UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA
INTERNATIONAL DOCTORATE IN STRUCTURAL BIOLOGY

CICLO XXXV

COORDINATOR Prof. Lucia Banci

Saccharidic epitopes of biological relevance: proteins binding and molecular recognition

Settore Scientifico Disciplinare CHIM/03

PhD student
Dott. Francesco Milanesi

Tutor
Dott. Oscar Francesconi

Coordinator
Prof. Lucia Banci

November 2019 - October 2022

Table of Content

Chapter 1

Roles of carbohydrates in recognition processes

1.1 Carbohydrates and their recognition

1.1.1 Carbohydrates structure and organization

Among all the biomolecules, carbohydrates are the most abundant class. Carbohydrates have been largely investigated between 1884 and 1902 by Emil Fisher that discovered that carbohydrates exist in different structures with different stereochemistry. To opportunely describe the structure of a carbohydrate several representations are accepted by the scientific community, but the chair conformation is the most suitable for cyclic carbohydrates. Sugars can be structured as 6-membered rings (pyranosidic form) or in 5-membered rings (furanosidic form); the interconversion occur *via* the open aldehydic form (**Figure 1**). For D-sugars, the hemi acetal or hemi ketal in position 1, called anomeric position, can exist in the α isomer in which the hydroxyl is in the axial position and in the β isomer, with the hydroxyl in equatorial configuration. (**Figure 1**).[1]

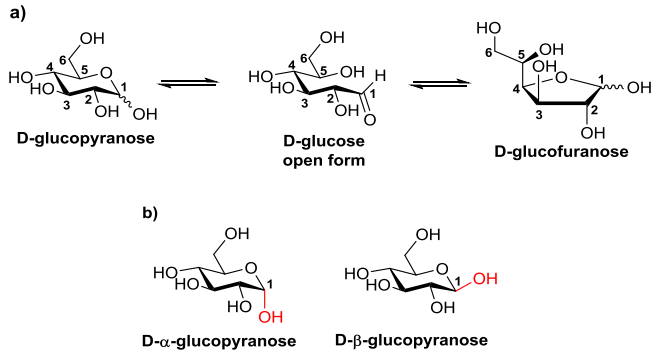


Figure 1. a) pyranosidic, open and furanosidc form of D-glucose; b) α and β isomers of D-glucopyranose.

The difference between axial and equatorial can be applied to all the hydroxyls of the sugar and the different configuration of these groups results in different carbohydrates. When two sugars differ only from the absolute configuration of a stereocenter are called epimers; D-galactose is the epimer in position 4 of D-glucose, since the hydroxyls in that position have an opposite configuration (**Figure 2**).

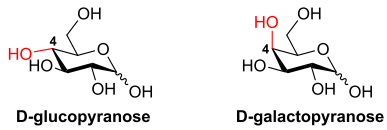


Figure 2. Structures of D-glucopyranose and D-galactopyranose.

From the linkage of monosaccharides is possible to obtain oligo- and polysaccharides. In these complex sugars only the terminal emiacetal saccharide can switch from the α to the β isomer. Several biologically relevant polysaccharides are obtained from the linkage of repeated units of saccharides.[2] Cellulose consists of the repetition of cellobiose units; cellobiose is a disaccharide composed by two glucopyranose connected by a β glycosidic linkage between position 1 of the first sugar and 4' of the second sugar. Starch is composed by units of maltose, a disaccharide in which two units of glucopyranose are linked through an α glycosidic linkage between position 1 and 4'. Oligo- and polysaccharides can be composed by different sugar units. Lactose, a disaccharide present in milk, is constituted by a galactose unit linked through a β glycosidic bond to the hydroxyl in position 4 of glucose (**Figure 3**).

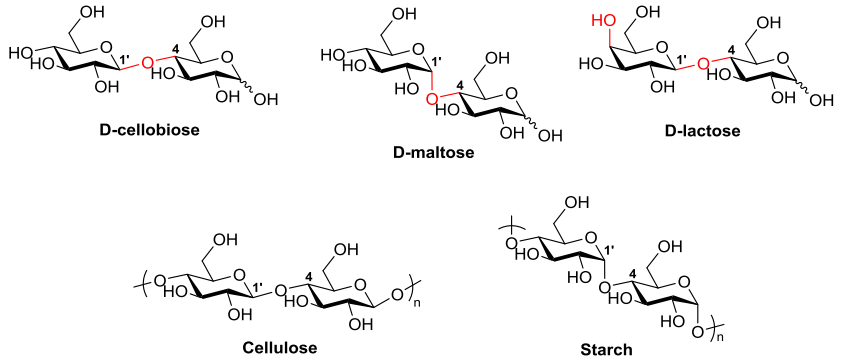


Figure 3. Structures of D-cellobiose, D-maltose, D-lactose and structures of the polymers cellulose and starch.

Furthermore, mono-, oligo- and polysaccharides can be linked to other biomolecules such as lipids or proteins through a process called glycosilation. Glycosylation is the enzymatic process that produces glycosidic bonds between saccharides and other biomolecules in producing glycoconjugates.[3] Glycoconjugates are defined by the type of linkage between the aglycone, the non-sugar portion of the molecule, and the sugar. Usually, conjugation occurs via nitrogen or oxygen linkage, giving *N*-glycans and O-glycans. Among all the glycans expressed on cell surface, glycosphingolipids are the most abundant class; these are ceramide-linked glycans made of different oligosaccharides often functionalized with a terminal sialic acid.[4] Concerning glycoproteins when *N*-glycosylation occurs, the saccharide chain is linked to an asparagine residue, while in O-glycosylation the sugar is linked to serine or threonine in the protein backbone. O-glycans are frequently found in secreted or membrane-associated glycoproteins and the initial saccharide is represented by D-*N*-acetyl-galactosamine (GalNAc).[5] *N*-glycans share a common pentasaccharide core region called $Man_3GlcNAc_2$ (**Figure 4**). This core can be further elongated, sialylated and fucosilated. Some glycoproteins can be found linked to phosphatidylinositol (GPI)- anchored proteins. Linear glycans can also be found; glycosaminoglycans are an example of linear glycans and of note is hyaluronic acid, a linear polymer of a disaccharide composed by D-glucuronic acid and D-*N*-acetyl-glucosamine.[6]

Changes in glycosylation pattern were associated with oncogenic transformation seven decades ago.[7,8] The tumour cells have a wide range of glycosylation alterations that are ascribable at two types, as postulated by Hakomori and Kannagi.[9] The first type of alteration consist of an incomplete synthesis of glycans, as has been seen for the truncated glycans Thomsen antigen (**Tn**), Thomsen-Friedenreich antigen (**TF**) and Sialyl-Tn antigen (**STn**) (**Figure 5**) expressed on the surface of cancer cells in breast and gastrointestinal cancer; this transformation is generally associated to cancers in early stage.[10,11] The second type of modification is a neo-synthesis, commonly observed in advanced stages, such as the expression of the antigens sialyl-Lewis a (**sLe^a, Figure 5**) and sialyl-Lewis x (**sLe^x, Figure 5**) on many types of cancer cells.[12]

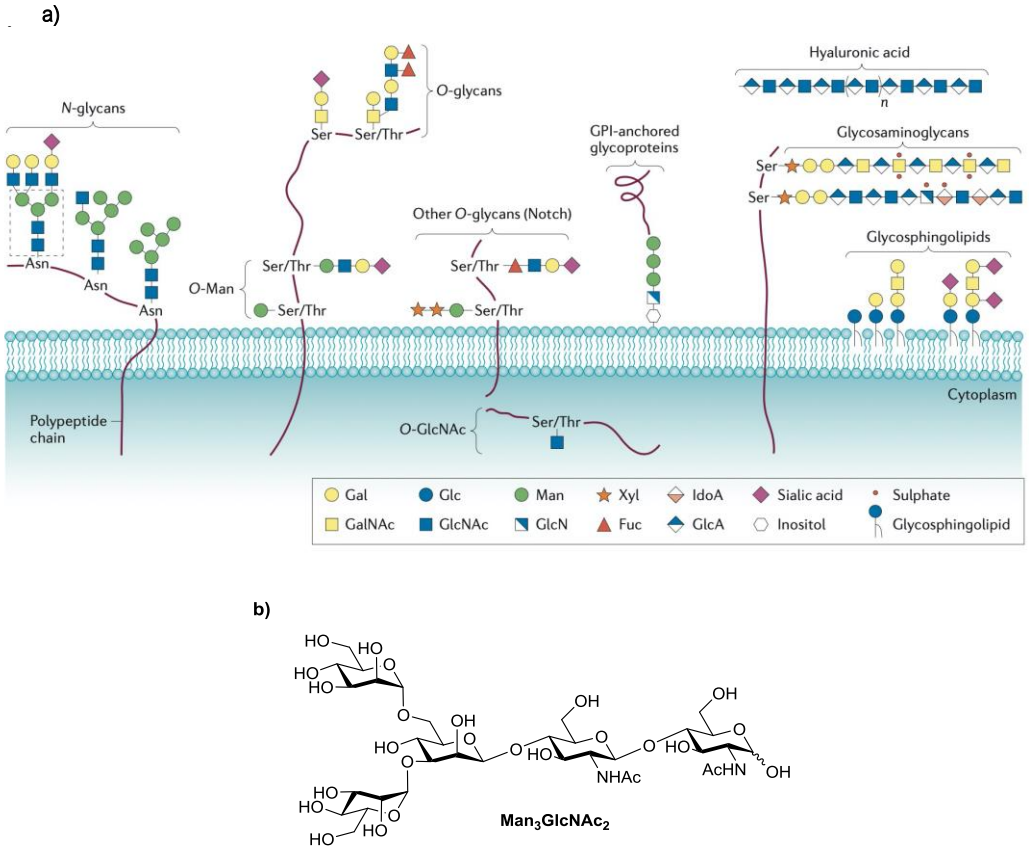


Figure 4. a) Schematic representation of common glycans; b) structure of the $Man_3GlcNAc_2$ core.

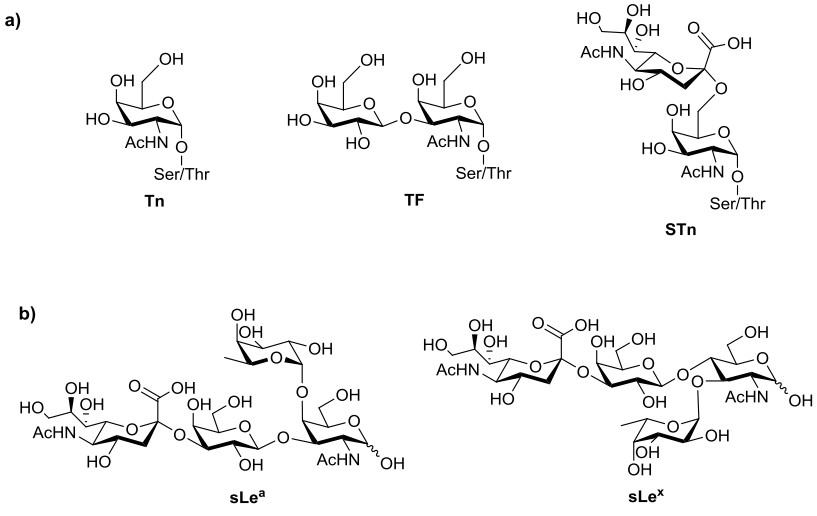


Figure 5. a) antigens Tn, TF and STn; b) sialyl-Lewis a (**sLe^a**) and sialyl-Lewis x (**sLe^x**).

1.1.2 Carbohydrates recognition

Glycosylation is strongly implicated in physiological and pathological processes. Enzymes and receptors can recognize saccharidic epitopes and trigger a biological response. A well-known example is that of blood groups. Indeed, on red blood cells it's possible to find different saccharidic antigens that identifies the blood group, as reported in **Figure 6**. [13]

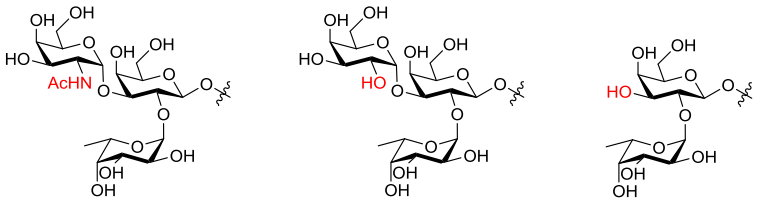


Figure 6. antigens of blood groups.

The family of sialic acids (**Figure 7**) is strongly implicated in recognition processes since its one of the most common terminal residues of glycans. Neu5Ac is normally found linked through an α-2,3 or α-2,6 glycosidic linkage between sialic acid and galactose or in dimeric structures bounded through an α-2,8 linkage.[14]

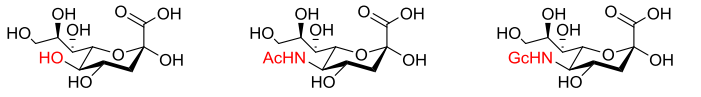


Figure 7. Family of sialic acids.

Among the proteins that recognizes sialic acid, a prominent role is that of Siglecs that are immunoglobulin-like lectins expressed on the surface of immune cells. Sialic acid is present on mammalian cells and it works as an immune check; the recognition by Siglecs ends in a down regulation of the immune system that helps to prevent reactions towards ''self'' antigens.[15]

As shown, carbohydrate recognition is crucial in many biological processes. Among all the classes of proteins that are capable of recognizing carbohydrates, lectins are one of the most abundant in animals, plants and bacteria. Recognition of sugars in solution is a difficult task even for nature since sugars are highly polar molecules. A solvated sugar surround itself of a solvation shell of water molecules through hydrogen bonds. Before the recognition occurs all these interactions must be broken to expose the sugar at the proteins binding site.[16] The mechanisms of ligand recognition have evolved in different families of lectins, but some key features are shared. The affinity between a carbohydrate and a lectin is generally assessed in the millimolar–micromolar range; these, if compared to standard protein-ligand interactions are weak affinities and this phenomenon could be ascribed to the surface interactions that sugars have with lectins, which are not endowed with a deep pocket for ligand recognition.[17]

Sugars recognition is complex, but interactions that a carbohydrate could have with a protein are found among the following:

a. Hydrogen bonding
b. Nonpolar interactions
c. Electrostatic interactions
d. Coordination to metal ions

Sugars are highly polar molecules, endowed with several hydroxyls, amino groups, amides and carboxylic groups. For this reason, direct hydrogen bonding is one of the most important recognition strategies acted by proteins.[17] Galectins are a class of lectins that recognizes β galactosides. The carbohydrates recognition domain (CRD) is conserved among the isoforms and the binding towards galactosides has been largely explored. Human galectin-1 (Gal-1) to recognize lactose has several H-bonds with the Gal such as those from Arg48, His44 and Asn61. Besides polar interactions, sugars have the possibility of hydrophobic effects between the aliphatic skeleton of the sugar, represented by the CH and non-polar aminoacids. In the case of *h*Gal-1, protons in positions 1,3 and 5 are in proximity of Trp68, an aromatic aminoacid that provides both hydrophobic contacts and CH-π interactions (**Figure 8**).[18]

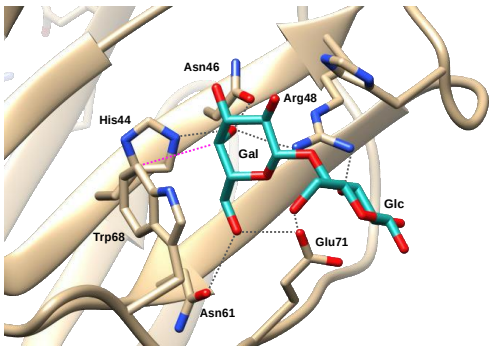


Figure 8. Recognition of lactose by *h*Gal-1 (PDB entry: 1GZW). Hydrogen bonds are reported as grey lines, CH- π as purple lines.

The most extensive structural description of charged sugars recognition are for sialic acid – binding proteins. In the recognition of sialic acid is crucial to anchor the sugar in the binding site through a salt bridge with the carboxylic acid. In **Figure 9** is shown the binding mode of 3'-sialyllactose (Neu5Ac-α-2,3-Gal-β-1,4-Glc) towards Siglec-1.[19] As most of sialic-acid binding proteins the recognition is guided by sialic acid that is often the most involved residue in the interaction. In the case of Siglec-1 the carboxylic function interacts through a salt bridge with Arg87; this interaction is crucial for sialic acid recognition. Other key interactions are the CH- π between the CH_3 of the NHAc in position 5 and Trp2, and the one between the protons in position 9 and Trp108. Beside Neu5Ac, only the Gal unit has an H-bond that involves Tyr44.

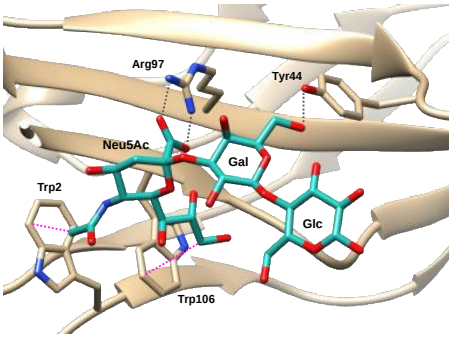


Figure 9. Recognition of 3'-sialyllactose by Siglec-1 (PDB entry: 1QFO). Hydrogen bonds and salt bridges are reported as grey lines, metal coordination as dashed lines.

Several proteins require ions for their functions and lectins are no exception. The E-selectin is a lectin expressed on endothelial cells and acts as a checkpoint for extravasation. If a cell is endowed with its ligands, the tetrasaccharides sialyl-Lewis a (**sLe^a, Figure 5**) and sialyl-Lewis x (**sLe^x, Figure 5**) the extravasation is promoted.[20] E-selectin is one of the most involved lectins in the metastasis process and for this reason it has been largely investigated. In **Figure 10** is shown the recognition of **sLe^x**. E-slectin shown a Ca^{2+} ion in the binding pocket, chelated by Glu80, Asn82, Asn83, Asn105 and Asp106. The Ca^{2+} ion has the function of interacting with the *cis* hydroxyls in positions 3 and 4 of fucose (Fuc). The chelation of 1,2-*cis*-diols and 1,2-*cis*-diols is a common strategy of metal containing lectins to recognize sugars.

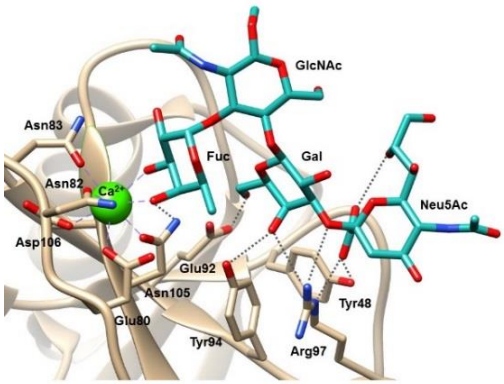


Figure 10. Recognition of **sLe^x** by E-selectin (PDB entry: 1G1T). Hydrogen bonds and salt bridges are reported as grey lines, metal coordination as dashed lines.

Given the high biological relevance of carbohydrates recognition, during years several research groups have investigated chemical tools to interfere in these processes. The first approach was the design of analogues and mimetics of saccharidic epitopes recognized by a specific protein. Analogues and mimetics have an improved stability and improved pharmacological properties in respect to natural products. Furthermore, the affinity for the target protein is generally equal or higher than those of natural products, making possible for these ligands to compete for the binding site.

A complementary approach has been developed in the last three decades. This approach aims at mimicking the protein with a small synthetic molecule that can recognize a carbohydrate in solution. These compounds are synthetic receptors, a class of non-proteic carbohydrates binding agents (CBAs)

1.2 Sialylated glycans recognition

1.2.1 Sialic acids

Sialic acids are a family of 9-termed carbohydrates of the L series (**Figure 7**), of which Neu5Ac is the most representative member. The anomeric position of sialic acids is position 2 and the absolute configuration of this stereocenter is determined by the configuration of the carboxylic function in position 1 in respect to the position of the hydroxyl in position 8. If the carboxylic function is in the axial position, the configuration of the anomeric position is α, if instead, the carboxylic center is in equatorial position, the configuration is β (**Figure 11**).

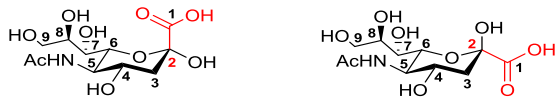


Figure 11. Neu5Ac in α and β configuration and atom labelling.

Neu5Ac is always found in glycans as α anomer since it is the most stabilized form by anomeric effect. Sialic acid is mostly found in glycans as terminal residue linked to Gal through an α-2,6 or an α-2,3 glycosidic linkage or to another sialic acid molecule through an α-2,8 linkage (**Figure 12**).[14]

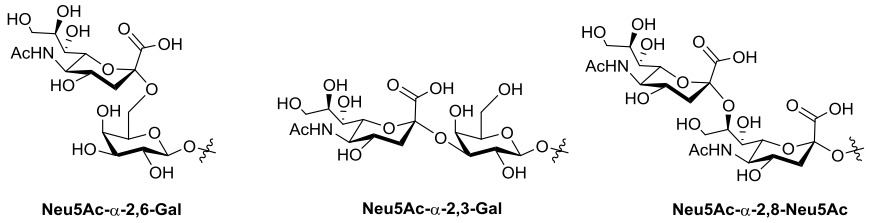


Figure 12. Examples of sialylated epitopes.

Sialylated epitopes are abundant in glycoproteins where they act as ligands for several proteins. Sialic acids are highly present on the surfaces of mammalian cells and for this reason the recognition of sialylated epitopes is crucial for the activation of immune system and its regulation.[15] Among sialylated glycans, of note are **sLe^a** and **sLe^x** (**Figure 10**); saccharidic epitopes involved in tumor metastasis. Sialic acids are also abundant in brain as part of gangliosides. Gangliosides are glycosphingolipids in which an oligosaccharide is linked to a ceramide or a sphingoid lipid tail that is anchored to the plasma membrane. The oligosaccharide moiety is variable, and gangliosides are classified according to the number of sialic acids contained in the sugar portion (M: one; D: two; T: three; Q: four); **GM1** is shown in **Figure 13** as example. Gangliosides have the function of regulating nervous transmission and to work as ''self/non-self'' discriminating epitopes in neurons.[21]

As seen the role of sialic acids in mammals is that of immune checkpoints. Several pathogens adapted to express sialylated glycans to trick the immune system. Moreover, sialylated residues on cell surfaces are used as anchors from viruses, such as Influenza virus, that through the envelope proteins haemagglutinin and neuraminidase, can bind to sialic acid residues and perform cell entry.[22] Thus, sialylated glycans have been largely studied and several chemical tools have been developed to interfere in their recognition processes.

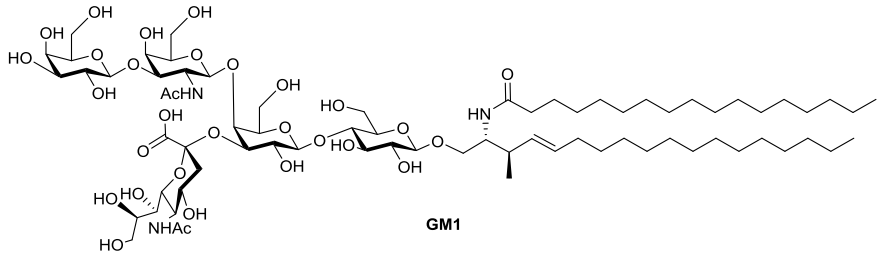


Figure 13. Structure of the ganglioside GM1.

1.2.2 Sialylated glycans and immune response

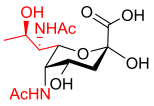
A family of proteins that regulates both innate and adaptive immune processes in mammals is those of Siglecs *(sialic acid binding immunoglobulin-like lectins)*. The Siglecs family is composed of 14 members. Generally, they are expressed on immune cells; the exceptions are Siglec-4, found on oligodendrocytes and Schwann cells, and Siglec-6, expresses on placental trophoblast.[23] Depending on sequence evolution is possible to classify Siglecs in classics, that are related to Siglec-1 and those related to Siglec-3 (CD33). Classic Siglecs are conserved among the species, while those related to Siglec-3 are less conserved.[23] The Ig-like domain is extracellular and through recognition of sialylated glycans this family of proteins can regulate the activation or the inactivation of the immune response. The type of response depends on the intracellular motif associated to the protein. ITIM (*immunoreceptor tyrosine-based inhibitory motifs*) and ITIM-like are associated with a depletion of the immune system. ITAM (*immunoreceptor tyrosine-based activatory motifs*) and ITAM-like have instead an activation of the immune response as final effect of the biochemical cascade. Siglec-4 has a different intracellular motif, a FYN kinase site that regulates cell growth and mobility. Despite Siglec-4, most of Siglecs are endowed with an ITIM or ITIM-like motifs while Siglec-14, -15 and -16 are associated with ITAM and ITAM-like motifs.[15]

Siglecs can interact with *N*-glycans of nearby proteins through *cis* interactions.[24] This binding is crucial to regulate the signaling functions of Siglecs by maintaining inactivated activatory receptors. When Siglecs binds a soluble glycoconjugate or a ligand expressed on another cell the interaction is *trans*. *Trans* interactions are used by immune cells to discriminate endogenous entities, decorated with sialylated glycans, from ''non-self'' entities.

Despite the complex mechanism of Siglecs, some human pathogens, such as group B *Streptococcus* (GBS) can target these proteins to promote the infection. GBS is coated with a capsule of Neu5Ac-α-2,3-Gal-β-1,4-GlcNAc that is recognized by Siglec-9; this recognition suppresses the action of neutrophiles and contributes to pathogen survival.[25]

The opportunistic human pathogen *Canpylobacter jejuni* display gangliosides-like ligands that can be recognize by Siglec-1 and other isoforms, moreover, on the flagella of *C. jejuni* residues of pseudaminic acid (Figure 14) are present. Pseudaminic acid is a sialic acid – related compound that is recognized by Siglec-10 on human dendritic cells (DCs); this interaction causes the release of several inflammatory factors, such as IL-10, that may be related to the inflammatory responses to this and other enteric pathogens.[26]

The ligand specificity for most Siglecs is still poorly known; in Table 1 are reported the known ligand preferences for the known members of the family together with the localization, the type of intracellular domain and the role in specific pathologies.[15]



Pseudaminic acid

Figure 14. Structure of pseudaminic acid; differences with Neu5Ac are highlighted in red.

The binding mode for sialic acid is conserved among the isoforms, as shown in **Figure 9** with the complex between Siglec-1 and 3'-sialyllactose. Normally, the non-sialic portions of ligands are limited in the interaction with the binding site, with the second saccharidic unit being only slightly involved in the recognition process.

Siglec	Expression	Associated motif	Sisalosides preference	Disease implication
Siglec-1 (sialoadhesin)	Macrophages and activated monocytes	ITIM and ITIM-like	α3 β4	Autoimmunity, HIV-1, GBS, *C. jejuni*
Siglec-2 (CD22)	B cells	ITIM and ITIM-like	α6 β4 α6 6S β4	Lymphoma, leukaemia, rheumatoid arthritis
Siglec-3 (CD33)	Macrophages, monocytes, microglia and granulocytes	ITIM and ITIM-like	α6 β4	Leukemia, Alzheimer
Siglec-4 (MAG)	Oligodendrocytes and Schwann cells	FYN kinase	α3 β4	Neurodegeneration
Siglec-5	Neutrophils, monocytes and B cells	ITIM and ITIM-like	α8 α6	GBS infection
Siglec-6	Trophoblast and B cells	ITIM and ITIM-like	α6	Pre-eclampsia
Siglec-7	NK cells, monocytes and mast cells	ITIM and ITIM-like	α8 α3 β4	Cancer and *C. jejuni* infection
Siglec-8	Eosinophils, mast cells and basophils	ITIM and ITIM-like	6S α6 β4	Asthma and eosinophilia
Siglec-9	Neutrophils, monocytes, DCs and NK cells	ITIM and ITIM-like	6S α3 β4 α3	Chronic lung inflammation
Siglec-10	B cells, monocytes and eosinophilis	ITIM and ITIM-like	α6 β4 α6 β4	Lymphoma, leukaemia, eosinophilia, allergy
Siglec-11	Macrophages and microglia	ITIM and ITIM-like	α8 α3	Neuroinflammation
Siglec-14	Neutrophils and monocytes	ITAM and ITAM-like	α8 α3 α6	GBS infectiomn
Siglec-15	Osteoclasts, macrophages and DCs	ITAM and ITAM-like	α6	Osteopetrosis
Siglec-16	Macrophages and microglia	ITAM and ITAM-like	α8 α3	Unknown

Table 1. Human Siglecs, localization in the human body, type of intracellular domain, ligand specificity and pathological role. NK, natural killer cells, DCs, dendritic cells. Glycans are reported schematically: Neu5Ac (purple diamond), Neu5Gc (cyan diamond), Gal (yellow circle) GalNAc (yellow square), GlcNAc (blue square), Fuc (red triangle), sulphation indicated with "S".

1.2.3 Sialylated glycans and viral infection

Siglecs are targeted also by viruses to overcome the immune response. The enveloped virus HIV is coated with sialic acids that can be recognized by Siglec-1 that promotes uptake and infection. It has been also demonstrated that the presence on the envelope of GM3 is crucial for the infection of DCs *in vitro*.[27]

Despite expressing sialic acids on the surface, viruses can use mammal and birds sialylated glycans as anchor for cell entry. Influenza virus is an RNA virus that can engage α-2,3 and α-2,6 linked sialic acids; in human influenza, the most targeted are α-2,6.[28] Hemagglutinin is a trimeric protein anchored to the envelope and composed by the HA1 domain, that recognize sialic acid and the HA2 domain that facilitates membrane fusion. The carbohydrate-binding site is conserved in all influenza subtypes (**Figure 15**).[28]

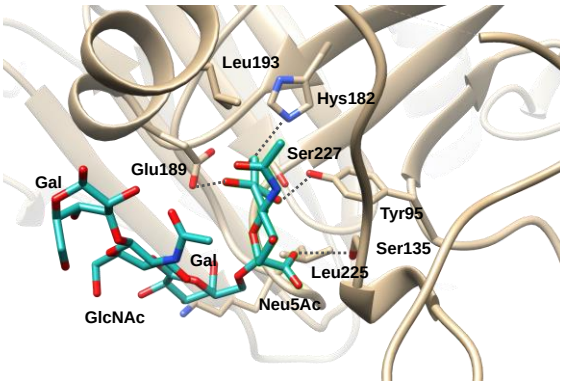
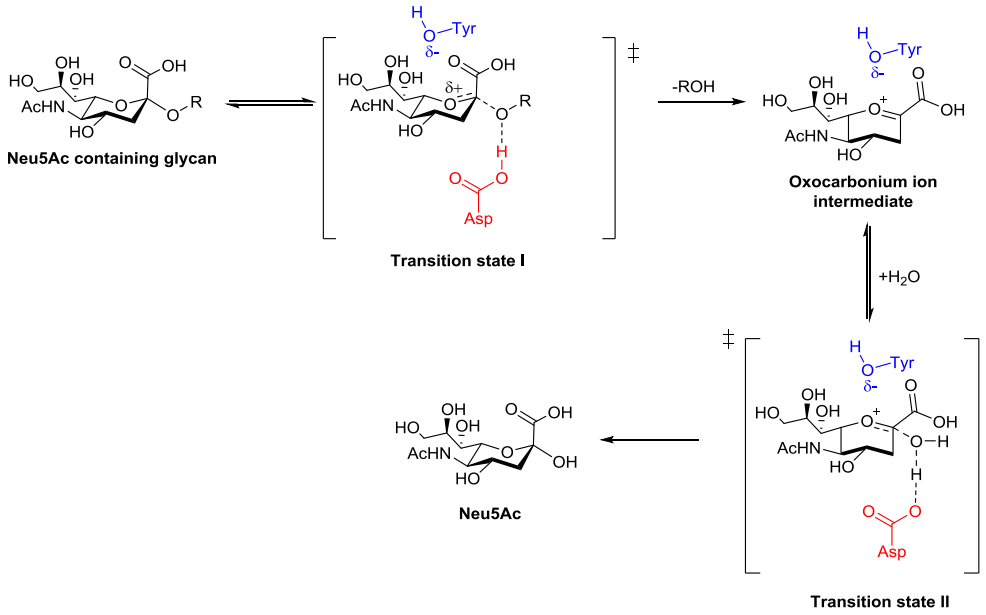


Figure 15. H2 influenza haemagglutinin in complex with α-2,6 linked sialic acid ligand (PDB entry: 2WR7). Hydrogen bonds and salt bridges are reported as grey lines.

After replication of viral material inside the mammalian cells, the final step of the cell infection is the release of the virus particles after budding from the membrane of the host cell. This action is mediated by neuraminidases (NA), envelope proteins that hydrolyzes sialic acid residues from glycans on host membrane. The hydrolysis mechanism involves a Tyr and an Asp residue in the catalytic site; details on the hydrolysis mechanism are explained in **Scheme 1**.

Several viruses can target sialylated glycans, of which influenza viruses are the most relevant example. Among other sialic acid – targeting viruses, of note is Mumps virus. Mumps is a member of *Paramyxoviridae* family, and it is an enveloped RNA virus. Before routine vaccination, 95% of adults had serological markers of exposure to the virus. Despite vaccination, some large Mumps outbreaks continues to appear even among vaccinated subjects. The main symptom of the infections is a painful swelling of parotitis glands, but other tissues are involved with a wild range of inflammatory reactions such as encephalitis, meningitis, orchitis, myocarditis, pancreatitis, and nephritis.[29] Humans are the only host of Mumps virus, and it is transmitted by inhalation or oral contact. The initial infection involves the upper respiratory tract, and the epithelial cells are the first target. On the envelope of mumps virus, the most represented protein is Mumps virus haemagglutinin-neuraminidase (MuV-HN). Like what has been told for influenza viruses, MuV-HN has the role of anchoring the viral particle to the host membrane and facilitate the release of viral material inside the cell.[30] Differently from influenza, the MuV-HN prefers α-2,3 linked sialylated glycans. The binding mode of the protein towards α-2,3-sialyllactose has been elucidate from Kubota and coworkers by X-ray crystallography (**Figure 16**).[31]



Scheme 1. Hydrolytic mechanism of neuraminidases.

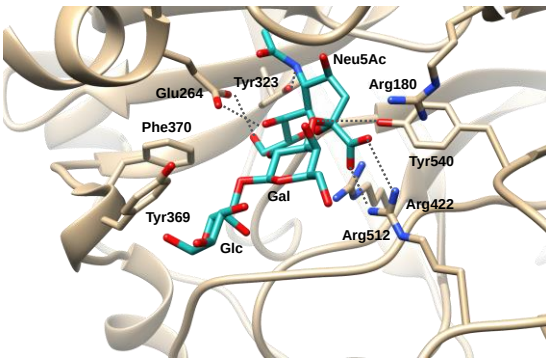


Figure 16. crystal structure of MuV-HN in presence of α-2,3-sialyllactose. Hydrogen bonds and salt bridges are reported as grey lines.[31]

1.2.4 Chemical tools to target sialylated glycans recognition

Siglec-2: Siglec-2 is the most explored of the Siglecs family in the literature, it is expressed on B-cells and is crucial in pathologies with high social impact like lymphoma, leukemia, and rheumatoid arthritis. It is an ITIM endowed Siglec and the elective ligand is represented by Neu5Ac-α-2,6-Gal-β-1,4-GlcNAc recognized with an affinity of K_d = 100–200 μM (**Figure 17**).[32]

The therapeutic targeting of Siglec-2 has been largely investigated with the use of antibodies. Epratuzumab is a murine antibody that targets Siglec-2 on B cells. It has been tested clinically in lymphoma and non-Hodgkin lymphoma cases and has an acceptable safety profile for patients.[33]

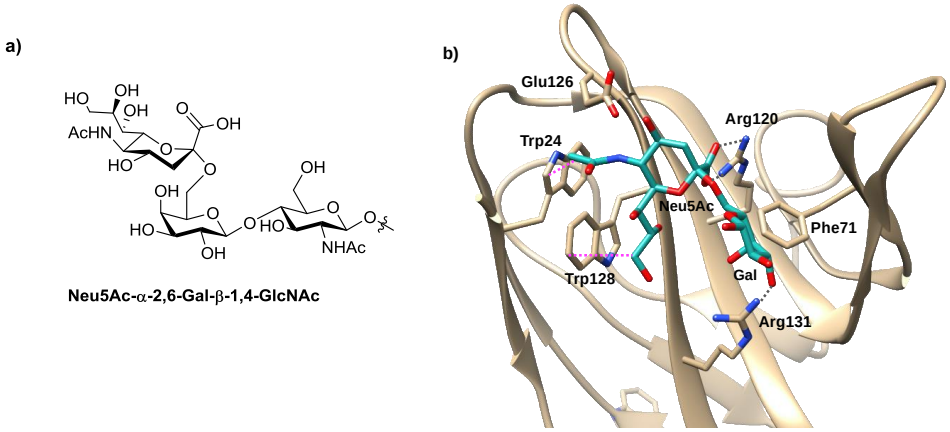


Figure 17. a) Epitope Neu5Ac-α-2,6-Gal-β-1,4-GlcNAc (α-2,6-sialyllactose); b) crystal structure of Ig domains 1-3 of Siglec-2 in complex with Neu5Ac-α-2,6-Gal (PDB entry: 5VKM). Hydrogen bonds and salt bridges are reported as grey lines, CH-π as purple lines.

Despite antibodies are promising therapeutics for Siglec-2 related pathologies, the protein nature of these compounds hampers their development. For this reason, synthetic ligands for Siglec-2 have been largely explored during the last two decades. Since the binding site has only slightly interactions with the second and the third saccharidic unit of the natural ligand, the general strategy to design a new molecule to target this protein is to modify sialic acid. A crucial feature to maintain of the natural saccharide is the carboxylic acid in position 1, key for the salt bridge interaction with Arg97.[32] In 2002 Kelm and coworkers highlighted the role of an aromatic scaffold in position 9.[34] Compound **1** (**Figure 18**) is endowed with a biphenyl-amide moiety in position 9 and shown an IC_{50} = 4.0 μM. The reason of this result is ascribable to the positioning of the biphenyl in a lipophilic pocket close to position 9.[34] A structural description of the binding is not available in literature but a crystal structure of **1** in complex with Siglec-1 has been solved (**Figure 18**).Since Siglec-1 and Siglec-2 are evolutionary related and the binding site is mostly conserved, the binding mode is reasonably similar in both proteins.

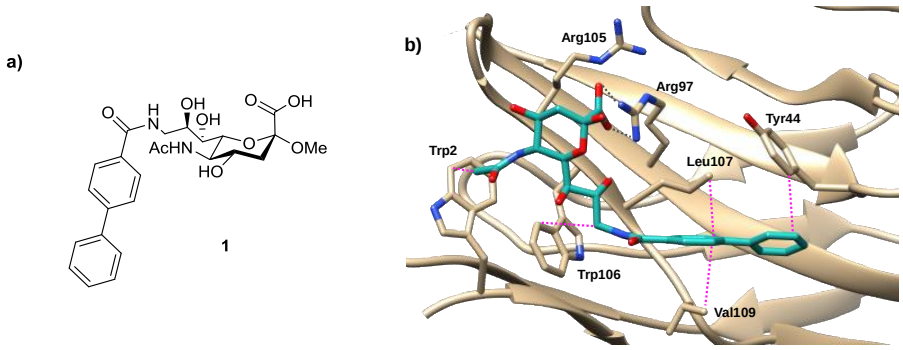


Figure 18. a) Structure of analogue **1**; b) recognition by Siglec-1 (PDB entry: 1ODA). Hydrogen bonds and salt bridges are reported as grey lines, CH- π as purple lines.

From 2009 to 2011 Kiso and coworkers explored the role of the substituent in position 2. The authors reported compound **2** (**Figure 19**) that presents a slightly modified biphenyl moiety and an α-2,6 glycosylation with a Gal that mimic the natural saccharide.[35,36] Despite the addition of the Gal residue, the reported IC_{50} was 3.8 μM. Furthermore, the authors proposed a series of compounds in which they performed a substitution of the Gal residue with a methyl-aryl residue. Compound **3** (**Figure 19**) is the most relevant of the series and presents a benzyl glycoside in position 2, together with an IC_{50} = 0.24 μM. This improvement was justified by the authors with an increased binding enthalpy, due to inserting of the *p*-OH group, that makes more energetically favorable the recognition by Siglec-2.[35,36]

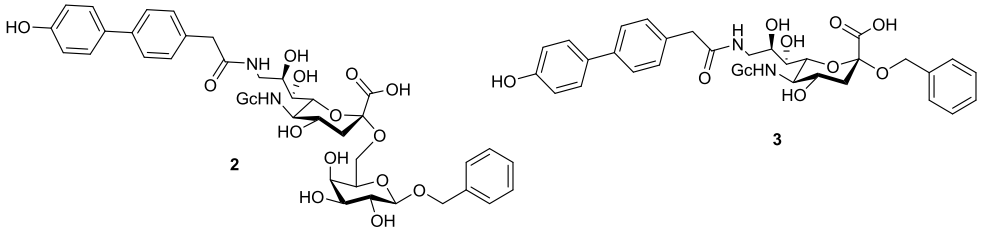


Figure 19. Structures of ligands **2** and **3**.

In 2012, Ernst and coworkers developed a series of Siglec-2 ligands starting from a library of MAG antagonists. The most effective compound resulted in **4** (**Figure 20**), a sialic acid derivative endowed with a dichloro-benzyl residue in position 2, a 4-(4-hydroxy-phenyl)benzamide in position 9 and an aryl-sulfonamide in position 5.[37] The compound reported an affinity of K_d = 60 nM for Siglec-2 but the parallel artificial membrane permeability assay (PAMPA) shown that the molecule is not capable of passing a cell membrane. Thus, despite the high affinity, the compound cannot be administrated orally, and further development is necessary.

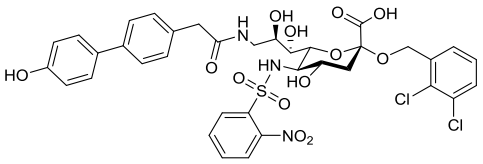


Figure 20: Structure of **4**.

More recently, Brossmer and coworkers explored different substituents in position 4 and in position 3. The authors developed a large family of compounds of which **6** and **7** (**Figure 21**) are the most promising, with an outstanding IC_{50} of 8 nM and 2 nM respectively.[38] Compound **6** presents, beside the biphenyl amide in position 9, a methyl sulfonamide in position 4, a hydroxy-group in position 3 and a spacer bearing a terminal carboxylic group in position 2. Compound **7** differs for the absence of substituents in position 3, a longer spacer in position 2 and a simple sulfonamide in position 4. The authors cannot determine the exact energy gain that grants the binding results of **6** and **7** but extra hydrogen bonds from the substituents in position 4 are the most reliable hypothesis.

Compounds **6** and **7**, together with compound **1** were also tested for their inhibition *in vitro* of Siglec-2 through ELISA assay. The human B-lymphocyte cell line Daudi was used to measure the release of Ca^{2+}; this biological response is a consequence B-cell receptor (BCR) activation. BCR is inhibited by Siglec-2 action, thus, if a compound can inhibit Siglec-2, a consequential increase in the Ca^{2+} concentration should be detected. While compound

1 lead to an increase of the Ca^{2+} concentration at 160 μM, compounds **6** and **7** strongly increased Ca^{2+} signaling at a concentration of 400 nM.

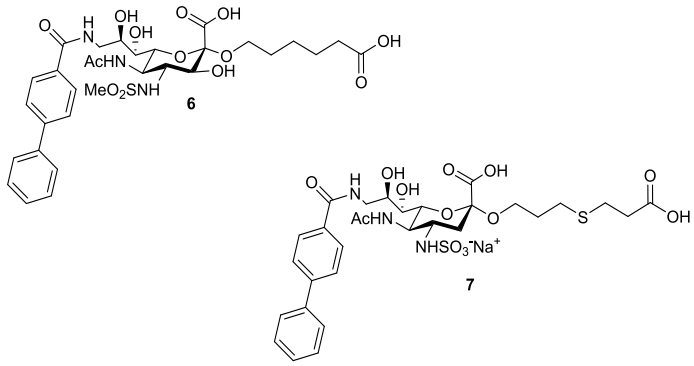
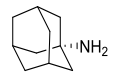


Figure 21. Structures of compounds **6** and **7**.

Neuraminidases inhibitors: The viral neuraminidase has the function of cleaving sialic acid residues from the surface of the cell to enable the release of newly formed virions. Thus, by blocking the neuraminidase action is possible to stop the spreading of the virus. The first compound that shown inhibition of the influenza virus replication was amantadine **8** (**Figure 22**).[39] **8** and related compounds can block the migration of H^+ ions into the virions, a process needed for the uncoating and the release of the viral genetic material. These compounds were used for several years but because a rapid drug-resistance emerged, their use is nowadays limited.



Amantadine
8

Figure 22. Structure of amantadine **8**.

Kim and coworkers in 1997 and Von Itzstein and coworkers in 2001 developed zanamivir **9**[40] and osteltamivir **10**[41] respectively (**Figure 23**). The two compounds present an $IC_{50} < 3$ nM towards the neuraminidases of influenza A and B and are currently used in the therapy against these viruses as active principles of Relenza® (zanamivir, GlaxoSmithKline) and Tamiflu® (oseltamvir, Roche).[42]

These two compounds can block the hydrolytic process by mimicking the transition state of the reaction (**Scheme 1**) and stabilizing the enzyme. The binding mode has been elucidated by X-rays crystallography as shown in **Figure 23**.

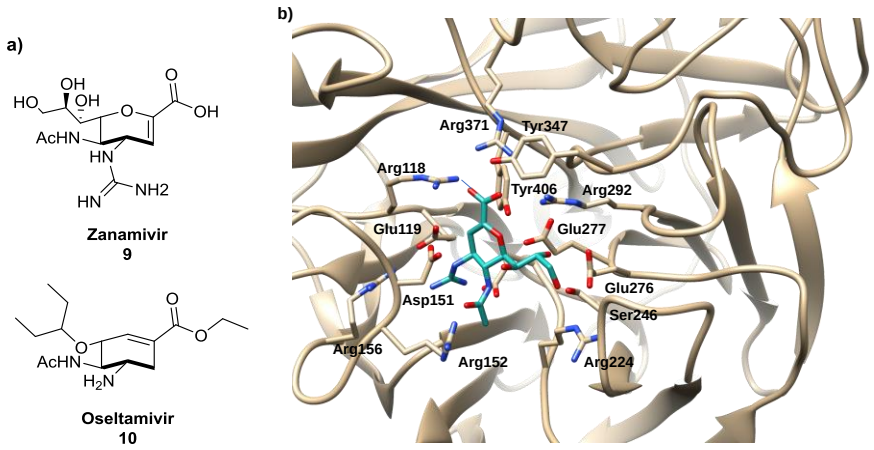


Figure 23. a) Structures of the inhibitors zanamivir **9** and oseltamivir **10**; b) binding mode of **9** to neuraminidase N1 of influenza A (PDB entry: 3CKZ).

Following the approach of mimicking the transition state, also cyclopentane derivatives have been developed, such as peramivir **11** (**Figure 24**).[43]

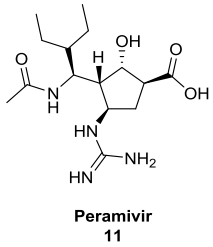


Figure 24. Structure of peramivir **11**.

Peramivir was approved in 2014 by the FDA as therapeutic and preventing agent against influenza A and B. It shown efficacy towards zanamivr and oseltamivir resistant variants of influenza A and B.[42]

Interesting results were reported also with covalent inhibitor, as shown from Vavricka and coworkers in 2013. The authors demonstrated that a 2,3-difluorinated-Neu5Ac (**12**, **Figure 23**) can inhibit the neuraminidase of influenza A by forming a covalent bond with a catalytic Tyr in the binding site.[44]

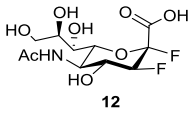


Figure 25. Structure of the covalent inhibitor **12**.

Despite effective compounds were obtained for influenza viruses, the development of new compounds is still of great interest because of the rapid drug resistance that viruses are capable to exploit. Moreover, the development of effective molecules against neuraminidases of different viruses, such as Mumps virus, is hampered by the lack of structural information on the enzymes.

1.3 Carbohydrates binding agents (CBAs)

The activity of recognizing carbohydrates has been ascribed for long time only to proteins. Despite the high value for medicine of tools capable of recognizing specific saccharidic epitopes, the development of lectins or other proteins as therapeutic treatment is hampered by several factors. Indeed, the production and the purification have high costs and the protein nature of these entities give them poor stability, which is not compatible with a use as therapeutics or diagnostic tools. Although protein have poor pharmacological properties, with a general low bioavailability and a high risk of immune response against the therapeutic agent. For this reason, the interest in carbohydrates binding agents (CBAs) of non-protein origin is continuously growing. Carbohydrates binding agents are an heterogenous class of compounds that is emerging in the last two decades as a complementary approach to analogues and mimetics in modulating carbohydrates recognition processes. CBAs have the ability of recognizing saccharidic epitopes in solution and they are structurally ascribable to different chemical classes, from peptides, such as lectins, to those of non-protein origin. The structures of the carbohydrates cited in this chapter are reported in **Figure A1** of the appendix section.

1.3.1 BC_{50}^{0} as a generalized affinity descriptor

Measuring binding affinities is a crucial step in the development of synthetic CBAs. Any time a new molecule is developed, there's the necessity of quantifying the affinity for a guest and to compare the values with literature data. The determination of binding constants is unavoidable to quantitative assets a host-guest affinity. When the 1:1 association is the only detectable equilibrium in solution, the system can be described by a single binding constant; normally a K_A or a K_D is used. This is the case of some biological processes such as the interaction between an enzyme and its substrate or between a receptor with its ligand. Despite these examples, most of the host-guest interactions are more opportunely described by a multiple equilibria system. When two species A and B generates multiple complexes with different stoichiometry (**Figure 25**), all the complexes contribute to the binding capability of one specie toward the other, so it is clear that, a single binding constant for the formation of a single complex is not sufficient to describe the affinity.[45] Moreover, since each host-guest system has its own complexity in term of number of complexes and stoichiometry of the complexes, it is not possible to compare the single formation constants of different systems.

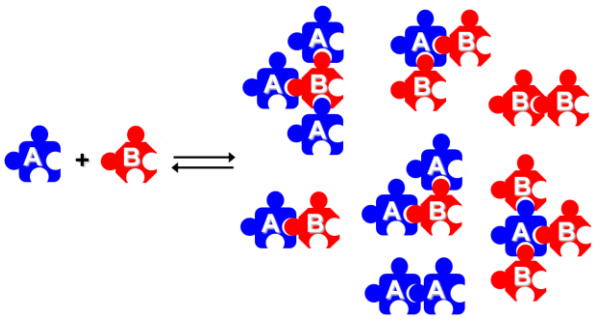


Figure 25. Schematic representation of multiple complexes formation from reacting species A and B.

To cope this issue our research group proposed in 2012 the generalized affinity descriptor BC_{50}, called *median binding concentration.*[45] BC_{50} is defined as the total

concentration of receptor (R) that is necessary to bind the 50% of guest (G). Like for IC_{50}, the higher is the affinity and lower is the BC_{50}. The BC_{50} is a conditional parameter that depends on the cumulative association constants of the different complexes and from the concentration of the ligand that must be specified together with the solvent used for the measurement and the temperature. The BC_{50} parameter has always the dimension of a concentration, making possible the comparison between affinities for different guests of a single host and between different guests' affinities for the same host.

When the total concentration of the guest (G) tents to 0, the value of BC_{50} became constant as shown in **Equation 1**

$$\lim_{G \to 0} BC_{50} = K = BC_{50}^{0}$$

Equation 1. Definition of the BC_{50}^{0} parameter.

BC_{50}^{0} is defined as *intrinsic median binding concentration* and it depends on all the formation constants of the system, but it is independent from the concentration of G. It can be demonstrated that in presence of the 1:1 association as single equilibrium, the value of the K_D is equal to the BC_{50}^{0}. Thus, in presence of a complicated system with multiple association constants, BC_{50}^{0} can be imagined as a global K_D.

1.3.2 Non-proteic CBAs of natural origin

In non-proteic CBAs is possible to distinguish between those of natural and non-natural origin. Non-protein based natural CBAs are a class of compound structurally related to pradimicin A (**Figure 26**).[46]

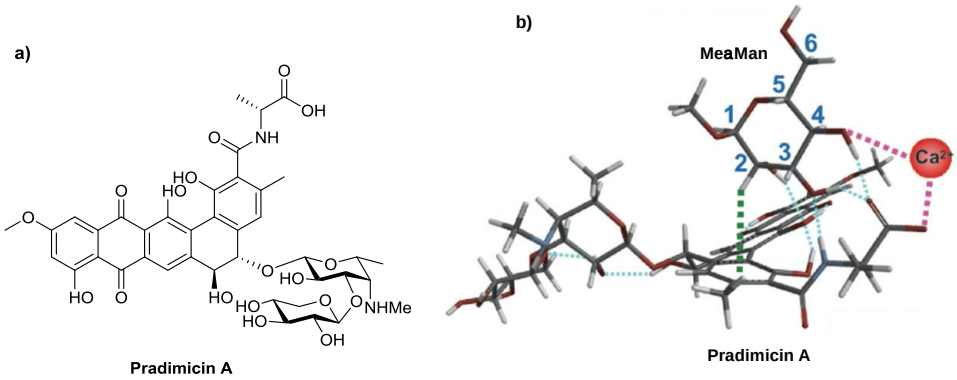


Figure 26. a) structure of pradimicin A; b) Ca^{2+}-dependent binding mode towards MeαMan.

Pradimicin A has been isolated from actinomycetes and has a strong antimicrobial activity *in vitro* and *in vivo* against fungi and yeasts like *Candida albicans*. It has C-type lectin properties; in presence of Ca^{2+} ions, the glycine moiety chelates the metal ion that can interact with the hydroxyl in position 4 of Man. The recognition of mannofuranosides on bacterial membrane causes an alteration of the permeability, with a consequential death of the bacterium.[47] Pradimicin A shown also blocking of HIV entry by recognition of mannofuranosides on gp120.[48]

1.3.3 Synthetic covalent CBAs

Non-proteic CBAs of non-natural origin are structurally heterogeneous and among them is possible to distinguish between covalent and noncovalent. Covalent synthetic CBAs are mostly represented by aryl-boronic acids that have the capability of binding in a reversible manner to 1,2 and 1,3 diols (**Figure 27**). Boronic acids acts as Lewis acids and this makes difficult the use of these compounds in conditions of neutral pH. To cope this issue a dialkylmethylammino residue can be inserted nearby the boronic function (**Figure 27**). An effective example of boron acid-endowed receptor is compound **13**, developed by Hall and coworkers. The receptor is a peptide structure that possess two boronic acids and that can recognize the TF antigen.[49]

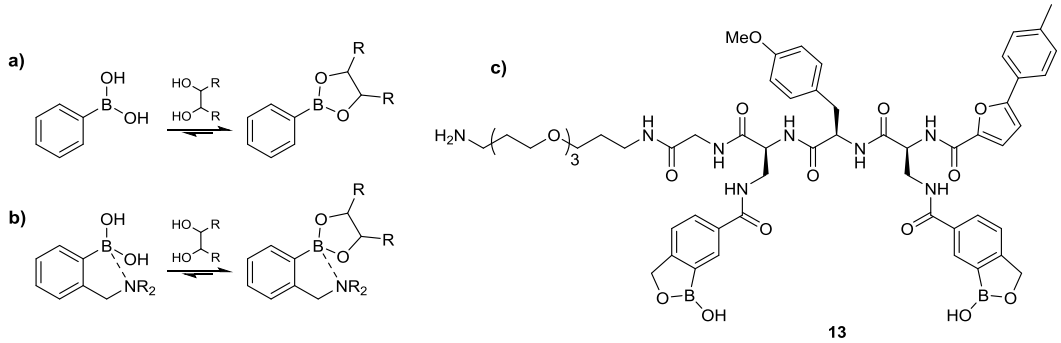


Figure 27. a) covalent bonds formation with a 1,2 diol of aryl-boronic acid; b) covalent bonds formation with a 1,2 diol dialkylmethylammino aryl-boronic acid; c) structure of compound **13**.

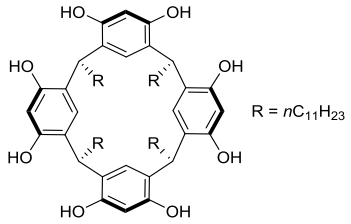
Despite the high efficacy, the use of these receptors *in vivo* is strongly hampered by the covalent recognition mechanism.

1.3.4 Synthetic non-covalent CBAs

Non-covalent CBAs are compounds that through H-bonds, CH-π and coordination of metal ions are capable of recognizing carbohydrates in solution. These molecules use the same interactions used by lectins and are called for this reason biomimetic receptors. The term ''synthetic lectins'' can also be found in literature but since it is somewhat misleading it will not be used.

To overcome the difficulties in mimicking the protein action with a synthetic molecule, the development of synthetic CBAs is explored in a first stage in non-competitive media, such as organic solvents with low polarity. In these media, polar groups are poorly solvated and thus, polar interaction with guests are enhanced, playing a central role in recognition. Hydrophobic effects are in contrary sacrificed in these media and despite the disadvantage in terms of binding, the enhancement of polar interactions greatly simplifies the development of new binding motifs.

Preorganization/Adaptivity: In the design of an effective CBA, all the binding groups must be properly organized in space to afford ligand recognition. The first example of sugar recognition by proper disposition of binding groups was the calixarene **14** (**Figure 28**), of Aoyama and coworkers that extracted D-glucose and D-ribose in CCl_4 from water.[50]



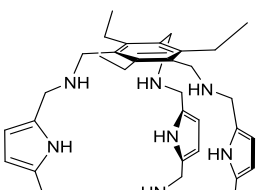
14

Figure 28. Structure of the calixarene **14**.

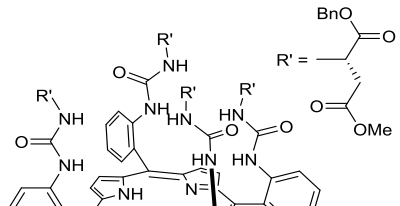
During years two different principles were followed to design synthetic CBAs, the preorganization and the adaptivity principle. A preorganized structure minimize the energy requirements to reach the correct positioning by placing the binding groups in a rigid scaffold. If the requirements to recognize a specific guest are known, a preorganized structure is advantageous, although, without knowing the structural requirements it takes the risk of not being well designed. An adaptable structure can reach the correct orientation of binding groups to recognize a specific ligand and the energy requirement for this process can be compensated by the energy gain from the ligand binding. Because of the flexibility and of the reaching of the correct binding conformation upon ligand recognition, adaptive structures are less selective than preorganized structure.[51]

Macrocycles are the most common architecture for preorganized structures. Receptor **15** (**Figure 29**) is a macrocyclic molecule reported by our research group and it is the first example of selective recognition of a single anomer.[52] The receptor is based on 2,4,6 trisubstituted benzene as aromatic scaffold and methylamino pyrrole moieties as binding groups. The receptor shown an affinity of 20 μM for octyl-β-glucose (OctβGlc) in $CDCl_3$.

Acyclic architectures are also effective for the design of preorganized structures; receptor **16** (**Figure 29**), proposed by Hong and coworkers is a successful example of this strategy.[53] Receptor **16** is endowed with an αααα-5,10,15,20-tetrakis-(o-amminophenyl)porphyrin that works both as a π-donor for CH-π interactions and as a rigid spacer for the correct positioning of the H-bonding groups, represented by ureas. Receptor **16** shown an affinity for OctβGlc of 50 nM in $CDCl_3$ and is effective in recognizing the sugar even in presence of a 9% of methanol, with an affinity of 3 mM. Because of the flexibility of the four aspartic acid arms, receptor **16** has a lack of selectivity among octyl-glycosides, showing affinities for octyl-α-gluycose (OctαGlc) and octyl-α-galctose (OctαGal) of 100 nM and 250 nM respectively.



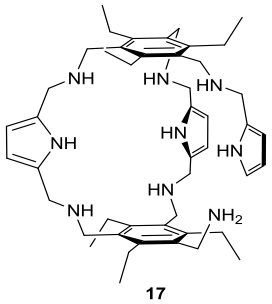
15



16

Figure 29. Structures of receptors **15** and **16**.

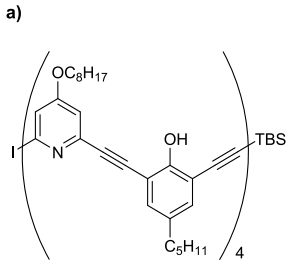
Adaptivity can be reached by using different architectures, the most effective and used strategy is to organize binding groups in a less preorganized structure. Receptor **17** (**Figure 30**), developed by our research group is a more flexible analogue of **15** where only a single bond was cleaved.[54] Receptor **17** presents a 3-fold increase in the affinity for OctβGlc (8 μM in $CDCl_3$) in respect to **15**. From the comparison of the thermodynamic parameters for OctβGlc recognition from **15** (ΔH = -69 KJ mol-1; -TΔS = 43 KJ) and **17** (ΔH = -81 KJ mol^{-1}; -TΔS = 52 KJ mol^{-1}) is clear that the increase in the affinity is of enthalpic origin, highlighting the role of bonding group adjustment upon recognition of the ligand. Moreover, it is possible to notice a mayor entropy loss of **17** in respect to **15** upon ligand recognition, due to the loss of more degrees of freedom.



17

Figure 30. Structure of receptor **17**.

Folded oligomeric structures have also been used for carbohydrates recognition. Abe and Inouye reported in 2015 a highly flexible pyridine-phenol alternated oligomer **18** (**Figure 31**) in which the units are linked through acetylene bonds.[55] The oligomer folds upon binding to a saccharidic ligand and shows nanomolar affinities in dichloroethane. Because of the high flexibility the receptor suffers of a lack of selectivity among α and β glycosides.



18

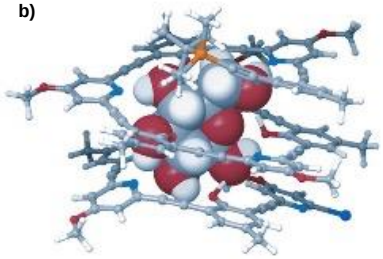
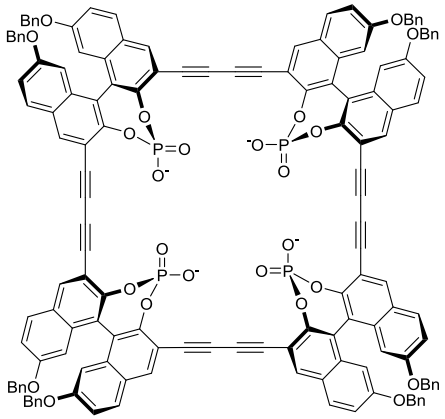


Figure 31. a) Structure of the oligomeric receptor **18**; b) proposed binding mode with β-D-glucose (DFT calculation).

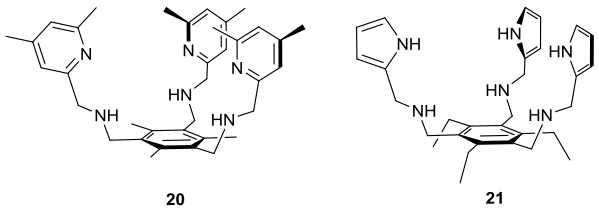
Binding groups: Hydrogen bonds and CH-π interactions are crucial for carbohydrate recognition acted by proteins. Mimicking nature, biomimetic synthetic CBAs are designed in order to present hydrogen bonding donor and acceptor and aromatic scaffolds for CH- π interactions. Exploring different H-bonding groups, those knowing to establish strong interactions like phosphonates or carboxylates were investigate. Receptor **19** (**Figure 32**) developed by Diederich and coworkers bears phosphonates as bidentate hydrogen bonding groups.[56] The receptor exploits 8 hydrogen bonding groups in the receptor cavity and it strongly bind OctβGlc with an affinity of 192 μM in a competitive media like methanol 8% in acetonitrile.



9

Figure 32: Structure of receptor **19**.

Most of carbohydrates-proteins recognition processes are driven by non-charged H-bonds, thus, the use non-charged hydrogen bonding groups in synthetic CBAs has been largely explored. To properly act as hydrogen-bonding donor or acceptor, the binding group has to be correctly positioned in space with an appropriate architecture. Of note, a privileged scaffold for bonding groups organization is hexasubstituted benzene and this scaffold has been used by several research groups to produce tripodal receptors. Receptor **20**[57] of Mazik and coworkers (**Figure 33**) is endowed with pyridine-based hydrogen-bonding groups and it strongly binds OctβGlc (94 μM in $CDCl_3$) with a α/β selectivity ratio of 15. Another effective example of hexasubstituted benzene architecture are the structures reported by our research group, of which receptor **21** is an example (**Figure 33**).[58]



20 **21**

Figure 33: Structures of receptors **20** and **21**.

Receptor **21** strongly binds OctβGlc (24 μM) in $CDCl_3$ with a α/β selectivity of 24 and it shown an interesting affinity for OctβGlc also in a competitive media such as CD_3CN (11.5 mM).

To identify the correct combinations for carbohydrate recognition, several H-bonding groups have been explored on benzene-based tripodal: ureas,[59a] sulphonamides, nitrones, ethers,[59b] amide, amine and imine[60] motifs were combined with indole, imidazole,[59b] pyridine, pyrrole[59c] and acetal.[59d]

Charged sugars are a group of carbohydrates with high biological relevance. Anionic carbohydrates are the most abundant *in vivo* and are mostly represented by phosphorylated sugars, glucuronic and galacturonic acids and from sialic acids. Cationic sugars are rarer *in vivo*, with glucosamine as starting substrate for the synthesis of *N*-glycans, being one of the few examples. Despite that, aminosugars are promising excipients in the pharmaceutical industry. To recognize a charged sugar the most effective strategy is salt bridge; in 1998 Schrader reported receptor **22** (**Figure 34**) that can recognize β-D-glucosammonium in DMSO with an affinity of 11 μM.[61] The bidentate phosphonate group forms, not only a salt

bridge with the ammonium group but it works also as H-bonding acceptor for the vicinal hydroxyls. Unfortunately, because of the simplicity of the system, receptor **22** cannot discriminate well between α and β glycosides.

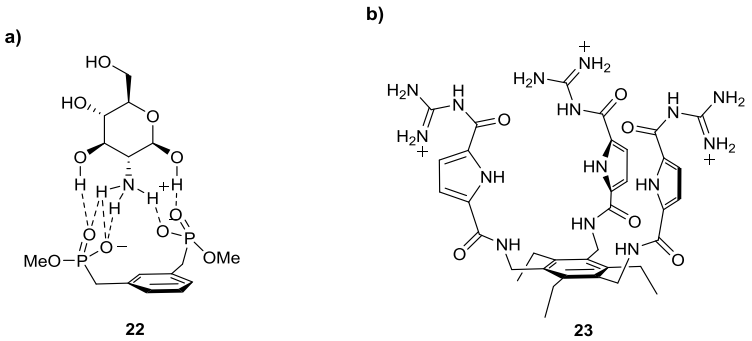


Figure 34: a) Schematic representation of glucosammonium recognition by **22**; b) structure of receptor **23**.

Receptor **23**, reported by Schmuck, is a tripodal receptor endowed with guanidine groups and suitable for recognition of anionic sugars.[62] The receptor shown an affinity for α-D-glucose-1-phosphate of 39 μM in 10% DMSO in water at pH 4. Although, a pH increases to pH 7.4 strongly reduces the affinity because of lower protonation degree of guanidine groups.

Metal ions are often present in proteins structure and these atoms can have an indirect or a direct role in the recognition of a ligand. Metals have been used also in synthetic CBAs; in 2018 Huc and coworkers proposed an aromatic oligoamide metallofoldamer (**24**, **Figure 35**) as receptor for carbohydrates.[63] The oligoamide is flexible and poorly organized while in solution and it forms the receptor cavity upon addition of Cu^{2+}. The receptor is reported to bind D-mannose over D-glucose, D-galactose and D-fructose with an affinity of 2.2 mM in a competitive media such as 20% DMSO in chloroform.

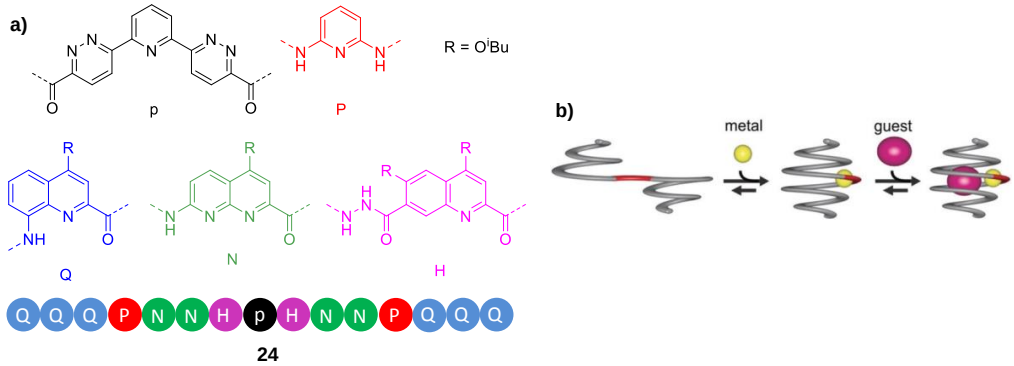


Figure 35: a) Schematic representation of receptor **24;** b) Representation of helical-capsule folding upon metal coordination followed by metal-assisted ligand recognition.

Chirality: Carbohydrates are chiral molecules and following the criteria that the best way to properly bind a chiral molecule is to use a chiral receptor, several molecules have been reported in literature. Our research group reported a family of chiral receptors based on 1,2-*trans*-diaminocycloexane, of which receptors **25** and **26** reported in **Figure 36** are the most relevant examples.[64] These compounds were prepared in the all-*S* and all-*R* enantiomerically pure variants and reported enantioselective recognition of mannosides in polar solvents, in addition with selectivity for mannosides among other monosaccharides.

Receptors (*S*)-25 and (*R*)-25 shown the best affinities for octyl-β-mannose (OctβMan), with an affinity of 83 μM in acetonitrile reported for octyl-β-mannose (*S*)-25, together with the highest R/S discrimination (15:1). Receptor (*R*)-26 was instead, the best of the series in recognizing OctαMan (127 μM in acetonitrile), with a significant α/β selectivity of 1:7.

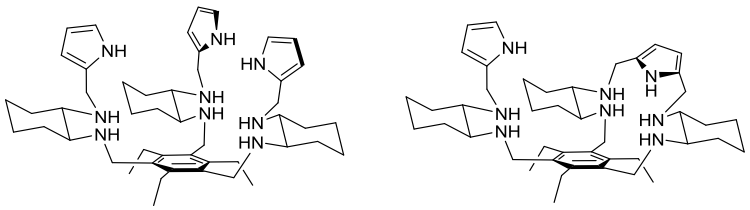


Figure 36: Structures of receptors 25 and 26.

1.3.5 Carbohydrates recognition in water

Recognition of sugars in water is a difficult task also for Nature. A carbohydrate in solution has all the polar groups exposed to the solvent and it is so ascribable to a cluster of water molecules. Moreover, since water is a strongly competitive media for hydrogen bonding, low affinities are generally detected between carbohydrates and lectins, in the micro-millimolar range. For the design of synthetic CBAs working in this solvent, hydrophobic interactions with the sugar are crucial to exploit the recognition and the architecture of the synthetic receptor must be designed properly. Only in the last two decades synthetic receptors for effective recognition of carbohydrates in water have been reported. In 2009, Davis and coworkers reported receptor 27 (Figure 37), a ''temple-structured'' receptor.[65] The ''roof'' and the ''floor'' of the temple are represented by biphenyl moieties, apolar scaffolds for CH-π interactions. The polar ''pillars'' of the temple are isophtalamides and solubility in water is granted by several carboxylic functions. Receptor 27 shown measurable affinities for ''all-equatorial'' carbohydrates such as D-glucose in water; the most relevant data was detected for methyl-β-*N*-acetylglucosamine (MeβGlcNAc) with an affinity of 1.6 mM that is comparable with the affinity for this saccharidic unit of Wheat Germ Agglutinin (1.4 mM). In 2016, the same authors reported an improved ''temple-structured'' receptor (28, Figure 37) endowed with pyrene scaffolds as ''roof'' and ''floor'' of the temple.[66] The enlarged aromatic scaffolds grant to receptor 28 the outstanding affinity for MeβGlcNAc of 55 μM in water.

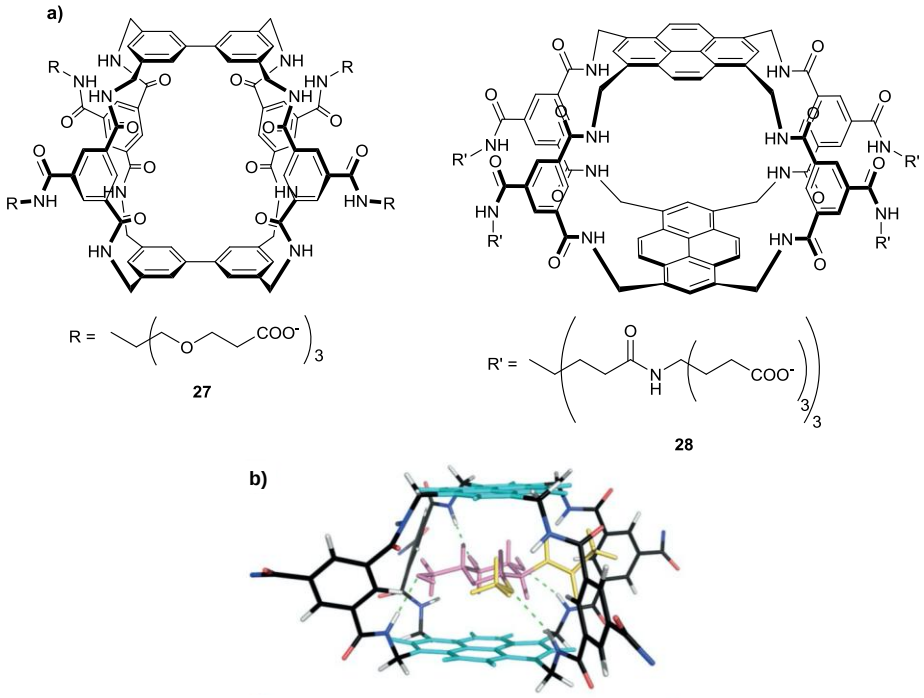


Figure 37: a) Structures of receptors **27** and **28**; b) proposed binding mode of receptor **28** towards MeβGlcNAc.

In 2018, our research group reported the synthetic receptor **29** (**Figure 38**).[67] The receptor is constituted by 3 structural elements: a) a large aromatic scaffold represented by anthracenes; b) diamonocarbazole as a rigid scaffold for H-bonds and c) phosphonates as hydro solubilizing groups. Besides an interesting affinity for MeβGlc of 1.3 mM measured in water at pH 7.4, the receptor shown to accommodate preferentially 1/4-axial and even better 1,4-diaxial hydroxy groups. Indeed, an affinity of 3.1 mM was reported for MeαGlc and of 1.2 mM for MeαGal. Moreover, the receptor recognizes with an affinity of 360 μM L-methyl-α-fucose (MeαFuc), a monosaccharide found usually as α anomer in many glycans, in blood groups antigens and in **sLe^x**. From thermodynamic parameters and NOESY-based modelling calculation, a description of binding modes towards ''all-equatorial'' and ''non-all-equatorial'' saccharides have been performed. While the recognition of MeβGlc is driven by entropy, suggesting strong interactions of the aromatic scaffolds with the hydrophobic faces of glucose, the binding to MeαFuc is enthalpic driven. This evidence highlights the importance of hydrogen bonding in the receptor-fucose complex (**Figure 38**).

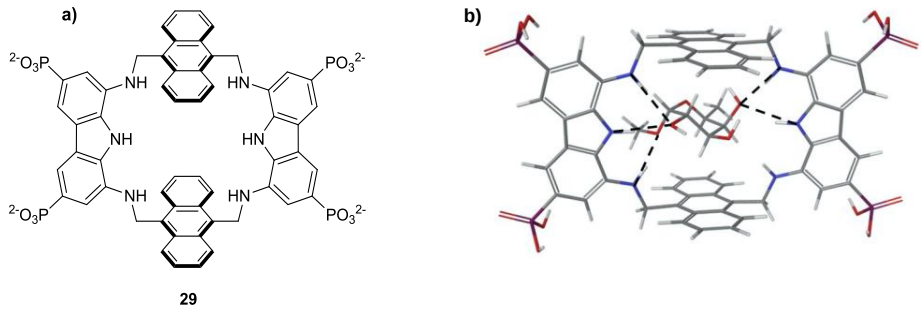


Figure 38: a) Structure of receptor **29**; b) binding mode of receptor **29** towards MeαFuc elucidated through NMR and modelling.

