi-antennary oligosaccharides, based on their position relative to the junction. (**Figure 20**)

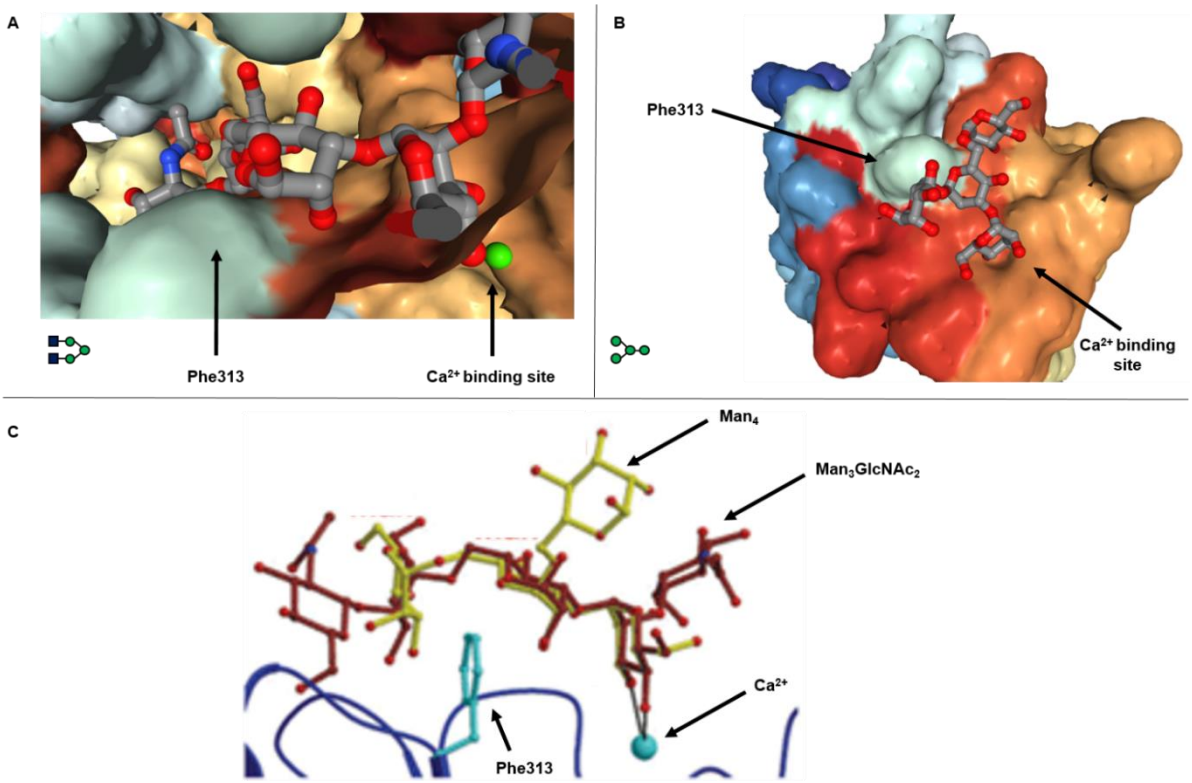



***Figure 20.*** *Binding mode of branched mannosylated oligosaccharides - $Man_3GlcNAc_2$ pentasaccahride and $Man_4$ - in the CRD of DC-SIGN. A: $Man_3GlcNAc_2$ is ligated at the $Ca^{2+}$ binding site by an internal Man epitope. Phe313 interacts with the Manα1-6 moiety and presumably blocks the binding of linear mannose oligosaccharides. (PDB code: 1K9I) B: Binding mode of $Man_4$ from top view. (PDB code: 1SL4) The two ligands have analogous binding modes, as seen in panel C. Panel C modified from Guo et al.,[227] the rest imported from PDB by using NGL viewer.*

The high Man family progenitor $Man_9$ displayed the highest affinity towards DC-SIGN on glycan microarrays.[227] The oligosaccharide is present in multiple copies on the surface of pathogens (linked to glycoproteins), thus DC-SIGN targeting for drug development demanded the identification of key binding elements and minimal recognition epitopes in the glycan structure. Once more, microarray technology proved to be a useful tool for screening the characteristic structural features of various Man-oligosaccharides. Adams *et al.* printed smaller, linear fragments of the native glycan, as well as a branched structure, and found that the Manα1-2Manα1-2Man trisaccharide and Manα1-2Manα1-6Manα1-6Man tetrasaccharide displayed similarly good affinities towards DC-SIGN as $Man_9$, on-chip.[232] This result strongly implied that the Manα1-2Man epitope at the nonreducing termini of larger Man-oligosaccharides is a key mediator in establishing contacts with the CRD. This idea was further supported by the fact that a pentasaccharide - based on $Man_9$, but lacking terminal Manα1-2Man moieties – exhibited significantly poorer affinity towards DC-SIGN, than the full oligosaccharide. The Manα1-2Man epitope, that alone is 4-times more potent than mannose (similarly to the Manα1-3[Manα1-6]Man trisaccharide), was therefore assumed to dramatically enhance affinity.

As it has already been discussed in a previous section, X-ray crystal structures were obtained for $Man_6$ and Manα1-2Man, but they did not reveal the complete molecular arrangement of the two (major and minor) bound conformations.[85] The observed dual binding mode helps to reduce the entropy debt of ligation, but additional interactions between the sugar and different sites on the protein surface were also suspected to contribute to the affinity difference. STD-NMR experiments coupled with computational studies were able to provide a description of both ligand conformations and highlight key-residues in the binding site.[86] In $Man_6$, the major conformer allows a direct contact between the nonreducing terminus Manα1-2Man and Val351. On the other hand, the minor conformation is swapped by a 180° axis around the C-3/C-4 bisector of the principal

$Ca^{2+}$-bound Man, and the 1-2 linked terminus is positioned to form H-bonds with Glu358 and Ser360, as well as interacting with Phe313. The two binding modes of Manα1-2Man reflects $Man_6$ in its two orientations in the primary binding site. In this case, the major conformation nonreducing end is in contact with Phe313, and the NMR experiments and computational model brought to light the engagement of the same phenylalanine with the minor conformer, although through a reversed ligation at the $Ca^{2+}$-site. (**Figure 20**)

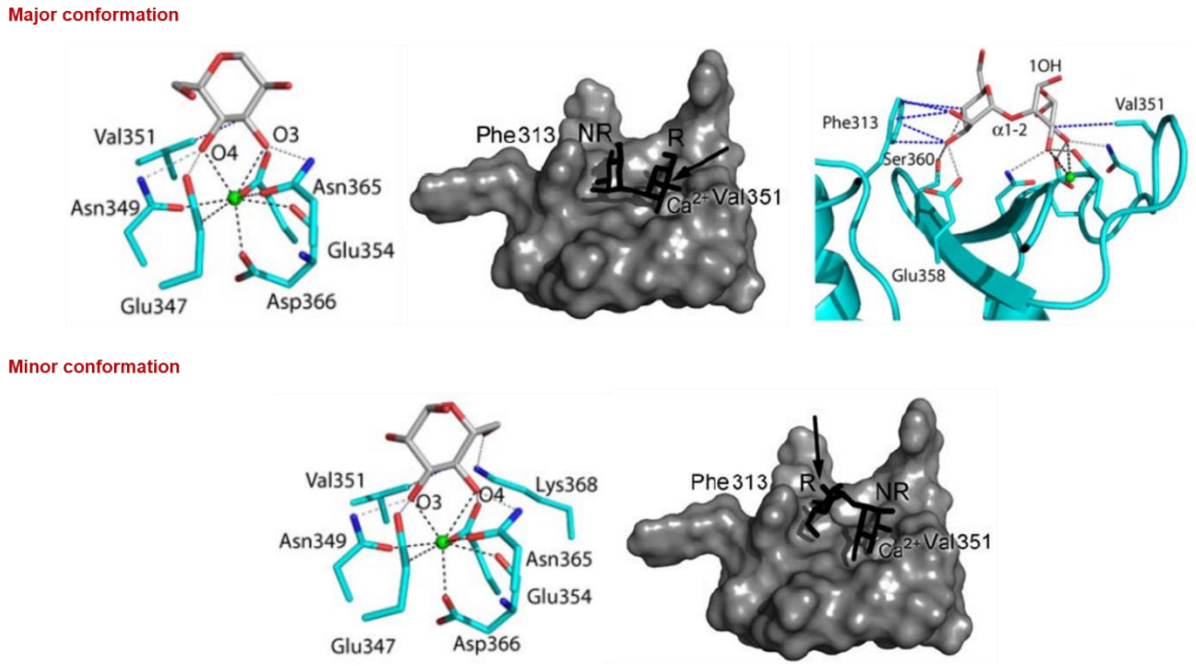



***Figure 21.*** *Manα1-2Man conformations in the CRD of DC-SIGN. Top: The major conformation uses the reducing end mannose epitope for $Ca^{2+}$ ligation and this orientation is stabilized by Val351 through a H-bond. At the same time, the nonreducing moiety directly contacts Phe313 and forms H-bonds with Ser360 and Glu358. (PDB code: 2IT6) Bottom: Minor conformation of Manα1-2Man, revealed partially by X-ray crystallography and by computational studies and STD-NMR. In this case, the nonreducing terminus is found at the principal $Ca^{2+}$ binding site, and the reducing end reaches Phe313. Modified from: Feinberg et al.[85] and Angulo et al.[86]*

The other most prominent class of natural DC-SIGN ligands, the Lewis-type antigens, benefit from their preformed conformation when binding to the receptor. The $Le^X$ epitope consistently assumes the same conformation in solution and crystals, as well as when bound to different receptors (shown by X-ray structures and NMR experiments). This rigid structure can be attributed to the van der Waals mediated packing of the Fuc ad Gal residues and an unusual C-H---O hydrogen bond found between these sugar rings.[233] This stable conformation substantially reduces the entropy cost of binding to the similarly preformed CRD of DC-SIGN. The CLR recognized all 4 Lewis-type antigens and blood antigens A and B on the glycan microarrays, but, surprisingly, not all Fuc-terminated structures.[227] Thus, identification of additional factors playing role in ligand binding became the object of investigations, and eventually Guo *et al.* succeeded in crystallizing lacto-*N*-fucopentaose III (LNFP-III) with the DC-SIGN CRD (PDB code: 1SL5). This oligosaccharide contains the $Le^X$ structure and engages with $Ca^{2+}$ through the fucose 3-OH and 4-OH groups. (**Figure 22**) These hydroxyl groups orientate the Fuc-ring differently than Man, however, as one of them (4-OH) is axial. As a consequence, the tilted sugar ring interacts with Val351 through its 2-OH. The valine creates a hydrophobic site for strong interactions with Fucα1-3/4GlcNAc moieties, where GalNAc is projected vertically, away from the CRD surface. (**Figure 22**) Val351 is crucial for the recognition of Lewis-type structures, and mutation of the amino acid residue can completely abrogate binding to this class of ligands. The terminal galactose interacts with a

secondary binding site, and notably, the stacked, compact conformation of fucose and galactose does not allow Phe313 to reach and interact with the ligand.

It would be tempting to connect the distinct binding mechanism of mannose and fucose containing oligosaccharides with the differential signalling activation described by Gringhuis and co-workers [194] The lack of any evidence of conformational differences in the overall CRD structure in unligated, mannose- and fucose-bound complexes, however, does not support this hypothesis. Additionally, synthetic glycomimetic ligands (discussed later on), which interact with the same secondary site (Val351) in the CRD as $Le^X$, were not described to elicit the same biological response, and so, this explanation needs to be handled with caution.

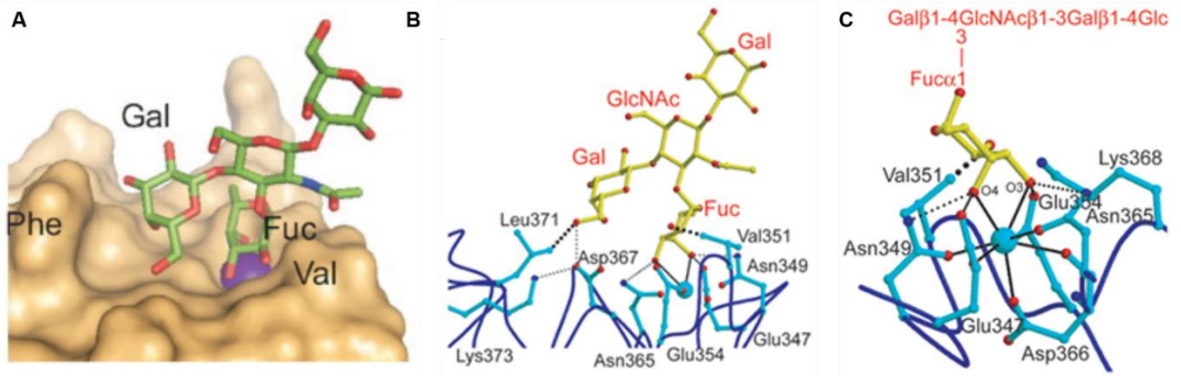



***Figure 22.*** *Crystal structure of lacto-N-fucopentaose in the CRD of DC-SIGN. (PDB code: 1SL5) A and B: The rigid $Le^X$ core forms interactions through the stacked Fuc and Gal epitopes, while the GlcNAc residue "sticks out" from the CRD. C: Fucose is coordinated at the primary $Ca^{2+}$ site by the 3-OH and 4-OH groups, one of which (4-OH) is axial. Consequently, the sugar ring tilts toward Val351 and interacts with the sidechain by 2-OH. Modified from: Taylor et al.[230] and Guo et al.[227]*

In summary, the multiple binding modes can, on one hand, statistically enhance binding (for example to diversely arranged epitopes on pathogens), and on the other, engage in interactions with various binding sites in the CRD. In the latter case, Phe313 and Val351 are pivotal residues for ligand accommodation in the DC-SIGN CRD.

Microbial adhesion, the activation of T-cells by dendritic cells and shaped immune response can be achieved by blocking DC-SIGN.[140a] The lectin is efficiently blocked by DC-SIGN specific antibodies, $Ca^{2+}$ chelators, mannan and Man/Fuc-based oligosaccharides, but undoubtedly, the most straightforward strategy in DC-SIGN targeting is the design and synthesis of small-molecule inhibitors that exploit various interactions in the CRD or multiple CRDs (multivalent constructs). The synthetic difficulties associated with larger oligosaccharides, as well as their complexity that prevents their presentation on multivalent systems for medical applications, put forward smaller glycan-based molecules. Glycomimetic antagonists can achieve higher affinity than the natural monosaccharides and smaller oligosaccharides, without the cumbersome synthetic effort requirement of complex oligosaccharides. Furthermore, their advantages include enhanced metabolic stability and the relative ease of chemical modifications for their presentation on multivalent systems.

## 1.7.3 Small molecule inhibitors targeting DC-SIGN

### 1.7.3.1 Non-carbohydrate inhibitors

Based on the modest affinities displayed by natural saccharides, the Kiessling group turned towards non-carbohydrate based DC-SIGN antagonists. High-throughput fluorescence-based competition assays of two commercially available small-molecule libraries (16 000 membered Chembridge DIVERSet and 20 000 membered Chemical Diversity Labs - ChemDiv) allowed the estimation of $K_i$ values.[234] The assay was performed on immobilized tetrameric DC-SIGN, with a fluorescent mannosylated-BSA competitor against the non-carbohydrate molecules. As opposed to the reference compounds, $Man_9$ ($K_i$ = 0.21 mM) and ManNAc ($K_i$ = 6.9 mM), the screening resulted in 7 outstanding hits (**1.1** – **1.7**), with $K_i$ values ranging from 1.6 - 32 μM. Comparing these values to monosaccharide affinities towards DC-SIGN, the ligands show an1000-fold increase in affinity. These molecules do not only inhibit the binding of carbohydrates by DC-SIGN, but they also block DC-SIGN mediated cell-adhesion, although the affinity values determined by these assays are ca. 10-fold lower than in the fluorescent-competition experiments (7.3 to 227 μM). These findings highlight two main messages. Firstly, carbohydrate-based structure is not essential in targeting DC-SIGN, and therefore sugar-related drawbacks can be completely circumvented during drug-candidate development for the CLR. Secondly, the lack of obvious $Ca^{2+}$ chelators (like 3-OH and 4-OH in mannose) raise the question that a binding site, other than the principal $Ca^{2+}$ pocket is involved in binding these ligands. This suspicion is dampened by the fact that amide carbonyl groups present in the structures are capable of interacting with the metal ion. Additionally, the hits contain aromatic rings in their structure and these could contact aromatic sidechains around the shallow, but rather extended groove in the CRD.

The identified ligands belong to two major classes: molecules that contain a pyrazolone scaffold, and structures featuring a quinoxalinone core.[235] The electrophilic nature of the first group prevents their study in cellular assays. Quinoxalinones, on the other hand, undergo degradation due to the oxidizable thioether found in these compounds. Nonetheless, the later issue was successfully addressed by replacement of the sensitive sulfur-atom by a methylene functionality. This optimized structure was developed into a 20 member library by late-stage derivatization during the synthetic route. The divergent strategy significantly increases the efficiency of a library synthesis, and the incorporation of diverse elements in the final steps allows the evaluation of distinct features contributing to DC-SIGN binding. Ca. 25% of the tested compounds showed good affinities towards the lectin in competitions assays (highest affinity: **1.8**: 0.31 – 10 μM). A sensitive flow cytometry technique was developed to evaluate the ligands' cell surface binding and inhibitory capacity of Man-Fl-BSA (fluorescent mannosylated BSA) uptake in DC-SIGN$^+$ Raji cells. The compounds were able to block both the binding of glycoconjugates and their DC-SIGN-mediated internalization.

### 1.7.3.2 Glycomimetic DC-SIGN ligands

<u>1.7.2.3.1 Shikimic acid-based inhibitors</u>

The same research group has also developed carbohydrate-inspired inhibitors of DC-SIGN, starting from the structure of Shikimic-acid.[236] In this cyclohexene-based structure hydroxyl groups in positions 2, 3 and 4 resemble the configurations of the respective mannose and fucose OH groups. These glycomimetic molecules (**1.9**) were derivatized at 3 focal-points by solid-phase synthesis. Although aliphatic substituents were also implemented in one of these positions, the main focus was placed on benzyl groups, because these could interact with the aromatic residues lining the CRD of the lectin. From the rationally optimized common

scaffold, a ligand-library of 192 molecules was developed and screened in a fluorescence-based high-throughput format. (**Figure 23**)

Competition between the synthetic ligands and immobilized mannan surface was evaluated against fluorescently-tagged DC-SIGN ECD in solution. The $IC_{50}$ value of the highest-affinity glycomimetic was determined as 3.2 mM (in the previously described assays of immobilized DC-SIGN ECD against Man-Fl-BSA) which is superior to the 11.2 mM value of ManNAc in the same assay. Furthermore, the ligands showed good selectivity towards DC-SIGN over MBL – a CLR inducing innate inflammatory immune responses and activating the complement system.

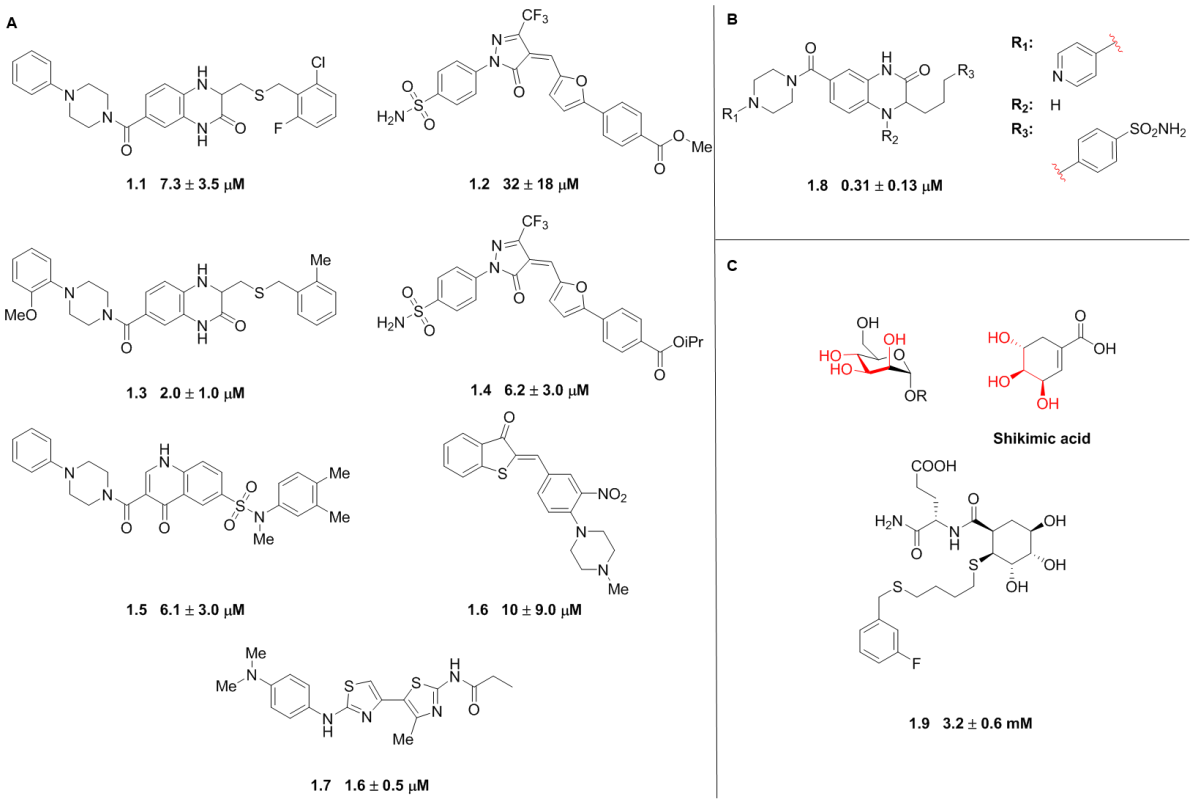



***Figure 23.*** *Non-carbohydrate based DC-SIGN inhibitors developed by the Kiessling group. A. Screening commercially available small molecule libraries yielded 7 hits with $K_i$ values in the low micromolar range.*

*B. Based on the library screening a quinoxalinone structure was developed, and some optimized ligands exhibited affinities in the nM range.*

*C. Shikimic acid derivatives. The mannose configuration on C2, C3 and C4 is mimicked by the Shikimic acid hydroxyl groups. The best ligands from a library of 192 compounds reached affinity values of 3.2 mM. These ligands can be multivalently presented after the synthesis of artificial glycopolymers by the ROMP method (ring-opening metathesis polymerization).*

1.7.2.3.2 Fucosylated inhibitors

The screening results of 130 glycans by the Consortium for Functional Glycomics suggested a key role of fucose-containing glycans in selectivity between DC-SIGN and other CLRs.[227] 14 glycans containing a terminal fucose and displaying Lewis-type structures were exclusively recognized by this lectin, but it was also clear, that a reducing-end fucose alone does not determine binding to DC-SIGN. The highly conserved, rigid structure of $Le^X$, however, serves as a well-defined starting point in glycomimetic approach. Based on the three-dimensional structure of Galβ1-4[Fucα1-3]GlcNAc, where Fuc is stacked on the Gal-ring, a set of glycomimetics were designed with an α-fucosylamide anchor (**1.10**).[229] The fucose-residue was used to take part in the primary coordination with $Ca^{2+}$, and the rest of the molecule was built to resemble the structure of

the natural glycan with enhanced affinity and metabolic stability. In the first series of ligands, the fucose was linked to a galactose-mimic through a cyclic *cis*-β-amino acid ((1S,2R)-2-amino-cyclohexanecarboxylic acid) that ensures a conformation almost identical to that of $Le^X$. The nature of linkages bear crucial importance, as α-fucosides are enzymatically, as well as chemically unstable. α-Glycosylamides are basically nonexistent in natural glycans, implying that they will be resistant to enzymatic and chemical degradation. Moreover, the formation of amide-bonds simplifies the chemical route to these ligands, although the DeShong synthesis of α-glucopyranosylamides[237] is a peculiar glycosylation method.

SPR competition-assay studies against a mannosylated-BSA gold-chip revealed an increased affinity of glycomimetic ligands towards DC-SIGN. In contrast to the weak binding of the natural saccharide references (D-Man: 1.8 mM, L-Fuc: 1.2 mM and $Le^X$: 0.8 mM), the $IC_{50}$ value of the synthetic LeX mimic reached 0.35 mM, and even more interestingly, the Gal-lacking ligand (**1.11**) showed almost the same affinity (0.5 mM). This similarity in inhibitory concentrations indicates that contribution from the galactose-like moiety is negligible and therefore other functional groups could be introduced in this position.

An NMR and molecular modelling based evaluation of the pivotal interactions between two aromatic α-fucosylamides confirmed that the main mediator in binding is the fucose-residue which is coordinated at the primary binding-site. STD-NMR spectra clearly showed that the α-fucosylamides assume the same chair conformation as natural α-fucosides. Further interactions were detected from the aromatic substituents, and, in consistence with the hairpin-like conformation of $Le^X$, the cyclohexane ring's contribution in the STD-NMR spectrum is minor – suggesting that this moiety stands out from the CRD surface. Notably, the aromatic phenol (e.g. **1.12a** and **b**) and pyridine groups interact with Phe313, a residue that was previously only accounted for enhancing or diminishing binding towards mannosylated oligosaccharides, as proven by X-ray-based docking studies and NMR analysis. Encouraged by these results, other derivatives of the α-fucosylamide glycomimetics were synthesized,[238] for example by systematic variation of the β-amino acid and by using a more flexible alanine linker for a simpler synthesis, but the affinity of these compounds (**1.13**) did not outperform the one of the original α-fucosylamide. (**Figure 24**)

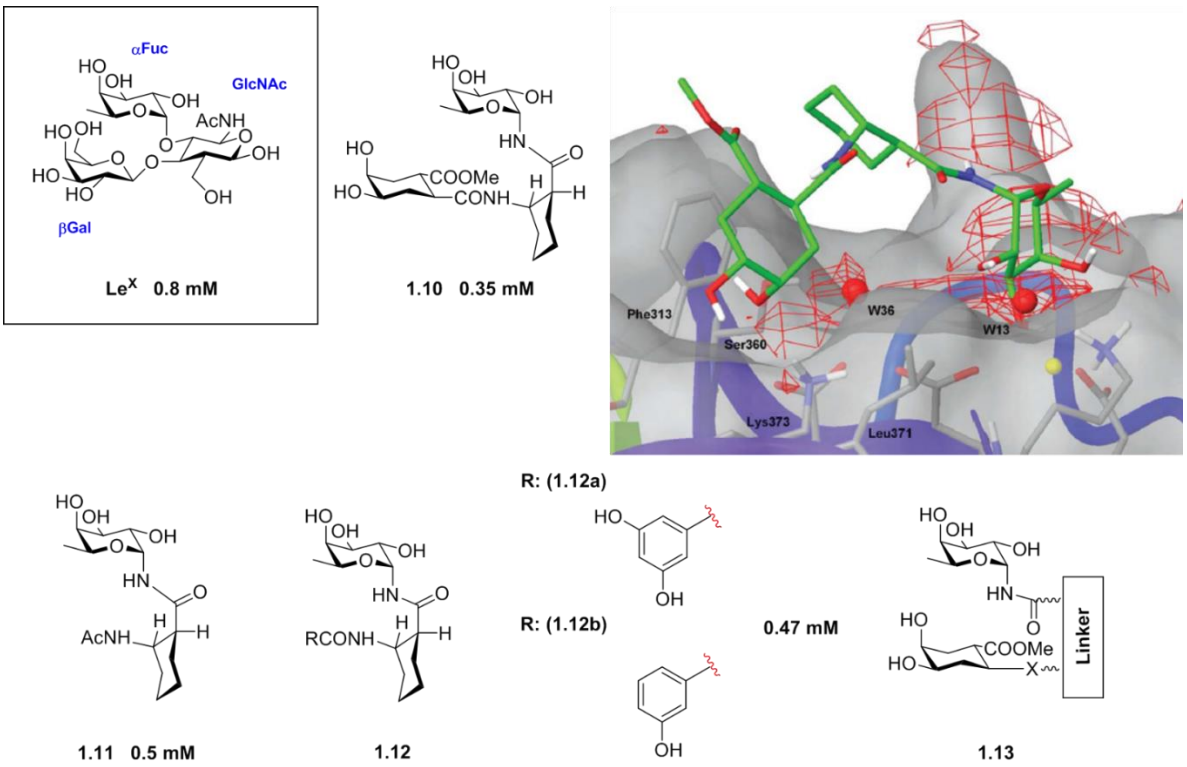



***Figure 24**. α-Fucosylamide-based DC-SIGN inhibitors. The structure of **1.10** was based on the natural ligand $Le^X$ and docked into the available crystal structure of the $Le^X$-containing lacto-N-fucopentaose. (PDB code: 1SL5). The original trisaccharide mimetic **1.10**, and surprisingly the same structure lacking Gal (**1.11**) exhibited ca. 2-fold higher affinity towards DC-SIGN than $Le^X$. Further glycomimetic structures were synthesized, based on the same α-fucosylamide anchor. X-ray structure based docking studies suggested that aromatic groups (like phenyl moieties in **1.12a** and **1.12b**) interact with Phe313 – an amino acid earlier only associated with mannosylated oligosaccharides. Modified from: Andreini et al.*[238]

1.7.2.3.3 Ligands based on a monosaccharide mannose ore and aglycone elements

The number of mannose-based monovalent glycomimetic inhibitors considerably exceed the fucosylated ones, as most natural ligands of DC-SIGN bind through this epitope. An early paper from Mitchell *et al.* describes a single modified monosaccharide ligand whose affinity is 48-times higher compared to mannose.[239] Substituents in position 2 on the sugar-ring, like Me, $CH_2OH$, $CH_2N_3$ and in particular $CH_2NH_2$ (**1.14**) dramatically increased the affinity towards the lectin, as determined by SPR competition assays, against immobilized gp120 ($K_i$(2-C-aminomethyl-D-Man) = 0.35 mM; $K_i$(mannose) = 17.1 mM). The μM affinity of such a small molecule suggests that the appropriate modifications can create striking differences in binding, by forming strong interactions with the CRD and that these ligands can become potential lead structures for further development.

Similarly, modifications in the anomeric position of mannose can yield highly potent monovalent glycomimetics. Tomasic *et al.* saw Phe313 as an optimal target that could be exploited to form additional interactions between the ligands and the CRD.[240] As discussed earlier, this amino acid can be accounted for hindering the binding of linear Man-oligosaccharides, and enhancing the affinity of branched ones. As the sidechain creates hydrophobic pockets both in the direction of the $Ca^{2+}$ ion and behind the residue (hydrophobic groove), they believed that aromatic groups could engage in hydrophobic or π-π interactions in this region - presumably even two aryl groups simultaneously on both sides. Docking studies predicted 1,3-diaminopropan-2-ol and glycerol as linkers of the appropriate length that could span the distance between the primary sugar binding site and Phe313. The two most potent ligands (dinaphthyl-substituted *O*-aryl ethers, e.g. **1.15**) reached $IC_{50}$ values of 40 and 50 μM in inhibition assays with DC-SIGN ECD against gp120 (reference

compound L-Fuc: 2.95 mM). The ligands were also evaluated in DC-adhesion inhibition assays, as well as molecular docking and molecular dynamics simulations, where the hydrophobic interactions with Phe313 were partially confirmed, although one ligand exhibited high flexibility in the binding site.

Mannosylated glycolipid mimetics (**1.16**) were inspired by the structure of ManLAM, a natural ligand of DC-SIGN and numerous other CLRs.[241] In these glycoconjugates, the lipid chains are suspected to help supramolecular organization of the ligands and promote micelle formation which leads to increased avidity effects in solutions. The glycolipid-mimetics contain a polar mannose head that is linked to a hydrophilic linker, crucial for aqueous solubility. The ligands end in a saturated lipid tail of various length that was optimized to exhibit the highest affinity values in the low micromolar range, as measured by SPR direct binding assays and DC-mediated HIV-1 *trans*-infection experiments. (**Figure 25**)

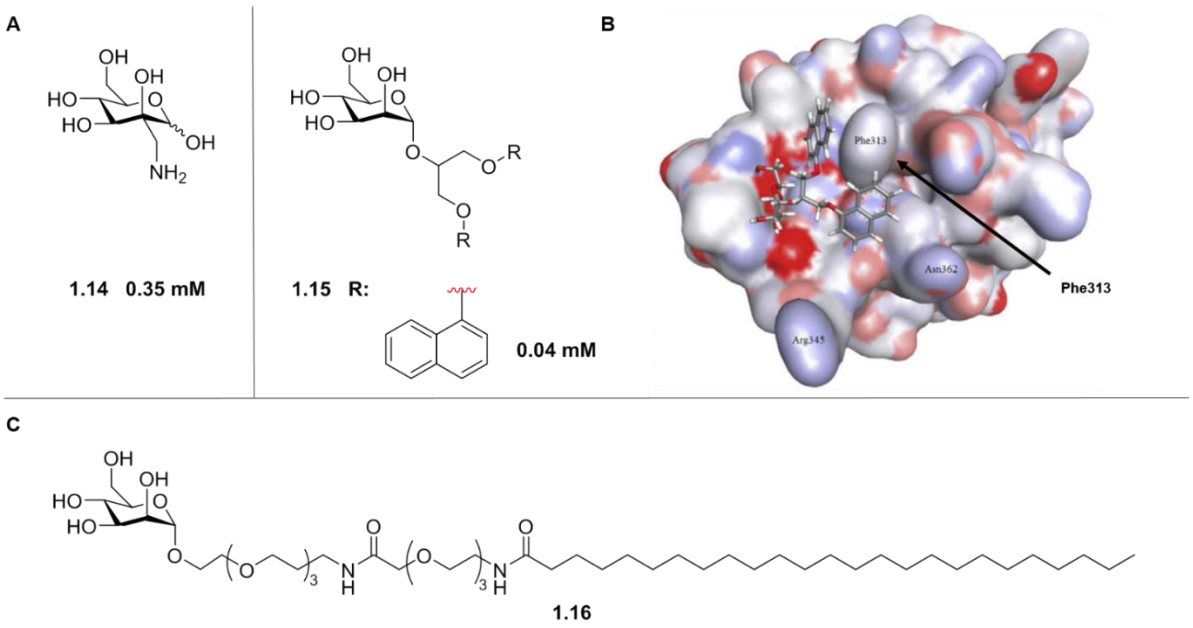



***Figure 25.*** *Monovalent mannose-based glycomimetic ligands as DC-SIGN inhibitors. A. C-2 modified mannose glycomimetic developed by the Fleet group.[239] B. The anomerically functionalized mannose glycomimetics developed in the Anderluh group target Phe313. The best results were reached by naphtyl-derivatives. (FlexX-calculated binding mode shown on the right). Modified from: Tomasic et al.[240] C. Mannosylated glycolipids contain a polar sugar-head and a hydrophilic linker that is conjugated with a long lipid chain. The molecules supposedly assemble into micelles, and this creates a multivalent platform to interact with the receptor and enhance avidity.*

### 1.7.4 Pseudo-disaccharide glycomimetic ligands

Based on the high-density array findings of Adams *et al.*,[232] that smaller, linear fragments of $Man_9$ are able to reach similar affinity towards DC-SIGN as the native *N*-glycan; and that a crucial epitope in these oligosaccharides is the terminal Manα1-2Man motif (**1.17**); Bernardi *et al.* have been designing and optimizing glycomimetic inhibitors that resemble this structure.

The progenitor glycomimetic pseudo-mannobioside **1.18a** was built on a mannose-anchor for coordination at the primary binding site (based on the terminal Manα1-2Man disaccharide, the nonreducing-end sugar was replaced by a conformationally locked diol that resembles the second mannose.[242] This structure was predicted to retain the three-dimensional arrangement of the natural disaccharide, based on the X-ray structure of $Man_4$-DC-SIGN complex (PDB code: 1SL4), as predicted by docking studies.[227] The mimetic cyclohexane moiety serves several purposes: it can enhance metabolic stability against mannosidase activity and its modifications enable a straightforward incorporation of functional groups for additional interactions with the CRD, or linkers for conjugations to multivalent constructs, for example. The 1,2 diaxial dioxygen array of this cyclohexane mimics the 1,2-substituted mannose conformation in the natural saccharide, MM3* calculations showed that compared to the $^4C_1$ chair, assuming the $^1C_4$ conformation would be accompanied by a +2.7 kcal/mol energy penalty.[243]

Molecular modelling studies highlighted that both Manα1-2Man and pseudo-mannobioside populate two low energy conformations rapidly interconverting into each other – the major extended conformation and the minor stacked structure.[242b] Both the synthetic and the natural disaccharide exhibit similar φ and ψ angles around the glycosidic linkage that are characterized by typical NOE cross peaks – for the pseudo-disaccharide: extended conformation – exclusive H1(M)–H3eq(C) and stacked conformation – exclusive H1(M)–H1(C). The two conformers are rotating around the C1-O1-C2'-C1' glycosidic bond in fast dynamic equilibrium and they are detectable in solution by trNOESY.

STD-NMR experiments with DC-SIGN had confirmed that the mannose coordinates the $Ca^{2+}$ ion – in accordance with expectations based on docking studies – and additionally, the cyclohexane resides close to the protein surface. As described by Feinberg *et al.*[85] and Angulo *et al.*,[86] natural Manα1-2Man has two distinct binding modes in the CRD of the lectin, one of which uses the reducing end mannose for ligation to the metal and another that is anchored by the nonreducing sugar, both interacting furthermore with Phe313. (**Figure 21**) Although considered as a possible alternative, a binding mode in which the reducing end mannose is ligated at the principal $Ca^{2+}$ site and the second Man leans over Val351 was experimentally (STD-NMR) shown not to contribute to the multimodal binding of the disaccharide in the CRD. Based on these observations with Manα1-2Man, the pseudo-mannobioside mimic was anticipated to use the same binding epitopes and CRD sites, and possibly exhibit several distinct binding modes.

Resolution of the crystal structure of pseudo-mannobioside in the DC-SIGN CRD has, however, unveiled a surprising result.[244] Although almost identical in shape and conformation to the natural disaccharide, the glycomimetic uses a third, unique binding mode that is closer to the one described for $Le^X$ than to any of the binding modes employed by mannose oligosaccharides. Differences can be seen already at the principal $Ca^{2+}$ coordination, as 3-OH and 4-OH are swapped by 180° compared to the major Manα1-2Man binding mode, so that 3-OH contacts Glu354 and 4-OH interacts with Glu347, as opposed to the natural ligand using 4-OH and 3-OH to contact the same residues. The asymmetrical pattern of H-bonding with Glu347, Asn349, Glu354 and Asn365 is not different from the previously described H-bond networks for mannose-epitopes though; this alone would not be enough to account for the essentially distinct orientation.

The X-ray image of the complex revealed that the ligand is bound in the CRD in its major, extended conformation (φ O5–C1–O2′–C2′ : 66.3°; ψ′ C1–O2′–C2′–C1′ : –141.11°), with the cyclohexane ring lying on top of Val351 – the key residue for $Le^X$ binding. The apolar sugar-mimic ring forms van der Waals

interactions with the sidechain, significantly contributing to the affinity of the ligand. The short ethanolamine tether points away from the CRD surface, implying that using this linker for multivalent presentations would not significantly affect the binding mode. From the three water molecules visible in the crystal structure, one links 2-OH indirectly with Asp367.

Further STD-NMR experiments assigned a key role of H6ax(C) (the axial hydrogen atom in position 6 on the cyclohexane) that is located the closest to the methyl groups of Val351. The poor STD-NMR "visibility" of the ethanolamine linker, and the dimethyl ester functionalities indicate that these moieties remain flexible in solution. The NMR studies proved that the binding mode in the crystallized complex is identical to the binding mode detected in solution and therefore the structure is a good representation of the ligand binding in both phases. (**Figure 26**)

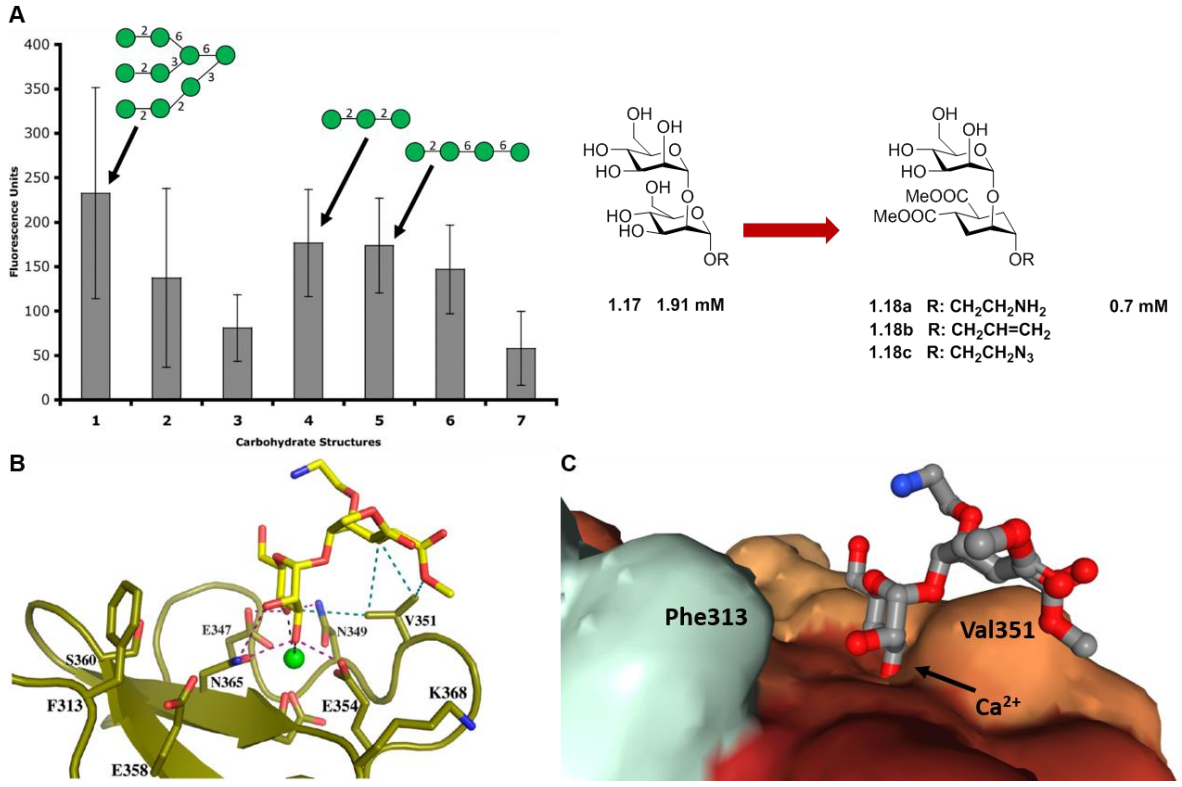



***Figure 26.*** *The development, structure and binding mode of pseudo-mannobioside* ***1.18*** *in the CRD of DC-SIGN. A: Microarray screenings from Adams et al. showed that simpler, linear fragments of the progenitor* $Man_9$ *structure have similar affinity towards DC-SIGN as the large, complex oligosaccharide and that Manα1-2Man terminal epitopes have crucial role in binding. Modified from: Adams et al.*[232]

*B and C (similar point of view): The Bernardi group developed glycomimetic ligands where the nonreducing-end sugar was replaced with a conformationally constrained cyclohexane-ring. The synthetic ligand assumes an extended conformation in the CRD and surprisingly, the cyclohexane ring leans on Val351 – a residue that typically interacts with the fucosylated ligands, but not with mannosylated oligosaccharides. The reducing end terminus is ligated at the* $Ca^{2+}$ *binding site. The short linker aminoethanol is not involved in binding, but is rather oriented towards the solvent. This suggests that using this arm for multivalent presentation of the glycomimetic molecule would not interfere with the binding mode. The methyl-ester groups on the sugar-mimetic ring are extending towards the protein surface, and presumably larger groups can be fitted in this position to form additional contacts with the CRD. B modified from Thépaut et al.*[244]

The antiviral activity of the pseudo-mannobioside was evaluated in Ebola infections models.[242b] The glycomimetic ligand inhibited pseudotype Ebola virus uptake in DC-SIGN$^+$ Jurkat cells three-times more efficiently than the natural disaccharide ($IC_{50}$ glycomimetic **1.18a**: 0.62 mM; Manα1-2Man: 1.91 mM). Additionally, the higher concentrations of Manα1-2Man, required to record the inhibition curve, caused poor dose-response of the cells, due to cytotoxicity at these concentration levels.

One of the major purposes behind including an unnatural cyclohexane ring in the ligand structure was the improvement of metabolic stability, and indeed, the pseudo-mannobioside was six-fold more resistant to the activity of commercial jack-bean mannosidase that the natural counterpart. Along the same lines, the thio-glycosidic analogue **1.19** was synthesized to further enhance enzymatic stability in the pseudo-disaccharide based glycomimetics.[245] This ligand was found to exhibit a very similar affinity towards DC-SIGN – $IC_{50}$ = 0.78 mM, as opposed to 0.72 mM of the *O*-glycoside in SPR experiments against a Man-BSA surface.

The protective role of langerin and the importance of selectivity between DC-SIGN and langerin has been stressed before. Langerin is capable of binding the Manα1-2Man epitope (crystal structure PDB code: 3P5F),[80] and thus the selectivity of the pseudo-mannobioside needed to be addressed. The two CLRs

coordinate different mannose residues in natural oligosaccharides – DC-SIGN has a preference for internal epitopes, while langerin binds terminal motifs. The natural disaccharide exhibits a 4-fold enhanced selectivity towards DC-SIGN over langerin, and the synthetic glycomimetic's selectivity was studied in comparison with this ligand. An increased selectivity of the synthetic ligand was observed in SPR experiments on mannosylated BSA surface compared to the natural disaccharide.

The crystal structure of Manα1-2Man complexed with the CRD of langerin offers an explanation for this selectivity. Similarly to DC-SIGN, the ligand has two binding modes, depending on which sugar ring is engaged in the $Ca^{2+}$ ligation, but the favoured orientation of the molecule is the opposite of the one in DC-SIGN CRD, due to a steric clash with Lys299. In this direction DC-SIGN displays Val351, but in langerin this residue is replaced with Ala289. The absence of additional methyl groups compared to valine that could contact alanine, and the lack of substituents on C6 of the glycomimetic cyclohexane ring prevent the formation of van der Waals interactions in this region, and cause a low affinity between the pseudo-mannobioside and langerin.

The D3 arm of $Man_9$ exhibits a linear Manα1-2Manα1-6Man trisaccharide that was also the basis of glycomimetic structures.[246] By replacing the central mannose with a conformationally constrained carbocyclic diol, the pseudo-trisaccharide **1.20** was synthesized. NMR experiments indicated a high degree of flexibility, mainly due to the 1-2 and 1-6 glycosidic linkages, but the ligand inhibits DC-SIGN binding to mannosylated-BSA surface with an increased affinity of 125 μM, as compared to one order of magnitude higher $IC_{50}$ of pseudo-disaccharide (1.0 mM) and mannose (1.8 mM) in the same SPR assay. Moreover, the pseudo-trisaccharide successfully inhibits HIV infection of DC-SIGN$^+$ B-THP-1 cells.

Some equivocal observations about the pseudo-trisaccharide, however, raised enough doubt about its inhibitory potency to promote further investigations.[247] (**Figure 27**) Firstly, the order of magnitude difference in affinity between the pseudo-di and trisaccharide was not confirmed by the X-ray structure of the two ligands bound in the CRD (PDB codes: 2XR6, 2XR5). No additional interactions were identified and the binding mode, ligand epitopes and protein sites involved in the binding essentially did not differ in the two cases. Additionally, this difference in inhibitory potency was lost as soon as the monovalent glycomimetics were tethered to multivalent systems. Isothermal titration calorimetry (ITC), analytical ultracentrifugation (AUC), static light-scattering (SLS), and dynamic light scattering (DLS) experiments finally unveiled the reason behind the increased affinity of the tested pseudo-trisaccharide. The relatively small molecule can function as a bivalent ligand that clusters two DC-SIGN tetramers and this results in an artificial overestimation of affinity at low ligand-to protein ratios. This phenomenon was not observed in multivalent systems, when pseudo-trisaccharide is tethered to larger scaffolds. This negative feature and considering that the pseudo-trisaccharide is synthetically more demanding than the pseudo-disaccharide, the second ligand became the lead structure for further glycomimetic development.

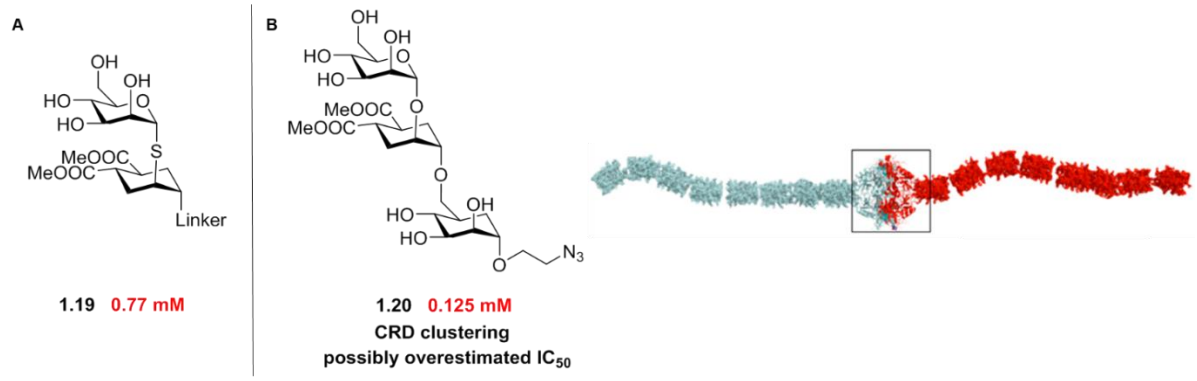



***Figure 27. A.:*** *Pseudo-thiosaccharide. Replacement of the O-glycosidic bond with a sulfur atom can lead to enhanced enzymatic stability, while the three-dimensional structure of the ligand is retained. The affinities of the O- and S-glycosides were almost identical (0.72 and 0.77 mM, respectively).*

*B. Pseudo-trisaccharide. Although the order of magnitude affinity increase for the monovalent trisaccharide-mimetic compared to the parent pseudo-mannobioside* ***1.18*** *was encouraging, this difference was lost as soon as the ligands were presented on multivalent scaffolds. Neither could the X-ray structure of both ligands complexed with the CRD of DC-SIGN explain the inhibitory potency increase, as apparently the two binding modes are almost identical. Further studies by ITC, AUC, SLS and DLS unveiled the CRD clustering capability of 1.20, which leads to artificial overestimation of the affinity of the monovalent trisaccharide in solution-phase. Proposed model from Sutkeviciute et al.*[247]

The X-ray structure of the pseudo-disaccharide, as well as the cyclohexane ring's close proximity to the CRD surface suggested that functionalization in the dimethyl ester positions could establish additional interactions with the protein. The position of the methyl ester substituents pointed towards an area on the CRD surface where larger groups could be accommodated, and presumably contact the surface. In particular, the introduction of lipophilic groups was suspected to contribute to ligand potency, by engaging in hydrophobic interactions with the CRD surface in areas which were not occupied by natural ligands. From the variously tethered (Me, $CH_2$-$CH_2$-$NH_2$, $CH_2$-CH=CH, $CH_2$-$CH_2$-$N_3$) pseudo-mannobiosides, a small library of bisamide derivatives (**1.21**) was developed with the allylic linker.[248] From these ligands, the bis(benzylamide)-substituted molecules - with or without further functional groups on the aromatic ring - inhibited DC adhesion to mannan-coated plates with the highest potency ($IC_{50}$ values in the mM range). Compared to pseudo-mannobioside, this meant a 2 order of magnitude affinity enhancement. On the other hand, preliminary SPR studies showed that tertiary amides in this position resulted in poor affinities towards DC-SIGN.[249]

The bis(benzylamide) series served as a promising starting point for further studies. The allylic tether was replaced by azido-ethanol linker because the first was found to slightly influence the measured affinity by aspecific interactions with the protein. Furthermore, by featuring an azido group, the monovalent ligands can be readily "clicked" on multimeric scaffolds in copper(I)-catalyzed alkyne-azide cycloadditions (CuAAC).

Docking studies indicated that the presence of H-bond donors or acceptors out of the plane of the aromatic ring could establish H-bonds with amino acid residues in the CRD and contribute to the affinity of the glycomimetic ligands. Thus, a library of 15 bis(benzylamide) ligands was synthesized,[250] based on considerations like participation in H-bonding, solubility and synthetic accessibility – all of these structures exhibited polar, O-containing functional groups on the two benzyl substituents. STD-NMR confirmed that the conformation of the bis(benzylamide) molecules in the binding site was similar to the pseudo-disaccharide, and that the aromatic rings formed lipophilic interactions with the protein surface. Notably, with the exception of 3 ligands, the 15 membered library showed increased activity towards DC-SIGN compared to the dimethyl ester **1.18** in SPR competition assays on a mannosylated BSA surface. Most of the ligands exhibited an $IC_{50}$ of ca. 0.3-0.4 mM, while in the same assay, the structurally significantly more complex $Man_9$ exhibited an inhibitory potency of 130 μM.

The role in H-bonding was deduced by comparing the ligand, bearing 4-hydroxymethelene groups on the aryl-ring (**1.21a**, also referred to as Man030) and the one with methyl esters in the same position. Since these ligands show similar affinities, it was postulated that the functional groups participate as H-bond acceptors in the interactions. Fluorine atoms and lipophilic groups attached to the phenyl ring virtually did not influence the potency.

A suitable balance between affinity ($IC_{50}$: 325 μM), selectivity, water solubility and synthetic accessibility was established in the case of Man030 (**1.21a**) which was the lead for further optimizations and the construction of multivalent systems, like glycodendrimers. Docking studies and STD-NMR experiments served as a proof that the ligand is bound by DC-SIGN in its main conformation, in a similar binding mode as the pseudo-mannobioside dimethyl-ester **1.18**.

Similarly to the pseudo-mannobioside, the bis(benzylamide) derivatives were also evaluated for their selectivity against langerin and displayed low affinity towards the lectin in single-point inhibitory assays, even at 100% DC-SIGN inhibition levels. This underlines that glycomimetic ligands can exhibit high affinity towards a target lectin, while maintaining selectivity towards another, with a highly similar CRD structure.

Achieving selectivity can be approached from various directions, though. In the classical strategy, ligands are modified and optimized to form additional contacts, or stronger interactions with the protein. During rational-differential design, elements are incorporated in the ligand structure to hinder, abolish binding to proteins which should be specifically avoided by the bioactive molecule. The first approach – although the previous examples of glycomimetic inhibitors all followed this one – is challenging, especially in the case of shallow, extended, accessible binding sites, like the CRD of DC-SIGN. The CRD of langerin, however, contains a certain residue and binds to an uncharacteristic carbohydrate-epitope that inspired a set of pseudo-mannobioside derivatives, with enhanced selectivity between the two lectins.

The natural ligands of langerin include 6-$SO_4$-Gal terminated glycans (see non-optimal binding mode **1.4.1**), as well as sulfated GlcNAc-containing GAGs. The crystal structure of langerin CRD complexed with the disulfated ligand GlcNS6S (PDB code: 5G6U),[251] together with site-directed mutagenesis identified K313 as the key mediator of accommodating negatively charged moieties in the binding site. Mutations in this residue, like K313A, can lead to a complete loss of binding. On the contrary, inclusion of a positive charge in position 6 could result in electrostatic repulsion, or steric hindrance with the amino acid sidechain. The 6-amino modified pseudo-disaccharide derivative **1.22a** served as the proof of this theory, as the ligand had drastically increased selectivity towards DC-SIGN over langerin, compared to the parent glycomimetic. Additionally, its relative affinity for the first CLR has also improved ($IC_{50}$ = 481 μM, as opposed to 956 μM); analogously to 6-amino-Man030 ($IC_{50}$ = 254 μM, as opposed to 329 μM). (**Figure 28**)

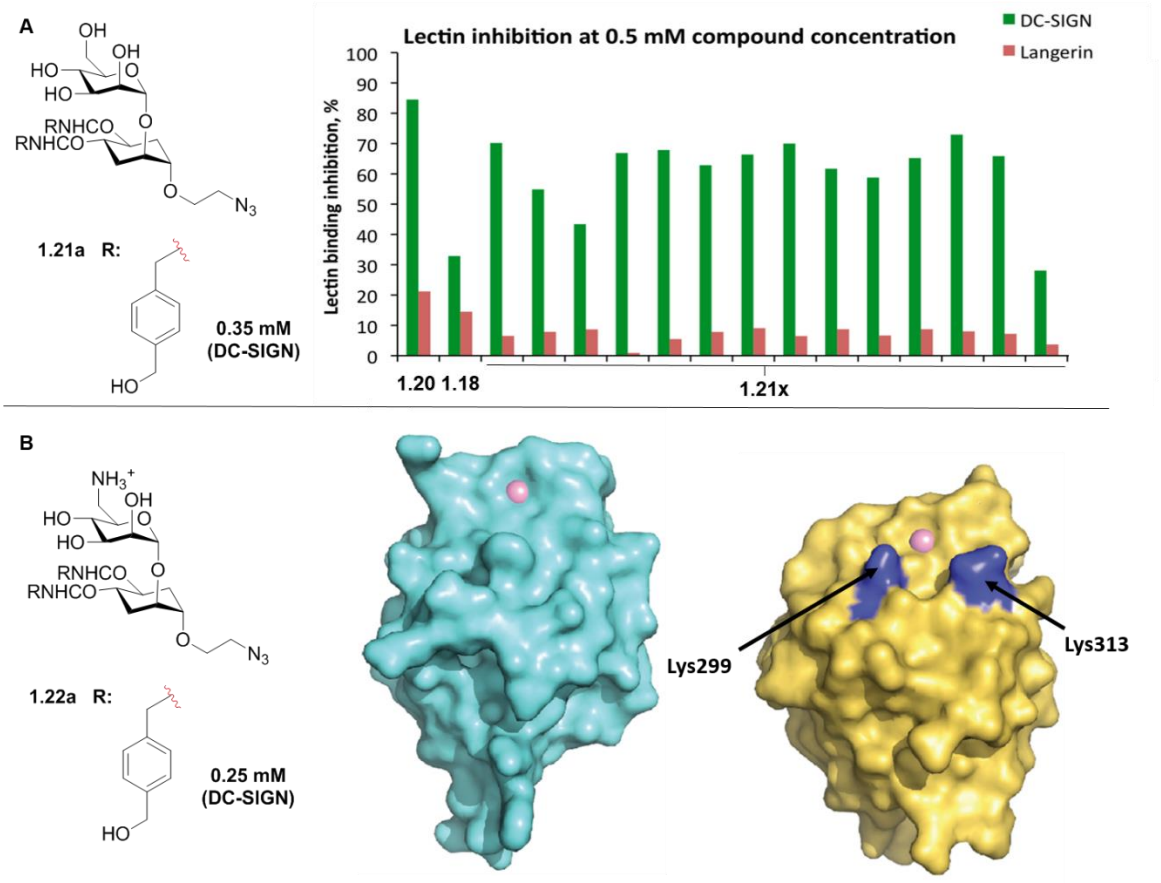



***Figure 28.*** *A. Bis(benzylamide) derivatives of pseudo-disaccharide* ***1.18****. Based on the X-ray crystal structure of* ***1.18*** *complexed with the CRD, various amide derivatives were synthesized to form additional interactions with the proteins surface. In particular, benzylamides showed superior affinity towards DC-SIGN, compared to* ***1.18****. Additionally, they were selective against langerin, a CLR that promotes HIV eradication by inducing the formation of Birbeck granules.* ***1.21a****, also known as Man030 offered an optimal balance between good affinity, selectivity, solubility in water and straightforward synthesis. The ligand's docked pose in the CRD of DC-SIGN (right) qualitatively accounted for the STD-NMR data. Modified from Varga et al.*[250]

*B. Highly selective monovalent DC-SIGN inhibitor conceived by rational-differential design. The crystal structure of langerin CRD (right, unligated) complexed with the disulfated ligand GlcNS6S (PDB code: 5G6U) and mutagenesis studies identified K313 as the key mediator for interacting with negatively charged groups in C6. Positive charges in this position were suspected to cause electrostatic repulsion and/or steric clashes, and therefore blocked binding by langerin. Compound* ***1.22a*** *showed not only exceptional selectivity towards DC-SIGN (left, unligated CRD) over langerin, but also enhanced affinity towards the first CLR. Modified from Porkolab et al.*[251]

# Part I

## Chapter I

# 2.1 SYNTHESIS OF A GLYCOMIMETIC LIGAND LIBRARY

## 2.1.1 DESIGN AND SYNTHESIS OF THE LIGAND LIBRARY

### 2.1.1.1 Selectively targeted CLRs could shape the immune response for therapeutic purposes

The fundamental role of CLRs in the immune response has been extensively discussed in the introduction. Depending on the receptor involved, CLRs are capable of activating different APC signalling pathways, induce cytokine secretion or the expression of co-stimulatory molecules. Furthermore, it has also been demonstrated that an early intervention in the immune signalling can steer the response towards medically desired directions. Selective targeting of specific C-type lectins can lead to the initiation of cellular or humoral immune pathways, like macrophage and B-cell activation, or elicit pro-/anti-inflammatory and allergic responses. These roles endow CLRs with an enormous potential for applications in areas, such as cancer immunotherapy, the treatment of autoimmune diseases or allergy, and vaccine development.

However, many aspects of the vital sugar-lectin interactions remain to be understood and described, with regards to the CLRs' structure, functions and carbohydrate specificity, as well as the in-depth exploration of key-epitopes and CRD sites involved in glycan recognition events. These are central points to address for the therapeutic exploitation of CLRs. Therefore, the main objective of the Immunoshape European Training Network is the development of highly selective lead structures as glycan-based immunotherapeutics for the treatment of cancer, autoimmune diseases and allergy.

CLRs do not always show high selectivity against natural glycans, and *vice versa*, pathogens are often recognized by several C-type lectin receptors. Selective inhibitors targeting exclusively one type of CLR on APCs, will thus have improved chances to induce only the desired responses (e.g immune-stimulatory or immune-suppressive), without any cross-reactivity. A powerful method to identify high-affinity and selective ligands for these receptors is the screening of glycan or glycomimetic libraries against different CLRs.

This concept is radically distinct from the previously seen approach of ligand optimization for a single lectin (e.g. DC-SIGN, see Introduction) and benefits greatly from high-throughput technologies, such as microarray-based binding assays. Several partners in the Immunoshape Network study immunologically-relevant glycan-based structures and employ state-of-the-art synthetic methods, involving chemoenzymatic diversification and automated solid-phase carbohydrate synthesis.

Our research group has been focusing on the synthesis of a glycomimetic ligand library via classical, solution-phase methods. This approach offers more freedom for major and minor modifications in the structure of glycomimetics, and facilitates tailored introduction of pharmacophores and other functional elements, like a linker moiety. In **1.5.2**, the numerous advantages of glycomimetics over natural glycans were discussed in details – these include improved (or altered) affinity and selectivity towards specific CLRs, enhanced metabolic stability, "drug-like" PK/PD profile and simpler synthesis than complex, natural oligosaccharides.

### 2.1.1.2 General structure of the glycomimetic ligands

As discussed earlier, CLRs sharing the EPN motif in their CRD can be expected to recognize both Man and Fuc epitopes in the principal sugar-binding site, while secondary structural elements can significantly enhance binding to certain lectins and lead to selectivity over others. Therefore, we envisioned a glycomimetic ligand library based on mannose and fucose, equipped with structurally diverse aglycon motifs. We anticipated that by screening various CLRs against this library, we would be able to identify active ligands that can selectively target specific CLRs on APCs.

During the initial library development, an already established synthetic route for DC-SIGN inhibitors was followed and expanded, in order to prepare a diverse set of mannosylated glycomimetics. Later, the same synthetic strategy was adapted to the synthesis of fucose-based ligands. Members of the library were numbered as **ISh0XX**, based on the chronological order of the synthesis. Subsequently, different CLRs available through the Immunoshape Network were screened against this ligand library that comprises both mannosylated and fucosylated glycomimetics.
The comparative assessment of binding-affinities on a large number of ligands can be extremely resource-demanding. The problems are rooted in the low affinities of monovalent ligands towards CLRs, while the construction of multivalent systems, like dendrimers, requires relatively large amounts of the ligand and increases the burden of synthetic effort. Moreover, the limited access to pure CLRs narrows down the applicability of techniques like SPR and NMR that use significant amounts of the protein.
Microarray technology is a robust, miniaturized binding-assay with accordingly low protein and ligand demand. This facile method provides a platform for the rapid identification of high affinity library ligands towards specific CLRs. During screening, the appropriately functionalized glycomimetic compounds are printed on functionalized glass-slide surfaces and the chips are incubated with fluorescently tagged CLRs. The read-out of the binding profile is fast and straightforward, and complementary assays (like SPR) can be used to validate the results on a smaller set of compounds. Cross-comparison of these profiles over different screenings with various CLRs requires standard, optimized microarray fabrication and incubation conditions, which also enable to determination of experimental reproducibility (more on this subject in a following section). (**Figure 29**)

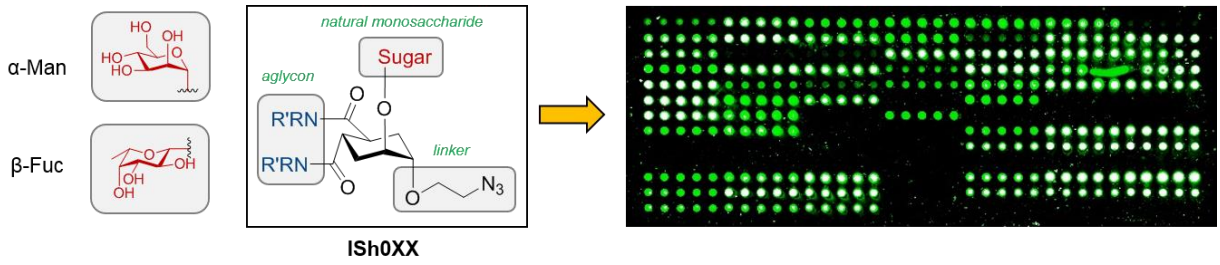



***Figure 29.*** *Glycomimetic approach to target C-type lectins. A mannose and fucose-based ligand library has been synthesized and diversified with bisamide-moieties. The molecules are equipped with an azido-group terminated linker that enables their presentation on multivalent platforms. As a robust binding assay, we chose microarray technology to screen various CLRs against the glycomimetic library, to identify high-affinity and potentially selective ligands.*

The orthogonal "handles" incorporated in the structure of our library members serve several purposes. On one hand, these linkers were used as anchors for ligand immobilization the microarray slides. Furthermore, once array screening and further binding assays (e.g. SPR) provide a selection of the most potent monovalent ligands, the glycomimetics can be tethered to core scaffolds, in order to build multivalent glycoconjugates (such as dendrons

and dendrimers, glycopolymers, GNPs or NGPs). These constructs typically exhibit increased affinity compared to the same number of monomers individually presented, because of enhanced avidity effects (see Introduction). Quantitative SPR assays and crystallographic data can further drive the rational design of high affinity inhibitors for specific CLRs. These methods are used to map crucial binding epitopes and exploitable sites on the surface of the CRD, to form new interactions between the CLRs and ligands.

As a future, late-stage evaluation option, the immunomodulatory properties of glycomimetic ligands can be assessed in functional immune assays. The top candidates may be conjugated with known tumor-antigens, anticipating that their presence facilitates the CLR-mediated antigen-uptake into DCs in human *ex-vivo* skin models. The subsequent antigen processing, stimulatory or tolerance-inducing response depends on the type of CLR and APC recognizing these antigens. Eventually, the selective targeting of DCs may improve anti-tumor immunity.

### 2.1.1.3 A convergent synthetic route leading to a common glycomimetic scaffold and its diversification

#### 2.1.3.1 Synthesis of the mannosylated scaffold

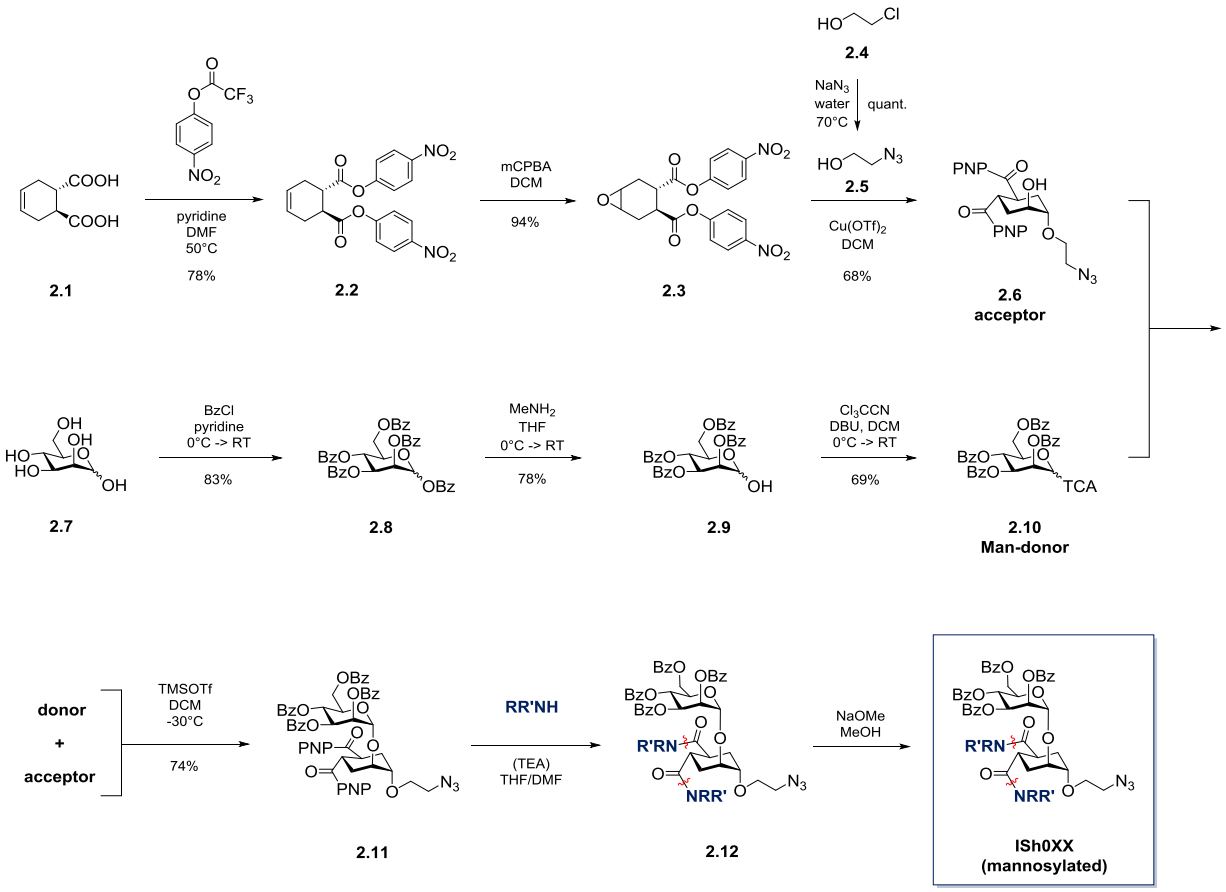
The previously described small library of pseudo-mannobioside-based bis(benzylamide) inhibitors (**1.21**) showed improved affinity towards DC-SIGN over the natural disaccharide and simultaneously, enhanced selectivity over langerin.[250] The results suggested that the set of target CLRs could be expanded to other mannose-/fucose-specific lectins by using the major monosaccharide epitope as an anchor in the primary $Ca^{2+}$ binding site, while modifications of secondary structural elements, such as the amide-moieties can tune the activity and selectivity of glycomimetic ligands. This is supported by important observations that are the basic pillars of our library development.

First, the presence of Glu-Pro-Asn motif in the vicinity of the principal binding site determines a preference for mannose and fucose. These sugars will therefore be used to ligate the glycomimetics at this site. Second, the conformationally constrained cyclohexane ring enables a unique, well-defined conformation of the mimic and favors a well-deifned binding mode in the CRD of DC-SIGN by interacting with Val351, while it also improves the enzymatic resistance of these compounds. Additionally, the X-ray structure of the pseudo-disaccharide **1.18** (**Figure 26**) in complex with DC-SIGN indicated that the methyl-ester groups appended to this sugar-mimic ring are projected towards the protein surface and indeed, their replacement with amide moieties significantly influenced the affinity towards DC-SIGN, based on the nature of these amide groups (see **1.7.3**). Lipophilic bis(benzylamides), for example, could enhance $IC_{50}$ values by an order of magnitude, while tertiary amides virtually abrogated binding to the receptor.[249] Likewise, varying solely the amide-moieties resulted in various levels of selectivity towards DC-SIGN over langerin.

On the whole, we can expect that by modifying the pseudo-disaccharide structure with diverse amide groups, the **ISh0XX** ligands can achieve increased affinity towards specific CLRs and they can be selectively recognized by certain receptors and not by others. Moreover, decorating the glycomimetic structures with both mannose and fucose can broaden the range of potential CLR targets and/or interactions in the primary binding site; as well as help the evaluation of the epitope's role.

These considerations allow us to rely on the previously established synthetic route to prepare the mannosylated library. The pathway serves our purpose particularly well, as the library diversification takes place during the penultimate step, and so, from there we can follow a divergent synthetic strategy by scaling-up the common scaffold **2.11** and appending the amide moieties at a late stage. The very last step is a straightforward Zemplén-

deprotection[252] that – in most of the cases - can be performed immediately after the amide coupling, leading directly to the mannosylated library members **ISh0XX**, equipped with the linker arm. (**Scheme 1**)




***Scheme 1.*** *Synthesis of the mannosylated glycomimetic library. Starting from the enantiomerically pure diacid* ***2.1****, activated ester groups and a short azidoethanol-linker were incorporated in the structure of the acceptor* ***2.6****. This trans-diaxially substituted cyclohexane ring mimics the conformation of 1,2-substituted mannose. The mannose-donor* ***2.10*** *was prepared through benzoylation, selective deprotection at the anomeric center and activation for the glycosylation step as a trichloroacetoimidate. After formation of the glycosidic linkage, a common scaffold was obtained (****2.11****) that readily reacts with diverse amines. In the last step, the mannobioside mimetics were debenzoylated under Zemplén conditions.*

The disaccharide scaffold is built from a natural mannose residue and an unnatural sugar-mimic ring. The orientation of the functional groups on the cyclohexane ring are crucial, as their proper configuration imitates the *trans*-diaxial hydroxyl groups on C1 and C2 positions of the nonreducing terminus mannose ring in Manα1-2Man. Therefore, the position of these substituents has an influence on the final conformation and the binding characteristics of the final glycomimetic ligands.

For this reason, the required configuration of two functionalities of the four is already implemented in our starting material - (1S,2S)-4-cyclohexene-1,2-dicarboxylic acid **2.1** - that was custom-synthesized by Alchemy, following a route described by our group.[253]

The diacid functionalities were transformed into the corresponding *p*-nitrophenyl esters; on one hand, to protect the acid functionalities in the subsequent steps, and on the other, to pave the way later for a facile insertion of the amide moieties. Pentafluoro-phenyl esters had also been tested previously with the same purpose, but this activating-group was discarded due to the lower yields obtained during the glycosylation step.[248] Reacting diacid **2.1** overnight at 50°C, with 4-nitrophenyl trifluoroacetate in the presence of pyridine allows us to crystallize the pure product from cold diethyl-ether without any column purification step. This procedure has also reached higher yields (>70%) than the previously reported ester activation step with 4-nitrophenol, EDC and DIPEA

(62%). The formation of tetrahydrophtalic-anhydride had been described earlier at elevated temperatures, therefore the reaction rate could not be accelerated by heating.[254] Nonetheless, the reaction had satisfactory yields to scale-up from an expensive starting material, in order to produce a larger amount of the common scaffold **2.11**.

The carboxylic acid protection/activation step was followed by the epoxidation of the double bond with mCPBA. The reaction has excellent yields and again, further purification of the isolated product was not needed.

The acid-catalyzed hydrolysis of the epoxide **2.3** leads to the *trans*-diaxial product **2.6** – the acceptor in the following glycosylation reaction. The selective opening of the epoxide is a key step that eventually determines the appropriate conformation of the glycomimetic ligands, and introduces the linker tail in the structure. The alcoholysis of the epoxide was catalyzed by $Cu(OTf)_2$ – a Lewis acid that offered the highest yields and selectivity during a previously reported Lewis acid and electron-transfer catalyst screening for the opening of this epoxide.[255]

An appropriately chosen alcohol in the epoxide opening step can simplify the future conjugation of the ligands to multivalent supports, such as the surface of microarray chips or dendrimers. The functional group at the end of the tether also defines the type of chemistry that can be employed to link the compounds to multimeric systems. Although numerous methods have been described, in the majority of the cases, coupling relies on the practical and prosperous chemistry of amine functionalities, or the copper(I)-catalyzed alkyne-azide cycloadditions (CuAAC) between an azido-group and a triple-bond. Additionally, these strategies are applied by other partners of the Immunoshape Network, and protocol uniformity among collaborating research groups, i.e. standard coupling chemistry enables a more efficient cooperation among different sites. The inclusion of an azide-moiety is suitable for the exploitation of both strategies, as the group can be easily reduced with palladium catalyst, to afford an amino-functionality, if necessary.

2.1.3.2 Trials to elongate the linker arm

Individual glycomimetic ligands and libraries of fewer members were typically synthesized in our group by opening the epoxide with allyl alcohol, 2-bromoethanol or 2-chloroethanol.[242b, 248, 256] Using allyl alcohol as the linker is inconvenient, because the double bond does not allow a straightforward coupling with the previously mentioned multivalent platforms. Moreover, the same functional group is suspected to interact non-specifically with DC-SIGN, as indicated by saturated signals in STD-NMR experiments.[250] 2-Bromoethanol could not entirely dissolve the copper-promoter, and thus, dichloromethane had to be added to the reaction, which, on the contrary decreased the conversion and the yield (ca. 40%) of the reaction.[242b] Furthermore, the purification of the crude was difficult and degradation of the reagent was often observed over time. Using 2-chloroethanol, both as reagent and solvent – both the catalytic amount of copper, as well as the substrate are dissolved - significantly improved the yield of the epoxide opening, compared to 2-bromoethanol. The replacement of the chlorine atom with sodium azide in DMF, however, could only be performed one-by-one for each ligand after the transformation of the activated ester groups into the bisamide substituents. This cumbersome protocol was omitted by Varga *et al.*,[250] who directly installed the azide-group at the end of the linker - prone to react with alkynes for multivalent presentation of the ligands - by using 2-azidoethan-1-ol in the epoxide-alcoholysis step. Ring opening by 2-azidoethan-1-ol can drastically simplify the required synthetic steps during our gram-scale preparation of the common scaffold **2.11** and thus, this alcohol was considered first as a linker for the preparation of our library.

This small organic azide compound can be potentially hazardous and explosive, however, and therefore, it needs to be handled with care. As a rule of thumb, the following equation indicates threshold for organic azide safety:[257]

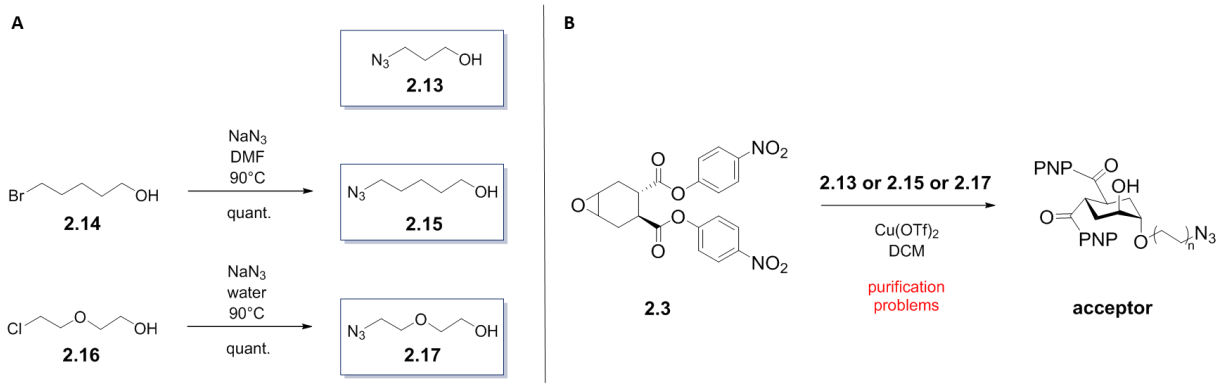
$$\frac{N(carbons) + N(oxigens)}{N(nitrogens)} \geq 3$$

As 2-azidoethan-1-ol does not fulfill this criterion (left side equals 1), it is not safe to isolate the liquid in its pure form, and its distillation is essentially forbidden. Safety measurements include the storage of small amounts, only in solution, at low temperatures, limited to short periods of time. The compound was synthesized from 2-chloroethanol, with sodium azide – metal spatulas (and generally any contact with metals, more specifically heavy metals) should be avoided during weighing of the salt, and extra attention must be paid not to leave any powder in ground glass joints. Wastes of azide-containing solutions must never be disposed in drains (toxicity of azide salts, metal pipelines) and have to be collected in dedicated containers, in order to prevent the release of the highly toxic hydrazoic acid ($pK_a$ = 4.6) under acidic conditions. Alternatively, less than 5% sodium azide content in aqueous solutions may be destroyed by using freshly prepared nitrous acid, following strict safety rules during the procedure (see details in the experimental section).

In our case, the tether was intended to be used primarily for ligand immobilization during the microarray-based CLR screening process, and this role could later be extended to linking the most promising candidates to multivalent scaffolds. The micro-scale surface of commercially available microarray slides is not two-dimensional, however, the glass-chips are coated with an appropriately functionalized, permeable polymer that has an increased immobilization capacity. To ensure the CLRs' sufficient access to our ligands, and provide the glycomimetics a certain degree of freedom to assume conformations which can be recognized by the lectins, we attempted to incorporate longer linker arms at this stage of the synthesis. We reasoned that a more extended, flexible spacer, already attached to the molecules, could project the glycomimetics away from the glass microarray surface and the three-dimensional polymer layer, so the solution phase CLRs could reach their ligands more easily. Furthermore, this slightly more distant orientation would help to maintain the native three-dimensional structure and functionality of the ligands.

Trials were conducted to open the epoxide **2.5** with azidoalcohols of different length which are also safer to synthesize and use (C : N = 1). The most extensive attempts were focused on equipping the cyclohexane moiety with commercially available 3-azidopropan-1-ol (**2.13**), however the product acceptor could not be entirely purified and was always contaminated with significant amounts of the alcohol. This is due to the fact that while 2-azidoethan-1-ol can be wholly removed from the reaction crude by evaporation (b.p.: 77°C at 32 mbar), 3-azidopropan-1-ol is not nearly as volatile as the shorter alcohol (b.p.: 62°C at 3-4 mbar). During column chromatography, purification of the product was hampered by the lengthy streaking of 3-azidopropan-1-ol. Similarly, our attempts failed with 5-azidopentan-1-ol (**2.15**), as well as with the more polar 2-(2-azidoethoxy)ethanol (**2.17**), whose additional aim was to increase the overall water-solubility of the glycomimetic ligands. These azide-terminated linkers were synthesized from 5-bromopentan-1-ol (**2.14**) and 2-(2-chloroethoxy)ethanol (**2.16**), respectively, with sodium azide. (**Scheme 2**)

As the following step of the synthetic route is the particularly sensitive glycosylation, even minimal amounts of residual nucleophile from the epoxide-opening could spoil the central reaction in our synthetic pathway. Instead of the continued search for another alcohol, we eventually decided to follow the established epoxide-opening with 2-azidoethan-1-ol and to elongate the linker with a double-functionalized spacer right before printing our glycomimetic library on the microarray chips (see Chapter II). This strategy also leaves room for various modifications on the short tether, or the introduction of various motifs in this position - like conjugation of the monovalent ligands to multivalent dendrons, without granting the constructs too much flexibility (which can also be disadvantageous in certain cases).




***Scheme 2.*** *Epoxide-opening trials with various azido-alcohols, in order to incorporate a longer, safer-to-handle linker than 2-azidoethan-1-ol. A. 3-azidopropan-1-ol is commercially available, 5-azidopentan-1-ol and 2-(2-azidoethoxy)ethanol were prepared with $NaN_3$ from 5-bromopentan-1-ol and 2-(2-chloroethoxy)ethanol, respectively.*
*B. The product acceptor could not be purified from residual azidoalcohols which were added in excess to the reaction. While leftover 2-azidoethan-1-ol can be evaporated from the reaction crude after the epoxide opening, longer linkers have too high boiling points. During column chromatography. The streaking alcohols are collected together with the product. Nucleophile impurities, even in small amounts, can represent serious problems during the following glycosylation step in the synthetic pathway.*

The *trans*-diaxial carbocyclic diester **2.6** plays the role of the acceptor in the following glycosylation step. The donor **2.10**, on the other hand, is a trichloroacetoimidate(TCA)-activated and benzoyl-protected mannose which was prepared from D-Man through well-known and straightforward transformations. All 5 hydroxyl-groups on the sugar were benzoylated with BzCl in pyridine (**2.8**), followed by the selective deprotection of the anomeric position with $MeNH_2$ (**2.9**). During previous trials in our group, acetyl-protected donors had been used for the glycosylation, but the product was obtained in low yields after a difficult purification. Using benzoates significantly improved the yield of this reaction and reduced the formation of orthoesters which are common byproducts of glycosylations in the presence of a participating group on C-2.
TCA activating groups are popular among carbohydrate chemists, thanks to their ease of synthesis – the anomeric hydroxyl group is readily converted into the imidate with trichloroacetonitrile, under basic conditions (here: DBU) – and the fact that this functionality is an excellent leaving group during the course of glycosylation, with satisfying stereoselective outcome of the newly formed glycosidic linkage. The downside of activating the protected mannose is that the TCA-donor is not stable during long periods of storage, even at -20°C, thus, the activation has to be performed not long before the glycosylation.
The connection of the unnatural acceptor moiety and the appropriately functionalized (protected and activated) mannose donor forms the link in the convergent first half of the synthetic pathway, leading to the common scaffold **2.11**. This is arguably the most delicate reaction in our sequence – traces of moisture need to be removed from the reactant mixture by co-evaporation with toluene, and by adding 4Å molecular sieves. The reaction is initiated by a Lewis acid activator (in our case TMSOTf), used in catalytic amounts, which promotes the cleavage of the leaving group. The so-formed oxocarbenium ion is stabilized as a benzoxonium ion - with the help of the axially oriented neighbouring group on C-2 – that is attacked in an $S_N2$-type reaction by the free OH group of the acceptor. The anchimeric-effect that defines the side of attack on the intermediate, and the strong anomeric-effect of mannose collectively lead to the exclusively 1,2-*trans*-glycosylated α-product. The glycosylation reaction is remarkably fast, and the step is carried out at -30°C in order to prevent the decomposition of the reactants and minimize the chances of possible side-reactions. The residual Lewis acid is quenched with TEA at the end of the reaction.
In summary, the convergent synthetic branches leading to the donor and acceptor were joint in the glycosylation step that yielded **2.11**, a common scaffold for the derivatization of the mannose-glycomimetic library. This

compound can be readily conjugated with different amine moieties as a result of the PNP-activation of the ester groups, and was therefore scaled-up to gram-scale. From here, we followed a divergent concept, to attach structurally diverse amines to the common pseudo-disaccharide scaffold – these bisamide-groups are the variable elements of the ligand library.

Based on the previous results with DC-SIGN and langerin,[250] we anticipated that the glycomimetic constructs' affinity and selectivity will be modified against various CLRs, depending on the nature of the amide groups (see **2.1.1**). At this point, we faced the challenging problem of having to select a diverse set of amines that will be appended on our molecules and screened against various CLRs. As we lacked a specific lectin target and because the library was synthesized with the main goal of screening - at that point - unspecified CLRs (that we would later have access to through our partners in the Immunoshape Network), rational choices could only originate from results with DC-SIGN. This lectin showed high affinity towards bis(benzylamide) moieties, and could not bind tertiary amides attached to the cyclohexane sugar-mimetic ring.[249] Furthermore, many lectin receptors have not yet been characterized, with respect to the structure of their binding pockets and preferred epitopes, therefore we turned to chemoinformatic tools to select a pool of structurally diverse amines which could contribute to the binding interactions between the pseudo-disaccharide ligands and the CRD of lectins.

2.1.3.3 Selection of diverse amines

With the help of Dr. Fabio Zuccotto from the University of Dundee, primary and secondary amines were collected from the compound databases available at eMolecules (>6.5 million compounds in August 2013) and Molport (>10 million compounds in December 2013), by using Markush structures. The findings were refined by applying physicochemical filters ($150 \leq MW \leq 275$; $9 \leq$ number of heavy atoms $\leq 20$; number of rotable bonds $< 6$; number of rings $> 0$, no undefined stereocentres), in agreement with "lead-like" drug properties. Structures containing potentially reactive species[258] and molecules that had been identified as pan-assay interference compounds (PAINS)[259] were removed from the lot. These amines could carry a risk of showing activity against a range of array-platforms, as well as proteins, typically by the means of metal chelation, chemical aggregation, redox activity, fluorescence, cysteine oxidation and binding promiscuity. These properties could significantly influence the results of binding assays, both *in vitro* and *in vivo*, and thus, should be avoided.

The remaining structures were clustered and the cluster centres defined by using chemoinformatic descriptors, for example extended connectivity fingerprint 4 (ECFP4), and by defining the maximum distance between cluster members (Tanimoto = 0.6). To allow visual inspection of the still vast number of amines, the original number of compounds in the cluster centres was narrowed down to 20%, while the molecular diversity was maximized by functional-class fingerprint 4 (FCFP4). These steps yielded 1116 primary amines (630 aliphatic and 486 aromatic) and 796 secondary amines (472 aliphatic and 324 aromatic), altogether 1912 structures that could be used for our glycomimetics. Representatives of the cluster centres were selected based on commercial availability and intuition.

The amide coupling reaction between the primary and secondary amines (added in excess) and the PNP-esters is straightforward, although insufficient reactivity or decomposition of the amine partner, and steric hindrance can slow down or block the formation of the two amide bonds. In cases, where passivity of the amine reactant was suspected, or when the ammonium salt was used for the reaction, TEA was added. Attempts to increase the reaction rate by heating, lead to the decomposition of the scaffold **2.11** even at 50°C, and was henceforward avoided.

Ideally, after completion, the reaction crude contained only the bisamide product, the *p*-nitrophenolate byproduct, the excess of amine (typically added in 3-5 equivalents) and possibly TEA. Where possible, i.e. the product did not contain any sensitive functional groups (that could significantly alter the polarity of the product upon protonation/deprotonation), the *p*-nitrophenolate was removed by basic workup, while the amine species

by acidic aqueous extraction. To facilitate the time-consuming and laborious workup steps, we installed Biotage ISOLERA® HM-N liquid-liquid extraction systems, in order to rapidly extract several reaction mixtures simultaneously. (**Figure 30**) The supported liquid extraction (SLE) method consists of dissolving the reaction crude in a compatible organic solvent and passing the solution through connected cartridges. These are packed with a solid matrix which is pre-conditioned with 1M HCl or 1M NaOH solutions in the different column elements, and a last (bottom) cartridge contains pure water, as a final washing step. Addition of a Vacmaster™ unit can speed up the descent of the organic phase by applying vacuum sucking at the end of the column. Optionally, any of the cartridges may be omitted (e.g. the acid-conditioned one, in case of undesired protonation of the product), and several extractions of the same type can be performed if necessary. The tetra-*O*-benzoylated bisamides (**2.12**) were only purified by flash chromatography where significant amounts of the amine or other by-products could not be removed by the SLE system.

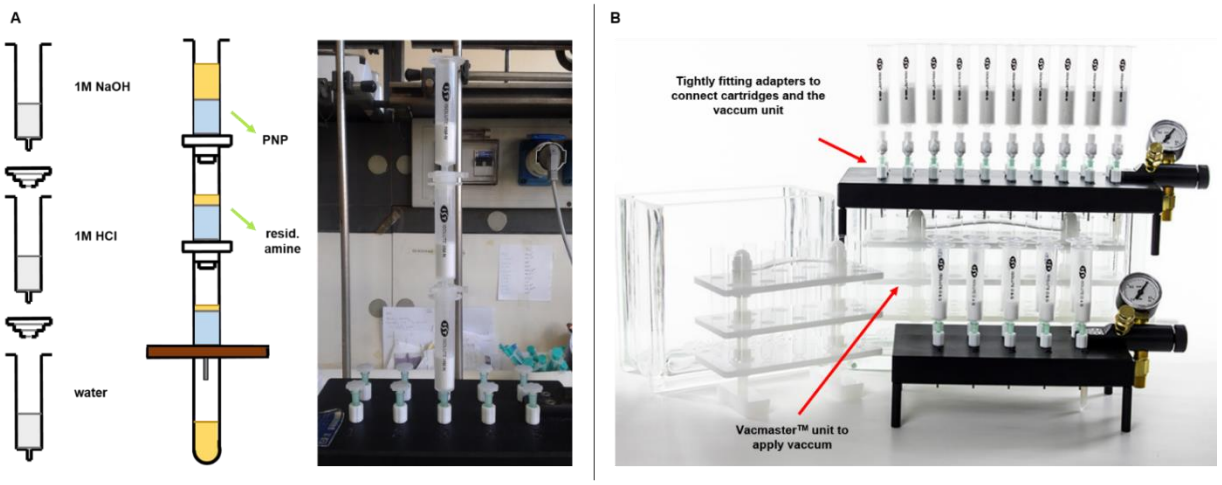



***Figure 30.*** *Rapid workup of the amide coupling reactions with Biotage ISOLERA® HM-N liquid-liquid extraction system. A. Solid matrix-packed cartridges were pre-conditioned with 1M NaOH, 1M HCl solutions and water and connected in a column system. The p-nitrophenol byproduct is removed on the basic cartridge, the excess amine is extracted in the acidic phase and the organic phase is washed on the final column with water. B. Vacuum can facilitate the descent of the organic phase which is then collected into test tubes in the Vacmaster™ unit. 10 samples can be extracted simultaneously – a great help for time-efficient preparation of a molecule library.*

The final debenzoylation under Zemplén-conditions is rapidly achieved by NaOMe in MeOH, and alternatively, this deprotection step can be carried out one-pot after the amide coupling. This way, even the simplified workup of the previous reaction might be skipped, and the two steps require only one chromatographic purification. The high boiling point of methyl-benzoate (199.6°C) coupled with the heat-sensitivity of the products does not allow the removal of the byproduct by evaporation.

### 2.1.3.4 β-fucosyalted glycomimetic ligands

As it has already been foreshadowed at the beginning of this section, we set out to expand the glycomimetic library to fucosylated pseudo-disaccharides, in addition to the mannose bearing ligands. According to our expectations, selective and/or high affinity CLR ligands could be identified among fucose-decorated glycomimetics whose structure was based on a natural mannobioside ligand of DC-SIGN. This prospect was based on the fact that often both of these sugar epitopes are recognized by lectins who share structural similarities in the CRD (EPN motif, see above and Introduction), and furthermore, that many mannosylated and fucosylated oligosaccharides (high Man structures, Lewis-antigens) can be found among the ligands of C-type lectins. The Glu-Pro-Asn residues

define the architecture in the direct vicinity of the $Ca^{2+}$ binding site where they primarily allow the ligation of two equatorial hydroxyl groups – 3-OH and 4-OH of mannose and 2-OH and 3-OH of fucose. (**Figure 31**) Some ligands are ligated in a noncanonical manner though, such as $Le^X$ in the CRD of DC-SIGN – the trisaccharide is tilted towards Val351 because of the atypical coordination by the equatorial 3-OH and axial 4-OH of the fucose moiety (see Introduction).

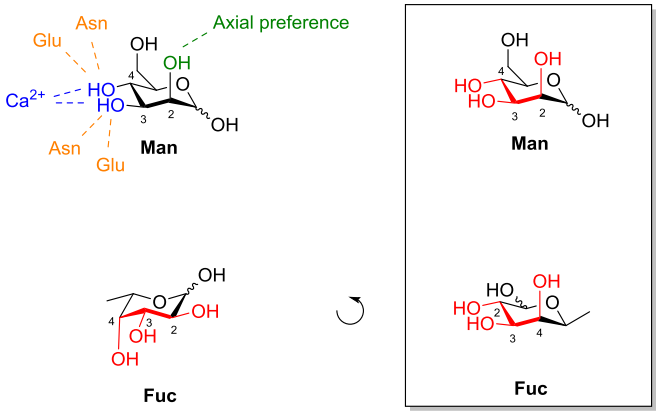



***Figure 31.*** *Minimal sugar epitopes for C-type lectins with an EPN motif in their CRD. These CLRs show the highest affinity towards mannose and fucose among natural monosaccharides, although they can often recognize glucose as well. In the case of Man, 3-OH and 4-OH are directly coordinated by the $Ca^{2+}$ and the surrounding Asn and Glu residues, while 2-OH is axially oriented. The same architecture is exhibited by fucose, although the metal-coordinated hydroxyl groups are 2-OH and 3-OH, while 4-OH is found in axial position.*

Fucose is among the few common monosacchride building blocks found essentially only in L-configuration in Nature. Although natural oligosaccharides contain almost exclusively α-linked fucose residues, from our glycomimetic point of view, the synthetic challenges associated with the different preparation routes leading to the two types of glycosidic linkages (α and β-Fuc) are incomparable. While the β-linked ligands could be prepared analogously to the mannosylated structures - by applying the same protecting group pattern and activation as in the case of the Man-donor - the substantially more difficult 1,2-*cis*-glycosylation of the donor could require a major effort during the synthetic pathway. 1,2-*cis*-glycoside linkage-formations are particularly problematic, and require carefully planned protecting strategies, activation methods (e.g. halogenation) and often special conditions during the glycosylation (e.g. solvent-effect), in order to achieve good yields and stereocontrol.
Additionally, for the same reason, it is reasonable to hypothesize that fucosidase enzymes encountering essentially only α-linked fucose in Nature, will less rapidly hydrolyze the β-fucoside structures.
In an attempt to predict how the synthetically less demanding β-fucosylated glycomimetics could interact with a CLR, we returned to the crystal structures and binding modes of $Le^X$ and pseudo-mannobioside **1.18** in the CRD of DC-SIGN. (**Figure 22, Figure 26**) As a starting point, a conformational search was performed for β-1,2-pseudofucobioside, and the findings were ranked. The 8$^{th}$ highest-scoring ligand exhibited extended conformation – similar to the binding mode of α-1,2-pseudo-mannobioside **1.18**. The $Ca^{2+}$ coordination observed for lacto-*N*-fucopentaose indicated that in spite of the configurational analogy between 3-OH and 4-OH of mannose and 2-OH and 3-OH of fucose, the oligosaccharide is tethered at the primary binding site through 3-OH (equatorial) and 4-OH (axial) of fucose. Therefore, the β-fucoside moiety of the glycomimetic ligand was rigidly overlapped with the fucose-residue in the LNFP-III-DC-SIGN CRD complex model (PDB: 1SL5). Then, this binding mode at the principal binding site was adopted to the crystal structure obtained with pseudo-mannobioside **1.18**, where the cyclohexane ring gets in close contact with Val351. Similarly to the crystallized glycomimetic ligand, β-fucosylated glycomimetic molecules are oriented towards this region, where

the cyclohexane ring and its substituents have the potential to interact with the protein surface, enhancing the affinity of the ligand. (**Figure 32**)

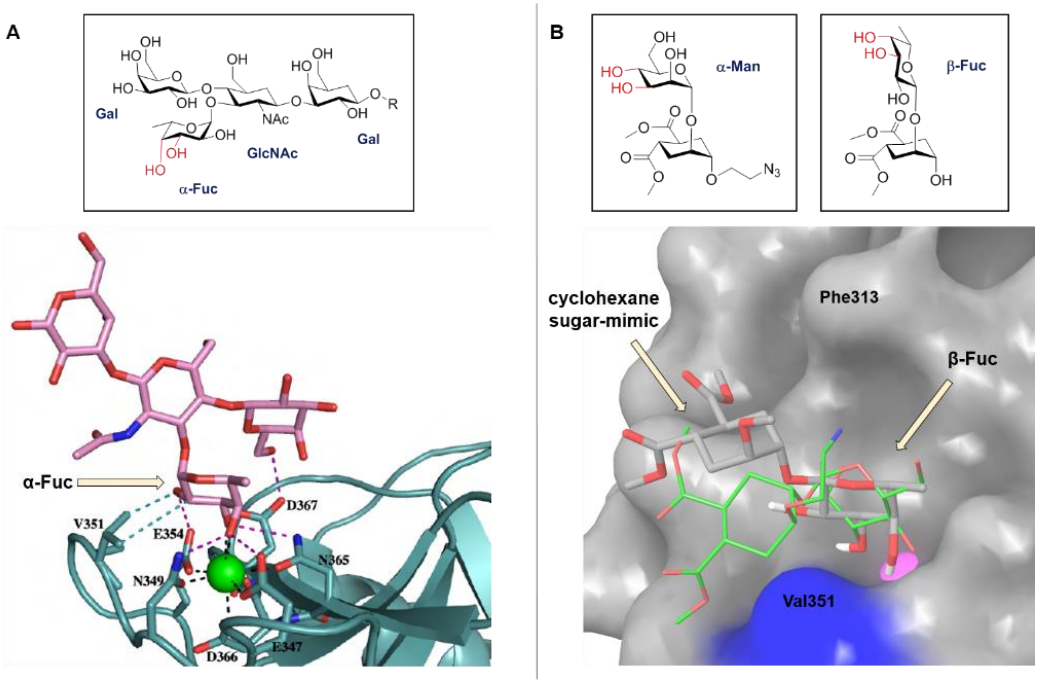
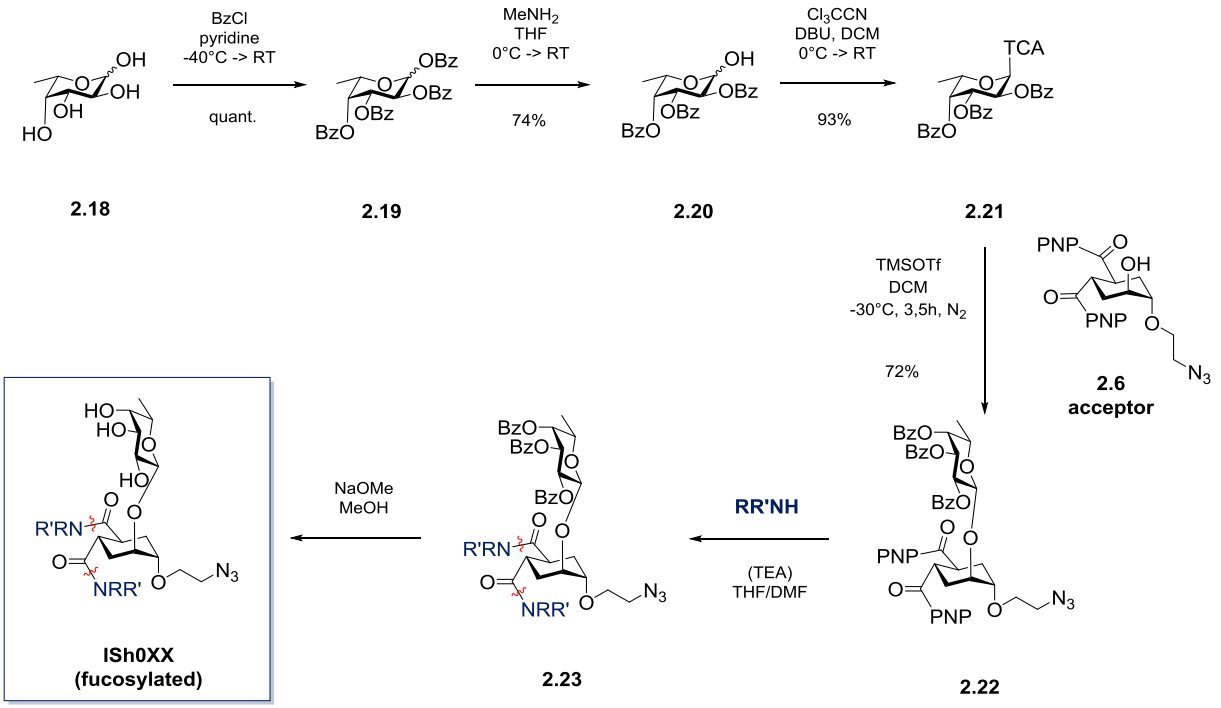



***Figure 32.*** *A. Crystal structure of lacto-N-fucopentaose in the CRD of DC-SIGN (PDB code: 1SL5). The α-Fuc residue is ligated by $Ca^{2+}$ through its 3-OH and 4-OH and engages in H-bonding with Val351. Modified from: Guo et al.*[227]

*B. Green ligand: 1,2-pseudo-mannobioside* ***1.18*** *binding mode in the CRD of DC-SIGN, based on the crystal structure of the complex (PDB code: 2XR5). Grey ligand: A conformational search had been performed for β1,2-pseudofucobioside, analogous to α1,2-pseudomannobioside* ***1.18****. The highest-scoring extended conformation ($8^{th}$ among the top conformations) was used to predict the binding mode of β-fucosylated glycomimetics in the CRD of DC-SIGN. Based on the lacto-N-fucopentaose structure, the 3 and 4 hydroxyl groups are assumed to engage in coordination with $Ca^{2+}$, while the cyclohexane ring, mimicking the second sugar residue leans towards Val351, similarly to the crystallized glycomimetic* ***1.18****. This suggests that – in the case of DC-SIGN, at least – glycomimetic ligands containing the same cyclohexane structure have similar potential to form additional interactions with the CRD through the sugar mimetic moiety and its substituents. Pseudo-fucobioside does not contain the azidoethanol linker moiety. The $Ca^{2+}$ coordinating OH-groups are highlighted in red. Image generated in Maestro, Schrödinger, LLC, New York, NY, 2018. Courtesy of Dr. Monica Civera.*

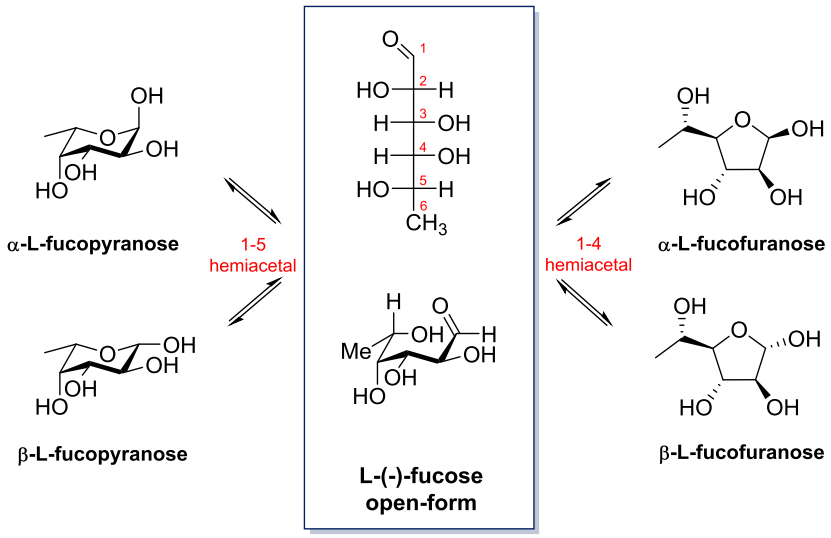
Based on this estimated binding orientation towards Val351, we concluded that β-fucoside glycomimetic ligands could potentially be recognized by the CRD of various C-type lectins and thus, we synthesized a number of β-fucose-bearing molecules.

Similarly to the Man-based library, the β-fucosylated ligands consist of a carbocyclic ring that can be linked to the monosaccharide in a 1,2-*trans*-glycosylation, using the previously described common acceptor **2.6**. The mannosylated donor-preparation route was adopted and adjusted to the synthesis of the 2,3,4-tri-*O*-benzoyl-L-fucopyranosyl-trichloroacetoimidate donor **2.21**. (**Scheme 3**)




***Scheme 3.*** *Synthesis of the fucosylated glycomimetic library. The fucose-donor* ***2.21*** *was prepared through benzoylation, selective deprotection at the anomeric center and activation for the glycosylation step as a trichloroacetoimidate. The synthetically less challenging 1,2-trans-glycosylation step leads to the β-fucosylated common scaffold (****2.22****) that readily reacts with diverse amines. The bisamide derivatives were debenzoylated under Zemplén conditions.*

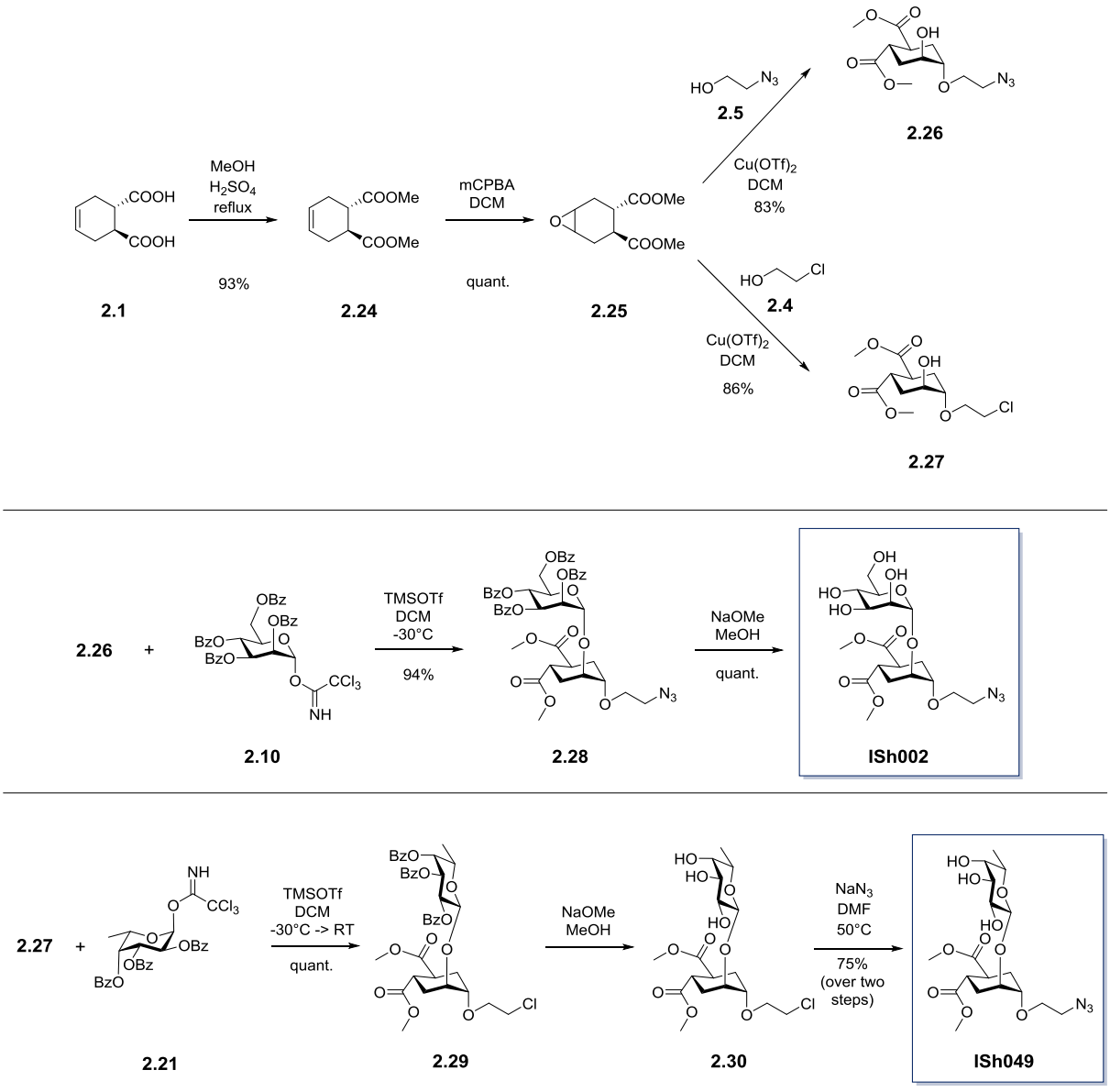
Based on our observations that the benzoyl-protected mannose donor proved to be more efficient during glycosylation than the acetylated one, BzCl (and initially also the less reactive $Bz_2O$) was used during the first step of the synthesis for the protection of the 4 hydroxyl groups on L-(-)-Fuc. Surprisingly, after the isolation of seemingly two different compounds with close $R_f$ values (0.26 and 0.24 in toluene : EtOAc = 97 : 3), which were suspected to be the α and β anomers, the NMR spectra indicated altogether four different, but similar structures with the same m/z value (ESI-MS calcd. for $C_{34}H_{28}O_9Na(M + Na^+)$: 603.16; found: 603.35). This is due to the formation of 4 isomers: α and β 1,2,3,4-tetra-*O*-benzyol-fucopyranose and α and β 1,2,3,4-tetra-*O*-benzoyl-fucofuranose. (**Scheme 4**) Separation of the latter, 5-membered ring form - and later its derivatives - is challenging throughout the entire synthetic sequence and reduces the atom-efficiency of the procedure, therefore its formation had to be minimized in the first step. We found that dissolution of L-(-)-Fuc in pyridine at low temperature (-40°C) and immediate addition of the reagent BzCl can reduce the time allowed to reach equilibrium between the two forms in solution. Furthermore, running the benzoylation reaction at this temperature, as long as the free-sugar is present (more polar compound than the mono-, di- and tri-benzoylated intermediates, or the product; lowest $R_f$ value) helped to limit the formation of the fucofuranoside form to less than 1%. A practical issue arises from the fact that benzoyl chloride freezes out from the reaction mixture over time (melting point in pure form: -1°C), and this makes stirring more difficult. Initially, this is not observed because the reaction is exothermic, but by adding more reagent, particular care needs to be taken on larger scales to avoid the formation of local hot spots in the mixture. Vigorous stirring facilitates the rapid benzoylation of the anomeric center, and continuous monitoring can help to indicate the point when the free sugar is consumed. At this point, the temperature can be elevated, as sugar ring opening and re-closure in the furanose-form is not a threat anymore.




***Scheme 4.*** *L-Fucose chain-ring tautomerism. In solution, the open form aldose closes into 5- or 6-membered rings. Depending on which hydroxyl group is involved in the hemiacetal formation, there are 4 various outcomes: closure between the aldehyde (C-1) and 5-OH leads to α and β anomers of the pyranoside-form, while α and β furanosides are formed if the carbonyl reacts with 4-OH. The formation of furanosides can be reduced by low temperatures and minimizing the time that the free-sugar spends in solution, i.e. rapidly protecting the anomeric hydroxyl group after dissolution.*

The next steps of anomeric deprotection and activation were carried out analogously to mannose, and by the end of the sequence, we obtained the 2,3,4-tri-*O*-benzoyl-fucopyranosyl-trichloroacetoimidate donor **2.21**. This activated sugar is significantly less reactive in the glycosylation, than its mannosylated counterpart. The reaction was initiated by the same TMSOTf promoter at low temperature (-30°C), but even after one hour, only partial conversion was observed. The temperature was therefore allowed to warm up - first until 0°C, then to room temperature - and after 3.5 hours, we obtained the product **2.22** in satisfactory yield (72%). This fucosylated scaffold was used for the derivatization of a smaller library, analogously to the Man-ligands (bisamide formation, deprotection).