While monosaccharides recognition has been achieved with different structures in last years, di- and oligosaccharides recognition is a more challenging goal. The first example was reported by Davis and coworkers in 2007 with receptor **30** (**Figure 39**).[68] Receptor **30** is a tetracyclic ''temple-structured'' receptor with a binding site suitable for recognition of ''all-equatorial'' disaccharides such as D-cellobiose for which the receptor shown an affinity of 1.7 mM in water. In 2015, the same authors proposed a simplified temple structure (**31**, **Figure 39**). In receptor **31** the tetracyclic architecture has been modified in a byciclic structure based on pyrene as aromatic scaffolds.[69] The receptor shown an affinity of 256 μM for D-cellobiose in water and even an improved affinity for cellotriose (192 μM) and cellotetrose (83 μM). Remarkably, the receptor strongly recognizes D-chitobiose, the β-1,4 dimer of D-glucosamine, with an affinity of 53 μM.

Beside binding of ''all-equatorial'' di- and oligosaccharides, few examples of effective structures for ''non-all-equatorial'' di- and oligosaccharides are reported. In 2017 Yoshizawa and coworkers reported the polyaromatic nanocapsule **32** (**Figure 40**) as receptor for mono- and disaccharides.[70] The receptor is a Pt coordination capsule of 1nM of diameter that is suitable for the recognition of a globular disaccharide such as sucrose, that the receptor recognizes with an affinity of 854 μM. Despite the lack of hydrogen bonding groups, the receptor can encapsulate through CH-π interactions and hydrophobic effects the disaccharide, as shown in **Figure 40**.

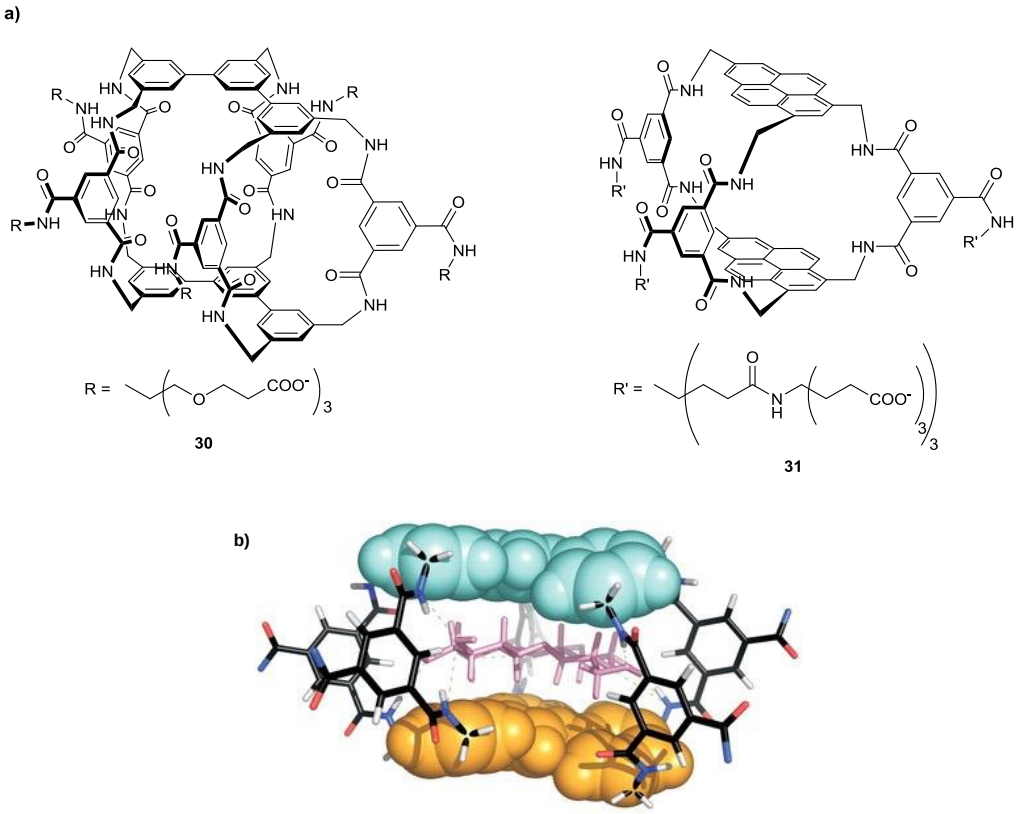


Figure 39: a) Structures of receptors **30** and **31**; b) proposed binding mode of **30** towards D-cellobiose.

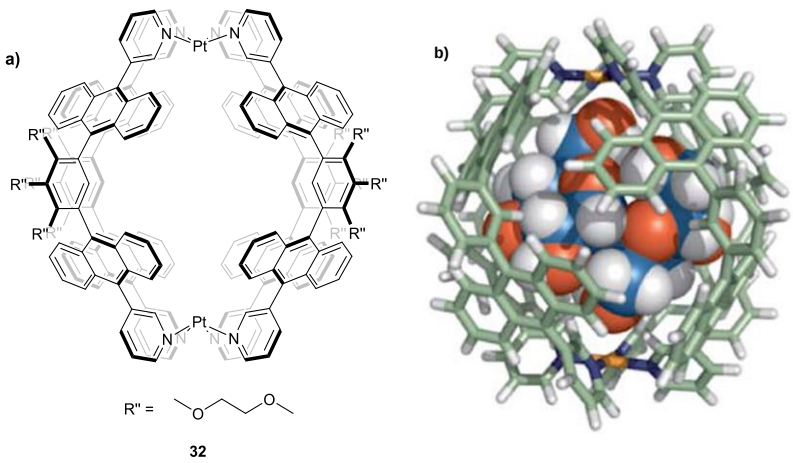


Figure 40: a) Structure of receptor **32**; b) encapsulation of sucrose.

Self-assembly by metal coordination is also an elegant strategy to overcome synthetic steps with low yields, such as macrocyclizations. In 2021, Mooibroek and coworkers reported an example of metal-assisted assembly of a water-soluble cage receptor.[71] Receptor **33** (**Figure 41**) is a Pd_2L_4 cage, where L = dipyridyl ligand. The guanidinium terminal functions of R groups are responsible of the water solubility and the receptor shown to recognize Neu5Ac in water with an affinity of 42 mM among other monosaccharides such as D-glucose, D-mannose and D-galactose.

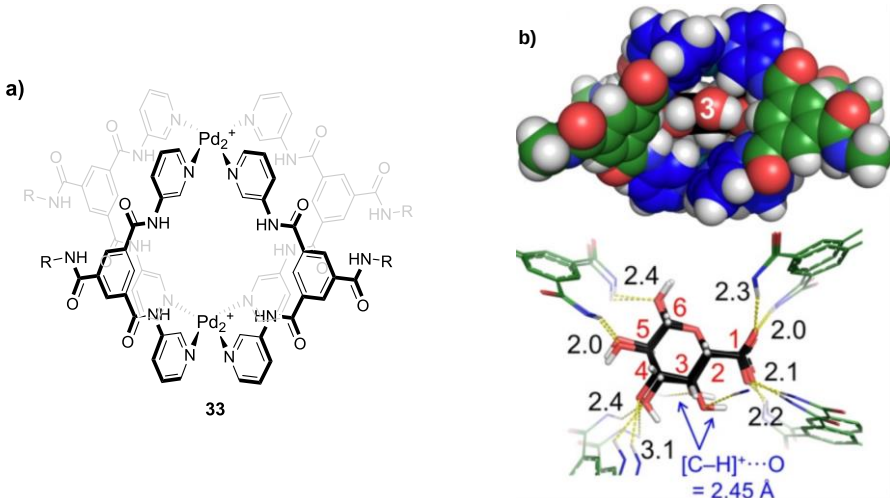
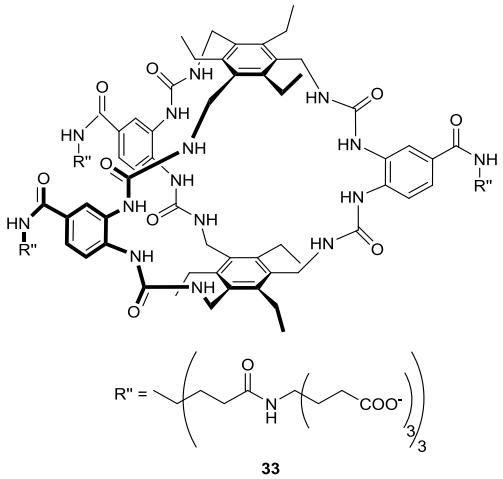


Figure 41. a) Structure of the Pd_2L_4 receptor **33**; b) proposed binding mode of receptor **25** towards Neu5Ac.

1.3.6 Biological activities

D-glucose: D-glucose is one of the most targeted ''all-equatorial'' monosaccharides, for which many receptors, especially from Davis and coworkers, have been developed with the purpose of creating a sensor for D-glucose in blood. Monitoring the concentration of glucose is a daily necessity for diabetics and enzymatic methodologies suffers from poor long-term stability of protein material. To cope this issue, synthetic CBAs are promising tools to substitute enzymes in daily tests. To properly act as a sensor a synthetic CBAs must recognize D-glucose in a physiological condition, such as an aqueous media at pH 7.4, with an affinity of at least 2-10 mM, concentration normally found in blood. In 2018, Davis and coworkers reported the last evolution of temple architectures for glucose sensing with receptor **34** (**Figure 42**), a tricyclic temple structure with 2,4,6 trisubstituted benzene as aromatic scaffold.[72] The receptor shown an unprecedent affinity for D-glucose of 54 μM. The six ureidic hydrogen bonding groups are convergently oriented to bind 1,2-*cis*-dihydroxy groups of the sugar in the binding site.



33

Figure 42. Structure of receptor **34.**

Mannosides: D-mannose is one of the most common saccharides found on the surface of pathogens. Mannans are present on the cell walls of yeasts and bacteria and on the surface of enveloped viruses, such as HIV and coronaviruses. Moreover, high-type oligosaccharides are present on first stages cancer cells upon oncogenic transformation. Since high-mannose type oligosaccharides are not present in healthy cells, these represent a promising therapeutic target.

The first CBA that shown antifungal activity was pradimicin and related compounds that through mannose recognition can alter the permeability of fungal membrane.[47] Our research group has been largely involved in developing synthetic receptors for mannosides, as shown with receptors **21**, **25** and **26** (**Figure 33** and **36**). In a preliminary screening, **21**, **25** and **26**, together with a family of tripodal receptors (**Figure 43**) were tested towards *Candida tropicalis*, *Pichia norvegiensis, Prototheca wickerhamii* and *Prototheca zopfii.* Compound **21** shown in particular a *minimum inhibitory concentration* (MIC) of 2 μg mL^{-1} towards *Pichia norvegiensis* and *Prototheca wickerhamii*, comparable to those of known antibiotics such as amphotericin B that has a MIC of 1 μg mL^{-1} and 2 μg mL^{-1} respectively for the two species.[73,74]

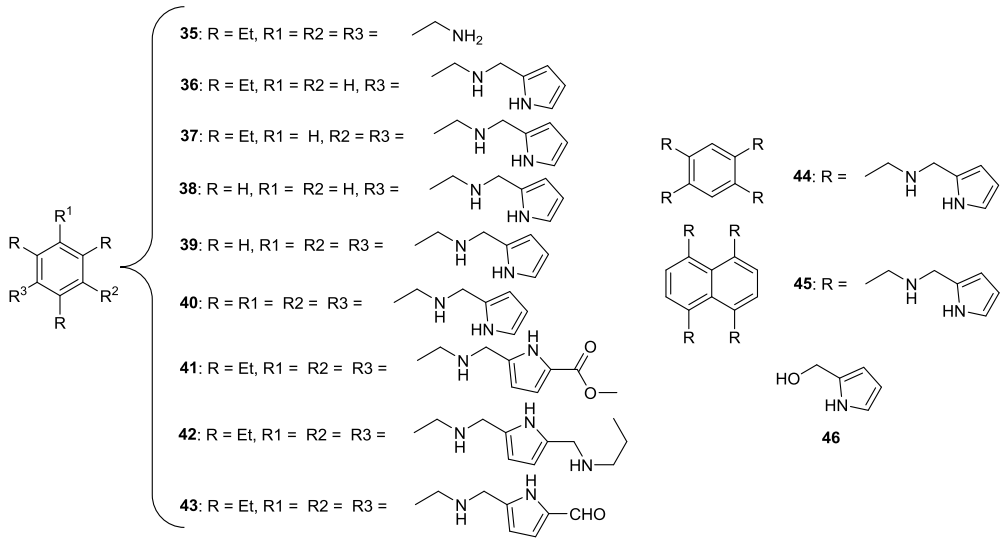
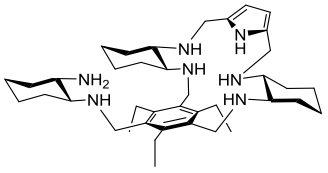


Figure 43. Receptors **35-46**.

Mannans on fungal walls are structurally similar to that of HIV gp120 glycans, thus, our research group has explored tripodal receptors for mannosides also in inhibiting HIV entry. The set of receptors tested included all the compounds screened for antifungal activity and receptor **47** (**Figure 44**).[75]



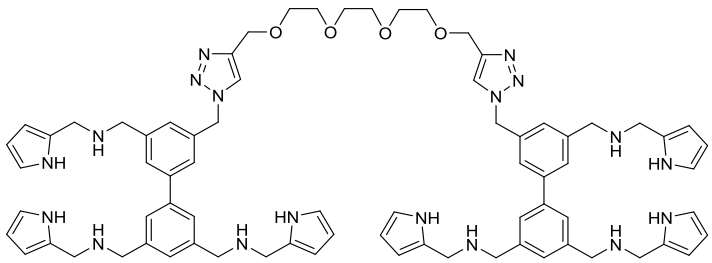
(*R*)-**47**
(*S*)-**47**

Figure 44. Structure of receptor **47**.

Receptor **(*R*)-47** and **(*S*)-47** shown an affinity for OctαMan of 1.16 mM and 0.65 mM, respectively, and for OctβMan of 1.34 mM and 0.42 mM, respectively. An evaluation of binding properties towards gp120 was performed by SPR elucidating the connection between binding to gp120 and affinity for mannosides. Indeed, compounds with submillimolar affinity, such as **25**, **42** and **47** shown strong binding to gp120, suggesting that interaction with target protein is driven by recognition of mannosides. Compounds were investigated in their capability in protecting human T lymphoblastic leukemia cells against HIV-1 entry. Despite some of the strongest binders of gp120 like **25** and **47** were toxic for cells, compounds **40** and **42** that are endowed with the best binding efficacy with the lowest cytotoxicity, shown strong inhibition of HIV-1 entry. Compound **42** shown an EC_{50} = 1.7 μM, comparable to those of pradimicin A (EC_{50} = 3.33 μM).

In 2019 Braunschweig and coworkers published a library of receptors for mannosides with inhibitory activity against Zika virus (ZIKV), a highly mannosilated enveloped virus.[76] Among the tested compounds, receptor **48** emerged as the most promising. The receptor is structured with two biarenyl cores, functionalized with aminopyrroles and liked through a tetraethylen glycol spacer. The authors evaluated the inhibitory activity towards the entry of viral particles of ZIKV in HeLa and Vero cell lines. Compound **48** reported an IC_{50} = 0.24 μM

(HeLa) and IC_{50} = 0.16 μM (Vera) respectively, together with a negligible toxicity even at higher concentrations.



48

Figure 45. Structure of receptor **48**.

High-mannose type glycans expressed on cancer cells have been pharmacologically targeted by CBAs. Mannose-binding lectins, such as Concavalin A (ConA), *Sophora avescens* lectin (SFL) and *Polygonatum cyrtonema* lectin (PCL) have shown cytotoxic activity towards highly cancer cells by inducing apoptosis in a mannose recognition – dependent manner.[77] The family of tripodal receptors for mannosides developed by our research group has been screened also for cytotoxicity with a set of human cancer cell lines: human cervical cancer cells (HeLa), human hepatocellular carcinoma cells (PCL/PRF/5), human lung adenocarcinoma epithelial cells (A549), the cell line with high degree of mannosilation and myeloid leukemia cells (KG-1), presenting a lower level of mannosilation. The tested compounds showed a cytotoxic activity dependent on mannosilation degrees of different cell lines and on the affinity for mannosides. Compounds **(*R*)-25** and **(*S*)-25**, the compounds with the highest affinity for mannosides, shown high cytotoxic effects with an IC_{50} = 1.3 μM and IC_{50} = 2.7 μM for HeLa cells and low cytotoxicity for KG-1 cells. It has also been demonstrated that the binding of ConA to to cells is drastically reduced in presence of **(*R*)-25** or **(*S*)-25** (**Figure 44**). Moreover, induction of apoptosis by mannosides recognition by **(*R*)-25** and **(*S*)-25** was also observed.[78]

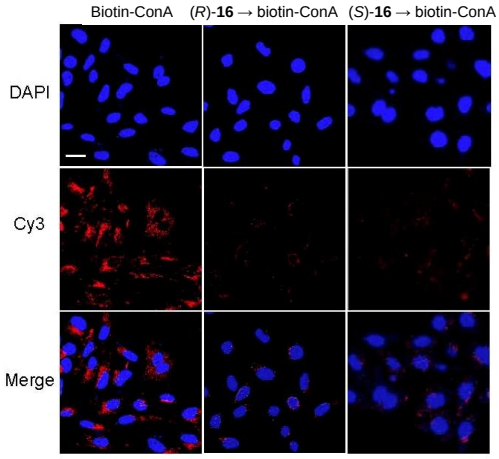


Figure 44. HeLa cells pre-incubated with 10 μM of **25**, then treated with 5 μM biotin-ConA. Treated cells were then stained with Cy3-streptavidin. DAPI colorant was used to stain the nucleus (scale bar = 20 μm).

1.4 Appendix

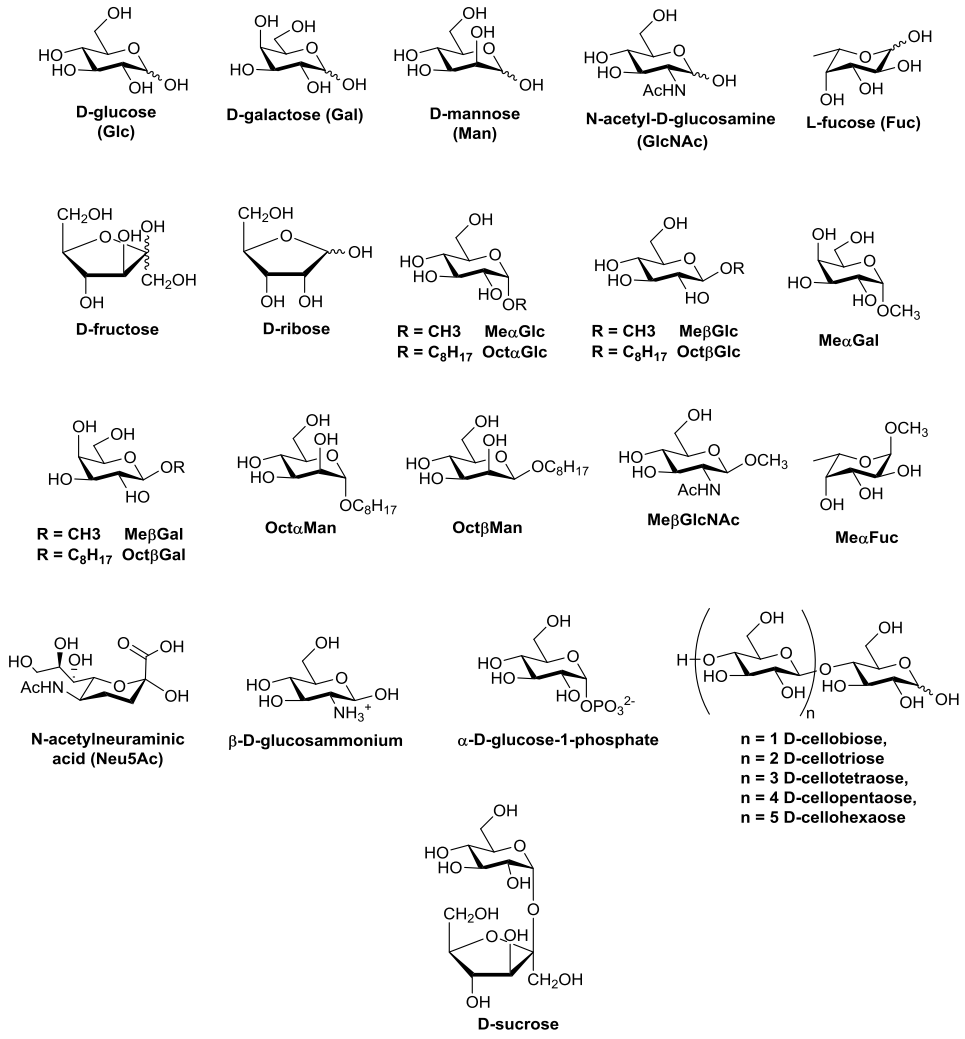


Figura 1A. Carbohydrates cited in text and abbreviations

1.5 Thesis outline

By taking inspiration from the advancements in the field of molecular recognition of carbohydrates, this thesis aims at exploring new molecules capable of modulating carbohydrates recognition processes and the resulting biological effects. Two different approaches will be used, the first one consists of synthetic CBAs (see **Chapter 1**, **1.3**). In **Chapter 2** a new tweezer shaped acyclic receptor for disaccharides will be presented. The receptor shown to be strongly selective in the recognition of disaccharides among monosaccharides and presents the highest affinity ever reported for the disaccharide $GlcNAc_2$.

To further understand the reasons behind the selectivity of the receptor exposed in **Chapter 2** for disaccharides, in **Chapter 3** we will discuss the comparison between this latter and a parent macrocyclic receptor in disaccharides recognition.

In **Chapter 4** we will present a second-generation acyclic receptor and its capability to recognize the core disaccharide $GlcNAc_2$ in a complex glycopeptide. Affinity and structural data will be provided *via* NMR, ITC, CD and modelling.

The second approach that will be considered to interfere in carbohydrates recognition will be those of analogues and mimetics of sialic acid containing glycans.

As explained in **Chapter 1** (**1.2**) sialic acid is strongly involved in immune response and sialylated glycans recognition represent an interesting therapeutic target to treat several pathologies. For this reason, in **Chapter 5** new constrained mimetics will be explored for their binding capabilities with *h*-CD22 (Siglec-2), a protein strongly involved in immune response.

In **Chapter 6** a study on Mumps virus neuraminidase will be exposed; by using a natural and a synthetic trisaccharides, through modeling and NMR techniques an in-depth description of ligand recognition will be reported, representing a promising basis for the design of new inhibitors of this enzyme.

1.6 Bibliohraphy

[1]Varki A., Sharon N., *Essentials of Glycobiology. 2nd edition*, **2009**, Chapter 1: *Historical Background and Overview.*

[2]Bertozzi R.C., Rabuka D., *Essentials of Glycobiology. 2nd edition*, **2009**, Chapter 2: *Structural Basis of Glycan Diversity.*

[3]Pinho S.S., Reis C.A., *Nat. Rev. Cancer*, **2015**, 15, 540-555.

[4]Schnaar R.L., Suzuki A., Stanley P., *Essentials of Glycobiology. 2nd edition*, **2009**, Chapter 10: *Glycosphingolipids.*

[5]Brockhausen I., Schachter H., Stanley P., *Essentials of Glycobiology. 2nd edition*, **2009**, Chapter 9: O-*GalNAc Glycans.*

[6]Stanley P., Schachter H., Taniguchi N., *Essentials of Glycobiology. 2nd edition*, **2009**, Chapter 8: *N-Glycans.*

[7]Ladenson R.P., Schwartz S.O., Ivy A.C., *Am. J. Med. Sci*., **1949**, 217, 194-197.

[8]Hakomori S.I., Murakami W.T., *Proc. Natl Acad. Sci. USA*, **1968**, 59, 254-261.

[9]Hakomori S.I., Kannagi R., J*. Natl Cancer Inst*., **1983**, 71, 231-251 (1983).

[10]Marcos N.T. *et al*., *Front. Biosci*. *(Elite Ed.),* 2011, 3, 1443-1455.

[11]Julien, S. *et al., Glycobiology,* **2006**, 16, 54-64.

[12]Kannagi R.I. Yin J., Miyazaki K., Izawa M., *Biochim. Biophys. Acta*, **2008**, 1780, 525-531.

[13]Stanley P., Wuhrer M., Lauc G., Stowell S.R., Cummings R.D., *Essentials of Glycobiology. 2nd edition*, **2009**, Chapter 13: *Structures Common to Different Glycans.*

[14]Varki A., Schauer R., *Essentials of Glycobiology. 2nd edition*, **2009**, Chapter 14: *Sialic Acids.*

[15]Macauley S.M., Crocker P.R., Paulson J.C, *Nat. Rev.,* **2014**, 14, 653-666.

[16]Ernst B., Hart G.W., Sinay P., *Carbohydrates in chemistry and biology*, **2008.**

[17]Lee C.Y., Lee R.T., *Acc. Chem. Res.,* **1995**, 28(*8*), 321-327.

[18]Bertuzzi S., Quintana J.I., Ardá A., Gimeno A., Jiménez-Barbero J., *Front. Chem.,* **2020**, 8, article 593.

[19]May A.P., Robinson R.C., Vinson, Crocker P.R., Jones E.Y., *Molecular Cell*, **1998**, 1, 719-728.

[20]Ernsy B., Magnani J.L:, *Nat. Rev.,* **2009**, 8, 661-667.

[21]Cutillo G., Saariaho A.-H., Meri S., *Cell. Mol. Immunol.,* **2020**, 17(*4*), 313-322.

[22]Varki A., *Trends in Molecular Medicine*, **2008**, 14(*8*), 351-360.

[23]Lenza M.P., Atxabal U., Oyenarte I., Jiménez-Barbero J., Ereño-Orbea J., *Cells*, **2020**, 9, article 2691.

[24]Hutzler S. *et al.*, *J. Immunol*., **2014**, 192, 5406-5414.

[25]Chang Y.C., Nizet V., *Glycobiology,* **2014**, 24, 818-825.

[26]Klaas M., Oetke C., Lewis L.E., Erwig L.P., Heikema A.P., Easton A., Willison H.J., Crocker P.R., *J. Immunol.,* **2012**, 189(*5*), 2414,2422.

[27]Puryear W.B., Yu X., Ramirez N.P., Reinhard B.M., Gummuluru S., *Proc. Natl. Acad. Sci. U.S.A*., **2012,** 109(*19*):7475-7480.

[28]Stencel-Baerenwald J.E., Reiss K., Reiter D.M, Stehle T., Dermody T.S., **2014**, *Nat. Rev.,* 12, 739-749.

[29]Rubin S., Eckhaus M., Rennick L.J., Bamford C.G., Duprex W.P., *J. Pathol*., **2015**, 235(*2*), 242-252,.

[30]Lamb R.A., *Virology*, **1993**, 197, 1-11.

[31]Forgione R.E., Di Carluccio C., Kubota M., Manabe Y., Fukase K., Molinaro A., Hashiguchi T., marchetti R., Silipo A., *Sci. Rep.,* **2020**, 10, 1589.

[32]Büll C., Heise T., Adema G.J., Noltje T.J., *Trends in Biochemical Sciences*, **2016**, 41(*6*), 519-531.

[33]Leonard J.P., Coleman M., Ketas J.C., Chadburn A., Furman R., Schuster M.W., Feldman E.J., Ashe M., Schuster S.J., Wegener W.A., *et al.*, *Clin. Cancer Res*., **2004**, 10, 5327-5334.

[34]Kelm S., Gerlach J., Brossmer R., Danzer C.-P-, Nitschke L., *J. Exp. Med.*, **2002**, 195(*9*), 1207-1213.

[35]Abdu-Allah H.H.M., Watanabe K., Hayashizaki K., Takaku C., Tamanaka T., Takematsu H., Kozutsumi Y., Tsubata T., Ishida H., Kiso M., *Bioorganic & Medicinal Chemistry Letters*, **2009**, 19(*19*), 5573-5575.

[36]Abdu-Allah H.H.M., Watanabe K., Completo G.C., Sadagopan M., Hayashizaki K., Takaku C., Tamanaka T., Takematsu H., Kozutsumi Y., Paulson J.C., Tsubata T., Ando H., Ishida H., Kiso M., *Bioorganic & Medicinal Chemistry Letters*, **2011**, 19(*6*), 1966-1971.

[37]Mesch S., Lemme K., Wittwer M., Koliwer-Brandl H., Schwardt O., Kelm S., Ernst B., **2012**, *ChemMedChem*, 7, 134-143.

[38]Prescher H., Schweizer A., Kuhfeldt E., Nitschke L., Brossmer R., A*CS Chem. Biol.*, **2014**, 9(*7*), 1444-1450.

[39]Davies W.L., Grunert R.R., Haff R.F., McGahen J.W., Neumayer E.M., Paulshock M., Watts J.C., Wood T.R., Hermann E.C., Hoffmann C.E., *Science*, **1964**, 144(*3620*), 862-863.

[40] Kim C.U., Lew W., Williams M.A., Liu H., Zhang L., Swaminathan S., Bischofberger N., Chen M.S., Mendel D.B., Tai C.Y., Laver W.G., Stevens R.C., *J. Am. Chem. Soc.*, **1997**, 119, 681-690.

[41]von Itzstein M., Wu W.Y., Kok G.B., Pegg M.S.m Dyason J.C., Jin B., Phan T.V., Smythe M.L., White H.F., Oliver S.W., *Nature*, **1993**, 363, 418-423.

[42]Clerq E.D., *Nat. Rev.*, **2006**, 5, 1015-1025.

[43]Sidwell R.W., Smee D.F., Huffman J.H., Bernard D.L., Bailey K.W., Morrey J.D., Babi Y.S., *Antimicrob. Agents Chemtoter.*, **2001**, 45, 749-757.

[44]Vavricka C., Liu Y., Kiyota H. *et al.*, *Nat Commun.,* **2013**, 4, 1491.

[45]Vacca A., Francesconi O., Roelens S., *Chem. Rec.*, **2012**, 12, 544-566.

[46]Oki T., Konishi M., Tomatsu K., Tomita K., Saitoh K., Tsunakawa M., Nishio M., Miyaki T., Kawaguchi H., *J. Antibiot*., **1988,** 41, 1701-1704.

[47]a) Fukagawa Y., Ueki T., Numata K., Oki T., *Actinomycetologia*, **1993**, 7, 1-22; b) Nakagawa Y., Masuda Y., Yamada K., Doi T., Takegoshi K., Igarashi Y., Ito Y., *Angew. Chem. Int. Ed.*, **2011**, 50, 6084-6088; c) Nakagawa Y., Doi T., Masuda Y., Takegoshi K., Igarashi Y., Ito Y., *J. Am. Chem. Soc.*, **2011**, 133, 17485-17493.

[48]Tanabe A., Nakashima H., Yoshida H., Yamamoto N., Tenmyo O., Oki T., *J. Antibiot. (Tokyo)*, **1988**, 41, 1708-1710.

[49]Pal. A., Bérubé M., Hall D.G., *Angew. Chem. Int. Ed.*, **2010**, 49, 1492-1495.

[50]a) Aoyama Y., Tanaka Y., Toi H., Ogoshi H., *J. Am. Chem. Soc.*,**1988**, 110, 634-635; b) Kikuchi Y., Tanaka Y., Sutaro S., Kobayashi K., Toi H., Aoyama Y., *J. Am. Chem. Soc.*, **1992**, 114, 10302-10306.

[51]Francesconi O., Roelens S., *ChemBioChem*, **2019**, 20, 1329.

[52]Francesconi O., Ienco A., Moneti G., Nativi C., Roelens S., *Angew. Chem. Int. Ed.*, **2006**, 45, 6693-6696.

[53]Kim Y.H., Hong J.I., *Angew. Chem., Int. Ed.*, **2002**, *41*, 2947-2950.

[54]Francesconi O., Gentili M., Nativi C., Arda A., Cañada F.J., Jiménez-Barbero J., Roelens S., *Chem. Eur. J.*, **2014**, 20, 6081-6091.

[55]Ohishi Y., Abe H., Inouye M., *Chem. Eur. J.*, **2015**, 21, 16504-16511.

[56]a) Anderson S., Neidlein U., Gramlich V., Diederich F., *Angew. Chem., Int. Ed.*, **1995**, 34, 1596-1600; b) Droz A.S., Neidlein U., Anderson S., Seiler P., Diederich F., *Helv. Chim. Acta*, **2001**, 84, 2243-2289.

[57]Mazik M., Radunz W., Sicking W., *Org. Lett.*, **2002**, 4, 4579-4582.

[58]Nativi C., Cacciarini M., Francesconi O., Vacca A., Moneti G., Ienco A., Roelens S., *J. Am. Chem. Soc.*, **2007**, 129, 4377-4385.

[59]Vacca A., Nativi C., Cacciarini M., Pergoli R., Roelens S., *J. Am. Chem. Soc.*, **2004**, 126, 16456-16465; b) Cacciarini M., Nativi C., Norcini M., Staderini S., Francesconi O., Roelens S., *Org. Biomol. Chem.*, **2011**, 9, 1085-1091; c) Francesconi O., Gentili M., Roelens S., *J. Org. Chem.*, 2012, 77, 7548-7554; d) Nativi C., Cacciarini M., Francesconi O., Moneti G., Roelens S., *Org. Lett.*, **2007**, 9, 4685-4688.

[60]Nativi C., Cacciarini M., Francesconi O., Vacca A., Moneti G., Ienco A., Roelens S:, *J. Am. Chem. Soc.*, **2007**, 129, 4377-4385.

[61]a) Schrader T., *J. Am. Chem. Soc.*, **1998**, 120, 11816-11817; b) Schrader T., *J. Org. Chem.*, **1998**, 63, 264-272.

[62]Schmuck C., Schwegmann M., *Org. Lett.*, **2005**, 7, 3517-3520.

[63]Mateus P., Wicher B., Ferrand Y., Huc I., *Chem. Comm.*, **2018**, 54, 5078-5081.

[624a)Ardá A., Venturi C., Nativi C., Francesconi O., Gabrielli G., Cañada F.J., Jiménez-Barbero J., Roelens S., *Chem. Eur. J.*, **2010**, 16, 414-418; b) Nativi C., Francesconi O., Gabrielli G., Vacca A., Roelens S., *Chem. Eur. J.*, **2011**, 17, 4814-4820; c) Ardá A., Cañada F.J., Nativi C., Francesconi O., Gabrielli G., Ienco A., Jiménez-Barbero J., Roelens S., *Chem. Eur. J.*, **2011**, 17, 4821-4829; d) Vila-Vicosa D., Francesconi O., Machuqueiro M., *Beilstein J. Org. Chem.*, **2014**, 10, 1513-1523.

[65]Ferrand Y., Klein E., Barwell N.P., Crump M.P., Jiménez-Barbero J., Vicent C., Boons G.J., Ingale S., Davis A.P., *Angew. Chem. Int. Ed.*, **2009**, 48, 1775-1779.

[66]Rios P., Carter T.S., Mooibroek T.J., Crump M.P., Lisbjerg M., Pittelkow M., Supekar N.T., Boons G.J., Davis A.P., *Angew. Chem. Int. Ed.*, **2016**, 55, 3387-3392.

[67]Francesconi O., Martinucci M., Badii L., Nativi C., Roelens S., *Chem. Eur. J.*, **2018**, 24, 6828-6836

[68]Ferrand Y., Crump M.P., Davis A.P., *Science*, **2007**, 318, 619-622.

[69]Mooibroek T.J., Casas-Solvas J.M., Harniman R.L., Renney C.M., Carter T.S., Crump M.P., Davis A.P., *Nat. Chem.*, **2016**, 8, 69-74.

[70]Yamashina M., Akita M., Hasegawa T., Hayashi S., Yoshizawa M., *Sci. Adv.*, **2017**, 3, e1701126.

[71]Schaapkens X., van Sluis R.N., Bobylev E.O., Reek J.N.H., Mooibroek T.J., *Chem. Eur. J.*, **2021,** 27, 13719

[72]Tromans R.A., Carter T.S., Chabanne L., Crump M.P., Li H., Matlock J.V., Orchard M.G., Davis A.P., **2018**, 11(*1*), 52-56.

[73]Nativi C., Francesconi O., Gabrielli G., De Simone I., Turchetti B., Mello T., Di Cesare Mannelli L., Ghelardini C., Buzzini P., Roelens S., *Chem. Eur. J.* **2012**, 18, 5064-5072.

[74]Ardá A., Venturi C., Nativi C., Francesconi O., Cañada F.J., Jiménez-Barbero J., Roelens S., *Eur. J. Org. Chem.*, **2010**, 1, 64-71.

[75]Francesconi O., Nativi C., Gabrielli G., De Simone I., Noppen S., Balzarini J., Liekens S., Roelens S., *Chem. Eur. J.*, **2015**, 21, 10089-10093.

[76]Palanichamy K., Joshi A., Mehmetoglu-Gurbuz T., Fernando Bravo, Shlain M.A., Schiro S., Naeem Y., Garg H., Braunschweig A.B., *J. Med. Chem.,* **2019**, 62, 4110-4119

[77]a)Shi Z., An N., Zhao S., Li X., Bao J.K., Yue B.S., *Cell Proliferation*, **2013**, 46, 86-96; b) Shi Z., Chen J., Li C.Y., An N., Wang Z.J., Yang S.L., Huang K.F., Bao J.K., *Acta Pharmacol. Sin.*, **2014**, 35, 248-256; c) Liu Z., Liu B., Zhang Z.T., Zhou T.T., Bian H.J., Min M.W., Liu Y.H., Chen J., Bao J.K., *Phytomedicine*, **2008**, 15, 867-875; d) Ruhul Amin A.R.M., Oo M.L, Senga T., Suzuki N., Feng G.S., Hamaguchi M., *Cancer Res.*, **2003**, 63, 6334–6339; e) Wang S.Y., Yu Q.J., Bao J.K., Liu B., *Biochem. Biophys. Res. Commun.*, **2011**, 406, 497–500.

[78]Park S.H., Choi Y.P., Park J., Share A., Francesconi O., Nativi C., Namkung W., Sessler J.L., Roelens S., Shin I., *Chem. Sci.*, **2015**, 6, 7284-7292.

Chapter 2

A Simple Biomimetic Receptor Selectively Recognizing the $GlcNAc_2$ Disaccharide in Water

Adapted from:

Francesconi O., Milanesi F., Nativi C., Roelens S.,
Angew. Chem. Int. Ed., 2021, *60*, 11168-11172

Abstract

$GlcNAc_2$ is the core disaccharide fragment present in *N*-glycans exposed on the surface of enveloped viruses of high health concern, such as coronaviruses. Because *N*-glycans are directly involved in the docking of viruses to host cells, recognition of $GlcNAc_2$ by a biomimetic receptor may be a convenient alternative to the use of lectins to interfere with viral entry and infection. Herein, we describe a simple biomimetic receptor recognizing the methyl-β-glycoside of $GlcNAc_2$ in water with an unprecedented affinity of 160 μM, exceeding that of more structurally complex receptors reported in the literature. The tweezers-shaped acyclic structure exhibits marked selectivity among structurally related disaccharides, and complete discrimination between mono- and disaccharides. Molecular modelling calculations supported by NOE data provided a three-dimensional description of the binding mode, shedding light on the origin of the affinities and selectivities exhibited by the receptor.

2.1 Introduction

Enveloped viruses are a broad class of highly glycosylated viroids of high health concern, including coronaviruses (SARS-CoV-2, SARS-CoV and MERS), retroviruses (HIV and hepatitis B), orthomyxoviruses (influenza A-C), flaviviruses (dengue, hepatitis C, yellow fever, Zika) and filoviruses (Ebola and Marburg fever), among others.[1] Viral adhesion to host cells is often mediated by specific carbohydrate-protein interactions, which occur through the glycans exposed on the surface of the viral envelope.[2] Biomimetic receptors for carbohydrates targeting these saccharides may inhibit virus-cell interaction, thereby preventing viral entry and infection.[3,4] In this context, among biologically relevant oligosaccharides, *N,N'*-diacetylchitobiose ($GlcNAc_2$) holds a pivotal role, because is part of the highly conserved $GlcNAc_2Man_3$ core fragment of *N*-glycans present on the surface of enveloped viruses, constituting the disaccharidic unit *N*-linked to membrane proteins through an asparagine residue, which often get exposed by mannoside deletions due to virus mutations.[5,6] Unsurprisingly, GlcNAc-binding lectins, such as NICTABA from *Nicotiana tabacum* and *Urtica dioica* Agglutinin (UDA), which target $GlcNAc_2$ at the stem of *N*-glycosilation sites, possess a broad-spectrum activity against several families of enveloped viruses.[7] Thus, effective molecular recognition of $GlcNAc_2$ in water by a simple and easily accessible biomimetic receptor can potentially represent a convenient alternative to natural lectins, because of advantages in terms of availability, molecular weight, purity, stability and immunogenicity.

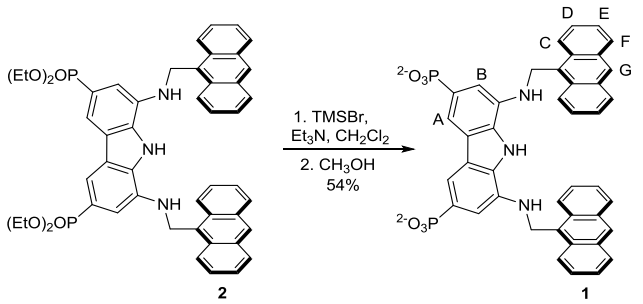
Selective recognition of neutral glycans by biomimeticreceptors in physiological medium represents a major challenge of current research,[4,8] because water is a strong competitor for recognition of polar molecules such as carbohydrates.[9] Nevertheless, in the last few years, significant steps forward have been made in the design of receptors for mono- and oligosaccharides, mainly developing appropriately sized macrocyclic architectures.[8,10] The main drawback of the latter strategy is that macrocyclic architectures must be precisely tailored on specific saccharidic targets and often require lengthy multistep syntheses with relatively low overall yields, due to the critical macrocyclization step.

2.2 Results and discussion

We have recently developed a new generation biomimetic receptor for monosaccharides in water, by assembling into a macrocyclic architecture a tridentate hydrogen binding motif (1,8-diaminocarbazole endowed with two hydrosolubilizing phosphonate groups) with anthracene moieties providing extended CH-π interactions with the saccharidic backbone.[11,12] Interestingly, the corresponding adaptive tweezers-shaped liposoluble receptor proved to effectively recognize biologically relevant xanthines in organic solvents.[13] Following the idea that an acyclic adaptive structure may accommodate disaccharides more effectively than its macrocyclic counterpart, we tested receptor **1**, the hydrosoluble version of the parent receptor **2**,[13] *vs*. a set of mono- and disaccharides in water, in the belief that effective recognition of disaccharides may be achieved by a simple, easily available structure.[14] We report here that this is true indeed, and that **1** is not only a simple and easily accessible receptor for disaccharides, but also the most effective biomimetic receptor for $GlcNAc_2$ in the literature up to date.

Compound **1** was easily obtained by hydrolysis of the previously reported ester **2** (**Scheme 1**).[13] 1 is freely soluble in water under both, mild alkaline (pH = 11) and physiological (pH = 7.4) conditions, whereas precipitates at acidic pH, due to high degree of protonation of phosphonate groups. Receptor **1** shows sharp ^1H NMR signals at low concentration values (5×10^{-4} M), broadening at higher values at pH = 7.4, but not at pH =

11, suggesting concentration-dependent self-association, supported by chemical shift changes.



Scheme 1. Synthesis of receptor **1** with its protons labelling.

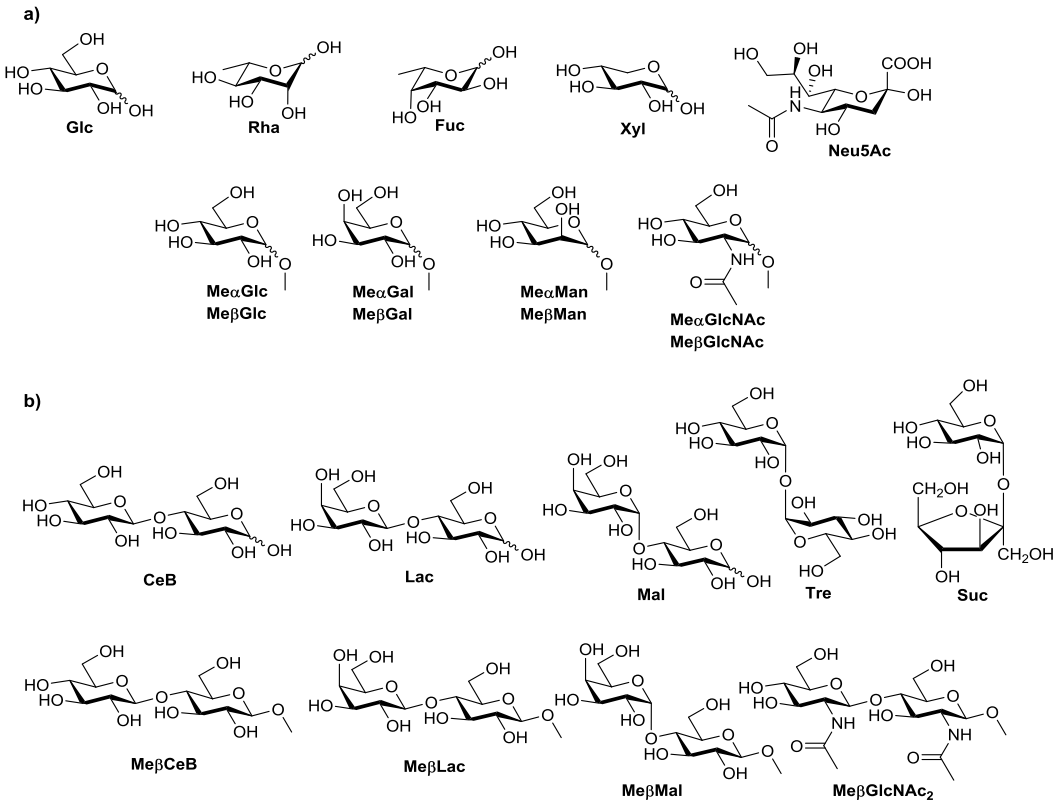


Figure 1. Structure of the investigated monosaccharides (a) and disaccharides (b) and their abbreviations.

The binding properties of receptor **1** were qualitatively screened by ^{1}H NMR spectroscopy toward a set of monosaccharides, including glucose, rhamnose, fucose, xylose, sialic acid, α and β methyl glucosides, galactosides, mannosides, and *N*-acetylglucosamine (**Figure 1a**), together with a set of disaccharides, including sucrose (Suc), trehalose (Tre), cellobiose (CeB), maltose (Mal) and lactose (Lac) (**Figure 1b**) by monitoring the shifts of the proton signals of the sugar upon addition of an equimolar amount of **1**. Surprisingly, little ($\Delta\delta \leq 0.03$ ppm) or no variations were observed for all the investigated monosaccharides and for Suc and Tre, whereas a marked upfield shift was observed for

CeB, Mal and Lac, which were larger for the β than for the α anomer (**Figures S3-S5**), reasonably due to the shielding effect of the anthracene moieties in the binding cavity. A concomitant broadening of signals, larger for the CeB, indicated slow chemical exchange, most likely due to strong binding, suggesting a preference for all-equatorial disaccharides.

A quantitative investigation was then carried out by NMR spectroscopy, extending the study to the all-equatorial $GlcNAc_2$. Because in *N*-glycans the $GlcNAc_2$ glycoside unit is present as the β anomer, methyl-β-glycosides of $GlcNAc_2$ ($MeβGlcNAc_2$), cellobiose (MeβCeB), maltose (MeβMal) and lactose (MeβLac) (**Figure 1a**) were employed, to avoid interconversion equilibria between anomers. Dilution experiments of receptor **1** were preliminary carried out at pH = 7.4, fitting a self-association model featuring two clusters, in which the dimer was the predominant species at low concentrations. The fit gave a dimerization constant of $log\beta_{dim} = 2.65 \pm 0.07$, most likely due to π stacking of the aromatic moieties, which was set invariant in the nonlinear regression analysis of the glycoside binding experiments. The cumulative association constants, reported in **Table 1**, were measured by ^1H NMR titrations in D_2O (pD 7.4) at 298 K, simultaneously fitting the complexation induced shifts of all the available signals to the appropriate association model by non- linear regression analysis. Because multiple binding constants were measured, affinities were assessed through the intrinsic median binding concentration parameter BC_{50}^0,[15] calculated from the measured binding constants and reported in **Table 1**. Amazingly, results show that receptor **1** binds $MeβGlcNAc_2$ with an affinity of 160 μM which is unprecedented in the literature for a synthetic receptor. Indeed, to the best of our knowledge, the highest affinity reported to date for $MeβGlcNAc_2$ by a biomimetic receptor is that observed by Davis and coworkers with a bicyclic polyamidic receptor, showing a 3-fold lower affinity than **1** (BC_{50}^0 = 455 μM, K_a=2200 M^{-1}).[16] The affinity of **1** for $MeβGlcNAc_2$ exceeds that of some lectin-like proteins, such as hevein from *Hevea brasiliensis*, which shows for $GlcNAc_2$ an affinity one order of magnitude smaller (BC_{50}^0 = 1.61 mM, $K_a = 620 \pm 50$).[17] Moreover, receptor **1** exhibits a marked selectivity, showing an affinity for MeβCeB nearly one order of magnitude smaller, and a 200-fold drop of affinity for MeβMal and MeβLac. ^1H NMR titrations with MeβCeB were also duplicated at pD 11 (**Table S1**) and fitted to the association model obtained at pD 7.4. The closely comparable affinities confirmed that the degree of protonation of the phosphonate groups does not affect the binding ability of receptor **1**. Most remarkably, recognition of monosaccharides appears completely depleted, as appreciated from the preliminary screening and from the titration of MeβGlcNAc, in which no significant variation of chemical shifts was observed (**Figure S11**). Somewhat counterintuitively, a flexible acyclic structure exhibits excellent affinities and selectivities, overriding those of more structurally complicated macrocyclic architectures.[8]

The binding affinities obtained by NMR were further confirmed by ITC in H_2O at physiological pH. Data from two to three independent titrations run at different reactant concentrations were combined and simultaneously fitted to remove ambiguities in the definition of the binding model. The dimerization constant obtained by NMR dilution experiments at pD 7.4 was set invariant in the nonlinear regression analysis of ITC data. Cumulative binding constants, together with affinity values, were reported in **Table 1** for a direct comparison with NMR results. The good agreement between the two independent techniques confirmed the observed affinities. Unfortunately, because of the strong self-association of receptor **1**, ITC measurements did not provide reliable thermodynamic parameters.

To shed light on the origin of the affinities and selectivities exhibited by receptor **1**, a three-dimensional description of the binding mode was attempted for the 1:1 complex of **1** with $MeβGlcNAc_2$, using molecular modelling calculations supported by NOE data from NMR spectroscopy. NOESY spectra carried out on an equimolar mixture of **1** and $MeβGlcNAc_2$

showed unambiguous intermolecular NOE contacts between the H-2 and H'-6' protons, belonging to the two monosaccharide units, and the H-C protons of the anthracenes (**Figure S18**). In addition, NOE contacts were found between the methyl protons of the *N*-acetyl group of the methylglycosidic unit with the H-C protons of the anthracene and the H-A and H-B protons of the carbazole. In contrast, no NOE contacts could be found for the second *N*-acetyl group.

		NMR		ITC	
Glycoside	R:G	*logβ*	BC_{50}^0	*logβ*	BC_{50}^0
MeβCeB	1:1	2.53±0.07	0.94±0.10	2.58±0.03	3.52±2.35
	2:1	6.33±0.06		5.28±0.29	
MeβMal	1:1	2.27±0.01	31.0±4.4	2.24±0.01	34.4±4.9
MeβLac	1:1	2.27±0.02	30.8±4.7	2.31±0.02	26.1±4.1
$Me\beta GlcNAc_2$	1:1	3.55±0.04	0.16±0.01	3.49±0.07	0.12±0.03
	2:1	7.35±0.09		7.71±0.22	

Table 1. Cumulative formation constants ($log\beta n$)[a] and intrinsic median binding concentration (BC_{50}^0, mM)[b] for receptor to glycoside (R:G) complexes of **1** with methyl glycosides, measured at 298 K from NMR data in D_2O at pD 7.4 and from ITC data in H_2O at pH 7.4.[c] [a] Formation constants were obtained by nonlinear least-square regression analysis of NMR and ITC data. [b] Calculated from the *logβ* values using the "BC_{50} Calculator" program.[15] [c] Receptor **1** dimerization constant ($log\beta_{dim}$ = 2.65±0.07) was set invariant in the nonlinear regression analysis of NMR and ITC data.

A conformational search carried out on the 1:1 complex returned a single family of minimum energy conformers within 10.0 kJ mol^{-1} from the global minimum. The minimum energy structure depicted in **Figure 2** shows the disaccharide entirely located inside the binding cleft, within the two anthracene faces, with the H-2 and H'-6' protons pointing toward the H-C protons of the anthracenes, in agreement with the proximities inferred by NOE contacts (**Figure 2a**). In addition, the *N*-acetyl group of the methylglycosidic unit faces the diaminocarbazole moiety, pointing the methyl protons toward the H-A and H-B protons, and to one of the H-C protons of the anthracene. From the above model, all O···H interatomic distances shorter than the sum of the van der Waals radii and compliant with hydrogen bonding criteria were calculated and four hydrogen bonding interactions were found between **1** and $Me\beta GlcNAc_2$ (**Figure 2b**): the carbazole NH and one of the aminic NH behave as donors toward OH-3 of the methylglycosidic unit, whereas the other aminic NH of the receptor acts as both, donor to the OH-3 and acceptor from the amidic NH of the *N*-acetyl group. In addition to CH-π interactions with the anthracene units,[18] the *N*-acetyl group contributes to stabilize the complex through CH-π interactions with the carbazole unit. Most likely, the substantial participation to binding of the *N*-acetyl group may account for the observed selectivity over other disaccharides.

In this context, it is worth mentioning that, interestingly, the binding mode of 1 with MeβGlcNAc2 is reminiscent of that between hevein and the corresponding trisaccharide chitotriose, as reported by the group of J. Jiménez-Barbero (**Figure 3**).[17,19] A close similarity can be appreciated between the binding mode of chitotriose to Tyr30 of hevein and of chitobiose to the aminocarbazole of **1**. Indeed, both receptors engage hydrogen-bonding with the saccharidic OH-3 and with the acetamidic NH, while in both cases the methyl group stabilizes the complex through CH-π interactions with the aromatic ring.

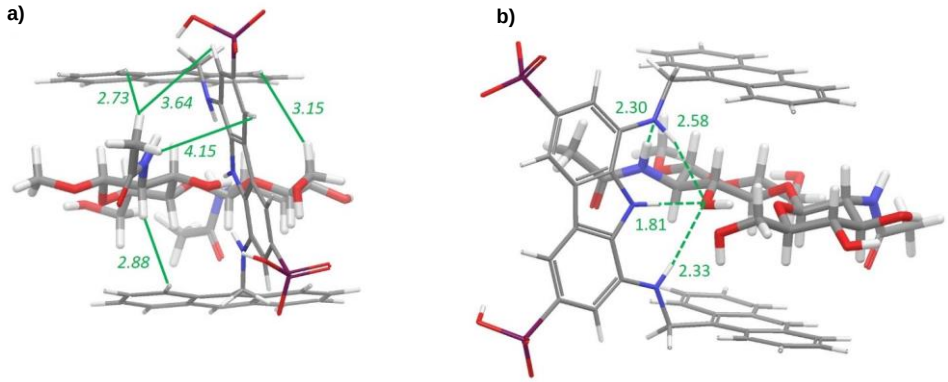


Figure 2. Global minimum structure of the **1**•MeβGlcNAc$_2$ complex in two different projections. a) The strongest intermolecular NOEs found between **1** and MeβGlcNAc$_2$ are indicated as solid lines with corresponding distances [Å] calculated from the lowest energy conformer. b) Intermolecular hydrogen-bonding interactions found in the calculated structure are indicated as dashed lines with corresponding oxygen/hydrogen and nitrogen/hydrogen distances [Å].

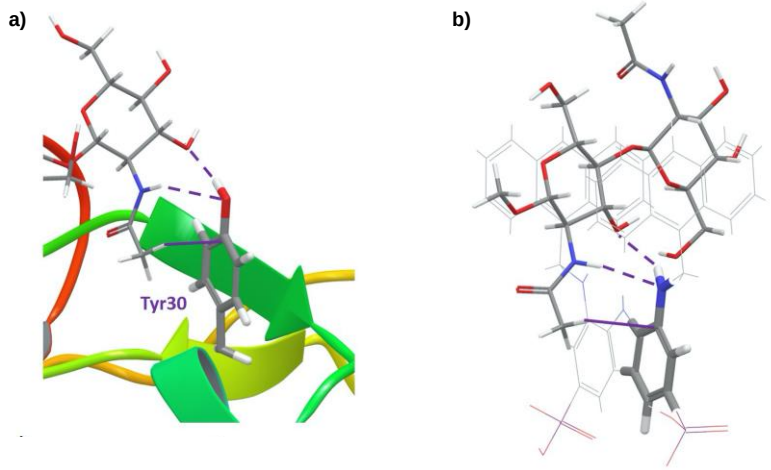


Figure 3. Comparison of the global minimum structures of a) hevein•chitotriose complex and b) **1**•MeβGlcNAc$_2$ complex. Intermolecular hydrogen-bonding and CH-π interactions involving Tyr30 and the arylamine moiety of **1** are indicated as dashed and solid lines, respectively. Protein backbone was drawn as a ribbon for clarity.

2.3 Conclusions

Altogether, the present work shows that molecular recognition of a disaccharide can effectively and selectively be achieved with a well-designed acyclic host featuring an adaptive architecture. The tweezers-shaped receptor **1** fully discriminates disaccharides from monosaccharides, selectively recognizes all-equatorial from non all-equatorial disaccharides, and shows an unprecedented affinity for GlcNAc$_2$, the core glycosidic fragment of viral *N*-glycans. The hydrogen-bonding and CH-π interactions established by receptor **1** with the *N*-acetyl group most likely account for the selectivity observed for MeβGlcNAc$_2$ over other all-equatorial disaccharides. Because of simple structure, easy synthetic availability, and accessible structure modifications, receptor **1** stands as a promising tool for saccharide recognition.

2.4 Supporting information

2.4.1 Synthesis and characterization of chemical materials

General: ESI-MS analyses were performed in negative ion mode and were recorded on an LCQ-Fleet Ion Trap equipped with a standard Ionspray interface. HRMS were performed on a Triple-TOF with a resolution of 35000 (FWHM). Chemical shifts are reported in part per million (δ) relative to 4,4-Dimethyl-4-silapentane-1-sulfonic acid (DSS) for D_2O, using the residual solvent line as secondary internal reference (4.79 ppm for spectra run in D_2O). ^{13}C NMR spectra were obtained at 125 MHz in D_2O. Chemical shifts are reported in δ relative to DSS for D_2O.

Materials: Reagents were purchased from commercial suppliers and used without purification. Compound **2** was prepared according to known methods.[11]

Synthesis of receptor 1: To a solution of **2** (809 mg, 0.950 mmol) in 6.4 mL of dry CH_2Cl_2, Et_3N (784 mg, 7.62 mmol) was added under a nitrogen atmosphere. The solution was cooled to 0 °C and TMSBr (1.17 g, 7.62 mmol) was slowly added. The solution was stirred at room temperature for 22 h, then diluted with 25 mL of CH_2Cl_2 and cooled to 0 °C. After addition of 50 mL of MeOH, the solution was stirred at room temperature for 30 minutes and the solvent evaporated to give 1.84 g of crude **2** as a pale brown solid. The solid was dissolved in 50 mL of NaOH 0.3 M under an ice bath and the aqueous phase washed with AcOEt (4 x 30 mL). EtOH (120 mL) was then added to the aqueous phase and the resulting suspension was centrifugated at 3000 rpm for 20 minutes. The liquids were removed, the solid was dissolved in water (40 mL), then, NaOH 0.5 M was added under an ice bath until pH = 12 was reached. Then, HCl 0.5 M was added (160 mL) and the resulting suspension was centrifugated at 3000 rpm for 20 minutes. The liquids were removed and the solid was suspended in 160 mL of HCl 0.1 M: The suspension was centrifugated at 3000 rpm for 20 minutes, the liquids were removed and the solid was suspended in MeOH (2 x 40 mL) and the resulting suspension centrifugated at 3000 rpm for 20 minutes. The liquids were removed, and the yellow solid was dried in vacuo to obtain pure **2** (380 mg, 54%). Mp: 145.2 (dec) ºC; ^1H NMR (500 MHz, D_2O, DSS as internal reference, δ = 0.0): δ 8.02 (d, *J* = 11.9 Hz, 2H, C*H*-4); 7.86 (d, *J* = 9.0 Hz, 2H, C*H*-14); 7.53 (d, *J* = 11.6 Hz, 2H, C*H*-2); 7.16-7.13 (m, 6H, C*H*-15, C*H*-19); 7.00 (bs, 8H, C*H*-16, C*H*-17); 4.93 (s, 4H, CH_2-11) ; ^{13}C NMR (125 MHz, D_2O, DSS as internal reference, δ = 0.0): δ 135.93 (d, *J* = 143.2 Hz); 135.2 (d, *J* = 10.7 Hz); 133.8; 133.29; 132.35; 132.28, 131.31 (*C*H-17); 129.99 (*C*H-19); 128.95 (*C*H-15); 127.60 (*C*H-16); 126.33 (*C*H-14); 125.93 (d, *J* = 16.3 Hz); 117.81 (*C*H-4); 117.73 (*C*H-2); 45.05 (*C*H_2-11); ^{31}P NMR (202 MHz, D_2O): δ 12.73; ESI-MS *m/z* (%): 367.67 (100%) $[M-2H]^{2-}$, 736.25 (86%) $[M-1H]^-$; HRMS (*m/z*): $[M-2H]^{2-}$ calcd. for $C_{42}H_{33}N_3O_6P_2$, 367.58496; found, 367.58727.

2.4.2 Binding studies

Preliminary screenings (298 K, 500 MHz) were performed in D_2O at pD 11 in presence of DSS as internal reference. Solution of reducing carbohydrates were prepared in D_2O and kept overnight at room temperature before the screening experiments, to ensure equilibration of the anomers. The spectra of the free sugars at 1 mM concentration were compared to the spectra of the equimolar mixture of sugars with receptor **1** (1 mM each) and chemical shift differences were evaluated.

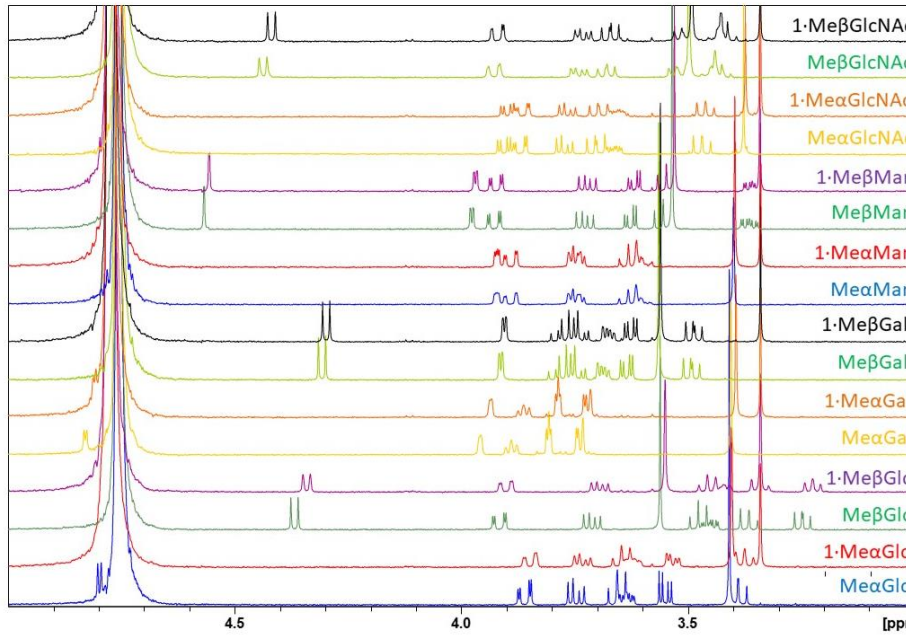


Figure S1. ^{1}H NMR spectra (500 MHz, D_2O) of a 1 mM solution of methyl glycosides and an equimolar mixture of methyl glycosides and **1** (1 mM each) at pD 11.

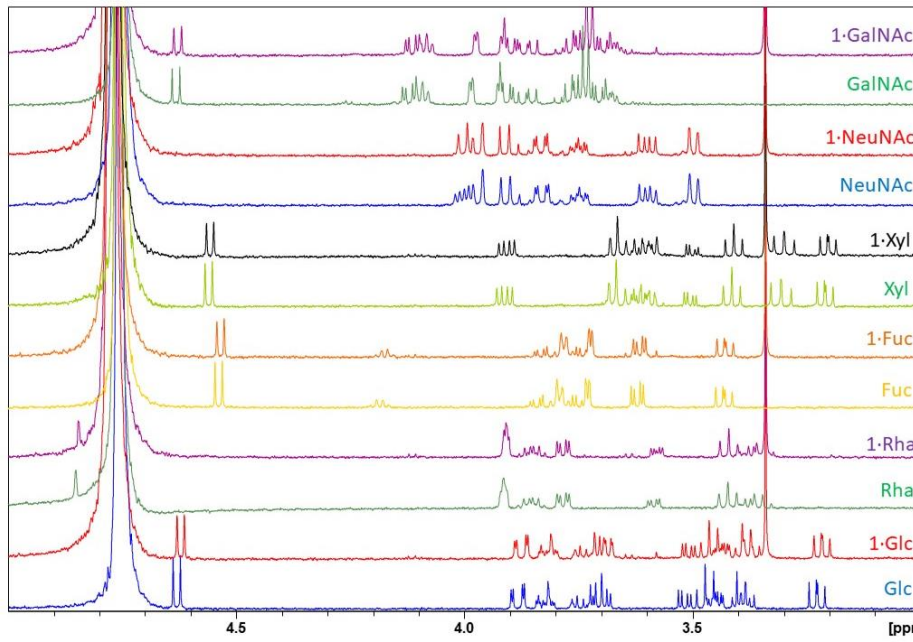


Figure S2. ^{1}H NMR spectra (500 MHz, D_2O) of a 1 mM solution of monosaccharides and an equimolar mixture of monosaccharides and **1** (1 mM each) at pD 11.

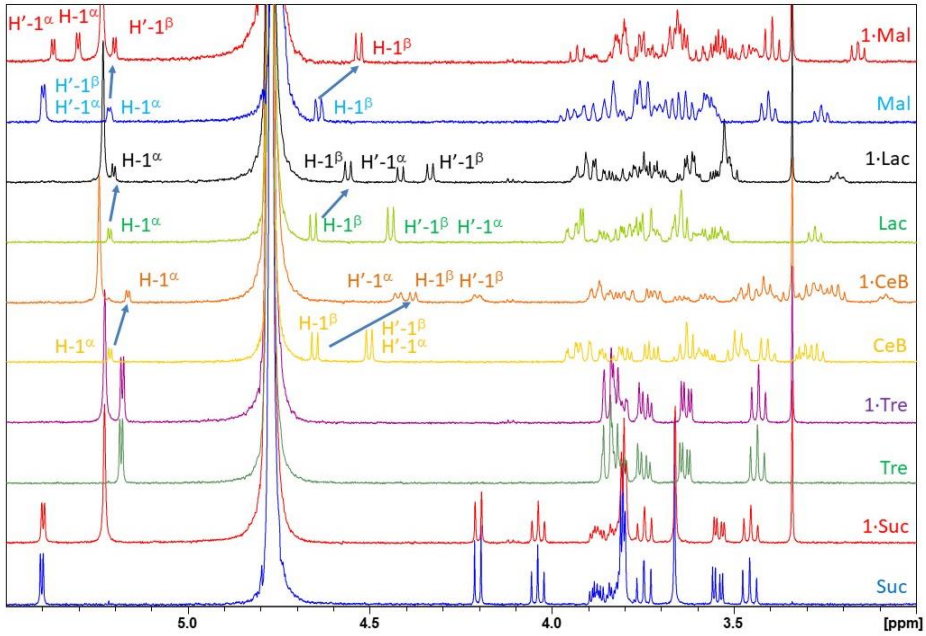


Figure S3. ^{1}H NMR spectra (500 MHz, D_2O) of a 1 mM solution of disaccharides and an equimolar mixture of disaccharides and **1** (1 mM each) at pD 11. Variations of the H-1 proton signal shifts of CeB, Lac, Mal are Δδ = 0.05, 0.01, 0.03 ppm for the α and Δδ = 0.23, 0.10, 0.11 ppm for the β anomer respectively.

NMR titrations and data analysis: Titrations were performed at 298 K, 500 MHz in 5 mm NMR tubes using microsyringes, following a previously described technique.[20] Concentration of the receptor was measured using an internal standard ($DMSO_2$) preliminarily to each binding measurement and maintained constant during the titrations with glycosides to avoid changes in ionic strength. The stock solutions of **1** were prepared in D_2O adjusting the pD with a diluted NaOH solution in D_2O. A correction factor of +0.4 was applied to the pH values measured by the pH meter to determine the pD values (pD = pH + 0.4). DSS was used as internal reference. The alkaline stock solution of **1** was stored during the titrations under nitrogen atmosphere to avoid acid/base reactions with atmospheric CO_2. Following this strategy, constant values of pD were observed during titrations and dilution experiments. DSS was used as internal reference. Dimerization constants of receptor 1 at pD 7.4 and at pD 11, were set invariant in the non-linear regression analysis of receptor-glycosides binding data measured at pD 7.4 and pD 11, respectively. Mathematical analysis of data and graphic presentation of results was performed using the HypNMR 2006 program.[21] Results pages and Plots of experimental and calculated shifts are reported hereafter.

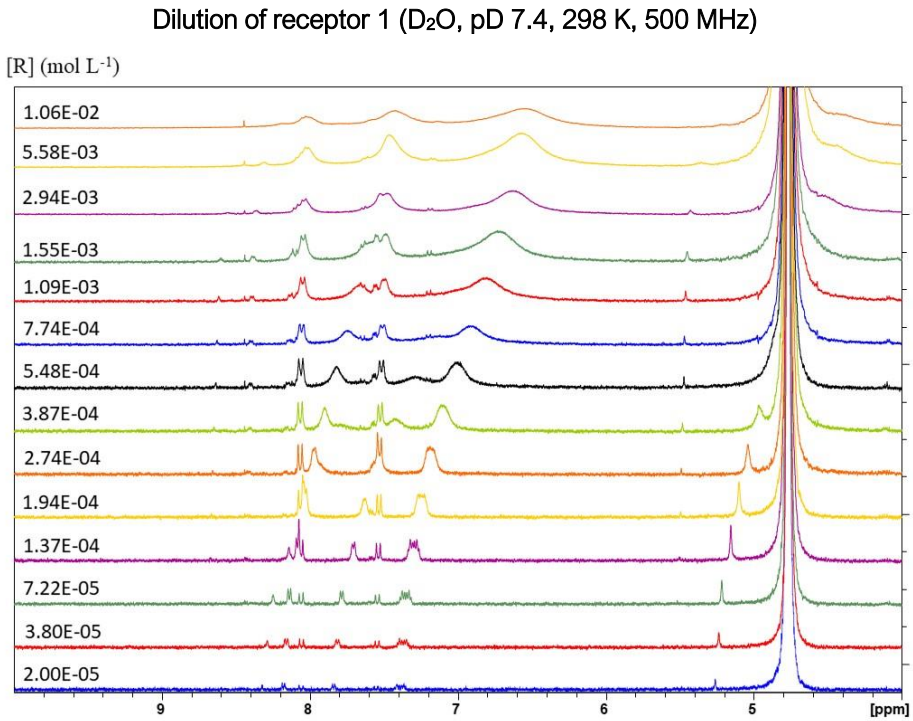


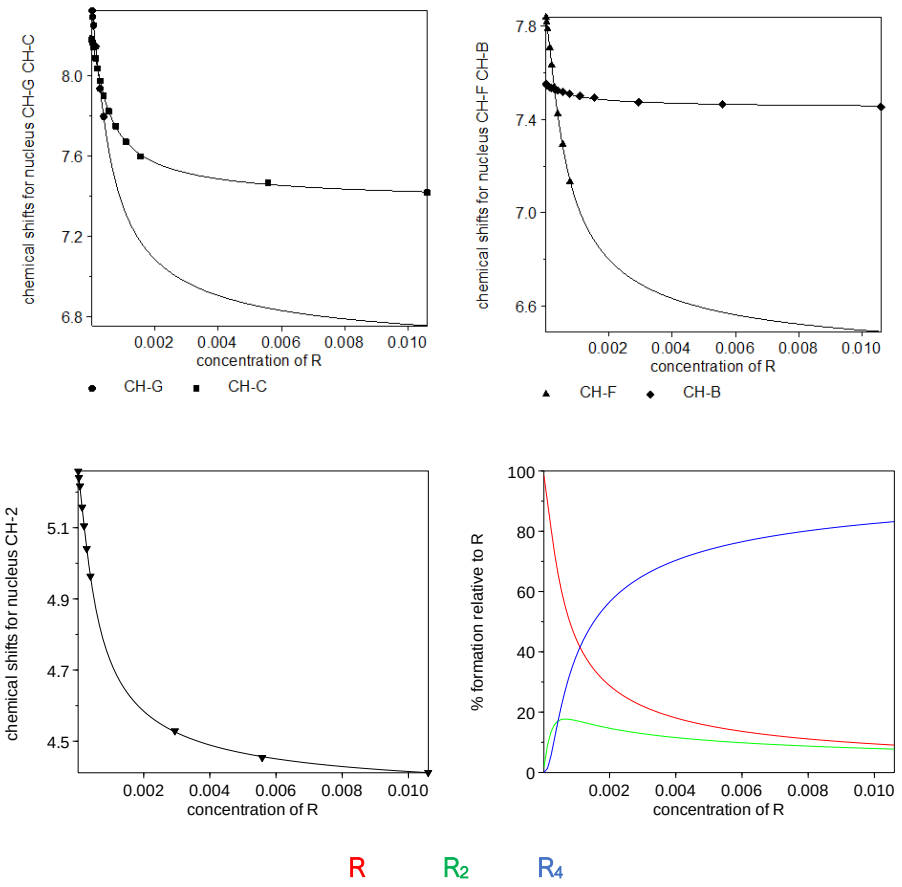
Figure S4. ^{1}H NMR spectroscopic spectra (500 MHz, D_2O, pD 7.4, 298 K) of receptor **1** (R) at different concentrations in dilution expertiment.

Results page

```
 no. of spectra            14
 no. of resonance values   52
 no. of resonant nuclei     5

Chi-squared =   7.38

   sigma =   0.00472549283          RMS weighted residual =   0.00387685411

      stoich          value   relative    log    standard
      coeff                   std devn    beta   deviation

 Beta  2 refined  4.4211E+002  0.1557    2.6455   0.0676  (   R2  )
 Beta  4 refined  2.5653E+009  0.2599    9.4091   0.1129  (   R4  )

Individual chemical shifts

                    R                          2
 ========================================================
        +        value    error          value    error
CH-G    +       8.3574   0.0047          6.4205   0.3035
CH-C    +       8.2006   0.0039          7.1043   0.1514
CH-F    +       7.8633   0.0045          6.4857   0.2426
CH-B    +       7.5491   0.0032          7.5032   0.0267
CH-2    +       5.2787   0.0039          4.1586   0.1697
        +
                    4
 ================================
        +        value    error
CH-G    +       6.6141   0.1198
CH-C    +       7.3655   0.0109
CH-F    +       6.3413   0.0454
CH-B    +       7.4446   0.0051
CH-2    +       4.3405   0.0119

Correlation coefficients*1000


       1    2
   1
   2  834

 Parameters are numbered as follows
    1 beta 2
    2 beta 4
```

Titration Plots

Chemical shifts (δ, ppm) vs. concentration of R (mol L^{-1})
experimental (symbols) and calculated (lines) values



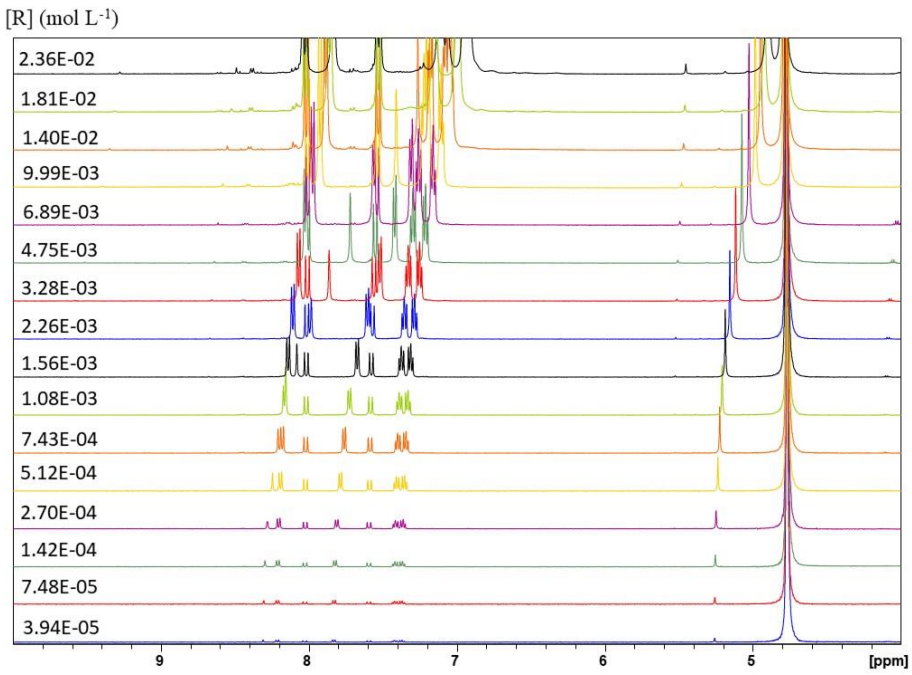


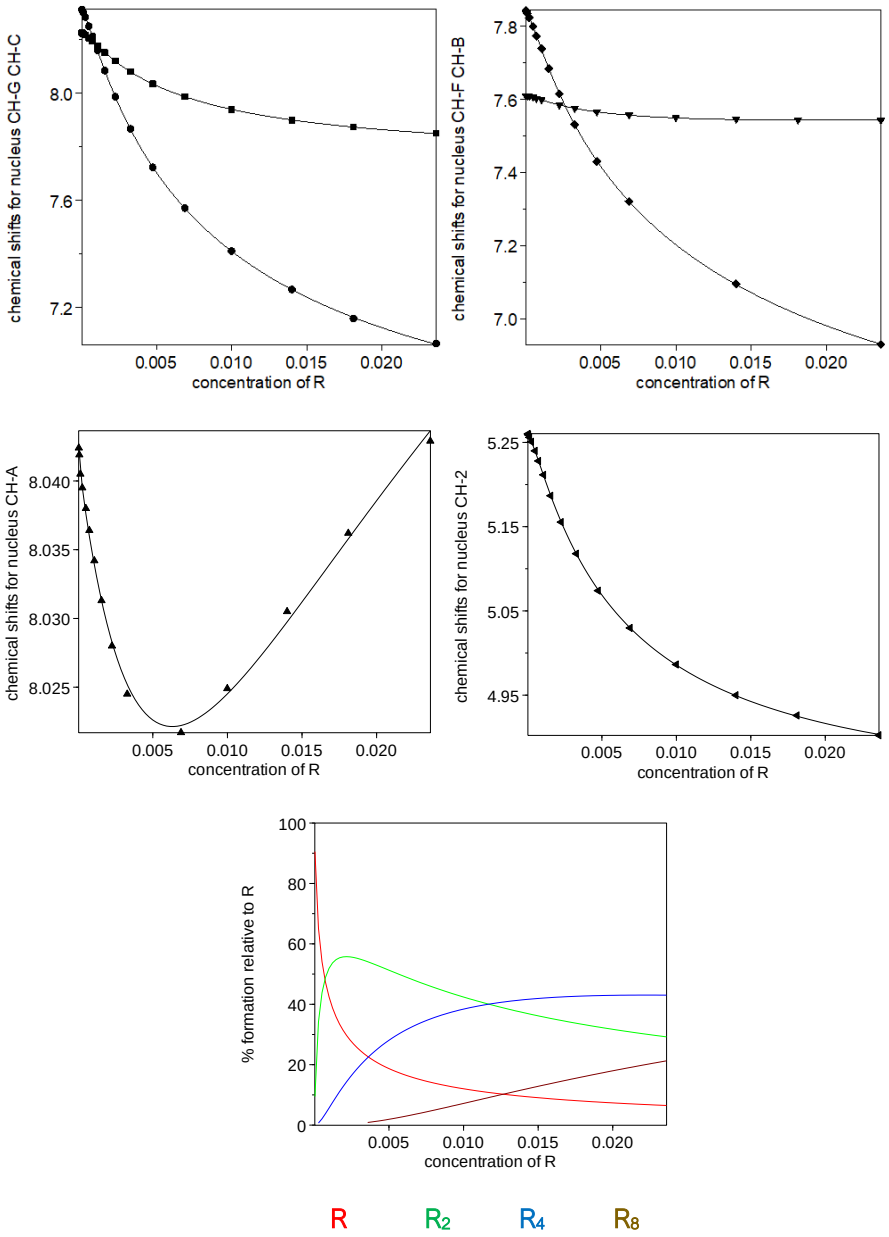
Figure S5. ^{1}H NMR spectroscopic spectra (500 MHz, D_2O, pD 11, 298 K) of receptor **1** (R) at different concentrations in dilution experiment.

Results page

```
 no. of spectra            16
 no. of resonance values   92
 no. of resonant nuclei     6

   sigma =   0.00105982572          RMS weighted residual =   0.00089083489

      stoich          value   relative    log    standard
      coeff                   std devn    beta   deviation

 Beta  2 refined  1.4667E+003  0.1658    3.1663   0.0720  (   R2   )
 Beta  4 refined  4.5958E+008  0.2472    8.6624   0.1074  (   R4   )
 Beta  8 refined  2.0568E+019  0.4825   19.3132   0.2096  (   R8   )

Individual chemical shifts

                    R                         2
 ============================================================
        +        value    error          value    error
CH-G    +       8.3177   0.0011         8.2536   0.0114
CH-C    +       8.2258   0.0010         8.2093   0.0038
CH-A    +       8.0434   0.0009         8.0333   0.0018
CH-F    +       7.8479   0.0010         7.7994   0.0084
CH-B    +       7.6101   0.0009         7.6054   0.0020
CH-2    +       5.2626   0.0009         5.2400   0.0039
        +
                    4                         8
 ============================================================
        +        value    error          value    error
CH-G    +       6.3761   0.0719         6.4302   0.0307
CH-C    +       7.5794   0.0219         7.7824   0.0250
CH-A    +       7.9777   0.0046         8.1913   0.0221
CH-F    +       6.5057   0.0526         6.3138   0.0242
CH-B    +       7.4546   0.0057         7.6127   0.0181
CH-2    +       4.6419   0.0208         4.8604   0.0263

 Correlation coefficients*1000


       1    2    3
   1
   2  990
   3  890  939

 Parameters are numbered as follows
    1 beta 2
    2 beta 4
    3 beta 8
```

Titration Plots

Chemical shifts (δ, ppm) vs. concentration of R (mol L^{-1})
experimental (symbols) and calculated (lines) values



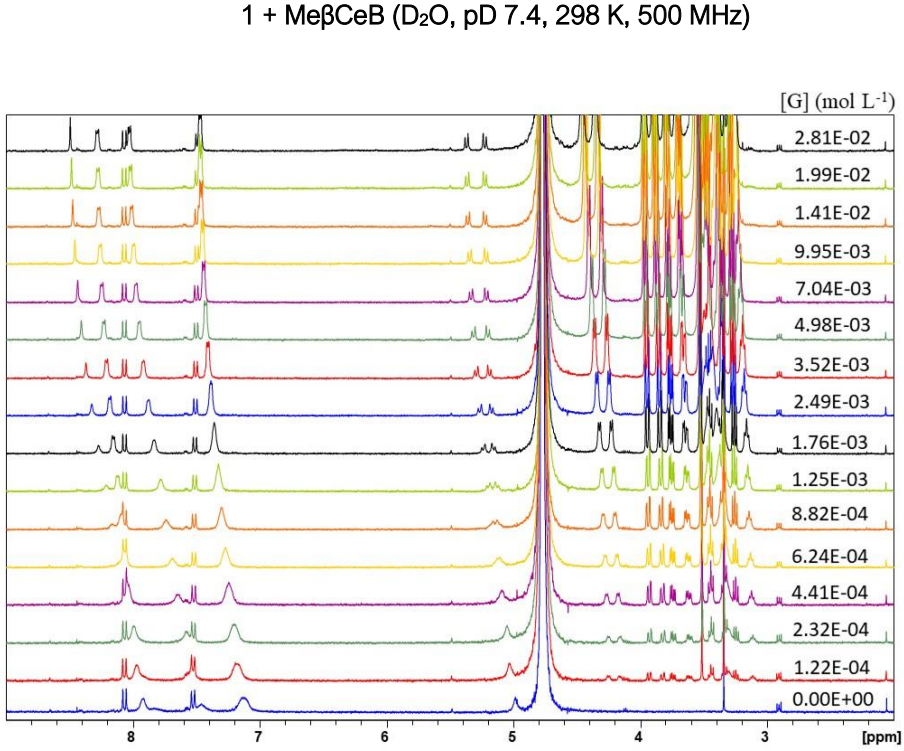


Figure S6. ^{1}H NMR spectroscopic titration (500 MHz, D_2O, pD 7.4, 298 K) of receptor **1** ($5.06\ 10^{-4}$ mol L^{-1}) with incremental concentrations of MeβCeB (G).

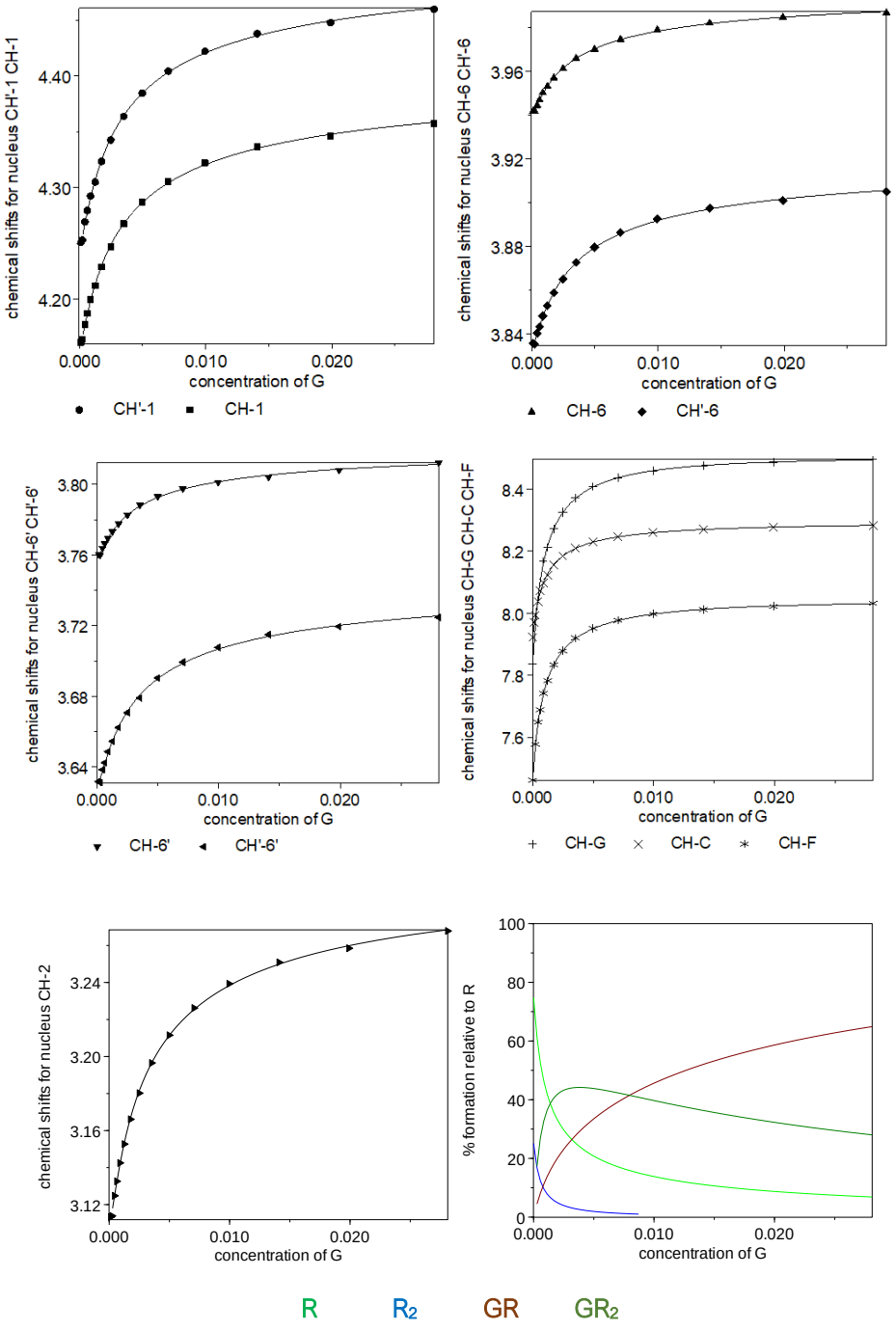
Results page

```
 no. of spectra            16
 no. of resonance values  148
 no. of resonant nuclei    10

   sigma =   0.00181807032          RMS weighted residual =   0.00158861640

       stoich            value   relative    log    standard
       coeff                     std devn    beta   deviation

 Beta  0  2 constant 4.4208E+002             2.6455           (   R2   )
 Beta  1  1 refined  3.4165E+002  0.1680     2.5336   0.0730  (   GR   )
 Beta  1  2 refined  2.1265E+006  0.1447     6.3277   0.0628  (   GR2  )

Individual chemical shifts

                     G                          R
 ============================================================
        +        value    error          value    error
CH'-1   +       4.5038   0.0044
CH-1    +       4.3992   0.0042
CH-6    +       3.9969   0.0019
CH'-6   +       3.9205   0.0022
CH-6'   +       3.8219   0.0020
CH'-6'  +       3.7458   0.0026
CH-2    +       3.3007   0.0035
CH-G    +                                7.5399   0.1320
CH-C    +                                7.7988   0.0778
CH-F    +                                7.2183   0.1163
        +
                    0,2                        1,1
 ============================================================
        +        value    error          value    error
CH'-1   +                                0.7438   0.6846
CH-1    +                                0.8403   0.6484
CH-6    +                                3.1554   0.1679
CH'-6   +                                2.6206   0.2457
CH-6'   +                                2.8986   0.1814
CH'-6'  +                                1.9671   0.3298
CH-2    +                                0.4787   0.5160
CH-G    +       8.7195   0.3943          8.4876   0.0163
CH-C    +       8.3034   0.2330          8.2915   0.0082
CH-F    +       8.1956   0.3475          8.0330   0.0134
        +
                    1,2
 ================================
        +        value    error
CH'-1   +       4.8811   0.0342
CH-1    +       4.7639   0.0344
CH-6    +       4.0873   0.0315
CH'-6   +       4.0621   0.0319
CH-6'   +       3.9142   0.0314
CH'-6'  +       3.9525   0.0331
CH-2    +       3.5928   0.0334
CH-G    +       8.7434   0.0238
CH-C    +       8.3830   0.0199
CH-F    +       8.2233   0.0227

 Correlation coefficients*1000

       1    2
   1
   2 -255
```

```
Parameters are numbered as follows
   1 beta 1,1
   2 beta 1,2
```

Titration Plots

Chemical shifts (δ, ppm) *vs*. concentration of G (mol L^{-1})
experimental (symbols) and calculated (lines) values



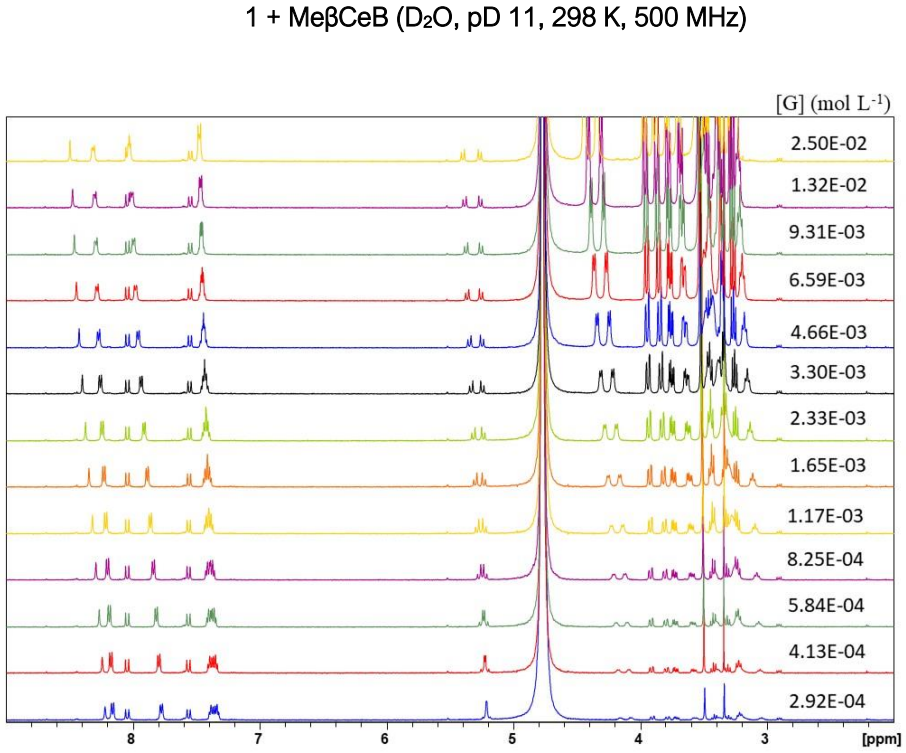


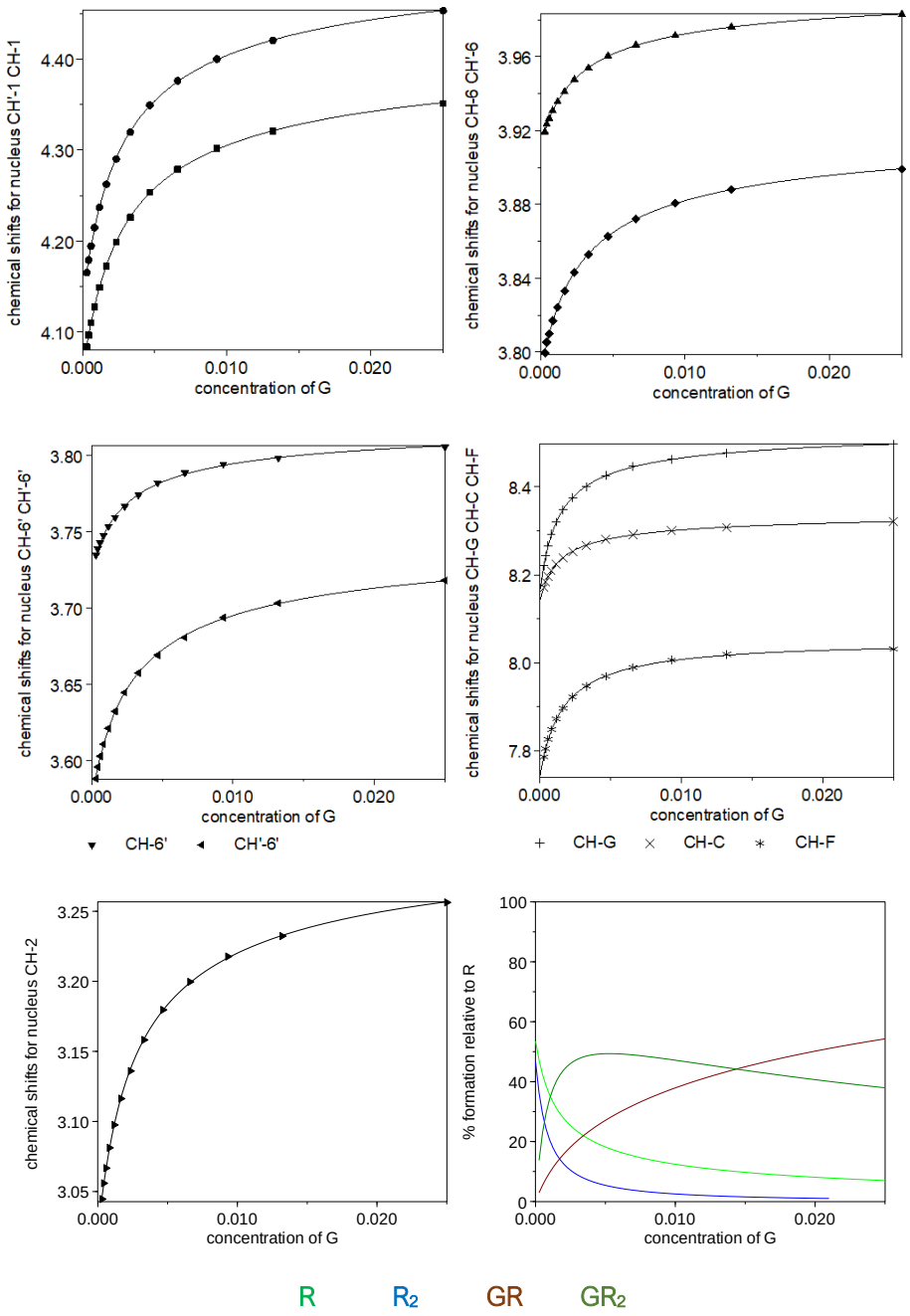
Figure S7. ^{1}H NMR spectroscopic titration (500 MHz, D_2O, pD 11, 298 K) of receptor **1** ($5.55\ 10^{-4}$ mol L^{-1}) with incremental concentrations of MeβCeB (G).

Results page

```
 no. of spectra            14
 no. of resonance values  130
 no. of resonant nuclei    10

Chi-squared =   6.98

   sigma =   0.00097912641          RMS weighted residual =   0.00083700662
       stoich            value   relative    log    standard
       coeff                     std devn    beta   deviation
 Beta  0  2 constant 1.4666E+003             3.1663           (   R2   )
 Beta  1  1 refined  3.1598E+002  0.1775     2.4997   0.0771  (   GR   )
 Beta  1  2 refined  2.8496E+006  0.2383     6.4548   0.1035  (   GR2  )

Individual chemical shifts
                     G                          R
 ========================================================
        +        value    error         value    error
CH'-1   +       4.5196   0.0058
CH-1    +       4.4154   0.0055
CH-6    +       3.9975   0.0019
CH'-6   +       3.9225   0.0024
CH-6'   +       3.8215   0.0020
CH'-6'  +       3.7481   0.0029
CH-2    +       3.3058   0.0044
CH-G    +                                 7.9930   0.2024
CH-C    +                                 8.0909   0.1167
CH-F    +                                 7.3215   0.1554
                   0,2                       1,1
 ========================================================
        +        value    error         value    error
CH'-1   +                                -0.9746   1.1777
CH-1    +                                -0.8523   1.1225
CH-6    +                                 2.8082   0.2650
CH'-6   +                                 1.9944   0.4170
CH-6'   +                                 2.5673   0.2828
CH'-6'  +                                 1.2173   0.5421
CH-2    +                                -0.7689   0.8729
CH-G    +       8.3548   0.2306          8.5330   0.0093
CH-C    +       8.1952   0.1337          8.3491   0.0071
CH-F    +       8.2206   0.1799          8.0364   0.0121
                   1,2
 ================================
        +        value    error
CH'-1   +       4.5074   0.1530
CH-1    +       4.4573   0.1320
CH-6    +       3.9878   0.0435
CH'-6   +       3.9309   0.0559
CH-6'   +       3.7791   0.0527
CH'-6'  +       3.7742   0.0659
CH-2    +       3.3105   0.1110
CH-G    +       8.5420   0.0368
CH-C    +       8.3256   0.0253
CH-F    +       8.1530   0.0192

 Correlation coefficients*1000
       1    2
   1
   2  -79

 Parameters are numbered as follows
    1 beta 1,1
    2 beta 1,2
```

Titration Plots

Chemical shifts (δ, ppm) *vs*. concentration of G (mol L^{-1})
experimental (symbols) and calculated (lines) values



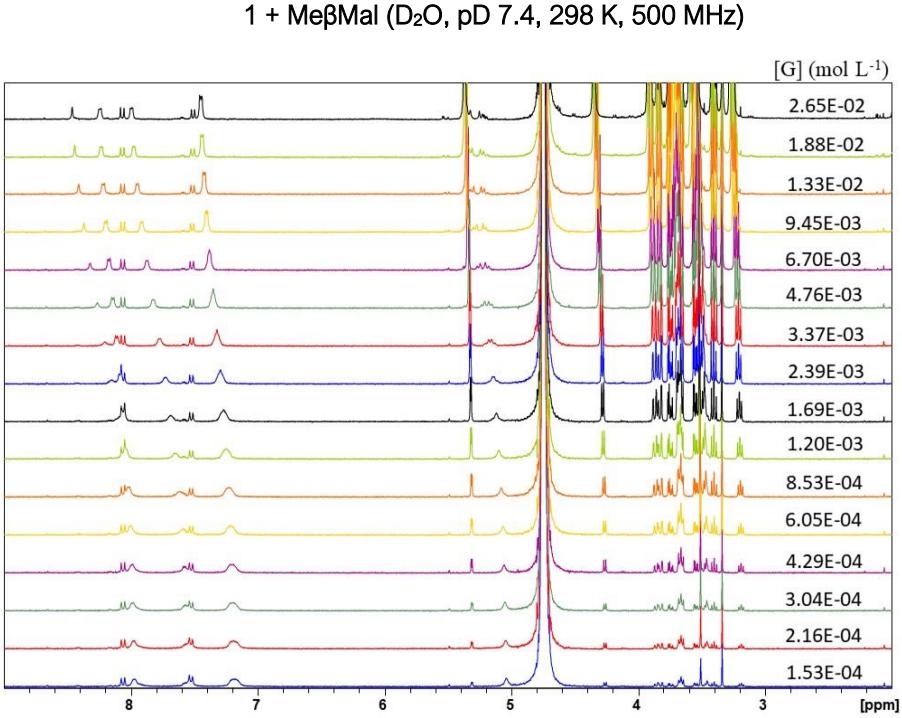


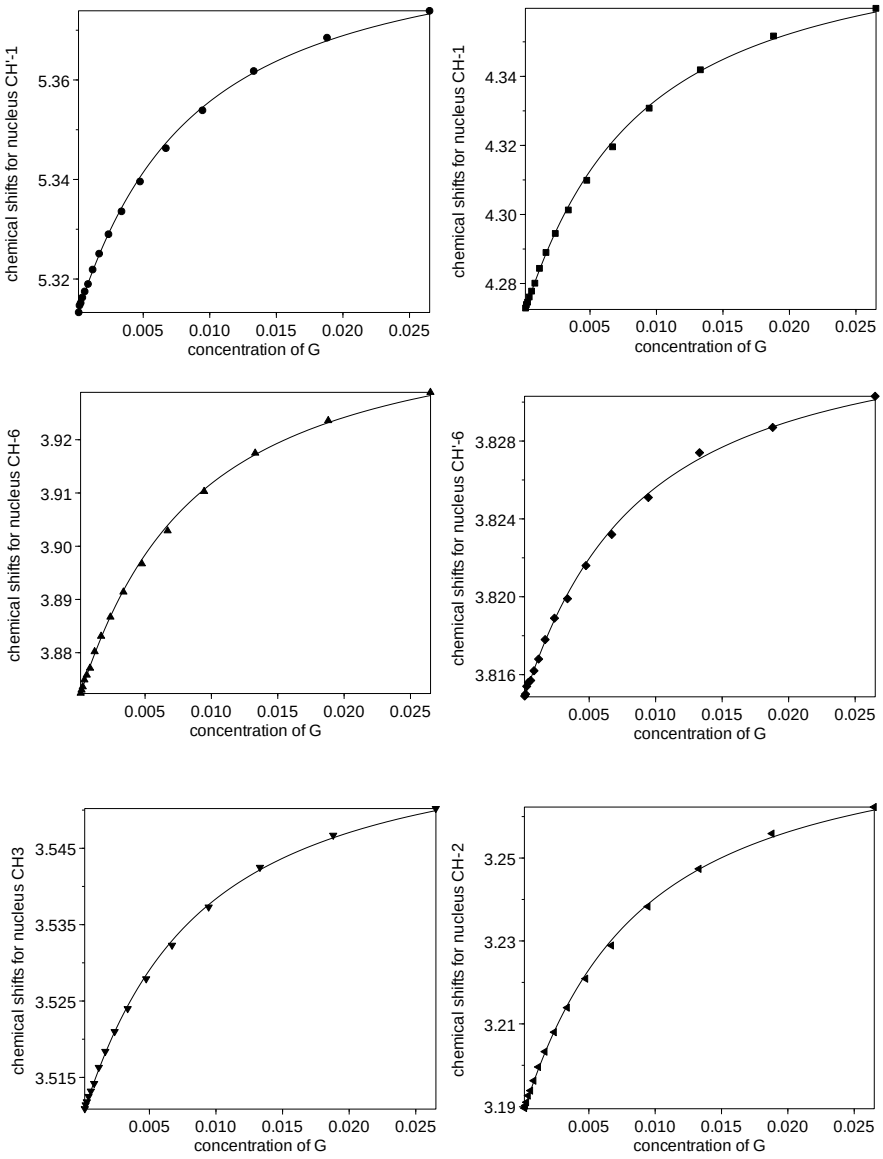
Figure S8. ^{1}H NMR spectroscopic titration (500 MHz, D_2O, pD 7.4, 298 K) of receptor **1** (4.01 10^{-4} mol L^{-1}) with incremental concentrations of MeβMal (G)

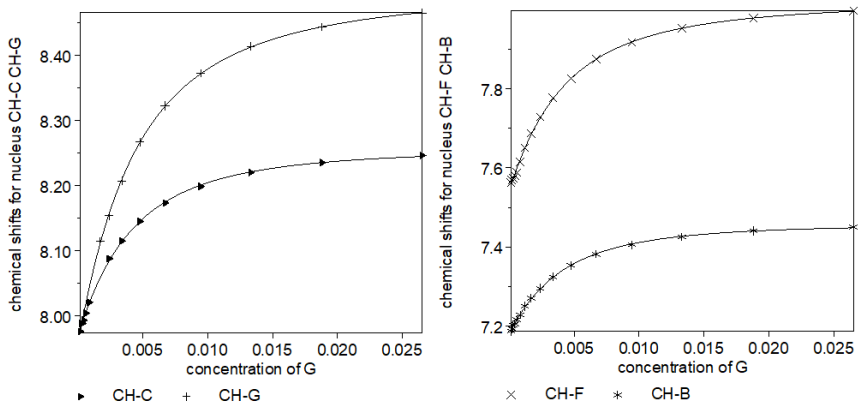
Results page

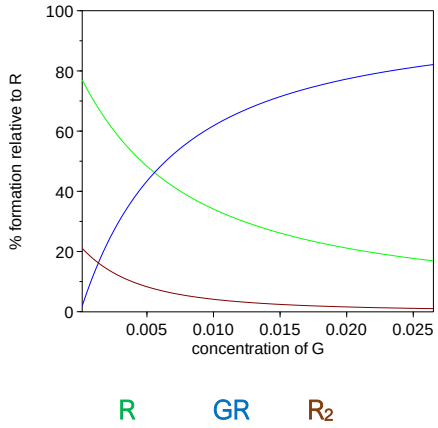
```
 no. of spectra            16
 no. of resonance values  151
 no. of resonant nuclei    10

   sigma =   0.00194509826          RMS weighted residual =   0.00177679918

       stoich            value   relative    log    standard
       coeff                     std devn    beta   deviation

 Beta  1  1 refined  1.8579E+002  0.0334    2.2690   0.0145  (   GR   )
 Beta  0  2 constant 4.4208E+002            2.6455           (   R2   )

Individual chemical shifts

                    G                          R
 ============================================================
        +        value    error          value    error
CH'-1   +       5.3910   0.0017
CH-1    +       4.3842   0.0019
CH-6    +       3.9449   0.0017
CH'-6   +       3.8347   0.0015
CH3     +       3.5615   0.0016
CH-2    +       3.2830   0.0018
CH-C    +                                8.2716   0.0258
CH-G    +                                8.3829   0.0383
CH-F    +                                7.9889   0.0387
CH-B    +                                7.4950   0.0258
        +
                   1,1                        0,2
 ============================================================
        +        value    error          value    error
CH'-1   +       3.9648   0.0662
CH-1    +       2.3238   0.0890
CH-6    +       2.6062   0.0632
CH'-6   +       3.4697   0.0365
CH3     +       2.6269   0.0502
CH-2    +       1.5599   0.0766
CH-C    +       8.2561   0.0048          6.8802   0.0934
CH-G    +       8.5083   0.0059          6.4217   0.1399
CH-F    +       8.0230   0.0061          5.9277   0.1386
CH-B    +       7.4560   0.0048          6.0455   0.0934
```

Titration Plots

Chemical shifts (δ, ppm) *vs*. concentration of G (mol L^{-1})
experimental (symbols) and calculated (lines) values




chemical shifts for nucleus CH-C CH-G
8.40
8.30
8.20
8.10
8.00
0.005
0.010
0.015
0.020
0.025
concentration of G
CH-C
CH-G
chemical shifts for nucleus CH-F CH-B
7.8
7.6
7.4
7.2
0.005
0.010
0.015
0.020
0.025
concentration of G
CH-F
CH-B


% formation relative to R
100
80
60
40
20
0
0.005
0.010
0.015
0.020
0.025
concentration of G
R
GR
R2

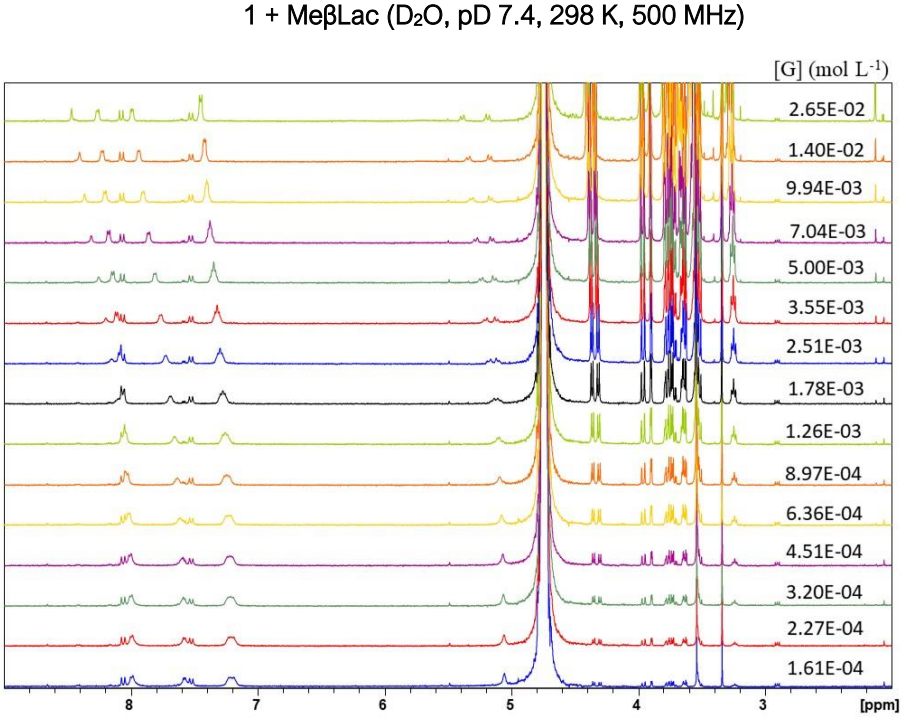


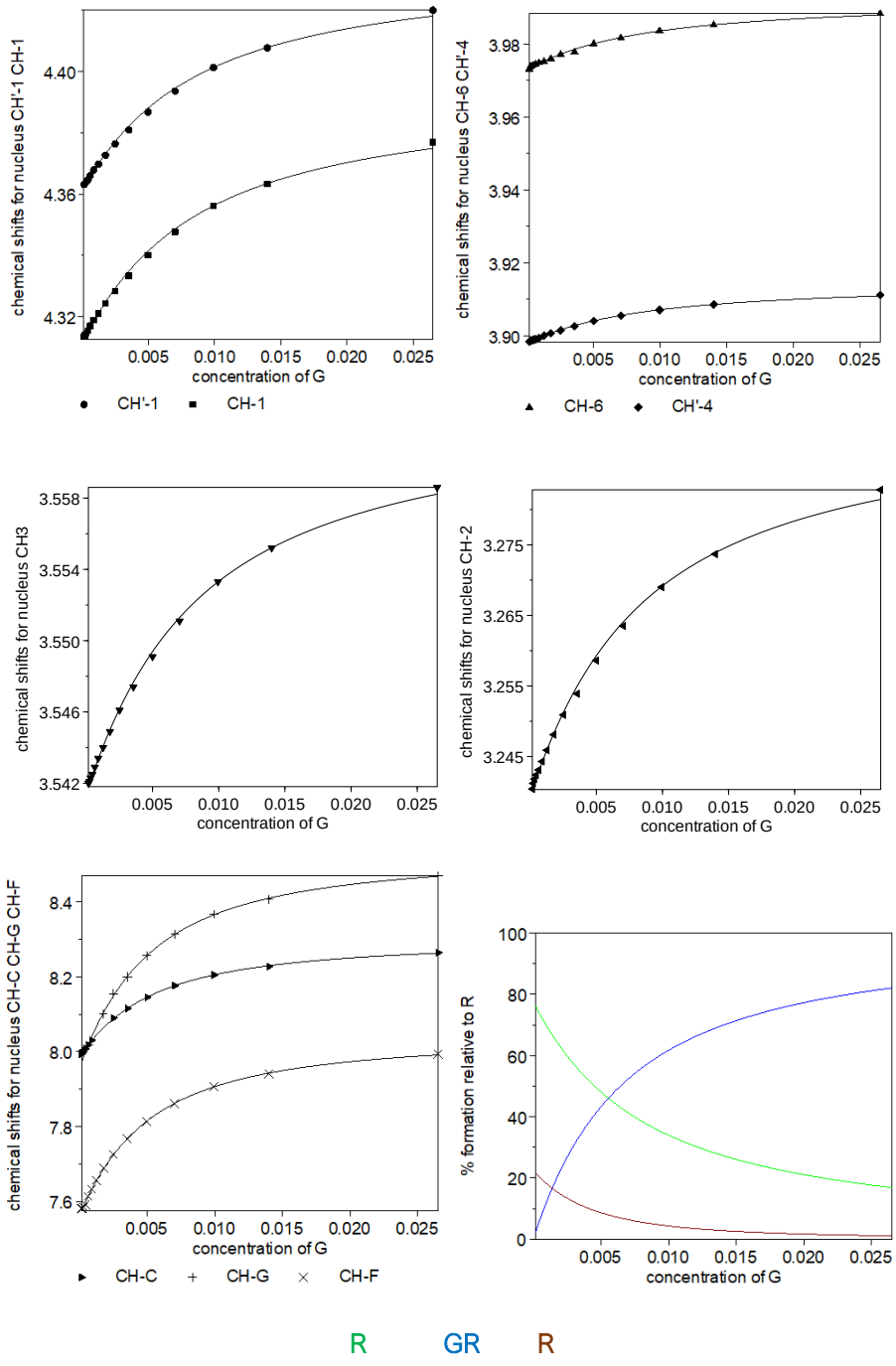
Figure S9. ^{1}H NMR spectroscopic titration (500 MHz, D_2O, pD 7.4, 298 K) of receptor **1** ($4.22\ 10^{-4}$ mol L^{-1}) with incremental concentrations of MeβLac (G).

Results page

```
 no. of spectra            15
 no. of resonance values  126
 no. of resonant nuclei     9

   sigma =   0.00148779121          RMS weighted residual =   0.00135167844

       stoich            value   relative    log    standard
       coeff                     std devn    beta   deviation

 Beta  1  1 refined  1.8642E+002  0.0450    2.2705   0.0195  (   GR   )
 Beta  0  2 constant 4.4208E+002            2.6455           (   R2   )

 Individual chemical shifts

                     G                          R
 ============================================================
        +        value    error          value    error
CH'-1   +       4.4349   0.0017
CH-1    +       4.3936   0.0018
CH-6    +       3.9924   0.0013
CH'-4   +       3.9147   0.0013
CH3     +       3.5632   0.0013
CH-2    +       3.2938   0.0015
CH-C    +                                8.1092   0.0283
CH-G    +                                8.2328   0.0459
CH-F    +                                7.8012   0.0435
        +
                    1,1                        0,2
 ============================================================
        +        value    error          value    error
CH'-1   +       3.1512   0.0752
CH-1    +       2.9625   0.0828
CH-6    +       3.6562   0.0326
CH'-4   +       3.6244   0.0313
CH3     +       3.1858   0.0339
CH-2    +       2.3484   0.0583
CH-C    +       8.3035   0.0038          7.5589   0.0978
CH-G    +       8.5352   0.0052          7.0000   0.1556
CH-F    +       8.0467   0.0052          6.7354   0.1496
```

Titration Plots

Chemical shifts (δ, ppm) *vs*. concentration of G (mol L^{-1})
experimental (symbols) and calculated (lines) values



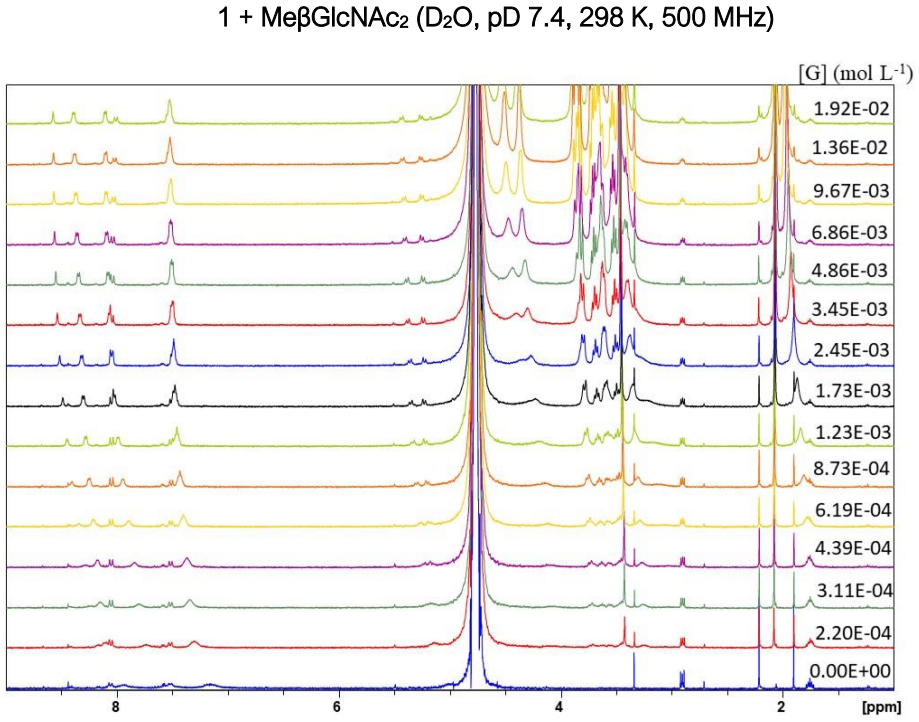


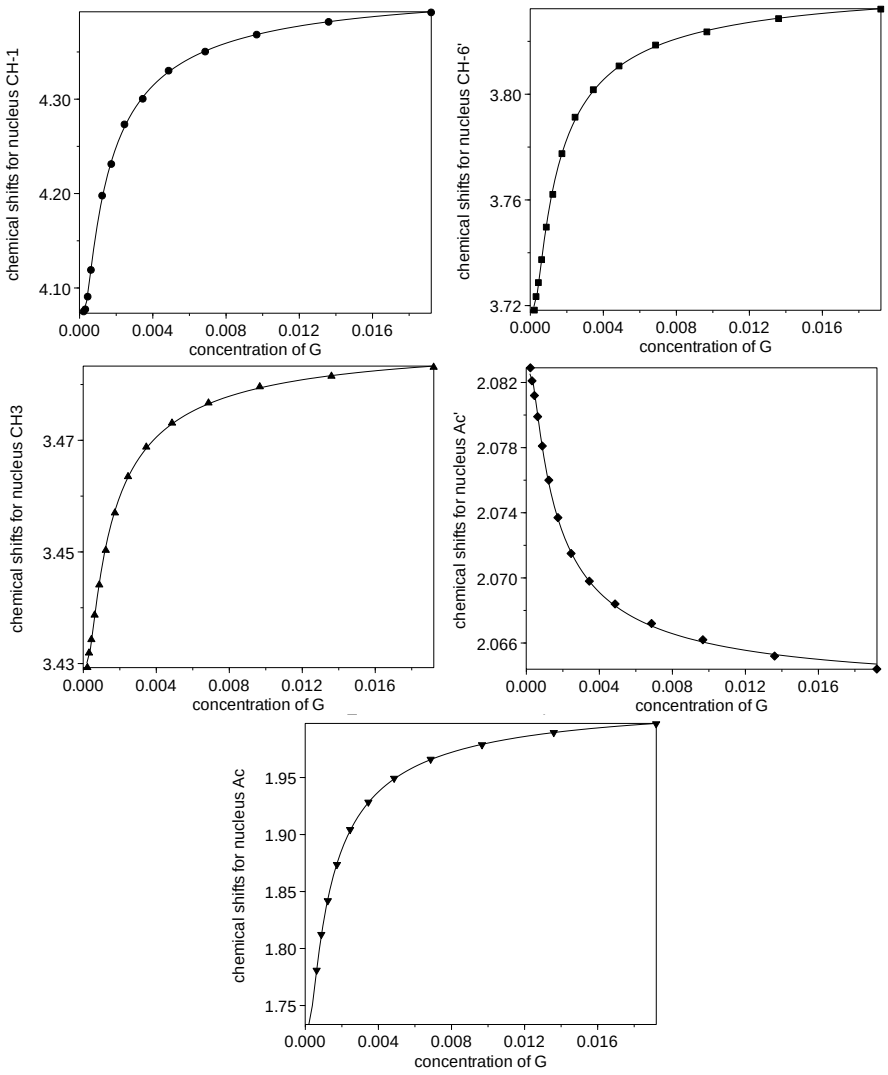
Figure S10. ^{1}H NMR spectroscopic titration (500 MHz, D_2O, pD 7.4, 298 K) of receptor **1** ($5.46\ 10^{-4}$ mol L^{-1}) with incremental concentrations of MeβGlcNAc$_2$ (G).

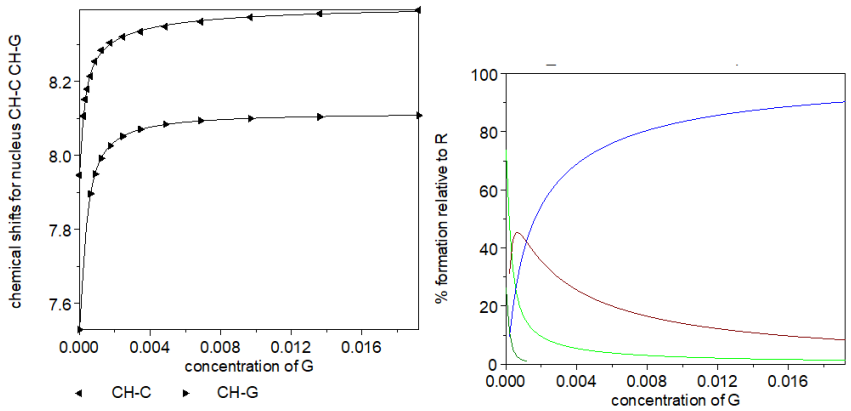
Results page

```
 no. of spectra            15
 no. of resonance values   93
 no. of resonant nuclei     7

   sigma =   0.00208276193          RMS weighted residual =   0.00178095364

       stoich            value   relative    log    standard
       coeff                     std devn    beta   deviation
 Beta  1  1 refined  3.5677E+003  0.0963    3.5524   0.0418  (   GR   )
 Beta  1  2 refined  2.2306E+007  0.2059    7.3484   0.0894  (   GR2  )
 Beta  0  2 constant 4.4208E+002             2.6455           (   R2   )

Individual chemical shifts
                     G                          R
 ============================================================
        +        value    error          value    error
CH-1    +       4.4214   0.0026
CH-6'   +       3.8422   0.0015
CH3     +       3.4880   0.0014
Ac'     +       2.0631   0.0013
Ac      +       2.0194   0.0022
CH-C    +                                7.9737   0.0781
CH-G    +                                6.7497   0.1138
        +
                    1,1                        1,2
 ============================================================
        +        value    error          value    error
CH-1    +       3.3006   0.0740          4.3181   0.0164
CH-6'   +       3.4590   0.0280          3.7986   0.0104
CH3     +       3.3064   0.0171          3.4678   0.0099
Ac'     +       2.1262   0.0130          2.0686   0.0098
Ac      +       1.1726   0.0543          1.8893   0.0340
CH-C    +       8.4080   0.0029          8.2519   0.0267
CH-G    +       8.1160   0.0023          8.2420   0.0203
        +
                    0,2
 ================================
        +        value    error
CH-1    +
CH-6'   +
CH3     +
Ac'     +
Ac      +
CH-C    +       7.8710   0.2208
CH-G    +       9.7161   0.3201

 Correlation coefficients*1000
       1    2
   1
   2  399

 Parameters are numbered as follows
    1 beta 1,1
    2 beta 1,2
```

Titration Plots

Chemical shifts (δ, ppm) *vs*. concentration of G (mol L^{-1})
experimental (symbols) and calculated (lines) values




chemical shifts for nucleus CH-C CH-G
8.2
8.0
7.8
7.6
0.000
0.004
0.008
0.012
0.016
concentration of G
CH-C
CH-G
% formation relative to R
100
80
60
40
20
0
0.000
0.004
0.008
0.012
0.016
concentration of G

R GR GR_2 R_2

1 + MeβGlcNAc (D_2O, pD 7.4, 298 K, 500 MHz).
R = **1** G = MeβGlcNAc

Titration
[R] = 4.97 10^{-4} mol L^{-}

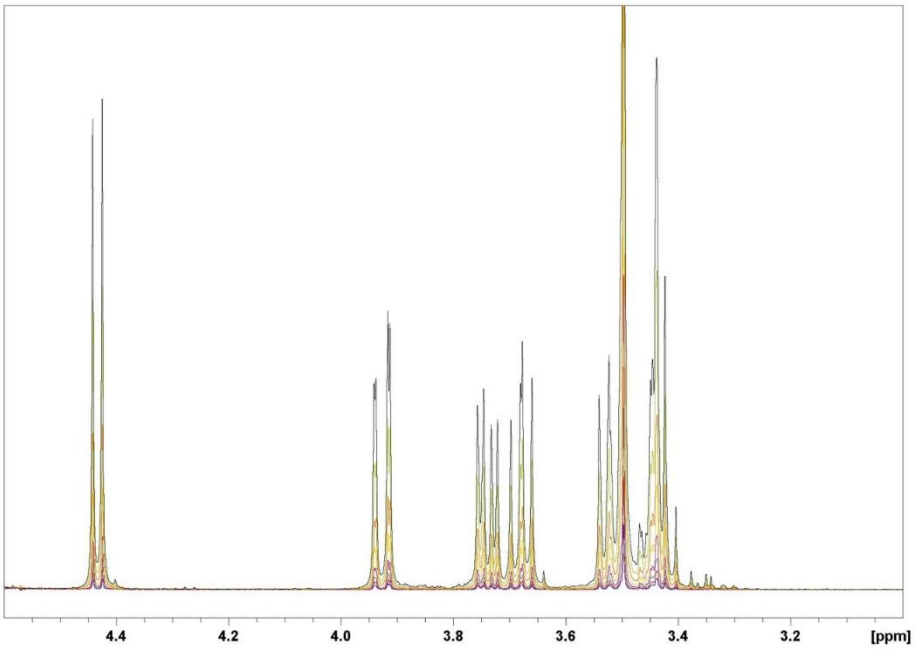


Figure S11. Superposition of ^{1}H NMR spectra registered at incremental concentrations of MeβGlcNAc (0.329 mM, 0.623 mM, 1.18 mM, 2.23 mM, 4.21 mM, 7.57 mM, 15.1 mM, 28.5 mM) in a 0.497 mM solution of **1**. Expansion of the saccharide region.

R:G	$log\beta_n$	BC_{50}^0
1:1	2.50±0.08	1.39±0.29
2:1	6.46±0.10	

Table S1. Cumulative formation constants ($log\beta_n$)[a] and intrinsic median binding concentration (BC_{50}^0, mM)[b] for receptor **1** to MeβCeB (R:G) complexes, measured at 298 K from NMR data in D_2O at pD 11.[c] [a] Formation constants were obtained by nonlinear least-square regression analysis of NMR data. [b] Calculated from the $log\beta$ values using the "BC_{50} Calculator" program.[15] [c] Receptor dimerization constant at pD 11 (**1**: $log\beta_{dim}$ = 3.17±0.07) was set invariant in the nonlinear regression analysis of NMR data.

Calorimetric titrations and data analysis: Isothermal Titration Microcalorimetry experiments were performed at 298 K with a Nano-ITC instrument. After an initial injection of 3 μL, which was excluded from data analysis, aliquots of the titrant solution, containig the glycoside, were injected stepwise into the sample cell containing a solution of the titrate **1**. Titrate solutions containing **1** were prepared in H_2O adjusting the pH with a diluted NaOH solution. All experiments were performed in H_2O at pH 7.4. Heats of dilution were measured by injecting the titrant solution into neat H_2O and then subtracted from the binding heats. To remove ambiguities in the definition of binding models of receptor **1**, data from indipendent titrations performed at different concentrations of the titrate were simultaneously fitted to measure the cumulative association constants and the thermodynamic parameters using the HypCal software package.[22] The dimerization constant $log\beta_{dim}$ 2.65±0.07 of receptor **1** was set invariant in the non linear regression analysis of receptor-glycosides binding data.

1 + MeβCeB (H_2O, pH 7.4, 298 K)

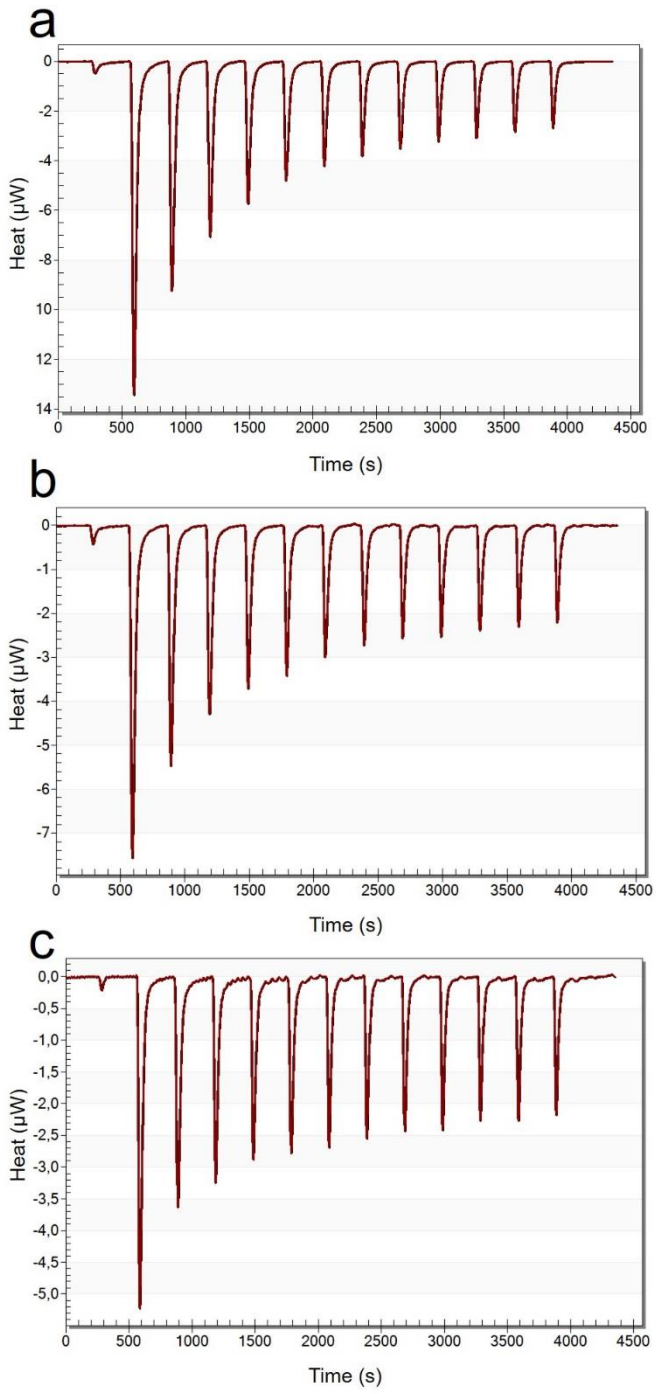


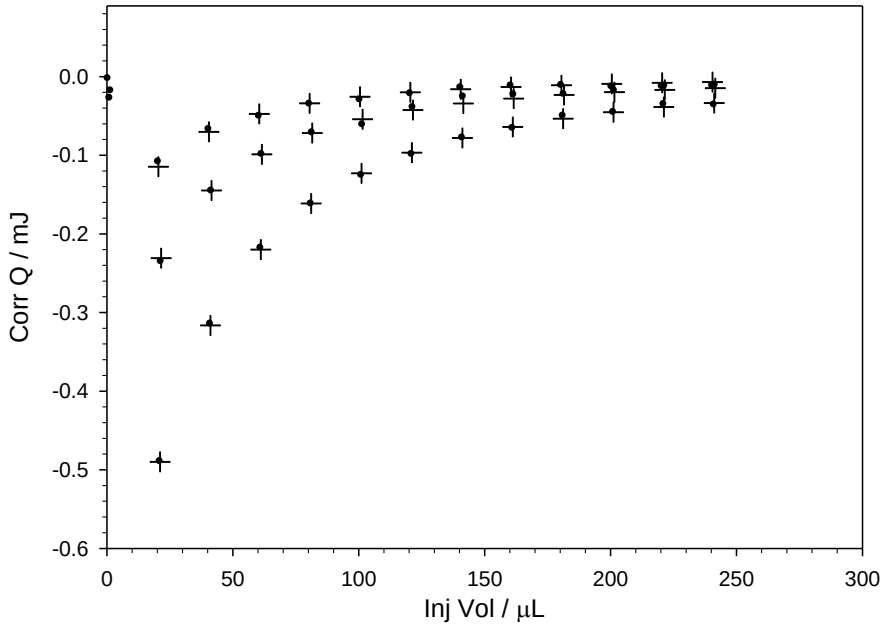
Figure S12. ITC results of MeβCeB with **1** in H_2O, pH 7.4, at 298 K: a) Titration of **1** ($1.25\ 10^{-4}$ mol L^{-1}) with MeβCeB ($4.04\ 10^{-2}$ mol L^{-1}); b) Titration of **1** ($6.25\ 10^{-5}$ mol L^{-1}) with MeβCeB ($4.04\ 10^{-2}$ mol L^{-1}); c) Titration of **1** ($3.13\ 10^{-5}$ mol L^{-1}) with MeβCeB ($4.04\ 10^{-2}$ mol L^{-1}).

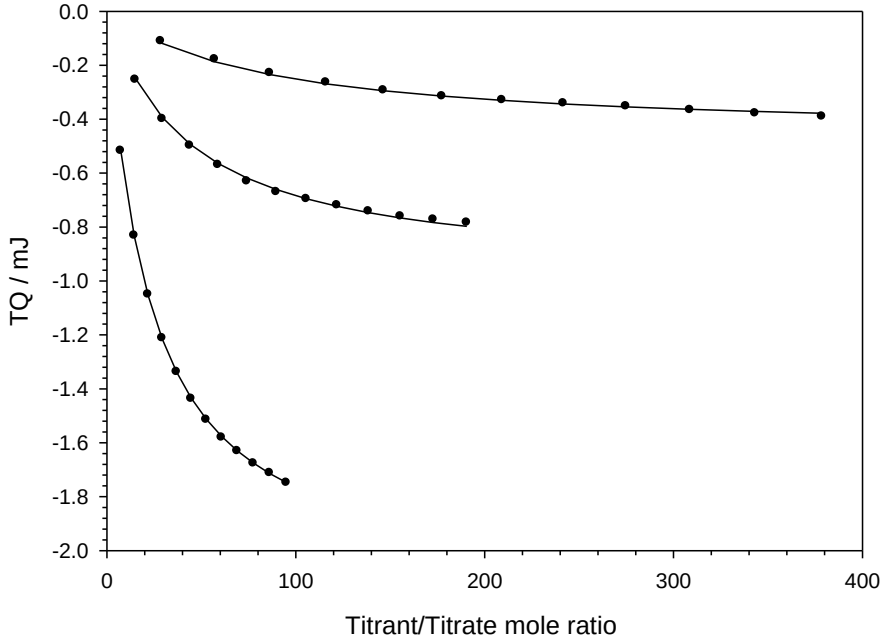
Results page

```
Reagent Reagent
  number   name
    1      R
    2      G

   sigma =   0.00363

 Formation constants   Value    relative    log    standard
                                std devn    beta   deviation
 Beta A refined     0.3790E+03  0.0722    2.5787    0.0314   1   1
 Beta B refined     0.1899E+06  0.6580    5.2786    0.2857   2   1
 Beta C constant    0.4421E+03            2.6455    0.0676   2   0


++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Thermodynamic Functions, kJ/mol


          - DeltaG°          - DeltaH°          T DeltaS°
   A   14.7191  0.1790    15.5995  0.6640     -0.8803  0.8176
   B   30.1304  1.6311  -109.3421  46.6052   139.4725  47.5953
   C   15.1006  0.1055  -108.9897  20.9161   124.0903  20.9163
```

Titration Plots

Experimental (symbols) and calculated (cross and lines) heats





1 + MeβMal (H_2O, pH 7.4, 298 K)

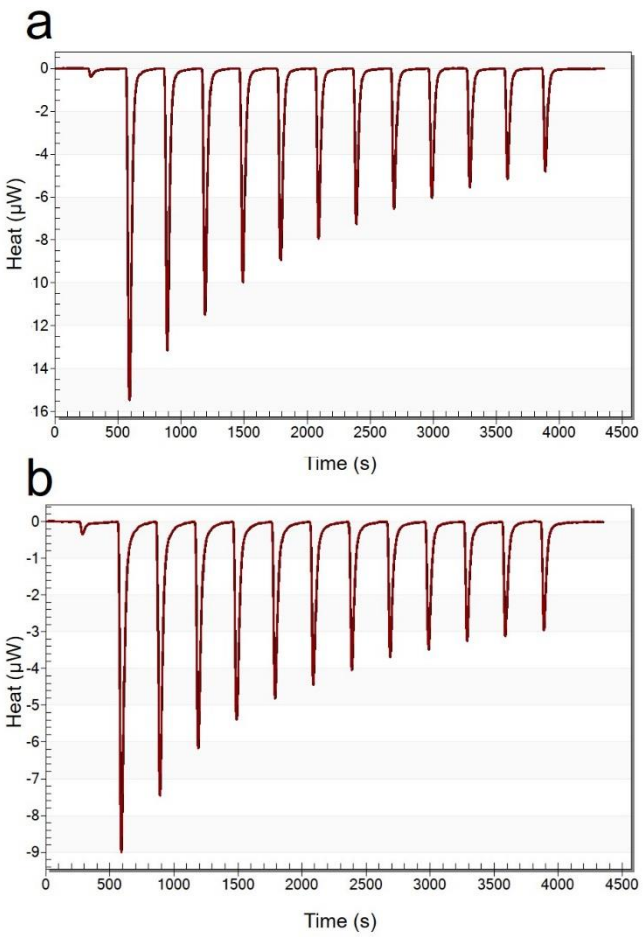
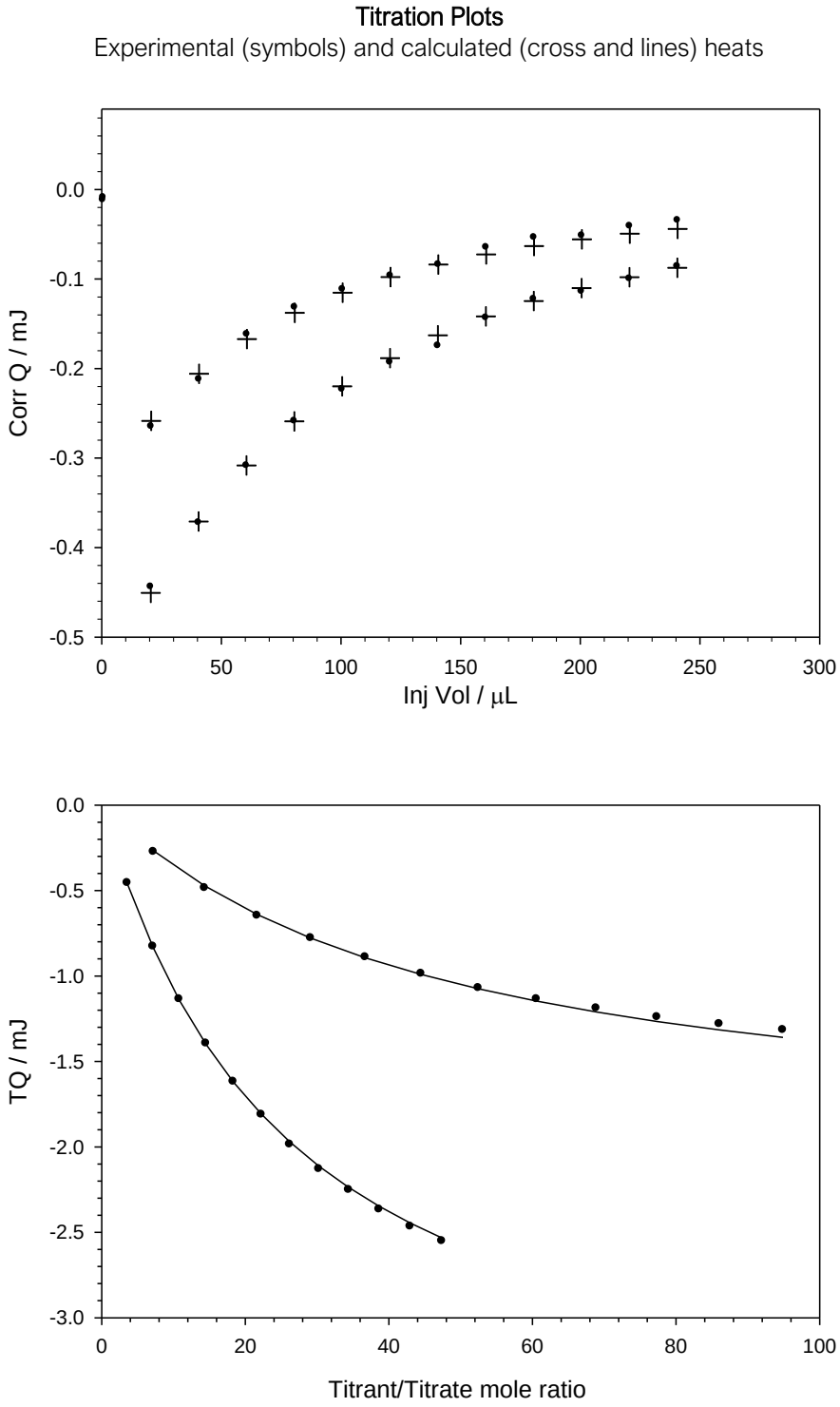


Figure S13. ITC results of MeβMal with **1** in H_2O, pH 7.4, at 298 K: a) Titration of **1** ($2.50\ 10^{-4}$ mol L^{-1}) with MeβMal ($4.04\ 10^{-2}$ mol L^{-1}); b) Titration of **1** ($1.25\ 10^{-4}$ mol L^{-1}) with MeβMal ($4.04\ 10^{-2}$ mol L^{-1}).

Results page

```
Reagent Reagent
  number   name
    1      R
    2      G

   sigma =   0.00596

 Formation constants   Value    relative    log    standard
                                std devn    beta   deviation
 Beta A refined     0.1756E+03  0.0309    2.2444    0.0134   1 1
 Beta B constant    0.4421E+03            2.6455    0.0676   2 0


++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
 Thermodynamic Functions, kJ/mol


          - DeltaG°          - DeltaH°          T DeltaS°
   A   12.8113  0.0766    20.7240  0.4795    -7.9127  0.4721
   B   15.1006  0.0574    23.7075  6.1583    -8.6070  6.1585
```


Titration Plots
Experimental (symbols) and calculated (cross and lines) heats
Corr Q / mJ
Inj Vol / μL
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
0
50
100
150
200
250
300
TQ / mJ
Titrant/Titrate mole ratio
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
0
20
40
60
80
100

1 + MeβLac (H_2O, pH 7.4, 298 K)

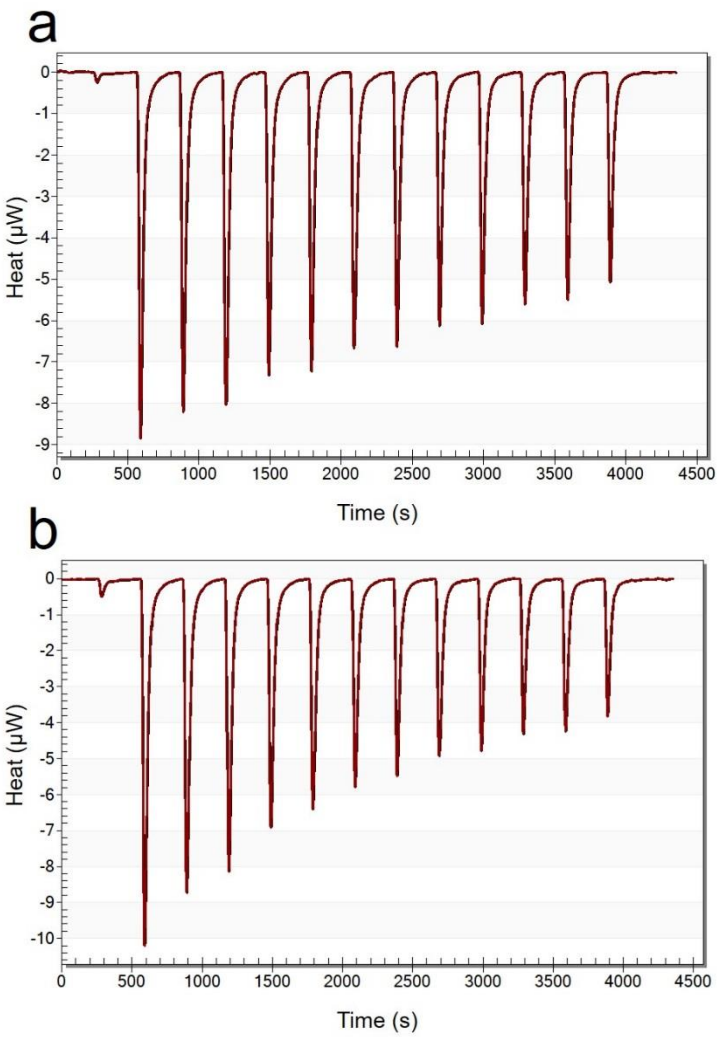


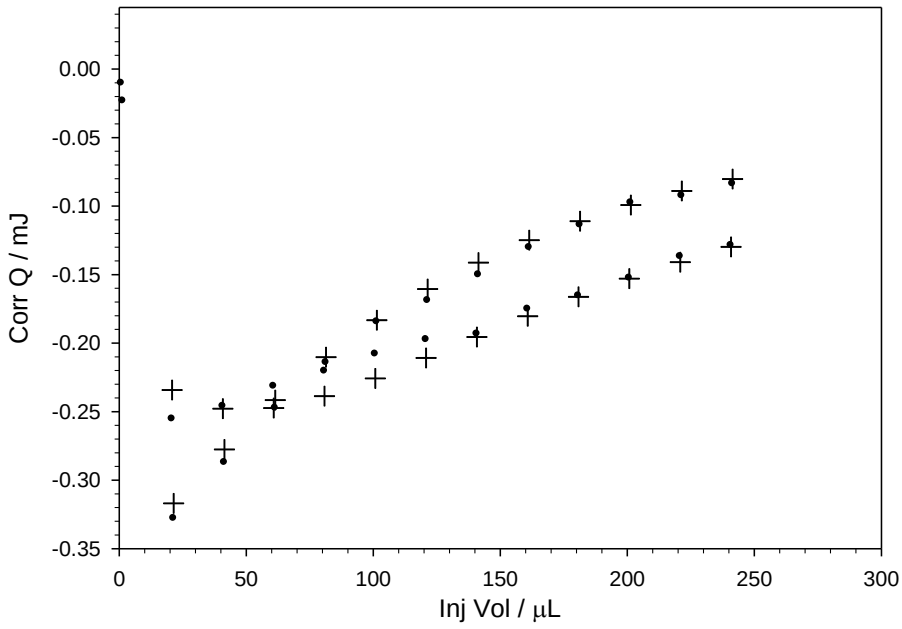
Figure S14. ITC results of MeβLac with **1** in H_2O, pH 7.4, at 298 K: a) Titration of **1** ($5.00\ 10^{-4}$ mol L^{-1}) with MeβLac ($4.05\ 10^{-2}$ mol L^{-1}); b) Titration of **1** ($2.50\ 10^{-4}$ mol L^{-1}) with MeβLac ($4.05\ 10^{-2}$ mol L^{-1})

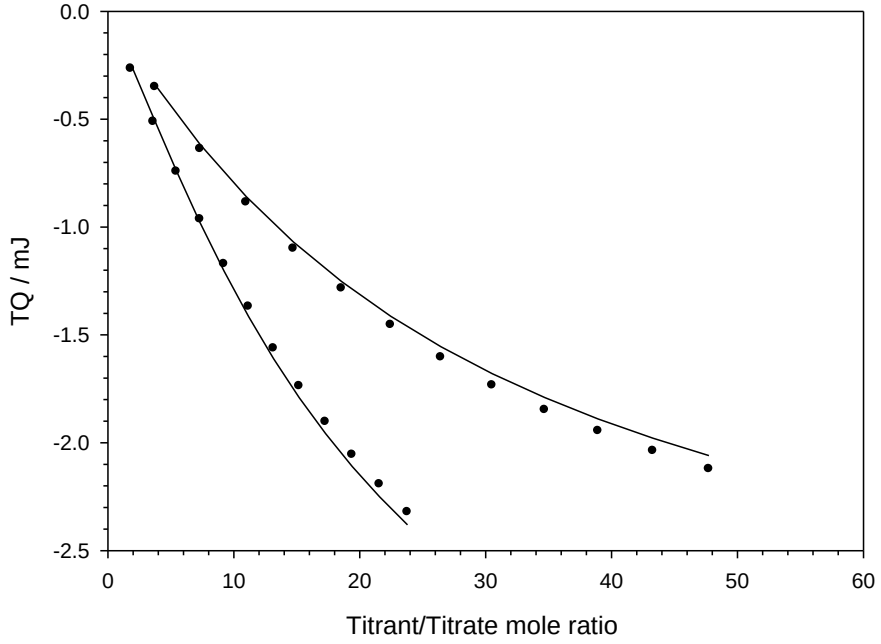
Results page

```
Reagent Reagent
  number   name
    1      R
    2      G

   sigma =   0.00992

 Formation constants   Value    relative    log    standard
                                std devn    beta   deviation
 Beta A refined     0.2043E+03  0.0517    2.3103    0.0225   1 1
 Beta B constant    0.4421E+03            2.6455    0.0676   2 0


++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
 Thermodynamic Functions, kJ/mol


          - DeltaG°          - DeltaH°          T DeltaS°
   A   13.1871  0.1283    23.8721  0.6050    -10.6850  0.6996
   B   15.1006  0.0628   101.0108  2.8812    -85.9102  2.8819
```

Titration Plots

Experimental (symbols) and calculated (cross and lines) heats





1 + MeβGlcNAc$_2$ (H_2O, pH 7.4, 298 K)

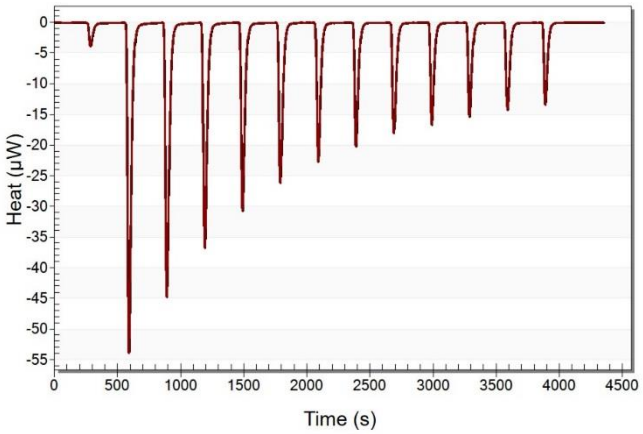


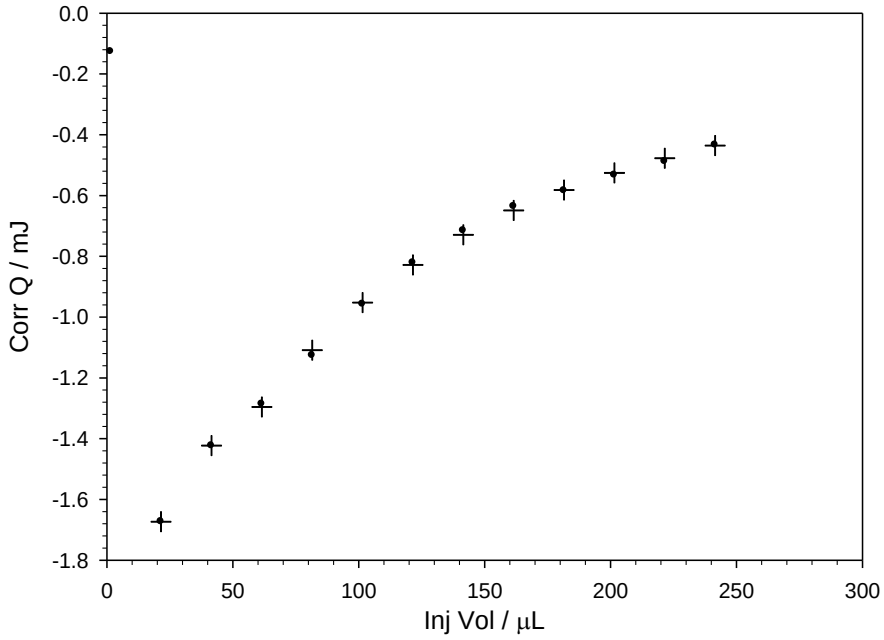
Figure S15. ITC results the titration of **1** ($4.04\ 10^{-4}$ mol L^{-1}) with MeβGlcNAc$_2$ ($1.06\ 10^{-2}$ mol L^{-1}) in H_2O, pH 7.4, at 298 K

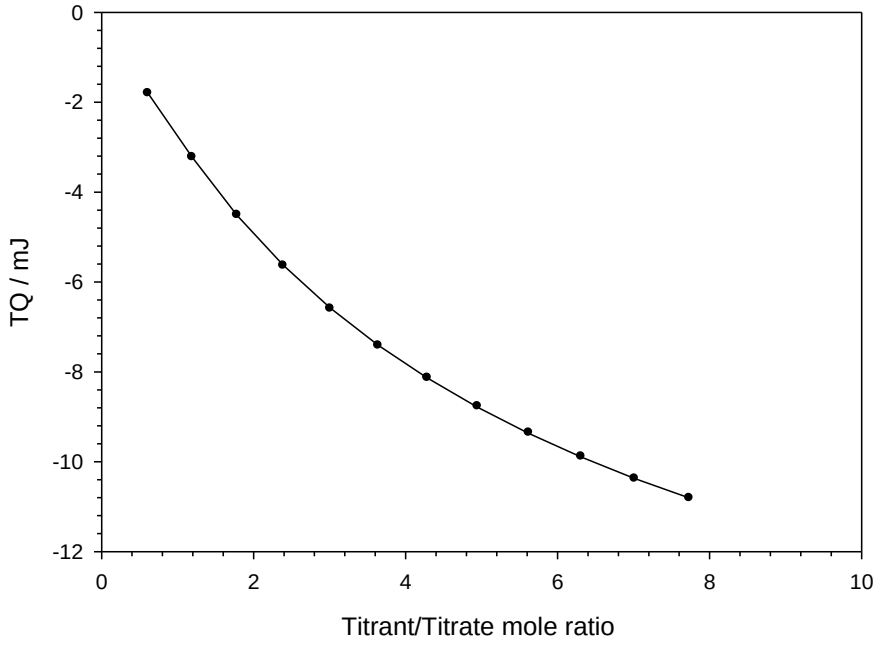
Results page

```
Reagent Reagent
  number   name
    1      R
    2      G

   sigma =   0.01205

 Formation constants   Value    relative    log    standard
                                std devn    beta   deviation
 Beta A refined     0.3066E+04  0.1560    3.4865    0.0677   1 1
 Beta B refined     0.5192E+08  0.5175    7.7153    0.2247   2 1
 Beta C constant    0.4421E+03            2.6455    0.0676   2 0

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

 Thermodynamic Functions, kJ/mol

          - DeltaG°          - DeltaH°          T DeltaS°
   A   19.9011  0.3866    42.4924  4.5528    -22.5913  4.4236
   B   44.0392  1.2829   -98.6733  7.2566    142.7125  8.2500
   C   15.1006  0.2656  -293.3921  16.2613   308.4927  16.2634
```

Titration Plots

Experimental (symbols) and calculated (cross and lines) heats





2.4.3 Structural studies

NMR methods: NMR experiments were performed at 500 MHz in D_2O at pD 11 at 298 K. The experiments on the complex were performed using an equimolar solution of **1** (R) and MeβGlcNAc$_2$ (G). Because of broad signals observed for the complex, due to a medium-exchange equilibrium in the NMR time scale, related to the strong binding, a concentration of 15 mM for both **1** and MeβGlcNAc$_2$ was used to obtain an acceptable signal to noise ratio for NOESY experiments. Alkaline pD was used to obtain sharper signal of the receptor at operative 15 mM concentration. In addition to standard 1D ^{1}H NMR spectra, COSY, TOCSY, HSQC and NOESY experiments (500 ms mixing time) were also acquired to assign the resonances of all the molecular entities and to detect the relevant intramolecular and intermolecular contacts.