Both the mannosylated dimethyl-ester **1.18** (from now on **ISh002**) and fucosylated ligands, carrying dimethyl-ester groups (**ISh049**), had been synthesized and added to the library. The acceptor, leading to these ligands, differs from the PNP-activated diester **2.6**, as this one is equipped with the two methylester groups at the start of the synthesis (Fischer-esterification). Later the already discussed epoxidation, epoxide opening, glycosylation and deprotection steps were followed – in the case of the fucosylated ligand, the epoxide was opened with 2-chloroethan-1-ol, and the nucleophilic substitution with $NaN_3$ took place after the Zemplén-debenzoylation. (**Scheme 5**)




***Scheme 5.*** *Preparation of the dimethyl-ester substituted ligands. The acceptor is prepared from the enantiomerically pure diacid* ***2.1****, through Fischer-esterification, epoxidation with mCPBA and the epoxide-opening step. In the case of the fucosylated ligand* ***ISh049****, the alcoholysis of the epoxide was performed with 2-chloroethan-1-ol* ***2.4****, while for the mannosylated ligand* ***ISh002*** *the previously discussed 2-azidoethanol* ***2.5*** *was used. The acceptor was glycosylated with the properly functionalized mannose- and fucose-donors* ***2.10*** *and* ***2.21****, respectively, to form the glycosidic linkage. The debenzoylation step leads directly to the α-mannoside ligand* ***ISh002****, while β-fucosylated* ***ISh049*** *still required the displacement of the chlorine atom with* $NaN_3$ *during the final step.*

Eventually, our glycomimetic ligand library consisted of 50 members - 39 mannosylated and 11 fucosylated compounds[a] - which also included **1.21.a** (Man030) as **ISh001**, and pseudo-mannobioside **1.18a** as **ISh002**. Furthermore, appropriately functionalized α-D-mannose (**ISh044**) was also added to the library, as well as the aminoethanol-functionalized counterpart of the two best-characterized ligands: **ISh046** (Man030) and **ISh045** (pseudo-mannobioside) as amino-derivatives of **ISh001** and **ISh002**, respectively (reasons are discussed in Chapter II). (**Table 1**)

[a] Note: 7 ligands were synthesized by Norbert Varga (**ISh004**, **ISh012**, **ISh014**, **ISh015**, **ISh018**, **ISh019**, **ISh039**) and 2 by Alessio Maria Caramiello (**ISh016**, **ISh017**), and added to the library. The spectral data of each library ligand are added here for convenience.

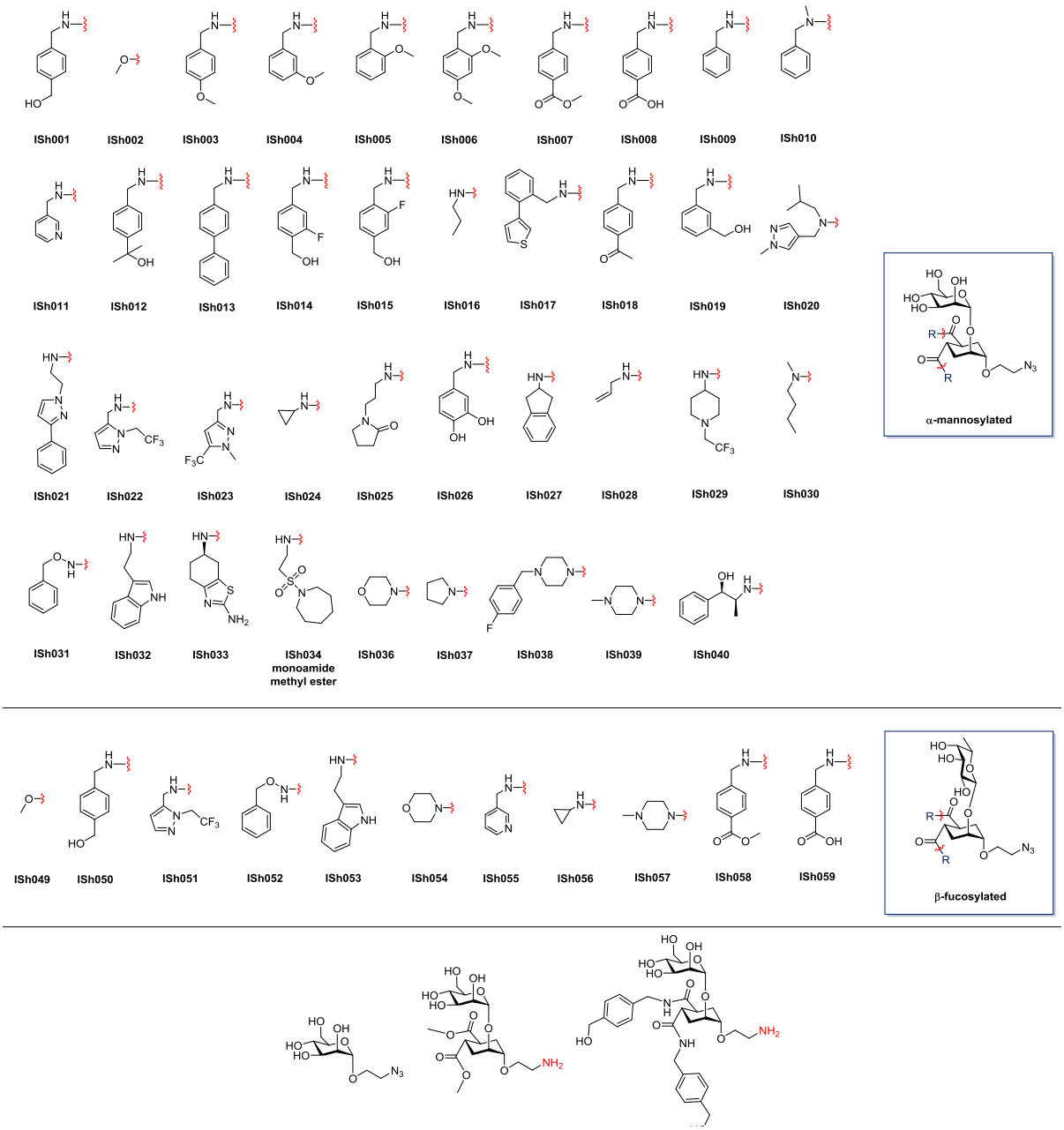



***Table 1.*** *Glycomimetic ligand library. 40 α-mannosylated and 11 β-fucosylated ligands have been prepared and diversified in bisamide positions. The compounds are terminated with an azidoethanol linker, and in two cases the aminoethanol-equipped ligands have also been included (**ISh001** – **ISh046** and **ISh002** – **ISh045**). Additionally, mannose has also been added to the library.*

Ligands **ISh035**, **ISh041**, **ISh042**, **ISh043, ISh047** and **ISh048** – although printed on the microarray chips and screened against C-type lectins - are not part of this discussion. The original numbering of ligands is retained here for convenience.

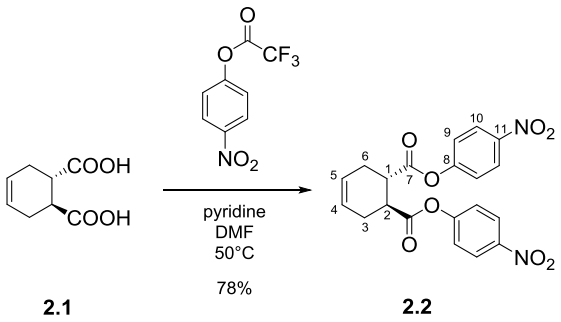
## 2.1.2 EXPERIMENTAL SECTION

### 2.1.2.1 General

Chemicals were purchased from commercial sources (Sigma-Aldrich or Acros Organics, and specific amines from Key Organics, Crea-Chim, Vitas M Labs, Life Chemicals, Alinda Chemicals, Chem Bridge or Enamine BB - source indicated at the bisamide derivatives characterization) and used without further purification, unless otherwise indicated. When anhydrous conditions were required, the reactions were performed under nitrogen or argon atmosphere, the latter indicated in the procedure. Anhydrous solvents were purchased from Sigma-Aldrich® with water content ≤ 0.005%. TEA, DCM, MeCN and MeOH were dried over calcium hydride, THF was dried over sodium/benzophenone and freshly distilled. Reactions were monitored by analytical thin-layer chromatography (TLC) performed on Silica Gel 60 F254 plates (Merck) or Silica Gel 60 RP-18F254s (Merck) plates with UV detection (254 nm and 365 nm) and/or staining with ammonium molybdate acid solution, potassium permanganate alkaline solution or ninhydrin. Flash column chromatography was performed according to the method of Still and co-workers[260] using Silica gel 60 (40-63 μm) (Merck). Automated flash chromatography was performed on a Biotage Isolera Prime system; unless otherwise indicated, Biotage SNAP KP-Sil cartridges were used. NMR experiments were recorded on a Bruker AVANCE-400 MHz instrument at 298 K. Chemical shifts (δ) are reported in ppm. The $^{1}H$ and $^{13}C$ NMR resonances of compounds were assigned with the assistance of COSY and HSQC experiments. Multiplicities are assigned as s (singlet), d (doublet), t (triplet), q (quartet), quint (quintet), m (multiplet). Mass spectra were recorded on a ThermoFischer LCQ apparatus (ESI ionization); high resolution mass spectra (HRMS) were acquired on a Waters SYNAPT G2 Si ESI QTof instrument, (Waters, Manchester, UK) and processed with Waters MassLynx software. The instrument was operated with a capillary voltage of 1.0 – 3.0 kV; sample infusion flow rate: 5 μl/min; source temperature: 100°C; sampling cone: 24.0000 V; desolvation temperature: 150°C. Desolvation gas flow was set to 500.0 L/hr; and the nebulizer gas flow to 7.0 Bar. Furthermore, MALDI-Tof mass measurements were performed on an Ultraflextreme III time-of-flight mass spectrometer, equipped with a pulsed Nd:YAG laser (λ 355 nm) and controlled by FlexControl 3.3 software (Bruker Daltonics, Bremen , Germany). The acquisitions (total of 2000-3000) were carried out in linear positive ion mode with pulsed ion extraction of 450 ns and laser frequency of 1000 Hz. Laser fluence was set up to 60-80% and the *m/z* range was chosen according to the mass of the sample. The accumulated spectra were processed using the Bruker software FlexAnalysis 3.3. Specific optical rotation values were measured using a Perkin-Elmer 241 polarimeter, at 589 nm in a 1 dm cell.

## 2.1.2.2 Synthesis of intermediates

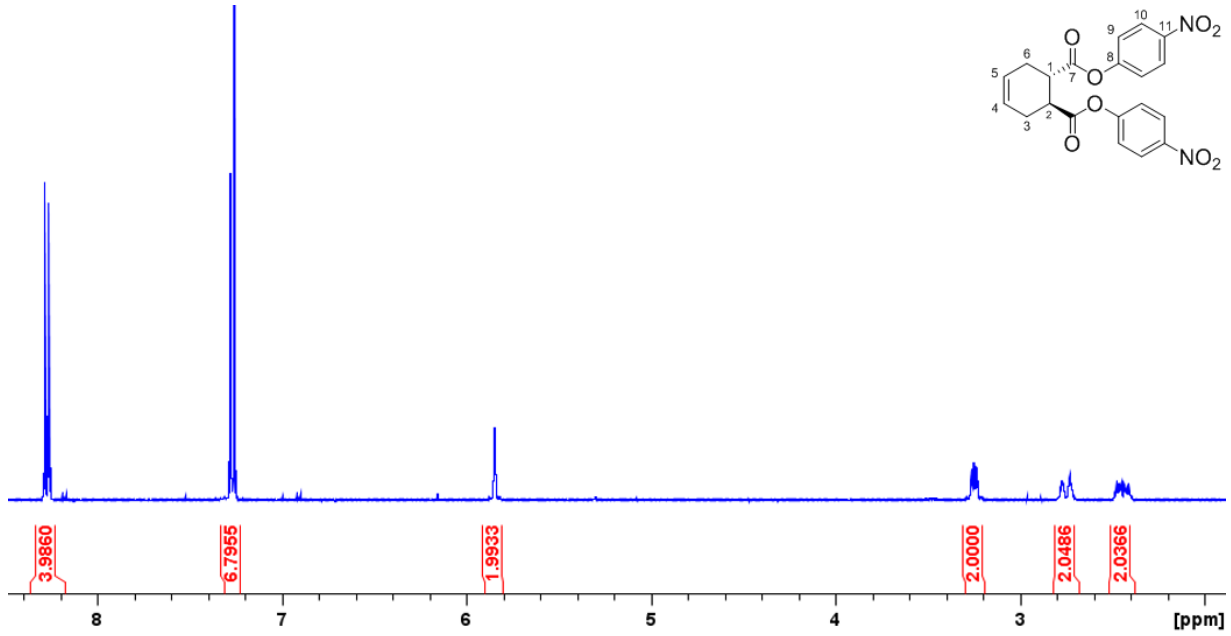
Bis(4-nitrophenyl) (1S,2S)-cyclohex-4-ene-1,2-dicarboxylate **2.2**



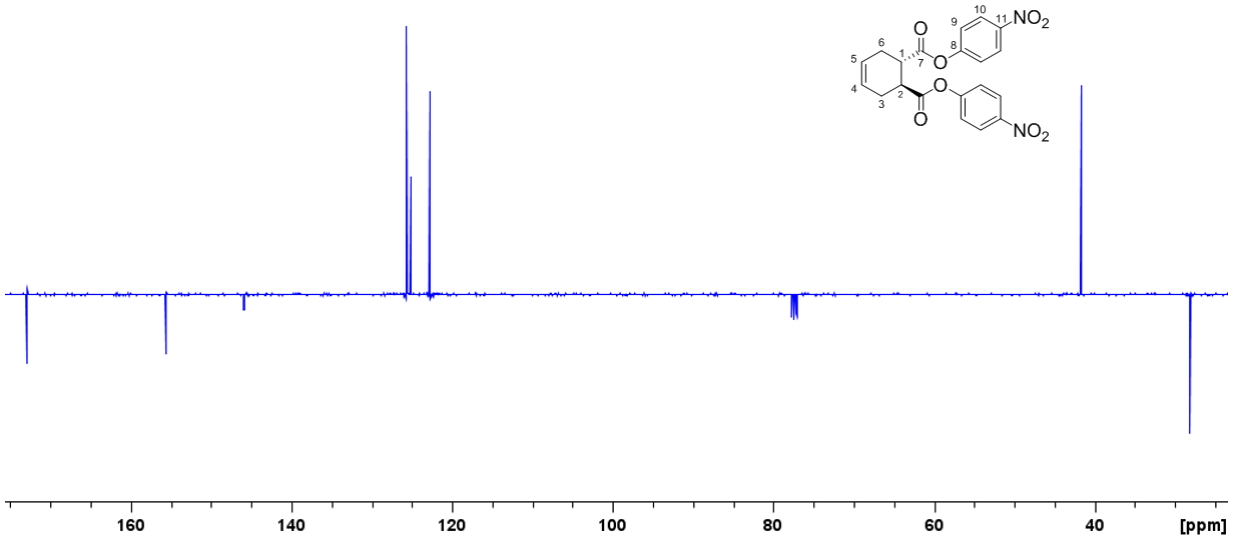

4.23 g (24.86 mmol, 1 equiv.) diacid **2.1** ($[\alpha]_D^{20}$: +153° (c = 1 in EtOH)) was dissolved in dry DMF under $N_2$ and 5.23 ml (64.63 mmol, 2.6 equiv.) pyridine was added dropwise to the solution. 14.03 g (59.66 mmol, 2.4 equiv.) 4-nitrophenyl trifluoroacetate was added to the mixture and the reaction was stirred overnight at 50°C. After completion ($R_f$ (product): 0.58 in toluene : EtOAc = 8 : 2 + 0.1% acetic acid), the reaction was diluted with 200 ml dichloromethane and washed twice with 100 ml 0.5M HCl, twice with 50 ml cold, saturated $NaHCO_3$ and twice with 50 ml water. The organic phase was dried over $Na_2SO_4$, filtered and concentrated in vacuo. The obtained dirty white crystals were washed with cooled diethyl ether and filtered to yield the pure product **2.2** as white crystals.

Yield: 78%
$[\alpha]_D^{20}$: +129.6° (c = 1 in $CHCl_3$)
$^1$H NMR (400 MHz, $CDCl_3$): δ = 8.28 – 8.22 (m, 4H; $H_{10}$), 7.27 – 7.22 (m, 4H; $H_9$), 5.83 (app d, J=2.8 Hz, 2H; $H_4$, $H_5$), 3.27 – 3.19 (m, 2H; $H_1$, $H_2$), 2.78 – 2.68 (m, 2H; $H_{3ps-eq}$, $H_{6ps-eq}$), 2.48 – 2.37 (m, 2H; $H_{3ps-ax}$, $H_{6ps-ax}$)
$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 172.8 ($C_7$); 155.4 ($C_{11}$); 145.7 ($C_8$); 125.5 ($C_{10}$); 124.9 ($C_4$, $C_5$); 122.5 ($C_9$); 41.5 ($C_1$, $C_2$); 28.0 ($C_3$, $C_6$)
MS (ESI): m/z calculated for $[C_{20}H_{16}N_2O_8Na]^+$: 435.08 $[M+Na]^+$; found: 435.33
M. p.: 171°C

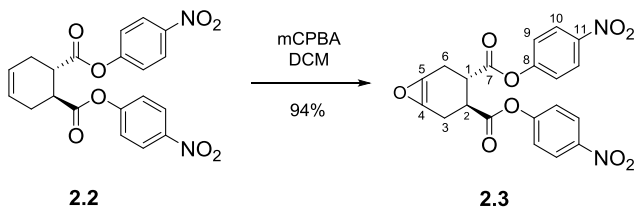
**$^{1}$H NMR spectrum of 2.2 in $CDCl_3$ (400 MHz)**




**$^{13}$C NMR spectrum of 2.2 in $CDCl_3$ (100 MHz)**




Bis(4-nitrophenyl) (3S,4S)-7-oxabicyclo[4.1.0]hept-1(6)-ene-3,4-dicarboxylate **2.3**




**2.2** **2.3**

To a solution of the PNP ester **2.2** (1367 mg, 3.32 mmol, 1 equiv.) in dry $CH_2Cl_2$ (11 ml) 891 mg mCPBA (3.98 mmol, 1.2 equiv.) was added under stirring. The reaction was stirred under nitrogen at room temperature for 16 h. The solvent was evaporated at reduced pressure and then the reaction mixture was diluted with EtOAc and washed three times with saturated $NaHCO_3$ and with water. The organic phase was dried over $Na_2SO_4$, filtered and the solvent was evaporated under reduced pressure to yield 94% of the pure product **2.3** as a white solid.

Yield: 94%

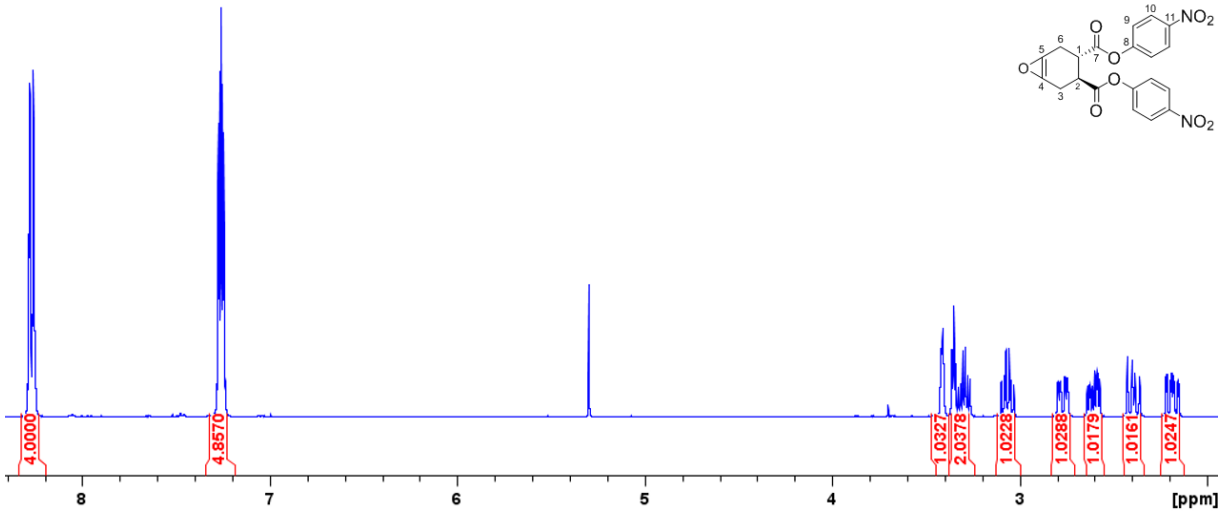
$[\alpha]_D^{20}$: +82.2° (c = 1.1 in $CHCl_3$)

$^1$H NMR (400 MHz, $CDCl_3$): δ = 8.28 – 8.22 (m, 4H; $H_{10}$), 7.27 – 7.21 (m, 4H; $H_9$), 3.43 – 3.36 (m, 1H; $H_4$ or $H_5$), 3.35 – 3.31 (m, 1H; $H_5$ or $H_4$), 3.32 – 3.24 (m, 1H; $H_1$ or $H_2$), 3.08 – 3.00 (m, 1H; $H_2$ or $H_1$), 2.79 – 2.71 (m, 1H, $H_{3eq}$ or $H_{6eq}$), 2.63 – 2.54 (m, 1H; $H_{6eq}$ or $H_{6eq}$), 2.42 – 2.33 (m, 1H; $H_{3ax}$ or $H_{6ax}$), 2.21 – 2.12 (m, 1H; $H_{6ax}$ or $H_{3ax}$)

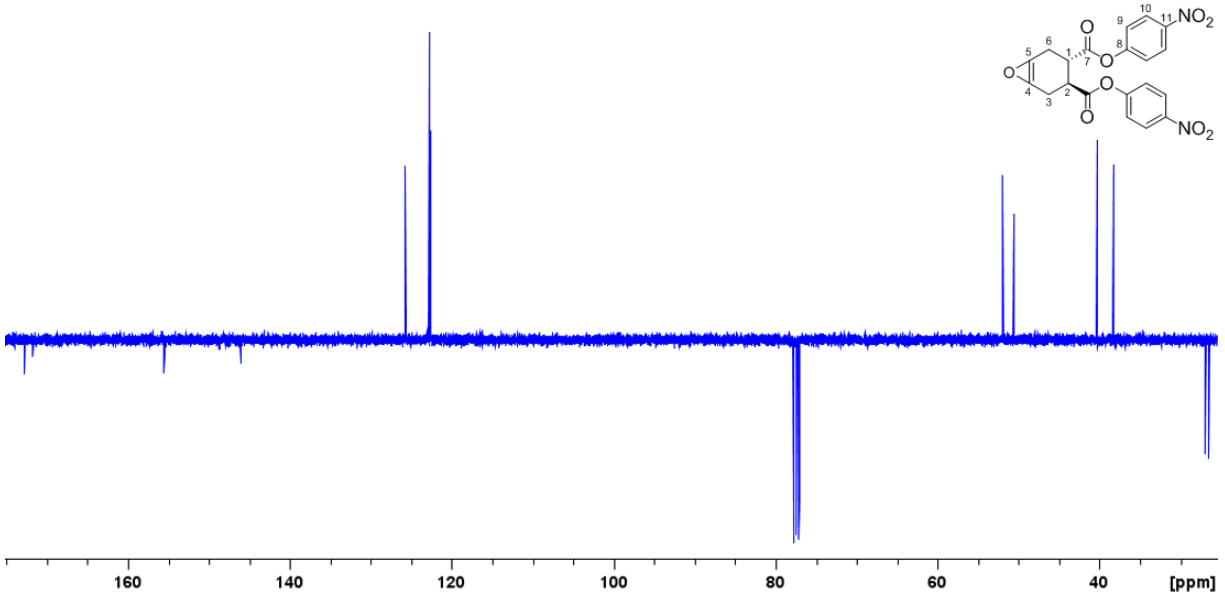
$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 172.6, 171.5 ($C_7$); 155.4, 155.3 ($C_8$); 149.9 ($C_{11}$); 125.5 ($C_{10}$); 122.5, 122.6 ($C_9$); 51.8, 50.3 ($C_4$, $C_5$); 40.1, 38.1 ($C_1$, $C_2$); 26.8, 26.3 ($C_3$, $C_6$)

HRMS: m/z calculated for $[C_{20}H_{16}N_2O_9Na]^+$: 451.07480; found: 451.07525

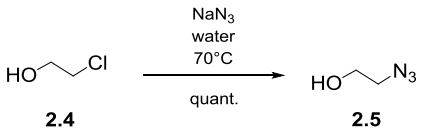
**$^1$H NMR spectrum of 2.2 in $CDCl_3$ (400 MHz)**




**$^{13}$C NMR spectrum of 2.3 in $CDCl_3$ (100 MHz)**



2-Azidoethan-1-ol[261] **2.5**



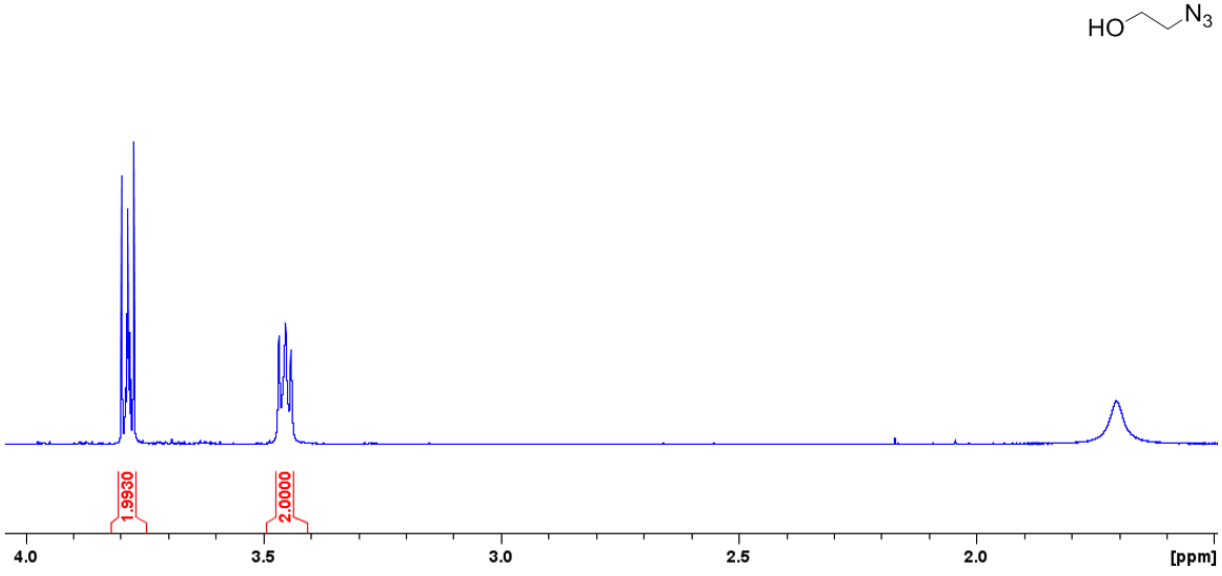

2.68 g (41.22 mmol, 1.18 equiv.) $NaN_3$ was dissolved in 6 ml water. To this mixture 2.35 ml (34.9 mmol, 1 equiv.) 2-chlorothan-1-ol was added and the reaction was stirred for 1 hour at RT, then at 70°C for 3 days. After completion, the reaction was cooled to RT, saturated with NaCl and extracted three times with DCM. The combined organic layers were dried over $Na_2SO_4$, filtered and concentrated *in vacuo* to ca. 20 ml (**2.5** in DCM solution).

Destruction of residual $NaN_3$ solution with less than 5% azide-content:
The operation must be carried out under a working fumehood. The aqueous $NaN_3$ solution is loaded into a three-necked flask, equipped with a dropping funnel, gas outlet and stirbar. 20% aqueous solution of sodium nitrite containing 1.5 g (ca. 40% excess) of sodium nitrite per gram of sodium azide is added with stirring. Then, 20% aqueous solution of sulfuric acid is added dropwise via the dropping funnel, until the reaction mixture is acidic. The reaction is accompanied by intense gas formation - these gases include toxic nitric oxide and should not be inhaled.

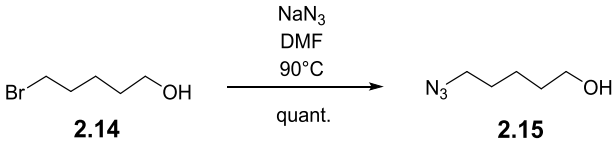
$$2NaNO_2 + H_2SO_4 \rightarrow 2HNO_2 + Na_2SO_4$$
$$2NaN_3 + HNO_2 \rightarrow 3N_2 + 2NO + 2NaOH$$

Yield: quant.
$^1$H NMR (400 MHz, $CDCl_3$): δ = 3.77 (t, 2H, HO-C*H*$_2$-, *J* = 5 Hz), 3.46 (t, 2H, -C*H*$_2$-$N_3$, *J* = 5 Hz), 1.71 (br s, 1H, O*H*)
$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 60.7 (HO-$CH_2$-); 52.9 (-$CH_2$-$N_3$)

**$^1$H NMR spectrum of 2.5 in $CDCl_3$ (400 MHz)**



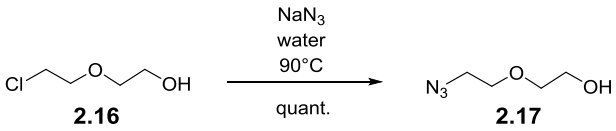
5-azidopentan-1-ol[262] **2.15**




200 mg (1.19 mmol, 1 equiv.) 5-bromopentan-1-ol (**2.14**) and 234 mg (3.59 mmol, 3 equiv.) $NaN_3$ were dissolved in 2 ml dry DMF. The reaction was warmed to 90°C and stirred overnight. After completion, the mixture was cooled to RT and concentrated, then diluted with 10 ml EtOAc and washed three times with water and twice with brine. The organic phase was dried over $Na_2SO_4$, filtered and concentrated.

Yield: quant.
$^1$H NMR (400 MHz, $CDCl_3$): δ = 3.66 (t, 2H, HO-C*H*$_2$-, *J* = 6.5 Hz), 3.29 (t, 2H, -C*H*$_2$-$N_3$, *J* = 6.8 Hz), 1.67-1.41 (m, 6H, HO-$CH_2$-C*H*$_2$- C*H*$_2$- C*H*$_2$- $CH_2$-$N_3$)
$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 61.5 (HO-$CH_2$-); 50.8 (-$CH_2$-$N_3$); 31.5, 28.0, 22.4 (HO-$CH_2$-C*H*$_2$- $CH_2$- C*H*$_2$-$CH_2$-$N_3$)

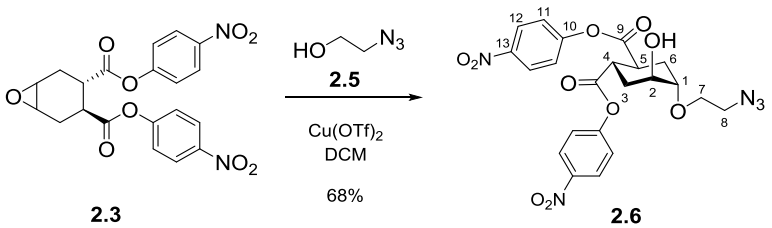
2-(2-azidoethoxy)ethanol[263] **2.17**




616 µl (5.84 mmol, 1 equiv.) 2-(2-chloroethoxy)ethanol (**2.16**) and 759 mg (11.67 mmol, 2 equiv.) $NaN_3$ were dissolved in 2.5 ml water. The reaction was stirred at 90°C overnight with a condenser. The mixture was cooled to RT and extracted three times with DCM. The combined organic layers were dried over $MgSO_4$, filtered and concentrated.

Yield: quant.
$^1$H NMR (400 MHz, $CDCl_3$): δ = 3.74 (t, 2H, HO-C*H*$_2$-, *J* = 4.5 Hz), 3.67 (t, 2H, HO-$CH_2$-C*H*$_2$-O, *J* = 5 Hz), 3.61 (t, 2H, O-C*H*$_2$-$CH_2$-$N_3$, *J* = 4.6 Hz), 3.41 (t, 2H, -C*H*$_2$-$N_3$, *J* = 5 Hz)
$^{13}$C NMR (100 MHz, $D_2O$): δ = 72.7 (HO-$CH_2$-); 70.3 (HO-$CH_2$-$CH_2$-O); 62.1 (O-$CH_2$-$CH_2$-$N_3$); 51.0 (-$CH_2$-$N_3$)

1,2-Cyclohexanedicarboxylic acid, (1S,2S,4S,5S)-5-(2-azidoethoxy)-4-hydroxy-1,2-bis-p-nitrophenyl ester **2.6**

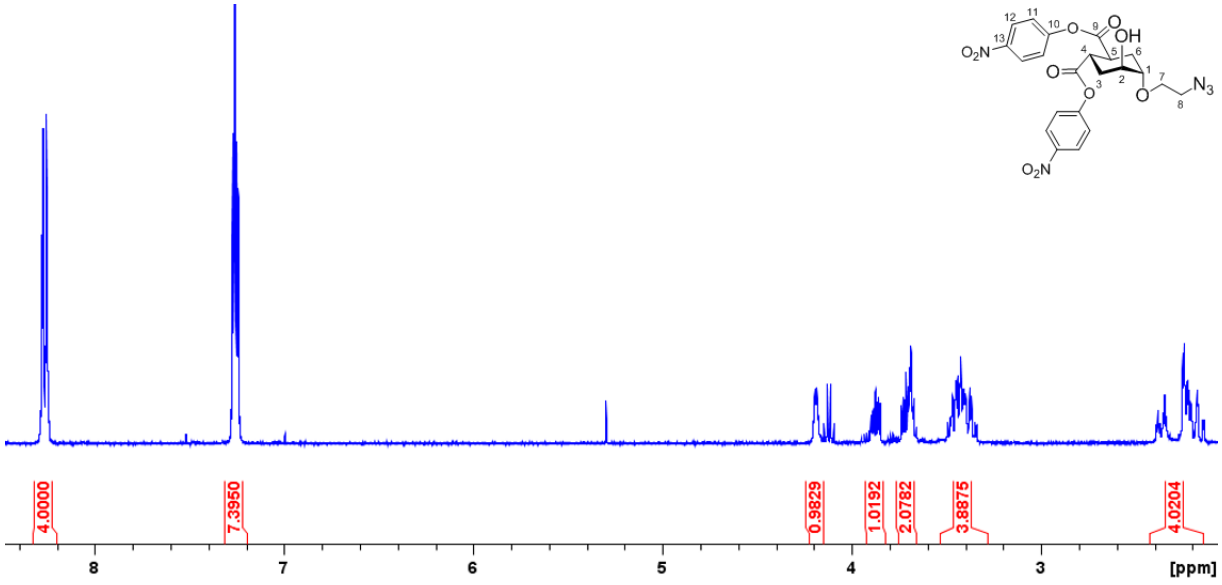



1.1 g (2.57 mmol, 1 equiv.) epoxide **2.3** and 186 mg $Cu(OTf)_2$ (0.51 mmol, 0.2 equiv.) were added to a solution of 2-azidoethan-1-ol **2.5** (ca. 38.5 mmol, 15 equiv) in dichloromethane (20 ml) and the mixture was stirred under

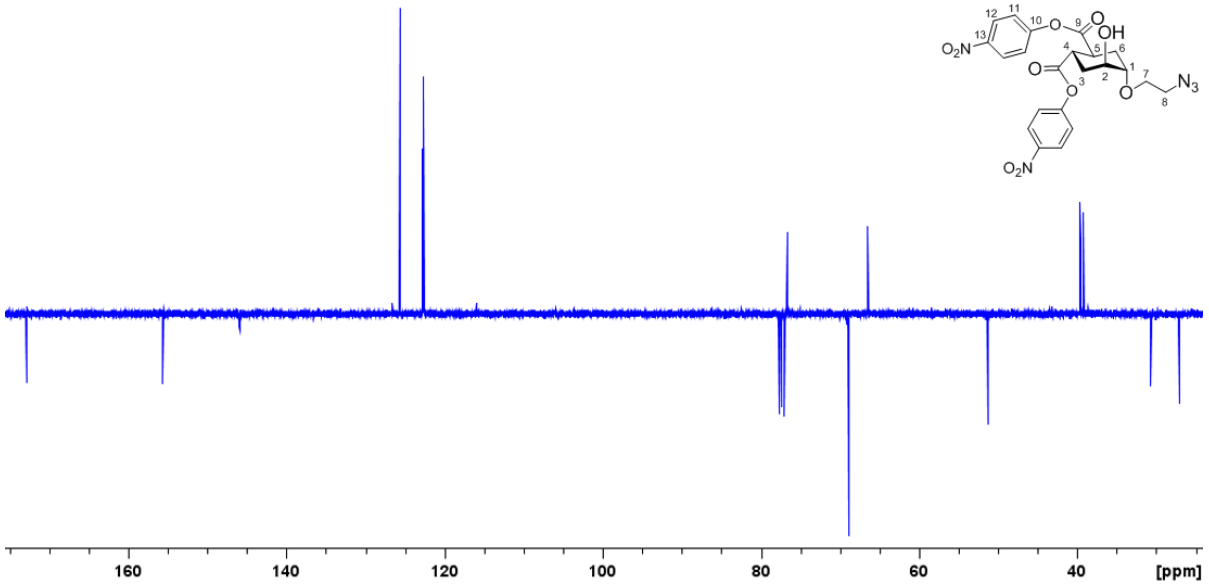
nitrogen at room temperature. After completion (16 h), the solvents were removed under reduced pressure and the crude was purified by flash chromatography (hexane with gradient of ethyl acetate from 20 to 50%) to yield 68% of pure product 2.6 as a colorless wax.

Yield: 68%
$[\alpha]_D^{20}$: +33.8° (c = 1.1 in $CHCl_3$)
$^1$H NMR (400 MHz, $CDCl_3$): δ = 8.27 – 8.17 (m, 4H; $H_{12}$), 7.31 – 7.23 (m, 4H; $H_{11}$), 4.19 – 4.14 (m, 1H; $H_2$), 3.88 – 3.82 (m, 1H; $H_{7a}$), 3.73 – 3.63 (m, 2H; $H_{7b}$, $H_1$), 3.49 – 3.29 (m, 4H; $H_4$, $H_5$, $H_{8a,b}$), 2.40 – 2.10 (m, 4H; $H_3$, $H_6$)
$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 172.7, 172.6 ($C_9$); 155.5 ($C_{10}$); 145.8, 145.7 ($C_{13}$); 125.5 ($C_{12}$); 122.6, 122.6 ($C_{11}$); 76.5 ($C_1$); 68.7 ($C_7$); 66.3 ($C_2$); 51.1 ($C_8$); 39.4, 39.1 ($C_4$, $C_5$); 30.5, 27.0 ($C_3$, $C_6$)
MS (ESI): m/z calculated for $[C_{22}H_{21}N_3O_{10}Na]^+$: 512.4 $[M\text{-}N_2\text{+}Na]^+$; found: 512.0.

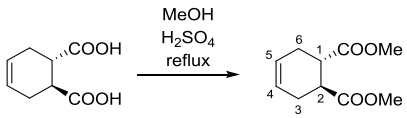
**$^1$H NMR spectrum of 2.6 in $CDCl_3$ (400 MHz)**




**$^{13}$C NMR spectrum of 2.6 in $CDCl_3$ (100 MHz)**



(1S,2S)-dimethyl cyclohex-4-ene-1,2-dicarboxylate **2.24**

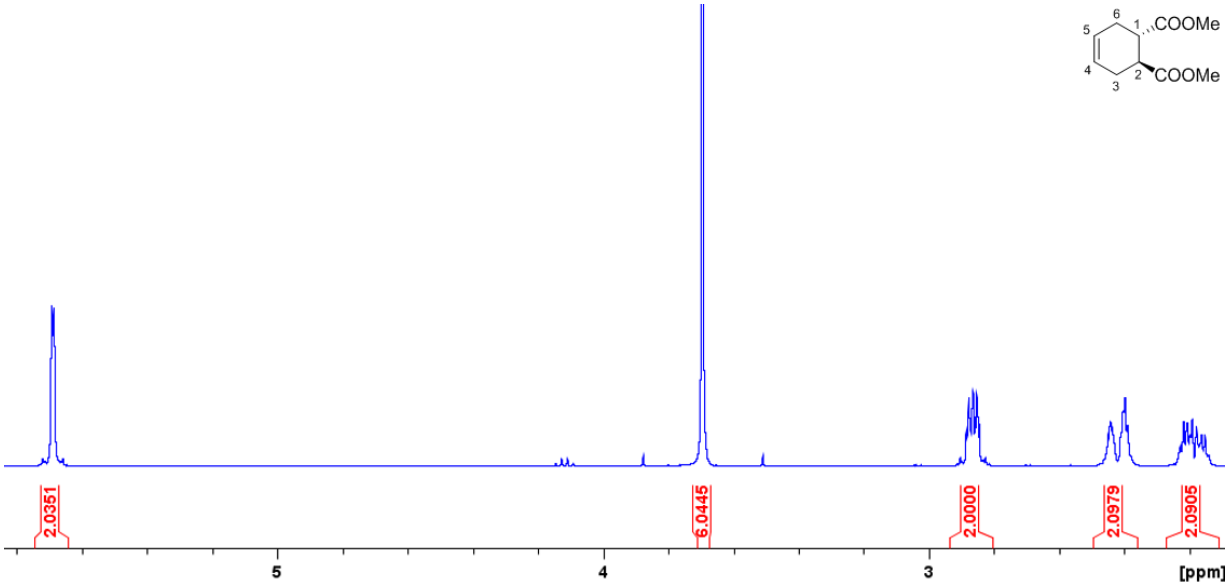



**2.1** **2.24**

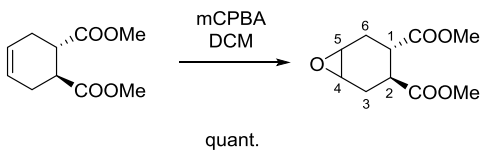
To a solution of diacid **2.1** (500 mg, 2.94 mmol, 1 equiv.) in 7.35 ml dry MeOH (0.4 M), 31 µl sulfuric acid (0.59 mmol, 0.2 equiv.) was added. The mixture was refluxed overnight under nitrogen atmosphere. Upon completion ($R_f$ = 0.33 in Hex : EtOAc = 1 : 1), the mixture was diluted with EtOAc and washed with $NaHCO_3$ until the pH was neutral, then twice with water. The organic phase was dried over anhydrous $Na_2SO_4$, filtered and the solvent was evaporated at reduced pressure to afford the pure product as a colourless oil.

Yield: 93%
$[\alpha]_D^{20}$: +115.1° (c = 1.31 in $CHCl_3$)
$^1$H NMR (400 MHz, $CDCl_3$): δ = 5.65 (m, 2H, $H_4$, $H_5$), 3.7 (s, 6H, OMe), 2.85 (m, 2H, $H_1$, $H_2$), 2.0 - 2.5 (m, 4H, $H_{3ax}$, $H_{3eq}$, $H_{6ax}$, $H_{6eq}$).
$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 174.8 (C=O); 124.7 ($C_4$, $C_5$); 51.5 (COOMe); 41.0 ($C_1$, $C_2$); 24.5 ($C_3$, $C_6$)

**$^1$H NMR spectrum of 2.24 in $CDCl_3$ (400 MHz)**




Synthesis of (3S,4S)-dimethyl 7-oxabicyclo[4.1.0]heptane-3,4-dicarboxylate **2.25**



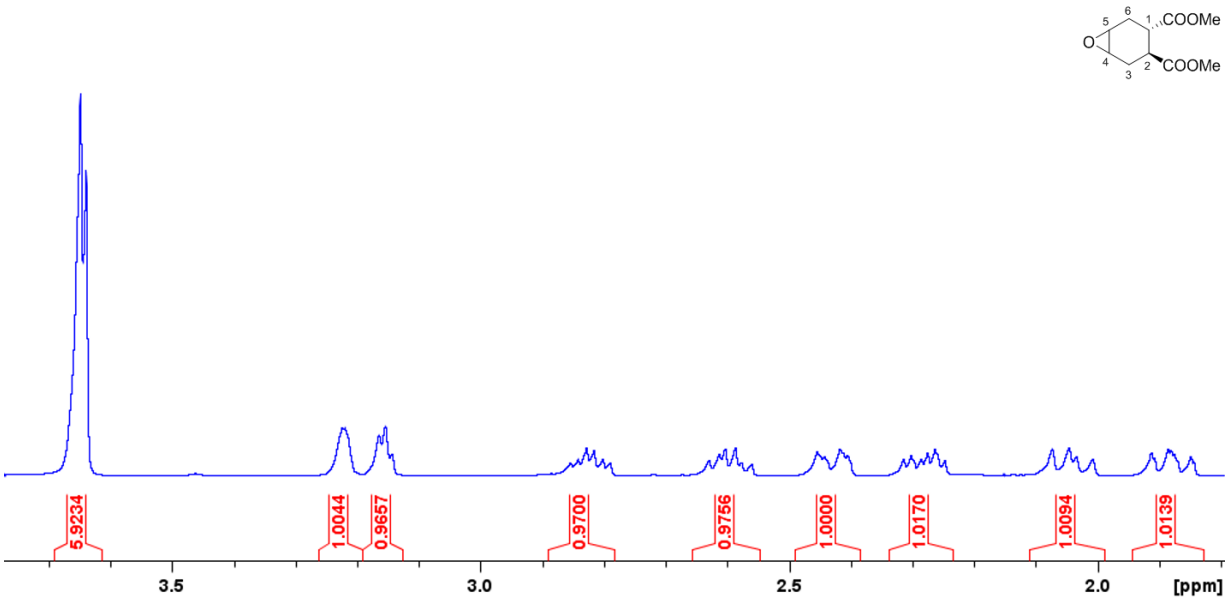

**2.24** **2.25**

Compound **2.24** (542 mg, 2.74 mmol, 1 equiv.) was dissolved in 6.8 ml dry DCM and 708.4 mg mCPBA (4.11 mmol, 1.5 equiv.) was added. The reaction was stirred at room temperature for 3 hours under nitrogen

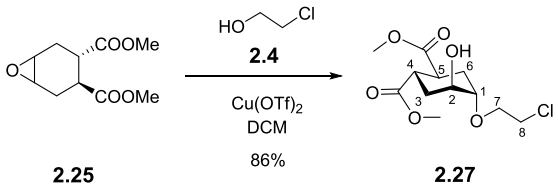
atmosphere. Upon completion ($R_f$ = 0.5 in Hex : EtOAc = 1 : 1), the mixture was diluted with DCM and washed twice with sat. $NaHCO_3$ and with water. The organic phase was dried over anhydrous $Na_2SO_4$, filtered and the solvent was removed at reduced pressure. The crude was purified by automated flash chromatography to afford the pure product **2.25**.

Yield: quant.
$[\alpha]_D^{20}$: +71.6° (c = 1.275 in $CHCl_3$)
$^1$H NMR (400 MHz, $CDCl_3$): δ = 3.66 - 3.65 (2s, 6H, OMe), 3.23 (m, 1H, $H_4$), 3.16 (m, 1H, $H_5$), 2.84 (dt, 1H, $H_1$, $J_{1\text{-}6eq}$ = 4.8 Hz, $J_{1\text{-}2}$ = $J_{1\text{-}6ax}$ = 10.7 Hz), 2.62 (dt, 1H, $H_2$, $J_{2\text{-}3eq}$ = 4.8 Hz, $J_{2\text{-}3ax}$ = 10.7 Hz), 2.42 (ddd, 1H. $H_{6eq}$, $J_{6eq\text{-}5}$ = 1.8 Hz, $J_{6eq\text{-}6ax}$ = 15 Hz,), 2.27 (ddd, 1H, H3eq, $J_{3eq\text{-}4}$ = 6.7 Hz, $J_{3eq\text{-}3ax}$ = 15.5 Hz), 2.06 - 2.02 (m, 1H, $H_{3ax}$), 1.90 (ddd, 1H, $H_{6ax}$, $J_{6ax\text{-}5}$ = 2.2 Hz).

**$^1$H NMR spectrum of 2.25 in $CDCl_3$ (400 MHz)**




(1S,2S,4S,5S)-1-(2-chloroethyl)-2-hydroxycyclohexen-4,5-carboxylic acid bis-methylester **2.27**



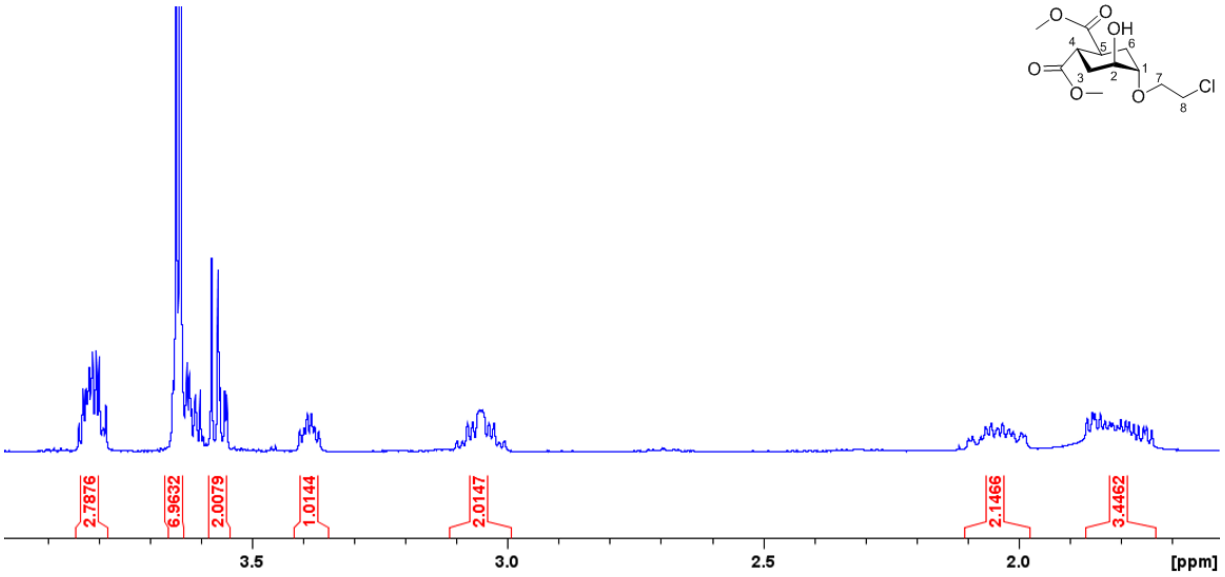

**2.25** **2.27**

To a solution of epoxide **2.25** (572.8 mg, 2.67 mmol, 1 equiv.) in 2-chloroethanol (5.4 ml, 80.2 mmol, 30 equiv.) $Cu(OTf)_2$ (96.7 mg, 0.27 mmol, 0.1 equiv.) was added. The light blue mixture was stirred at room temperature for one day under nitrogen atmosphere. Upon completion ($R_f$ = 0.33 in Hex : EtOAc = 6 : 4), the reaction mixture was diluted with EtOAc and washed once with $NH_3$/$NH_4Cl$ solution and twice with water. The organic phase was dried over anhydrous $Na_2SO_4$, filtered and the solvent was evaporated at reduced pressure to yield the pure product **2.27** as a colourless oil.

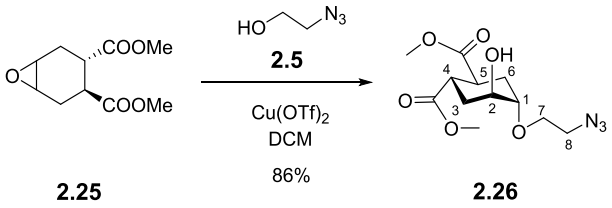
Yield: 86%

$^{1}$H NMR (400 MHz, $CDCl_3$): δ = 3.84 – 3.75 (m, 2H, $H_2$, $H_{7a}$), 3.67 - 3.58 (m, 7H, OMe, $H_{7a}$), 3.58 - 3.52 (m, 2H, $H_{8a,b}$), 3.38 (dd, 1H, $H_1$, $J$ = 5.6 Hz, $J$ = 3.2 Hz,), 3.10 - 2.95 (m, 2H, $H_4$, $H_5$), 2.07 - 1.94 (m, 2H, $H_{3ax}$, $H_{6ax}$), 1.88 - 1.72 (m, 3H, $H_{3eq}$, $H_{6eq}$, OH)
$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 174.7 (C=O); 77.7 ($C_1$); 69.3 ($C_7$); 67.6 ($C_2$); 52.0 (OMe); 43.2 ($C_8$); 39.5, 39.2 ($C_4$, $C_5$); 30.5 ($C_3$); 27.2 ($C_6$)
MS (ESI): m/z calculated for $[C_{12}H_{19}ClO_6Na]^+$: 317.7 $[M+Na]^+$; found: 317.1

**$^{1}$H NMR spectrum of 2.24 in $CDCl_3$ (400 MHz)**




(1S,2S,4S,5S)-1-(2-azidoethyl)-2-hydroxycyclohexen-4,5-carboxylic acid bis-methylester **2.26**



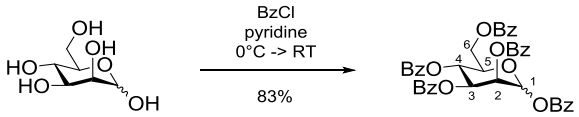

**2.25** **2.26**

To a solution of epoxide **2.25** (113.3 mg, 0.53 mmol, 1 equiv.) in 1 ml DCM 2-azidoethanol (10.6 mmol, 20 equiv. in 5 ml DCM) and $Cu(OTf)_2$ (96.7 mg, 0.27 mmol, 0.1 equiv.) was added. The light blue mixture was stirred at room temperature for one day under nitrogen atmosphere. Upon completion ($R_f$ = 0.33 in Hex : EtOAc = 6 : 4), the reaction mixture was diluted with EtOAc and washed once with $NH_3/NH_4Cl$ solution and twice with water. The organic phase was dried over anhydrous $Na_2SO_4$, filtered and the solvent was evaporated at reduced pressure to yield the pure product **2.26** as a colourless oil.

Yield: 83%
$[\alpha]_D^{20}$: +19.2° (c = 1.15 in $CHCl_3$)
$^{1}$H NMR (400 MHz, $CDCl_3$): δ = 3.75 (br s, 2H, $H_2$), 3.48 (s, 3H, OMe), 3.47 (s, 3H, OMe), 3.50 – 3.30 (m, 2H, $H_4$, $H_5$), 3.28 – 3.20 (m, 2H, $H_1$, $H_{7b}$), 3.03 (ddd, 1H, $H_{7a}$, $J_{gem}$ = 10.4 Hz, $H_{7a\text{-}8a}$ = 3.6 Hz, $H_{7a\text{-}8b}$ = 7.6 Hz), 2.78 (m, 1H, $H_{8b}$), 2.68 (m, 1H, $H_{8a}$), 2.18 (dt, 1H, $H_{3ax}$, $J_{ax\text{-}2}$ = 2.8 Hz, $J_{gem}$ = $J_{ax\text{-}4}$ = 13.6 Hz), 2.10 - 1.95 (m, 2H, $H_6$), 1.97 (dt, 1H, $H_{3eq}$, $J_{gem}$ = 13.6 Hz, $J_{3eq\text{-}4}$ = $J_{3eq\text{-}2}$ = 4.4 Hz)
$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 174.9 (C=O); 77.7 ($C_1$); 66.7 ($C_7$); 66.6 ($C_2$); 55.9 ($C_8$); 51.2 (OMe); 51.1 (OMe); 39.4 ($C_4$); 39.0 ($C_5$); 30.8 ($C_3$); 26.5 ($C_6$)

HRMS (ESI): m/z calculated for $[C_{12}H_{19}O_6N_3Na]^+$: 324.11661 $[M+Na]^+$; found: 324.11624

1,2,3,4,6-penta-*O*-benzoyl-mannopyranose **2.8** – Crude prepared by Roberto Russo[254]

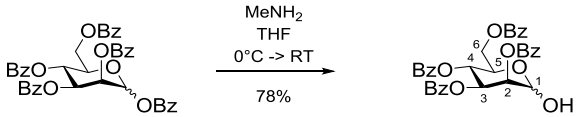



**2.7** **2.8**

10 g (55.6 mmol, 1 equiv.) D-mannose was dissolved in 114 ml dry pyridine at 0°C and 62.4 g (444.4 mmol, 8 equiv.) BzCl was added dropwise to the solution. The reaction was stirred at RT, under $N_2$ atmosphere overnight. After completion ($R_f$ = 0.35 in Hex : EtOAc = 7 : 3), the reaction was cooled again to 0°C and 100 ml water was added dropwise to the stirred reaction. The mixture was diluted with additional 100 ml water and extracted with 200 ml EtOAc. The organic phase was washed twice with 200 ml 1M HCl solution, dried over $Na_2SO_4$, filtered and concentrated *in vacuo*.

Yield: 83%, α : β = 2.6 : 1
$^1$H NMR (400 MHz, $CDCl_3$): α-anomer: δ = 8.24 – 7.22 (m, 25H, BzO), 6.44 (s, 1H, $H_1$), 6.17 (t, 1H, $H_4$), 6.12 (d, 1H, $H_2$), 5.80 (dd, 1H, $H_3$), 4.77 (dd, 1H, $H_{6b}$), 4.58 (dd, 1H, $H_{6a}$), 4.38 (dt, 1H, $H_5$)
β-anomer: δ = 8.24 – 7.22 (m, 25H, BzO), 6.40 (s, 1H, $H_1$), 6.28 (t, 1H, $H_4$), 5.92 (d, 1H, $H_2$), 5.80 (dd, 1H, $H_3$), 4.73 (dd, 1H, $H_{6b}$), 4.54 (dt, 1H, $H_5$), 4.48 (dd, 1H, $H_{6a}$)
MS (ESI): m/z calculated for $[C_{41}H_{33}O_{11}Na]^+$: 724.19 $[M+H+Na]^+$; found: 724.09

2,3,4,6-tetra-*O*-benzoyl-D-mannopyranose **2.9**

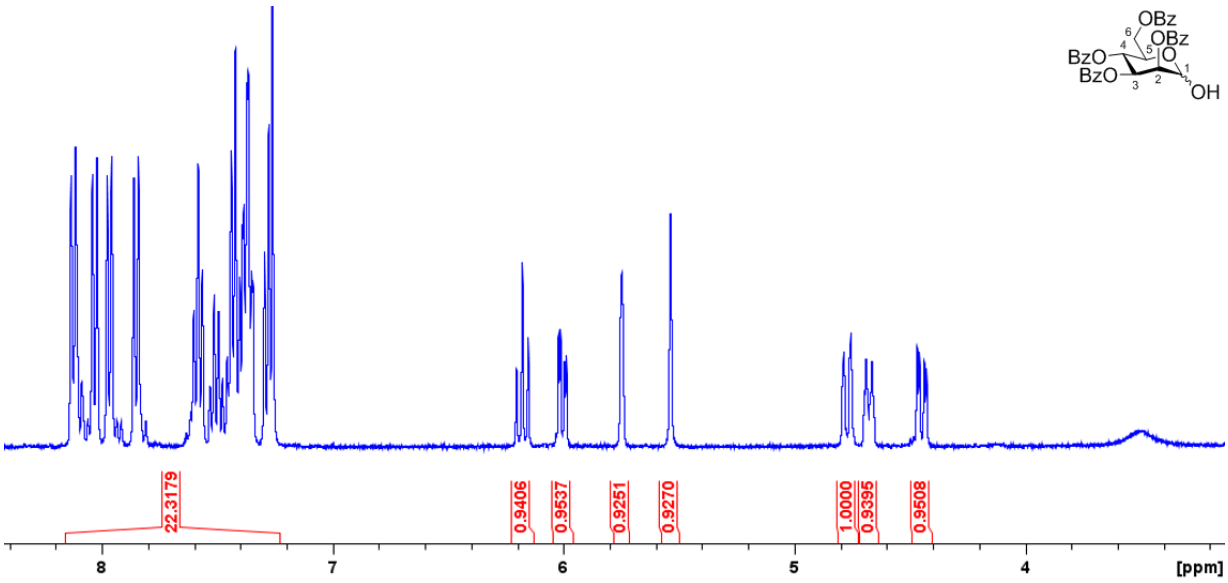



**2.8** **2.9**

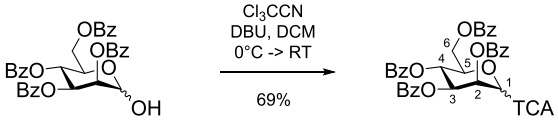
5 g (7.14 mmol, 1 equiv.) 1,2,3,4,6-penta-*O*-benzoyl-mannopyranose **2.8** was dissolved in 4 ml dry THF and cooled to 0°C. 10.7 ml 2M $MeNH_2$ in THF (21.4 mmol, 3 equiv.) was added to the solution and the reaction was stirred at RT for 7 hours. After completion, the mixture was concentrated and the crude was purified by flash chromatography ($R_f$: 0.44 in Hex : EtOAc = 6 : 4).

Yield: 78%
$^1$H NMR (400 MHz, $CDCl_3$): δ = 8.20 – 7.40 (m, 20H, BzO), 6.19 (t, 1H, $H_4$, $J$ = 10 Hz), 6.04 (dd, 1H, $H_3$, $J$ = 11 Hz, $J$ = 3.2 Hz), 5.76 (m, 1H, $H_2$), 5.56 (m, 1H, $H_1$), 4.79 (dd, 1H, $H_{6b}$, $J$ = 12.2 Hz, 2.3 Hz), 4.71 (dt, 1H, $H_5$, $J$ = 9.7 Hz, $J$ = 3.0 Hz), 4.50 (dd, 1H, $H_{6a}$, $J$ = 12.2 Hz, $J$ = 3.5 Hz)
MS (ESI): m/z calculated for $[C_{34}H_{28}O_{10}Na]^+$: 619.16 $[M+Na]^+$; found: 619.60

**$^1$H NMR spectrum of 2.9 in $CDCl_3$ (400 MHz)**




2,3,4,6-tetra-*O*-benzoyl-D-Mannose-1-trichloroacetimidate **2.10**

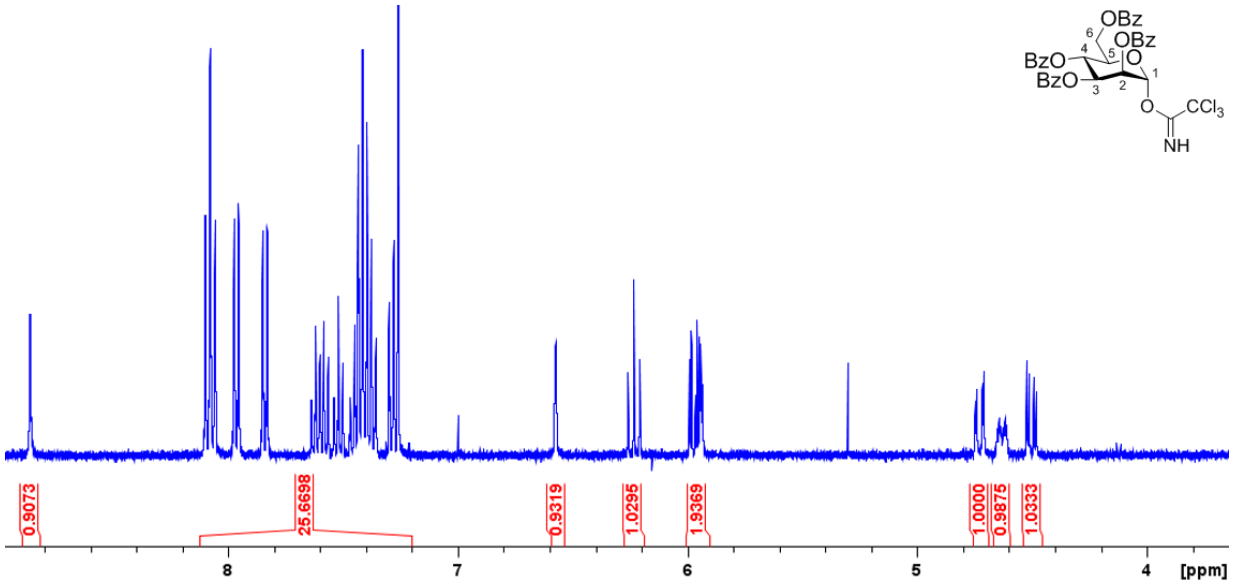



**2.9** **2.10**

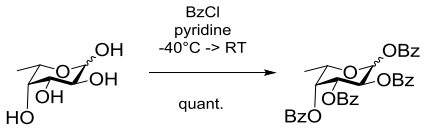
2.5 g (4.19 mmol, 1 equiv.) 2,3,4,6-tetra-*O*-benzoyl-D-mannopyranose (**2.9**) was dissolved in 10 ml DCM, cooled to 0°C and 1.2 ml (20.9 mmol, 5 equiv.) trichloroacetonitrile was added to the solution. Then, 125 µl (0.84 mmol, 0.2 equiv.) DBU was added and the reaction was stirred at RT under $N_2$ overnight. After completion, the mixture was concentrated and the crude was purified by flash chromatography ($R_f$: 0.38 in Hex : EtOAc = 8 : 2) to yield to pure product as a white foam.

Yield: 69% (Exclusively α-product)
$^1$H NMR (400 MHz, $CDCl_3$): δ = 8.9 (s, 1H, NH), 8.00 – 7.20 (m, 20H, BzO), 6.57 (d, 1H, $H_1$, *J* = 1.7 Hz), 6.23 (t, 1H, $H_4$, *J* = 10 Hz), 6.00 – 5.94 (m, 1H, $H_2$, $H_3$), 4.74 (dd, 1H, $H_{6b}$, *J* = 12.4 Hz, *J* = 2.5 Hz), 4.66 (m, 1H, $H_5$), 4.51 (dd, 1H, $H_{6a}$, *J* = 12.3 Hz, *J* = 4.14 Hz)
MS (ESI): m/z calculated for $[C_{36}H_{28}O_{10}Na]^+$: 762.07 $[M+Na]^+$; found: 762.90

**$^1$H NMR spectrum of 2.10 in $CDCl_3$ (400 MHz)**




1,2,3,4-tetra-O-benzoyl-L-fucopyranose **2.19**

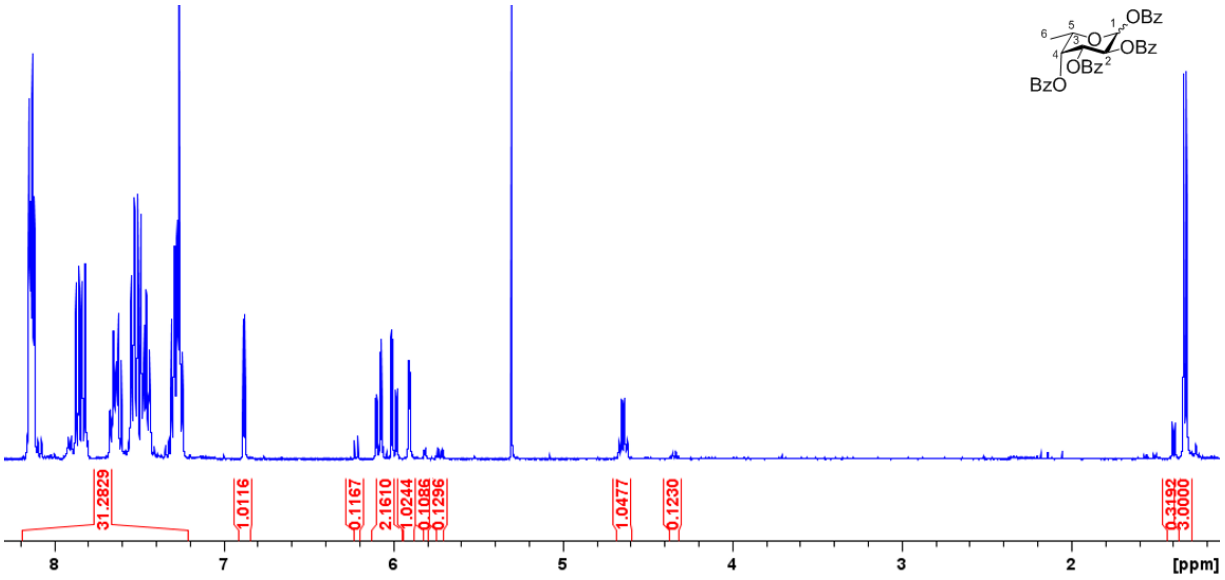



**2.18** **2.19**

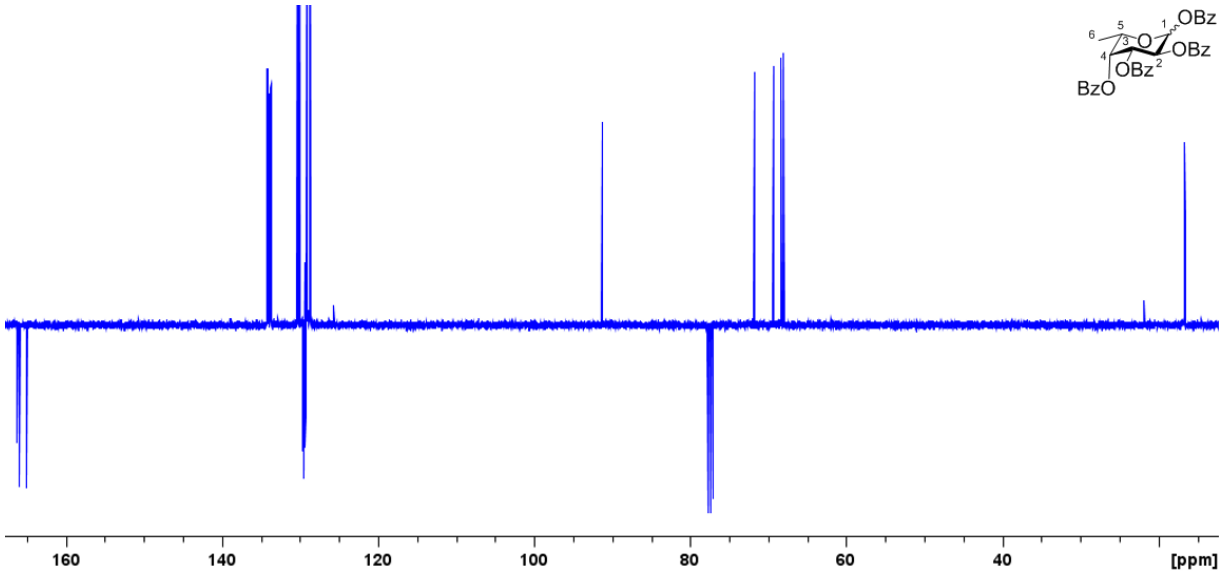
200 mg (1.22 mmol, 1 equiv.) L-(-)-fucose (**2.18**) was dissolved at -40°C in 1 ml pyridine under $N_2$ atmosphere and 636 µl BzCl (5.48 mmol, 4.5 equiv.) was added dropwise to the solution. The reaction was left to warm to room temperature, as soon as the starting unprotected sugar was not detected by TLC anymore. The reaction was stirred until there was only one major spot visible on the TLC plate ($R_f$ = 0.24 in toluene : EtOAc = 97 : 3). Upon completion, the reaction mixture was diluted with water and extracted with DCM. The joint organic phases were dried over $Na_2SO_4$, filtered and concentrated *in vacuo*, to yield the product as a white foam. At low furanoside ratio no further purification was needed.

Yield: quant (α : β = 9 : 1; ca. 1% α and β-fucofuranoside-form).
$^1$H NMR (400 MHz, $CDCl_3$): α-anomer: δ = 8.17 – 7.21 (m, 20H, OBz), 6.87 (d, 1H, $H_1$, $J_{1-2}$ = 3.7 Hz), 6.08 (dd, 1H, $H_3$, $J_{3-4}$ = 3.2 Hz, $J_{3-2}$ = 10.8 Hz), 5.99 (dd, 1H, $H_2$, $J_{2-3}$ = 10.8 Hz, $J_{2-1}$ = 3.7 Hz), 5.90 (dd, 1H, $H_4$, $J_{4-5}$ = 1.1 Hz, $J_{4-3}$ = 3.2 Hz), 4.64 (dd, 1H, $H_5$, $J_{5-6}$ = 6.6 Hz, $J$ = 12.9), 1.32 (d, 3H, $H_6$, $J$ = 6.4 Hz)
β-anomer: δ = 8.17 – 7.21 (m, 20H, OBz), 6.21 (d, 1H, $H_1$, $J_{1-2}$ = 8.3 Hz), 6.06 (m, 1H, $H_2$), 5.81 (dd, 1H, $H_4$, $J_{4-5}$ = 1.0 Hz, $J_{4-3}$ = 3.4 Hz), 5.72 (m, 1H, $H_3$), 4.34 (dd, 1H, $H_5$, $J_{5-6}$ = 5.7 Hz, $J$ = 12.5 Hz), 1.39 (d, 3H, $H_6$, $J$ = 6.4 Hz)
$^{13}$C NMR (100 MHz, $CDCl_3$): α-anomer: δ = 166.3, 166.1, 165.9, 165.0 (C=O); 134.1, 133.9, 133.7, 133.6, 130.3, 130.2, 130.0 (BzO); 129.6, 129.5, 129.3, 129.2 ($C_{quatBz}$); 129.1, 129.0, 128.7, 128.7 (BzO); 91.2 ($C_1$); 71.7 ($C_4$); 69.3 ($C_3$); 68.3 ($C_5$); 68.0 ($C_2$); 16.6 ($C_6$)
MS (ESI): m/z calculated for $[C_{34}H_{28}O_9Na]^+$(M + $Na^+$): 603.16; found: 603.35

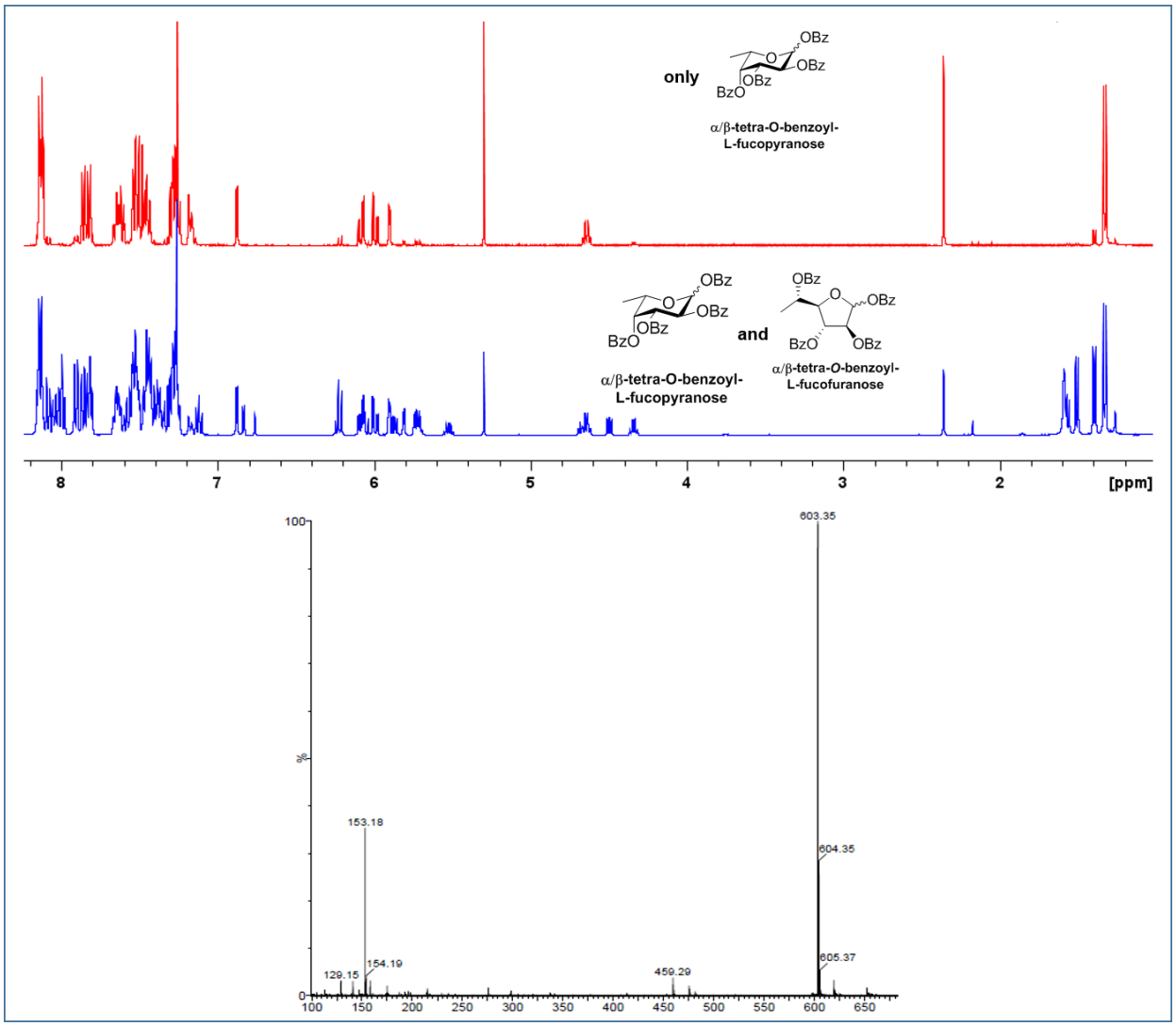
**$^{1}$H NMR spectrum of 2.19 in $CDCl_3$ (400 MHz)**



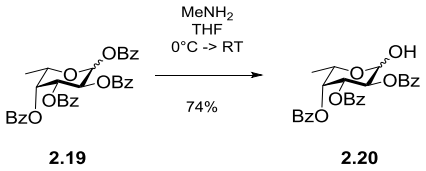

**$^{13}$C NMR spectrum of 2.6 in $CDCl_3$ (100 MHz)**



**$^1$H NMR spectrum of α and β tetra-*O*-benzoyl-fucopyranose and fucofuranose in $CDCl_3$ (400 MHz) and ESI MS spectrum of the four isomers**




2,3,4-tri-O-benzoyl-L-fucopyranose **2.20**

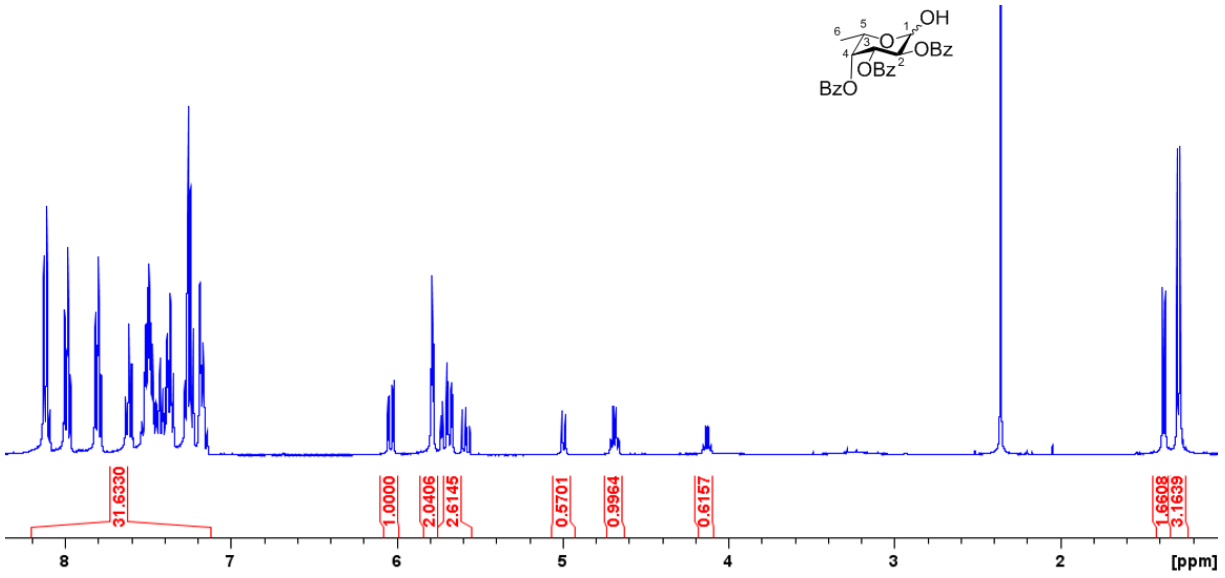



707.3 mg (1.218 mmol, 1 equiv.) 1,2,3,4-tetra-O-benzoyl-L-fucopyranose was dissolved in dry THF and cooled to 0°C under $N_2$ atmosphere. 731 µl (1.462 mmol, 1.2 equiv) 2M $MeNH_2$ in THF was added dropwise to the solution. After one hour, the major spot on the TLC had slightly lower $R_f$ than the starting material and a new spot started to appear under these two spots. The solution was concentrated *in vacuo*, and the product was purified by flash chromatography ($R_f$ = 0.18 and 0.15 for α and β anomers in toluene : EtOAc = 9 : 1) to yield the pure product as a colourless oil.

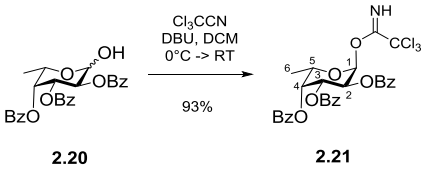
Yield: 74% (α : β = 2 : 1)

$^1$H NMR (400 MHz, $CDCl_3$): α-anomer: δ = 8.18 – 7.23 (m, 15H, OBz), 6.04 (dd, 1H, $H_3$, $J_{3\text{-}2}$ = 10.7 Hz, $J_{3\text{-}4}$ = 3.2 Hz), 5.81 – 5.77 (m, 2H, $H_1$, $H_4$), 5.68 (m, 1H, $H_2$), 4.69 (dd, 1H, $H_5$, $J_{5\text{-}6}$ = 6.5 Hz, $J_{5\text{-}4}$ = 13.1 Hz), 1.30 (d, 3H, $H_6$, $J_{6\text{-}5}$ = 6.5 Hz)
β-anomer: δ = 8.18 – 7.23 (m, 15H, OBz), 5.73 (dd, 1H, $H_4$, $J_{4\text{-}5}$ = 1.0 Hz, $J_{3\text{-}4}$ = 3.5 Hz), 5.71 – 5.66 (m, 1H, $H_3$), 5.58 (dd, 1H, $H_2$, $J_{2\text{-}1}$ = 7.9 Hz, $J_{2\text{-}3}$ = 10.4 Hz), 4.99 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 7.9 Hz), 4.13 (dd, 1H, $H_5$, $J_{5\text{-}6}$ = 6.1 Hz, $J_{5\text{-}4}$ = 12.8 Hz), 1.39 (d, 3H, $H_6$, $J_{6\text{-}5}$ = 6.5 Hz)
MS (ESI): m/z calculated for $[C_{27}H_{24}O_8Na]^+$(M + $Na^+$): 499.14; found: 499.84

**$^1$H NMR spectrum of 2.20 in $CDCl_3$ (400 MHz)**




2,3,4-tri-O-benzoyl-α-L-fucopyranosyl-1-trichloroacetoimidate **2.21**

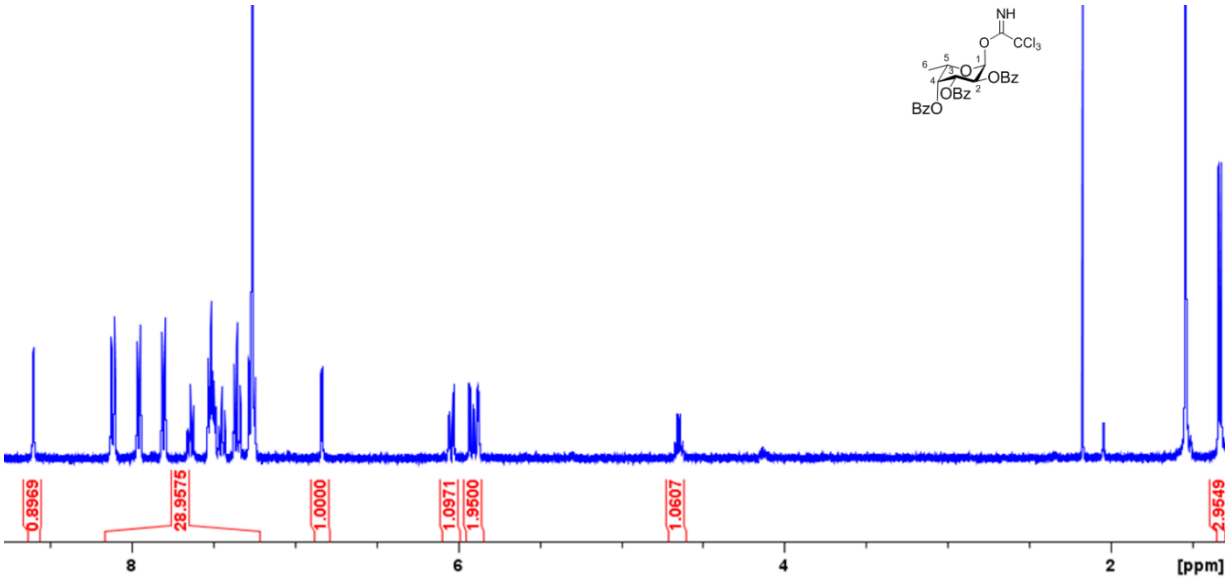



**2.20** **2.21**

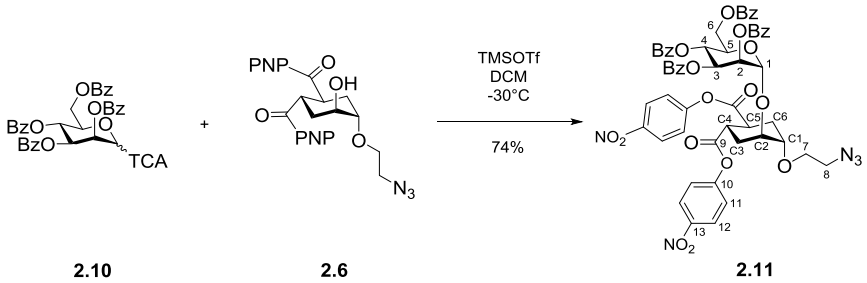
460 mg (0.969 mmol, 1 equiv.) 2,3,4-tri-O-benzoyl-L-fucopyranose **2.20** was dissolved in dry dichloromethane and cooled to 0°C. 58 µl (0.388 mmol, 0.4 equiv.) DBU and 972 µl (9.694 mmol, 10 equiv.) trichloroacetonitrile were added to the solution which was left to warm to room temperature and stirred until the starting material disappeared. The solution was concentrated *in vacuo*, and the product was purified by flash chromatography ($R_f$ = 0.34 in hexane : EtOAc = 9 : 1) to yield the pure product as a colourless oil (only α-anomer).

Yield: 93%.
$^1$H NMR (400 MHz, $CDCl_3$): δ = 8.60 (s, 1H, NH), 8.18 – 7.25 (m, 15H, OBz), 6.83 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 3.5 Hz), 6.04 (dd, 1H, $H_3$, $J_{3\text{-}2}$ = 10.9 Hz, $J_{3\text{-}4}$ = 3.5 Hz), 5.91 (dd, 1H, $H_2$, $J_{2\text{-}3}$ = 10.9 Hz, $J_{2\text{-}1}$ = 3.5 Hz), 5.89-5.86 (m, 1H, $H_4$), 4.65 (dd, 1H, $H_5$, $J_{5\text{-}6}$ = 6.5 Hz, $J_{5\text{-}4}$ = 6.3 Hz), 1.30 (d, 3H, $H_6$, $J_{6\text{-}5}$ = 6.5 Hz)
MS (ESI): m/z calculated for $[C_{29}H_{24}O_8ClNa]^+$(M + $Na^+$): 642.05; found: 642.37

**$^1$H NMR spectrum of 2.6 in $CDCl_3$ (400 MHz)**




1,2-Cyclohexanedicarboxylic acid, (1S,2S,4S,5S)-4-(2-azidoethoxy)-5-[(2,3,4,6-tetra-*O*-benzoyl-α-D-mannopyranosyl)oxy]-1,2-bis(p-nitrophenylester) **2.11**




**2.10** **2.6** **2.11**

A mixture of the acceptor **2.6** (1203 mg, 2.334 mmol, 1 equiv.) and the donor **2.10** (2076 mg, 2.801 mmol, 1.2 equiv.) was co-evaporated with toluene three times. Powdered and activated 4Å molecular sieves (acid washed) were added. The mixture was kept under vacuum for a few hours and then dissolved in dry DCM (10 ml). After cooling to -30°C, TMSOTf (84 µl, 0.467 mmol, 0.2 equiv.) was added to the stirred mixture. The solution was stirred at -30°C for 1 hour and upon completion, the reaction was quenched with TEA. The mixture was warmed to room temperature and filtered over a celite pad. The filtrate was evaporated at reduced pressure and the crude product was purified by flash chromatography (toluene with a gradient of ethyl acetate from 0 to 10%) to yield the pure product **2.11** as a white foam.

Yield: 74%

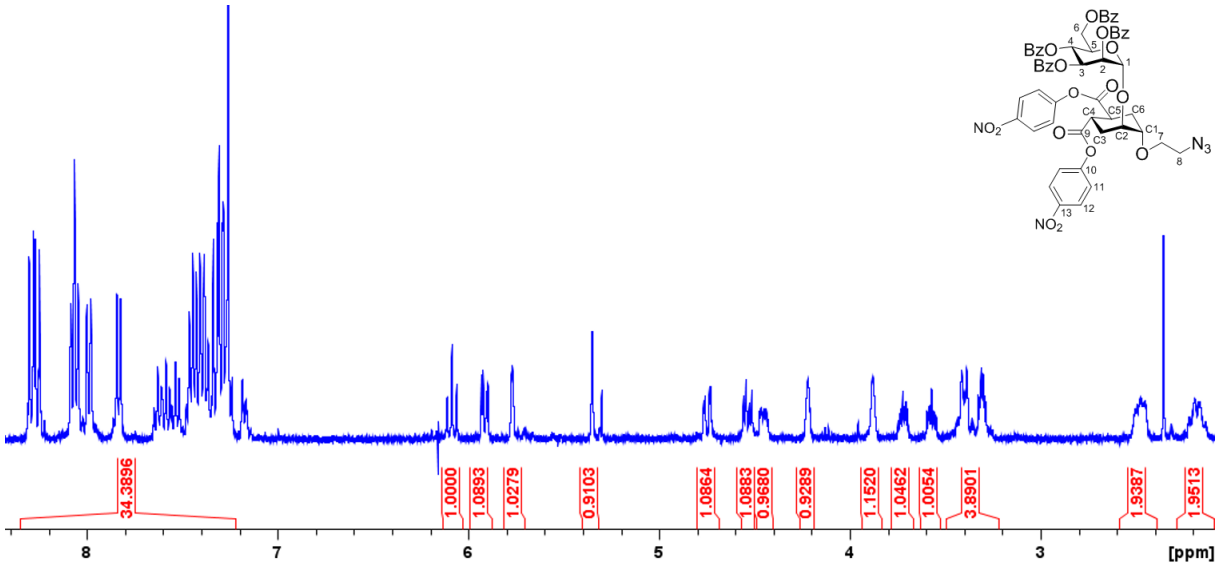
$[\alpha]_D^{20}$: -18.0° (c = 0.5 in $CHCl_3$)

$^1$H NMR (400 MHz, $CDCl_3$): δ = 8.30 – 8.20 (m, 4H; $H_{12}$), 8.09 – 8.00 (m, 4H; OBz), 7.99 – 7.94 (m, 2H; OBz), 7.83 – 7.79 (m, 3H; OBz), 7.63 – 7.48 (m, 3H; OBz), 7.46 – 7.20 (m, 12H; OBz, $H_{11}$), 6.07 (t, 1H; $H_4$, $J_{4\text{-}3} = J_{4\text{-}5}$ = 10.0 Hz), 5.90 (dd, 1H; $H_3$, $J_{3\text{-}4}$ = 10.0, $J_{3\text{-}2}$ = 3.3 Hz), 5.75 (dd, 1H; $H_2$, $J_{2\text{-}1}$ = 1.7, $J_{2\text{-}3}$ = 3.3 Hz), 5.34 (d, 1H; $H_1$, $J_{1\text{-}2}$ = 1.7 Hz), 4.73 (dd, 1H; $H_{6b}$, $J_{6b\text{-}5}$ = 2.9, $J_{6a\text{-}6b}$ = 12.0 Hz), 4.53 (dd, 1H; $H_{6a}$, $J_{6a\text{-}5}$ = 5.3, $J_{6a\text{-}6b}$ = 12 Hz), 4.48 – 4.40 (m, 1H; $H_5$), 4.23 – 4.18 (m, 1H; $C_2$), 3.89 – 3.84 (m, 1H; $C_1$), 3.74 – 3.67 (m, 1H; $H_{7a}$), 3.60 – 3.52 (m, 1H; $H_{7b}$), 3.44 – 3.34 (m, 2H; $C_4$, $C_5$), 3.35 – 3.23 (m, 2H; $H_8$), 2.52 – 2.41 (m, 2H; $C_{3eq}$,$C_{6eq}$), 2.25 – 2.09 (m, 2H; $C_{3ax}$, $C_{6ax}$)

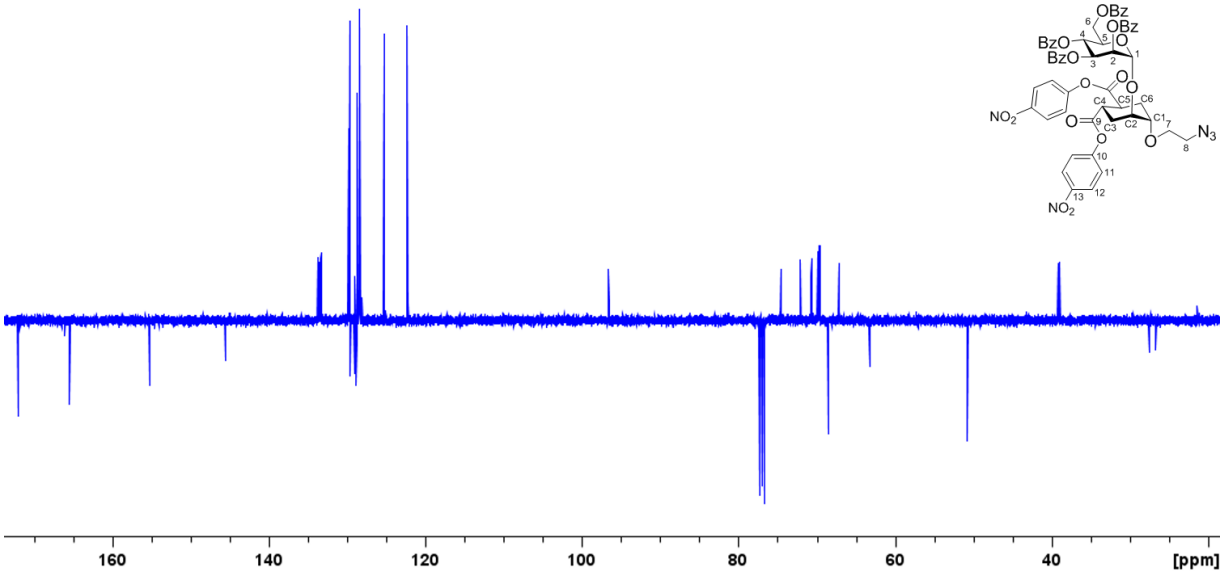
$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 172.3, 172.2 ($C_9$); 166.3, 165.8, 165.7 (COBz); 155.5, 155.4 ($C_{10}$); 145.8 ($C_{13}$); 133.9, 133.8, 133.6, 133.5 (CHBz); 130.1, 130.1, 129.9, 129.9, 129.9(CHBz); 129.3, 129.2, 129.1, 129.0 ($C_{quatBz}$); 128.9, 128.7, 128.4 (CHBz); 125.5, 125.4 ($C_{12}$); 122.7, 122.6 ($C_{11}$); 96.8 ($C_1$); 74.9 ($C_{C1}$); 72.3 ($C_{C2}$); 70.9 ($C_2$); 70.1 ($C_5$); 69.9 ($C_3$); 68.7 ($C_7$); 67.4 ($C_4$); 63.5 ($C_6$); 51.0 ($C_8$); 39.4, 39.2 ($C_{C4}$,$C_{C5}$); 27.8 ($C_{C3}$); 27.0 ($C_{C6}$)

MS (ESI): m/z calculated for $[C_{56}H_{47}N_5O_{19}Na]^+$: 1117.0; found 1116.1; HRMS: m/z calculated for [C56H47N5O19Na]$^+$: 1116.27575; found: 1116.27734.

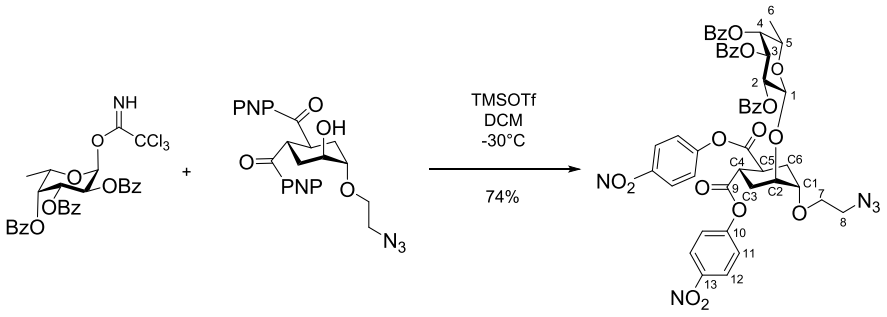
**$^1$H NMR spectrum of 2.11 in $CDCl_3$ (400 MHz)**




**$^{13}$C NMR spectrum of 2.11 in $CDCl_3$ (100 MHz)**



1,2-Cyclohexanedicarboxylic acid, (1S,2S,4S,5S)-4-(2-azidoethoxy)-5-[(2,3,4-tri-O-benzoyl-β-D-fucopyranosyl)oxy]-1,2-bis(p-nitrophenylester) **2.22**




A mixture of the acceptor **2.6** (250 mg, 0.4013 mmol, 1 equiv.) and the donor **2.21** (207 mg, 0.4013 mmol, 1 equiv.) was co-evaporated with toluene three times. Powdered and activated 4Å molecular sieves (acid washed) were added. The mixture was kept under vacuum for a few hours and then dissolved in dry DCM (4 ml). The solution was cooled to -30°C and TMSOTf (9 µl, 0.0401 mmol, 0.2 equiv.) was added to the stirred mixture. The solution was stirred at -30°C for 3.5 hours and upon completion, the reaction was quenched with TEA. The mixture was filtered over a celite pad. The solution was concentrated *in vacuo*, and the product was purified by flash chromatography ($R_f$ = 0.22 in hexane : EtOAc = 6 : 4) to yield the pure product as a colourless foam.

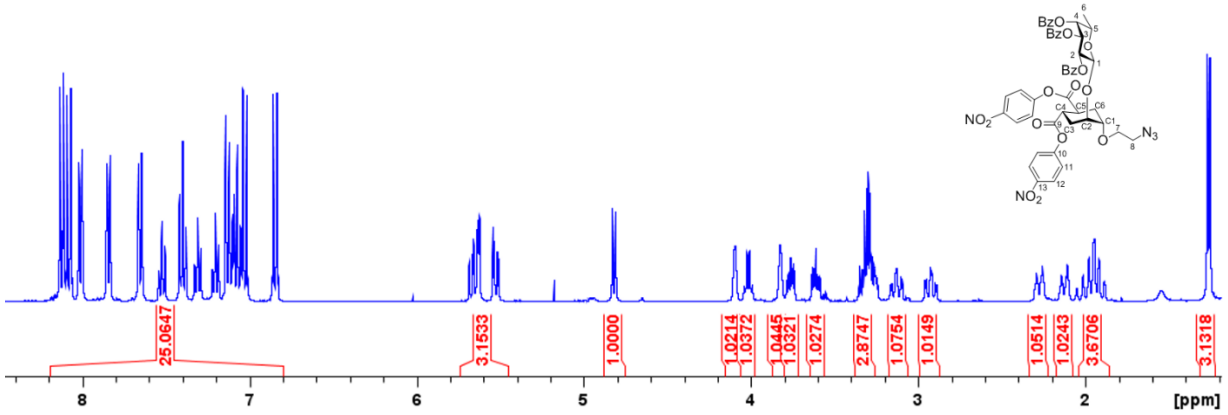
Yield: 72%.

$^1$H NMR (400 MHz, $CDCl_3$): δ = 8.23 – 6.83 (m, 23H, HAr), 5.75 – 5.63 (m, 2H, $H_2$, $H_4$), 5.57 (dd, 1H, $H_3$, $J_{3\text{-}2}$ = 10.5 Hz, $J_{3\text{-}4}$ = 3.4 Hz), 4.86 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 8.0 Hz), 4.14 (m, 1H, $C_2$), 4.06 (q, 1H, $H_5$, $J_{5\text{-}4}$ = 12.9 Hz, $J_{5\text{-}6}$ = 6.3 Hz), 3.87 (m, 1H, $C_1$), 3.84 – 3.77 (m, 1H, $H_{7a}$), 3.69 – 3.62 (m, 1H, $H_{7b}$), 3.41 – 3.31 (m, 2H, $H_{8a,b}$), 3.22 – 3.13 (m, 1H, $C_4$), 3.02 – 2.92 (m, 1H, $C_5$), 2.36 – 2.27 (m, 1H, $C_{3\text{ or }6eq}$), 2.21 – 2.12 (m, 1H, $H_{6\text{ or }3eq}$), 2.08 – 1.90 (m, 2H, $H_{3ax,\ 6ax}$)

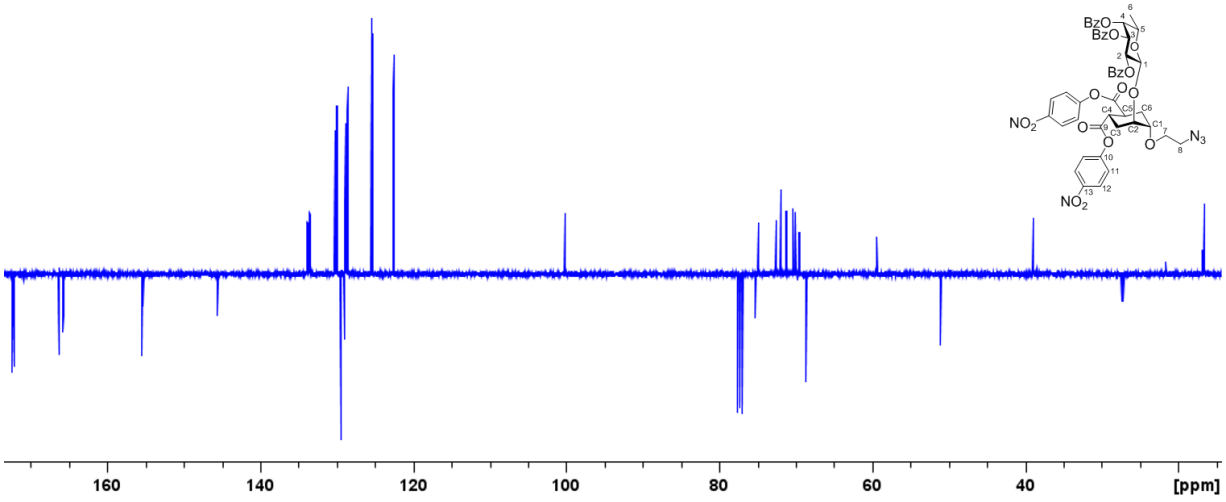
$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 172.5, 172.3 ($C_9$); 166.3, 165.9, 165.7 (COBz); 155.5, 155.3 ($C_{10}$); 145.8, 145.7 ($C_{13}$); 133.9, 133.7, 133.5 (CHBz); 130.3, 130.0, 130.0 (CHBz); 129.5, 129.1 ($C_{quatBz}$); 128.9, 128.7, 128.6 (CHBz); 125.5, 125.4 ($C_{12}$); 122.7, 122.6 ($C_{11}$); 100.2 ($C_1$); 75.0 ($C_{C1}$); 72.7 ($C_{C2}$); 72.0 ($C_3$); 71.3 ($C_2$); 70.4 ($C_5$); 70.1 ($C_7$); 69.6 ($C_4$); 51.2 ($C_8$); 39.0, 39.0 ($C_{C4}$,$C_{C5}$); 27.5, 27.2 ($C_{C3}$, $C_{C6}$); 16.9, 16.7 ($C_6$)

MS (ESI): m/z calculated for $[C_{49}H_{43}N_5O_{17}Na]^+$: 996.25; found: 996.86

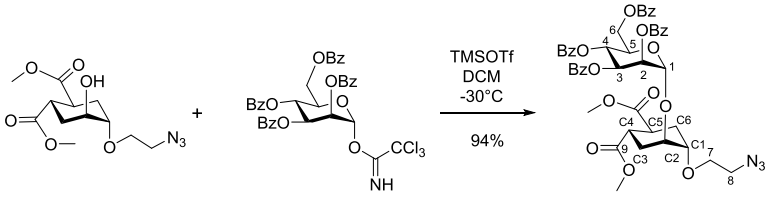
**$^1$H NMR spectrum of 2.22 in $CDCl_3$ (400 MHz)**



**$^{13}$C NMR spectrum of 2.22 in $CDCl_3$ (100 MHz)**




2,3,4,6-tetra-*O*-benzoyl-α(1,2)pseudomannobioside (azidoethanol linker) **2.28**




**2.26** **2.10** **2.28**

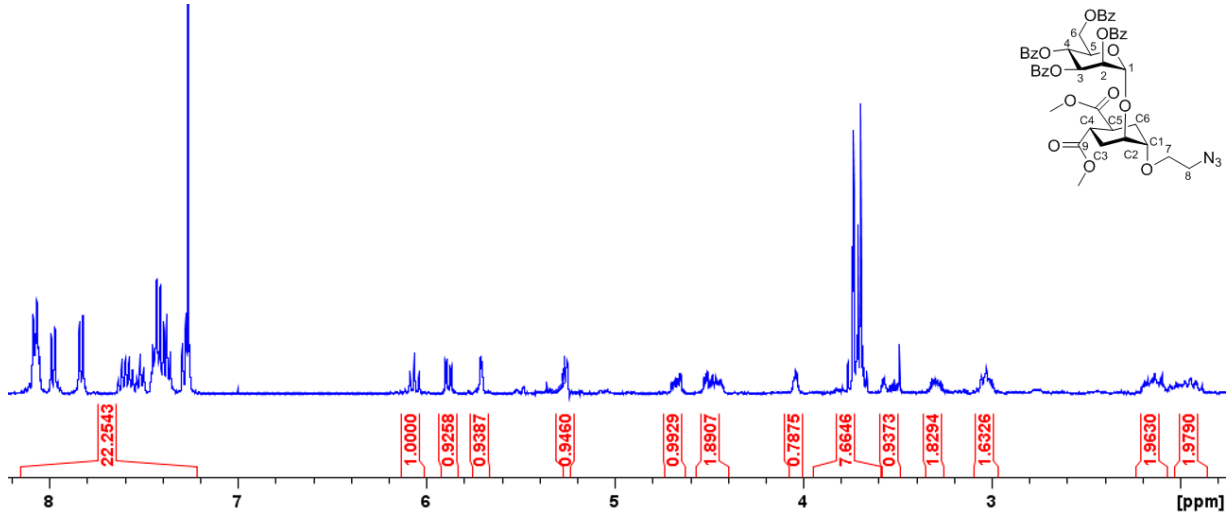
A mixture of the acceptor **2.26** (88.9 mg, 0.30 mmol, 1 equiv.) and the donor **2.10** (218.6 mg, 0.295 mmol, 1 equiv.) was co-evaporated with toluene three times. Powdered and activated 4Å molecular sieves (acid washed) were added. The mixture was kept under vacuum for a few hours and then dissolved in dry DCM (3 ml). The solution was cooled to -30°C and TMSOTf (14 µl, 0.06 mmol, 0.2 equiv.) was added to the stirred mixture. The solution was stirred at -30°C for 1 hour and upon completion, the reaction was quenched with TEA. The mixture was filtered over a celite pad. The solution was concentrated *in vacuo*, and the product was purified by flash chromatography ($R_f$ = 0.24 in hexane : EtOAc = 7 : 3) to yield the pure product as a colourless foam.

Yield: 94%.

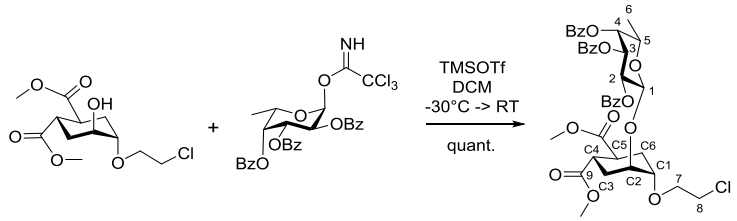
$^1$H NMR (400 MHz, $CDCl_3$): δ = 8.08 – 8.02 (m, 4H, OBz), 7.99 – 7.94 (m, 2H, OBz), 7.84 – 7.79 (m, 2H, OBz), 7.62 – 7.46 (m, 3H, OBz), 7.45 – 7.32 (m, 8H, OBz), 7.28 – 7.24 (m, 1H, OBz), 6.04 (t, 1H, $H_4$, $J_{4\text{-}5}$ = $J_{4\text{-}3}$ = 10 Hz), 5.86 (dd, 1H, $H_3$, $J_{3\text{-}2}$ = 3.6 Hz), 5.69 (dd, 1H, $H_2$, $J_{2\text{-}1}$ = 1.2 Hz), 5.25 (d, 1H, $H_1$), 4.66 (dd, 1H, $H_{6b}$, $J_{gem}$ = 12 Hz, $J_{6a\text{-}5}$ = 5.2 Hz), 4.45 – 4.40 (m, 1H, $H_5$), 4.05 – 3.99 (m, 1H, $C_2$), 3.72 (s, 3H, OMe), 3.68 (s, 3H, OMe), 3.74 – 3.62 (m, 2H, $H_{7b}$, $C_1$), 3.55 (ddd, 1H, $H_{7a}$, $J_{7a\text{-}8a}$ = 6.4 Hz, $J_{7a\text{-}8b}$ = 3.6 Hz, $J_{gem}$ = 10.4 Hz), 3.33 – 3.19 (m, 2H, $H_8$), 3.09 – 2.97 (m, 2H, $C_4$, $C_5$), 2.17 – 2.06 (m, 2H, $C_{3eq}$, $C_{6eq}$), 2.04 – 1.91 (m, 2H, $C_{3ax}$, $C_{6ax}$)

$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 174.9, 174.8 ($C_9$); 166.3, 165.8, 165.7, 165.6 (COBz); 133.8, 133.7, 133.5, 133.4 (CHBz); 130.1, 130.1, 129.9, 129.9 (CHBz); 130.0, 129.5, 129.2, 129.1 ($C_{quatBz}$); 128.9, 128.7, 128.7, 128.6 (CHBz); 96.4 ($C_1$); 75.4 ($C_{C1}$); 73.0 ($C_{C2}$); 70.9 ($C_2$); 70.0 ($C_3$); 69.8 ($C_5$); 68.3 ($C_7$); 67.3 ($C_4$); 63.3 ($C_6$); 52.3, 52.2 (OMe); 51.0 ($C_8$); 39.3, 39.1 ($C_{C4}$,$C_{C5}$); 27.8, 27.2 ($C_{C3}$, $C_{C6}$)

**$^1$H NMR spectrum of 2.28 in $CDCl_3$ (400 MHz)**




2,3,4,6-tetra-*O*-benzoyl-β(1,2)pseudofucobioside (chloroethanol linker); dimethyl (1S,2S,4S,5S)-4-(2-chloroethoxy)-5-(((2S,3R,4S,5S,6R)-3,4,5-tris(benzoyloxy)-6-methyltetrahydro-2H-pyran-2-yl)oxy)cyclohexane-1,2-dicarboxylate **2.29**

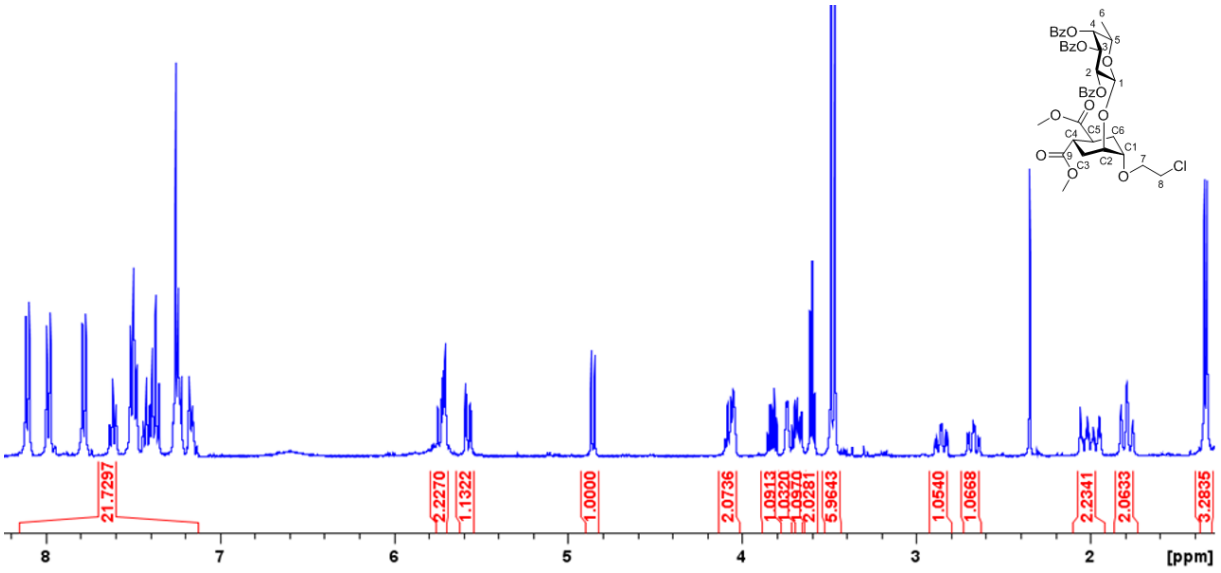



**2.27** **2.21** **2.29**

A mixture of the acceptor **2.27** (60.8 mg, 0.10 mmol, 1 equiv.) and the donor **2.21** (28.8 mg, 0.10 mmol, 1 equiv.) was co-evaporated with toluene three times. Powdered and activated 4Å molecular sieves (acid washed) were added. The mixture was kept under vacuum for a few hours and then dissolved in dry DCM (1 ml). The solution was cooled to -30°C and TMSOTf (2.2 µl, 0.01 mmol, 0.1 equiv.) was added to the stirred mixture. The solution was stirred at -30°C for 3 hours and then left to warmt to RT. Upon completion, the reaction was quenched with TEA. The mixture was filtered over a celite pad. The solution was concentrated in vacuo, and the product was purified by flash chromatography ($R_f$ = 0.16 in toluene : EtOAc = 9 : 1) to yield the pure product as a colourless foam.

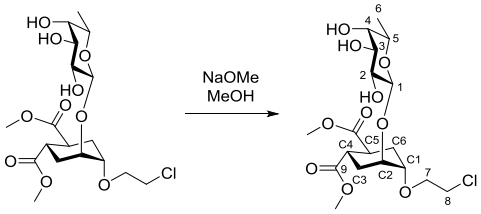
Yield: quant.
$^1$H NMR (400 MHz, $CDCl_3$): δ = 8.16 – 7.21 (m, 15H, OBz), 5.79 – 5.69 (m, 2H, $H_2$, $H_4$), 5.57 (dd, 1H, $H_3$, $J_{3\text{-}2}$ = 10.3 Hz, $J_{3\text{-}4}$ = 3.4 Hz), 4.86 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 8.0 Hz), 4.11 – 4.03 (m, 2H, $C_2$, $H_5$), 3.87 – 3.80 (m, 1H, $C_1$), 3.76 – 3.72 (m, 1H, $H_{7a}$), 3.72 – 3.65 (m, 1H, $H_{7b}$), 3.60 (t, 2H, $H_{8a,b}$, $J$ = 5.5 Hz), 3.49 (s, 3H, OMe), 3.47 (s, 3H, OMe), 2.86 (td, 1H, $C_4$, $J$ = 11.9 Hz, $J$ = 2.8 Hz), 3.02 – 2.92 (td, 1H, $C_5$, $J$ = 11.9 Hz, $J$ = 3.2 Hz), 2.07 – 1.92 (m, 2H, $C_{3eq}$, $C_{6eq}$), 1.85 – 1.74 (m, 2H, $H_{3ax}$, $H_{6ax}$), 1.64 (d, 1H, $H_6$, $J_{6\text{-}5}$ = 6.5 Hz)
MS (ESI): m/z calculated for $[C_{39}H_{41}ClO_{13}Na]^+$: 775.21, found: 775.23

**$^1$H NMR spectrum of 2.29 in $CDCl_3$ (400 MHz)**




<u>β(1,2)pseudofucobioside (chloroethanol linker); dimethyl (1S,2S,4S,5S)-4-(2-chloroethoxy)-5-(((2S,3R,4S,5R,6R)-3,4,5-trihydroxy-6-methyltetrahydro-2H-pyran-2-yl)oxy)cyclohexane-1,2-dicarboxylate</u>

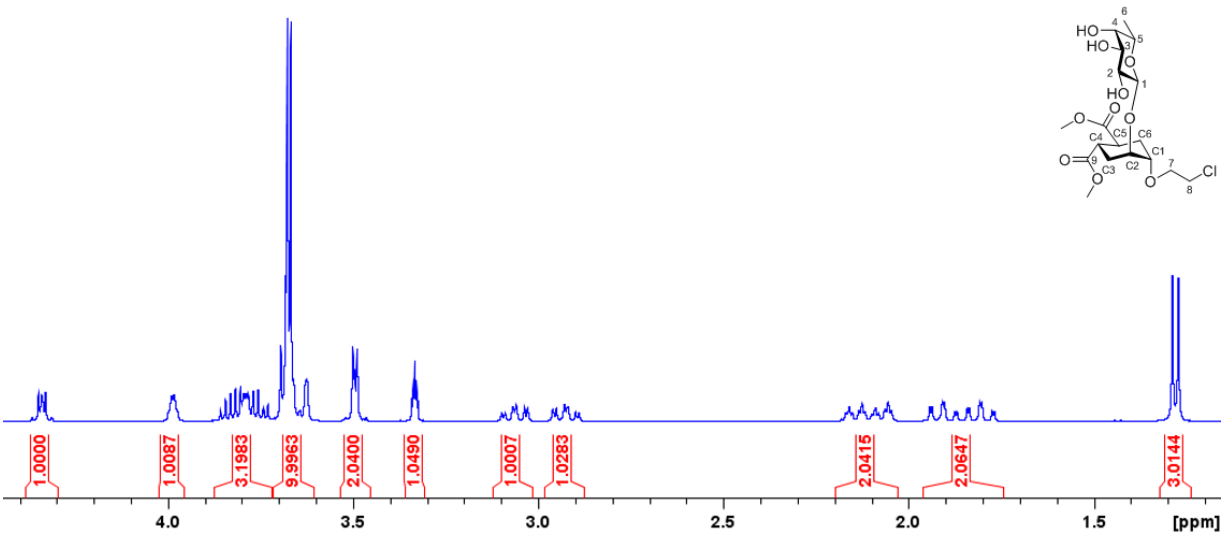
**<u>2.30</u>**



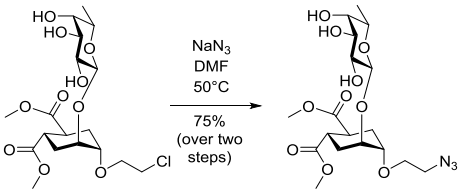

According to general procedure 2 (Zemplén-debenzoylation, see below).

$^1$H NMR (400 MHz, $CD_3OD$): δ = 4.36 – 4.32 (m, 1H, $H_1$), 4.00 – 3.97 (m, 1H, $C_2$), 3.86 – 3.72 (m, 3H, $C_1$, $H_{7a,b}$), 3.72 – 3.61 (m, 9H, OMe, $H_{8a,b}$, $H_4$), 3.52 – 3.46 (m, 2H, $H_2$, $H_3$), 3.33 (q, 1H, $H_5$, $J$ = 1.6 Hz), 3.06 (td, 1H, $C_4$, $J$ = 12.3 Hz, $J$ = 3.3 Hz), 2.92 (td, 1H, $C_5$, $J$ = 12.3 Hz, $J$ = 3.2 Hz), 2.11 (qt, 2H, $C_{3eq}$, $C_{6eq}$, $J$ = 13.8 Hz, $J$ = 3.6 Hz), 1.85 (dtd, 2H, $H_{3ax}$, $H_{6ax}$, $J$ = 40.1 Hz, $J$ = 13.1 Hz, $J$ = 1.9 Hz), 1.28 (d, 1H, $H_6$, $J_{6\text{-}5}$ = 6.4 Hz)
MS (ESI): m/z calculated for $[C_{39}H_{41}ClO_{13}Na]^+$: 775.21; found: 775.51

**$^1$H NMR spectrum of 2.30 in $CD_3OD$ (400 MHz)**




β(1,2)pseudofucobioside (chloroethanol linker); dimethyl (1S,2S,4S,5S)-4-(2-azidoethoxy)-5-(((2S,3R,4S,5R,6R)-3,4,5-trihydroxy-6-methyltetrahydro-2H-pyran-2-yl)oxy)cyclohexane-1,2-dicarboxylate **2.31**




25.5 mg (0.06 mmol, 1 equiv.) **2.30** was dissolved in dry DMF and 26.3 mg (0.40 mmol, 7 equiv.) $NaN_3$ was added to the stirred solution. The reaction was stirred at 50°C for 5 days, then cooled to RT, concentrated and purified by flash chromatography ($R_f$ = 0.13 in DCM : MeOH = 95 : 5), to yield the pure product as a white foam. Characterization reported at the library ligands (**ISh049**).

## 2.1.2.3 Characterization of the library members

General procedure yielding bisamides **ISh0XX**:

1. Coupling with amines

Scaffold **2.11** or **2.22** (1 equiv.) was dissolved in dry THF or DMF under $N_2$ and 3 equiv. amine was added to the solution (and additional 3.2 equiv. TEA, for amines sold as ammonium salts and amines with low reactivity). The mixture was stirred at RT from 1 hour to 2 days, and monitored by TLC or NMR. Upon completion, the solution was washed with 1M HCl, 1M NaOH and water on supported liquid extraction cartridges (Biotage ISOLUTE® HM-N). The crude was purified by flash chromatography (DCM with gradient of methanol from 0 to 20%) or used without purification in the following Zemplén-deprotection if the purity was satisfying.

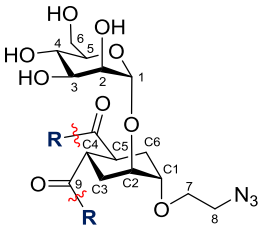
2. Zemplén-debenzoylation

1 equiv. protected bisamide was dissolved in distilled MeOH and 1M freshly prepared NaOMe in MeOH was added to the solution (1.5 mol equiv NaOMe) to 0.1M final concentration). After completion, the reaction was neutralized with Amberlite® IR120 hydrogen form ion-exchange resin, filtered and concentrated *in vacuo*. The crude was purified by direct or reverse phase flash chromatography, yielding the pure product.

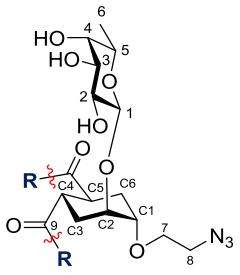
3. Azido-group reduction to amine

The azide (1 equiv.) was dissolved in dry MeOH and a catalytic amount (5-10%) of Pd/C was added. The suspension was placed under $H_2$ atmosphere and stirred at RT. Upon completion (monitoring by TLC), the mixture was filered over a celite pad and the solvent was evaporated *in vacuo*. The crude was purified by direct or reverse phase flash chromatography, yielding the pure product.

The unusual numbering used for pseudo-disaccharide and pseudo-trisaccharide derivatives are shown below. The cyclohexanediol unit is indicated as 'C'. Numbering of this ring does not follow IUPAC rules, but it was adopted in analogy to the numbering of mannose.

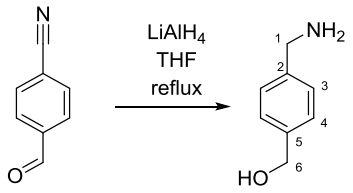


**ISh0XX (mannosylated)**



**ISh0XX (fucosylated)**

**ISh001 amine**




**2.31** **2.32**

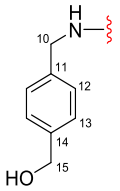
596.9 mg (4.55 mmol, 1 equiv.) nitrile **2.31** and 691.0 (18.21 mmol, 4 equiv.) $LiAlH_4$ were separately dissolved in distilled THF and cooled to 0°C. The nitrile solution was added dropwise to the stirred and cooled $LiAlH_4$ solution. The reaction was refluxed for 3 hours. Upon completion, the reaction was cooled to 0°C and quenched according to the Fieser-method (for X gr $LiAlH_4$ X ml water was slowly added at 0°C, then X ml 15% aqueous NaOH solution was added, followed by 3*X ml water. The reaction was stirred at RT for 15 minutes and $MgSO_4$ was added. The suspension was filtered on a celite pad). The solvent was evaporated and the crude was purified by flash chromatography (DCM : MeOH = 7 : 3)

Yield: 93%

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.32 (s, 4H, $H_3$, $H_4$), 4.58 (s, 2H, $H_6$), 3.78 (s, 2H, $H_1$).

MS (ESI): m/z calculated for $[C_8H_{12}NO]^+$: 138.09, found: 138.11

**ISh001**




Yield: 74% over two steps

$[\alpha]_D$: +12.1° (c = 0.81, MeOH)

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.29 (d, 4H, $H_{12}$, $J_{12\text{-}13}$ = 8 Hz), 7.23 (d, 4H, $H_{13}$, $J_{13\text{-}12}$ = 8 Hz), 4.96 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.6 Hz), 4.58 (s, 4H, $H_{15a,b}$), 4.31 (s, 4H, $H_{10a,b}$), 4.08 – 4.03 (m, 1H, $C_2$), 3.93 – 3.89 (m, 1H, $H_2$), 3.89 – 3.84 (m, 1H, $H_{6a}$), 3.84 – 3.65 (m, 5H, $C_1$, $H_{6b}$, $H_{7a,b}$, $H_3$), 3.64 – 3.54 (m, 2H, $H_4$, $H_5$), 3.47 – 3.35 (m, 2H, $H_{8a,b}$), 3.02 – 2.85 (m, 2H, $C_4$, $C_5$), 2.06 – 1.86 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.2, 177.0 ($C_9$); 141.7 ($C_{14}$); 139.2 ($C_{11}$); 128.7, 128.6 ($C_{12}$); 128.3 ($C_{13}$); 100.4 ($C_1$); 76.6 ($C_3$); 75.7 ($C_5$); 72.7 ($C_{C1}$); 72.5 ($C_2$); 72.4 ($C_{C2}$); 69.3 ($C_7$); 68.9 ($C_4$); 65.1 ($C_{15}$); 63.2 ($C_6$); 52.1 ($C_8$); 43.8 ( $C_{10}$); 42.1, 41.9 ($C_{C4}$, $C_{C5}$); 29.9, 29.0 ($C_{C3,}$ $C_{C6}$)

**MS** (HRMS) calculated for $[C_{32}H_{43}N_5O_{11}Na]^+$: 696.28568; found: 696.28423

**ISh002**


Yield: 78% over two steps

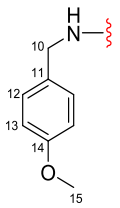
$[\alpha]_D$: +80.0° (c = 0.65, MeOH)

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 4.95 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.6 Hz), 4.02 (m, 1H, $C_2$), 3.92 – 3.82 (m, 2H, $H_2$, $H_{6a}$), 3.82 - 3.69 (m, 4H, $C_1$, $H_{6b}$, $H_3$, $H_4$); 3.66 (s, 3H, $H_{10}$); 3.65 (s, 3H, $H_{10}$), 3.63 - 3.51 (m, 2H, $H_{8a,b}$), 3.38 (m, 2H, $H_{7a,b}$), 3.32 (m, 1H, $H_5$), 3.00 - 2.82 (m, 2H, $C_4$, $C_5$), 2.20 - 2.04 (m, 2H, $C_{3eq}$, $C_{6eq}$), 1.90 - 1.75 (m, 2H, $C_{3ax}$, $C_{6ax}$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 176.8; 176.7 ($C_9$); 100.5 ($C_1$); 75.7 ($C_5$); 75.6 ($C_3$); 72.5 ($C_{C1}$); 72.4 ($C_2$); 72.1 ($C_{C2}$); 68.6 ($C_4$); 68.4 ($C_7$); 63.1 ($C_6$); 52.4 ($C_{10}$); 51.8 ($C_8$); 40.2, 40.1 ($C_{C4}$, $C_{C5}$); 28.9, 28.2 ($C_{C3}$, $C_{C6}$)

**MS** (HRMS) calculated for $[C_{18}H_{29}N_3O_{11}Na]^+$: 486.16943; found: 486.16870

**ISh003**




Yield: 75% over two steps

$[\alpha]_D$: +1.81° (c = 0.15, EtOH)

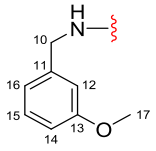
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.20 (d, 4H, $H_{12}$, $J_{12\text{-}13}$ = 8 Hz), 6.88 (d, 4H, $H_{13}$, $J_{13\text{-}12}$ = 8 Hz), 4.99 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.6 Hz), 4.26 (d, 4H, $H_{10a,b}$, $J_{10a\text{-}10b}$ = 2.4 Hz), 4.10 – 4.06 (m, 1H, $C_2$), 3.95 (dd, 1H, $H_2$, $J_{2\text{-}1}$ = 1.6 Hz, $J_{2\text{-}}$

$_3$ = 3.2 Hz), 3.93 – 3.87 (m, 1H, $H_{6a}$), 3.86 – 3.69 (m, 11H, $C_1$, $H_{6b}$, $H_{7a,b}$, $H_3$, $H_{15}$), 3.70 – 3.51 (m, 2H, $H_4$, $H_5$), 3.49 – 3.35 (m, 2H, $H_{8a,b}$), 3.01 – 2.86 (m, 2H, $C_4$, $C_5$), 2.08 – 1.90 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.0, 176.8 ($C_9$); 160.4 ($C_{14}$); 132.2 ($C_{11}$); 129.9, 129.8 ($C_{12}$); 115.0 ($C_{13}$); 100.4 ($C_1$); 76.6 ($C_3$); 75.6 ($C_5$); 72.5 ($C_2$); 72.5 ($C_{C2}$); 72.2 ($C_{C1}$); 69.3 ($C_7$); 68.9 ($C_4$); 63.2 ($C_6$); 55.8 ($C_{15}$); 52.2 ($C_8$); 43.6 ( $C_{10}$); 42.1, 42.0 ($C_{C4}$, $C_{C5}$); 29.8, 29.0 ($C_{C3,}$ $C_{C6}$)

**MS** (HRMS) calculated for $[C_{32}H_{43}N_5O_{11}Na]^+$: 696.28568; found: 696.28465

**ISh004**




Prepared by Norbert Varga

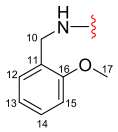
$[\alpha]_D$: +16.2° (c = 0.78, EtOH)

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.20 (t, 2H, $H_{15}$, $J_{15\text{-}16} = J_{15\text{-}14}$ =8 Hz), 6.85 - 6.75 (m, 6H, $H_{12}$, $H_{14}$, $H_{16}$), 4.96 (br s, 1H, $H_1$), 4.28 (s, 4H, $H_{10a,b}$, $J_{10a\text{-}10b}$ = 2.4 Hz), 4.09 – 4.03 (m, 1H, $C_2$), 3.94 - 3.90 (m, 1H, $H_2$), 3.90 – 3.84 (m, 1H, $H_{6a}$), 3.83 – 3.65 (m, 11H, $C_1$, $H_{6b}$, $H_{7a,b}$, $H_3$, $H_{17}$), 3.64 – 3.54 (m, 2H, $H_4$, $H_5$), 3.47 – 3.31 (m, 2H, $H_{8a,b}$), 3.03 – 2.85 (m, 2H, $C_4$, $C_5$), 2.08 – 1.90 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.2, 177.0 ($C_9$); 161.5 ($C_{13}$); 141.8 ($C_{11}$); 130.6 ($C_{15}$); 120.7, 120.7 ($C_{16}$); 114.0, 113.9, 113.7, 113.7 ($C_{14}$, $C_{12}$); 100.4 ($C_1$); 76.6 ($C_3$); 75.7 ($C_5$); 72.7 ($C_{C1}$); 72.5 ($C_2$); 72.4 ($C_{C2}$); 69.3 ($C_7$); 68.9 ($C_4$); 63.2 ($C_6$); 55.8 ($C_{17}$); 52.1 ($C_8$); 44.0 ( $C_{10}$); 42.1, 41.9 ($C_{C4}$, $C_{C5}$); 29.9, 29.0 ($C_{C3,}$ $C_{C6}$)

**MS** (HRMS) calculated for $[C_{32}H_{43}N_5O_{11}Na]^+$: 696.28568; found: 696.28599

**ISh005**




Amine purchased from Sigma-Aldrich. Reaction performed in THF overnight, with the addition of 3.2 equiv. $Et_3N$.

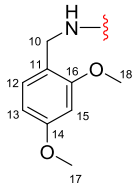
Yield: 65% over two steps

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.26 - 7.20 (t, 2H, $H_{13}$, $J_{13\text{-}14}$ = 7.5 Hz, $J_{13\text{-}12}$ = 7.9 Hz), 7.19 - 7.14 (d, 2H, $H_{15}$, $J_{15\text{-}14}$ = 7.4 Hz), 6.97 - 6.91 (d, 2H, $H_{12}$, $J_{12\text{-}13}$ = 7.9 Hz), 6.91 - 6.84 (t, 2H, $H_{14}$, $J_{14\text{-}15}$ = 7.4 Hz, $J_{14\text{-}13}$ = 7.5 Hz), 4.95 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.4 Hz), 4.35 (dd, 2H, $H_{10a}$, $J_{10a\text{-}NH}$ = 4.3 Hz, $J_{10a\text{-}10b}$ = 15.3 Hz), 4.23 (dd, 2H, $H_{10b}$, $J_{10b\text{-}NH}$ = 3.9 Hz, $J_{10b\text{-}10a}$ = 15.3 Hz), 4.07 – 4.03 (m, 1H, $C_2$), 3.92 (dd, 1H, $H_2$, $J_{2\text{-}1}$ = 1.4 Hz, $J_{2\text{-}3}$ = 3.1 Hz), 3.89 – 3.84 (m, 7H, $H_{6a}$, $H_{17}$), 3.83 – 3.65 (m, 5H, $C_1$, $H_{6b}$, $H_{7a,b}$, $H_3$), 3.62 – 3.57 (m, 2H, $H_4$, $H_5$), 3.43 – 3.37 (m, 2H, $H_{8a,b}$), 2.99 – 2.84 (m, 2H, $C_4$, $C_5$), 2.05 – 1.88 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.8, 177.6 ($C_9$); 159.5 ($C_{16}$); 130.3, 130.3 ($C_{13}$); 130.1, 130.0 ($C_{15}$); 128.3 ($C_{11}$); 122.3, 122.3 ($C_{14}$); 112.1, 112.1 ($C_{12}$); 101.1 ($C_1$); 77.4 ($C_3$); 76.4 ($C_5$); 73.4 ($C_2$); 73.2 ($C_{C1}$); 73.1 ($C_{C2}$); 70.6 ($C_7$); 69.6 ($C_4$); 63.9 ($C_6$); 56.6 ($C_{17}$); 52.8 ($C_8$); 42.9, 42.8 ( $C_{10}$); 40.2, 40.1 ($C_{C4}$, $C_{C5}$); 30.5, 29.6 ($C_{C3,}$ $C_{C6}$)

**MS** (MALDI-TOF) calculated for $[C_{32}H_{43}N_5O_{11}Na]^+$: 696.285; found: 696.561

**ISh006**




Amine purchased from Sigma-Aldrich. Reaction performed in THF overnight, with the addition of 3.2 equiv. $Et_3N$.

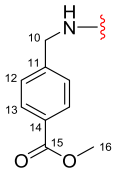
Yield: 60% over two steps.

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.02 (d, 2H, $H_{12}$, $J_{12\text{-}13}$ = 8.3 Hz), 6.48 (d, 2H, $H_{15}$, $J_{15\text{-}13}$ = 2.2 Hz), 6.38 (dt, 2H, $H_{13}$, $J_{13\text{-}15}$ = 2.2 Hz, $J_{13\text{-}12}$ = 8.3 Hz), 4.91 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.5 Hz), 4.26 (dd, 2H, $H_{10a}$, $J_{10a\text{-}NH}$ = 3.3 Hz, $J_{10a\text{-}10b}$ = 15.0 Hz), 4.15 (dd, 2H, $H_{10b}$, $J_{10b\text{-}NH}$ < 1.0 Hz, $J_{10b\text{-}10a}$ = 15.0 Hz), 4.01 (m, 1H, $C_2$), 3.89 - 3.86 (m, 2H, $H_2$, $H_{6a}$), 3.79 (s, 6H, $H_{18}$), 3.73 (s, 6H, $H_{17}$), 3.86 – 3.60 (m, 5H, $C_1$, $H_{6b}$, $H_{7a,b}$, $H_3$), 3.60 – 3.50 (m, 2H, $H_4$, $H_5$), 3.39 – 3.32 (m, 2H, $H_{8a,b}$), 2.91 – 2.76 (m, 2H, $C_4$, $C_5$), 2.00 – 1.81 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.6, 177.5 ($C_9$); 172.6 ($C_{16}$); 160.5 ($C_{14}$); 131.1, 131.0 ($C_{12}$); 120.6 ($C_{11}$); 106.1, 106.0 ($C_{13}$); 101.0 ($C_1$); 100.0 ($C_{15}$); 77.4 ($C_3$); 76.4 ($C_5$); 73.4 ($C_2$); 73.2 ($C_{C1}$); 73.1 ($C_{C2}$); 70.0 ($C_7$); 69.6 ($C_4$); 63.9 ($C_6$); 56.7, 56.6 ($C_{17}$, $C_{18}$); 52.8 ($C_8$); 42.9, 42.8 ( $C_{10}$); 39.9, 39.8 ($C_{C4}$, $C_{C5}$); 30.5, 29.5 ($C_{C3,}$ $C_{C6}$)

**MS** (MALDI-TOF) calculated for $[C_{34}H_{47}N_5O_{13}Na]^+$: 756.306; found: 756.654

**ISh007**




Yield: 62%

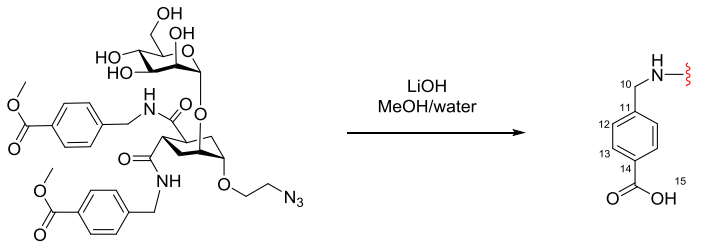
$[\alpha]_D$: +54.3° (c = 0.55, MeOH)

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.81 (d, 4H, $H_{13}$, $J_{13\text{-}12}$ = 8 Hz), 7.28 (d, 4H, $H_{12}$, $J_{12\text{-}13}$ = 8 Hz), 4.96 (s, 1H, $H_1$), 4.58 (dd, 2H, $H_{10a}$, $J_{10a\text{-}NH}$ < 1 Hz, $J_{10a\text{-}10b}$ = 16.0 Hz), 4.36 (dd, 2H, $H_{10b}$, $J_{10b\text{-}NH}$ = 2.3 Hz, $J_{10b\text{-}10a}$ = 15.8 Hz), 4.08 – 4.03 (m, 1H, $C_2$), 3.92 – 3.89 (m, 1H, $H_2$), 3.88 – 3.83 (m, 7H, $H_{6a}$, $H_{16}$), 3.83 – 3.64 (m, 5H, $C_1$, $H_{6b}$, $H_{7a,b}$, $H_3$), 3.62 – 3.52 (m, 2H, $H_4$, $H_5$), 3.45 – 3.33 (m, 2H, $H_{8a,b}$), 3.05 – 2.90 (m, 2H, $C_4$, $C_5$), 2.52 (s, 3H, $H_{16}$), 2.06 – 1.91 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.5, 177.3 ($C_9$); 168.5 ($C_{15}$); 145.9 ($C_{11}$); 130.8 ($C_{13}$); 130.0 ($C_{14}$); 128.2, 128.1 ($C_{12}$); 100.5 ($C_1$); 76.6 ($C_3$); 75.7 ($C_5$); 72.7 ($C_{C1}$); 72.6 ($C_2$); 72.5 ($C_{C2}$); 69.3 ($C_7$); 68.9 ($C_4$); 63.2 ($C_6$); 52.7 ($C_{16}$); 52.2 ($C_8$); 43.6 ( $C_{10}$); 42.0, 41.9 ($C_{C4}$, $C_{C5}$); 30.0, 29.2 ($C_{C3,}C_{C6}$)

**MS** (HRMS) calculated for $[C_{34}H_{43}N_5O_{13}Na]^+$: 752.27551; found: 752.27418

**ISh008**




**ISh007**

14.1 mg (0.02 mmol, 1 equiv.) **ISh007** was dissolved in dry MeOH and 3 equiv. aqueous LiOH solution was added to the mixture. Upon completion ($R_f$ = 0.66 in water : MeCN = 1 : 1 + 0.1% TFA), the reaction was concetrated *in vacuo*.

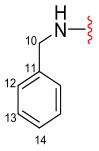
Yield: quantitative.

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 10.04 (s, 2H, $H_{15}$), 7.82 (d, 4H, $H_{13}$, $J_{13\text{-}12}$ = 7.4 Hz), 7.2 (d, 4H, $H_{12}$, $J_{12\text{-}13}$ = 7.4 Hz), 4.99 (s, 1H, $H_1$), 4.60 (dd, 2H, $H_{10a}$, $J_{10a\text{-}NH}$ < 1 Hz, $J_{10a\text{-}10b}$ = 16.1 Hz), 4.38 (dd, 2H, $H_{10b}$, $J_{10b\text{-}NH}$ < 1 Hz, $J_{10b\text{-}10a}$ = 16.1 Hz), 4.08 – 4.04 (m, 1H, $C_2$), 3.95 – 3.92 (m, 1H, $H_2$), 3.88 – 3.54 (m, 8H, $H_{6a,b}$, $C_1$, $H_{7a,b}$, $H_3$, $H_4$, $H_5$), 3.47 – 3.34 (m, 2H, $H_{8a,b}$), 3.07 – 2.91 (m, 2H, $C_4$, $C_5$), 2.08 – 1.91 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.3, 178.0 ($C_9$); 176.6 ($C_{15}$); 146.9 ($C_{14}$); 137.8 ($C_{11}$); 130.6 ($C_{13}$); 129.2 ($C_{12}$); 101.1 ($C_1$); 77.3 ($C_3$); 76.3 ($C_5$); 73.4 ($C_2$); 73.3 ($C_{C2}$); 73.2 ($C_{C1}$); 70.1 ($C_7$); 69.6 ($C_4$); 63.7 ($C_6$); 52.9 ($C_8$); 44.4 ( $C_{10}$); 42.7, 42.6 ($C_{C4}$, $C_{C5}$); 31.6, 30.7 ($C_{C3,}$ $C_{C6}$)

**MS** (ESI-TOF) calculated for $[C_{32}H_{37}N_5O_{13}Na_2]^+$: 745.2183; found: 745.0625 $[M\text{-}2H^+\text{+}2Na^+]$

**ISh009**




Amine purchased from Sigma-Aldrich. Reaction performed in THF in 1 hour.

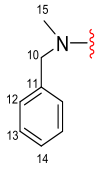
Yield: 88% over two steps

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.29 – 7.16 (m, 10H, $H_{12}$, $H_{13}$, $H_{14}$), 4.93 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.5 Hz), 4.32 – 4.22 (m, 4H, $H_{10a,b}$), 4.04 – 4.00 (m, 1H, $C_2$), 3.90 – 3.86 (m, 1H, $H_2$), 3.86 – 3.80 (m, 1H, $H_{6a}$), 3.80 – 3.62 (m, 5H, $C_1$, $H_{6b}$, $H_{7a,b}$, $H_3$), 3.60 – 3.51 (m, 2H, $H_4$, $H_5$), 3.43 – 3.30 (m, 2H, $H_{8a,b}$), 2.98 – 2.83 (m, 2H, $C_4$, $C_5$), 2.01 – 1.86 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.9, 177.7 ($C_9$); 140.9 ($C_{11}$); 130.3, 129.3, 129.2 ($C_{12}$, $C_{13}$, $C_{14}$); 101.1 ($C_1$); 77.3 ($C_3$); 76.4 ($C_5$); 73.4 ($C_{C1}$); 73.3 ($C_2$); 73.2 ($C_{C2}$); 70.0 ($C_7$); 69.6 ($C_4$); 63.9 ($C_6$); 52.8 ($C_8$); 44.8, 44.7 ( $C_{10}$); 42.8, 42.6 ($C_{C4}$, $C_{C5}$); 30.6, 29.8 ($C_{C3,}$ $C_{C6}$)

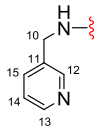
**MS** (MALDI-TOF) calculated for $[C_{30}H_{39}N_5O_9Na]^+$: 636.264; found: 636.504

**ISh010**



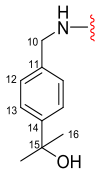

Amine purchased from Sigma-Aldrich. Reaction performed in THF in 2 hours.
Yield: 74% over two steps
**$^{1}$H NMR** (400 MHz, $CD_3OD$): δ = 7.42 – 7.16 (m, 10H, $H_{12}$, $H_{13}$, $H_{14}$), 4.98 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.5 Hz), 4.82 – 4.65 (m, 4H, $H_{10a,b}$), 4.15 – 4.10 (m, 1H, $C_2$), 4.07 – 3.98 (m, 1H, $H_2$, $H_{6a}$), 3.89 – 3.44 (m, 5H, $C_1$, $H_{6b}$, $H_{7a,b}$, $H_3$), 3.43 – 3.33 (m, 2H, $H_4$, $H_5$), 3.33 – 3.27 (m, 2H, $H_{8a,b}$), 3.27 – 3.13 (m, 6H, $H_{15}$), 2.91 – 2.83 (m, 2H, $C_4$, $C_5$), 2.11 – 1.76 (m, 4H, $C_3$, $C_6$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.4, 178.1 ($C_9$); 139.2 ($C_{11}$); 130.5, 129.4, 129.1 ($C_{12}$, $C_{13}$, $C_{14}$); 101.6 ($C_1$); 77.4 ($C_3$); 76.6 ($C_5$); 73.5 ($C_{C2}$); 73.4 ($C_2$, $C_{C1}$); 70.7 ($C_7$); 69.6 ($C_4$); 64.0 ($C_6$); 53.0, 52.8 ($C_8$, $C_{10}$); 49.9 ($C_{15}$); 39.3, 36.3 ($C_{C4}$, $C_{C5}$); 30.3, 29.1 ($C_{C3,}C_{C6}$)
**MS** (HRMS) calculated for $[C_{32}H_{43}N_5O_9Na]^+$: 664.2958; found: 664.3009

**ISh011**




Amine purchased from Sigma-Aldrich. Reaction performed in THF in 4 hours.
Yield: 54% over two steps
**$^{1}$H NMR** (400 MHz, $D_2O$): δ = 8.37 (br s, 4H, $H_{12}$, $H_{13}$), 7.71 (d, 2H, $H_{14}$, $J_{14\text{-}13}$ = $J_{14\text{-}15}$ = 7.6 Hz), 7.39 (br s, 2H, $H_{15}$), 5.02 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.6 Hz), 4.43 – 4.24 (m, 4H, $H_{10}$), 4.10 – 4.05 (m, 1H, $C_2$), 4.04 – 4.00 (m, 1H, $H_2$), 3.91 – 3.85 (m, 7H, $H_{6b}$, $H_{7a,b}$, $H_3$, $H_4$, $H_5$, $C_1$), 3.48 (t, 2H, $H_{8a,b}$, $J$ = 4.9 Hz), 2.94 – 2.83 (m, 2H, $C_4$, $C_5$), 2.10 – 1.85 (m, 4H, $C_3$, $C_6$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.4, 177.1 ($C_9$); 147.2 ($C_{15}$); 147.0 ($C_{13}$); 145.9 ($C_{11}$); 140.1 ($C_{14}$); 127.2 ($C_{15}$); 101.9 ($C_1$); 76.9 ($C_3$); 75.9 ($C_5$); 74.0 ($C_{C1}$); 73.8 ($C_2$); 72.5 ($C_{C2}$); 71.8 ($C_7$); 69.0 ($C_4$); 63.3 ($C_6$); 52.8 ($C_8$); 43.8, 43.6 ($C_{C4}$, $C_{C5}$); 42.7 ($C_{10}$); 30.2, 30.0 ($C_{C3,}C_{C6}$)
**MS** (MALDI-TOF) calculated for $[C_{28}H_{37}N_7O_9Na]^+$: 638.254; found: 638.508

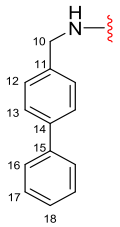
**ISh012**




Prepared by Norbert Varga
$[\alpha]_D$: +5.3° (c = 0.48, MeOH)

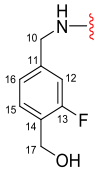
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.40 (d, 4H, $H_{12}$, $J_{12\text{-}13}$ = 8.1 Hz), 7.18 (d, 4H, $H_{13}$, $J_{13\text{-}12}$ = 8.1 Hz), 4.93 (br s, 1H, $H_1$), 4.26 (s, 4H, $H_{10a,b}$), 4.04 – 3.00 (m, 1H, $C_2$), 3.88 – 3.86 (m, 1H, $H_2$), 3.86 – 3.79 (m, 1H, $H_{6a}$), 3.78 – 3.61 (m, 5H, $C_1$, $H_{6b}$, $H_{7a,b}$, $H_3$), 3.59 – 3.50 (m, 2H, $H_4$, $H_5$), 3.43 – 3.32 (m, 2H, $H_{8a,b}$), 2.98 – 2.82 (m, 2H, $C_4$, $C_5$), 2.00 – 1.84 (m, 4H, $C_3$, $C_6$), 1.47 (s, 12H, $H_{16}$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.1, 176.9 ($C_9$); 149.9 ($C_{14}$); 138.2 ($C_{11}$); 128.3, 128.2 ($C_{12}$); 125.9 ($C_{13}$); 100.4 ($C_1$); 76.6 ($C_3$); 75.7 ($C_5$); 73.0 ($C_{15}$); 72.7 ($C_{C1}$); 72.5 ($C_2$); 72.4 ($C_{C2}$); 69.3 ($C_7$); 68.9 ($C_4$); 63.2 ($C_6$); 52.1 ($C_8$); 43.8 ( $C_{10}$); 42.1, 42.0 ($C_{C4}$, $C_{C5}$); 32.1 ($C_{16}$); 29.9, 29.0 ($C_{C3,}$ $C_{C6}$)
**MS** (HRMS) calculated for $[C_{34}H_{43}N_5O_{13}Na]^+$: 752.34828; found: 752.34702

**ISh013**




Amine purchased from Sigma-Aldrich. Reaction performed in THF overnight.
Yield: 59% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.60 (t, 8H, $H_{13}$, $H_{16}$, $J_{13\text{-}12}$ = $J_{16\text{-}17}$ = 7.4 Hz), 7.48 (t, 4H, $H_{12}$, $J_{12\text{-}13}$ = 7.4 Hz), 7.40 (d, 6H, $H_{17}$, $H_{18}$, $J_{17\text{-}18}$ = 7.8 Hz), 4.98 – 4.96 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.3 Hz), 4.40 (dd, 2H, $H_{10a}$, $J_{10a\text{-}NH}$ = 4.0 Hz, $J_{10a\text{-}10b}$ = 15.3 Hz), 4.30 (dd, 2H, $H_{10b}$, $J_{10b\text{-}NH}$ < 1 Hz, $J_{10b\text{-}10a}$ = 15.3 Hz), 4.09 – 4.05 (m, 1H, $C_2$), 3.94 – 3.90 (m, 1H, $H_2$), 3.90 – 3.84 (m, 1H, $H_{6a}$), 3.84 – 3.63 (m, 5H, $C_1$, $H_{6b}$, $H_{7a,b}$, $H_3$), 3.62 – 3.56 (m, 2H, $H_4$, $H_5$), 3.47 – 3.37 (m, 2H, $H_{8a,b}$), 3.05 – 2.90 (m, 2H, $C_4$, $C_5$), 2.09 – 1.91 (m, 4H, $C_3$, $C_6$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.1, 177.9 ($C_9$); 143.1, 142.2 ($C_{14}$, $C_{15}$); 140.1 ($C_{11}$); 130.8 ($C_{12}$); 129.8 ($C_{17}$, $C_{18}$), 129.2, 129.0, 128.8 ($C_{13}$, $C_{16}$), 101.2 ($C_1$); 77.4 ($C_3$); 76.4 ($C_5$); 73.5 ($C_{C1}$); 73.3 ($C_2$); 73.3 ($C_{C2}$); 70.1 ($C_7$); 69.7 ($C_4$); 64.0 ($C_6$); 53.0 ($C_8$); 44.6 ( $C_{10}$); 42.9, 42.8 ($C_{C4}$, $C_{C5}$); 30.7, 29.9 ($C_{C3,}$ $C_{C6}$).
**MS** (MALDI-TOF) calculated for $[C_{42}H_{47}N_5O_9Na]^+$: 788.327; found: 788.689

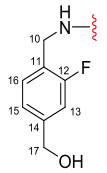
**ISh014**




Prepared by Norbert Varga
$[\alpha]_D$: +6.9° (c = 0.33, MeOH)
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.22 (t, 2H, $H_{15}$, $J_{15\text{-}16}$ = 7.7 Hz, $J_{15\text{-}F}$ = 7.7. Hz), 7.12 (d, 2H, $H_{16}$, $J_{16\text{-}15}$ = 7.7 Hz), 7.04 (d, 2H, $H_{12}$, $J_{12\text{-}F}$ = 11.5 Hz), 5.04 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.6 Hz), 4.69 (s, 4H, $H_{17a,b}$), 4.44 – 4.33 (m, 4H, $H_{10a,b}$), 4.15 – 4.11 (m, 1H, $C_2$), 3.99 (dd, 1H, $H_2$, $J_{2\text{-}1}$ = 1.6 Hz, $J_{2\text{-}3}$ = 3.1 Hz), 3.96 – 3.91 (m, 1H, $H_{6a}$), 3.90 – 3.73 (m, 5H, $C_1$, $H_{6b}$, $H_{7a,b}$, $H_3$), 3.73 – 3.62 (m, 2H, $H_4$, $H_5$), 3.53 – 3.41 (m, 2H, $H_{8a,b}$), 3.08 – 2.93 (m, 2H, $C_4$, $C_5$), 2.13 – 1.96 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.3, 177.1 ($C_9$); 162.0 (d, $C_{13}$, $^1J_{13\text{-}F}$ = 245.0 Hz); 142.1 (d, $C_{11}$, $^3J_{11\text{-}F}$ = 7.3 Hz); 130.6 (d, $C_{15}$, $^3J_{15\text{-}F}$ = 5.0 Hz); 128.4, 128.3 (d, $C_{14}$, $^2J_{14\text{-}F}$ = 15.2 Hz), 124.2, 124.1 (d, $C_{16}$, $^4J_{16\text{-}F}$ = 4.7 Hz), 115.1, 115.0 (d, $C_{12}$, $J_{12\text{-}F}$ = 22.5 Hz); 100.5 ($C_1$); 76.6 ($C_3$); 75.7 ($C_5$); 72.8 ($C_{C1}$); 72.6 ($C_2$); 72.5 ($C_{C2}$); 69.3 ($C_7$); 69.0 ($C_4$); 63.2 ($C_6$); 58.8 (d, $C_{17}$, $^3J_{17\text{-}F}$ = 4.3 Hz); 52.2 ($C_8$); 43.4 ( $C_{10}$); 42.1, 42.0 ($C_{C4}$, $C_{C5}$); 29.9, 29.0 ($C_{C3}$, $C_{C6}$)
**MS** (HRMS) calculated for $[C_{32}H_{41}F_2N_5O_{11}Na]^+$: 732.26683; found: 732.26529

**ISh015**

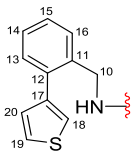



Prepared by Norbert Varga
$[\alpha]_D$: +9.3° (c = 0.44, MeOH)
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.22 (t, 2H, $H_{16}$, $J_{16\text{-}15}$ = 7.7 Hz, $J_{16\text{-}F}$ = 7.7. Hz), 7.05 (d, 2H, $H_{13}$, $J_{13\text{-}F}$ = 11.5 Hz), 7.04 (d, 2H, $H_{15}$, $J_{15\text{-}16}$ = 7.7 Hz), 4.94 (br s, 1H, $H_1$), 4.55 (s, 4H, $H_{17a,b}$), 4.37 – 4.26 (m, 4H, $H_{10a,b}$), 4.04 – 4.00 (m, 1H, $C_2$), 3.89 (dd, 1H, $H_2$, $J_{2\text{-}1}$ = 1.6 Hz, $J_{2\text{-}3}$ = 3.1 Hz), 3.87 – 3.81 (m, 1H, $H_{6a}$), 3.81 – 3.63 (m, 5H, $C_1$, $H_{6b}$, $H_{7a,b}$, $H_3$), 3.62 – 3.51 (m, 2H, $H_4$, $H_5$), 3.44 – 3.31 (m, 2H, $H_{8a,b}$), 2.99 – 2.83 (m, 2H, $C_4$, $C_5$), 2.01 – 1.86 (m, 4H, $C_3$, $C_6$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.3, 177.1 ($C_9$); 162.1 (d, $C_{12}$, $^1J_{12\text{-}F}$ = 250.0 Hz); 144.8, 144.7 (d, $C_{14}$, $^3J_{14\text{-}F}$ = 4.3 Hz); 130.6, 130.5 (d, $C_{16}$, $^3J_{16\text{-}F}$ = 4.5 Hz); 125.5 (d, $C_{11}$, $^2J_{11\text{-}F}$ = 13.8 Hz), 123.5 (d, $C_{15}$, $^4J_{15\text{-}F}$ = 3.1 Hz), 114.3 (d, $C_{13}$, $^2J_{13\text{-}F}$ = 22.1 Hz); 100.4 ($C_1$); 76.6 ($C_3$); 75.7 ($C_5$); 72.7 ($C_{C1}$); 72.5 ($C_2$); 72.5 ($C_{C2}$); 69.3 ($C_7$); 68.9 ($C_4$); 64.3 ($C_{17}$); 63.1 ($C_6$); 52.1 ($C_8$); 42.0, 41.8 ($C_{C4}$, $C_{C5}$); 37.7 (d, $C_{10}$, $^3J_{10\text{-}F}$ = 4.6 Hz); 29.8, 29.9 ($C_{C3}$, $C_{C6}$)
**MS** (HRMS) calculated for $[C_{32}H_{41}F_2N_5O_{11}Na]^+$: 732.26683; found: 732.26691

**ISh016**


Prepared by Alessio Caramiello
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 4.98 (d, 1H, $H_1$), 4.10 - 4.06 (m, 1H, $C_2$), 3.94 (s, 1H, $H_2$), 3.87 (s, 2H, $H_{6a,b}$), 3.83 - 3.64 (m, 6H, $C_1$, $H_{7a,b}$, $H_3$, $H_4$, $H_5$, $H_{6b}$), 3.46 – 3.39 (m, 2H, $H_{8a,b}$), 3.20 – 3.04 (m, 4H, $H_{10}$), 2.91 – 2.79 (m, 2H, $C_4$, $C_5$), 2.05 – 1.84 (m, 4H, $C_3$, $C_6$), 1.57 – 1.46 (sex, 4H, $H_{11}$, $J_{11\text{-}12}$ = 7.3 Hz), 0.97 – 0.89 (t, 6H, $H_{12}$, $J_{12\text{-}11}$ = 7.3 Hz)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 176.9 ($C_9$); 101.3 ($C_1$); 77.4 ($C_3$); 76.5 ($C_5$); 73.4 ($C_2$); 73.2 ($C_{C1}$, $C_{C2}$); 70.1 ($C_7$); 69.8 ($C_4$); 63.9 ($C_6$); 53.0 ($C_8$); 33.2 ($C_{10}$); 33.0 ($C_{C4}$, $C_{C5}$); 30.7, 29.8 ($C_{C3}$, $C_{C6}$); 24.5 ($C_{11}$); 12.7 ($C_{12}$)
**MS** (MALDI-TOF) calculated for $[C_{22}H_{39}N_5O_9Na]^+$: 540.264; found: 540.585

**ISh017**



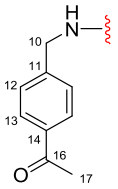

Prepared by Alessio Caramiello

**$^{1}$H NMR** (400 MHz, $CD_3OD$): δ = 8.04 – 7.97 (m, 4H, $H_{16}$, $H_{17}$), 7.63 – 7.15 (m, 10H, $H_{12}$, $H_{13}$, $H_{18}$), 4.93 – 4.86 (m, 1H, $H_1$), 4.42 – 4.27 (dd, 4H, $H_{10a,b}$), 4.05 (s, 1H, $C_2$), 3.94 (s, 1H, $H_2$), 3.85 (m, 2H, $H_{6a,b}$), 3.78 - 3.65 (m, 4H, $H_3$, $C_1$, $H_{7a,b}$), 3.60 – 3.50 (m, 2H, $H_4$, $H_5$), 3.40 – 3.35 (m, 2H, $H_{8a,b}$), 2.94 – 2.79 (m, 2H, $C_4$, $C_5$), 1.93 – 1.80 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.8 ($C_9$); 143.2 ($C_{15}$); 138.4 ($C_{14}$); 137.8 ($C_{11}$); 130.7, 130.0, 129.8, 129.5, 129.0, 127.5, 124.9 ($C_{12}$, $C_{13}$, $C_{16}$, $C_{17}$, $C_{18}$); 101.0 ($C_1$); 77.3 ($C_3$); 76.3 ($C_5$); 73.4 ($C_{C1}$); 73.2 ($C_2$); 73.0 ($C_{C2}$); 70.0 ($C_7$); 69.6 ($C_4$); 63.8 ($C_6$); 52.8 ($C_8$); 43.2 ($C_{10}$); 42.5 ($C_{C4}$, $C_{C5}$); 30.4, 29.6 ($C_{C3}$, $C_{C6}$)

**MS** (ESI) calculated for $[C_{38}H_{43}N_5O_9S_2Na]^+$: 800.239; found: 800.774

**ISh018**




Prepared by Norbert Varga

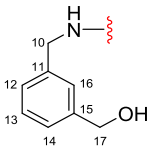
$[\alpha]_D$: -42.8° (c = 0.1, MeOH)

**$^{1}$H NMR** (400 MHz, $CD_3OD$): δ = 7.82 (d, 4H, $H_{13}$, $J_{13\text{-}12}$ = 7.3 Hz), 7.33 (d, 4H, $H_{12}$, $J_{12\text{-}13}$ = 8.1 Hz), 4.98 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.6 Hz), 4.54 – 4.26 (m, 4H, $H_{10a,b}$), 4.10 - 4.05 (m, 1H, $C_2$), 3.94 – 3.90 (m, 1H, $H_2$), 3.90 – 3.84 (m, 1H, $H_{6a}$), 3.84 – 3.66 (m, 5H, $H_{6b}$, $C_1$, $H_{7a,b}$, $H_3$), 3.64 – 3.57 (m, 2H, $H_4$, $H_5$), 3.47 – 3.35 (m, 2H, $H_{8a,b}$), 3.08 – 2.92 (m, 2H, $C_4$, $C_5$), 2.52 (s, 6H, $H_{16}$), 2.08 – 1.92 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 200.2 ($C_{15}$); 177.5, 177.3 ($C_9$); 146.2 ($C_{14}$); 137.1 ($C_{11}$); 129.8 ($C_{13}$); 128.3 ($C_{12}$); 100.5 ($C_1$); 76.6 ($C_3$); 75.7 ($C_5$); 72.8 ($C_{C1}$); 72.6 ($C_2$); 72.6 ($C_{C2}$); 69.4 ($C_7$); 68.9 ($C_4$); 63.2 ($C_6$); 52.2 ($C_8$); 43.6 ($C_{10}$); 42.0, 41.9 ($C_{C4}$, $C_{C5}$); 30.0, 29.2 ($C_{C3}$, $C_{C6}$); 26.8 ($C_{16}$)

**MS** (HRMS) calculated for $[C_{34}H_{43}N_5O_{11}Na]^+$: 720.28568; found: 720.28552

**ISh019**

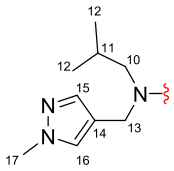



Yield: 59% over two steps

$[\alpha]_D$: +21.5° (c = 0.33, MeOH)

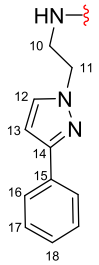
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.35 – 7.18 (m, 8H, $H_{12}$, $H_{13}$, $H_{14}$, $H_{16}$), 5.01 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.6 Hz), 4.61 (m, 4H, $H_{17a,b}$), 4.41 – 4.31 (m, 4H, $H_{10a,b}$), 4.13 - 4.08 (m, 1H, $C_2$), 3.96 (dd, 1H, $H_2$, $J_{2\text{-}1}$ = 1.6 Hz, $J_{2\text{-}3}$ = 3.2 Hz), 3.94 – 3.89 (m, 1H, $H_{6a}$), 3.88 – 3.71 (m, 5H, $H_{6b}$, $C_1$, $H_{7a,b}$, $H_3$), 3.71 – 3.59 (m, 2H, $H_4$, $H_5$), 3.51 – 3.38 (m, 2H, $H_{8a,b}$), 3.06 – 2.92 (m, 2H, $C_4$, $C_5$), 2.09 – 1.95 (m, 4H, $C_3$, $C_6$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.2, 177.0 ($C_9$); 143.1 ($C_{15}$); 140.3 ($C_{11}$); 129.7 ($C_{16}$); 127.7, 127.1, 126.8 ($C_{12}$, $C_{13}$, $C_{14}$); 100.5 ($C_1$); 76.7 ($C_3$); 75.7 ($C_5$); 72.8 ($C_{C1}$); 72.6 ($C_2$, $C_{C2}$); 69.3 ($C_7$); 69.0 ($C_4$); 65.3 ($C_{17}$); 63.2 ($C_6$); 52.2 ($C_8$); 43.6 ($C_{10}$); 42.1, 41.9 ($C_{C4}$, $C_{C5}$); 29.9, 29.1 ($C_{C3}$, $C_{C6}$)
**MS** (HRMS) calculated for $[C_{32}H_{43}N_5O_{11}Na]^+$: 696.28568; found: 696.28578

**ISh020**




Amine purchased from Chem Bridge. Reaction performed in DMF in 2 days, with the addition of 3.2 equiv. $Et_3N$. Yield: 56% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.75 – 7.27 (m, 4H, $H_{15}$, $H_{16}$), 4.91 (s, 1H, $H_1$), 4.78 – 4.21 (m, 8H, $H_{10}$, $H_{13}$), 4.08 – 3.98 (m, 1H, $C_2$), 3.91 – 2.83 (m, 19H, $H_2$, $H_{6a,b}$, $C_1$, $H_{7a,b}$, $H_3$, $H_4$, $H_5$, $H_{8a,b}$, $H_{17}$, $C_4$, $C_5$), 2.18 – 1.73 (m, 6H, $C_3$, $C_6$, $H_{11}$), 1.01 – 0.92 (m, 6H, $H_{12a}$), 0.88 – 0.77 (m, 6H, $H_{12b}$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.1, 177.7 ($C_9$); 140.7, 140.0 ($C_{15}$); 132.5, 132.3 ($C_{16}$); 120.3, 120.2 ($C_{14}$); 101.9 ($C_1$); 77.4 ($C_3$); 76.6 ($C_5$); 73.9 ($C_{C2}$); 73.6 ($C_2$); 73.4 ($C_{C1}$); 70.6 ($C_7$); 69.7 ($C_4$); 64.0 ($C_6$); 57.2, 54.4 ($C_{C4}$, $C_{C5}$); 53.1 ($C_8$); 44.3, 44.2, 43.1, 42.9 ($C_{10}$, $C_{13}$); 39.9, 39.7, 38.6, 39.1 ($C_{17}$); 31.2, 30.4 ($C_3$, $C_6$); 30.1, 28.8 ($C_{11}$); 21.7, 21.5, 21.4, 21.2 ($C_{12}$)
**MS** (MALDI-TOF) calculated for $[C_{34}H_{55}N_9O_9Na]^+$: 756.401; found: 756.741

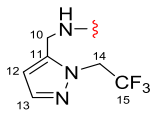
**ISh021**




Amine purchased from Vitas M Labs. Reaction performed in DMF overnight, with the addition of 3.2 equiv. $Et_3N$. Yield: 62% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.76 – 7.72 (m, 4H, $H_{16}$), 7.59 (dd, 2H, $H_{12}$, $J_{12\text{-}13}$ = 2.2 Hz, $J$ = 5.3 Hz), 7.39 – 7.33 (m, 4H, $H_{17}$), 7.29 – 7.23 (m, 2H, $H_{18}$), 6.60 (t, 2H, $H_{13}$, $J_{13\text{-}12}$ = 2.2 Hz), 4.93 – 4.91 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.6 Hz), 4.30 – 4.13 (m, 4H, $H_{10}$), 3.93 – 3.88 (m, 1H, $C_2$), 3.87 – 3.77 (m, 2H, $H_2$, $H_{6a}$), 3.71 – 3.42 (m, 11H, $H_3$, $H_4$, $H_5$, $H_{7a,b}$, $H_{11}$, $C_1$, $H_{6a}$), 3.34 – 3.21 ($H_{8a,b}$), 2.82 – 2.68 (m, 2H, $C_4$, $C_5$), 1.95 – 1.68 (m, 4H, $C_3$, $C_6$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.5, 178.3 ($C_9$); 154.2, 154.1 ($C_{15}$); 153.6, 153.6 ($C_{14}$); 134.4, 134.2 ($C_{12}$); 130.5 ($C_{17}$); 129.6 ($C_{18}$); 127.5 ($C_{16}$); 104.9, 104.8 ($C_{13}$); 101.1 ($C_1$); 77.3 ($C_3$); 76.3 ($C_5$); 73.4 ($C_2$); 73.3

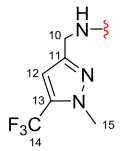
($C_{C2}$); 73.2 ($C_{C1}$); 69.9 ($C_7$); 69.6 ($C_4$); 63.8 ($C_6$); 52.8, 52.8 ($C_8$, $C_{10}$); 42.7, 42.1 ($C_{C4}$, $C_{C5}$); 41.7 ($C_{11}$); 30.6, 30.0 ($C_{C3}$, $C_{C6}$)
**MS** (MALDI-TOF) calculated for $[C_{38}H_{47}N_9O_9Na]^+$: 796.339; found: 796.710

**ISh022**




Amine purchased from Crea-Chim. Reaction performed in DMF overnight, with the addition of 3.2 equiv. $Et_3N$.
Yield: 73% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.41 (t, 2H, $H_{13}$, $J_{13\text{-}12}$ = 1.7 Hz), 6.26 (d, 2H, $H_{12}$, $J_{13\text{-}12}$ = 1.7 Hz), 5.03 – 4.89 (m, 5H, $H_1$, $H_{14}$), 4.36 – 4.30 (m, 4H, $H_{10}$), 4.03 – 3.98 (m, 1H, $C_2$), 3.88 – 3.85 (m, 1H, $H_2$), 3.85 – 3.79 (m, 1H, $H_{6a}$), 3.79 – 3.49 (m, 5H, $H_3$, $H_{6b}$, $H_{7a,b}$, $C_1$), 3.42 – 3.31 (m, 4H, $H_4$, $H_5$, $H_{8a,b}$), 2.92 – 2.78 (m, 2H, $C_4$, $C_5$), 1.98 – 1.81 (m, 4H, $C_3$, $C_6$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.1, 177.9 ($C_9$); 144.4 ($C_{11}$); 142.1 ($C_{13}$); 125.8 ($C_{15}$, $^1J_{15\text{-}F}$ = 279.3 Hz); 108.9 ($C_{12}$); 101.2 ($C_1$); 77.2 ($C_3$); 76.4 ($C_5$); 73.5 ($C_{C2}$); 73.3 ($C_2$); 73.2 ($C_{C1}$); 70.1 ($C_7$); 69.6 ($C_4$); 63.9 ($C_6$); 52.9 ($C_8$); 51.8, 51.5, 51.1 ($C_{14}$); 42.4, 42.3 ($C_{C4}$, $C_{C5}$); 35.3 ($C_{10}$); 30.5, 29.7 ($C_{C3}$, $C_{C6}$)
**MS** (MALDI-TOF) calculated for $[C_{28}H_{37}F_6N_9O_9Na]^+$: 780.251; found: 780.600

**ISh023**



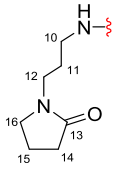

Amine purchased from Vitas M Labs. Reaction performed in DMF in 2 days, with the addition of 3.2 equiv. $Et_3N$.
Yield: 65% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 6.57 (s, 2H, $H_{12}$), 4.95 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.4 Hz), 4.34 – 4.19 (m, 4H, $H_{10}$), 4.05 – 4.01 (m, 1H, $C_2$), 3.91 (s, 6H, $H_{15}$), 3.89 – 3.52 (m, 7H, $H_2$, $H_3$, $H_{6a,b}$, $H_{7a,b}$, $C_1$), 3.44 – 3.52 (m, 4H, $H_4$, $H_5$, $H_{8a,b}$), 2.97 – 2.82 (m, 2H, $C_4$, $C_5$), 2.02 – 1.86 (m, 4H, $C_3$, $C_6$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.1, 177.9 ($C_9$); 151.6 ($C_{13}$); 134.2 ($C_{11}$); 122.3 ($C_{14}$, $^1J_{14\text{-}F}$ = 269.1 Hz); 108.1 ($C_{12}$); 101.1 ($C_1$); 77.3 ($C_3$); 76.4 ($C_5$); 73.5 ($C_{C1}$); 73.3 ($C_2$); 73.1 ($C_{C2}$); 70.1 ($C_7$); 69.6 ($C_4$); 63.9 ($C_6$); 52.9 ($C_8$); 42.6, 42.5 ($C_{C4}$, $C_{C5}$); 39.1 ($C_{15}$); 38.3 ($C_{10}$); 30.5, 29.7 ($C_{C3}$, $C_{C6}$)
**MS** (MALDI-TOF) calculated for $[C_{28}H_{37}F_6N_9O_9Na]^+$: 780.251; found: 780.604

**ISh024**


Amine purchased from Sigma-Aldrich. Reaction performed in THF in 1 hour.
Yield: 89% over two steps

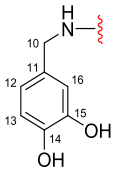
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 4.92 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.6 Hz), 4.01 – 3.97 (m, 1H, $C_2$), 3.89 – 3.86 (m, 1H, $H_2$), 3.86 – 3.80 (m, 1H, $H_{6a}$), 3.80 – 3.50 (m, 5H, $C_1$, $H_{6b}$, $H_3$, $H_4$, $H_5$, $H_{7a,b}$), 3.43 – 3.31 (m, 2H, $H_{8a,b}$), 2.79 – 2.64 (m, 2H, $C_4$, $C_5$), 2.61 – 2.52 (sept d, 2H, $H_{10}$, $J$ = 3.7 Hz, $J$ = 1.0 Hz), 1.94 – 1.76 (m, 4H, $C_3$, $C_6$), 0.71 – 0.63 (m, 4H, $H_{11a}$), 0.47 – 0.40 (m, 4H, $H_{11b}$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 179.5, 179.3 ($C_9$); 101.2 ($C_1$); 77.3 ($C_3$); 76.4 ($C_5$); 73.5 ($C_{C2}$); 73.3 ($C_2$); 73.2 ($C_{C1}$); 70.0 ($C_7$); 69.6 ($C_4$); 63.9 ($C_6$); 52.9 ($C_8$); 42.5, 42.4 ($C_{C4}$, $C_{C5}$); 30.3, 29.5 ($C_{C3}$, $C_{C6}$); 24.1 ($C_{10}$), 7.4, 7.3 ($C_{11}$)
**MS** (MALDI-TOF) calculated for $[C_{22}H_{35}N_5O_9Na]^+$: 536.233; found: 536.435

**ISh025**



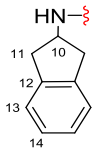

Amine purchased from Sigma-Aldrich. Reaction performed in THF in 1 hour.
Yield: 87% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 4.94 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.4 Hz), 4.04 – 3.99 (m, 1H, $C_2$), 3.92 – 3.88 (m, 1H, $H_2$), 3.87 – 3.81 (m, 1H, $H_{6a}$), 3.81 – 3.62 (m, 5H, $C_1$, $H_{6b}$, $H_3$, $H_{7a,b}$), 3.61 – 3.53 (m, 2H, $H_4$, $H_5$), 3.45 (t, 4H, $H_{16}$, $J_{16\text{-}15}$ = 7.1 Hz), 3.41 – 3.35 (m, 2H, $H_{8a,b}$), 3.29 – 3.22 (m, 4H, $H_{12}$), 3.11 (t, 4H, $H_{10}$, $J_{10\text{-}11}$ = 6.8 Hz), 2.87 – 2.72 (m, 2H, $C_4$, $C_5$), 2.38 (t, 4H, $H_{14}$, $J_{14\text{-}15}$ = 8 Hz), 2.04 (q, 4H, $H_{15}$), 1.98 – 1.85 (m, 4H, $C_3$, $C_6$), 1.72 – 1.62 (m, 4H, $H_{11}$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.7 ($C_9$); 178.0, 177.8 ($C_{13}$), 101.2 ($C_1$); 77.3 ($C_3$); 76.4 ($C_5$); 73.4 ($C_{C1}$); 73.3 ($C_2$, $C_{C1}$); 70.1 ($C_7$); 69.7 ($C_4$); 63.9 ($C_6$); 52.9 ($C_8$); 49.6 ($C_{16}$); 42.8, 42.7 ($C_{C4}$, $C_{C5}$); 42.1 ($C_{12}$); 38.5 ($C_{10}$); 32.9 ($C_{14}$); 30.7, 29.8 ($C_{C3}$, $C_{C6}$); 28.8 ($C_{11}$); 19.3 ($C_{15}$)
**MS** (MALDI-TOF) calculated for $[C_{30}H_{49}N_7O_{11}Na]^+$: 706.338; found: 706.646

**ISh026**



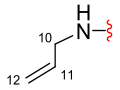

Amine purchased from Sigma-Aldrich. Reaction performed in DMF overnight, with the addition of 3.2 equiv. $Et_3N$.
Yield: 70% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 8.09 – 7.91 (m, 2H, $H_{16}$), 7.45 (q, 2H, $H_{13}$, $J_{13\text{-}12}$ = 8.2 Hz), 6.75 – 6.61 (m, 2H, $H_{12}$), 5.00 – 4.65 (m, 5H, $H_1$, $H_{10}$), 4.23 – 4.13 (m, 1H, $C_2$), 4.13 – 3.06 (m, 11H, $H_2$, $H_{6a,b}$, $C_1$, $H_3$, $H_{7a,b}$, $H_4$, $H_5$, $H_{8a,b}$), 3.06 – 2.73 (m, 2H, $C_4$, $C_5$), 2.10 – 1.71 (m, 4H, $C_3$, $C_6$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.4 ($C_9$); 164.9, 164.0 ($C_{14}$, $C_{15}$); 141.0 ($C_{11}$); 131.6, 130.6 ($C_{16}$, $C_{13}$); 116.7 ($C_{12}$); 100.3 ($C_1$); 76.6 ($C_5$); 76.3 ($C_3$); 73.5 ($C_2$); 73.3 ($C_{C2}$); 73.2 ($C_{C1}$); 69.7 ($C_7$); 69.1 ($C_4$); 63.5 ($C_6$); 52.1 ($C_8$); 45.1, 44.2 ($C_{10}$); 42.5 ($C_{C4}$); 42.2 ($C_{C5}$); 31.4, 29.9 ($C_{C3}$, $C_{C6}$)
**MS** (MALDI-TOF) calculated for $[C_{30}H_{39}N_5O_{13}Na]^+$: 700.244; found: 700.559
Unstable – slow decomposition in solution.

**ISh027**



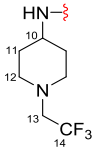

Amine purchased from Sigma-Aldrich. Reaction performed in THF in 1 hour.
Yield: 79% over two steps
**$^{1}$H NMR** (400 MHz, $CD_3OD$): δ = 7.22 – 7.15 (m, 4H, $H_{14}$), 7.15 – 7.09 (m, 4H, $H_{13}$), 4.91 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.5 Hz), 4.53 – 4.44 (m, 2H, $H_{10}$), 4.02 – 3.98 (m, 1H, $C_2$), 3.88 – 3.85 (m, 1H, $H_2$), 3.86 – 3.79 (m, 1H, $H_{6a}$), 3.79 – 3.60 (m, 5H, $C_1$, $H_{6b}$, $H_3$, $H_{7a,b}$), 3.59 – 3.50 (m, 2H, $H_4$, $H_5$), 3.42 – 3.31 (m, 2H, $H_{8a,b}$), 3.22 (quint, 4H, $H_{11a}$, $J$ = 7.8 Hz), 2.86 – 2.71 (m, 6H, $C_4$, $C_5$, $H_{11b}$), 1.96 – 1.79 (m, 4H, $C_3$, $C_6$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.9, 177.7 ($C_9$); 143.1, 143.0 ($C_{12}$); 128.6 ($C_{13}$); 126.4 ($C_{14}$); 101.1 ($C_1$); 77.3 ($C_3$); 76.4 ($C_5$); 73.4 ($C_{C2}$); 73.2 ($C_2$); 73.1 ($C_{C1}$); 70.0 ($C_7$); 69.6 ($C_4$); 63.9 ($C_6$); 52.8 ($C_8$); 52.7 ($C_{10}$); 42.7, 42.5 ($C_{C4}$, $C_{C5}$); 41.3, 41.2, 41.1 ($C_{11}$); 30.3, 29.4 ($C_{C3}$, $C_{C6}$)
**MS** (MALDI-TOF) calculated for $[C_{34}H_{43}N_5O_9Na]^+$: 688.295; found: 688.601

**ISh028**




Amine purchased from Sigma-Aldrich. Reaction performed in THF in 1 hour.
Yield: 90% over two steps
**$^{1}$H NMR** (400 MHz, $CD_3OD$): δ = 5.84 - 5.71 (app oct, 2H, $H_{11}$, $J$ = 5.4 Hz), 5.17 (dd, 2H, $H_{12a}$, $J$ = 17.1 Hz, $J$ = 1.5 Hz), 5.08 (dt, 2H, $H_{12b}$, $J$ = 10.3 Hz, $J$ = 1.3 Hz), 4.94 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.4Hz), 4.06 – 3.99 (m, 1H, $C_2$), 3.92 – 3.77 (m, 7H, $H_2$, $H_{6a,b}$, $C_1$, $H_{7a,b}$, $H_3$), 3.77 – 3.63 (m, 2H, $H_{10}$), 3.61 – 3.50 (m, 2H, $H_4$, $H_5$), 3.44 – 3.31 (m, 2H, $H_{8a,b}$), 3.05 – 2.73 (m, 2H, $C_4$, $C_5$), 2.00 – 1.80 (m, 4H, $C_3$, $C_6$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.9, 177.7 ($C_9$); 136.4 ($C_{11}$); 116.8 ($C_{12}$); 101.1 ($C_1$); 77.3 ($C_3$); 76.4 ($C_5$); 73.4 ($C_2$); 73.3, 73.2 ($C_{C1,}$ $C_{C2}$); 70.0 ($C_7$); 69.6 ($C_4$); 63.9 ($C_6$); 52.9 ($C_8$); 43.4 ($C_{10}$); 42.8, 42.7 ($C_{C4}$, $C_{C5}$); 30.6, 29.7 ($C_{C3}$, $C_{C6}$)
**MS** (ESI) calculated for $[C_{22}H_{35}N_5O_9Na]^+$: 536.233; found: 536.457

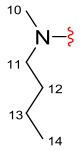
**ISh029**




Amine purchased from Enamine BB. Reaction performed in DMF in two days, with the addition of 3.2 equiv. $Et_3N$.
Yield: 60% over two steps

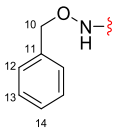
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 4.96 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.5 Hz), 4.05 – 4.01 (m, 1H, $C_2$), 3.94 – 3.91 (m, 1H, $H_2$), 3.89 – 3.83 (m, 1H, $H_{6a}$), 3.83 – 3.64 (m, 5H, $H_3$, $H_{6b}$, $H_{7a,b}$, $C_1$), 3.64 – 3.53 (m, 4H, $H_4$, $H_5$, $H_{10}$), 3.47 – 3.34 (m, 2H, $H_{8a,b}$), 3.05 (q, 4H, $H_{13}$, $J_{13\text{-}12}$ = 9.9 Hz, $J$ = 19.6 Hz), 2.99 – 2.90 (m, 4H, $H_{12a}$), 2.86 – 2.72 (m, 2H, $C_4$, $C_5$), 2.46 (td, 4H, $H_{12b}$, $J_{12\text{-}13}$ = 9.9 Hz, $J_{12\text{-}11}$ = 3.0 Hz), 1.98 - 1. 85 (m, 4H, $C_3$, $C_6$), 1.80 (t, 4H, $H_{11a}$, $J$ = 15.2 Hz), 1.59 – 1.46 (m, 4H, $H_{11b}$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.3, 177.1 ($C_9$); 128.0 ($C_{14}$, $^1J_{14\text{-}F}$ = 277.7 Hz); 101.2 ($C_1$); 77.4 ($C_3$); 76.4 ($C_5$); 73.4 ($C_{C2}$); 73.3 ($C_2$); 73.2 ($C_{C1}$); 70.0 ($C_7$); 69.6 ($C_4$); 63.9 ($C_6$); 59.7 ($C_{13}$, $^2J_{13\text{-}F}$ = 30.6 Hz); 55.0, 54.9 ($C_{12}$); 52.9 ($C_8$); 48.3, 48.2 ($C_{10}$); 42.7, 42.6 ($C_{C4}$, $C_{C5}$); 33.6, 33.5, 33.4, 33.4 ($C_{11}$); 30.6, 29.7 ($C_{C3}$, $C_{C6}$)
**MS** (MALDI-TOF) calculated for $[C_{30}H_{47}F_6N_7O_9Na]^+$: 786.323; found: 786.718

**ISh030**



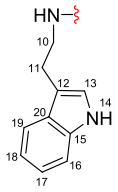

Amine purchased from Sigma-Aldrich. Reaction performed in THF in 2 hours.
Yield: 75% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 4.99 – 4.92 (m, 1H, $H_1$), 4.12 – 4.01 (m, 1H, $C_2$), 3.90 – 3.81 (m, 2H, $H_2$, $H_{6a}$), 3.81 – 3.62 (m, 5H, $C_1$, $H_{6b}$, $H_3$, $H_{7a,b}$), 3.62 – 3.53 (m, 2H, $H_4$, $H_5$), 3.53 – 3.20 (m, 9H, $H_{8a,b}$, $H_{10a}$, $H_{11}$), 3.16 – 3.05 (m, 2H, $C_4$, $C_5$), 2.87 (s, 3H, $H_{10b}$), 1.99 – 1.80 (m, 4H, $C_3$, $C_6$), 1.80 – 1.21 (m, 8H, $H_{12}$, $H_{13}$), 0.94 (t, 6H, $H_{14}$, $J_{14\text{-}13}$ = 7.3 Hz)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.9, 177.5 ($C_9$); 101.4 ($C_1$); 77.5 ($C_3$); 76.6 ($C_5$); 73.8 ($C_2$); 73.5 ($C_{C1}$); 73.3 ($C_{C2}$); 70.6 ($C_7$); 69.6 ($C_4$); 64.0 ($C_6$); 53.0 ($C_8$); 52.0 ($C_{11}$); 39.1 ($C_{10}$); 36.8 ($C_{C4}$); 36.6 ($C_{C5}$); 35.2 ($C_{10}$); 32.8, 31.3 ($C_{12}$); 30.2 ($C_{C3}$); 29.0 ($C_{C6}$); 21.8 ($C_{13}$); 15.1 ($C_{14}$)
**MS** (HRMS) calculated for $[C_{26}H_{47}N_5O_9Na]^+$: 596.3271; found: 596.3281

**ISh031**




Amine purchased from Sigma-Aldrich. Reaction performed in DMF in 4 hours.
Yield: 47% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.43 (d, 4H, $H_{12}$, $J_{12\text{-}13}$ = 7.0 Hz), 7.40 – 7.32 (m, 6H, $H_{13}$, $H_{14}$), 4.92 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.4 Hz), 4.85 – 4.76 (m, 4H, $H_{10}$), 4.04 – 4.00 (m, 1H, $C_2$), 3.90 – 3.83 (m, 2H, $H_2$, $H_{6a}$), 3.79 – 3.49 (m, 7H, $H_3$, $H_4$, $H_5$, $H_{6b}$, $H_{7a,b}$, $C_1$), 3.45 – 3.35 (m, 2H, $H_{8a,b}$), 2.80 – 2.64 (m, 2H, $C_4$, $C_5$), 2.01 – 1.74 (m, 4H, $C_3$, $C_6$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 176.8, 176.6 ($C_9$); 138.01 ($C_{11}$); 131.5, 131.4 ($C_{12}$); 130.6, 130.4 ($C_{13}$, $C_{14}$); 101.3 ($C_1$); 80.1 ($C_{10}$); 77.2 ($C_3$); 76.5 ($C_5$); 73.5 ($C_{C1}$); 73.4 ($C_2$); 73.1 ($C_{C2}$); 70.1 ($C_7$); 69.7 ($C_4$); 64.0 ($C_6$); 52.9 ($C_8$); 39.4, 39.4 ($C_{C4}$, $C_{C5}$); 30.5, 29.7 ($C_{C3}$, $C_{C6}$)
**MS** (MALDI-TOF) calculated for $[C_{30}H_{39}N_5O_{11}Na]^+$: 668.254; found: 668.554

**ISh032**




Amine purchased from Sigma-Aldrich. Reaction performed in DMF in two days, with the addition of 3.2 equiv. $Et_3N$.

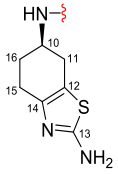
Yield: 55% over two steps

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 8.61 – 8.56 (m, 2H, $H_{14}$), 7.55 (d, 2H, $H_{19}$, $J_{19\text{-}18}$ = 7.7 Hz), 7.32 (d, 2H, $H_{16}$, $J_{16\text{-}17}$ = 7.7 Hz), 7.12 – 7.04 (m, 4H, $H_{13}$, $H_{17}$), 7.02 – 6.95 (m, 2H, $H_{18}$), 4.93 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.4 Hz), 4.02 – 3.98 (m, 1H, $C_2$), 3.92 – 3.84 (m, 2H, $H_2$, $H_{6a}$), 3.78 – 3.53 (m, $C_1$, $H_{6b}$, $H_{7a,b}$, $H_3$, $H_4$, $H_5$), 3.53 – 3.35 (m, 6H, $H_{11}$, $H_{8a,b}$), 2.92 (oct, 4H, $H_{10}$, $J$ = 7.4 Hz), 2.84 – 2.70 (m, 2H, $C_4$, $C_5$), 1.97 – 1.74 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.9 ($C_9$); 139.1 ($C_{15}$); 139.8 ($C_{20}$); 124.5, 123.2 ($C_{13}$, $C_{17}$); 120.5 ($C_{18}$); 120.2 ($C_{19}$); 113.8 ($C_{12}$); 113.2 ($C_{16}$); 101.3 ($C_1$); 77.5 ($C_3$); 76.5 ($C_5$); 73.6 ($C_{C1}$); 73.4 ($C_{C2}$); 73.4 ($C_2$); 70.1 ($C_7$); 69.8 ($C_4$); 64.0 ($C_6$); 53.0 ($C_8$); 43.1, 42.9 ($C_{C4}$, $C_{C5}$); 42.2 ($C_{11}$); 30.4, 29.4 ($C_{C3}$, $C_{C6}$); 27.2 ($C_{10}$)

**MS** (MALDI-TOF) calculated for $[C_{36}H_{45}N_7O_9Na]^+$: 742.317; found: 742.661

**ISh033**




Amine purchased from Key Organics. Reaction performed in DMF in two days, with the addition of 3.2 equiv. $Et_3N$.

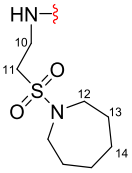
Yield: 25% over two steps

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 4.94 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.4 Hz), 4.13 – 4.04 (m, 2H, $H_{10}$), 4.04 – 4.00 (m, 1H, $C_2$), 3.93 – 3.88 (m, 2H, $H_2$, $H_{6a}$), 3.80 – 3.64 (m, 5H, $H_3$, $H_{6b}$, $H_{7a,b}$, $C_1$), 3.62 – 3.54 (m, 2H, $H_4$, $H_5$), 3.48 – 3.33 (m, 2H, $H_{8a,b}$), 2.91 – 2.76 (m, 4H, $C_4$, $C_5$, $H_{11a}$), 2.72 – 2.65 (m, 8H, $H_{15}$, $H_{16}$), 2.61 – 2.52 (m, 1H, $H_{11b1}$), 2.48 – 2.39 (m, 1H, $H_{11b2}$), 2.00 – 1.90 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.6, 177.5 ($C_9$); 145.5, 145.5 ($C_{12}$, $C_{13}$); 115.6 ($C_{14}$); 101.2 ($C_1$); 77.4 ($C_3$); 76.4 ($C_5$); 73.5 ($C_{C1}$); 73.3 ($C_2$); 73.2 ($C_{C2}$); 70.1 ($C_7$); 69.6 ($C_4$); 63.9 ($C_6$); 52.9 ($C_8$); 47.9, 47.8 ($C_{10}$); 42.7, 42.7 ($C_{C4}$, $C_{C5}$); 41.3 ($C_{15}$, $C_{16}$); 30.5, 30.3 ($C_{C3}$, $C_{C6}$), 26.2 ($C_{11}$)

**MS** (MALDI-TOF) calculated for $[C_{30}H_{43}N_9O_9S_2Na]^+$: 760.252; found: 760.608

**ISh034**



Amine purchased from Alinda Chemicals. Reaction performed in DMF in two days, with the addition of 3.2 equiv. $Et_3N$.

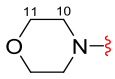
Yield: 40% over two steps (incomplete amide coupling – monoamide methyl ester)

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 4.97 (s, 1H, $H_1$), 4.08 – 4.01 (m, 1H, $C_2$), 3.92 – 3.84 (m, 2H, $H_2$, $H_{6a}$), 3.84 – 3.49 (m, 12H, $C_1$, $H_3$, $H_4$, $H_5$, $H_{6b}$, $H_{7a,b}$, $H_{11}$, $H_{15}$), 3.43 – 3.35 (m, 6H, $H_{8a,b}$, $H_{12}$), 3.27 – 3.12 (m, 2H, $H_{10}$), 3.08 – 2.91 (m, 1H, $C_{4\text{ or }5}$), 2.86 – 2.70 (m, 1H, $C_{5\text{ or }4}$), 2.20 – 1.80 (m, 4H, $C_3$, $C_6$), 1.80 – 1.72 (m, 4H, $H_{13}$), 1.72 – 1.63 (m, 4H, $H_{14}$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$) (monoamide substituent in position C4 or C5): δ = 178.5, 178.3, 177.9, 177.6 ($C_9$); 102.2, 101.4 ($C_1$); 77.4, 77.0 ($C_3$); 76.5 76.5 ($C_5$); 73.5, 73.5 ($C_{C1}$), 73.4, 73.3 ($C_2$), 73.3 73.0 ($C_{C2}$); 70.2, 70.1 ($C_7$); 69.6 ($C_4$); 63.9 ($C_6$); 53.1 ($C_{15}$); 52.9 ($C_8$); 50.7 ($C_{10}$); 49.9 ($C_{12}$); 42.4, 42.2, 41.1, 40.9 ($C_{C4}$, $C_{C5}$); 36.1 ($C_{11}$); 31.6 ($C_{13}$); 30.6, 29.9, 29.8, 28.9 ($C_{C3}$, $C_{C6}$); 28.8 ($C_{14}$)

**MS** (MALDI-TOF) calculated for $[C_{25}H_{43}N_5O_{12}SNa]^+$: 660.252; found: 660.542

**ISh036**




Amine purchased from Sigma-Aldrich. Reaction performed in THF in 2 hours.

Yield: 43% over two steps

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 4.92 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.5Hz), 4.09 – 4.05 (m, 1H, $C_2$); 3.90 – 3.31 (m, 29H, $C_1$, $C_4$, $C_5$, $H_2$, $H_3$, $H_4$, $H_5$, $H_{6a,b}$, $H_{7a,b}$, $H_{8a,b}$, $H_{10}$, $H_{11}$), 1.97 – 1.77 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 176.6, 176.3 ($C_9$); 101.7 ($C_1$); 77.3 ($C_3$); 76.6 ($C_5$); 73.4 ($C_{C1}$); 73.4 ($C_2$); 73.3 ($C_{C2}$); 70.8 ($C_7$); 76.6 ($C_4$); 68.7, 68.6 ($C_{11}$); 63.9 ($C_6$); 53.0 ($C_8$); 44.3, 48.3 ($C_{10}$); 38.4, 38.3 ($C_{C4}$, $C_{C5}$); 30.8, 29.5 ($C_{C3}$, $C_{C6}$)

**MS** (HRMS) calculated for $[C_{24}H_{39}N_5O_{11}Na]^+$: 596.2544; found: 596.2544

**ISh037**


Amine purchased from Sigma-Aldrich. Reaction performed in THF in 2 hours.

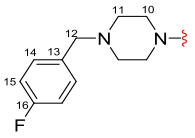
Yield: 80% over two steps

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 4.95 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.5 Hz), 4.09 – 4.05 (m, 1H, $C_2$), 3.86 – 3.81 (m, 2H, $H_2$, $H_{6a}$), 3.78 – 3.31 (m, 17H, $C_1$, $H_3$, $H_4$, $H_5$, $H_{6b}$, $H_{7a,b}$, $H_{8a,b}$, $H_{10}$), 3.28 – 3.10 (m, 2H, $C_4$, $C_5$), 2.02 – 1.78 (m, 12H, $C_3$, $C_6$, $H_{11}$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 176.5, 176.3 ($C_9$); 101.3 ($C_1$); 77.4 ($C_3$); 76.5 ($C_5$); 73.5 ($C_2$); 73.4 ($C_{C1}$); 73.1 ($C_{C2}$); 70.6 ($C_7$); 69.6 ($C_4$); 63.9 ($C_6$); 53.0 ($C_8$); 48.7, 48.7, 48.5, 47.8 ($C_{10}$); 40.7, 40.6 ($C_{C4}$, $C_{C5}$); 29.9, 28.7 ($C_{11}$); 27.8, 27.7 ($C_{C3}$, $C_{C6}$)

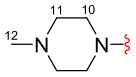
**MS** (MALDI-TOF) calculated for $[C_{24}H_{39}N_5O_9Na]^+$: 564.264; found: 564.546

**ISh038**



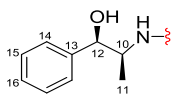
Amine purchased from Sigma-Aldrich. Reaction performed in DMF, overnight.
Yield: 54% over two steps
**$^{1}$H NMR** (400 MHz, $CD_3OD$): δ = 7.44 – 7.33 ($H_{15}$), 7.12 – 7.04 ($H_{14}$), 4.92 (s, 1H, $H_1$), 4.11 – 4.06 (m, 1H, $C_2$), 3.89 – 3.33 (m, 25H, $C_1$, $C_4$, $C_5$, $H_2$, $H_3$, $H_4$, $H_5$, $H_{6a,b}$, $H_{7a,b}$, $H_{8a,b}$, $H_{10}$, $H_{12}$), 2.69 – 2.40 (m, 8H, $H_{11}$), 1.98 – 1.77 (m, 4H, $C_3$, $C_6$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 176.3, 176.1 ($C_9$); 164.7 ($C_{16}$, $^{1}J_{C\text{-}F}$ = 245.9Hz); 134.7, 134.5 ($C_{13}$); 133.3 ($C_{15}$); 117.1, 116.9 ($C_{14}$); 101.6 ($C_1$); 77.3 ($C_3$); 76.6 ($C_5$); 73.4 ($C_{C1}$); 73.3 ($C_2$); 73.3 ($C_{C2}$); 70.7 ($C_7$); 69.6 ($C_4$); 64.0 ($C_6$), 63.4 ($C_{12}$); 54.8, 54.7, 54.5, 54.4 ($C_{11}$); 53.0 ($C_8$); 47.3, 43.4 ($C_{10}$); 38.4, 38.4 ($C_{C4}$, $C_{C5}$); 30.9, 29.6 ($C_{C3}$, $C_{C6}$)
**MS** (MALDI-TOF) calculated for $[C_{38}H_{51}F_2N_7O_9Na]^+$: 810.361; found: 810.815

**ISh039**



Amine purchased from Sigma-Aldrich. Reaction performed in THF in 2 hours.
Yield: 75% over two steps
**$^{1}$H NMR** (400 MHz, $CD_3OD$): δ = 4.95 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.4 Hz), 4.11 – 4.07 (m, 1H, $C_2$); 3.89 – 3.51 (m, 17H, $C_1$, $H_2$, $H_3$, $H_4$, $H_5$, $H_{6a,b}$, $H_{7a,b}$, $H_{10}$), 3.46 – 3.32 (m, 4H, $C_4$, $C_5$, $H_{8a,b}$); 2.94 – 2.59 (m, 8H, $H_{11}$), 2.54 (d, 6H, $H_{12}$, $J$ = 8.8 Hz), 2.05 – 1.77 (m, 4H, $C_3$, $C_6$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 176.5, 176.2 ($C_9$); 101.7 ($C_1$); 77.3 ($C_3$); 76.7 ($C_5$); 73.5 ($C_{C1}$); 73.4 ($C_2$); 73.3 ($C_{C2}$); 70.8 ($C_7$); 69.6 ($C_4$); 64.0 ($C_6$); 56.1, 55.8 ($C_{11}$); 53.0 ($C_8$); 46.4, 46.2 ($C_{10a,b}$); 46.0, 45.8 ($C_{12}$); 42.5, 42.4 ($C_{10c,d}$); 38.5, 38.5 ($C_{C4}$, $C_{C5}$); 30.8, 29.6 ($C_{C3}$, $C_{C6}$)
**MS** (MALDI-TOF) calculated for $[C_{26}H_{45}N_7O_9Na]^+$: 622.317; found: 622.625

**ISh040**



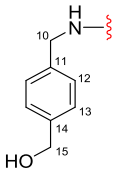
Prepared by Norbert Varga
$[\alpha]_D$: +17.8° (c = 0.5, MeOH)
**$^{1}$H NMR** (400 MHz, $CD_3OD$): δ = 7.36 (d, 4H, $H_{14}$, $J_{14\text{-}15}$ = 7.2 Hz), 7.27 (t, 4H, $H_{15}$, $J_{15\text{-}14}$ = $J_{15\text{-}16}$ = 7.2 Hz), 7.18 (t, 2H, $H_{16}$, $J_{16\text{-}15}$ = 7.2 Hz), 4.96 (br s, 1H, $H_1$), 4.80 (d, 2H, $H_{12}$, $J_{12\text{-}10}$ = 1.7 Hz), 4.12 – 4.00 (m, 3H, $C_2$, $H_{10}$), 3.94 (dd, 1H, $H_2$, $J_{2\text{-}1}$ = 1.6 Hz, $J_{2\text{-}3}$ = 3.1 Hz), 3.90 – 3.80 (m, 2H, $C_1$, $H_{6a}$), 3.82 – 3.63 (m, 5H, $C_1$, $H_3$, $H_{6b}$, $H_{7a,b}$), 3.63 – 3.59 (m, 2H, $H_4$, $H_5$), 3.49 – 3.36 (m, 2H, $H_{8a,b}$), 2.97 – 2.82 (m, 2H, $C_4$, $C_5$), 1.96 – 1.84 (m, 4H, $C_3$, $C_6$), 0.95 (t, 6H, $H_{11}$, $J_{11\text{-}10}$ = 6.4 Hz)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.2, 177.1 ($C_9$); 143.4, 143.3 ($C_{13}$); 129.2 ($C_{15}$); 125.2 ($C_{16}$); 127.4 ($C_{14}$); 100.5 ($C_1$); 76.7 ($C_3$); 76.1, 76.0 ($C_{12}$); 75.7 ($C_5$); 72.8 ($C_{C2}$); 72.5 ($C_2$, $C_{C1}$); 69.4 ($C_7$); 68.9 ($C_4$); 63.2 ($C_6$); 52.5 ($C_8$); 52.2 ($C_{10}$); 42.1, 41.9 ($C_{C4}$, $C_{C5}$); 30.1, 29.1 ($C_{C3}$, $C_{C6}$); 13.4, 13.3 ($C_{11}$)
**MS** (HRMS) calculated for $[C_{34}H_{47}N_5O_{11}Na]^+$: 724.31698; found: 724.31698

**ISh049**

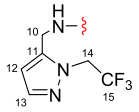

Yield: 77% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 4.34 – 4.27 (m, 1H, $H_1$), 4.00 – 3.93 (m, 1H, $C_2$), 3.80 – 3.71 (m, 2H, $C_1$, $H_{7a}$), 3.69 – 3.55 (m, 9H, $H_{7b}$, $H_4$, $H_5$, $H_{10}$), 3.50 - 3.42 (m, 2H, $H_2$, $H_3$), 3.35 (t, 2H, $H_{8a,b}$, $J_{8\text{-}7}$ = 4.9 Hz), 3.07 (td, 1H, $C_4$, $J$ = 12.5 Hz, $J$ = 3.7 Hz), 2.93 (td, 1H, $C_4$, $J$ = 12.5 Hz, $J$ = 3.7 Hz), 2.17 – 1.06 (m, 2H, $C_{3ax}$, $C_{6ax}$), 1.91 (td, 1H, $C_{6eq}$, $J$ = 13.6 Hz, $J$ = 2.4 Hz), 1.82 (td, 1H, $C_{3eq}$, $J$ = 13.6 Hz, $J$ = 2.4 Hz), 1.25 (d, 1H, $H_6$, $J_{6\text{-}5}$ = 6.4 Hz)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.2, 178.0 ($C_9$); 104.8 ($C_1$); 77.2 ($C_2$); 75.9 ($C_4$); 75.1 ($C_{C2}$); 73.4 ($C_5$); 73.1 ($C_2$); 72.8 ($C_3$); 70.2 ($C_7$); 53.2, 53.1 ($C_{10}$); 52.9 ($C_8$); 41.0 ($C_{C4}$, $C_{C5}$); 29.5, 29.2 ($C_{C3}$, $C_{C6}$); 17.7 ($C_6$)
**MS** (MALDI-TOF) calculated for $[C_{18}H_{29}N_3O_{10}Na]^+$: 470.175; found: 470.134

**ISh050**




Yield: 68% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.25 (dd, 4H, $H_{13a,b}$, $J$ = 8.0 Hz, $J$ = 2.9 Hz), 7.19 (dd, 4H, $H_{12a,b}$, $J$ = 8.0 Hz, $J$ = 2.9 Hz), 4.53 (d, 4H, $H_{15a,b}$, $J$ = 3.2 Hz), 4.31 (s, 1H, $H_1$), 4.32 – 4.26 (m, 4H, $H_{10a,b}$), 4.02 – 3.98 (m, 1H, $C_2$), 3.79 – 3.72 (m, 1H, $H_{7a}$), 3.72 – 3.69 (m, 1H, $H_3$), 3.65 – 3.56 (m, 3H, $H_{7b}$, $H_2$, $C_1$), 3.52 – 3.42 (m, 2H, $H_4$, $H_5$), 3.42 – 3.30 (m, 2H, $H_{8a,b}$), 2.98 – 2.86 (m, 2H, $C_4$, $C_5$), 2.03 – 1.83 (m, 4H, $C_3$, $C_6$), 1.24 (d, 3H, $H_6$, $J_{6\text{-}5}$ = 1.24)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.4, 177.8 ($C_9$); 142.5, 142.4 ($C_{14}$); 139.9, 139.8 ($C_{11}$); 129.3, 129.2, 129.0, 129.0 ($C_{12}$, $C_{13}$); 103.7 ($C_1$); 77.6 ($C_3$); 75.8 ($C_5$); 74.4 ($C_{C2}$); 73.8 ($C_{C1}$); 73.3 ($C_4$); 72.8 ($C_2$); 69.9 ($C_7$); 65.8 ($C_{15}$); 52.9 ($C_8$); 44.6, 44.5 ( $C_{10}$); 42.3, 42.3 ($C_{C4}$, $C_{C5}$); 30.0, 29.9 ($C_{C3}$, $C_{C6}$); 17.6 ($C_6$)
**MS** (HRMS) calculated for $[C_{32}H_{43}N_5O_{10}Na]^+$: 680.2908; found: 680.2918

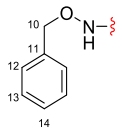
**ISh051**




Amine purchased from Crea-Chim. Reaction performed in THF overnight, with the addition of 3.2 equiv. $Et_3N$.
Yield: 36% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.50 (q, 2H, $H_{13}$, $J_{13\text{-}12}$ = 1.8 Hz), 6.31 (d, 2H, $H_{12}$), 5.05 – 4.89 (m, 4H, $H_{14}$), 4.46 – 4.35 (m, 4H, $H_{10}$), 4.34 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 7.0 Hz), 4.05 – 4.00 (m, 1H, $C_2$), 3.83 – 3.75 (m, 1H, $H_{7a}$), 3.75

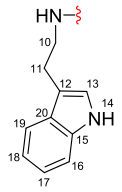
– 3.71 (m, 1H, $C_1$), 3.97 – 3.59 (m, 3H, $H_{7b}$, $H_4$, $H_5$), 3.57 – 3.46 (m, 2H, $H_2$, $H_3$), 3.46 – 3.34 (m, 2H, $H_{88a,b}$), 2.98 – 2.86 (m, 2H, $C_4$, $C_5$), 2.01 – 1.81 (m, 4H, $C_3$, $C_6$), 1.27 (d, 1H, $H_6$, $J_{6\text{-}5}$ = 6.4 Hz)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.6, 178.2 ($C_9$); 144.3 ($C_{11}$); 142.1 ($C_{13}$); 125.8 ($C_{15}$, $^1J_{15\text{-}F}$ = 278.8 Hz); 108.9 ($C_{12}$); 103.8 ($C_1$); 77.5 ($C_3$); 75.9 ($C_5$); 74.4 ($C_{C2}$); 73.9 ($C_{C1}$); 73.3 ($C_4$); 72.9 ($C_2$); 69.9 ($C_7$); 52.9 ($C_8$); 51.8, 51.5 ($C_{14}$); 42.0, 41.9 ($C_{C4}$, $C_{C5}$); 35.4 ($C_{10}$); 30.0, 29.9 ($C_{C3}$, $C_{C6}$); 17.6 ($C_6$)
**MS** (HRMS) calculated for $[C_{28}H_{37}N_9O_8F_6Na]^+$: 764.2567; found: 764.2571

**ISh052**




Amine purchased from Sigma-Aldrich. Reaction performed in DMF overnight, with the addition of 3.2 equiv. $Et_3N$.
Yield: 55% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.45 – 7.29 (m, 10H, $H_{12}$, $H_{13}$, $H_{14}$), 4.88 – 4.73 (m, 4H, $H_{10}$), 4.28 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 7.0 Hz), 4.01 – 3.94 (m, 1H, $C_2$), 3.79 – 3.66 (m, 2H, $H_{7a}$, $C_1$), 3.66 – 3.56 (m, 3H, $H_{7b}$, $H_4$, $H_5$), 3.53 – 3.43 (m, 2H, $H_3$, $H_2$), 3.42 – 3.32 (m, 2H, $H_{8a,b}$), 2.78 – 2.63 (m, 2H, $C_4$, $C_5$), 2.00 – 1.74 (m, 4H, $C_3$, $C_6$), 1.26 (d, 1H, $H_6$, $J_{6\text{-}5}$ = 6.4 Hz)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 174.7, 174.4 ($C_9$); 137.8, 137.7 ($C_{11}$); 131.4 ($C_{12}$); 130.5, 130.4, 130.3 ($C_{13}$, $C_{14}$); 103.6 ($C_1$); 80.1, 80.0 ($C_{10}$); 77.4 ($C_{C1}$); 75.9 ($C_4$); 74.1 ($C_{C2}$); 73.9 ($C_3$); 73.7 ($C_2$); 72.9 ($C_5$); 69.9 ($C_7$); 52.9 ($C_8$); 39.0, 38.9 ($C_{C4}$, $C_{C5}$); 29.8, 29.7 ($C_{C3}$, $C_{C4}$); 17.6 ($C_6$)
**MS** (MALDI-TOF) calculated for $[C_{30}H_{39}N_5O_{10}Na]^+$: 652.259; found: 652.160

**ISh053**

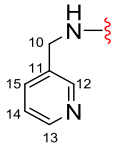



Amine purchased from Sigma-Aldrich. Reaction performed in THF overnight.
Yield: 61% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 8.82 (s, 2H, $H_{14}$), 7.55 (t, 2H, $H_{19}$, $J_{19\text{-}18}$ = 7.4 Hz), 7.32 (d, 2H, $H_{16}$, $J_{16\text{-}17}$ = 8.0 Hz), 7.07 (t, 4H, $H_{13}$, $H_{17}$, $J$ = 7.4 Hz), 6.98 (t, 2H, $H_{18}$, $J$ = 7.4 Hz), 4.26 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 7.3 Hz), 4.00 – 3.96 (m, 1H, $C_2$), 3.74 – 3.34 (m, 13H, $H_{7a,b}$, $C_1$, $H_2$, $H_3$, $H_4$, $H_5$, $H_{8a,b}$, $H_{11}$), 3.00 – 2.84 (m, 4H, $H_{10}$), 2.84 – 2.74 (m, 2H, $C_4$, $C_5$), 1.96 – 1.74 (m, 4H, $C_3$, $C_6$), 1.28 (d, 1H, $H_6$, $J_{6\text{-}5}$ = 6.28 Hz)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.4, 177.9 ($C_9$); 139.1 ($C_{15}$); 129.8, 129.7 ($C_{20}$); 124.6, 124.5 ($C_{13}$); 123.2 ($C_{17}$); 120.5 ($C_{18}$); 120.3, 120.2 ($C_{19}$); 114.1, 114.0 ($C_{12}$); 113.2 ($C_{16}$); 103.7 ($C_1$); 77.7 ($C_{C1}$); 75.9 ($C_4$); 74.4 ($C_{C2}$); 73.9 ($C_3$); 73.4 ($C_2$); 73.9 ($C_5$); 69.9 ($C_7$); 52.9 ($C_8$); 42.6, 42.5 ($C_{C4}$, $C_{C5}$); 42.2, 42.1 ($C_{11}$); 29.9, 29.8 ($C_{C3}$, $C_{C6}$); 27.3, 27.2 ($C_{10}$); 17.7 ($C_6$)
**MS** (MALDI-TOF) calculated for $[C_{36}H_{45}N_7O_8Na]^+$: 726.332; found: 726.177

**ISh054**


Amine purchased from Sigma-Aldrich. Reaction performed in THF overnight.
Yield: 67% over two steps
**$^{1}$H NMR** (400 MHz, $CD_3OD$): δ = 4.32 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 7.0 Hz), 4.10 – 4.05 (m, 1H, $C_2$), 3.79 – 3.33 (m, 27H, $C_1$, $H_2$, $H_3$, $H_4$, $H_5$, $H_{7a,b}$, $H_{8a,b}$, $H_{10}$, $H_{11}$, $C_4$, $C_5$), 2.00 – 1.73 (m, 4H, $C_3$, $C_6$); 1.27 (d, 1H, $H_6$, $J_{6\text{-}5}$ = 6.5 Hz)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.0, 176.5 ($C_9$); 103.9 ($C_1$); 77.2 ($C_{C1}$); 76.0 ($C_4$); 73.9 ($C_{C2}$); 73.8 ($C_3$); 73.0 ($C_2$); 72.8 ($C_5$); 70.6 ($C_7$); 69.0, 68.7, 68.7, 68.6 ($C_{11}$); 53.0 ($C_8$); 48.4, 44.4 ($C_{10}$); 38.2, 38.1 ($C_{C4}$, $C_{C5}$); 30.2, 29.2 ($C_{C3}$, $C_{C6}$); 17.7 ($C_6$)
**MS** (MALDI-TOF) calculated for $[C_{24}H_{39}N_5O_{10}Na]^+$: 580.259; found: 580.182

**ISh055**

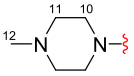



Amine purchased from Sigma-Aldrich. Reaction performed in DMF overnight.
Yield: 57% over two steps
**$^{1}$H NMR** (400 MHz, $CD_3OD$): δ = 8.35 (br s, 2H, $H_{12}$), 8.29 (t, 2H, $H_{13}$, $J$ = 4.0 Hz), 7.65 – 7.58 (m, 2H, $H_{14}$), 7.24 (dt, 2H, $H_{15}$, $J$ = 7.8 Hz, $J$ = 4.7 Hz); 4.34 – 4.18 (m, 5H, $H_1$, $H_{10}$), 3.97 – 3.90 (m, 1H, $C_2$), 3.75 – 3.62 (m, 2H, $H_{7a}$, $C_1$), 3.59 – 3.49 (m, 3H, $H_{7b}$, $H_4$, $H_5$), 3.47 – 3.36 (m, 2H, $H_2$, $H_3$), 3.36 – 3.21 (m, 2H, $H_{8a,b}$), 2.93 – 2.79 (m, 2H, $C_4$, $C_5$), 1.94 – 1.74 (m, 4H, $C_3$, $C_6$), 1.22 – 1.13 (d, 1H, $H_6$, $J_{6\text{-}5}$ = 6.4 Hz)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.8, 178.3 ($C_9$); 150.2, 149.6 ($C_{12}$, $C_{13}$); 138.3 ($C_{14}$); 137.7 ($C_{11}$); 126.1 ($C_{15}$); 103.9 ($C_1$); 77.7 ($C_{C1}$); 76.0 ($C_4$); 74.6 ($C_2$); 73.9 ($C_3$); 73.4 ($C_2$); 72.9 ($C_5$); 70.1 ($C_7$); 53.0 ($C_8$); 42.4, 42.3 ($C_{C4}$, $C_{C5}$); 42.3 ($C_{10}$); 30.2 ($C_{C3}$, $C_{C6}$); 17.7 ($C_6$)
**MS** (HRMS) calculated for $[C_{28}H_{37}N_7O_8Na]^+$: 622.2601; found: 622.2617

**ISh056**

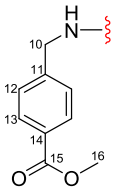

Amine purchased from Sigma-Aldrich. Reaction performed in THF in 1 hour.
Yield: 75% over two steps
**$^{1}$H NMR** (400 MHz, $CD_3OD$): δ = 4.32 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 7.0 Hz), 4.09 - 4.04 (m, 1H, $C_2$), 3.81 – 3.37 (m, 11H, $H_2$, $H_3$, $H_4$, $H_5$, $H_{7a,b}$, $H_{8a,b}$, $C_1$, $C_4$, $C_5$), 2.61 – 2.52 (m, 6H, $H_{10}$), 1.98 – 1.71 (m, 4H, $C_3$, $C_6$), 1.26 (d, 1H, $H_6$, $J_{6\text{-}5}$ = 6.4 Hz), 0.72 – 0.61 (m, 4H, $H_{11a}$), 0.47 – 0.38 (m, 4H, $H_{11b}$)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 179.1, 178.6 ($C_9$); 102.9 ($C_1$); 76.7 ($C_{C1}$); 75.0 ($C_4$); 73.5 ($C_{C2}$); 73.0 ($C_3$); 72.4 ($C_2$); 71.9 ($C_5$); 69.0 ($C_7$); 52.0 ($C_8$); 41.2, 41.1 ($C_{C4}$, $C_{C5}$); 28.9, 28.8 ($C_{C3}$, $C_{C6}$); 23.2, 23.1 ($C_{10}$); 16.8 ($C_6$); 6.6, 6.5, 6.4 ($C_{11}$)
**MS** (MALDI-TOF) calculated for $[C_{22}H_{35}N_5O_8Na]^+$: 520.238; found: 520.177

**ISh057**



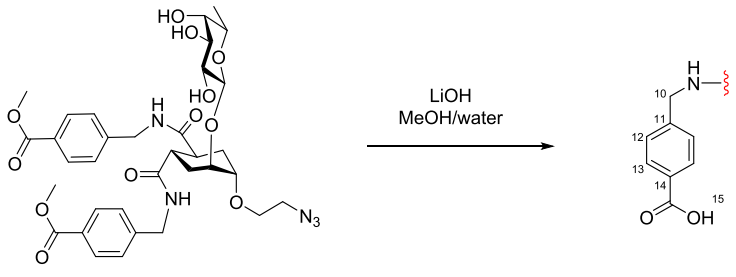
Amine purchased from Sigma-Aldrich. Reaction performed in THF in 6 hours.
Yield: 68% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 4.32 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 7.0 Hz), 4.09-4.04 (m, 1H, $C_2$), 3.81 – 3.37 (m, 11H, $H_2$, $H_3$, $H_4$, $H_5$, $H_{7a,b}$, $H_{8a,b}$, $C_1$, $C_4$, $C_5$), 3.60 – 2.27 (m, 8H, $H_{10}$, $H_{11}$), 1.98 – 1.71 (m, 4H, $H_3$, $H_6$), 1.26 (d, 1H, $H_6$, $J_{6\text{-}5}$ = 6.4 Hz)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 176.8, 176.3 ($C_9$); 103.9 ($C_1$); 77.7 ($C_{C1}$); 76.0 ($C_4$); 73.8 ($C_{C2}$); 73.7 ($C_3$); 73.0 ($C_2$); 72.8 ($C_5$); 70.6 ($C_7$); 57.1, 56.9, 56.5, 56.4 ($C_{10}$); 53.0 ($C_8$); 50.7 ($C_{12a}$); 47.3, 47.2 ($C_{11a}$); 46.8 ($C_{12b}$); 43.4 ($C_{11b}$); 38.3, 38.2 ($C_{C4}$, $C_{C5}$); 30.2, 29.2 ($C_{C3}$, $C_{C6}$); 17.7 ($C_6$)
**MS** (MALDI-TOF) calculated for $[C_{26}H_{45}N_7O_8Na]^+$: 606.332; found: 606.235

**ISh058**



Amine purchased from Sigma-Aldrich. Reaction performed in THF in 6 hours.
Yield: 66% over two steps
**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.82 (dd, 4H, $H_{13}$, $J_{13\text{-}12}$ = 8.4 Hz, $J$ = 5.4 Hz), 7.19 (dd, 4H, $H_{12}$, $J_{12\text{-}13}$ = 8.4 Hz, $J$ = 5.4 Hz), 4.35 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 7.3 Hz), 4.61 – 4.24 (m, 4H, $H_{10}$), 4.07 – 4.02 (m, 1H, $C_2$), 3.86 (d, 6H, $H_{16}$, $J$ = 3.4 Hz),3.78 – 3.74 (m, 2H, $H_{7a}$, $C_1$), 3.69 – 3.58 (m, 3H, $H_{7b}$, $H_4$, $H_5$), 3.56 – 3.46 (m, 2H, $H_2$, $H_3$), 3.46 – 3.28 (m, 2H, $H_{8a,b}$), 3.05 – 2.96 (m, 2H, $C_4$, $C_5$), 2.06 – 1.89 (m, 4H, $C_3$, $C_6$), 1.27 (d, 1H, $H_6$, $J_{6\text{-}5}$ = 6.5 Hz)
**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 176.4 ($C_9$); 175.9 ($C_{15}$); 166.9 ($C_{11}$); 129.3, 129.2 ($C_{13}$); 128.5, 128.4 ($C_{14}$); 126.6 ($C_{12}$); 101.5 ($C_1$); 75.4 ($C_{C1}$); 73.6 ($C_4$); 72.3 ($C_{C2}$); 71.6 ($C_3$); 71.1 ($C_2$); 70.6 ($C_5$); 76.7 ($C_7$); 51.1 ($C_{16}$); 50.6 ($C_8$); 42.1 ($C_{10}$); 40.1, 40.0 ($C_{C4}$, $C_{C5}$); 27.9, 27.8 ($C_{C3}$, $C_{C6}$); 15.4 ($C_6$)
**MS** (MALDI-TOF) calculated for $[C_{34}H_{43}N_5O_{12}Na]^+$: 736.280; found: 736.158

**ISh059**




**ISh058**

14.1 mg (0.02 mmol, 1 equiv.) **ISh058** was dissolved in dry MeOH and 3 equiv. aqueous LiOH solution was added to the mixture. Upon completion ($R_f$ = 0.60 in water : MeCN = 1 : 1 + 0.1% TFA), the reaction was concetrated *in vacuo*.

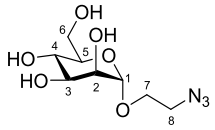
Yield: 95%

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 8.00 – 7.79 (br s, 4H, $H_{13}$), 7.27 (t, 4H, $H_{12}$, $J$ = 7.3 Hz), 4.49 – 4.19 (m, 4H, $H_{10}$), 4.36 – 4.31 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 7.3 Hz), 4.07 – 3.98 (m, 1H, $C_2$), 3.83 – 3.70 (m, 2H, $H_{7a}$, $C_1$), 3.70 – 3.56 (m, 3H, $H_{7b}$, $H_4$, $H_5$), 3.56 – 3.44 (m, 2H, $H_2$, $H_3$), 3.44 – 3.31 (m, 2H, $H_{8a,b}$), 3.04 – 2.91 (m, 2H, $C_4$, $C_5$), 2.07 – 1.85 (m, 4H, $C_3$, $C_6$), 1.25 (d, 1H, $H_6$, $J_{6\text{-}5}$ = 6.4 Hz)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 178.6, 178.1 ($C_9$); 175.8 ($C_{15}$); 145.2, 145.1 ($C_{11}$); 131.9 ($C_{13}$); 130.6 ($C_{14}$); 128.9 ($C_{12}$); 103.8 ($C_1$); 77.6 ($C_{C1}$); 75.9 ($C_4$); 74.5 ($C_{C2}$); 73.9 ($C_5$); 73.3 ($C_2$); 72.9 ($C_3$); 70.0 ($C_7$); 52.9 ($C_8$); 44.6 ($C_{10}$); 42.3, 42.2 ($C_{C4}$, $C_{C5}$); 30.1 ($C_{C3}$, $C_{C6}$), 17.1 ($C_6$)

**MS** (MALDI-TOF) calculated for $[C_{18}H_{29}N_3O_{11}Na]^+$: 708.249; found: 708.129

**ISh044**



Prepared by Norbert Varga

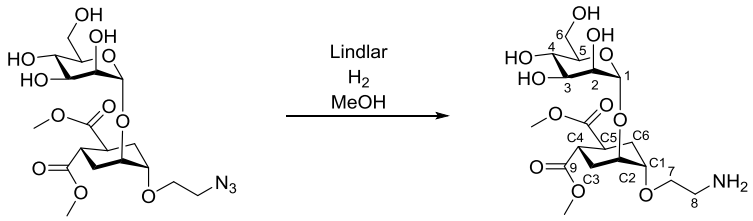
$[\alpha]_D^{20}$ = +55.64° (c = 1.2, MeOH).

**$^1$H NMR** (400 MHz, $D_2O$): δ = 4.95 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.6 Hz), 4.01 (dd, 1H, $H_2$, $J_{2\text{-}3}$ = 3.2 Hz), 3.98 - 3.92 (m, 1H, $H_{7b}$), 3.95 - 3-91 (m, 1H, $H_{6b}$), 3.87 (dd, 1H, $H_3$, $J_{3\text{-}4}$ = 9.4 Hz), 3.82 – 3.77 (m, 1H, $H_{6a}$), 3.78 - 3.73 (m, 1H, $H_{7a}$), 3.72 - 3.68 (m, 2H, $H_4$, $H_5$), 3.58 (ddd, 1H, $H_{8b}$, $J_{gem}$ = 13.6 Hz, $J_{8b\text{-}7a}$ = 6.4 Hz, $J_{8b\text{-}7b}$ = 3.2 Hz), 3.52 (ddd, 1H, $H_{8a}$, $J_{8a\text{-}7a}$ = 6.0 Hz, $J_{8a\text{-}7b}$ = 3.2 Hz).

**$^{13}$C NMR** (100 MHz, $D_2O$): 99.6 ($C_1$); 72.7 ($C_4$); 70.2 ($C_3$); 69.7 ($C_2$); 66.5 ($C_5$); 66.1 ($C_7$); 60.7 ($C_6$); 50.0 ($C_8$).

**MS** (ESI) calculated for $[C_8H_{15}N_3O_6Na]^+$: 272.1; found: 272.6

**ISh045**




**ISh002**

100 mg (0.22 mmol, 1 equiv.) azide **ISh002** was dissolved in dry MeOH and catalytic amount of Lindlar-catalyst was added to the solution. The suspension was stirred under hydrogen atmosphere at RT and monitored by TLC. Upon completion ($R_f$ = 0.37 in water : MeOH = 1 : 1 + 0.1% TFA), the catalyst was filtered and the solvent evaporated. The product was purified by flash chromatography.

Yield: 97%

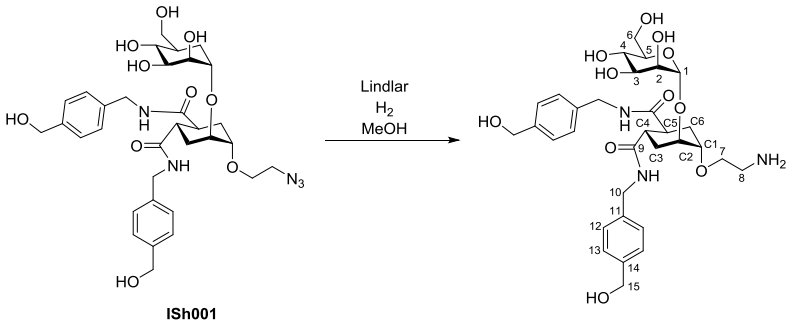
$[\alpha]_D^{20}$ = +48.3° (c = 0.60, MeOH)

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 4.96 (s, 1H, $H_1$), 4.05 (br s, 1H, $C_2$), 3.90 - 3.83 (m, 2H, $H_{6b}$, $H_2$), 3.74 – 3.64 (m, 4H, $H_{6a}$, $H_3$, $H_{7a}$, $C_1$), 3.68 (s, 6H, OMe), 3.64 – 3.60 (m, 1H, $H_{7a}$), 3.60 - 3.52 (m, 2H, $H_4$, $H_5$), 2.98 – 2.83 (m, 4H, $H_{8a,b}$, $C_4$, $C_5$), 2.17 – 2.03 (m, 2H, $C_{3eq}$, $C_{6eq}$), 1.87 – 1.74 (m, 2H, $C_{3ax}$, $C_{6ax}$)

**$^{13}$C NMR** (400 MHz, $CD_3OD$): 177.0, 176.9 (C=O); 100.3 ($C_1$); 76.0 ($C_{C2}$); 75.6 ($C_4$); 72.5, 72.4 ($C_2$, $C_3$); 72.1 ($C_{C1}$); 69.8 ($C_7$); 68.7 ($C_5$); 63.1 ($C_6$); 52.5, 52.4 (OMe); 41.9 ($C_8$); 40.4, 40.2 ($C_{C4}$, $C_{C5}$); 29.0 ($C_{C6}$); 28.2 ($C_{C3}$)

**MS** (HRMS) calculated for $[C_8H_{31}NO_{11}Na]^+$: 460.17893; found: 460.17826

**ISh046**




ISh001

60.4 mg (0.076 mmol, 1 equiv.) azide **ISh002** was dissolved in dry MeOH and catalytic amount of Lindlar-catalyst was added to the solution. The suspension was stirred under hydrogen atmosphere at RT and monitored by TLC. Upon completion ($R_f$ = 0.27 in water : MeOH = 1 : 1 + 0.1% TFA), the catalyst was filtered and the solvent evaporated. The product was purified by flash chromatography.