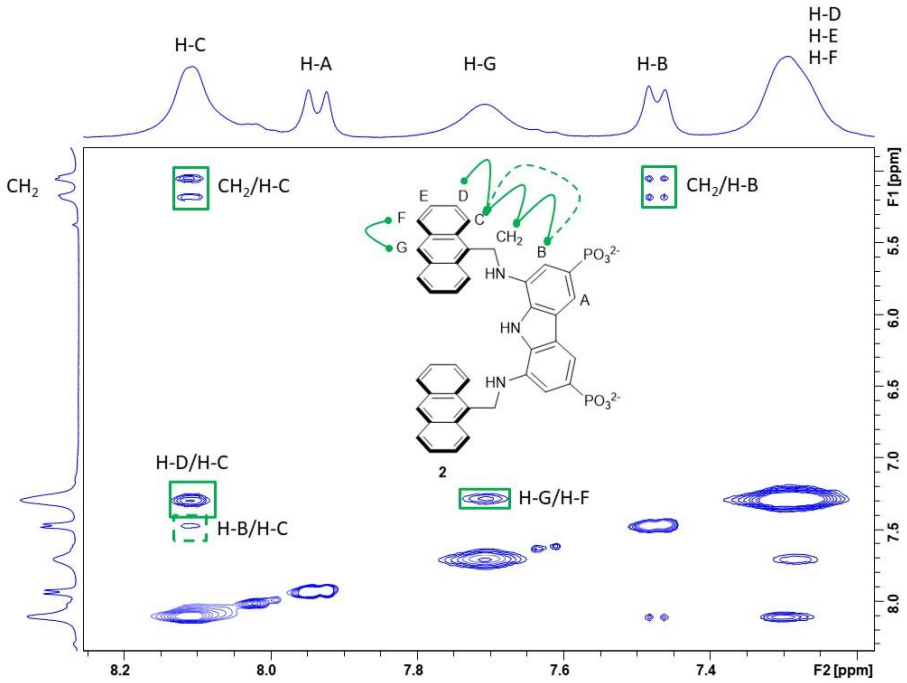


Figure S16. 500 MHz NOESY spectrum of an equimolar mixture of MeβGlcNAc$_2$ and **1** (15 mM each) in D_2O at 298 K. Intramolecular NOE cross peaks of the receptor **1** are indicated by squares and schematically represented (solid and dashed squares and lines were used to indicate strong and medium NOEs).

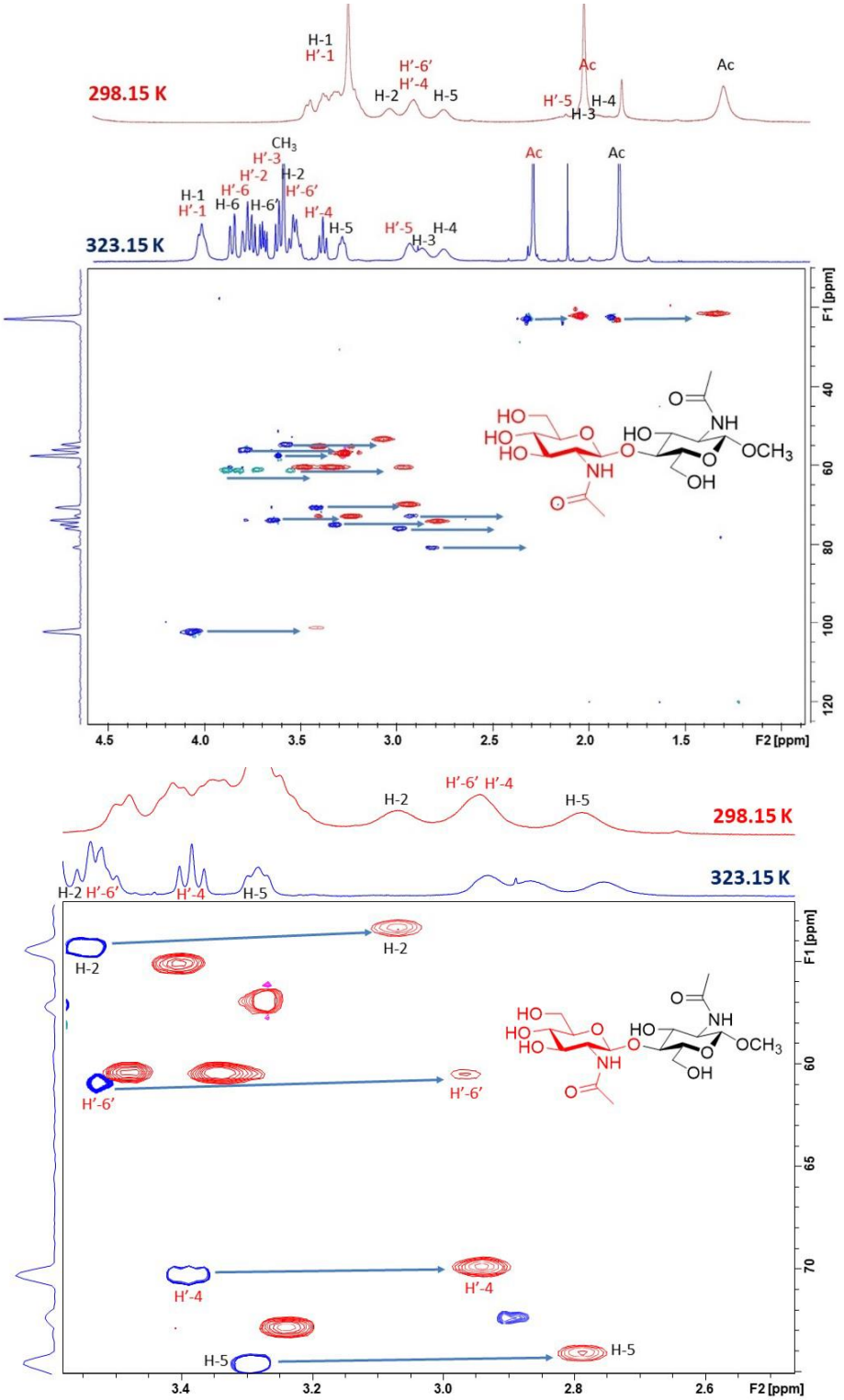


Figure S17. Superposition of 500 MHz HSQC and 1H-NMR spectra acquired at 298.15 K (in red) and 323.15 K (in blue) of an equimolar mixture of MeβGlcNAc$_2$ and **1** (15 mM each) in D_2O at 298 K. Temperature-dependent chemical shift variations are indicated by arrows and were used to assign proton signals of broad peaks at 298.15 K.

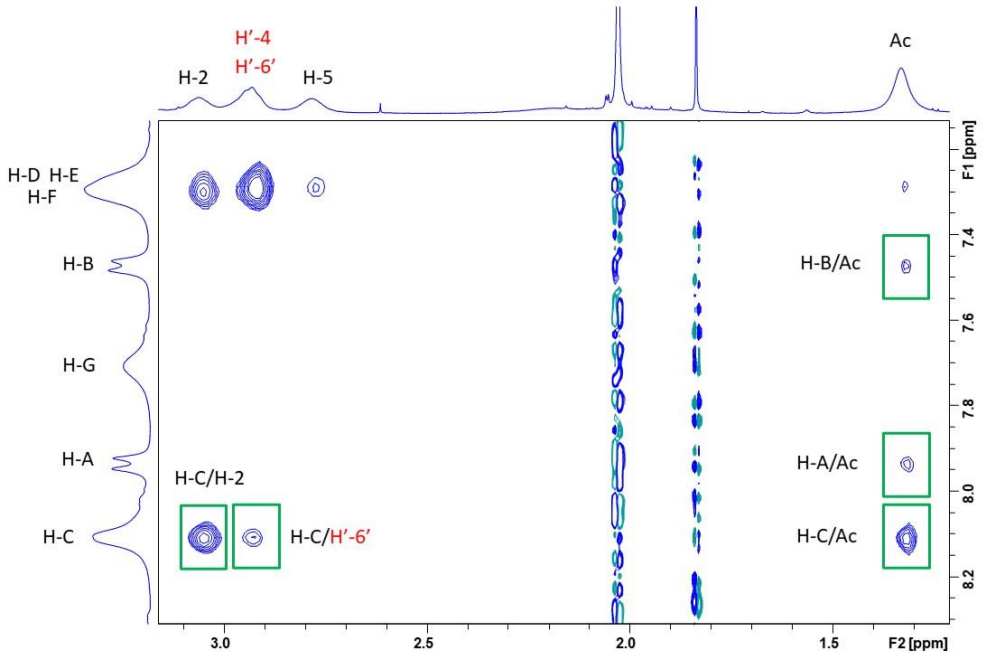


Figure S18. 500 MHz NOESY spectrum of an equimolar mixture of MeβGlcNAc2 and **1** (15 mM each) in D_2O at 298 K. Unambiguous intermolecular NOE cross peaks are indicated by squares.

Molecular modeling methods: Initial structures of MeβGlcNAc$_2$ and receptor **1** were built and minimized using conjugate gradients with the OPLS_2005 force field, water was set as solvent and an extended cutoff was used to treat remote interactions. Each phosphonate group of the receptor was considered monoprotonated, accordingly with the degree of protonation observed for a structurally correlated receptor at neutral pH.[11] A maximum number of 5000 iterations were employed with the Polak-Ribiere Conjugate Gradient (PRCG) scheme, until the convergence energy threshold was 0.05. Once the optimum geometries had been achieved, a conformational search protocol was adopted for the receptors, using a Monte Carlo torsional sampling method (MCMM) with automatic setup during the calculation, energy window of 21 kJ mol^{-1}, 1000 maximum number of steps, and 100 steps per torsion of the bond to be rotated. The best structures obtained from this calculation in terms of energy were chosen and then, the disaccharide was manually docked within the receptors cleft with different starting relative orientations and further minimized. Minimization results afford different structures which were employed as input for further conformational search protocols without any constraints. Several complexes were found to be stable, in which the sugar was located inside the receptor cleft. The lowest energy structures were analyzed to check the agreement with experimental NMR data. The protocol returned a family of structures, containing the minimum energy structure of the conformational search, in agreement with the observed NOE data.

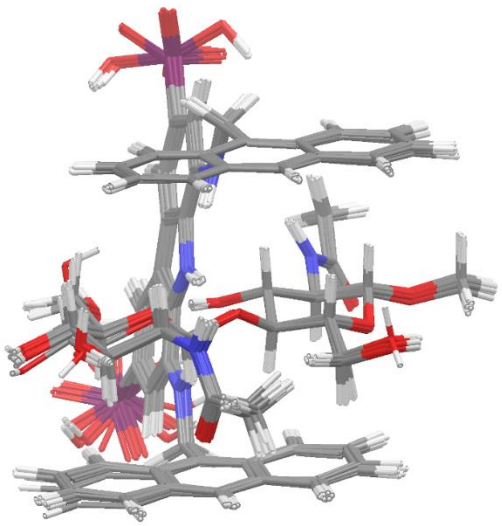


Figure S19. Molecular modelling results from conformational search for the complex of **1** with MeβGlcNAc$_2$. Superposition of the 16 energy minimum structures, within an energy window of 10.0 kJ mol^{-1}, identified among the 156 structures obtained from the calculation.

2.5 Bibliography

[1]Thompson A.J., de Vries R.P., Paulson J.C., *Curr. Opin. Virol*., **2019**, 34, 117-129.

[2]a) Balzarini J., *Nat. Rev. Microbiol.*, **2007**, 5, 583–597; b) Bagdonaite I., Wandall H.H., *Glycobiology*, **2018**, 28, 443-467; c) Keyaerts E., Vijgen L., Pannecouque C., Van Damme E., Peumans W., Egberink H., Balzarini J., Van Ranst M., *Antivir. Res*., **2007**, 75, 179-187; d) Mitchell C.A., Ramessar K., O'Keefe B.R., *Antivir. Res*., **2017**, 142, 37-54.

[3]a) Bravo M.F., Lema M.A., Marianski M., Braunschweig A.B., *Biochemistry*, 60, 999-1018; b) Tommasone S.; Allabush F., Tagger Y.K., Norman J., Köpf M., Tucker J.H.R., Mendes P.M., *Chem. Soc. Rev.*, **2019**, 48, 5488-5500.

[4]Francesconi O., Roelens S., *ChemBioChem*, **2019**, 20, 1329-1346.

[5]Stanley P., Schachter H., Taniguchi N, *Essentials of Glycobiology*, Chapter 8: *N*-Glycans. (Eds: A. Varki, R.D. Cummings, J.D. Esko, et al.) 2nd ed., Cold Spring Harbor, New York, USA, **2009**.

[6]Burrell C.J., Howard C.R., Murphy F.A., *Fenner and White's Medical Virology*, Chapter 4: Virus Replication. (Eds: C. J. Burrell, C. R. Howard, F. A. Murphy) 5th ed., pp 39-55, Academic Press, London, **2017**.

[7]a) Gordts S.C., Renders M., Férir G., Huskens D., Van Damme E.J.M., Peumans W., Balzarini J., Schols D., *J. Antimicrob. Chemother*., **2015**, 70, 1674-1685; b) Balzarini J., Van Laethem K., Hatse S., Foreyen M., Peumans W., Van Damme E., Schols D., *J. Biol. Chem*., **2005**, 49, 41005-41014.

[8]Davis A.P., *Chem. Soc. Rev*., **2020**, 49, 2531-2545.

[9]a) Kataev E.A. and Muller C., *Tetrahedron*, **2014**, 70, 137-167; b) Oshovsky G.V., Reinhoudt D.N., Verboom W., *Angew. Chem., Int. Ed.*, **2007**, 46, 2366-2393.

[10]Selected examples from recent literature: a) Zhang D.W., Martinez A. and Dutasta J.P., *Chem. Rev*., **2017**, 117, 4900-4942; b) Hayashi T., Ohishi Y., Abe H., Inouye M., *J. Org. Chem*. **2020**, 85, 1927-1934; c) Ohishi Y., Yamamoto N., Abe H., Inouye M., *J. Org. Chem*., **2018**, 83, 5766-5770; d) Yamashina M., Akita M., Hasegawa T., Hayashi S., Yoshizawa M., *Sci. Adv*., **2017**, 3, e1701126; e) Schaapkens X., Bobylev E.O., Reek J.H.H., Mooibroek T., *Org. Biomol. Chem*., **2020**, 18, 4734-4738.

[11]Francesconi O., Martinucci M., Badii L., Nativi C., Roelens S., *Chem. Eur. J.*, **2018**, 24, 6828-6836.

[12]Francesconi O., Cicero F., Nativi C., Roelens S., *ChemPhysChem*, **2020**, 21, 257-262.

[13]Francesconi O., Ienco A., Nativi C., Roelens S., *ChemPlusChem*, **2020**, 85, 1369-1373.

[14]Examples of adaptive architectures in carbohydrate recognition: a) Kim Y.H., Hong J.I., *Angew. Chem. Int. Ed.,* **2002**, 41, 2947-2950; b) Mazik M., *RSC Adv*., **2012**, 2, 2630-2642; c) Mateus P., Wicher P., Ferrand Y., Huc I., *Chem. Commun*., **2018**, 54, 5078-5081; d) Ohishi Y., Masuda K., Kudo K., Abe H., Inouye M., *Chem. Eur. J.*, **2021**, 27, 785-793.

[15]Vacca A., Francesconi O., Roelens S., *Chem. Rec.*, **2012**, 12, 544-566.

[16]Mooibroek T.J., Casas-Solvas J.M., Harniman R.L., Renney C.M., Carter T.S., Crump M.P., Davis A.P., *Nat. Chem*., **2016**, 8, 69-74.

[17]Asesio J.L., Cañada F.J., Bruix M., Rodriguez-Romero A., Jiménez-Barbero J., *Eur. J. Biochem.*, **1995**, 230, 621-633.

[18]Observed short intermolecular distances to anthracene planes: H-1, 2.64 Å; H-2, 2.86 Å; H-3, 2.69 Å; H-4, 2.81 Å; H'-1, 2.96 Å; H'-2, 2.77 Å; H'-5, 2.67 Å. Me (Ac) to Carbazole plane: 2.92 Å.

[19]Fernández-Alonso M.C., Díaz D., Berbis M.A., Marcelo F., Cañada F.J., Jiménez-Barbero J*., Curr. Protein Pept. Sc.*, **2012**,13,816-830.

[20]Vacca A., Nativi C., Cacciarini M., Pergoli R., Roelens S., *J. Am. Chem. Soc.*, **2004**, 126, 16456-16465.

[21]Frassineti C., Ghelli S., Gans P., Sabatini A., Moruzzi M.S., Vacca A., *Anal. Biochem*. **1995**, 231, 374-382.

[22]Gans P., Sabatini A., Vacca A., *J. Solution Chem*. **2008**, 37, 467-476.

Chapter 3

Molecular Recognition of Disaccharides in Water: Preorganized Macrocyclic or Adaptive Acyclic?

Adapted from:

Francesconi O., Milanesi F., Nativi C., Roelens S.,
Chem. Eur. J., 2021, *27*, 10456-10460

Abstract

When facing the dilemma of following a preorganized or adaptive design approach in conceiving the architecture of new biomimetic receptors for carbohydrates, shape-persistent macrocyclic structures were most often chosen to achieve effective recognition of neutral saccharides in water. In contrast, acyclic architectures have been seldom explored, even though potentially simpler and more easily accessible. In this work, comparison of the binding properties of two structurally related diaminocarbazolic receptors, featuring a macrocyclic and an acyclic tweezers-shaped architecture, highlighted the advantages provided by the acyclic receptor in terms of selectivity in the recognition of 1,4-disaccharides of biological interest. Selective recognition of GlcNAc2, the core fragment of *N*-glycans exposed on the surface of enveloped viruses, stands as an emblematic example. NMR spectroscopic data and molecular modeling calculations were used to ascertain the differences in binding mode and to shed light on the origin of recognition efficacy and selectivity.

3.1 Introduction

Among the plethora of biologically relevant oligosaccharides, those connected by a glycosidic 1,4-linkage are plentiful in nature. Lactose and maltose are two of the most common 1,4-disaccharides, whereas cellulose and chitin, constituted by repeating units of cellobiose (Glc_2, CeB) and of *N,N'*-diacetylchitobiose ($GlcNAc_2$), respectively, both connected by 1,4 glycosidic linkages, are among the most abundant biopolymers in nature.[1,2] Glycosidic 1,4-linkages are also very common in glycan structures. For example, the disaccharide $GlcNAc_2$ is a part of the $GlcNAc_2Man_3$ fragment, highly conserved in the core of *N*-glycans exposed on the surface of enveloped viruses, some of which are particularly hazardous for human health, including, among others, coronaviruses and retroviruses.[3,4]

Molecular recognition of disaccharides of biomedical relevance by biomimetic receptors in physiological media represents a major challenge of current research,[5,6] because selective recognition of neutral saccharides in water must cope with a highly competitive solvent.[7] Nevertheless, in the last few years significant steps forward have been made by developing biomimetic receptors based on rigid macrocyclic architectures.[8]

Although this approach has been quite successful for the recognition of several mono- and oligosaccharides, it is hampered by lengthy multistep syntheses of low overall yields, due to the critical macrocyclization step.[9] On the other hand, examples of effective recognition of neutral saccharides in water by acyclic receptors are extremely rare in the literature,[10,11] even though acyclic flexible architectures can take advantage of being more easily adaptable to the guest, while featuring simpler structures suitable for further optimization.

We have recently reported two biomimetic receptors (**1**[12] and **2**[ch.2,13] Figure 1) effectively recognizing carbohydrates in water. The two receptors share a common tridentate diaminocarbazole hydrogen binding motif, equipped with phosphonate hydrosolubilising groups, and two anthracene groups, providing extended CH-π interactions with the saccharidic backbone.[14] Receptor **1** features a preorganized macrocyclic structure possessing a hydrophobic cavity lined with H-bonding groups, whereas receptor **2** possesses a flexible, acyclic, tweezers-shaped architecture featuring analogous binding motifs. Receptor **1**, easily available in six steps with 30% overall yield, effectively binds monosaccharides in water, selectively recognizing the beta anomer of glucose with a 1.3 mM affinity (expressed as intrinsic median binding concentrations, BC_{50}^{0}), and the alpha anomers of glucose, galactose, and fucose with affinities of 3.12 mM, 1.19 mM and 360 μM, respectively.[12] Although extensively investigated toward monosaccharides, the binding properties of receptor **1** toward disaccharides were not yet explored. On the other hand, receptor **2**, which has been shown to effectively recognize 1,4-disaccharides, with a marked affinity (160 μM) and selectivity for the methyl β glycoside of $GlcNAc_2$,[13] did not bind to monosaccharides at all.

In order to assess the role of the architecture in saccharide recognition, in the present work we investigated the binding affinities of receptor **1** toward the set of glucose-containing disaccharides used to test receptor **2**, to compare their binding properties and ascertain the effect of macrocyclic (preorganized) vs. acyclic (adaptive) structures on recognition ability.[15] NMR-based molecular modelling calculations were used to give a three-dimensional description of the complexes of the two receptors with a common guest, which revealed the substantial role of CH-π interactions.

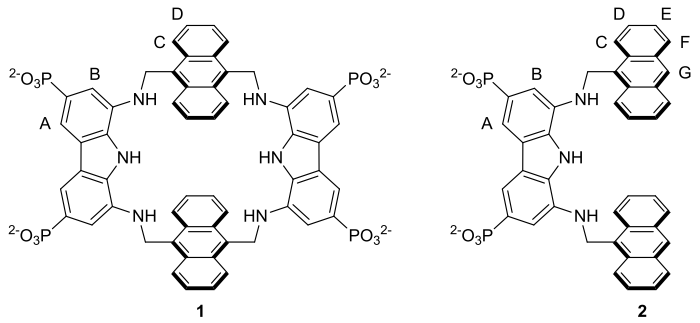


Figure 1. Structure of receptors **1** and **2** with proton labelling.

3.2 Results and discussion

In a preliminary screening by ^{1}H NMR spectroscopy, the binding ability of **1** was tested toward a set of disaccharides constituted by at least one glucose unit, for which the receptor showed good affinities, including cellobiose (CeB), lactose (Lac), maltose (Mal), trehalose (Tre), and sucrose (Suc) (**Figure 2**). Binding ability was qualitatively evaluated by monitoring the shifts of the proton signals of the sugar upon addition of an equimolar amount of **1**. Although for Suc and Tre no variations were observed, a marked upfield shift was detected for CeB, Mal and Lac, reasonably due to the shielding effect of the anthracene moieties in the binding cavity, larger for the β than for the α anomers. A concomitant broadening of signals, larger for the β anomers, indicated slow chemical exchange, most likely due to strong binding (**Figures S1-S3**).

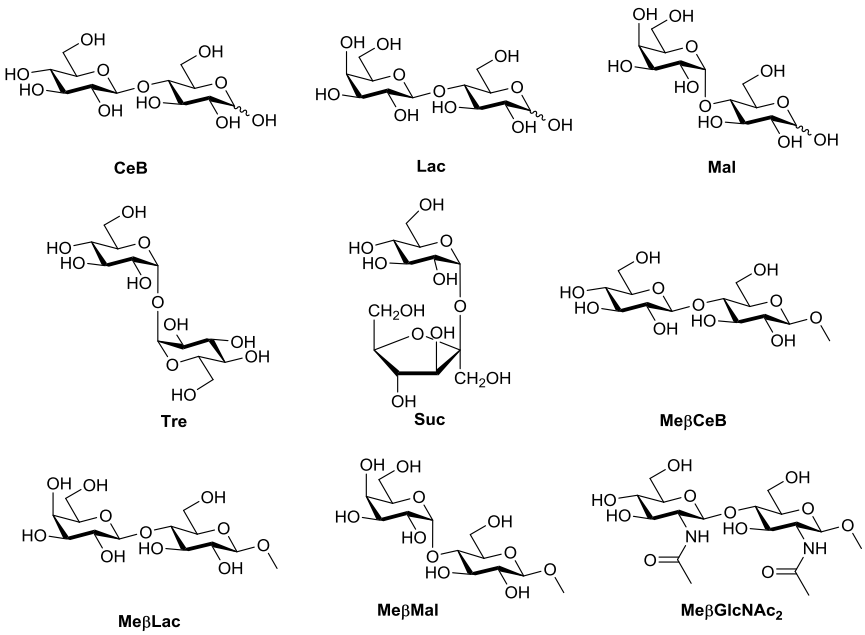


Figure 2. Structure of the investigated disaccharides and their abbreviations.

For a quantitative determination of the binding ability of **1**, ^{1}H-NMR titrations of methyl-β-glycosides of cellobiose (MeβCeB), lactose (MeβLac), and maltose (MeβMal) were carried

out in D_2O (pD 7.4) at 298 K, additionally including MeβGlcNAc$_2$, for which receptor **2** showed high affinity (**Figure 2**). To avoid ambiguities in the definition of the binding model, the cumulative association constants reported in **Table 1** were obtained by the simultaneous fit of all available signals from two independent titrations, run at different reactant concentrations. Because multiple complex species were found for all systems, the overall affinities reported in **Table 1** were determined by the intrinsic median binding concentration parameter (BC_{50}^{0}),[16] which was calculated from the measured binding constants. ^{1}H NMR titrations with MeβMal were also duplicated at pD 11 (**Table S1**) and fitted to the association model obtained at pD 7.4. While protonation of the aromatic amino groups is not expected in the investigated range of pD, the degree of protonation of the phosphonate groups does not affect the binding ability of receptor **1**, as previously observed for binding to monosaccharides[12] and confirmed by the comparable affinities obtained at different pD values.

As with monosaccharides, **Table 1** shows multiple binding constants for receptor **1** with disaccharides. Strong self-association, with a dimerization constant of $log\beta_{dim}$ = 3.84±0.20, favors complex species in which the receptor is dimeric, featuring two binding cavities and giving rise to complexes with stoichiometries higher than 1:1. Results show that receptor **1** effectively binds to MeβCeB, MeβMal, and MeβLac with good affinities, though with lack of selectivity, but does not recognize MeβGlcNAc$_2$, for which no significant variations of chemical shifts were detected (**Figure S8**). Thus, receptor **1** can distinguish 1-4 from 1-1' disaccharides (Suc/Tre), which are not bound at all, and MeβCeB from the *N*-acetylated amino-analogue MeβGlcNAc$_2$, but cannot discriminate among glucose containing 1-4 disaccharides, proving to be insensitive to the configuration of the anomeric linkage (MeβCeB/ MeβMal) and to the presence of axial substituents (MeβLac). Surprisingly, these 1-4 disaccharides are bound with an affinity very close to that previously observed for MeβGlc, indicating lack of selectivity between mono- and disaccharides.

Comparison of binding properties between receptors **1** and **2**, as obtained by ^{1}H NMR titrations, quantifies the selectivity advantage achieved through the adaptive architecture. Indeed, in contrast to **1**, receptor **2** not only strongly binds to MeβGlcNAc$_2$, but also discriminates among the investigated glucose containing 1,4-disaccharides. Receptor **2** shows preference for the all-equatorial MeβCeB, which is bound with an affinity very close to that observed for **1**, whereas MeβMal and MeβLac are bound with an affinity more than one order of magnitude smaller.

To shed light on the origin of such unexpected difference between **1** and **2**, a description of the binding mode characterizing the receptor-disaccharide complexes in solution was attempted by combining NMR techniques with molecular modelling calculations, following the approach previously adopted to study the complex between MeβGlcNAc$_2$ and receptor **2**, which provided an informative picture of the interaction.[13] Because MeβGlcNAc$_2$ did not bind to **1**, the investigation was carried out on MeβCeB, MeβMal and MeβLac.

Chemical shift variations of the anomeric protons H-1 and H'-1 upon formation of the 1:1 complex with receptor **1**, as calculated by non-linear regression analysis of titration data, showed an upfield shift for all three disaccharides caused by the aromatic shielding effect (**Figure 3a** and **Table S2**), which is more pronounced on the H proton of the methyl-β-glucoside unit (Δδ = 0.80-0.82) than on the H'-1 proton (Δδ = 0.24-0.32), suggesting a closer contact of the former to the aromatic moieties. On the other hand, from the analysis of chemical shift variations in the 1:1 complexes with receptor **2** (**Figure 3b** and **Table S3**), a stronger shielding effect is clearly apparent, more marked for MeβCeB, indicating a closer proximity of the entire disaccharide to the aromatic rings. This evidence suggests that **2** can adapt better than **1** to the disaccharidic guests, showing preference for the all-equatorial MeβCeB. The chemical shift differences (CSDs) for the H-1 and H'-1 protons of the latter

are very similar, suggesting a fit of the entire disaccharide into the cleft of the receptor. The CSDs of MeβMal and MeβLac, significantly smaller and with the H-1 CSD predominant, suggest a less comfortable fit into the cleft, in agreement with the corresponding lower affinities.

Receptor			1		2
Glycoside	R:G	$log\beta$	BC_{50}^0	$log\beta$	BC_{50}^0
MeβCeB	1:1	3.27±0.02	1.15±0.04	2.53±0.07	0.94±0.10
	1:2	4.92±0.03			
	2:1	6.81±0.02		6.33±0.06	
	2:2	8.91±0.06			
MeβMal	1:1	3.29±0.05	1.06±0.07	2.27±0.01	31.0±4.4
	1:2	4.77±0.05			
	2:1	6.82±0.04			
	2:2	9.21±0.09			
MeβLac	1:1	3.19±0.01	1.43±0.05	2.27±0.02	30.8±4.7
	1:2	4.42±0.02			
	2:1	6.22±0.04			
	2:2	8.04±0.13			
MeβGlcNAc$_2$	1:1	n.d.[d]		3.55±0.04	0.16±0.01
	2:1			7.35±0.09	

Table 1. Cumulative formation constants ($log\beta_n$)[a] and intrinsic median binding concentration (BC_{50}^0, mM)[b] for receptor to glycoside (R:G) complexes of **1** and 2 with methyl glycosides, measured at 298 K from NMR data in D_2O at pD 7.4.[c] [a] Formation constants were obtained by nonlinear least-square regression analysis of NMR data. [b] Calculated from the $log\beta$ values using the "BC_{50} Calculator" program.[16] [c] Receptor dimerization constants at pH 7.4 (**1**: $log\beta_{dim}$ = 3.84±0.20; **2**: $log\beta_{dim}$ = 2.65±0.07) were set invariant in the nonlinear regression analysis of NMR data. [d] not detectable.

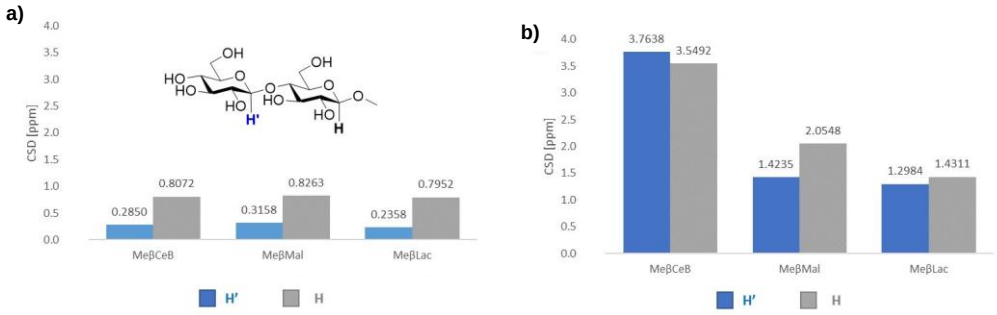


Figure 3. Plot of the chemical shift differences (CSD, ppm) between free and bound state of the anomeric H and H' protons for MeβCeB (shown with proton labelling), MeβMal and MeβLac when bound to **1** (a) and to **2** (b) in the 1:1 complexes in D_2O (pD 7.4) at T=298 K.

The complexes of receptor **1** and **2** with MeβCeB were then selected as representative examples, and their binding modes were studied by NOESY spectroscopy at pD 11, a medium in which the receptors are fully deprotonated species. From NOESY spectra run on the 1:1 mixture of **1** and MeβCeB, a strong intramolecular NOE contact was found between the H'-1 and the H-4 protons (**Figure S10**), suggesting that in the complex the disaccharide is in the conformation usually found in solution. Unambiguous intermolecular NOE contacts were also identified (**Figure S11, Figure S12**), the strongest of which were those between the OCH_3 protons and both the H-C and H-D protons of the anthracene ring (**Figure 1**), and between the H'-1/H'-5 protons and the H-D protons.

NOESY spectra performed on an equimolar mixture of **2** and MeβCeB showed unambiguous intermolecular NOE contacts between both saccharidic units of MeβCeB and the anthracene protons of **2** (**Figure S15**). The NOESY map shows a strong NOE cross peak between H-2 and H-C, and a NOE contact of H'-2 with the H-F located on the opposite side of the anthracene ring. Moreover, the OCH_3 protons show NOE contacts with the H-C, H-D and H-E protons.

Based on NOESY NMR evidence, molecular mechanics calculations were carried out on the 1:1 complex of **1** with MeβCeB, on the assumption that, although prevalently dimeric, the receptor would feature two independent binding sites. A conformational search, using a well-tested unconstrained molecular mechanics protocol,[17] returned a family of conformers within 5.19 kJ mol^{-1} from the global minimum that was in very good agreement with NMR spectroscopic data. The minimum energy structure depicted in **Figure 4** (**a** and **b**) shows MeβCeB partially located inside the receptor cavity, with the methyl glycoside unit nested inside the cavity and the other unit protruding outward, in a geometry that agrees with the strongest NOE contacts observed in NOESY maps (**Table S4**) and with the shift differences observed from titration experiments (**Figure 4a**). All O···H interatomic distances shorter than the sum of the van der Waals radii and compliant with hydrogen bonding criteria were calculated from the above model, and several hydrogen-bonding interactions were found involving the methyl glucoside unit exclusively (**Figure 4b**). Additional contribution to binding is provided by several CH-π interactions showing short distances, established between the methyl glucoside unit and the anthracene rings (**Table S6**).

The binding geometry obtained from calculations supports the observed affinities. Indeed, because the disaccharide is bound through the methyl glycoside unit exclusively, lack of selectivity among the investigated set of disaccharides can be easily anticipated, irrespective of the α/β glycosidic linkage to the second unit. This evidence also explains the closely similar affinities observed between the disaccharides and the monosaccharide MeβGlc. Likewise, the 1-1' disaccharides, featuring a bulky substituent in place of the methyl group, and $MeβGlcNAc_2$ featuring the *N*-acetyl groups, can hardly fit into the receptor cavity. Thus, despite the good affinities observed, lack of selectivity between glucose containing 1,4-disaccharides can be ascribed to the size of the macrocyclic cavity, unable to accommodate the entire disaccharide.

The conformational search carried out on the 1:1 complex between **2** and MeβCeB resulted in a single family of minimum energy conformers within 8.73 kJ mol^{-1} from the global minimum. The minimum energy structure depicted in Figure 4c shows the MeβCeB entirely located inside the binding cleft between the two anthracene faces, in a geometry closely similar to that previously observed in the complex with $MeβGlcNAc_2$,[13] and in agreement with the proximities inferred by strong NOE contacts (**Table S5**).

Hydrogen bonding interactions could be calculated from the above model (**Figure 4d**) and, analogously to **1**, four hydrogen bonds were found between the diaminocarbazole unit and MeβCeB. However, in contrast to **1**, a significant enhancement to binding could result

from the extensive network CH-π interactions that can be established between the axial protons of both the saccharidic units and the anthracenes (Table S6).

The above three-dimensional descriptions clearly show that the acyclic structure of **2** can adapt to the disaccharidic guest better than the macrocyclic structure of **1**, giving rise to increased affinity despite the lack of a hydrogen-bonding unit. Evidence indicates that the latter is effectively compensated by a tighter fit and by extensive CH-π interactions. Such a compensation is not fully achieved with MeβMal and MeβLac because axial substituents hamper a tight fit into the cleft, causing an affinity drop. On the contrary, lack of preorganization and absence of a hydrogen-bonding unit cause a severe drop of affinity of **2** for monosaccharides, not compensated by additional interactions, which results in undetectable binding. Thus, the macrocyclic receptor **1** appears to be well preorganized for binding a monosaccharidic but not a disaccharidic guest, whereas the acyclic receptor **2** can take advantage of its adaptive structure to establish more extensive attractive interactions with respect to its macrocyclic counterpart.

Eventually, the enhanced binding of **2** to MeβGlcNAc$_2$ compared to MeβCeB can be explained by the additional hydrogen bonding and CH-π interactions involving the *N*-acetyl group that the former can establish with the receptor.

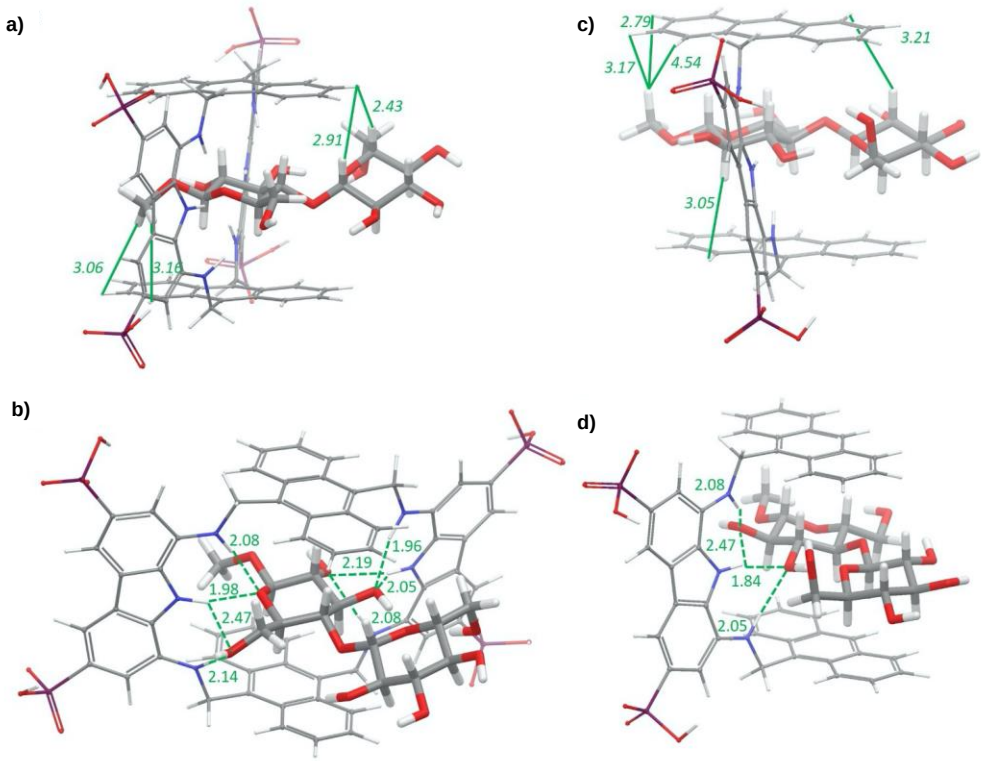


Figure 4. Global minimum structures of the **1**•MeβCeB, (a) and (b), and **2**•MeβCeB, (c) and (d), complexes in two different projections. The strongest intermolecular NOEs found between **1** and MeβCeB (a) and **2** and MeβCeB (c) are indicated as solid lines, with the corresponding distances [Å] calculated for the lowest energy conformer. Intermolecular hydrogen-bonding interactions found in the calculated structures are indicated as dashed lines in (b) and (d), together with the corresponding oxygen/hydrogen distances [Å].

3.3 Conclusions

Altogether, the results presented demonstrate that a flexible acyclic structure can be an effective alternative to the widely studied macrocyclic architectures for molecular recognition of neutral disaccharides in water, provided that a suitably designed combination of hydrogen bonding and CH-π interactions can be established with the saccharidic guest. The structurally simple tweezers-shaped receptor **2** presents significant advantages over its macrocyclic counterpart **1**, accommodating the disaccharidic guest within the binding cleft and selectively recognizing the methyl-β-glycoside of $GlcNAc_2$ over a set of monosaccharides and structurally related 1-4 disaccharides. Because of the simple structure, the easy synthetic availability, and the potential structural modifications, the tweezers-shaped architecture of receptor **2** opens the way to the design of acyclic receptors for the recognition of saccharides in water.

3.4 Supporting information

3.4.1 Binding studies

NMR preliminary screening. Preliminary screenings (298 K, 500 MHz) were performed in D_2O at pD 11 in presence of DSS as internal reference. Solution of reducing carbohydrates were prepared in D_2O and kept overnight at room temperature before the screening experiments, to ensure equilibration of the anomers. The spectra of the free sugars at 1 mM concentration were compared to the spectra of the equimolar mixture of sugars with receptor **1** (1 mM each) and chemical shift differences were evaluated.

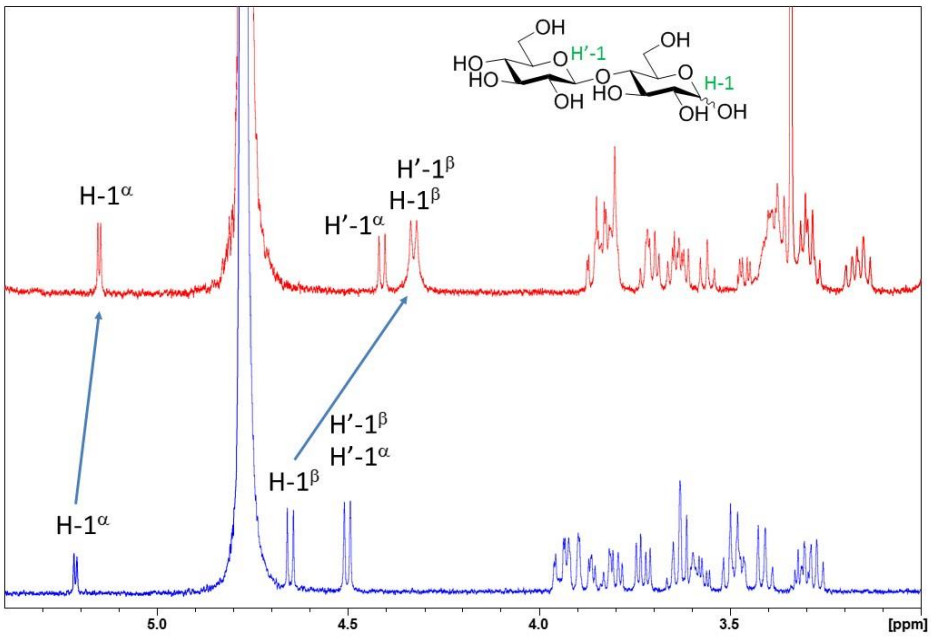


Figure S1. ^{1}H NMR spectra (500 MHz, D_2O) of a 1 mM solution of CeB (bottom) and of an equimolar mixture of CeB and **1** (top, 1 mM each) at pD 11. Variations of the H-1 proton signals are $\Delta\delta$ = 0.06 ppm for the α and $\Delta\delta$ = 0.32 ppm for the β anomer.

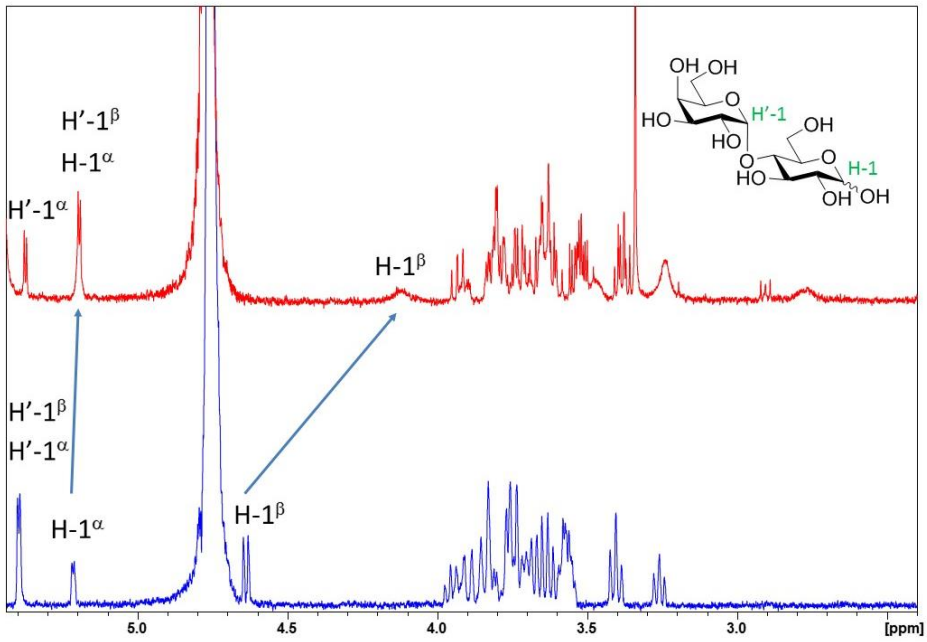


Figure S2. ^{1}H NMR spectra (500 MHz, D_2O) of a 1 mM solution of Mal (bottom) and of an equimolar mixture of Mal and **1** (top, 1 mM each) at pD 11. Variations of the H-1 proton signals are Δδ = 0.02 ppm for the α and Δδ = 0.52 ppm for the β anomer.

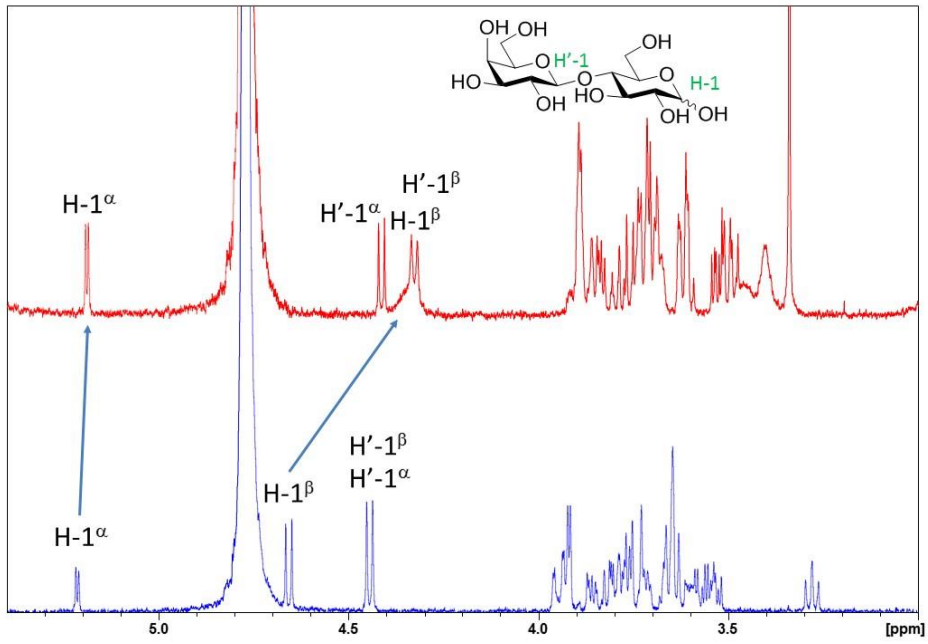


Figure S3. ^{1}H NMR spectra (500 MHz, D_2O) of a 1 mM solution of Lac (bottom) and of an equimolar mixture of Lac and **1** (top, 1 mM each) at pD 11. Variations of the H-1 proton signals are Δδ = 0.03 ppm for the α and Δδ = 0.31 ppm for the β anomer.

NMR titrations and data analysis: Titrations were performed at 298 K, 500 MHz in 5 mm NMR tubes using microsyringes, following a previously described technique.[18] Concentration of the receptor was maintained constant during the titrations with glycosides to avoid changes in ionic strength. The stock solutions of **1** was prepared in D_2O adjusting the pD with a diluted NaOH solution in D_2O. A correction factor of + 0.4 was applied to the pH values measured by the pH meter to determine the pD values (pD = pH + 0.4). The alkaline stock solution of **1** was stored during the titrations under nitrogen atmosphere to avoid acid/base reactions with atmospheric CO_2. Following this strategy, constant values of pD were observed during titrations and dilution experiments. DSS was used as internal reference. Dimerization constants of receptor **1** at pD 7.4 and at PD 11, were set invariant in the non-linear regression analysis of receptor-glycosides binding data measured at pD 7.4 and pD 11, respectively. To avoid any ambiguities in the definition of the equilibrium model with receptor **1**, independent titrations were performed at significantly different receptor concentrations and all set of data were simultaneously fitted through a nonlinear least-square regression analysis, including in the fit all the available signals from both reactants at pD 11 and from the glycoside only at pD 7.4. Mathematical analysis of data and graphic presentation of results was performed using the HypNMR 2006 program.[19] Results pages and Plots of experimental and calculated shifts are reported hereafter.

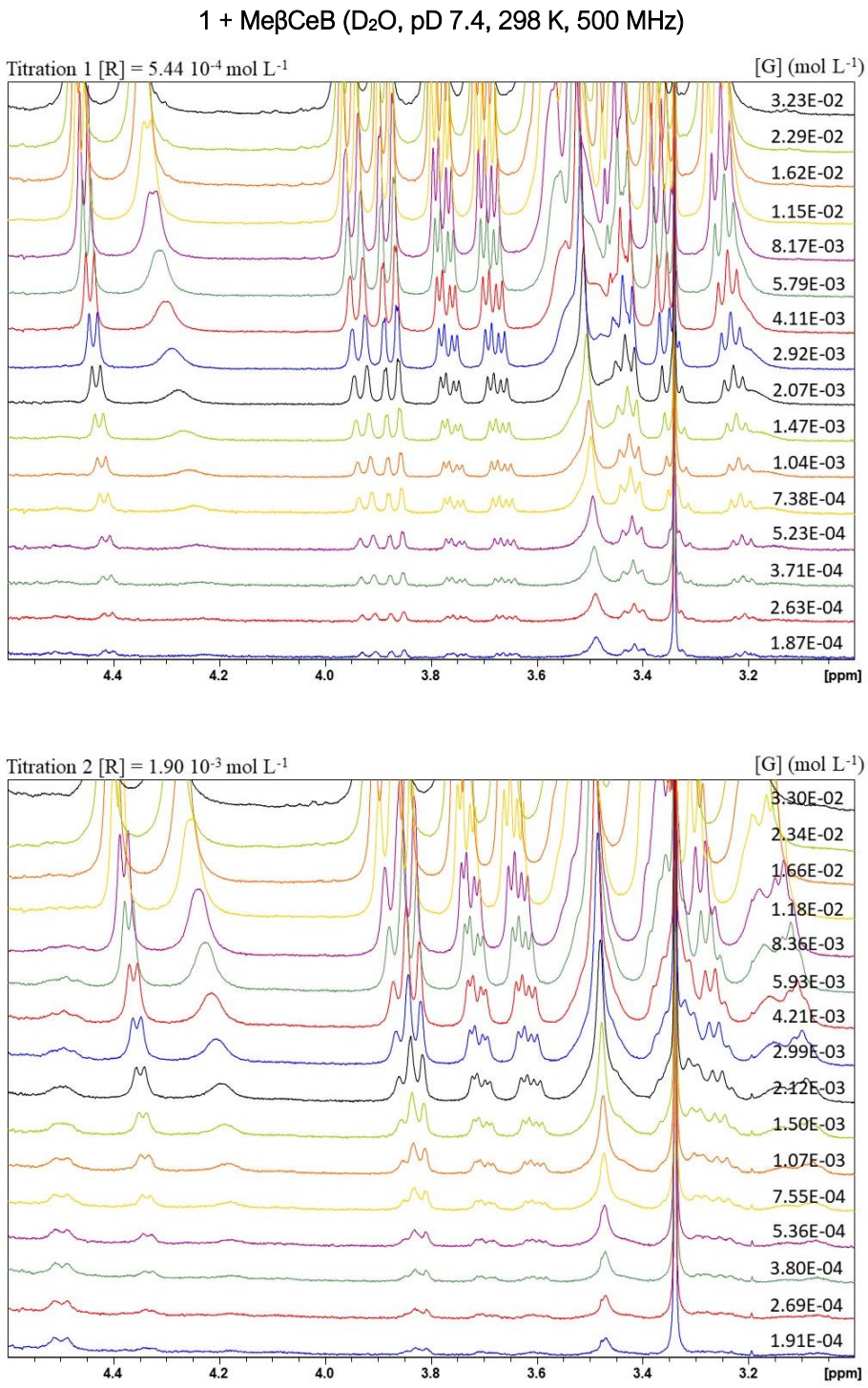


Figure S4. ^{1}H NMR spectroscopic titrations (500 MHz, D_2O, pD 7.4, 298 K) of receptor **1** (R) with incremental concentrations of MeβCeB (G).

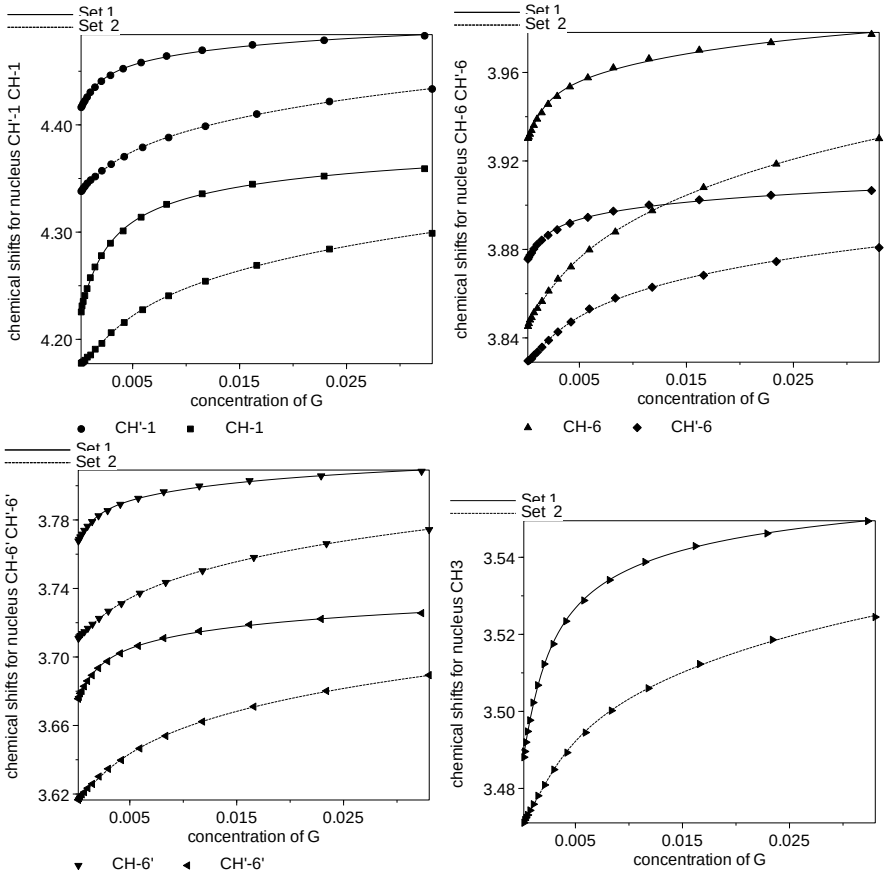
Results page

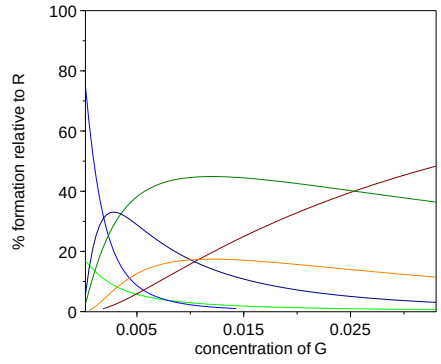
```
 no. of spectra            32
 no. of resonance values  224
 no. of resonant nuclei     7

   sigma =   0.00057332333          RMS weighted residual =   0.00052102849

       stoich            value   relative    log    standard
       coeff                     std devn    beta   deviation

 Beta  0  2 constant 6.9231E+003            3.8403           (   R2   )
 Beta  2  1 refined  8.2635E+004  0.0659    4.9172   0.0286  (   G2R  )
 Beta  1  1 refined  1.8808E+003  0.0452    3.2743   0.0196  (   GR   )
 Beta  1  2 refined  6.5153E+006  0.0514    6.8139   0.0223  (   GR2  )
 Beta  2  2 refined  8.0522E+008  0.1470    8.9059   0.0638  (   G2R2 )

Individual chemical shifts
                    G                          R
 ============================================================
        +        value    error          value    error
CH'-1   +       4.5078   0.0006
CH-1    +       4.3897   0.0008
CH-6    +       4.0000   0.0007
CH'-6   +       3.9185   0.0005
CH-6'   +       3.8252   0.0006
CH'-6'  +       3.7431   0.0006
CH3     +       3.5620   0.0006
                  0,2                        2,1
 ============================================================
        +        value    error          value    error
CH'-1   +                                3.3473   0.0206
CH-1    +                                3.0539   0.0263
CH-6    +                                2.8858   0.0195
CH'-6   +                                3.3329   0.0135
CH-6'   +                                3.0258   0.0158
CH'-6'  +                                2.9043   0.0163
CH3     +                                3.0191   0.0148
                  1,1                        1,2
 ============================================================
        +        value    error          value    error
CH'-1   +       4.2228   0.0122          4.2036   0.0097
CH-1    +       3.5825   0.0336          4.3950   0.0099
CH-6    +       3.8675   0.0068          3.6140   0.0132
CH'-6   +       3.8206   0.0055          3.7115   0.0073
CH-6'   +       3.6775   0.0071          3.5818   0.0083
CH'-6'  +       3.5388   0.0091          3.5103   0.0076
CH3     +       3.1740   0.0166          3.6015   0.0062
                  2,2
 ================================
        +        value    error
CH'-1   +       4.0410   0.0426
CH-1    +       4.5829   0.0468
CH-6    +       3.3057   0.0679
CH'-6   +       3.5649   0.0349
CH-6'   +       3.4060   0.0408
CH'-6'  +       3.3391   0.0381
CH3     +       3.7465   0.0328

 Correlation coefficients*1000
       1    2    3    4
   1
   2  775
   3  152 -388
```

```
   4  762  328  525

 Parameters are numbered as follows
    1 beta 2,1
    2 beta 1,1
    3 beta 1,2
    4 beta 2,2
```

Titration Plots

Chemical shifts (δ, ppm) vs. concentration of G (mol L^{-1})
experimental (symbols) and calculated (lines) values





R R_2 GR G_2R GR_2 G_2R_2

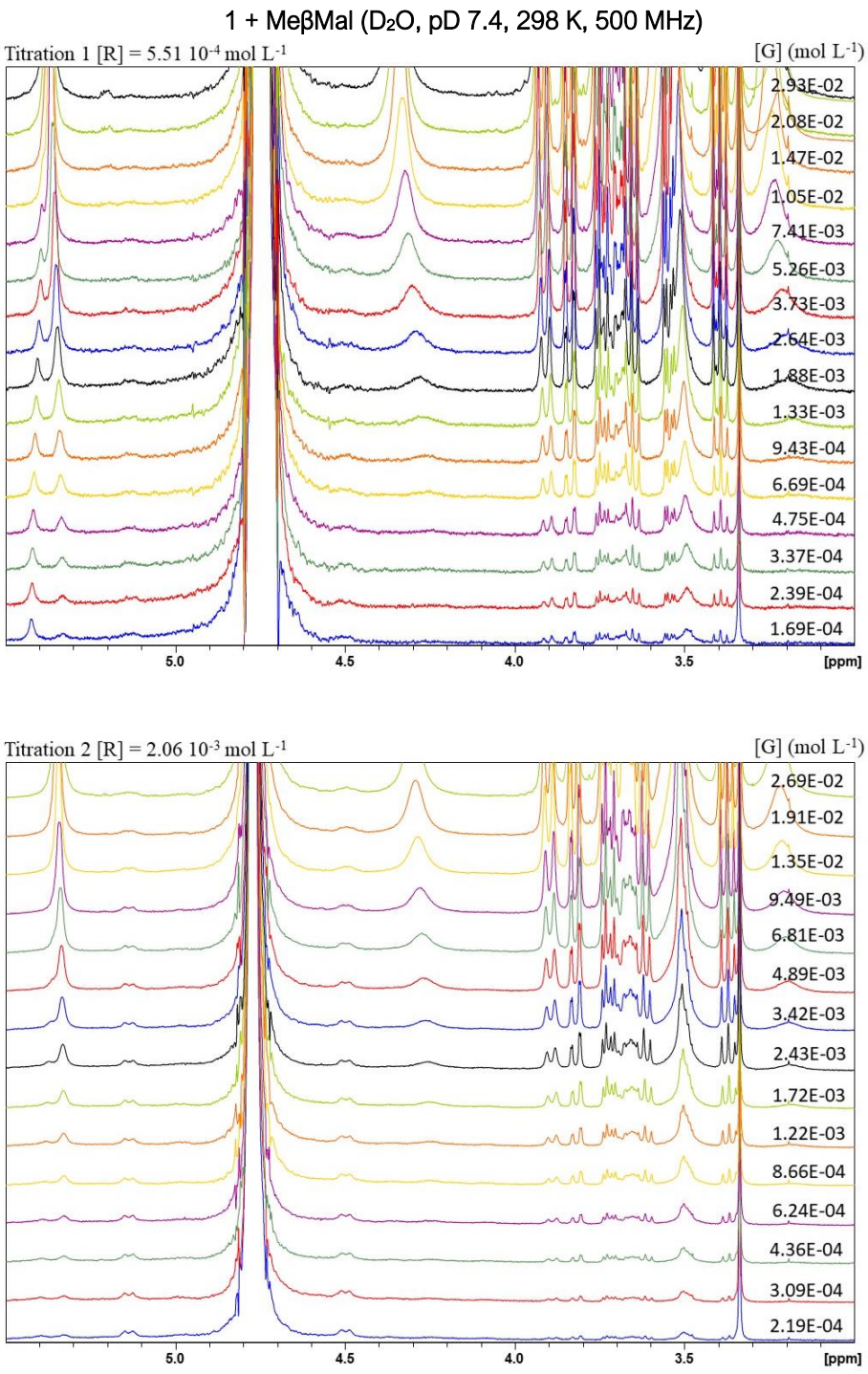


Figure S5. ^{1}H NMR spectroscopic titrations (500 MHz, D_2O, pD 7.4, 298 K) of receptor **1** (R) with incremental concentrations of MeβMal (G).

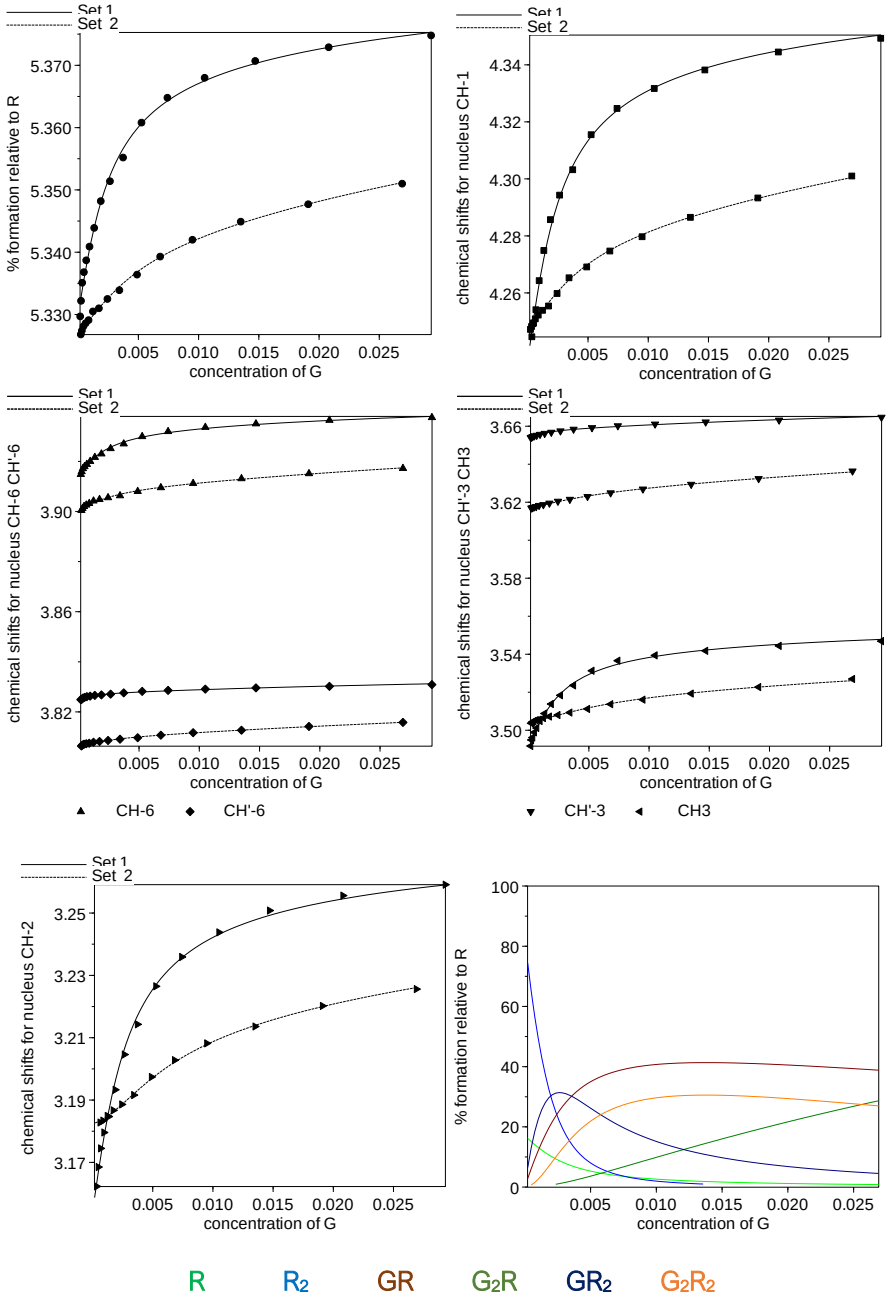
Results page

```
 no. of spectra            31
 no. of resonance values  210
 no. of resonant nuclei     7

   sigma =   0.00083497849          RMS weighted residual =   0.00075346606

       stoich            value   relative    log    standard
       coeff                     std devn    beta   deviation

 Beta  0  2 constant 6.9231E+003            3.8403           (   R2   )
 Beta  1  1 refined  1.9372E+003  0.1056    3.2872   0.0459  (   GR   )
 Beta  2  1 refined  5.8809E+004  0.1241    4.7694   0.0539  (   G2R  )
 Beta  1  2 refined  6.6479E+006  0.0851    6.8227   0.0370  (   GR2  )
 Beta  2  2 refined  1.6284E+009  0.2169    9.2118   0.0942  (   G2R2 )

Individual chemical shifts
                    G                          R
 ============================================================
        +        value    error          value    error
CH'-1   +       5.3883   0.0012
CH-1    +       4.3786   0.0021
CH-6    +       3.9485   0.0011
CH'-6   +       3.8386   0.0009
CH'-3   +       3.6789   0.0014
CH3     +       3.5604   0.0011
CH-2    +       3.2783   0.0016
                  0,2                        1,1
 ============================================================
        +        value    error          value    error
CH'-1   +                                5.0725   0.0325
CH-1    +                                3.5523   0.0835
CH-6    +                                3.8024   0.0166
CH'-6   +                                3.8227   0.0062
CH'-3   +                                3.6586   0.0075
CH3     +                                3.1328   0.0429
CH-2    +                                2.4938   0.0782
                  2,1                        1,2
 ============================================================
        +        value    error          value    error
CH'-1   +       4.6967   0.0758          5.4627   0.0103
CH-1    +       2.9497   0.1530          4.6601   0.0289
CH-6    +       3.3377   0.0670          3.9243   0.0053
CH'-6   +       3.3752   0.0524          3.7540   0.0075
CH'-3   +       2.8109   0.0915          3.5066   0.0132
CH3     +       2.9507   0.0690          3.7434   0.0180
CH-2    +       2.4045   0.0983          3.6365   0.0324
                  2,2
 ================================
        +        value    error
CH'-1   +       5.5079   0.0327
CH-1    +       4.8008   0.0997
CH-6    +       3.9535   0.0117
CH'-6   +       3.7510   0.0172
CH'-3   +       3.4855   0.0336
CH3     +       3.7978   0.0551
CH-2    +       3.7073   0.0972

 Correlation coefficients*1000
       1    2    3    4
   1
   2  421
   3 -529  320
```

```
  4 -739   -5  619
Parameters are numbered as follows
   1 beta 1,1
   2 beta 2,1
   3 beta 1,2
   4 beta 2,2
```

Titration Plots

Chemical shifts (δ, ppm) vs. concentration of G (mol L^{-1})
experimental (symbols) and calculated (lines) values



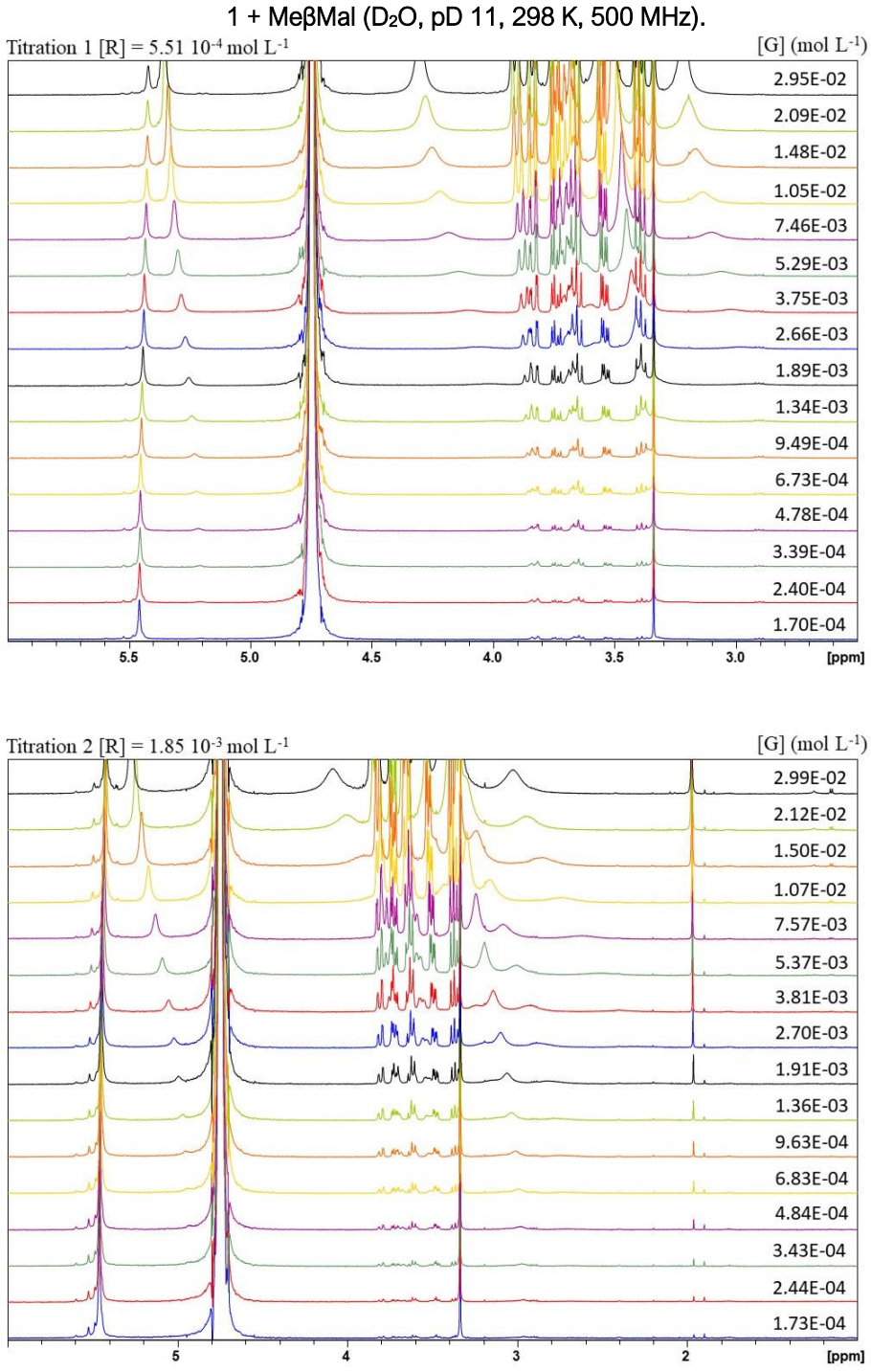


Figure S6. ^{1}H NMR spectroscopic titrations (500 MHz, D_2O, pD 11, 298 K) of receptor **1** (R) with incremental concentrations of MeβMal (G).