Yield: 94%

**$^1$H NMR** (400 MHz, $CD_3OD$): δ = 7.31 – 7.16 (m, 8H, $H_{12}$, $H_{13}$), 4.94 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.56 Hz), 4.55 (s, 4H, $H_{15a,b}$), 4.30 (s, 4H, $H_{10a,b}$), 4.06 – 4.02 (m, 1H, $C_2$), 3.91 – 3.82 (m, 2H, $H_2$, $H_{6a}$), 3.75 – 3.50 (m, 7H, $C_1$, $H_{6b}$, $H_{7a,b}$, $H_3$, $H_4$, $H_5$), 2.97 – 2.70 (m, 4H, $C_4$, $C_5$, $H_{8a,b}$), 2.02 – 1.85 (m, 4H, $C_3$, $C_6$)

**$^{13}$C NMR** (100 MHz, $CD_3OD$): δ = 177.1 ($C_9$); 141.6 ($C_{14}$); 139.0 ($C_{11}$); 128.4, 128.1 ($C_{12}$, $C_{13}$); 100.2 ($C_1$); 76.5 ($C_3$); 75.5 ($C_5$); 72.6 ($C_{C1}$); 72.4 ($C_2$); 72.2 ($C_{C2}$); 70.8 ($C_7$); 69.6 ($C_4$); 68.8 ($C_8$); 64.9 ($C_{15}$); 63.1 ($C_6$); 43.7 ( $C_{10}$); 41.8, 41.7 ($C_{C4}$, $C_{C5}$); 29.7, 28.8 ($C_{C3,}C_{C6}$)

**MS** (MALDI-TOF) calculated for $[C_{32}H_{43}N_5O_{11}Na]^+$: 668.315; found: 668.115

# Part I

## Chapter II

## 2.2 MICROARRAY-BASED SCREENINGS WITH CLRS

### 2.2.1 THE EVOLUTION OF MICROARRAYS

The Human Genome Project was completed two years ahead of the previously estimated date, mainly due to the revolutionary innovations in research and technology that accompanied its evolution. The project promoted the development of a large number of inventions, novel tools and methods - not only in the field of genomics, but in molecular biology, physics, chemistry, robotics and computation as well, where the progress opened new perspectives and offered new platforms for research.
While the vast sequence information obtained in the HGP permits the identification of almost all theoretical genes in the genome, it does not indicate the actual gene expression patterns and the expressing cell types. These questions were answered by two new technologies that emerged parallel with the HGP – the quantitative reverse transcription-polymerase chain reaction method (RT-PCR)[264] and DNA microarrays.

#### 2.2.1.1 DNA microarrays

In particular, DNA microarray chips enabled the simultaneous studying of mRNA levels and monitor gene expression profiles "real-time" for hundreds to thousands of genes in biological samples. In their generative work,[265] Shena *et al.* based the quantitative gene expression measurements on the hybridization of complementary sequences - the same principle as how nucleic acids encode genetic information. Hybridization between reverse transcribed cDNA from a biological sample and short, synthetic nucleic acid probes attached to a solid surface (typically glass or silicone chips; earlier filter paper[266]) provide insight into gene expression levels in the sample. (**Figure 33.A**) The technique enabled the analysis of differential gene expression levels e.g. in healthy or diseased tissue, by using multi-coloured fluorescence or biotin/radioactive-labelled reverse-transcribed cDNA by competitive hybridization. The uniqueness of this small and high-density array format lied in its capacity to analyze simultaneously thousands of spotted samples or *in situ* constructed sequence probes in designated locations with the help of computer-based image processing. These are, indeed, the principles of microarray technology.
Although this work was the starting point of the still-to-date exponentially growing number of microarray-related publications from the mid '90s, and it shed light to the versatile applicability of the technique, microarrays had already been described in earlier papers and patents. In one of the earliest works from 1983, Chang studied cell-surface antigen interactions with an antibody chip, although binding events were "evaluated simply by the naked eye".[267] A pioneer in the field of DNA microarrays, Affimetrix, developed high density peptide and oligonucleotide arrays in the early '90s, and described photolithography-based on-chip synthesis and fluorescent-based detection. Early on in the HGP, the company recognized the potential of GeneChips® as a tool to detect specific genes, and analyze genetic information in a high-throughput format.[268] DNA microarrays catalyzed the evolution of functional genomics – a discipline that determines the distinct function and role of genes, thus helping to understand protein expression and their patterns in different cellular processes.

### 2.2.1.2 Lectin microarrays

Not surprisingly, microarray technology was soon adopted in other research areas as well. Protein microarrays were initially developed to obtain a more realistic image of protein expression than what DNA chips provided, as mRNA quantities in the cell are not always mirrored by the actual protein expression levels. Furthermore, post-translational modifications (such as glycosylation, see Introduction) are not indicated by DNA microarrays, but are fundamental for the understanding of protein functions. Nowadays, proteomics benefits greatly from microarrays, due to the minuscule amount of protein and affinity reagent (e.g. antibodies) demand, while large numbers of binding assays can be performed rapidly in a parallel setup – a welcome replacement of techniques like two-dimensional gel electrophoresis and affinity chromatography. In addition to proteomics, protein chips are widespread in diagnostics and are used to detect antigens, antibodies and other biomarkers, furthermore, to monitor disease states and response to treatment. Alternatively, microarrays can be used for the characterization, specificity determination and epitope mapping of proteins. (**Figure 33.B**)

As one of the main classes of GBPs and as a result of their previously discussed biological importance, lectin arrays soon became inevitable and widespread platforms to analyze carbohydrate-lectin interactions after their first appearance in 2005.[269] In this format, lectins are immobilized on a solid support in a highly-organized fashion, and complex analytes – usually disease-associated species - exhibiting various glycosylation patterns are probed on the chip.

The enormous advantage of lectin arrays is that different cell-surface glycosylation signatures, as well as glycoprotein mixtures from cell-lysates can be simultaneously analyzed in a fast assay.[185b, 269b] This can drastically accelerate the glycan-profiling of different cell types – examples include the screening of fluorescently labelled mammalian cells against 43 lectins (Chinese hamster ovary, splenocytes from wild-type and β1-3-*N*-acetylglycosaminyltransferase II knockout mice and K562 leukemia cells during differentiation);[270] enteropathogenic *E. coli*-adhesion studies with intestinal Caco-2 cells[271] and competition assays on a siglec-array to investigate the cellular uptake of adipose-derived mesenchymal stem cell exosomes.[272] Lectin microarrays allow the observation of bacterial surface glycosylation and have been used to identify closely related strains, based on these glyco-patterns. The technology demonstrated that pathogenic and non-pathogenic *E. coli* strains of neonatal meningitis show dynamic alterations in surface glycosylation as a result of environmental stimuli, and these modulations, bacterial fingerprints can be effectively monitored on the microarray chips.[273] Similarly, surface glycans on mammalian Jurkat T-cells were studied on lectin-arrays. Carbohydrate profiling of the cells was performed by immobilizing plant and fungal lectins of various glycan-preferences (*Concanavalin A* (ConA), *Ricinus communis* agglutinin 120 (RCA120), *Sambucus nigra* agglutinin (SNA), *Aleuria aurantia* lectin (AAL) and wheat germ agglutinin (WGA)) and then the adherence of Jurkat T-cells was measured.[274] The same study described the on-chip evaluation of binding affinities and the determination of $IC_{50}$ values between WGA and carbohydrates.

Lectin-based microarrays have facilitated the identification and detection of glycan biomarkers that indicates a great perspective for their use in cancer research and diagnosis. As numerous lectins with different carbohydrate-specificity are commercially available, these chips can be relatively easily prepared and they enable the rapid screening of aberrant glycan and glycoprotein structures in complex samples.[275] The tumoral glycoproteome shows distinct features at different stages (see Introduction), and these are identifiable in the high-throughput microarray-format.

A complementary method, the antibody-lectin sandwich array[276] is used to evaluate the role and modifications of glycosylation patterns on glycoproteins. In this case, immobilized antibodies capture glycoproteins from

complex biological fluids and in the following step, labelled lectins recognize the antibody-bound carbohydrates – the amount of lectin reflects the amount of specific glycans on the glycoprotein attached to the antibody.
These application areas – at least partially - overlap with another main type of microarrays, the glycan microarrays.

### 2.2.1.3 Glycan microarrays

Drickamer and Taylor saw the immense potential in the microarray technology as early as 2002[93] when they estimated the number of all endogenous mammalian glycans around 500. As all these ligands could fit on a tiny microarray chip, they forecasted that the implementation of this new tool could give a boost to glycomics – and be used to investigate or exploit interactions between carbohydrate-based ligands and GBPs: lectins, antibodies and enzymes, with a main focus on the first group. According to their expectations, by printing every mammalian glycan on the chips, the approximately 100 mammalian lectins could be fully characterized with regards to their carbohydrate-specificity, including confirmation of already known interactions and the identification of new or unexpected ligands. At the same time, they proposed that glycan microarrays will be useful for other purposes as well, for example to study interactions of glycans and lectins from non-mammalian sources (e.g. microorganisms), or to address specific issues by custom-designed glycan-chips. (**Figure 33.C**)

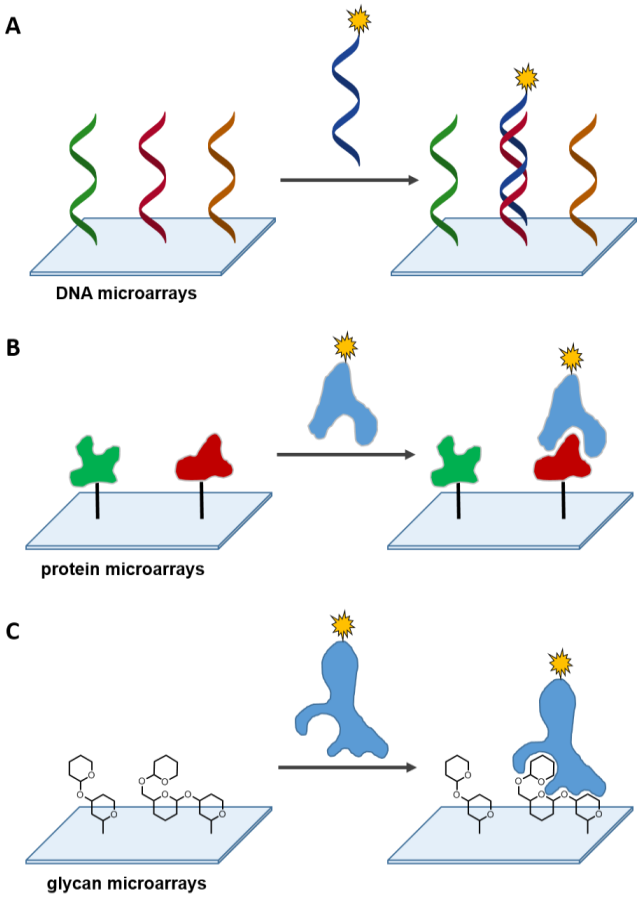



***Figure 33.*** *Three major classes of microarrays. A. DNA microarrays. Reverse transcribed cDNA hybridizes with synthetic oligonucleotides printed on the microarrays chips. B. Protein microarrays. Immobilized proteins can be probed with proteins (e.g. antibodies), cells, pathogenic species (e.g. bacteria) or different biomarkers, such as glycan antigens. C. Glycan microarrays. Carbohydrate structures are anchored to the solid surface in order to analyze glycan-GBP interactions, with minute amounts of the biological target, like lectins, antibodies and enzymes.*

In accordance with the predictions, glycan microarrays are presumably the fastest, most powerful tools to study the binding preference of lectins and to identify lead structures from a larger pool of ligands. A well-designed, accurate and reproducible screening setup offers a platform to evaluate and quantify interactions between carbohydrates and lectins. Moreover, glycan chips have proven to be useful to profile binding specificities of newly isolated lectins, identify new GBPs and their functions, as well as to reveal unexpected, not-obvious carbohydrate ligands of these proteins. As many parameters are well-controlled and tunable, microarrays are often used to observe and highlight finer differences in ligand preference. Large natural and synthetic oligosaccharide libraries have been screened against different classes of lectins, and the obtained results have in numerous cases paved the way for further studies of individual ligands. The identified high affinity binders can, for example, suggest ideal starting points for glycomimetic structure design and optimization.

#### 2.2.1.3.1 The Consortium for Functional Glycomics

It is important to stress that differences in microarray fabrication – from the preparation of the carbohydrate-derivatives and their attachment to the solid support, followed by incubation with target molecules, up until detection and evaluation - greatly affect the obtained results, affinity and specificity. Therefore, an international initiative – similar to the MicroArray Quality Control[277] - was introduced for a standardized protocol system in glycomics. The Consortium for Functional Glycomics (CFG) was founded in 2001, to address procedural

guidelines, quality control and assistance for the investigation of protein-carbohydrate interactions. The CFG provides readily available resources for carbohydrate chemists – among these the largest public mammalian glycan microarray (both natural and synthetic structures), with over 600 well-defined glycans; a microbial glycan library (>300 targets), as well as the associated databases and an extended networking forum. These accessible libraries cover the majority of carbohydrate epitopes found on glycoproteins and glycolipids. Sadly, the glue grant that brought to life this ambitious project finished in 2011 and, as the glycan resources are finishing, the Consortium has been in need of new contributors and funding to be fully operational again. Yet, during those ten years, the CFG accomplished to profile hundereds of lectins available from collaborators or commercial sources.

#### 2.2.1.3.2 Lectin profiling on glycan microarrays

Plant and microbial lectins are of particular importance in biological research and diagnostics (see Introduction). Numerous plant lectins have been screened to reveal their carbohydrate-specificity to the full extent, and isolated bacterial lectins have also been tested for their ligand specificity.[273a, 278] A few examples of bacterial lectins probed on glycan microarrays are *E. coli* adhesins (FedF, FimH),[279] *Burkholderia cenocepacia* lectin (soluble BC2L-C)[280] and bacterial adhesins LecA[281] and LecB[282] (see Introduction) from *Pseudomonas aeruginosa*. The ligand specificity of LecB for fucosylated oligosaccharides, in particular $Le^X$ was confirmed on glycan microarrays.

In the broader class of mammalian lectins, successful applications of microarrays include specificity determination of cation-independent and cation-dependent mannose-6-phosphate receptors on a phosphorylated glycan-chip,[283] screening studies of galectins against diverse oligosaccharides[228, 284] and lead identification in the siglec-family.[228, 285] In particular, the last group of lectins were probed on microarrays displaying not only natural, but non-natural motifs as well, because the natural sialosides show only poor affinity towards siglecs. For a better understanding of siglec functions and in search of high-affinity and selective ligands, several sialoside libraries have been generated directly on-chip. From a few parent structures modified sialic-acid libraries of more than 1000 members have been prepared with the help of chemoenzymatic transformations[286] or CuAAC-reactions ("click and pick" strategy)[287] and the compounds were analyzed against various siglec-family members. Interestingly, this glycomimetic approach yielded several lead structures, some of which were identified as selective siglec-7 ligands and were employed to functionalize liposomal nanoparticles to enhance binding affinity.[288]

#### 2.2.1.3.3 Previous examples of microarray-based CLR screenings

The purpose of our glycomimetic library and its screening against C-type lectin receptors on-chip is analogous to this approach: chemical synthesis of diverse glycomimetic ligands, microarray-based screening against various target lectins to select the most potent and possibly selective monovalent lead structures and further improvement of their affinity by presentation in multivalent systems.

Murine C-type lectin family members have been probed on glycan microarrays, including rat asialoglycoprotein receptor, Kupffer cell receptor and MGL;[289] murine SIGNR-1, SIGNR-3 and langerin;[290] murine MGL1 and MGL2[291] and the entire mouse SIGNR family (see Introduction).[292] A novel screening platform was described for CLEC-2, CLEC-4a (DCIR), CLEC-4b (DCAR), CLEC-4d (MCL), CLEC-4e (Mincle), CLEC-7a (dectin-1), CLEC-9a, CLEC-12a (MICL), CLEC-12b, MGL-1, and SIGNR3 in order to identify carbohydrate ligands of immunologically relevant CLRs of the dectin-1 and DCIR family. This platform is based on CLR-Fc fusion proteins that display 2 CRDs on the Fc fragment of human IgG1, and screened against a glycan microarray. Detection required a tagged-secondary antibody, and the most promising ligands were used to prepare neoglycoconjugates on the model antigen ovalbumin.[293]

Numerous examples of microarray-based screening of human C-type lectins can be found from the recent years. Profiling campaigns have been conducted, for example, with SRCL,[88a, 289] MGL,[294] BDCA2,[81] dectin-1,[295]

LSECtin,[296] and Mincle[297] and identified specific recognition epitopes, or indicated slight differences in binding to similar residues.

DC-SIGN was in several cases the target of microarray-based assays. Glycan microarrays helped to confirm that $Man_9$ was indeed the key motif in DC-SIGN's binding to gp120 and one of the highest-affinity natural ligands of the lectin. At the same time, ligands such as other high mannose structures and fucosylated moieties (the 4 Lewis-type and blood type A and B antigens) were identified as good leads for further studies. The arrays preparation included the immobilization of biotinylated sugars on streptavidin-covered wells and their incubation with fluorescein-tagged DC-SIGN.[227] (**Figure 19**) The results obtained with DC-SIGN-Fc chimera[228] were consistent with the first study, various mannosylated and fucosylated glycans were recognized on a microarray chip displaying more than 200 natural and synthetic carbohydrates. In this case, interactions were detected by a secondary incubation with fluorescently tagged anti-human IgG that bound to the Fc-region of the recombinant fusion protein. Furthermore, an expanded library of glycans and neoglycoconjugates (biotinylated polyacrylamide containing mono- and oligosaccharides) was screened against DC-SIGN-Fc and the $IC_{50}$ was determined by on-plate titration.[186] Microarray screenings indicated that linear fragments of this oligosaccharide exhibit binding potency similar to the progenitor and crucial epitopes were determined (see Introduction).[232] High density $Man_4$-coated and $Man_9$ dendron-functionalized slides were prepared for competition assays with recombinant dimeric DC-SIGN and the HIV-neutralizing monoclonal antibody 2G12 to measure the $IC_{50}$ values of multivalent oligomannoside dendrons.[298] In a unique approach, a nucleic acid-encoded glycan library was prepared on-chip and probed with DC-SIGN. Repeated chemical derivatization and split and mix steps were followed by peptide-nucleic acid encoding and combinatorial self-assembly into DNA templates which all contained the Manα1-2Man epitope.[299] A 33-membered oligosaccharide library containing core-modified (fucosylated, xylosylated) *N*-glycans was screened against DC-SIGN-Fc, and demonstrated the negative effect of core xylose on binding to the CLR.[300]

DC-SIGN's closely related homologue, DC-SIGNR has also been the subject of microarray screenings – as the two proteins exhibit highly similar features in their CRD, microarray technology is an excellent method to compare finer ligand specificities and identify potentially discriminative epitopes. During the first, and by far most influential glycan microarray analysis of the two (fluorescently-labeled) ECDs,[227] Guo *et al.* observed critical differences in their ligand preferences – while both lectins recognize highly mannosylated *N*-glycans, above all the $Man_9GlcNAc_2Asn$ glycopeptide, DC-SIGNR lacks the ability to bind to Lewis-type and blood group antigens, as well as to charged, bulky sialylated oligosaccharides. The results also underline the importance of presentation, as identical epitopes displayed on various spacer arms lead to different recognition. DC-SIGNR's carbohydrate specificity was confirmed by glycan microarray assays performed with DC-SIGNR-Fc chimeras, and additionally an agalactosylated complex-type *N*-glycan ligand was identified.[301]

As discussed earlier (see Introduction), DC-SIGN and langerin pose an important challenge for selective ligand binding because of their respective promoter and preventive role in HIV-infections. Once more, microarrays offer an ideal platform to comprehensively study the binding characteristics of hundreds of ligands simultaneously. If possible, the straightforward comparison of binding profiles (obtained under the same conditions) could facilitate the identification of selective lead candidates. The first array-based assay of langerin was performed with the murine homologue in 2004,[290] and the human lectin was used for the first time only in 2010. The native trimeric form of langerin was described in the same paper by Feinberg *et al.* who labeled a truncated fragment with fluorescein to run the binding assays on glycan microarrays from the CFG.[97] Shortly after, langerin-Fc chimeras were prepared and analyzed on a glycoconjugate microarray,[97] and both studies confirmed langerin's affinity towards 6-sulfated Gal-containing $Le^X$ structures and some high mannose-family members (see Introduction). In a further paper, bacterial lipopolysaccharide-derived O-specific polysaccharides were printed

on microarray chips, and in addition to the previous ligands, blood group B antigen (Galα1-3(Fucα1-2)Gal) was identified as a langerin ligand.[80]

Similarly to the DC-SIGN homologues, dectin-2 recognizes high mannose structures on glycan arrays, primarily $Man_9GlcNAc_2$ and to some extent $Man_8GlcNAc_2$ and $Man_7GlcNAc_2$.[302] The screening was performed according to the conventions of the CFG with dectin-2-Fc chimeras, although not on glass microarray chips, but on a streptavidin-biotin surface in an ELISA-format assay. Up to date, DC-SIGN has been known to recognize a broader set of oligosaccharides than dectin-2, while for the latter, only common high-Man glycans and glycoconjugates were identified as ligands (see **2.2.3.4**).

#### 2.2.1.3.4 Other applications of glycan microarrays

Apart from their application field in profiling lectins, glycan microarrays have offered new solutions in clinical diagnostics, vaccine development and pathogen-related disease monitoring.[303] (**Figure 34**) In addition, quantitative measurements of carbohydrate-protein interactions can be carried out on glycan microarrays. In solidly established, well-characterized microarray systems, binding affinity can be determined as $IC_{50}$, $K_D$ or $K_{app}$ (apparent dissociation constant).[304]

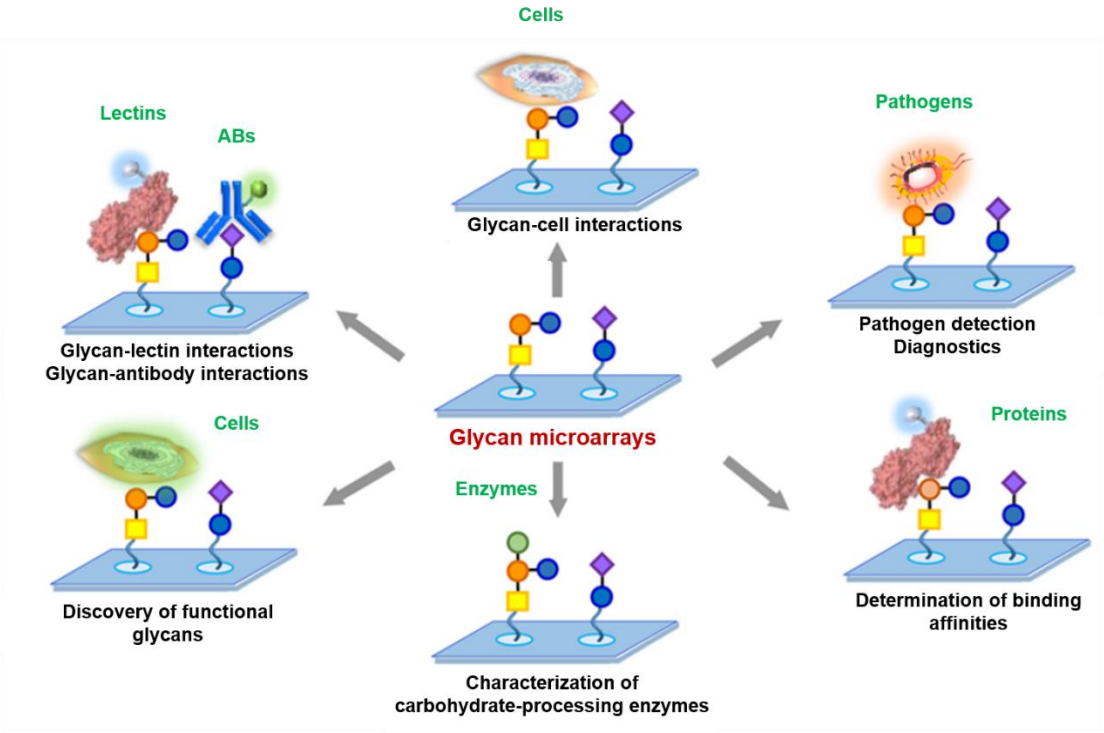



***Figure 34.*** *Main applications of glycan microarrays. Modified from Hyun et al.*[303]

## 2.2.2 MICROARRAY FABRICATION

### 2.2.2.1 Elements of the microarray system

In a typical microarray experiment, a substantial number of compounds of interest and known identity are immobilized in a well-defined, tunable arrangement on a suitably functionalized solid support. Pico- or nanoliters of the probe solutions are disposed in a spatially organized and identifiable manner by a robotic spotter, so that each compound is displayed in a high-density, individual spot (diameters of hundreds μm). Several thousands of spots can fit on a single slide, but in practice, the chip is divided into identical subarrays with the help of a silicone or plastic gasket that allows parallel screenings on the same chip with small volumes of the protein solution. After the allotted immobilization time, the slides are washed and unreacted functionalities are blocked. The array surface is then probed with the solution of the biological target which, in many cases, already contains a tag (e.g. fluorescent or radio-labelled), or has high affinity towards a labelled secondary reagent (e.g. secondary antibodies or biotin-avidin pairs). After removal of the unbound protein, the results may be read-out by various detection methods, depending on the employed revelation strategy. The different intensity values obtained for each spot directly reflect the number of bound targets by the compound printed at that exact position on the slide.
Numerous glycan microarray systems have been described in the past 16 years, and several alternatives are available for each component of the system. The purpose and utility of building our microarray platform determines (among other details) the nature of glycans printed; the solid surface and its functionalization; the printing technique, chemistry and concentration; as well as the detection method we intend to employ.

#### 2.2.2.1.1 Glycan preparation for immobilization

In the majority of cases, the carbohydrate ligands are appropriately modified for immobilization on the solid support. If the glycan was obtained though chemical synthesis or chemoenzymatic reactions, the reducing end residue typically contains a linker carrying a functionality from the early stages of the pathway, which is then exploited for attachment to the chip. This moiety has to be compatible with the synthetic route and not interfere with other groups on the molecule, furthermore be able to form covalent or non-covalent bonds with the surface rapidly and specifically.
Oligosaccharides obtained from natural sources rarely contain any aglycon functional groups prone to help their immobilization. They are, however, often found in nature as glycoproteins, glycolipids, or polysaccharides and these conjugates open other perspectives. These molecules are often adsorbed on nitrocellulose surfaces, although the level of immobilization can vary significantly. Soon, the glycoconjugate approach was adopted to synthetic oligosaccharides, and NGPs/neoglycopeptides, NGLs, glycopolymers and glycodendrimers were generated analogously.[228, 232, 271, 295, 305] Glycolipids have been applied on microarray slides in liposomal systems that resemble cell-surface environment and endow these glycoconjugates with mobility over the surface.[305e]
When the introduction of new moieties or functional groups is not possible, derivatization of the free reducing end residue can offer a solution to immobilize unmodified oligosaccharides. In this case, however, one important point is the preservation of the native sugar conformation – or attempting not to distort the original structure too significantly. The closest architecture to the unmodified glycan is achieved, when the reducing end ring is retained in its closed-form – for example, by linking the saccharide with *N*-substituted *N*-hydroxylamine,[306] or through the acylation of glycosylamines, followed by their reaction with acyl chlorides.[307]
The equilibrium between the hemiacetal and the aldehyde also allows the exploitation of the electrophilic aldehyde group, for example, with nucleophiles like amines, hydrazides and oxyamines. During reductive

amination, the formation of solely the open ring product leads to a loss of glycomic information encoded by this residue. Unsubstituted hydroxylamines yield mainly the open-ring glycan (oxime ligation),[308] while *N,O*-alkyl-substituted[306a] and hydrazide-derivatives[309] form predominantly the closed-ring β-product.
The derivatization can be performed prior to separation of a mixture of glycans from a natural source, in which case the modification can even facilitate the following purification. During a reductive amination, for instance, the introduced fluorescent 2,6-diaminopyridine served both the purpose of separation by multidimensional HPLC and the later immobilization via the second amino-group.[284b, 310]
Similar requirements for derivatization and conformational preservation apply to glycans that have been assembled by automated solid-phase synthesizers[311] or by enzymatic transformations.[312] In the latter case, simple monosaccharide building blocks, or larger oligosaccharides can be subjected to glycosyltransferases, and if the substrate and acceptor specificities of the enzyme allow, even unnatural structures carrying linkers and aglycon motifs can be elongated by glycosylation.[313]

### 2.2.2.2 Presentation of the carbohydrate epitopes

The microarray chip is extremely useful for studies of glycan-lectin interactions because the high-throughput and miniaturized setup offers a fast, simultaneous analysis of binding events between glycan probes and their lectin targets, providing a uniquely large amount of results per volume of biological sample. On one hand, huge volumes of data can be collected rapidly on a small surface, from tiny amounts of the precious proteins and oligosaccharides (synthetic, or obtained from natural sources). On the other hand, the inherently weak affinity of monovalent glycans towards lectins is amplified through their high-density, multivalent presentation on the solid platform, leading to an easier detection of these stronger, reversible interactions with the help of the "velcro"-effect (see Introduction).
Indeed, the strength of binding events is enhanced on the multivalent microarrays as compared to monovalent ligands. However, proteins must have proper access to interact with these ligands. Too dense presentation is disfavoured, as saturation on the surface, and thus in the obtained signals, causes loss in the assay sensitivity. The concentration and presentation of ligands spotted onto the chips, as well as the concentration of the proteins and the read-out technique's sensitivity are therefore key parameters to optimize during the fabrication of glycan microarrays. The surface density and presentation of printed glycans and glycoconjugates has been evaluated in different systems and its effect on lectin and antibody specificity was found to be fundamental in some cases.[314]
This is due to the fact that in optimal case, the ligands' distribution on the surface has to match the network of binding sites on the proteins (e.g. the CRDs on the lectin).[315] In one example, distinct antibody subpopulations could effectively discriminate glycans printed at varying density levels.[314a]
Three dimensional presentation is more advantageous than two dimensional, but the spacing between individual ligands is rarely well-controlled. Similar problems arise from using multivalent glycoconjugates, as the local density of ligands, as well as their orientation is difficult to characterize at the molecular level, and yet, the enhanced avidity of the ligand-protein complex compensates these uncertainties.[304b, 314b, 316] In ideal case, this matrix should mimic the natural structure of glycan presentation, and the density and positions of glycan epitopes on the multivalent platform would match the binding site (CRD) arrangement on the protein. The collective effects of multimeric carbohydrate-presentation and lectin oligomerization can enhance affinity through avidity, but the glycan orientation - apart from some specific cases (e.g. noncovalently immobilized glycolipids and self-assembling monolayers) – is typically inordinate; the spacing between individual ligands is not well defined.
This issue can be circumvented by the preparation of artificial glycoconjugates with well-established valency.[317]
A quantitative approach to emulate cell-surface glycan presentation and its effects on their interactions with *E.*

*coli* was attempted by using fluidic glycan microarrays.[304c] In addition to ligand accessibility, the most important aspect of carbohydrate presentation and density on the chip is the homogeneous distribution of ligands over the spot for an accurate software-based analysis of binding signal intensity. These parameters need to be optimized during the setup of a new microarray system.

### 2.2.2.3 Types of solid supports used with different detection methods

The choice of solid support is determined by the main function of the microarray and the detection system we intend to employ. Currently, many slide platforms are commercially available, and display a high degree of reproducibility and well-controlled surface-modification parameters for production. In addition to their composition and functionalization, the main characteristics of microarray slides include: stability, chemical homogeneity, polarity, wettability, chemical resistance and reactivity towards various functional groups, and susceptibility to aspecific binding.

The most widespread and commercially available chips are the glass slides, thanks to their low cost, transparency and straightforward chemical modification. Transparency is essential for optical detection methods, such as fluorescence- or colorimetry-based read-out systems. Additionally, glass slides are attractive solid supports because of their low-porosity, flat surface and rigidity. During their preparation, glass slides are treated with highly corrosive reagents, such as piranha solution, NaOH solution or oxygen plasma, to activate silanol-groups, which are then silanized with the appropriate organofunctional silanes.[318] Often, glycan microarrays are covered with a covalently-attached, permeable three-dimensional matrix (e.g. polymeric hydrogel, polyacrylamide) that is densely functionalized and imitates the glycan-coat of natural species. The reactive functional groups can be linked through a flexible spacer, allowing more spatial freedom to the ligands. This structure also allows the diffusion of ligands into the matrix, taking full advantage of the enhanced capacity compared to two-dimensional surfaces.

The popularity of glass slides can be attributed to the above mentioned advantages, but other applications of glycan microarrays can demand different supports. During the advent of glycan microarray technology, plastic microtiter plates were often used for immobilization.[319] Conductive gold surfaces, on the other hand, allow a label-free, real-time SPR or MALDI-Tof[320] monitoring of binding events, providing insight both on the affinity of the protein towards the ligand and the kinetics of the given interactions as well. The drawbacks of SPR microarrays are the limited number of interactions that can be studied simultaneously, the complex detection system and the necessity to appropriately functionalize the gold-surface, typically through thiol-chemistry.[321] Conductive, and transparent ITO (indium tin oxide) and aluminum oxide covered glass chips can be an ideal option to follow enzymatic reactions on-chip by MALDI-MS.[322] Both SPR chips and ITO/aluminum oxide slides offer a more organized ligand presentation than three-dimensional polymer matrix coated glass slides, where glycans are randomly oriented on the surface.

### 2.2.2.4 Immobilization of the ligands

Carbohydrate ligands can be attached to the microarray surface through plentiful methods, but the choice of immobilization chemistry highly depends on the nature or functionalization of the ligands, the surface and the purpose of the experiment. The incorporated functional groups (if there is any) on the glycans can be selected based on the method by which the ligands are linked to the surface, and *vice versa*, occasionally the limited options

of ligand modifications narrow down the selection of attachment methods. The most basic criteria of immobilization is that the applied reactions are rapid, specific and have high conversion rates – keeping in mind that these aspects can fundamentally differ between solution phase reactions, and at the interface of the solid support and the printed ligand solution. Naturally, the majority of linking methods have been adapted from DNA and protein microarrays, but there are a number of techniques developed specifically for glycans. Immobilization can be divided into two main classes: non-covalent and covalent attachment of glycans.

#### 2.2.2.4.1 Non-covalent immobilization

Non-covalent adsorption relies on the adherence of carbohydrate ligands to hydrophobic surfaces and the interactions are governed by van der Waals, electrostatic or other forces, without the formation of any covalent bond. Since these interactions are generally weaker than covalent linkages, their strength, and therefore the stability of ligand attachment depends on the extent of contact area. As a consequence of weak adsorption, ligands can be desorbed from the surface, for example, during washing steps.

Among the first glycan microarrays, unmodified polysaccharide, glycolipid, glycosphingolipid and glycoprotein immobilization on nitrocellulose or oxidized black polystyrene surfaces were described.[305a, 323] The same method could not be applied to smaller glycans, however, because the lower molecular weight resulted in insufficiently weak binding. This issue was effectively solved by the preparation of NGLs (elongating smaller glycans with a long, hydrophobic tail) that could later adhere to nitrocellulose or polyvinylidenedifluoride matrices.[305c-e, 324]

Similarly, immobilization relying on electrostatic interactions is exclusively employed for high MW carbohydrate-molecules. GAG chains, such as unmodified heparin, as well as natural or chemically modified dextran can engage in such interactions through their sulfate groups with amine-based surfaces (for instance, with positively charged poly-L-lysine) or semicarbazide glass slides.[325]

Fluorous-fluorous interactions can also be exploited to immobilize fluorous-tagged carbohydrates on fluorous-functionalized (e.g. fluoroalkylated, fluorophosphonated) solid surfaces.[326] An additional advantage of this method is that during ligand derivatization, fluorous-labelled glycans can be purified by fluorous chromatography. As in the case of glycolipid adsorption on hydrophobic surfaces, fluorous-based immobilization is more resistant to washing than other physisorption-based microarrays.

The uniquely strong binding between biotin and streptavidin ($K_D \sim 10^{-15}$ M) provides an attractive tool for the strong, but non-covalent conjugation of biomolecules - these tags are utilized both on protein and glycan microarrays. On one hand, their interaction allows the stable immobilization of ligands, and on the other, they can be used to attach fluorescent tags to the targets (discussed later on). Biotin-tagged glycans have been anchored to streptavidin-covered solid platforms on several occasions, including the early glycan array of the CFG.[227, 285a, 327]

Glycan immobilization is possible through the hybridization of glycan functionalized complementary oligonucleotides. Diverse glycans can be conjugated to single stranded oligonucleotides and anchored to the complementary oligonucleotide-covered surface, then probed with the target proteins.[316] (**Figure 35**)

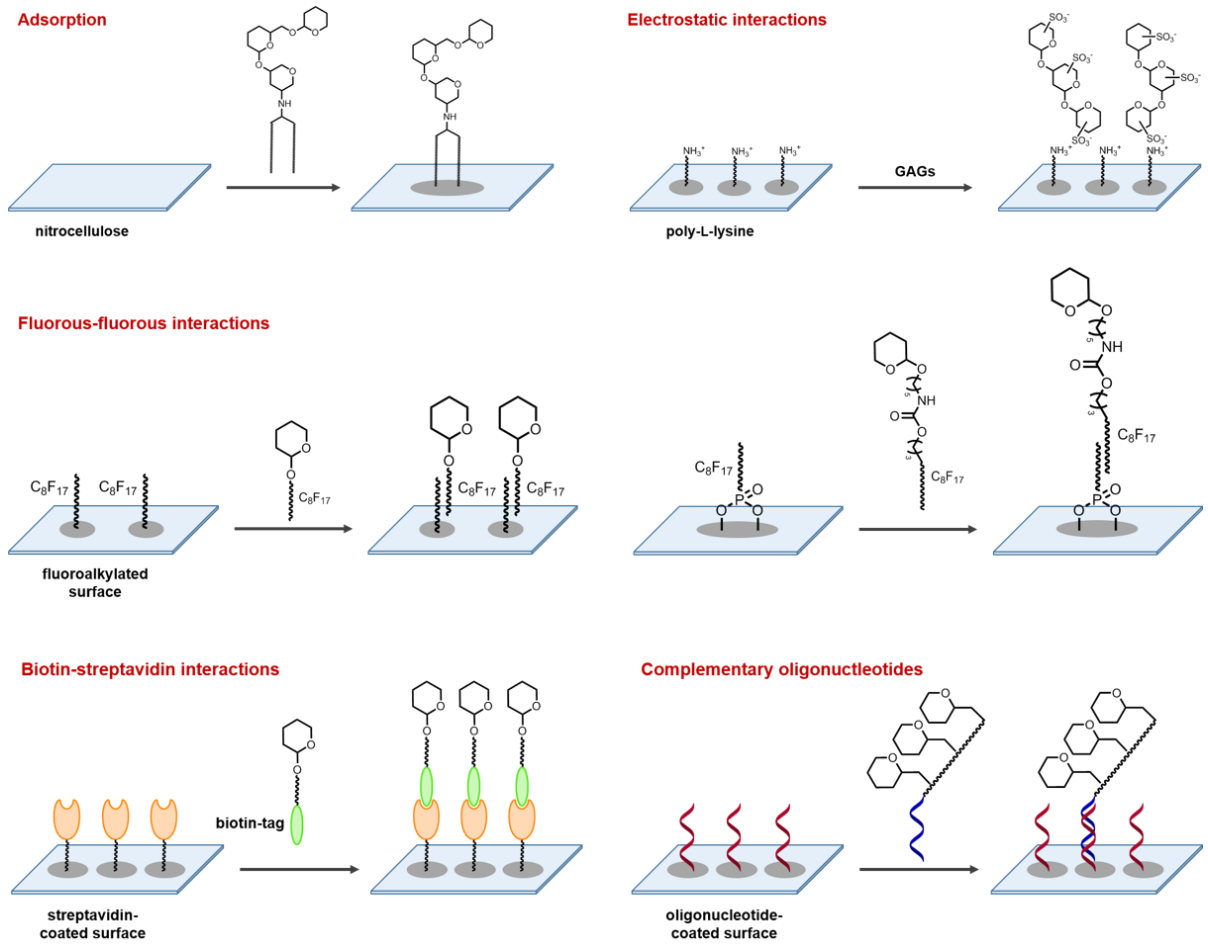



***Figure 35.*** *Non-covalent immobilization of glycans on functionalized or non-functionalized glass microarray surfaces. Non-covalent interactions are significantly weaker than covalent linkages, therefore carbohydrates anchored through some of these methods are more likely to be removed during washing steps. On the other hand, non-covalent immobilization, for example through van der Waals interactions are ideal for MS-based detection methods. Glycans are often derivatized with long lipophilic chains and absorbed on lipophilic surfaces in order to study enzyme activity and detect modifications by MALDI analysis.*[328]

### 2.2.2.4.2 Covalent immobilization

Covalent immobilization methods involve properly modified glycans and/or microarray chips, which are equipped with reactive groups – a reaction between these functionalities leads to the formation of a covalent-bond and a strong, stable anchoring of the ligand. The functional groups have to be appropriately reactive, chemoselective and possibly biorthogonal, depending on the studied biological system. Furthermore, in case of synthetic carbohydrates, these groups need to be compatible with the followed chemical strategy.

Although a great number of immobilization chemistries have been described, many researchers opt for the robust and widely-explored amine chemistry. Its components, such as pre-functionalized chips and linkers, are often readily available and the parameters (e.g. concentrations) can be simply optimized for the target system. Amine-decorated glycans can be easily tethered to *N*-hydroxysuccinimide(NHS) ester surfaces, optimally at slightly alkaline pH.[228] Amino- and aminooxy-groups both on glycans and on the microarray surface can also react with epoxide,[321b, 329] aldehyde[330] and cyanuric chloride moieties.[331]

Likewise, another popular immobilization method is based on thiol-ene type reactions, primarily between maleimide-functionalized chips and sugars, equipped with a thiol group (or in reverse combination).[232, 314b, 332] Similarly to amino-tethered sugars, thiol-bearing ligands can also be attached to epoxide surfaces.[333]

Epoxide-coated slides are versatile in their reactivity towards various functionalities and glycoconjugates. In addition to amino and thiol groups, they covalently link hydrazide-functionalized sugars[334] and are able to anchor NGPs to the chip-surface.[314a, 335]

Ligand immobilization on gold microarray chips equipped with an SPR detection system can be performed by the formation of disulfide bonds.[336] In the two step process, surface-bound sulfhydryl groups are reacted with 2,2'-dipyridyl disulfide solution, then thiol-equipped glycan ligands are anchored to the surface by a thiol-disulfide exchange. Furthermore, the coupling reaction between glycomethanethiosulfonate ligands and a thiol-functionalized polymer surface was used for the preparation of rewritable microarrays, from which the attached carbohydrates could later be stripped.[337]

Cycloaddition reactions of various types have been applied to fabricate microarray chips, as the reactions are highly chemoselective and rapid. Cyclopentadiene-terminated glycan ligands undergo Diels-Alder type reactions with benzoquinone and penta(ethylene glycol) decorated self-assembled monolayers on the chip-surface.[338] The same cycloaddition with inverse-electron-demand was employed to ligate carbohydrate-dienophiles on tetrazine-functionalized glass slides.[339]

CuAAC or click reactions have offered a convenient, highly selective solution for coupling biomolecules without the formation of side-products, since their first description in the early 2000s.[340] The reaction is compatible with other functional groups present, and can be performed in different solvents – thus, its potential for the fabrication of glycan microarrays has been investigated by several groups. The Cu(I)-catalyzed reaction between an azide and an alkyne partner was used to "click" various azido-group containing oligosaccharides,[319b, 341] aminoglycosides[342] or glycopolymers[343] onto microarrays. An alkyne-functionalized polymeric coating was developed specifically for glycan microarrays.[344] The drawback of CuAAC reactions in bioanalytical assays is the presence of the heavy metal catalyst, which has been reported to promote the degradation of viruses and oligonucleotide strands *in vitro*.[345]

A unique type "click" reaction, the strain-promoted alkyne-azide cycloaddition (SPAAC) could offer a catalyst-free, biorthogonal solution to prepare glycan microarrays, but so far, the reaction has only been exploited for the fabrication of peptide chips.[346] The advantages of this type of chemistry will be discussed later on.

Azido-groups are also useful for Staudinger-ligations with phosphanes,[347] and the photolabile 4-azido-2,3,5,6-tetrafluorophenyl group can be employed to attach glycans to polymer monolayer surfaces or gold SPR chips by photo-induced reactions.[348]

As discussed earlier among the carbohydrate ligand derivatization methods, smaller and larger unmodified glycans can be immobilized on appropriately functionalized microarray slides via condensation reactions between their reducing end and reactive groups on the chip. A number of site-nonspecific photochemical reactions have also been described to link glycans to solid supports with the help of light irradiation. By addition of unmodified sugars to photoreactive-group coated surfaces, such as commercially available aryl-trifluoromethyldiazirine dextran chips,[271, 349] 4-azido-2,3,5,6-tetrafluorophenyl-coated slides[348a] or phthalimido-activated self-assembled monolayers,[350] covalent linkages can be formed after photoactivation. In the first cases, the irradiated reaction goes through highly reactive carbene and nitrene species, while in the last, aromatic carbonyl groups in the phthalimide moiety absorb a photon from UV light (300 nm) and abstract a hydrogen from the spotted free sugars. Radical recombination leads to the formation of a covalent bond between the carbohydrate and the surface.

Another site-nonspecific method has been reported to immobilize 1,2- or 1,3-diol containing sugars (e.g. Glu, Gal, Fru) on boronic-acid covered surfaces.[351] The straightforward fabrication method relies on the complexation of boronic acid with the diol, thus tethering the sugar covalently by the formation of a boronate, but the undefined position of the diol reduces the applicability of this technique. (**Figure 36**)

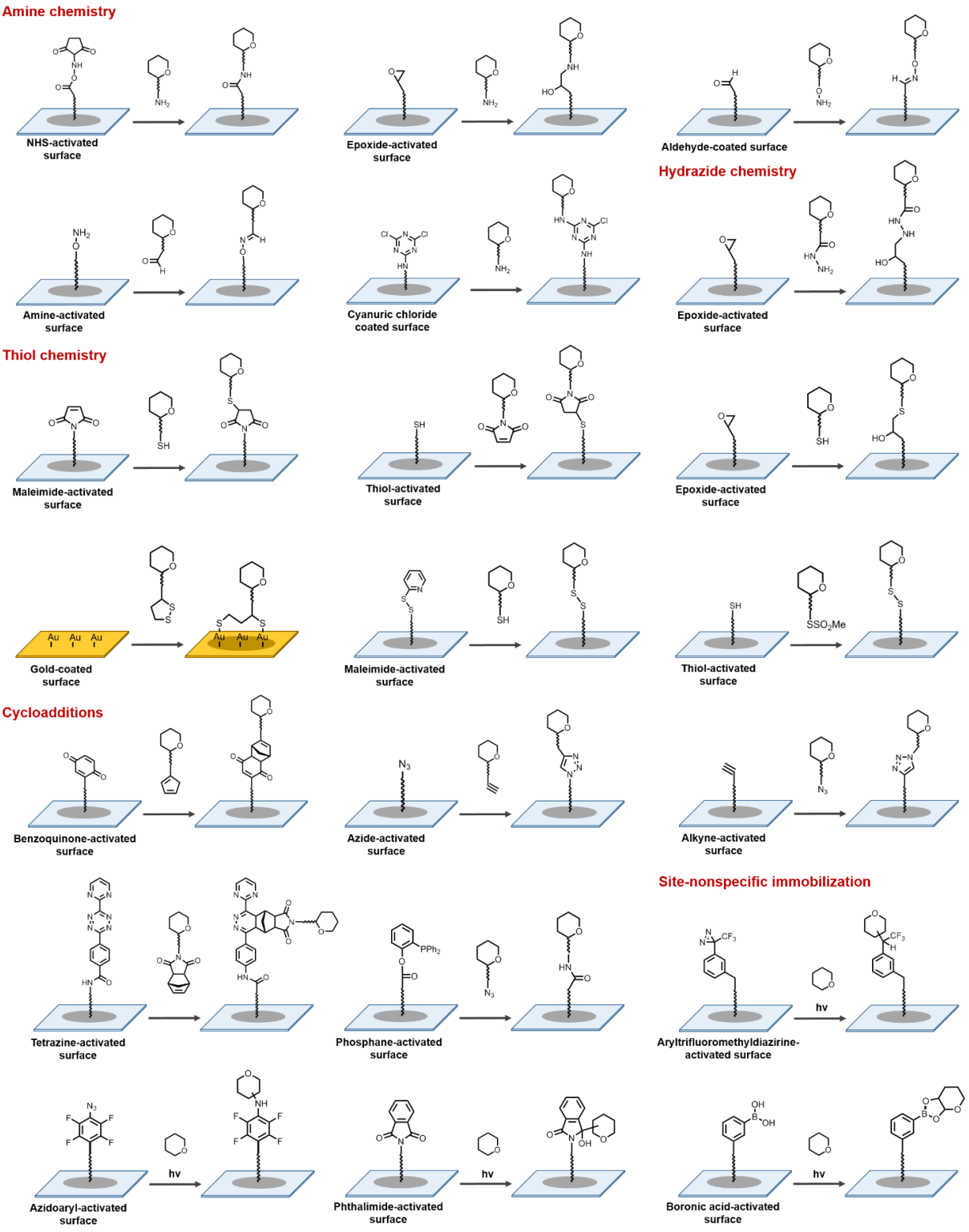



***Figure 36.*** *Covalent immobilization methods. Ideal linking chemistry is based on functional groups easily installed during the synthetic route of the glycan preparation that are also compatible with other functionalities in the assay. At the same time, the immobilization has to be rapid (appropriately reactive), selective and consistent with the biological target and screening system.*

Once the chemical means of immobilization have been established, the glycan solutions are prepared in designated wells of a microplate. During printing, pico or nanoliters of these solutions are transferred to the solid surface and deposited in a pre-arranged, orderly format.

### 2.2.2.5 Ligand spotting

Two common spotting techniques are in use: contact and non-contact printing. In the first case, steel pins are dipped into the ligand solution wells, then moved over the slide and shortly pressed against its surface. The duration of the contact is a key parameter. Up to 64 needles can be used simultaneously; however, the uncontrolled shape of the spots is a main drawback of this method.

Non-contact spotting consists of a substantially more complex printing system, based on inverse piezoelectricity.[352] The original (direct) effect was discovered by the Curie brothers – upon mechanical stress electric charge accumulates in certain solid materials. Inverse piezoelectricity, on the other hand, describes the phenomenon when applied electric current leads to deformation in the material, and this effect is exploited in piezo-electric microarray printers. A thin glass capillary (nozzle), surrounded by a piezoelectric quartz crystal, is filled with the printing solution. When the nozzle is moved to the microarray slide, controlled electric signals are applied on the crystal, which, in turn, flexes in proportion to the current. The deformation of this crystal applies pressure on the capillary and so, squeezes the solution out of the nozzle. The drop size and thus the volume of the deposed solution can be tuned by varying the applied current. (**Figure 37**)

Non-contact printing is significantly more precise, and yields consistent, well-controlled drop size which is crucial for reproducibility. At the same time, the ligand distribution is more homogeneous across the spots, and the leftover solution that was sucked into the nozzle can be later returned to the sourceplate. On the downside, this system requires a more elaborate technology and the number of nozzles is a limiting factor - maximum 4-8 needles can work parallel - which greatly increases the printing time.

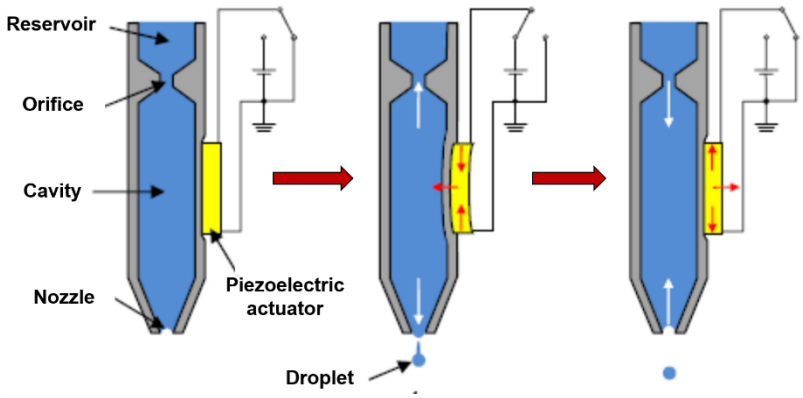



***Figure 37.*** *Non-contact deposition of printing solutions. First, the glycan solution is sucked into the cavity from the source microplate. Once the nozzle is positioned over the slide, electric current is applied on the piezo-electric quartz crystal surrounding the cavity. As the crystal flexes, droplets of the carbohydrate solution are squeezed out of the tip and dripped on the solid support. Printing by non-contact nozzles results in better-defined drop volume, as well as spot size and shape. Modified from: Faulkner et al.*[353]

### 2.2.2.6 Detection techniques

After disposing the glycan solutions on the surface, the slides are left to dry, usually under controlled conditions with respect to temperature and humidity. In case it is necessary, unreacted functional groups are quenched, followed by the washing of the chips. Ideally, the printed slides can be stored at this stage for longer periods and later microarray screenings can be performed when required. At this point, the glycan microarrays are ready to be interrogated with GBPs, and the binding events can be analyzed with a variety of read-out methods.

#### 2.2.2.6.1 Fluorescence-based read-out

Without doubt, the most widespread optical detection method is based on fluorescence. There are three major techniques to visualize the proteins bound to the printed spots. In the most straightforward case, the proteins are directly labeled with fluorophores. In the alternative approaches, the slide is first incubated with the target GBP, then washed and incubated either with a fluorescently-tagged secondary reagent that recognizes the protein (e.g. fluorescent antibody); or with a fluorescent secondary reagent that binds to a tag previously attached to the target (e.g. biotin-streptavidin pairs or His-tags). (**Figure 38**)

The commercial availability of tags and well-established labeling protocols are among the most attractive features of this detection method. Furthermore, high resolution (sub-micrometer precision) microarray scanners are available in many laboratories, and these can detect femtomoles of bound proteins. The fluorescent signals are proportional to the number of bound proteins, and their quantification is done with specialized softwares. These typically create a pixel-by-pixel map of the chip surface and identify different spots, based on the signal/background ratio in the vicinity of each pixel. The local background signal is then subtracted from the mean/median fluorescent intensity across these fitted circles and the intensity values are organized into graphs for a better visualization and easier comparison of binding events in each spot.

Ideally, several screenings should be carried out at different GBP concentrations, in order to optimize the obtained signals. As mentioned earlier, in case of saturated signals, smaller differences in affinity cannot be appreciated, whereas decreased protein concentration allows us to highlight and select only the most potent ligands.

The major uncertainty surrounding fluorescent-based detection is the labeling step because this can affect (reduce or alter) the biological target protein's binding properties, in case the fluorescent tag is attached around the CRD and hampers binding interactions. The degree of labeling is rarely established, although it is an important parameter of detections. Furthermore, secondary reagents are not always available to recognize the target GBP and they can, in some cases, show affinity towards the glycan ligands. Because of these considerations, appropriate controls should be used at all times and evaluated carefully. Additionally, fluorophore tags are often light- and oxidation-sensitive, therefore their intensities can fade over time. As an alternative, fluorescent linkers, such as the earlier discussed 2,6-diaminopyridine have been introduced to tag the glycan-partner.[304e, 310]

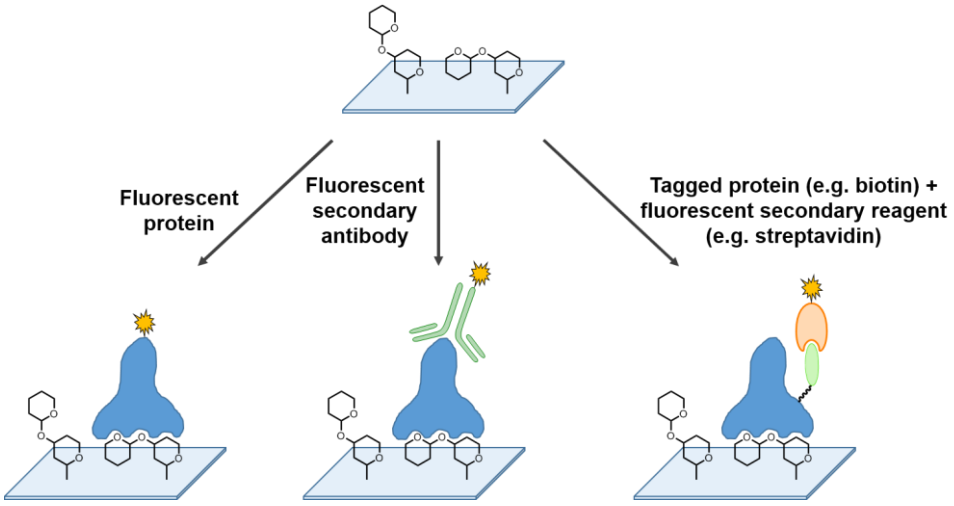



***Figure 38.*** *Fluorescence-based detection methods. Binding events can be detected by probing the microarray slides either with fluorescently-tagged lectins, or by performing a following incubation with fluorescently-labelled secondary reagents. Such high-affinity agents include antibodies or biotin/streptavidin pairs.*

After incubation, the slides are washed again and dried, before scanning. These steps could be in some cases disadvantageous, particularly when the GBP and its ligand dissociate rapidly. Evanescent-field fluorescence scanning can offer a solution when sensitive, real-time detection is required, because the detection is performed from under the chip, without any further washing.[354] Other optical read-out techniques are based on colorimetry[355] and chemiluminescence.[356]

#### 2.2.2.6.2 SPR detection systems

In order to prevent any disadvantageous effects in recognition caused by different tags on the glycan or protein partners, two main label-free methods have been developed to detect events on the glycan microarrays. SPR chips belong to the optical detection systems that allow detection of binding events deduced from the mass change of a thin, glycan-functionalized gold-chip (see **2.2.6.3**). Upon interactions between the proteins and their ligands attached to the surface, changes in the resonance angle are recorded in sensograms as a function of time, and so, these systems provide a real-time insight to the affinity, specificity and kinetic aspects of binding events.[321a] At the same time, however, they require functionalized gold-platforms and a complex, elaborate optical system, as opposed to the simple fluorescent scanners. SPR microarrays have been applied to identify non-natural galactoside ricin ligands,[321b, 357] and to detect glycan-specific antibodies in serum from *S. mansoni*-infected population.[321b]

#### 2.2.2.6.3 MS-based detection systems

The other main label-free detection technique involves analysis by mass-spectrometry.[322a, 358] Changes in the mass of glycan structures which are non-covalently attached to target plates (e.g. by self-assembling monolayers, SAMs) can be followed by MALDI-Tof analysis. The method is widely applied to monitor on-chip modifications of the immobilized sugars, such as their enzymatic transformations (e.g. by glycosyltransferases);[358-359] but it has also been used to identify GBPs from complex mixtures.[328]

Other detection techniques

Additional, but less common examples include detection based on radioactivity,[349, 360] ellipsometry,[361] integrated complementary metal-oxide semiconductors,[362] cantilever array sensors,[363] or oblique-incidence reflectivity difference microscopy.[364]

As the previous discussion demonstrates, microarray systems can be constructed by combining a large number of available options, or even by developing new platforms that are suitable for the objectives of our research. In all cases, choosing the appropriate technology and its elements, adjusting the parameters to achieve optimized performance or to establish a standardized protocol requires careful planning and experimenting. Furthermore, the results must be reproducible, consistent and – if possible - validated, for example, with data obtained in alternative assays.[314b]

## 2.2.3 SETUP DESIGN AND OPTIMIZATION OF A GLYCOMIMETIC MICROARRAY CHIP TO SCREEN FLUORESCENTLY LABELLED LECTINS[b]

The benefits of glycan microarrays were detailed in the previous chapter, with emphasized focus on running hundreds of binding assays in parallel, accompanied by an extremely low ligand and protein demand. Our choice fell on this technique because of its ability to highlight a lectin's preferential binders among a larger pool of ligands. At the same time, cross-array comparison could indicate possibly selective ligands that are recognized by a CLR, but not by another. These are, in fact, the main goals of our project – to find high affinity and selective binders of the available CLRs, in order to further investigate these findings. Once leads are identified by on-chip screening, validation of the results is performed by complementary techniques, and the most promising candidates can be used for the construction of multivalent systems.

It is of crucial importance, however, to be aware of each technique characteristics, limitations and restrictions, especially, when comparing results among different array-based screenings or with other binding assays (e.g. SPR). The results of microarray screenings are determined by numerous factors, most of them having main influence on the reliability and reproducibility of the obtained data.

In spite of the randomness of immobilized ligands on the solid support, the affinity of lectins towards the multivalently presented structures is higher (surface dissociation constant - $K_{D\ surface}$ - in the nanomolar range) than the affinity measured for monovalent ligands in solution ($K_D$ in the micromolar range).[304b] An SPR competition experiment where the monovalent ligands are studied in solution phase is therefore not directly comparable to the results of a microarray-based screening, the two assays should be used to complement each other. In theory, the results of a microarray screening could help to pick ligands to build other multivalent constructs, such as glycodendrimers, whereas SPR analysis is an ideal assay to select high affinity monovalent ligands. The latter technique, however, requires substantially higher amounts of both ligands and proteins, and a complex optical detection system (see **2.2.6.3**).

The presentation of carbohydrate structures is crucial, as this can affect not only the binding affinity, but the selectivity as well. This is influenced by the carrier platform, as well as by the carbohydrate structures' and additional linkers' size and flexibility.[223a, 365]

### 2.2.3.1 Ligand immobilization on the microarray surface

Having emphasized the importance of ligand presentation, the appropriate immobilization chemistry had to be selected in order to anchor our glycomimetics on the microarrays chips. Based on the previously established glycan-microarray setup in the laboratory of Dr. Niels C. Reichardt,[300] we intended to covalently attach the library in a compatible format on commercially available NHS-functionalized glass slides with a robotic non-contact piezoelectric microarray spotter, and then to probe the arrays with fluorescently labeled lectins. NHS-groups on the chip surface could react with amino-functionalities – and, in principle, the terminal azides could be simply and rapidly reduced into amines, for example in H-cube® flow hydrogenators. However, as it has already been anticipated in Chapter I, it was uncertain whether the short azidoethanol-linker attached directly to the sugar-mimic cyclohexane ring will provide sufficient distance between the three-dimensional, permeable hydrogel coating on the solid glass surface and the ligands. This can hinder the CLRs proper access to these ligands and furthermore, it may affect the conformation of the glycomimetics – in conclusion, it could lead to

[b] Microarray experiments and related analyses were carried out during a secondment in the laboratory of Dr. Niels C. Reichardt, at CIC biomaGUNE, San Sebastián, Spain, with the help of Dr. Sonia Serna.

biased results. This negative effect has, in fact, been demonstrated earlier by Park *et al.* who tested linkers of different length, and found that the performance of the microarrays was the lowest in the case of the shortest spacers (weak signals below 5 mM printing concentration).[332a, 332b] Additionally, the presence of hydrophilic moieties in the linker - e.g. oligo- or poly(ethylene glycol) (PEG) – improves the overall binding properties compared to hydrophobic linkers, and minimizes aspecific protein binding.[338]

As our attempts to install a longer, or more hydrophilic linker during the synthetic route failed (see **2.1.3**), we decided to elongate the tether with a hetero-bifunctional spacer. Several aspects had been considered for the selection of this spacer. An optimal choice consists of a flexible, extended structure that contains a reactive group ready to form a covalent bond with the NHS-activated surface and an additional functionality that will link the glycomimetic ligands with the help of the azido-group at the end of the shorter tail. Additionally, the solubility and hydrophilicity of the system could benefit from, for instance, the earlier mentioned oligo ethylene glycol moiety in the spacer's structure. On the other hand, the spacer should not negatively affect the functionality of the glycomimetic ligands, their recognition on-chip and the outcome of the screening – including any aspecific interaction with the proteins (i.e. false positive results).

The already incorporated azido-group on the glycomimetic compounds can be directly exploited in CuAAC-reactions with alkynes, however, the reaction requires copper(I)-catalyst.[366] This method has several disadvantages: on one hand, the oxidation state of Cu(I) needs to be maintained, typically by the addition of the reducing agent sodium ascorbate and the chelating agent TBTA (tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine) that stabilizes Cu(I) without interfering with the catalytic cycle. The reactions are sensitive to the presence of oxygen, both in the atmosphere and in the solvents and isolation of the product often requires a purification step. On the other hand, the added heavy metal catalyst is potentially cytotoxic, and even the smallest amounts of residual copper could limit the possible future utility of the library in cellular assays.

Azides can also react with phosphines in the Staudinger reaction, and this has been used to covalently conjugate glycan structures to biopolymers through an amide bond.[367] The Staudinger ligation is one of the first completely biocompatible reactions described that can be carried out traceless without any catalyst, but the reaction rate is slow.

#### 2.2.3.1.1 The strain-promoted alkyne-azide[3+2]cyloaddition (SPAAC)

In search of further biorthogonal reactions with more attractive kinetics, the Bertozzi group reported another highly selective reaction a few years later, involving an azido-group and a triple bond.[368] Contrary to the original Huisgen 1,3-dipolar cycloaddition, or the Sharpless-type CuAAC reactions, these "click"-type reactions are not initiated by heating or by any added metal catalyst, but are accelerated by a strained triple bond. This is an alternative route to overcome the activation barrier of the reaction by deforming the alkyne's bond angle and these reactions are therefore called strain-promoted alkyne-azide cycloadditions (SPAAC). The simplest method to strain the bond is its inclusion in a monocycle, and the smallest stable ring containing a triple bond is the cyclooctyne where the bond angles of the sp-hybridized carbons are around 160°.

SPAAC reactions are driven by the interaction between the lowest unoccupied molecular orbital (LUMO) of the alkyne and the highest occupied molecular orbital (HOMO) of the azide. According to the distortion-interaction theory, the transition state (TS) occurs when the orbital overlap between bimolecular fragments is maximal, and the gaps between frontier molecular orbitals (FMO) is narrow enough for stabilizing interactions to overcome destabilizing distortions.[369] (**Figure 39**)

$$\Delta E^{\ddagger} = \Delta E_d^{\ddagger}\,(destabilizing) + \Delta E_i^{\ddagger}\,(stabilizing)$$

It was shown that - compared to strain-free acetylene - in cyclooctynes the barrier heights are rather controlled by differences in $\Delta E_d^{\ddagger}$ (the energy required for the distortion of the 1,3 dipole and the dipolarophile), than by FMO interactions or thermodynamics. The transition state is formed earlier, already at longer C-N bond lengths and larger N-N-N dipole angles (173°). Cyclooctyne in its ground state is already more bent (C-C-C: 153° and 155°) than the distorted transition state of acetylene (H-C-C: 158° and 166°). In accordance with this, the reaction rate is substantially increased ($10^6$), leading to the strain-promoted cycloaddition where no additional catalyst or heating is necessary.

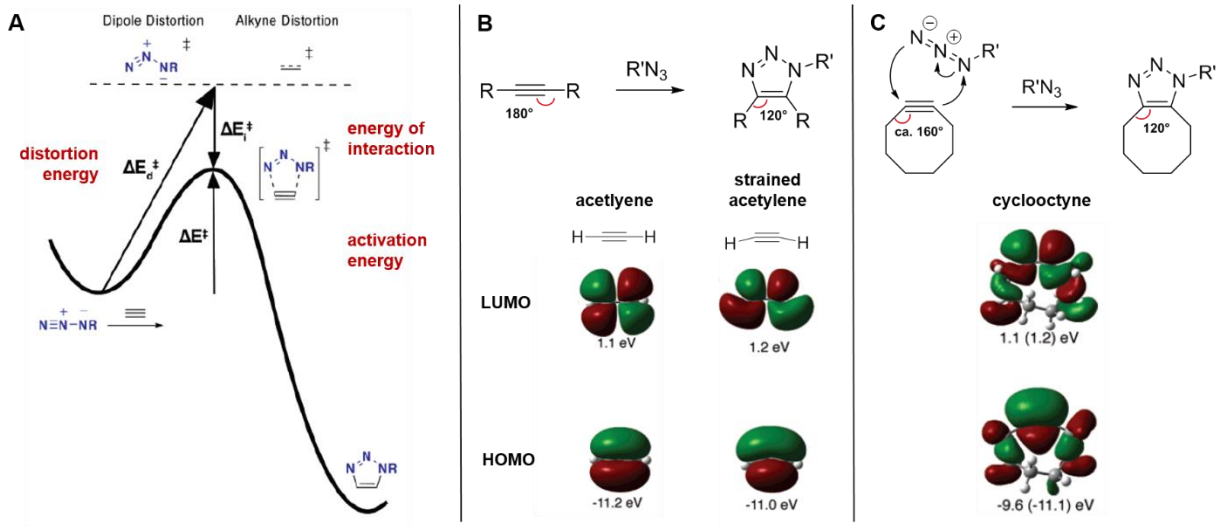



***Figure 39.*** *The strain-promoted alkyne-azide cycloaddition (SPAAC). A: Energy levels of undistorted dipole and dipolarophile reactants, the transition state and the triazole product. SPAAC reactions are driven by the strained triple bond (lowered distortion energy) and, to a smaller extent, by frontier molecular orbital (FMO) interactions. B: Orbital energies (eV) and geometries in relaxed and strained acetylene. C: The distortion of ground state cyclooctyne exceeds that of the distorted transition state of acetylene. Modified from: Ess et al.*[369a]

Another advantage of this system is that the reactivity can be tuned by steric and electronic effects. Different strategies have been employed to vary the properties of the cyclooctyne system, for example, the alkyne LUMO can be lowered (and HOMO increased significantly) by the addition of electron withdrawing groups to the ring,[370] or increased strain energy can lead to enhanced reaction rate (dibenzocyclooctyne derivatives).[371]Installation of two fluorine atoms (DIFO – difluorinated cyclooctyne),[370] bicyclic constructs (BCN - bicyclononyne),[372] as well as benzo-fusion and amide nitrogen containing (ADIBO - azadibenzocyclooctyne)[373] rings have been reported, their reactivity increasing from the first to the last. (**Figure 40**)

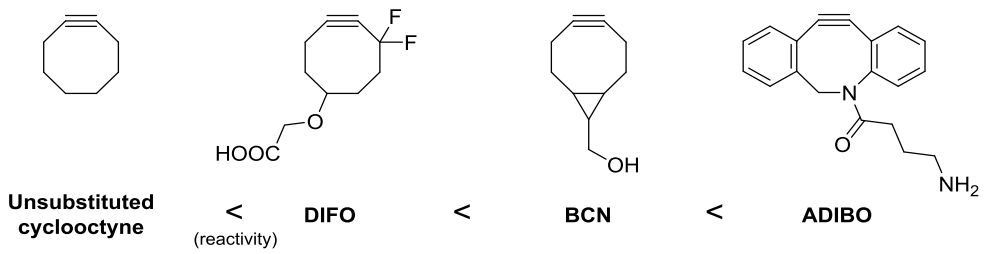



***Figure 40.*** *Differently activated cyclooctyne structures. Difluorinated (DIFO), bicyclic (BCN - bicyclononyne), as well as benzo-fused and amide nitrogen containing (ADIBO - azadibenzocyclooctyne) cyclooctynes were described. By modifying the substituents in the vicinity of the triple bond, for example by the addition of electron-withdrawing groups, the ring-strain can be tuned and optimal reactivity can be achieved.*

SPAAC reactions are particularly beneficial for our purpose: in addition to the absence of catalysts and additives, they are extremely selective and high-yielding, can be performed in water and are not sensitive to air. The conversion is rapid even at room temperature and does not require heating, pressure, or high concentrations, moreover, there are no toxic reagents involved.

At the time of our microarray experiments, two main types of cyclooctynes were commercially available: BCN and ADIBO-type spacers. Both could be purchased with a terminal amino-group and in many cases, this functionality was found at the end of a PEG-linker. Extended lipophilic systems, such as the dibenzocyclooctyne moiety, however, could have a higher risk of interacting aspecifically with different sites on the proteins, and so, we selected N-[(1R,8S,9s)-bicyclo[6.1.0]non-4-yn-9-ylmethyloxycarbonyl]-1,8-diamino-3,6-dioxaoctane (Sigma-Aldrich®, from now on: **BCN-amine**) as the spacer that will link our glycomimetic ligands to the NHS-functionalized glass slides. (**Figure 41**)

Previous studies demonstrated that even the smallest differences in presentation – e.g. linker length – can have major effect on the presentation, accessibility and overall recognition of glycomimetics,[332a, 374] and that the best results are typically obtained at high printing concentrations.[304b, 314a] Scenarios where the compound is not accessible, or not available at the appropriate concentrations to detect the otherwise weak carbohydrate-lectin interactions, can easily lead to false negative results. Any adverse effects caused by the linker could influence the screening results, therefore a careful evaluation of the spacer's role, supported by different experiments, concentrations and controls was indispensable.

#### 2.2.3.1.2 Ligand immobilization trials

Immobilization of the ligands through the doubly functionalized linker requires two reactions - clicking the glycomimetics on the cyclooctyne ring in a SPAAC reaction, and forming of a covalent bond between the amines at the end of the BCN-amine tail and the NHS-groups on the glass slide. These can be performed in different order: we can either first coat the entire microarray surface with the spacer moiety and subsequently spot the ligands at allotted positions on the BCN-modified slides, or, we can first form the triazole-linked adduct of the ligands and the spacer, and then directly print these conjugates on the chips. (**Figure 41**)

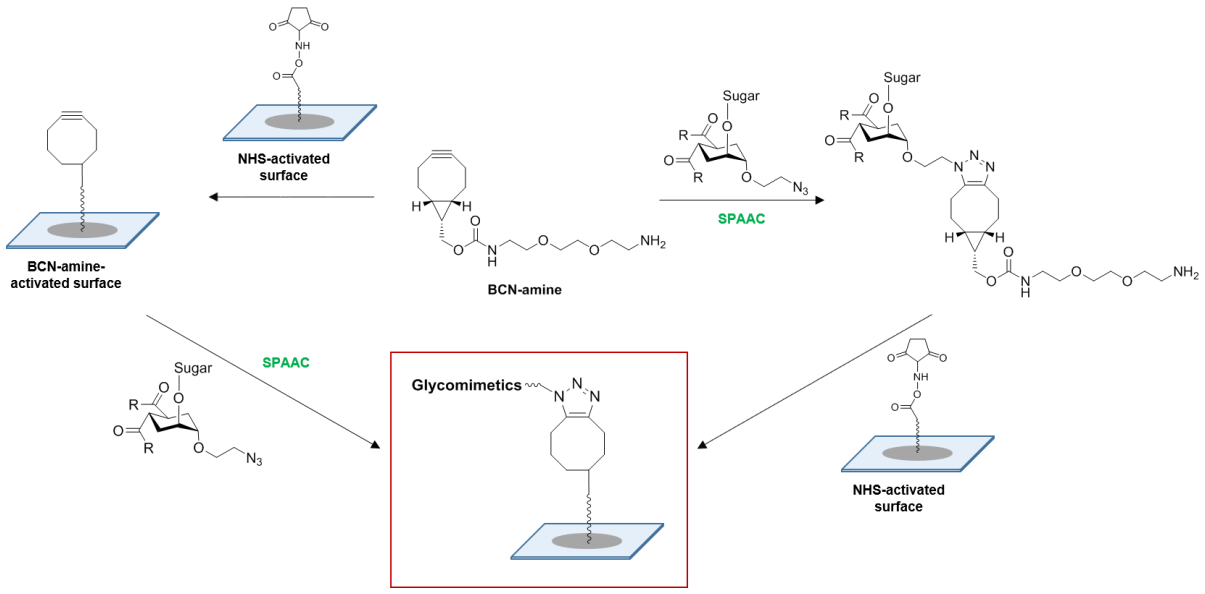



***Figure 41.*** *Glycomimetic library immobilization through the additional BCN-amine spacer. Left route: NHS-functionalized glass slides are coated with BCN-amine and the cyclooctyne-activated surface is ready to react with the terminal amino-groups found on the ligands in SPAAC reactions. Immobilization through this strategy failed, most likely due to the thermal instability of cyclooctynes during the microarray fabrication. Right route: Individual coupling of ligands and BCN-amine linkers, followed by spotting on NHS-coated slides. As the conjugation step is fast, and the product can be analyzed rapidly by MALDI-Tof spectroscopy, we chose this pathway to prepare the glycomimetic chips.*

In the first case, by functionalizing the whole slide with the cyclooctynes, we could avoid the individual conjugation of every ligand separately, but, on the other hand, the coating step needs to be well-controlled for good reproducibility. Surface modification with SPAAC-reactive cyclooctynes has been demonstrated in a few cases before, primarily with ADIBO-type cyclooctynes,[375] and with BCN-amine on NHS- and pentafluorophenol-activated surfaces.[376]

In our trials, NHS-activated glass slides were incubated overnight with buffered BCN-amine solution at pH ~ 8.5. The concentration of BCN-amine was chosen in order to have an excess number of molecules to react with all NHS-groups found on the surface, based on previous calculations for these glass slides ($10^{14}$ molecules/cm$^2$).[304b] This number was estimated by Liang *et al.* who printed fluorescein isothiocyanate cadaverine (FITC) at varying concentrations and performed fluorescence-wash off experiments to quantify the maximum loading. The number is similar to the sugar (or peptide) functionalization capacity of SPR biosensors.

In order to evaulate the viability of this strategy, after functionalization of the slides with BCN-amine and blocking unreacted sites with ethanolamine solution, we spotted Lac-$CH_2$-$N_3$ at 500 μM with the help of the automated microarray printer. The slides were incubated for 1 hour with fluorescently labeled peanut agglutinin (PNA-AF555) - a lactose recognizing plant lectin – washed and scanned.

The scanned images showed that although the lectin was functional, no recognition was observed on the BCN-amine modified slides where azido-lactose was printed. We suspected, that in spite of the previously reported successful surface modifications with cyclooctynes and following on-chip SPAAC reactions, our surface functionalization failed presumably due to the low thermal stability of the BCN-cyclooctyne ring.[372]

The second strategy based on the conjugation of the ligands with BCN-amine, followed by spotting the adducts on the NHS-activated slides, was evaluated in a similar system. The SPAAC reaction between equimolar amounts of lactose$(CH_2)_2N_3$ and BCN-amine (5 mM each) was tested in three different solvents – water, DMSO and methanol - and the conversion was monitored by MALDI-Tof mass spectrometry. The click-reaction, yielding lactose-BCN-amine (**LacBCN**) was completed after 1.5 hours in water, essentially without any residual starting

materials or side-products, therefore this was later used as the primary solvent during coupling of the glycomimetic ligands. Furthermore, the product did not require any additional purification step, and the lactose-BCN-amine conjugate aqueous solution was directly used for printing.

Different concentrations (10, 50, 100 and 250 μM) of **LacBCN** were printed on NHS-activated glass slides. On the same slide, **Lac$(CH_2)_5NH_2$** was directly printed serving as positive control, whereas the non-glycosylated linker (azidopropanol-clicked BCN-amine, **AZBCN**) was used as negative control, to check any aspecific interaction between the bulky cyclooctyne moiety (without sugar moiety, only with the alkyne blocked) and the lectins. (**Figure 42**)

After printing, the slides were left to dry, the unreacted functional groups were quenched with ethanolamine, and the chips were washed with water. The microarrays were incubated with fluorescently-labelled lectin for an allotted time, washed again, scanned and the fluorescent signals were quantified by software-based analysis. The fluorescence intensity value of individual spots directly correlates with the number of ligand-bound proteins and gives an approximation of the relative binding strength of the interaction between the lectin (in solution phase) and the surface-immobilized ligands.

After the incubation with PNA-AF555 solutions at two different concentrations (0.25 and 0.063 mg/mL), an increasing intensity in the fluorescent signals was observed at higher protein concentration.

Clear and well-defined spots and basically no aspecific interaction with the **AZBCN** control was observed on the scanned image. Interestingly, BCN-conjugated lactose resulted in higher fluorescent signals than lactose equipped with only the aminopentanol linker, suggesting that the ligand is more accessible for the lectin when presented with the additional spacer.

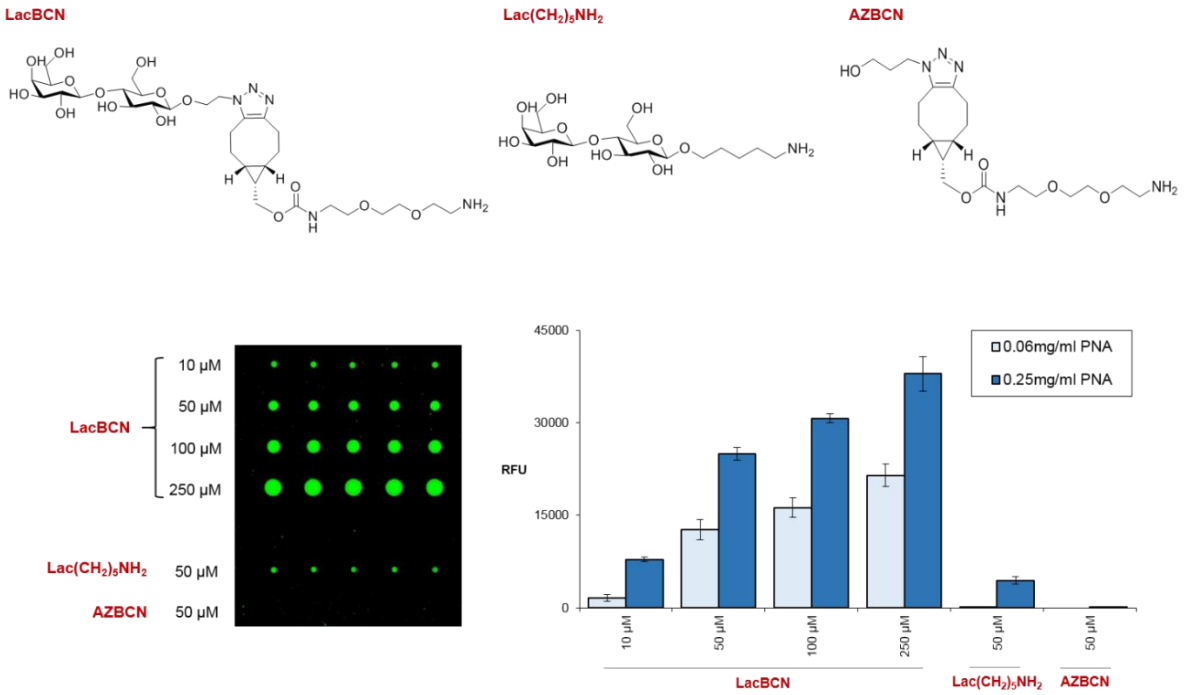



***Figure 42.*** *Printing trials with lactose and peanut agglutinin (PNA). Lactose$(CH_2)_2N_3$ was coupled with BCN-amine and the conjugate was printed on the NHS-activated slides at different concentrations. Two incubations were performed: at 0.06 μg/ml and 0.025 μg/ml protein concentrations.*

#### 2.2.3.1.3 Controls

In order to rule out any disadvantageous effects of the BCN-linker on the recognition of glycomimetic ligands by other lectins, (CLRs), two compounds (**ISh045** and **ISh046**) had been included in the library (see Chapter I). In order to evaluate their presentation, these compounds were printed on the NHS-modified surface directly functionalized with aminoethanol linker (without spacer) and indirectly conjugated to BCN-amine (with spacer). In ideal case, the introduced spacer should ensure a better access to the proteins to reach the ligands anchored to the three-dimensional polymer coating on the chips, and should not interfere with the lectins' binding properties. These ligands are well-characterized and their binding properties towards some CLRs (DC-SIGN, langerin) had been studied, thus, together with **AZBCN**, they could serve as ideal controls during the validation of the microarray setup. (**Figure 43**)

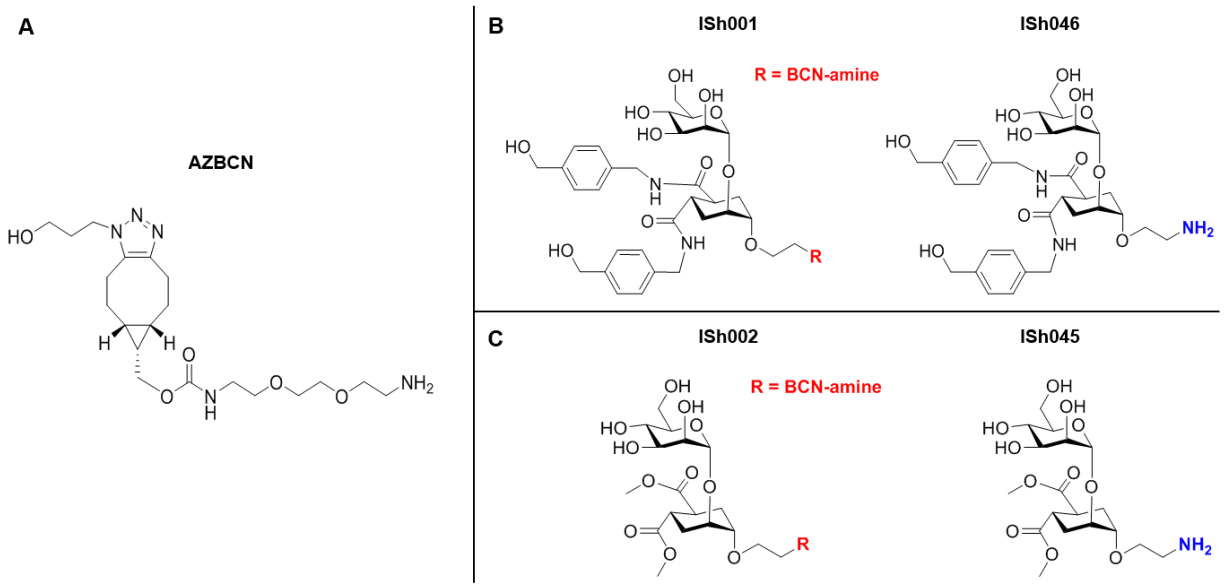



***Figure 43.*** *Controls included in the microarray experiments. A. Non-glycosylated BCN-amine spacer (clicked with azidopropanol,* ***AZBCN****) was used to test aspecific binding between the lectins and the additional spacer. B and C. Two previously studied ligands were both indirectly (****ISh001*** *and* ***ISh002****) and directly (****ISh046*** *and* ***ISh045****, respectively) printed on the microarray slides, in order to evaluate their presentation*

As the second immobilization strategy – clicking ligands on BCN-amine in solution through SPAAC reaction, followed by immobilization on the activated solid platform - was proven to be viable, we performed the conjugation on every glycomimetic ligand in our library. MALDI-Tof analysis before and after the conjugation, and further experimental conditions are detailed in the experimental section. SPAAC reactions were performed primarily in water, but due to the poor solubility of some glycomimetic structures (**ISh006**, **ISh013**, **ISh018**, **ISh021**, **ISh027**, **ISh032**, **ISh033**), around 5% DMSO was added to these reactions. The resulting solutions after conjugation were diluted with printing buffer and were prepared in a 384-well microplate that can be stored over longer periods at -20°C.[c]

Immobilization validation was performed by probing the printed glycomimetic array with common lectins of known (and broad) specificity. These included plant and fungal lectins, acquired from commercial source. These lectins were labeled with Alexa Fluor® fluorescent dyes (see experimental section for details) and their degree of labeling (DOL; an approximation of the ratio of dye and protein in the labeled proteins) was estimated by

[c] Note: In the following part of this chapter, unless stated otherwise, ligands labeled as **ISh0XX** refer to the full (BCN-amine conjugated) structures.

absorbance analysis. Although the DOL is often not specified in literature, rate of fluorescent tagging can have major effects on the efficiency of detection. Fluorescence tagging poses a number of uncertain factors, apart from the variable DOL values, for example, labelling can occur in the CRD of the receptor and this can severely influence the binding interactions.

### 2.2.3.2 Microarray setup optimization

The first printing setup designed to include the entire glycomimetic library consisted of spotting 5 drops of ca. 250 pl of 50 μM solutions in 5 replicates on each subarray on a ca. 7.5 cm long, 2.5 cm wide functionalized glass slide that was divided into 7*2, altogether 14 subarrays. This subdivision of the slide helps to minimize the volume of protein solutions needed to cover the whole printed surface - in this case 100 μl protein solution was sufficient to probe 15*16=240 individual spots on each subarray that allows the parallel assaying of 48 different ligands printed in 5 replicates. (**Figure 44**)

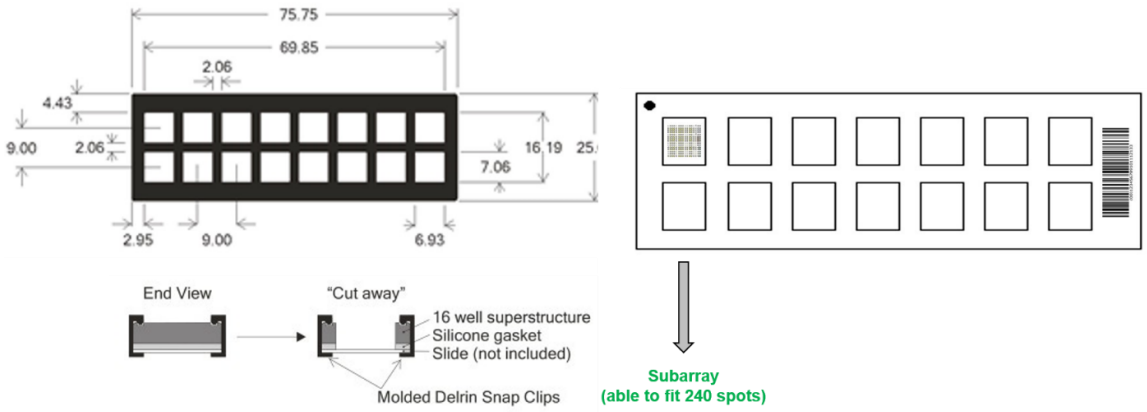



***Figure 44.*** *Microarray slide and plastic/silicone gasket. The chip is divided into smaller subunits to minimize the protein consumption, and to perform several screenings on the same slide. Depending on the design, the setup and gasket is selected to fit all spots (typically several replicates in the same ligand) on a minimal surface. The gaskets are fixed on the glass slides with steel clips and the incubating solutions are then loaded into the gasket wells.*

In order to validate the printing setup, assess the effects of the included spacer and confirm the homogeneous immobilization of ligands, we first used the mannose specific jack bean (*Canavalia ensiformis*) lectin ConA-AF555 to interrogate the glycomimetic chip.

#### 2.2.3.2.1 The coffee-stain effect

Two things were apparent on the scans already at first sight: the spots were relatively too large on the subarray, positioned too close to each other, and the majority were bordered with blurred rims. Even more importantly, an uneven distribution of the glycomimetic ligands over some spots was observed. Further inspection revealed, that this effect occurs most prominently in the case of fluorine-atom and trifluoromethyl-group containing glycomimetics, or extended lipophilic groups (**ISh013**, **ISh022**, **ISh023**, **ISh027**, **ISh029**, **ISh033**, **ISh038**). (**Figure 45**)

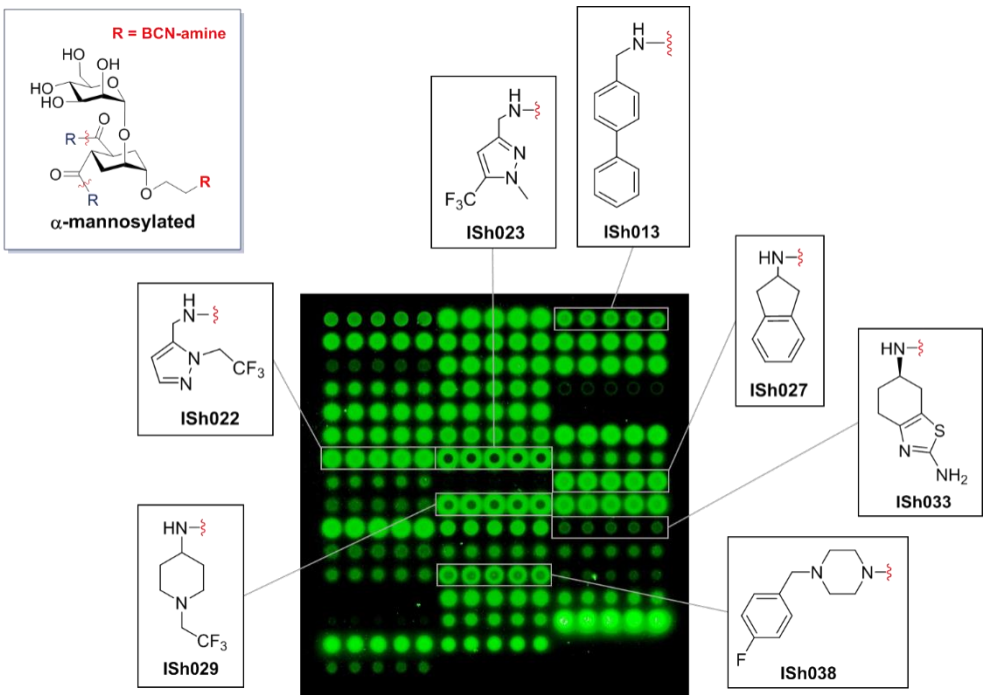



***Figure 45.*** *First screening with mannose-specific Concanavalin A. Three major issues are apparent: the spots are not well-defined, but have blurred edges; they are too large and in some cases almost overlap and most importantly, the ligand distribution is not homogenous over the spot. Fluorine-atom-, or trifluoromethyl-group containing and highly lipophilic ligands tend to accumulate around the rim of the spots, leaving the center less concentrated and therefore darker*

The printing solutions already contained a minimal amount of TWEEN®-20 surfactant and the printing and drying steps took place in humidity-controlled environments (60%), precisely to avoid rapid evaporation and to promote the homogeneous immobilization of ligands. This phenomenon, known as coffee-stain, coffee-spot or doughnut-pattern effect, can represent a major issue during quantitative translation of fluorescence signals into intensity values and needs to be solved before proceeding to other lectins. This is necessary because computer-based image processing (as it has been described above) generates a pixel-by-pixel image over the detected fluorescent spot, and calculates intensity values based on this matrix. Inhomogeneous signal intensity could result in difficult or inaccurate quantification, therefore elimination or minimization of this effect became our next goal during the setup optimization.

The observed coffee stains are a result of outward material flow during the evaporation of the drop and at the end, the solute accumulates in the peripheries where the ligands become immobilized. There are several factors contributing to the inhomogeneous disposal of particles: pinning of the three-phase contact line, high evaporation rate at the droplet boundary, surface tension gradients and a convective flux caused by the uneven evaporation rate.[377] In particular, the latent heat of evaporation and the surface tension gradients would lead to a recirculatory flow towards the drop centre (Marangoni effect) that could reverse the ligands distribution along the contact line, and thus prevent the coffee-staining.[378] The fact that water droplets and the interfaces are rarely clear of contaminants suppresses this inward flux. (**Figure 46**)

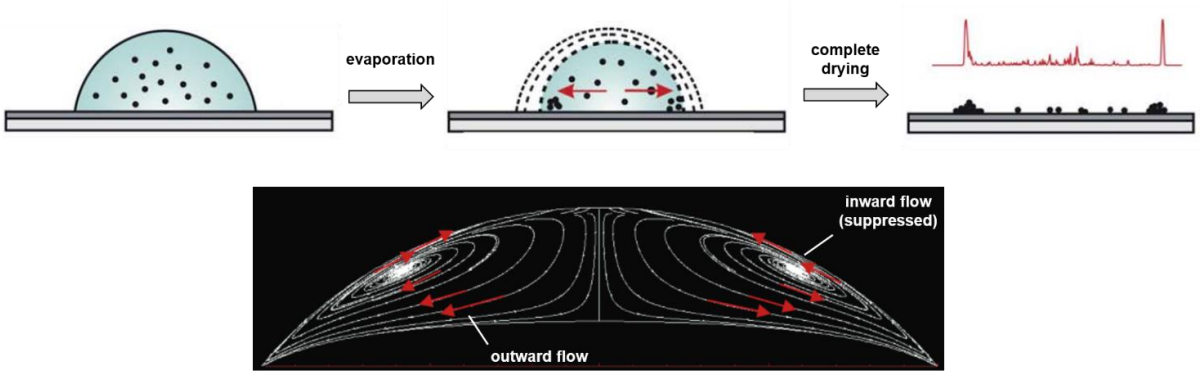



***Figure 46.*** *Formation of "coffee stains" over the printed spots. Surface-tension gradient caused by concentration- or temperature-gradient (evaporating solvent) will induce an outward mass flow over the spot. Eventually, the ligands are depositied and immobilized at the rim of these spots. The fluorescent lectins will therefore bind mainly at the edge, and the inhomogenously distributed fluorescent signals will be inaccurately quantified. Modified from Eral et al.*[378a] *and Hu et al.*[378b]

Coffee-stains are well-known obstacles during microarray fabrication, and solving the problem often relies on acquiring a better control over the evaporation process.[379] We tested several conditions to improve the spot appearance:

- Reduction of the number of spots: after the initial 5 drops of ca. 250 pl solution, we disposed only 3 drops of the same volume on the microarray surface. The smaller volume could yield spots of a smaller diameter, where the mass-transfer to the edge is less pronounced.
- Added glycerol: 0.02% glycerol in the printing buffer could reduce the evaporation rate of the drop and the enhanced viscosity can minimize the inner flow within the droplet.
- Added DMSO: 10% DMSO in the printing buffer could to reduce the evaporation rate and thus, increase the homogeneity of the drying droplet. The additive could also improve the solubility of more hydrophobic structures in the reducing volume of solution during drying.

The scans obtained in each case are shown below. (**Figure 47**)

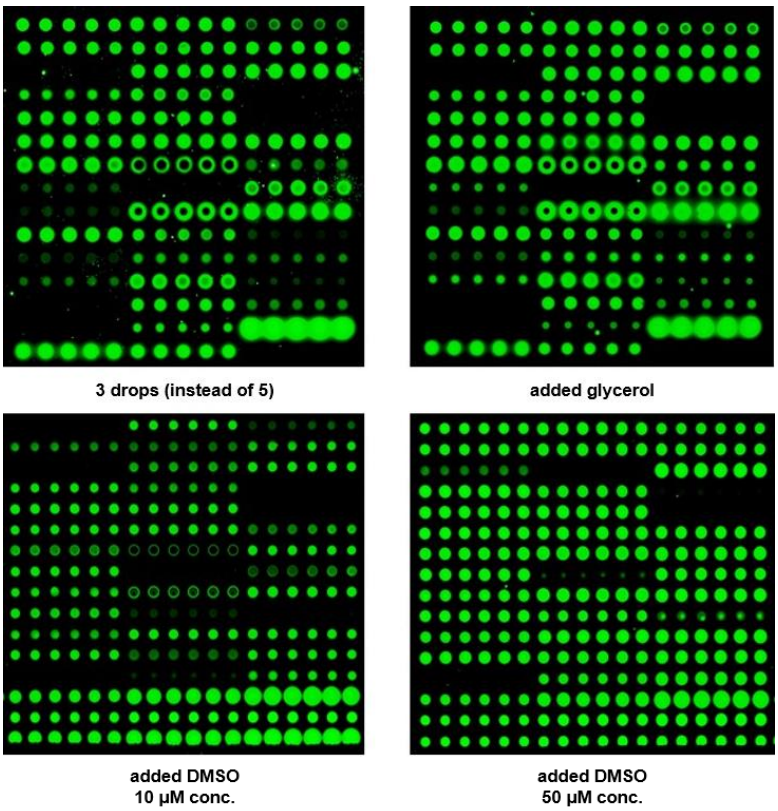



***Figure 47.*** *Effect of additives in the printing buffer on the coffee stain effect. To achieve a better control over the evaporation process, we tried several conditions. Reducing the number of spots (from 5 to 3), additives, like glycerol, or surfactants (TWEEN®-20) present in the printing solution and printing at lower concentrations did not solve the problem of lignad distribution inhomogeneity. Homogeneous, well-defined spots were finally obtained by added DMSO that presumably reduces the rate of evaporation and helps the solution homogeneity of more lipophilic glycomimetics.*

The addition of DMSO in the printed solution clearly improves the homogeneity of solute distribution over the spot and the reduced number of drops (from 5 to 3) helps to separate the spots from each other. These conditions were implemented in our sourceplate preparation and printing protocol from this point, and new microarray slides were prepared with the optimized printing solutions (details in the experimental section).
During the initial microarray screenings, compounds: **ISh016**, **ISh017** (mannosylated bisamides) and **ISh049**-**ISh059** (fucosylated glycomimetics) had not been added to the library yet. These ligands were synthesized after the optimization of the microarray setup, and were later printed along the mannosylated structures, but the screening results containing these structures were obtained by Silvia Achilli and Dr. Sonia Serna. In this thesis, mainly the original results – obtained during my secondment - are reported, but in other cases, when discussion or conclusion requires this, the results obtained by our collaborators are included here. During these assays, the printing and screening conditions established by us were employed, and incubations were performed with lectins that became available only later. Two main printing arrangements had been designed to fit the glycomimetic library in subarrays: design 1 used during screenings without ligands **ISh016**, **ISh017** and **ISh049** to **ISh059** (mannosides and controls) and design 2 – used for screenings performed on the whole glycomimetic library (design details provided in the experimental section).

## 2.2.3.3 Screenings with plant, fungal and bacterial lectins

### 2.2.3.3.1 Concanavalin A (ConA) and other plant lectins

The binding profile obtained at 10 μg/ml ConA concentration is shown below (**Figure 48**)

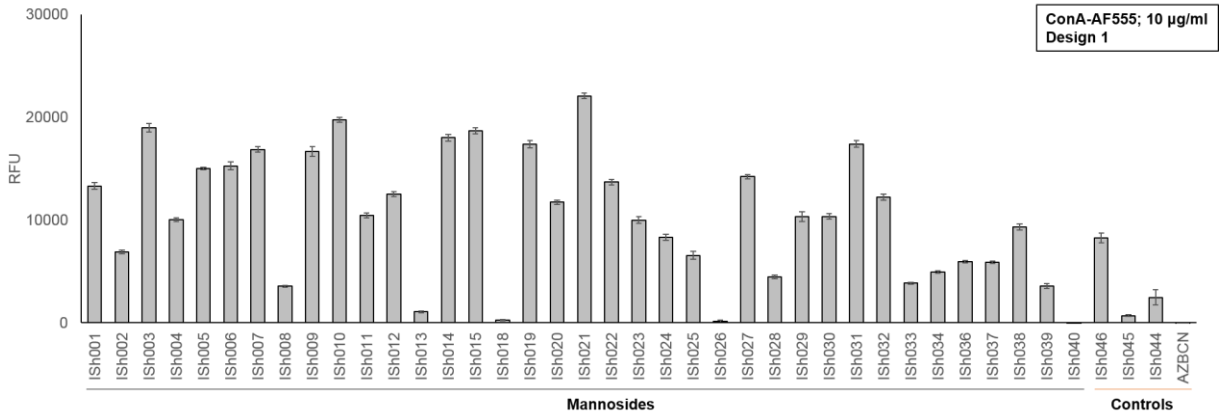



***Figure 48.*** *Screening with 10 μg/ml ConA-AF555 (not full library; only mannosides and controls)*

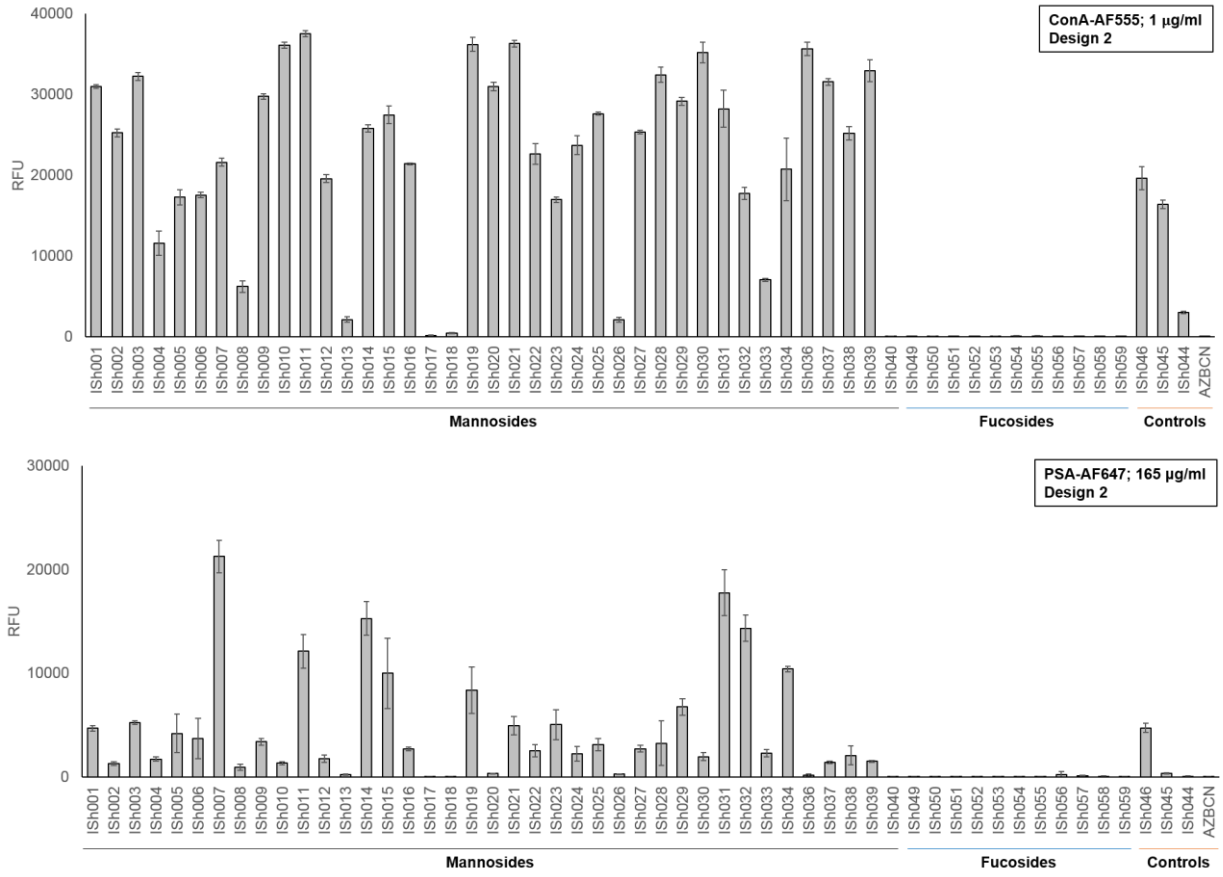
First, the intensity values of altogether 5 control compounds were evaluated in order to assess the role of the additional BCN-amine linker. **AZBCN**, the non-glycosylated spacer moiet blocked with simple 3-azido-1-propanol, was essentially not recognized by the lectin, therefore, we can state that ConA does not bind the linker aspecifically. Comparison of the signals of **ISh001** with **ISh046** (Man030), and **ISh002** with **ISh045** (pseudo-disaccharide dimethylester) clearly indicates that the ligands immobilized through the spacer are better recognized by ConA, than the ligands directly attached to the slide. This confirms that the elongated distance between the chip and ligands improves the binding affinity of the same ligands.

Other important conclusions can be drawn by comparing the intensity of the mannose-control (**ISh044**) to the glycomimetic library, ConA shows stronger binding towards the majority of the mimetic compounds than towards the natural mannose ligand, suggesting that the glycomimetic bisamides are capable of increasing affinity not only in the case of DC-SIGN, but towards other lectins as well. This is a key point, as it demonstrates that our initial hypothesis - being able to target different lectins with a glycomimetic template that had been designed for one concrete CLR – is viable. Moreover, the library of ligands show diverse intensity values towards the protein, and in some cases, they are barely recognized, or not recognized at all. The differential binding indicates that, based on the nature of the bisamide substituents (the only variable structural element in the glycomimetics), the affinity towards different lectins can be tuned, and discriminative ligand recognition can be achieved. This serves as a validation of our glycomimetic ligand library that was conceived in order to identify good affinity and selective CLR ligands in the first place.

Additionally, the results verified the applicability of the designed microarray setup, as gradually increasing concentrations of the incubating lectin solution lead to increasing signal intensities. At high concentrations, the signals reached saturation and the array lost its sensitivity. Ideally, a series of experiments at various protein concentration levels are necessary to identify both high potency ligands (picked at low concentrations where they "stand out" among the other ligands) and structures that are not recognized by the lectin (identified when they do not show binding even at high lectin concentrations). Unfortunately, in many later cases, the limited amount of CLRs available did not allow us to perform microarray screenings at more than a few concentrations.

To test the stability of the chips, we printed several slides under the same conditions and incubated some of them immediately after drying, while others were stored at -20°C for two weeks, before interrogation (at the same ConA concentration). The results were identical in both cases, thus we concluded, that the time-consuming printing step can be carried out in advance and the slides can be used at a later point.

Once the entire library (including **ISh016**, **ISh017** and the fucosides – **ISh049**-**ISh059**) was printed on the microarray chips, the slides were probed again with ConA. As expected, the mannose-binding lectin did not recognize any of the fucosylated structures. The glycomimetic chips were interrogated with other mannose-specific plant lectins, such as *Pisum sativum* agglutinin (PSA) and *Lens culinaris* agglutinin (LCA), and similarly to ConA, they recognized the mannosylated structures at various levels, but not the fucosylated ones. These lectins are common diagnostic tools as they detect core α-(1,6)-linked fucose in *N*-glycans – a modification typically upregulated in different cancer types -[380] but apparently, they cannot bind the monosaccharide epitope in the glycomimetic ligands. The controls supported the fact that the spacer does not influence negatively the binding properties of the ligands. The binding profile of ConA, PSA and LCA probed on the full library are shown below. (**Figure 49**)



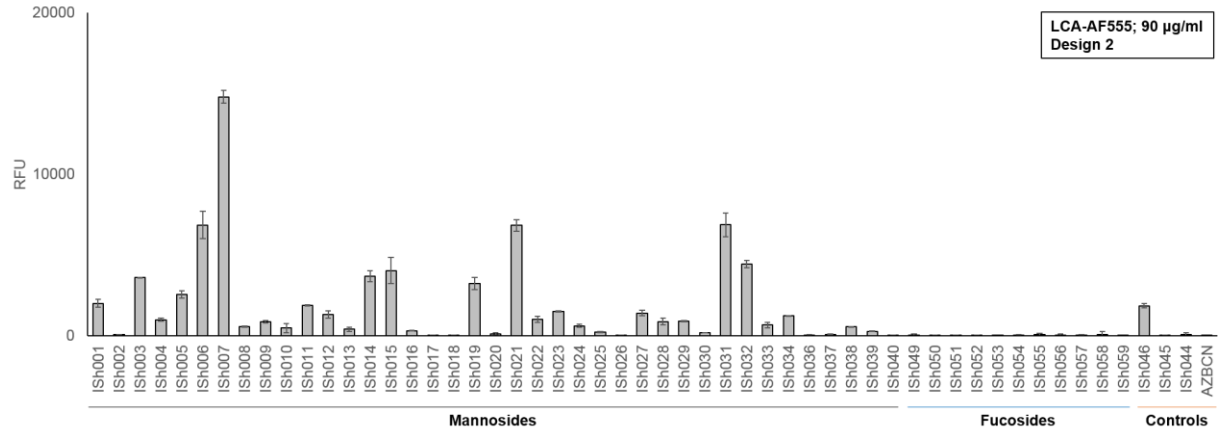



***Figure 49.*** *Screening with 1 µg/ml ConA-AF555, 165 µg/ml PSA-AF647 and 90 µg/ml LCA-AF555 (full library)*

### 2.2.3.3.2 Aleuria aurantia lectin (AAL)

The proper immobilization of the β-fucosides was verified by the fucose-specific fungal *Aleuria aurantia* lectin (AAL).[381] (**Figure 50**) This protein bound exclusively to the fucosylated structures; neither the mannosides, nor the controls showed any fluorescent signals on the chips. The similar intensity values of the β-fucoside ligands suggests, however, that in this case, the affinities are not differentiated by the bisamide substituents on the sugar-mimic ring.

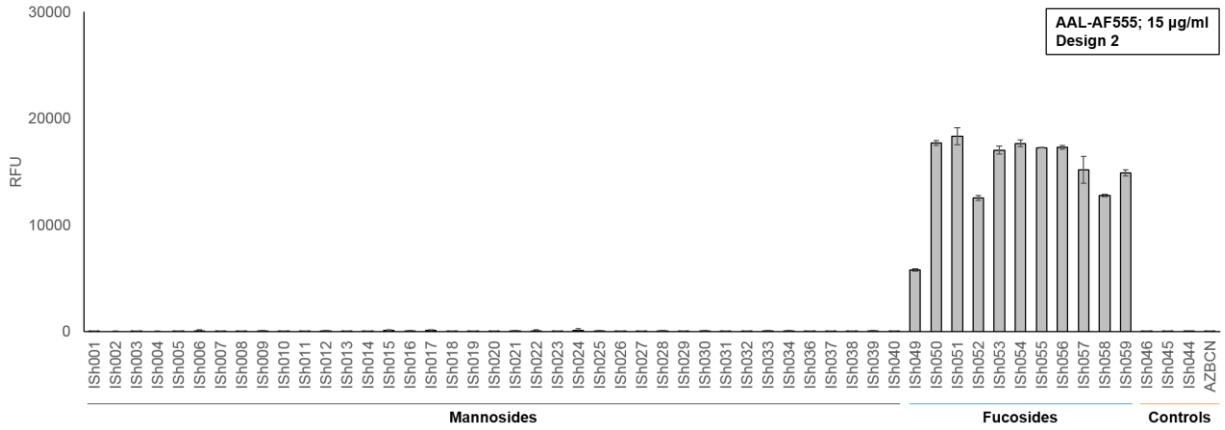



***Figure 50.*** *Screening with 15 µg/ml AAL-AF555 (full library)*

### 2.2.3.3.3 LecB (PA-IIL) from *Pseudomonas aeruginosa*

The array was also interrogated with the *Pseudomonas aeruginosa* lectin B (PA-IIL or LecB). PA-IIL is one of the two lectins produced by the antibiotic-resistant bacterium *P. aeruginosa* and primarily recognizes fucose and α-fucosides, but it shows affinity towards D-Man as well.[178, 382] The pathogen is particularly threatening in hospital environment for patients of reduced immune-defense and cystic fibrosis because the formation of thick biofilm layers shields the bacterial colonies from treatment and host response.[174b, 383] In this adhesive process, PA-IIL plays a crucial role by binding to fucosylated glycoproteins of mucin and the respiratory epithelium and has been a viable target in *Pseudomonas aeruginosa* treatment strategies (by fucose-mimetics, among others, see Introduction).[179, 384]

The screening results with LecB are shown below (**Figure 51**) – the lectin was provided to the Reichardt laboratory by the group of Prof. Anne Imberty (IBS, Grenoble, France).

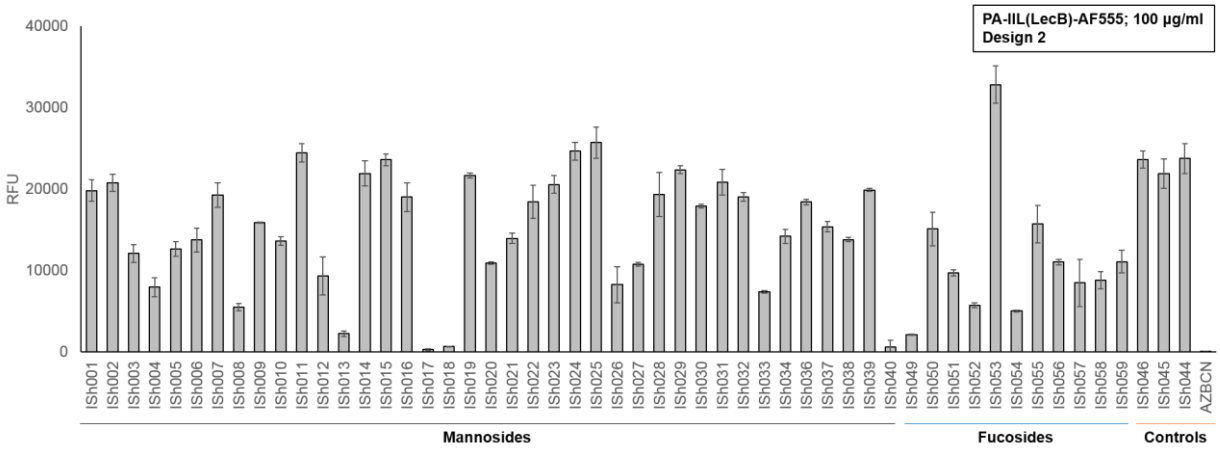



***Figure 51.*** *Screening with 100 µg/ml PA-IIL (LecB)-AF555 (full library)*

The lectin recognizes many of the glycomimetic structures of our microarray - both mannosylated and fucosylated compounds. Almost all the mannosides display some degree of affinity, with a few exceptions (**ISh013**, **ISh017**, **ISh018**, **ISh040**). The β-fucosylated parent methyl ester **ISh049** interacts weakly with PA-IIL, but, in accordance with the mannose-specific lectins, the affinity is tuned by variations in the amide residue. The β-fucoside **ISh053**, carrying two tryptamine moieties, has superior affinity compared to mannose. These data confirm once more that the differential amide substitution is indeed generating diversity that can be read-out by the lectin partner.

### 2.2.3.4 Screening with recombinant human and murine CLRs

Following the optimization of the glycomimetic microarray system and testing its functionality with common plant, fungal and bacterial lectins, we began screening experiments with human CLRs, provided by the laboratory of Dr. Franck Fieschi (IBS, Grenoble, France). The cloning, expression and purification protocols are reported in the Appendix. In some cases, these procedures were established during the course of the IMMUNOSHAPE project and so, CLRs that became available or were reproduced only after the establishment of the microarray setup were probed on the glycomimetic chips (microarray B) – as stated above - by Silvia Achilli and Dr. Sonia Serna.

Human CLRs used during the first microarray assays were previously labelled by Silvia Achilli with Alexa Fluor®-555, while lectins probed later against the whole library were tagged with fluorescent Cy3 (cyanine 3). Single aliquots of the lectins stored at -80 °C were thawed on ice immediately before incubations (the preparation of buffered incubation solutions are detailed in the experimental section, respectively). For incubations, 200 µL of each lectin solution was applied to each subarray using 8 Well ProPlate™ Module incubation chambers. The microarray was incubated under gentle shaking overnight in the dark at 4°C. The slides were washed using TBS containing 4 mM $CaCl_2$ and water, dried with slide spinner and the fluorescence was analyzed with a microarray scanner.

As a control for the fluorescence data, the emission spectra of the different solutions used for incubations were measured. The labeling protocols, as well as a comprehensive table of all lectins screened against the glycomimetic chip with regards to their fluorescent properties and carbohydrate specificities are reported in the Appendix.

#### 2.2.3.4.1 DC-SIGN ECD and validation of the assay with SPR data

As we already had information about the binding properties of several library members towards DC-SIGN, we deemed useful and reasonable to start the CLR screenings with this lectin. The lectin was applied in its tetrameric ECD form on the glycomimetics microarray. (**Figure 52**)

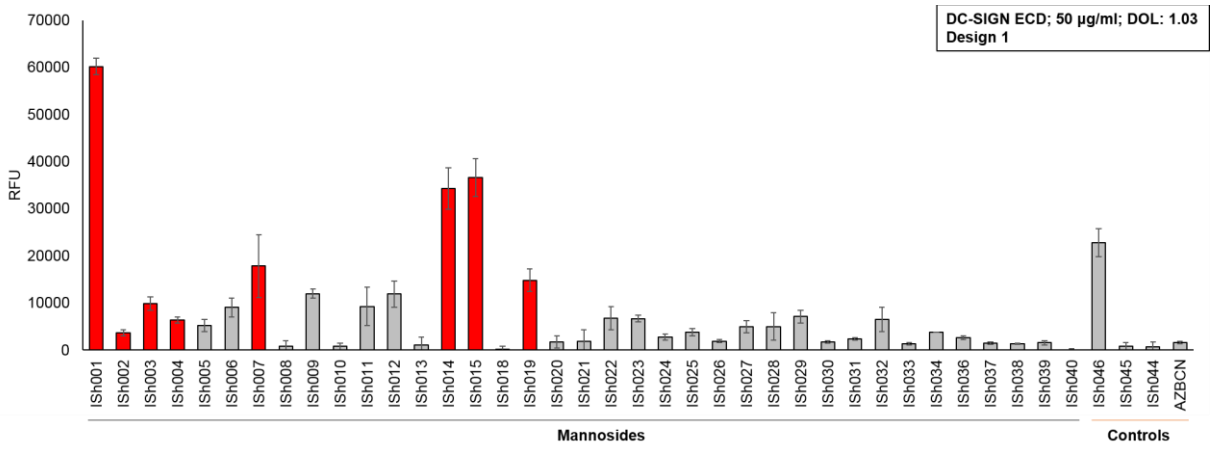



***Figure 52.*** *Screening with 50 µg/ml DC-SIGN ECD-AF555 (not full library; only mannosides and controls).*

The controls are used again to assess the presentation of the ligands with the additional BCN-linker. The lectin does not bind aspecifically to the spacer alone (AZBCN), and recognizes the same glycomimetic ligands immobilized with the extra linker (**ISh001**, **ISh002**) slightly better than the directly linked ones (**ISh046**, **ISh045**). This effect is more pronounced in the case of Man030 (**ISh001** and **ISh046**) because the pseudo-disaccharide dimethyl ester (**ISh002**, **ISh045**) is barely bound on the microarray chips. Binding is not observed towards the natural monosaccharide ligand mannose (**ISh044**), that shows in solution phase SPR competition assays an order of magnitude lower affinity ($IC_{50}$: 10-20 mM) towards the CLR than the dimethyl ester glycomimetic **ISh002** ($IC_{50}$: 0.9 mM).[242b] **ISh001**, a ligand designed based on the crystal structure of **ISh002**, exhibits a three-fold affinity enhancement compared to **ISh002** in SPR tests ($IC_{50}$: 0.3 mM)[250] and indeed, this ligand has the most intense signal on the chip (see Introduction).

DC-SIGN is known to recognize highly mannosylated oligosaccharides abundantly present on the surface of viral and bacterial pathogens; the four Lewis-type blood group antigens ($Le^x$ $Le^y$, $Le^a$, $Le^b$), mannan-capped LAM-structures and mannosylated phosphatidylinositol expressed by mycobacteria. On the glycomimetic chip, the CLR binds to some extent to almost all mannose-based ligands.

Some bisbenzylamide derivatives described by Varga *et al*[250] were designed to target DC-SIGN and had been studied earlier in SPR experiments, along with **ISh001**. Their inclusion in the microarray screenings offered an opportunity to assess the established microarray setup's ability to perform rapid binding activity analysis in order to identify lead structures. These known ligands are highlighted by red bars (**ISh001**, **ISh002**, **ISh003**, **ISh004**, **ISh007**, **ISh014**, **ISh015**, **ISh001**, **ISh018**, **ISh019**) in **Figure 52**, and contain all of the highest fluorescence intensity signals on the chip.

The different heights of these bars indicate that DC-SIGN exhibits variable affinity towards these ligands anchored to the solid support, however the reported $IC_{50}$ values obtained by SPR analysis fall in a narrow range (ca. 300 µM, with the exception of **ISh002**,), similar to the $IC_{50}$ of **ISh001**. (**Figure 53**) In these assays, the ligands exhibited highly similar affinities ($IC_{50}$: 300-400 µM, with the exception of **ISh002**), whereas the binding profile obtained in the first microarray screening showed significant differences in signal intensities. This may be the result of the intrinsically different format of the two assays – on the microarray chips, ligands are presented on a

solid support in a multivalent fashion, while SPR tests measure their ability to inhibit the CLR's binding to a functionalized surface in solution phase.

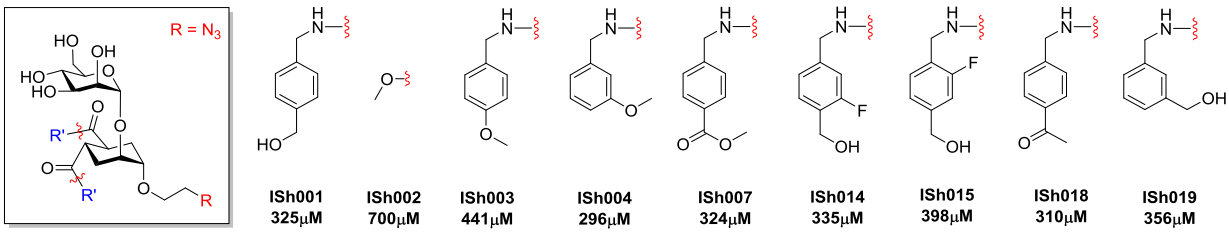



***Figure 53.*** *$IC_{50}$ values of bisbenzylamide-substituted glycomimetic ligands obtained in SPR inhibition assays by Varga et al. on a Man-BSA surface*[250]

The affinity of these ligands showed a closer correlation with SPR results in a later microarray screening with DC-SIGN ECD - this time against the whole glycomimetic library. (**Figure 54**) The slightly different binding profile in the two screenings might be a result of the employed CLR batch. Proteins are sensitive to long-term storing and can denaturate during freeze-thaw cycles, therefore the date of expression and the screening campaign may have effects on the results. Furthermore, the CLR was labelled under different conditions in the first (Alexa Fluor®-555) and in the second case (Cy3), displaying different fluorescent properties (varying DOL). The results are shown in **figure 54**.

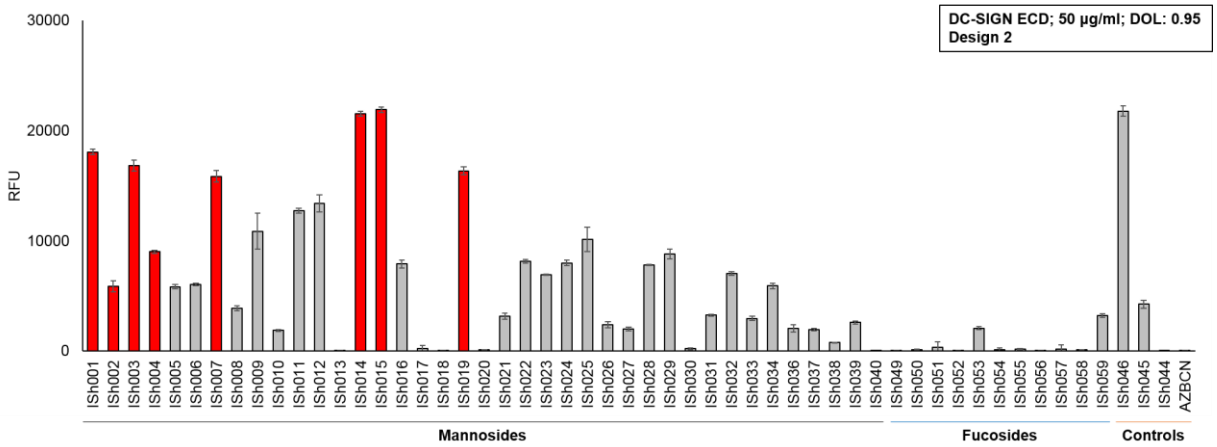



***Figure 54.*** *Screening with 50 µg/ml DC-SIGN ECD-Cy3 (full library)*

In this case, the majority of ligands show similar affinity, except for **ISh004** and **ISh018**. The highest fluorescent signals all belong to already known bisbenzylamides, specifically designed as DC-SIGN antagonists. None of the other bisamides in the library were able to exceed the good affinity of these ligands towards the CLR.
At the same time, previous SPR results with bisamide-substituted pseudo-dimannoside derivatives (carrying an allylic linker) suggested that tertiary amides on the mannosylated glycomimetic scaffold significantly decreased the binding affinity towards DC-SIGN.[249] The tertiary amide ligands on the microarray chips were barely, or not recognized by DC-SIGN (**ISh010**, **ISh020**, **ISh030**, **ISh036**, **ISh037**, **ISh038**, **ISh039**). (**Figure 55**)

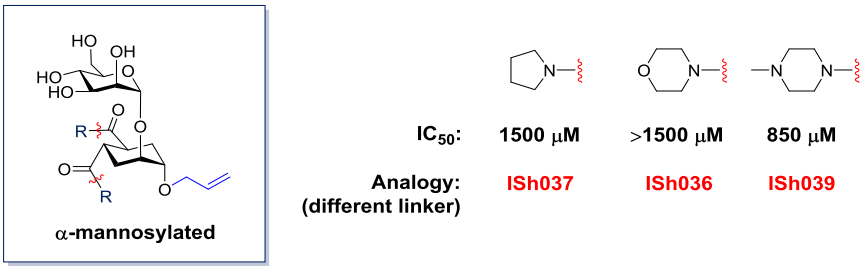



***Figure 55.*** *$IC_{50}$ values of tertiary amide glycomimetic ligands, equipped with an allylic linker, as determined by SPR inhibition assays on a Man-BSA surface. Compared to secondary amides, tertiary amide substituents on the cyclohexane ring significantly lower the affinity of ligands. The depicted structures (tethered with the azidoethanol-BCN-amine linker) were also poorly recognized on the glycomimetic chips, similarly to other tertiary amides.*

The data obtained by the two different assays suggest that tertiary amide-substituents actively hinder the glycomimetics' binding to DC-SIGN and they may offer an opportunity to generate selectivity against this CLR. Because the results of two different assays, microarray screenings and SPR experiments strongly correlated, we drew the conclusion that the microarray technique is reliable and adequate to assess the binding properties of the printed glycomimetic ligands. These results validate the established screening setup as a practical tool which is able to distinguish between better and poorer ligands.

As it has been already mentioned briefly, the constitutionally different microarray experiments and SPR assays are not expected to yield identical results, but to collectively provide an insight into carbohydrate-lectin interactions. Ideally, for our purpose - to select monovalent, good affinity and possibly selective ligands for further elaboration – a similar ranking of the binding affinity obtained by the various techniques would be satisfactory. We have to stress, however, that while glycan microarrays analyze the multimeric CLR's affinity towards a multivalent platform, where the carbohydrate structures are tethered to a solid surface, the SPR-based competition assays described by Varga *et al.* were used to measure the potency of a solution phase monovalent ligand to inhibit binding of a solution phase CLR to a functionalized surface. These experiments can indicate selectivity and kinetic properties of the interactions, do not require labeling - a time- and cost-demanding step that can interfere with the binding sites and binding events – and are sensitive to detect small differences in ligand potency. A brief description of SPR analysis can be found in **2.2.6.3**.

### 2.2.3.4.2 DC-SIGNR ECD

The DC-SIGN related molecule DC-SIGNR, otherwise known as L-SIGN, is expressed together with its closely related homologue, but on a different subset of cells. In humans, DC-SIGNR is found on sinusoidal endothelial cells in the lymph node, the liver, and lungs; on capillary endothelial cells in the placenta and in the endothelium of the gastrointestinal tract. [202, 385]

The two lectins show remarkable similarity, 73% identity on the nucleic acid level and 77% identity in their amino acid sequences which is likely a result of a duplication of a single common ancestral gene.[202, 386] Not surprisingly, similarly to DC-SIGN, DC-SIGNR also functions primarily as an adhesion molecule and receptor for pathogens like HIV-1, [190, 387] Ebola,[198] hepatitis C virus (HCV),[200a, 388] SARS-CoV,[385b, 385c] *M. tuberculosis*,[389] *M. bovis*,[390] influenza virus A,[391] furthermore mediates lymphocyte transport.[181, 385a], On liver cells, DC-SIGNR captures HCV from the blood flow and facilitates its internalization into hepatic cells where the virus escapes neutralization and degradation at low pH.[392]

Although similar to DC-SIGN in its tetrameric organization and CRD structure (and thus carbohydrate specificity), the two CLRs exhibit differences in the coiled-coil neck domains – a significant factor affecting the spatial arrangement of the four CRDs.[190, 386, 393] While DC-SIGN exists in its single form and contains seven

complete and one incomplete copies of a 23-aminoacid sequence, the number of repeats of the same sequence varies between four to nine in DC-SIGNR.[365, 394] This can be attributed to a genetic polymorphism observed in DC-SIGNR and affects the stability of oligomerization.[395] As a consequence, the CRDs of DC-SIGN are arranged in a more closed structure, with their long axis alligned roughly parallel to the axis of the neck, whereas DC-SIGNR presents its CRDs in a more open conformation. (**Figure 56**) This leads to different avidity effects in the two CLRs (DC-SIGN's interaction with ligands often benefit from the more suitably oriented binding sites).[96, 189] Eventually, these structural differences can influence the ligand-binding affinity or efficiency and cause differing susceptibility of the two CLRs, for example, towards HIV-1 and West-Nile virus among other pathogens.[396]

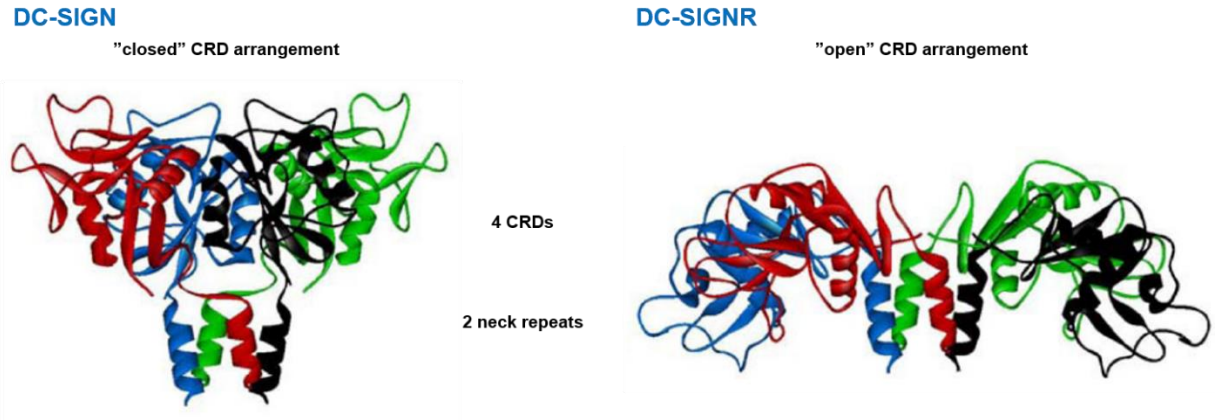



***Figure 56.*** *Oligomerization of DC-SIGN and DC-SIGNR CRDs. In both proteins, the tetrameric structure is achieved by oligomerization of the neck domains, but the different structure of the coiled-coil region leads to a more "open" spatial arrangement of the four CRDs in the case of DC-SIGNR. Modified from: Tabarani et al.*[189]

Other significant deviation from DC-SIGN's carbohydrate-specificity can be observed in DC-SIGNR's recognition of Lewis-type epitopes and some high mannose oligosaccharides, as it was revealed in microarray screenings (see Introduction). A crucial factor in the interactions between Lewis$^X$ and DC-SIGN is the presence of a valine (Val351), a residue which is missing from DC-SIGN. Val351 in DC-SIGN is a central element not only for the recognition of natural fucosylated ligands, but plays a key role in the binding of the pseudo disaccharide dimethyl ester **ISh002** as well.[244]

Different DC-SIGNR constructs were expressed and provided by the Fieschi group. During the initial screenings, the tetrameric ECD was probed on the glycomimetic chips at 50 and 100 μg/ml concentrations – the results are shown below. (**Figure 57**)

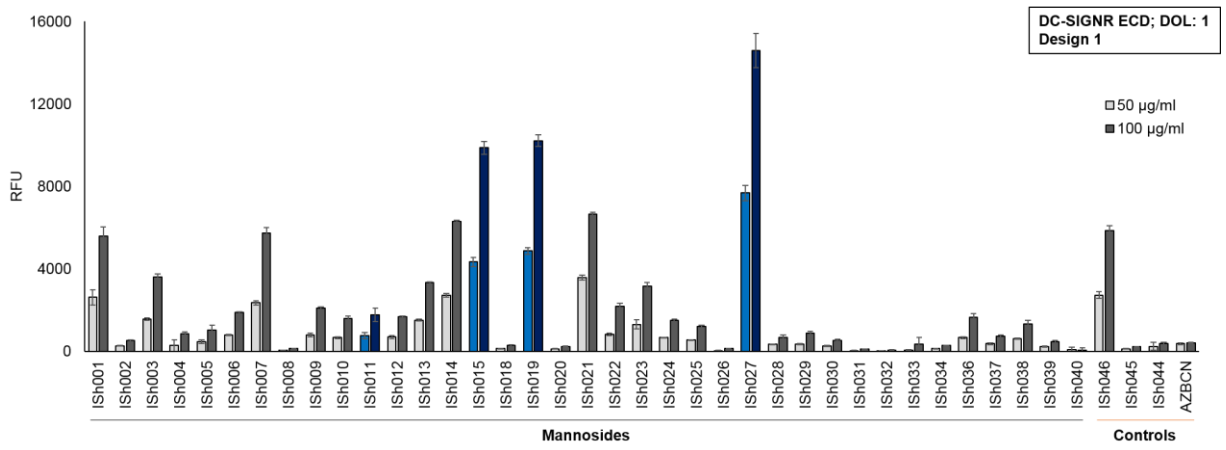



***Figure 57.*** *Screening with 50 and 100 µg/ml DC-SIGNR ECD-AF555 (not full library; only mannosides and controls)*

Starting our evaluation with the controls, it has been confirmed once more, that the spacer does not show any aspecific binding of the CLR (**AZBCN**) or interfere with the recognition of individual ligands (**ISh001** and **ISh046**; **ISh002** and **ISh045**). Again, the majority of glycomimetics show higher fluorescent signals than the natural ligand mannose. Interestingly, the pseudo-disaccharide dimethyl ester **ISh002** exhibits similar affinity towards the lectin as the natural monosaccharide.

Remarkably different ranking is shown from the binding profile of DC-SIGN ECD, however, as the fluorescent intensity of **ISh027** is superior to that of every other ligand, at both concentrations. On the other hand, **ISh015** and **ISh019** are – again - among the top binding ligands.

Based on these results, 4 ligands, **ISh011** - a medium scoring ligand and **ISh015**, **ISh019** and **ISh027** were selected for SPR testing, and were sent to the laboratory of Prof. Franck Fieschi (IBS, Grenoble, France). The SPR analyses were performed by Silvia Achilli and Dr. Corinne Deniaud, on a setup identical to that of the previosuly studied DC-SIGN ligands, against a Man-BSA functionalized chip (protocol in Appendix). (**Figure 58**Figure 58)

In the SPR inhibition assays, the glycomimetic recognition elements were tested (azidoethanol linker, no BCN-motif) against both DC-SIGNR ECD and DC-SIGN ECD, in order to simultaneously evaluate and compare affinity towards both CLRs. The results are shown below as $IC_{50}$ values.

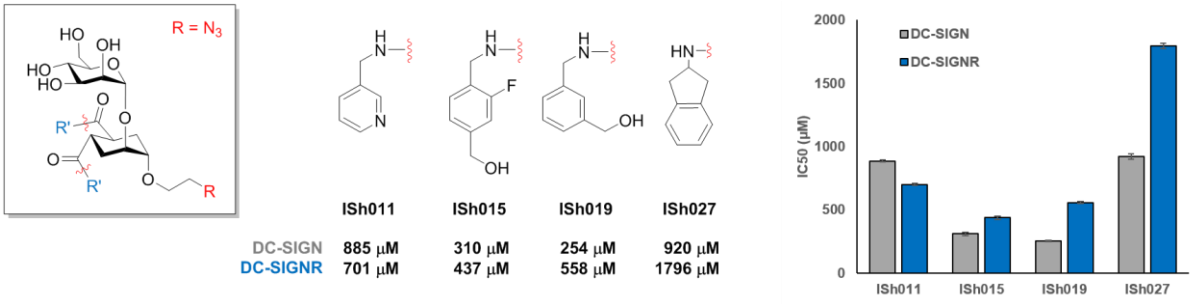



***Figure 58.*** *DC-SIGN and DC-SIGNR inhibitory potency of selected glycomimetic ligands. $IC_{50}$ values were determined on a Man-BSA coated surface in SPR inhibition assays. Ligand ranking for DC-SIGNR obtained from the first microarray screenings and SPR experiments did not correlate, as* ***ISh027*** *was identified as a poor ligand in the solution phase SPR tests, although it showed the strongest signals on the microarray chips. $IC_{50}$ values and fluorescent signals were more in accordance with each other in the case of DC-SIGN.*

In this case, apparent disagreements can be observed between the affinities on the microarray chips and the $IC_{50}$ values obtained in solution phase by SPR. As for DC-SIGNR, **ISh015** and **ISh019** show similar affinities in both

assays ($IC_{50}$: 437 and 558 μM respectively), while **ISh011** exhibits more modest results, but its relative affinity, compared to the two previous ligands, is in accordance on the microarrays and in SPR tests ($IS_{50}$: 701 μM).
**ISh027**, however, reached a poor $IC_{50}$ value (ca. 1.8 mM) in the solution phase inhibition assays, in spite of its high intensity signals when the microarrays were probed with DC-SIGNR ECD at two concentrations. This may be the result of several possible factors: either the fundamentally different physical characteristics of the assays affect severely the ligand's recognition (see **2.2.1** and **2.2.2**), or the ligand binds to the CLR aspecifically, and the lower intensity binding profile obtained with the ECD apparently amplifies this phenomenon.
Regarding DC-SIGN's affinity towards the selected 4 ligands, we see good correlation between the first microarray (**Figure 52**) and SPR scores. During the second screening (**Figure 54**), however, **ISh011** appeared to be closer in affinity to **ISh015** and **ISh019** than to **ISh027**, whereas the results obtained by SPR assays reveal that **ISh011** and **ISh027** have $IC_{50}$ values in the same range (ca. 900 μM), well below the other two glycomimetics. (Sensorgrams in appendix) $IC_{50}$ values of glycomimetic ligands were determined for the first time against DC-SIGNR.
Once the entire glycomimetic library became available, further screenings were performed on the microarray chips with another batch of the ECD. These data are shown below. (**Figure 59**)

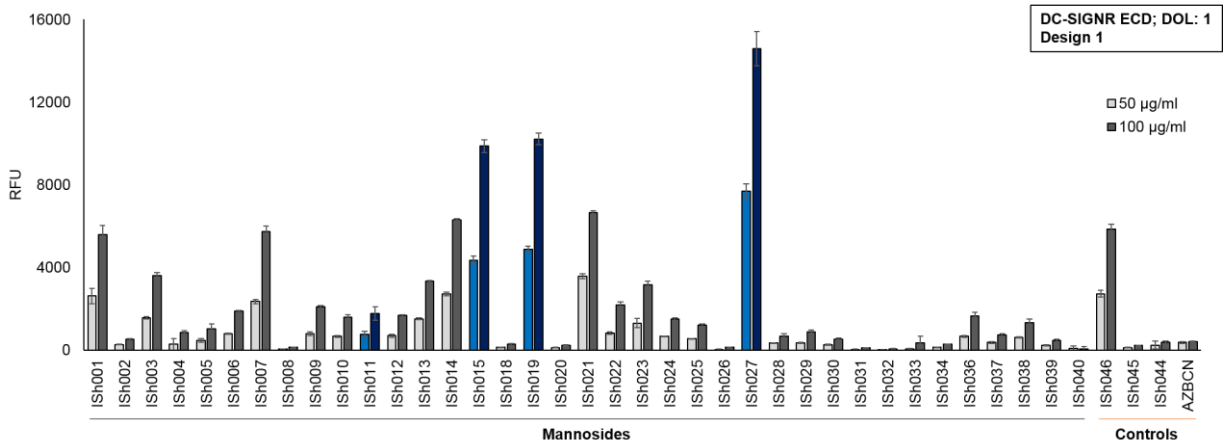



***Figure 59.*** *Screening with 150 μg/ml DC-SIGNR ECD-Cy3 (full library)*

In this screening, the ranking of the four selected ligands show higher correlation with the previous SPR results and highlights **ISh007** as one of the highest affinity ligands. The SPR inhibition test has not been performed with this ligand yet.
It is intriguing to consider whether the overall weak signals of DC-SIGNR are results of the less organized oligomerization of the CRDs. Too flexible arragement (compared to DC-SIGN for example, see discussion above), could – in theory - lead to impaired affinity towards the solid-surface immobilized ligands, instead of enhancing avidity effects through oligomerization as it would be expected.
The examples of DC-SIGN and DC-SIGNR clearly demonstrate that to evaluate binding interactions between CLRs and their carbohydrate ligands, ideally multiple complementary analytical platforms need to be used, so that the results can be integrated, and used to determine the ligands' activity and selectivity. From this complex information, it is, in principle, possible to identify the key structural elements that define the best lead compounds.

#### 2.2.3.4.3 Langerin

Langerin is a endocytic type II integral transmembrane receptor of 37.5 kDa expressed abundantly on the surface of Langerhans cells – specialized dendritic cells in the dermis and mucosal tissue - in the epidermis, but the CLR

is also present, although at lower levels, on intestinal lamina propria in the human colon and on CD1c$^+$ dendritic cells of the myeloid lineage.[101, 397]

Langerin recognizes a broad range of viral, bacterial and fungal pathogens. In addition to HIV-1 (see Introduction),[224, 398] the CLR binds measles virus[399] and *Mycobacterium leprae*.[400] Langerin strongly interacts with many *Candida* species, such as *C. albicans, C. dubliensis, C. glabrata, C. guilliermondii, C. krusei, C. parapsilosis, C. tropicalis, C. lusitaniae, C. nivariensis, C. orthopsilosis* and *C. metapsilosis*; furthermore with *Malessezia furfur* and with all *Saccharomyces* species, including *Saccharomyces cerevisiae*.[401]

Langerin is organized into trimers through the formation of α-helical coiled-coils by the oligomerization of the extracellular neck domain stem regions.[97] (**Figure 60.A**) *This* structure is more rigidly embedded in the membrane and the orientation of the CRDs is more fixed than on the flexible DC-SIGN tetramers, for example. Furthermore, langerin is equipped with a short intracellular tail.

The CRD of the lectin features the conserved EPN motif in the primary $Ca^{2+}$ binding site which is an indication for mannose and fucose specificity, and in accordance with this, langerin binds to high mannose oligosaccharides.[402] Furthermore, the CLR recognizes β-glucans,[401] GlcNAc[403] and - as it has been detailed in the Introduction - 6-sulfated galactose via a noncanonical binding mode.[80, 404] Additionally, this glycan preference can be shifted by commonly observed polymorphism in the CRD.[405] Other endogeneous antigens include heparin (see Introduction),[89b] hyaluronic acid[406] and blood group antigens.[80]

In contrast with the mainly overlapping carbohydrate specificity, the recognition of fucosylated Lewis-type antigens significantly differs in langerin and DC-SIGN. While the latter lectin exhibits good affinity towards $Le^X$, $Le^Y$, $Le^A$ and $Le^B$, langerin's recognition of these structures is limited to the difucosylated $Le^Y$ and $Le^B$ antigens.[80]

In spite of the high number of shared glycan ligands between langerin and DC-SIGN, the distinct location of their expression, the different structural organization and the fact that langerin is more rigidly projected to the extracellular media, suggest fundamentally different biological activating role of langerin and DC-SIGN.

The primary langerin expressing cells, the Langerhans cells contain a unique set of organelles, called Birbeck granules whose formation and activation is associated with langerin.[94, 101, 407] (**Figure 60.B**) It was demonstrated on several occasions that the CLR and its correct oligomerization is indispensable for the proper architecture and functioning of Birbeck granules (see Introduction).[100] These organelles are important mediators of HIV eradication, since after langerin captures the virus by interacting with the gp120 glycoprotein, the viral particles are internalized into the Birbeck granules where they are submitted to degradation under not fully understood conditions.[224]

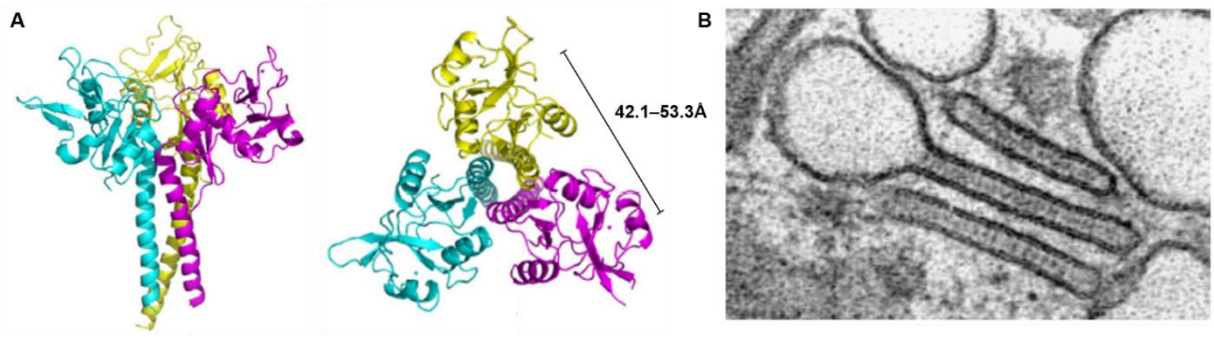



***Figure 60.*** *A. Trimeric oligomerization of langerin CRDs. The three CRDs are arranged in a triangular shape, extended towards the exterior of Langerhans-cells. Modified from Feinberg et al.*[97] *B. Birbeck granules. Tennis-racket and rod-like shapes are visible in the transmission electronmicroscopy (TEM) images, taken of dermal Langerhans-cells. These granules are crucial in HIV-virus eradicartion. The formation and function of Birbeck granules is highly dependent on the expression and appropriate oligomerization of langerin. Modified from Romani et al.*[408]

The immunological role of langerin and DC-SIGN in HIV infections is probably the most known challenge of selective CLR targeting up to date. An efficient anti-adhesive antiviral strategy based on these lectins must be able to specifically inhibit viral capture by DC-SIGN, but, at the same, ensure that langerin is fully functional to help the destruction of HIV in Langerhans cells.

As it has been described in the Introduction, the selectivity of the pseudo disaccharide dimethyl ester **ISh002** is superior to the natural disaccharide Manα1-2Man which is also a good affinity ligand of langerin.[250] The low affinity of the pseudo disaccharide ligands might be attributed to the lack of the valine methyl groups and the absence of C6 substituents on the cyclohexane ring (see Introduction). Another lysine residue (Lys313) has been targeted to generate rationally designed selectivity between the CLRs.[251]

The first microarray screening with langerin was performed with an older batch of the protein that went through several freeze-thaw cycles and showed traces of precipitation. The screening was later repeated on the whole microarray with a new batch, labeled with Cy3, the results are shown in **figure 61**.

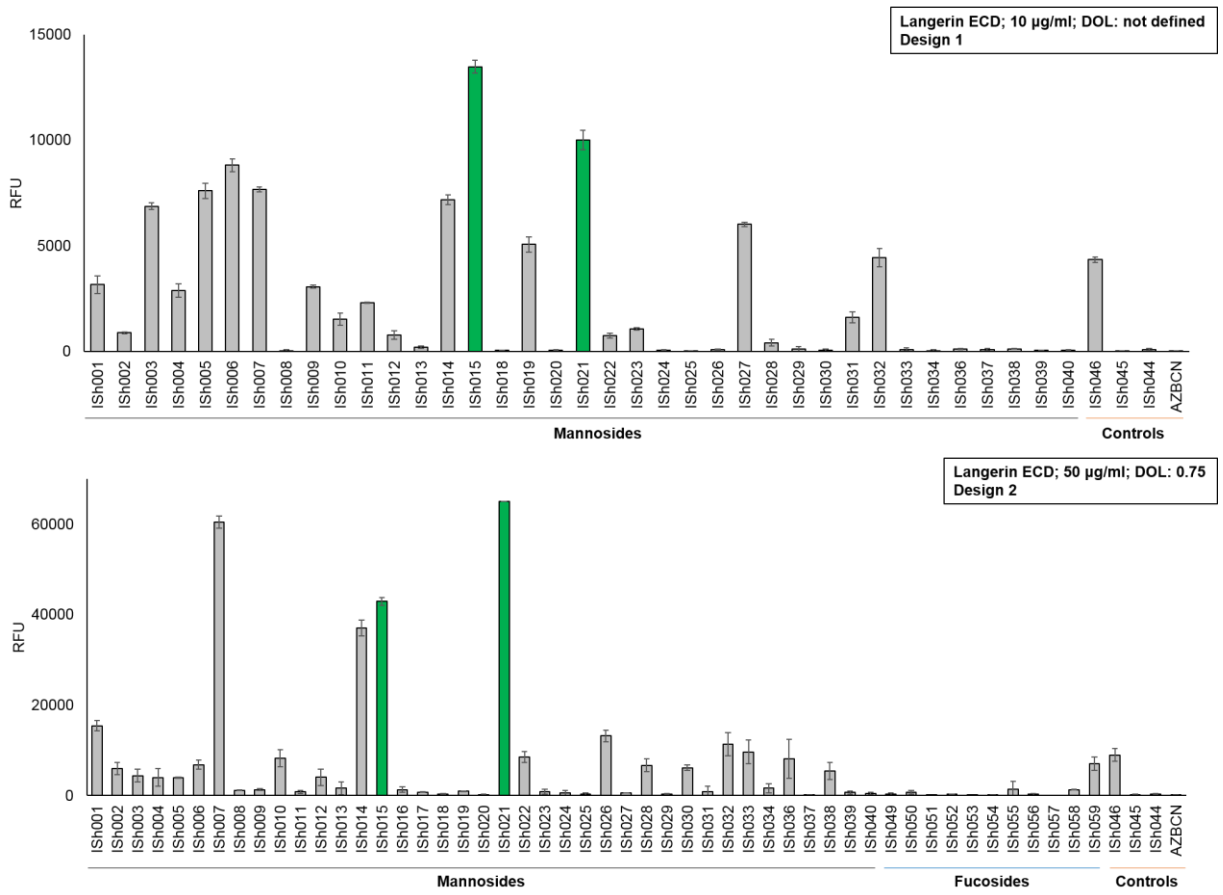



***Figure 61.*** *Screening with 10 µg/ml langerin-AF555 (not full library; only mannosides and controls) and 50 µg/ml langerin-Cy3 (full library)*

The controls provide another evidence that the additionally linker moiety doesn't impair recognition of the glycomimetics ligands (**ISh046** and **ISh045** as opposed to **ISh001** and **ISh002**, respectively), and that the spacer alone (**AZBCN**) does not form aspecific interactions with the CLR.

We already had in our hands solution phase single point SPR results about numerous bisbenzylamide ligands of the library, described by Varga *et al.*[250] These experiments measured inhibitions at fixed concentration levels on a mannosylated BSA surface, and while the selectivity between DC-SIGN and langerin was modest for the dimethyl ester **ISh002**, the other ligands exhibited good discriminative binding towards DC-SIGN over langerin. From the collective binding profile, obtained in two campaigns, 4 ligands (**ISh007**, **ISh014**, **ISh015**, **ISh019**) could be identified as well-recognized structures on the microarray chips. After the first screening on the mannoside library, we chose two ligands, **ISh015** and **ISh021** for further testing in similar SPR inhibition experiments (without the additional BCN amine spacer). (**Figure 62**)

Neither **ISh015**, nor **ISh021** could inhibit langerin binding to the mannosylated BSA surface, even at high millimolar concentrations. When the high affinity natural competitor Man-BSA was replaced with the weaker $Le^A$-BSA functionalized surface, the full inhibition curve of **ISh015** could be recorded and the $IC_{50}$ value determined at 1.8 mM. This serves as a confirmation that the few high intensity glycomimetic compounds indicated by the microarray screenings are in fact, very poor ligands of the lectin. The reason behind their strong signals on-chip remains unknown, but as langerin exhibits very low tolerance to the bisamide substituted glycomimetic ligands, these findings may be false positive results.

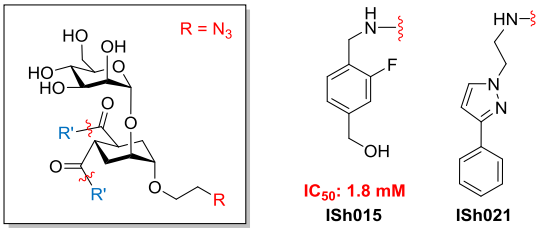



***Figure 62.*** *Selected glycomimetic ligands, tested in SPR experiments against langerin. None of the compounds could inhibit the binding of langerin to Man-BSA surface. When the gold chip was functionalized with the weaker natural ligand Le$^A$-BSA, the IC$_{50}$ of ISh015 was determined as 1.8 mM. This may indicate that the ligands showing strong signals on the microarray chips are only false positive results.*

#### 2.2.3.4.4 MBL

The mannose binding lectin (MBL) is a soluble C-type collectin that is secreted into the serum primarily by hepatocytes, to reach an average concentration of 1 μg/ml.[409] As a pattern recognition receptor, it triggers the lectin pathway of the complement system during the innate immune response.[410] As MBL is present in the circulatory system, the stimulation of the complement system is one of the very first mechanisms triggered when the lectin encounters infectious pathogens (PAMPs) or altered self-structures (DAMPs). Upon these interactions, the lectin initiates the autoactivation of associated proteases.[411] When MBL recognizes apoptotic and necrotic cells, it activates macrophage-mediated phagocytosis to destroy the dying cells without stimulating inflammation.[412] Furthermore, MBL's disadvantageous involvement at the site of ischaemic stroke was demonstrated.[413]

MBL consists of an N-terminal collagen-like domain, a hydrophobic neck region and the C-type lectin like domain (C-terminus).[414] The lectin oligomerizes into homotrimers that are further organized into higher oligomers, in humans typically from dimers to hexamers (2/6 * 3 CRDs).[415]

Similarly to other EPN motif containing CLRs, MBL's carbohydrate specificity includes Man, Fuc, Glc and GlcNAc. These structures are often present on PAMP and DAMP fingerprints, but the signalling pathway initiated by their recognition differs. In addition to triggering the complement system, MBL can promote apoptosis, modulate inflammation or activate opsonophagocytosis.[416]

The lectin recognizes different species of viruses, fungi and bacteria, for example, herpes simplex virus-2, influenza A virus, *Saccharomyces cerevisiae, Aspergillus fumigatus, Candida albicans* and *Staphylococcus aureus, Mycobacterium avium.*[417] Most often, these recognition events occur through MBL's interaction with high glycans, like lipoarabinomannan-type structures, peptidoglycan and lipoteichoic acid.

We purchased recombinant human MBL (rhMBL) from a commercial source, and labelled the lectin with Alexa Fluor-555. The result at 12.5 μg/ml is shown in **figure 63**. (Labelling and screening performed by Dr. Sonia Serna)

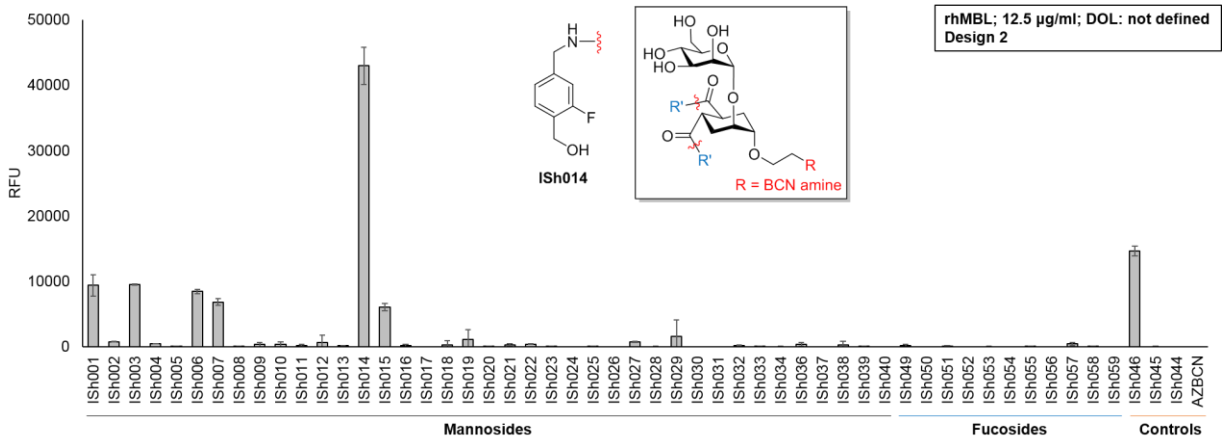



***Figure 63.*** *Screening with 12.5 µg/ml hMBL-AF555 (full library)*

The only ligand displaying significant fluorescence signal is **ISh014**, while the structurally similar **ISh015** and a few other glycomimetics exhibit only weak binding. This may be another case of a false positive signal (see langerin), the result has not been assessed yet.
The controls confirm again that the additional spacer does not influence negatively the screening results.

### 2.2.3.4.5 Murine Fc-fusion CLRs

Murine homologues of C-type lectin receptors are often subject of investigations, even though the divergence between CLRs of different mammalian species can be significant, and can restrain the transferability of results (see Introduction).
The Lepenies laboratory (University of Veterinary Medicine, "TiHo", Hannover, Germany) has recently described a comprehensive murine CLR-Fc fusion protein library that covers the immunologically most relevant dectin-1 and DCIR family members of C-type lectin receptors.[293]
These fusion proteins consist of the respective CLR's extracellular part which is fused in two copies to the Fc fragment of human IgG1. The obtained fusion proteins display the CRD in a dimeric form on the Fc region of the sufficiently glycosylated IgG molecule. This fragment can be recognized with a fluorescently tagged secondary antibody which enables the detection of binding events on microarray chips. (**Figure 64**)

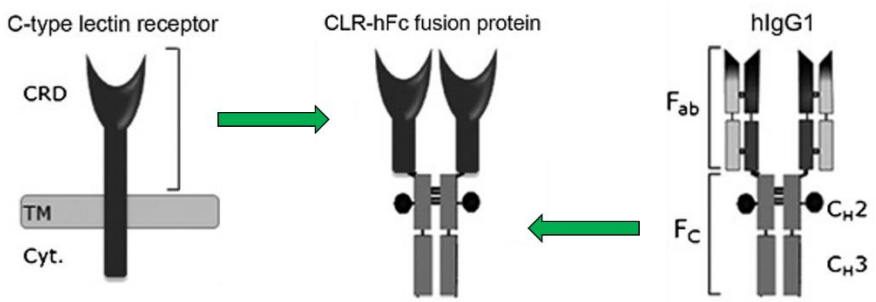



***Figure 64.*** *Murine Fc-fusion proteins. CLR ECDs are fused with human IgG1 Fc-region, to obtain a dimeric form of the lectins. The human Fc-portion can later be detected on the microarray chips during a following incubation with a fluorescently-tagged secondary antibody. Modified from Maglinao et al.*[293]

Murine CLR-Fc fusion proteins were provided by the Lepenies group expressed and purified by João Monteiro. The available CLRs and their carbohydrate specificities are listed in **table 2.**

| Murine CLR | Name | Carbohydrate specificity |
|---|---|---|
| DCAR | Dendritic cell immunoactivating receptor | Carbohydrate specificity not defined |
| Dectin-1 | CLEC7A, Dendritic cell-associated C-type lectin-1 | β-1,3 and β-1,6-linked glucans |
| DCL-1 | DEC-205-associated C-type lectin-1 | The receptor does not contain the amino acids that coordinate $Ca^{2+}$-dependent sugar binding in other C-type lectins, suggesting that it does not have the classical sugar binding motif. No endogenous or exogenous ligands have been identified yet |
| CLEC9A | | Carbohydrate specificity not defined |
| MGL-1 | Macrophage galactose lectin | $Lewis^X$ and $Lewis^A$, terminal Gal and GalNAc epitopes |
| SIGNR-3 | ICAM-3-grabbing nonintegrin-3 | High mannose and fucose structures |
| SIGNR-5 | ICAM-3-grabbing nonintegrin-5 | High mannose and fucose structures |
| MICL | myeloid inhibitory C-type *lectin-l*ike receptor | Carbohydrate specificity not defined |

***Table 2.*** *Murine CLR-Fc fusion proteins and their carbohydrate specificities*

All of these CLRs had been incubated at 50 μg/ml concentration overnight at 4°C. The slides were washed and incubated for 1 hour with anti-human IgG (Fc-specific)-Cy3 antibody produced in goat as secondary antibody. As negative controls, one subarray was first incubated with human IgG Fc and later with the secondary antibody, and another "empty" well was incubated with goat Anti-human IgG-Cy3.

None of the incubations yielded quantifiable binding profiles, as all of them showed essentially only two fluorescent signals: **ISh021** and **ISh027** appeared on all of the microarray slides probed with murine CLR-Fc fusion proteins, but not on the two negative controls.

These two ligands, both of them having had apparent interactions with human CLRs in earlier screenings (**ISh021** – langerin; **ISh027** – DC-SIGNR) were eventually deemed aspecific binders and their results evaluated critically from this point on. A typical scanned image of a murine CLR-Fc fusion protein (in this case DCL-1) probed slide is shown in **figure 65**. Fluorescent spots belong to the 5-5 replicates of **ISh021** and **ISh027**. Top left corner is marked with fluorescent BSA - printed on each subarray of the slide to indicate the starting position of spotting.

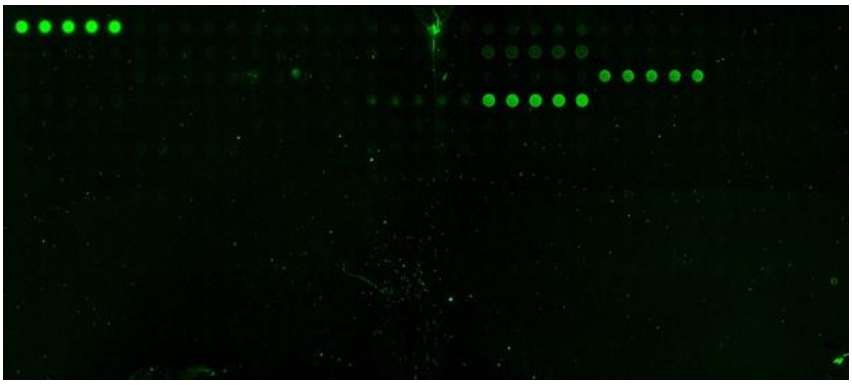


***Figure 65.*** *Scanned image of microarrays incubated with different murine Fc-fusion CLRs. Top left corner: fluorescent BSA, signalling the start of the subarray. The two set of replicate fluorescent spots belong to* ***ISh021*** *and* ***ISh027*** *– two ligands deemed to bind aspecifically to all of the murine CLR-Fc lectins and to some of the human CLRs as well. No evaluable screening results were obtained for any of the murine Fc-fusion proteins.*

#### 2.2.3.4.6 Dectin-2

The dendritic cell–associated C-type lectin-2, dectin-2 is a transmembrane receptor expressed primarily on tissue macrophages, peripheral blood monocytes, several subsets of dendritic cells, Langerhans cells, Kupffer cells, lymphocytes and neutriphils as well.[418] The lectin belongs to the CLR family of BDCA-2, DCAR, DCIR, CLECSF8 and Mincle.[419] Dectin-2 and Mincle (dectin-3) have been found to organize into heterooligomers (see Introduction).[99]

The CLR exhibits high specificity for mycobacteria (e.g. capping mannose in lipoarabinomannan and lipopolysaccharides) and fungi (terminal mannose residues *O*-linked to glycoproteins, or capsular polysaccharides)[420] and pathogens recognized by the lectin include *M. tuberculosis, M. bovis, Klebsiella pneumoniae, Hafnia alvei, Saccharomyces cerevisiae, Candida albicans, Malassezia, Paracoccoides brasiliensis, Histoplasma capsulatum, Cryptococcus neoformans,* schistosome egg antigen, as well as dust mites (*Dermatophagoides pteronyssinus, Dermatophagoides farinae*).[420a, 421]

Evidence of the CLR's protective anti-fungal activity was obtained in animal models, furthermore, dectin-2 shows anti-metastatic activity by mediating uptake and clearance of cancer cells in the liver (Kupffer-cell phagocytosis by TNF-α, ROS (reactive oxigen species) and stimulated NK-cell activity).

Dectin-2 consists of a C-terminal CRD, a transmembrane region and a short cytoplasmic domain that lacks typical signalling motifs, but can associate with the common Fc receptor γ subunit FcRγ which induces cytokine production.[422] (**Figure 66.A**) The immunotyrosine activation motif in FcRγ interacts with Syk-kinase and indirectly stimulates cytokine secretion (e.g. TNF, IL-2, IL-10, IL-23, IL-1β, IL-6 and IL-12 production) eventually initiating T-cell - for example Th17 or Th2 - immune responses. In particular, recognition of dust mites trigger a dectin-2 mediated allergic response.[109, 137b, 423]

As an EPN-motif displaying CLR, dectin-2 shows affinity – although very weak – towards mannose.[302] Feinberg *et al.* measured inhibition constants ($K_i$) in a solid-phase competition assay – which binding indicators are similar to $IC_{50}$ values is SPR inhibition tests.[302] In these experiments, $^{125}$I-Man-BSA was used as a reporter ligand - well below saturation concentrations - whose binding to a dectin-2 coated surface was inhibited by added ligand solutions. Among the monosaccharide ligands, Man and GlcNAc were found to exhibit similar affinity towards the lectin ($K_i$(αMe-Man): 17.1 mM), while Glc and Fuc are recognized much weaker in the assays. Gal and GalNAc displayed very poor affinity in the assays. (**Figure 66.A**)

A number of disaccharide compounds were also probed and achieved significantly lower $K_i$ values. The highest affinity, 2.6 mM inhibition constant was reached by Manα1-2Man, indicating the presence of a secondary binding site to accommodate both mannose residues. This assumption was further supported by a microarray-based screening of more than 600 glycans ligands which revealed a high specificity towards Manα1-2Man containing oligosaccharides, such as $Man_9$. (**Figure 66.B**)

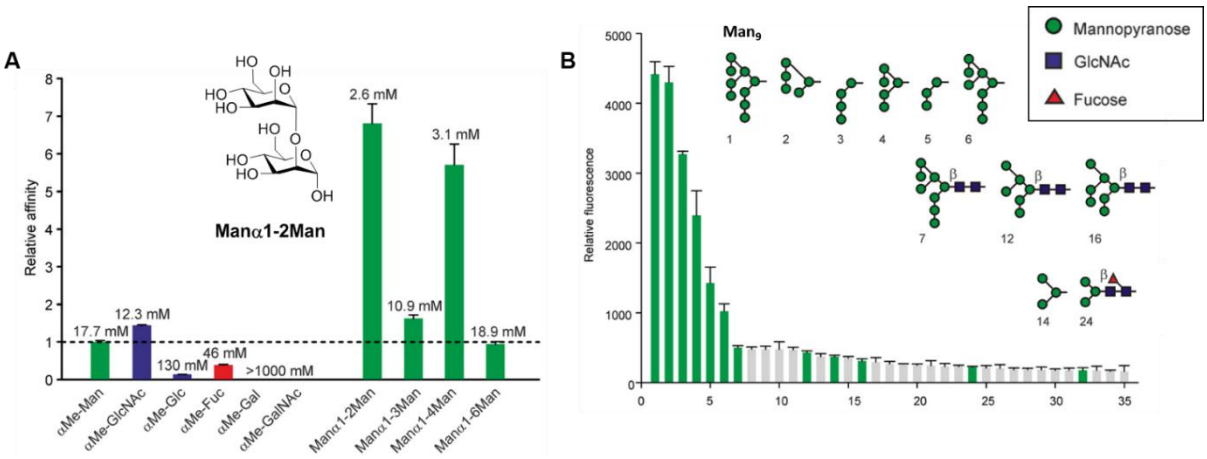



***Figure 66.*** *Natural ligands of dectin-2. A. Inhibition constants of natural ligands, as determined in solid-phase competition assay with $^{125}$I-Man-BSA tracer. Manα1-2Man, the natural disaccharide that our glycomimetic ligands are based upon, exhibited the highest affinity towards dectin-2. Its $IC_{50}$ was estimated at 2.8 mM, as opposed to 17.7 mM of the mannose monosaccharide. B. Microarray screenings with dectin-2 revealed a high preference for Manα1-2Man motif in mannosylated glycans. As in the case of DC-SIGN, $Man_9$ is also the best natural ligand of this CLR. Modified from Feinberg et al.*[424]

This high preference for Manα1-2Man - the structure which our glycomimetic compounds were designed upon - suggested that the lectin may be able to recognize library members on the microarray chips. The ECD of dectin-2 was expressed, purified and labelled with Cy3 by Silvia Achilli in the laboratory of Prof. Franck Fieschi (IBS, Grenoble, France), although the exact oligomerization state of the CLR is unknown. The binding profile at two concentrations – 25 and 50 μg/ml - is shown in **figure 67**.

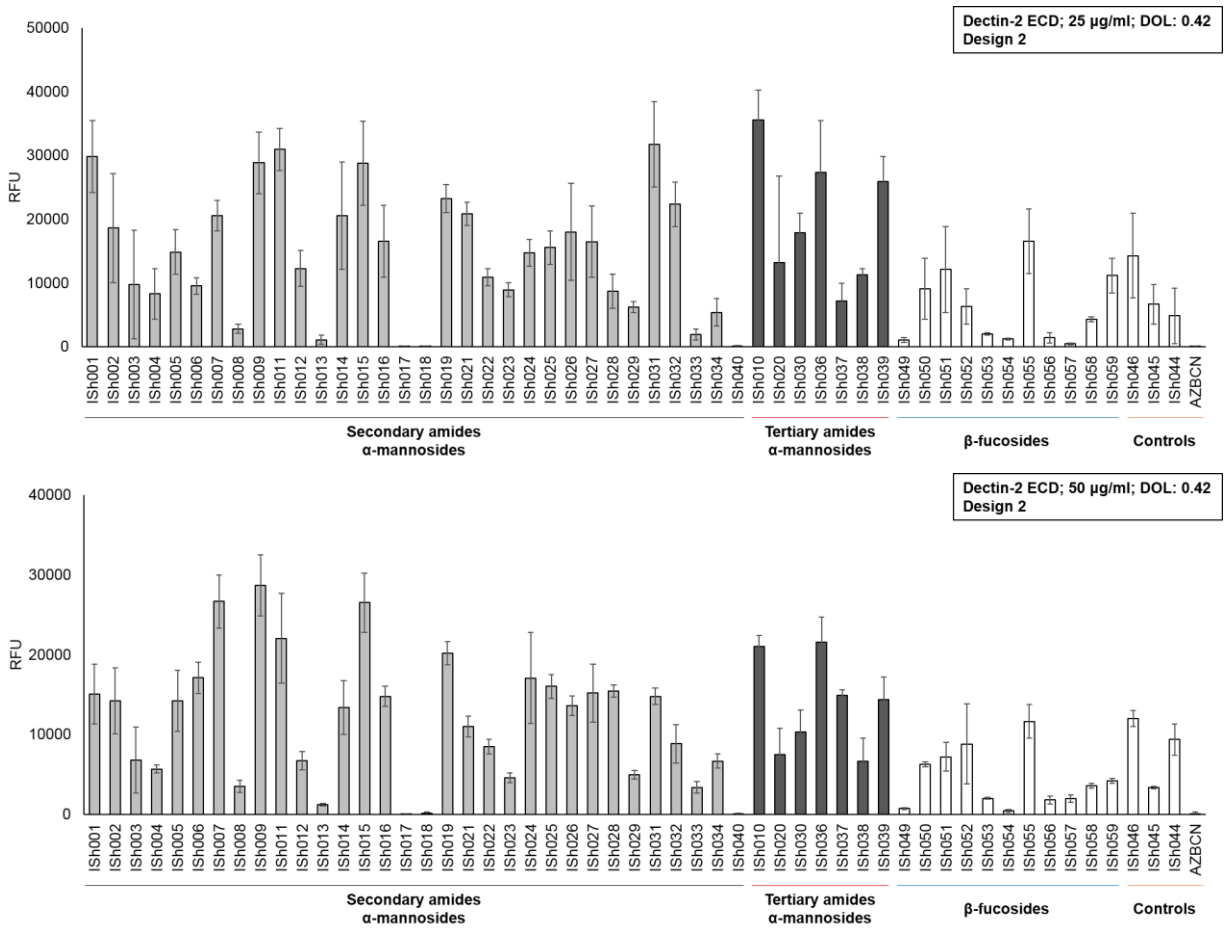



***Figure 67.*** *Screening with 25 and 50 µg/ml dectin-2-Cy3 (full library)*

First, the evaluation of the controls proves again that the added BCN-amine spacer helps the monovalent ligands' presentation on the microarray chips (**ISh001** and **ISh046**; **ISh002** and **ISh045**), and does not bind aspecifically to the lectin (**AZBCN**).

Comparing the glycomimetic binding profile to mannose (**ISh044**), the only known natural ligand printed on the slide, it is notable that many synthetic ligands show higher affinity towards the CLR. This intensity enhancement is more pronounced at 25 µg/ml, but can be appreciated at 50 µg/ml as well.

Although the overall binding profile shows resemblance to that of DC-SIGN, especially among the mannosylated secondary bisamides; and high intensity fluorescence values were obtained for **ISh001**, **ISh007**, **ISh015**, **ISh019** - suggesting that these are common ligands of the two CLRs, some striking differences can be observed.

In particular, dectin-2 recognizes on the microarrays tertiary bisamide α-mannosylated glycomimetics with good affinity, furthermore it also binds to compounds of the β-fucoside series. These structures were not or only barely recognized by DC-SIGN on chip. A comparative histogram, shown in **figure 68**, highlights this discriminative recognition in the above mentioned groups.

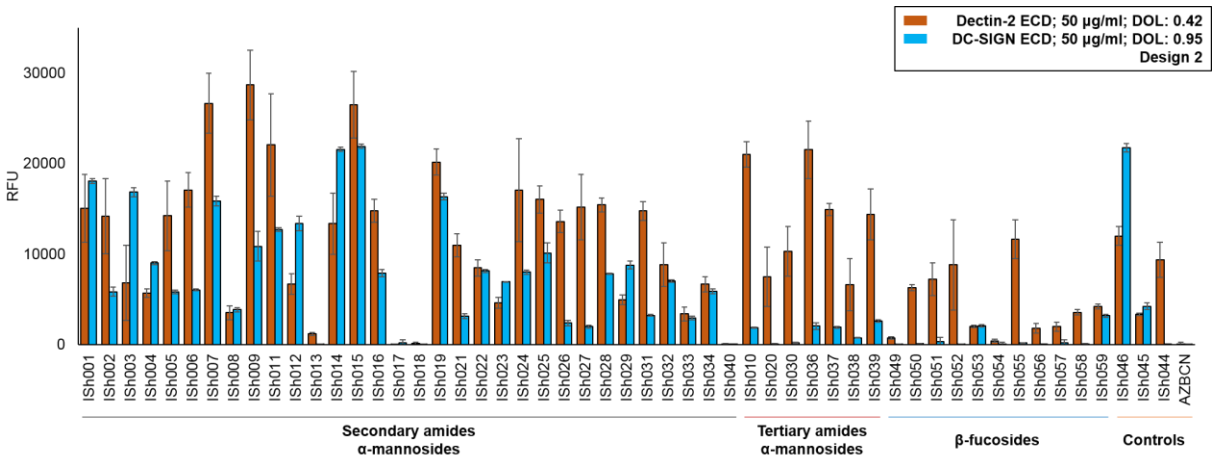



***Figure 68.*** *Comparison of binding profiles obtained with dectin-2 (orange) and DC-SIGN (blue) at 50 µg/ml. The mannosylated secondary amides display similar affinities towards the two CLRs. Significant differences can be observed, however, in the recognition of tertiary amide-substituted mannosides which are barely bound by DC-SIGN, but interact with dectin-2 – in many cases with similar affinity as the secondary mannosides. Interestingly, dectin-2 also exhibits some tolerance towards a few β-fucosylated compounds that are not recognized by DC-SIGN.*

#### 2.2.3.4.7 Proposed structural explanation of the observed discriminative binding

The differential binding of these particular sets of bisamide structures implies the possibility of an unprecedented selectivity between dectin-2 and DC-SIGN towards glycomimetic ligands. To explore the structural basis of this potentially selective recognition, we started to further examine the binding site of DC-SIGN and the recently resolved X-ray image of the dectin-2 CRD,[302] complexed with $Man_9$.[424] Interestingly, the oligosaccharide cross-links two dectin-2 CRDs in the crystal by binding different Manα1-2Man containing branches in the two binding sites (D1 and D2 arm; Manα1-2Manα1-2Man and Manα1-2Manα1-3Man, respectively).

The conserved primary binding site in the CRD of dectin-2 is lined by 5 amino acid residues where mannose can be accommodated in two orientations (swapped around a 180° axis). Similarly to DC-SIGN complexed with mannosylated oligosaccharides, the terminal sugar is oriented towards the centre of the CRD; in dectin-2, however, this area is less enclosed, due to the absence of the Phe313 residue present in DC-SIGN. The nonreducing end mannose positioned over this surface contacts a tryptophan (Trp182), an arginine (Arg198) and a valine residue (Val192). Additionally, apart from the typical metal-coordination and H-bonding around the principal $Ca^{2+}$ binding site, the subterminal mannose residue also forms van der Waals interactions with the same Trp182 residue. (**Figure 69.B**)

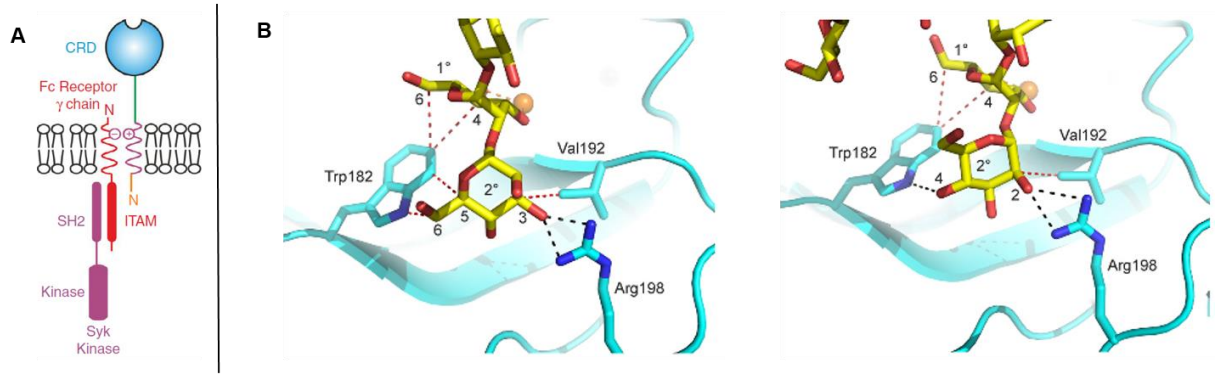



***Figure 69.*** *The structure of dectin-2. A. The CLR consists of a CRD, a short neck, transmembrane region and an ITAM-containing cytoplasmic tail. This domain can associate with the common Fc receptor γ subunit FcRγ to trigger cytokine production. B. Mannose oligosaccharide binding mode at the primary $Ca^{2+}$-site. The sugar residue coordinated by the metal can be oriented in two directions (swapped around a 180° axis). The nonreducing end sugar-ring, when positioned over the surface of the CRD center can form interactions with Trp182, Arg198 and Val192. Modified from Drickamer et al.[94] and Feinberg et al.[424]*

In the opposite direction of the primary binding site in DC-SIGN, we can find Val351 – a residue that plays key role in the binding mode of the parent pseudo-disaccharide **ISh002** and the bisbenzylamide **ISh001** (see discussion of the crystal structure and NMR evidences in the Introduction). This valine contacts the cyclohexane ring in the two glycomimetic ligands, and in the case of **ISh001** (Man030), the bisbenzylamide substituents face the CRD surface to establish lipophilic interactions.

By overlaying the dectin-2 and DC-SIGN CRD structures (courtesy of Dr. Monica Civera, University of Milan, Milan, Italy), we could observe the high similarity of the two tertiary structures, but at the same time, differences in the crucial interaction points became visible. The valine found in DC-SIGN is replaced by a histidine residue in dectin-2 (H171), and additionally, it is located on a differently oriented loop. This loop adopts a more open orientation in dectin-2, and could offer more available space in the region where the cyclohexane substituents are positioned in DC-SIGN. This difference in the CRD structure may be able to provide the answer for the discriminative recognition of the tertiary bisamide mannosylated glycomimetics.

The overlaid images of the CRDs of DC-SIGN (blue, PDB code: 2XR5) and dectin-2 (orange, PDB code: 5VYB) are shown in **figure 70** (image generated by the Fieschi Group, IBS, Grenoble, France). For the purpose of clarity, the dectin-2 ligand has been removed, while the pseudo disaccharide has been retained in the aligned figure.

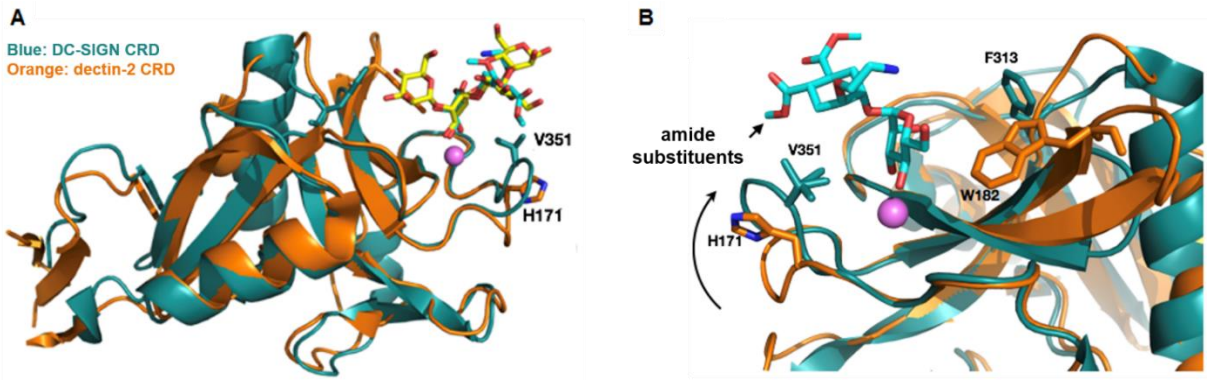



***Figure 70.*** *Overlaid images of the CRD of dectin-2 (orange) and DC-SIGN (blue). A: The aligned images of dectin-2 complexed with* $Man_9$ *(PDB code: 5VYB; a portion of the oligosaccharide is shown in yellow) and DC-SIGN complexed with the pseudo-disaccharide* ***ISh002*** *(PDB code: 2XR5, ligand is shown in cyan) show a high degree of structural similaritiy in the two CRDs. The natural ligand in dectin-2 points one mannose residue towards the center of the CRD (where Trp182, Arg198 and Val192 are found) – similarly to DC-SIGN-natural mannose oligosaccharide complexes. B: Zoom on the primary binding sites from the opposite direction (dectin-2 ligand removed for a clearer view). In DC-SIGN, the area of Val351 is occupied by the pseudo-disaccharide ligand that contacts the amino acid by the cyclohexane sugar mimetic. The two methyl ester groups are oriented towards the protein surface, and amide moieties in this position presumably point towards the same loop (small arrow). This loop could effectively prevent accommodation of larger groups, such as tertiary amide moieties, in this area. The different orientation of this loop in dectin-2 (presenting His171), however, leaves more space in the vicinity of the primary binding site, as it is highlighted by the curved arrow. This difference in loop orientation may provide an explanation for the discriminative recognition of tertiary amide ligands in the CRD of dectin-2. Courtesy of the Fieschi group.*

With these results in hand, we reached a point where the microarray screening and a computational hypothesis clearly indicated a structural feature in the glycomimetic ligands that could be potentially expoited to generate selectivity among different CLRs. This was one of the main goals of the preparation of the ligand library and setting up the microarray-based assays.

A practical issue has been delaying this further assessment, however. While during a cycle of CLR production 50 mg/l culture DC-SIGN and 10 mg/l DC-SIGNR are obtained, dectin-2 expression yields only 0.5 mg/l culture of the purified ECD.

In each case, the CLRs are first overexpressed in inclusion bodies which are then solubilized and protein refolding takes place through flash dilution. For purification, a His or streptavidin tag is incorporated at the N-termini and the CLRs are purified based on their interactions with carbohydrates, for example, by using a mannan column. Conditions throughout these steps require fine tuning for each CLR to obtain good yields, but in the case of dectin-2, the buffer used in the protein refolding step has not been fully optimized yet.

The low amount of available CLR and the numerous, possibly selective candidates (**ISh010**, **ISh020**, **ISh026**, **ISh030**, **ISh036**, **ISh037**, **ISh038**, **ISh039**, **ISh050**, **ISh051**, **ISh055; figure 71**) did not enable the evaluation of their affinity through the earlier described SPR setup. Instead, we chose a multivalent system to enhance the weak interactions between the protein and glycomimetic ligands, in order to facilitate the detection and quantitatve analysis of binding events.

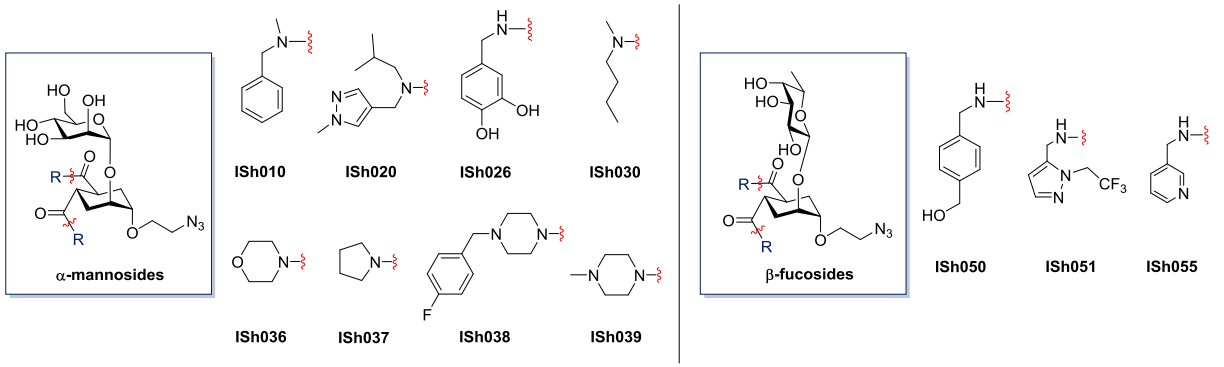



***Figure 71.*** *Possibly dectin-2 selective ligands chosen for further studies. A group of 11 glycomimetic ligands – 8 tertiary α-mannosides and 3 β-fucosides were selected, as they were recognized by dectin-2, but not by the other CLRs on the microarray-chips.*

### 2.2.3.4.8 Multivalent constructs with improved affinity: neoglycoproteins (NGPs) and gold-nanoparticles (GNP)

On a rationally designed multivalent construct, the monovalent carbohydrate ligands' density, topology of presentation and orientation should match the binding sites (CRDs) found on the lectin. As it has been discussed before, these parameters have major influence on the affinity and selectivity of sugar-CLR interactions. For DC-SIGN - a well-characterized lectin with respect to the spatial arrangement and geometry of its CRDs - several glycoconjugates, including glycopolymers,[34c, 425] glycodendrons,[298, 426] glycodendrimers,[140a] calixarenes[427] and gold-nanoparticles[428] have been built in the recent years to amplify these interactions.[231]

From these, our group has focused on glycomimetic-presenting multivalent dendrons and dendrimers of various valency, length and rigidity.[157d, 429] Notably, a hexavalent "rod" structure was found to be one of the most potent DC-SIGN-mediated *trans*-HIV infection inhibitors ($IC_{50}$: 24 nM). In this glycodendrimer, the central spacer moiety extends the two ends of the construct at the appropriate distance to bridge two CRDs of the four, found in DC-SIGN ECD. The 3-3 monovalent glycomimetics (Man030 or **ISh001**, structurally optimized for this CLR), displayed at the two ends, represent high local ligand density and can therefore further enhance avidity. (**Figure 72**) This dendimer of controllable size and valency is a proof that a fine-tuned multivalent platform, carrying an appropriately selected, high affinity monovalent glycomimetic ligand, can increase ligand potency to the levels of optimized pharmaceutical leads (in contrast to the weak monovalent ligand affinities) to effectively target CLRs.

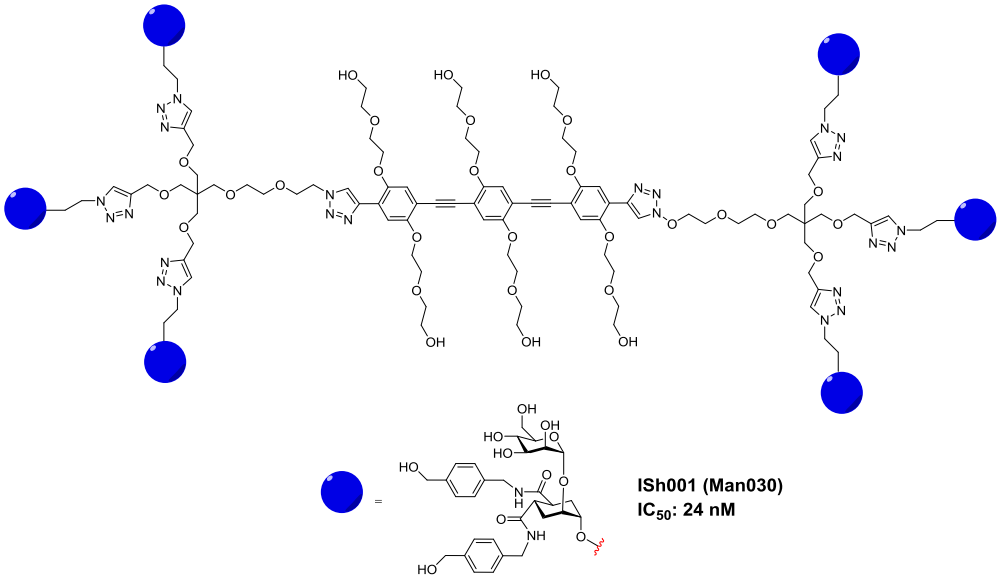



***Figure 72.*** *A hexavalent "rod"-like dendrimer, carrying 6 copies of ISh001 (Man030). This ligand is up to date one of the most potent inhibitors of DC-SIGN-mediated trans-HIV infection ($IC_{50}$: 24 nM). Multivalent constructs, such as this one, can span the distance between different CRDs in the DC-SIGN tetramer, and flexibly present the glycomimetic structures at high local density at the two ends of a rigid rod-core. This leads to increased avidity effects, and enhances affinity towards the CLR more significantly, than 6 individual monovalent ligands would.*

Due to the lack of information regarding the binding site organization and oligomerization in dectin-2, we could not rationally design a multivalent system to simultaneously target several CRDs. Instead, we opted for a less well-controlled, but typically higher-valency platform to choose the most potent candidates. The above mentioned 11 compounds were sent to Blanka Didak, GLYcoDiag, Orléans, France, to be presented on neoglycoproteins (NGPs).[430]

At GLYcoDiag, NGPs are built on a BSA-core which is first densely functionalized with alkyne-terminated linkers. Once more, the azido-group found at the end of the glycomimetic ligands' linker tail is used to tether the ligands to the protein via CuAAC reactions. The length of the additional linker arms can help to tune the valency of the constructs - ideally up to 30 monovalent ligands can be linked to BSA. Neoglycoproteins of all 11 ligands have been successfully prepared by Blanka Didak, but their affinities towards dectin-2 is yet of be analyzed.

The $IC_{50}$ values of these NGPs will later be determined, for example, by SPR. Affinity values of NGP constructs fall typically in the nanomolar range.

The enhanced affinity of monovalent glycomimetic ligands towards DC-SIGN, as compared to mannose and the natural disaccharide Manα1-2Man suggested that their presentation on gold nanoparticles could improve GNP recognition by the receptor and possibly help their uptake into DCs. Therefore, the two most extensively studied DC-SIGN ligands – **ISh001** and **ISh002** (Man030 and pseudo-disaccharide) – were re-synthesized and sent to Midatech Pharma (Derio, Spain) to functionalize gold nanoparticles. Glycomimetic ligand-functionalized GNP preparation and related assays were performed by Valentin Cognet, in collaboration with the laboratory of Dr. N. Reichardt (CIC biomaGUNE, San Sebastián, Spain).

GNPs are of particular interest as targeted drug-delivery systems in cancer research and in medical imaging (immunolabeling, X-ray imaging, radiotherapy), but they exhibit an increased risk of cytotoxicity.[431] Furthermore, their production has to comply with regulatory requirements, in terms of quality, characterization and reproducibility, i.e. the general GMP (good manufacturing practice) standards.

The proprietary synthetic platform developed by Midatech Pharma has high controllability of synthesis and undergoes a comprehensive characterization. Ultrasmall nanoparticles, with a gold core of 100-110 atoms are

obtained through a modified Brust-Schiffrin method.[432] The one-pot synthesis allows the reduction of a gold salt in the presence of thiol ligands using sodium borohydride to form divalent particles. GNPs are made of a non-immunogenic monosaccharide and a polyethylene glycol chain bearing a terminal carboxylic acid. In a second step, glycomimetics are linked through an amide bond to the PEG chain. The glycomimetic ligand-functionalized GNPs were used in microarray-based quenching assays where fluorescent intensity signals of fluorescently labeled lectins is diminished upon binding to carbohydrate coated GNPs. Additionally, cellular uptake experiments are planned with the glycomimetic gold-nanoconstructs.

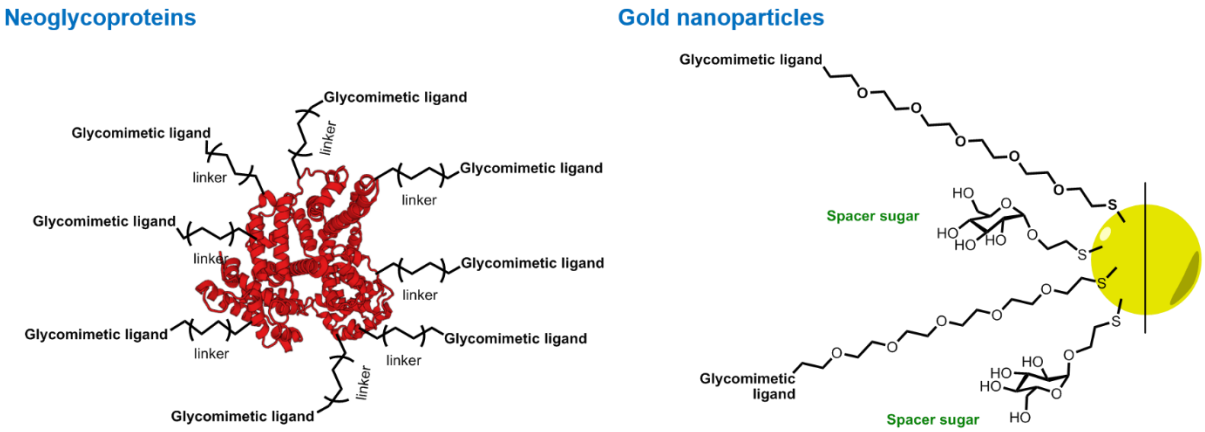



***Figure 73*** *Multivalent constructs of glycomimetic structures. Neoglycoproteins are built on a BSA-core that can present up to 30 monovalent ligands, tethered to the protein by linkers of various length. Similarly, gold nanoparticles present high numbers of monovalent glycomimetic ligands, anchored to a gold-core by PEG-linkers. Ligand density and GNP solubility can be tuned by the addition of monosaccharide spacers during the functionalization of GNPs.*

## 2.2.4 Conclusions to Part I (Chapters I and II)

In summary, we prepared an α-mannose and β-fucose-based glycomimetic ligand library, derived from the structure of the natural disaccharide Manα1-2Man. These ligands feature a cyclohexane sugar-mimic ring that is diversified by two amide groups, in order to selectively target different C-type lectin receptors.

The library was printed on microarray slides and a screening platform was set up to test the glycomimetics' binding to mannose- and fucose-specific CLRs. For an appropriate accessibility, a doubly functionalized cyclooctyne-linker was introduced as a spacer moiety. Plant and fungal lectins of known carbohydrate-specificity were used to validate the microarray system, with additional controls included and evaluated. The introduced linker did not show any negative effect during any of the screenings; and the ligand recognition on-chip was consistent with the carbohydrate-preference of the examined lectins.

The glycomimetic arrays were probed with a set of four human CLRs, including DC-SIGN, DC-SIGNR, langerin and dectin-2. All of these lectins are known to interact with Manα1-2Man, a common natural ligand, while they differ significantly in their ability to recognize fucosylated oligosaccahrides. From the tested human CLRs, only dectin-2 showed affinity towards the β-fucoside structures, although lower than towards the α-mannosides.

Variations in the nature of the bisamide substituents were recognized differently by the CLRs, or, in other words: the diversification of bisamide appendages on the glycomimetics generated differing recognition profiles. DC-SIGN and DC-SIGNR, two closely related receptors display good tolerance towards the pseudo-mannobioside bisamide structures, with similar binding profiles. Langerin, on the other hand, shows relatively weak affinity towards the majority of the glycomimetic ligands.

The binding interactions of a few selected glycomimetic compounds were tested towards DC-SIGNR and langerin in additional SPR inhibition assays. Ligand ranking in microarray and SPR experiments did not correlate in every case, however – which only underlines the fact that the assay format has major influence on studying sugar-lectin interactions. The binding affinities and selectivities obtained by different assays often differ significantly, and thus delay the discovery process.

Most importanty, dectin-2 was found to interact with tertiary amide α-mannosides – a group of ligands which were poorly recognized by the other three CLRs. In a preliminary attempt to explain this discrimination, we overlaid the known crystal structure of DC-SIGN and dectin-2 CRDs, and found that the orientation of a loop in the vicinity of the primary binding site may be accountable for the differential binding. Dectin-2 may be able to accommodate larger substituents (e.g tertiary amides) in this region, while the proximity of the same loop in DC-SIGN might effectively hinder binding of these ligands. Additionally, a few β-fucosylated glycomimetic structures also exhibited selectivity towards dectin-2 over DC-SIGN and DC-SIGNR on the microarrays.

These results indicate potential key features for the design and optimization of selective dectin-2 ligands. Characterization of the binding properties of dectin-2 and selected glycomimetic ligands in other binding assays is underway, and the structures' will be used for the construction of multivalent systems.

In conclusion, the optimized microarray sytem allows the rapid screening of C-type lectin receptors against the glycomimetic library and offers an ideal tool to select promising structures to selectively target CLRs. At the same time, the method has an extremely low resource demand compared to other binding assays, and can therefore be used as the primary analysis, before performing complementary tests.

## 2.2.5 EXPERIMENTAL SECTION

### 2.2.5.1 General

#### 2.2.5.1.1 Conjugation of the ligands with the bifunctional linker

The originally lyophilized compounds were dissolved in water to a concentration of 10 mM. In a few cases, the ligands were first dissolved in a minimal amount of DMSO (<5%, Thermo Scientific Molecular Probes™), due to their low aqueous solubility, then diluted to the same concentration (**ISh006**, **ISh013**, **ISh018**, **ISh021**, **ISh027**, **ISh032**, **ISh033**). Aqueous solutions were prepared from nanopure water produced with a Diamond UV water purification system (Branstead International, IA, Madrid, Spain). The SPAAC-reactions were performed by stirring 10 mM solution of the ligands overnight at room temperature with equimolar amounts of BCN-amine (Sigma Aldrich) in water. The conversion was monitored by MALDI-Tof mass spectrometry using 2,5-dihydroxybenzoic acid (DHB) as matrix (5 mg/ml in MeCN : water = 3 : 7 + 0.1% aqueous TFA, containing 0.005% NaCl). MALDI-Tof mass measurements were performed on an Ultraflextreme III time-of-flight mass spectrometer equipped with a pulsed Nd:YAG laser (λ 355 nm) and controlled by FlexControl 3.3 software (Bruker Daltonics, Bremen, Germany). The acquisitions (total of 2000-3000) were carried out in positive or negative reflector ion extraction of 450 ns and laser frequency of 1000 Hz. Laser fluence was set up to 60-80% and the *m/z* range was chosen according to the mass of the sample. The accumulated spectra were processed using Bruker software FlexAnalysis 3.3.

#### 2.2.5.1.2 MALDI-Tof results of SPAAC reactions with BCN-amine

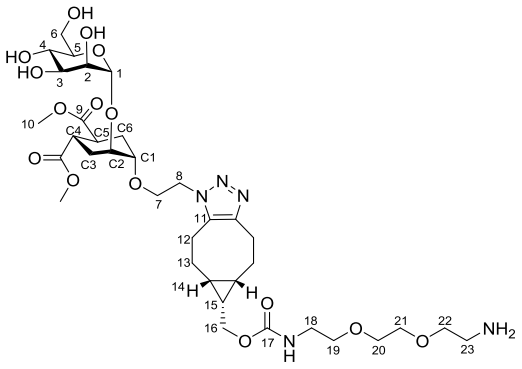
The MALDI-Tof *m/z* values before and after conjugation with BCN-amine (MW = 324.42) are reported in the table below. (**Table 3**)

| Ligand (before conjugation) | MS calculated $[M+Na]^+$ | MS (MALDI-TOF) before conjugation | MS (MALDI-TOF) after conjugation |
|---|---|---|---|
| **ISh001** | 696.286 | 696.564 | 1020.981 |
| **ISh002** | 486.169 | 486.330 | 810.749 |
| **ISh003** | 696.286 | 696.580 | 1020.976 |
| **ISh004** | 696.286 | 696.576 | 1020.969 |
| **ISh005** | 696.286 | 696.561 | 1020.956 |
| **ISh006** | 756.306 | 756.654 | 1081.046 |
| **ISh007** | 752.276 | 752.580 | 1076.973 |
| **ISh008** | 745.218* | 745.0625* | 1044.960 |
| **ISh009** | 636.264 | 636.504 | 960.908 |
| **ISh010** | 664.295 | 664.555 | 988.952 |
| **ISh011** | 638.254 | 638.508 | 962.912 |
| **ISh012** | 752.348 | 752.778 | 1077.195 |
| **ISh013** | 788.327 | 788.689 | 1113.082 |
| **ISh014** | 732.267 | 732.661 | 1057.071 |
| **ISh015** | 732.267 | 732.643 | 1057.054 |
| **ISh016** | 540.264 | 540.585 | 865.058 |
| **ISh017** | 800.239 | 800.774 | 1125.236 |
| **ISh018** | 720.286 | 720.609 | 1045.005 |
| **ISh019** | 696.286 | 696.586 | 1020.992 |
| **ISh020** | 756.401 | 756.741 | 1081.135 |
| **ISh021** | 796.339 | 796.710 | 1121.105 |
| **ISh022** | 780.251 | 780.600 | 1104.993 |
| **ISh023** | 780.251 | 780.604 | 1104.999 |
| **ISh024** | 536.233 | 536.435 | 860.847 |
| **ISh025** | 706.338 | 706.646 | 1031.055 |
| **ISh026** | 700.244 | 700.559 | 1024.960 |
| **ISh027** | 688.295 | 688.601 | 1013.009 |
| **ISh028** | 536.233 | 536.457 | 860.880 |
| **ISh029** | 786.323 | 786.718 | 1111.120 |
| **ISh030** | 596.327 | 596.587 | 921.001 |
| **ISh031** | 668.254 | 668.554 | 992.966 |
| **ISh032** | 742.317 | 742.661 | 1067.061 |
| **ISh033** | 760.252 | 760.608 | 1085.012 |
| **ISh034** | 660.252 | 660.542 | 984.942 |
| **ISh036** | 596.254 | 596.580 | 921.000 |
| **ISh037** | 564.264 | 564.546 | 888.979 |
| **ISh038** | 810.361 | 810.815 | 1135.218 |
| **ISh039** | 622.317 | 622.625 | 947.049 |
| **ISh040** | 724.317 | 724.688 | 1049.093 |
| **ISh049** | 470.175 | 470.134 | 794.258 |
| **ISh050** | 680.290 | 680.185 | 1004.246 |
| **ISh051** | 764.256 | 764.131 | 1088.303 |
| **ISh052** | 652.259 | 652.160 | 976.320 |
| **ISh053** | 726.332 | 726.177 | 1050.353 |
| **ISh054** | 580.259 | 580.182 | 904.330 |
| **ISh055** | 622.260 | 622.173 | 946.326 |
| **ISh056** | 520.238 | 520.177 | 844.316 |
| **ISh057** | 606.332 | 606.235 | 930.388 |
| **ISh058** | 736.280 | 736.158 | 1032.297 |
| **ISh059** | 708.249 | 708.129 | 1032.297 |

* **MS** (ESI-TOF) calculated for $[C_{32}H_{37}N_5O_{13}Na_2]^+ = [M\text{-}2H^+\text{+}2Na]^+$

***Table 3.*** *MALDI-TOF m/z values before and after conjugation with BCN-amine*

Conjugate **ISh002** was fully characterized for analytical purposes and its characterization is reported below.



**$^1$H NMR** (400 MHz, $D_2O$): $\delta$ = 4.95 (s, 1H, $H_1$), 4.55 – 4.47 (m, 2H, $H_8$), 4.19 (d, 2H, $H_{16}$, $J_{16\text{-}15}$ = 7.6 Hz), 4.02 – 3.54 (m, 21H, $C_2$, $H_2$, $H_{7a,b}$, $C_1$, $H_3$, $H_4$, $H_{10}$, $H_{19}$, $H_{20}$, $H_{21}$, $H_{22}$), 3.14 – 2.97 (m, 4H, $H_{18}$, $H_{23}$), 2.91 – 2.75 (m, 2H, $C_4$, $C_5$), 2.12 – 2.32 (m, 4H, $C_3$, $C_6$), 2.09 – 1.84 (m, 2H, $H_{12a}$), 1.74 – 1.51 (m, 4H, $H_{12b}$, $H_{13a}$), 1.47 – 1.20 (m, 3H, $H_{15}$, $H_{13b}$), 1.15 – 0.94 (m, m, 2H, $H_{14}$)
**$^{13}$C NMR** (100 MHz, $D_2O$): $\delta$ = 177.4, 177.0 ($C_9$); 158.8 ($C_{17}$); 144.9 ($C_{11a}$); 136.0 ($C_{11b}$); 98.5 ($C_1$); 73.6, 73.4 ($C_5$, $C_3$); 70.7, 70.4 ($C_{C1}$, $C_2$, $C_{C2}$); 69.5, 69.3, 69.2 ($C_{19}$, $C_{20}$, $C_{21}$ $C_{22}$); 66.7 ($C_4$); 66.6 ($C_7$); 63.7 ($C_6$); 60.9 ($C_{16}$); 52.6, 52.5 ($C_{10}$); 47.7 ($C_8$); 39.9, 39.5 ($C_{23}$, $C_{18}$); 38.7, 38.6 ($C_{C4}$, $C_{C5}$); 28.5 ($C_{12}$, $C_{13}$); 26.6, 26.3 ($C_3$, $C_6$); 20.7 ($C_{12}$, $C_{13}$); 19.7 ($C_{15}$); 17.0 ($C_{14}$)
**MS** (MALDI-TOF) calculated for $[C_{35}H_{57}N_5O_{15}Na]^+$: 810.375; found: 810.749

## 2.2.5.2 Microarray screenings

### 2.2.5.2.1 Source plate preparation

Conjugate stock solutions were diluted to a ligand concentration of 1 mM in water. From these, 50 µM buffered solutions were prepared by dilution with sodium phosphate buffer (300 mM, pH ~ 8.5, 0.005 % Tween® 20, 10 % DMSO) to a final concentration of 50 µM. 40 µl of each solution was transferred into a 384-well source microplate (Scienion, Berlin, Germany) that could stored at -20°C and reused if necessary.

### 2.2.5.2.2 Ligand immobilization on the microarray surface

Glycomimetic ligands were immobilized on commercially available NHS-functionalized glass slides (7.5 cm long, 2.5 cm wide, Nexterion® Slide H - Schott AG, Mainz, Germany). The solutions (750 pL, 3 drops of 250 pL) were spotted onto the activated slides with robotic non-contact piezoelectric SciFLEXARRAYER microarray spotter S11 (Scienion, Berlin, Germany) in 4 replicates (9 different ligands per row), establishing the complete microarray that was printed in 7 copies onto each slide. After printing, the slides were placed in a 75% humidity chamber (saturated NaCl solution) at room temperature to dry overnight. Unreacted functional groups were quenched with ethanolamine solution (50 mM in 50 mM sodium borate at pH ~ 9 for 1 hour) and the slides were washed.
The dry microarray chips were then divided into subarrays by placing silicone-lined plastic gaskets on the glass surface, and fixing these with metal clamps. Protein incubations were performed using 16 or 8 Well ProPlate™ Slide Modules from Electron Microscopy Sciences, Hatfield, UK.

#### 2.2.5.2.3 Incubation with fluorescently labelled lectins and signal detection

The subarrays were then incubated with fluorescently-tagged lectins. After the allotted incubation time, the chips were washed again and scanned. Fluorescence measurements were performed in an Agilent G2565BA microarray scanner system (Agilent Technologies, Santa Clara, USA) with 10 μm resolution, using 2 lasers (532 nm, 633 nm). Signal saturation can be reduced in the scanner by adjusting the photomultiplier tube (PMT) sensitivity. Quantification was performed by ProScanArray® Express software (Perkin Elmer, Shelton, USA), employing an adaptive circle quantitation method from 50 μm (minimum spot diameter) to 200 μm (maximum spot diameter). Average RFU (relative fluorescence unit) values with local background subtraction and standard deviation of the mean were reported as histograms. Intensity values estimated for smeared, visibly contaminated or wrongly defined spots were manually removed from the quantification. Fluorescence intensity values over the spots directly correlate with the amount of bound protein and give a close approximation of the relative binding strength of the interaction between the lectin and the surface-immobilized ligands.
Unless stated otherwise, this protocol was followed during the microarray screenings discussed below.

#### 2.2.5.2.4 Trial to functionalize the glass slides with BCN-amine

NHS-activated glass slides were incubated overnight with 0.33 μg/ml of BCN-amine in 300 mM phosphate buffer at pH ~ 8.5. The concentration of the BCN-amine was chosen in order to have an excess number of molecules to react with all functional groups found on the surface, based on previous calculations for NHS-activated glass slides ($10^{14}$ molecule/$cm^2$). Unreacted sites were then blocked with ethanolamine solution, the slides were washed with water and dried. 2, 4 and 8 drops of 500 μM Lac-$(CH_2)_3$-$NH_2$ in PBS and TWEEN®-containing buffer was printed in 6 replicates on the slides, but no fluorescent signals were detected.

#### 2.2.5.2.5 Array design for screenings

Design 1 – Microarray design employed during the first screenings (without ligands **ISh016**, **ISh017** and **ISh049-ISh059**)
From the 1 mM stock solutions, 50 and 10 μM solutions were prepared in a 384-well source microplate by dilution with sodium phosphate buffer (300 mM, pH 8.5, 0.005 % Tween® 20, 10% DMSO). 3 drops of ca. 250 pl printing solutions were disposed on 7*1 subarrays on every slide (these subarrays could be incubated with 200 μl lectin solutions), each being able to fit 36*18 spots (6 replicates of the same compound) to include all ligands at both concentrations. The 10 μM spots eventually proved to be too weak to detect, therefore the results are obtained by quantification of the 50 μM series.

Design 2 – Microarray design employed in screenings performed on the whole glycomimetic library
From the 1 mM stock solutions, 50 μM solutions were prepared in a 384-well source microplate by dilution with sodium phosphate buffer (300 mM, pH 8.5, 0.005 % Tween® 20, 10 % DMSO). 3 drops of ca. 250 pl printing solutions were disposed on 7*1 subarrays on every slide (7.5 cm long, 2.5 cm wide NHS-activated glass slide), each being able to fit 36*17 spots (4 replicates of the same compound).

#### 2.2.5.2.6 Method validation with plant and fungal lectins of known specificity

Immobilization validation was performed by probing the printed glycomimetic array with common lectins of known (and broad) carbohydrate-specificity. Trials consisted of screening with plant and fungal lectins (ConA, PSA, LCA, AAL), acquired from commercial source (Vector Laboratories). These lectins were labeled with Alexa Fluor® fluorescent dyes (Alexa Fluor® 555 and 647 NHS-ester (succinimidyl ester) (Thermo Fisher Scientific)), and their degree of labelling (DOL) was estimated. The emission spectra of the different solutions used for

incubations were measured. The labeling protocols, as well as a comprehensive table of all lectins screened against the glycomimetic chip with regards to their fluorescent properties and carbohydrate specificities are reported in the Appendix.

#### 2.2.5.2.7 Fluorescent labelling with Alexa Fluor® fluorescent dyes

A typical fluorescent labelling procedure was performed on 1-2 mg/ml lectin in 50 mM sodium phosphate buffered solutions at pH ~ 8.5, by the addition of 1µl Alexa Fluor® fluorescent dye. After 40 minutes, the solutions were dialyzed, and their absorbance was measured. Based on the Lambert-Beer law, the ratio of the dye and protein in the labeled proteins could be determined, and eventually the DOL was established. Although the DOL is often not specified in literature, the rate of fluorescent tagging can have major effects on the efficiency of detection.

#### 2.2.5.2.8 Screening with human CLR

Human CLRs used during the first microarray assays were expressed and previously labelled by Silvia Achilli with Alexa Fluor®-555, while lectins probed later against the whole library were tagged with fluorescent Cy3 (cyanine 3). Single aliquots of the lectins stored at -80 °C were thawed on ice, immediately before incubations.
Solutions of the proteins at working concentrations were prepared by dilution in binding buffer (25 or 50 mM tris-buffer, 150 mM NaCl, 4 mM $CaCl_2$, 0,005% Tween®-20, 0.5% IgG-free BSA, pH ~ 8).

Recombinant human MBL (rhMBL) was purchased from R&D Systems. Human CLRs were expressed and provided by the laboratory of Prof. Franck Fieschi (IBS, Grenoble, France) and murine Fc-fusion CLRs were expressed and provided by Prof. Bernd Lepenies (University of Veterinary Medicine, "TiHo", Hannover, Germany) to the laboratory of Dr. Niels Reichardt (CIC biomaGUNE, San Sebastián, Spain).

## 2.2.6 APPENDIX

### 2.2.6.1 Human CLR preparation and labelling

Human CLR labellings were performed by Silvia Achilli, IBS, Genoble, France.

#### 2.2.6.1.1 Cloning

Standard pUC57 plasmids containing optimized synthetic human genes encoding human DC-SIGNR ECD (amino acids 78-399) and dectin-2 ECD (amino acids 46-209), designed for the efficient production in *E. coli*, were manufactured by GeneCust Europe (Luxembourg). PCR amplification using suitable primers and restriction enzyme digestion were used to sub-cloned DC-SIGNR ECD into the pET30-b (Novagen) between the NdeI and HindIII restriction sites and N-terminally strep tagged dectin-2 ECD into pET30-b (Novagen) between the XbaI and HindIII ones. The sequencing of each construction was done by Genewiz (Takeley, UK).

#### 2.2.6.1.2 Human CLR expression and purification

DC-SIGN extracellular domain (DC-SIGN ECD) and langerin extracellular domain (Langerin ECD) constructs were produced and purified as previously described.[189]

DC-SIGNR ECD and dectin-2 ECD was expressed in *E. coli* BL21(DE3) in 1 liter of LB medium supplemented with 50 µg/ml kanamycin at 37 °C. Expression was induced by addition of 1 mM isopropyl 1-thio-D-galactopyranoside (IPTG) when the culture had reached an A600 nm of 0.8 and maintained for 3h. The protein was expressed in the cytoplasm as inclusion bodies. Cells were harvested by a 20-min centrifugation at 5000 g at 4°C. The pellet was resuspended in 30 ml of a solution containing 0.150 mM NaCl, 25 mM Tris-HCl, pH ~ 8 and one anti-protease mixture tablet (Complete EDTA free, Roche). Cells were disrupted by sonication and cell debris eliminated by centrifugation at 100.000 g for 45 min at 4 °C in a Beckman 45Ti rotor. The pellet was solubilized in 30 ml of 6 M guanidine-HCl containing 25 mM Tris-HCl pH ~ 8, 150 mM NaCl and 0,01% beta-mercaptoethanol. The mixture was centrifuged at 100.000 g for 45 min at 4°C and the supernatant was diluted 5-fold, by slow addition with stirring, with 1.25 M NaCl, 25 mM $CaCl_2$ and 25 mM or 200 mM Tris-HCl pH ~ 8 for DC-SIGNR and dectin-2 ECD respectively. The diluted mixture was dialyzed against 10 volumes of 25 mM Tris-HCl, pH ~ 8, 150 mM NaCl, 4 mM $CaCl_2$ (buffer A) with 3 buffer changes. After dialysis, insoluble precipitate was removed by centrifugation at 100.000 g for 1h at 4°C.

The supernatant containing DC-SIGNR ECD was loaded on mannan agarose column (Sigma) for purification by affinity chromatography equilibrated with buffer A. After loading, DC-SIGNR ECD was tightly bound to the column and eluted in the same buffer without $CaCl_2$ but supplemented with 1 mM EDTA (buffer B). This step was followed by SEC (Size Exclusion Chromatography) using a Superose 6 column (GE Heathcare) equilibrated with buffer A. Fractions were analyzed by SDS-PAGE (12%) and DC-SIGNR ECD containing fractions were pooled and concentrated by ultrafiltration (YM10 membrane from Amicon).

The supernatant containing the Strep tagged dectin-2 ECD was loaded onto a StrepTrap HP column (GE Heathcare) at 4°C. Unbound proteins were washed away with buffer A before dectin-2 ECD was eluted with buffer C (150 mM NaCl, 25 mM Tris-HCl, pH ~ 8, 4 mM $CaCl_2$, 2.5 mM D-desthiobiotin). Eluted fractions were analyzed by SDS-PAGE (15%) and the dectin-2 ECD containing fractions were pooled and concentrated by ultrafiltration (YM10 membrane from Amicon).

Each protein construct was checked by N-terminal amino acid sequencing and mass spectrometry.

#### 2.2.6.1.3 Cy3 Labeling of DC-SIGN ECD, langerin ECD, DC-SIGNR ECD and dectin-2 ECD

1 mg of DC-SIGN, langerin, and DC-SIGNR ECD and 0.25 mg of dectin-2 ECD in 25 mM HEPES pH ~ 7,25, 4 mM $CaCl_2$ were prepared. 1 µL of 10 mg/ml Cy3-NHS ester (Gene Copoeia) were added per 0.3 mg of protein for the ECDs of DC-SIGN, langerin, DC-SIGNR, while 3 µL of Cy3-NHS were added per 0.3 mg of dectin-2 ECD. The reactions were gently shaken at RT for 2 h and then at 4°C for 4 h. Excess dye was removed by two dialysis (3.5k Z-lyser from Thermo Scientific) of 3h against 25 mM Tris pH ~ 8, 150 mM NaCl, 4 mM $CaCl_2$. The amount of attached Cy3 was estimated spectrophotometrically based on the dye molar extinction coefficient ($\varepsilon_{555}$ = 150 000 $cm^{-1}M^{-1}$) and protein extinction coefficients (DC-SIGN ECD $\varepsilon_{280}$ = 70400$cm^{-1}M^{-1}$; langerin ECD $\varepsilon_{280}$ = 56170, DC-SIGNR ECD $\varepsilon_{280}$ = 60890 $M^{-1}cm^{-1}$, dectin-2 ECD $\varepsilon_{280}$ = 65930 $cm^{-1}M^{-1}$).
The obtained degree of labeling (DOL) was 0.95 for both DC-SIGN ECD and DC-SIGNR ECD, 0.75 for langerin ECD and 0.42 for dectin-2 ECD.

#### 2.2.6.1.4 Screenings with Cy3 labelled human CLRs (performed by Dr. Sonia Serna and Silvia Achilli - CIC biomaGUNE, San Sebastián, Spain and IBS, Grenoble, France)

The immobilized ligands were probed with solutions of fluorescently labeled C-type lectins. Solutions of Cy3-labelled DC-SIGN ECD-Cy3 (50 µg/ml, DOL: 0.3) and DC-SIGNR ECD (150 µg/ml, DOL: 0.95), langerin-Cy3 (25 µg/ml, DOL: 0.75) and dectin-2-Cy3 (50 µg/ml, DOL: 0.4) were diluted in TBS (50 mM Tris·HCl, 150 mM NaCl, pH ~ 8.0) containing 4 mM $CaCl_2$, 0.5% BSA and 0.005% Tween®-20. For incubations, 200 µl of each lectin solution was applied to each subarray using 8 Well ProPlate™ Module incubation chambers. The microarray was incubated under gentle shaking overnight in the dark at 4°C. The slides were washed using TBS containing 4 mM $CaCl_2$ and water, dried with argon and the fluorescence was analyzed with a microarray scanner.

#### 2.2.6.1.5 Labelling and screening with rhMBL (performed by Dr. Sonia Serna, CIC biomaGUNE, San Sebastián, Spain)

50 µL of 0.5 mg/ml lectin (purchased from R&D Systems) was labelled with Alexa Fluor®-555 in PBS buffer at 4°C overnight. Excess of dye was quenched by the addition of 2 µl of 1M Tris pH ~ 7.5. Solutions of rhMBL (50 and 12.5 µg/ml) in TSM buffer containing 4 mM $CaCl_2$, 0.005% Tween®-20 and 0.5% BSA were incubated at 4°C, under gentle shaking overnight. The slides were then washed with TSM buffer (20 mM Tris-HCl (pH ~ 7.4), 150 mM NaCl and 1 mM $CaCl_2$) and water, dried in a slide spinner and the fluorescent intensity valies were measured with a fluorescent microarray scanner.

#### 2.2.6.1.6 Production of CLR-Fc fusion proteins (performed by João Monteiro, University of Veterinary Medicine, "TiHo", Hannover, Germany)

RNA was isolated from C57BL/6 mouse splenocytes and was reversely transcribed into cDNA using Reverse Transcriptase (New England Biolabs, Ipswich, MA). Polymerase chain reaction (PCR) was performed to amplify the cDNA encoding the extracellular part of each CLR using specific primers. The cDNA was then ligated into the pFuse-hIgG1-Fc expression vector (InvivoGen, San Diego, CA). The CLR-Fc vector constructs were either stably transfected into CHO cells or transiently transfected using the FreeStyle Max CHO-S Expression System (Life Technologies, Darmstadt, Germany). Intracellular staining with antihuman IgG-PE (Dianova, Hamburg, Germany) followed by flow cytometry analysis was done to verify the stable expression of CLR-Fc. The cell supernatant containing the CLR-Fc fusion proteins was collected and each fusion protein was purified using HiTrap Protein G HP columns (GE Healthcare, Piscataway, NJ). Sodiumdodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) with subsequent Coomassie stain confirmed the purity of the CLR-Fc fusion proteins. Western blot using anti-human IgG-HRP (Dianova) was performed to detect the presence of the fusion

proteins. Protein concentrations were determined using Micro BCA Protein Assay Kit (Thermo Scientific, Rockford, IL).

### 2.2.6.2 Lectin specificities of CLRs used during the microarray screenings

| | Lectin | DOL | Carbohydrate specficity | MW | Absorbance 0.1% | ε |
|---|---|---|---|---|---|---|
| CLRs probed in the initial trials (not containing the whole library) | DC-SIGN ECD - AF555 | 1.03 | High Man, Fuc | 38845 | 1.812 | |
| | DC-SIGNR ECD - AF555 | 1 | High Man | 36.996 | | 60390 |
| | Langerin ECD - AF555 | not defined | High Man N-linked oligosaccharides, terminal 6-sulfated Gal | 28359.1 | 1.972 | |
| CLRs probed on the whole library | DC-SIGN ECD - Cy3 | 0.95 | | | | |
| | DC-SIGNR ECD - Cy3 | 0.95 | | | | |
| | Langerin ECD - Cy3 | 0.75 | | | | |
| | Dectin-2 ECD - Cy3 | 0.42 | Man, GlcNAc, Manα1-2Man, Manα1-4Man | | | |
| | hMBL - AF555 | not defined | Man | | | |
| Plant, fungal and bacterial lectins | ConA - AF555 | 1 | Man | | | |
| | AAL - AF555 | 15 | Fuc | | | |
| | LCA - AF555 | 90 | Man | | | |
| | PSA - AF647 | 165 | Man | | | |
| | PA-IIL - AF555 | 100 | Man and Fuc | | | |
| Murine Fc-fusion CLRs | DCAR | | ND | | | |
| | Dectin-1 | | β-1,3-glycans | | | |
| | DCL-1 | | ND | | | |
| | CLEC9A | | ND | | | |
| | MGL-1 | | $Le^x$, $Le^a$, terminal Gal, GalNAc | | | |
| | SIGN-R3 | | High Man, Fuc | | | |
| | SIGN-R5 | | High Man, Fuc | | | |
| | MICL | | ND | | | |

***Table 4.*** *Comprehensive table of CLRs used in the microarray-based screenings*

### 2.2.6.3 SPR inhibition assays

#### 2.2.6.3.1 SPR inhibition assays – background

SPR biosensors are widely used to study interactions between receptors and small molecules, therefore they are popular in drug discovery. Similarly to other optical detection systems used, for example, in microscopy, the physical basis of this technique is the surface plasmon polariton.[433] These infrared or visible light electromagnetic waves propagate along the surface of a conductor, typically a metal (in this case gold), and are basically light waves trapped at the surface due to their interaction with the free electrons found in the conductor. As a result, the light wave and the electrons start a synchronized movement, and the charge oscillations' and the electromagnetic field's collective resonance creates the so-called surface plasmon. As the wave travels at the very boundary of the conductor and external medium's interface, the surface plasmon can be exploited to detect even the smallest changes in its environment.

When a beam of polarized light is focused on the conductor's surface, a fraction of the light will excite the free electrons around the surface, while the rest will be reflected from the surface. By changing the angle of incidence, and monitoring the intensity of reflected light, a minimum intensity will be observed, where the proportion of light exciting the electrons is the highest. This happens when the wave vector of the beam matches the wavelength at which the electrons oscillate, and this induces a wave-like resonance among the free electrons of the conductor. The incident angle, at which the intensity minimum is reached due to the loss of absorbed light is called the surface plasmon resonance angle ($\Theta_{SPR}$), or SPR-dip. The created electromagnetic field penetrates the thin metal, as well as its surrounding medium, and thus, the resonance frequency of the surface plasmon, and so the SPR angle depend on, among other things, the refractive indices of media on both sides of the gold chip. Another fundamental parameter affecting the reflected light is the temperature, therefore SPR experiments require good temperature control.

The surface plasmon's characteristics are very sensitive to changes in the conductor material. If the gold chip is functionalized, for example with a ligand, and proteins bind to this coated surface, the cumulative mass of the thin gold layer will lead to a change in the surface plasmon, i.e. the refractive index in the vicinity of the metal will be affected. This causes a shift in the angle at which the SPR-dip can be observed and can be recorded by a photo-detector array. Every $10^{-4}$ degree in angle shift is translated into 1 RU (resonance or response units).

Based on the sensor-chip functionalization, there are three main types of strategies used: immobilization of the ligand, the receptor, or a binding competitor. The last surface modification is most useful when the ligand has low molecular weight (such as smaller oligosaccharides) because its weak interactions on the surface would be hard to detect. Instead, in this case, a competitor, for instance, a natural glycan ligand is used to coat the gold surface. Buffered protein solution is passed through a microfluidic flow cell on the gold chip at fixed concentrations and the proteins adhere to the chip as they bind to and release the natural ligand immobilized on the surface. Then, the studied ligand is added at increasing concentrations to the flow. The solution phase carbohydrate ligands, if recognized by the similarly solution phase protein, will inhibit the receptor's binding to the surface (the competitor). This will change the mass of the gold surface and thus, the surface plasmon resonance, until the interaction reaches a stable association rate constant. Equilibrium occurs when the number of ligands associating with and dissociating from the receptor is equal. (**Figure 74**Figure 74)

Having recorded this maximum at a certain concentration, the analyte ligand solution is replaced again with the buffer. When the analyte is added to the flow by constantly increasing the concentration, a full $IC_{50}$ curve can be recorded, whereas at fixed concentrations, single-point SPR data is obtained.