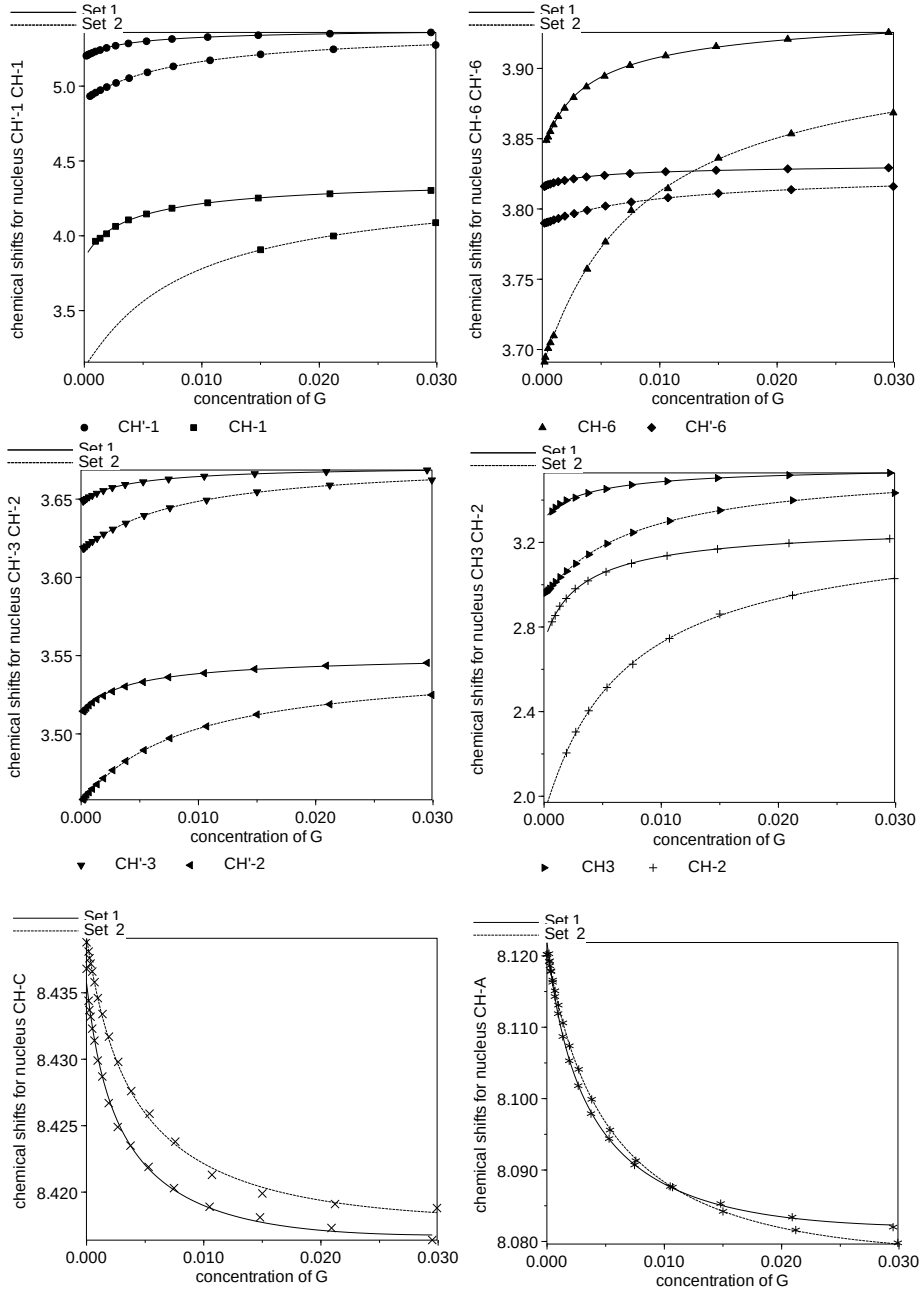
Results page

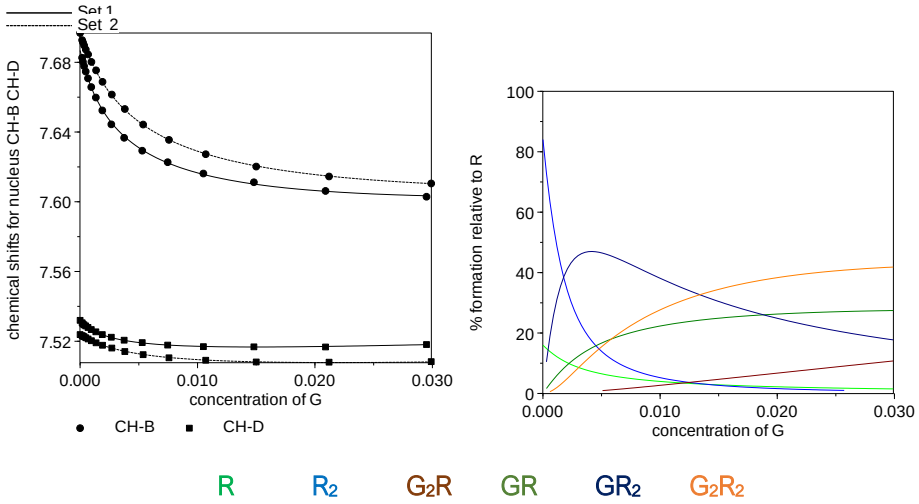
```
 no. of spectra            34
 no. of resonance values  348
 no. of resonant nuclei    12

   sigma =   0.00146502824          RMS weighted residual =   0.00131412110

       stoich            value   relative    log    standard
       coeff                     std devn    beta   deviation

 Beta  0  2 constant 8.9084E+003            3.9498           (   R2   )
 Beta  2  1 refined  9.0798E+003  0.8722    3.9581   0.3788  (   G2R  )
 Beta  1  1 refined  6.5012E+002  0.1017    2.8130   0.0442  (   GR   )
 Beta  1  2 refined  7.5419E+006  0.0851    6.8775   0.0370  (   GR2  )
 Beta  2  2 refined  6.3443E+008  0.1676    8.8024   0.0728  (   G2R2 )

Individual chemical shifts

                    G                         R
 ============================================================
        +        value    error          value    error
CH'-1   +       5.3908   0.0035
CH-1    +       4.3951   0.0074
CH-6    +       3.9551   0.0026
CH'-6   +       3.8373   0.0019
CH'-3   +       3.6693   0.0019
CH'-2   +       3.5553   0.0019
CH3     +       3.5598   0.0042
CH-2    +       3.2794   0.0076
CH-C    +                                8.4123   0.0062
CH-A    +                                8.1310   0.0062
CH-B    +                                7.6202   0.0067
CH-D    +                                7.5865   0.0063
                   0,2                       2,1
 ============================================================
        +        value    error          value    error
CH'-1   +                                3.4381   0.5478
CH-1    +                               -0.7745   1.2849
CH-6    +                                1.1418   1.6634
CH'-6   +                                2.9332   0.6466
CH'-3   +                                3.8785   0.3460
CH'-2   +                                2.6531   0.5380
CH3     +                                2.4954   0.6445
CH-2    +                                1.3365   1.3485
CH-C    +       8.4442   0.0018          8.4276   0.0204
CH-A    +       8.1186   0.0018          8.0955   0.0264
CH-B    +       7.7109   0.0019          7.6310   0.0367
CH-D    +       7.5120   0.0018          7.5411   0.0335
                   1,1                       1,2
 ============================================================
        +        value    error          value    error
CH'-1   +       5.1468   0.0613          4.1173   0.0632
CH-1    +       3.2900   0.2378          1.1400   0.1865
CH-6    +       3.6697   0.0528          3.2897   0.0351
CH'-6   +       3.7560   0.0409          3.7259   0.0114
CH'-3   +       3.6359   0.0401          3.5337   0.0115
CH'-2   +       3.4431   0.0412          3.3117   0.0152
CH3     +       3.2921   0.0702          1.9282   0.0826
CH-2    +       2.9469   0.1411         -0.4437   0.1868
CH-C    +       8.4075   0.0059          8.4318   0.0026
CH-A    +       8.0830   0.0061          8.1019   0.0027
CH-B    +       7.5704   0.0105          7.6715   0.0029
CH-D    +       7.5361   0.0063          7.5047   0.0026
```

```
                    2,2
 ================================
        +        value    error
CH'-1   +       2.4223   0.2188
CH-1    +      -3.4904   0.5299
CH-6    +       2.4048   0.1251
CH'-6   +       3.5691   0.0390
CH'-3   +       3.3451   0.0389
CH'-2   +       2.9732   0.0544
CH3     +      -0.1605   0.2732
CH-2    +      -4.3302   0.5151
CH-C    +       8.4174   0.0044
CH-A    +       8.0614   0.0047
CH-B    +       7.6033   0.0053
CH-D    +       7.4799   0.0055

 Correlation coefficients*1000

      1    2    3    4
   1
   2  724
   3  -72 -553
   4   90 -450  957

 Parameters are numbered as follows
    1 beta 2,1
    2 beta 1,1
    3 beta 1,2
    4 beta 2,2
```

Titration Plots

Chemical shifts (δ, ppm) vs. concentration of G (mol L^{-1})
experimental (symbols) and calculated (lines) values




Set 1
Set 2
chemical shifts for nucleus CH-B CH-D
7.68
7.64
7.60
7.56
7.52
0.000
0.010
0.020
0.030
concentration of G
CH-B
CH-D
% formation relative to R
100
80
60
40
20
0
0.000
0.010
0.020
0.030
concentration of G
R
R2
G2R
GR
GR2
G2R2

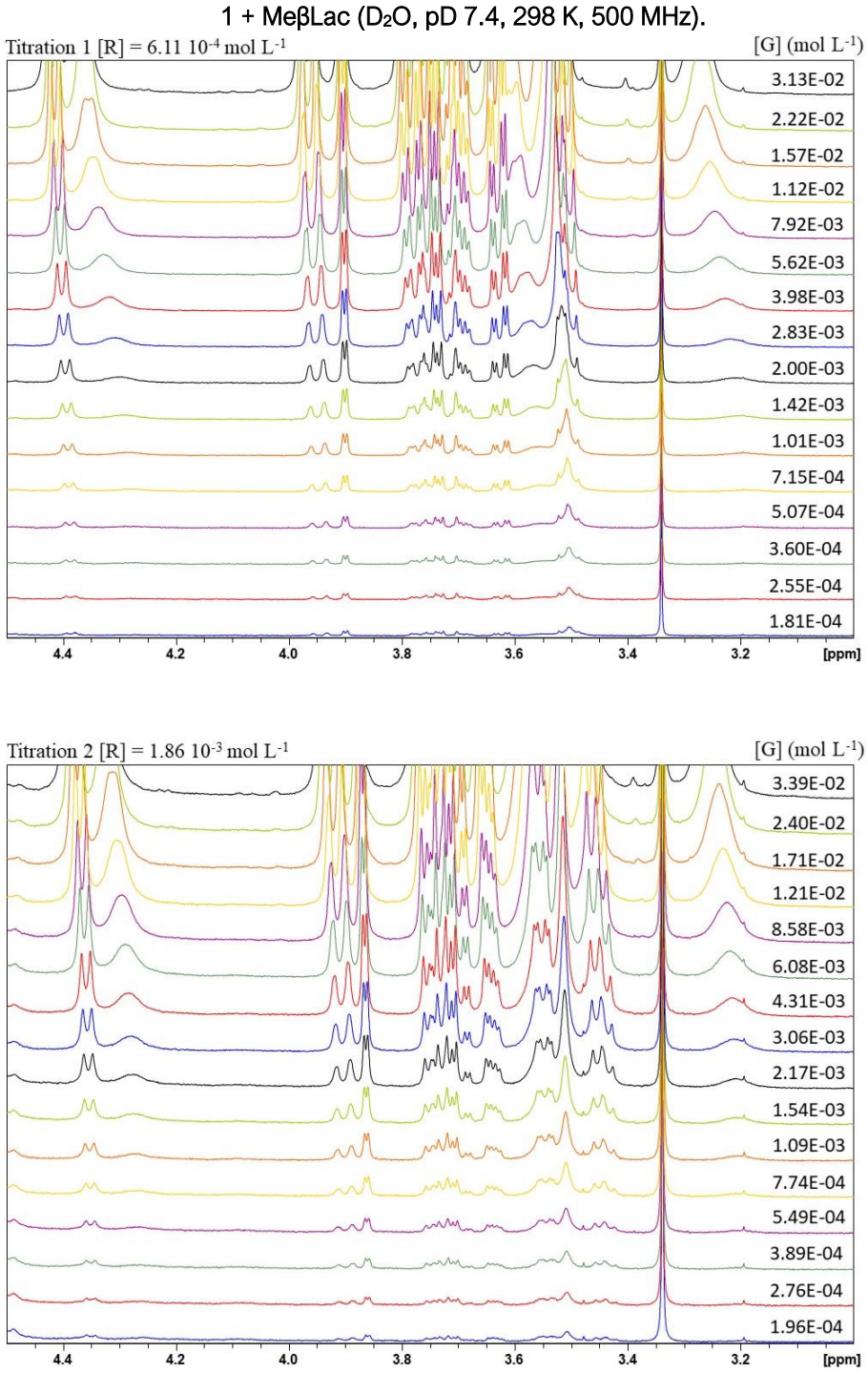


Figure S7. ^{1}H NMR spectroscopic titrations (500 MHz, D_2O, pD 7.4, 298 K) of receptor **1** (R) with incremental concentrations of MeβLac (G).

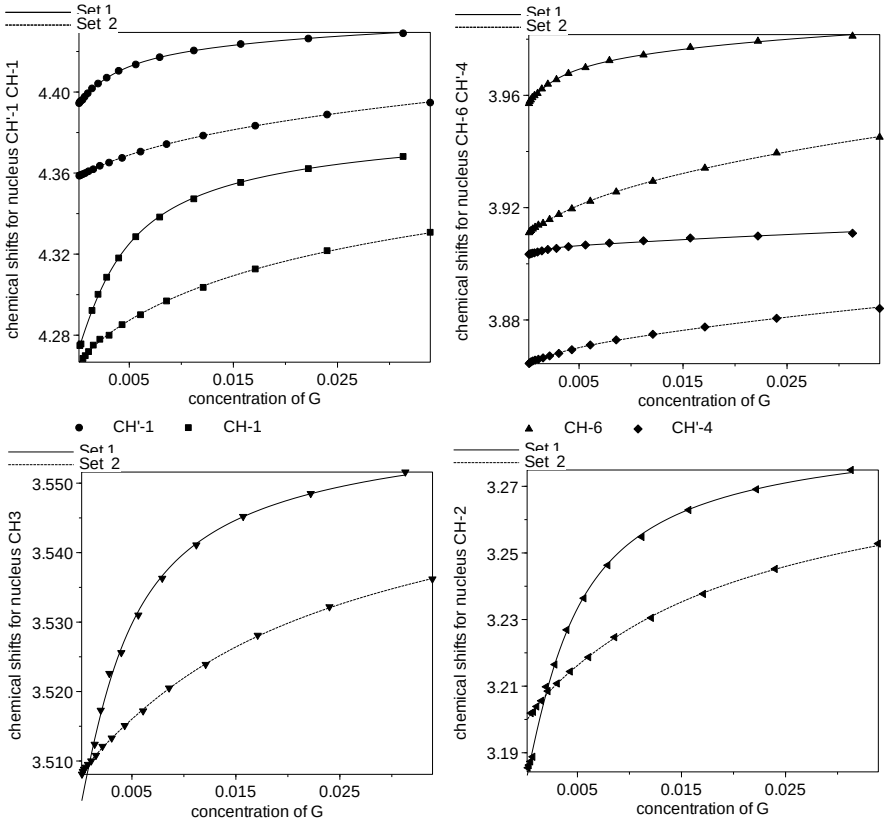
Results page

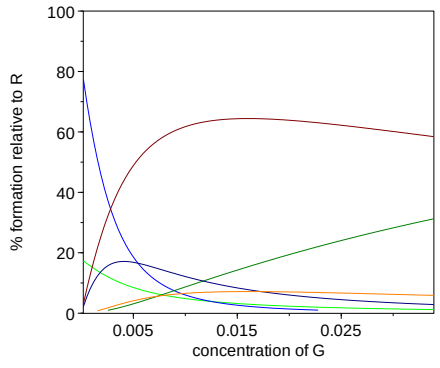
```
 no. of spectra            32
 no. of resonance values  173
 no. of resonant nuclei     6

   sigma =   0.00045085220          RMS weighted residual =   0.00040412763

       stoich            value   relative    log    standard
       coeff                     std devn    beta   deviation

 Beta  0  2 constant 6.9231E+003             3.8403           (   R2   )
 Beta  1  1 refined  1.5380E+003  0.0287     3.1870   0.0125  (   GR   )
 Beta  2  1 refined  2.6093E+004  0.0524     4.4165   0.0228  (   G2R  )
 Beta  1  2 refined  1.6783E+006  0.0883     6.2249   0.0383  (   GR2  )
 Beta  2  2 refined  1.0965E+008  0.3043     8.0400   0.1322  (   G2R2 )

Individual chemical shifts

                    G                         R
 ========================================================
        +        value    error          value    error
CH'-1   +       4.4496   0.0007
CH-1    +       4.3936   0.0008
CH-6    +       4.0022   0.0007
CH'-4   +       3.9261   0.0006
CH3     +       3.5620   0.0005
CH-2    +       3.2913   0.0007
                   0,2                       1,1
 ========================================================
        +        value    error          value    error
CH'-1   +                                4.2138   0.0084
CH-1    +                                3.5984   0.0237
CH-6    +                                3.8750   0.0060
CH'-4   +                                3.9186   0.0040
CH3     +                                3.1514   0.0132
CH-2    +                                2.4855   0.0237
                   2,1                       1,2
 ========================================================
        +        value    error          value    error
CH'-1   +       3.0887   0.0478          4.2617   0.0163
CH-1    +       3.0443   0.0499          5.4141   0.0908
CH-6    +       2.5299   0.0511          3.5179   0.0393
CH'-4   +       2.8070   0.0402          3.3825   0.0451
CH3     +       3.0565   0.0248          4.1991   0.0558
CH-2    +       2.6018   0.0322          4.6695   0.1197
                   2,2
 ================================
        +        value    error
CH'-1   +       4.4177   0.0499
CH-1    +       6.8388   0.7318
CH-6    +       3.4126   0.1684
CH'-4   +       3.1621   0.2201
CH3     +       4.8938   0.3952
CH-2    +       6.1712   0.8524

 Correlation coefficients*1000

       1    2    3    4
   1
   2  280
   3 -647  208
   4 -544  236  498
```

```
Parameters are numbered as follows
   1 beta 1,1
   2 beta 2,1
   3 beta 1,2
   4 beta 2,2
```

Titration Plots

Chemical shifts (δ, ppm) vs. concentration of G (mol L^{-1})
experimental (symbols) and calculated (lines) values





R R_2 GR G_2R GR_2 G_2R_2

1 + MeβGlcNAc$_2$ (D_2O, pD 7.4, 298 K, 500 MHz).
R = **1** G = MeβGlcNAc$_2$
Titration
[R] = 8.54 10^{-4} mol L^{-}

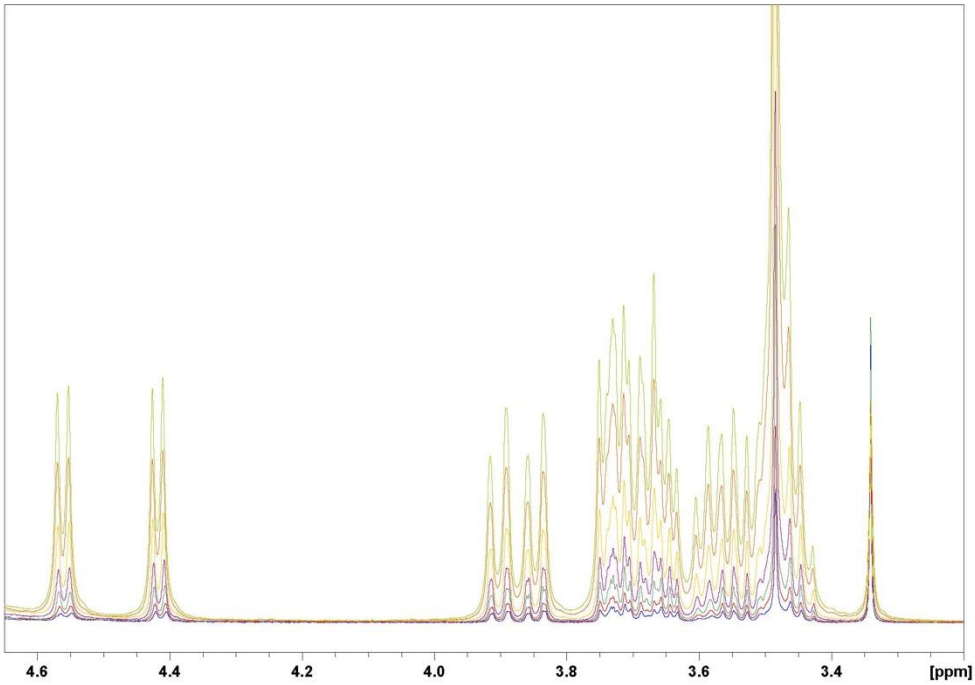


Figure S8. Superposition of ^{1}H NMR spectra registered at incremental concentrations of MeβGlcNAc$_2$ (0.904 mM, 1.64 mM, 2.99 mM, 5.45 mM, 10.8 mM, 21.6 mM, 30.4 mM) in a 0.854 mM solution of **1**. Expansion of the saccharide region.

R:G	$log\beta_n$	BC_{50}^0
1:1	2.81±0.04	1.93±0.40
1:2	3.96±0.38	
2:1	6.88±0.04	
2:2	8.80±0.07	

Table S1. Cumulative formation constants ($log\beta_n$)[a] and intrinsic median binding concentration (BC_{50}^0, mM)[b] for receptor **1** to MeβMal (R:G) complexes, measured at 298 K from NMR data in D_2O at pD 11.[c] [a] Formation constants were obtained by nonlinear least-square regression analysis of NMR data. [b] Calculated from the $log\beta$ values using the "BC_{50} Calculator" program.[16] [c] Receptor dimerization constant at pD 11 (1: $log\beta_{dim}$ = 3.95±0.11) was set invariant in the nonlinear regression analysis of NMR data.

3.4.2 Structural studies

Chemical shift difference (CSD) analysis:

	Nucleus	free	bound	CSD
MeβCeB	H'-1	4.5078	4.2228	0.2850
	H-1	4.3897	3.5825	0.8072
MeβMal	H'-1	5.3883	5.0725	0.3158
	H-1	4.3786	3.5523	0.8263
MeβLac	H'-1	4.4496	4.2138	0.2358
	H-1	4.3936	3.5984	0.7952

Table S2. Chemical shifts and chemical shift differences (CSD) [ppm] of the anomeric protons of MeβCeB, MeβMal and MeβLac in free and 1:1 bound state with **1** in D_2O (pD 7.4) at T=298 K[a] [a] Chemical shifts were obtained by nonlinear regression of the experimental data from titration of MeβCeB, MeβMal and MeβLac with **1**.

	Nucleus	free	bound	CSD
MeβCeB	H'-1	4.5078	0.7440	3.7638
	H-1	4.3897	0.8405	3.5492
MeβMal	H'-1	5.3883	3.9648	1.4235
	H-1	4.3786	2.3238	2.0548
MeβLac	H'-1	4.4496	3.1512	1.2984
	H-1	4.3936	2.9625	1.4311

Table S3. Chemical shifts and chemical shift differences (CSD) [ppm] of the anomeric protons of MeβCeB, MeβMal and MeβLac in free and 1:1 bound state with **2** in D_2O (pD 7.4) at T=298 K[a] [a] Chemical shifts were obtained by nonlinear regression of the experimental data from titration of MeβCeB, MeβMal and MeβLac with **2**.

NMR methods: NMR experiments were performed at 500 MHz in D_2O at pD 11 at 298 K. The experiments on the complex were performed using an equimolar solution of **1** or **2** (R) and MeβCeB (G). In addition to standard 1D ^1H NMR spectra, COSY, TOCSY, HSQC and NOESY experiments (500 ms mixing time) were also acquired to assign the resonances of all the molecular entities and to detect the relevant intramolecular and intermolecular contacts.

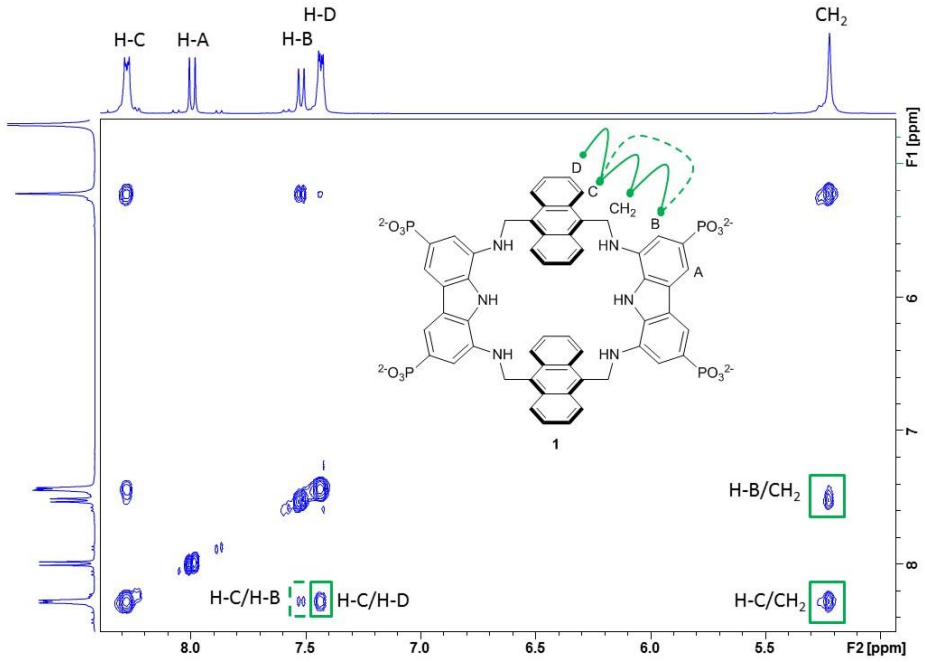


Figure S9. 500 MHz NOESY spectrum of an equimolar mixture of MeβCeB and **1** (30 mM each) in D_2O at 298 K. Intramolecular NOE cross peaks of the receptor **1** are indicated by squares and schematically represented (solid and dashed squares and lines were used to indicate strong and medium NOEs found).

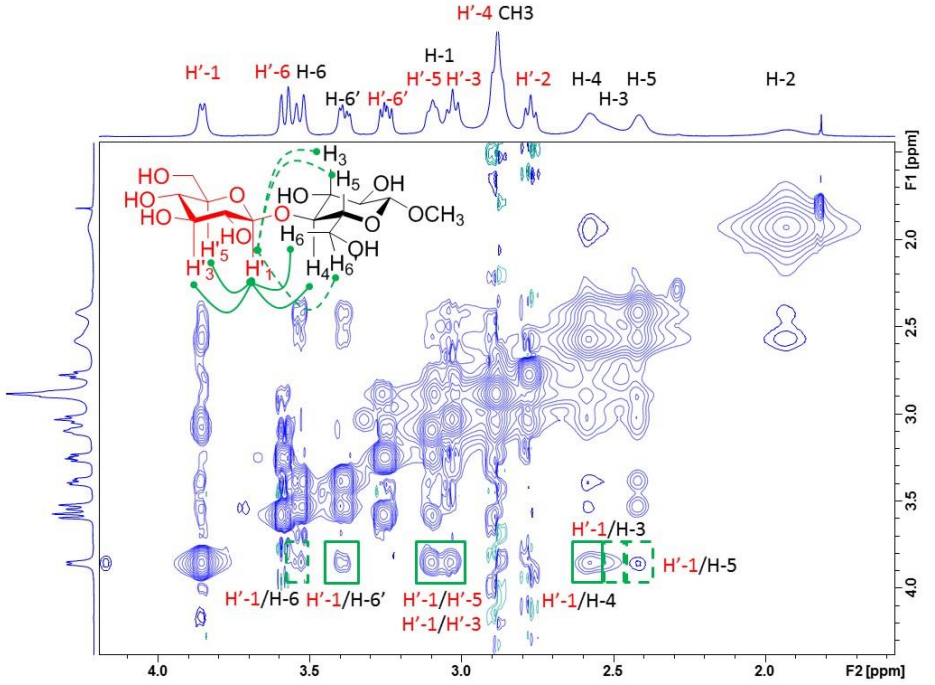


Figure S10. 500 MHz NOESY spectrum of an equimolar mixture of MeβCeB and **1** (30 mM each) in D_2O (pD 11) at 298 K. Intramolecular NOE cross peaks of MeβCeB relating to H'-1 proton are indicated by squares and schematically represented (solid line and dashed lines were used to indicate strong and weak NOEs found).

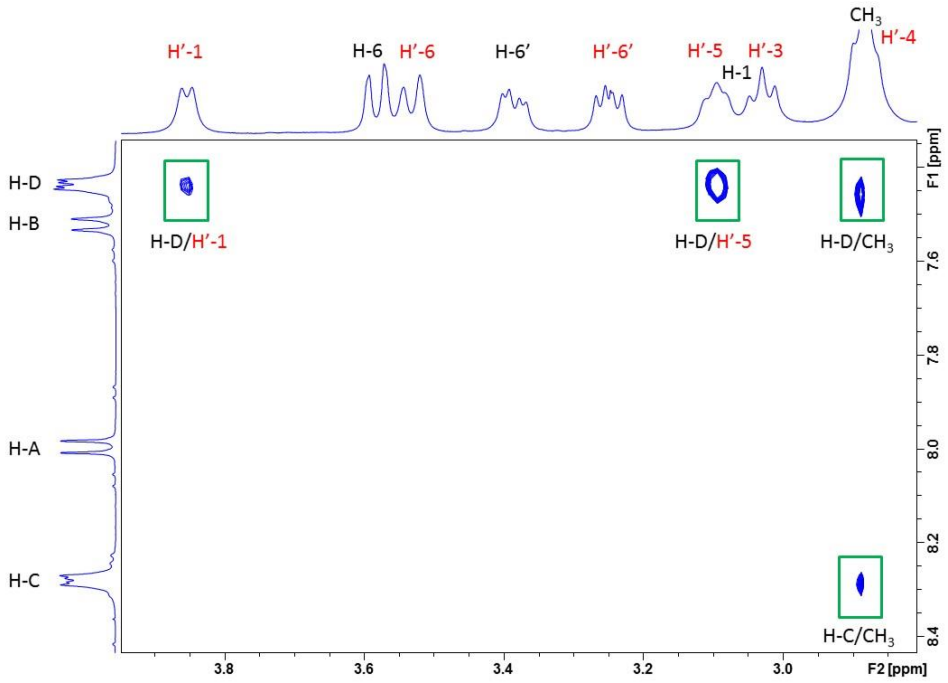


Figure S11. 500 MHz NOESY spectrum of an equimolar mixture of MeβCeB and **1** (30 mM each) in D_2O (pD 11) at 298 K. Strong intermolecular NOE cross peaks are indicated by solid squares.

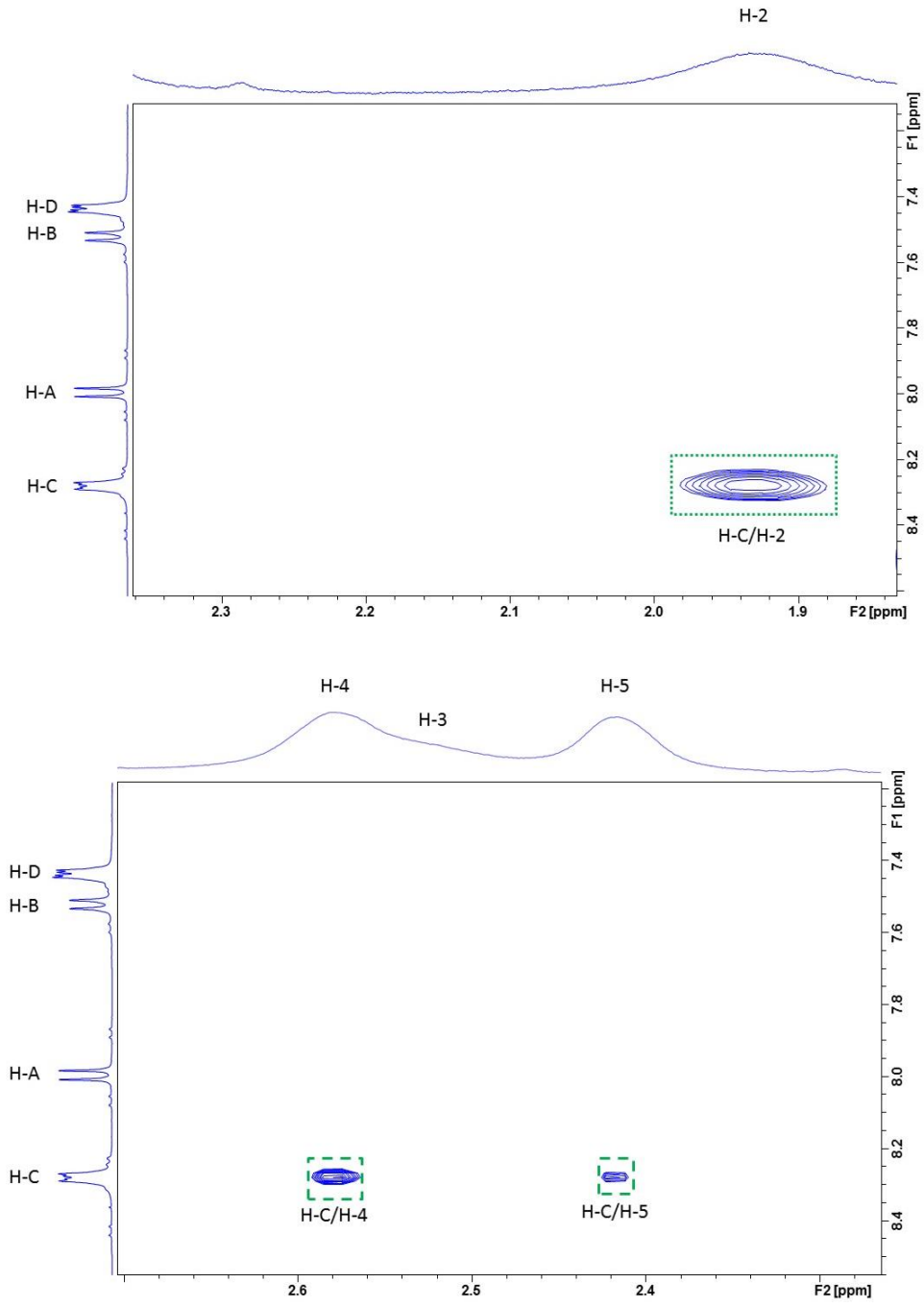


Figure S12. 500 MHz NOESY spectrum of an equimolar mixture of MeβCeB and **1** (30 mM each) in D_2O at (pD 11) 298 K. Medium and weak intermolecular NOE cross peaks are indicated by dashed and dotted squares respectively.

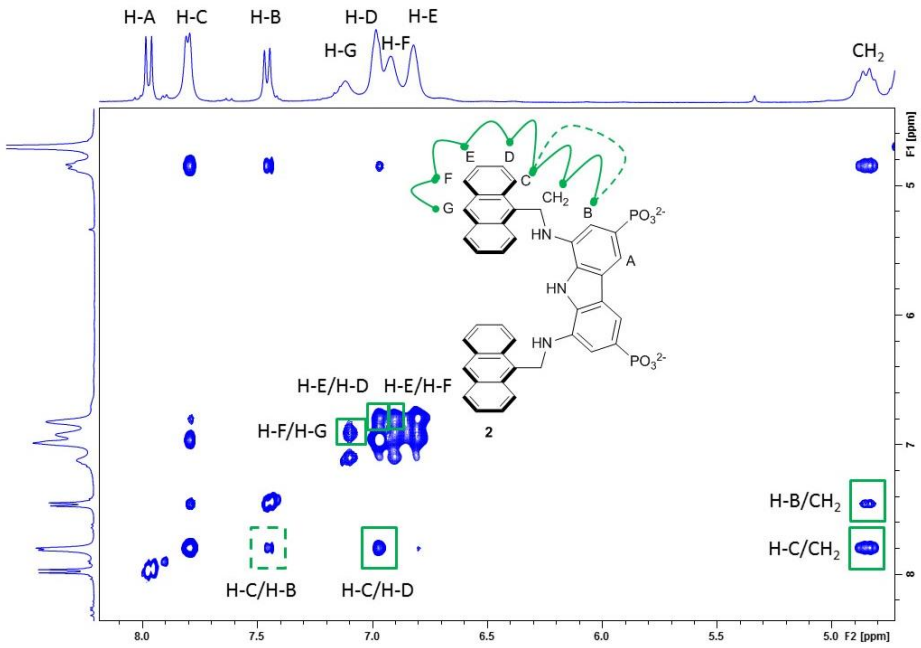


Figure S13. 500 MHz NOESY spectrum of an equimolar mixture of MeβCeB and **2** (20 mM each) in D_2O at 298 K. Intramolecular NOE cross peaks of the receptor **2** are indicated by squares and schematically represented (solid and dashed squares and lines were used to indicate strong and medium NOEs).

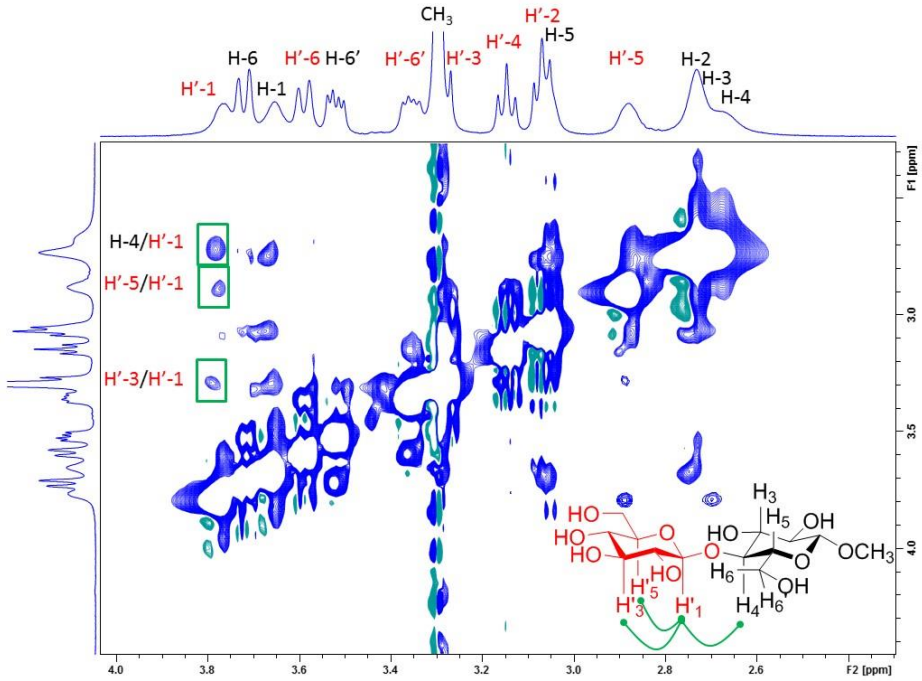


Figure S14. 500 MHz NOESY spectrum of an equimolar mixture of MeβCeB and **2** (20 mM each) in D_2O (pD 11) at 298 K. Intramolecular NOE cross peaks of MeβCeB relating to H'-1 proton are indicated by squares and schematically represented.

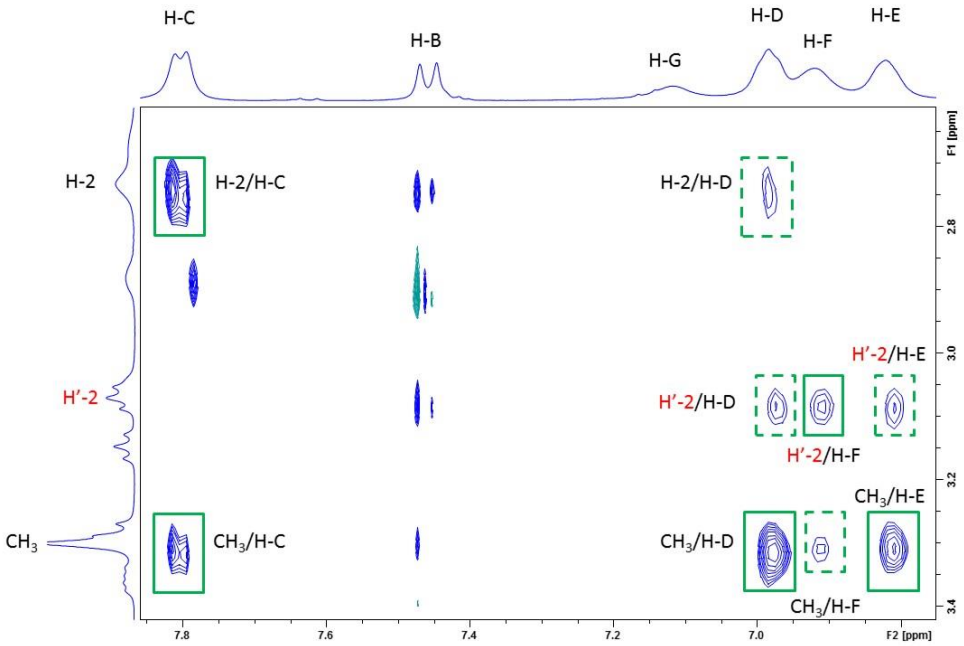


Figure S15. 500 MHz NOESY spectrum of an equimolar mixture of MeβCeB and **2** (20 mM each) in D_2O (pD 11) at 298 K. Strong and medium intermolecular NOE cross peaks are indicated by solid and dashed squares respectively.

Molecular modeling methods: Initial structures of MeβCeB, receptor **1** and receptor **2** were built and minimized using conjugate gradients with the OPLS_2005 force field, water was set as solvent and an extended cutoff was used to treat remote interactions. A maximum number of 5000 iterations were employed with the Polak-Ribiere Conjugate Gradient (PRCG) scheme, until the convergence energy threshold was 0.05. Once the optimum geometries had been achieved, a conformational search protocol was adopted for the receptors, using a Monte Carlo torsional sampling method (MCMM) with automatic setup during the calculation, energy window of 21 kJ mol^{-1}, 10000 maximum number of steps, and 100 steps per torsion of the bond to be rotated. The best structures obtained from this calculation in terms of energy were chosen and then, the disaccharide was manually docked within the receptor cleft with different starting relative orientations and further minimized. Minimization results afford different structures which were employed as input for further conformational search protocols without any constraints. Several complexes were found to be stable, in which the sugar was located inside the receptor cleft. The lowest energy structures were analyzed to check the agreement with experimental NMR data. The protocol returned a family of structures, containing the minimum energy structure of the conformational search, in agreement with the observed NOE data.

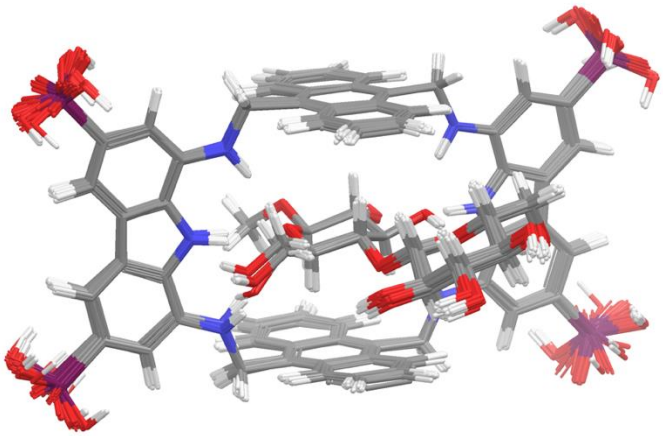


Figure S16. Molecular modelling results from conformational search for the complex of **1** with MeβCeB. Superposition of the 41 energy minimum structures, within an energy window of 5.19 kJ mol^{-1}, identified among the 1152 structures obtained from the calculation.

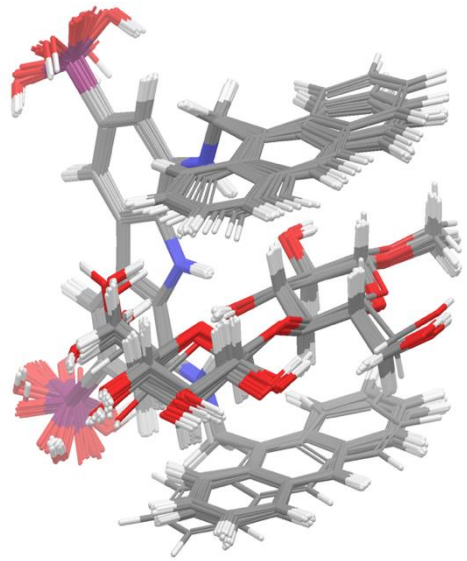


Figure S17. Molecular modelling results from conformational search for the complex of **2** with MeβCeB. Superposition of the 31 energy minimum structures, within an energy window of 8.73 kJ mol^{-1}, identified among the 653 structures obtained from the calculation.

	H'-1	H'-5	CH_3	H-4	H-5	H-2
H-C			s, 3.16	m, 3.59	m, 3.03	w, 3.69
H-D	s, 2.91	s, 2.43	s, 3.06			-

Table S4. Intermolecular NOE cross-peaks found in NOESY spectrum of an equimolar mixture of MeβCeB and **1**, and corresponding distances (Å) obtained from the calculated minimum energy structures.[a] [a] (s), strong; (m), medium; (w), weak

	H-2	H'-2	OCH_3
H-C	s, 3.05		s, 4.54
H-D	m, 3.56	m, 3.92	s, 3.17
H-F		s, 3.21	m, 4.02
H-E		m, 2.67	s, 2.79

Table S5. Intermolecular NOE cross-peaks found in NOESY spectrum of an equimolar mixture of MeβCeB and **2**, and corresponding distances (Å) obtained from the calculated minimum energy structures.[a] [a] (s), strong; (m), medium.

	OCH_3	H-1	H-2	H-3	H-4	H-5	H'-1	H'-2	H'-5
1	2.76	2.72	2.63	2.74	2.71	2.71	-	-	-
2	-	2.63	2.74	2.74	2.73	2.97	2.83	2.82	2.80

Table S6. Distances (Å) measured between MeβCeB protons and the ring planes of receptors anthracenes compliant with CH-π interactions obtained from the minima energy structures of the complexes between MeβCeB and receptors **1** and **2**.

3.5 Bibliography

[1]Habibi Y., Lucia L:A., Rojas O.J., *Chem. Rev.*, **2010**, 110, 3479-3500.

[2] Martínez J.P., Falomir M.P., Gozalbo D., Chitin: *A Structural Biopolysaccharide with Multiple Applications. In: eLS. John Wiley & Sons*, **2014**.

[3]Thompson A.J., de Vries R.P., Paulson J.C., *Curr. Opin. Virol*., **2019**, 34, 117-129.

[4]a) Balzarini J., *Nat. Rev. Microbiol*. **2007**, *5*, 583–597; b) Bagdonaite I., Wandall H.H., *Glycobiology*, **2018**, 28, 443-467.

[5]a) Bravo M.F., Lema M.A., Marianski M., Braunschweig A.B., *Biochemistry*, **2021**, 60, 999-1018;
b) Tommasone B., Allabush F., Tagger Y.K., Norman J., Köpf M., Tucker J.H.R., Mendes P.M., *Chem. Soc. Rev.* **2019**, 48, 5488-5500.

[6]a) Francesconi O., Roelens S., *ChemBioChem,* **2019**, 20, 1329-1346; b) Francesconi O., Nativi C., Gabrielli G., De Simone I., Noppen S., Balzarini J., Liekens S., Roelens S., *Chem. Eur. J.*, **2015**, 21, 10089-10093.

[7]a) Whitehead M., Turega S., Stephenson A., Hunter C.A., Ward M.D., *Chem. Sci*., **2013**, 4, 2744-2751;
b) Ferguson Johns H.P., Harrison E.E., Stingley K.J., Waters M.L., *Chem. Eur. J.*, **2021**, 27, 6620-6644.

[8]Davis A.P., *Chem. Soc. Rev.*, **2020**, 49, 2531-2545.

[9]Martí-Centelles V., Pandey M.D., Burguete M.I., Luis S.V., *Chem. Rev.*, **2015**, 115, 8736-8834.

[10]a) Pal A., Berube M. and Hall D.G., *Angew. Chem. Int. Ed.*, **2010**, *49*, 1492-1495; *Angew. Chem*., **2010**, 122, 1534-1537; b) Mazik M. and Cavga H., *J. Org. Chem*., **2006**, 71, 2957-2963.

[11]Recent examples of acyclic biomimetic receptors effective in organic solvents: a) Palanichamy K., Bravo M.F., Shlain M.A., Schiro F., Naem Y., Marianski M., Braunschweig A.B., *Chem. Eur. J.*, **2018**, 24, 13971-13982;
b) Mateus P., Chandramouli N., Mackereth C.D., Kauffmann B., Ferrand Y., Huc I., *Angew. Chem. Int. Ed*., **2020**, *59*, 5797-5805; *Angew. Chem.*, **2020**, 132, 5846-5854.

[12]Francesconi O., Martinucci M., Badii L., Nativi C, Roelens S., *Chem. Eur. J.*, **2018**, 24, 6828-6836.

[13]Francesconi O., Ienco A., Nativi C., Roelens S., *ChemPlusChem*, **2020**, 85, 1369-1373.

[14]Francesconi O., Cicero F., Nativi C, Roelens S., *ChemPhysChem*, **2020**, 21, 257-262.

[15]Examples of adaptive architectures in carbohydrate recognition: a) Kim Y.H., Hong J.I., *Angew. Chem. Int. Ed.*, **2002**, *41*, 2947-2950; *Angew. Chem.*, **2002**, 114, 3071-3074; b) Mazik M., *RSC Adv.*, **2012**, 2, 2630-2642;
c) Gentili M., Nativi C., Francesconi O., Roelens S., *Carbohydr. Chem*., **2016**, 41, 149-186; d) Francesconi O., Gentili M., Nativi C., Ardá A., Cañada F.J., Jiménez-Barbero J., Roelens S., *Chem. Eur. J*., **2014**, 20, 6081-6091; e) Mateus P., Wicher B., Ferrand Y., Huc I., *Chem. Commun*., **2018**, 54, 5078-5081; f) Ohishi Y., Abe H., Inouye M., *Chem. Eur. J.*, **2015**, 21, 16504-16511; g) Ohishi Y., Masuda K., Kudo K., Abe H., Inouye M., *Chem. Eur. J.*, **2021**, 27, 785-793.

[16]Vacca A., Francesconi O., Roelens S., *Chem. Rec.*, **2012**, 12, 544-566.

[17]Ardá A., Venturi C., Nativi C, Francesconi O., Cañada F.J., Jiménez-Barbero J., Roelens S., *Eur. J. Org. Chem.*, **2010**, 64-71.

[18]Vacca A., Nativi C., Cacciarini M., Pergoli R., Roelens S., *J. Am. Chem. Soc*., **2004**, 126, 16456-16465.

[19]Frassineti C., Ghelli S., Gans P., Sabatini A., Moruzzi M.S., Vacca A., *Anal. Biochem*., **1995**, 231, 374-382.

Chapter 4

A biomimetic receptor for glycans. Selective recognition of the core $GlcNAc_2$ disaccharide of the sialylglycopeptide SGP

Adapted from:

Milanesi F., Unione L., Ardá A., Nativi C., Jiménez-Barbero J., Roelens S., Francesconi O., *manuscript in preparation,* 2022

Abstract

In recent years, glycomics has shown how pervasive is the role of carbohydrates in biological systems and how chemical tools are essential to investigate glycan function and modulate carbohydrate-mediated processes. Biomimetic receptors for carbohydrates can carry out this task but, although significant affinities and selectivities toward simple saccharides have been achieved, targeting complex glycoconjugates still remains unattained. In this work we report the unprecedented recognition of a biantennary sialoglycopeptide (SGP) by a tweezers-shaped biomimetic receptor, which selectively binds to the core $GlcNAc_2$ disaccharide of the *N*-glycan with an affinity of 170 μM. Because of the simple structure and the remarkable binding ability, this biomimetic receptor can represent a versatile tool for glycoscience, opening the way to useful applications.

4.1 Introduction

Oligosaccharides are widely expressed on the surface of eukaryotic cells as glycoconjugates of proteins and lipids.[1,2] Their selective recognition by different classes of proteins, such as lectins, triggers essential physio- and pathological processes, like cell adhesion and infection, cell differentiation, inflammation, and immune response.[3-5] Glycosylation is a ubiquitous post-translational modification of proteins that leads to conjugation with O- and N-linked oligosaccharides. In particular N-glycans, that is, N-linked oligosaccharides, all share the common pentasaccharidic structure $Man_3GlcNAc_2$ [Manα1-3(Manα1-6)Manβ1-4GlcNAcβ1–4GlcNAcβ] N-linked to the asparagine residue of the Asn-X-Thr/Ser (X is any amino acid except proline) sequence in the protein backbone.[6] Depending on the structural diversity of the oligosaccharides extending beyond the common structure, N-glycans are classified as high-mannose type, complex type, and hybrid type. *N*-glycans play essential roles in the chemical and physicochemical behaviour of proteins, including protein folding and stability,[7,8] solubility,[9] resistance to proteolysis,[10] intracellular traffic,[11] antigenicity,[12] activity, and recognition by carbohydrate-binding proteins.[13] Because of glycan ubiquity, agents targeting *N*-glycans can lead to a deeper understanding of their function and to the development of new diagnostic and prognostic strategies.[14]

Typically, recognition of specific saccharides to drive function is achieved in nature by lectins, which are glycan binding proteins specialized in recognizing particular sugar epitopes.[15] However, their pharmaceutical development has remained largely elusive, probably to their susceptibility to proteolysis, their ability to elicit immune response, and other features related to their protein nature.

Biomimetic receptors for carbohydrates are a heterogeneous class of molecules developed for selective recognition of carbohydrates of biological interest.[16-21] To mimic the recognition function of lectins, the binding ability of biomimetic receptors rely on non-covalent interactions exclusively. However, achieving effective recognition of polar species like saccharides in water relying on a biomimetic approach is a non-trivial task.[22,23] Indeed, significant milestones in the field have been achieved only recently, mainly by means of the use of smart designs of sophisticated macrocyclic architectures.[24–30] In contrast, the use of acyclic receptors has been largely unexplored,[31–35] although they may present distinct advantages over macrocyclic organizations in terms of simpler synthesis, more easily modifiable structure, adaptability to different shapes and, most of all, to the possibility of binding non-terminal saccharide entities.

We have recently reported the acyclic, tweezers-shaped receptor **1** (**Figure 1**),[36,ch.2] which effectively recognizes disaccharides in water at physiological pH. Receptor **1** is based on a very simple structure, consisting of a diaminocarbazole hydrogen-bonding unit[ch.3] and two anthracene rings establishing CH-π interactions with the aliphatic backbone of carbohydrates. Receptor **1** selectively recognized the methyl β glycoside of *N*,*N*'-diacetylchitobiose ($MeβGlcNAc_2$), the core disaccharide of the conserved pentasaccharidic structure of *N*-glycans, over a wide range of mono- and disaccharides, showing the remarkable affinity of 160 μM. The latter, not only exceeds the affinity exhibited by more complex macrocyclic receptors, but also that shown by the pseudo-lectin hevein (from *Hevea brasiliensis*) by more than one order of magnitude (K_D = 1.61 mM).[37] Receptor **1** stands therefore as the ideal candidate to be tested vs. N-glycans of biological interest.

Among *N*-glycans, sialylated *N*-glycans are of particular interest, because are the target of a class of lectins named Siglecs (Sialic acid binding immunoglobulin type lectins).[38,39] Siglecs are cell surface immunomodulatory receptors, which selectively recognize specific sialylated glycans exposed on all mammalian cells and help immune cells to distinguish between self and non-self. One of the most studied sialylated N-glycans is the complex

biantennary sialylundecasaccharide SGP, which is obtained in large scale from egg yolk as the glycopeptide (**Figure 1**).[4,40]

Despite the above remarkable features of the receptor, the evaluation of the binding properties of **1** *vs*. more complex *N*-glycans is hampered by the strong self-association in water ($log\beta_{dim}$ = 2.65 ± 0.07),[ch.2] mainly due to the extended aromatic surfaces of anthracenes, stacking intermolecularly under the pressure of hydrophobic effects.[41] Unfortunately, such a strong dimerization hampers not only the determination of binding affinities, but also a structural definition of the binding modes with complex glycans.

In this work we present a second-generation tweezers-shaped receptor **2** (**Figure 1**) which, in addition to addressing the self-association issues of the progenitor, shows the ability, unprecedented for a biomimetic receptor, of recognizing the natural *N*-glycan SGP by selectively binding to the $GlcNAc_2$ core disaccharide.

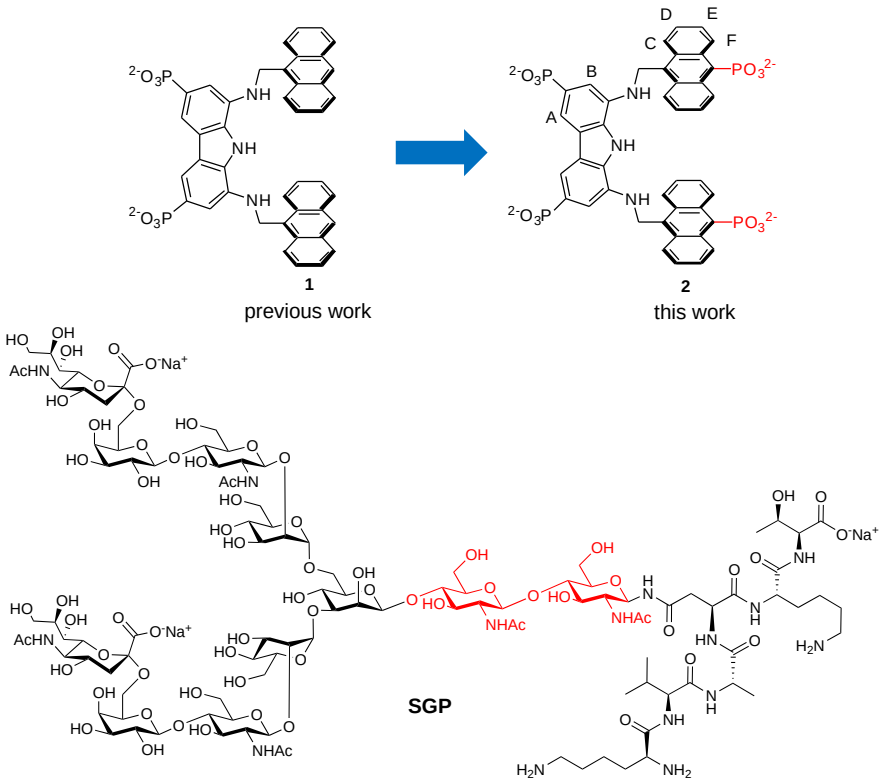


Figure 1. Concept of the work: conversion of the biomimetic receptor **1** to **2** (with proton labelling) to be tested vs the sialoglycopeptide SGP. The $GlcNAc_2$ disaccharide is highlighted in red.

4.2 Results and Discussion

Based on the recently demonstrated evidence that introduction of charged substituents on the anthracene rings inhibited self-association of tweezers-shaped receptors, we have functionalized the para position of the anthracene rings of **1** with two additional phosphonate groups to give receptor **2** (**Figure 1**).[42]

Receptor **2** was prepared from commercially available 9,10-dibromoanthracene (**3**) in six steps (**Figure 2a**). After formylation of **3** using *n*-BuLi followed by DMF, the carbaldehyde **4** was protected as the dimethylpropylene acetal **5** by the *p*TsOH catalyzed condensation with neopentyl glycol. The bromo-substituent of the acetal intermediate was then replaced by a

diethylphosphonate group using *n*-BuLi followed by $PO(OEt)_2Cl$. The acetal **6** was then hydrolyzed by HCl in acetone, to give the aldehyde **7**, which was condensed with the diaminocarbazole **8**, prepared according to a previously reported method, to provide the corresponding bis-imino derivative. Subsequent reduction with $NaBH_4$ gave the liposoluble receptor **9**. Hydrolysis of the phosphonate esters with TMSBr, followed by treatment with methanol gave the desired receptor **2**.

In contrast to its progenitor, dilution experiments carried out on receptor **2** at pH 7.4 revealed sharp signals in a concentration range from 24 to 1 mM, together with a modest variation of ^{1}H-NMR chemical shifts, suggesting a weak self-association (**Figure S4**). Indeed, nonlinear regression of the corresponding data gave a dimerization constant ($\log\beta_{dim}$ = 0.98 ± 0.05) that, although still detectable, is more than one order of magnitude smaller than that observed with parent receptor 1. As expected, additional hydrosolubilising groups on the anthracene moieties disrupted π-stacking interactions and, consequently, self-association phenomena.

A qualitative screening of the binding properties of receptor **2** was then carried out by ^{1}H NMR spectroscopy (**Figures S1-S3**) toward a set of mono- and disaccharides *vs* which the parent receptor **1** was previously tested, including monosaccharides (glucose, N-acetylgalactosamine, rhamnose, fucose, xylose, sialic acid, α and β methyl glucosides, galactosides, mannosides, and *N*-acetylglucosamine) (**Figure 2b**), together with disaccharides [sucrose (Suc), trehalose (Tre), cellobiose (CeB), maltose (Mal) and lactose (Lac)] (**Figure 2c**), by monitoring the shifts of the proton signals of the sugar upon addition of an equimolar amount of **2**. Except for the methyl β glucoside (MeβGlc), for which an appreciable variation was observed ($\Delta\delta_{H\text{-}1}$ = 0.05 ppm), little ($\Delta\delta$ < 0.04 ppm) or no variations were observed for the other monosaccharides and for Suc and Tre, whereas a marked shift was observed for the β anomer of CeB and, to a smaller extent, for that of Lac and Mal, analogously to what was previously observed with the parent receptor **1**, suggesting a similar binding profile.

A quantitative investigation was thus carried out by NMR spectroscopy on the methyl β glycosides of glucose (MeβGlc), cellobiose (MeβCeB), maltose (MeβMal), lactose (MeβLac), and *N,N'*-diacetylchitobiose ($Me\beta GlcNAc_2$). Methyl glycosides were used to avoid interconversion equilibria between anomers. Cumulative association constants, reported in **Table 1**, were measured by ^{1}H NMR titrations in D_2O (pD 7.4) at 298 K, simultaneously fitting the complexation induced shifts of all the available signals to the appropriate association model by nonlinear regression analysis. Because multiple binding constants were detected, affinities were assessed though the intrinsic median binding concentration parameter BC_{50}^0,[43] calculated from the measured binding constants and reported in **Table 1**.

At first glance, the data reported in **Table 1** confirm a closely similar binding profile to that observed with the progenitor **1**. Indeed, receptor 2 effectively discriminates disaccharides from monosaccharides, with a significantly lower affinity for MeβGlc. Moreover, receptor 2 exhibits the preference for all-equatorial disaccharides observed with **1**, binding to MeβCeB and $Me\beta GlcNAc_2$ with larger affinities than to MeβLac and MeβMal. Finally, the higher affinity of the set was observed for $Me\beta GlcNAc_2$, (270 μM), a value quite close to that of 160 μM observed for **1**, showing that the additional phosphonate groups only marginally affect binding of the investigated sugars.

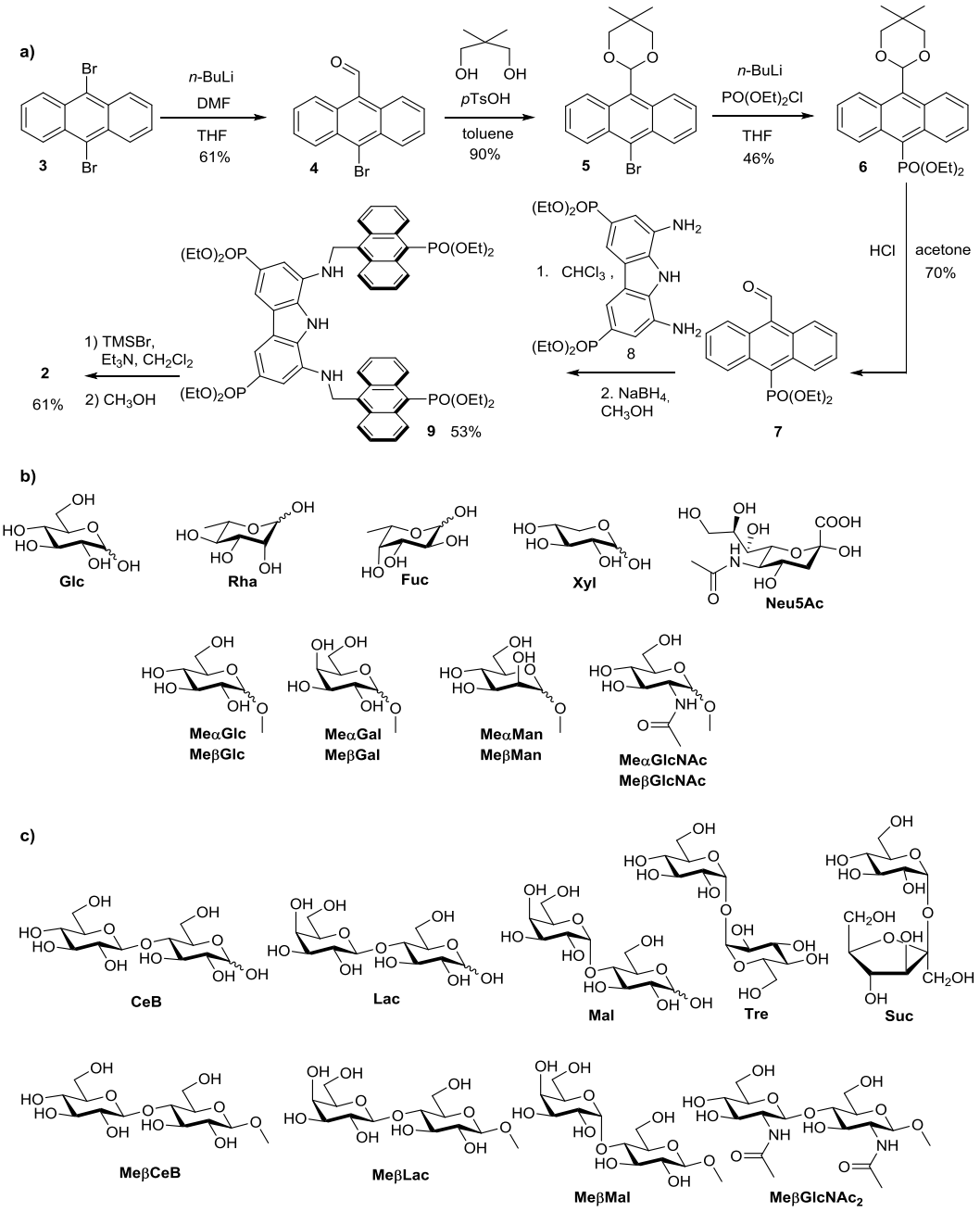


Figure 2. a) Synthetic scheme of the biomimetic receptor 2. Investigated monosaccharides (b) and disaccharides (c) and their abbreviations.

Receptor		2 (NMR)		2 (ITC)	
Glycoside	R:G	*logβ*	BC_{50}^{0}	*logβ*	BC_{50}^{0}
MeβGlc	1:1	1.59±0.01	38.4±1.7		
MeβCeB	1:1	2.40±0.01	1.04±0.06	2.49±0.07	1.32±0.43
	2:1	5.79±0.07		5.53±0.38	
	2:2	8.00±0.09			
MeβMal	1:1	2.00±0.01	11.9±0.3	1.78±0.02	16.6±0.7
MeβLac	1:1	2.11±0.01	9.01±016	1.97±0.01	10.5±0.3
$Me\beta GlcNAc_2$	1:1	3.47±0.02	0.27±0.01	3.45±0.10	0.18±0.04
	2:1	6.48±0.05		7.16±0.23	
SGP	1:1			3.44±0.26	0.17±0.05
	2:1			7.30±0.26	

Table 1. Cumulative formation constants (log β_n)[a] and intrinsic median binding concentration (BC_{50}^{0}, mM)[b] for receptor to glycoside (R:G) complexes of **2** with methyl glycosides and SGP, measured at 298 K from NMR data in D_2O at pD 7.4 and from ITC data in H_2O at pH 7.4.[c] [a] Formation constants were obtained by nonlinear least-square regression analysis of NMR data. [b] Calculated from the log β values using the "BC_{50} Calculator" program.[16] [c] Receptor dimerization constants at pH 7.4 (**2**: $log\beta_{dim}$ = 0.98 ± 0.05) was set invariant in the nonlinear regression analysis of NMR.

To shed light on the energetics of binding and to support the affinities measured for 1,4-disaccharides by ^{1}H-NMR spectroscopy, binding properties were also evaluated by ITC in H_2O at pH 7.4. Because the dimerization constant of the receptor **2** is negligible, dimerization was not included in the nonlinear regression analysis of ITC data. Cumulative binding constants, together with affinity values, were reported in **Table 1** for a direct comparison with NMR results. The good agreement between the two independent techniques confirmed the reliability of the observed affinities. In contrast to receptor **1**, for which strong dimerization prevented the determination of thermodynamic parameters, reliable data for the formation of 1:1 complexes could be obtained for receptor **2**, and corresponding thermodynamic parameters are reported in **Table 2**. It can be appreciated that the formation of all 1:1 complexes with disaccharides are enthalpically driven, with an adverse entropic contribution, which is consistent with a prominent role of polar interactions over solvophobic effects.

Glycoside	-ΔG0	-ΔH0	TΔS0
MeβCeB	14.2±0.4	26.1±1.4	-11.9±1.0
MeβMal	10.2±0.1	17.9±0.5	-7.7±0.6
MeβLac	11.3±0.1	23.4±0.5	-12.1±0.6
MeβGlcNAc2	19.7±0.6	47.3±2.2	-27.7±2.7
SGP	19.6±1.5	14.5±5.0	5.2±6.5

Table 2. Thermodynamic parameters (kJ mol^{-1}) for the formation of receptor **2** to glycoside 1:1 complexes of MeβCeB. MeβMal. MeβLac, $Me\beta GlcNAc_2$, and SGP in H_2O. pH 7.4 at 298 K.

Based on the ascertained selectivity of receptor **2** for $Me\beta GlcNAc_2$, we have preliminarily attempted measurements of the binding affinity toward SGP by NMR techniques. Unfortunately, because of the strong signal overlap in the ^{1}H-NMR titration spectra (see **Figure S10**), we were not able to collect chemical shift data useful for analysis. Therefore, the binding ability of **2** toward SGP was measured by the ITC technique. ITC titrations were carried out in H_2O at pH 7.4, using neutralized solutions of both the reactants, and results are reported in **Table 1** and **Figure 3a**. Gratifyingly, a strong binding of **2** to SGP was observed, showing 170 μM affinity, very close to that observed for $Me\beta GlcNAc_2$. In contrast,

thermodynamic parameters were substantially different (see **Table 2**): while the free energy of binding is essentially the same with respect to MeβGlcNAc$_2$, the enthalpic contribution of SGP recognition is significantly reduced, whereas a positive entropy indicates an entropically driven binding. These data suggest a drastic change in the solvation features of SGP with respect to MeβGlcNAc$_2$.

The affinity of **2** for SGP measured by ITC was further confirmed by Circular Dichroism (CD) as a complementary technique. Because both, the UV chromophore **2** and the chiral SGP, are silent on CD, the technique is particularly convenient, because it monitors the unambiguous formation of the complex between the two partners. CD titrations were carried out on pre-neutralized reactants in H_2O at pH 7.4. Upon addition of SGP to receptor **2**, the CD intensity increased at 254.5, 260 and 265.5 nm, featuring negative sign for the first two wavelength and positive for the latter (**Figure 3b**). Nonlinear least-square regression analysis of CD data fitted a 1:1 stoichiometry model, giving $log\beta$ = 3.737 ± 0.003, corresponding to BC_{50}^{0} = 183 ± 1 μM, closely comparable to the affinity obtained by ITC. The highly diluted conditions used for the CD titration (63 μM) account for the different model, as higher stoichiometry complexes are undetectable.

It should be emphasized that recognition of *N*-glycans by biomimetic receptors is, to our knowledge, unprecedented in the literature and that the affinity achieved with receptor **2** is fully comparable to that of natural lectins.

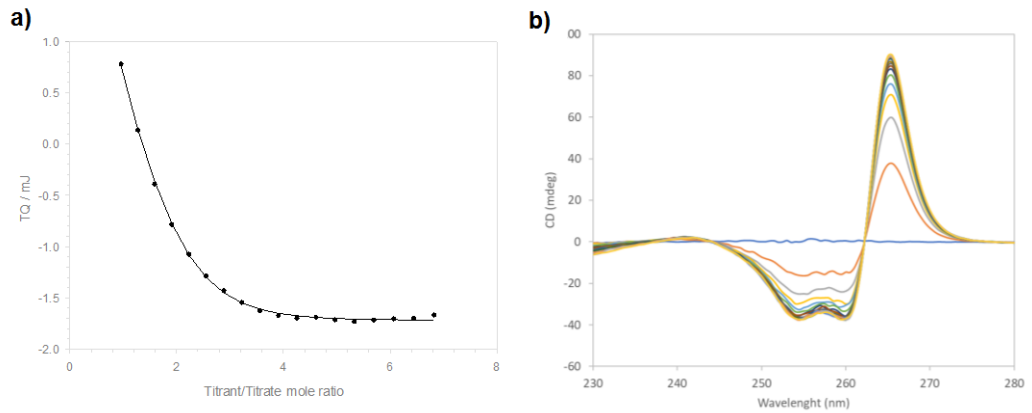


Figure 3. Titration results of **2** with SGP in H_2O (pH 7.4) at 298 K: a) ITC plot of experimental data points, [black circles - total heat (TQ) *vs* **2**/SGP mole ratio] and fitted curve [solid line] calculated by non-linear regression analysis; b) CD titration spectra of **2** (63 μM) with SGP

Prompted by these exciting results, we have carried out an investigation on the binding mode of **2** with SGP by combining NMR experimental data with molecular modeling calculations. The HSQC 2D-NMR experiments reported in **Figure 4** showed significant changes in the glycopeptide signals upon addition of an equimolar amount of receptor **2**. Indeed, binding induces a variation of chemical shift together with line broadening of proton signals of the guest, most likely due to complexation induced restrictions of conformational freedom. Despite this phenomenon prevents the use of chemical shift data for binding constant determination, it conveniently allows identifying the residues most directly involved in binding. As can be appreciated in **Figure 4**, it is immediately evident that signals of proton in proximity (saccharide c and (partially) d, and amino acids N, A, and K), and those belonging to the GlcNAc$_2$ disaccharide (residues a and b), disappear upon addition of **2**, whereas variations of chemical shift are detectable on protons belonging to residues d', d (partially) and on amino acid V. No variation in chemical shift was instead detected for apical residues e/e', f/f', g/g' and for N and C terminal amino acids K' and T. This behaviour strongly

suggests that recognition selectively occurs on the core $GlcNAc_2$ disaccharide, leaving terminal saccharide residues and distant amino acids unperturbed.