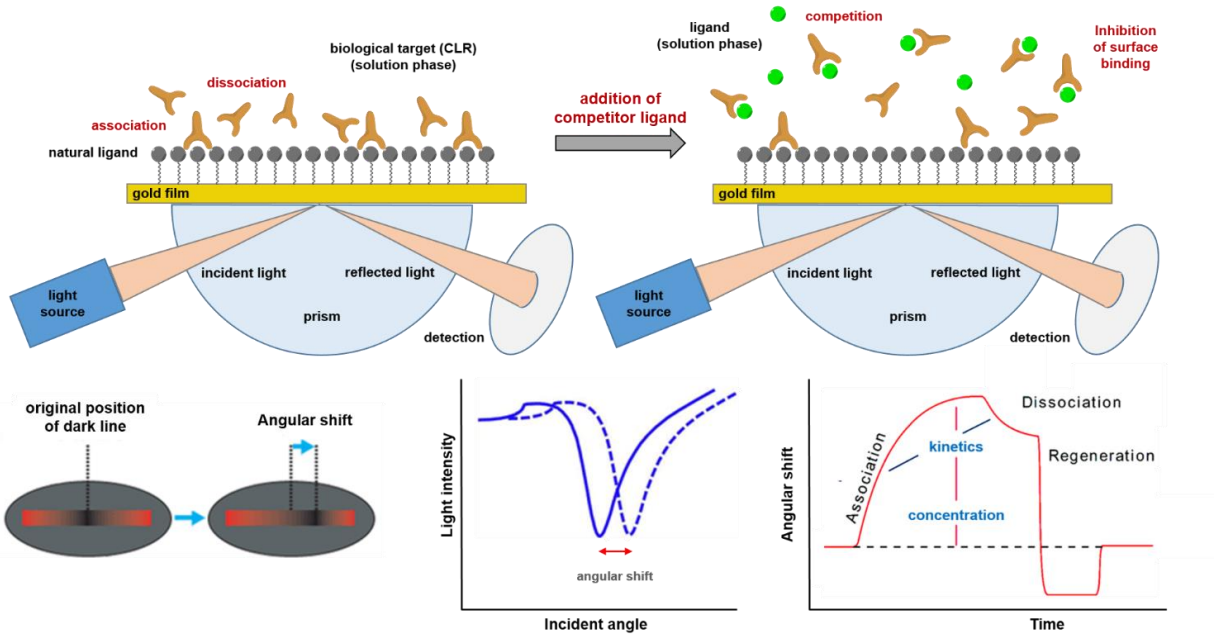



***Figure 74.*** *Surface plasmon resonance-based inhibition assay. A beam of polarized light is focused on a natural ligand coated gold film. The biological target is passed through a flow cell over the surface, associating on and dissociating from the natural ligand. In the reflected light, a dark line, caused by the surface plasmon phenomenon can be detected. The SPR dip is recorded in equilibrium, then the analyte ligand is injected into the flow. As this ligand competes with the natural, immobilized ligand, the proteins' binding to the surface is inhibited. The weight of the metal film changes, which affects the surface plasmon, and, as a consequence, the angle of SPR-dip shifts, until reaching the new equilibrium. Once the full sensorgram is recorded at different concentrations, the $IC_{50}$ values can be calculated.*

The affinity of glycomimetic ligands in **figure 53, figure 55** and **figure 58** were measured on a mannosylated-BSA functionalized surface. The glycoprotein displays the Manα1-3[Manα1-6]Man trisaccharide at 15 glycosylation sites, and exhibits μM affinity towards DC-SIGN ECD as determined by surface titration. The full $IC_{50}$ curve was recorded for each ligand.

### 2.2.6.3.2 SPR inhibition assays with selected ligands, based on microarray screening results

The SPR experiments were performed on a BIAcore T200 using a CM3 sensor chip. Flow cells were activated as previously described.[201] Flow cell 1 was functionalized with BSA, blocked with ethanolamine and subsequently used as a control surface. Flow cells 2 and 3 were treated with BSA-Manα1-3[Manα1-6]Man (Dextra) (60 μg/ml) in 10 mM NaOAc pH 4 to reach different binding densities and blocked with ethanolamine. The final densities on flow cells 2 and 3 were 2126 and 2085 RU, respectively. The affinity of the various compounds for DC-SIGN ECD were evaluated via an established inhibition assay[238] in which DC-SIGN ECD was injected at 20 μM alone or in the presence of increasing concentrations of inhibitors (ranging from 0 to 5 mM). Injections were performed at 5 μl/min using 25 mM Tris-HCl pH ~ 8, 150 mM NaCl, 4 mM $CaCl_2$, 0,05% P20 surfactant as running buffer. The surface was regenerated by the injection of 50 mM EDTA.
Titration curves of SPR assays against DC-SIGN and DC-SIGNR; as well as the inhibition test results with langerin are reported below. (**Figure 75** and **Figure 76**)

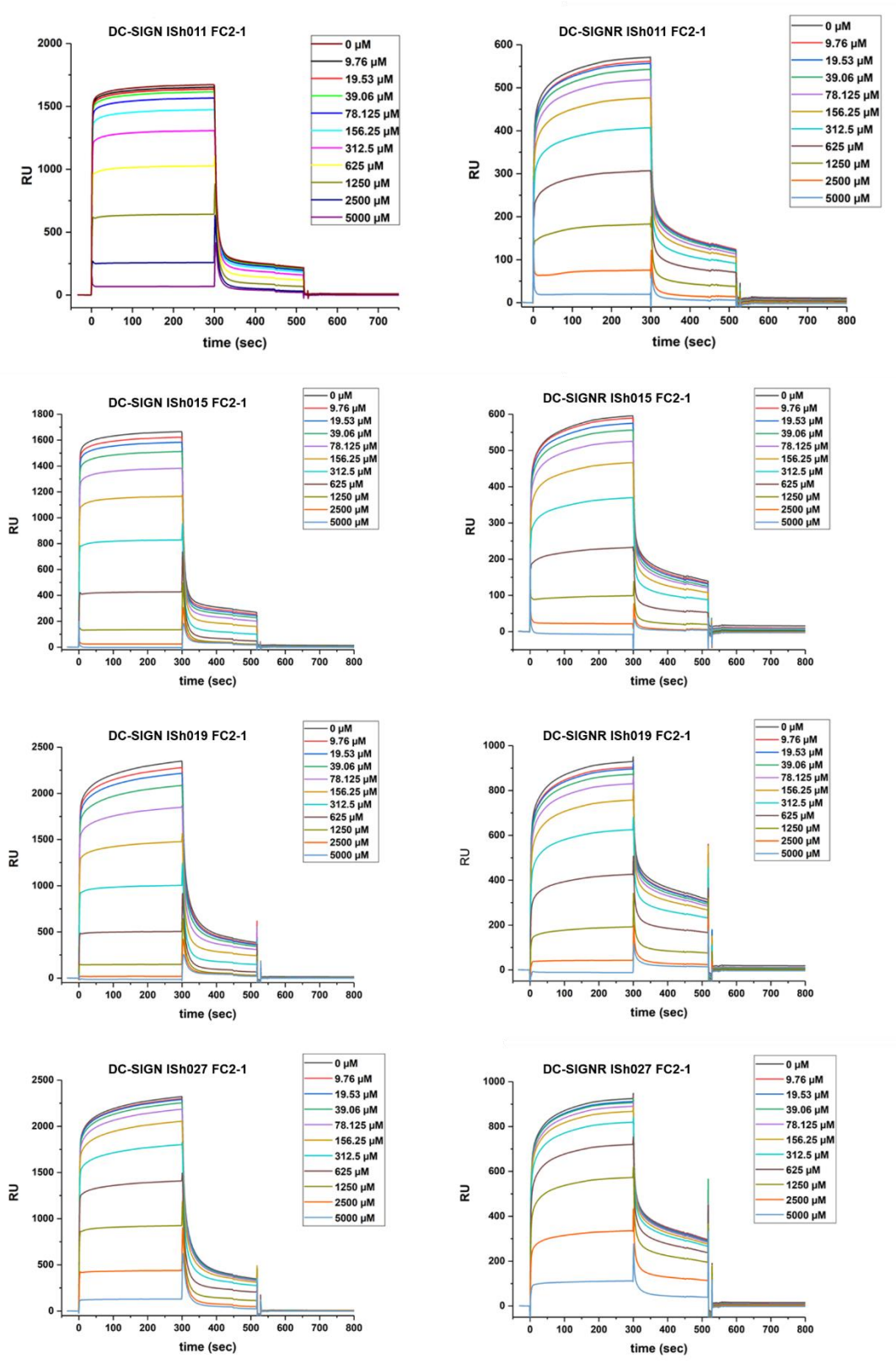



***Figure 75.*** *Sensorgrams for the titration of ISh011, ISh015, ISh019 and ISh27*

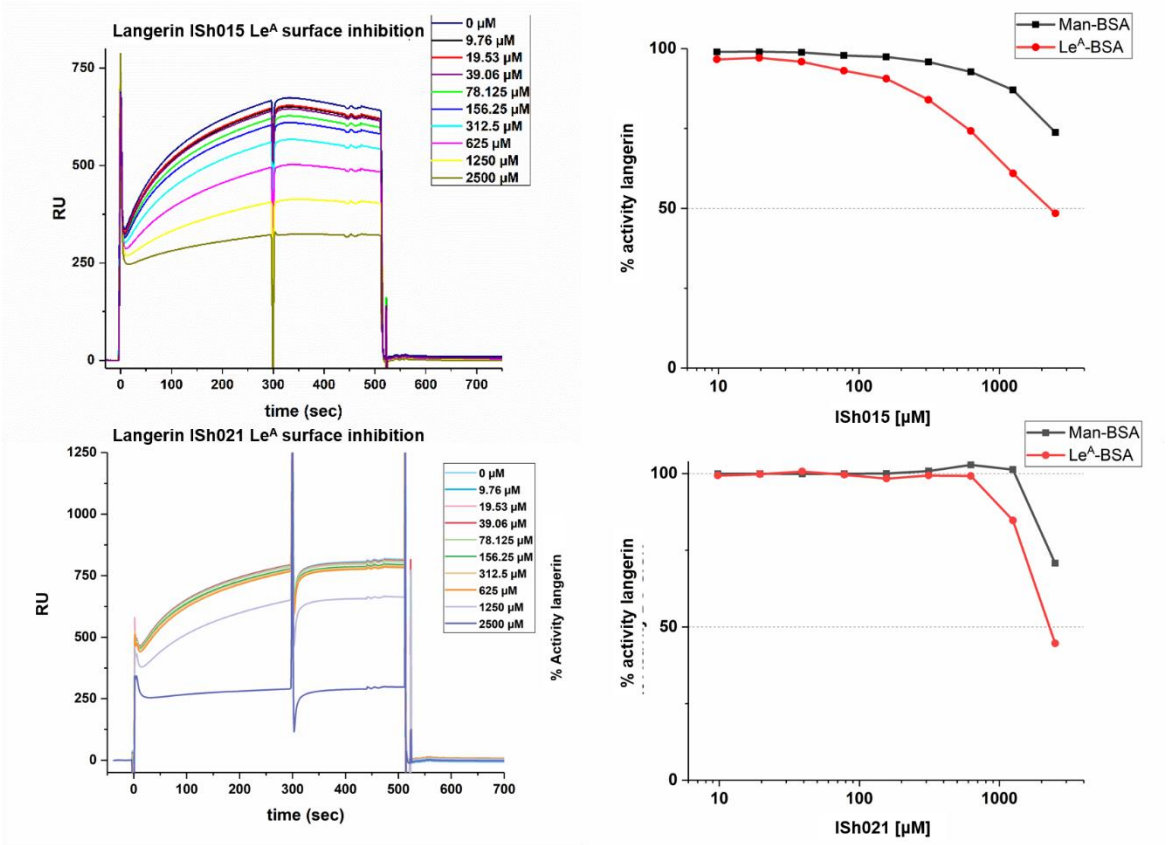



***Figure 76.*** *Left panel: SPR competition assay of ISh015 against langerin ECD/Lea-BSA interaction. Right panel: Comparison of inhibitory potency of ISh015 towards langerin interaction with Man-BSA and Lea-BSA.*

# Part II

## Chapter III

# 3. Design and synthesis of high affinity DC-SIGN inhibitors

## 3.1 Ligand binding sites in the CRD of DC-SIGN

Until this point we have discussed many aspects and challenges of C-type lectins as potential drug targets. The involvement of CLRs in pathogen recognition, immune cell regulation and signalling renders these proteins promising targets for anti-adhesive therapies and as vaccine adjuvants in drug-delivery systems. In spite of these pharmaceutically relevant application areas, small molecule inhibitors of lectin-carbohydrate interactions are sparse.

The underlying reasons can be traced back, on one hand, to the shallow, hydrophilic and largely solvent-exposed CRDs that show structural resemblance across different CLRs; and, on the other, to the similarly hydrophilic ligands (both natural and synthetic) that contradict the requirements of oral availability. As a result, small-molecules (i.e. monosaccharides, smaller oligosaccharides) exhibit a very low affinity towards these receptors, with $IC_{50}$ values well-below pharmaceutical relevance.

As a prime example of a thoroughly-investigated C-type lectin, DC-SIGN is an attractive target for drug discovery, owing to its role in pathogen uptake, innate immune-signalling and viral transmission (HIV, Ebola, Dengue, etc., see Introduction). Additionally, DC-SIGN is almost exclusively expressed on dendritic cells, and exhibits therefore a high potential for targeted drug-delivery.

Pathogen capture is driven by carbohydrate coordination at the principal $Ca^{2+}$-binding site in the CRD, and endosomal ligand release is presumably mediated by the pH dependent change in the CLR oligomerization state, as well as by the active $Ca^{2+}$ transport from the endosome.

### 3.1.1 Druggable sites in the CRD of DC-SIGN

Natural DC-SIGN ligands are recognized by two main orientations in the primary sugar-binding site: in addition to the metal-chelation, mannosylated oligosaccharides form interactions in the central region, lined by Phe313, while fucosylated structures tilt towards the outer area of the CRD, where Val351 can be found (for further details, refer to the Introduction).

DC-SIGN, similarly to other CLRs, binds monovalent monosaccharides very poorly (mM range; $K_D$ (Man): 3.5 mM);[434] while multivalently presented epitopes can display affinities in the low nanomolar range.[157d, 435] Therefore, ligands for DC-SIGN are developed mainly by following two strategies: either weak affinity ligands are presented in multivalent systems to enhance avidity, or, glycomimetic structures are designed to maximize interactions with the CRD surface and improve binding properties.

In particular, glycomimetic structure optimization relies - in the majority of the cases – on the structure of native oligosaccharides. Glycomimetics are often designed to form additional contacts with secondary binding sites in the close proximity of the $Ca^{2+}$, but this means that rationally built structures have been mostly limited to form contacts around the primary sugar-binding site, often inspired by the binding pose and interactions of known ligands. Examples of such ligands include those described by Tomasic *et al.*[240] and Mitchell *et al.*[239] (mannose-

based **1.14** and **1.15**, **figure 25**), the fucosylated glycomimetics shown in **figure 24**; and the Shikimic acid-analogues (**1.9**) reported by Garber *et al.*[236]

Surprisingly, the pseudo-disaccharide glycomimetic ligand **1.18**, derived from the structure of Manα1-2Man, assumes a unique pose in the CRD of DC-SIGN, and interacts with the protein surface around Val351 – a residue that had earlier been associated only with fucosylated ligands.[227]

Although rational design offers an ideal template to install beneficial structural elements - for example, to tune pharmacokinetic properties or affinity, these ligands have been restricted to interact with the lectin only in the close proximity of the principal $Ca^{2+}$ binding site.

To intentionally exploit secondary sites, new binding pockets or areas need to be defined (for example, based on X-ray structures) which have to be accessible for interactions with the ligand, or, alternatively, ligand discovery can originate from high-throughput small molecule screenings. Borrok *et al.* described a fluorescence-based competition assay to identify non-carbohydrate DC-SIGN inhibitors from commercially available ligand libraries (**figure 23**). Although the $IC_{50}$ values of these small molecules fell in the low μM range (as determined by the competition tests on a DC-SIGN ECD functionalized surface and by DC-SIGN mediated cell-adhesion assays), the studies lack any structural explanation for binding interactions or binding sites. It was proposed that in the absence of an obvious $Ca^{2+}$ coordinating motif, allosteric mechanisms could be accountable for the ligands' high potency.

During the extremely costly drug discovery campaigns, it is of crucial importance that early-stage failures - for example in biological target identification - are avoided. Therefore, a number of tools were developed to assess the druggability of therapeutically relevant proteins, and forecast risks of failure prior to significant investment. Druggability is a term used to indicate the presence of at least one suitable binding site which could interact with small-molecule drugs that induce an alteration in the protein function.

The evaluation of druggability consists of two main steps: binding pockets are identified on the protein surface (based on resolved crystal-structures), and these sites are then scored, according to their structural properties and accessability for ligand accommodation.

Aretz *et al.*[436] assessed the druggability of 22 human lectin CRDs, and found that *in silico*, almost all of these lectins had low druggability scores and would be deemed unattractive drug-discovery targets, based solely on the computational analysis. During these studies, the crystal structure of different CLRs was superimposed in a molecular operating environment and they displayed moderate overall similarity - although the C-type lectin-like domain is structurally conserved among them.

Then, binding sites were determined by DoGSite software, according to structural and physicochemical properties, and the findings were grouped, based on their position relative to the primary or secondary $Ca^{2+}$ ions. The authors suggested that the binding sites in the vicinity of the principal $Ca^{2+}$ are associated with carbohydrate recognition.

Curiously, upon a following experimental fragment screening by $^{19}$F-NMR, high hit rates were observed, contrary to the poor druggability values of computational predictions. For the NMR study, 281 fragments were selected from commercial sources, based on chemoinformatic tools and physicochemical properties (< 23 non-hydrogen atoms; 150 < MW < 300; number of rings > 0; PAINS exclusion – note analogy to the amine selection in Chapter I). These low-complexity small-molecules have better chances at binding to receptors than larger structures, and indeed, 13.5% hit rate was achieved (moderate druggability). Almost half of these fragments were confirmed by SPR-experiments. Additionally, three fragment hits were identified which constitute the previously described, low-micromolar non-carbohydrate DC-SIGN inhibitor **1.3**.[234] (**Figure 77**, **Figure 23**)

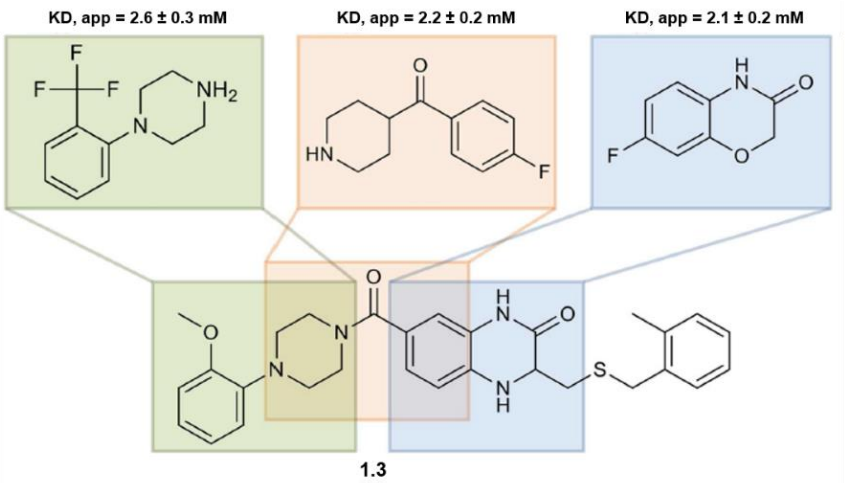



***Figure 77.*** *Small molecule DC-SIGN inhibitor fragments identified by $^{19}$F-NMR screening. These three fragments are present in a previously described non-carbohydrate DC-SIGN antagonist (**1.3**) validating the viability of the screening experiments. Modified from Aretz et al.*[436]

By experimentally confuting computational pocket prediction results, the studies shed light to the limitations of software-based druggability scoring. High numbers of false negative results during *in silico* binding site predictions can lead to discarded, but in reality druggable biological targets. This issue may arise from the software, as the algorithm was not optimized for carbohydrate ligands or metal-binding sites, and neither includes protein flexibility, conformational shifts or allosteric effects.

$^{19}$F-NMR experiments with fragment mixtures have several advantages, as they do not require any labelling step and the nuclear spin of fluorine is highly sensitive to changes in the chemical environment. Overall however, the analysis exhibits low sensitivity and has high protein consumption; furthermore fragments need stable isotope labelling. Yet, NMR-based fragment screenings are useful tools in small-molecule drug research, and can provide insight into carbohydrate-protein interactions as well (see Introduction).

Further experimental fragment-based screenings were performed with DC-SIGN, in order to determine correlations among different screening platforms, used to predict druggability. Other techniques included protein co-crystallization with organic solvents and array-based fragment screenings (similar setup to the earlier described microarray screenings with fluorescently-labelled CLRs).

Upon a follow-up study, 986 fragments were screened against DC-SIGN in SPR and reporter displacement assays (RDA) and hits were further validated by additional SPR tests and $^{1}$H-$^{15}$N HSQC NMR experiments, in order to identify specific binding sites on the CLR.[437] It had been estimated previously, that 1000 structurally diverse fragments cover a chemical space similar to 10 trillion drug-like molecules. The fragments used in the assays all comply with the rule-of three requirements for drug-fragments.[438]

After the initial screenings in SPR and RDA tests, 28 fragments were further tested by NMR and SPR. The most potent compound exhibited an $IC_{50}$ value of 0.3 mM (similar to the bisbenzylamide **1.21a**, developed in our group). Most importantly, $^{1}$H-$^{15}$N HSQC NMR experiments with $^{15}$N-labelled DC-SIGN CRD and computational binding pocket predictions identified 5 potential binding sites, in addition to the primary $Ca^{2+}$ site. (**Figure 78**)

From the 5 binding sites, 2 are located close to the $Ca^{2+}$ ion (site I close to the CRD center, near Phe313; site II "behind" the primary sugar binding pocket). Sites III and IV are located near the neck region, while pocket V is found between α-helices 1 and 2 on the opposite side of the CRD. (**Figure 18**) Sites II, III, IV and V might be able to open only for interactions with ligands, and are therefore not visible in CRD crystal structures.

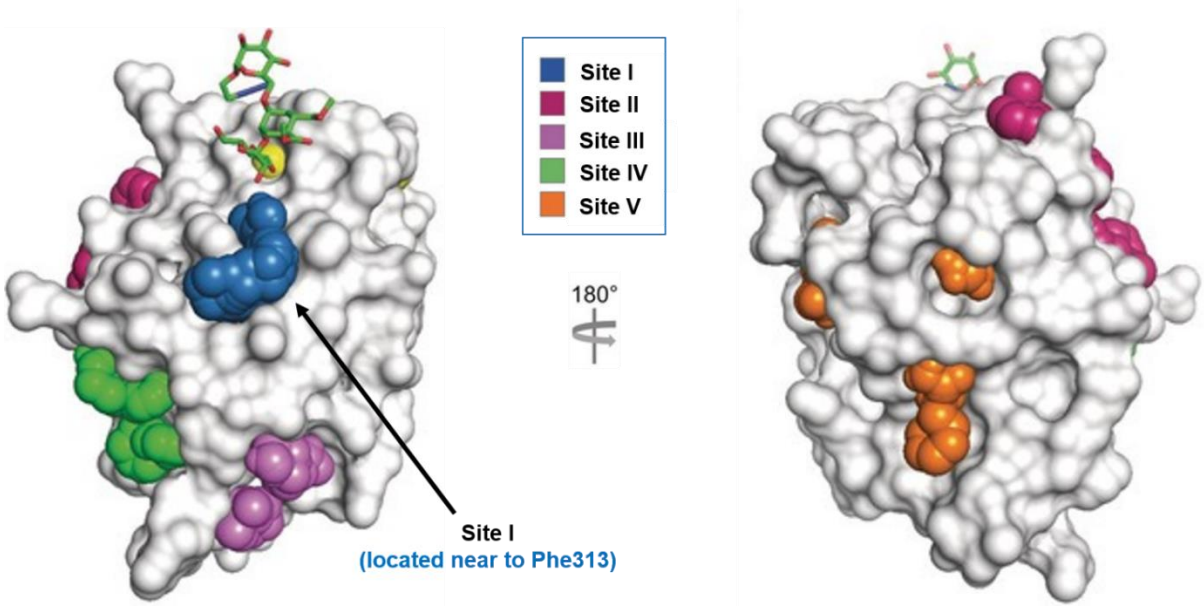



***Figure 78.*** *Predicted binding sites in the CRD of DC-SIGN. Site I is in the vicinity of Phe313 – a key residue in mannose-oligosaccharide binding. Site II is similarly close to the principal $Ca^{2+}$ binding site, while sites III and IV are near the neck region. Site V is defined by two α-helices on the opposite site of the domain. Modified from Aretz et al.*[437]

In particular, site I has moderate druggability results, and was earlier successfully targeted by glycomimetic ligands.[240] The phenylalanine-residue (Phe313) emerges as a good target for lipophilic interactions, and the reported ligand **1.15**, carrying two naphtyl substituents, is among the highest affinity monovalent ligands reported so far. Based on a computational model, these appendages reach towards Phe313 over the flat surface between the $Ca^{2+}$ site and the residue, and interact with the phenyl-ring from both sides (π- π interactions). (**Figure 79.A**) This region is also contacted by mannose-oligosaccharides, as a complementary binding site.[83, 85-86]

Most interestingly (from our point of view), the X-ray crystal structure of pseudo-mannobioside **1.18**, as well as the computational-model and NMR-data obtained for **1.21a** indicate that this area is unoccupied in the glycomimetic/DC-SIGN CRD complex. At the same time, the 2-OH group on the nonreducing mannose ring points directly towards this site, therefore it is logical to consider, whether we could reach this area and form favourable interactions by functionalizing the glycomimetic ligands in the C-2 position. (**Figure 79.B**) Establishing new contacts in the CRD can improve the – otherwise weak - affinity of monovalent molecules toward the CLR and open up new perspectives for selective DC-SIGN targeting.

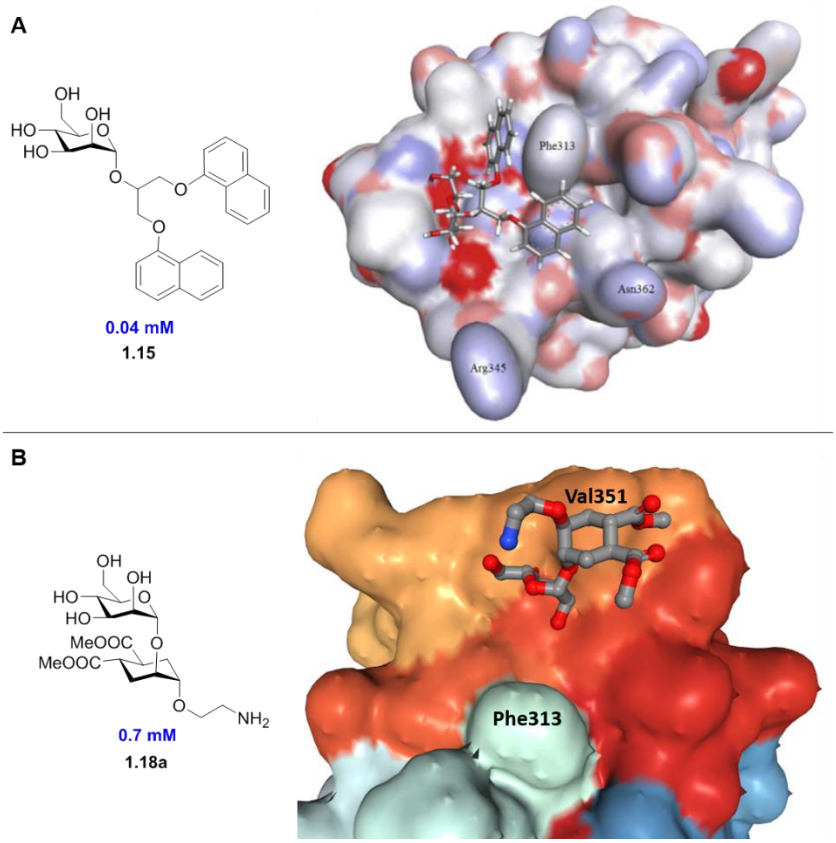



***Figure 79.*** *A. Glycomimetic ligand targeting Phe313 in the CRD of DC-SIGN. The phenylalanine residue can engage in lipophilic interactions, and the ligands decorated with two aromatic groups, such as naphtyl-moieties (**1.15**), show enhanced affinity towards the lectin ($IC_{50}$: 40 μM) Modified from Tomasic et al.*[240]

*B. Top view of the glycomimetic ligand **1.18** (developed by our group) in the CRD of DC-SIGN. The X-ray structure clearly shows that the region of Phe313 is left unoccupied because the ligand is oriented towards Val351. At the same time, the hydroxyl group in the second position of the mannose ring points towards the phenylalanine. Imported from the PDB using NGL viewer.*

## 3.2 Results and Discussion

### 3.2.1 Virtual fragment screening in the CRD of DC-SIGN

Monovalent glycomimetic ligands exhibit modest affinities towards DC-SIGN, typically in the micromolar range. Inhibitors developed in the Bernardi group reached $IC_{50}$ values of 200-300 μM by modifications of the parent pseudo-disaccharide **1.18** which shows similar affinity towards DC-SIGN as the natural disaccharide Manα1-2Man ($IC_{50}$: 700-900 μM). Although inhibitory potency was determined by different assay formats, **1.21a** and **1.22a**, two compounds modified in the carbomethoxy and mannose 6-OH groups of **1.18**, are clearly among the strongest monovalent DC-SIGN binders up to date, along with **1.14**[239] (Fleet group) and **1.15** (Anderluh group).[240] Unlike the two latter structures, however, our ligand template contains a linker moiety which can be readily used for multivalent presentation of these structures, and the dendrimeric constructs are known to reach low nanomolar levels of inhibition (see Introduction).

Yet, additional interactions in the CRD are continuously sought to increase the affinity of monovalent ligands. The region of Phe313 - which is known not to be exploited by the pseudo-disaccharide derived glycomimetics - emerges as an ideal target to contact.

To explore this area from the ligand point of view, the group of Dr. Sonsoles Martín-Santamaría performed a virtual screening in the complex of DC-SIGN/**1.18** (PDB code: 2XR5), with more than 1000 drug-like fragments (Maybridge library) according to the GLIDE protocol. The fragments were scored, based on their predicted binding energies and ranked.

The fragment-based approach identified several interesting structures that could favourably interact with the CRD in this region. Remarkably, all the top-scoring fragments (**929, 903, 261, 250**) present an ammonium-group and at least one aromatic moiety (heteroaromatics).

A well-defined pocket - lined with Ser360, Phe313 and Glu358 – that appears to be the main mediator for the accommodation of the highest affinity fragments was identified. The positively charged ammonium ion can engage in hydrogen-bonding with Ser360, cation-π interactions with Phe313 and ionic interactions with Glu358. (**Figure 80**)

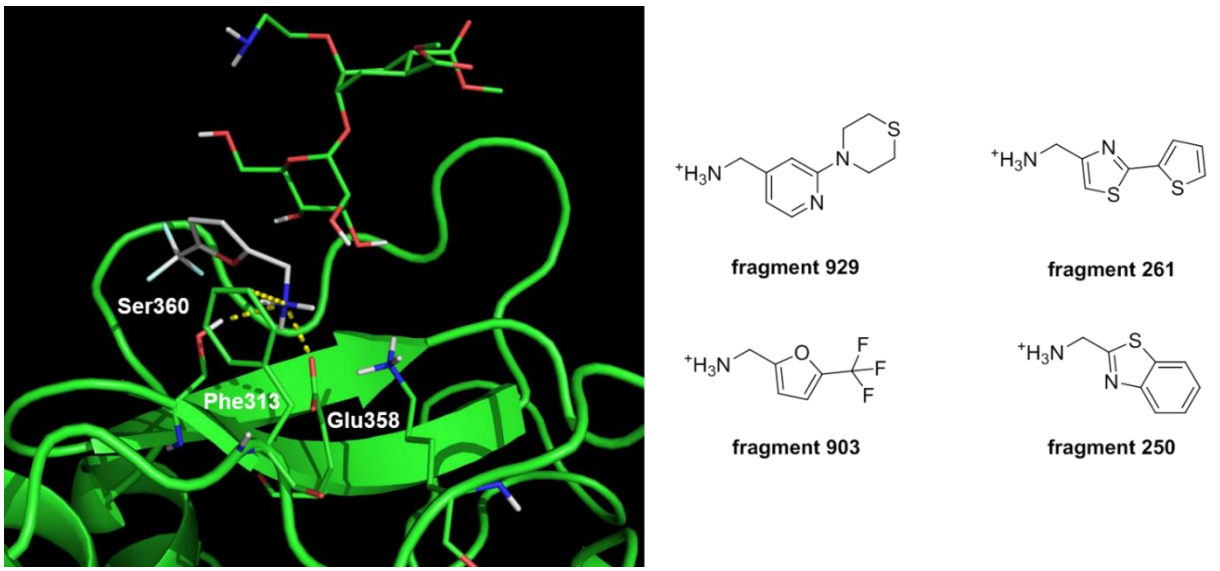



***Figure 80.*** *Virtual fragment screening in the CRD of DC-SIGN. The Martín-Santamaría group (CIB-CSIC, Madrid, Spain) screened more than a thousand drug-like fragments in the area of Phe313, using the complex of* ***1.18****/DC-SIGN CRD (PDB code: 2XR5). The top-ranking fragments featured two common structural elements: an ammonium group and aromatic moieties. The studies identified a novel ammonium binding pocket in the vicinity of Phe313.*

Between Phe313 and the $Ca^{2+}$ binding site (where the natural mannose ring is ligated), a flat area extends towards the ammonium binding pocket. This shallow surface, fringed with the hydrophobic groove of Phe313, may be, in principle, spanned by an appropriate modification in position 2 of the mannose ring (C-2). Therefore, we can rationally design glycomimetic ligands that exhibit structures similar to the above discussed fragments in the C-2 position and evaluate their contribution to binding. These conclusions could only be valid, if the glycomimetic core retains its binding mode in the CRD, and substituents do not interfere with its pose.

### 3.2.2 Glycomimetic structure functionalization in position C-2

As seen on the images above, the fragments are optimally oriented for conjugation to the C-2 position of mannose. The glycomimetic scaffold, however, has to be appropriately functionalized to link the fragment structures and the sugar ring. The most straightforward modification in this position for further derivatization is the installation of a N-atom in the place of the 2-OH group. Two major synthetic strategies were envisioned that make use of N-atom containing functional groups (**figure 81**):

- Amino groups can be coupled with amino acids and form a stable amide bond
- Azido-groups can undergo CuAAC reactions with alkynes

Azido-groups are ideal functionalities for both pathways, as they can mask the otherwise sensitive amino-group during the synthetic route, but can be rapidly converted into amines by reduction. **3.1** is therefore a common ligand for both synthetic routes which can be readily linked with fragment-like structures.

Some α- and β-amino acids were used for coupling through the first pathway by Luca Senaldi and Marco Minneci - primarily, in order to evaluate the role of the amino- and aromatic-groups. This route is synthetically demanding, as it requires protection and deprotection steps. Additionally, SPR inhibition tests demonstrated that coupling with differently-substituted α-amino acids leads to slightly decreased affinities towards DC-SIGN. The only exception is the glycine derivative **3.2** that exhibited slightly improved affinities ($IC_{50}$: 0.5 M) compared to the unmodified pseudo-disaccharide **1.18** ($IC_{50}$: 0.7 mM). This might suggest a minor contribution from the

ammonium group, but instead of further investigating this route, the attention was rather turned towards the second strategy.

Remarkably, the simplest triazole-functionalized derivative, **3.3** - scaffold **3.1** clicked with propargylamine (synthesized by Luca Senaldi) - showed an affinity increase by almost an order of magnitude ($IC_{50}$: 0.1 mM) compared to **1.18** ($IC_{50}$: 0.7 mM). This effect suggests an insufficient spacing between the glycomimetic scaffold and the ammonium group in amide-derivatives (C-α distance: 3.8 - 3.9 Å) which ultimately cannot reach the targeted binding site. This issue is overcome by functionalization through the larger triazole-group (5.0-5.1 Å),[439] which, additionally, is completely stable to enzymatic cleavage.

As a conclusion, we decided to pursue the click-reaction based derivatization of the azido-scaffold **3.1** with differently substituted alkynes.

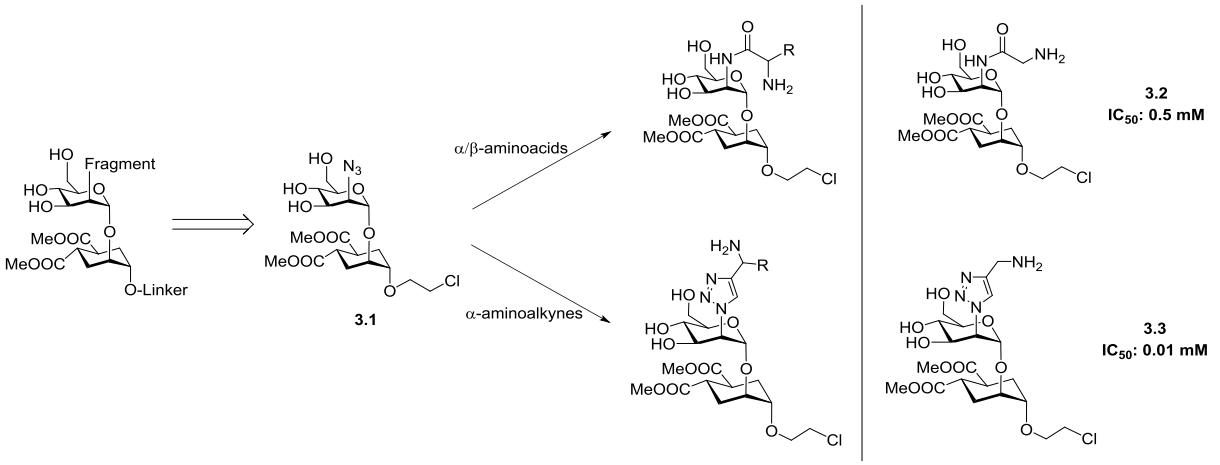



***Figure 81.*** *Synthetic strategies to functionalize the glycomimetic disaccharide scaffold in the C-2 position of the mannose residue. An azide-group can offer an ideal amine-masking group to form amide-linkages, but it can also be directly used in CuAAC reactions.*

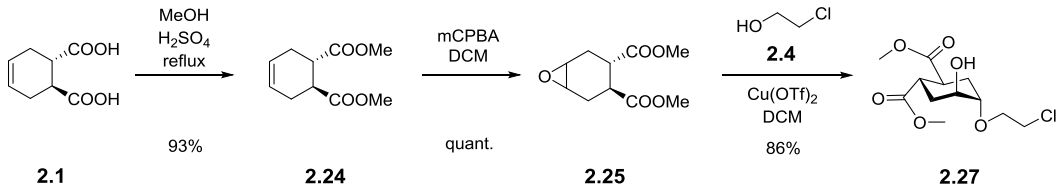
*The progenitor ligands synthesized through the two pathways demonstrate that the triazole linkage is superior to the amide bond, most likely owing to the larger distance between the terminal amine and the mannose ring. The simplest propargylamine derivative was able to enhance affinity almost by an order of magnitude ($IC_{50}$: 0.01 mM), compared to the unmodified glycomimetic dimethyl ester ($IC_{50}$: 0.9 mM)*

## 3.2.3 Synthetic route optimization of the azido-scaffold 3.1

Similarly to the library synthesis discussed in **2.1.1**, a general scaffold can be used to generate structurally diverse ligands, derivatized in further steps with alkynes. Analogously to the common mannosylated- and fucosylated-intermediates (**2.11** and **2.22**), **3.1** can be obtained through a convergent synthetic strategy by glycosylation reaction between a dimethyl ester sugar-mimic acceptor and a 2-azido mannosyl-donor.

### 3.2.3.1 Synthesis of the acceptor

Acceptor **2.27** is ideal for this purpose, as the linker terminates in a chlorine atom - instead of an azide (**2.26**) (otherwise the scaffold would contain two azido-groups which could both participate in the following CuAAC reaction). Later, after reacting the C-2 azido group **3.1** with alkynes, this chlorine can be easily exchanged for an azide using $NaN_3$ to allow "clicking" these ligands on multivalent constructs. (**Figure 72**) The high yields throughout the synthesis (starting from the enantiomerically pure diacid **2.1**) were reproduced on a larger scale.



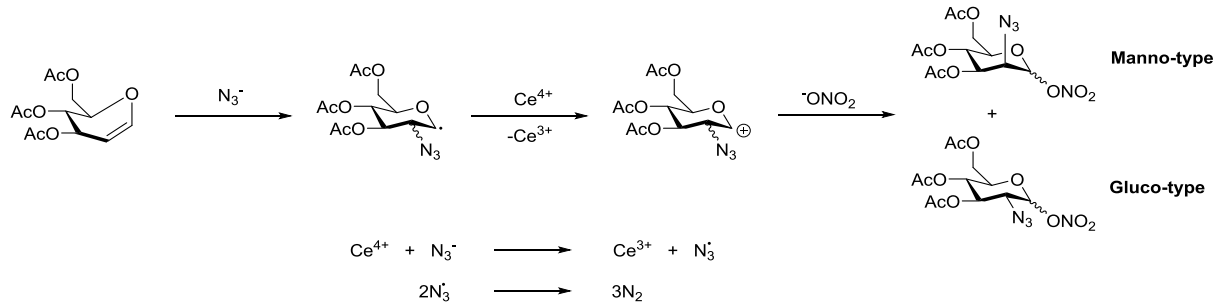

***Scheme 6.*** *Synthesis of the dimethyl ester glycomimetic acceptor, starting from the enantiomerically pure diacid* ***2.1***

### 3.2.3.2 Synthesis of the donor

Installation of the azide-functionality on the mannose-donor requires an entirely different approach, however. The synthesis of 2-azido-2-deoxy sugars is extensively studied in literature, as N-acetyl-sugars are important motifs in natural glycans (e.g bacterial polysaccharides), and the non-participating azido-group is an optimal amine masking group during the synthesis of oligosaccharides.

#### 3.2.3.2.1 Azidonitration

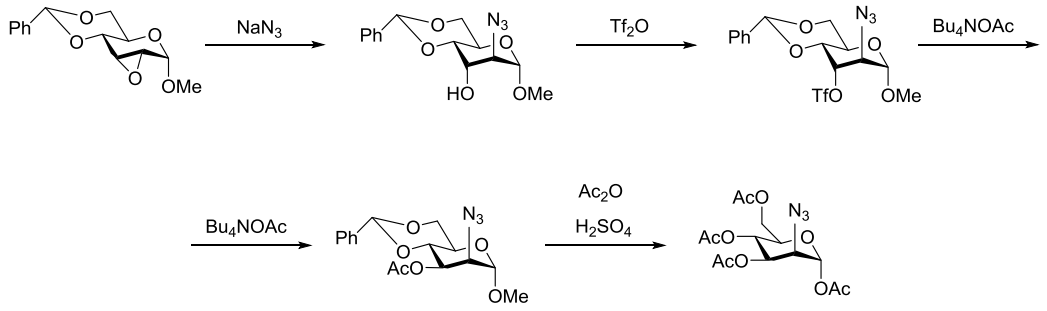
Several multi-step syntheses have been described to prepare 2-azido-2-deoxy derivatives, decorated with various protecting groups. A common approach, first reported by Lemieux and Ratcliffe,[440] relies on the azidonitration of D-glucals with $NaN_3$ and cerium ammonium nitrate (CAN). (**Scheme 7**) Although the reaction mechanism is not clear, it is probably initiated by the formation of an azide radical[441] which is then added to the double bond on C-2 (anti-Markovnikov) to form the anomeric radical. CAN promoted oxidation leads to the oxocarbenium cation that will readily react with the nitrate ion. As a consequence, stereoselectivity is poorly controlled, and depends on the protecting group present on the glucal, the configuration at C-4 (steric hindrance), the ratio of reactants and the temperature. Azidonitration of D-glucal yields therefore both the manno- and gluco-product at varying ratios, but predominantly the latter.[152a, 442] Further disadvantages are the low overall yields (protecting group installation, azidonitration, acetylation), toxic reagents and difficult purification steps. Kinfe and Paulsen both noted, that performing the reaction at low temperatures promotes a higher mannoside ratio.[443] Likewise, treatment of glucals with halogenoazides,[444] or their azido phenylselenylation results in mixtures of 2-azido-2-deoxy glucoside and mannoside.[445]




***Scheme 7.*** *Synthesis of 2-azido-2-deoxy mannose from D-glucals by azidonitration, described by Lemieux et al.*[440]

3.2.3.2.2 Epoxide opening with $NaN_3$

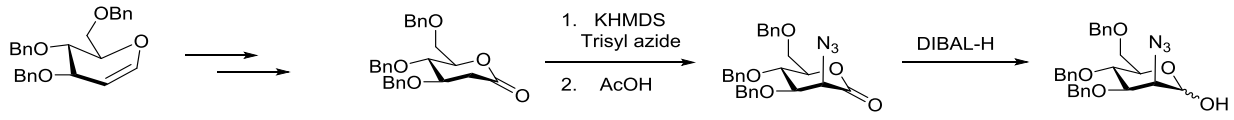
Alternatively, 1,3,4,6-tetra-*O*-acetyl-2-azido-2-deoxy-α-D-mannopyranose can be synthesized through 2,3-anhydro-4,6-*O*-benzylidene-(D-glucopyranoside), by opening the epoxide with $NaN_3$.[446] (**Scheme 8**) Again, the high number of steps disfavours this route to obtain the azido-mannosyl donor.




***Scheme 8.*** *Synthesis of 2-azido-2-deoxy mannose from 2,3-anhydro-4,6-O-benzylidene-(D-glucopyranoside) through stereoselective epoxide-opening with $NaN_3$*

3.2.3.2.3 Electrophilic azidation

In a highly stereoselective strategy, the electrophilic azidation established by Evans *et al.*[447] was applied to prepare 2-azido-2-deoxy-D-mannono-1,5-lactone.[448] (**Scheme 9**) Selective reduction of the lactone product with DIBAL-H afforded the aldose analogues and no C-2 epimerization was observed during the synthetic pathway. This approach circumvents the earlier reported issue of complete epimerization at the C-2 centre during the oxidation of 2-azido-3,4,6-tri-*O*-benzyl-2-deoxy-D-mannopyranose.[449] On the other hand, this synthetic sequence is complicated and the yields remain below satisfactory.

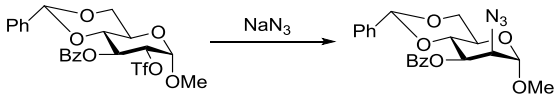



***Scheme 9.*** *Synthesis of 2-azido-2-deoxy mannose from the 1.5 lactone through electrophilic azidation*

3.2.3.2.4 Nucleophilic displacement

$S_N2$-type displacement reactions on C-2 were successfully employed to yield 2-azido manno-type products, in spite of the well-known difficulties of nucleophilic substitutions on glycosides.[450] (**Scheme 10**) 2-Triflate

activated, appropriately protected glucosides can be converted into the 2-azido manno-analogues by the linear azide nucleophile, with complete stereoselectivity. The application of this method is limited by the long synthetic pathways,[451] even though careful workup can improve the generally low yields.

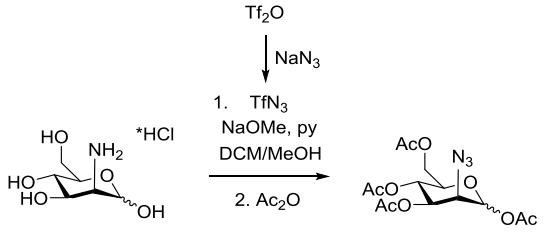



***Scheme 10.*** *Synthesis of 2-azido-2-deoxy mannose through an $S_N2$-type trifate-displacement from a gluco-type starting material*

3.2.3.2.5 Diazo transfer

The preparative sequence was considerably simplified by Vasella *et al.* who adapted the direct transformation of amino-groups into azides[452] to unprotected 2-amino-2-deoxy aldoses.[453] (**Scheme 11**) The reaction, otherwise known as the diazo transfer, proceeds through mild conditions with triflic azide and sodium methoxide, with retention of the configuration and reasonable yield for 1,3,4,6-tetra-*O*-acetyl-2-azido-2-deoxy-D-mannopyranose (65% over two steps, according to lit.). On the downside, the potentially explosive $TfN_3$ is a small molecular weight organic azide reagent that requires extra caution and needs to be handled in solution.

The original procedure starts from commercially available D-mannosamine hydrochloride that is treated with NaOMe in MeOH to liberate the amine *in situ*, then pyridine is added, and the free amine is subjected to diazo transfer by $TfN_3$. Acetylation of the crude yields the protected azido-mannose derivative.



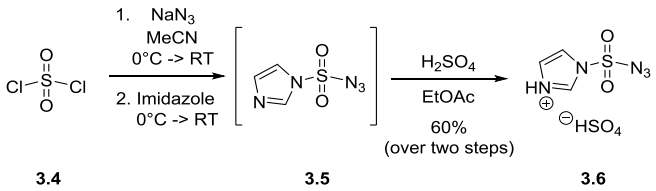

***Scheme 11.*** *Diazo transfer-based synthesis of 2-azido-2-deoxy mannose from mannosamine hydrochloride with triflic azide, reported by Vasella et al.[453] The reaction is completely stereoselective and affords the product in acceptable yield.*

Yields were slightly improved by Alper *et al.* who noted, that $Cu^{2+}$, $Ni^{2+}$, and $Zn^{2+}$ divalent cations catalyze the diazo transfer reaction.[454] The authors used $K_2CO_3$ as the base to deprotonate the starting hydrochloride salt, and obtained 72% 2-azido mannoside over two steps.

The hazards associated with triflic azide and the relatively high cost of the starting triflic anhydride prevent the use of $TfN_3$ on larger scale or in industrial environment. Switching the solvent from dichloromethane to a more homogeneous, three-component system (methanol/water/toluene) decreased risks by avoiding the formation of azido-chloromethane and diazidomethane, but the yield for the manno-type azide dropped significantly compared to the original procedure.[455] Therefore, several studies focused on its replacement with more stable diazo-donors, improving both safety and the reagent's shelf-life.

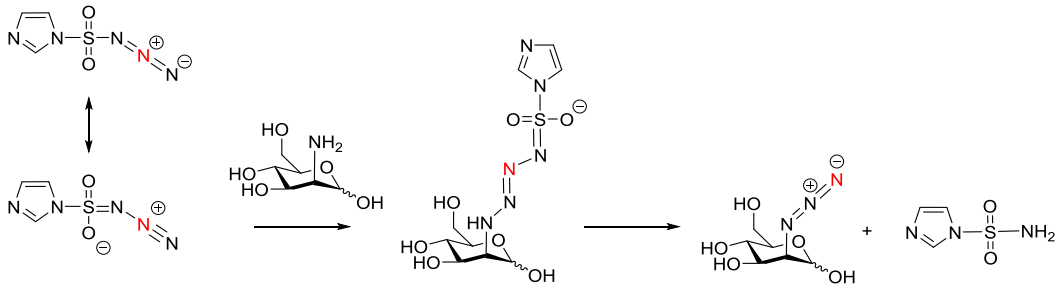
In order to replace the trifluoromethanesulfonyl moiety with another electron withdrawing group, which, at the same time, improves safety and stability, Goddard-Borger and Stick (inspired by the works of Hanessian and Vatèle)[456] developed imidazole-1-sulfonyl azide hydrochloride,[457] as an ideal alternative. Imidazole-1-sulfonyl azide was prepared by treating sulfonyl chloride with $NaN_3$ to yield chlorosulfonyl azide, followed by the addition of imidazole in acetonitrile.

The hydrochloride salt could be isolated as a crystalline and worked as efficiently in diazo transfer reactions with glycosamine hydrochloride as $TfN_3$. The solid was later found to be impact sensitive, however, and concentration of the mother liquors could lead to accumulation of explosive sulfonyl diazide or hydrazoic acid.[457]
The use of other salt forms, in particular $HSO_4$ was proposed. [458] Imidazole-1-sulfonyl azide hydrogen sulfate proved to be a similarly inexpensive, but safer-to-handle diazotransfer reagent, owing to its superior tolerance to impact, friction, temperature and electrostatic discharges. Furthermore, an improved procedure to obtain this salt form has been reported recently, with increased yield and an overall safer handling.[459] (**Scheme 12**)




***Scheme 12.*** *Preparation of the diazotransfer reagent imidazole-1-sulfonyl azide hydrogen sulfate. This reagent is more stable and safer than the previously reported triflic azide, and imidazole-1-sulfonyl azide hydrogenchloride salt*

Based on the satisfactory yields between $TfN_3$ and $ManNH_2$*HCl (reported by Vasella and Alper), and further encouraged by the safer procedure of Goddard-Borger (described for $GlcNH_2$*HCl), we decided to adapt the imidazole-1-sulfonyl azide hydrogen sulfate salt-mediated diazo transfer reaction to $ManNH_2$*HCl.
The mechanism of the diazo transfer reaction has been investigated in $^{15}N$ NMR studies with labeled amine and azide species.[460] (**Scheme 13**) According to these studies, the reaction proceeds through an attack on the amine by the terminal azide nitrogen, then the two terminal nitrogen atoms are transferred from the azide to the amine, which, in turn retains its original nitrogen. The site of attack is not influenced by the presence of the metal catalyst.



***Scheme 13.*** *Plausible mechanism of the diazo transfer reaction, supported by isotope labelling studies (red nitrogen atoms)*

As the starting sugar is relatively costly, the reaction sequence required optimization before scale-up. The first trials, carried out by Luca Senaldi, afforded 40-50% acetylated 2-azido mannose (2 : 1 = α : β anomeric ratio) in two steps. In a typical setup, 2-azido-2-deoxy mannosamine hydrochloride (1 equiv.) was dissolved in dry MeOH, $K_2CO_3$ (2.7 equiv.) and $CuSO_4$ (0.01 equiv.) was added to yield a suspension, which was then subjected to imidazole-1-sulfonyl azide hydrogen sulfate (1.2 equiv.). After completion, the crude was dissolved in pyridine and treated with acetic anhydride. It was demonstrated that benzoyl protecting groups lead to even lower yields. Additionally, the formation of the glucose-analogue was noted, but this minor side-product was attributed to glucosamine contamination in the starting material.
During my first attempts, even these moderate yields could not be reproduced - instead varying results were achieved, lacking any obvious explanation (yields between 15 and 70%). At first, the inhomogeneous suspension

and partial reactant dissolution was suspected to stand behind the random outcomes, therefore we switched to a water/MeOH solvent system, without any apparent improvement.

After several small-scale trials under slightly different conditions (concentration, reaction time), and using a new batch of mannosamine hydrochloride free of contamination, the gluco-type analogue was still observed in the crude of the diazo transfer step. NMR spectra of the manno-type and gluco-type products were in agreement with those reported by Vasella and coworkers:[453] $J_{1\text{-}2}$ (α azido-mannoside; eq-eq) = 1.9 Hz; $J_{1\text{-}2}$ (β azido-mannoside; ax-eq) = 1.4 Hz; $J_{1\text{-}2}$ (α azido-glucoside; eq-ax) = 3.7 Hz and $J_{1\text{-}2}$ (β azido-glucoside; ax-ax) = 8.6 Hz. (**Figure 82**)

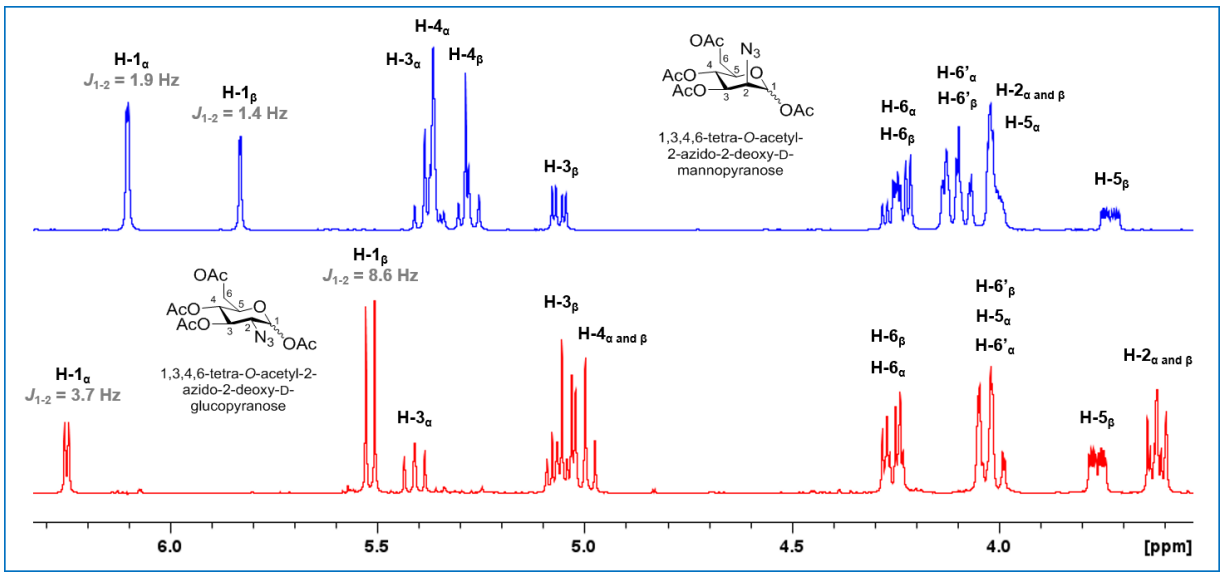



***Figure 82.*** *$^1$H-NMR spectra of the manno- and gluco-type diazotransfer products. The anomeric protons have much larger coupling constants in the α and β gluco-derivatives (ax-eq; ax-ax) than in the respective mannose products (eq-ax; eq-eq).*

More importantly, the gluco side-product also formed at variable proportions, sometimes up to 40% compared to the theoretical yield, other times being only a minor contaminant. This indicated that the side reaction was also related to the conditions, but the experimentally confirmed mechanism of the diazo transfer step did not explain configurational changes on C-2.

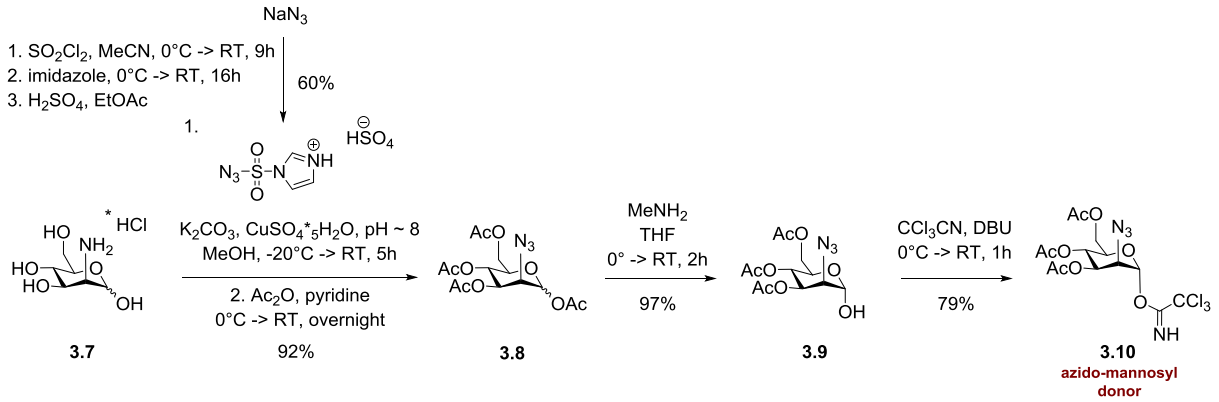
The only logical explanation originates from the addition of the base ($K_2CO_3$, pKa of the conjugate acid: 10.25) to liberate the free amine. Under basic conditions, a proton removal from the C-2 position leads to a planar enolate anion and loss of configuration at this centre. Upon re-protonation, the new stereo centre (α position to the sugar carbonyl group) can become either (*R*) or (*S*) configured and afford the *manno*- and *gluco*-type isomer. In other words, base-catalysis can lead to epimerization at the C-2 centre, and due to the less hindered equatorial position, gives rise to the formation of 2-azido-2-deoxy-D-glucopyranose.

To prevent the base-promoted epimerization, we implemented tighter pH control, slower addition of the base ($K_2CO_3$) and lower temperatures. (**Scheme 14**) In the first step, the free mannosamine is liberated from the hydrochloric salt by 1.05 equiv. base at -20°C. $CuSO_4$ is added, then, after 5 minutes, another 1.05 equiv. $K_2CO_3$, followed immediately by the addition of the imidazole-1-sulfonyl azide hydrogen sulfate salt. Due to risks of hydrazoic acid formation, the pH is controlled throughout the reaction (pH ~ 8).

This method afforded reproducibly high yields (~ 90%), and no formation of the gluco-type product was observed. The optimized conditions allowed us to perform the reaction on a gram scale for the preparation of a larger stock of the azido scaffold **3.1**.

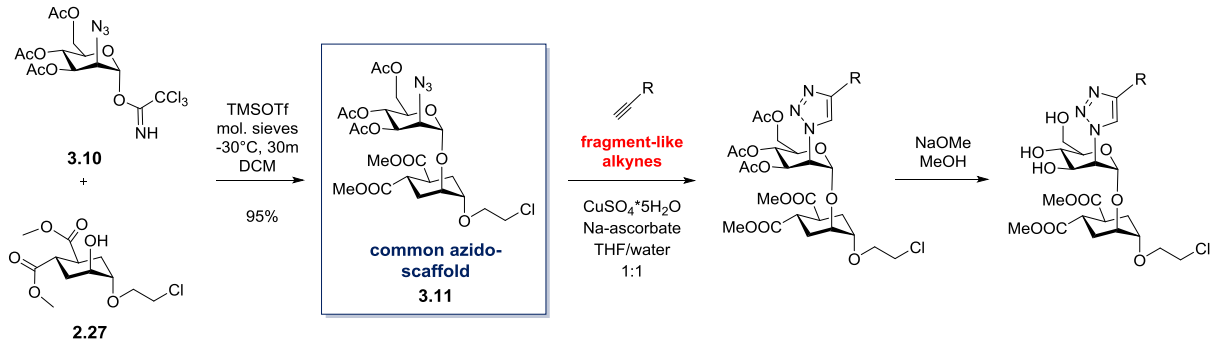
### 3.2.3.3 Synthesis of the common azido scaffold

The next steps in the synthetic sequence are the anomeric deprotection with $MeNH_2$, and the TCA-leaving group installation with trichloroacetonitrile and DBU. Finally, the azido-mannosyl donor could be obtained in 69% yield over four steps from mannosamine. (**Scheme 14**)




***Scheme 14.*** *Synthesis of the azido-mannosyl donor* ***3.10*** *from mannosamine hydrochloride. The diazo transfer reaction is performed with imidazole-1-sulfonyl azide and the product is directly protected without isolation. Anomeric deprotection and leaving group installation lead to* ***3.10*** *in good yields (70% over three steps).*

The glycosylation between the acceptor **2.27** and the donor **3.10**, initiated by a catalytic amount of TMSOTf afforded the desired 2-azido-glycomimetic scaffold **3.1** that could readily react with alkynes in CuAAC click reactions. The final step in the sequence is the removal of acetyl-groups under Zemplén-conditions. (**Scheme 15**)




***Scheme 15.*** *Glycosylation and click reaction with fragment-like alkyne structures. Reaction of the donor* ***3.10*** *and acceptor* ***2.27*** *yield the common scaffold* ***3.11****. CuAAC reactions with alkynes, followed by the final Zemplén-deprotection step afford C-2 modified glycomimetic ligands that are used to exploit the ammonium binding site near Phe313 in the CRD of DC-SIGN.*

### 3.2.3.4 Previous results witb amino acid and propargylic derivatives

A summary of previously obtained results with different fragment-like structures is shown in **figure 83**. The $IC_{50}$ values of all C-2 functionalized ligands were determined in SPR inhibition assays on a mannosylated-BSA chip in the laboratory of Prof. Franck Fieschi (IBS, Grenoble, France).

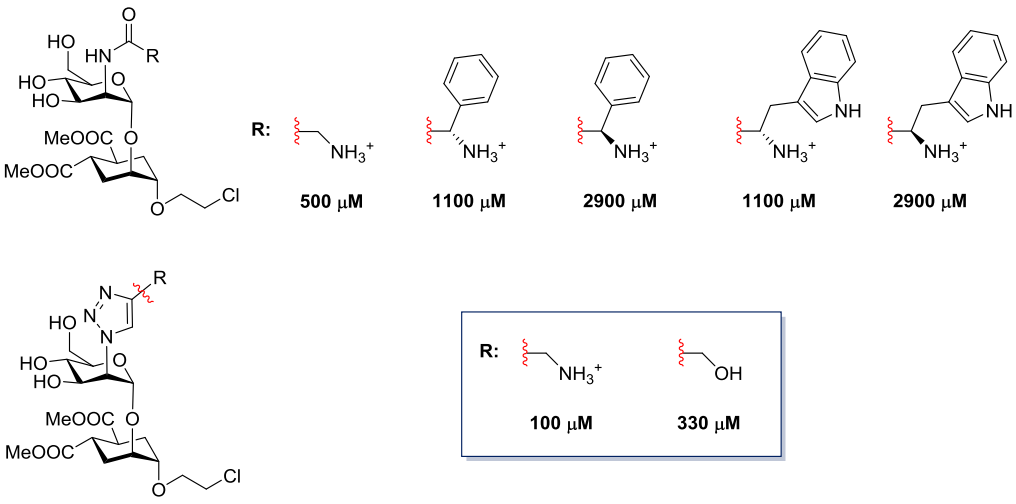



***Figure 83.*** *Previous SPR inhibition test results with C-2 functionalized glycomimetic ligands. From the amide-coupled series only the glycine-derivative showed improved affinity compared to the unmodified pseudo-disaccharide* ***1.18****. Clicking propargylamine on the common scaffold* ***3.11****, however, increased the affinity by a factor of seven. Interestingly, also the propargyl alcohol derivative exhibited a 2-fold increase in inhibitory potency, suggesting a H-bonding contribution to the interactions with the CRD.*

The affinity enhancement observed in the propargylamine derivative **3.3** ($IC_{50}$: 100 μM) clearly indicated a beneficial effect of the ammonium ion (SPR experiments are carried out at pH ~ 8), but even the propargyl alcohol derivative exhibited superior affinity ($IC_{50}$: 330 μM), compared to the parent pseudo-disaccharide **1.18** ($IC_{50}$: 700 μM). Although not as potent as the ammonium ion, this information suggested that the OH group could engage in H-bonds around the Phe313 area, possibly with the same Ser360 residue as predicted for the virtual fragments. At the same time, aromatic and heteroaromatic substituents on the amide derivatives could not increase the affinity, and were therefore suspected not to be optimally positioned to establish interactions with the phenylalanine residue.

With these data in hand, we set off to explore the effects of various substituents on the binding interactions with DC-SIGN. The encouraging results of propargyl-derivatives stimulated the evaluation of a set of propargylamine and propargyl alcohol derivatives.

If the ammonium ion on **3.3** is indeed able to reach the ammonium binding pocket, hydrophobic groups around this ion could protect and stabilize the charged cation, by preventing disruption of the interaction network by water molecules. At the same time, the lipophilic substituents may interact with other lipophilic residues in this region in CH-π type interactions,[461] for example, with Phe313.

### 3.2.4 *N*-methylation on the ammonium group

We have compiled a series of methylated propargylamines to click on scaffold **3.1**, and compare the influence of methyl groups to the free ammonium group. (**Figure 84**) *N*-methylpropargylamine and 3-dimethylamino-1-propyne were commercially available, while *N,N,N*-trimethylprop-2-yn-1-aminium bromide could be synthesized by the reaction of propargyl bromide and trimethylamine.[462]

Comparison of **3.3** and the mono-, di- and trimethyl-ammonium derivatives showed a gradual loss of affinity towards DC-SIGN. More specifically: the installation of even one methyl group significantly decreased the affinity, which further declined (although less drastically) by the addition of other Me-groups.

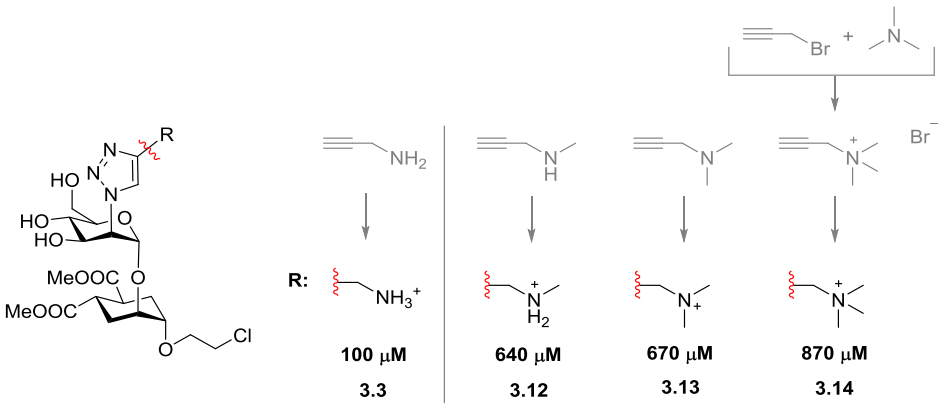



***Figure 84.*** *Affinity of a differently methylated series of propargylamine-derivatives. The $IC_{50}$ values are gradually increasing with the addition of methyl groups. Already the first methyl group reduced the affinity of these ligands, suggesting a steric effect hindering the proper accommodatin of the ammonium group.*

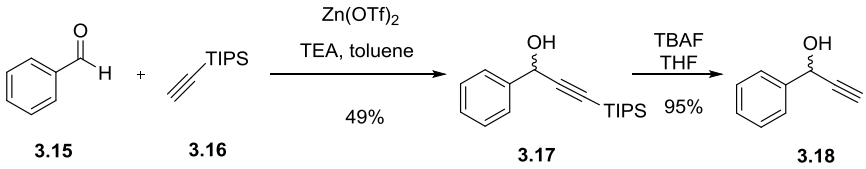
We deduced two main reasons that could explain this observation. On one hand, replacement of hydrogen atoms with methyl groups on the nitrogen would preclude engagement in H-bonding as a donor, however, the fact that already the mono-methyl derivative loses potency suggests that there is rather a steric effect in operation. If the added methyl groups force the ammonium group in a non-optimal position in the pocket, binding interactions, and thus the overall affinity would be weakened, compared to the unmodified propargyl amine derivative. The ammonium ion's role in the H-bonding network is not clear, but the importance of this bond is underlined by the alcohol derivative that exhibits improved inhibitory potency compared to the unmodified glycomimetic.

## 3.2.5 Installation of aromatic groups

In addition to the ammonium ion, the virtual fragment screening also suggested that aromatic groups could possibly interact with Phe313 by π-π stacking. Although in the amino acid series aryl substituents did not show any positive contribution to binding affinities, this may be the result of the shorter amide linkage.
Aromatic functionalization of the propargyl derivatives could position the aryl moieties closer to the phenylalanine ring to establish contacts. The simplest structure, comprising an amino group, the aromatic ring and the triple bond for the subsequent CuAAC reaction is 1-phenylprop-2-yn-1-amine, a molecule with a stereo centre. This alkyne, however, is not commercially available, and so, we first decided to perform the click reaction with the synthetically less challenging alcohol derivative to test its inhibitory potency. Although we had no model available whether the ammonium and alcohol derivatives share similar binding modes, the two alcohol diastereoisomers could still provide valuable information about the effects of aryl rings near the binding site.

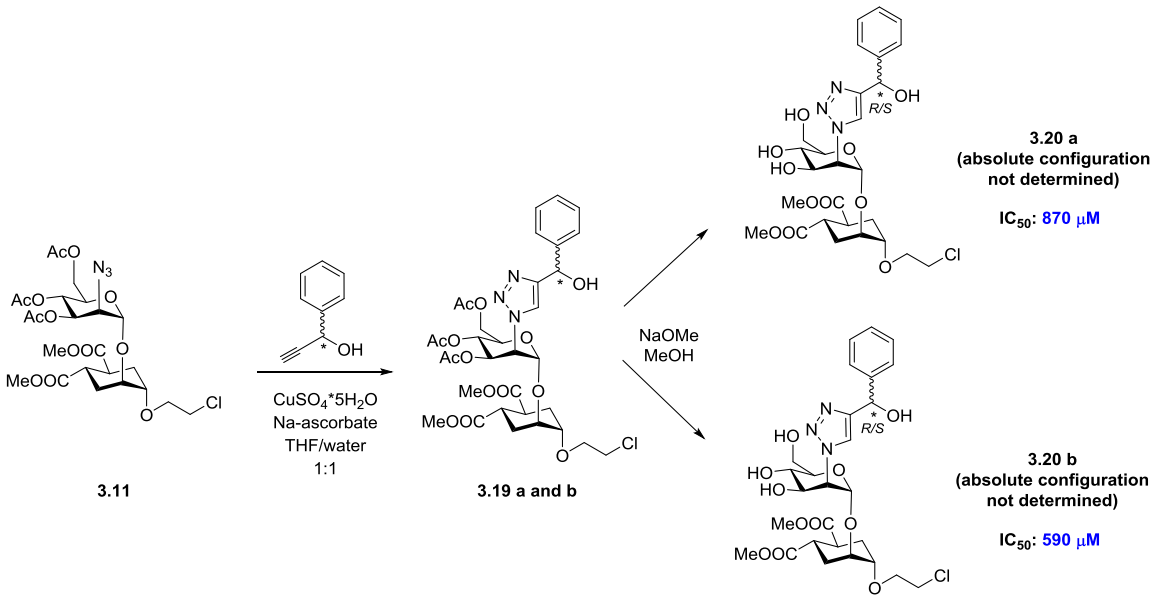
### 3.2.5.1 Phenyl-propargyl alcohol derivatives

Phenyl-propargyl alcohols can be synthesized through stereoselective methods, however, as we lacked indications on which could be the more potent configuration, both enantiomers were used for the CuAAC reaction. The racemic phenyl-propargyl alcohol could be synthesized in moderate yields from benzaldehyde and (triisopropylsilyl/TIPS)-acetylene in the presence of $Zn(OTf)_2$ and trimethylamine,[463] but the compound is also available at commercial sources. (**Scheme 16**)




***Scheme 16.*** *Synthesis of 1-phenyl-2-propyn-1-ol with zinc triflate, in the presence of trimethylamine with modest yields*

After coupling on scaffold **3.1**, the two diastereomers were separated following the Zemplén-deprotection step ($R_f$(**3.20.a**): 0.26; $R_f$(**3.20.b**): 0.22 in DCM : MeOH = 94 : 6). (**Scheme 17**)




***Scheme 17.*** *Synthesis of phenyl-propargyl alcohol derivatives. The two diastereomers,* ***3.20a*** *and* ***3.20b*** *were separated after the deprotection step*

The $IC_{50}$ values of **3.20a** shows a two-fold, **3.20b** an almost three-fold increase compared to the unmodified propargyl alcohol derivative, lacking the phenyl-group – another clear indication that aromatic-substituents did not improve binding affinity, as already observed for the amino acid derivatives. (**Figure 83**) The inhibitory potency of the two diastereomers did not differ significantly from each other or exceed that of the alcohol, therefore their absolute configuration was not determined.[464]

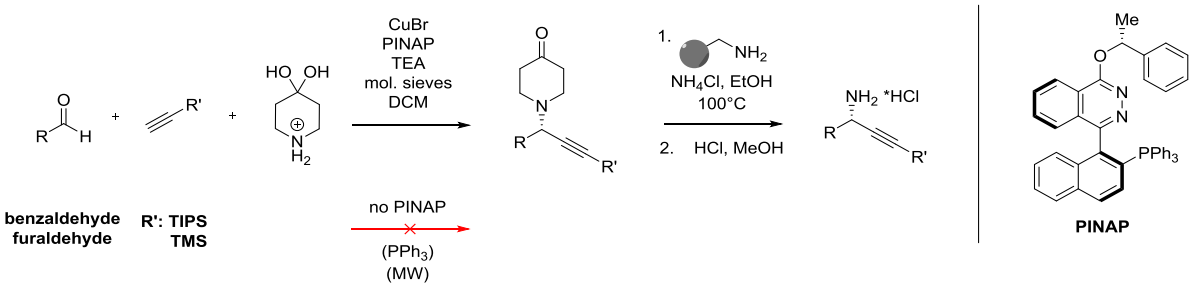
### 3.2.5.2 Phenyl-propargylamine derivatives

The preparation of 1-phenylpropargylamines is not straightforward, and only a few examples were reported that do not involve expensive transition metal catalysis, complex ligands or long synthetic sequences. Many of these methods were developed as stereoselective tools, but for similar reasons as in the case of the alcohol, an ideal synthesis would serve us best if it yielded both enantiomers that we would separate at a later stage of the synthetic route.

During the first trials, we attempted to simplify the stereoselective, Cu(I)-catalyzed method of Aschwanden *et al.*,[465] by omitting the chiral P,N-ligand PINAP. (**Scheme 18**) The originally reported three component reaction can occur between aromatic aldehydes, appropriately protected alkynes and 4-piperidone hydrochloride, in the presence of Cu(I)Br and PINAP. Later, the amine protecting group can be removed by ammonia-saturated

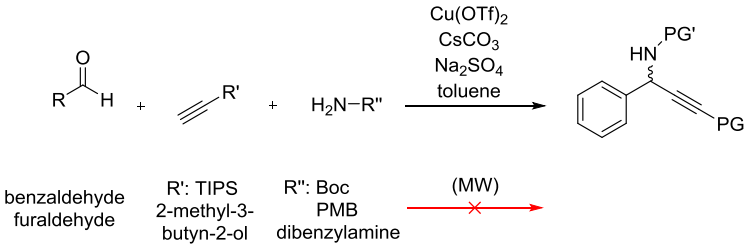
EtOH, or, alternatively, on a polymer-supported scavenger amine. This direct preparation of modified propargylamines was not described for phenyl-substituents.

We performed the reaction in the absence of PINAP on furfural and benzaldehyde, but the product was not observed in the crude. We suspected, that in addition to ensuring stereoselectivity, the complex ligand plays a role in the catalytic effect or helps the solubility of the catalyst, therefore a trial was conducted with added $PPh_3$ as a P-ligand, but this did not lead to any positive result.



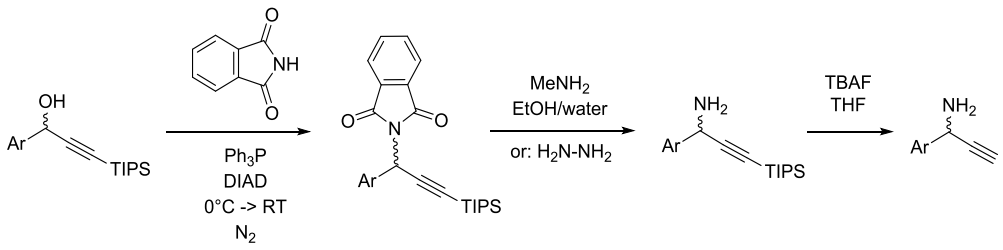

***Scheme 18.*** *Synthesis of phenyl-propargylamine derivatives with Cu(I)-catalysis and a P,N-ligand (PINAP) through the synthetic route reported by Aschwanden et al.*[465]

Another three-component reaction between an aryl-aldehyde, protected alkyne and protected amine was described by Meyet *et al.*[466] (**Scheme 19**) Alkynylation of the imine can be catalyzed by copper(II), most efficiently by $Cu(OTf)_2$, and the reaction was demonstrated on a number of various aldehyde, alkyne and amine substrates. We tried to perform the coupling with several different reagents - including benzaldehyde and furfural; TIPS-acetylene and 2-methyl-3-butyn-2-ol; and *tert*-butyl carbamate, *p*-methyoxybenzyl (PMB)-amine and dibenzylamine – but apart from observing traces of the desired product, our efforts were unsuccessful.



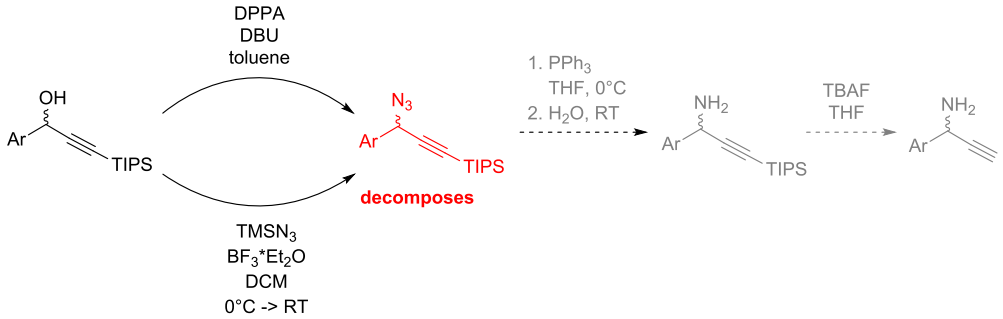

***Scheme 19.*** *Three component coupling of aryl-aldehydes, protected amines and protected acetylene with Cu(II)-catalysis. The reaction afforded only traces of products on our substrates*

A further alternative is based on the Mitsonobu-reaction of the protected propargyl alcohol with phthalimide. (**Scheme 20**) The free amine was obtained by treatment with hydrazine-hydrate, or methylamine, but the poor yields throughout the reaction sequence and difficult purification steps deterred us from this strategy (the isolated phenyl-propargylamine was insufficient for the following click reaction; furyl-propargylamine was not isolated in its final form).



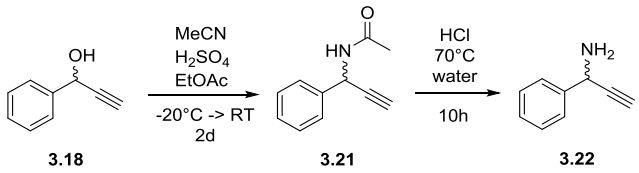

***Scheme 20.*** *Synthesis of phenyl-propargylamine derivatives through the Mitsonobu-reaction*

Installation of an azide-group at the chiral carbon by DPPA and DBU, or $TMSN_3$ and $BF_3$*$Et_2O$ was successful, but after isolation of the intermediate and during the following reduction step - Staudinger reaction under mild conditions to avoid the saturation of the alkyne during hydrogenation – the compound underwent thermal decomposition, likely due to the simultaneous presence of an azido-group and a triple bond. (**Scheme 21**)




***Scheme 21.*** *Phenyl-propargylamine synthetic route by installation of an azido-group with DPPA or $TMSN_3$. The red intermediate slowly decomposed, likely due to the simultaneous presence of an azido-group and an alkyne.*

Probably the simplest (and least expensive) preparation of aryl-propargylamines can be performed in two steps: the Ritter-reaction of aryl-propargyl alcohol with concentrated $H_2SO_4$ and MeCN yields the acetamide **3.22**, and then hydrolysis in the presence of HCl at elevated temperatures affords the free amine. (**Scheme 22**) This reaction offers a straightforward option to synthesize 1-phenyl-propargylamine, and the procedure is being scaled up to obtain sufficient amount of the alkyne for the following click reaction.




***Scheme 22.*** *Synthesis of phenyl-propargylamine through the Ritter-reaction and acetamide hydrolysis*

Affinity assay of the glycomimetic ligands carrying both the ammonium-group and an aromatic ring will offer a decisive evidence on the role of aromatic substituents.

A phenyl- or furyl-substituent at the second carbon of propargyl-amine is not the only option to include a nitrogen atom and an aromatic-system in the vicinity of the predicted binding pocket. As the ammonium-ion's role as a H-bond donor or acceptor has not been confirmed yet, a N-atom on a pyridine-ring could engage in these

interactions in the latter role. At the pH of the SPR assays, the pyridine-moiety is supposedly not protonated (pKa of the conjugate acid ~ 5.2).

### 3.2.5.3 Picolylamine-derivatives

To our surprise, the 2-picolylamine derivative **3.23** exhibited almost as good affinity towards DC-SIGN, as the ammonium ion containing **3.3** (propargylamine derivative). (**Figure 85**) This observation may suggest that the aromatic system is able to positively contribute to the binding strength - for example through T-type π-π interactions, even in the absence of the cationic ammonium group. The 3-picolylamine derivative **3.24** was also synthesized, but its inhibitory potency has not been tested yet. Further experiments are underway to confirm the effects of aromatic interactions, including the synthesis of the 1-aryl-propargylamine derivative stereoisomers.

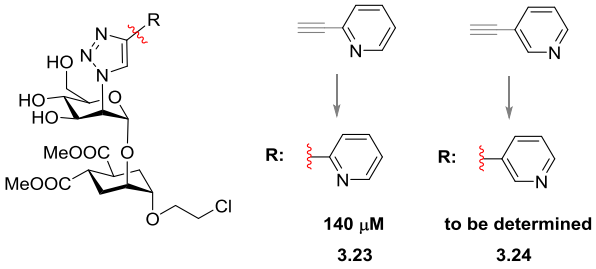



***Figure 85.*** *Picolylamine derivatives and SPR inhibition test results with* ***3.23****. To our surprise,* ***3.23*** *exhibited almost as good affinity towards DC-SIGN as* ***3.3****, even though it lacks the positively charged nitrogen. In spite of the previous results with aromatic amino acid derivatives and phenyl-propargyl alcohol derivatives (****3.20a*** *and* ***3.20b****), π-π-type contributions may be almost as significant as the ammonium group.*

The above discussed efforts focused on the modification of the glycomimetic dimethyl ester **1.18** to target a recently identified ammonium binding pocket. The appended structures resemble the top-scoring fragments of a virtual screening campaign and could exploit new interactions in this area. The ammonium binding site – fringed with Ser360, Phe313 and Glu358 – is left unoccupied in the crystal complex of the parent pseudo-disaccharide **1.18** that contacts the region of Val351 instead, in the opposite direction from the $Ca^{2+}$ ion. The affinity improvement of triazole-functionalized ligands, such as **3.3**, demonstrated, however, that the ammonium pocket could be reached by modifications in the C-2 position of the natural mannose residue. The simplest propargyl amine derivative **3.3** increased the affinity towards the CLR by almost an order of magnitude.

## 3.2.6 A combined strategy of C-2 modifications and bisbenzylamide substituents

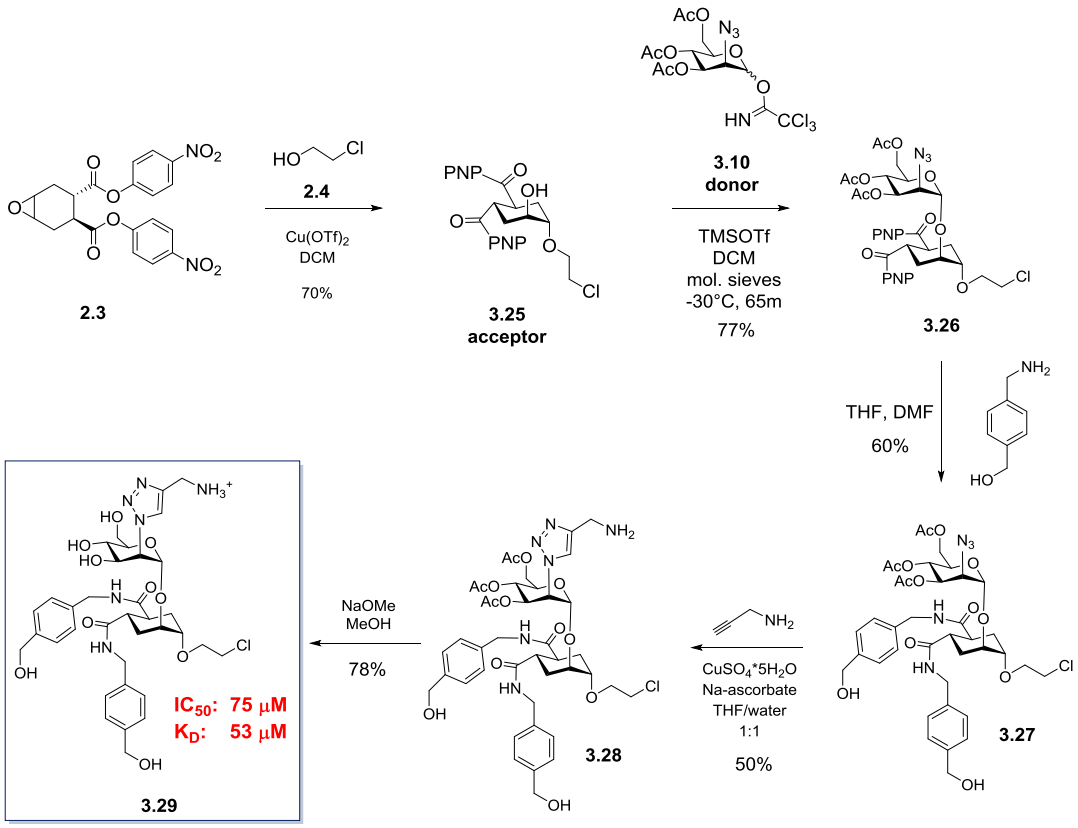
### 3.2.6.1 Synthesis of 3.29

We knew from earlier results that replacing the two methyl ester groups on the cyclohexane ring by lipophilic substituents improved the affinity towards DC-SIGN by a factor of three. Based on a computational model supported by STD-NMR data, the docked pose of the *p*-hydroxymethylenebenzylamide derivative (Man030, **1.21a**) indicated van der Waals interactions between the aromatic moieties and the outer regions of the CRD surface. The NMR observations could qualitatively account for the computationally proposed binding mode, but the actual ligand binding pose and ligand-CRD complex structure has never been resolved by X-ray crystallography.

We hypothesized that lipophilic contacts near Val351 and the ammonium-ion interactions around Phe313 could be simultaneously exploited to increase affinity towards the lectin. To demonstrate this, we envisioned ligand

**3.29** that comprises the mannose C-2 functionalization strategy with bisbenzylamide substituents on the cyclohexane ring.

The synthetic sequence leading to **3.29** was based on the glycosylation between the azido-mannosyl donor **3.10** and the slightly modified preparation of the PNP-activated acceptor **3.25** (analogous to **2.6**, but the linker contains a terminal chlorine, instead of the azide, for compatibility reasons). The glycosylation was followed by the amide coupling with *p*-hydroxymethylenebenzylamine, CuAAC with propargylamine and in the final step by the Zemplén-deprotection of the mannose residue. (**Scheme 23**)




***Scheme 23.*** *Synthesis of* ***3.29****. The ligand comprises two glycomimetic strategies: functionalization at the C-2 position of the mannose residue to reach an ammonium binding pocket, and bisbenzylamide substituents on the cyclohexane sugar-mimic ring to form lipophylic interactions with DC-SIGN CRD.*

### 3.2.6.2 Affinity results of 3.29

Remarkably, ligand **3.29** reached an $IC_{50}$ value of 75 μM in SPR inhibition assay on a Man-BSA surface, while in the same campaign the affinity of compound **3.3** and **3.23** was determined as 160 μM and 180 μM, respectively. The inhibitory potency of **3.29** was superior to every glycomimetic compound synthesized in our group and it is up to date among the best monovalent DC-SIGN inhibitors. Due to the lack of a standardized assay system and universal reference, we cannot directly compare its affinity to ligands developed in other research groups, but it likely falls in a similar range as the binaphyl-modified glycomimetic ligand (**1.15**) of the Anderluh group. Unlike **1.15**, however, ligand **3.29** can be readily presented on multivalent constructs in two steps (chlorine-azide exchange and click on dendrons/dendrimers, for example).

The outstanding $IC_{50}$ values determined in SPR tests prompted the further investigation of the binding properties of **3.29**. These experiments and related analyses were performed in the laboratory of Prof. Franck Fieschi (Silvia Achilli, Dr. Michel Thépaut, IBS, Grenoble, France).

The dissociation constant was measured in isothermal titration microcalorimetry (ITC) with DC-SIGN ECD, and determined as $K_D$ = 53 μM. ITC is one of the most precise techniques available to determine equilibrium binding constants, as it measures the heat changes in a cell upon addition of increasing concentrations of the ligand solution to a fixed concentration of the lectin solution (compared to a reference cell). These changes evolve from the binding event in the analyte cell, i.e. relate to the enthalpy (ΔH) of binding, and by plotting the data, the main thermodynamic information (e.g. $K_D$) of the interaction can be obtained.

### 3.2.6.3 X-ray crystal structure of 3.29

More importantly, ligand **3.29** was successfully co-crystallized with the CRD of DC-SIGN and provided invaluable information of the binding mode and established interactions. (**Figure 86**)

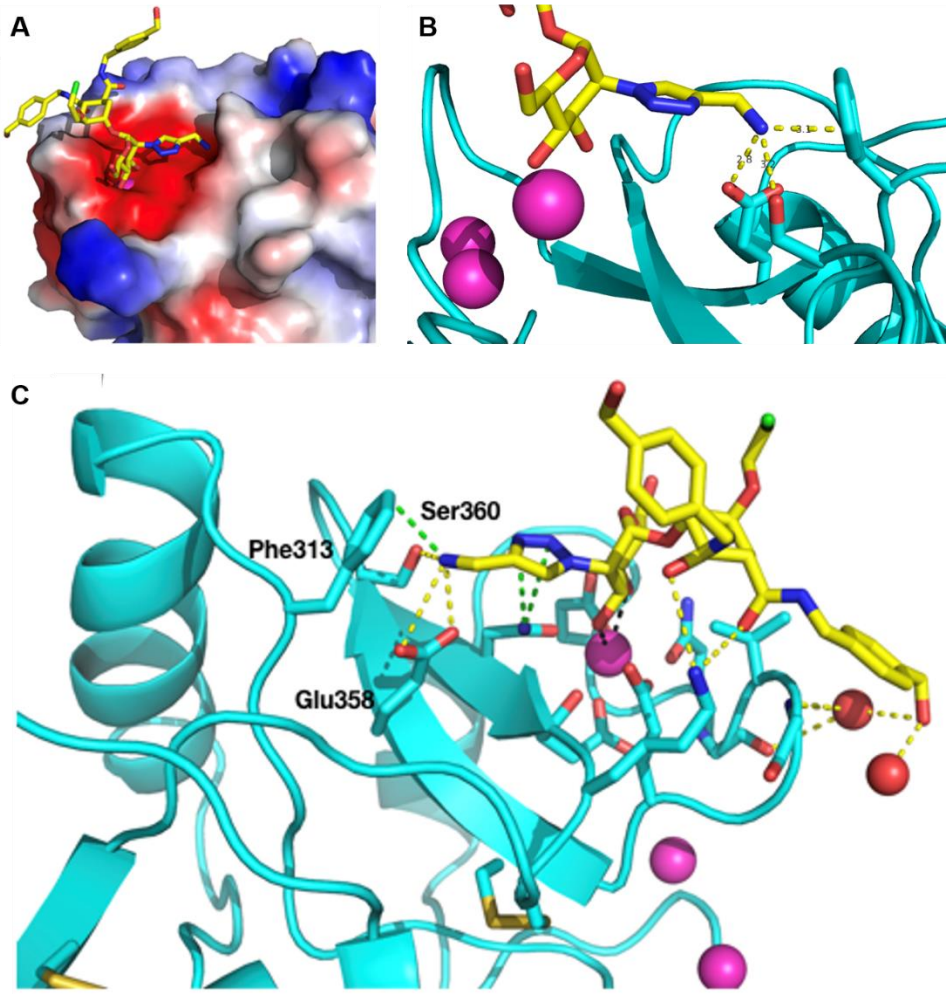



***Figure 86.*** *X-ray crystal structure of* ***3.29*** *complexed in the CRD of DC-SIGN. As anticipated, the ligand assumes a similar pose to the parent glycomimetic pseudo-disaccharide* ***1.18*** *(PDB code: 2XR5). Additionally, the computationaly predicted contacts are formed in the ammonium binding pocket, more specifically, the ammonium group engages in interactions with Ser360, Phe313 and Glu358. This is a direct confirmation of the existence and exploitability of this binding site. At the same time, one of the bisbenzylamide groups packs against the protein surface around Val351 and forms indirect H-binds through two water molecules. The exact position of the lipophilic amides in* ***1.21a*** *has not been resolved by X-ray crystallography before, the ligand was only studied in a computational model and STD-NMR experiments. On the other hand, one of these aromatic groups appears to interact with another CRD - further studies of this observation are underway.*

The first confirmation we obtained by this crystal structure is that ligand **3.29** assumes a similar pose as the pseudo-disaccharide **1.18** in the binding site, and this demonstrates that the rationally designed glycomimetic ligands of different modifications (C-2 functionalization, bisamide substituents) were indeed establishing interactions in the originally predicted regions.

The ammonium group is accommodated in the computationally identified binding pocket in the central region of the CRD. The cation is positioned at an ideal distance from the amino acid residues to establish H-bonds with Ser360 (d: 3.2 Å); cation-π type interactions with Phe313 (d: 3.1 Å) and electrostatic interactions with Glu358 (d: 2.8 Å). We can therefore confirm that the *in silico* identified binding site exists and the contacts predicted computationally are established in reality.

The triazole ring connecting the methyleneammonium moiety with the rest of the glycomimetic scaffold lies on the flat surface between the $Ca^{2+}$ binding site and the ammonium binding pocket, contacting an asparagine residue. The mannose ring is coordinating the metal ion at the primary binding site, while the cyclohexane sugar mimic ring is pointing towards Val351, in the same extended conformation as seen for **1.18**.

Analogously to the docked pose of **1.21a** (Man030), one of the bisbenzylamide substituents (on C5 of the cyclohexane ring, according to the conventional numbering of the glycomimetic molecule) forms van der Waals interactions by packing against the CRD surface around Val351. Curiously, H-bonds are established between the terminal hydroxyl-group and two water molecules crystallized in the complex. The position of this lipophilic group verifies the binding mode not only of **3.29**, but also the members of the **1.21** bisbenzylamide series, including Man030, and it is in accordance with the computational model and STD-NMR results.

In addition to providing a stable linkage between the aromatic substituents and the cyclohexane ring, the amide bonds seem to play an additional role in binding.

The role of the other *p*-hydroxymethylenebenzylamide moiety (on C4 of the cyclohexane) raises important questions, however. On the 1 : 1 ligand/CRD image this substituent seems to be "floating" in the solvent, distant from the CRD surface. In the crystal cell, however, this substituent is visibly interacting with another CRD at a site far from the principal sugar binding site. This casts doubt on the real affinity of the ligand, because CRD-clustering would artificially enhance the observed potency, as it was demonstrated for the pseudo-trisaccharide **1.20** (see Introduction). This effect would not be detected in solution-phase SPR and ITC experiments - as the clustering could occur at a 2 : 2 ligand/CRD ratio – but would cause an overestimation of affinity values.

This phenomenon does not necessarily occur in solution phase, though. The interactions between the benzylamide-group and the other CRD might be an artefact of co-crystallization, due to the compact arrangement of the complex. This can be determined by analytical ultracentrifugation - a technique that separates the large molecules based on their hydrodynamic radius. This technique supplied the evidence of the bivalent nature and DC-SIGN clustering activity of **1.20**. These studies are underway in the group of Prof. F. Fieschi, IBS, Grenoble, France.

The X-ray structure of **3.29** implies a number of options to improve the design of our future glycomimetics. A closer look at the ammonium binding site indicates that lipophilic substituents on the methylene-group next to the ammonium ion could perhaps shield the cation from the intrusive effects of the solvent. Carrying one or two methyl-groups in this position, 1-methyl-prop-2-ynylamine or 2-methyl-3-butyne-2-amine could be clicked on the azide scaffold **3.1**, in order to verify the efficiency of this strategy. Additionally, the orientation of the ammonium group and the position of the amino acid residues suggests that from the two 1-phenyl-propargyl amine enantiomers, the glycomimetic derivative displaying (*R*)-configuration on the chiral centre has higher chances to interact favourably with Phe313.

### 3.2.7 Selectivity analyses of 3.3, 3.23 and 3.29

The importance of selectivity has been emphasized numerous times in the previous chapters, particularly in connection with selectivity towards DC-SIGN over langerin (see Introduction and **2.2.3.4**). Ligand selectivity between the two CLRs was analyzed in SPR inhibition tests (Silvia Achilli, Fieschi laboratory, IBS, Grenoble, France), for ligands **3.3**, **3.23** and **3.29**. Once more, the results of **3.29** are superior among the other glycomimetics, as the ligand is highly selective towards DC-SIGN. The pyridine derivative **3.23** shows insignificant affinity towards langerin, and **3.3** is able to interact with the lectin at high concentrations, but this cannot be considered effective binding. These experiments underline that **3.29** is an outstanding monovalent DC-SIGN ligand, with excellent affinity and selectivity towards the CLR. (See Appendix)
To further confirm the ligand's promising lead-like properties, metabolic stability analysis in mouse serum is planned in the near future.

## 3.3 Conclusions

In the third chapter of this thesis, the rational design and synthesis of highly potent DC-SIGN antagonists has been described. Virtual fragment screening, conducted by the Martín-Santamaría group, identified a novel binding site in the CRD of the lectin that could accommodate an ammonium-group in the vicinity of the Phe313 residue. This residue plays a key role in the binding mode of several natural mannosides, as well as a binaphtyl glycomimetic ligand **1.15** from the Anderluh group, but the pseudo-disaccharide **1.18** does not take advantage of it, as it forms contacts in the opposite direction of the CRD. The crystal structure of **1.18** indicated that the hydroxyl group in position 2 of the natural mannose residue points towards the direction of the predicted binding pocket – lined with Ser360, Phe313 and Glu358. We hypothesized therefore, that drug-like fragments - predicted *in silico* to interact around Phe313 - could be linked with the glycomimetic scaffold at C-2 and increase affinity.
Two main synthetic strategies were conceived for this functionalization and both could benefit from a N-atom in the second position of the mannosyl donor **3.10**. On one hand, an amino group could be used to form amide bonds between the fragments and the general glycomimetic scaffold **3.1**, and on the other, an azido-group could participate in Cu(I)-catalyzed cycloadditions with alkynes. Previous results indicated that the amide linkage did not provide appropriate spacing to cover the distance between the primary and the predicted binding site, and so, we investigated the second strategy.
The common scaffold **3.1** can be obtained by the glycosylation of the dimethyl ester cyclohexane acceptor **2.27** and an acetyl-protected, TCA-activated 2-azido mannosyl donor (**3.10**). We started the synthesis of the donor from *N*-acetyl mannosamine hydrochloride by a diazo transfer reaction, to install the azide. Our initial trials resulted in low, variable yields that could be tracked back to the base-catalyzed epimerization at the C-2 centre, leading to the formation of the 2-azido gluco-analogue. By carrying out the reaction at lower temperatures, under a tighter pH control, this side reaction could be effectively prevented and high, reproducible yields were obtained. The other steps in the sequence were straightforward deprotection and activation steps, well-known in sugar chemistry. The glycosylation afforded the azido-scaffold **3.1** in good yields.
Exploration of favourable interactions around the ammonium binding site was initially based on the affinity data available about the propargylamine derivative **3.3**. In an attempt to stabilize the positive charge, and possibly form additional lipophilic interactions, a series of mono-, di- and trimethylated ammonium derivatives were prepared - their affinity, however, remained below the unsubstituted ligand.
Propargyl alcohol derivatization had also been known to improve the affinity, although to a lesser extent than propargylamine. The role of aryl-substituents was evaluated by linking the 1-phenyl propargyl alcohol derivatives

to the glycomimetic scaffold (**3.20a** and **3.20b**), but none of the two diastereomers were able to enhance affinity. Efforts were made to functionalize the ligands with synthetically more challenging 1-phenyl propargylamine, but the role of these substituents are yet to be elucidated. To our surprise, another aromatic system lacking the positive charge – the pyridine containing **3.23** - exhibited almost as good inhibitory potency as **3.3**.

The most important results were obtained after we combined the C-2 functionalization strategy with the bisbenzylamide substituents on the cyclohexane ring. These two modifications were meant to exploit interactions both around Phe313 and near Val351 – two residues involved in the binding of a different set of ligands. In accordance with our expectations, ligand **3.29** - compressing an ammonium group and *p*-hydroxymethylenebenzylamide moieties - displayed the highest affinity among the pseudo-disaccharide-based glycomimetics ($IC_{50}$: 75 μM, $K_D$: 53 μM), and this is one of the strongest monovalent DC-SIGN ligands, up to date. (**Figure 87**)

The X-ray crystal structure of **3.29** in the CRD of DC-SIGN offered unique insight on the structural details of binding interactions in the ammonium pocket, as well as on the earlier not resolved binding mode of bisamide ligands. The structure of the complex validated the predicted interactions between the ammonium ion and the serine-phenylalanine-glutamic acid residues, and served as a proof that the bisbenzylamide group establishes van der Waals interactions with the CRD around Val351.

Additionally, **3.3**, **3.23** and **3.29** are highly selective DC-SIGN ligands that do not inhibit langerin. Further studies are underway to evaluate the metabolic stability of **3.29**, and to install other favourable modifications in the glycomimetic structure.

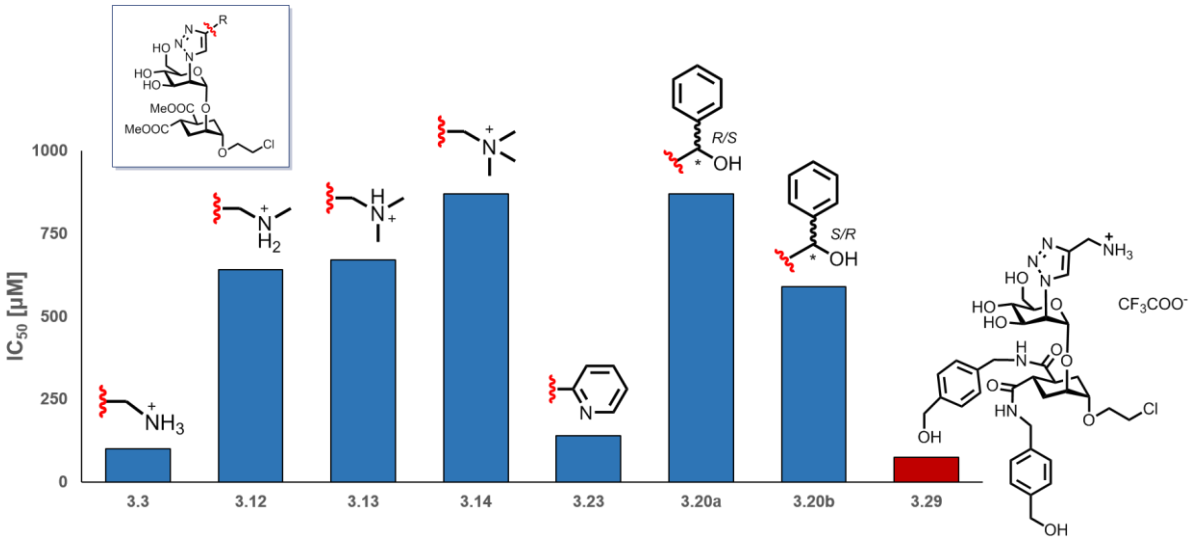


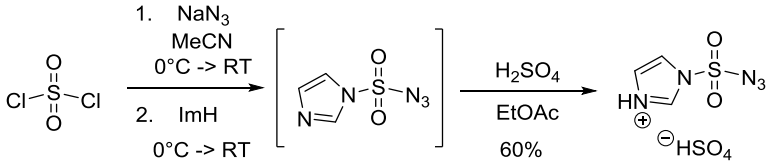

***Figure 87.*** *Affinity values of C-2 modified glycomimetic ligands, determined in SPR inhibition assays. Three ligands stand out of these results: the simple propargylamine derivative* **3.3** *($IC_{50}$: 100 μM), the 2-picolylamine derivative* **3.23** *($IC_{50}$: 160 μM) and* **3.29**, *a ligand comprising two modification startegies.* **3.29** *simultaneously targets the in silico predicted ammonium binding pocket near Phe313, and engages in lipophilic interactions with the CRD surface in the vicinity of Val351*

## 3.4 EXPERIMENTAL SECTION

See general experimental details in **2.1.2.1**.

### 3.4.1 Synthesis of intermediates

Imidazole-1-sulfonyl azide hydrogen sulfate **3.6**

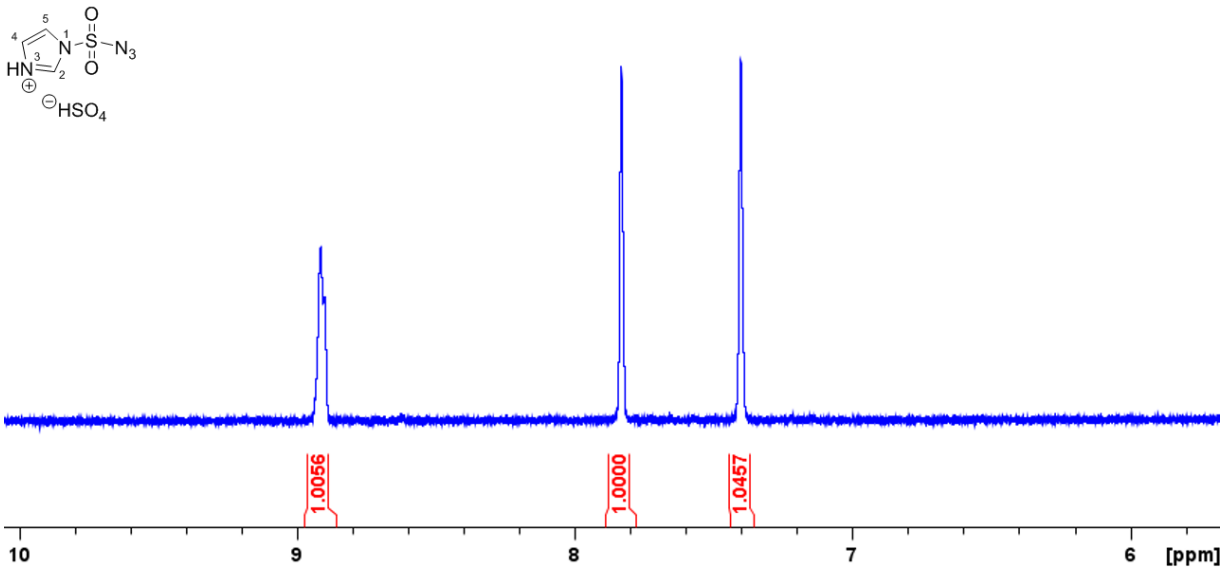



**3.4** **3.5** **3.6**

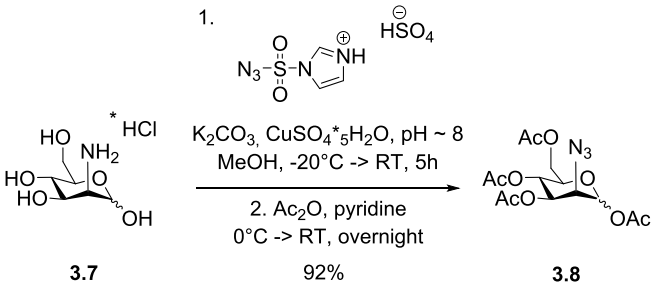
$NaN_3$ (1.04 g, 16 mmol, 1 equiv.) was suspended in dry MeCN under nitrogen atmostphere and the mixture was cooled to 0°C. Sulfuryl chloride (1.293 ml, 16 mmol, 1 equiv.) was added dropwise to the stirred mixture, and the suspension turned light pink. The reaction was left to warm to RT and after 9 hours cooled again to 0°C when imidazole (2.07 g, 30.4 mmol, 1.9 equiv.) was slowly added. The mixture was stirred overnight at RT, and turned light brown. Th next day, the reaction was concentrated to 2/3 volume, diluted with 20 ml EtOAc and washed with 2 * 20 ml water, 2 * 20 ml saturated $NaHCO_3$ solution and finally with 20 ml water. The aqueous phase was treated according to the standard azide-quenching procedure (see **2.1.2.2**). The organic phase was dried over $Na_2SO_4$, filtered and concentrated to 1/2 volume *in vacuo*. The solution was cooled to 0°C and cc. $H_2SO_4$ (896 µl, 16 mmol, 1 equiv.) in 32 ml EtOAc was added dropwise. After 1.5 hours, the precipitate was filtered on a Büchner-funnel and washed with cold EtOAc to yield the product as white crystals. WARNING: Imidazole-1-sulfonyl azide and its salts are potentially explosive and hazardous, due to their sensitivity to heat, impact, friction, and electrostatic discharge.

Yield: 60%
$^1$H NMR (400 MHz, DMSO-$d_6$): $\delta$ = 8.91 (dd, 1H, $H_2$, $J$ = 2.4 Hz, $J$ = 5.6 Hz), 7.83 (dd, 1H, $H_5$, $J$ = 1.4 Hz, $J$ = 1.8 Hz), 7.40 (dd, 1H, $H_4$, $J$ = 1.0 Hz, $J$ = 1.8 Hz)

**$^1$H NMR spectrum of 3.6 in $D_2O$ (400 MHz)**




1,3,4,6-tetra-*O*-acetyl-2-azido-2-deoxy-D-mannopyranose (**3.8**)




Mannosamine hydrochloride (1.2050 g, 5.59 mmol, 1 equiv.) was suspended in dry MeOH (50 ml) and cooled to -20°C. $K_2CO_3$ (811 mg, 5.87 mmol, 1.05 equiv.) was dissolved in 4 ml water and added to the stirred suspension. $CuSO_4$*$5H_2O$ (14 mg) was added to the mixture. Another portion of $K_2CO_3$ (811 mg, 5.87 mmol, 1.05 equiv.) in 4 ml water was added, immediately followed by the addition of imidazole-1-sulfonyl azide (1.8190 g, 6.7080 mmol, 1.2 equiv.).The pH was continuously controlled and the reactions was stirred at -20°C for 2 hours, when it was left to warm to RT (light blue suspension). After 4 hours, only one major spot was detected by TLC ($R_f$ = 0.4 in DCM : MeOH = 8 : 2). The solvent was evaporated *in vacuo* and the crude was suspended in dry pyridine (20 ml). The mixture was cooled to 0°C and $Ac_2O$ (4.22 ml, 44.72 mmol, 8 equiv.) was added dropwise. The reaction was stirred overnight at RT and the next day only one major spot was detected on TLC ($R_f$ = 0.36 in toluene : EtOAc = 9 : 1). The crude was purified by flash chromatography (toluene : EtOAc = 9 : 1) to yield the pure product as a white foam.