2D-NMR NOESY spectra carried out on the 1:1 mixture of **2** and SGP showed several NOE contacts, confirming that binding occurs on the $GlcNAc_2$ core (**Figure 5**). Similarly to what previously observed for **2** with $Me\beta GlcNAc_2$, the methyl of the N-acetyl group (CH_3-a) shows a NOE contact with the carbazole H-B proton. Moreover, the H-2-b of the $GlcNAc_2$ disaccharide shows a NOE contact with the anthracene H-F, clearly indicating that the disaccharide is located inside the receptor cleft between the two anthracene rings. Additional proximities were inferred by NOE contacts involving the H-D and H-E protons of the anthracene and the H-5-c of the central mannose residue, and with H-6-d of the 6-linked mannose residues, suggesting that the glycan extends outside the binding cleft of 2 beyond the anthracene moieties. Eventually, carbazole proton H-B shows a contact with H-5-c.

To give a three-dimensional description of the binding mode, computational studies are currently ongoing.

4.3 Conclusions

In conclusion, in the present work the structure of a recently reported receptor **1** has been optimized into a second-generation receptor **2**, in which the self-association phenomena of the progenitor have been substantially avoided, while leaving the remarkable binding properties toward $Me\beta GlcNAc_2$ unaffected. In contrast to macrocyclic architectures, the tweezers-shaped biomimetic receptor **2** demonstrated that an acyclic structure can effectively recognize the $GlcNAc_2$ core of an *N*-glycan with high selectivity. Because of the simple and easily accessible/modifiable structure, receptor **2** can hopefully become a new tool for glycoscience.

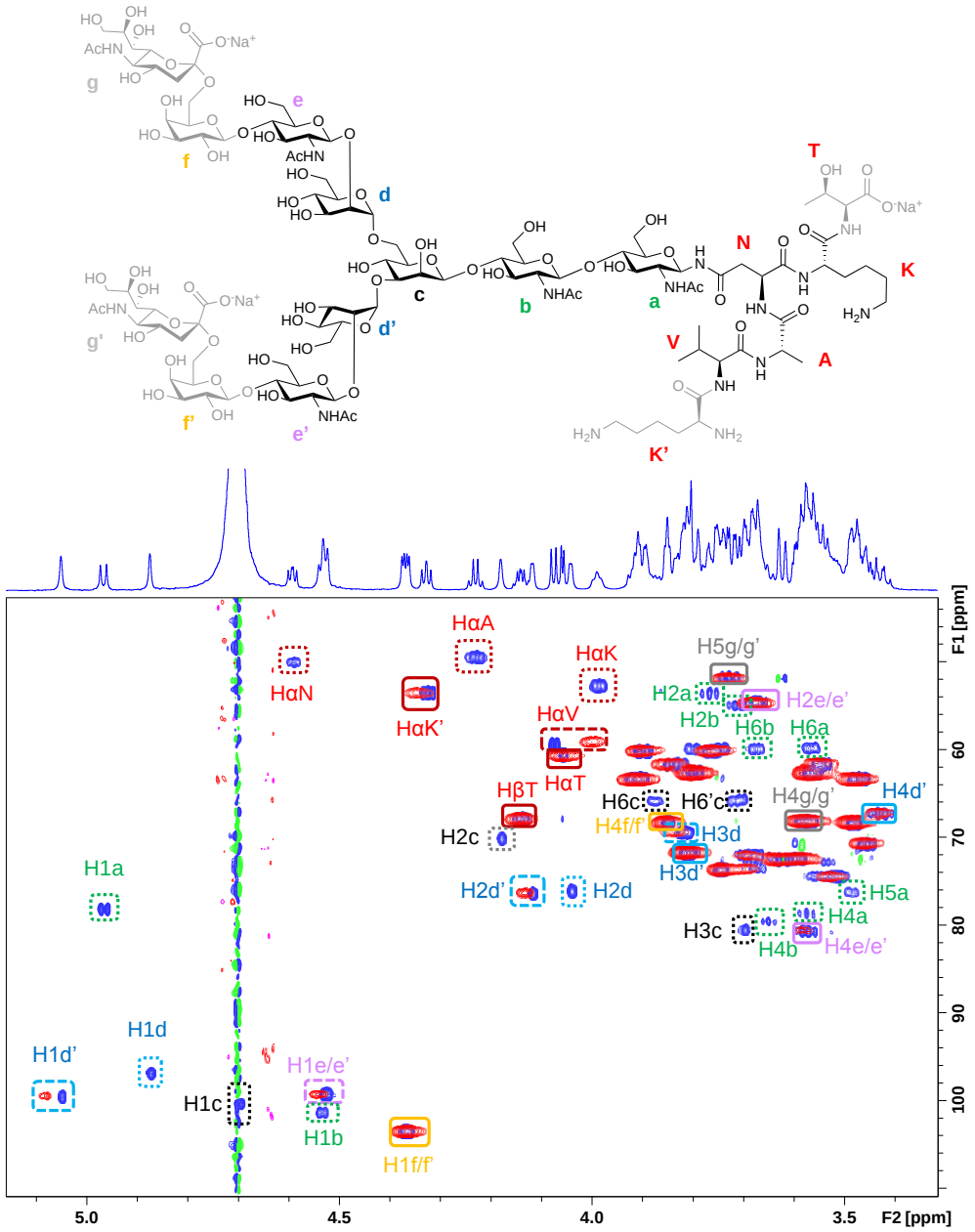


Figure 4. Superimposition of HSQC spectra of a 2 mM solution of SGP before (red peaks) and after (blue peaks) the addition of an equimolar amount of **2**. Selected peaks were squared and labelled: peaks that present chemical shift variation were squared using a dashed line, peaks that disappear after addition of **2** were squared using a dotted line.

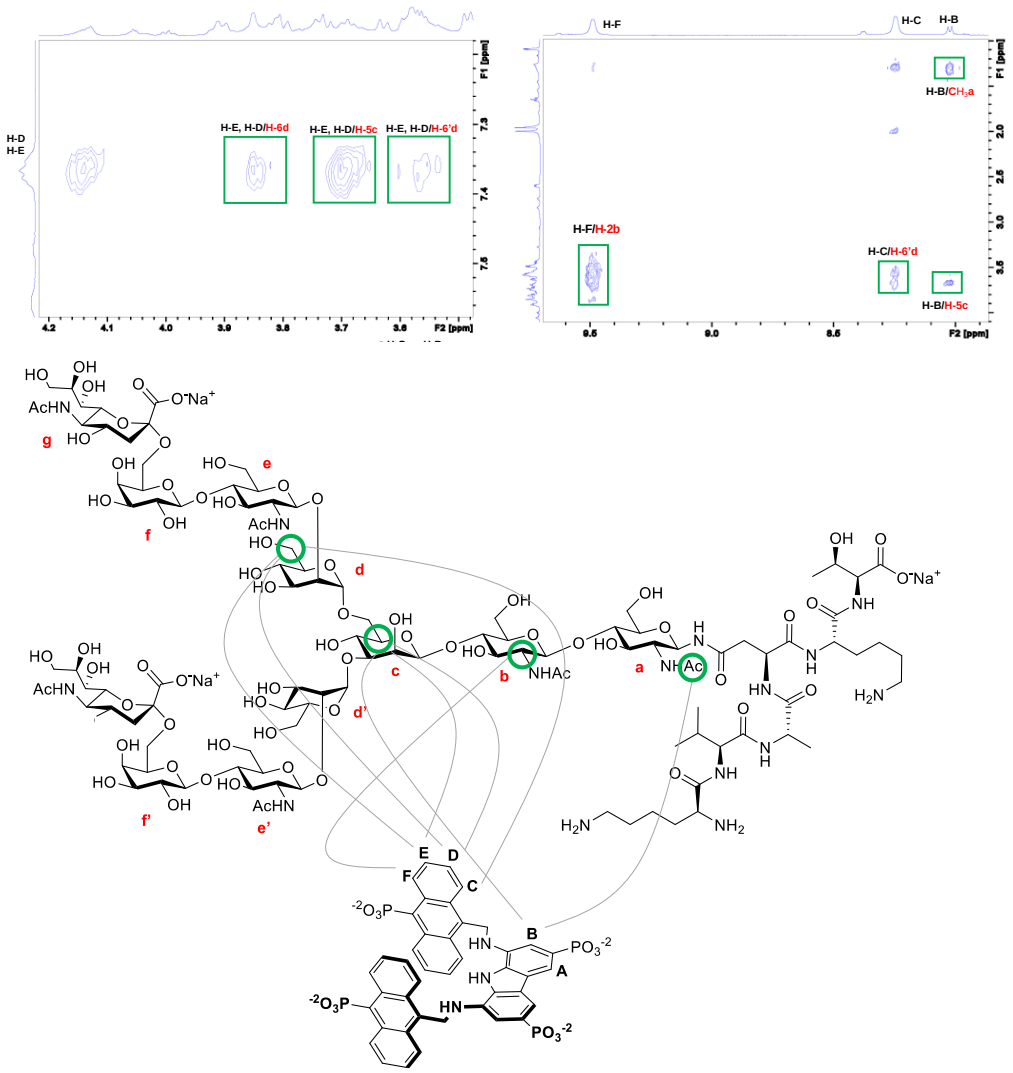


Figure 5. 800 MHz NOESY (500 ms) spectrum of an equimolar mixture of SGP and **2** (2 mM each) in D2O at 298 K. Unambiguous intermolecular NOE cross peaks are indicated by squares. A schematic representation of the intermolecular NOEs found between **2** and SGP is depicted by side.

4.4 Supporting information

4.4.1 Synthesis and characterization of chemical materials

General: ESI-MS analyses were performed in negative ion mode and were recorded on an LCQ-Fleet Ion Trap equipped with a standard Ionspray interface. HRMS were performed on an LTQ-IT-Orbitrap with a spray voltage of 2.10 kV and a resolution of 100000 (FWHM). Chemical shifts are reported in part per million (δ) relative to tetramethylsilane (TMS) for $CDCl_3$ and 4,4-Dimethyl-4-silapentane-1-sulfonic acid (DSS) for D_2O, using the residual solvent line as secondary internal reference (7.26 ppm for spectra run in $CDCl_3$ and 4.79 ppm for spectra run in D_2O). ^{13}C NMR spectra were obtained at 125 MHz in $CDCl_3$ and D_2O. Chemical shifts are reported in δ relative to TMS for $CDCl_3$, using the central solvent line as secondary internal reference at 77.16 ppm for $CDCl_3$ and reported in δ relative to DSS for D_2O.

Materials. Reagents were purchased from commercial suppliers and used without purification. Compound **10** was prepared according to known methods.[25]

Synthesis of 10-bromoanthracene-9-carbaldehyde (4): To a suspension of 9,10-dibromoanthracene **3** (4.00 g, 11.9 mmol) in 200 mL of dry THF, cooled at -78 ºC, *n*BuLi 1.6 M in hexane (9.0 mL, 14.3 mmol) was added dropwise under nitrogen atmosphere. The mixture was stirred for 30 minutes at -78 ºC and dry DMF was added (2.22 mL, 28.4 mmol). The mixture was let return slowly to RT and then diluted with 400 mL of water. Aqueous phase was the extracted with CH_2Cl_2 (2 x 300 mL) and combined organic layers washed with water (2 x 300 mL), dried over anhydrous $NaSO_4$, filtered and solvent evaporated, to give 3.18 g of the crude product that was purified by flash chromatography (CH_2Cl_2/petroleum ether 60:40; R_f = 0.58) to afford pure **4** as a yellow solid (2.06 mg, 61%). Mp: 200.8 - 201.7 ºC; ^1H NMR (500 MHz, $CDCl_3$, δ = 7.26): δ 11.51 (s, 1H); 8.91-8.89 (m, 2H); 8.70-8.68 (m, 2H); 7.73-7.66 (m, 4H); ^{13}C NMR (125 MHz, $CDCl_3$, δ = 77.16): δ 193.44; 132.11; 131.97; 130.46; 129.18; 129.07; 127.56; 125.88; 124.00

Synthesis of 2-(10-bromoanthracen-9-yl)-5,5-dimethyl-1,3-dioxane (5): A solution of **4** (2.11 g, 7.40 mmol), neopentyl glycol (2.08 g, 20.0 mmol) and a catalytic amount of *p*TsOH in 15 mL of toluene was refluxed for 5 h. The solution was cooled to RT and diluted with 40 mL of toluene and organic phase was washed with a saturated solution of $NaHCO_3$ (3 x 20 mL) and with water (3 x 20 mL). The organic layers were dried over anhydrous Na_2SO_4, filtered and the solvent evaporated to give 3.46 g of crude that was purified by flash chromatography (petroleum ether/ethyl acetate 90:10; R_f = 0.45) to afford pure **5** as a yellow solid (2.50 g, 90%). Mp: 148.8 - 149.6 ºC; ^1H NMR (500 MHz, $CDCl_3$, δ = 7,26): δ 8.86-8.84 (m, 2H); 8.62-8.60 (m, 2H); 7.60-7.55 (m, 4H); 6.76 (s, 1H); 4.02 (d, *J* = 11.5 Hz, 2H); 3.92 (d, *J* = 11.5 Hz, 2H); 1.63. (s, 3H); 0.95 (s, 3H); ^{13}C NMR (125 MHz, $CDCl_3$, δ = 77.16): δ 150.60; 130.59; 128.66; 128.58; 126.82; 126.33; 126.21; 125.41; 101.74; 79.50; 30.83; 22.56.

Synthesis of diethyl (10-(5,5-dimethyl-1,3-dioxan-2-yl)anthracen-9-yl)phosphonate (6): To a solution of **5** (1.6 g, 4.3 mmol) in 100 mL of dry THF, cooled at -78 ºC, *n*BuLi 1.6 M in hexane (3.00 mL, 4.80 mmol) was added dropwise under a nitrogen atmosphere. The solution was stirred for 30 minutes at -78 ºC and then heated to -40 ºC and stirred for other 30 minutes. The solution was cooled again to -78 ºC and $PO(OEt)_2Cl$ (1.90 mL, 13.1 mmol) was added dropwise. The mixture was let return slowly to RT and then diluted with 150 mL

of CH_2Cl_2. The organic phase was washed with a saturated solution of $NaHCO_3$ (3 x 100 mL) and water (3 x 100 mL). The organic layers were dried over anhydrous Na_2SO_4, filtered and the solvent evaporated to give 2.60 g of crude that was purified by flash chromatography (petroleum ether/ethyl acetate 50:50; R_f = 0.48) to afford pure **6** as a yellow oil (860 mg, 46%). ^{1}H NMR (500 MHz, $CDCl_3$, δ = 7,26): δ 9.37-9.35 (m, 2H); 8.87-8.85(m, 2H); 7.59-7.53 (m, 4H); 6.77 (s, 1H); 4.25-4.17 (m, 2H); 4.03 (d, *J* = 11.5 Hz, 2H); 4.00-3.95 (m, 2H); 3.92 (d, *J* = 11.5 Hz, 2H); 1.64 (s, 3H); 1.24 (t, *J* = 7.3 Hz, 6H); 0.95(s, 3H); ^{13}C NMR (125 MHz, $CDCl_3$, δ = 77.16): δ 135.14 (d, *J* = 11.2 Hz); 134.54 (d, *J* = 4.5 Hz); 129.42 (d, *J* = 15.2 Hz); 127.88 (d, *J* = 4.7 Hz); 126.64; 125.64; 125.28; 101.87; 79.48; 62.15 (d, *J* = 5.0 Hz); 30.91; 24.79; 22.55; 16.42 (d, *J* = 6.9 Hz); ^{31}P NMR (202 MHz, $CDCl_3$): δ 19.12; ESI-MS *m/z* (%): 451.50 (100) $[M+Na]^+$; 879.42 (37) $[2M+Na]^+$; 467.42 (23) $[M+K]^+$.

Synthesis of 9: A solution of **8** (236 mg, 0.460 mmol), **7** (389 mg, 1.00 mmol) and AcOH (0.230 mL, 4.00 mmol) in $CHCl_3$ (9 mL) was heated to reflux and stirred for 17 h. The solution was cooled to RT, neutralized with Et_3N and diluted with 40 mL of $CHCl_3$. The organic phase was washed with a saturated solution of $NaHCO_3$ (1 x 20 mL) and with water (1 x 20 mL) and the organic layers dried over anhydrous Na_2SO_4, filtered and the solvent evaporated to give 556 mg of crude as an orange solid. To the solid a freshly prepared suspension of $NaBH_4$ (226 mg, 5.97 mmol) in 10 mL of methanol was added. The mixture was stirred for 1.5 h at RT then diluted with 65 mL of $CHCl_3$. The organic phase was washed with water (2 x 30 mL), with brine (2 x 30 mL) and the organic layers dried over anhydrous Na_2SO_4, filtered and the solvent evaporated to give 456 mg of crude that was purified by flash chromatography ($CHCl_3$/THF 50:50; R_f = 0.43) to afford pure **9** as a yellow solid (300 mg, 53%). Mp: >170 ºC (dec.); ^{1}H NMR (500 MHz, $CDCl_3$ + 5% MeOH, δ = 7,26): δ 10.7 (s, 1H); 9.17 (d, *J* = 8.9 Hz, 4H); 8.22 (d, *J* = 8.9 Hz, 4H); 8.09 (d, *J* = 14.2 Hz, 4H); 7.44-7.36 (m, 10H); 5.28 (s, 4H); 4.26-4.09 (m, 12H); 3.96-3.88 (m, 4H); 1.40 (t, *J* = 7.0 Hz, 12H); 1.19 (t, *J* = 7.2 Hz, 12H); ^{13}C NMR (125 MHz, $CDCl_3$, δ = 77.16): δ 136.06 (d, *J* = 4.9 Hz); 134.70 (d, *J* = 18.8 Hz); 134.28 (d, *J* = 11.3 Hz); 131.63 (d, *J* = 2.7 Hz); 129.99 (d, *J* = 15.2 Hz); 127.55; 126.67; 126.17; 124.42; 123.97 (d, *J* = 20.2 Hz); 120.53; 119.02; 116.18(d, *J* = 10.2 Hz); 108.65 (d, *J* = 14.0 Hz); 62.24 (d, *J* = 5.2 Hz); 61.94 (d, *J* = 5.2 Hz); 42.20; 16.64 (d, *J* = 6.7 Hz); 16.12 (d, *J* = 6.3 Hz); ^{31}P NMR (202 MHz, CDCl3): δ 22.36; 19.12 ESI-MS *m/z* (%): 1144.75 (100) $[M+Na]^+$; 584.05 (64) $[M+2Na]^{2+}$; 1160.67 (60) $[M+K]^+$.

Synthesis of receptor 2: To a solution of **9** (301 mg, 0.268 mmol) in 1.8 mL of dry CH_2Cl_2, Et_3N (0.610 mL, 4.28 mmol) was added under a nitrogen atmosphere. The solution was cooled to 0 °C and TMSBr (0.560 mL, 4.28 mmol) was slowly added. The solution was stirred at room temperature for 22 h, then diluted with 10 mL of CH_2Cl_2 and cooled to 0 °C. After addition of 9 mL of MeOH, the solution was stirred at room temperature for 30 minutes and the solvent evaporated to give 357 mg of crude **3** as a pale brown solid. The solid was suspended in MeOH (5 mL x 2) and the suspension centrifugated at 3000 rpm for 20 minutes. The liquids were removed and the solid was suspended in HCl 0.5 M (5 mL x2). After centrifugation at 3000 rpm for 20 minutes, liquids were removed and the obtained solid dried under high vacuum to obtain pure **2** (147 mg, 0.102 mmol, 61%) in its acid form as a yellow powder. Mp: >150 ºC (dec.); ^{1}H NMR (500 MHz, D_2O, DSS as internal reference, δ = 0.0): δ 9.46 (d, *J* = 8.8 Hz, 4H, C*H*-17); 8.32 (d, *J* = 8.4 Hz, 4H, C*H*-14); 8.12 (d, *J* = 12.9 Hz, 2H, C*H*-4); 7.60 (d, *J* = 12.3 Hz, 2H, C*H*-2); 7.48-7.42 (m, 8H, C*H*-15, C*H*-16); 5.39 (s, 4H, CH_2-11); ^{13}C NMR (125 MHz, D_2O): δ 135.16; 133.88; 133.61 (d, *J* = 16.3 Hz); 132.76 (d, *J* = 9.2 Hz); 132.15 (d, *J* = 4.1 Hz); 130.46 (d, *J* = 2.6 Hz); 129.76 (d, *J* = 7.3 Hz); 129.68 (s, *C*H-17); 125.81 (s, *C*H-16); 124.69 (s, *C*H-15); 123.95 (s, CH-14); 123.02 (d, *J* = 17.5 Hz); 114.19 (d, *J* = 10.0 Hz, *C*H-4); 110.85 (d, *J* = 12.2 Hz, *C*H-2); 41.69 (s, *C*H_2-11); ^{31}P

NMR (202 MHz, D_2O): δ 14.33; 9.77; ESI-MS *m/z* (%): 896.42 (100) $[M-H]^-$, 447.83 (25) $[M-2H]^{2-}$, 918.42 (24) $[M-2H+Na]^-$. HRMS (*m/z*): $[M-H]^-$ calcd. for $C_{42}H_{34}N_3O_{12}P_4$, 896.10984; found, 896.10669.

4.4.2 Binding studies

NMR preliminary screening: Preliminary screenings (298 K, 500 MHz) were performed in D_2O at pD 7.4 in presence of DSS as internal reference. Solution of reducing carbohydrates were prepared in D_2O and kept overnight at room temperature before the screening experiments, to ensure equilibration of the anomers. The spectra of the free sugars at 1 mM concentration were compared to the spectra of the equimolar mixture of sugars with receptor **2** (1 mM each) and chemical shift differences were evaluated.

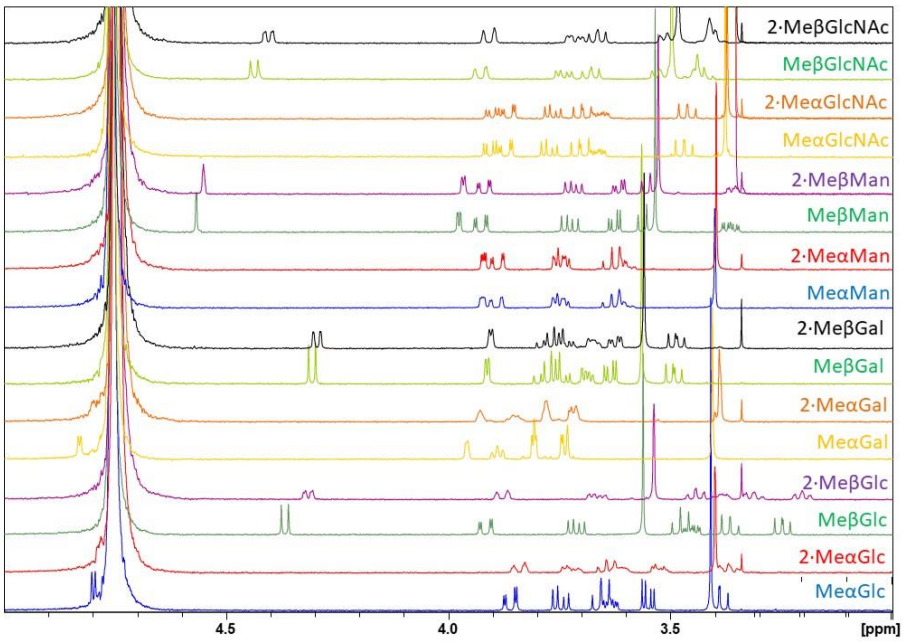


Figure S1. ^1H NMR spectra (500 MHz, D_2O) of a 1 mM solution of methyl glycosides and an equimolar mixture of methyl glycosides and **2** (1 mM each) at pD 7.4. Variations of the H-1 proton signal shifts of MeβGlc is $\Delta\delta$ = 0.05 ppm.

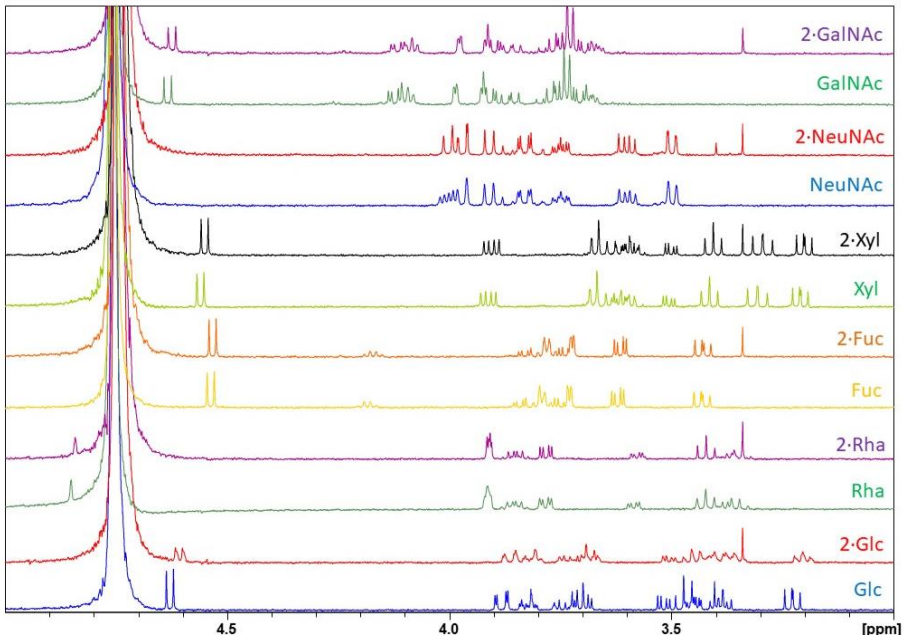


Figure S2. ^{1}H NMR spectra (500 MHz, D_2O) of a 1 mM solution of monosaccharides and an equimolar mixture of monosaccharides and **2** (1 mM each) at pD 7.4.

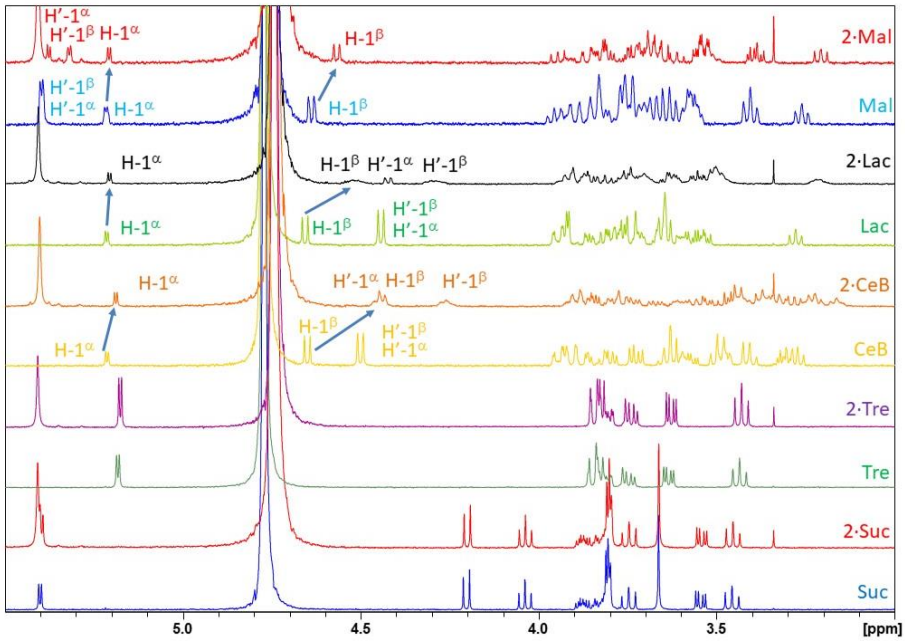


Figure S3. ^{1}H NMR spectra (500 MHz, D_2O) of a 1 mM solution of disaccharides and an equimolar mixture of disaccharides and **2** (1 mM each) at pD 7.4. Variations of the H-1 proton signal shifts of CeB, Lac, Mal are $\Delta\delta$ = 0.02, 0.01, 0.01 ppm for the α and $\Delta\delta$ = 0.21, 0.14, 0.07 ppm for the β anomer respectively.

NMR titrations and data analysis: Titrations were performed at 298 K, 500 MHz in 5 mm NMR tubes using microsyringes, following a previously described technique.[44] The stock solutions of **2** were prepared in D_2O adjusting the pD to a value of 7.4 with a diluted solution of NaOH. A correction factor of +0.4 was applied to the pH values measured by the pH meter to determine the pD values (pD = pH + 0.4). DSS was used as internal reference. A dilution experiment of the free receptor **2** was performed to independently measure the dimerization constant at pD 7.4, which was set invariant in the non-linear regression analysis of receptor-glycosides binding data. To avoid any ambiguities in the definition of the equilibrium model with receptor **2**, independent titrations were performed, when appropriate, at significantly different receptor concentrations and all set of data were simultaneously fitted through a nonlinear least-square regression analysis, including in the fit all the available signals from both reactants. Mathematical analysis of data and graphic presentation of results was performed using the HypNMR 2006 program.[45] Results pages and Plots of experimental and calculated shifts are reported hereafter.

Dilution of receptor 2 (D_2O, pD 7.4, 298 K, 500 MHz)

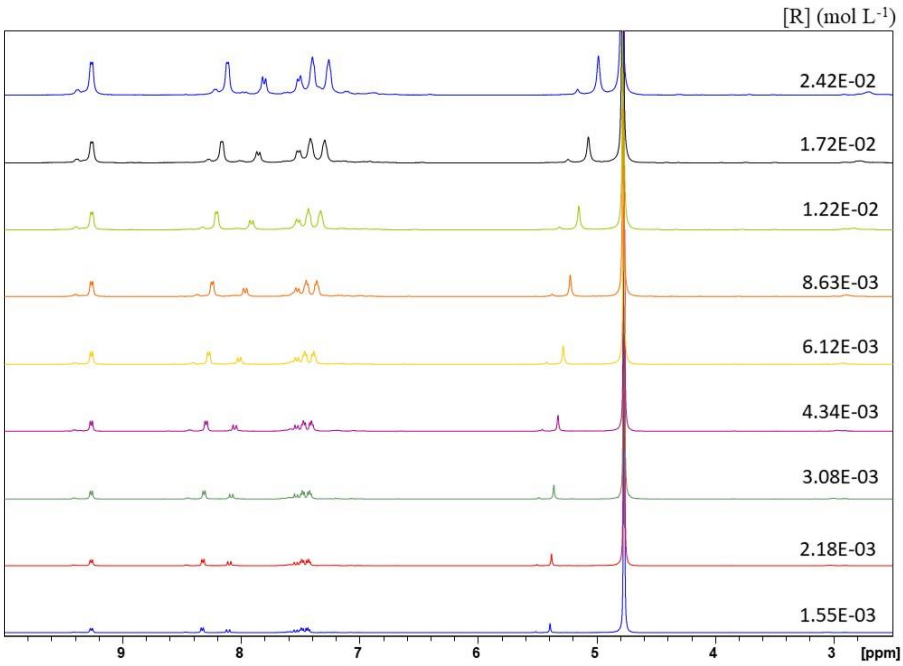


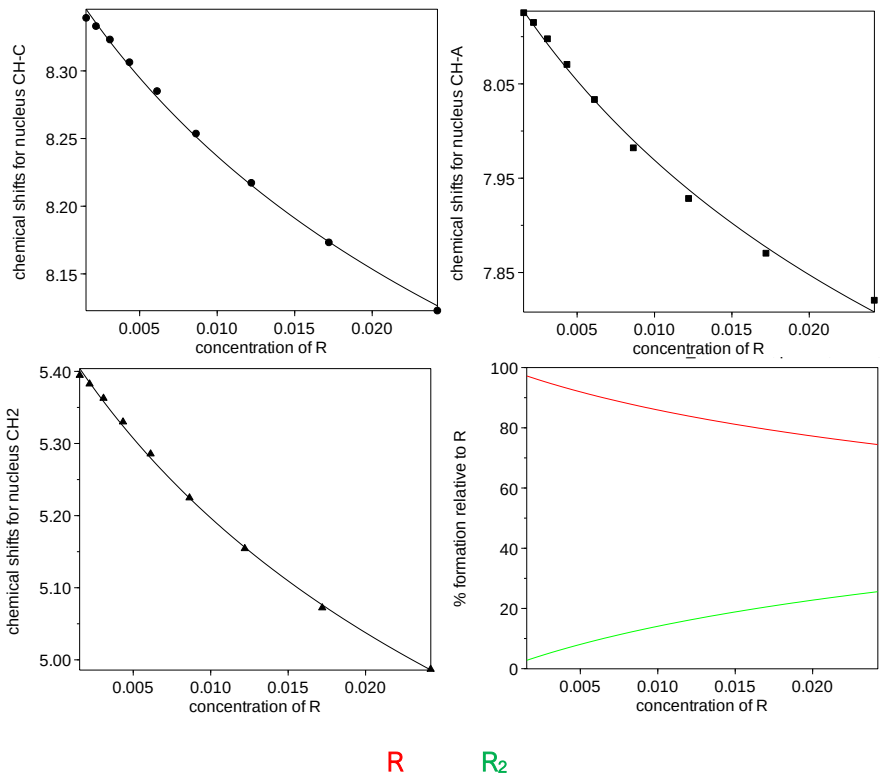
Figure S4. ^{1}H NMR spectroscopic spectra (500 MHz, D_2O, pD 7.4, 298 K) of receptor **2** (R) at different concentrations in dilution expertiment.

Results page

```
 no. of spectra             9
 no. of resonance values   27
 no. of resonant nuclei     3

   sigma =   0.00591425500          RMS weighted residual =   0.00509018025

      stoich          value   relative    log    standard
      coeff                   std devn    beta   deviation

 Beta  2 refined     9.5341    0.1224    0.9793   0.0532  (   R2  )

Individual chemical shifts

                    R                          2
 ============================================================
        +        value    error         value    error
CH-C    +       8.3724   0.0043         7.4103   0.0720
CH-A    +       8.1662   0.0049         6.7660   0.1018
CH2     +       5.4551   0.0056         3.6205   0.1319
```

Titration Plots

Chemical shifts (δ, ppm) *vs*. concentration of R (mol L^{-1})
experimental (symbols) and calculated (lines) values



2 + MeβGlc (D_2O. pD 7.4, 298 K, 500 MHz).

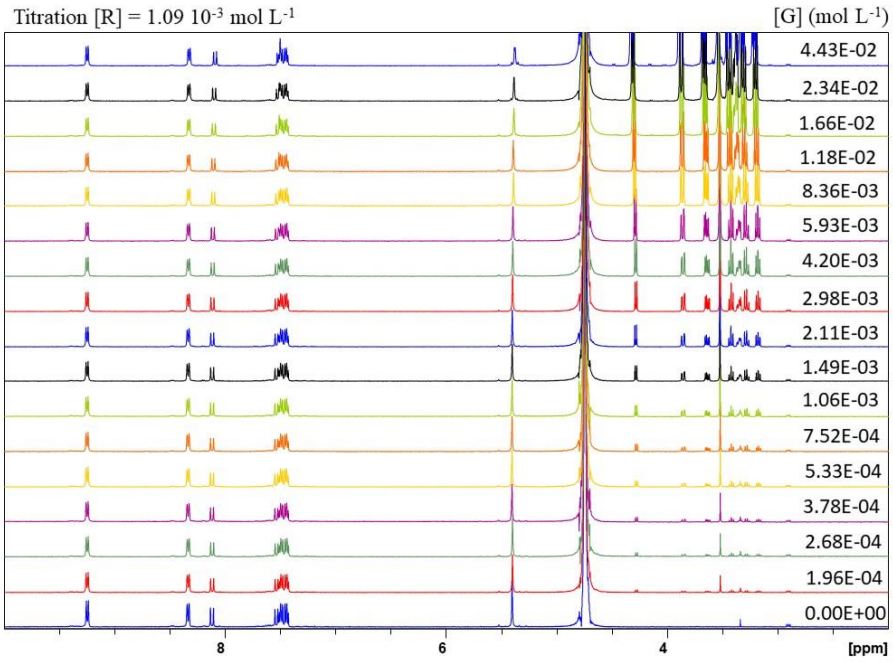


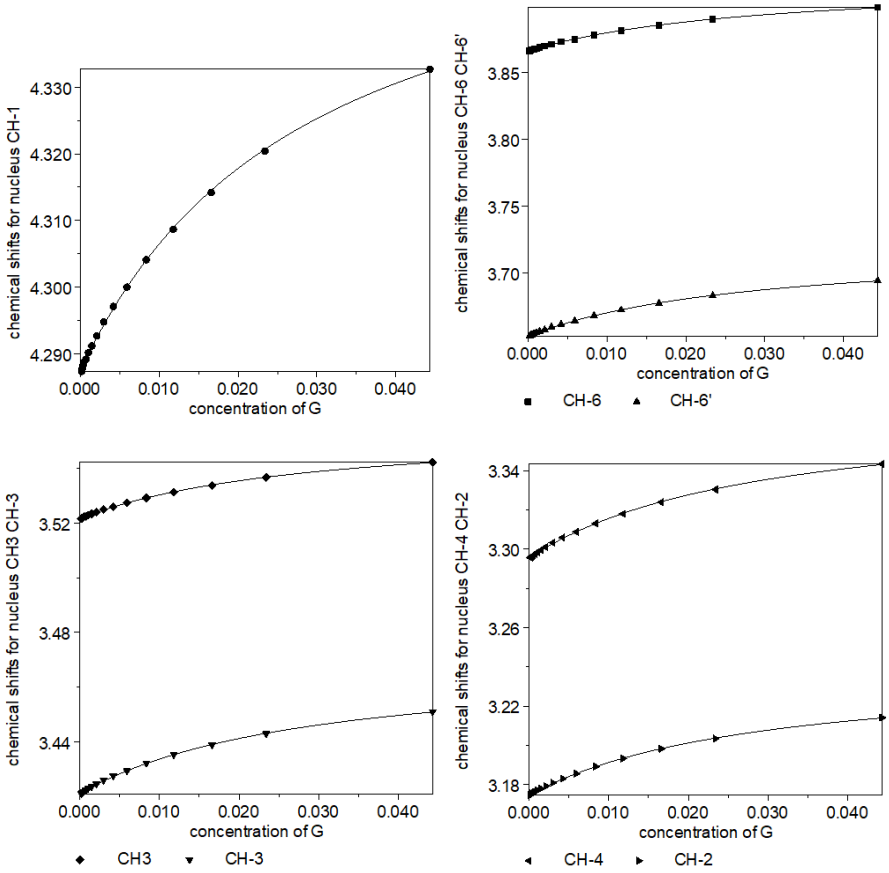
Figure S5. ^1H NMR spectroscopic titration (500 MHz, D_2O, pD 7.4, 298 K) of receptor **2** (R) with incremental concentrations of MeβGlc (G)

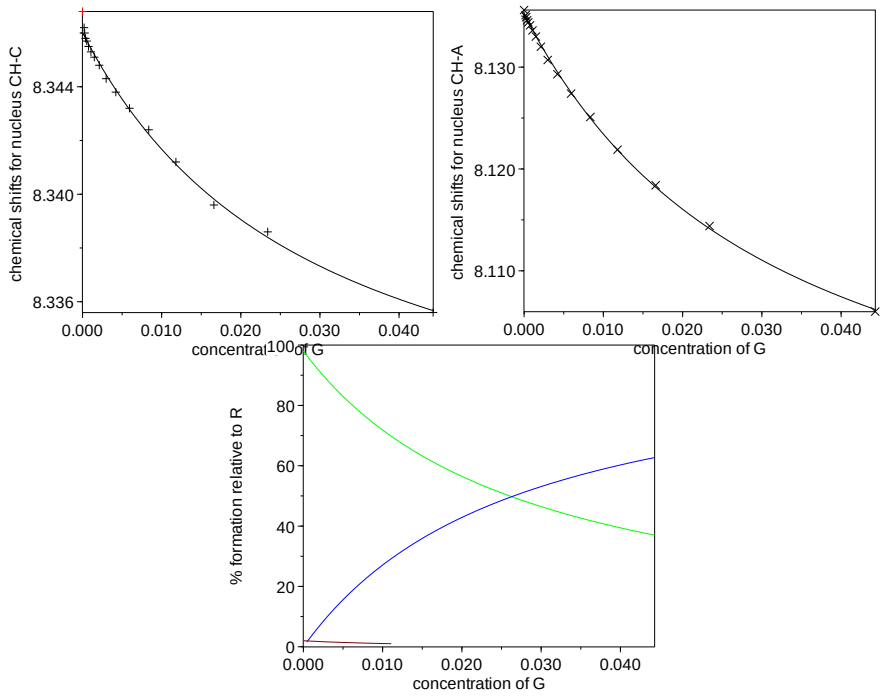
Results page

```
 no. of spectra            17
 no. of resonance values  145
 no. of resonant nuclei     9

Chi-squared =   4.57

   sigma =   0.00023165762          RMS weighted residual =   0.00021422661

       stoich            value   relative    log    standard
       coeff                     std devn    beta   deviation

 Beta  1  1 refined  3.8855E+001  0.0127    1.5894   0.0055  (   GR   )
 Beta  0  2 constant    9.5345              0.9793            (   R2   )

Individual chemical shifts

                     G                        R
 ============================================================
        +        value    error         value    error
CH-1    +       4.3608   0.0005
CH-6    +       3.9189   0.0004
CH-6'   +       3.7200   0.0005
CH3     +       3.5552   0.0003
CH-3    +       3.4697   0.0004
CH-4    +       3.3732   0.0005
CH-2    +       3.2388   0.0005
CH-C    +                                8.3474   0.0019
CH-A    +                                8.1433   0.0019
        +
                    1,1                      0,2
 ============================================================
        +        value    error         value    error
CH-1    +       2.5200   0.0333
CH-6    +       2.5930   0.0246
CH-6'   +       2.0279   0.0308
CH3     +       2.7145   0.0168
CH-3    +       2.2464   0.0229
CH-4    +       1.4242   0.0352
CH-2    +       1.6295   0.0294
CH-C    +       8.3290   0.0009         8.2769   0.0941
CH-A    +       8.0860   0.0009         7.7394   0.0946
```

Titration Plots

Chemical shifts (δ. ppm) *vs*. concentration of G (mol L^{-1})
experimental (symbols) and calculated (lines) values




chemical shifts for nucleus CH-C
8.344
8.340
8.336
0.000
0.010
0.020
0.030
0.040
concentration of G
chemical shifts for nucleus CH-A
8.130
8.120
8.110
0.000
0.010
0.020
0.030
0.040
concentration of G
% formation relative to R
100
80
60
40
20
0
0.000
0.010
0.020
0.030
0.040
concentration of G
R
GR
R2

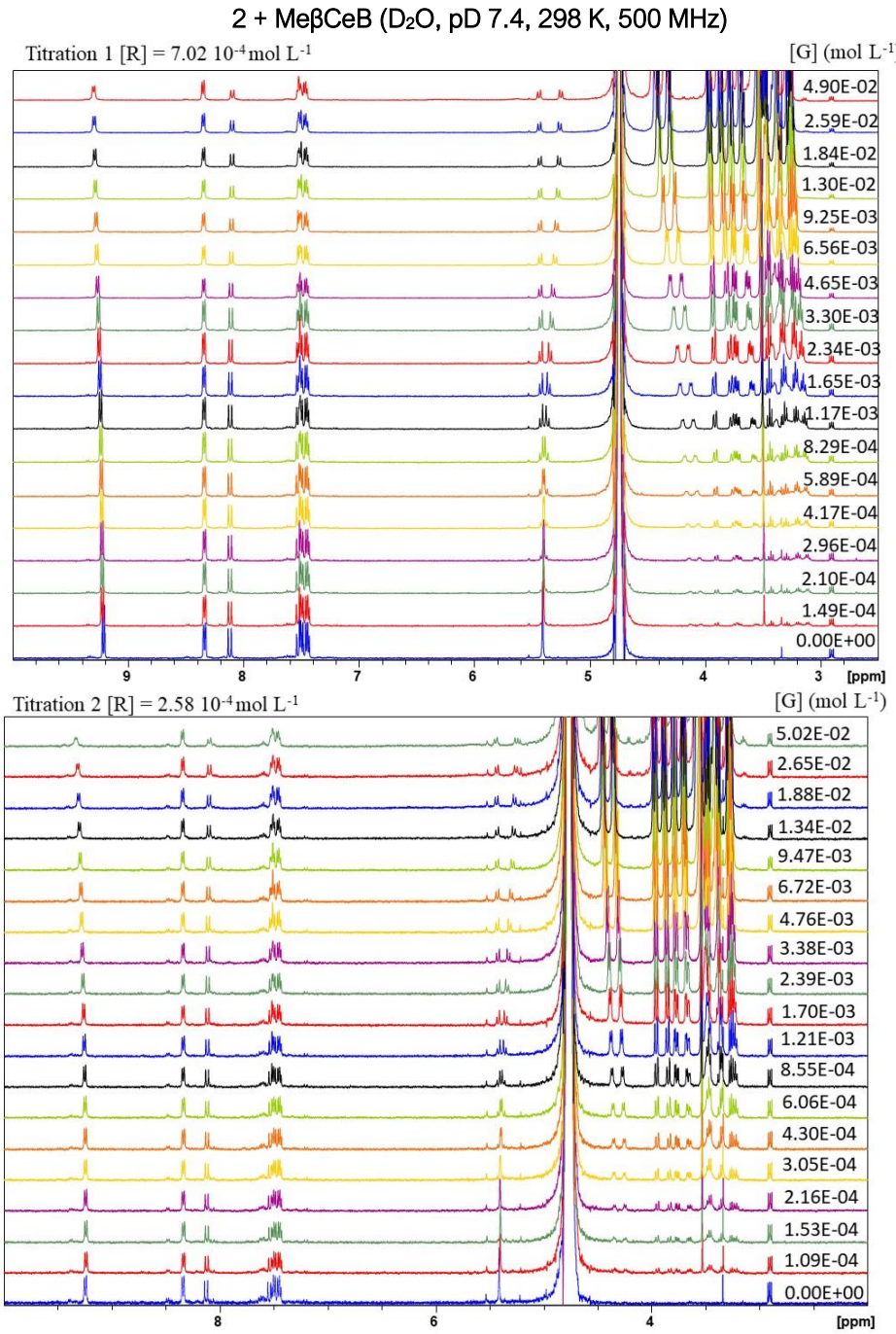


Figure S6. ^{1}H NMR spectroscopic titrations (500 MHz, D_2O, pD 7.4, 298 K) of receptor **2** (R) with incremental concentrations of MeβCeB (G).

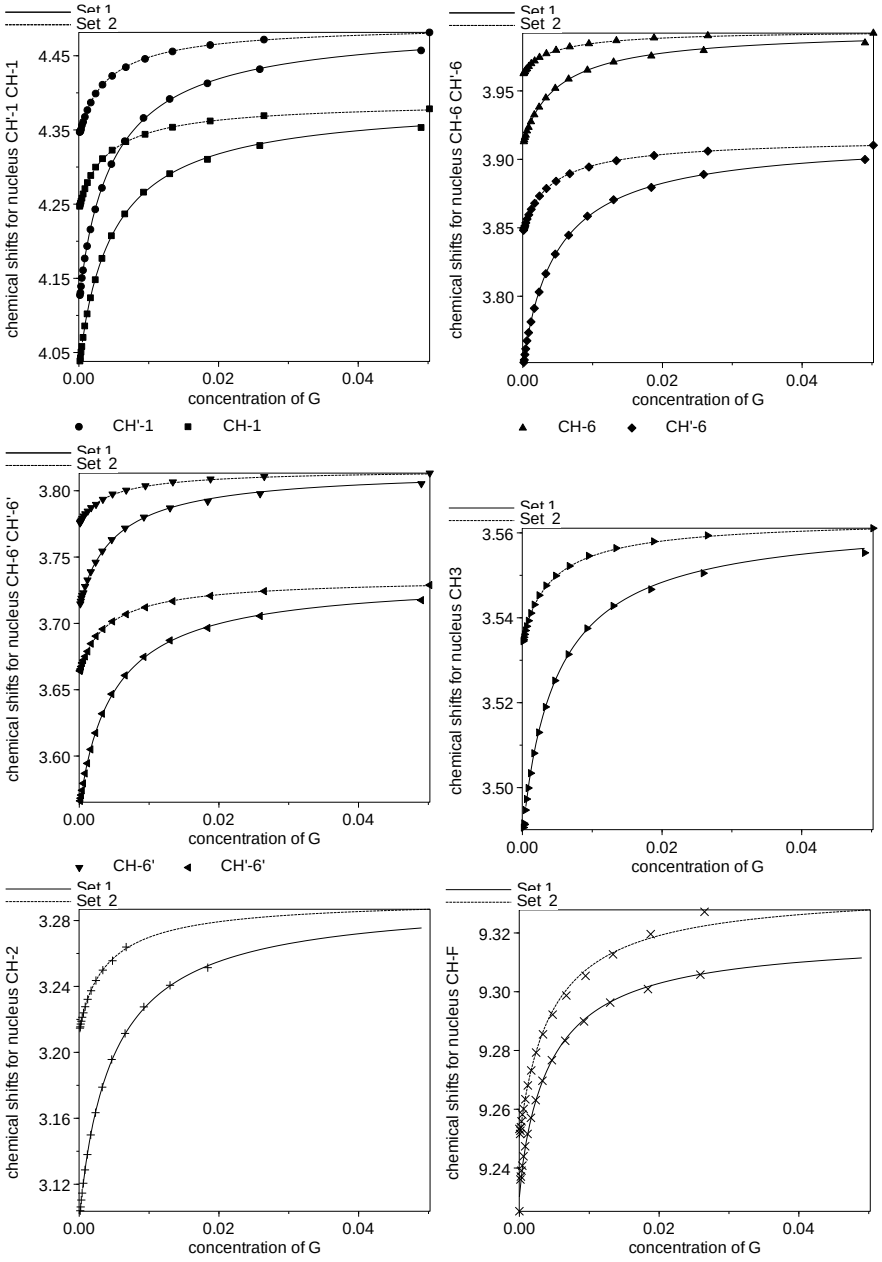
Results page

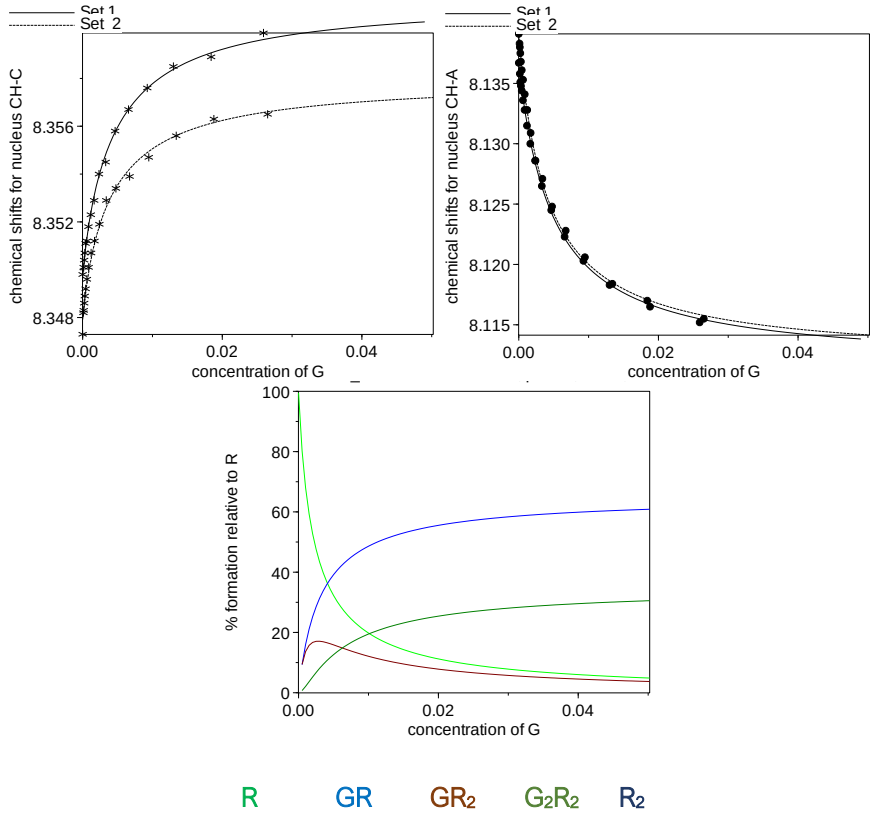
```
 no. of spectra            37
 no. of resonance values  378
 no. of resonant nuclei    11

   sigma =   0.00078399674          RMS weighted residual =   0.00073030669

       stoich            value   relative    log    standard
       coeff                     std devn    beta   deviation

 Beta  1  1 refined  2.5027E+002  0.0281    2.3984   0.0122  (   GR   )
 Beta  1  2 refined  6.0952E+005  0.1525    5.7850   0.0662  (   GR2  )
 Beta  2  2 refined  1.0001E+008  0.1979    8.0001   0.0860  (   G2R2 )
 Beta  0  2 constant    9.5345              0.9793           (   R2   )

Individual chemical shifts
                     G                          R
 ============================================================
        +        value    error          value    error
CH'-1   +       4.4920   0.0005
CH-1    +       4.3882   0.0005
CH-6    +       3.9943   0.0003
CH'-6   +       3.9154   0.0004
CH-6'   +       3.8160   0.0003
CH'-6'  +       3.7340   0.0004
CH3     +       3.5633   0.0003
CH-2    +       3.2930   0.0007
CH-F    +                                 9.2624   0.0006
CH-C    +                                 8.3468   0.0006
CH-A    +                                 8.1407   0.0006
        +
                    1,1                        1,2
 ============================================================
        +        value    error          value    error
CH'-1   +       2.2092   0.0587          3.9454   0.0143
CH-1    +       2.2156   0.0560          3.8552   0.0147
CH-6    +       3.5278   0.0177          3.8479   0.0103
CH'-6   +       2.8622   0.0294          3.6892   0.0092
CH-6'   +       3.2087   0.0202          3.6516   0.0099
CH'-6'  +       2.6231   0.0308          3.5175   0.0090
CH3     +       3.1374   0.0170          3.4349   0.0098
CH-2    +       2.0541   0.0391          3.0436   0.0125
CH-F    +       9.3682   0.0050          9.2033   0.0029
CH-C    +       8.3511   0.0016          8.3587   0.0016
CH-A    +       8.1129   0.0015          8.1308   0.0016
        +
                    2,2                        0,2
 ============================================================
        +        value    error          value    error
CH'-1   +       1.5916   0.0862
CH-1    +       1.5592   0.0871
CH-6    +       3.2716   0.0307
CH'-6   +       2.6382   0.0391
CH-6'   +       2.9418   0.0341
CH'-6'  +       2.4141   0.0401
CH3     +       2.9331   0.0269
CH-2    +       1.8463   0.0420
CH-F    +       9.2731   0.0032          6.7872   0.0742
CH-C    +       8.3709   0.0024          8.5692   0.0662
CH-A    +       8.1104   0.0023          7.7992   0.0665

 Correlation coefficients*1000
```

```
      1    2    3
  1
  2 -503
  3 -508  995

 Parameters are numbered as follows
    1 beta 1,1
    2 beta 1,2
    3 beta 2,2
```

Titration Plots

Chemical shifts (δ, ppm) *vs*. concentration of G (mol L^{-1})
experimental (symbols) and calculated (lines) values




Set 1
Set 2
chemical shifts for nucleus CH-C
8.356
8.352
8.348
0.00
0.02
0.04
concentration of G
chemical shifts for nucleus CH-A
8.135
8.130
8.125
8.120
8.115
% formation relative to R
100
80
60
40
20
0
R
GR
GR2
G2R2
R2

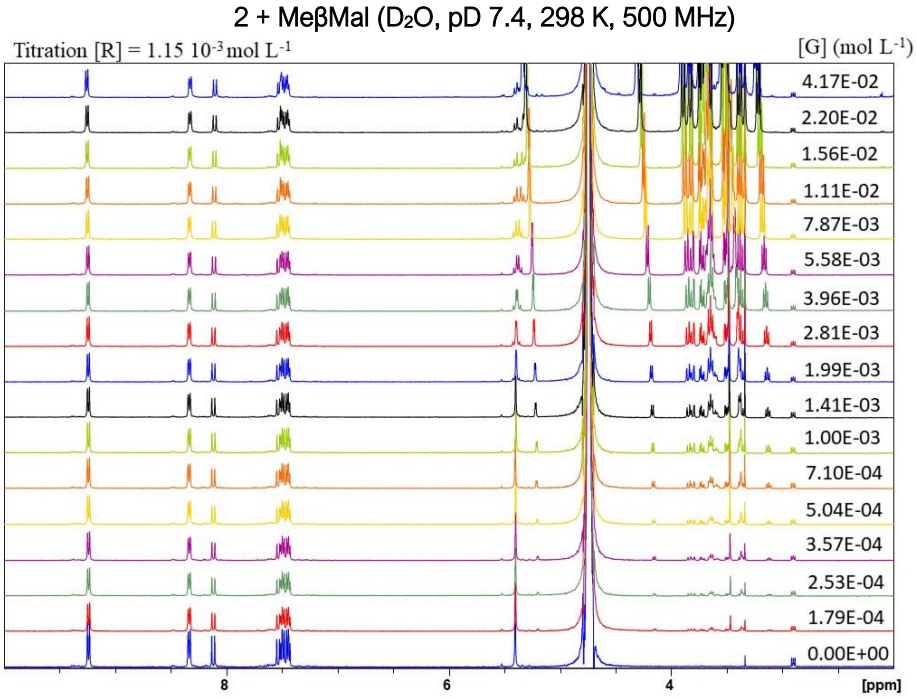


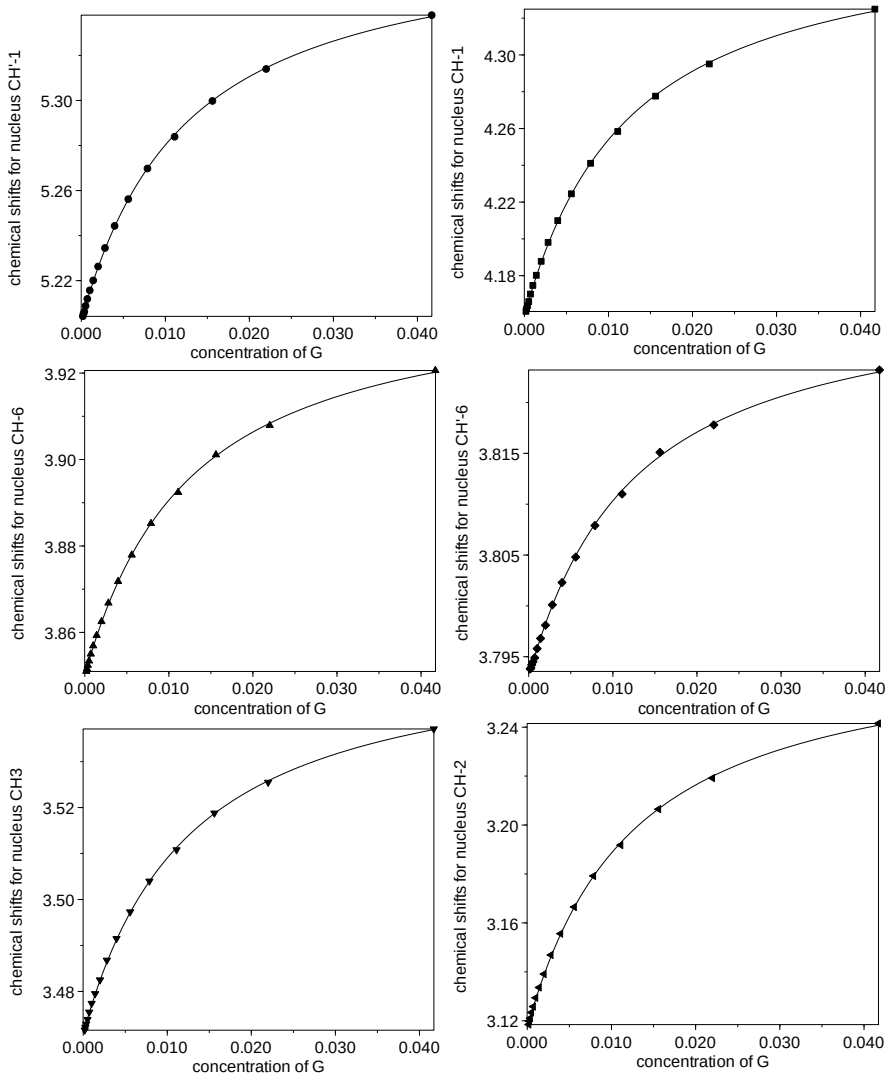
Figure S7. ^{1}H NMR spectroscopic titration (500 MHz, D_2O, pD 7.4, 298 K) of receptor **2** (R) with incremental concentrations of MeβMal (G).

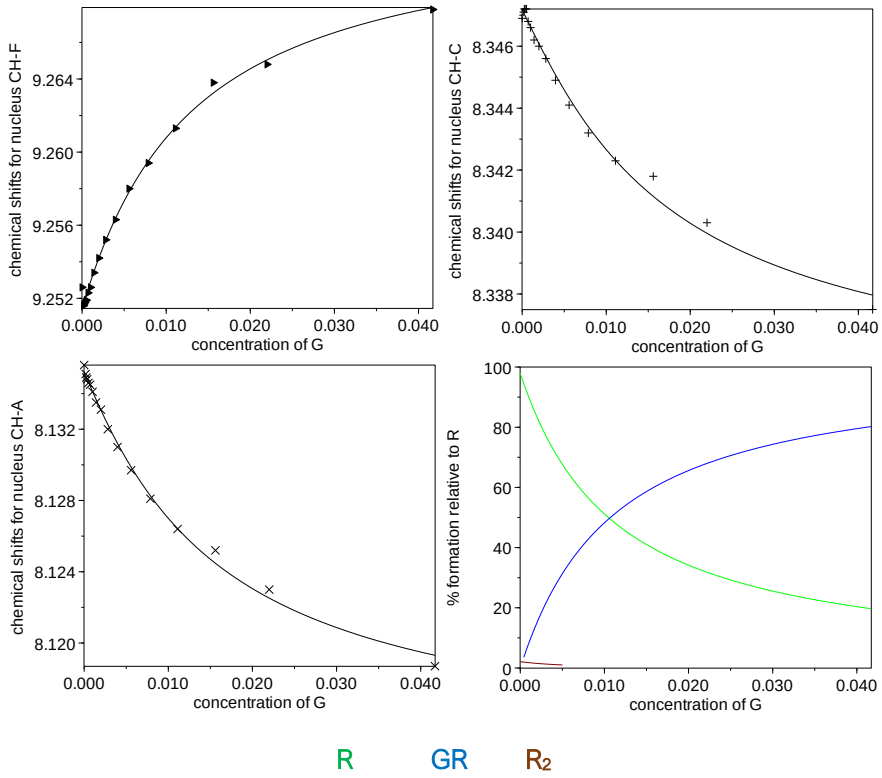
Results page

```
 no. of spectra            17
 no. of resonance values  147
 no. of resonant nuclei     9

Chi-squared =   6.31

   sigma =   0.00045931299          RMS weighted residual =   0.00042355037

       stoich            value   relative    log    standard
       coeff                     std devn    beta   deviation

 Beta  1  1 refined  9.9792E+001  0.0082    1.9991   0.0036  (   GR   )
 Beta  0  2 constant    9.5345              0.9793           (   R2   )

Individual chemical shifts
                    G                          R
 ========================================================
        +        value    error          value    error
CH'-1   +       5.3750   0.0006
CH-1    +       4.3701   0.0007
CH-6    +       3.9399   0.0004
CH'-6   +       3.8314   0.0004
CH3     +       3.5555   0.0004
CH-2    +       3.2754   0.0005
CH-F    +                                9.2476   0.0021
CH-C    +                                8.3540   0.0021
CH-A    +                                8.1442   0.0021
        +
                  1,1                        0,2
 ========================================================
        +        value    error          value    error
CH'-1   +       3.6664   0.0164
CH-1    +       2.2751   0.0199
CH-6    +       3.0514   0.0093
CH'-6   +       3.4516   0.0056
CH3     +       2.7181   0.0089
CH-2    +       1.7110   0.0152
CH-F    +       9.2727   0.0008          9.4288   0.1044
CH-C    +       8.3343   0.0008          8.0269   0.1042
CH-A    +       8.1136   0.0008          7.7150   0.1044
```

Titration Plots

Chemical shifts (δ, ppm) *vs*. concentration of G (mol L^{-1})
experimental (symbols) and calculated (lines) values




chemical shifts for nucleus CH-F
9.264
9.260
9.256
9.252
0.000
0.010
0.020
0.030
0.040
concentration of G
chemical shifts for nucleus CH-C
8.346
8.344
8.342
8.340
8.338
0.000
0.010
0.020
0.030
0.040
concentration of G
chemical shifts for nucleus CH-A
8.132
8.128
8.124
8.120
0.000
0.010
0.020
0.030
0.040
concentration of G
% formation relative to R
100
80
60
40
20
0
0.000
0.010
0.020
0.030
0.040
concentration of G
R
GR
R2

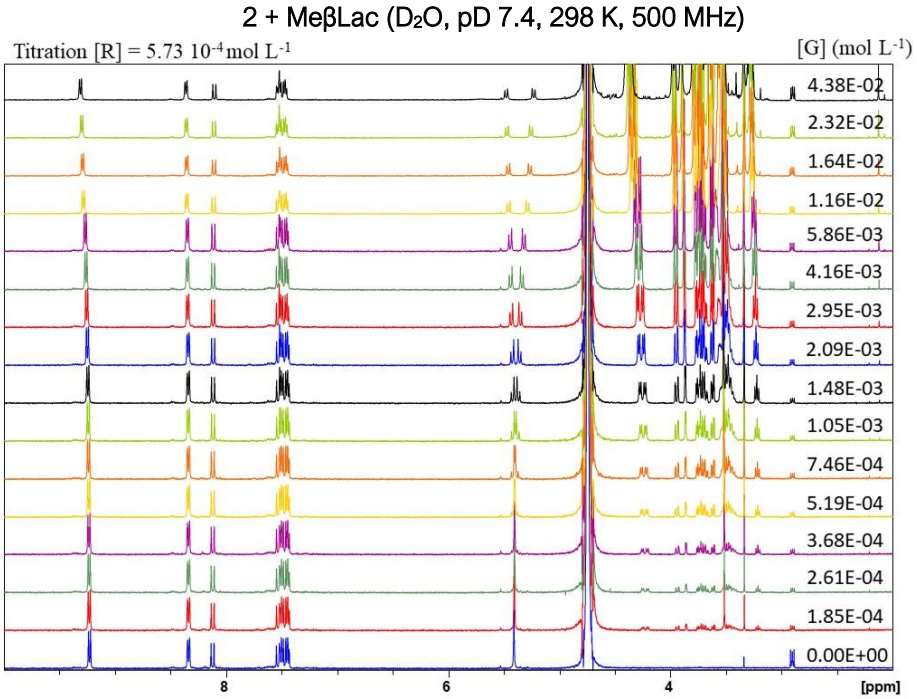


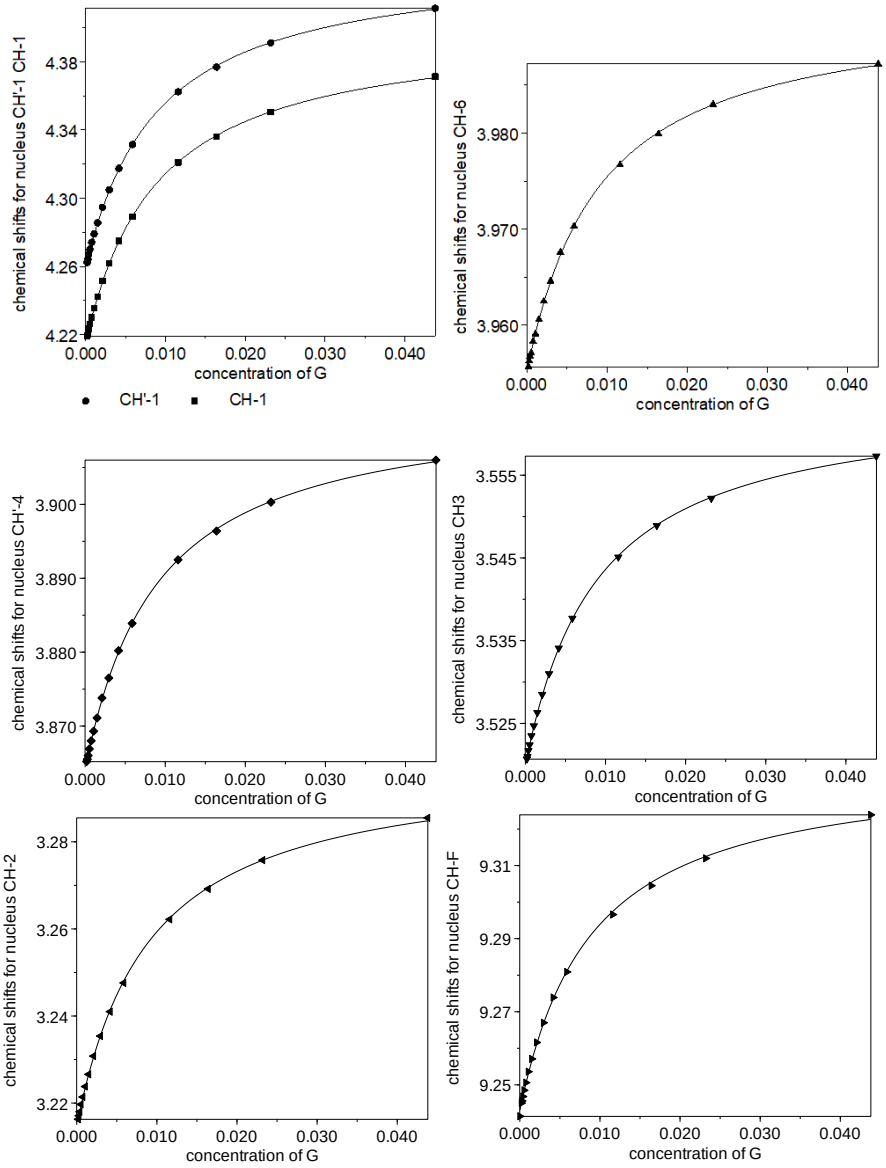
Figure S8. ^{1}H NMR spectroscopic titration (500 MHz, D_2O, pD 7.4, 298 K) of receptor **2** (R) with incremental concentrations of MeβLac (G).

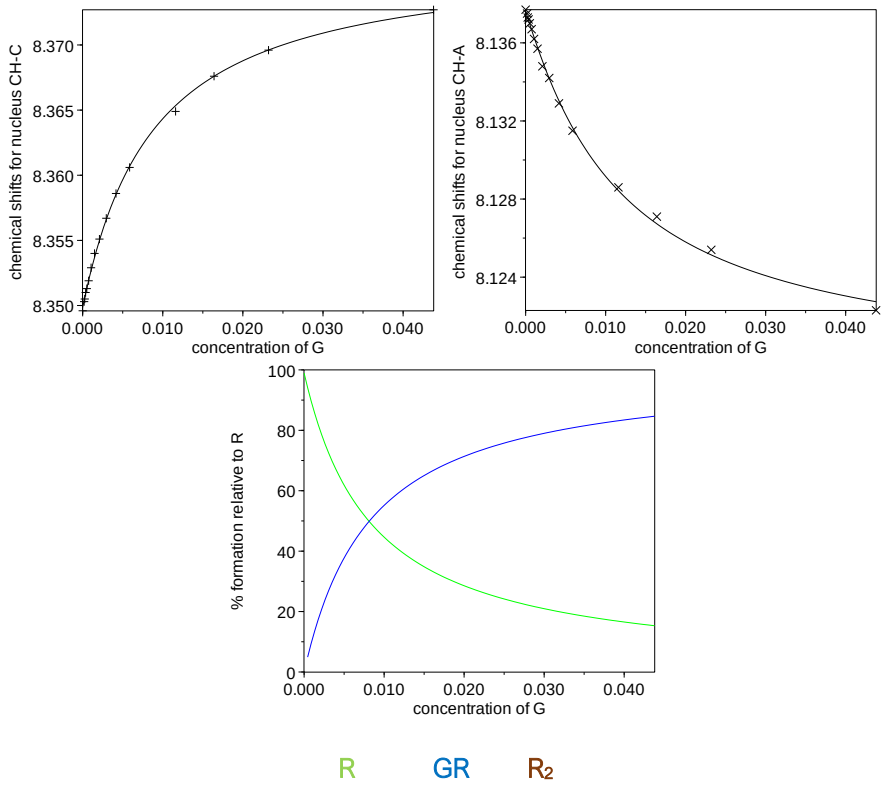
Results page

```
 no. of spectra            16
 no. of resonance values  138
 no. of resonant nuclei     9


   sigma =   0.00041122663          RMS weighted residual =   0.00037702546

       stoich            value   relative    log    standard
       coeff                     std devn    beta   deviation

 Beta  1  1 refined  1.2760E+002  0.0077    2.1059   0.0033  (   GR   )
 Beta  0  2 constant    9.5345              0.9793           (   R2   )

Individual chemical shifts
                    G                         R
 ============================================================
         +        value    error         value    error
CH'-1    +       4.4403   0.0005
CH-1     +       4.4009   0.0005
CH-6     +       3.9933   0.0003
CH'-4    +       3.9140   0.0003
CH3      +       3.5644   0.0003
CH-2     +       3.2986   0.0003
CH-F     +                             9.2370   0.0019
CH-C     +                             8.3482   0.0017
CH-A     +                             8.1453   0.0017
         +
                   1,1                      0,2
 ============================================================
         +        value    error         value    error
CH'-1    +       1.7690   0.0236
CH-1     +       1.6625   0.0242
CH-6     +       3.4285   0.0074
CH'-4    +       3.1766   0.0084
CH3      +       2.9061   0.0079
CH-2     +       2.0574   0.0120
CH-F     +       9.3380   0.0006         9.8078   0.1818
CH-C     +       8.3769   0.0006         8.5042   0.1712
CH-A     +       8.1189   0.0006         7.4280   0.1705
```

Titration Plots

Chemical shifts (δ, ppm) *vs*. concentration of G (mol L^{-1})
experimental (symbols) and calculated (lines) values




chemical shifts for nucleus CH-C
8.370
8.365
8.360
8.355
8.350
0.000
0.010
0.020
0.030
0.040
concentration of G
chemical shifts for nucleus CH-A
8.136
8.132
8.128
8.124
0.000
0.010
0.020
0.030
0.040
concentration of G
% formation relative to R
100
80
60
40
20
0
0.000
0.010
0.020
0.030
0.040
concentration of G
R
GR
R2

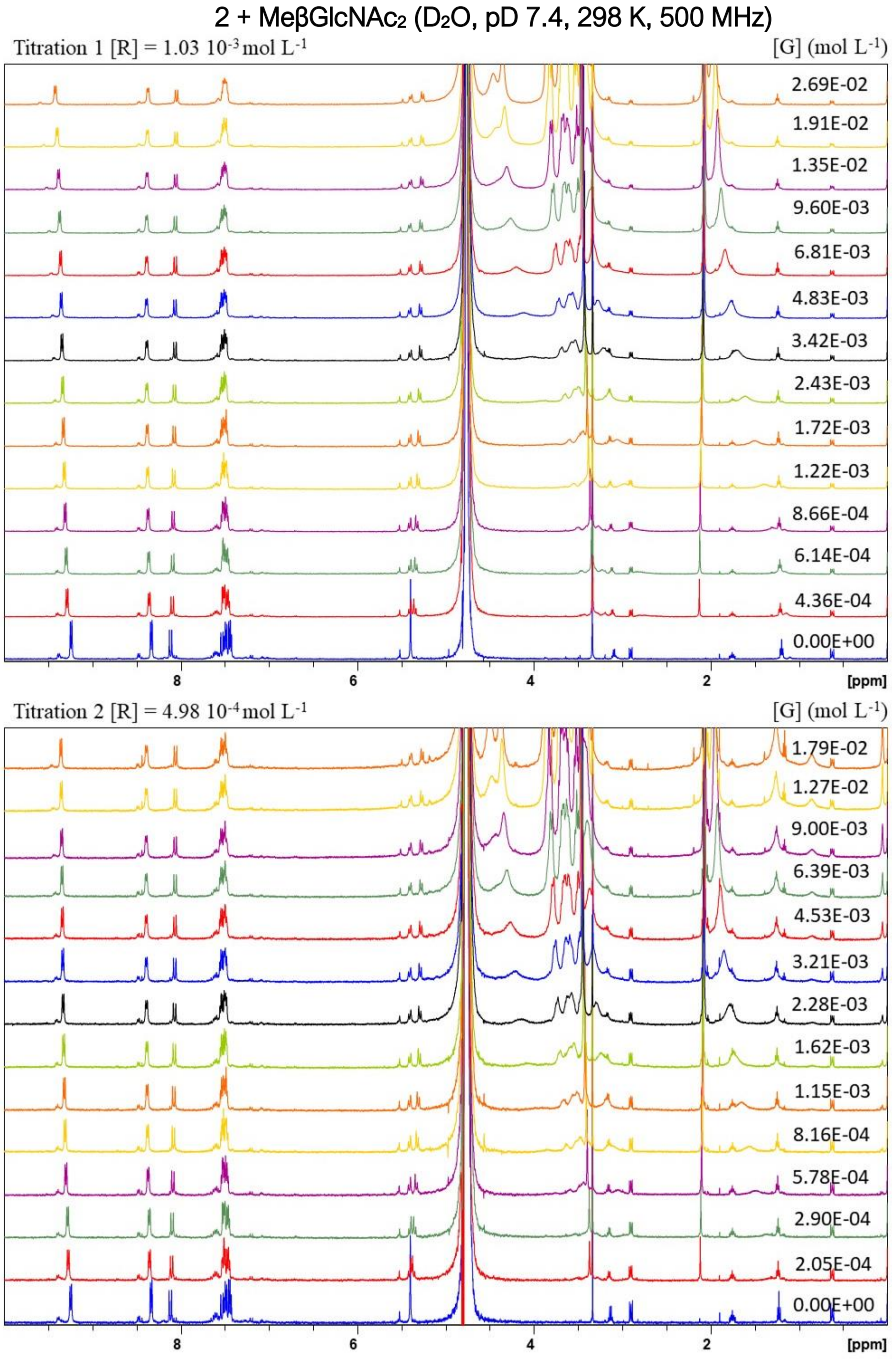


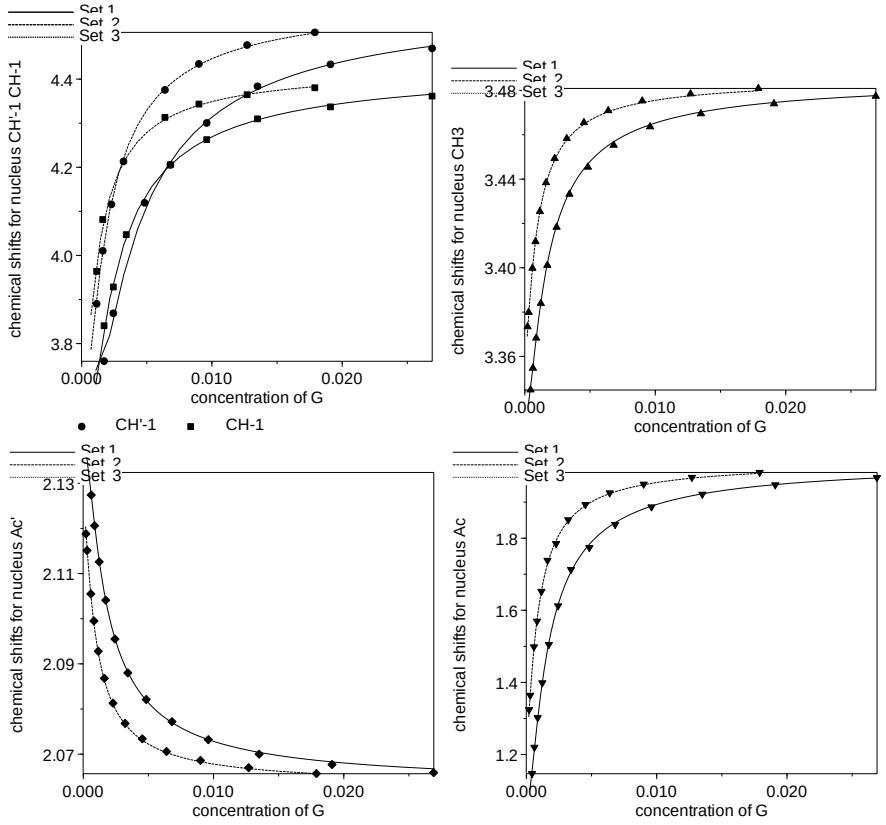
Figure S9. ^{1}H NMR spectroscopic titrations (500 MHz, D_2O, pD 7.4, 298 K) of receptor **2** (R) with incremental concentrations of MeβGlcNAc$_2$ (G).

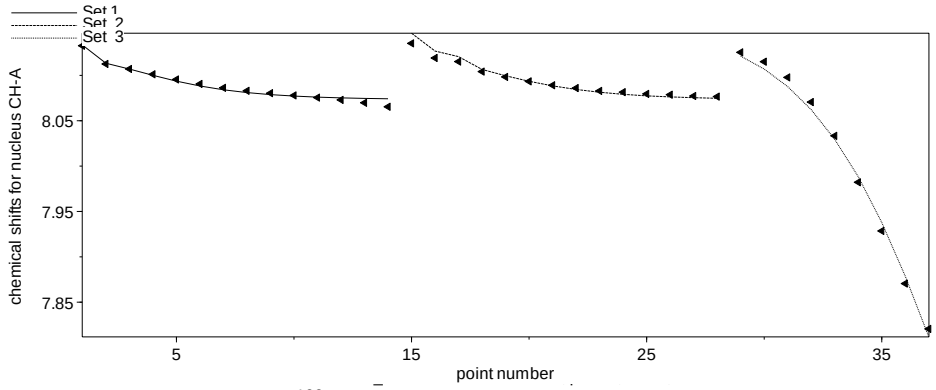
Results page

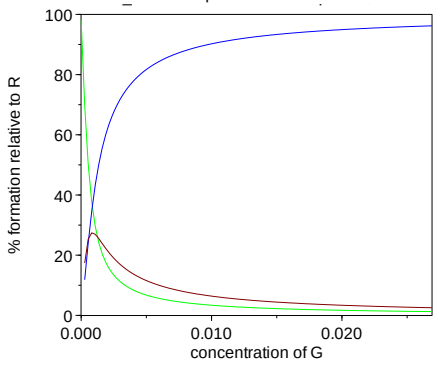
```
 no. of spectra            37
 no. of resonance values  145
 no. of resonant nuclei     6


   sigma =   0.00509559649          RMS weighted residual =   0.00471217999

       stoich            value   relative    log    standard
       coeff                     std devn    beta   deviation

 Beta  1  1 refined  2.9612E+003  0.0526    3.4715   0.0228  (   GR   )
 Beta  1  2 refined  3.0222E+006  0.1074    6.4803   0.0467  (   GR2  )
 Beta  0  2 constant    9.5345              0.9793           (   R2   )

Individual chemical shifts
                    G                          R
 ============================================================
        +        value    error          value    error
CH'-1   +       4.5845   0.0042
CH-1    +       4.4287   0.0034
CH3     +       3.4856   0.0022
Ac'     +       2.0629   0.0021
Ac      +       2.0137   0.0028
CH-A    +                                8.1592   0.0023
        +
                   1.1                        1.2
 ============================================================
        +        value    error          value    error
CH'-1   +       1.5624   0.0820          7.7181   0.5640
CH-1    +       2.7163   0.0511          4.3282   0.2684
CH3     +       3.2730   0.0137          3.3332   0.0297
Ac'     +       2.1674   0.0127          2.1392   0.0296
Ac      +       0.7558   0.0352          0.9914   0.0377
CH-A    +       8.0732   0.0018          8.0710   0.0098
        +
                   0.2
 ================================
        +        value    error
CH'-1   +
CH-1    +
CH3     +
Ac'     +
Ac      +
CH-A    +       6.8010   0.0174

 Correlation coefficients*1000


       1    2
   1
   2 -407

 Parameters are numbered as follows
    1 beta 1.1
    2 beta 1.2
```

Titration Plots

Chemical shifts (δ. ppm) *vs*. concentration of G (mol L^{-1})
experimental (symbols) and calculated (lines) values




Set 1
Set 2
Set 3
chemical shifts for nucleus CH-A
8.05
7.95
7.85
5
15
25
35
point number


% formation relative to R
100
80
60
40
20
0
0.000
0.010
0.020
concentration of G

R GR GR_2 R_2

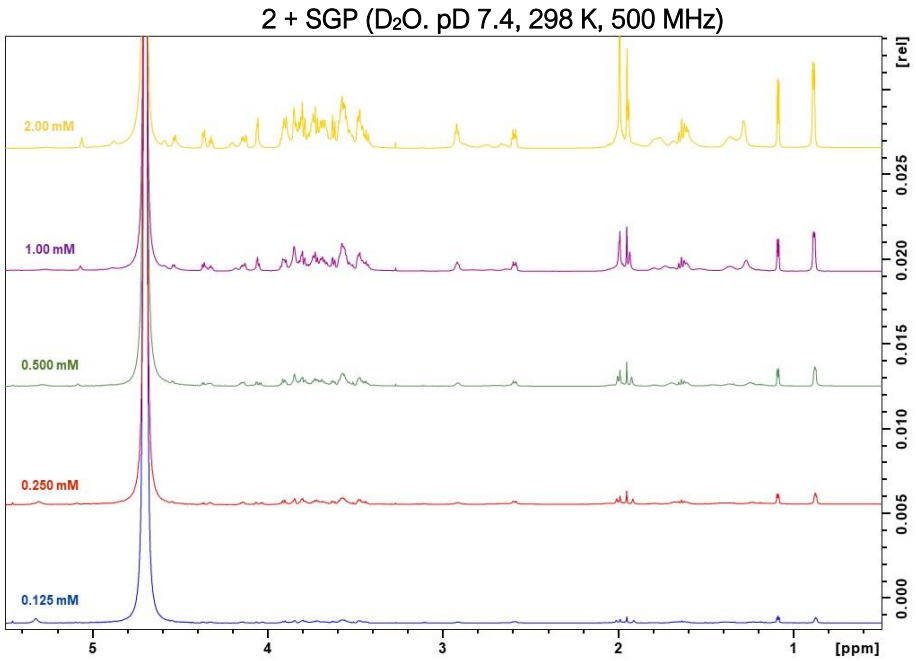


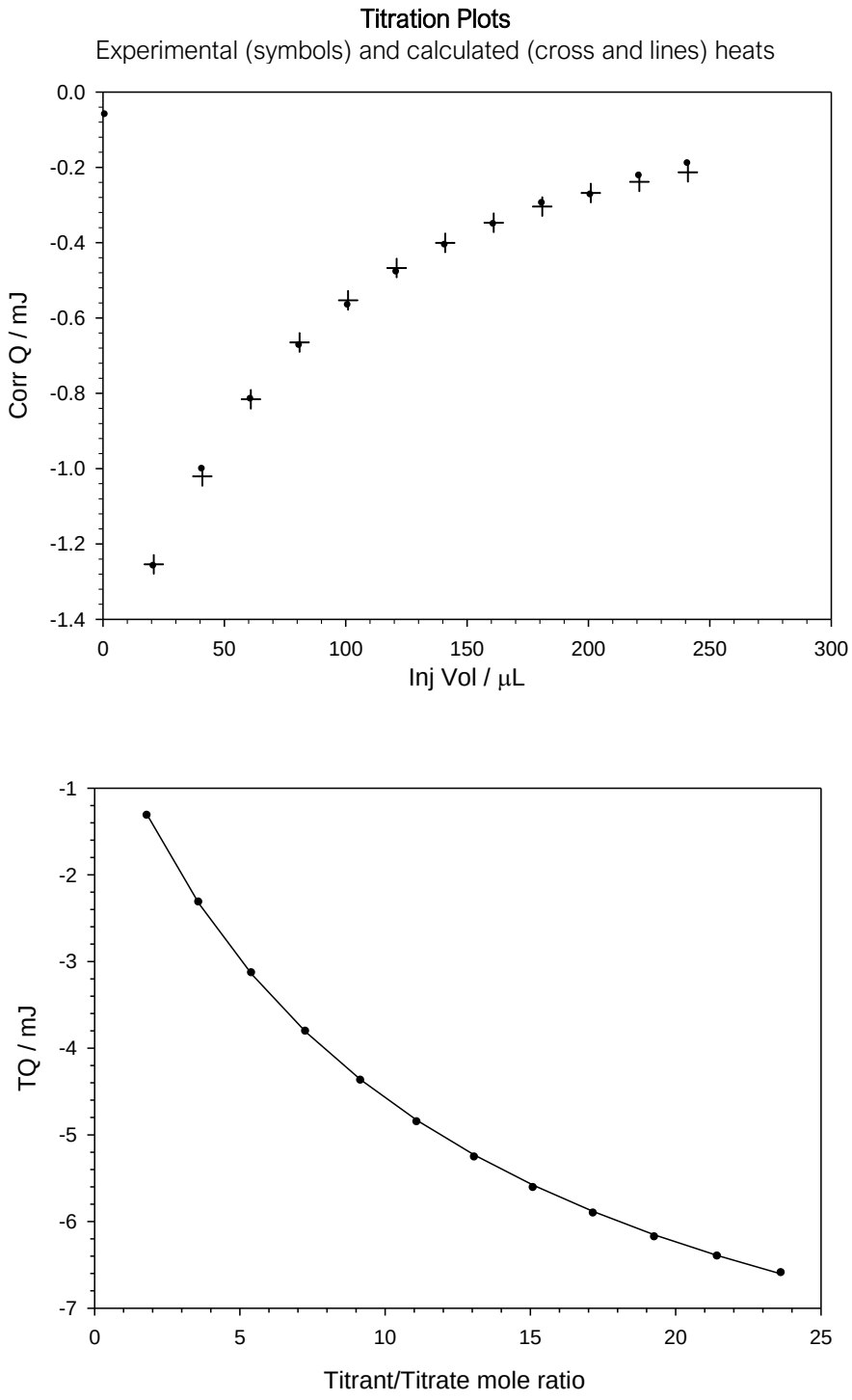
Figure S10. ^{1}H NMR spectroscopic titration (500 MHz, D_2O, pD 7.4, 298 K) of receptor **2** (0.5 mM) with incremental concentrations of SGP (0.125, 0.250, 0.500, 1.00, 2.00 mM).

Calorimetric titrations and data analysis: Isothermal Titration Microcalorimetry experiments were performed at 298 K with a Nano-ITC instrument. After an initial injection of 3 μL, which was excluded from data analysis. aliquots of the titrant solution, containig the glycoside, were injected stepwise into the sample cell containing a solution of the titrate **2**. For SGP binding measurement, **2** was use as the titrant (in syring) and SGP as the titrate (in sample cell) Solutions containing **2** and SGP were prepared in H_2O adjusting the pH to a value of 7.4. All experiments were performed in H_2O at pH 7.4. Heats of dilution were measured by injecting the titrant solution into neat H_2O and then subtracted from the binding heats. Titrations data were fitted to measure the cumulative association constants and the thermodynamic parameters using the HypCal software package.[46]

2 + MeβCeB (H_2O. pH 7.4. 298 K)

Results page

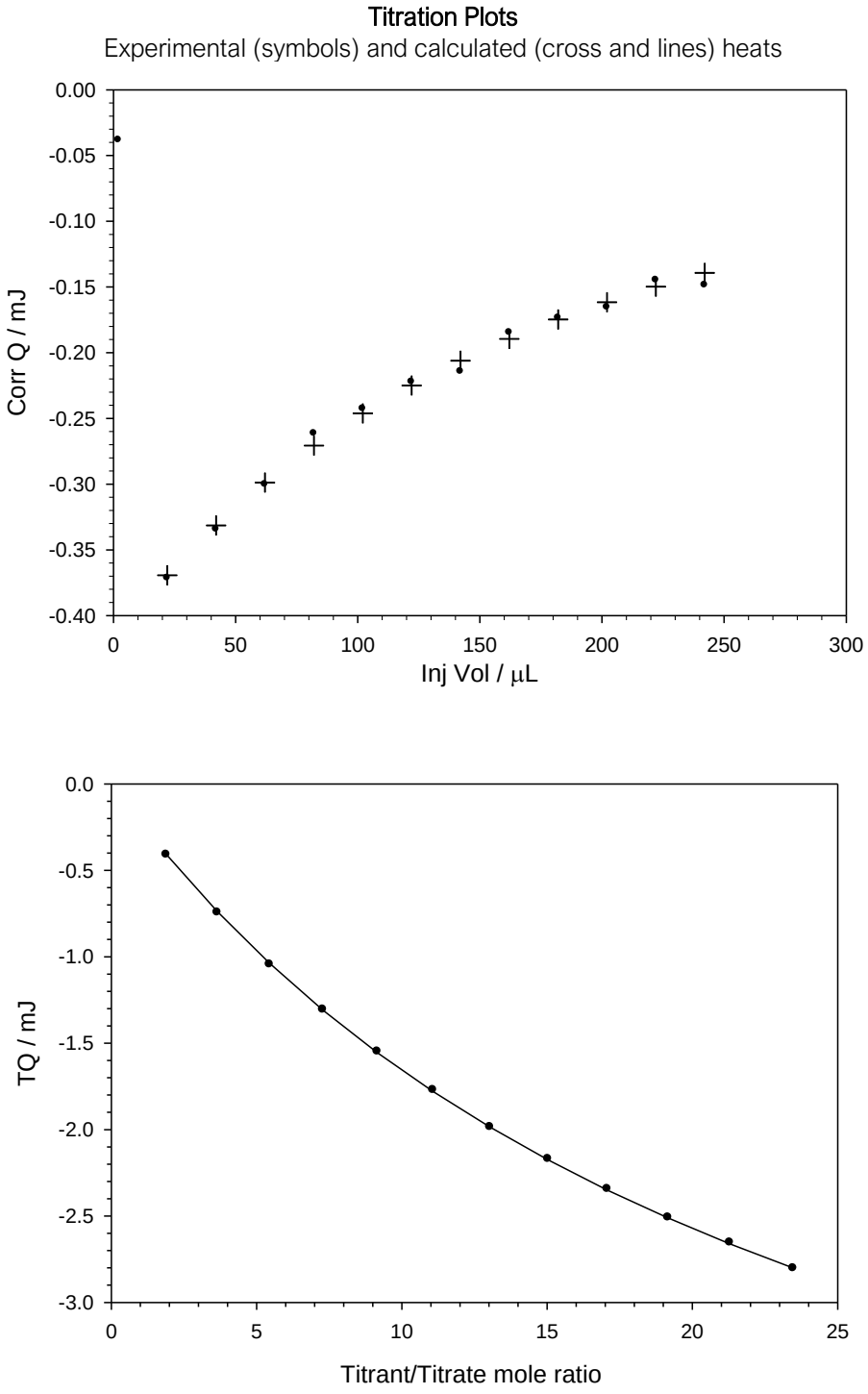
```
Reagent Reagent
  number   name
    1      R
    2      G

 Temperature: 25.000 Celsius. Excessive limit 0.99

   sigma =   0.01442

 Formation constants   Value    relative    log    standard
                                std devn    beta   deviation
 Beta A refined     0.3113E+03  0.1586    2.4932    0.0689   1 1
 Beta B refined     0.3380E+06  0.8799    5.5290    0.3821   2 1


++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

 Thermodynamic Functions. kJ/mol


          - DeltaG°          - DeltaH°          T DeltaS°
   A   14.2312  0.3931    26.1484  1.4425    -11.9172  1.2086
   B   31.5594  2.1813   -24.2222  11.0113    55.7816  13.1525
```


Titration Plots
Experimental (symbols) and calculated (cross and lines) heats
Corr Q / mJ
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-1.4
0
50
100
150
200
250
300
Inj Vol / μL
TQ / mJ
-1
-2
-3
-4
-5
-6
-7
0
5
10
15
20
25
Titrant/Titrate mole ratio

2 + MeβMal (H_2O. pH 7.4. 298 K)

Results page

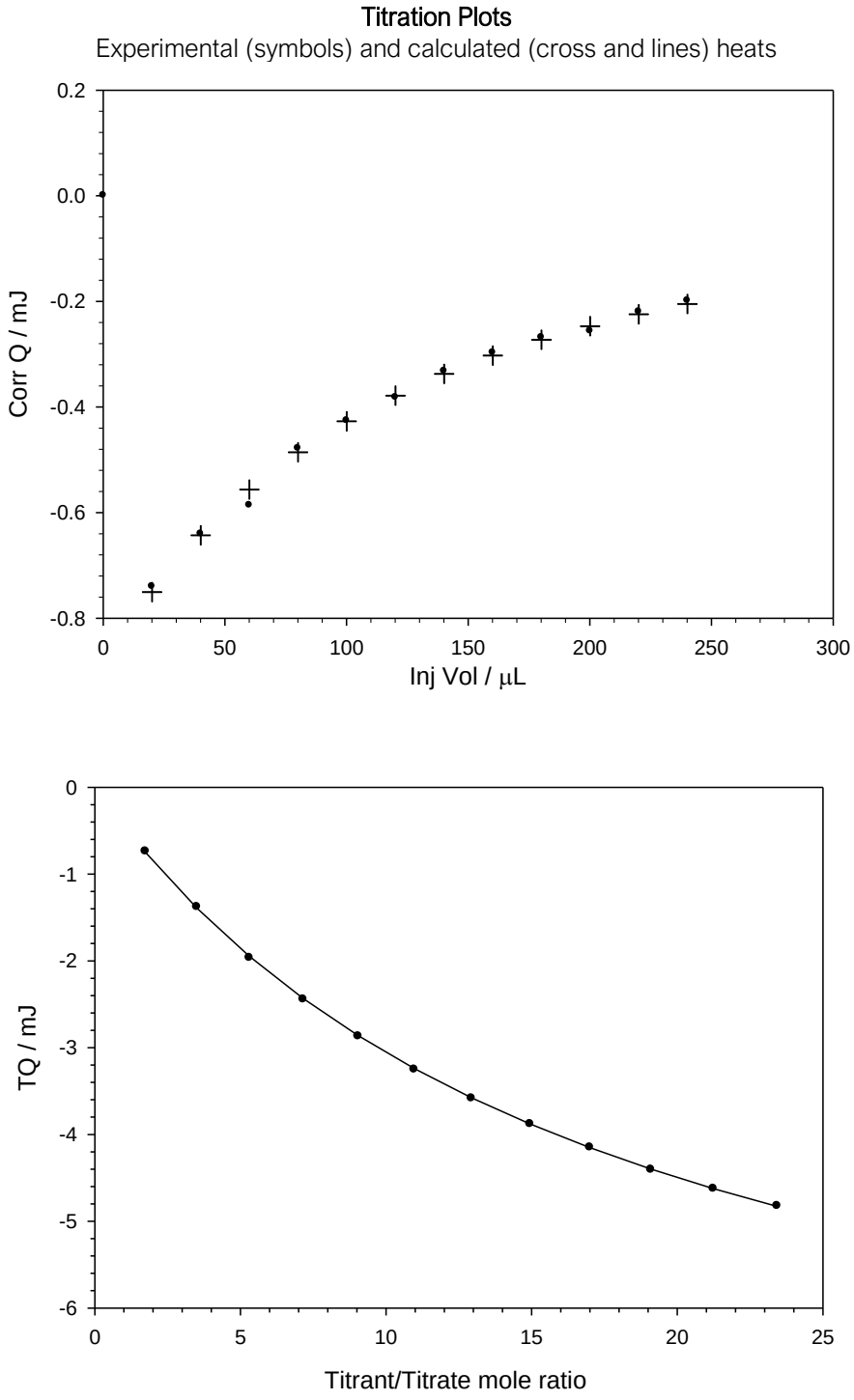
```
Reagent Reagent
  number   name
    1      R
    2      G

   sigma =   0.00583

 Formation constants   Value    relative    log    standard
                                std devn    beta   deviation
 Beta A refined     0.6007E+02  0.0393    1.7787    0.0171   1 1


++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
 Thermodynamic Functions. kJ/mol


          - DeltaG°          - DeltaH°           T DeltaS°
   A   10.1526  0.0974    17.8674  0.5063     -7.7148  0.6012
```


Titration Plots
Experimental (symbols) and calculated (cross and lines) heats
Corr Q / mJ
0.00
-0.05
-0.10
-0.15
-0.20
-0.25
-0.30
-0.35
-0.40
0
50
100
150
200
250
300
Inj Vol / µL
TQ / mJ
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
0
5
10
15
20
25
Titrant/Titrate mole ratio

2 + MeβLac (H_2O. pH 7.4. 298 K)

Results page

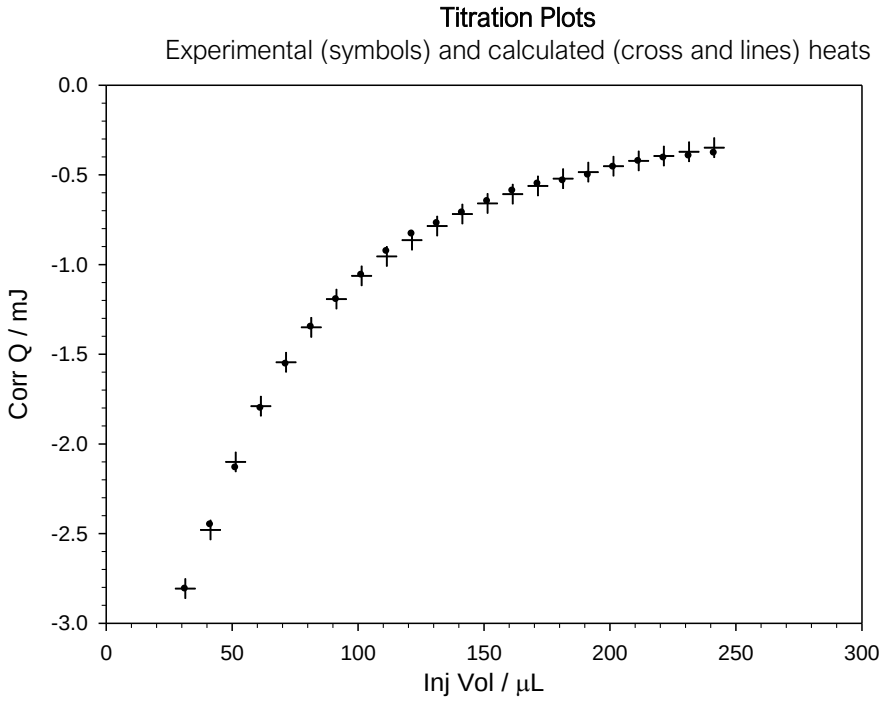
```
Reagent Reagent
  number   name
    1      R
    2      G

   sigma =   0.01137

 Formation constants   Value    relative    log    standard
                                std devn    beta   deviation
 Beta A refined     0.9535E+02  0.0331    1.9793    0.0144   1 1


++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
 Thermodynamic Functions. kJ/mol


          - DeltaG°          - DeltaH°           T DeltaS°
   A   11.2981  0.0821    23.3893  0.5200    -12.0912  0.5978
```

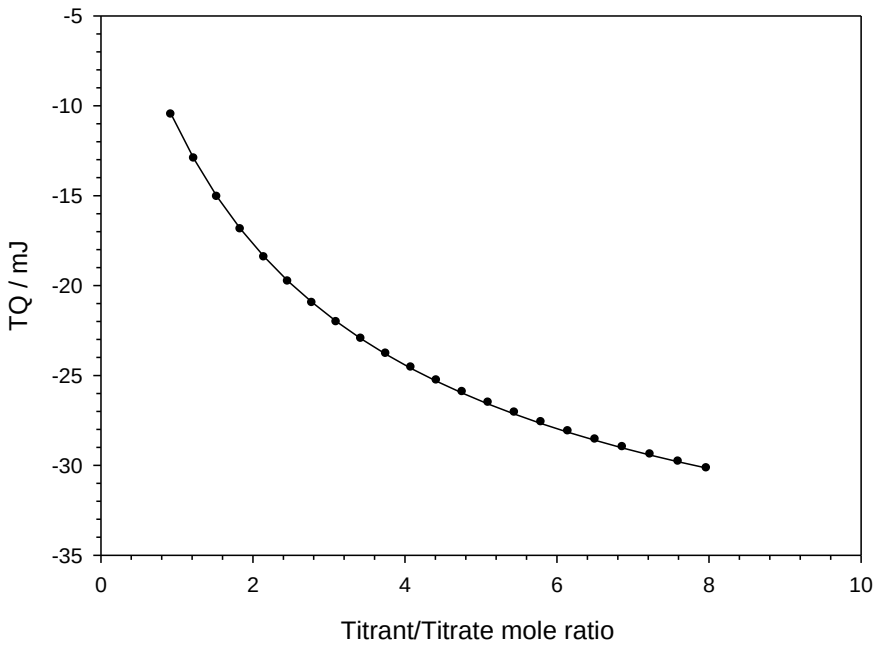

Titration Plots
Experimental (symbols) and calculated (cross and lines) heats
Corr Q / mJ
Inj Vol / μL
TQ / mJ
Titrant/Titrate mole ratio

2 + MeβGlcNAc$_2$ (H_2O. pH 7.4. 298 K)

Results page

```
Reagent Reagent
  number   name
    1      R
    2      G

   sigma =   0.01978

 Formation constants   Value    relative    log    standard
                                std devn    beta   deviation
 Beta A refined     0.2794E+04  0.2230    3.4462    0.0969   1 1
 Beta B refined     0.1456E+08  0.5385    7.1632    0.2339   2 1


++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

 Thermodynamic Functions. kJ/mol


          - DeltaG°          - DeltaH°          T DeltaS°
   A   19.6711  0.5529    47.3247  2.1591    -27.6535  2.6886
   B   40.8877  1.3350   -27.4006  5.0372     68.2883  3.8008
```

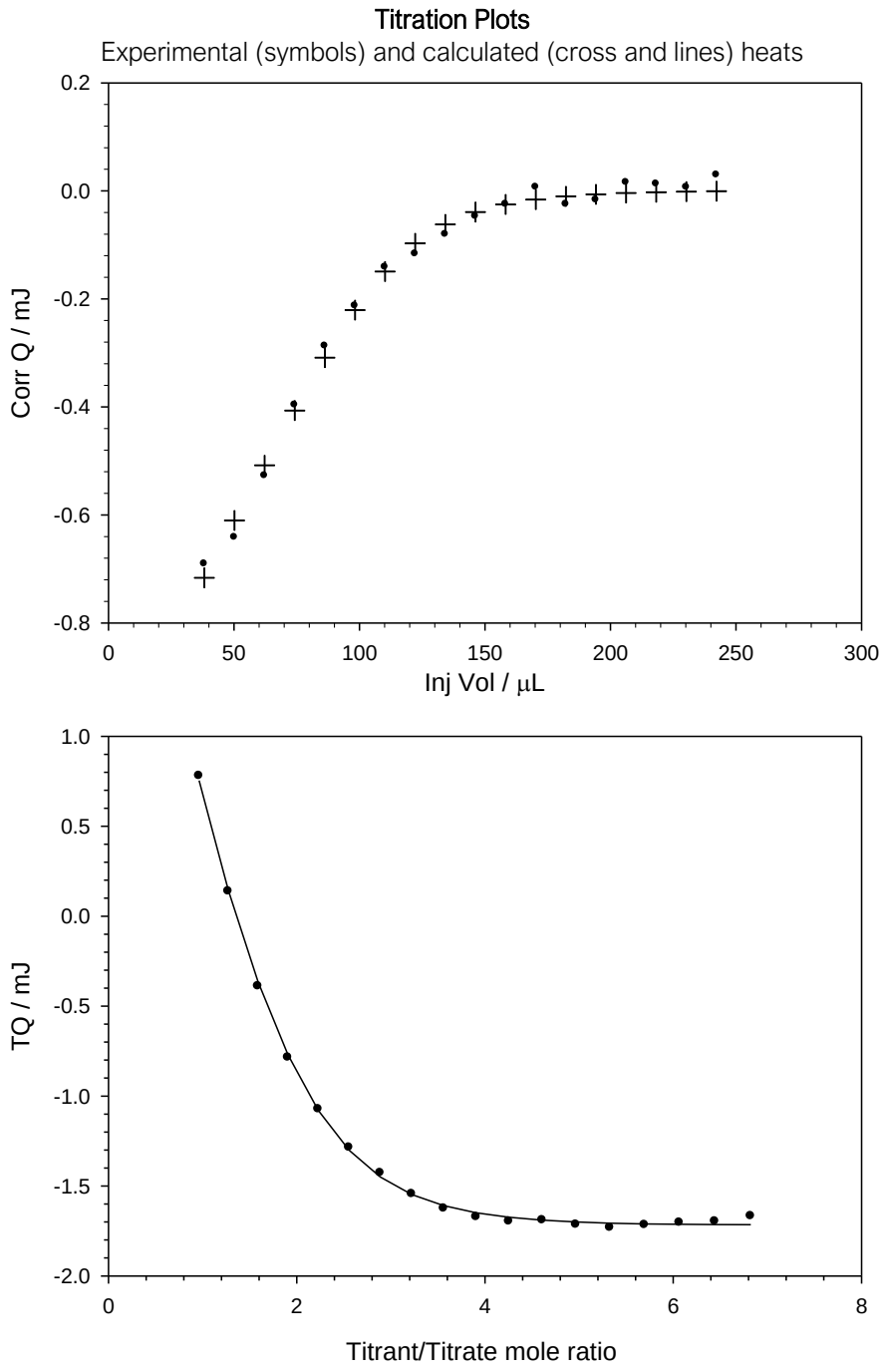

Titration Plots
Experimental (symbols) and calculated (cross and lines) heats
Corr Q / mJ
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
0
50
100
150
200
250
300
Inj Vol / μL


TQ / mJ
-5
-10
-15
-20
-25
-30
-35
0
2
4
6
8
10
Titrant/Titrate mole ratio

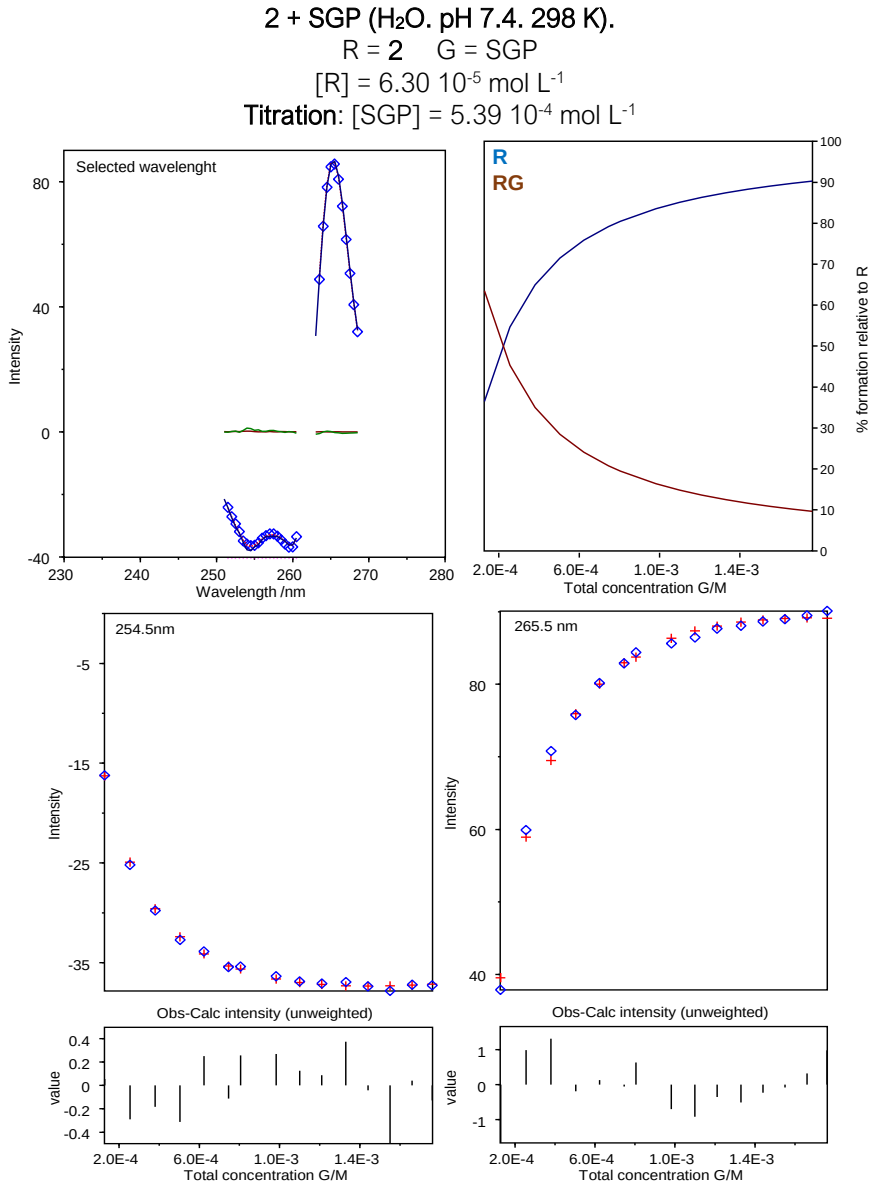
2 + SGP (H_2O. pH 7.4. 298 K)

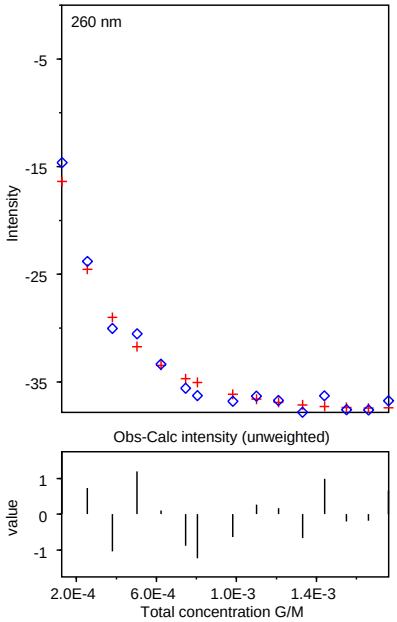
Results page

```
Reagent Reagent
  number   name
    1      G
    2      R

 Temperature: 25.000 Celsius. Excessive limit 0.99

   sigma =   0.02054

 Formation constants   Value    relative    log    standard
                                std devn    beta   deviation
 Beta A refined     0.2735E+04  0.6023    3.4370    0.2616    1
1
 Beta B refined     0.1986E+08  0.6000    7.2981    0.2606    1
2


++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

 Thermodynamic Functions, kJ/mol


          - DeltaG°          - DeltaH°          T DeltaS°
   A   19.6185  1.4931    14.4588  5.0028      5.1598  6.4810
   B   41.6575  1.4874    10.6189  0.2919     31.0386  1.4303
```


Titration Plots
Experimental (symbols) and calculated (cross and lines) heats
Corr Q / mJ
Inj Vol / μL
TQ / mJ
Titrant/Titrate mole ratio

Circular dichroism data analysis: Circular dichroism experiments were performed at 298 K using a 0.1 cm cuvette. To a diluted solution of receptor **2** aliquots of a concentrated solution of SGP were added and CD spectra recorded. Titrations data were fitted to measure the cumulative association constants using the HypSpec2014 software package.[47]






```
HypSpec refinement output. Version number 1.2.35
Refinement concluded at 05/08/2022 11:27:10
Converged in 3 iterations with sigma = 0.28996

                        standard
Log beta    value       deviation
RG          3.7371      0.0031
```

4.4.3 Structural studies:

NMR methods. NMR experiments were performed at 800 MHz in D_2O at pD 7.4 at 298 K. The experiments on the complex were performed using a 2 mM equimolar solution of **2** and SGP. In addition to standard 1D ^1H NMR spectra, COSY, TOCSY, HSQC and NOESY experiments (500 ms mixing time) were also acquired to assign the resonances of all the molecular entities and to detect the relevant intramolecular and intermolecular contacts.

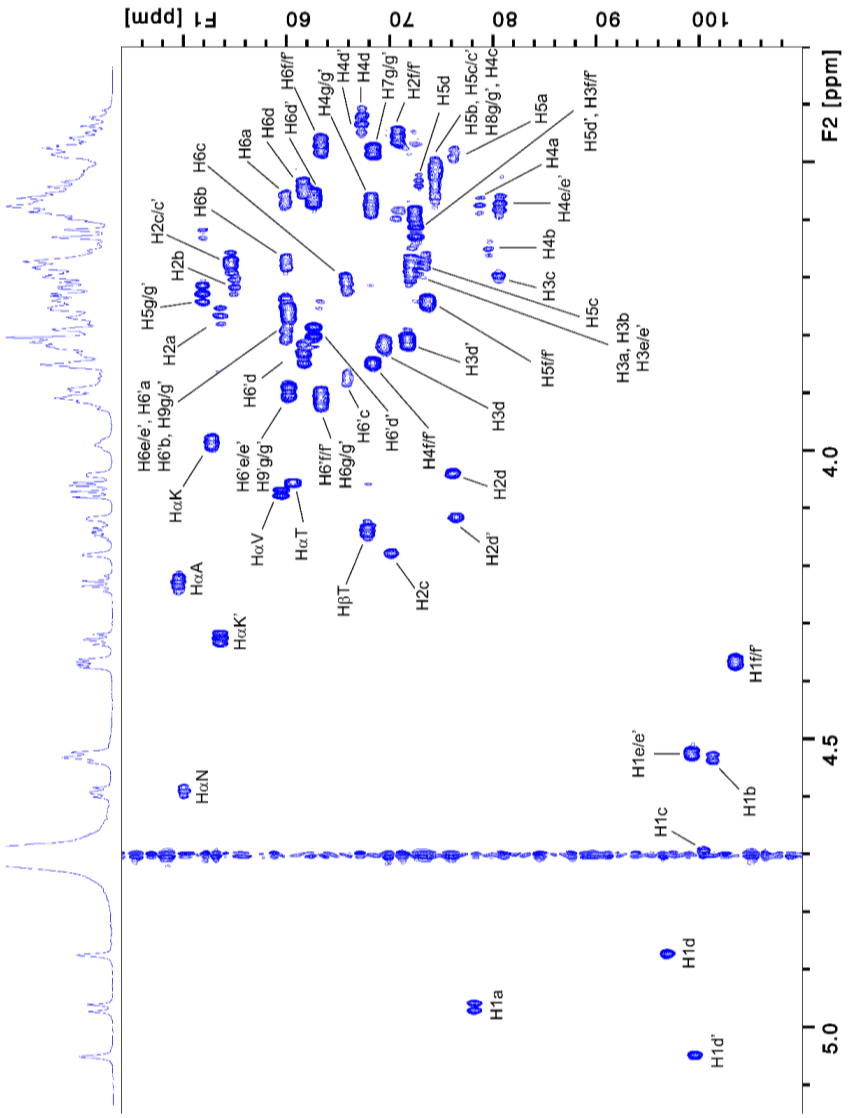


Figure S11. Saccharidic region of HSQC (800 MHz) spetrum of SGP (2 mM). Assigned protons are labelled.

4.5 Bibliography

[1]Ernst B., Hart G.W., Sinay P., *Carbohydrates in chemistry and biology*, **2008**, vols 1–4.

[2]Varki A., *Glycobiology*, **2017**, 27, 3-49.

[3]Smith B.A.H., Bertozzi C.R., *Nature Reviews Drug Discovery*, **2021**, 20, 217-243.

[4]Reily C., Stewart T.J., Renfrow M.B., Novak J., *Nat. Rev. Nephrol.*, **2019**, 15, 346-366

[5]Taniguchi N., Kizuka Y., *Advances in Cancer Research*, **2015**, 126, 11-51.

[6]Stanley P., Schachter H., Taniguchi N., *Essentials of Glycobiology. 2nd edition*, **2009**, Chapter 8: *N-Glycans*.

[7]Shental-Bechor D., Levy Y., *Proc Natl Acad Sci U.S.A.*, **2008**, 105(*24*), 8256-8261.

[8]Jayaprakash N.G., Surolia A., *Biochemical Journal*, **2017**, 474(*14*), 2333-2347.

[9]Paul A., Segal D., Zacco E., *Neural Regeneration Research*, **2021**, 16(*11*), 2215-2216.

[10]Amore A., *et al*., *Proc Natl Acad Sci U.S.A.*, **2017**, 114(*52*), 13667-13672.

[11]Hirayama S., Hori Y., Benedek Z., Suzuki T., Kikuchi K., *Nat. Chem. Biol.*, **2016**, 12, 853-859.

[12]Dowling W., *et al.*, *J Virol*, **2007**, 81(*4*), 1821-1837.

[13]Schnaar R.L., *Journal of Allergy and Clinical Immunology*, **2015**, 135(*3*), 609-615.

[14]Balzarini J., *Nat. Rev. Microbiol.*, **2007**, 5, 583-597.

[15]Sharon N., Lis H., *Glycobiology*, **2004**, 14(*11*), 53R-62R.

[16]Davis A.P., *Chem. Soc. Rev.*, **2020**, 49, 2531-2545.

[17]Francesconi O., Roelens S., *ChemBioChem*, **2020**, 20, 1329-1346.

[18]Francesconi O. *et al.*, *iScience*, **2022**, 25(*5*), 104239.

[19]Bravo M.F., Lema M.A., Marianski M., Braunschweig A.B., *Biochemistry*, **2021**, 60, 999-1018.

[20]Francesconi O. *et al*., *Chem. Eur. J.*, **2015**, 21, 10089-10093.

[21]Park S.H., *et al.*, *Chem. Sci.,* **2015**, 6, 7284-7292.

[22]Kubik S., *Supramolecular chemistry in water*, **2019**.

[23]Yao H. *et al.*, *J. Am. Chem. Soc.*, **2018**, 140, 13466-13477.

[24]Yamashina M., Akita M., Hasegawa T., Hayashi S., Yoshizawa M., *Sci. Adv*., **2017**, 3, 2-8.

[25]Francesconi O., Martinucci M., Badii L., Nativi C., Roelens S., *Chem. Eur. J.*, **2018**, 24, 6828-6836.

[26]Zhang D., Martinez A., Dutasta J.P., *Chem. Rev*., **2017**, 117, 4900-4942.

[27]Hayashi T., Ohishi Y., Abe H., Inouye M., *Journal of Organic Chemistry,* **2020**, 85(*4*), 1927-1934.

[28]Timmer B.J.J., Kooijman A., Schaapkens X., Mooibroek T.J., *Angew. Chem. Int. Ed*., **2021**, 60, 16178-16783.

[29]Mooibroek T.J. *et al.*, *Nat. Chem.*, **2016**, 8, 69-74.

[30]Tromans R.A. *et al.,* Nat. Chem., **2019**, 11, 52-56.

[31]Ohishi Y., Masuda K., Kudo K., Abe H., Inouye M., *Chem. Eur. J.*, **2021**, 27, 785-793.

[32]Mateus P., Wicher B., Ferrand Y., Huc, I., *Chem. Comm*., **2018**,54, 5078-5081.

[33]Mazik M., *Chem. Soc. Rev*., **2009**, 38, 935-956.

[34]Carter T.S. *et al*., *Angew. Chem. Int. Ed*., **2016**, 55, 9311-9315.

[35]Nativi, C. *et al.*, *Chem. Eur J*., **2012**, 18, 5064-5072.

[36]Francesconi O., Cicero F., Nativi C., Roelens S., ChemPhysChem, 2020, 21, 257-262.

[37]Asensio J.L., Cañada F.J., Bruix M., Rodriguez-Romero A., Jiménez-Barbero J., *Eur. J. Biochem*., **1995**, 230, 621-633.

[38]Schnaar R.L., *J. Leukoc. Biol*., **2016**, 99, 825-838.

[39]Duan S., Paulson J.C., *Annual Review of Immunology*, **2020**, 38, 365-395.

[40]Kajihara Y., *et al.*, *Chem. Eur. J*., **2004**, 10, 971-985.

[41]Francesconi O., Ienco A., Nativi C., Roelens S., *ChemPlusChem*, **2020**, 85, 1-6.

[42]Francesconi O. *et al*., *J. Org. Chem.*, **2022**, 87(*5*) 2662-2667.

[43]Vacca A., Francesconi O., Roelens S., *Chem. Rec.*, **2012**, 12, 544-566.

[44]Vacca A., Nativi C., Cacciarini M., Pergoli R., Roelens S., *J. Am. Chem. Soc*., **2004**, *126*, 16456-16465.

[45]Frassineti C., Ghelli S., Gans P., Sabatini A., Moruzzi M.S., Vacca A., *Anal. Biochem*., **1995**, *231*, 374-382.

[46]Gans P., Sabatini A., Vacca A., *J. Solution Chem*. **2008**, 37, 467-476.

[47]Gans P., Sabatini A., Vacca A., *Talanta*, **1996**, 43, 1739-1753.

Chapter 5

Conformationally constrained sialyl analogues as new potential binders of *h*-CD22

Adapted from:

Forgione R.E., Nieto F.F., Di Carluccio C., Milanesi F., Fruscella M., Papi F., Nativi C., Molinaro A., Palladino P., Scarano S., Minunni M., Montefiori M., Civera M., Sattin S., Francesconi O., Marchetti R., Silipo A., *ChemBioChem.,* **2022**, *23*, e202200076

Abstract

Here, two conformationally constrained sialyl analogues were synthesized and characterized in their interaction with the inhibitory Siglec, human CD22 (*h*-CD22). An orthogonal approach, including biophysical assays (SPR and fluorescence), ligand-based NMR techniques, and molecular modelling, was employed to disentangle the interaction mechanisms at a molecular level. The results showed that the Sialyl-TnThr antigen analogue represents a promising scaffold for the design of novel *h*-CD22 inhibitors. Our findings also suggested that the introduction of a biphenyl moiety at position 9 of the sialic acid hampers the canonical accommodation of the ligand in the protein binding pocket, even though the affinity with respect to the natural ligand is increased. Our outcomes address the search of novel modifications of the Neu5Ac-α-2,6-Gal epitope, outlining new hints for the design and synthesis of high-affinity *h*-CD22 ligands, offering new prospects for therapeutic intervention to prevent autoimmune diseases and B-cells malignancies.

5.1 Introduction

CD22 (Siglec-2) is a transmembrane receptor belonging to the evolutionary conserved class of Siglecs (Sialic acid binding immunoglobulin type lectins) and is selectively expressed on B-lymphocytes and, to a lesser extent, to other hematopoietic system cells.[1-3]

Upon sialoglycans engagement, CD22 negatively modulates B Cell Receptor (BCR) signalling with significant implications in maintaining tolerance to self-antigens, mandatory to prevent autoimmune diseases and B cells related malignancies.[4,5] The mechanism of BCR modulation involves the formation of CD22 homo-oligomers on resting B cells by means of *cis* interactions. Subsequently, specific antigens provoke a conformational change of BCR, leading to the activation of immune response and to the recruitment of CD22 clusters; ultimately, trans sialoglycan binding of autologous self-cells triggers BCR inhibition.[6]

The accommodation in the binding pocket of CD22 of sialic acids α-2,6-linked to galactose epitopes (Neu5Ac- α-2,6-Gal), common terminus of mammalian surface glycoproteins, has been recently described and occurs with a K_D = 281±10 μM.[7] Furthermore, the conformational behavior of natural complex-type biantennary *N*-glycans interacting with CD22 was undertaken[8] thus proving that the terminal Neu5Ac- α-2,6-Gal is the sole epitope recognized by and interacting with CD22, differently from other lectins-glycans interactions[9], and also showing the key role of glycan conformation in triggering CD22 homo-oligomers formation.

Given its role in health and diseases, therapeutic tools based on CD22 as candidate target in immunomodulation and tumor therapies have been conceived, particularly focusing on the development of antibody-based approaches.[10-12] Nonetheless, alternative lines are also under evaluation, including glycan-based therapies that rely on the design of high affinity ligands with the capability to take over endogenous glycans.

In this context, low-molecular-weight Siglec ligands and their conjugates, especially those based on synthetic sialic acid derivatives, showed promising results in clinical trials as therapeutics for the nerve regeneration, oncology, and immunology.[13] The majority of sialic acid analogues are characterized by the modifications of functional groups, as at carbons 2, 4, 5, and 9. In particular, substitutions at the hydroxyl/*N*-acetyl groups at position 9 and 5 respectively have shown to significantly enhance the binding affinity to Siglecs.[13-16] In parallel, the design of polyvalent constructs that include Siglecs specific- high affinity sialylated probes, to be conjugated to multivalent carriers, is a further step for biomedical applications.[13-17]

In a recent paper,[18] the synthesis and biological activity of a Neu5Ac- α-2,6-Gal derivative, developed as an analogue of the Sialyl TnThr tumor antigen, were reported. From the structural viewpoint, the synthetic compound is a constrained tricyclic glycoside, that retains the α-O-glycosidic linkage at the Gal moiety and in which the terminal lactam ring mimics the Thr residue of the Tn/STn natural antigens. The Sialyl TnThr analogue belongs to a class of compounds exhibiting interesting inhibitory activity also toward other biological targets.[19-20]

With the believe that the rigid galactoside moiety would affect the binding specificity of CD22, here we report the description of the molecular recognition by human CD22 of the Sialyl-TnThr analogue and of a newly synthesized derivative, with the aim to identify novel ligands with therapeutic and diagnostic potential. The binding mode of such sialylated analogues was depicted by using NMR, fluorescence, Surface Plasmon Resonance (SPR), and molecular modelling techniques. The obtained results might contribute to rational design of CD22 potent inhibitors and/or regulators.

5.2 Results and discussion

Based on the structure of the previously synthesized sialo-derivative **1** (**Figure 1**),[18-20] the synthesis of **2** has been conceived. The sialo-derivative **2** (**Figure 1**), having a similar scaffold than **1** but bearing a biphenyl substituent at position 9, was synthetized based on the observation that a similar sialo-analogue increased the bound potency and selectivity for CD22.[21]

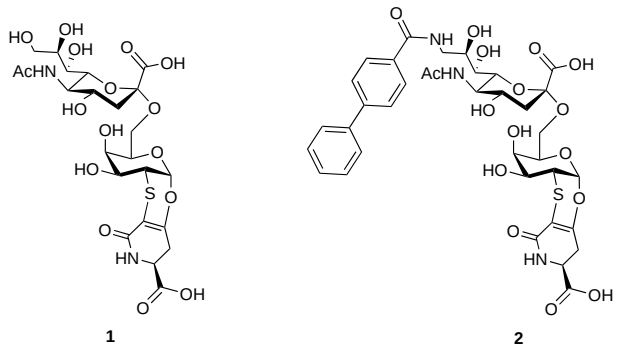


Figure 1. Structures of the sialic acid analogues used in this study.

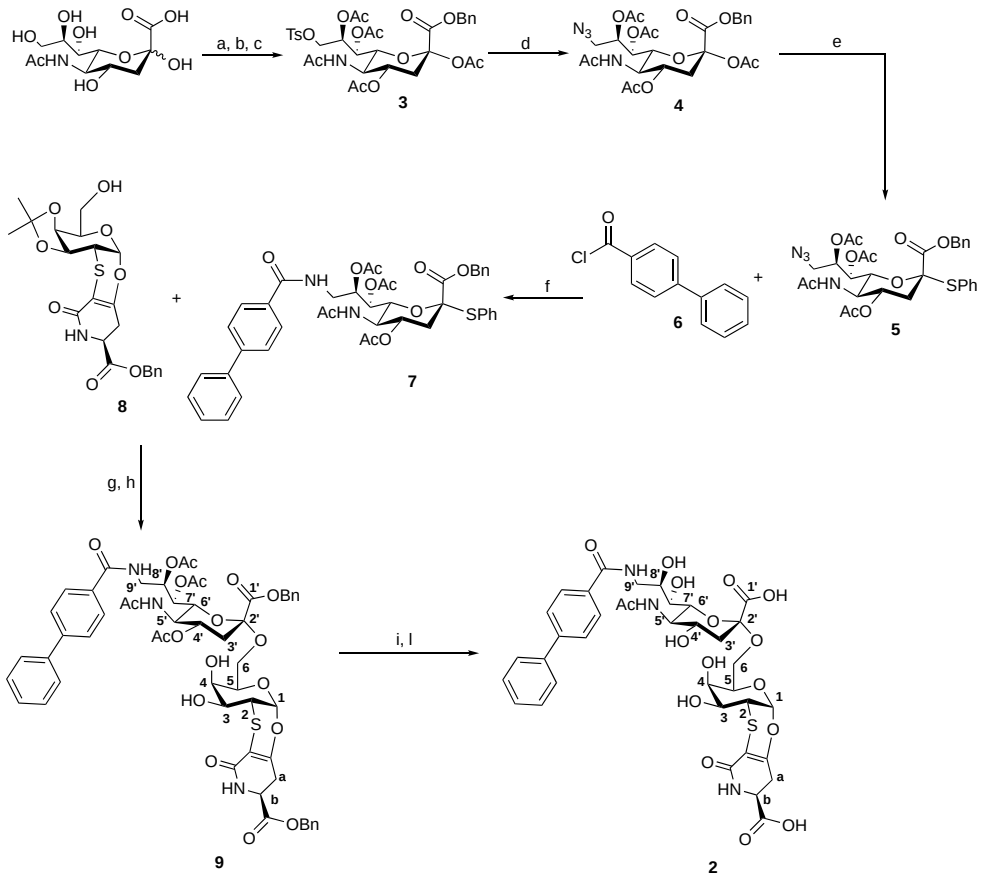
In detail, analogue **2** was prepared (**Scheme 1**) starting from *N*-acetylneuraminic acid. Esterification of the carboxylic group using benzyl bromide was followed by tosylation of the hydroxylic group in position 9 and acetylation of the remaining positions to give the protected sialic derivative 3 with a 40% yield on three steps. Tosylate group of **3** was then substituted with NaN_3 to give the azido derivative **4** with a 50% yield. Reaction of **4** with thiophenol in presence of BF_3•Et_2O gives the *S*-phenyl sialoside **5** with a 92% yield. Intermolecular Staudinger ligation between **5** and [1,1'-biphenyl]-4-carbonyl chloride **6**[22] in presence of PPh_3 gave the biphenyl substituted sialoside **7** with a 30% yield. Glycosylation reaction between the glycosidic donor **7** and the acceptor **8**, prepared according to known literature[23] , was carried out activating the *S*-phenyl group of 7 with *N*-iodosuccinimide (NIS) and catalyzing the reaction with **8** using TfOH. After removal of the isopropylidene group with AcOH 80%, the sialo derivative **9** was obtained with a 40% yield in two steps. Removal of the benzyl ester by hydrogenation and of the acetylic groups with an ammonia solution in methanol, gave the sialo derivative **2** with a yield of 30% on two steps.

5.2.1 Molecular recognition of a sialic acid analogue 1 by CD22

The binding features of the sialo-derivative **1** (**Figure 1**) in the interplay with *h*-CD22 were described by means of fluorescence studies, ligand-based NMR techniques and molecular modelling.

Intrinsic fluorescence studies of analogue **1** in the interaction with *h*-CD22 showed that the tryptophane residues of the receptor were quenched by the ligand addition, thus proving the complex formation. The interpolation of the fluorescence data provided the corresponding dissociation constant K_D= 2.0 ± 0.1 μM (**Figure 2a** and **Figure S1**).

Furthermore, STD NMR[24-25] and tr-NOESY[26] have been employed to dissect at molecular level the interaction between analogue **1** and *h*-CD22 and to map the ligand's interacting epitope and bioactive conformation (**Figure 3**). The STD NMR spectrum highlighted that the sialyl derivative was recognized by *h*-CD22 (**Figure 3**) as shown by the relative enhancements in the STD spectrum: a strong contribution from the neuraminic acid (Neu5Ac) moiety was detected, with the highest STD effect observed for its acetyl group.



Scheme 1. Synthesis of sialo derivative **2** with atom labelling. Reagents and conditions: a) BnBr, DBU, DMF, room temperature, 18h; b) TsCl, Py, room temperature, 18h; c) Ac_2O, DMAP, Py, room temperature, 5h; d) NaN_3, dry DMF, 70 °C, 5h; e) PhSH, $BF_3{\cdot}Et_2O$, dry CH_2Cl_2, room temperature, 18h; f) PPh_3, CH_2Cl_2, room temperature, 48h; g) NIS, TfOH, dry CH_3CN/CH_2Cl_2 (10:1), -40 °C, 4h; h) AcOH 80%, 40 °C, 18h; i) Pd/C, H_2, MeOH, room temperature, 72h; l) NH_3 4M in MeOH, room temperature, 120h.

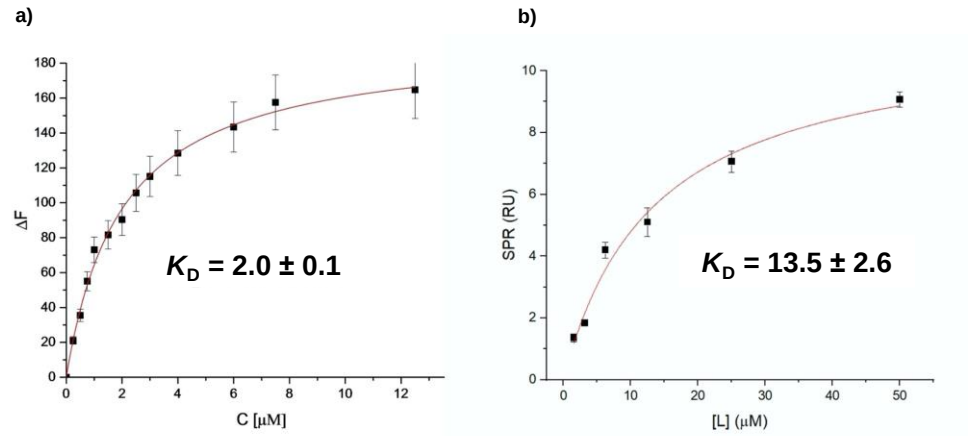


Figure 2. a) Fluorescence titration of *h*-CD22 upon the addition of analogue **1**. Each emission spectrum was recorded at the excitation wavelength of 285 nm and a temperature of 10°C. The relative binding isotherm and the value of the dissociation constant (K_D) is reported. For each data point, 10 % Y error bars are shown. b) SPR binding curve for analogue **2** vs *h*-CD22 on Protein A. Each point is representative of 3 replicates (RU_{mean} ± SD).

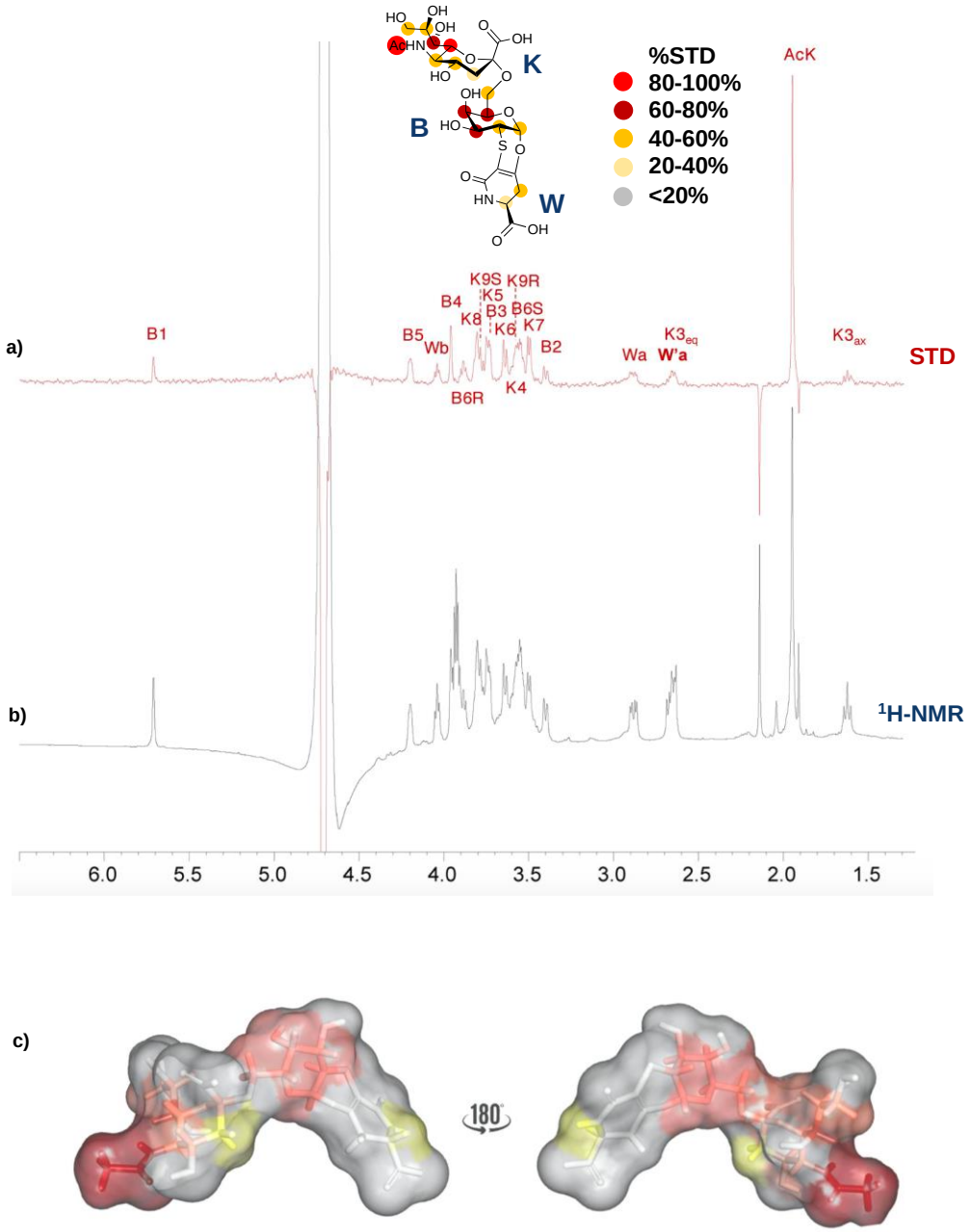


Figure 3. STD NMR analysis of analogue **1** in the interaction with *h*-CD22. Superimposition of the STD NMR spectrum a) and the 1H NMR spectrum b) of *h*-CD22/analogue **1** mixture with a molecular ratio of 1:100, at 298 K. The interacting epitope map of analogue **1** as derived by STD NMR data is also reported. c) 3D representation of the analogue **1** in the bioactive conformation obtained by tr-NOESY with molecular surface colored according to STD enhancements.

Furthermore, the involvement of the Neu5Ac lateral chain was deduced from the strong enhancements (between 50% and 70%) observed for K5, K6, K7, K8 and K9R/S protons. Significant STD effects were observed also for the galactose (Gal) residue, in particular for protons B3, B4 and B5. Interestingly, also the diastereoisotopic protons of the lactam moiety

Wa/a' showed low STD signals, thus suggesting that also the aglycon moiety was involved in the binding.

With the aim to explore the dynamic behavior of the ligand into the *h*-CD22 binding site, NOE-based experiments were combined with molecular dynamics (MD) studies (**Figure S2**). The analogue **1** exhibited a similar conformational behavior in free and bound states, as evident by comparing the NOE and tr-NOE contacts. In detail, φ torsion angle around the glycosidic linkage between Neu5Ac-α-(2-6)-Gal was located around -60° (**Figure S2a, S3b**), as demonstrated by the absence of the NOE contacts between B6 and K3 (**Figure S2c,d**) and confirmed by the stability of MD trajectories in both free and bound states (**Figure S3**). Despite a similar conformational behavior, analogue **1** adopted a more extended conformation due to the rigid nature of the lactam ring (W) if compared to 6'SLN umbrella-like topology.[27]

Computational studies also allowed to construct a 3D model of the protein-ligand complex (**Figure 4**, **S4**). Docking calculations showed that in the most populated binding mode, the analogue **1** is accommodated in the canonical binding site of *h*-CD22 (**Figure 4**). Such docked pose was then considered as starting geometry for MD simulations. The root-mean square deviation (RMSD) analysis showed that the ligand remained stable in the binding pocket of the receptor for the entire simulation time (**Figure S3a**).

The analysis of the interactions of representative models obtained from MD trajectory clustering highlighted that the sialic acid moiety interacted with Arg120 to form the carboxylate bridge essential determinant for Siglecs' recognition,[4] and that this interaction was maintained for 100% of the simulated time. Moreover, the lateral glycerol protons of Neu5Ac chain established stacking interactions with Trp128 residue, as well as hydrogen bonds with Arg131. Additionally, the *N*-acetyl group formed a hydrogen bond with Lys127. The Gal pyranose ring stacked with the Tyr64 residue thus explaining the significant STD effects. Notably, the aglycon moiety was in contact with the CC' loop of *h*-CD22 and engaged a strong hydrogen bond between its carboxylate group and the Lys66 residue of the loop, implying a significant involvement of W moiety in the binding without remarkable alteration of the minimally recognized epitope.

5.2.2 Molecular recognition of a sialic acid analogue 2 by CD22

The binding features of the sialo-derivative **2** (**Figure 1**) were investigated in the interaction with *h*-CD22 by means of SPR, NMR and molecular modelling.

In detail, the interaction between a dilution series of **2** in HBS-EP buffer and the CD22 protein immobilized on a working cell of SPR gold chip modified with protein A was evaluated against a reference cell with only protein A to compensate for matrix interferences, refractive index effects and non-specific binding of the analyte (see Supporting information). Small but reproducible SPR signals were fitted according to a 1:1 binding model (**Figure 2b**), indicating an apparent dissociation constant in the micromolar range (K_D = 13.5 ± 2.6 μM). Differently, no significative SPR signals were detected upon application of the same method to sialo-derivative **1**. We tentatively ascribe these different results to the limitations of the applied method that resulted sensitive enough for to sialo-derivative **2** only, probably for the higher hydrophobicity and larger mass of **2** with respect to sialo-derivative **1** (802.21 vs 623.14 $gmol^{-1}$). The interacting epitope of **2** upon *h*-CD22 binding was described by STD NMR (**Figure 5**). Interestingly remarkable STD enhancements from the aromatic moiety were detected, suggesting a major involvement of the biphenyl ring in the recognition process. The highest STD effect belonged to the Z' aromatic ring of the biphenyl group with the intensity of proton h at 100%, although considerable STD enhancements were detected also for the Z ring (between 40–60 %).

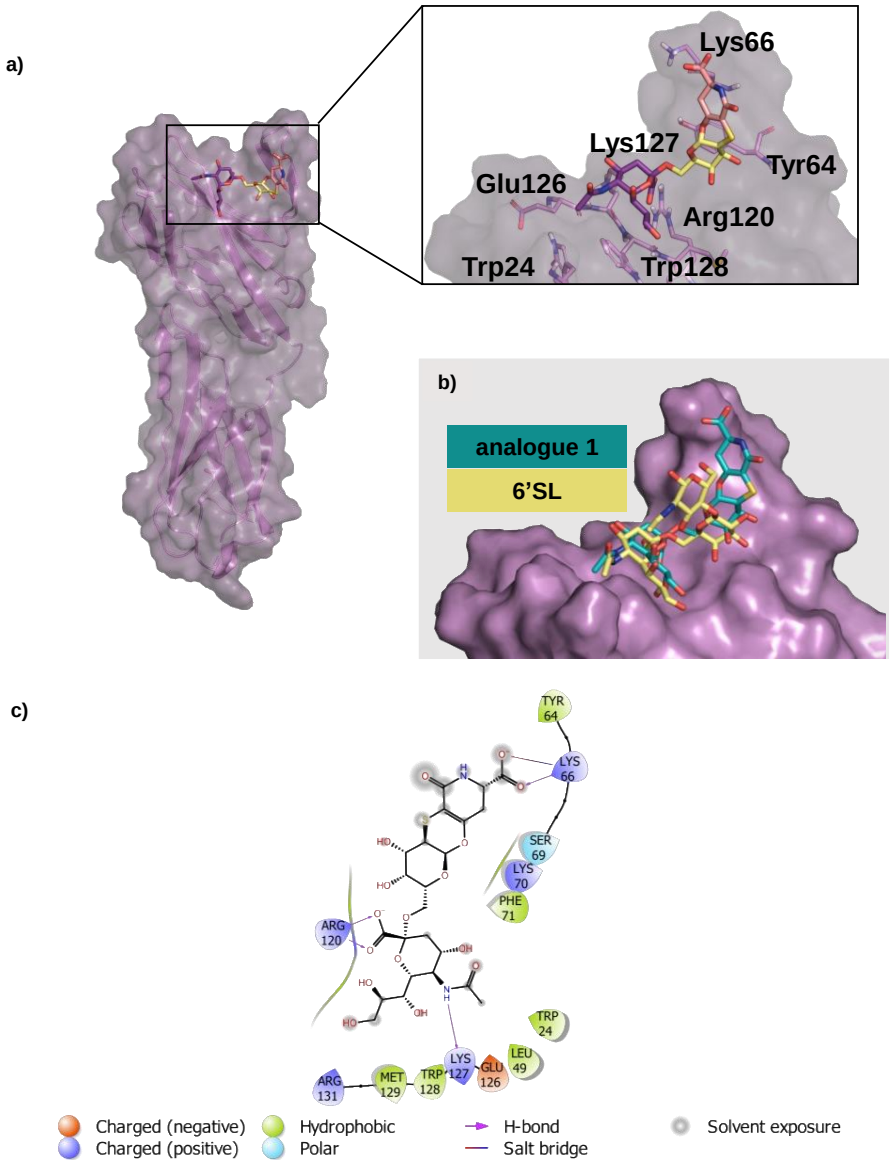


Figure 4. Interaction between *h*-CD22 and analogue *1* by molecular modelling. a) 3D model derived by docking and MD simulations for the analogue *1* bound to *h*-CD22 (PDB entry: 5VKM); the representative frame of the most populated MD cluster, obtained by K-means algorithm, was considered to depict the complex; b) superimposition of the previously obtained X-ray complex of *h*-CD22/6'sialyllactose (6'SL) and the analogue **1** bound model; c) two-dimensional plots representing the interactions between the analogue **1** and the binding site residues of *h*-CD22.

Differently from **1**, the aglycon moiety on the Gal unit did not contribute to the binding event. Concerning the saccharide moieties, very weak STD effects were observed for H1, H2, H4 protons of galactose unit (B). Moreover, medium (40-60 %) STD effects were detected for Neu5Ac (K) unit, specifically for H4, H5, H6, H8, H9R/S protons. Low STD effects were observed for H7 and the *N*-acetyl group of Neu5Ac residue. This suggested that, conversely from the natural ligand, the recognition is not driven by the Neu5Ac unit, but rather by the biphenyl moiety.

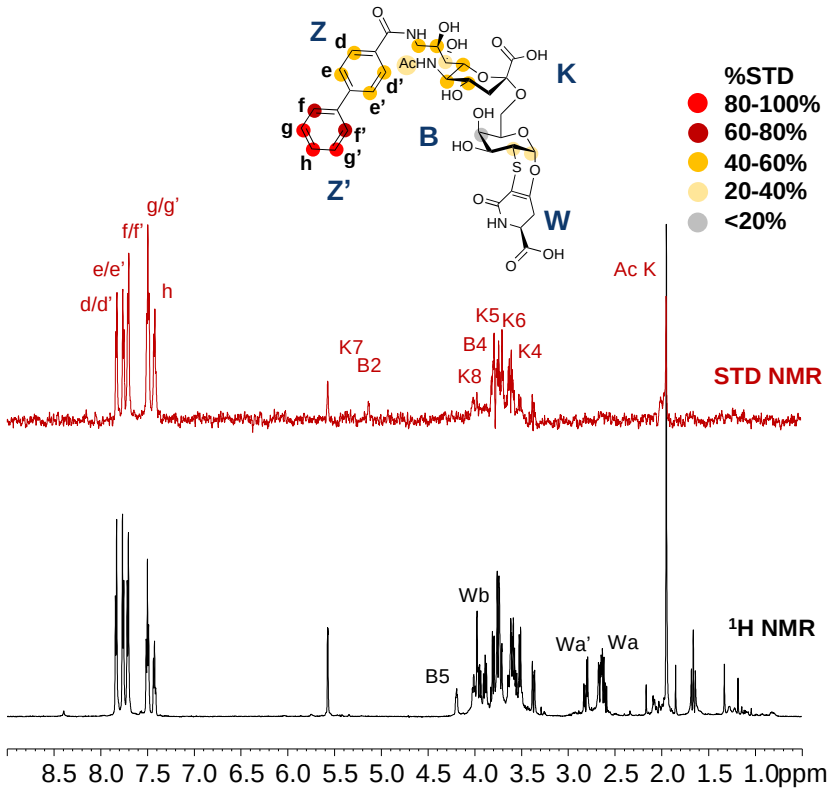


Figure 5. STD NMR analysis of analogue **2** in the interaction with CD22. Interacting epitope map of analogue **2** as derived by STD-NMR data (top panel). ^{1}H NMR and STD NMR spectra of *h*-CD22/analogue **2** mixture with a molecular ratio of 1:100, at 298 K (bottom panel).