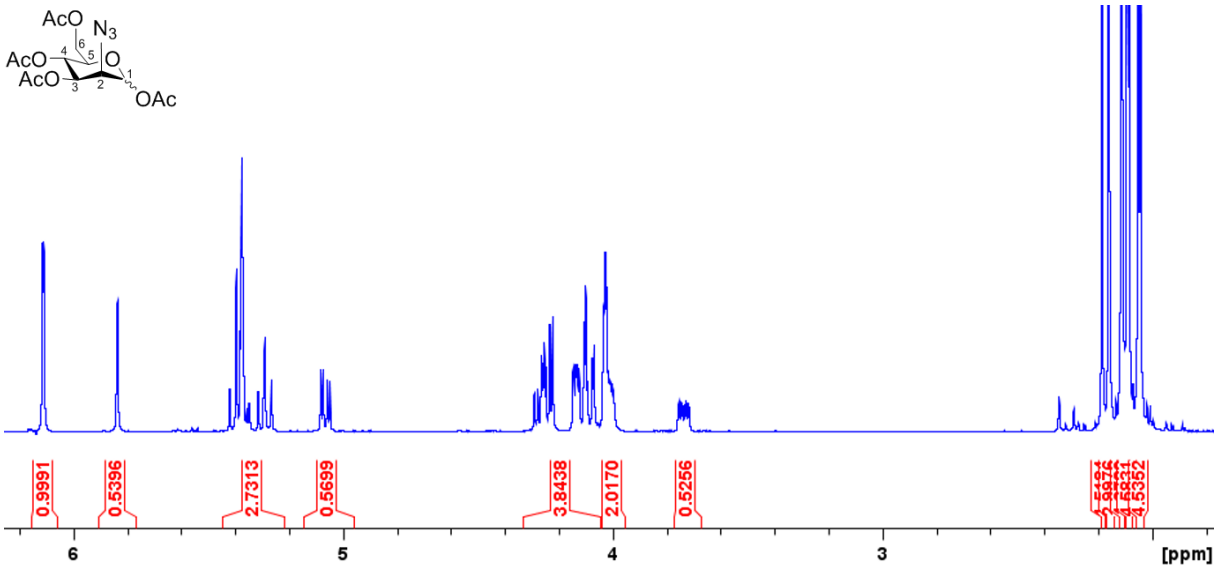
Yield: 92%, α : β = 2 : 1
α-anomer: $^1$H NMR (400 MHz, $CDCl_3$): δ = 6.09 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.9 Hz), 5.43 – 5.34 (m, 2H, $H_3$, $H_4$), 4.21 (dd, 1H, $H_{6a}$, *J* = 4.6 Hz, *J* = 12.4 Hz), 4.08 (dd, 1H, $H_{6b}$, *J* = 2.2 Hz, *J* = 12.4 Hz), 4.04 – 3.98 (m, 2H, $H_2$, $H_5$), 2.16 (s, 3H, AcO), 2.11 (s, 3H, AcO), 2.09 (s, 3H, AcO), 2.05 (s, 3H, AcO)

$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 170.6 (C=O (AcO)); 169.9 (C=O (AcO)); 169.3 (C=O (AcO)); 168.1 (C=O (AcO)); 91.7 ($C_1$); 71.1 ($C_3$); 71.1 ($C_5$); 65.6 ($C_4$); 62.1 ($C_6$); 60.1 ($H_2$); 21.2 (Me (OAc)); 20.9 (Me (OAc)); 20.9 (Me (OAc)); 20.8 (Me (OAc))

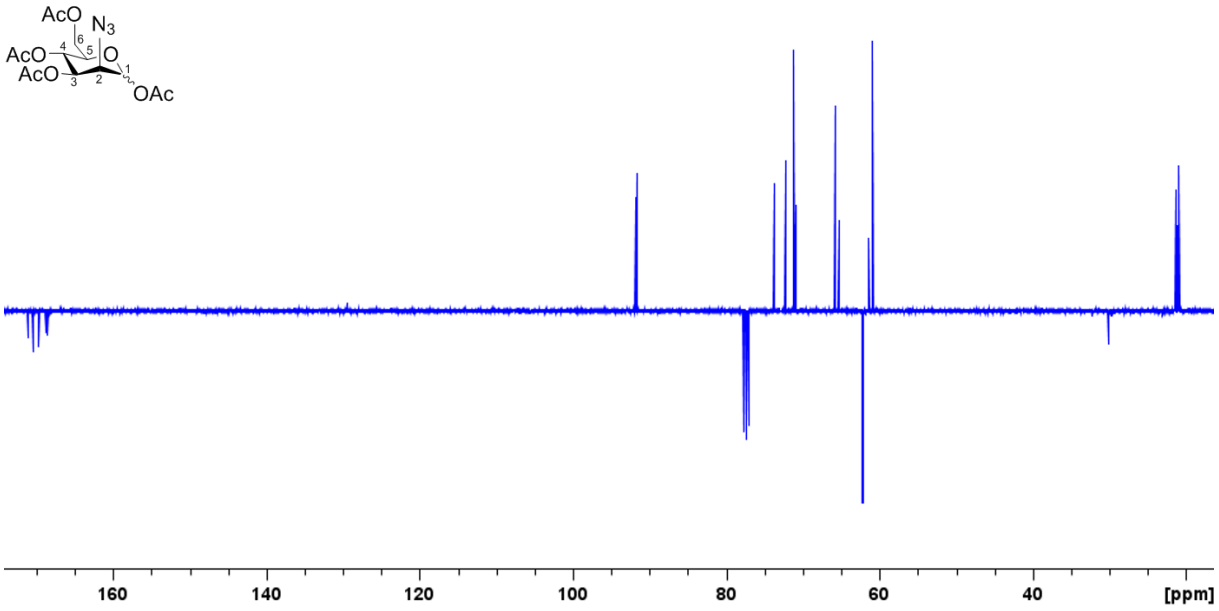
β-anomer: δ = 5.81 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.4 Hz), 5.27 (t, 1H, $H_4$), 5.04 (dd, 1H, $H_3$, *J* = 3.7 Hz, *J* = 9.8 Hz), 4.23 (dd, 1H, $H_{6a}$, *J* = 4.9 Hz, *J* = 12.4 Hz), 4.06 (dd, 1H, $H_{6b}$, *J* = 2.4 Hz, *J* = 12.4 Hz), 3.70 (ddd, 1H, $H_5$, *J* = 2.4 Hz, *J* = 4.9 Hz, *J* = 9.9 Hz), 2.18 (s, 3H, AcO), 2.11 (s, 3H, AcO), 2.08 (s, 3H, AcO), 2.04 (s, 3H, AcO)

$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 170.6 (C=O (AcO)); 169.9 (C=O (AcO)); 169.3 (C=O (AcO)); 168.1 (C=O (AcO)); 91.5 ($C_1$); 73.6 ($C_5$); 72.2 ($C_3$); 65.3 ($C_4$); 62.1 ($C_6$); 61.4 ($H_2$); 21.1 (Me (OAc)); 21.1 (Me (OAc)); 21.0 (Me (OAc)); 20.9 (Me (OAc))

**$^1$H NMR spectrum of 3.8 in $CDCl_3$ (400 MHz)**



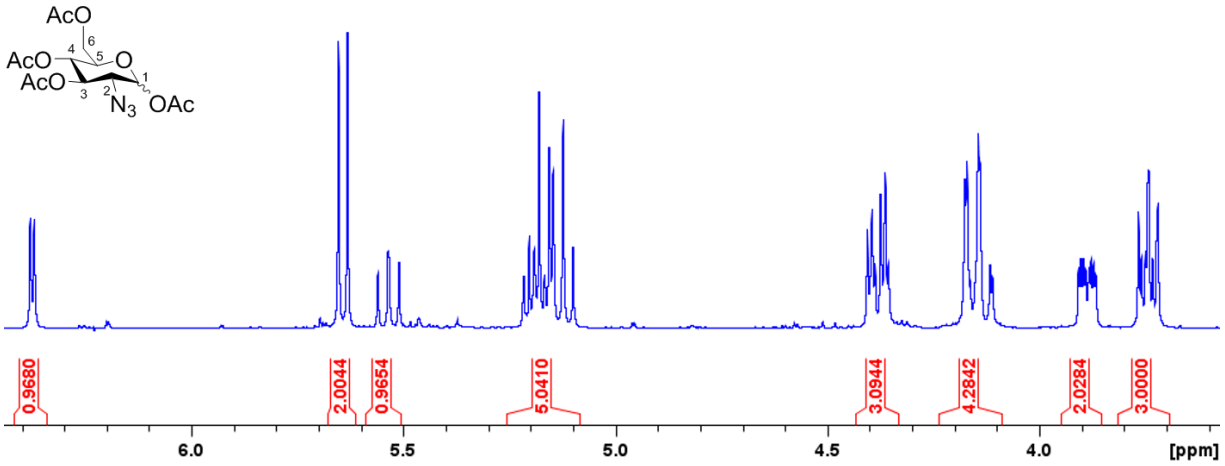

**$^{13}$C NMR spectrum of 3.8 in $CDCl_3$ (100 MHz)**



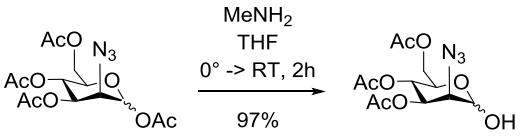
During the trials, a significant amount of 1,3,4,6-tetra-*O*-acetyl-2-azido-2-deoxy-D-glucopyranose formed, the NMR data are reported below.

$^1$H NMR (400 MHz, $CDCl_3$): α-anomer: δ = 6.25 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 3.7 Hz), 5.41 (dd, 1H, $H_3$, *J* = 9.4 Hz, 10.4 Hz), 5.1 – 4.9 (m, 1H, $H_4$), 4.25 (dd, 1H, $H_{6a}$, *J* = 4.1 Hz, *J* = 12.5 Hz), 4.12 – 3.98 (m, 2H, $H_5$, $H_{6b}$), 3.62 (dd, 1H, $H_2$, *J* = 3.7 Hz, 10.5 Hz), 2.15 (s, 3H, AcO), 2.06 (s, 3H, AcO), 2.03 (s, 3H, AcO), 2.00 (s, 3H, AcO)
β-anomer: δ = 5.51 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 8.6 Hz), 6.1 – 4.9 (m, 2H, $H_3$, $H_4$), 4.24 (dd, 1H, $H_{6a}$, *J* = 4.5 Hz, *J* = 12.4 Hz), 4.04 (dd, 1H, $H_{6a}$, *J* = 2.2 Hz, *J* = 12.4 Hz), 3.76 (ddd, 1H, $H_5$, *J* = 2.2 Hz, *J* = 4.5 Hz, *J* = 9.7 Hz), 3.62 (dd, 1H, $H_2$, *J* = 8.6 Hz, *J* = 9.8 Hz), 3.14 (s, 3H, AcO), 2.04 (s, 3H, AcO), 2.03 (s, 3H, AcO), 1.98 (s, 3H, AcO)

**$^1$H NMR spectrum of 1,3,4,6-tetra-*O*-acetyl-2-azido-2-deoxy-D-glucopyranose in $CDCl_3$ (400 MHz)**




3,4,6-tri-*O*-acetyl-2-azido-2-deoxy-D-mannopyranose (**3.9**)



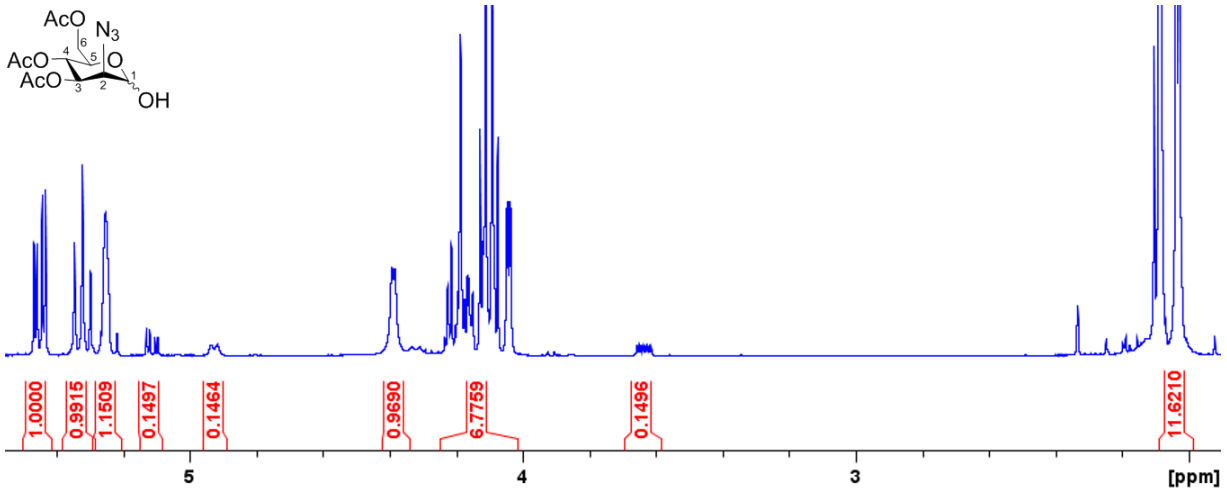

**3.8** **3.9**

1,3,4,6-tetra-*O*-acetyl-2-azido-2-deoxy-D-mannopyranose **3.8** (506 mg, 1.36 mmol, 1 equiv.) was dissolved in 1 ml dry THF and cooled to 0°C. 813 µl 2M $MeNH_2$ in THF (1.626 mmol, 1.2 equiv.) was added to the solution and the reaction was stirred at RT for 1.5 hours. After completion, the mixture was concentrated and the crude was purified by flash chromatography ($R_f$: 0.27 (α) and 0.24 (β) in toluene : EtOAc = 6 : 4).

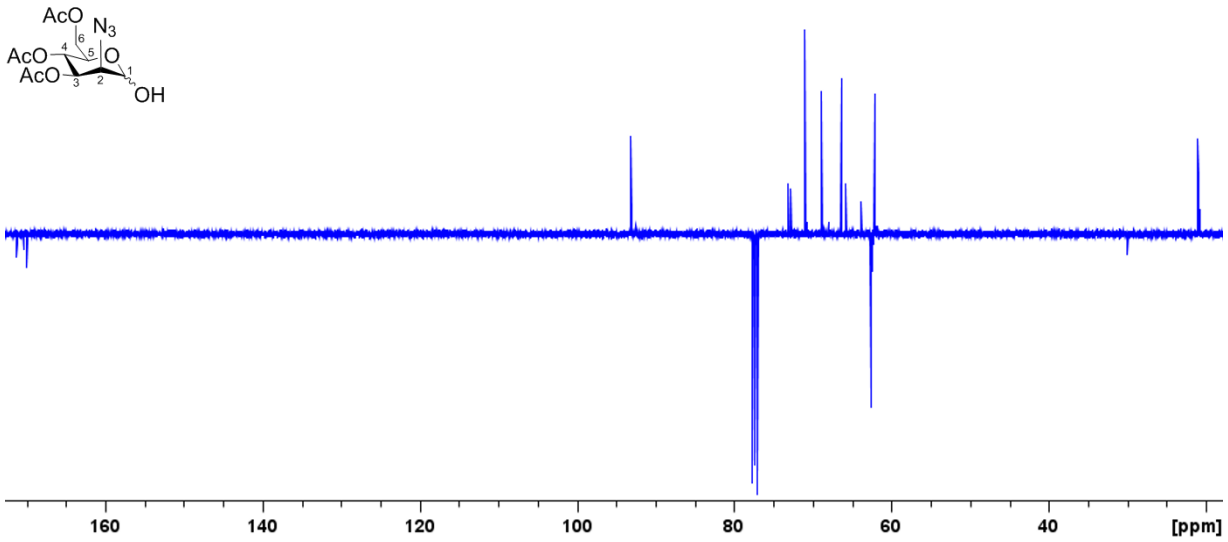
Yield: 97%, α : β = 87 : 13
$[\alpha]_D^{20}$: +92.2° (c = 0.7 in $CH_2Cl_2$)
$^1$H NMR (400 MHz, $CDCl_3$): α-anomer: δ = 5.45 (dd, 1H, $H_3$, $J_{2\text{-}3}$ = 3.8 Hz, $J_{3\text{-}4}$ = 9.8 Hz), 5.35 (t, 1H, $H_4$, $J_{4\text{-}5}$ = 9.5 Hz), 5.27 (s, 1H, $H_1$), 4.26 – 4.09 (m, 3H, $H_{6a,b}$, $H_5$), 4.06 (dd, 1H, $H_2$, $J_{1\text{-}2}$ = 1.8 Hz), 2.10 (s, 3H, OAc), 2.09 (s, 3H, OAc), 2.05 (s, 3H, OAc)
β-anomer: δ = 5.29 – 5.23 (m, 1H, $H_4$), 5.13 (dd, 1H, $H_3$, $J_{2\text{-}3}$ = 3.8 Hz, $J_{2\text{-}3}$ = 9.9 Hz), 4.91 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 6.4 Hz), 4.26 – 4.10 (m, 2H, $H_{6a,b}$), 3.67 – 3.60 (m, 2H, $H_5$, $H_2$), 2.12 (s, 3H, OAc), 2.10 (s, 3H, OAc), 2.05 (s, 3H, OAc)

$^{13}$C NMR (100 MHz, $CDCl_3$): α-anomer: δ = 170.8 (C=O (AcO)); 170.0 (C=O (AcO)); 169.6 (C=O (AcO)); 92.8 ($C_1$); 70.7 ($C_3$); 68.6 ($C_5$); 65.9 ($C_4$); 62.2 ($C_6$); 61.7 ($C_2$); 20.8 (Me (OAc)); 20.7 (Me (OAc)) ; 20.6 (Me (OAc))
β-anomer: δ = 170.8 (C=O (AcO)); 170.0 (C=O (AcO)); 169.6 (C=O (AcO)); 92.7 ($C_1$); 72.8 ($C_3$); 72.4 ($C_5$); 65.4 ($C_4$); 63.5 ($C_2$); 62.1 ($C_6$); 20.8 (Me (OAc)); 20.7 (Me (OAc)) ; 20.6 (Me (OAc))
MS (ESI): m/z calculated for $[C_{12}H_{17}N_3O_8Na]^+$: 354.10 $[M+Na]^+$; found: 354.78

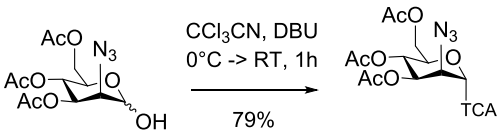
**$^1$H NMR spectrum of 3.9 in $CDCl_3$ (400 MHz)**




**$^{13}$C NMR spectrum of 3.9 in $CDCl_3$ (100 MHz)**



2-azido-2-deoxy-3,4,6-tri-*O*-acetyl-D-mannose trichloroacetimidate (**3.10**)

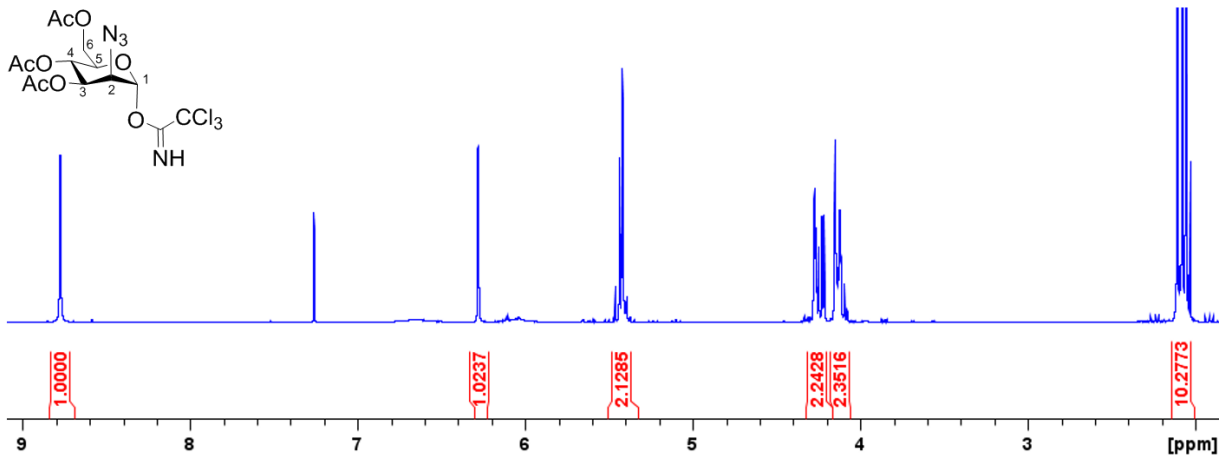



**3.9** **3.10**

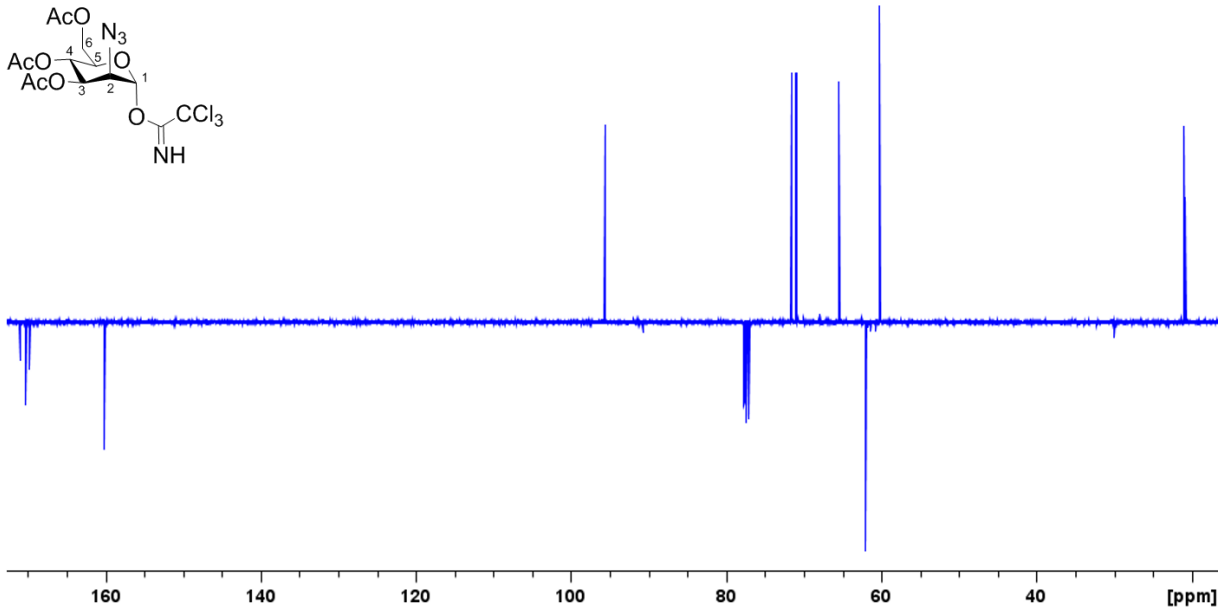
3,4,6-tri-*O*-acetyl-2-azido-2-deoxy-D-mannopyranose **3.9** (400 mg, 1.20 mmol, 1 equiv.) was dissolved in 5 ml DCM, cooled to 0°C and trichloroacetonitrile (1.33 ml, 13.29 mmol, 11 eqiv.) was added to the solution. Then, DBU (45 µl, 0.30 mmol, 0.25 equiv.) was added and the reaction was stirred at RT under $N_2$ overnight. After completion, the mixture was concentrated and the crude was purified by flash chromatography ($R_f$: 0.38 in H : EA = 75 : 25) to yield the pure product as a white foam.

Yield: 77% (Exclusively α-product)
$[\alpha]_D^{20}$: +88.9° (c = 0.6 in $CH_2Cl_2$)
$^1$H NMR (400 MHz, $CDCl_3$): δ = 8.77 (s, 1H, NH), 6.29 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.8 Hz), 5.45 (t, 1H, $H_4$, $J_{4\text{-}3}$ = $J_{4\text{-}5}$ = 10.0 Hz), 5.43 (m, 1H, $H_3$), 4.28 (dd, 1H, $H_2$, $J_{2\text{-}3}$ = 3.3 Hz), 4.25 (dd, 1H, $H_6$, $J_{6a\text{-}6b}$ = 12.7 Hz, $J_{6a\text{-}5}$ = 4.8 Hz,), 4.18 - 4.12 (m, 2H, $H_5$, $H_{6b}$), 2.12 (s, 3H, OAc), 2.09 (s, 3H, OAc), 2.07 (s, 3H, OAc)
$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 170.8 (C=O (AcO)); 170.1 (C=O (AcO)); 169.6 (C=O (AcO)); 160.0 (C=NH); 95.5 (C1); 90.42 (CCl3); 71.4 (C5); 70.8 (C3); 65.2 (C4); 61.8 (C6); 60.0 (C2); 20.9 (OAc); 20.8 (OAc); 20.7 (OAc)
MS (ESI): m/z calculated for $[C_{14}H_{17}Cl_3N_4O_8Na]^+$: 497.01 $[M+Na]^+$; found: 496.88

**$^1$H NMR spectrum of 3.10 in $CDCl_3$ (400 MHz)**



**$^{13}$C NMR spectrum of 3.10 in $CDCl_3$ (100 MHz)**

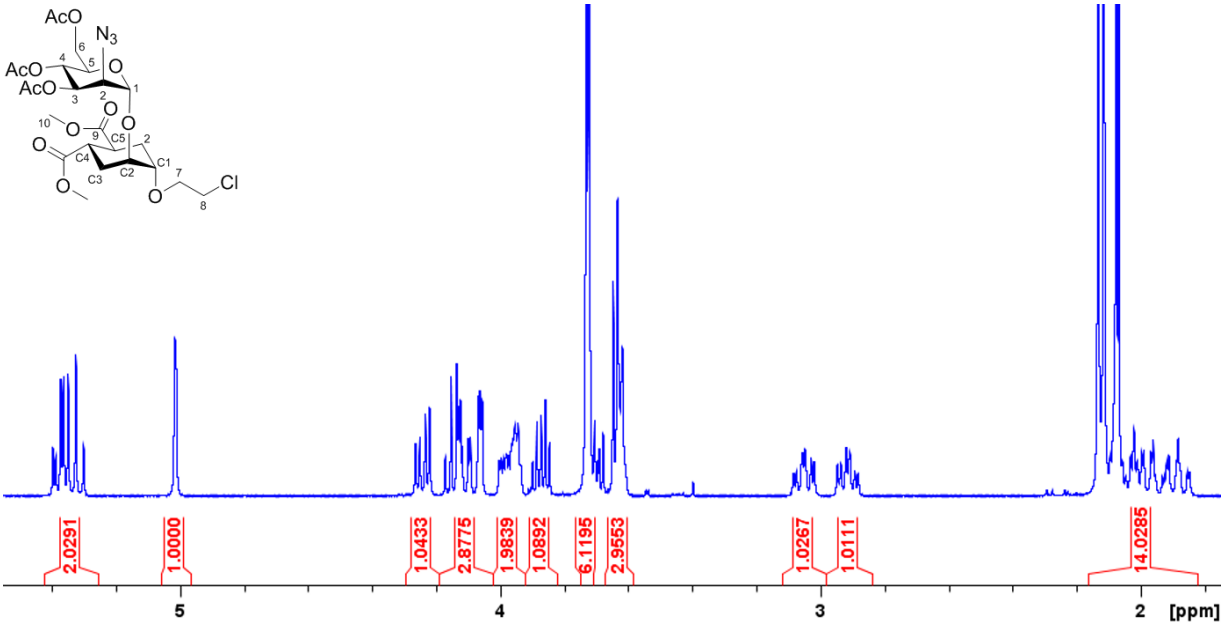



1,2-cyclohexanedicarboxylic acid, 4-(2-chloroethoxy)-5-((3,4,6-*O*-triacetyl)-2-azido-α-D-2-deoxymannopyranosyloxy)-1,2-dimethylester (1*S*,2*S*,4*S*,5*S*) (**3.11**)

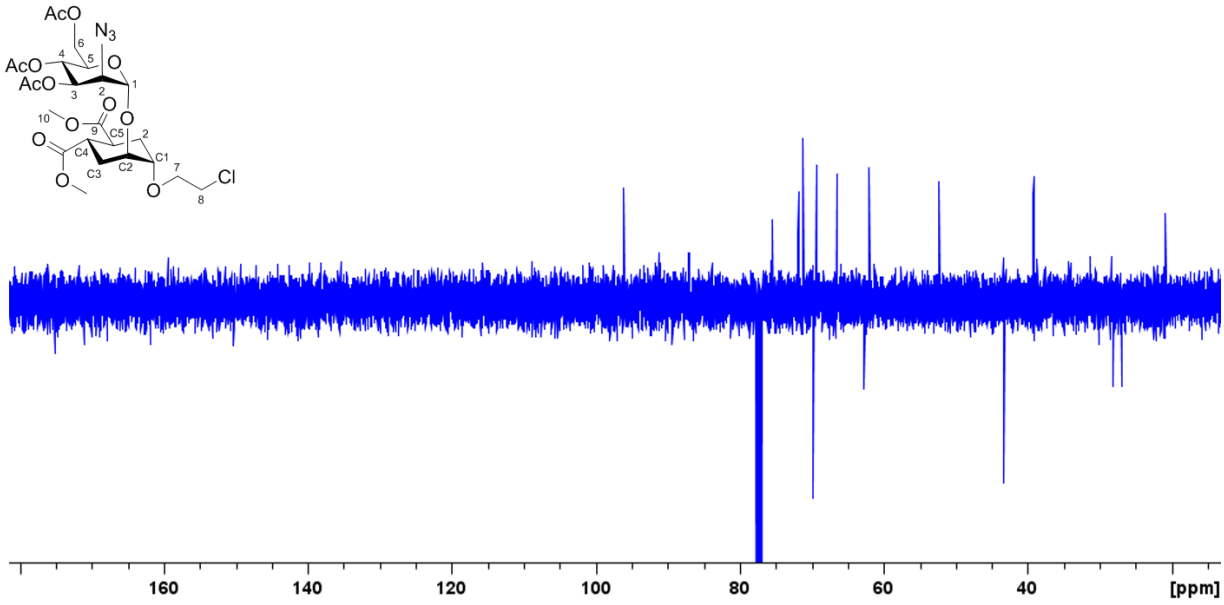
A mixture of the acceptor **2.27** (200 mg, 0.42 mmol, 1 equiv.) and the donor **3.10** (123.8 mg, 0.42 mmol, 1 equiv.) was co-evaporated with toluene three times. Powdered and activated 4Å molecular sieves (acid washed) were added. The mixture was kept under vacuum for a few hours and then dissolved in dry DCM (10 ml). After cooling to -30°C, TMSOTf (15.2 µl, 0.084 mmol, 0.2 equiv.) was added to the stirred mixture. The solution was stirred at -30°C for 1 hour and upon completion, the reaction was quenched with TEA. The mixture was warmed to room temperature and filtered over a celite pad. The filtrate was evaporated at reduced pressure and the crude product was purified by flash chromatography (Hex : EtOAc = 72 : 28) to yield the pure product **3.11** as a white foam.

Yield: 94% (exclusively α-product)
$[\alpha]_D^{20}$: +73.4° (c = 0.68 in $CH_2Cl_2$)
$^1$H NMR (400 MHz, $CDCl_3$): δ = 5.35 (dd, 1H, $H_3$, $J_{3\text{-}2}$ = 3.6 Hz, $J_{3\text{-}4}$ = 9.6 Hz), 5.29 (dd, 1H, $H_4$, $J_{4\text{-}5}$ = 9.4 Hz), 4.98 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.5 Hz), 4.21 (dd, 1H, $H_{6a}$, $J_{6\text{-}5}$ = 5.4 Hz, $J_{6\text{-}6b}$ = 12.2 Hz), 4.08 (dd, 1H, $H_{6b}$, $J_{6b\text{-}5}$ = 2.3 Hz), 4.03 (m, 1H, $H_2$), 3.98 - 3.94 (m, 1H, $H_5$), 3.94 - 3.90 (m, 1H, $H_2$), 3.87 - 3.80 (m, 1H, $H_{7a}$), 3.70 (s, 3H, $H_{10}$), 3.69 (s, 3H, $H_{10}$), 3.68 - 3.64 (m, 1H, $H_{7b}$), 3.62 - 3.59 (m, 2H, 2H, $H_{8a,b}$), 3.59 - 3.58 (m, 1H, $C_1$), 3.02 (dt, 1H, $C_4$, $J_{C4\text{-}C3eq}$ = 4.0 Hz, $J_{C4\text{-}C3ax}$ = $J_{C4\text{-}C5}$ = 12.0 Hz,), 2.88 (dt, 1H, $H_{C5}$, $J_{C5\text{-}C6eq}$ = 4.0 Hz, $J_{C5\text{-}C6ax}$ = 12.0 Hz), 2.10 (s, 3H, OAc), 2.08 (s, 3H, OAc), 2.04 (s, 3H, OAc), 2.10 - 1.80 (m, 4H, $H_{C3ax}$, $H_{C3eq}$, $H_{C6eq}$, $H_{C6ax}$)
$^{13}$C NMR (100 MHz, $CDCl_3$): δ = 174.9 ($C_9$); 174.6 ($C_9$); 170.8 (C=O (AcO)); 170.0 (C=O (AcO)); 169.6 (C=O (AcO)); 95.9 ($C_1$); 75.3 ($C_{C1}$); 71.7 ($C_{C2}$); 71.0 ($C_3$); 69.6 ($C_7$); 69.2 ($C_5$); 66.3 ($C_4$); 62.6 ($C_6$); 61.9 ($C_2$); 52.2 ($C_{10}$); 43.2 ($C_8$); 39.2 ($C_{C5}$); 39.0 ($C_{C4}$); 28.0, 26.8 ($C_{C3}$, $C_{C6}$); 20.9 (Me (AcO)); 20.8 (Me (AcO)); 20.7 (Me (AcO))
MS (ESI): m/z calculated for $[C_{24}H34ClN_3O_{13}Na]^+$: 630.17, found: 630.51

**$^{1}$H NMR spectrum of 3.11 in $CDCl_3$ (400 MHz)**




**$^{13}$C NMR spectrum of 3.11 in $CDCl_3$ (100 MHz)**




General conditions for CuAAC reactions

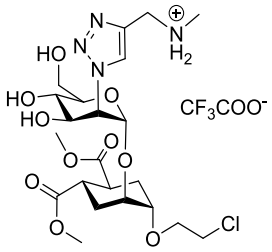
1M $CuSO_4$*$5H_2O$ and Na-ascorbate solutions were prepared in degassed water. The alkyne (1 equiv.) was dissolved in degassed THF and 0.1 equiv. $CuSO_4$*$5H_2O$ solution and 0.4 equiv. Na-ascorbate solutions were added, under nitrogen atmosphere. The azide was also dissolved in degassed THF and added to the mixture. The reaction was stirred under nitrogen, then, upon completion, QuadraSil™ MP metal scavenger was added to remove the copper, the mixture was filtered and concentrated *in vacuo*. If the purity of the protected product was satisfactory, the crude was used directly in the following Zemplén-deprotection step.

General conditions for Zemplén-deacetylation

1 equiv. Acetylated compound was dissolved in distilled MeOH and 1M freshly prepared NaOMe in MeOH was added to the solution (1.5 mol equiv NaOMe) to 0.1M final concentration of the substrate. After completion, the reaction was neutralized with Amberlite® IR120 hydrogen form ion-exchange resin, filtered and concentrated *in vacuo*. The crude was purified by direct (DCM : MeOH) or reverse phase flash chromatography (water: MeOH/MeCN, for propargyl amine derivatives + 0.01% TFA was added), yielding the pure product. Compounds **3.12**, **3.13**, **3.14** and **3.29** were isolated as ammonium trifluoroacetate salts.

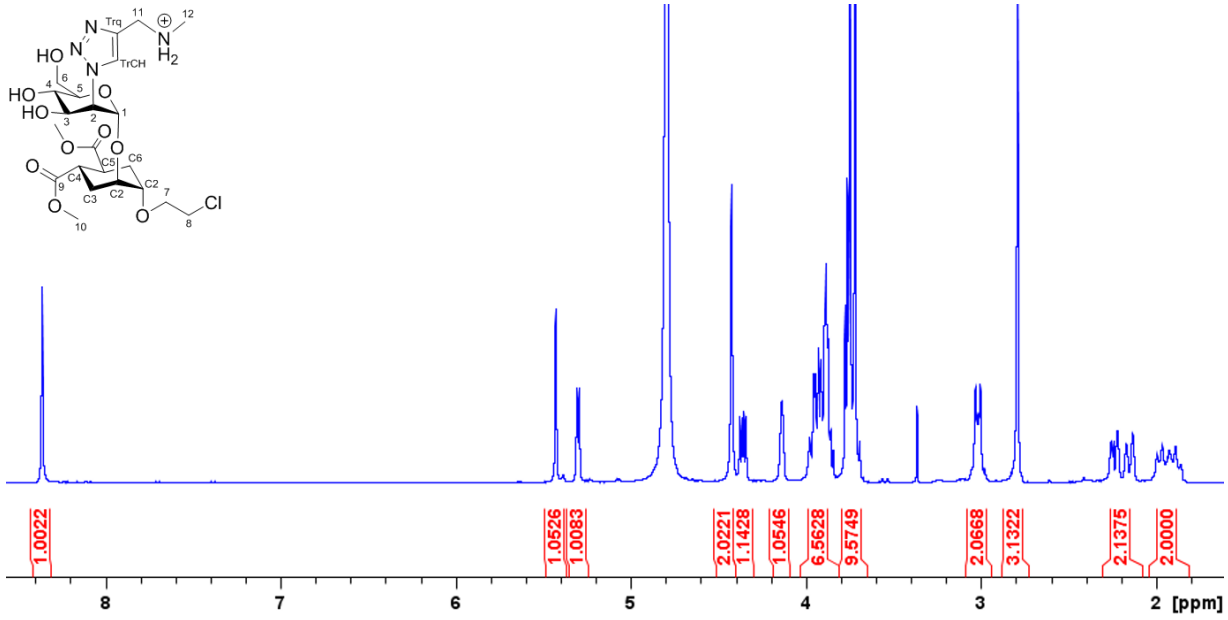
### 3.4.2 Characterization of C-2 modified ligands

Characterization of **3.12**

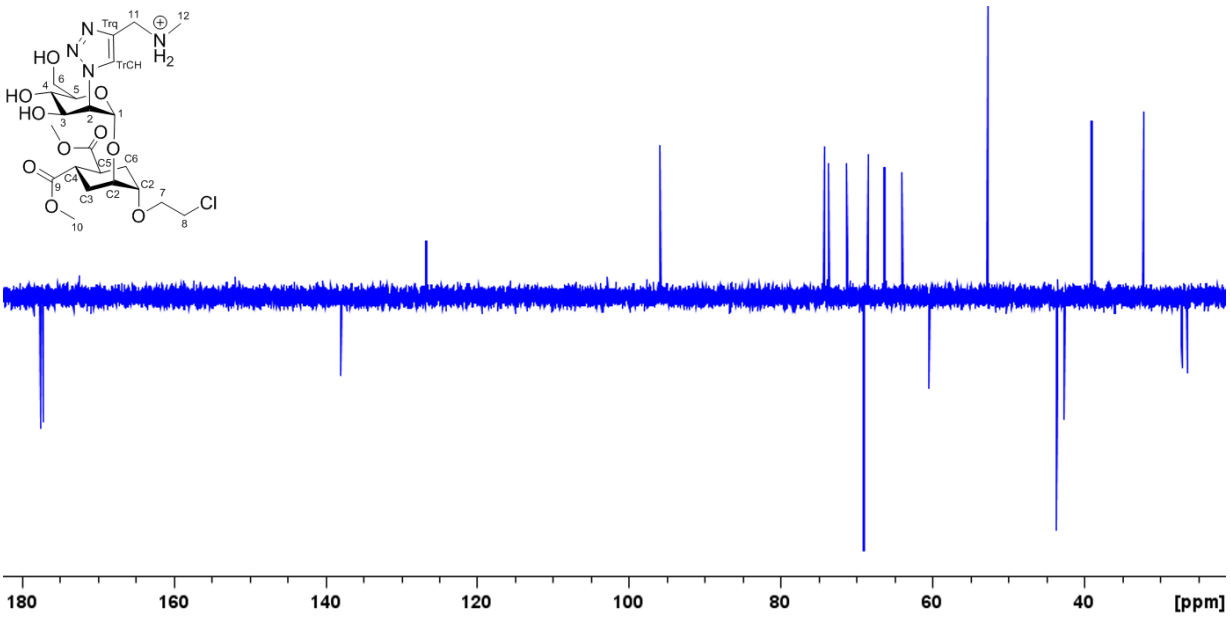



Yield: 75% (over two steps)
$R_f$: 0.06 in water : MeOH = 1 : 1 + 0.01% TFA
$^1$H NMR (400 MHz, $D_2O$): δ = 8.36 (s, 1H, $H_{TrCH}$), 5.42 (s, 1H, $H_1$), 5.30 (d, 1H, $H_2$, $J_{1\text{-}2}$ = 4.9 Hz), 4.42 (s, 2H, $H_{11}$), 4.36 (dd, 1H, $H_3$, $J_{3\text{-}4}$ = 9.5 Hz, $J$ = 4.9 Hz), 4.17 - 4.11 (m, 1H, $C_2$), 4.00 - 3.83 (m, 6H, $H_5$, $H_{6a,b}$, $H_{7a,b}$, $C_1$), 3.78 - 3.75 (m, 2H, $H_{8a,b}$), 3.75 (s, 3H, $H_{10}$), 3.72 (s, 3H, $H_{10}$), 3.75 - 3.72 (m, 1H, $H_4$), 3.06 - 2.96 (m, 2H, $C_4$, $C_5$), 2.79 (bs, 3H, NMe), 2.28 – 2.10 (m, 2H, $C_{3ax,6ax}$), 2.03 – 1.84 (m, 2H, $C_{3ax,6ax}$)
$^{13}$C NMR (100 MHz, $D_2O$): δ = 177.5 ($C_9$); 177.2 ($C_9$); 138.0 ($C_{Trq}$); 126.8 ($C_{TrCH}$); 96.0 ($C_1$); 74.3 ($C_{C1}$ or $C_5$); 73.8 ($C_5$ or $C_{C1}$); 71.4 ($C_{C2}$); 69.1 ($C_7$); 68.6 ($C_3$); 66.4 ($C_4$); 64.0 ($C_2$); 60.5 ($C_6$); 52.8 ($C_{10}$); 43.7 ($C_8$); 42.7 ($C_{11}$); 39.1 ($C_{C4}$, $C_{C5}$); 34.0 ($C_{NMe}$); 27.2 ($C_6$); 26.4 ($C_3$)

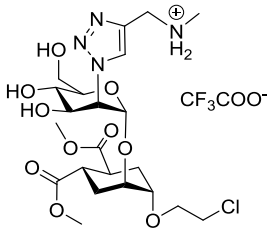
**$^{1}$H NMR spectrum of 3.12 in $D_2O$ (400 MHz)**




**$^{13}$C NMR spectrum of 3.12 in $D_2O$ (100 MHz)**



Characterization of **3.13**



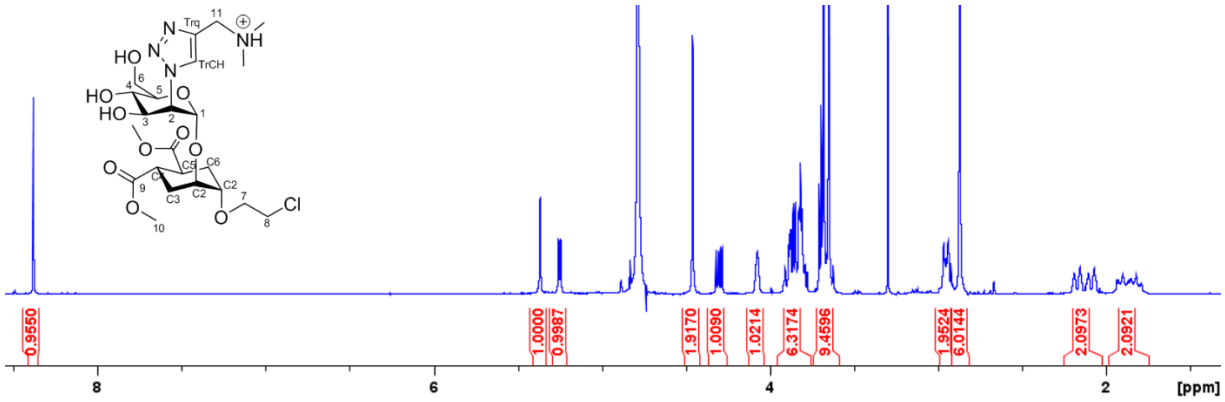
Yield: 84% (over two steps)

$R_f$: 0.1 in water : MeOH = 1 : 1 + 0.01% TFA

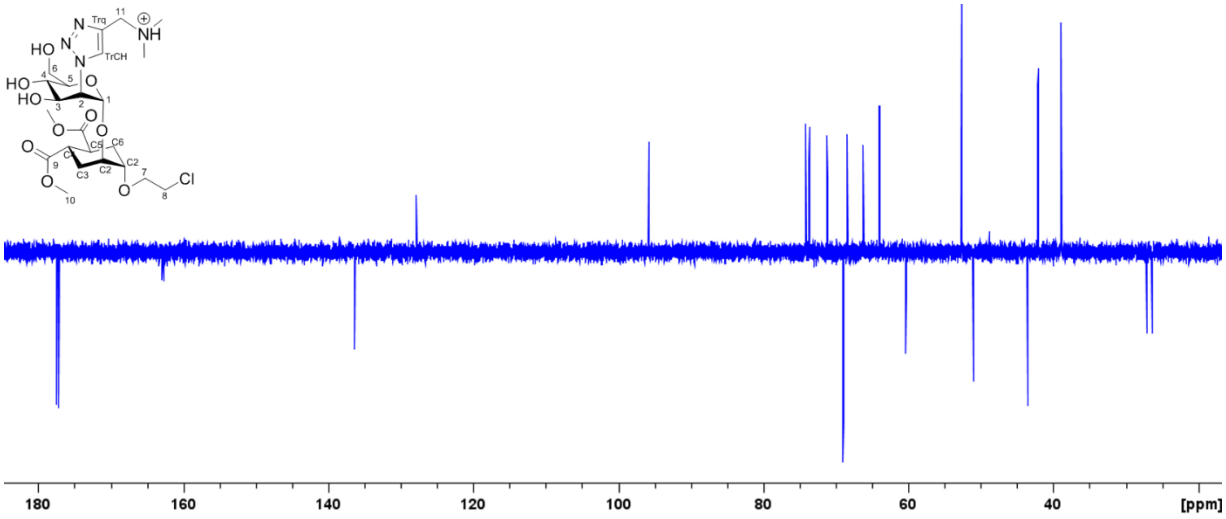
$^1$H NMR (400 MHz, $D_2O$): δ = 8.38 (s, 1H, $H_{TrCH}$), 5.36 (s, 1H, $H_1$), 5.25 (d, 1H, $H_2$, $J_{1\text{-}2}$ = 5.1 Hz, $J_{2\text{-}3}$ = 1.2 Hz), 4.46 (s, 2H, $H_{11}$), 4.33 (dd, 1H, $H_3$, $J_{3\text{-}4}$ = 9.5 Hz, $J$ = 5.0 Hz), 4.10 - 4.05 (m, 1H, $C_2$), 3.92 - 3.77 (m, 6H, $H_5$, $H_{6a,b}$, $H_{7a,b}$, $C_1$), 3.71 - 3.69 (m, 2H, $H_{8a,b}$), 3.68 (s, 3H, $H_{10}$), 3.65 (s, 3H, $H_{10}$), 3.68 - 3.65 (m, 1H, $H_4$), 3.00 - 2.90 (m, 2H, $C_4$, $C_5$), 2.79 (app d, 6H, $NMe_2$), 2.23 – 2.03 (m, 2H, $C_{3ax,6ax}$), 1.97 – 1.76 (m, 2H, $C_{3ax,6ax}$)

$^{13}$C NMR (100 MHz, $D_2O$): δ = 177.5 ($C_9$); 177.2 ($C_9$); 136.4 ($C_{Trq}$); 127.9 ($C_{TrCH}$); 95.9 ($C_1$); 74.2 ($C_{C1}$ or $C_5$); 73.7 ($C_5$ or $C_{C1}$); 71.2 ($C_{C2}$); 69.0 ($C_7$); 68.4 ($C_3$); 66.3 ($C_4$); 64.0 ($C_2$); 60.3 ($C_6$); 52.7 ($C_{10}$); 51.1 ($C_8$), 43.6 ($C_{11}$); 42.1 ($C_{C4}$, $C_{C5}$); 39.0 ($C_{NMe2}$); 27.2 ($C_6$); 26.4 ($C_3$)

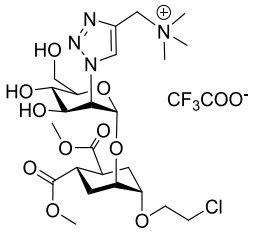
**$^1$H NMR spectrum of 3.13 in $D_2O$ (400 MHz)**



**$^{13}$C NMR spectrum of 3.13 in $D_2O$ (100 MHz)**



Characterization of **3.14**



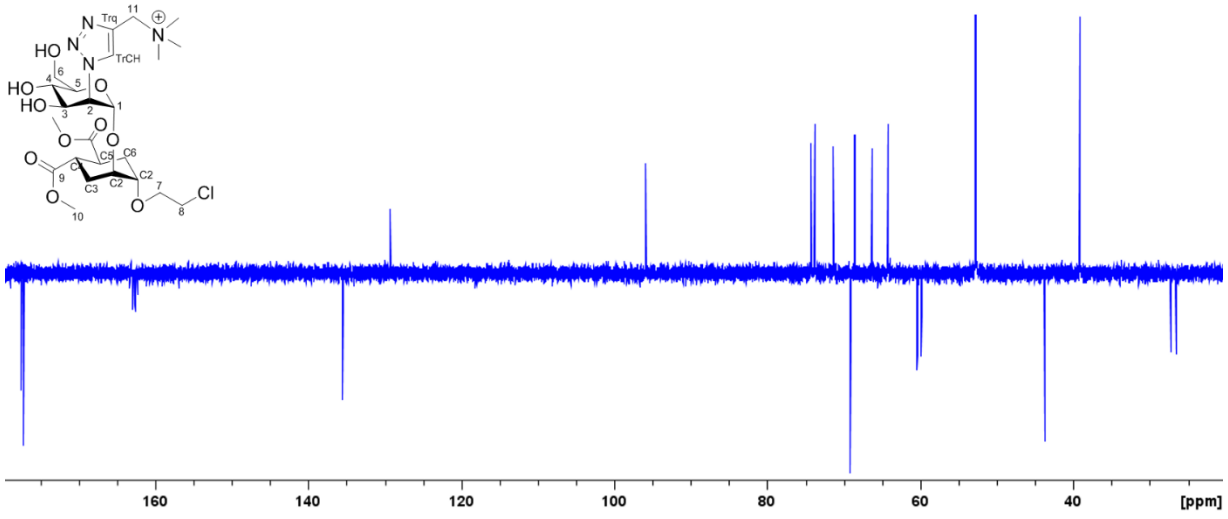
Yield: 53% (over two steps)

$R_f$: 0.26 in water : MeOH = 1 : 1 + 0.01% TFA

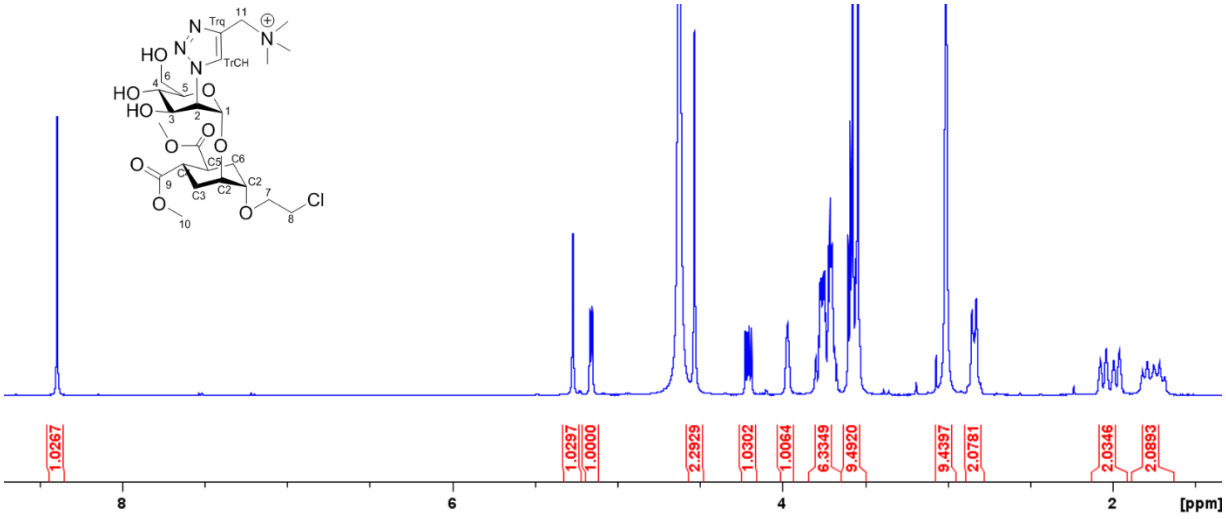
$^1$H NMR (400 MHz, $CD_3OD$): δ = 8.08 (s, 1H, $H_{TrCH}$), 7.52 – 7.27 (m, 5H, $H_{12}$, $H_{13}$, $H_{14}$), 5.27 (s, 1H, $H_1$), 5.26 (d, 1H, $H_2$, $J_{1\text{-}2}$ = 5.1 Hz, $J_{2\text{-}3}$ = 1.0 Hz), 4.53 (s, 2H, $H_{11}$), 4.21 (dd, 1H, $H_3$, $J_{3\text{-}4}$ = 9.6 Hz, $J$ = 5.3 Hz), 3.99 – 3.95 (m, 1H, $C_2$), 3.81 – 3.66 (m, 6H, $H_5$, $H_{6a,b}$, $H_{7a,b}$, $C_1$), 3.61 - 3.59 (m, 2H, $H_{8a,b}$), 3.58 (s, 3H, $H_{10}$), 3.55 (s, 3H, $H_{10}$), 3.58 – 3.55 (m, 1H, $H_4$), 3.01 (s, 9H, $NMe_3$), 3.00 - 2.90 (m, 2H, $C_4$, $C_5$), 2.01 – 1.94 (m, 2H, $C_{3ax,6ax}$), 1.85 – 1.67 (m, 2H, $C_{3ax,6ax}$)

$^{13}$C NMR (100 MHz, $CD_3OD$): δ = 177.5 ($C_9$); 177.2 ($C_9$); 135.5 ($C_{Trq}$); 129.3 ($C_{TrCH}$); 96.0 ($C_1$); 74.4 ($C_{C1}$ or $C_5$); 73.9 ($C_5$ or $C_{C1}$); 71.4 ($C_{C2}$); 69.2 ($C_7$); 68.6 ($C_3$); 66.4 ($C_4$); 64.3 ($C_2$); 60.4 ($C_6$); 59.9 ($C_{11}$); 52.8 ($C_8$); 52.7 ($C_{NMe3}$); 43.8 ($C_{10}$); 39.3 ($C_{C4}$, $C_{C5}$); 27.2 ($C_6$); 26.4 ($C_3$)

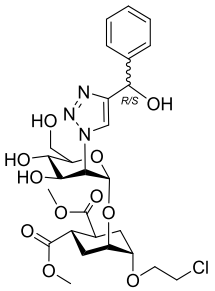
**$^1$H NMR spectrum of 3.13 in $CD_3OD$ (400 MHz)**



**$^{13}$C NMR spectrum of 3.13 in $CD_3OD$ (100 MHz)**




**Characterization of 3.20a and 3.20b**



Yield: 65% (over two steps)

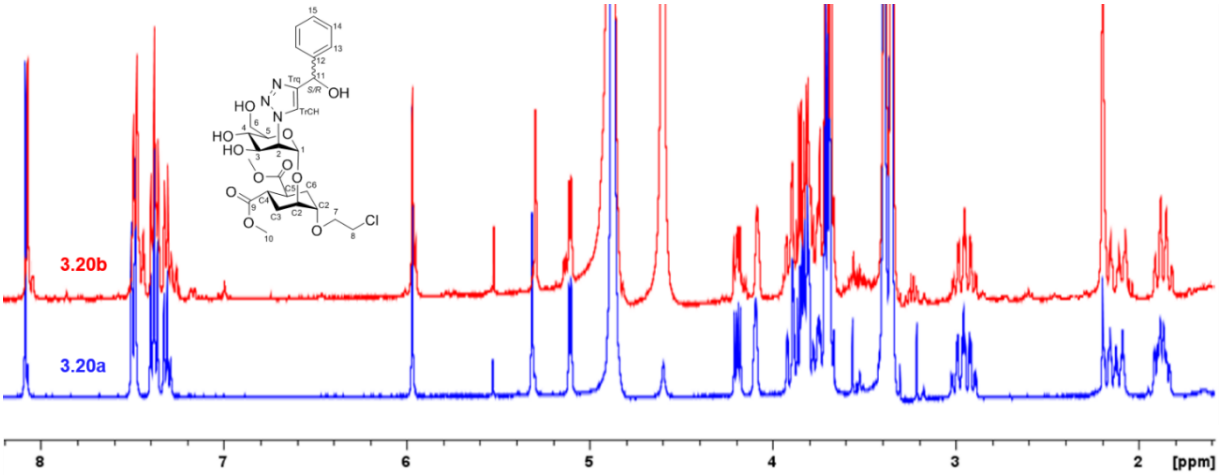
$R_f$ (**3.20.a**): 0.26; $R_f$ (**3.20.b**): 0.22 in DCM : MeOH = 94 : 6

$^1$H NMR (400 MHz, $CD_3OD$): δ = 8.08 (s, 1H, $H_{TrCH}$, (**3.20a**)), 8.06 (s, 1H, $H_{TrCH}$, (**3.20b**)), 5.87 (s, 1H, $H_{11}$, (**3.20a**)), 5.86 (s, 1H, $H_{11}$ , (**3.20b**)), 5.31 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.4 Hz, (**3.20a**)), 5.29 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.4 Hz(**3.20b**)),

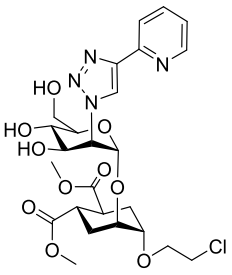
5.10 (dd, 1H, $H_2$, $J_{1\text{-}2}$ = 1.4 Hz, $J_{2\text{-}3}$ = 5.0 Hz), 4.19 (dd, 1H, $H_3$, $J_{3\text{-}4}$ = 9.0 Hz, $J$ = 5.0 Hz), 4.11 - 4.06 (m, 1H, $C_2$), 3.93 – 3.72 (m, 9H, $H_4$, $H_5$, $H_{6a,b}$, $H_{7a,b}$, $C_1$, $H_{8a,b}$), 3.71 (s, 3H, $H_{10}$), 3.68 (s, 3H, $H_{10}$), 3.03 - 2.87 (m, 2H, $C_4$, $C_5$), 2.21 – 2.06 (m, 2H, $C_{3ax,6ax}$), 1.93 – 1.81 (m, 2H, $C_{3ax,6ax}$)

$^{13}$C NMR (100 MHz, $CD_3OD$): δ = 177.5 ($C_9$); 177.2 ($C_9$); 138.0 ($C_{Trq}$); 130.4 ($C_{13}$); 129.7 ($C_{15}$); 128.7 ($C_{14}$); 126.8 ($C_{TrCH}$); 99.3 ($C_1$); 77.0 ($C_{C1}$ or $C_5$); 76.7 ($C_5$ or $C_{C1}$); 74.1 ($C_{C2}$); 71.7 ($C_7$); 71.4 ($C_3$); 71.2 ($C_{11}$); 69.0 ($C_4$); 66.6 ($C_2$); 63.2 ($C_6$); 53.4 ($C_{10}$); 45.3 ($C_8$); 41.4, 41.2 ($C_{C4}$, $C_{C5}$); 30.1 ($C_6$); 29.2 ($C_3$)

**$^1$H NMR spectrum of 3.12 in $CD_3OD$ (400 MHz)**




Characterization of 3.23



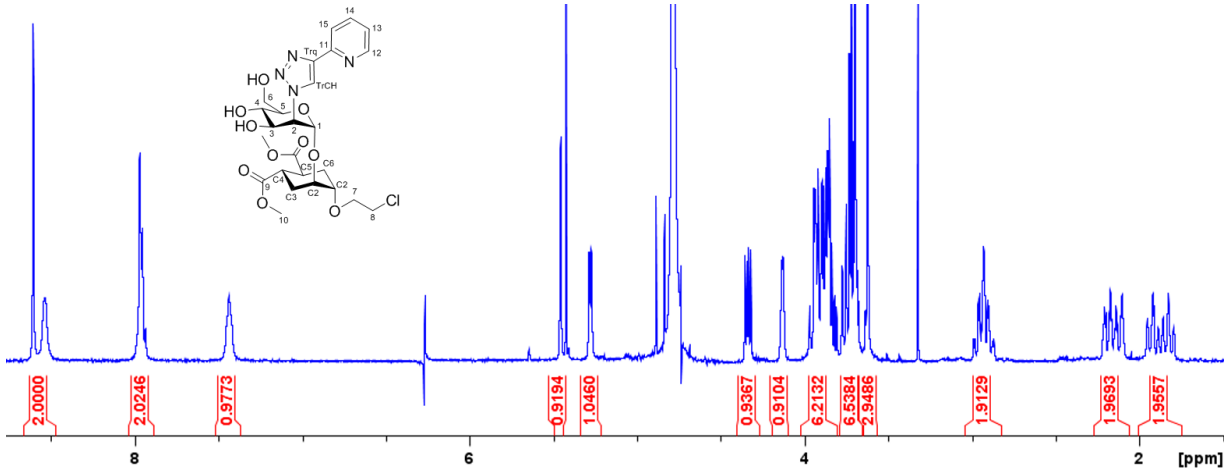
Yield: 77% (over two steps)

$R_f$: 0.11 in water : MeOH = 1 : 1

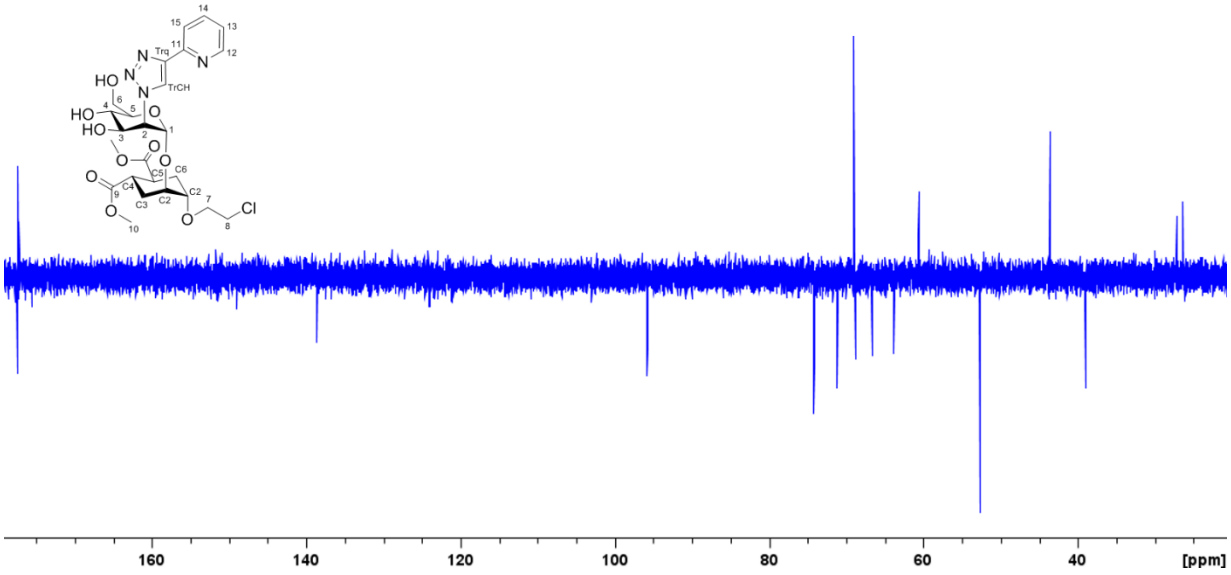
$^1$H NMR (400 MHz, $D_2O$): δ = 8.61 (s, 1H, $H_{TrCH}$), 8.51 (br s, 1H, $H_{12}$), 8.01 – 7.92 (m, 2H, $H_{13}$, $H_{15}$), 7.47 – 7.39 (m, 1H, $H_{14}$), 5.46 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.7 Hz), 5.28 (dd, 1H, $H_2$, $J_{1\text{-}2}$ = 1.7 Hz, $J_{2\text{-}3}$ = 5.1 Hz), 4.34 (dd, 1H, $H_3$, $J_{3\text{-}4}$ = 9.1 Hz, $J$ = 5.1 Hz), 4.15 - 4.11 (m, 1H, $C_2$), 3.98 - 3.79 (m, 6H, $H_5$, $H_{6a,b}$, $H_{7a,b}$, $C_1$), 3.79 - 3.70 (m, 3H, $H_{8ab}$, $H_4$), 3.70 (s, 3H, $H_{10}$), 3.62 (s, 3H, $H_{10}$), 3.01 - 2.86 (m, 2H, $C_4$, $C_5$), 2.24 – 2.08 (m, 2H, $C_{3ax,6ax}$), 1.98 – 1.78 (m, 2H, $C_{3ax,6ax}$)

$^{13}$C NMR (100 MHz, $D_2O$): δ = 177.4 ($C_9$); 177.2 ($C_9$); 149.0 ($C_{12}$); 138.7 ($C_{Trq}$); 138.7 ($C_{15}$); 124.0 ($C_{13}$); 123.6 ($C_{TrCH}$); 121.0 ($C_{14}$); 95.8 ($C_1$); 74.2 ($C_{C1}$ or $C_5$); 74.1 ($C_5$ or $C_{C1}$); 71.2 ($C_{C2}$); 69.0 ($C_7$); 68.8 ($C_3$); 66.7 ($C_4$); 63.9 ($C_2$); 60.6 ($C_6$); 52.6 ($C_{10}$); 43.6 ($C_8$); 39.0, 38.9 ($C_{C4}$, $C_{C5}$); 27.1 ($C_6$); 26.4 ($C_3$)

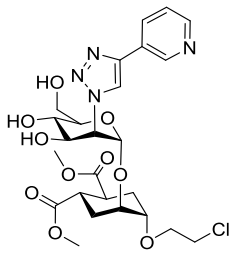
**$^{1}$H NMR spectrum of 3.12 in $D_2O$ (400 MHz)**




**$^{13}$C NMR spectrum of 3.12 in $D_2O$ (100 MHz)**




Characterization of 3.24



Yield: 45% (over two steps)

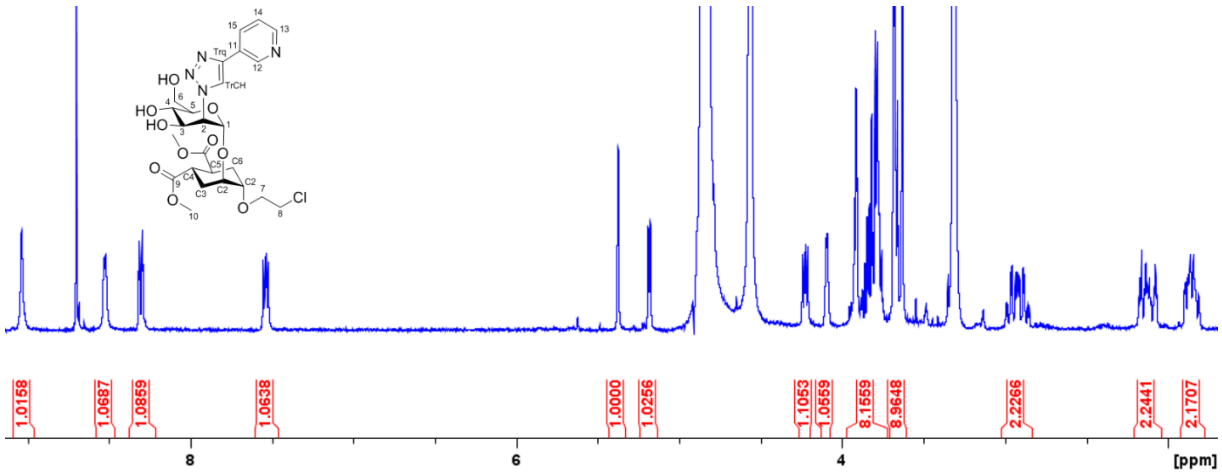
$R_f$: 0.19 in water : MeOH = 1 : 1

$^{1}$H NMR (400 MHz, $CD_3OD$): δ = 9.03 (br s, 1H, $H_{12}$), 8.70 (s, 1H, $H_{TrCH}$), 8.52 (d, 1H, $H_{13}$, $J_{13\text{-}14}$ = 4.3 Hz), 8.30 (dt, 1H, $H_{15}$, $J$ = 1.8 Hz, $J$ = 8.0 Hz), 7.53 (dd, 1H, $H_{14}$, $J$ = 4.8 Hz, $J$ = 8.0 Hz), 5.37 (d, 1H, $H_1$, $J_{1\text{-}2}$ = 1.7 Hz), 5.18 (dd, 1H, $H_2$, $J_{1\text{-}2}$ = 1.8 Hz, $J_{2\text{-}3}$ = 5.0 Hz), 4.22 (dd, 1H, $H_3$, $J_{3\text{-}4}$ = 8.5 Hz, $J$ = 5.0 Hz), 4.11 - 4.07 (m, 1H, $C_2$),

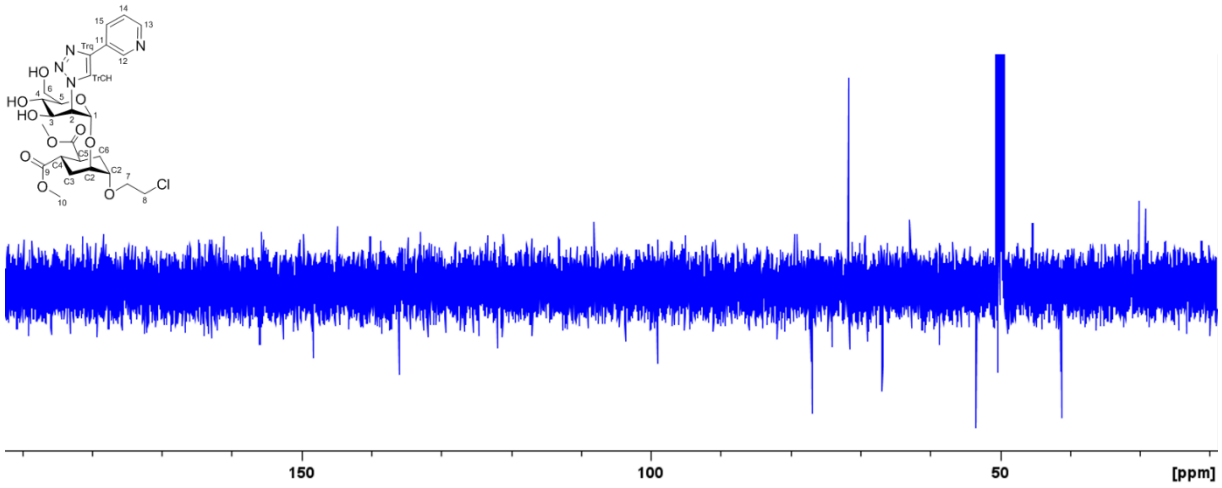
3.96 - 3.73 (m, 6H, $H_5$, $H_{6a,b}$, $H_{7a,b}$, $C_1$), 3.69 - 3.68 (m, 1H, $H_4$), 3.68 (s, 3H, $H_{10}$), 3.68 - 3.64 (m, 2H, $H_{8a,b}$), 3.63 (s, 3H, $H_{10}$), 3.00 - 2.84 (m, 2H, $C_4$, $C_5$), 2.19 – 2.06 (m, 2H, $C_{3ax,6ax}$), 1.91 – 1.79 (m, 2H, $C_{3ax,6ax}$)

$^{13}$C NMR (100 MHz, $CD_3OD$): δ = 177.4 ($C_9$); 177.2 ($C_9$); 148.2 ($C_{12}$); 138.7 ($C_{Trq}$); 136.1 ($C_{13}$); 123.6 ($C_{TrCH}$); 122.0 ($C_{15}$); 121.2 ($C_{14}$); 99.1 ($C_1$); 77.0 ($C_{C1}$ or $C_5$); 76.9 ($C_5$ or $C_{C1}$); 71.7 ($C_{C2}$); 71.5 ($C_7$); 68.9 ($C_3$); 66.8 ($C_4$); 62.9 ($C_2$); 58.6 ($C_6$); 53.4 ($C_{10}$); 45.3 ($C_8$); 41.3, 41.2 ($C_{C4}$, $C_{C5}$); 30.1 ($C_6$); 29.1 ($C_3$)

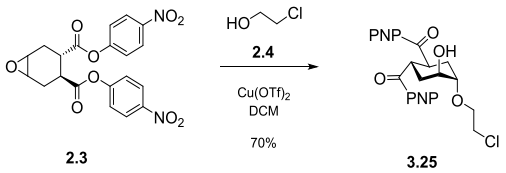
**$^1$H NMR spectrum of 3.12 in $CD_3OD$ (400 MHz)**




**$^{13}$C NMR spectrum of 3.12 in $CD_3OD$ (100 MHz)**



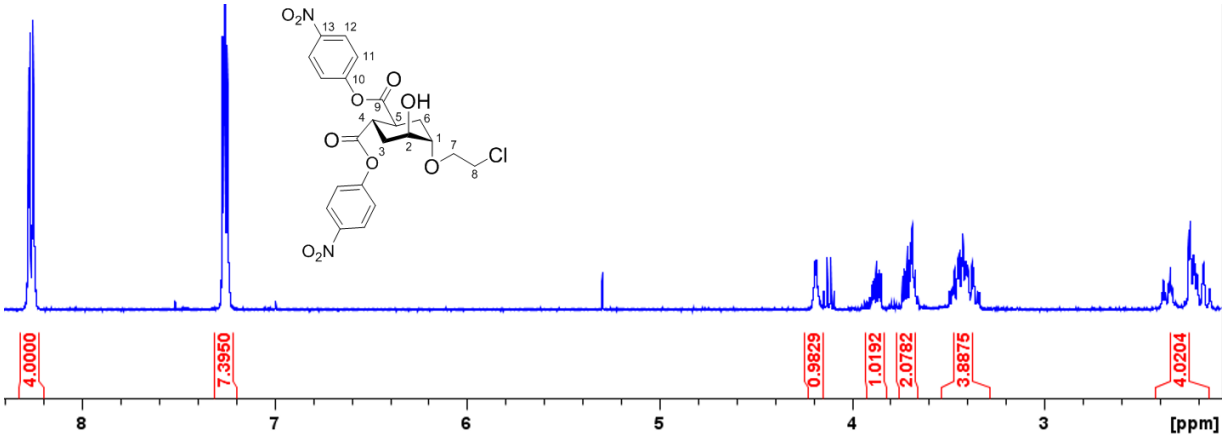
1,2-Cyclohexanedicarboxylic acid, (1S,2S,4S,5S)-5-(2-chloroethoxy)-4-hydroxy-1,2-bis-p-nitrophenyl ester **(3.25)**



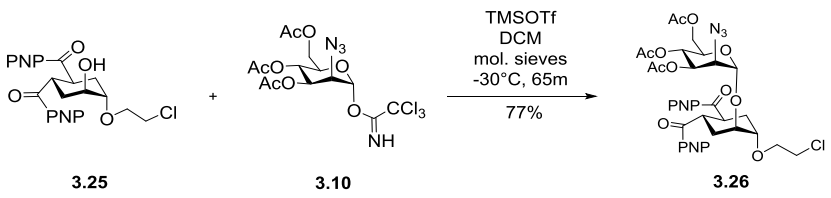

Epoxide **2.3** (1.2 g, mmol, 2.8014 mmol, 1 equiv.) and $Cu(OTf)_2$ (253 mg, 0.7 mmol, 0.25 equiv.) were added to 2-chloroethan-1-ol (30 ml) and the mixture was stirred under nitrogen at room temperature. After completion (16 h), the solvents were removed under reduced pressure and the crude was purified by flash chromatography (Hex with gradient of EtOAc from 20 to 50%) to yield 80% of pure product **3.25** as a colorless wax.

Yield: 80%
$R_f$: 0.4 in Hex : EtOAc = 1 : 1
$^1$H NMR (400 MHz, $CDCl_3$): δ = 8.30 – 8.24 (m, 4H; $H_{12}$), 7.29 – 7.23 (m, 4H; $H_{11}$), 4.21 – 4.16 (m, 1H; $H_2$), 3.91 – 3.84 (m, 1H; $H_{7a}$), 3.75 – 3.66 (m, 2H; $H_{7b}$, $H_1$), 3.50 – 3.33 (m, 4H; $H_4$, $H_5$, $H_{8a,b}$), 2.40 – 2.13 (m, 4H; $H_3$, $H_6$)

**$^1$H NMR spectrum of 3.25 in $CDCl_3$ (400 MHz)**




1,2-Cyclohexanedicarboxylic acid, (1S,2S,4S,5S)-4-(2-chloroethoxy)-5-[(3,4,6-tri-O-benzoyl-2-azido-2-deoxy-α-D-mannopyranosyl)oxy]-1,2-bis(p-nitrophenylester) **(3.26)**

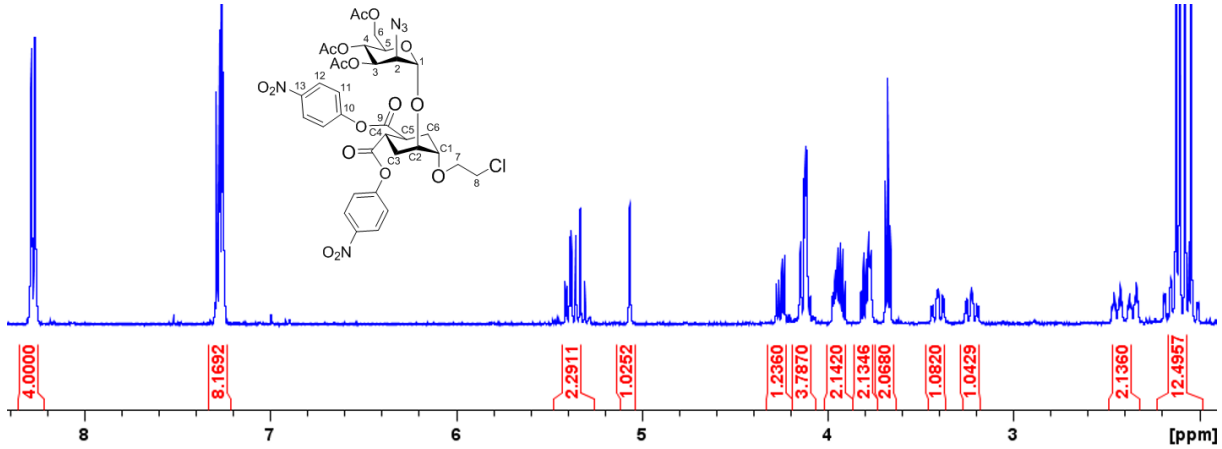



A mixture of the acceptor **3.25** (79 mg, 0.1549 mmol, 1 equiv.) and the donor **3.10** (74 mg, 0.1549 mmol, 1 equiv.) was co-evaporated with toluene three times. Powdered and activated 4Å molecular sieves (acid washed) were added. The mixture was kept under vacuum for a few hours and then dissolved in dry DCM (2 ml). After

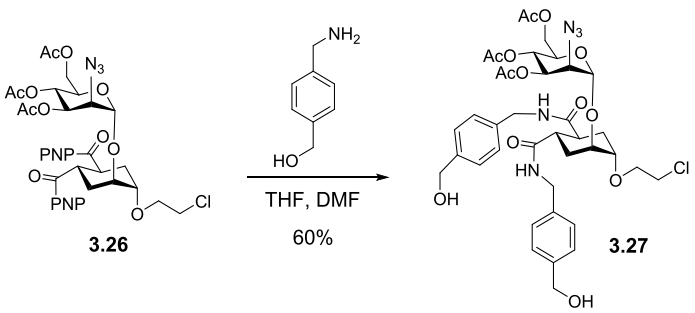
cooling to -30°C, TMSOTf (5.6 µl, 0.0310 mmol, 0.2 equiv.) was added to the stirred mixture. The solution was stirred at -30°C for 1 hour and upon completion, the reaction was quenched with TEA. The mixture was warmed to room temperature and filtered over a celite pad. The filtrate was evaporated at reduced pressure and the crude product was purified by flash chromatography (toluene with a gradient of EtOAc from 0 to 10%) to yield the pure product **3.26** as a white foam.

Yield: 82%
$R_f$: 0.25 in Hex : EtOAc = 6 : 4
$^1$H NMR (400 MHz, $CDCl_3$): δ = 8.30 – 8.25 (m, 4H; $H_{12}$), 7.30 –7.24 (m, 4H; $H_{11}$), 5.42 – 5.27 (m, 2H; $H_3$, $H_4$), 5.07 (d, 1H; $H_1$, $J_{1\text{-}2}$ = 1.6 Hz), 4.25 (dd, 1H; $H_{6b}$, $J_{6b\text{-}5}$ = 5.6 Hz, $J_{6a\text{-}6b}$ = 12.3 Hz), 4.16 – 4.09 (m, 4H, $H_2$, $H_{6b}$, $C_1$, $C_2$), 3.98 – 3.90 (m, 1H; $H_{7a,b}$), 3.83 – 3.75 (m, 1H, $H_5$), 3.68 (t, 2H, $H_{8a,b}$, $J$ = 5.5 Hz), 2.45 – 3.67 (m, 1H; $C_4$ or $C_5$), 3.27 – 3.18 (m, 1H; $C_5$ or $C_4$), 2.48 – 2.32 (m, 4H; $C_{3eq}$, $C_{6eq}$), 2.20 – 2.13 (m, 4H; $C_{3ax}$, $C_{6ax}$), 2.12 (s, 3H, OAc), 2.11 (s, 3H, OAc), 2.08 (s, 3H, OAc)

**$^1$H NMR spectrum of 2.11 in $CDCl_3$ (400 MHz)**




Synthesis of **3.27**



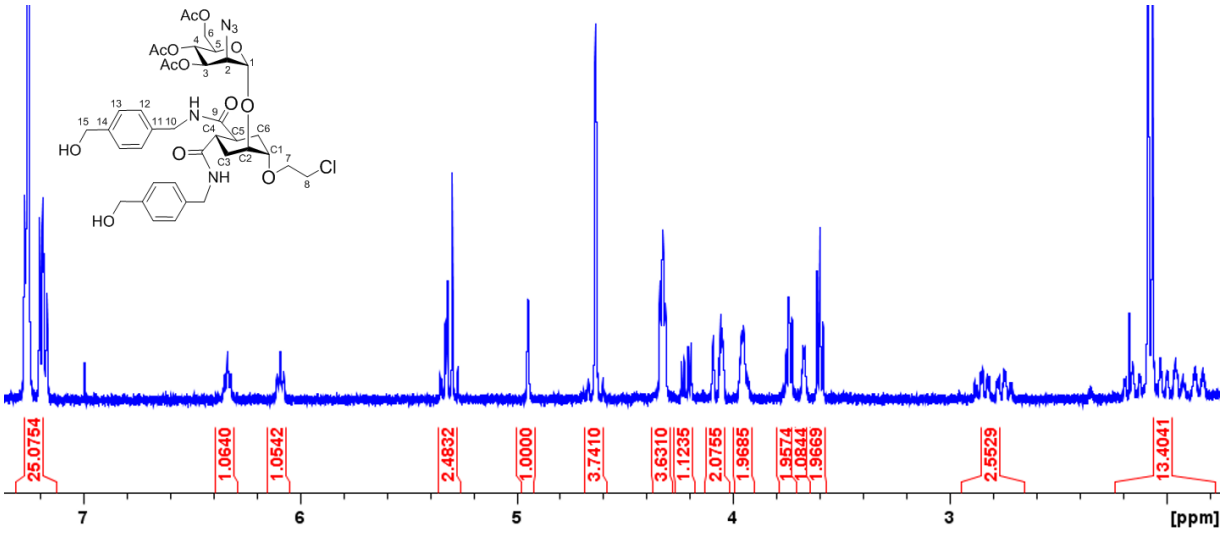

The PNP-ester **3.26** (104 mg, 0.1263 mmol, 1 equiv.) was dissolved in dry DMF (150 µl) and diluted with dry THF (1 ml) and (4-hydroxymethylene)benzylamine (52 mg, 0.38 mmol, 3 equiv.) was added. The reaction was stirred at RT overnight, then, upon completion, the solvent was evaporated. The crude product was purified by flash chromatography (DCM with a gradient of MeOH from 0 to 10%) to yield the pure product **3.26** as a white foam.

Yield: 86%

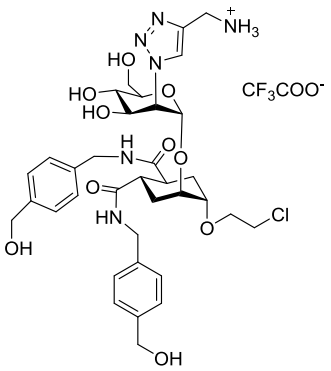
$R_f$: 0.13 in DCM : MeOH = 97 : 3

$^1$H NMR (400 MHz, $CDCl_3$): δ = 7.29 – 7.15 (m, 8H; $H_{12}$ $H_{13}$), 6.34 (m, 1H, NH), 6.09 (m, 1H; NH), 5.36 – 5.27 (m, 2H; $H_3$, $H_4$), 4.95 (d, 1H; $H_1$, $J_{1\text{-}2}$ = 1.6 Hz), 4.68 – 4.60 (m, 4H, $H_{15}$), 4.35 – 4.30 (m, 4H, $H_{10}$), 4.21 (dd, 1H; $H_{6b}$, $J_{6b\text{-}5}$ = 5.1 Hz, $J_{6a\text{-}6b}$ = 10.2 Hz), 4.11 – 4.03 (m, 2H, $H_2$, $C_2$), 3.98 – 3.92 (m, 2H; $H_{6b}$, $C_1$), 3.77 – 3.72 (m, 1H; $H_{7a,b}$), 3.69 – 3.65 (m, 1H, $H_5$), 3.60 (t, 2H, $H_{8a,b}$, $J$ = 5.6 Hz), 2.89 – 2.70 (m, 2H; $C_4$, $C_5$), 2.20 – 1.81 (m, 4H; $C_3$, $C_6$)

**$^1$H NMR spectrum of 2.11 in $CDCl_3$ (400 MHz)**




Characterization of **3.29**



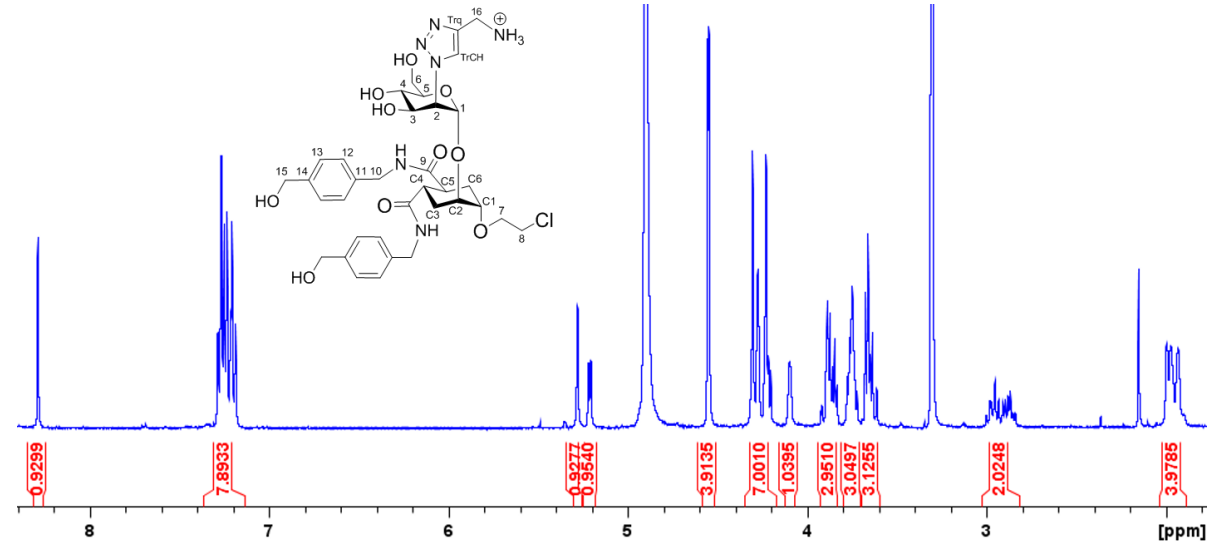

Yield: 76% (over two steps)

$R_f$: 0.38 in water : MeOH = 1 : 1 + 0.01% TFA

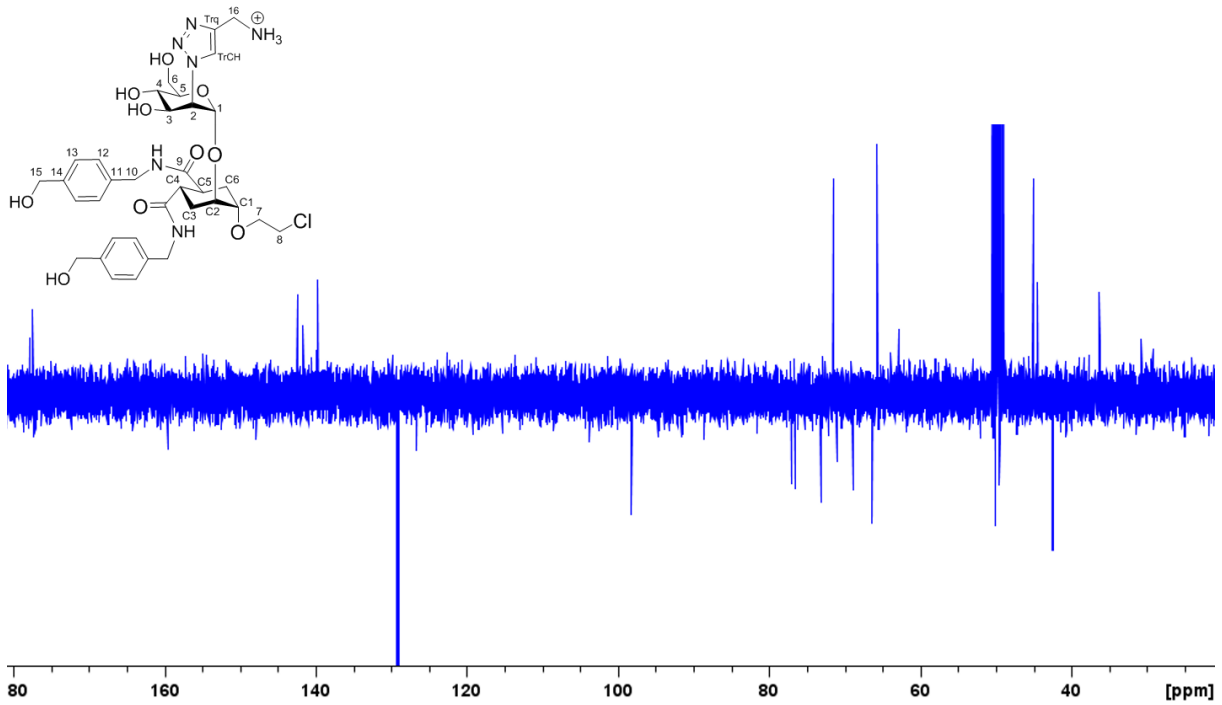
$^1$H NMR (400 MHz, $CD_3OD$): δ = 8.29 (s, 1H, $H_{TrCH}$), 7.30 – 7.18 (m, 8H, $H_{12}$, $H_{13}$), 5.28 (s, 1H, $H_1$), 5.22 (dd, 1H, $H_2$, $J_{1\text{-}2}$ = 1.0 Hz, $J_{2\text{-}3}$ = 5.1 Hz), 4.55 (d, 4H, $H_{10}$, $J$ = 2.6 Hz), 4.32 – 4-20 (m, 7H, $H_3$, $H_{15}$, $H_{16}$), 4.12 – 4.08 (m, 1H, $H_2$), 3.92 – 3.82 (m, 3H, $H_{6a,b}$, $H_{7a}$), 3.79 – 3.72 (m, 3H, $H_5$, $H_{7b}$, $C_1$), 3.69 – 3.60 (m, 3H, $H_{8a,b}$, $H_4$), 3.01 – 2.83 (m, 2H, $C_4$, $C_5$), 2.03 – 1.89 (m, 4H, $C_3$, $C_6$)

$^{13}$C NMR (100 MHz, $D_2O$): δ = 177.9 ($C_9$); 177.6 ($C_9$); 142.5, 142.4 ($C_{14}$); 141.7, 140.7 ($C_{13}$); 139.8 ($C_{Trq}$); 129.3, 129.2 ($C_{12}$); 126.8 ($C_{TrCH}$); 98.3 ($C_1$); 77.1 ($C_{C1}$ or $C_5$); 76.7 ($C_5$ or $C_{C1}$); 73.2 ($C_{C2}$); 71.6 ($C_7$); 71.1 ($C_3$); 69.0 ($C_4$); 65.8 ($C_{10}$);65.5 ($C_2$); 63.0 ($C_6$); 45.2 ($C_8$); 44.6, 44.5 ($C_{15}$); 42.5 ($C_{C4}$, $C_{C5}$); 36.3 ($C_{16}$); 30.8, 29.3 ($C_6$, $C_3$)

**$^{1}$H NMR spectrum of 3.29 in $CD_3OD$ (400 MHz)**




**$^{13}$C NMR spectrum of 3.29 in $CD_3OD$ (100 MHz)**



# 3.5 Appendix

## 3.5.1 Selectivity tests

SPR inhibition tests against DC-SIGN and langerin were performed by Silvia Achilli, IBS, Grenoble, France.

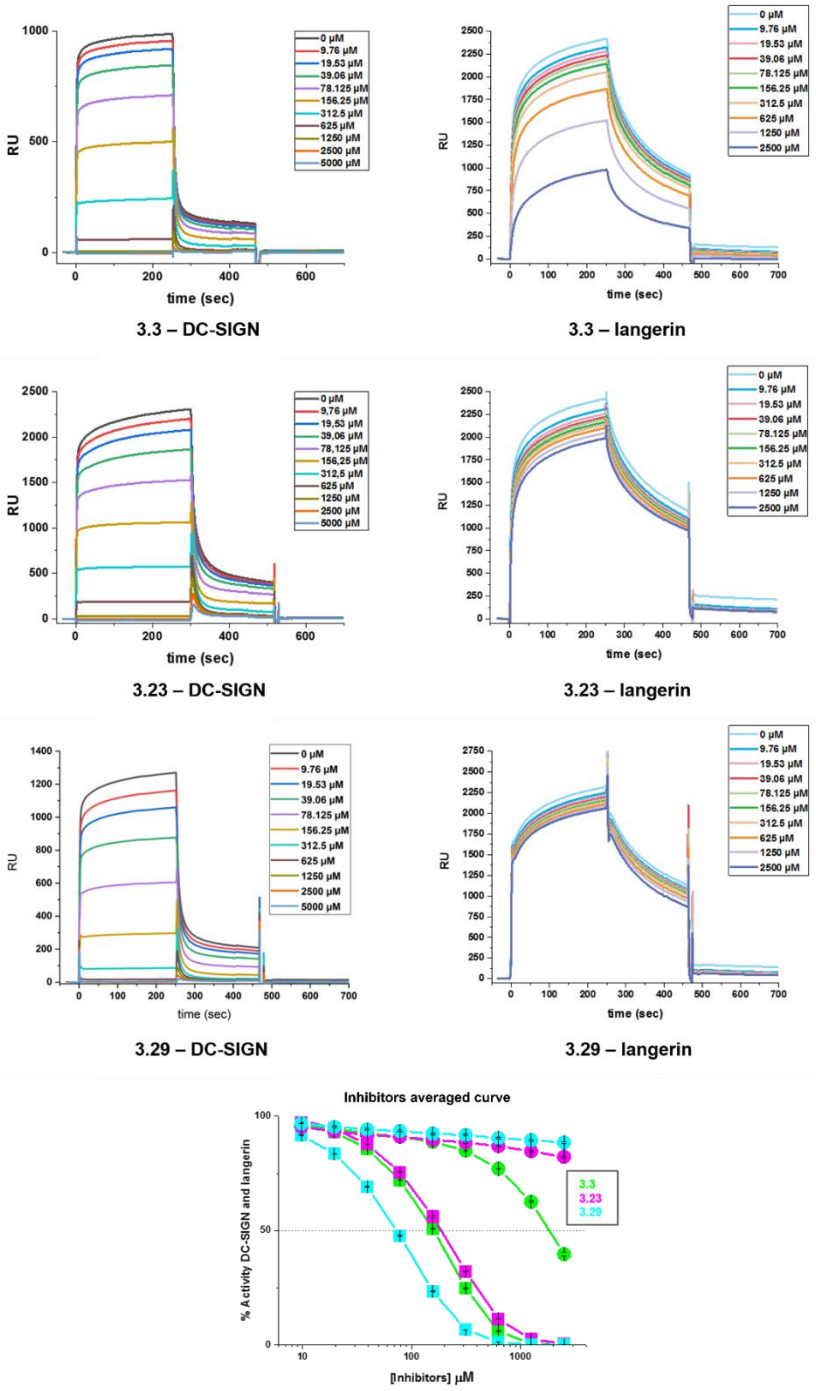



***Figure 88.*** *Sensorgrams of* ***3.3****,* ***3.23*** *and* ***3.29*** *obtained in SPR competition assays against DC-SIGN and langerin on a mannosylated-BSA surface*

### 3.5.2 ITC tests

Isothermal titration microcalorimetry (performed by Silvia Achilli, IBS, Grenoble, France)

Interaction between **3.29** and DC-SIGN ECD by ITC was additionally analysed by ITC. **3.29** (2.5 mM) was titrated into a lectin solution (100 μM). A one binding site model fitting of the data with an assumed stoichiometry value fixed to 1 yielded a $K_D$ of 52.08 ± 1.32 μM that was in agreement with the apparent affinity determined by SPR competition assay ($IC_{50}$: 76.25 ± 3.25 μM). A titration with **3.29** alone was performed to remove the glycomimetic's contribution of dilution in the $K_D$ determination. The experiment was repeated twice.

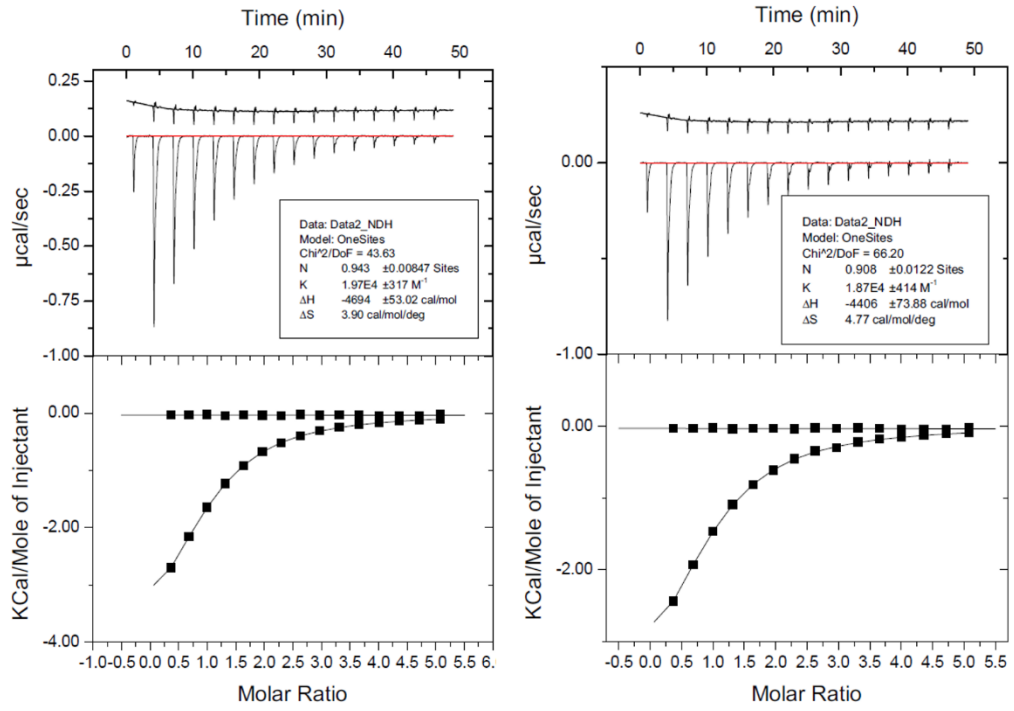



***Figure 89.*** *ITC analysis of interaction between* ***3.29*** *and DC-SIGN ECD. Titrations of* ***3.29*** *with 2.5 mM to DC-SIGN (100 μM). The upper panel shows the titration thermogram and the lower panel the data integration with fitted curves (1 : 1 binding model). Additionally, both panels contain also the result of* ***3.29*** *titration alone.*

### 3.5.3 Co-crystallization of 3.29 and DC-SIGN ECD

Co-crystallisation trials of DC-SIGN CRD and Man069 were performed by Dr. Michel Thépaut.

HTX crystallization platform (EMBL) was used to screen conditions of co-crystallization with the hanging-drop vapor-diffusion method at 293 K. The drop was composed of a protein/reservoir ratio of 1:1 with protein concentred at 5.54 mg/ml in 150 mM NaCl, 4 mM $CaCl_2$, 25 mM Tris pH ~ 8, 2% (v : v) DMSO buffer and 3.25 mM **3.29**. Among the crystallisation hits obtained, the condition F04 (200 mM $Mg(NO_3)_2$, 20% PEG 3350) from the kit PEGs-Suite Qiagen was chosen for manual optimisation screening with four different buffers (MES pH ~ 6, HEPES pH ~ 7, TRIS pH ~ 8 and Bicine pH 9), concentration of PEG 3350 (15 %-25%) and $Mg(NO_3)_2$ concentration (from 0.15 M to 0.2 M). Finally, the best crystals were obtained in the following condition: 20% PEG 3350, 200mM $Mg(NO_3)_2$, 100mM MES pH ~ 6.
Crystals were directly flash frozen in liquid nitrogen using Paratone-N as cryoprotectant. Data collection was performed at id30A-1 beamline (MASSIF-1), ESRF Grenoble. XDS/XSCALE (version 20180126) programs were used to process data resulting in a $P2_1$ space group, with unit cell parameters and a data set of 65455 reflections between 40 and 2.1Å resolution with 98.4% of completeness. Structure was solved by molecular replacement, using 1K9I DC-SIGN CRD structure and MOLREP (version 11.6.2).
Model building was made alternating refinement with REFMAC (version 5.8.0218) and manual construction with COOT (version 0.8.9 EL) resulting in a final structure with the following parameters. (**Table 4**Table 5)

| DC-SIGN CRD/**3.29** complex data collection and structure refinement statistics | |
|---|---|
| Data collection statistics | |
| Wavelength (Å) | 0.966 |
| Space group | $P2_1$ |
| Unit cell parameters (Å) | a=105.612, b=57.507, c=107.247 α=90, β=118.666, γ=90 |
| Resolution (Å) | 94.1-2.1 (2.2-2.1) |
| Unique reflections | 65455 (8427) |
| Completeness (%) | 98.4 (97.9) |
| $I/\sigma(I)$ | 8.47 (2.01) |
| $R_{merge}$[b] (%) | 9.9 (54.5) |
| Structure refinement statistics | |
| Resolution (Å) | 40-2.1 |
| Refinement factors | |
| Used reflections/free (%) | 62185/3261 |
| $R_{cryst}$[c] | 0.1759 |
| $R_{free}$[c] | 0.2336 |
| rmsd from ideality | |
| Bond angles (deg) | 1.6610 |
| Bond lengths (Å) | 0.0166 |
| Ramachandran plot (%) | |
| Most favoured regions | 97.8 |
| Additional allowed regions | 1.4 |
| Disallowed regions | 0.8 |
| Average B-factors (Å$^2$) | 33.681 |

***Table 5.*** *X-ray crystal structure parameters for* ***3.23*** *complexed with DC-SIGN CRD.*

# CONCLUSIONS AND OUTLOOK

The immunological importance of C-type lectins renders these receptors interesting therapeutic targets to shape the immune response. Until now, there are only a few pharmaceutically relevant ligands described, however - mainly as a result of the weak interactions between carbohydrate molecules and CLRs. On one hand, the receptors display a highly conserved carbohydrate recognition domain with a relatively flat and solvent-exposed surface that leads to challenges in finding high affinity and selective CLR ligands. On the other hand, carbohydrate molecules do not comply with the basic requirements of drug-likeness, as natural glycans are highly hydrophilic, flexible structures.
Therefore, two main strategies emerged as viable concepts to target CLRs: the construction of multivalent scaffolds can multiply the strength of weak individual binding events, while glycomimetic structures can exploit additional interactions, and improve their affinity towards the CLR.

Earlier results of our group have shown that modifications of a common glycomimetic disaccharide scaffold can both tune affinity towards certain CLRs and generate selectivity against others. We envisioned a glycomimetic library to be screened against various CLRs as a powerful tool for the identification of promising lead structures.
We took the earlier described bisamides (**1.21**) as the template to build this library, and diversified the structures by appending diverse amides on the pseudo-disaccharide core. As CLRs containing an EPN amino acid triplet often show preference for both mannose and fucose, we expanded the mannosylated library with β-fucosylated glycomimetic ligands.
The ligand library was printed on microarray chips, for a resource-efficient primary screening with CLRs. This primary screening can be used to select monovalent candidates for further studies or their presentation on multivalent constructs.
We used a doubly functionalized linker (BCN-amine) during the immobilization of ligands, in order to provide the solution-phase CLRs appropriate access to the surface-bound ligands. The printing setup was optimized and validated with the help of plant and fungal lectins of known specificity.
After establishing optimal screening conditions, we probed several human CLRs on the glycomimetic chip. Previous SPR results with DC-SIGN allowed us to critically evaluate the comparability of data obtained by the two, intrinsically different techniques, and, at the same time, confirm that the glycomimetic microarray screening method is indeed able to distinguish between better and poorer binders.
We obtained interesting results with dectin-2, a macrophage and dendritic cell-associated CLR, involved in bacterial and fungal infections. Some tertiary amide ligands, and additionally a few β-fucosylated structures were identified as possibly selective dectin-2 ligands which are not recognized by DC-SIGN/R and langerin. A preliminary study highlighted a structural element in the CRD of dectin-2 that may be accountable for the accommodation of tertiary amide substituents, uniquely among the other investigated CLRs. The interactions between dectin-2 and these ligands are further studied by other methods at the moment.

DC-SIGN is of particular therapeutic interest, due to its involvement in the uptake of numerous viral and bacterial pathogens (HIV, Dengue, HCV, Ebola, *M. tuberculosis*) and their *trans*-infection to T-cells. Inhibitors of the lectin bear a high potential as anti-microbial drugs or vaccine adjuvants, but the most potent monovalent inhibitors reach $IC_{50}$ values only the low micromolar range. Several glycomimetic ligands were designed to form additional interactions around the CRD, and thus enhance affinity, but druggability studies are constantly searching for new sites to exploit.

Upon binding, natural glycan ligands of DC-SIGN contact two main areas in the CRD. Mannose oligosaccharides are accommodated in the central region, around a phenylalanine residue (Phe313), while fucosylated molecules are known to lean more in the opposite direction, where Val351 can be found. Similarly, the pseudo-disaccharide ligand contacts Val351 which leaves the other side of the $Ca^{2+}$ binding site unoccupied.
A virtual fragment screening performed on the crystal complex of **1.18** in the CRD of DC-SIGN revealed a binding pocket lined with Ser360, Phe313 and Glu358 which could accommodate ammonium groups. Additionally, aromatic groups may forms beneficial interactions with the phenylalanine.
As the 2-OH group of the mannose residue in the glycomimetic molecules points directly towards the newly identified ammonium pocket, we expected that modification of the scaffold with fragment-like structures could reach the binding site.
From the two most straightforward synthetic strategies – amide coupling and CuAAC reactions in position C-2 – the first did not improve significantly the affinity compared to the parent pseudo-disaccharide **1.18**. The latter strategy, however, showed an almost order of magnitude affinity enhancement for the propargylamine derivative **3.3** ($IC_{50}$: 100 μM, compared to 700-900 μM).
After optimization of the synthetic route, a set of C-2 modified glycomimetic ligands was synthesized, in order to evaluate the effects of various structural elements. Methylation of the ammonium group derogated the binding affinity, and no improvement was shown with phenyl-propargyl alcohol derivatives either. The 2-picolylamine derivative **3.23**, however, exhibited similar affinity to **3.3**, suggesting an advantageous role of the aromatic ring.
In an attempt to combine the two glycomimetic modification strategies, ligand **3.29** was synthesized. The ligand's superior affinity ($IC_{50}$: 75 μM) compared to the individual modifications demonstrated a collective effect of the favourable interactions formed at different sites in the CRD.
The most important results came from the successful co-crystallization of **3.29** with the CRD of DC-SIGN. The structure verifies the position of the ammonium group in the computationally predicted binding site and the interactions with the three key residues. Additionally, the crystal structure also shows lipophilic interactions near Val351 – a contact that had earlier been indicated only in a computational model and STD-NMR studies. Evaluation of the results is underway.

The main achievements of this thesis work provide solid starting points for further steps. The glycomimetic microarray was designed to select potent and selective ligands, and did, indeed, indicate structural features that might be able to generate selectivity towards dectin-2 – one of the key receptors responsible for fungal infections caused by *Candida*. Multivalent constructs of the most promising ligands are being prepared, in order to assess the binding interactions of CLRs and glycomimetics (neoglycoproteins, gold-nanoparticles). The glycomimetic chip is available at CIC biomaGUNE, San Sebastián, Spain for new screenings with other (C-type) lectin receptors.
From **Part II**, **3.29** is among the highest affinity monovalent DC-SIGN inhibitors up to date, and due to the already incorporated linker moiety, multivalent systems can be functionalized with this ligand in one step (CuAAC). At the same time, the X-ray crystal structure provides a detailed insight to the molecular level contacts, and can be used to discover more about the nature of DC-SIGN's interactions with its ligands.

It is important to occasionally take a step back from our research and critically examine its purpose, viability and goals. Developing anti-adhesive drugs for pattern-recognition receptors of the immune-system - that would block, for example, HIV-uptake and *trans*-infection to T-cells - could be used in combination therapies with anti-retroviral drugs to minimize the viral load. Similarly, they could be applied in topical formulations, to minimize the risk of passing the virus on to others.

We know so little about the signalling pathways induced by CLRs upon ligand capture, however, that response evaluation and broad biological tests are inevitable before we can even consider CLR targeting drugs. Inhibiting the immune system's own receptors in order to prevent the uptake of certain pathogens carries the risk of exposing the individual to other harmful species, or cause deviations from the normal immunohomeostasis. These aspects must be thoroughly investigated during CLR drug discovery campaigns in the future.

At this point, however, we have only limited access to information about many aspects of CLR-ligand interactions. By exploring the molecular details of these binding events, or by discovering differences in ligand preference, we are taking steps towards the therapeutic exploitation of this important receptor family.

# Acknowledgements / Köszönetnyilvánítás

Mindenki más előtt Édesanyámnak szeretném kifejezni a köszönetemet, aki ezeket a sorokat soha nem fogja elolvasni.
Korábban, ha gondolkodtam is rajta, nem láttam át az életedet, és bár teljes egészében nem is fogom soha, most sokkal többet értek belőle. Jobban látom azt a rengeteg áldozatot, amit az én előrehaladásomért hoztál; azt, hogy mennyire összefonódott a sorsunk és sokkal jobban érzem annak a melegségnek, derűnek, együttérzésnek a hiányát, amit a körülötted lévők is gyakran éreztek és már nincs itt. Nem tudom, mennyiszer vettem fel a telefont, hogy elmondjam Neked, valami rettenetes ért, de ez a rettenetes éppen az, hogy már nem mondhatom ezt el. Időbe tellett, mire rájöttem, hogy belőlem is eltűnt valami.
Nem nagy dolog megemlítésre kerülni egy doktori dolgozat legvégén és nem sok marad egy ember után sem: néhány személyes tárgy, szomorúság, emlékek. Az én életem 2015 első felében kettétört, már csak távolról tudok tekinteni az azelőtti emlékeimre; de azt tudom, hogy jó tulajdonságaimat – ha vannak – Tőled kaptam, és azokat magamban hordozom most is.

I would like to say thank you to my supervisor, Prof. Anna Bernardi, who, apart from being a great scientist, is also a great person and a true teacher. I still do not know what I will appreciate more in my future life – the things I learned from her about chemistry and research; the chance to work on such a multidisciplinary project; or the role model she is to her students. We could all learn from her how to effortlessly motivate people; how to treat everyone with tactfulness, while at the same time also subtly delivering messages; and how to personally follow and promote the progress of everyone in her group.
I would like to express my gratitude to our collaborators of the Immunoshape Network - primarily to Dr. Niels Reichardt in CIC biomaGUNE (Spain) for hosting me in his laboratory and to Dr. Sonia Serna, whose guidance and help during this secondment was invaluable.
Thank you to the Fieschi Group at the IBS (France) for their significant contribution to the results described here and for the stimulating discussions during evaluation of the data.
I would like to acknowledge the Martín-Santamaría Group at CIB-CSIC (Spain) to propose a new direction in DC-SIGN inhibitors; and Prof. Bernd Lepenies and João Monteiro at the University of Veterinary Medicine (Germany) for the provided Fc-fusion proteins.
I am utterly grateful for all the help I received from Prof. Sara Sattin during these years; from the very first day when she helped me to get around Milano, until the last, stressful period of writing up my thesis. Additionally, I want to say thank you to Prof. Laura Belvisi, Prof. Donatella Potenza, Dr. Francesca Vasile, Dr. Monica Civera and Dr. Luca Pignataro at the University of Milan for being always at our disposal when we needed help.
Furthermore, thank you to past and present members of the Bernardi and Gennari Group who made me feel welcome when I arrived to a new, foreign environment. I believe that some of these friendships will last a lifetime and that is more than what I could have asked for.
I would like to mention the fellows of the Immunoshape Network who became genuine friends over the past three years, and thank them for the unforgettable and cheerful moments we spent together.

Meg szeretném köszönni a családomnak azt a rengeteg támogatást, amit Tőlük kaptam; nem csak ebben a három évben, de az ide vezető úton is. Soha nem fogom tudni viszonozni Géza türelmét és segítségét: a sok hajnali ébresztést, reptéri utat, utolsó pillanatokban szerkesztett ábrákat, előre beszerzett túró rudikat és mindent, amit már nagyon-nagyon régóta értem tesz. Nem vagyok a közelben, hogy a két nagymamámnak be tudjak segíteni most, amikor már igazán szükség lenne rá, pedig ők az első pillanatomtól kezdve ott voltak a számomra. Egy köszönöm nem fejezi ki, hogy mennyit jelent mindez, mennyi erőt merítettem a telefonhívásaikból és a találkozásainkból.

Testvéreimnek, Dorkának, Daninak és Zsófinak szívből kívánom, hogy olyan szakmát találjanak, amit annyira szeretnek majd, mint én ezt. Amikor úgy ébred reggel az ember, hogy siet a munkába, mert *érdekli*, akkor tudja, hogy jó úton jár.

Köszönöm a barátaimnak, hogy – bár távol vagyunk egymástól - továbbra is számíthatok rájuk és ugyanott vesszük fel a fonalat minden alkalommal, amikor találkozunk, mint amikor ez még naponta megtörtént.

I would like to say a special thank you to someone, without whom and whose support, help and motivation I could not have been able to get through the past years. The practical advices, constant encouragement and appreciation have been the pillars of my self-esteem in the darkest times and the happy and relaxed moments the reward after though periods. This continuous guidance helped to finish this thesiswork, *"step-by-step, bit-by-bit, stone-by-stone, NMR-by-NMR…"*

# References


[1] P. Guermonprez, J. Valladeau, L. Zitvogel, C. Théry, S. Amigorena, *Annu. Rev. Immunol.* **2002**, *20*, 621-667.

[2] OpenStax CNX, **2016**.

[3] H. Neumann, I. M. Medana, J. Bauer, H. Lassmann, *Trends Neurosci.* **2002**, *25*, 313-319.

[4] T. Kambayashi, T. M. Laufer, *Nat. Rev. Immunol.* **2014**, *14*, 719.

[5] Z. Wang, Z. S. Chinoy, S. G. Ambre, W. Peng, R. McBride, R. P. de Vries, J. Glushka, J. C. Paulson, G.-J. Boons, *Science* **2013**, *341*, 379.

[6] a) Y.-P. Cheon, C.-H. Kim, *Clin. Exp. Reprod. Med.* **2015**, *42*, 77-85; b) D. R. Tulsiani, H. Yoshida-Komiya, Y. Araki, *Biol. Reprod.* **1997**, *57*, 487-494; c) T. Eillen, G. Pascal, *Mol. Reprod. Dev.* **2015**, *82*, 635-650.

[7] a) T. Muramatsu, H. Muramatsu, *Glycoconj J.* **2004**, *21*, 41-45; b) M. E. Pedersen, G. Snieckute, K. Kagias, C. Nehammer, H. A. B. Multhaupt, J. R. Couchman, R. Pocock, *Science* **2013**, *341*, 1404.

[8] a) K. Ohtsubo, J. D. Marth, *Cell* **2006**, *126*, 855-867; b) J. W. Dennis, M. Granovsky, K. G. Warren, *BioEssays* **1999**, *21*, 412-421; c) S. W. Sophia, W. Bengt, T. L. O., *FEBS J.* **2014**, *281*, 46-62.

[9] a) E. Maverakis, K. Kim, M. Shimoda, M. E. Gershwin, F. Patel, R. Wilken, S. Raychaudhuri, L. R. Ruhaak, C. B. Lebrilla, *J. Autoimmun.* **2015**, *57*, 1-13; b) M. A. Wolfert, G.-J. Boons, *Nat. Chem. Biol.* **2013**, *9*, 776.

[10] a) R. Raman, S. Raguram, G. Venkataraman, J. C. Paulson, R. Sasisekharan, *Nat. Methods* **2005**, *2*, 817; b) J. C. Paulson, O. Blixt, B. E. Collins, *Nat. Chem. Biol.* **2006**, *2*, 238; c) H.-J. Gabius, S. André, J. Jiménez-Barbero, A. Romero, D. Solís, *Trends Biochem. Sci.* **2011**, *36*, 298-313; d) Richard D. Cummings, J. M. Pierce, *Chem. Biol.* **2014**, *21*, 1-15.

[11] a) E. J. Toone, *Curr. Opin. Struct. Biol.* **1994**, *4*, 719-728; b) Y. C. Lee, R. T. Lee, *Acc. Chem. Res.* **1995**, *28*, 321-327; c) T. K. Dam, C. F. Brewer, *Chem. Rev.* **2002**, *102*, 387-430.

[12] J. J. Lundquist, E. J. Toone, *Chem. Rev.* **2002**, *102*, 555-578.

[13] a) G. T. Noble, S. L. Flitsch, K. P. Liem, S. J. Webb, *Org. Biomol. Chem.* **2009**, *7*, 5245-5254; b) S. M. Dimick, S. C. Powell, S. A. McMahon, D. N. Moothoo, J. H. Naismith, E. J. Toone, *J. Am. Chem. Soc.* **1999**, *121*, 10286-10296.

[14] a) P. C. N. Rensen, L. A. J. M. Sliedregt, M. Ferns, E. Kieviet, S. M. W. van Rossenberg, S. H. van Leeuwen, T. J. C. van Berkel, E. A. L. Biessen, *J. Biol. Chem.* **2001**, *276*, 37577-37584; b) U. Boas, P. M. Heegaard, *Chem. Soc. Rev.* **2004**, *33*, 43-63.

[15] a) R. S. Haltiwanger, J. B. Lowe, *Annu. Rev. Biochem.* **2004**, *73*, 491-537; b) S. Reitsma, D. W. Slaaf, H. Vink, M. A. M. J. van Zandvoort, M. G. A. oude Egbrink, *Pflügers Arch.* **2007**, *454*, 345-359; c) A. H. Salmon, S. C. Satchell, *J. Pathol.* **2012**, *226*, 562-574.

[16] L. Ludwig, S. Sabine, T. Widmar, *Angew Chem. Int. Ed. Engl.* **2006**, *45*, 6802-6818.

[17] S. S. Pinho, C. A. Reis, *Nat. Rev. Cancer* **2015**, *15*, 540.

[18] S. Weinbaum, J. M. Tarbell, E. R. Damiano, *Annu. Rev. Biomed. Eng.* **2007**, *9*, 121-167.

[19] A. Martinez-Palomo, C. Braislovsky, W. Bernhard, *Cancer Res.* **1969**, *29*, 925.

[20] K. Amako, Y. Meno, A. Takade, *J. Bacteriol.* **1988**, *170*, 4960-4962.

[21] H. R. Gelderblom, E. H. S. Hausmann, M. Özel, G. Pauli, M. A. Koch, *Virology* **1987**, *156*, 171-176.

[22] A. Varki, R. D. Cummings, J. D. Esko, e. E. o. G. I. r. edition., *Cold Spring Harbor Laboratory Press* **2015-2017.**

[23] C. Abeijon, C. B. Hirschberg, *Trends Biochem. Sci.* **1992**, *17*, 32-36.

[24] G. Walsh, R. Jefferis, *Nat. Biotechnol.* **2006**, *24*, 1241.

[25] J. E. Scott, *FASEB J.* **1992**, *6*, 2639-2645.

[26] G. A. M. Cross, *Annu. Rev. Cell Biol.* **1990**, *6*, 1-39.

[27] D. Lingwood, K. Simons, *Science* **2010**, *327*, 46.

[28] T. Kinoshita, M. Fujita, Y. Maeda, *J. Biochem.* **2008**, *144*, 287-294.

[29] a) M. L. Huang, K. Godula, *Glycobiology* **2016**, *26*, 797-803; b) A. Restuccia, M. M. Fettis, G. A. Hudalla, *J. Mater. Chem. B* **2016**, *4*, 1569-1585.
[30] a) S. Cecioni, A. Imberty, S. Vidal, *Chem. Rev.* **2015**, *115*, 525-561; b) P. V. Murphy, S. André, H.-J. Gabius, *Molecules* **2013**, *18*, 4026-4053.
[31] C. Müller, G. Despras, T. K. Lindhorst, *Chem. Soc. Rev.* **2016**, *45*, 3275-3302.
[32] A. Ogura, A. Kurbangalieva, K. Tanaka, *Glycobiology* **2016**, *26*, 804-812.
[33] A. R. Aricescu, R. J. Owens, *Curr. Opin. Struct. Biol.* **2013**, *23*, 345-356.
[34] a) S. S. Ting, G. Chen, M. H. Stenzel, *Polym. Chem.* **2010**, *1*, 1392-1412; b) S. G. Spain, N. R. Cameron, *Polym. Chem.* **2011**, *2*, 60-68; c) Q. Zhang, J. Collins, A. Anastasaki, R. Wallis, D. A. Mitchell, C. R. Becer, D. M. Haddleton, *Angew Chem. Int. Ed. Engl.* **2013**, *125*, 4531-4535.
[35] Á. Martínez, C. O. Mellet, J. M. G. Fernández, *Chem. Soc. Rev.* **2013**, *42*, 4746-4773.
[36] S. André, C. Grandjean, F.-M. Gautier, S. Bernardi, F. Sansone, H.-J. Gabius, R. Ungaro, *ChemComm* **2011**, *47*, 6126-6128.
[37] S. P. Vincent, K. Buffet, I. Nierengarten, A. Imberty, J. F. Nierengarten, *Chem. Eur. J.* **2016**, *22*, 88-92.
[38] K. Barr, B. Kannan, E. Korchagina, I. Popova, I. Ryzhov, S. Henry, N. Bovin, *Biopolymers* **2016**, *105*, 787-794.
[39] J. Xie, N. Bogliotti, *Chem. Rev.* **2014**, *114*, 7678-7739.
[40] a) A. M. Eissa, A. Abdulkarim, G. J. Sharples, N. R. Cameron, *Biomacromolecules* **2016**, *17*, 2672-2679; b) L. Valtola, A. Rahikkala, J. Raula, E. I. Kauppinen, H. Tenhu, S. Hietala, *Eur. Polym. J.* **2014**, *59*, 282-289.
[41] D. Rabuka, M. B. Forstner, J. T. Groves, C. R. Bertozzi, *J. Am. Chem. Soc.* **2008**, *130*, 5947-5953.
[42] J. R. Kramer, B. Onoa, C. Bustamante, C. R. Bertozzi, *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 12574.
[43] O. Neth, D. L. Jack, A. W. Dodds, H. Holzel, N. J. Klein, M. W. Turner, *Infect. Immun.* **2000**, *68*, 688-693.
[44] H. Attrill, A. Imamura, R. S. Sharma, M. Kiso, P. R. Crocker, D. M. F. van Aalten, *J. Biol. Chem.* **2006**, *281*, 32774-32783.
[45] W. I. Weis, K. Drickamer, *Annu. Rev. Biochem.* **1996**, *65*, 441-473.
[46] G. R. Vasta, M. Nita-Lazar, B. Giomarelli, H. Ahmed, S. Du, M. Cammarata, N. Parrinello, M. A. Bianchet, L. M. Amzel, *Dev. Comp. Immunol.* **2011**, *35*, 1388-1399.
[47] B. Fiege, S. Rabbani, R. C. Preston, R. P. Jakob, P. Zihlmann, O. Schwardt, X. Jiang, T. Maier, B. Ernst, *ChemBioChem* **2015**, *16*, 1235-1246.
[48] T. K. Dam, C. F. Brewer, *Advances in Carbohydrate Chemistry and Biochemistry, Academic Press* **2010**, *63*, 139-164.
[49] a) P. H. v. Hippel, *Annu. Rev. Biophys.* **2007**, *36*, 79-105; b) T. K. Dam, T. A. Gerken, C. F. Brewer, *Biochemistry* **2009**, *48*, 3822-3827.
[50] C. F. Brewer, M. C. Miceli, L. G. Baum, *Curr. Opin. Struct. Biol.* **2002**, *12*, 616-623.
[51] N. Sharon, H. Lis, *Glycobiology* **2004**, *14*, 53R-62R.
[52] a) S. S. Komath, M. Kavitha, M. J. Swamy, *Org. Biomol. Chem.* **2006**, *4*, 973-988; b) E. d. S. Cândido, M. F. S. Pinto, P. B. Pelegrini, T. B. Lima, O. N. Silva, R. Pogue, M. F. Grossi-de-Sá, O. L. Franco, *FASEB J.* **2011**, *25*, 3290-3305.
[53] a) G. Vandenborre, G. Smagghe, E. J. M. Van Damme, *Phytochemistry* **2011**, *72*, 1538-1550; b) A. R. War, M. G. Paulraj, T. Ahmad, A. A. Buhroo, B. Hussain, S. Ignacimuthu, H. C. Sharma, *Plant Signal. Behav.* **2012**, *7*, 1306-1320.
[54] R. Loris, T. Hamelryck, J. Bouckaert, L. Wyns, *Biochim. Biophys. Acta - Protein Structure and Molecular Enzymology* **1998**, *1383*, 9-36.
[55] N. Sharon, H. Lis, *FASEB J.* **1990**, *4*, 3198-3208.
[56] J. Baenziger, D. Fiete, *J. Biol. Chem.* **1979**, *254*, 2400-2407.
[57] R. N. Knibbs, I. J. Goldstein, R. M. Ratcliffe, N. Shibuya, *J. Biol. Chem.* **1991**, *266*, 83-88.
[58] N. Shibuya, I. J. Goldstein, W. F. Broekaert, M. Nsimba-Lubaki, B. Peeters, W. J. Peumans, *J. Biol. Chem.* **1987**, *262*, 1596-1601.
[59] H. Rüdiger, H.-J. Gabius, *Glycoconj J.* **2001**, *18*, 589-613.
[60] F. Khan, R. H. Khan, A. Sherwani, S. Mohmood, M. A. Azfer, *Med. Sci. Monit.* **2002**, *8*, RA293-RA300.

[61] J. J. Skehel, D. C. Wiley, *Annu. Rev. Biochem.* **2000**, *69*, 531-569.
[62] L. M. Nilsson, W. E. Thomas, E. Trintchina, V. Vogel, E. V. Sokurenko, *J. Biol. Chem.* **2006**, *281*, 16656-16663.
[63] D. B. Williams, *J. Cell Sci.* **2006**, *119*, 615.
[64] Y. Oda, N. Hosokawa, I. Wada, K. Nagata, *Science* **2003**, *299*, 1394.
[65] a) Y. Kamiya, D. Kamiya, K. Yamamoto, B. Nyfeler, H.-P. Hauri, K. Kato, *J. Biol. Chem.* **2008**, *283*, 1857-1861; b) T. Anelli, R. Sitia, *EMBO J.* **2008**, *27*, 315.
[66] P. Ghosh, N. M. Dahms, S. Kornfeld, *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 202.
[67] S. J. Lee, S. Evers, D. Roeder, A. F. Parlow, J. Risteli, L. Risteli, Y. C. Lee, T. Feizi, H. Langen, M. C. Nussenzweig, *Science* **2002**, *295*, 1898.
[68] A. D. Luster, R. Alon, U. H. von Andrian, *Nat. Immunol.* **2005**, *6*, 1182.
[69] P. R. Crocker, J. C. Paulson, A. Varki, *Nat. Rev. Immunol.* **2007**, *7*, 255.
[70] a) T. L. O'keefe, G. T. Williams, F. D. Batista, M. S. Neuberger, *J. Exp. Med.* **1999**, *189*, 1307-1313; b) J. C. Poe, T. F. Tedder, *Trends Immunol.* **2012**, *33*, 413-420; c) P. Fiorina, A. Vergani, S. Dada, M. Jurewicz, M. Wong, K. Law, E. Wu, Z. Tian, R. Abdi, I. Guleria, *Diabetes* **2008**, *57*, 3013-3024.
[71] E. Ishikawa, T. Ishikawa, Y. S. Morita, K. Toyonaga, H. Yamada, O. Takeuchi, T. Kinoshita, S. Akira, Y. Yoshikai, S. Yamasaki, *J. Exp. Med.* **2009**, *206*, 2879.
[72] S. Mukherjee, H. Zheng, M. G. Derebe, K. M. Callenberg, C. L. Partch, D. Rollins, D. C. Propheter, J. Rizo, M. Grabe, Q.-X. Jiang, L. V. Hooper, *Nature* **2013**, *505*, 103.
[73] S. Pillai, I. A. Netravali, A. Cariappa, H. Mattoo, *Annu. Rev. Immunol.* **2012**, *30*, 357-392.
[74] F. T. Liu, G. A. Rabinovich, *Ann. N. Y. Acad. Sci.* **2010**, *1183*, 158-182.
[75] a) K. Drickamer, *Curr. Opin. Struct. Biol.* **1999**, *9*, 585-590; b) A. N. Zelensky, J. E. Gready, *FEBS J.* **2005**, *272*, 6179-6217.
[76] F. Yang, C. A. Bewley, J. M. Louis, K. R. Gustafson, M. R. Boyd, A. M. Gronenborn, G. M. Clore, A. Wlodawer, *J. Mol. Biol.* **1999**, *288*, 403-412.
[77] S. T. Iobst, K. Drickamer, *J. Biol. Chem.* **1994**, *269*, 15512-15519.
[78] T. Hatakeyama, T. Kamiya, M. Kusunoki, S. Nakamura-Tsuruta, J. Hirabayashi, S. Goda, H. Unno, *J. Biol. Chem.* **2011**, *286*, 10305-10315.
[79] A. Loukas, N. P. Mullin, K. K. A. Tetteh, L. Moens, R. M. Maizels, *Curr. Biol.* **1999**, *9*, 825-828.
[80] H. Feinberg, M. E. Taylor, N. Razi, R. McBride, Y. A. Knirel, S. A. Graham, K. Drickamer, W. I. Weis, *J. Mol. Biol.* **2011**, *405*, 1027-1039.
[81] S. A. F. Jégouzo, H. Feinberg, T. Dungarwalla, K. Drickamer, W. I. Weis, M. E. Taylor, *J. Biol. Chem.* **2015**, *290*, 16759-16771.
[82] M. Nagae, K. Yamanaka, S. Hanashima, A. Ikeda, K. Morita-Matsumoto, T. Satoh, N. Matsumoto, K. Yamamoto, Y. Yamaguchi, *J. Biol. Chem.* **2013**, *288*, 33598-33610.
[83] H. Feinberg, D. A. Mitchell, K. Drickamer, W. I. Weis, *Science* **2001**, *294*, 2163.
[84] M. Nagae, Y. Yamaguchi, *Curr. Opin. Struct. Biol.* **2015**, *34*, 108-115.
[85] H. Feinberg, R. Castelli, K. Drickamer, P. H. Seeberger, W. I. Weis, *J. Biol. Chem.* **2007**, *282*, 4202-4209.
[86] J. Angulo, I. Díaz, J. J. Reina, G. Tabarani, F. Fieschi, J. Rojo, P. M. Nieto, *ChemBioChem* **2008**, *9*, 2225-2227.
[87] F. Marcelo, F. Garcia-Martin, T. Matsushita, J. Sardinha, H. Coelho, A. Oude-Vrielink, C. Koller, S. André, E. J. Cabrita, H. J. Gabius, S. I. Nishimura, J. Jiménez-Barbero, F. J. Cañada, *Chem. Eur. J.* **2014**, *20*, 16147-16155.
[88] a) P. J. Coombs, S. A. Graham, K. Drickamer, M. E. Taylor, *J. Biol. Chem.* **2005**, *280*, 22993-22999; b) H. Feinberg, M. E. Taylor, W. I. Weis, *J. Biol. Chem.* **2007**, *282*, 17250-17258.
[89] a) E. Chabrol, A. Nurisso, A. Daina, E. Vassal-Stermann, M. Thepaut, E. Girard, R. R. Vivès, F. Fieschi, *PLoS One* **2012**, *7*, e50722; b) J. C. Muñoz-García, E. Chabrol, R. R. Vivès, A. Thomas, J. L. de Paz, J. Rojo, A. Imberty, F. Fieschi, P. M. Nieto, J. Angulo, *J. Am. Chem. Soc.* **2015**, *137*, 4100-4110.
[90] M. E. Taylor, K. Drickamer, *Curr. Opin. Struct. Biol.* **2014**, *28*, 14-22.
[91] N. Silva-Martín, Sergio G. Bartual, E. Ramírez-Aportela, P. Chacón, Chae G. Park, Juan A. Hermoso, *Structure* **2014**, *22*, 1595-1606.

[92] H. Feinberg, S. A. F. Jégouzo, T. J. W. Rowntree, Y. Guan, M. A. Brash, M. E. Taylor, W. I. Weis, K. Drickamer, *J. Biol. Chem.* **2013**, *288*, 28457-28465.
[93] K. Drickamer, M. E. Taylor, *Genome Biol.* **2002**, *3*, reviews1034.1031.
[94] K. Drickamer, M. E. Taylor, *Curr. Opin. Struct. Biol.* **2015**, *34*, 26-34.
[95] R. A. Ezekowitz, K. Sastry, P. Bailly, A. Warner, *J. Exp. Med.* **1990**, *172*, 1785.
[96] H. Feinberg, Y. Guo, D. A. Mitchell, K. Drickamer, W. I. Weis, *J. Biol. Chem.* **2005**, *280*, 1327-1335.
[97] H. Feinberg, A. S. Powlesland, M. E. Taylor, W. I. Weis, *J. Biol. Chem.* **2010**, *285*, 13285-13293.
[98] F. Teillet, B. Dublet, J.-P. Andrieu, C. Gaboriaud, G. J. Arlaud, N. M. Thielens, *J. Immunol.* **2005**, *174*, 2870.
[99] L.-L. Zhu, X.-Q. Zhao, C. Jiang, Y. You, X.-P. Chen, Y.-Y. Jiang, X.-M. Jia, X. Lin, *Immunity* **2013**, *39*, 324-334.
[100] E. Chabrol, M. Thépaut, C. Dezutter-Dambuyant, C. Vivès, J. Marcoux, R. Kahn, J. Valladeau-Guilemond, P. Vachette, D. Durand, F. Fieschi, *Biophys. J.* **2015**, *108*, 666-677.
[101] J. Valladeau, O. Ravel, C. Dezutter-Dambuyant, K. Moore, M. Kleijmeer, Y. Liu, V. Duvert-Frances, C. Vincent, D. Schmitt, J. Davoust, C. Caux, S. Lebecque, S. Saeland, *Immunity* **2000**, *12*, 71-81.
[102] A. Cambi, F. de Lange, N. M. van Maarseveen, M. Nijhuis, B. Joosten, E. M. H. P. van Dijk, B. I. de Bakker, J. A. M. Fransen, P. H. M. Bovee-Geurts, F. N. van Leeuwen, N. F. Van Hulst, C. G. Figdor, *J. Cell Biol.* **2004**, *164*, 145.
[103] M. J. Robinson, D. Sancho, E. C. Slack, S. LeibundGut-Landmann, C. R. e. Sousa, *Nat. Immunol.* **2006**, *7*, 1258.
[104] B. Ferwerda, G. Ferwerda, T. S. Plantinga, J. A. Willment, A. B. van Spriel, H. Venselaar, C. C. Elbers, M. D. Johnson, A. Cambi, C. Huysamen, *N. Engl. J. Med.* **2009**, *361*, 1760-1767.
[105] I. D. Iliev, V. A. Funari, K. D. Taylor, Q. Nguyen, C. N. Reyes, S. P. Strom, J. Brown, C. A. Becker, P. R. Fleshner, M. Dubinsky, J. I. Rotter, H. L. Wang, D. P. B. McGovern, G. D. Brown, D. M. Underhill, *Science* **2012**.
[106] N. Fujikado, S. Saijo, T. Yonezawa, K. Shimamori, A. Ishii, S. Sugai, H. Kotaki, K. Sudo, M. Nose, Y. Iwakura, *Nat. Med.* **2008**, *14*, 176.
[107] L. J. C., F. Line, E. Carina, B. Liselotte, R. Ulrica, G. J. Ping, S. Marina, D. Erik, S. Maria, B. A. J., A. Lars, K. Lars, P. Leonid, F. Sigbjörn, *Arthritis Rheumatol.* **2007**, *56*, 2620-2632.
[108] A. Dzionek, Y. Sohma, J. Nagafune, M. Cella, M. Colonna, F. Facchetti, G. Günther, I. Johnston, A. Lanzavecchia, T. Nagasaka, *J. Exp. Med.* **2001**, *194*, 1823-1834.
[109] N. A. Barrett, A. Maekawa, O. M. Rahman, K. F. Austen, Y. Kanaoka, *J. Immunol.* **2009**, *182*, 1119-1128.
[110] E. J. Lee, B. R. Brown, E. E. Vance, P. E. Snow, P. B. Silver, D. Heinrichs, X. Lin, Y. Iwakura, C. A. Wells, R. R. Caspi, H. L. Rosenzweig, *J. Immunol.* **2016**, *196*, 3148.
[111] a) A. Cambi, C. Figdor, *Curr. Biol.* **2009**, *19*, R375-R378; b) N. D. S. Rambaruth, M. V. Dwek, *Acta Histochem.* **2011**, *113*, 591-600; c) C. J. M. Melief, *Immunity* **2008**, *29*, 372-383; d) E. Rodríguez, S. T. T. Schetters, Y. van Kooyk, *Nat. Rev. Immunol.* **2018**, *18*, 204.
[112] M. N. Christiansen, J. Chik, L. Lee, M. Anugraham, J. L. Abrahams, N. H. Packer, *Proteomics* **2014**, *14*, 525-546.
[113] A. S. Carvalho, A. Harduin-Lepers, A. Magalhães, E. Machado, N. Mendes, L. T. Costa, R. Matthiesen, R. Almeida, J. Costa, C. A. Reis, *Int. J. Biochem. Cell Biol.* **2010**, *42*, 80-89.
[114] W. R. Alley, J. A. Vasseur, J. A. Goetz, M. Svoboda, B. F. Mann, D. E. Matei, N. Menning, A. Hussein, Y. Mechref, M. V. Novotny, *J. Proteome Res.* **2012**, *11*, 2282-2300.
[115] R. Saldova, Y. Fan, J. M. Fitzpatrick, R. W. G. Watson, P. M. Rudd, *Glycobiology* **2011**, *21*, 195-205.
[116] L. Muinelo-Romay, S. Villar-Portela, E. Cuevas, E. Gil-Martín, A. Fernández-Briera, *BMC Cancer* **2011**, *11*, 508.
[117] S. Zelenay, A. M. Keller, P. G. Whitney, B. U. Schraml, S. Deddouche, N. C. Rogers, O. Schulz, D. Sancho, C. Reis e Sousa, *J. Clin. Investig.* **2012**, *122*, 1615-1627.
[118] Y. Suzuki, Y. Nakano, K. Mishiro, T. Takagi, K. Tsuruma, M. Nakamura, S. Yoshimura, M. Shimazawa, H. Hara, *Sci. Rep.* **2013**, *3*, 3177.
[119] d. R. V. J. Carlos, B. F. J. III, B. A. F., A. O. F., B. M. Ross, d. R. V. J. Pablo, *J. Neurotrauma* **2015**, *32*, 228-236.

[120] M. Nonaka, B. Y. Ma, R. Murai, N. Nakamura, M. Baba, N. Kawasaki, K. Hodohara, S. Asano, T. Kawasaki, *J. Immunol.* **2008**, *180*, 3347.
[121] F. Xu, J. Liu, D. Liu, B. Liu, M. Wang, Z. Hu, X. Du, L. Tang, F. He, *Cancer Res.* **2014**, *74*, 3418.
[122] a) H. Irjala, K. Alanen, R. Grénman, P. Heikkilä, H. Joensuu, S. Jalkanen, *Cancer Res.* **2003**, *63*, 4671; b) Y. Zuo, S. Ren, M. Wang, B. Liu, J. Yang, X. Kuai, C. Lin, D. Zhao, L. Tang, F. He, *Gut* **2013**, *62*, 1169; c) H. Läubli, L. Borsig, *Semin. Cancer Biol.* **2010**, *20*, 169-177; d) D. Ding, Y. Yao, S. Zhang, C. Su, Y. Zhang, *Oncol. Lett.* **2017**, *13*, 13-21.
[123] S. Chiba, H. Ikushima, H. Ueki, H. Yanai, Y. Kimura, S. Hangai, J. Nishio, H. Negishi, T. Tamura, S. Saijo, Y. Iwakura, T. Taniguchi, *eLife* **2014**, *3*, e04177.
[124] a) K. Palucka, J. Banchereau, *Nat. Rev. Cancer* **2012**, *12*, 265; b) T. Yau, X. Dan, C. C. Ng, B. T. Ng, *Molecules* **2015**, *20*; c) D. Dinther, D. A. Stolk, R. Ven, Y. Kooyk, T. D. Gruijl, J. M. M. Haan, *J. Inherit. Metab. Dis.* **2017**, *102*, 1017-1034.
[125] a) J. Idoyaga, A. Lubkin, C. Fiorese, M. H. Lahoud, I. Caminschi, Y. Huang, A. Rodriguez, B. E. Clausen, C. G. Park, C. Trumpfheller, R. M. Steinman, *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 2384; b) K. Tewari, B. J. Flynn, S. B. Boscardin, K. Kastenmueller, A. M. Salazar, C. A. Anderson, V. Soundarapandian, A. Ahumada, T. Keler, S. L. Hoffman, M. C. Nussenzweig, R. M. Steinman, R. A. Seder, *Vaccine* **2010**, *28*, 7256-7266.
[126] P. J. Tacken, I. J. M. de Vries, R. Torensma, C. G. Figdor, *Nat. Rev. Immunol.* **2007**, *7*, 790.
[127] G. Schreibelt, K. F. Bol, H. Westdorp, F. Wimmers, E. H. J. G. Aarntzen, T. Duiveman-de Boer, M. W. M. M. van de Rakt, N. M. Scharenborg, A. J. de Boer, J. M. Pots, M. A. M. Olde Nordkamp, T. G. M. van Oorschot, J. Tel, G. Winkels, K. Petry, W. A. M. Blokx, M. M. van Rossum, M. E. B. Welzen, R. D. M. Mus, S. A. J. Croockewit, R. H. T. Koornstra, J. F. M. Jacobs, S. Kelderman, C. U. Blank, W. R. Gerritsen, C. J. A. Punt, C. G. Figdor, I. J. M. de Vries, *Clin. Cancer Res.* **2016**, *22*, 2155.
[128] W. Kastenmüller, K. Kastenmüller, C. Kurts, R. A. Seder, *Nat. Rev. Immunol.* **2014**, *14*, 705.
[129] M. M. Curtis, S. S. Way, *Immunology* **2009**, *126*, 177-185.
[130] S. Romagnani, *J. Allergy Clin. Immunol.* **2004**, *113*, 395-400.
[131] T. B. H. Geijtenbeek, S. I. Gringhuis, *Nat. Rev. Immunol.* **2009**, *9*, 465.
[132] S. I. Gringhuis, J. den Dunnen, M. Litjens, B. van het Hof, Y. van Kooyk, Teunis B. H. Geijtenbeek, *Immunity* **2007**, *26*, 605-616.
[133] a) N. Kanazawa, K. Tashiro, Y. Miyachi, *Immunobiology* **2004**, *209*, 179-190; b) W. Cao, L. Zhang, D. B. Rosen, L. Bover, G. Watanabe, M. Bao, L. L. Lanier, Y.-J. Liu, *PLOS Biol.* **2007**, *5*, e248.
[134] a) F. Meyer-Wentrup, D. Benitez-Ribas, P. J. Tacken, C. J. Punt, C. G. Figdor, I. J. M. de Vries, G. J. Adema, *Blood* **2008**, *111*, 4245-4253; b) K. Neumann, M. Castiñeiras-Vilariño, U. Höckendorf, N. Hannesschläger, S. Lemeer, D. Kupka, S. Meyermann, M. Lech, H.-J. Anders, B. Kuster, Dirk H. Busch, A. Gewies, R. Naumann, O. Groß, J. Ruland, *Immunity* **2014**, *40*, 389-399.
[135] G. D. Brown, *Nat. Rev. Immunol.* **2005**, *6*, 33.
[136] P. Kankkunen, L. Teirilä, J. Rintahaka, H. Alenius, H. Wolff, S. Matikainen, *J. Immunol.* **2010**, *184*, 6335.
[137] a) S. Yamasaki, E. Ishikawa, M. Sakuma, H. Hara, K. Ogata, T. Saito, *Nat. Immunol.* **2008**, *9*, 1179; b) M. J. Robinson, F. Osorio, M. Rosas, R. P. Freitas, E. Schweighoffer, O. Groß, J. S. Verbeek, J. Ruland, V. Tybulewicz, G. D. Brown, L. F. Moita, P. R. Taylor, C. Reis e Sousa, *J. Exp. Med.* **2009**, *206*, 2037.
[138] C. Plank, K. Zatloukal, M. Cotten, K. Mechtler, E. Wagner, *Bioconjugate Chem.* **1992**, *3*, 533-539.
[139] a) T. E. Warkentin, M. N. Levine, J. Hirsh, P. Horsewood, R. S. Roberts, M. Gent, J. G. Kelton, *N. Engl. J. Med.* **1995**, *332*, 1330-1336; b) J. Hirsh, T. E. Warkentin, S. G. Shaughnessy, S. S. Anand, *Chest* **2001**, *119*, S64.
[140] a) S. Sattin, A. Bernardi, *Trends Biotechnol.* **2016**, *34*, 483-495; b) B. Ernst, J. L. Magnani, *Nat. Rev. Drug Discov.* **2009**, *8*, 661.
[141] A. Beck, T. Wurch, C. Bailly, N. Corvaia, *Nat. Rev. Immunol.* **2010**, *10*, 345.
[142] J. W. Eschbach, J. C. Egrie, M. R. Downing, J. K. Browne, J. W. Adamson, *N. Engl. J. Med.* **1987**, *316*, 73-78.
[143] Y. Zhang, F. Wang, *Drug Discov. Ther.* **2015**, *9*, 79-87.
[144] M.-M. Samama, G. T. Gerotziafas, *Thromb. Res.* **2003**, *109*, 1-11.
[145] M. D. Privitera, *Ann. Pharmacother.* **1997**, *31*, 1164-1173.

[146] H. Bischoff, *Clin. Investig. Med.* **1995**, *18*, 303-311.
[147] A. Mitrakou, N. Tountas, A. E. Raptis, R. J. Bauer, H. Schulz, S. A. Raptis, *Diabetic Med.* **1998**, *15*, 657-660.
[148] D. Elstein, C. Hollak, J. M. F. G. Aerts, S. van Weely, M. Maas, T. M. Cox, R. H. Lachmann, M. Hrebicek, F. M. Platt, T. D. Butters, R. A. Dwek, A. Zimran, *J. Inherit. Metab. Dis.* **2004**, *27*, 757-766.
[149] a) F. G. Hayden, A. D. Osterhaus, J. J. Treanor, D. M. Fleming, F. Y. Aoki, K. G. Nicholson, A. M. Bohnen, H. M. Hirst, O. Keene, K. Wightman, *N. Engl. J. Med.* **1997**, *337*, 874-880; b) F. G. Hayden, R. L. Atmar, M. Schilling, C. Johnson, D. Poretz, D. Paar, L. Huson, P. Ward, R. G. Mills, O. S. Group, *N. Engl. J. Med.* **1999**, *341*, 1336-1343.
[150] S. J. Davies, G. Woodrow, K. Donovan, J. Plum, P. Williams, A. C. Johansson, H.-P. Bosselmann, O. Heimbürger, O. Simonsen, A. Davenport, *J. Am. Soc. Nephrol.* **2003**, *14*, 2338-2344.
[151] a) S. Muthana, H. Cao, X. Chen, *Curr. Opin. Chem. Biol.* **2009**, *13*, 573-581; b) K. M. Koeller, C.-H. Wong, *Glycobiology* **2000**, *10*, 1157-1169.
[152] a) P. H. Seeberger, S. Roehrig, P. Schell, Y. Wang, W. J. Christ, *Carbohydr. Res.* **2000**, *328*, 61-69; b) O. Calin, S. Eller, P. H. Seeberger, *Angew Chem. Int. Ed. Engl.* **2013**, *52*, 5862-5865.
[153] V. Roldós, F. J. Cañada, J. Jiménez-Barbero, *ChemBioChem* **2011**, *12*, 990-1005.
[154] a) F. T. Alberto, C. F. Javier, J. B. Jesús, *Chem. Eur. J.* **2015**, *21*, 10616-10628; b) S. B. Shuker, P. J. Hajduk, R. P. Meadows, S. W. Fesik, *Science* **1996**, *274*, 1531-1534; c) F. T. Alberto, C. F. Javier, J. B. Jesús, *ChemMedChem* **2015**, *10*, 1291-1295.
[155] A. Bernardi, J. Jiménez-Barbero, A. Casnati, C. De Castro, T. Darbre, F. Fieschi, J. Finne, H. Funken, K.-E. Jaeger, M. Lahmann, *Chem. Soc. Rev.* **2013**, *42*, 4709-4727.
[156] W. Lew, X. Chen, C. U. Kim, *Curr. Med. Chem.* **2000**, *7*, 663-672.
[157] a) Y. M. Chabre, R. Roy, *Advances in Carbohydrate Chemistry and Biochemistry, Academic Press* **2010**, *63*, 165-393; b) J. E. Gestwicki, C. W. Cairo, L. E. Strong, K. A. Oetjen, L. L. Kiessling, *J. Am. Chem. Soc.* **2002**, *124*, 14922-14933; c) V. Wittmann, R. J. Pieters, *Chem. Soc. Rev.* **2013**, *42*, 4492-4503; d) S. Ordanini, G. Goti, A. Bernardi, *Can. J. Chem.* **2017**, *95*, 881-890.
[158] L. McKerracher, S. David, D. L. Jackson, V. Kottis, R. J. Dunn, P. E. Braun, *Neuron* **1994**, *13*, 805-811.
[159] a) Y. J. Kim, L. Borsig, N. M. Varki, A. Varki, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9325; b) L. Borsig, R. Wong, R. O. Hynes, N. M. Varki, A. Varki, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2193; c) P. Brodt, L. Fallavollita, R. S. Bresalier, S. Meterissian, C. R. Norton, B. A. Wolitzky, *Int. J. Cancer* **1997**, *71*, 612-619.
[160] T. D. Farr, C.-H. Lai, D. Grünstein, G. Orts-Gil, C.-C. Wang, P. Boehm-Sturm, P. H. Seeberger, C. Harms, *Nano Lett.* **2014**, *14*, 2130-2134.
[161] S. I. van Kasteren, S. J. Campbell, S. Serres, D. C. Anthony, N. R. Sibson, B. G. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 18.
[162] T. Wun, M. J. Telen, L. Krishnamurti, T. L. McCavit, L. M. DeCastro, H. Flanner, F. A. Kuypers, S. K. Larkin, S. Rhee, J. L. Magnani, *American Society of Hematology* **2014**.
[163] J. Egger, C. Weckerle, B. Cutting, O. Schwardt, S. Rabbani, K. Lemme, B. Ernst, *J. Am. Chem. Soc.* **2013**, *135*, 9820-9828.
[164] I. U. Mysorekar, M. A. Mulvey, S. J. Hultgren, J. I. Gordon, *J. Biol. Chem.* **2002**, *277*, 7412-7419.
[165] N. Dreux, J. Denizot, M. Martinez-Medina, A. Mellmann, M. Billig, D. Kisiela, S. Chattopadhyay, E. Sokurenko, C. Neut, C. Gower-Rousseau, *PLOS Pathog.* **2013**, *9*, e1003141.
[166] A. Wellens, M. Lahmann, M. Touaibia, J. Vaucher, S. Oscarson, R. Roy, H. Remaut, J. Bouckaert, *Biochemistry* **2012**, *51*, 4790-4799.
[167] O. Schwardt, S. Rabbani, M. Hartmann, D. Abgottspon, M. Wittwer, S. Kleeb, A. Zalewski, M. Smieško, B. Cutting, B. Ernst, *Bioorganic Med. Chem.* **2011**, *19*, 6454-6473.
[168] L. Pang, S. Kleeb, K. Lemme, S. Rabbani, M. Scharenberg, A. Zalewski, F. Schädler, O. Schwardt, B. Ernst, *ChemMedChem* **2012**, *7*, 1404-1422.
[169] S. Kleeb, L. Pang, K. Mayer, D. Eris, A. Sigl, R. C. Preston, P. Zihlmann, T. Sharpe, R. P. Jakob, D. Abgottspon, A. S. Hutter, M. Scharenberg, X. Jiang, G. Navarra, S. Rabbani, M. Smiesko, N. Lüdin, J. Bezençon, O. Schwardt, T. Maier, B. Ernst, *J. Med. Chem.* **2015**, *58*, 2221-2239.
[170] T.-F. Mah, B. Pitts, B. Pellock, G. C. Walker, P. S. Stewart, G. A. O'Toole, *Nature* **2003**, *426*, 306.
[171] P. S. Stewart, J. William Costerton, *Lancet* **2001**, *358*, 135-138.

[172] a) A. Ilangovan, M. Fletcher, G. Rampioni, C. Pustelny, K. Rumbaugh, S. Heeb, M. Cámara, A. Truman, S. R. Chhabra, J. Emsley, *PLOS Pathog.* **2013**, *9*, e1003508; b) M. Hentzer, H. Wu, J. B. Andersen, K. Riedel, T. B. Rasmussen, N. Bagge, N. Kumar, M. A. Schembri, Z. Song, P. Kristoffersen, M. Manefield, J. W. Costerton, S. Molin, L. Eberl, P. Steinberg, S. Kjelleberg, N. Høiby, M. Givskov, *EMBO J.* **2003**, *22*, 3803.

[173] K. Winzer, C. Falconer, N. C. Garber, S. P. Diggle, M. Camara, P. Williams, *J. Bacteriol.* **2000**, *182*, 6401-6411.

[174] a) S. P. Diggle, R. E. Stacey, C. Dodd, M. Cámara, P. Williams, K. Winzer, *Environ. Microbiol.* **2006**, *8*, 1095-1104; b) D. Tielker, S. Hacker, R. Loris, M. Strathmann, J. Wingender, S. Wilhelm, F. Rosenau, K.-E. Jaeger, *Microbiology* **2005**, *151*, 1313-1323.

[175] G. Cioci, P. Mitchell Edward, C. Gautier, M. Wimmerová, D. Sudakevitz, S. Pérez, N. Gilboa-Garber, A. Imberty, *FEBS Lett.* **2003**, *555*, 297-301.

[176] E. P. Mitchell, C. Sabin, L. Šnajdrová, M. Pokorná, S. Perret, C. Gautier, C. Hofr, N. Gilboa-Garber, J. Koča, M. Wimmerová, A. Imberty, *Proteins: Structure, Function, and Bioinformatics* **2005**, *58*, 735-746.

[177] M. Whiteley, M. G. Bangera, R. E. Bumgarner, M. R. Parsek, G. M. Teitzel, S. Lory, E. P. Greenberg, *Nature* **2001**, *413*, 860.

[178] K. Marotte, C. Préville, C. Sabin, M. Moumé-Pymbock, A. Imberty, R. Roy, *Org. Biomol. Chem.* **2007**, *5*, 2953-2961.

[179] D. Hauck, I. Joachim, B. Frommeyer, A. Varrot, B. Philipp, H. M. Möller, A. Imberty, T. E. Exner, A. Titz, *ACS Chem. Biol.* **2013**, *8*, 1775-1784.

[180] R. Sommer, S. Wagner, K. Rox, A. Varrot, D. Hauck, E.-C. Wamhoff, J. Schreiber, T. Ryckmans, T. Brunner, C. Rademacher, R. W. Hartmann, M. Brönstrup, A. Imberty, A. Titz, *J. Am. Chem. Soc.* **2018**, *140*, 2537-2545.

[181] F. Zhang, S. Ren, Y. Zuo, *Int. Rev. Immunol.* **2014**, *33*, 54-66.

[182] J. J. Garcia-Vallejo, Y. van Kooyk, *Trends Immunol.* **2013**, *34*, 482-486.

[183] E. J. Soilleux, L. S. Morris, G. Leslie, J. Chehimi, Q. Luo, E. Levroney, J. Trowsdale, L. J. Montaner, R. W. Doms, D. Weissman, N. Coleman, B. Lee, *J. Leukoc. Biol.* **2002**, *71*, 445-457.

[184] M. Relloso, A. Puig-Kröger, O. M. Pello, J. L. Rodríguez-Fernández, G. de la Rosa, N. Longo, J. Navarro, M. A. Muñoz-Fernández, P. Sánchez-Mateos, A. L. Corbí, *J. Immunol.* **2002**, *168*, 2634.

[185] a) Y. van Kooyk, T. B. H. Geijtenbeek, *Nat. Rev. Immunol.* **2003**, *3*, 697; b) U. Švajger, M. Anderluh, M. Jeras, N. Obermajer, *Cell. Signal.* **2010**, *22*, 1397-1405.

[186] E. van Liempt, C. M. C. Bank, P. Mehta, J. J. Garcı´a-Vallejo, Z. S. Kawar, R. Geyer, R. A. Alvarez, R. D. Cummings, Y. v. Kooyk, I. van Die, *FEBS Lett.* **2006**, *580*, 6123-6131.

[187] I. van Die, S. J. van Vliet, A. K. Nyame, R. D. Cummings, C. M. C. Bank, B. Appelmelk, T. B. H. Geijtenbeek, Y. van Kooyk, *Glycobiology* **2003**, *13*, 471-478.

[188] a) D. A. Mitchell, A. J. Fadden, K. Drickamer, *J. Biol. Chem.* **2001**, *276*, 28939-28945; b) Q. D. Yu, A. P. Oldring, A. S. Powlesland, C. K. W. Tso, C. Yang, K. Drickamer, M. E. Taylor, *J. Mol. Biol.* **2009**, *387*, 1075-1080.

[189] G. Tabarani, M. Thépaut, D. Stroebel, C. Ebel, C. Vivès, P. Vachette, D. Durand, F. Fieschi, *J. Biol. Chem.* **2009**, *284*, 21229-21240.

[190] A. A. Bashirova, T. B. H. Geijtenbeek, G. C. F. van Duijnhoven, S. J. van Vliet, J. B. G. Eilering, M. P. Martin, L. Wu, T. D. Martin, N. Viebig, P. A. Knolle, V. N. KewalRamani, Y. van Kooyk, M. Carrington, *J. Exp. Med.* **2001**, *193*, 671.

[191] S. Menon, K. Rosenberg, S. A. Graham, E. M. Ward, M. E. Taylor, K. Drickamer, D. E. Leckband, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11524.

[192] a) L. Ping, W. Xiang, I. M. S., N. A. K., d. S. A. M., J. Ken, T. N. L., *Traffic* **2014**, *15*, 179-196; b) C. Manzo, J. A. Torreno-Pina, B. Joosten, I. Reinieren-Beeren, E. J. Gualda, P. Loza-Alvarez, C. G. Figdor, M. F. Garcia-Parajo, A. Cambi, *J. Biol. Chem.* **2012**, *287*, 38946-38955.

[193] T. B. H. Geijtenbeek, D. J. E. B. Krooshoop, D. A. Bleijs, S. J. van Vliet, G. C. F. van Duijnhoven, V. Grabovsky, R. Alon, C. G. Figdor, Y. van Kooyk, *Nat. Immunol.* **2000**, *1*, 353.

[194] S. I. Gringhuis, J. den Dunnen, M. Litjens, M. van der Vlist, T. B. H. Geijtenbeek, *Nat. Immunol.* **2009**, *10*, 1081.

[195] M. P. Bergman, A. Engering, H. H. Smits, S. J. van Vliet, A. A. van Bodegraven, H.-P. Wirth, M. L. Kapsenberg, C. M. J. E. Vandenbroucke-Grauls, Y. van Kooyk, B. J. Appelmelk, *J. Exp. Med.* **2004**, *200*, 979.
[196] T. B. H. Geijtenbeek, D. S. Kwon, R. Torensma, S. J. van Vliet, G. C. F. van Duijnhoven, J. Middel, I. L. M. H. A. Cornelissen, H. S. L. M. Nottet, V. N. KewalRamani, D. R. Littman, C. G. Figdor, Y. van Kooyk, *Cell* **2000**, *100*, 587-597.
[197] B. M. Curtis, S. Scharnowske, A. J. Watson, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 8356.
[198] C. P. Alvarez, F. Lasala, J. Carrillo, O. Muñiz, A. L. Corbí, R. Delgado, *J. Virol.* **2002**, *76*, 6841-6844.
[199] B. Tassaneetrithep, T. H. Burgess, A. Granelli-Piperno, C. Trumpfheller, J. Finke, W. Sun, M. A. Eller, K. Pattanapanyasat, S. Sarasombath, D. L. Birx, R. M. Steinman, S. Schlesinger, M. A. Marovich, *J. Exp. Med.* **2003**, *197*, 823.
[200] a) S. Pöhlmann, J. Zhang, F. Baribaud, Z. Chen, G. J. Leslie, G. Lin, A. Granelli-Piperno, R. W. Doms, C. M. Rice, J. A. McKeating, *J. Virol.* **2003**, *77*, 4070-4080; b) P.-Y. Lozach, H. Lortat-Jacob, A. De Lacroix De Lavalette, I. Staropoli, S. Foung, A. Amara, C. Houlès, F. Fieschi, O. Schwartz, J.-L. Virelizier, F. Arenzana-Seisdedos, R. Altmeyer, *J. Biol. Chem.* **2003**, *278*, 20358-20366.
[201] F. Halary, A. Amara, H. Lortat-Jacob, M. Messerle, T. Delaunay, C. Houlès, F. Fieschi, F. Arenzana-Seisdedos, J.-F. Moreau, J. Déchanet-Merville, *Immunity* **2002**, *17*, 653-664.
[202] S. Pöhlmann, E. J. Soilleux, F. Baribaud, G. J. Leslie, L. S. Morris, J. Trowsdale, B. Lee, N. Coleman, R. W. Doms, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2670.
[203] A. Marzi, T. Gramberg, G. Simmons, P. Möller, A. J. Rennekamp, M. Krumbiegel, M. Geier, J. Eisemann, N. Turza, B. Saunier, *J. Virol.* **2004**, *78*, 12090-12095.
[204] A.-R. Goncalves, M.-L. Moraz, A. Pasquato, A. Helenius, P.-Y. Lozach, S. Kunz, *J. Virol.* **2013**, *87*, 11504-11515.
[205] B. E. E. Martina, P. Koraka, P. van den Doel, G. F. Rimmelzwaan, B. L. Haagmans, A. D. M. E. Osterhaus, *Virus Res.* **2008**, *135*, 64-71.
[206] L. Tailleux, O. Schwartz, J.-L. Herrmann, E. Pivert, M. Jackson, A. Amara, L. Legres, D. Dreher, L. P. Nicod, J. C. Gluckman, P. H. Lagrange, B. Gicquel, O. Neyrolles, *J. Exp. Med.* **2003**, *197*, 121.
[207] L. Steeghs, S. J. Van Vliet, H. Uronen-Hansson, A. Van Mourik, A. Engering, M. Sanchez-Hernandez, N. Klein, R. Callard, J. P. M. Van Putten, P. Van Der Ley, Y. Van Kooyk, J. G. J. Van De Winkel, *Cell. Microbiol.* **2006**, *8*, 316-325.
[208] a) N. Maeda, J. Nigou, J.-L. Herrmann, M. Jackson, A. Amara, P. H. Lagrange, G. Puzo, B. Gicquel, O. Neyrolles, *J. Biol. Chem.* **2003**, *278*, 5513-5516; b) N. N. Driessen, R. Ummels, J. J. Maaskant, S. S. Gurcha, G. S. Besra, G. D. Ainge, D. S. Larsen, G. F. Painter, C. M. Vandenbroucke-Grauls, J. Geurtsen, *Infect. Immun.* **2009**, *77*, 4538-4547.
[209] M. Colmenares, A. L. Corbí, S. J. Turco, L. Rivas, *J. Immunol.* **2004**, *172*, 1186.
[210] A. Cambi, K. Gijzen, I. J. M. de Vries, R. Torensma, B. Joosten, G. J. Adema, M. G. Netea, B. J. Kullberg, L. Romani, C. G. Figdor, *Eur. J. Immunol.* **2003**, *33*, 532-538.
[211] D. Serrano-Gómez, J. Antonio Leal, A. L. Corbí, *Immunobiology* **2005**, *210*, 175-183.
[212] M. Emara, P.-J. Royer, J. Mahdavi, F. Shakib, A. M. Ghaemmaghami, *J. Biol. Chem.* **2012**, *287*, 5756-5763.
[213] B. J. Appelmelk, I. van Die, S. J. van Vliet, C. M. J. E. Vandenbroucke-Grauls, T. B. H. Geijtenbeek, Y. van Kooyk, *J. Immunol.* **2003**, *170*, 1635.
[214] F. Lasala, E. Arce, J. R. Otero, J. Rojo, R. Delgado, *Antimicrob. Agents Chemother.* **2003**, *47*, 3970-3972.
[215] D. S. Kwon, G. Gregorio, N. Bitton, W. A. Hendrickson, D. R. Littman, *Immunity* **2002**, *16*, 135-144.
[216] G. A. Snyder, J. Ford, P. Torabi-Parizi, J. A. Arthos, P. Schuck, M. Colonna, P. D. Sun, *J. Virol.* **2005**, *79*, 4589-4598.
[217] L. Burleigh, P.-Y. Lozach, C. Schiffer, I. Staropoli, V. Pezo, F. Porrot, B. Canque, J.-L. Virelizier, F. Arenzana-Seisdedos, A. Amara, *J. Virol.* **2006**, *80*, 2949-2957.
[218] J.-F. Arrighi, M. Pion, E. Garcia, J.-M. Escola, Y. van Kooyk, T. B. Geijtenbeek, V. Piguet, *J. Exp. Med.* **2004**, *200*, 1279.
[219] L. Wu, V. N. KewalRamani, *Nat. Rev. Immunol.* **2006**, *6*, 859.

[220] a) P.-Y. Lozach, L. Burleigh, I. Staropoli, E. Navarro-Sanchez, J. Harriague, J.-L. Virelizier, F. A. Rey, P. Desprès, F. Arenzana-Seisdedos, A. Amara, *J. Biol. Chem.* **2005**, *280*, 23698-23708; b) M. M. F. Alen, S. J. F. Kaptein, T. De Burghgraeve, J. Balzarini, J. Neyts, D. Schols, *Virology* **2009**, *387*, 67-75.
[221] Z. Elena, D. Marta, D. Stefano, *J. Leukoc. Biol.* **2009**, *86*, 1393-1401.
[222] J. Kleinnijenhuis, M. Oosting, L. A. Joosten, M. G. Netea, R. Van Crevel, *Clin. Dev. Immunol.* **2011**, *2011*.
[223] a) R. J. E. Li, S. J. van Vliet, Y. van Kooyk, *Curr. Opin. Biotechnol.* **2018**, *51*, 24-31; b) K. Palucka, J. Banchereau, *Immunity* **2013**, *39*, 38-48; c) Y. van Kooyk, W. W. J. Unger, C. M. Fehres, H. Kalay, J. J. García-Vallejo, *Mol. Immunol.* **2013**, *55*, 143-145.
[224] L. de Witte, A. Nabatov, M. Pion, D. Fluitsma, M. A. W. P. de Jong, T. de Gruijl, V. Piguet, Y. van Kooyk, T. B. H. Geijtenbeek, *Nat. Med.* **2007**, *13*, 367.
[225] E. Pokidysheva, Y. Zhang, A. J. Battisti, C. M. Bator-Kelly, P. R. Chipman, C. Xiao, G. G. Gregorio, W. A. Hendrickson, R. J. Kuhn, M. G. Rossmann, *Cell* **2006**, *124*, 485-493.
[226] T. B. H. Geijtenbeek, G. C. F. van Duijnhoven, S. J. van Vliet, E. Krieger, G. Vriend, C. G. Figdor, Y. van Kooyk, *J. Biol. Chem.* **2002**, *277*, 11314-11320.
[227] Y. Guo, H. Feinberg, E. Conroy, D. A. Mitchell, R. Alvarez, O. Blixt, M. E. Taylor, W. I. Weis, K. Drickamer, *Nat. Struct. Mol. Biol.* **2004**, *11*, 591.
[228] O. Blixt, S. Head, T. Mondala, C. Scanlan, M. E. Huflejt, R. Alvarez, M. C. Bryan, F. Fazio, D. Calarese, J. Stevens, N. Razi, D. J. Stevens, J. J. Skehel, I. van Die, D. R. Burton, I. A. Wilson, R. Cummings, N. Bovin, C.-H. Wong, J. C. Paulson, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17033.
[229] T. Gabriele, T. Georges, A. Marko, I. Donatella, V. Francesca, P. Donatella, N. P. M., R. Javier, F. Franck, B. Anna, *ChemBioChem* **2008**, *9*, 1921-1930.
[230] M. E. Taylor, K. Drickamer, *Glycobiology* **2009**, *19*, 1155-1162.
[231] M. Anderluh, G. Jug, U. Svajger, N. Obermajer, *Curr. Med. Chem.* **2012**, *19*, 992-1007.
[232] E. W. Adams, D. M. Ratner, H. R. Bokesch, J. B. McMahon, B. R. O'Keefe, P. H. Seeberger, *Chem. Biol.* **2004**, *11*, 875-881.
[233] a) E. Yuriev, W. Farrugia, A. M. Scott, P. A. Ramsland, *Immunol. Cell Biol.* **2005**, *83*, 709-717; b) T. Aeschbacher, M. Zierke, M. Smieško, M. Collot, J. M. Mallet, B. Ernst, F. H. T. Allain, M. Schubert, *Chem. Eur. J.* **2017**, *23*, 11598-11610.
[234] M. J. Borrok, L. L. Kiessling, *J. Am. Chem. Soc.* **2007**, *129*, 12780-12785.
[235] S. L. Mangold, L. R. Prost, L. L. Kiessling, *Chem. Sci.* **2012**, *3*, 772-777.
[236] K. C. Garber, K. Wangkanont, E. E. Carlson, L. L. Kiessling, *ChemComm* **2010**, *46*, 6747-6749.
[237] F. Damkaci, P. DeShong, *J. Am. Chem. Soc.* **2003**, *125*, 4408-4409.
[238] M. Andreini, D. Doknic, I. Sutkeviciute, J. J. Reina, J. Duan, E. Chabrol, M. Thepaut, E. Moroni, F. Doro, L. Belvisi, *Org. Biomol. Chem.* **2011**, *9*, 5778-5786.
[239] D. A. Mitchell, N. A. Jones, S. J. Hunter, J. M. D. Cook, S. F. Jenkinson, M. R. Wormald, R. A. Dwek, G. W. J. Fleet, *Tetrahedron: Asymmetry* **2007**, *18*, 1502-1510.
[240] T. Tomašić, D. Hajšek, U. Švajger, J. Luzar, N. Obermajer, I. Petit-Haertlein, F. Fieschi, M. Anderluh, *Eur. J. Med. Chem.* **2014**, *75*, 308-326.
[241] L. Dehuyser, E. Schaeffer, O. Chaloin, C. G. Mueller, R. Baati, A. Wagner, *Bioconjugate Chem.* **2012**, *23*, 1731-1739.
[242] a) S. Mari, H. Posteri, G. Marcou, D. Potenza, F. Micheli, F. J. Cañada, J. Jimenez-Barbero, A. Bernardi, *Eur. J. Org. Chem.* **2004**, *2004*, 5119-5225; b) J. J. Reina, S. Sattin, D. Invernizzi, S. Mari, L. Martínez-Prats, G. Tabarani, F. Fieschi, R. Delgado, P. M. Nieto, J. Rojo, A. Bernardi, *ChemMedChem* **2007**, *2*, 1030-1036.
[243] D. Arosio, I. Vrasidas, P. Valentini, R. M. Liskamp, R. J. Pieters, A. Bernardi, *Org. Biomol. Chem.* **2004**, *2*, 2113-2124.
[244] M. Thépaut, C. Guzzi, I. Sutkeviciute, S. Sattin, R. Ribeiro-Viana, N. Varga, E. Chabrol, J. Rojo, A. Bernardi, J. Angulo, P. M. Nieto, F. Fieschi, *J. Am. Chem. Soc.* **2013**, *135*, 2518-2529.
[245] A. Tamburrini, S. Achilli, F. Vasile, S. Sattin, C. Vivès, C. Colombo, F. Fieschi, A. Bernardi, *Bioorganic Med. Chem.* **2017**, *25*, 5142-5147.
[246] S. Mari, I. Sánchez-Medina, P. Mereghetti, L. Belvisi, J. Jiménez-Barbero, A. Bernardi, *Carbohydr. Res.* **2007**, *342*, 1859-1868.

[247] I. Sutkeviciute, M. Thépaut, S. Sattin, A. Berzi, J. McGeagh, S. Grudinin, J. Weiser, A. Le Roy, J. J. Reina, J. Rojo, M. Clerici, A. Bernardi, C. Ebel, F. Fieschi, *ACS Chem. Biol.* **2014**, *9*, 1377-1385.
[248] N. Obermajer, S. Sattin, C. Colombo, M. Bruno, U. Švajger, M. Anderluh, A. Bernardi, *Mol. Divers.* **2011**, *15*, 347-360.
[249] S. Sattin, *PhD thesis, Milano* **2009**.
[250] N. Varga, I. Sutkeviciute, C. Guzzi, J. McGeagh, I. Petit-Haertlein, S. Gugliotta, J. Weiser, J. Angulo, F. Fieschi, A. Bernardi, *Chem. Eur. J.* **2013**, *19*, 4786-4797.
[251] V. Porkolab, E. Chabrol, N. Varga, S. Ordanini, I. Sutkevičiūtė, M. Thépaut, M. J. García-Jiménez, E. Girard, P. M. Nieto, A. Bernardi, F. Fieschi, *ACS Chem. Biol.* **2018**, *13*, 600-608.
[252] G. Zemplén, E. Pacsu, *Ber. Dtsch. Chem. Ges.* **1929**, *62*, 1613-1614.
[253] A. Bernardi, D. Arosio, D. Dellavecchia, F. Micheli, *Tetrahedron: Asymmetry* **1999**, *10*, 3403-3407.
[254] R. Russo, *Master thesis, Milano* **2015**.
[255] A. Bernardi, D. Arosio, L. Manzoni, F. Micheli, A. Pasquarello, P. Seneci, *J. Org. Chem.* **2001**, *66*, 6209-6216.
[256] J. J. Reina, A. Bernardi, *Tetrahedron* **2011**, *67*, 5770-5775.
[257] S. Bräse, K. Banert, *Organic azides: syntheses and applications*, John Wiley & Sons, **2010**.
[258] a) R. F. Bruns, I. A. Watson, *J. Med. Chem.* **2012**, *55*, 9763-9772; b) R. Brenk, A. Schipani, D. James, A. Krasowski, I. H. Gilbert, J. Frearson, P. G. Wyatt, *ChemMedChem* **2008**, *3*, 435-444.
[259] J. B. Baell, G. A. Holloway, *J. Med. Chem.* **2010**, *53*, 2719-2740.
[260] W. C. Still, M. Kahn, A. Mitra, *J. Org. Chem.* **1978**, *43*, 2923-2925.
[261] Y. Dong, X. Liang, H. Yuan, S. Qi, F. Chen, D. Wang, *Green Chem.* **2008**, *10*, 990-994.
[262] M. Rubinshtein, C. R. James, J. L. Young, Y. J. Ma, Y. Kobayashi, N. C. Gianneschi, J. Yang, *Org. Lett.* **2010**, *12*, 3560-3563.
[263] V. Aucagne, I. E. Valverde, P. Marceau, M. Galibert, N. Dendane, A. F. Delmas, *Angew Chem. Int. Ed. Engl.* **2012**, *51*, 11320-11324.
[264] W. M. Freeman, S. J. Walker, K. E. Vrana, *BioTechniques* **1999**, *26*, 112-125.
[265] M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* **1995**, *270*, 467.
[266] F. Taub, E. B. Thompson, *Anal. Biochem.* **1982**, *126*, 222-230.
[267] C. Tse-Wen, *J. Immunol. Methods* **1983**, *65*, 217-223.
[268] S. P. Fodor, J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu, D. Solas, *Science* **1991**, *251*, 767.
[269] a) K. T. Pilobello, L. Krishnamoorthy, D. Slawek, L. K. Mahal, *ChemBioChem* **2005**, *6*, 985-989; b) A. Kuno, N. Uchiyama, S. Koseki-Kuno, Y. Ebe, S. Takashima, M. Yamada, J. Hirabayashi, *Nat. Methods* **2005**, *2*, 851; c) T. Zheng, D. Peelen, L. M. Smith, *J. Am. Chem. Soc.* **2005**, *127*, 9982-9983.
[270] H. Tateno, N. Uchiyama, A. Kuno, A. Togayachi, T. Sato, H. Narimatsu, J. Hirabayashi, *Glycobiology* **2007**, *17*, 1138-1146.
[271] S. Angeloni, J. L. Ridet, N. Kusy, H. Gao, F. Crevoisier, S. Guinchard, S. Kochhar, H. Sigrist, N. Sprenger, *Glycobiology* **2005**, *15*, 31-41.
[272] A. Shimoda, Y. Tahara, S.-i. Sawada, Y. Sasaki, K. Akiyoshi, *Biochem. Biophys. Res. Commun.* **2017**, *491*, 701-707.
[273] a) K.-L. Hsu, K. T. Pilobello, L. K. Mahal, *Nat. Chem. Biol.* **2006**, *2*, 153; b) K.-L. Hsu, L. K. Mahal, *Nat. Protoc.* **2006**, *1*, 543.
[274] M.-r. Lee, S. Park, I. Shin, *Bioorganic Med. Chem. Lett.* **2006**, *16*, 5132-5135.
[275] a) R. Zhao, W. Qin, R. Qin, J. Han, C. Li, Y. Wang, C. Xu, *Clin. Proteom.* **2017**, *14*, 20; b) P. Chen, Y. Liu, X. Kang, L. Sun, P. Yang, Z. Tang, *J. Cancer Res. Clin. Oncol.* **2008**, *134*, 851-860; c) S.-C. Tao, Y. Li, J. Zhou, J. Qian, R. L. Schnaar, Y. Zhang, I. J. Goldstein, H. Zhu, J. P. Schneck, *Glycobiology* **2008**, *18*, 761-769; d) J. Wu, X. Xie, Y. Liu, J. He, R. Benitez, R. J. Buckanovich, D. M. Lubman, *J. Proteome Res.* **2012**, *11*, 4541-4552; e) S. Fry, B. Afrough, A. Leathem, M. Dwek, in *Metastasis Research Protocols*, Springer, **2012**, pp. 267-272.
[276] a) S. Chen, T. LaRoche, D. Hamelinck, D. Bergsma, D. Brenner, D. Simeone, R. E. Brand, B. B. Haab, *Nat. Methods* **2007**, *4*, 437; b) C. Li, D. M. Simeone, D. E. Brenner, M. A. Anderson, K. A. Shedden, M. T. Ruffin, D. M. Lubman, *J. Proteome Res.* **2008**, *8*, 483-492.

[277] L. Shi, L. H. Reid, W. D. Jones, R. Shippy, J. A. Warrington, S. C. Baker, P. J. Collins, F. De Longueville, E. S. Kawasaki, K. Y. Lee, *Nat. Biotechnol.* **2006**, *24*, 1151.

[278] a) S. Park, J. C. Gildersleeve, O. Blixt, I. Shin, *Chem. Soc. Rev.* **2013**, *42*, 4310-4326; b) C. D. Rillahan, J. C. Paulson, *Annu. Rev. Biochem.* **2011**, *80*, 797-823; c) E. Lonardi, C. I. A. Balog, A. M. Deelder, M. Wuhrer, *Expert Rev. Proteomics* **2010**, *7*, 761-774; d) A. Geissner, C. Anish, P. H. Seeberger, *Curr. Opin. Chem. Biol.* **2014**, *18*, 38-45; e) A. Geissner, P. H. Seeberger, *Annu. Rev. Anal. Chem.* **2016**, *9*, 223-247.

[279] a) M. De Kerpel, I. Van Molle, L. Brys, L. Wyns, H. De Greve, J. Bouckaert, *Acta Crystallogr. Section F: Structural Biology and Crystallization Communications* **2006**, *62*, 1278-1282; b) J. Bouckaert, J. Mackenzie, J. L. De Paz, B. Chipwaza, D. Choudhury, A. Zavialov, K. Mannerstedt, J. Anderson, D. Piérard, L. Wyns, P. H. Seeberger, S. Oscarson, H. De Greve, S. D. Knight, *Mol. Microbiol.* **2006**, *61*, 1556-1568.

[280] a) E. Lameignere, L. Malinovská, M. Sláviková, E. Duchaud, Edward P. Mitchell, A. Varrot, O. Šedo, A. Imberty, M. Wimmerová, *Biochem. J.* **2008**, *411*, 307; b) O. Šulák, G. Cioci, M. Delia, M. Lahmann, A. Varrot, A. Imberty, M. Wimmerová, *Structure* **2010**, *18*, 59-72.

[281] B. Blanchard, A. Nurisso, E. Hollville, C. Tétaud, J. Wiels, M. Pokorná, M. Wimmerová, A. Varrot, A. Imberty, *J. Mol. Biol.* **2008**, *383*, 837-853.

[282] a) M. Karine, S. Charles, P. Cathy, M. P. Myriam, W. Michaela, M. E. P., I. Anne, R. René, *ChemMedChem* **2007**, *2*, 1328-1338; b) J. Topin, J. Arnaud, A. Sarkar, A. Audfray, E. Gillon, S. Perez, H. Jamet, A. Varrot, A. Imberty, A. Thomas, *PLoS One* **2013**, *8*, e71149.

[283] X. Song, Y. Lasanajak, L. J. Olson, M. Boonen, N. M. Dahms, S. Kornfeld, R. D. Cummings, D. F. Smith, *J. Biol. Chem.* **2009**, *284*, 35201-35214.

[284] a) S. R. Stowell, C. M. Arthur, M. Dias-Baruffi, L. C. Rodrigues, J.-P. Gourdine, J. Heimburg-Molinaro, T. Ju, R. J. Molinaro, C. Rivera-Marrero, B. Xia, D. F. Smith, R. D. Cummings, *Nat. Med.* **2010**, *16*, 295; b) X. Song, B. Xia, S. R. Stowell, Y. Lasanajak, D. F. Smith, R. D. Cummings, *Chem. Biol.* **2009**, *16*, 36-47; c) D. F. Smith, X. Song, R. D. Cummings, in *Methods Enzymol., Vol. 480* (Ed.: M. Fukuda), Academic Press, **2010**, pp. 417-444; d) V. Marková, K. Smetana, G. Jeníková, J. Láchová, V. Krejciríková, M. Poplstein, M. Fábry, J. Brynda, R. A. Alvarez, R. D. Cummings, *Int J Mol Med.* **2006**, *18*, 65-76; e) T. Horlacher, M. A. Oberli, D. B. Werz, L. Kröck, S. Bufali, R. Mishra, J. Sobek, K. Simons, M. Hirashima, T. Niki, P. H. Seeberger, *ChemBioChem* **2010**, *11*, 1563-1573.

[285] a) B. S. Bochner, R. A. Alvarez, P. Mehta, N. V. Bovin, O. Blixt, J. R. White, R. L. Schnaar, *J. Biol. Chem.* **2005**, *280*, 4307-4312; b) H. Tateno, P. R. Crocker, J. C. Paulson, *Glycobiology* **2005**, *15*, 1125-1135.

[286] X. Song, H. Yu, X. Chen, Y. Lasanajak, M. M. Tappert, G. M. Air, V. K. Tiwari, H. Cao, H. A. Chokhawala, H. Zheng, R. D. Cummings, D. F. Smith, *J. Biol. Chem.* **2011**, *286*, 31610-31622.

[287] R. C. D., S. Erik, M. Ryan, F. V. V., P. J. C., *Angew Chem. Int. Ed. Engl.* **2012**, *51*, 11014-11018.

[288] C. D. Rillahan, E. Schwartz, C. Rademacher, R. McBride, J. Rangarajan, V. V. Fokin, J. C. Paulson, *ACS Chem. Biol.* **2013**, *8*, 1417-1422.

[289] P. J. Coombs, M. E. Taylor, K. Drickamer, *Glycobiology* **2006**, *16*, 1C-7C.

[290] C. Galustian, C. G. Park, W. Chai, M. Kiso, S. A. Bruening, Y. S. Kang, R. M. Steinman, T. Feizi, *Infect. Immun.* **2004**, *16*, 853-866.

[291] S. K. Singh, I. Streng-Ouwehand, M. Litjens, D. R. Weelij, J. J. García-Vallejo, S. J. van Vliet, E. Saeland, Y. van Kooyk, *Mol. Immunol.* **2009**, *46*, 1240-1249.

[292] A. S. Powlesland, E. M. Ward, S. K. Sadhu, Y. Guo, M. E. Taylor, K. Drickamer, *J. Biol. Chem.* **2006**, *281*, 20440-20449.

[293] M. Maglinao, M. Eriksson, M. K. Schlegel, S. Zimmermann, T. Johannssen, S. Götze, P. H. Seeberger, B. Lepenies, *J. Control. Release* **2014**, *175*, 36-42.

[294] S. J. van Vliet, E. van Liempt, E. Saeland, C. A. Aarnoudse, B. Appelmelk, T. Irimura, T. B. H. Geijtenbeek, O. Blixt, R. Alvarez, I. van Die, Y. van Kooyk, *Int. Immunol.* **2005**, *17*, 661-669.

[295] A. S. Palma, T. Feizi, Y. Zhang, M. S. Stoll, A. M. Lawson, E. Díaz-Rodríguez, M. A. Campanero-Rhodes, J. Costa, S. Gordon, G. D. Brown, W. Chai, *J. Biol. Chem.* **2006**, *281*, 5771-5779.

[296] A. S. Powlesland, T. Fisch, M. E. Taylor, D. F. Smith, B. Tissot, A. Dell, S. Pöhlmann, K. Drickamer, *J. Biol. Chem.* **2008**, *283*, 593-602.

[297] S. Yamasaki, M. Matsumoto, O. Takeuchi, T. Matsuzawa, E. Ishikawa, M. Sakuma, H. Tateno, J. Uno, J. Hirabayashi, Y. Mikami, K. Takeda, S. Akira, T. Saito, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 1897.
[298] S.-K. Wang, P.-H. Liang, R. D. Astronomo, T.-L. Hsu, S.-L. Hsieh, D. R. Burton, C.-H. Wong, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3690-3695.
[299] M. Ciobanu, K.-T. Huang, J.-P. Daguer, S. Barluenga, O. Chaloin, E. Schaeffer, C. G. Mueller, D. A. Mitchell, N. Winssinger, *ChemComm* **2011**, *47*, 9321-9323.
[300] K. Brzezicka, B. Echeverria, S. Serna, A. van Diepen, C. H. Hokke, N.-C. Reichardt, *ACS Chem. Biol.* **2015**, *10*, 1290-1302.
[301] Y. Rikio, T. Hiroaki, H. Jun, *FEBS J.* **2010**, *277*, 4010-4026.
[302] E. P. McGreal, M. Rosas, G. D. Brown, S. Zamze, S. Y. C. Wong, S. Gordon, L. Martinez-Pomares, P. R. Taylor, *Glycobiology* **2006**, *16*, 422-430.
[303] J. Y. Hyun, J. Pai, I. Shin, *Acc. Chem. Res.* **2017**, *50*, 1069-1078.
[304] a) A. Goudot, G. Pourceau, A. Meyer, T. Gehin, S. Vidal, J.-J. Vasseur, F. Morvan, E. Souteyrand, Y. Chevolot, *Biosens. Bioelectron.* **2013**, *40*, 153-160; b) P.-H. Liang, S.-K. Wang, C.-H. Wong, *J. Am. Chem. Soc.* **2007**, *129*, 11177-11184; c) X. Y. Zhu, B. Holtz, Y. Wang, L.-X. Wang, P. E. Orndorff, A. Guo, *J. Am. Chem. Soc.* **2009**, *131*, 13646-13650; d) S. Park, M.-R. Lee, I. Shin, *Bioconjugate Chem.* **2009**, *20*, 155-162; e) X. Song, B. Xia, Y. Lasanajak, D. F. Smith, R. D. Cummings, *Glycoconj J.* **2008**, *25*, 15-25.
[305] a) D. Wang, S. Liu, B. J. Trummer, C. Deng, A. Wang, *Nat. Biotechnol.* **2002**, *20*, 275; b) H. Shinya, G. Sebastian, L. Yan, I. Akemi, K. A. Kyoko, T. Naoyuki, V. S. Daniel, F. Ten, S. P. H., Y. Yoshiki, *ChemBioChem* **2015**, *16*, 1502-1511; c) S. Fukui, T. Feizi, C. Galustian, A. M. Lawson, W. Chai, *Nat. Biotechnol.* **2002**, *20*, 1011; d) A. S. Palma, T. Feizi, R. A. Childs, W. Chai, Y. Liu, *Curr. Opin. Chem. Biol.* **2014**, *18*, 87-94; e) Y. Liu, R. A. Childs, A. S. Palma, M. A. Campanero-Rhodes, M. S. Stoll, W. Chai, T. Feizi, in *Carbohydrate Microarrays: Methods and Protocols* (Ed.: Y. Chevolot), Humana Press, Totowa, NJ, **2012**, pp. 117-136.
[306] a) O. Bohorov, H. Andersson-Sand, J. Hoffmann, O. Blixt, *Glycobiology* **2006**, *16*, 21C-27C; b) Y. Liu, W. Chai, R. A. Childs, T. Feizi, in *Methods Enzymol., Vol. 415*, Academic Press, **2006**, pp. 326-340.
[307] X. Song, Y. Lasanajak, B. Xia, D. F. Smith, R. D. Cummings, *ACS Chem. Biol.* **2009**, *4*, 741-750.
[308] Y. Liu, T. Feizi, M. A. Campanero-Rhodes, R. A. Childs, Y. Zhang, B. Mulloy, P. G. Evans, H. M. I. Osborn, D. Otto, P. R. Crocker, W. Chai, *Chem. Biol.* **2007**, *14*, 847-859.
[309] M.-r. Lee, I. Shin, *Org. Lett.* **2005**, *7*, 4269-4272.
[310] B. Xia, Z. S. Kawar, T. Ju, R. A. Alvarez, G. P. Sachdev, R. D. Cummings, *Nat. Methods* **2005**, *2*, 845.
[311] S. Deborah, S. Frank, G. Andreas, S. P. H., P. Fabian, *Chem. Eur. J.* **2015**, *21*, 5709-5713.
[312] T. J. Boltje, T. Buskas, G.-J. Boons, *Nat. Chem.* **2009**, *1*, 611.
[313] O. Blixt, K. Allin, O. Bohorov, X. Liu, H. Andersson-Sand, J. Hoffmann, N. Razi, *Glycoconj J.* **2008**, *25*, 59-68.
[314] a) O. Oyelaran, Q. Li, D. Farnsworth, J. C. Gildersleeve, *J. Proteome Res.* **2009**, *8*, 3529-3538; b) L. Wang, R. D. Cummings, D. F. Smith, M. Huflejt, C. T. Campbell, J. C. Gildersleeve, J. Q. Gerlach, M. Kilcoyne, L. Joshi, S. Serna, N.-C. Reichardt, N. Parera Pera, R. J. Pieters, W. Eng, L. K. Mahal, *Glycobiology* **2014**, *24*, 507-517.
[315] G. B. Thomas, L. H. Rader, J. Park, L. Abezgauz, D. Danino, P. DeShong, D. S. English, *J. Am. Chem. Soc.y* **2009**, *131*, 5471-5477.
[316] Y. Chevolot, C. Bouillon, S. Vidal, F. Morvan, A. Meyer, J. P. Cloarec, A. Jochum, J. P. Praly, J. J. Vasseur, E. Souteyrand, *Angew Chem. Int. Ed. Engl.* **2007**, *46*, 2398-2402.
[317] K. Godula, C. R. Bertozzi, *J. Am. Chem. Soc.* **2012**, *134*, 15732-15742.
[318] N. K. Kamisetty, S. P. Pack, M. Nonogawa, K. C. Devarayapalli, T. Kodaki, K. Makino, *Anal. Bioanal. Chem.* **2006**, *386*, 1649-1655.
[319] a) M. C. Bryan, O. Plettenburg, P. Sears, D. Rabuka, S. Wacowich-Sgarbi, C.-H. Wong, *Chem. Biol.* **2002**, *9*, 713-720; b) F. Fazio, M. C. Bryan, O. Blixt, J. C. Paulson, C.-H. Wong, *J. Am. Chem. Soc.* **2002**, *124*, 14397-14402.
[320] a) S. Jing, M. Milan, *Angew Chem. Int. Ed. Engl.* **2002**, *41*, 4715-4718; b) Z. Zheng-Liang, L. Nicolas, P. A. K., K. Rositsa, F. Margherita, V. Josef, W. Adam, B. J. M., C. P. R., R. D. A., F. Sabine, F. R. A., T. J. E., *ChemBioChem* **2008**, *9*, 1568-1575.

[321] a) R. Karamanska, J. Clarke, O. Blixt, J. I. MacRae, J. Q. Zhang, P. R. Crocker, N. Laurent, A. Wright, S. L. Flitsch, D. A. Russell, R. A. Field, *Glycoconj J.* **2008**, *25*, 69-74; b) A. R. de Boer, C. H. Hokke, A. M. Deelder, M. Wuhrer, *Glycoconj J.* **2008**, *25*, 75-84.

[322] a) B. Ana, C. Javier, S. Sonia, Y. Shi, W. I. B. H., M. L. Manuel, R. N. Christian, *Angew Chem. Int. Ed. Engl.* **2013**, *52*, 7477-7481; b) T. W. Powers, B. A. Neely, Y. Shao, H. Tang, D. A. Troyer, A. S. Mehta, B. B. Haab, R. R. Drake, *PLoS One* **2014**, *9*, e106255.

[323] a) W. G. T. Willats, S. E. Rasmussen, T. Kristensen, J. D. Mikkelsen, J. P. Knox, *Proteomics* **2002**, *2*, 1666-1671; b) J. L. Kanter, S. Narayana, P. P. Ho, I. Catz, K. G. Warren, R. A. Sobel, L. Steinman, W. H. Robinson, *Nat. Med.* **2005**, *12*, 138.

[324] a) A. S. Palma, Y. Liu, H. Zhang, Y. Zhang, B. V. McCleary, G. Yu, Q. Huang, L. S. Guidolin, A. E. Ciocchini, A. Torosantucci, *Mol. Cell. Proteomics* **2015**, mcp. M115. 048272; b) T. Feizi, W. Chai, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 582.

[325] a) E. L. Shipp, L. C. Hsieh-Wilson, *Chem. Biol.* **2007**, *14*, 195-208; b) C. Olivier, L. Julien, D. Guy, D. C. Latifa, C. José, M. Oleg, *ChemBioChem* **2006**, *7*, 817-826; c) C. J. Rogers, P. M. Clark, S. E. Tully, R. Abrol, K. C. Garcia, W. A. Goddard, L. C. Hsieh-Wilson, *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 9747.

[326] a) K.-S. Ko, F. A. Jaipuri, N. L. Pohl, *J. Am. Chem. Soc.* **2005**, *127*, 13162-13163; b) S. K. Mamidyala, K.-S. Ko, F. A. Jaipuri, G. Park, N. L. Pohl, *J. Fluorine Chem.* **2006**, *127*, 571-579; c) G.-S. Chen, N. L. Pohl, *Org. Lett.* **2008**, *10*, 785-788.

[327] O. Galanina, M. Mecklenburg, N. Nifantiev, G. Pazynina, N. Bovin, *Lab Chip* **2003**, *3*, 260-265.

[328] S. R. Antonio, S. Sonia, R. Nerea, M. L. Manuel, R. Niels-Christian, *Angew Chem. Int. Ed. Engl.* **2011**, *50*, 1801-1804.

[329] A. R. de Boer, C. H. Hokke, A. M. Deelder, M. Wuhrer, *Anal. Chem.* **2007**, *79*, 8107-8113.

[330] a) S. E. Tully, M. Rawat, L. C. Hsieh-Wilson, *J. Am. Chem. Soc.* **2006**, *128*, 7740-7741; b) J. L. de Paz, D. Spillmann, P. H. Seeberger, *ChemComm* **2006**, 3116-3118.

[331] a) M. Schwarz, L. Spector, A. Gargir, A. Shtevi, M. Gortler, R. T. Altstock, A. A. Dukler, N. Dotan, *Glycobiology* **2003**, *13*, 749-754; b) L. Nimrichter, A. Gargir, M. Gortler, R. T. Altstock, A. Shtevi, O. Weisshaus, E. Fire, N. Dotan, R. L. Schnaar, *Glycobiology* **2004**, *14*, 197-203.

[332] a) S. Park, M.-r. Lee, S.-J. Pyo, I. Shin, *J. Am. Chem. Soc.* **2004**, *126*, 4812-4819; b) P. Sungjin, S. Injae, *Angew Chem. Int. Ed. Engl.* **2002**, *41*, 3180-3182; c) B. T. Houseman, E. S. Gawalt, M. Mrksich, *Langmuir* **2003**, *19*, 1522-1531; d) B. M. A., D. M. D., S. P. H., *ChemBioChem* **2006**, *7*, 421-424.

[333] G. Sebastian, A. Nahid, T. Yu-Hsuan, G. Uwe, R. Anika, A. Chakkumkal, S. P. H., V. S. Daniel, *Angew Chem. Int. Ed. Engl.* **2014**, *53*, 13701-13705.

[334] a) L. Myung-ryul, S. Injae, *Angew Chem. Int. Ed. Engl.* **2005**, *44*, 2881-2884; b) S. Park, M.-R. Lee, I. Shin, *Nat. Protoc.* **2007**, *2*, 2747; c) S. Park, M.-R. Lee, I. Shin, in *Small Molecule Microarrays: Methods and Protocols* (Eds.: M. Uttamchandani, S. Q. Yao), Humana Press, Totowa, NJ, **2010**, pp. 195-208.

[335] Y. Zhang, C. Campbell, Q. Li, J. C. Gildersleeve, *Mol. BioSyst.* **2010**, *6*, 1583-1591.

[336] E. A. Smith, W. D. Thomas, L. L. Kiessling, R. M. Corn, *J. Am. Chem. Soc.* **2003**, *125*, 6140-6148.

[337] L. G. Harris, W. C. E. Schofield, K. J. Doores, B. G. Davis, J. P. S. Badyal, *J. Am. Chem. Soc.* **2009**, *131*, 7755-7761.

[338] B. T. Houseman, M. Mrksich, *Chem. Biol.* **2002**, *9*, 443-454.

[339] H. S. G. Beckmann, A. Niederwieser, M. Wiessler, V. Wittmann, *Chem. Eur. J.* **2012**, *18*, 6548-6554.

[340] K. H. C., F. M. G., S. K. Barry, *Angew Chem. Int. Ed. Engl.* **2001**, *40*, 2004-2021.

[341] a) M. C. Bryan, F. Fazio, H.-K. Lee, C.-Y. Huang, A. Chang, M. D. Best, D. A. Calarese, O. Blixt, J. C. Paulson, D. Burton, I. A. Wilson, C.-H. Wong, *J. Am. Chem. Soc.* **2004**, *126*, 8640-8641; b) O. Michel, B. J. Ravoo, *Langmuir* **2008**, *24*, 12116-12118; c) X.-L. Sun, C. L. Stabler, C. S. Cazalis, E. L. Chaikof, *Bioconjugate Chem.* **2006**, *17*, 52-57.

[342] O. J. Barrett, A. Pushechnikov, M. Wu, M. D. Disney, *Carbohydr. Res.* **2008**, *343*, 2924-2931.

[343] K. Godula, D. Rabuka, K. T. Nam, C. R. Bertozzi, *Angew Chem. Int. Ed. Engl.* **2009**, *48*, 4973-4976.

[344] C. Zilio, A. Bernardi, A. Palmioli, M. Salina, G. Tagliabue, M. Buscaglia, R. Consonni, M. Chiari, *Sens. Actuators B Chem.* **2015**, *215*, 412-420.

[345] a) Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless, M. G. Finn, *J. Am. Chem. Soc.* **2003**, *125*, 3192-3193; b) J. Gierlich, G. A. Burley, P. M. E. Gramlich, D. M. Hammond, T. Carell, *Org. Lett.* **2006**, *8*, 3639-3642.
[346] D. Prim, M. E. Pfeifer, in *Peptide Microarrays: Methods and Protocols* (Eds.: M. Cretich, M. Chiari), Springer New York, New York, NY, **2016**, pp. 157-166.
[347] K. Maja, W. Ron, P. Carsten, S. Hendrik, S. Laurent, B. Rolf, N. C. M., W. Herbert, *Angew Chem. Int. Ed. Engl.* **2003**, *42*, 5830-5834.
[348] a) Z. Pei, H. Yu, M. Theurer, A. Waldén, P. Nilsson, M. Yan, O. Ramström, *ChemBioChem* **2007**, *8*, 166-168; b) A. Tyagi, X. Wang, L. Deng, O. Ramström, M. Yan, *Biosens. Bioelectron.* **2010**, *26*, 344-350.
[349] M. Shipp, R. Nadella, H. Gao, V. Farkas, H. Sigrist, A. Faik, *Glycoconj J.* **2008**, *25*, 49-58.
[350] a) W. Denong, C. G. T., T. N. J., K. J. T., K. Pavol, S. Rina, A. Roberto, H. L. A., H. L. A., S. Lawrence, *Proteomics* **2007**, *7*, 180-184; b) G. T. Carroll, D. Wang, N. J. Turro, J. T. Koberstein, *Langmuir* **2006**, *22*, 2899-2905.
[351] H.-Y. Hsiao, M.-L. Chen, H.-T. Wu, L.-D. Huang, W.-T. Chien, C.-C. Yu, F.-D. Jan, S. Sahabuddin, T.-C. Chang, C.-C. Lin, *ChemComm* **2011**, *47*, 1187-1189.
[352] J. B. Delehanty, in *Protein Arrays: Methods and Protocols* (Ed.: E. T. Fung), Humana Press, Totowa, NJ, **2004**, pp. 135-143.
[353] A. Faulkner, W. Shu, *Collegium* **2012**, *35*.
[354] a) H. Tateno, A. Mori, N. Uchiyama, R. Yabe, J. Iwaki, T. Shikanai, T. Angata, H. Narimatsu, J. Hirabayashi, *Glycobiology* **2008**, *18*, 789-798; b) K. Takahara, T. Arita, S. Tokieda, N. Shibata, Y. Okawa, H. Tateno, J. Hirabayashi, K. Inaba, *Infect. Immun.* **2012**, *80*, 1699-1706.
[355] Y.-J. Chuang, X. Zhou, Z. Pan, C. Turchi, *Biochem. Biophys. Res. Commun.* **2009**, *389*, 22-27.
[356] E. Han, L. Ding, S. Jin, H. Ju, *Biosens. Bioelectron.* **2011**, *26*, 2500-2505.
[357] M. Fais, R. Karamanska, S. Allman, S. A. Fairhurst, P. Innocenti, A. J. Fairbanks, T. J. Donohoe, B. G. Davis, D. A. Russell, R. A. Field, *Chem. Sci.* **2011**, *2*, 1952-1959.
[358] L. Ban, N. Pettit, L. Li, A. D. Stuparu, L. Cai, W. Chen, W. Guan, W. Han, P. G. Wang, M. Mrksich, *Nat. Chem. Biol.* **2012**, *8*, 769.
[359] a) T. Horlacher, P. H. Seeberger, *Chem. Soc. Rev.* **2008**, *37*, 1414-1422; b) R. Šardzík, A. P. Green, N. Laurent, P. Both, C. Fontana, J. Voglmeir, M. J. Weissenborn, R. Haddoub, P. Grassi, S. M. Haslam, G. Widmalm, S. L. Flitsch, *J. Am. Chem. Soc.* **2012**, *134*, 4521-4524.
[360] O. Blixt, S. Han, L. Liao, Y. Zeng, J. Hoffmann, S. Futakawa, J. C. Paulson, *J. Am. Chem. Soc.* **2008**, *130*, 6680-6681.
[361] Y. Fei, Y.-S. Sun, Y. Li, H. Yu, K. Lau, J. Landry, Z. Luo, N. Baumgarth, X. Chen, X. Zhu, *Biomolecules* **2015**, *5*, 1480.
[362] J. Baader, H. Klapproth, S. Bednar, T. Brandstetter, J. Rühe, M. Lehmann, I. Freund, *Biosens. Bioelectron.* **2011**, *26*, 1839-1846.
[363] K. Gruber, T. Horlacher, R. Castelli, A. Mader, P. H. Seeberger, B. A. Hermann, *ACS Nano* **2011**, *5*, 3670-3678.
[364] Y. Y. Fei, A. Schmidt, G. Bylund, D. X. Johansson, S. Henriksson, C. Lebrilla, J. V. Solnick, T. Borén, X. D. Zhu, *Anal. Chem.* **2011**, *83*, 6336-6341.
[365] Y. Guo, C. Sakonsinsiri, I. Nehlmeier, M. A. Fascione, H. Zhang, W. Wang, S. Pöhlmann, W. B. Turnbull, D. Zhou, *Angew Chem. Int. Ed. Engl.* **2016**, *55*, 4738-4742.
[366] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew Chem. Int. Ed. Engl.* **2002**, *114*, 2708-2711.
[367] E. Saxon, C. R. Bertozzi, *Science* **2000**, *287*, 2007.
[368] a) N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126*, 15046-15047; b) E. M. Sletten, C. R. Bertozzi, *Angew Chem. Int. Ed. Engl.* **2009**, *48*, 6974-6998.
[369] a) D. H. Ess, G. O. Jones, K. N. Houk, *Org. Lett.* **2008**, *10*, 1633-1636; b) F. Schoenebeck, D. H. Ess, G. O. Jones, K. N. Houk, *J. Am. Chem. Soc.* **2009**, *131*, 8121-8133.
[370] N. J. Agard, J. M. Baskin, J. A. Prescher, A. Lo, C. R. Bertozzi, *ACS Chem. Biol.* **2006**, *1*, 644-648.
[371] X. Ning, J. Guo, M. A. Wolfert, G. J. Boons, *Angew Chem. Int. Ed. Engl.* **2008**, *47*, 2253-2255.

[372] D. Jan, S. Samuel, T. Rinske, H. L. J. A., R. F. P. J. T., v. H. J. C. M., L. D. J., F. Peter, v. D. F. L., *Angew Chem. Int. Ed. Engl.* **2010**, *49*, 9422-9425.
[373] A. Kuzmin, A. Poloukhtine, M. A. Wolfert, V. V. Popik, *Bioconjugate Chem.* **2010**, *21*, 2076-2085.
[374] O. C. Grant, H. M. K. Smith, D. Firsova, E. Fadda, R. J. Woods, *Glycobiology* **2014**, *24*, 17-25.
[375] a) S. Jung, H. Yi, *Biomacromolecules* **2013**, *14*, 3892-3902; b) D. Prim, F. Rebeaud, V. Cosandey, R. Marti, P. Passeraub, M. Pfeifer, *Molecules* **2013**, *18*, 9833; c) S. V. Orski, A. A. Poloukhtine, S. Arumugam, L. Mao, V. V. Popik, J. Locklin, *J. Am. Chem. Soc.* **2010**, *132*, 11024-11026.
[376] R. K. Manova, S. P. Pujari, C. A. G. M. Weijers, H. Zuilhof, T. A. van Beek, *Langmuir* **2012**, *28*, 8651-8663.
[377] a) Y. Deng, X. Y. Zhu, T. Kienlen, A. Guo, *J. Am. Chem. Soc.* **2006**, *128*, 2768-2769; b) R. D. Deegan, O. Bakajin, T. F. Dupont, G. Huber, S. R. Nagel, T. A. Witten, *Phys. Rev. E* **2000**, *62*, 756-765; c) E. Adachi, A. S. Dimitrov, K. Nagayama, *Langmuir* **1995**, *11*, 1057-1060; d) R. D. Deegan, O. Bakajin, T. F. Dupont, G. Huber, S. R. Nagel, T. A. Witten, *Nature* **1997**, *389*, 827.
[378] a) H. B. Eral, D. M. Augustine, M. H. Duits, F. Mugele, *Soft Matter* **2011**, *7*, 4954-4958; b) H. Hu, R. G. Larson, *J. Phys. Chem. B* **2006**, *110*, 7090-7094.
[379] a) M. Dufva, *Biomol. Eng.* **2005**, *22*, 173-184; b) K. Pappaert, H. Ottevaere, H. Thienpont, P. Van Hummelen, G. Desmet, *BioTechniques* **2006**, *41*, 609-618.
[380] a) D. Henri, D. Dominique, S. Gérard, S. Geneviève, M. Jean, *Eur. J. Biochem.* **1981**, *117*, 41-51; b) E. Miyoshi, K. Moriwaki, T. Nakagawa, *J. Biochem.* **2008**, *143*, 725-729.
[381] J. Olausson, L. Tibell, B.-H. Jonsson, P. Påhlsson, *Glycoconj J.* **2008**, *25*, 753.
[382] a) E. Mitchell, C. Houles, D. Sudakevitz, M. Wimmerova, C. Gautier, S. Pérez, A. M. Wu, N. Gilboa-Garber, A. Imberty, *Nat. Struct. Mol. Biol.* **2002**, *9*, 918; b) A. Imberty, M. Wimmerová, E. P. Mitchell, N. Gilboa-Garber, *Microbes and Infection* **2004**, *6*, 221-228.
[383] J. R. Govan, V. Deretic, *Microbiol. Rev.* **1996**, *60*, 539-574.
[384] a) P. Von Bismarck, R. Schneppenheim, U. Schumacher, *Klin. Pädiatr.* **2001**, *213*, 285-287; b) S. Wagner, R. Sommer, S. Hinsberger, C. Lu, R. W. Hartmann, M. Empting, A. Titz, *J. Med. Chem.* **2016**, *59*, 5929-5969.
[385] a) A. Engering, S. J. van Vliet, K. Hebeda, D. G. Jackson, R. Prevo, S. K. Singh, T. B. H. Geijtenbeek, H. van Krieken, Y. van Kooyk, *Am. J. Pathol.* **2004**, *164*, 1587-1595; b) V. S. F. Chan, K. Y. K. Chan, Y. Chen, L. L. M. Poon, A. N. Y. Cheung, B. Zheng, K.-H. Chan, W. Mak, H. Y. S. Ngan, X. Xu, G. Screaton, P. K. H. Tam, J. M. Austyn, L.-C. Chan, S.-P. Yip, M. Peiris, U.-S. Khoo, C.-L. S. Lin, *Nat. Genet.* **2005**, *38*, 38; c) S. A. Jeffers, S. M. Tusell, L. Gillim-Ross, E. M. Hemmila, J. E. Achenbach, G. J. Babcock, W. D. Thomas, L. B. Thackray, M. D. Young, R. J. Mason, D. M. Ambrosino, D. E. Wentworth, J. C. DeMartini, K. V. Holmes, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15748.
[386] a) E. J. Soilleux, R. Barten, J. Trowsdale, *J. Immunol.* **2000**, *165*, 2937; b) E. J. Soilleux, *Clin. Sci.* **2003**, *104*, 437.
[387] J. Auwerx, K. O. François, E. Vanstreels, K. Van Laethem, D. Daelemans, D. Schols, J. Balzarini, *Antivir. Res.* **2009**, *83*, 61-70.
[388] a) Y. Li, B. Hao, X. Kuai, G. Xing, J. Yang, J. Chen, L. Tang, L. Zhang, F. He, *Mol. Cell. Biochem.* **2009**, *327*, 183-190; b) J. P. Gardner, R. J. Durso, R. R. Arrigale, G. P. Donovan, P. J. Maddon, T. Dragic, W. C. Olson, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4498; c) W. K. Lai, P. J. Sun, J. Zhang, A. Jennings, P. F. Lalor, S. Hubscher, J. A. McKeating, D. H. Adams, *Am. J. Pathol.* **2006**, *169*, 200-208.
[389] E. A. Koppel, I. S. Ludwig, M. Sanchez Hernandez, T. L. Lowary, R. R. Gadikota, A. B. Tuzikov, C. M. J. E. Vandenbroucke-Grauls, Y. van Kooyk, B. J. Appelmelk, T. B. H. Geijtenbeek, *Immunobiology* **2004**, *209*, 117-127.
[390] M. V. Carroll, R. B. Sim, F. Bigi, A. Jäkel, R. Antrobus, D. A. Mitchell, *Protein Cell* **2010**, *1*, 859-870.
[391] S. L. Londrigan, S. G. Turville, M. D. Tate, Y.-M. Deng, A. G. Brooks, P. C. Reading, *J. Virol.* **2011**, *85*, 2990-3000.
[392] I. S. Ludwig, A. N. Lekkerkerker, E. Depla, F. Bosman, R. J. Musters, S. Depraetere, Y. van Kooyk, T. B. Geijtenbeek, *J. Virol.* **2004**, *78*, 8322-8332.
[393] K. E. A., V. G. K. P. J. M., G. T. B. H., V. K. Yvette, *Cell. Microbiol.* **2005**, *7*, 157-165.
[394] H. Feinberg, C. K. W. Tso, M. E. Taylor, K. Drickamer, W. I. Weis, *J. Mol. Biol.* **2009**, *394*, 613-620.

[395] a) D. Zhu, A. Kawana-Tachikawa, A. Iwamoto, Y. Kitamura, *Biochem. Biophys. Res. Commun.* **2010**, *393*, 598-602; b) T. Gramberg, T. Zhu, C. Chaipan, A. Marzi, H. Liu, A. Wegele, T. Andrus, H. Hofmann, S. Pöhlmann, *Virology* **2006**, *347*, 354-363; c) S. Liu, H. Weng, J. Wu, Z. Zhang, Y. Zeng, H. Tian, *Chinese Journal of Experimental and Clinical Virology* **2011**, *25*, 331-333; d) M. Wang, H. Han, J. Lu, S. Liu, Q. Jiang, *Chinese Journal of Hepatology* **2009**, *17*, 645-648; e) d. S. Ália, H. Andreas, O. Felipe, L. N. K., T. Aylin, T. M. E., D. Kurt, *Protein Sci.* **2017**, *26*, 306-316.

[396] a) H. Liu, M. Carrington, C. Wang, S. Holte, J. Lee, B. Greene, F. Hladik, D. M. Koelle, A. Wald, K. Kurosawa, C. R. Rinaldo, C. Celum, R. Detels, L. Corey, M. J. McElrath, T. Zhu, *J. Infect. Dis.* **2006**, *193*, 698-702; b) C. W. Davis, H.-Y. Nguyen, S. L. Hanna, M. D. Sánchez, R. W. Doms, T. C. Pierson, *J. Virol.* **2006**, *80*, 1290-1301; c) L. Psylvia, T. Marilou, Y. Berthe, C. Nicole, R. R. N., F. Marie, L. Pierre-Yves, *Traffic* **2016**, *17*, 639-656.

[397] a) N. E. Welty, C. Staley, N. Ghilardi, M. J. Sadowsky, B. Z. Igyártó, D. H. Kaplan, *J. Exp. Med.* **2013**, *210*, 2011; b) B. Venetia, M. Naomi, M. Paul, D. Rachel, P. Sarah, C. Sharon, H. Muzlifah, C. Matthew, *J. Leukoc. Biol.* **2015**, *97*, 627-634; c) C. G. Figdor, Y. van Kooyk, G. J. Adema, *Nat. Rev. Immunol.* **2002**, *2*, 77.

[398] V. Michiel, G. T. B. H, *Immunol. Cell Biol.* **2010**, *88*, 410-415.

[399] M. van der Vlist, L. de Witte, R. D. de Vries, M. Litjens, M. A. W. P. de Jong, D. Fluitsma, R. L. de Swart, T. B. H. Geijtenbeek, *Eur. J. Immunol.* **2011**, *41*, 2619-2631.

[400] H. J. Kim, P. J. Brennan, D. Heaslip, M. C. Udey, R. L. Modlin, J. T. Belisle, *J. Bacteriol.* **2015**, *197*, 615-625.

[401] M. A. W. P. de Jong, L. E. M. Vriend, B. Theelen, M. E. Taylor, D. Fluitsma, T. Boekhout, T. B. H. Geijtenbeek, *Mol. Immunol.* **2010**, *47*, 1216-1225.

[402] J. Hanske, J. Schulze, J. Aretz, R. McBride, B. Loll, H. Schmidt, Y. Knirel, W. Rabsch, M. C. Wahl, J. C. Paulson, C. Rademacher, *J. Biol. Chem.* **2017**, *292*, 862-871.

[403] N. S. Stambach, M. E. Taylor, *Glycobiology* **2003**, *13*, 401-410.

[404] H. Tateno, K. Ohnishi, R. Yabe, N. Hayatsu, T. Sato, M. Takeya, H. Narimatsu, J. Hirabayashi, *J. Biol. Chem.* **2010**, *285*, 6390-6400.

[405] H. Feinberg, T. J. W. Rowntree, S. L. W. Tan, K. Drickamer, W. I. Weis, M. E. Taylor, *J. Biol. Chem.* **2013**, *288*, 36762-36771.

[406] L. M. van den Berg, S. Cardinaud, A. M. G. van der Aar, J. K. Sprokholt, M. A. W. P. de Jong, E. M. Zijlstra-Willems, A. Moris, T. B. H. Geijtenbeek, *J. Immunol.* **2015**, *195*, 1763.

[407] R. M. Dermott, U. Ziylan, D. Spehner, H. Bausinger, D. Lipsker, M. Mommaas, J.-P. Cazenave, G. Raposo, B. Goud, H. d. l. Salle, J. Salamero, D. Hanau, S. R. Pfeffer, *Mol. Biol. Cell* **2002**, *13*, 317-335.

[408] R. Nikolaus, C. B. E., S. Patrizia, *Immunol. Rev.* **2010**, *234*, 120-141.

[409] I. W. K. Eddie, K. Takahashi, E. R. Alan, L. M. Stuart, *Immunol. Rev.* **2009**, *230*, 9-21.

[410] T. Fujita, *Nat. Rev. Immunol.* **2002**, *2*, 346.

[411] D. József, P. Gábor, C. László, G. Péter, *Immunol. Rev.* **2016**, *274*, 98-111.

[412] N. A. J., R. J. Nicoline, R. Anja, D. M. R., M. H. O., B. E. M. C., R. L. P., K. Claus, G. Peter, *Eur. J. Immunol.* **2003**, *33*, 2853-2863.

[413] G. Raffaella, S. Claudio, F. Alessandro, S. Matteo, Z. E. R., O. Franca, V. Helene, M. M. L. Maurice, Z. Bertjan, G. Marco, D. S. Maria-Grazia, *Ann. Neurol.* **2009**, *66*, 332-342.

[414] a) W. I. Weis, K. Drickamer, *Structure* **1994**, *2*, 1227-1240; b) S. Sheriff, C. Y. Chang, R. A. B. Ezekowitz, *Nat. Struct. Mol. Biol.* **1994**, *1*, 789.

[415] P. Garred, F. Larsen, H. O. Madsen, C. Koch, *Mol. Immunol.* **2003**, *40*, 73-84.

[416] N. Brouwer, K. M. Dolman, M. van Houdt, M. Sta, D. Roos, T. W. Kuijpers, *J. Immunol.* **2008**, *180*, 4124.

[417] a) D. P. Eisen, R. M. Minchinton, *Clin. Infect. Dis.* **2003**, *37*, 1496-1505; b) E. M. Anders, C. A. Hartley, D. C. Jackson, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 4485; c) M. C. Ghezzi, G. Raponi, S. Angeletti, C. Mancini, *J. Infect. Dis.* **1998**, *178*, 1743-1749; d) J. A. Willment, G. D. Brown, *Trends Microbiol.* **2008**, *16*, 27-32.

[418] M. J. Fernandes, A. A. Finnegan, L. D. Siracusa, C. Brenner, N. N. Iscove, B. Calabretta, *Cancer Res.* **1999**, *59*, 2709.

[419] I. M. Dambuza, G. D. Brown, *Curr. Opin. Immunol.* **2015**, *32*, 21-27.

[420] a) T. Ishikawa, F. Itoh, S. Yoshida, S. Saijo, T. Matsuzawa, T. Gonoi, T. Saito, Y. Okawa, N. Shibata, T. Miyamoto, S. Yamasaki, *Cell Host Microbe* **2013**, *13*, 477-488; b) A. Yonekawa, S. Saijo, Y. Hoshino, Y. Miyake, E. Ishikawa, M. Suzukawa, H. Inoue, M. Tanaka, M. Yoneyama, M. Oh-hora, K. Akashi, S. Yamasaki, *Immunity* **2014**, *41*, 402-413; c) A. Wittmann, D. Lamprinaki, K. M. Bowles, E. Katzenellenbogen, Y. A. Knirel, C. Whitfield, T. Nishimura, N. Matsumoto, K. Yamamoto, Y. Iwakura, S. Saijo, N. Kawasaki, *J. Biol. Chem.* **2016**, *291*, 17629-17638.

[421] a) B. Kerscher, J. A. Willment, G. D. Brown, *Int. Immunol.* **2013**, *25*, 271-277; b) D. C. Ifrim, J. M. Bain, D. M. Reid, M. Oosting, I. Verschueren, N. A. Gow, J. H. van Krieken, G. D. Brown, B.-J. Kullberg, L. A. Joosten, *Infect. Immun.* **2014**, *82*, 1064-1073; c) M. Ritter, O. Gross, S. Kays, J. Ruland, F. Nimmerjahn, S. Saijo, J. Tschopp, L. E. Layland, C. Prazeres da Costa, *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 20459; d) M. W. Parsons, L. Li, A. M. Wallace, M. J. Lee, H. R. Katz, J. M. Fernandez, S. Saijo, Y. Iwakura, K. F. Austen, Y. Kanaoka, N. A. Barrett, *J. Immunol.* **2014**, *192*, 1361; e) E. Vinogradov, E. Frirdich, L. L. MacLean, M. B. Perry, B. O. Petersen, J. Ø. Duus, C. Whitfield, *J. Biol. Chem.* **2002**, *277*, 25070-25081; f) E. Katzenellenbogen, N. A. Kocharova, G. V. Zatonsky, J. Kübler-Kielb, A. Gamian, A. S. Shashkov, Y. A. Knirel, E. Romanowska, *FEMS Immunol. Med. Microbiol.* **2001**, *30*, 223-227.

[422] a) M. J. Marakalala, L. M. Graham, G. D. Brown, *Clin. Dev. Immunol.* **2011**, *2010*; b) K. Sato, X.-l. Yang, T. Yudate, J.-S. Chung, J. Wu, K. Luby-Phelps, R. P. Kimberly, D. Underhill, P. D. Cruz, K. Ariizumi, *J. Biol. Chem.* **2006**, *281*, 38854-38866.

[423] a) S. Saijo, S. Ikeda, K. Yamabe, S. Kakuta, H. Ishigame, A. Akitsu, N. Fujikado, T. Kusaka, S. Kubo, S.-h. Chung, *Immunity* **2010**, *32*, 681-691; b) D. Sancho, C. R. e. Sousa, *Annu. Rev. Immunol.* **2012**, *30*, 491-529.

[424] H. Feinberg, S. A. F. Jégouzo, M. J. Rex, K. Drickamer, W. I. Weis, M. E. Taylor, *J. Biol. Chem.* **2017**, *292*, 13402-13414.

[425] C. R. Becer, M. I. Gibson, J. Geng, R. Ilyas, R. Wallis, D. A. Mitchell, D. M. Haddleton, *J. Am. Chem. Soc.* **2010**, *132*, 15130-15132.

[426] S. Munneke, K. Kodar, G. F. Painter, B. L. Stocker, M. S. Timmer, *RSC Adv.* **2017**, *7*, 45260-45268.

[427] I. Morbioli, V. Porkolab, A. Magini, A. Casnati, F. Fieschi, F. Sansone, *Carbohydr. Res.* **2017**, *453*, 36-43.

[428] a) M. Á. Olga, H. Karolin, M. Marco, C. Caroline, C. Colin, K. Charles, P. Soledad, *Chem. Eur. J.* **2009**, *15*, 9874-9888; b) F. Chiodo, M. Marradi, J. Park, A. F. Ram, S. Penadés, I. van Die, B. Tefsen, *ACS Chem. Biol.* **2013**, *9*, 383-389; c) D. Arosio, F. Chiodo, J. J. Reina, M. Marelli, S. Penadés, Y. van Kooyk, J. J. Garcia-Vallejo, A. Bernardi, *Bioconjugate Chem.* **2014**, *25*, 2244-2251.

[429] N. Varga, I. Sutkeviciute, R. Ribeiro-Viana, A. Berzi, R. Ramdasi, A. Daghetti, G. Vettoretti, A. Amara, M. Clerici, J. Rojo, *Biomaterials* **2014**, *35*, 4175-4184.

[430] M. Monsigny, A.-C. Roche, E. Duverger, S. Oruganti, Elsevier, Amsterdam, **2007**.

[431] a) N. S. Abadeer, C. J. Murphy, *J. Phys. Chem. C* **2016**, *120*, 4691-4716; b) W. G. Kreyling, A. M. Abdelmonem, Z. Ali, F. Alves, M. Geiser, N. Haberl, R. Hartmann, S. Hirn, D. J. De Aberasturi, K. Kantner, *Nat. Nanotechnol.* **2015**, *10*, 619; c) W. Zhou, X. Gao, D. Liu, X. Chen, *Chem. Rev.* **2015**, *115*, 10575-10636.

[432] M. Brust, M. Walker, D. Bethell, D. J. Schiffrin, R. Whyman, *ChemComm* **1994**, 801-802.

[433] a) J. Homola, S. S. Yee, G. Gauglitz, *Sens. Actuators B Chem.* **1999**, *54*, 3-15; b) W. L. Barnes, A. Dereux, T. W. Ebbesen, *Nature* **2003**, *424*, 824; c) R. B. Schasfoort, *Handbook of surface plasmon resonance*, Royal Society of Chemistry, **2017**.

[434] A. Holla, A. Skerra, *Protein Eng. Des. Sel.* **2011**, *24*, 659-669.

[435] M. Á. Olga, B. L. M., M. Marco, C. Caroline, A. José, P. Soledad, *ChemBioChem* **2009**, *10*, 1806-1809.

[436] J. Aretz, E.-C. Wamhoff, J. Hanske, D. Heymann, C. Rademacher, *Front. Immunol.* **2014**, *5*, 323.

[437] J. Aretz, H. Baukmann, E. Shanina, J. Hanske, R. Wawrzinek, V. A. Zapol'skii, P. H. Seeberger, D. E. Kaufmann, C. Rademacher, *Angew Chem. Int. Ed. Engl.* **2017**, *56*, 7292-7296.

[438] H. Jhoti, G. Williams, D. C. Rees, C. W. Murray, *Nat. Rev. Drug Discov.* **2013**, *12*, 644.

[439] H. C. Kolb, K. B. Sharpless, *Drug Discov. Today* **2003**, *8*, 1128-1137.

[440] R. U. Lemieux, R. M. Ratcliffe, *Can. J. Chem.* **1979**, *57*, 1244-1251.

[441] W. S. Trahanovsky, M. D. Robbins, *J. Am. Chem. Soc.* **1971**, *93*, 5256-5258.

[442] K. Briner, A. Vasella, *Helv. Chim. Acta* **1987**, *70*, 1341-1356.

[443] a) H. H. Kinfe, University of Cape Town **2006**; b) H. Paulsen, J. P. Lorentzen, W. Kutschker, *Carbohydr. Res.* **1985**, *136*, 153-176.
[444] N. V. Bovin, S. É. Zurabyan, A. Y. Khorlin, *Carbohydr. Res.* **1981**, *98*, 25-35.
[445] S. Czernecki, E. Ayadi, D. Randriamandimby, *J. Org. Chem.* **1994**, *59*, 8256-8260.
[446] a) P. Krist, M. Kuzma, I. F. Pelyvás, P. Simerská, V. Křen, *Collect. Czechoslov. Chem. Commun.* **2003**, *68*, 801-811; b) P. Hans, S. Wolfgang, *Ber. Dtsch. Chem. Ges.* **1978**, *111*, 2334-2347.
[447] D. A. Evans, T. C. Britton, J. A. Ellman, R. L. Dorow, *J. Am. Chem. Soc.* **1990**, *112*, 4011-4030.
[448] J. Dileep Kumar, F.-Y. Dupradeau, M. J. Strouse, M. E. Phelps, T. Toyokuni, *J. Org. Chem.* **2001**, *66*, 3220-3223.
[449] E. Ayadi, S. Czernecki, J. Xie, *Carbohydr. Chem.* **1996**, *15*, 191-199.
[450] a) Y. Ishido, N. Sakairi, *Carbohydr. Res.* **1981**, *97*, 151-155; b) M. Kiso, H. Ishida, A. Kawaide, A. Hasegawa, *Carbohydr. Res.* **1984**, *127*, 129-135; c) J. B. Schwarz, S. D. Kuduk, X.-T. Chen, D. Sames, P. W. Glunz, S. J. Danishefsky, *J. Am. Chem. Soc.* **1999**, *121*, 2662-2673.
[451] a) M. Turský, J. Veselý, I. Tišlerová, T. Trnka, M. Ledvina, *Synthesis* **2008**, *2008*, 2610-2616; b) A. Popelová, K. Kefurt, M. Hlaváčková, J. Moravcová, *Carbohydr. Res.* **2005**, *340*, 161-166.
[452] a) P. Teodorović, R. Slättegård, S. Oscarson, *Carbohydr. Res.* **2005**, *340*, 2675-2676; b) W. Fischer, J. P. Anselme, *J. Am. Chem. Soc.* **1967**, *89*, 5284-5285; c) C. Cavender, V. Shiner Jr, *J. Org. Chem.* **1972**, *37*, 3567-3569.
[453] V. Andrea, W. Christian, C. Jose-Luis, M. L. Manuel, *Helv. Chim. Acta* **1991**, *74*, 2073-2077.
[454] P. B. Alper, S.-C. Hung, C.-H. Wong, *Tetrahedron Lett.* **1996**, *37*, 6029-6032.
[455] A. Titz, Z. Radic, O. Schwardt, B. Ernst, *Tetrahedron Lett.* **2006**, *47*, 2383-2385.
[456] J.-M. Vatèle, S. Hanessian, *Tetrahedron* **1996**, *52*, 10557-10568.
[457] E. D. Goddard-Borger, R. V. Stick, *Org. Lett.* **2007**, *9*, 3797-3800.
[458] N. Fischer, E. D. Goddard-Borger, R. Greiner, T. M. Klapötke, B. W. Skelton, J. Stierstorfer, *J. Org. Chem.* **2012**, *77*, 1760-1764.
[459] G. T. Potter, G. C. Jayson, G. J. Miller, J. M. Gardiner, *J. Org. Chem.* **2016**, *81*, 3443-3446.
[460] a) A. K. Pandiakumar, S. P. Sarma, A. G. Samuelson, *Tetrahedron Lett.* **2014**, *55*, 2917-2920; b) M. Y. Stevens, R. T. Sawant, L. R. Odell, *J. Org. Chem.* **2014**, *79*, 4826-4831.
[461] M. Brandl, M. S. Weiss, A. Jabs, J. Sühnel, R. Hilgenfeld, *J. Mol. Biol.* **2001**, *307*, 357-377.
[462] E.-H. Ryu, Y. Zhao, *Org. Lett.* **2005**, *7*, 1035-1037.
[463] D. E. Frantz, R. Fässler, E. M. Carreira, *J. Am. Chem. Soc.* **2000**, *122*, 1806-1807.
[464] J. A. Dale, D. L. Dull, H. S. Mosher, *J. Org. Chem.* **1969**, *34*, 2543-2549.
[465] P. Aschwanden, C. R. J. Stephenson, E. M. Carreira, *Org. Lett.* **2006**, *8*, 2437-2440.
[466] C. E. Meyet, C. J. Pierce, C. H. Larsen, *Org. Lett.* **2012**, *14*, 964-967.