These data were in agreement with the CD22/**2** complex (**Figure 6**) obtained by docking calculations. Interestingly, a close-up view of the less energetic cluster (**Figures 6** and **S3b**) highlighted that **2** was placed at the protein V-set domain, directing the biphenyl group toward the CC' loop of the receptor, stacked against Lys66 and Tyr64 residues of the loop. The Neu5Ac displayed few polar contacts with Arg120, Trp24 and Trp128 whereas the *N*-acetyl moiety was not oriented toward the canonical hydrophobic cleft comprising Trp24 and Trp128 and lacked the hydrogen bond with Glu126. Interestingly, the orientation of **2** was different from that observed for the natural sialoglycans and for **1** as well (**Figure 7**). Consistently with the STD-NMR, Gal and the aglycone moieties were not in contact with the receptor surface. These findings suggested that the presence of the biphenyl moiety, that promoted the favorable stacking interactions at the CC' loop, leads to a different orientation of the Neu5Ac from its canonical binding site.

5.3 Conclusions

The discovery of the essential roles played by *h*-CD22 in the regulation of immune cells and its expression pattern restricted to B cells has made this receptor an attractive therapeutic target.[28,29] With the aim to exploit *h*-CD22 ability to inhibit specific cells, several strategies have been developed and much efforts have been directed toward the development of CD22 high affinity ligands.[30-33]

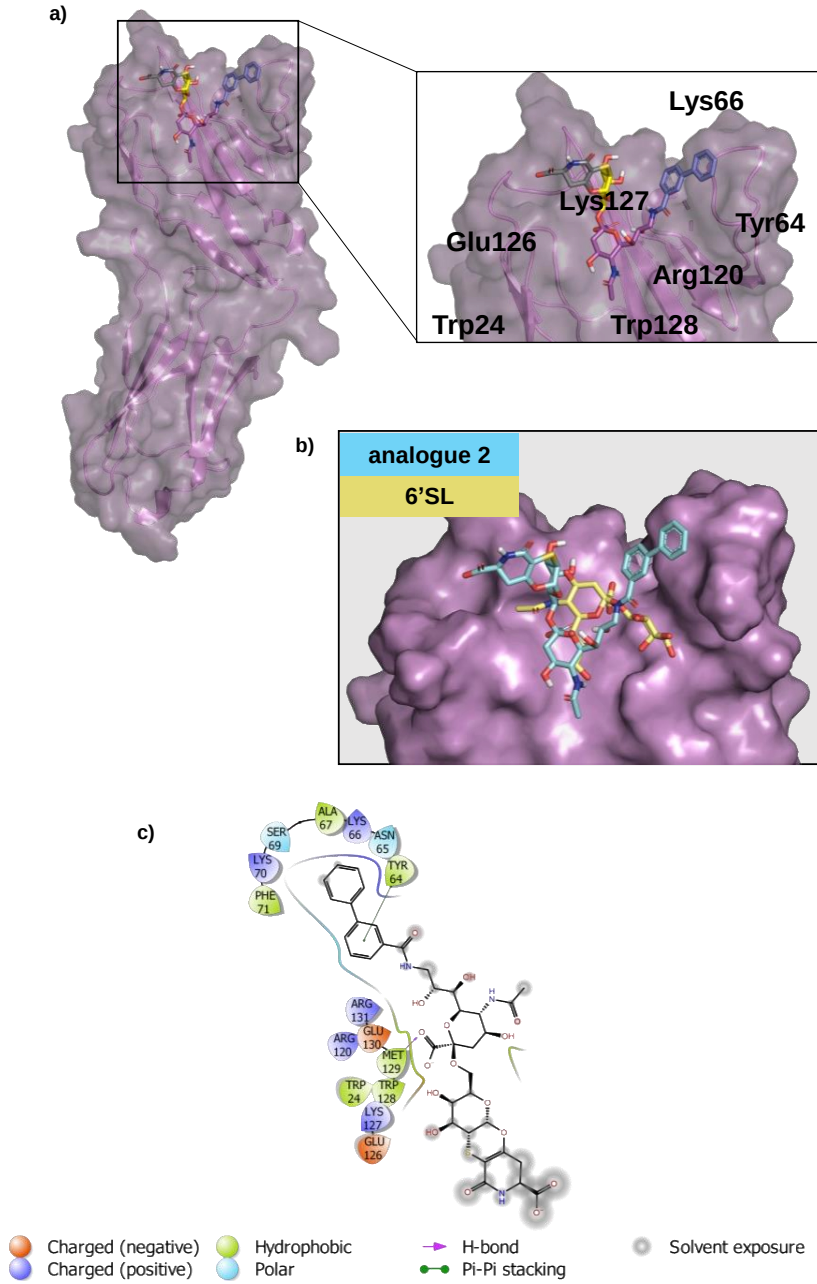


Figure 6. Interaction between CD22 and analogue **2** by molecular modelling. a) 3D model derived by docking calculations of **2** bound to *h*-CD22 (PDB entry: 5VKM). The lowest energy cluster binding mode was considered to depict the complex; b) superimposition of the previously obtained X-ray complex of *h*-CD22/6'SL and the analogue **2** bound model; c) two-dimensional plots representing the interactions between the analogue **2** and the binding site residues of *h*-CD22.

With the aim to further improve the design of CD22 modulators for therapeutic applications, we studied here two structurally constrained analogues, bearing functionalizable groups at the aglycon moiety allowing for the easy conjugation to bioavailable nanomaterials, in the interaction with *h*-CD22.

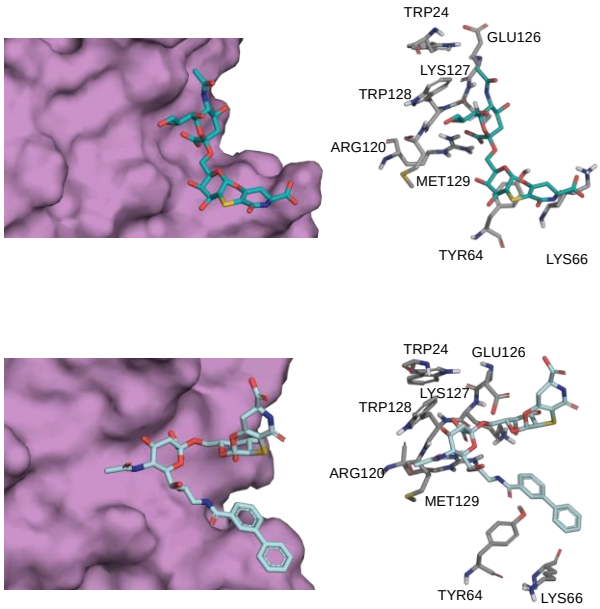


Figure 7. Comparison of the binding modes of analogues **1** and **2** upon interaction with *h*-CD22. Close up view of the analogues **1** (light blue sticks) and **2** (cyan sticks) into the *h*-CD22 binding site (violet molecular surface). The *h*-CD22 residues binding to the ligands are represented as grey sticks.

The characterization of the binding epitopes and affinities together with the determination of tridimensional models of the protein-ligand complexes have been carried out. In detail, the STD NMR investigation allowed to define the binding epitope of the sialyl derivatives; the bound conformation of **1** was also described by using NOE based data in combination with a computational approach.

Although the orientation around the Neu5Ac-α-2,6-Gal was the same as observed for natural CD22 ligands[7,8,34] also when interacting with other Siglecs, as Siglec-10,[35] our results suggested a more extended conformation of the analogue **1** if compared to the natural ligand due to the presence of the rigid aglycon moiety. Moreover, docking and modelling studies provided insights into the fine structural characteristics of the interacting interfaces. CD22 bound analogue **1** in the canonical sialic acid binding site, without relevant alterations of the interaction pattern and demonstrated that the rigid aglycone moiety, that was close to the CC' loop region of CD22 V-set domain, did not affect the presentation of the Neu5Ac-Gal epitope to the enzyme.[7]

Conversely, the STD NMR analysis carried out on **2**, bearing an aromatic extension at C9 of Neu5Ac, indicated that the strongest contribution to the binding was given by this moiety rather than the sialic acid unit itself. Docking studies allowed to identify a good model explaining and supporting the experimental results. Indeed, in the proposed model, the biphenyl group was accommodated at the V-set domain and oriented toward the CC' loop region where it formed aromatic interactions with Tyr64. This may be due to the steric hindrance given by the combination of the biphenyl and the aglycon moiety, that makes the sialic acid unit not optimally oriented within the receptor binding pocket, although it still established weaker polar interactions at the CD22 interface.

The interaction between *h*-CD22 and the sialo-analogues **1** and **2** were also proved by SPR and fluorescence assays. Interestingly, the slightly different K_D values align with the observations that in **1** the sialyl moiety was well oriented in the receptor binding pocket, without alteration of key interactions established by the natural epitope and with an increased

affinity thanks to the contribution of the rigid aglycon moiety. On the other hand, the binding mode of **2** was influenced by the presence of the biphenyl moiety that lead to a different accommodation of the sialic acid in the binding site of the protein with respect to the natural epitope. Despite this, a good increase of binding affinity was observed for **2** with respect to the natural ligand, mostly due to the π-π interactions between the ligand aromatic moiety and the receptor aromatic residues belonging to the CC' loop. Such interactions were added to the stabilizing interactions of the Neu5Ac moiety with polar the CD22 binding site polar residues.

The obtained results further highlighted the importance of the Siglecs CC' loop region for the recognition of sialoglycans,[36] thus suggesting that this region may be further exploited to design glycosyl analogues for CD22 immunomodulation and to fine tune its specificity and affinity. Overall, our structural and affinity data showed that the sialyl derivative **1** can strongly interact with *h*-CD22; thus, a class of related compounds may be designed as promising candidate for glycoimmunotherapy targeting CD22. Furthermore, the functionalization and characterization of suitable nanomaterials coated with **1**-like ligands may lead to interesting therapeutic applications, for example to efficiently probe the cell surfaces.[13,33]

5.4 Supporting information

5.4.1 Materials and methods

Protein expression and purification: The plasmids encoding for the three *N*-terminal Ig-like domains of human CD22 fused to the Fc region of mouse IgG2b was expressed in CHO cell lines and purified as described elsewhere.[34]

Synthesis and characterization of chemical materials: ESI-MS analyses were performed in negative ion mode and were recorded on an LCQ-Fleet Ion Trap equipped with a standard Ionspray interface. HRMS were performed on a Triple-TOF with a resolution of 35000 (FWHM). Chemical shifts are reported in part per million (δ) using the residual solvent line as secondary internal reference. ^{1}H NMR spectra were obtained at 500 MHz and 700 MHz. ^{13}C NMR spectra were obtained at 125 MHz and 175 MHz. Polarimetry analysis were performed on a Jasco DIP-370.

Synthesis of compound 3: To a solution of *N*-acetylneuraminic acid (6.00 g, 19.4 mmol) in 15 mL of DMF, cooled to 0 °C, 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) (4.40 mL, 29.5 mmol) and benzyl bromide (3.40 mL, 28.6 mmol) were added. The solution was stirred at room temperature for 18h and then poured into 1 L of cold CH_2Cl_2. The precipitate was collected by filtration, washed with CH_2Cl_2 (7 x 100 mL) and dried in vacuo, to afford 6.23 g of crude as a white solid. To a solution of crude (2.00 g, 5.20 mmol) in 20 mL of pyridine, cooled to 0 °C, TsCl (6.00 mL, 7.80 mmol) was added, and the solution was stirred at room temperature for 18h, then, Ac_2O (16 mL) and a catalytic amount of DMAP were added, and the solution was stirred for 5 h at room temperature. The reaction mixture was diluted with 200 mL of AcOEt and the organic layer washed with 1 M HCl and then with water. The organic phase was then collected, dried over anhydrous Na_2SO_4, filtered and evaporated, to afford 5.02 g of crude as a yellow oil. Crude was purified by flash-cromathography on silica gel (petroleum ether 10% in AcOEt) to afford **3** as a yellow amorphous solid (1.22 g, 40%). ^{1}H NMR (500 MHz, $CDCl_3$, δ = 7.26): δ 7.75 (d, *J* = 8.4 Hz, 2H, *Ar*); 7.38-7.29 (m, 7H, *Ar*); 5.56 (d, *J* = 9.7 Hz, N*H*Ac); 5.38 (dd, J_1 = 4.9 Hz, J_2 = 2.2 Hz, 1H, C*H*-7); 5.24 (td, J_1 = 10.9 Hz, J_2 = 4.9 Hz, 1H, C*H*-4); 5.16 (q, 2H, CH_2Ph); 5.03-5.00 (m, 1H, CH-6); 4.50 (dd, J_1 = 11.3 Hz, J_2 = 2.8 Hz, 1H, C*H*-9); 4.13-4.03 (m, 3H, C*H*-8, C*H*-5, C*H*-9); 2.51 (dd, J_1 = 13.3 Hz, J_2 = 5.1 Hz, 1H, C*H*-3$_{eq}$); 2.42 (s, 3H, SO2Ph*Me*); 2.08 (s, 3H, *Ac*); 2.06 (s, 3H, *Ac*); 2.03 (s, 4H, *Ac*, C*H*-3$_{ax}$); 2.00 (s, 3H, *Ac*); 1.95 (s, 3H, *Ac*); 1.87 (s, 3H, *Ac*). ^{13}C NMR (125 MHz, CDCl3, δ = 77.16): δ 171.32; 171.00; 170.58; 170.32; 168.37; 165.60; 144.83; 134.96; 133.10; 129.90; 128.71; 128.63; 128.3; 128.09; 97.53 (*C*-2); 72.75 (*C*H-8); 70.86 (*C*H-6); 68.29 (*C*H-4); 68.00 (*C*H_2Ph); 67.72 (*C*H_2-9); 67.54 (*C*H-7); 49.36 (*C*H-5); 36.02 (*C*H2-3); 23.26 (SO2Ph*Me*); 21.75; 21.16; 20.93; 20.78.

Synthesis of compound 4: To a solution of 3 in 15 mL of anhydrous DMF, NaN_3 (600 mg, 9.36 mmol) was added and the mixture was heated to 70 °C and stirred for 5h. Then, a second amount of NaN_3 (300 mg, 3.18 mmol) was added and the mixture was stirred at 70 °C for 1 h. The reaction mixture was diluted with 400 mL of AcOEt and the organic phase washed with brine. The organic layer was collected, dried over anhydrous Na_2SO_4, filtered and evaporated, to afford 1.10 g of crude as a brown solid. The crude was purified by flash-chromatography on silica gel (AcOEt) to afford pure 4 as an amorphous solid (450 mg, 50%). ^{1}H NMR (500 MHz, $CDCl_3$, δ = 7.26): δ 7.38-7.34 (m, 5H, *Ar*); 5.50 (d, *J* = 9.8 Hz, 1H, N*H*Ac); 5.37 (dd, J_1 = 3.4 Hz, J_2 = 2.5 Hz, 1H, C*H*-7); 5.26-5.14 (m, 3H, C*H*-4, CH_2Ph); 4.87 (dt, J_1 = 7.9 Hz, J_2 = 3.2 Hz, 1H, C*H*-8); 4.15 (q, *J* = 10.2 Hz, 1H, C*H*-6); 4.06 (dd, J_1

= 10.6 Hz, J_2 = 2.6 Hz, 1H, C*H*-5); 3.77 (dd, J_1 = 13.5 Hz, J_2 = 2.7 Hz, 1H, C*H*-9); 3.30 (dd, J_1 = 13.5 Hz, J_2 = 8.0 Hz, 1H, C*H*-9); 2.51 (dd, J_1 = 13.3 Hz, J_2 = 5.0 Hz, 1H, C*H*-3_{eq}); 2.14 (s, 3H, *Ac*); 2.11 (s, 3H, *Ac*); 2.05 (s, 3H, *Ac*); 2,04 (bs, 1H, C*H*-3_{ax}); 2.02 (s, 3H, *Ac*); 1.87 (s, 3H, *Ac*). ^{13}C NMR (125 MHz, $CDCl_3$, δ = 77.16): δ 171.11; 170.92; 170.51; 170.45; 168.53; 165.81; 134.87; 128.75; 128.48; 97.39 (*C*-2); 73.88 (*C*H-8); 73.46 (*C*H-5); 68.46 (*C*H-7); 68.35 (*C*H-4); 68.21 (CH_2Ph); 50.16 (CH_2-9); 49.26 (*C*H-6); 36.24 (CH_2-3); 23.36; 21.02; 20.95; 20.88; 20.80.

Synthesis of compound 5: To a solution of 4 (790 mg, 1.33 mmol) in 15 mL of anhydrous CH_2Cl_2, cooled to 0 °C, PhSH (160 µL, 1,55 mmol) and BF_3•Et_2O (402 µL, 3.26 mmol) were added. The solution was stirred at room temperature for 18h. The solution was diluted with 100 mL of CH_2Cl_2 and the organic layer washed with a saturated solution of $NaHCO_3$, dried over anhydrous Na_2SO_4, filtered, and evaporated, to give 820 mg of crude. Crude was purified by flash-chromatography on silica gel (petroleum ether 20% in AcOEt), to obtain 5 as a white amorphous solid (780 mg, 92%). ^{1}H NMR (500 MHz, $CDCl_3$, δ = 7.26): δ 7.36-7.24 (m, 10H, *Ar*); 5.89 (d, J = 9.8 Hz, 1H, N*H*Ac); 5.43 (t, *J* = 2.0 Hz, 1H, C*H*-7); 5.39 (td, , J_1 = 11.0 Hz, J_2 = 4.7 Hz, 1H, C*H*-4); 5.05 (dd, , J_1 = 63.38 Hz, J_2 = 12.2 Hz, 2H, CH_2Ph); 4.67 (dt, J_1 = 9.0 Hz, J_2 = 1.9 Hz, 1H, C*H*-8); 4.57 (dd, J_1 = 10.6 Hz, J_2 = 2.4 Hz, 1H, C*H*-6); 4.16-4.13 (m, 1H, C*H*-5); 3.52 (dd, J_1 = 13.6 Hz, J_2 = 1.4 Hz, 1H, C*H*-9); 3.20 (dd, J_1 = 13.6 Hz, J_2 = 9.5 Hz, 1H, C*H*-9); 2.68 (dd, J_1 = 13.8 Hz, J_2 = 4.8 Hz, 1H, C*H*-3_{eq}); 2.14 (bs, 1H, C*H*-3_{ax}); 2.09 (s, 3H, *Ac*); 2.06 (s, 3H, *Ac*); 2.03 (s, 3H, *Ac*); 1.87 (s, 3H, *Ac*). ^{13}C NMR (125 MHz, $CDCl_3$, δ = 77.16): δ 171.28; 171.05; 170.49; 170.29; 167.55; 135.88; 130.11; 129.47; 128.72; 128.69; 88.66 (*C*-2); 74.78 (*C*H-8); 73.24 (*C*H-6); 69.14 (*C*H-4); 68.88 (*C*H-7); 67.88 (CH_2Ph); 49.74 (*C*H-9); 49.37 (*C*H-5); 37.64 (CH_2-3); 23.18; 21.10; 20.96; 20.82.

Synthesis of compound 7: Under nitrogen atmosphere, to a solution of 5 (780 mg, 1.21 mmol) and 6[22] (1.08 g, 4.90 mmol) in 80 mL of CH_2Cl_2, PPh_3 was added (650 mg, 2.87 mmol). The solution was stirred for 48h and then the solvent was removed, to obtain 2.34 g of crude as a white solid. Crude was purified by flash-chromatography on silica gel (petroleum ether 20% in AcOEt), to obtain pure 7 as an amorphous white solid (278 mg, 30%). ^{1}H NMR (500 MHz, $CDCl_3$, δ = 7.26): δ 7.77 (d, *J* = 8.3 Hz, 2H, *Ar*); 7.64-7.62 (m, 4H, *Ar*); 7.48 (t, *J* = 8.0 Hz, 2H, *Ar*); 7.41-7.38 (m, 1H, *Ar*); 7.32 (d, *J* = 1.5 Hz, 5H, *Ar*); 7.27-7.25 (m, 2H, *Ar*); 7.16-7.13 (m, , 3H, *Ar*); 6.77 (t, *J* = 6.4 Hz, 1H, N*H*); 5.69 (d, *J* = 10.3 Hz, 1H, N*H*Ac); 5.45-5.40 (m, 2H, C*H*-4, C*H*-7); 5.11 (bs, 2H, CH_2Ph); 4.92 (q, *J* = 4.0 Hz, 1H, C*H*-8); 4.67 (dd, J_1 = 10.6 Hz, J_2 = 2.6 Hz, 1H, C*H*-6); 4.22 (q, *J* = 10.4 Hz, 1H, C*H*-5); 3.94 (ddd, J_1 = 15.0 Hz, J_2 = 6.8 Hz, J_3 = 3.3 Hz, 1H, C*H*-9); 3.31, (dt, J_1 = 15.0 Hz, J_2 = 5.4 Hz, 1H, C*H*-9); 2.67 (dd, J_1 = 13.8 Hz, J_2 = 4.8 Hz, 1H, C*H*-3_{eq}); 2.21-2.15 (m, 1H, C*H*-3_{ax}); 2.13 (s, 3H, *Ac*); 2.03 (s, 3H, *Ac*); 2.02 (s, 3H, *Ac*); 1.89 (s, 3H, *Ac*). ^{13}C NMR (125 MHz, $CDCl_3$, δ = 77.16): δ 171.33; 161.15; 171.07; 171.01; 167.76; 167.06; 144.21; 140.23; 136.13; 134.81; 132.26; 132.18; 132.12; 129.96; 129.37; 129.08; 128.80; 128.75; 128.71; 128.61; 128.11; 127.78; 127.30; 127.15; 88.62 (*C*-2); 73.31 (*C*H-8); 72.79 (*C*H-6); 69.23 (*C*H-4); 69.16 (*C*H-7); 67.95 (CH_2Ph); 49.62 (*C*H-5); 39.31 (*C*H-9); 37.36 (CH_2-3); 23.27; 21.26; 21.07; 21.00.

Synthesis of compound 9: Under nitrogen atmosphere, to a solution of 7 (65 mg, 0.140 mmol) and 8[23] (278 mg, 0.350 mmol) in 1.5 mL of an anhydrous mixture of CH_3CN/CH_2Cl_2 (10:1), cooled at -40 °C, NIS (132 mg, 0.580 mmol) and TfOH (25 µL, 0.280 mmol) were added. The reaction mixture was stirred at -40 °C for 4 h, then Et_3N was added (150 µL) and the mixture was let slowly return to room temperature. The reaction mixture was diluted with

20 mL of CH_2Cl_2 and the organic layer washed with a 10% $Na_2S_2O_3$ solution. The organic layer was collected, dried over anhydrous Na_2SO_4, filtered and solvent evaporated, to obtain 275 mg of crude. Crude was purified by flash-chromatography on silica gel (AcOEt), to obtain 175 mg of a partially purified mixture. The mixture was dissolved in 10 mL of AcOH 80% and heated at 45 °C for 18h. The solvent was removed and crude purified by flash-chromatography on silica gel (MeOH 5% in AcOEt), to obtain pure 9 as an amorphous white solid (60 mg, 40%) ^{1}H NMR (500 MHz, $CDCl_3$, δ = 7.26): δ 7.85 (d, *J* = 8.4 Hz, 2H, *Ar*); 7.65 (d, *J* = 8.4 Hz, 2H, *Ar*); 7.61-7.59 (m, 2H, *Ar*); 7.46 (t, *J* = 7.3 Hz, 2H, *Ar*); 7.39-7.31 (m, 11H, *Ar*); 6.99 (dd, , J_1 = 8.4 Hz, J_2 = 3.9 Hz, 1H, N*H*); 6.10 (s, 1H, N*H*); 5.56 (d, *J* = 2.56 Hz, 1H, C*H*-1); 5.40 (d, J = 10.0 Hz, 1H, N*H*Ac); 5.25-5.17 (m, 6H, C*H*-7', C*H*-8', CH_2Ph x 2); 4.88-4.83 (m, 1H, C*H*-4'); 4.34 (ddd, J_1 = 14.9 Hz, J_2 = 8.5 Hz, J_3 = 2.4 Hz, 1H, C*H*-9'); 4.26 (ddd, J_1 = 10.1 Hz, J_2 = 6.0 Hz, J_3 = 1.2 Hz, 1H, C*H*-b); 4.17 (q, 1H, *J* = 10.4 Hz, C*H*-5'); 4.09-4.06 (m, 1H, C*H*-6); 4.05-4.00 (m, 2H, C*H*-4, C*H*-5); 3.95 (bs, 1H, C*H*-6'); 3.72 (dd, J_1 = 9.5 Hz, J_2 = 4.7 Hz, 1H, C*H*-6); 3.64-3.61 (m, 2H, C*H*-3, O*H*); 3.47 (dd, J_1 = 10.8 Hz, J_2 = 2.5 Hz, 1H, C*H*-2); 3.47 (bs, 1H, O*H*); 3.09 (dt, J_1 = 14.9 Hz, J_2 = 4.6 Hz, 1H, C*H*-9'); 2.85-2.72 (m, 2H, CH_2-a); 2.67 (dd, J_1 = 12.7 Hz, J_2 = 4.9 Hz, 1H, C*H*-3'$_{eq}$); 2.22 (s, 3H, *Ac*); 2.10 (s, 3H, *Ac*); 2.02 (s, 4H, *Ac*, C*H*-3'$_{ax}$); 1.87 (s, 3H, *Ac*). ^{13}C NMR (125 MHz, $CDCl_3$, δ = 77.16): δ 171.95; 171.03; 170.89; 170.47; 169.52; 167.51; 167.43; 165.09; 155.28; 144.60; 140.02; 134.91; 134.82; 132.75; 129.09; 128.94; 128.90; 128.88; 128.66; 128.55; 128.20; 127.65; 127.41:127.31; 98.96 (*C*-2'); 97.07; 96.39 (*C*H-1); 73.10 (*C*H-4); 71.85 (*C*H-5); 70.19 (*C*H-5'); 69.09 (*C*H-7'); 68.41 (*C*H-4'); 68.25 (*C*H-8'); 68.03 (*C*H-6', CH_2Ph x 2); 65.88 (*C*H-3); 64.20 (*C*H-2); 51.56 (*C*H-b); 49.40 (*C*H_2-9'); 39.43 (*C*H_2-6); 37.49 (*C*H_2-3'); 30.97 (*C*H_2-a); 23.28; 21.27; 21.20; 20.98.

Synthesis of sialo-derivative 2: To a solution of 9 (20 mg, 0.017 mmol) in 4 mL of MeOH, Pd/C was added (24 mg). The suspension was stirred for 73h under hydrogen atmosphere, then, the suspension was filtered on an HPLC filter and the solution evaporated to afford 7 mg of crude. Crude was dissolved in 1 mL of NH_3 4M in MeOH and the solution was stirred at room temperature for 120h. The solution was evaporated and the crude was suspended in 1 mL of Et_2O/MeOH (6:4) and centrifugated (5' at 3000 rpm), the supernatant was eliminated and the procedure repeated 10 times. The purified product was dried under high vacuum to obtain pure 2 (4 mg, 30%). ^{1}H NMR (700 MHz, D_2O, δ = 4.79): δ 7.90 (d, *J* = 8.2 Hz, 2H, *Ar*); 7.83 (d, *J* = 8.2 Hz, 2H, *Ar*); 7.78 (d, *J* = 8.2 Hz, 2H, *Ar*); 7.57 (t, *J* = 7.6 Hz, 2H, *A*r); 7.49 (t, *J* = 7.3 Hz, 1H, *Ar*); 5.63 (d, *J* = 8.2 Hz, 2H, C*H*-1); 4.26 (dd, J_1 = 6.9 Hz, J_2 = 3.9 Hz, C*H*-8'); 4.09-4.07 (m, 1H, C*H*-5); 4.04 (d, *J* = 2.8 Hz, 1H, C*H*-7'); 4.02 (dd, J_1 = 9.5 Hz, J_2 = 6.7 Hz, 1H, C*H*-9'); 3.96 (dd, J_1 = 10.7 Hz, J_2 = 7.7 Hz, 1H, C*H*-a); 3.88 (t, *J* = 9.9 Hz, 1H, C*H*-5'); 3.82-3.78 (m, 3H, C*H*-3, C*H*-6, C*H*-6'); 3.70-3.63 (m, 3H, C*H*-4', C*H*-4, C*H*-9'); 3.59 (dd, J_1 = 8.7 Hz, J_2 = 1.3 Hz, 1H, C*H*-6); 3.44 (dd, J_1 = 11.4 Hz, J_2 = 2.9 Hz, 1H, C*H*-2); 2.87 (dd, J_1 = 16.8 Hz, J_2 = 6.6 Hz, 1H, C*H*-b); 2.73 (dd J_1 = 12.2 Hz, J_2 = 4.6 Hz, 1H, C*H*-3$_{eq}$); 2.68 (dd, J_1 = 17.2 Hz, J_2 = 9.9 Hz, 1H, C*H*-b); 2.02 (s, 3H, *Ac*); 1.73 (t, *J* = 12.0 Hz, 1H, C*H*-3'$_{ax}$). ^{13}C NMR (175 MHz, D_2O): δ 174.97; 173.40; 170.66; 167.36; 158,11; 144.12; 139.46; 132.44; 129.19; 128.39; 127.82; 127.17; 127.12; 100.67; 96.02 (*C*H-1); 94.01; 72.59 (*C*H-3); 72.19 (CH-8'); 70.23 (*C*H-5); 70.13 (*C*H-4); 68.54 (*C*H-7'); 68.18 (*C*H-b); 65.15 (*C*H-4'); 64.10 (*C*H-6'); 52.86 (*C*H-9'); 51.86 (*C*H_2-5'); 42.97 (*C*H_2-6); 40.21 (*C*H_2-3'); 38.13 (*C*H-2); 31.11(*C*H-a); 22.00. ESI-MS m/z (%): 400.83 (100%) [M-2H/]$^{2-}$, HRMS (m/z): [M-2H/]$^-$, calcd. for $C_{36}H_{40}O_{16}$ N_3S, 802.21348, found 802.21155; $[\alpha]_D^{21°C}$ = +6 (0.5 mg/mL in H_2O).

Surface Plasmon Resonance (SPR) analysis: The SPR measurements were performed on a Biacore X100 instrument (Cytiva, Global Life Sciences Solutions, Marlborough, USA). Protein A (10600-P07E, Sino Biological Inc., Beijing, China) was immobilized on both flow cells (FC1 and FC2) of a gold sensor chip (Cytiva) reaching ~1200 response units (RU) by using a protein solution of 30 mg L^{-1} in 10 mM acetate buffer pH 5.0 injected over the gold surface for 10 min at a flow rate of 10 µL min^{-1}. CD22 protein was captured on the sensor chip injecting 40 mg L^{-1} of CD22 in 10 mM acetate buffer pH 5 over FC2 at a flow rate of 5 µL min^{-1} for 3 min, and using FC1 as the reference surface; Both flow cells were equilibrated with HBS-EP buffer overnight at a flow rate of 5 µL min^{-1}, achieving for FC2 ~ 5000 RU. Twofold dilution series of the sialic acid analogues, the analytes, were freshly prepared in HBS-EP running buffer. All binding experiments were performed at 25 °C at a flow rate of 30 µL min^{-1}. The samples were injected for 1 min followed by 1 min dissociation. Each sample concentration was measured in triplicate. Double referencing was applied to correct for bulk effects and other systematic artifacts (subtraction of reference surface and blank injections). Data processing was performed by using the Biacore X100 evaluation software. The dissociation constant (K_D) for the analyte interaction with CD22 was determined according to the common 1:1 binding model described by the equation:

$$\mathrm{RU} = \mathrm{RU_{max}} \times \frac{[\mathrm{Analyte}]}{[\mathrm{Analyte}] + \mathrm{K_D}}$$

where RU_{max} is the maximum SPR response, and K_D corresponds to the analyte concentration that gives half of the maximum SPR response, i.e., $RU_{max}/2$. [37,38]

Fluorescence analysis: The experiments of steady-state fluorescence spectroscopy have been carried out on a Fluoromax-4 spectrofluorometer (Horiba, Edison, NJ, USA) at the fixed temperature of 5 °C; an excitation at 280 nm was used and emission spectra were recorded in the range of 290–500 nm. The slit widths were fixed at 4 nm for the excitation and 5 nm for the emission wavelength. A quartz cuvette with a path length of 1 cm and a chamber volume of 1 mL was used under constant stirring. 0.9 mL of CD22 solution at fixed concentration of 0.25 µM was titrated by adding small volumes (1–20 µL of a ligand stock solution of 500 µM) of analogue **1**. The PBS buffer at pH 7.4 was used for all solutions. The optical density of the solution at the excitation wavelength was kept less than 0.05.
The data were analyzed by non-linear regression with One Site- Specific Binding model for the determination of the dissociation constant (K_D) as implemented in OriginPro 2016, according to the following equation. [39]

$$\mathrm{Y} = \frac{\mathrm{B_{max}} * \mathrm{X}}{\mathrm{K_D} + \mathrm{X}}$$

where X stands for the ligand concentration, Y is the change of the fluorescence intensity at the maximum wavelength, B_{max} represents the maximum specific binding and K_D is the equilibrium dissociation constant.

NMR: NMR spectra were acquired on a Bruker 600-MHz Avance Neo instrument fitted with a cryo probe. NMR samples were dissolved in 50 mM deuterate phosphate buffer (NaCl 140 mM, Na_2HPO_4 10 mM, KCl 3 mM, pH 7.4) and the [D4](trimethylsilyl)propionic acid, sodium salt (TSP, 10 uM) was used as internal reference to calibrate all the spectra. Data acquisition and processing were analyzed using TOPSPIN 3.2 software. The chemical shifts of the glycan ligands were assigned by ^{1}H, COSY, TOCSY, NOESY and HSQC experiments.

Homonuclear 2D ^{1}H-^{1}H NOESY experiments were carried out by using data sets of 2048x512 points and mixing times of 300 ms.

^{1}H NMR spectra were registered by using 16 k and 32 k data points and zero-filled up to 64 k data points prior to processing. The 2D homonuclear spectra were recorded with data sets of 4096x512 (t1 x t2) points and the data matrix processed with zero-filling in the F1 dimension up to 4096x2048 points. In order to improve the resolution, a cosine-bell function was used before Fourier transformation in both dimensions. Heteronuclear single quantum coherence (HSQC) experiments were carried out by using the sequence "hsqcedetgpsisp" from the Bruker library, setting data points of 2048x256.

The STD NMR spectra were acquired with a number of 64 transients, in addition to 32 scans to allow the sample to come to equilibrium, and 64K data points. A protein:ligand ratio of 1:100 and a saturation time of 2 s were used with the on-resonance pulse at 6.5 ppm and the off-resonance at 40ppm. By using these conditions, no STD signals were observed in the control STD NMR spectrum of the ligand alone. A train of 50ms (field strength of 21 Hz) Gaussian shaped pulse with an attenuation of 60db has been used to saturate the protein.

The epitope mapping of ligands 1 and 2 was achieved by the calculation of the ratio (I0 – Isat)/I0, where (I0 – Isat) is the intensity of the signal in the STD NMR spectrum and I0 is the peak intensity referred to the unsaturated reference spectrum (off-resonance).

Docking:

1.Preparation of the macromolecules:

The crystal structure of *h*-CD22 extracellular domain was used for docking purposes (PDB entry: 5VKM). The structure was submitted to 100000 steps of steepest descent minimization with OPLS3 force field using MacroModel [40] before being used for docking calculations.

Building of ligands. The 3D coordinates of analogues 1 and 2 were built using Maestro program.[41] The ligands geometries were optimized by 100000 step of steepest descent minimization with OPLS3 force field using Macro Model.[40] Ligands were prepared for docking calculations using AutoDockTools setting all rotatable bonds free to move during the calculations except for the glycosidic bonds.

2.Docking calculations: Docking calculations of all compounds were performed using AutoDock 4.2.2.[42] Analysis of the docking poses was performed with AutoDockTools. The docking protocol was validated by carrying out the docking of CD22 crystallographic structure in complex with Neu5Ac-α-2,6-Gal ligand (PDB entry: 5VKM). The 3D structure of Neu5Ac-α-2,6-Gal was extracted from the crystallographic structure of CD22.The grid point spacing was set at 0.375 Å, and a hexahedral box was built with x, y, z dimensions: 64 Å, 46 Å, 56 Å centered in the centroid position among the binding site CD22 residues. A total of 200 runs using Lamarckian Genetic algorithm was performed, with a population size of 100, and 250000 energy evaluations. After docking, the 200 solutions were clustered in groups with root-mean-square deviation less than 1.0 Å. The clusters were ranked according to the lowest energy representative of each cluster.

3.MD simulations: Molecular dynamics calculations were performed within the AMBER 18 software package in explicit water using the following forcefields: Glycam06j-1 for the glycans, FF14SB for the proteins and Gaff for organic moieties.[43]

All the oligosaccharides were built up and minimized by using Maestro package and the carbohydrate builder utility of the glycam website (www.glycam.com),[44] and then the torsional angles were set to the values obtained through the molecular mechanics

calculations. For the protein preparation, missing hydrogen atoms were added, and protonation state of ionizable groups and cap termini were computed by using Maestro Protein Preparation Wizard.[45] For the preparation of analogue **1**, xleap antechamber[46] and parmchk2 modules implemented in AMBER were used in order to parametrize the molecule.

The input files were generated using the tleap modules of the AMBER package, the minimization steps were performed using Sander module and molecular dynamic calculations were performed using the PMEMD module.

The corresponding molecules (free and bound analogue **1**) were positioned within an octahedral box of TIP3P water of the proper size and the remote interactions were calculated using a cut off of 10 Angstroms and Counterions were added to neutralize the system. After the preparation of the input files, an energy minimization process was performed to refine the initial structure. The calculations employed SHAKE for the C-H bonds and 1 fs of integration step. Periodic boundary conditions were applied, as well as the smooth particle mesh Ewald method to represent the electrostatic interactions, with a grid space of 1 Å. The system was minimized, at first, holding the solute fixed, while a second minimization was performed on the entire system. Furthermore, the whole system was slowly heated from 0 to 300 K using a weak restrain on the solute and then, the system was equilibrated at 300 K using constant pressure and removing the restrains on the solute. The system coordinates were saved and used for the 100ns simulations using the PMEMD module implemented in AMBER. Coordinate trajectories were recorded each 2 ps throughout production runs, yielding an ensemble of 10000 structures for each complex, which were finally analyzed. The stability of energy, pressure, temperature and other thermodynamic parameters were monitored along the trajectory and then, RMSD, torsions, clusters distances and hydrogen bonds were extracted. Cpptraj [47] (Roe, D. R., & Cheatham, T. E., 3rd; 2013). was the utility used for analyzing and processing trajectories and coordinate files created from the MD simulations. VMD software was used to visualize the trajectory.[48].

5.4.2 Supporting figures

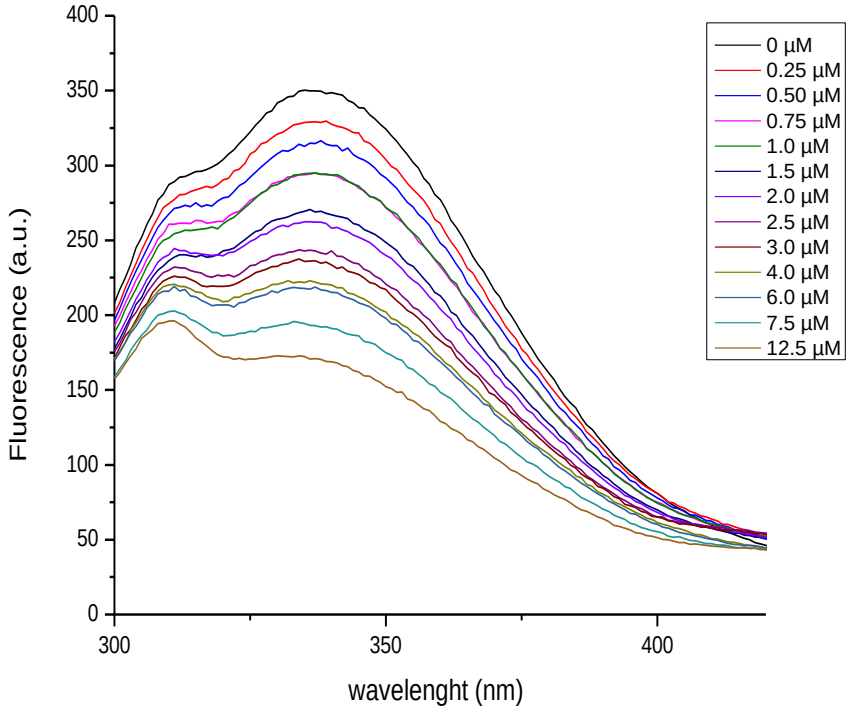


Figure S1. Fluorescence spectra of *h*-CD22 (black line) in the presence of increasing amounts of **1** (colored lines).

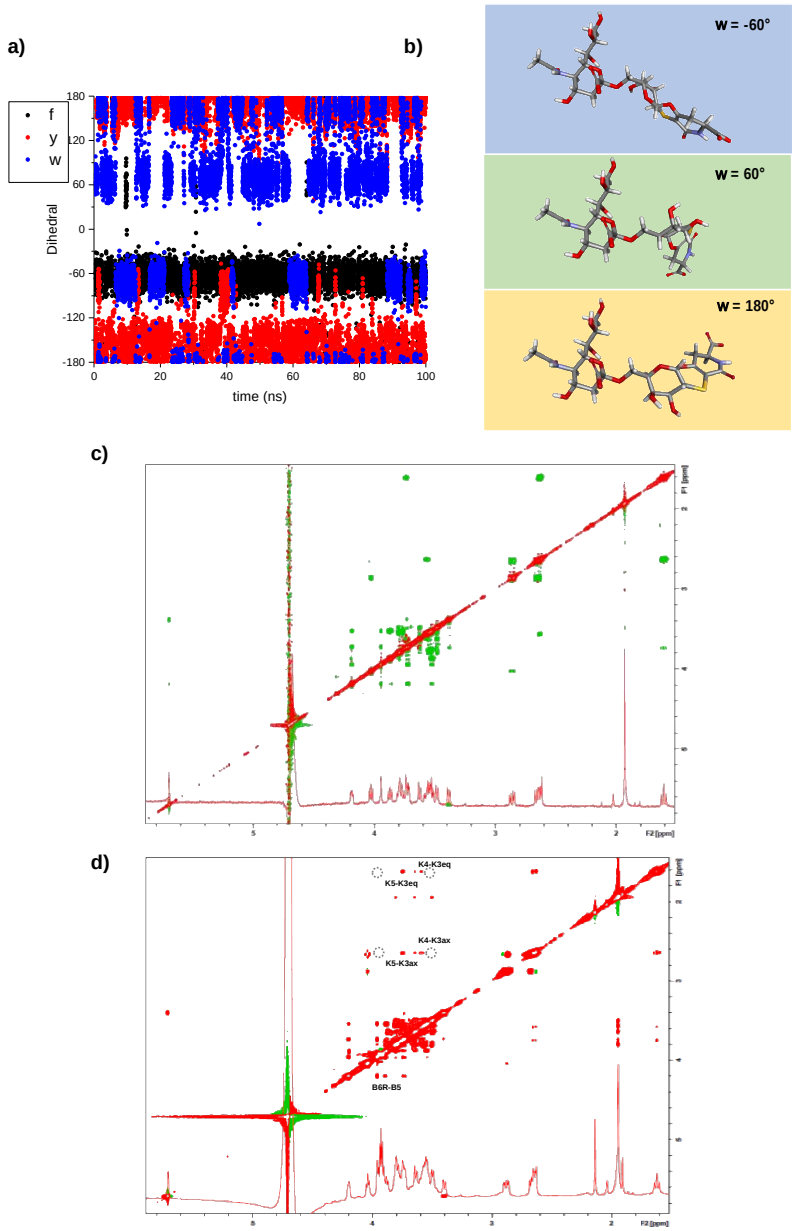


Figure S2. a) Dihedral angle fluctuations around the Neu5Ac-Gal linkage along the MD simulation of **1** in the free state; b) 3D representation of the main conformations of **1** according to the three different values of ω torsion; c) T-ROESY (mixing time: 300ms) of **1** in the free state; d) tr-NOESY of 1:30 *h*-CD22/**1** mixture (T=298K).

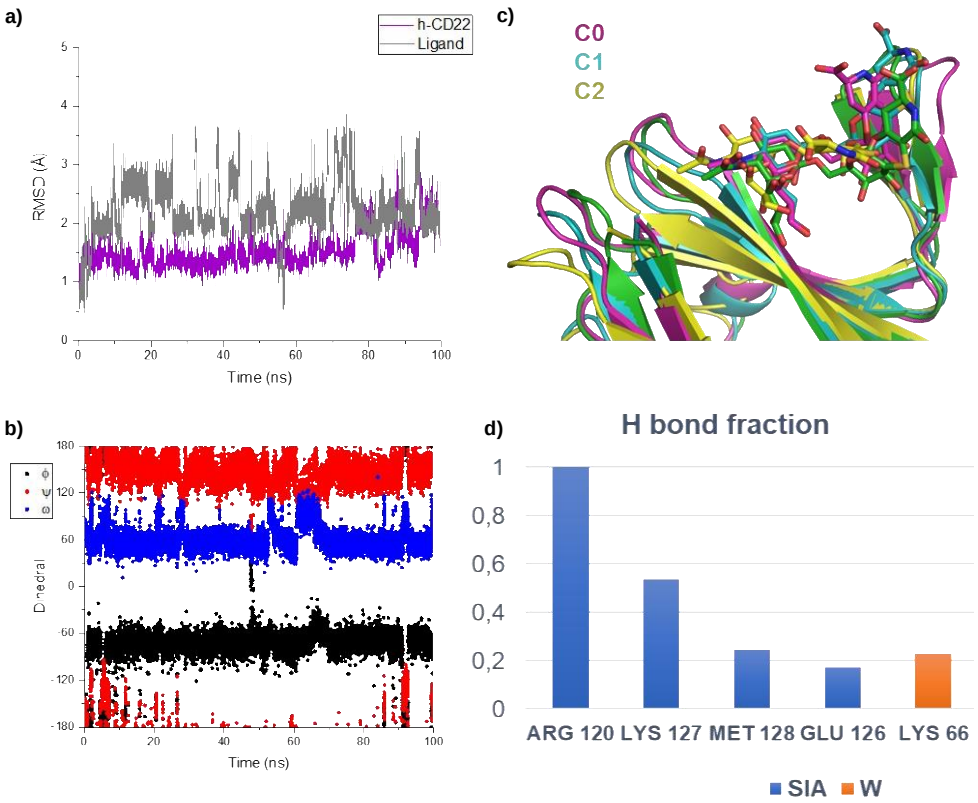


Figure S3. MD simulation analysis *h*-CD22/**1** complex. a) Protein and ligand RMSD. The ligand RMSD was calculated in reference to the protein; b) superimposition of the three most populated clusters from the MD simulation. K-mean algorithm was considered for clustering; c) fluctuation of Neu5Ac-Gal dihedral angles of **1** along the MD simulation; d) protein/ligand H bonds fraction.

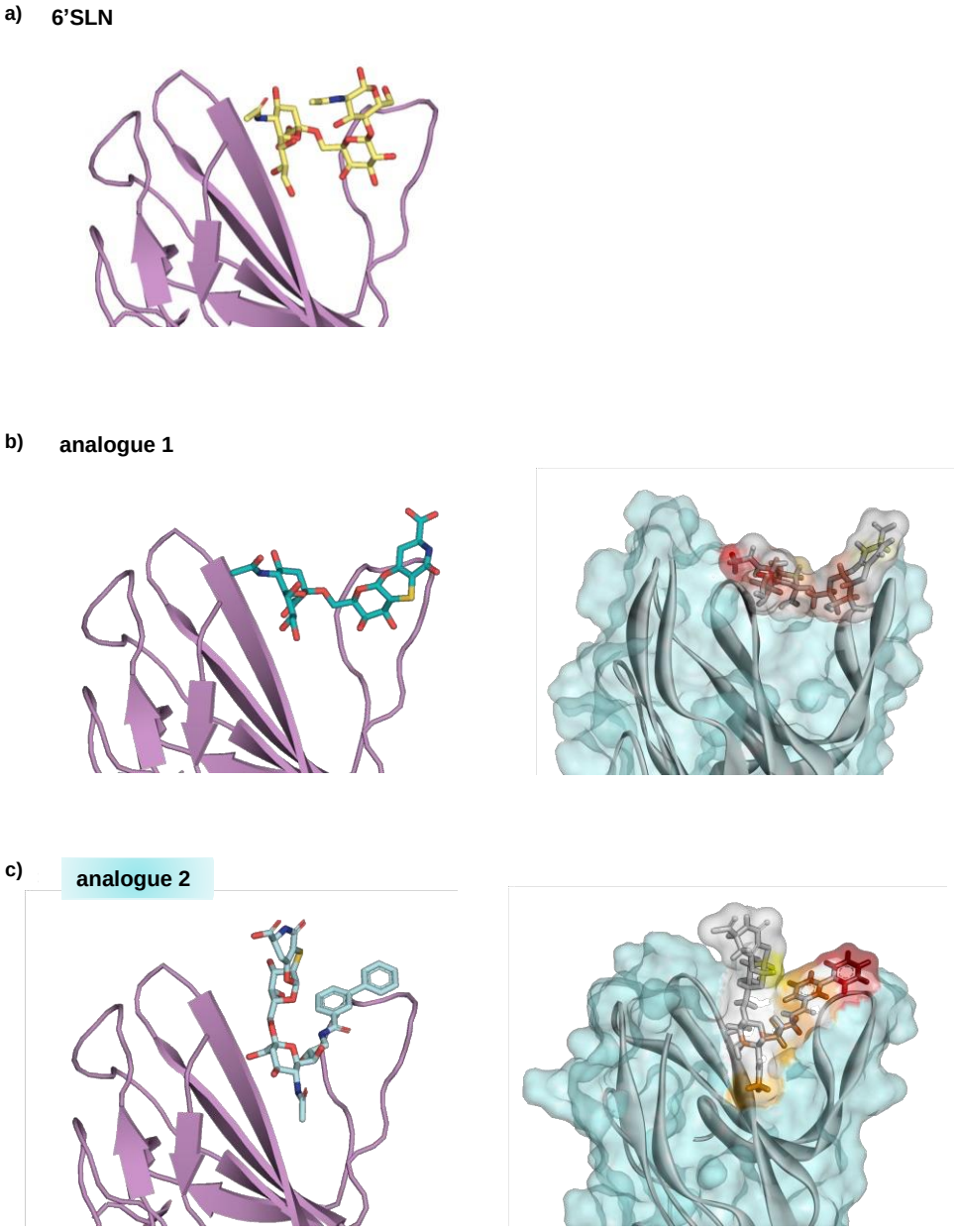


Figure S4. 3D view of the *h*-CD22/6'SLN, *h*-CD22/**1** and *h*-CD22/**2** complexes. On the right: Representations of the binding pose of 6'SLN (a), **1** (b) and **2** (c) in the binding pocket of *h*-CD22. On the left: 3D view of *h*-CD22/**1** (b) and *h*-CD22/**2** (c) complexes in accordance with STD, tr-NOESY and MD data. The ligand surface is colored according to the STD effects.

5.5 Bibliography

[1]Schnaar R.L., *J. Leukocyte Biol*., **2016**, 99, 825-838.

[2]Müller J., Nitschke L., *Nat. Rev. Rheumatol*., **2014**, 10, 422-428.

[3]Duan S., Paulson J.C., *Annu. Rev. Immunol*., **2020**, 38, 365-395.

[4]Crocker P.R., Varki A., *Immunology*, **2011**, 103, 137-145.

[5]Pillai S., Netravali I.A., Cariappa A., Mattoo H., *Annu. Rev. Immunol*., **2012**, 30, 357-392.

[6]Nitschke L.,*Glycobiology*, **2014**, 24, 807-817.

[7]Ereño-Orbea J., Sicard T., Cui H., Mazhab-Jafari M.T., Benlekbir,A.Guarné S., Rubinstein J.L., Julien J.P., *Nat. Commun*., **2014**, 8, 764.

[8]Di Carluccio C., Crisman E., Manabe Y., Forgione R.E., Lacetera A., Amato J., Pagano B., Randazzo A., Zampella A., Lanzetta R., Fukase K., Molinaro A., Crocker P.R., Martín-Santamaría S., Marchetti R., Silipo A., *ChemBioChem*, **2020**, 21, 129-140.

[9]Gimeno A., Reichardt N.C., Cañada F.J., Perkams L., Unverzagt C., Jiménez-Barbero J., Ardá A., *ACS Chem. Biol.,* **2017**,12, 1104-1112.

[10]Ai J., Advani A., *Br. J. Haematol*., **2015**,168, 471-480.

[11]Angata T., Nycholat C.M., Macauley M.S., *Trends Pharmacol.Sci*., **2015**, 36, 645–660.

[12]Gottenberg J.E., Dörner T., Bootsma H., Devauchelle-Pensec V., Bowman S.J., Mariette X., Bartz H., Oortgiesen M., Shock A., Koetse W., Galateanu C., Bongardt S., Wegener W.A., Goldenberg D.M., Meno-Tetang G., Kosutic G., Gordon C., *Arthritis Rheumatol*., **2018**, 70, 763-773.

[13]Büll C., Heise T., Adema G.J., Boltje T.J., *Trends Biochem. Sci*., **2016**, 41, 519-531.

[14]Mesch S., Lemme K., Wittwer M., Koliwer-Brandl H., Schwardt O., Kelm S., Ernst B., *ChemMedChem*, **2012**, 7, 134-143.

[15]Kelm S., Madge P., Islam T., Bennett R., Koliwer-Brandl H., Waespy M.,von Itzstein M., Haselhorst T., *Angew. Chem. Int. Ed.,* **2013**, 52, 3616-3620; *Angew. Chem*., **2013**, 125, 3704-3708.

[16]Prescher H., Schweizer A., Kuhfeldt E., Nitschke L., Brossmer R., *ACS Chem. Biol*., **2014**, 9, 1444-1450.

[17]Prescher H., Schweizer A., Kuhfeldt E., Nitschke L., Brossmer, *ChemBioChem*, **2017**, 18, 1216-1225.

[18]Papi F., Pâris A., Lafite P., Daniellou R., Nativi C., *Org. Biomol. Chem*., **2020**,18, 7366-7372.

[19]Ardá A., Bosco A., Sastre J., Cañada F.J., André S., Gabius H.-J.J., Richichi B., Jiménez-Barbero J., Nativi C., *Eur. J. Org. Chem*., **2015**, 6823–6831.

[20]Nativi C., Papi F., Roelens S., Chem.Commun.2019,55, 7729–7736.

[21]Kelm S., Gerlach J., Brossmer R., Danzer C.P., Nitschke L., J. *Exp. Med.*, **2002**, 195, 1207-1213.

[22]Lerchen A., Knecht T., Koy M., Daniliuc C.G., Glorius F., *Chem. Eur. J*., **2017**, 23, 12149-12152.

[23]Jiménez-Barbero J., Dragoni E., Venturi C., Nannucci F., Ardá A., Fontanella M., André S., Cañada F.J., Gabius H.J., Nativi C., *Chem. Eur. J.*, **2009**,15, 10423-10431.

[24]Angulo J., Nieto P.M., *Eur. Biophys. J*., **2011**, 40, 1357-1369.

[25]Di Carluccio C., Forgione R.E., Martini S., Berti F., Molinaro A, Marchetti R., Silipo A., *Carbohydr. Res*., **2021**, 503, 108313.

[26]Nieto P.M., *Front. Mol. Biosci*., **2018**, 5, 33.

[27]Chandrasekaran A., Srinivasan A., Raman R., Viswanathan K., Raguram S., Tumpey T.M., Sasisekharan V., Sasisekharan R., *Nat. Biotechnol*., **2008**, 26, 107-113.

[28]Sullivan-Chang L., O'Donnell R.T., Tuscano J.M., *Bio Drugs*, **2013**, 27, 293-304.

[29]Tu X., La Vallee T., Lechleider R., *J. Exp. Ther. Oncol*., **2011**, 9, 241–248

[30]Hevey R., *Pharmaceuticals*, **2019**, 12, 55.

[31]Lenza M.P., Atxabal U., Oyenarte I., Jiménez-Barbero J., Ereño-Orbea J., *Cells*, **2020**, 9, 2691.

[32]Collins B.E., Blixt O., Han S., Duong B., Li H., Nathan J.K., Bovin N., Paulson J.C., *J. Immunol*., **2006**, 177, 2994-3003.

[33]O'Reilly M.K., Paulson J.C., *Methods Enzymol*., **2010**, 478, 343-363.

[34]Di Carluccio C., Forgione R.E., Montefior M., Civera M., Sattin S., Smaldone G., Fukase K., Manabe Y., Crocker P.R., Molinaro A., Marchetti R., Silipo A., *iScience*, **2020**, 24, 101998.

[35]Forgione R.E., Di Carluccio C., Guzmán-Caldentey J., Gaglione R., Battista F., Chiodo F., Manabe Y., Arciello A., Del Vecchio P., Fukase K., Molinaro A., Martín-Santamaría S., Crocker P.R., Marchetti R., Silipo A., *iScience*, **2020**, 23, 101231.

[36]Attrill H., Imamura A., Sharma R.S., Kiso M., Crocker P.R., van Aalten D.M., *J. Biol. Chem*., **2006**, 281, 32774-3278

[37]Torrini F., Palladino P., Brittoli A., Baldoneschi V., Minunni M., Scarano S., *Anal. Bioanal. Chem*., **2019**, 411, 7709-7716.

[38]Torrini F., Palladino P., Baldoneschi V., Scarano S., Minunni M., *Anal. Chim. Acta*, **2021**, 1161, 338481.

[39]Sindrewicz P., Li X., Yates E.A., Turnbull J.E., Lian L.Y., Yu L.G., *Sci Rep*., **2019**, 14, 9, 11851.

[40]MacroModel, Schrödinger, LLC, New York, NY (USA), **2020**.

[41]Maestro, Schrödinger, LLC, New York, NY (USA), **2021**.

[42]Morris G.M., Huey R., Lindstrom W., Sanner M.F., Belew R.K., Goodsell D.S., Olson A.J., *J. Comput. Chem*., **2009**, 16, 2785-91.

[43]Case D.A., Ben-Shalom I.Y., Brozell S.R., Cerutti D.S., Cheatham III T.E., Cruzeiro V.W.D., Darden T.A., Duke R.E., Ghoreishi D., Gilson M.K., *et al*,. AMBER 2018, University of California, **2018**.

[44]Woods Group, GLYCAM Web. Complex Carbohydrate Research Center, University of Georgia, Athens, GA, can be found underhttp://www.glycam.com, **2005-2021**.

[45]Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY (USA), **2021**; Impact, Schrödinger, LLC, New York, NY (USA), **2021**; Prime, Schrödinger, LLC, New York, NY (USA), **2021**.

[46]Roe D.R., Cheatham T.E., 3rd PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data., *J. Chem. Theory Comput*., **2013**, 9, 3084–3095.

[47]Wang J., Wang W., Kollman P.A., Case D.A., *J. Mol. Graphics Modell*., **2006**, 25, 247-260.

[48]Humphrey W., Dalke A., Schulten K., J. *Mol. Graphics,* **1996**, 14, 33-28

Chapter 6

Characterization of natural and synthetic sialoglycans targeting the Hemagglutinin-Neuraminidase of mumps virus

Adapted from:

Marchetti R., Forgione R.E., Di Carluccio C., Milanesi F., Kubota M., Nieto F.F., Molinaro A., Hashiguchi T., Francesconi O., Silipo A., *Front. Chem.,* 2021, *9*, article 711346

Abstract

The inhibition of surface viral glycoproteins offers great potential to hamper the attachment of viruses to the host cells surface and the spreading of viral infection. Mumps virus (MuV) is the etiological agent of the mumps infectious disease and causes a wide spectrum of mild to severe symptoms due to the inflammation of the salivary glands. Here we focus our attention on the hemagglutinin-neuraminidase (HN) isolated from MuV SBL-1 strain. We describe the molecular features of host sialoglycans recognition by HN protein by means of NMR, fluorescence assays and computational studies. Furthermore, we also describe the synthesis of a *N*-acetylneuraminic acid-derived thiotrisaccharide targeting the viral protein, and the corresponding 3D-complex. Our results provide the basis to improve the design and synthesis of potent viral hemagglutinin-neuraminidase inhibitors.

6.1 Introduction

Mumps is an infectious disease with a high morbidity in non-immunized children, caused by a respiratory-droplets transmitted virus, known as mumps virus. Usually, it causes swollen salivary glands under the ears, referred to as parotitis; other symptoms include fever, headache, muscle aches and tiredness. Especially in adults, mumps can also cause orchitis, mastitis, pancreatitis, encephalitis and meningitis. Despite the advent of an effective vaccine, mumps virus continues to circulate throughout the world and although the majority of mumps infections are subclinical in vaccinated individuals, severe complications still occur especially in underdeveloped countries. Moreover, in the last years, mumps outbreaks among vaccinated young adults have been also reported in different countries.[1,2]

Mumps virus is a member of the Paramyxoviridae family, belonging to the genus Orthorubulavirus, related to parainfluenza and Newcastle disease viruses, containing nonsegmented negative-strand RNA genome. Different glycoproteins spikes protrude away from the viral surface, playing key roles in the different steps of the infection process as host receptor recognition, binding, and cleavage. Among the 8 gene-encoded viral glycoproteins, exhibiting hemagglutinin, neuraminidase and cell fusion activity, hemagglutinin-neuraminidase (HN) has attracted attention as potential target for the development of specific and potent inhibitors able to regulate the enzyme functions. This glycoprotein is a specific sialic-acid-binding lectin able to recognize host cell-surface glycans which serve as

attachment points for the virus, allowing not only its entry into the host cells but also the spreading of the infection. HN protein, indeed, also participates in the internalization of viral particles activating the fusion protein and permitting the fusion of viral membranes to the host cell. Finally, HN prevents selfagglutination of viral particles and favors the release of virions from the infected cells since it acts as a sialidase, removing the sialic acid moiety from viral progeny.[3] Thus, the disruption of the interactions between virus related HN protein and host glycans may prevent the virus attachment to host cells as well as its multiplication and release from the infected cells. With the goal to provide the basis for the design and development of novel HN inhibitors, we here report the characterization of the molecular interaction of MuV-HN protein from SBL-1 strain with α-2,3-linked sialic acid containing ligands. In detail, we described the kinetic parameters that characterize the hydrolysis of the sialyllactosamine, 3'SLN, catalyzed by the HN protein; we identified the epitope mapping of the ligand when interacting with the protein and we determined the three-dimensional structure of the protein-glycan complex. To date, the crystal structure of the MuV-HN from SBL-1 strain has not been solved yet; thus, with the aim to achieve a 3D perspective of the molecular recognition mechanism, the homology model of the protein has been built. Finally, we synthesized a potential inhibitor of HN neuraminidase and elucidated by NMR, fluorescence analysis and computational studies its interaction with the viral protein.

6.2 Results and discussion

Receptor motifs that can be recognized by MuV-HN (hemagglutinin-neuraminidase) protein from SBL-1 strain have been recently identified by glycan array screening[4] as specific host-cell sialoglycans which terminate with neuraminic acid (Neu5Ac) moiety. Among them, Neu5Acα2-3Galβ1-4GlcNAc (3'SLN), a sialoglycan broadly expressed in and exposed on various host tissues, was reported as one of the strongest binders; thus, we decided to investigate, at molecular level, its recognition by MuV-HN SBL-1 strain by means of NMR spectroscopy, biophysical and computational approaches. Given the enzymatic activity of MuV-HN SBL-1, upon addition of the sialoglycan substrate in solution, MuV-HN protein catalyzed the cleavage of the terminal Neu5Ac moiety, producing *N*-

acetyllactosammine (LacNAc, residues A' and B') and reducing sialic acid (red-Neu5Ac, K') (**Figure 1a**). Therefore, the real time kinetics of 3'SLN hydrolysis by MuV-HN was followed by 1D ^{1}H NMR, as reported in **Figure 1b**, where the overlapped NMR spectra detected over time progression are shown.

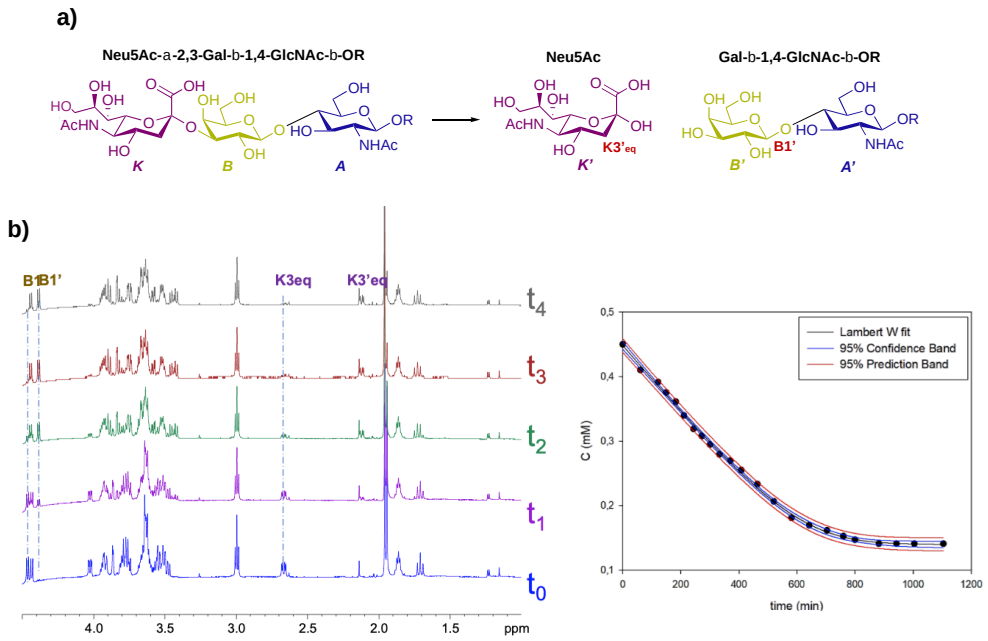


Figure 1. NMR analysis of MuV-HN from SBL-1 in complex with 3'SLN. a) Scheme of the mechanism of hydrolysis of 3'SLN catalyzed by MuV-HN (SBL-1strain); b) (left panel) section of ^{1}H NMR spectra at different time of the enzymatic reaction have been reported. The spectra were recorded using 13 μM of MuV-HN protein and 700 μM of 3'SLN in PBS deuterate buffer (pH 7). The NMR quantification of the substrate concentration was performed by integration of the well-dispersed resonances of the product (B1', K3'$_{eq}$). (right panel) Analysis of the MuV-HN kinetics toward 3'SLN by means of the explicit reformulation of the integrated form of MM equation with Lambert W fit as solution. The plot of the concentration of 3'SLN as a function of the time has been reported. The substrate concentration was evaluated from the K3$_{eq}$ resonance of Neu5Ac unit. The fit of the kinetic data by the Lambert-W fit produced a K_M value of 5 μM and Vmax of $6 \cdot 10^{-3}$ (mM/min). The blue dashed line represented the confidence interval of the fit and the red dashed lines are the prediction bands.

The progress of the hydrolysis was assessed by monitoring the decrease of the signals of the substrate (as B1 and K3eq) and the concurrent appearance and following enhancement of the intensity of the products resonances (as B'1 and K'3eq). Similarly to other viral sialidases[5], MuV-HN acts as retaining glycosidase, with a net retention of substrates configuration. The H3eq resonance of Neu5Ac was monitored to evaluate the variation of the substrate concentration and, by fitting the kinetic data through the Lambert W function[6], the values of the enzyme K_M value of 5 μM and Vmax of $6 \cdot 10^{-3}$ mM/min were determined (**Figure 1b**). The low value of the K_M corresponds to high affinity for 3'SLN substrate.

Notably, the K_M evaluated for 3'SLN/MuV-HN from SBL-1 strain was lower with respect to that observed by our group for the same substrate in the interplay with MuV-HN from Hoshino strain. (K_M = 12 μM)[7], suggesting a higher enzymatic affinity of the protein from SBL-1 strain for 3'SLN.

Despite the neuraminidase activity of MuV-HN induces the cleavage of the sialic acid within NMR time scale, saturation transfer difference (STD) NMR allowed to identify the

ligand binding epitope (**Figure 2**).[8] A qualitative analysis of the STD effects demonstrated that the trisaccharide was entirely accommodated into the binding pocket of the protein.

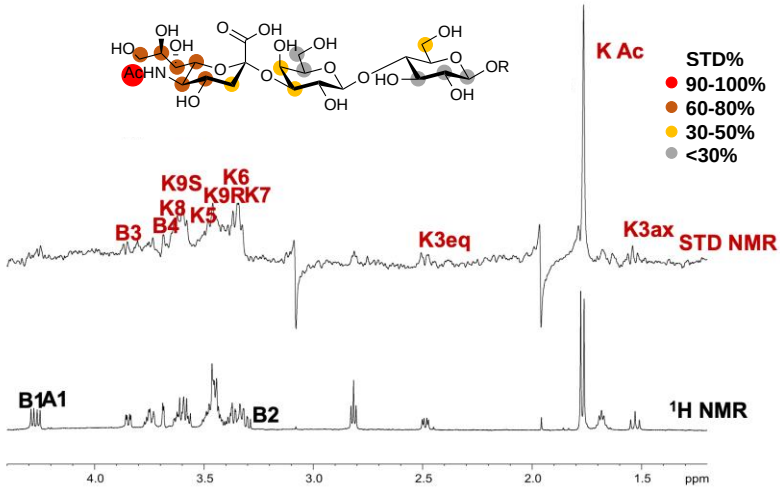


Figure 2. STD NMR analysis of 3'SLN in the interaction with MuV- HN from SBL-1 strain. (top panel) Interacting epitope map of 3'SLN as derived by STD NMR data. (bottom panel) The ^{1}H NMR spectrum and the STD NMR spectrum of MuV-SBL-1–3'SLN mixture with a molecular ratio of 1:70, at 283 K.

The moiety most contributing to the interaction was the neuraminic acid residue (K) which established several contacts with the protein, not only through its acetamide portion but also via the lateral chain; the highest STD effect, set at 100%, belonged indeed to the *N*-acetyl group; protons H7–H8–H9 also showed significant STD enhancements; finally, STD signals were detected for some protons of the Neu5Ac carbohydrate ring, as H4, H5 and H6. As already observed in the case of MuV-HN from Hoshino strain, also the other two residues (A and B) were involved in the interaction with the protein although to a lower extent. Furthermore, by evaluating the relative intensities of two *N*-acetyl groups of Neu5Ac K and GlcNAc A, it was clear that the contribution from the acetamide of the GlcNAc was not significant.

To get structural insights into sialoglycans recognition by MuV-HN (SBL-1 strain, genotype A), comparative homology modelling was carried out considering as template the crystal structure of MuV-HN head domain (Hoshino strain, genotype B), PDB entry: 5B2D, which possesses 95% of structural similarity to the target sequence (**Figure 3a**), in agreement with the low number of genomic mutations occurring among different MuV genotypes at HN encoding gene.[9] From the sequence analysis of the two different strains (**Figures 3a**, **3b** and **Figure S1**), the most relevant modifications of the amino acid sequences occurred at the known B-cell epitopes, particularly at 327-363; 375-403; 440-443; and 533 regions, as similarly detected when comparing other MuV-HN genotypes.[10]

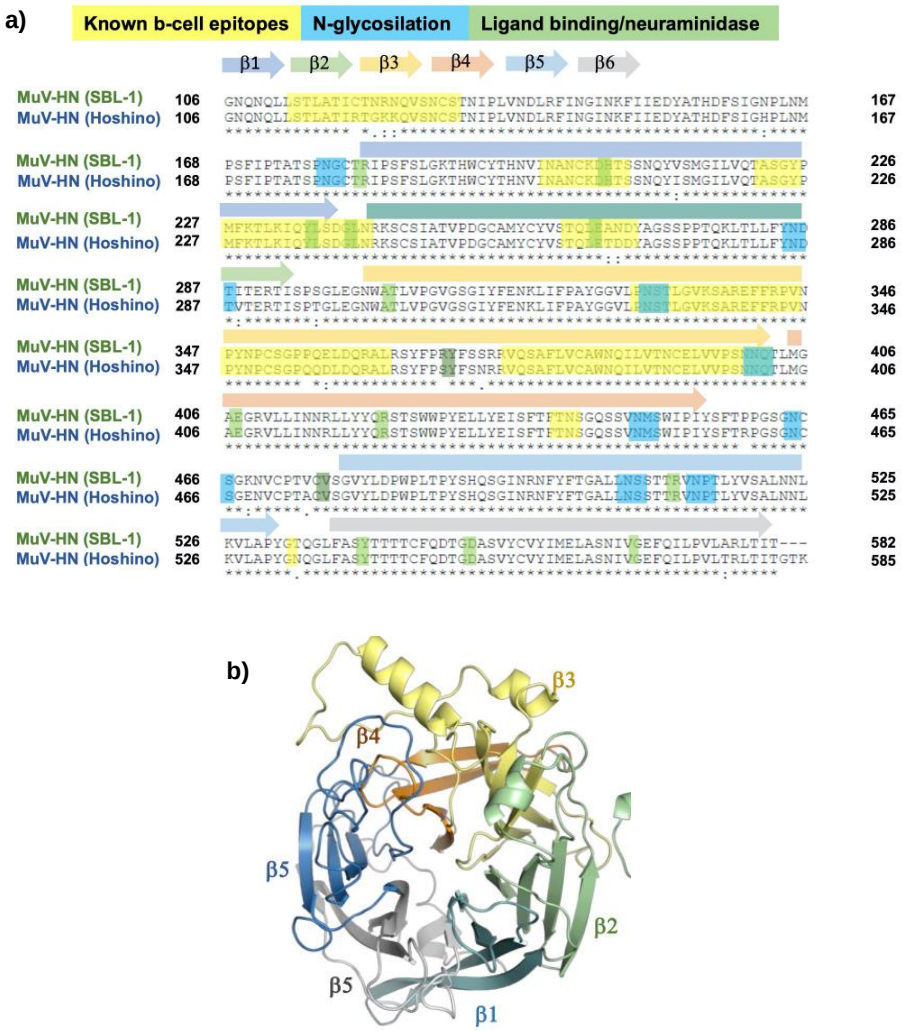


Figure 3. a) CLUSTAL alignment of the head domain of MuV-HN from different strains. Amino acids involved in molecular recognition and neuraminidase activity are highlighted in green, *N*-glycosylation sites are highlighted in blue and Known B cell epitopes in yellow. Sequences corresponding to the six four-stranded beta sheets are indicated with differently colored arrows; b) MuV-HN monomer six bladed propeller structure, whose constituent structural elements (β1–β6 sheets) are differently colored.

The homology model of HN protein from SBL-1 strain was then built using Swiss Model server[11] and, after refinement by means of Maestro[12], its quality was assessed through SAVES resource. As result, the refined model, analyzed using ERRAT plot, exhibited an overall quality factor of 87.11. The Ramachandran plots generated by means of the PROCHECK tool (**Figure S1b**) allowed to assess the stereochemical quality of the resulting structure[13], showing that 99% of the residues' dihedrals belonged to the core/allowed regions, hence confirming the fitness of the model. Moreover, the 3D profile of the model structure obtained by VERIFY 3D, scoring the compatibility of the 3D structure model with respect to amino acid sequence, showed a scoring above 0.2 for above 80% of the amino acid residues, further confirming the accuracy of the model (**Figure S1c**). Comparing our refined model with the target structure, a similar topology was noted, and the RMSD

calculation in reference to the MuV-HN crystal structure of 0.05 Å further denoted its remarkable quality, in full agreement with the high sequence identity between the homologous proteins from different MuV strains (**Figure S1a**). Indeed, as predicted by the sequence analysis, the changes in the amino acid sequence among the proteins from the two different strains did not affect neither the overall tridimensional structure of the sialic acid binding motif nor the accessibility of the protein to the glycans on the host cell surface. In detail, MuV-HN (SBL-1 strain) model showed a globular structure, characterized by six four-stranded antiparallel β-sheets (β1-β6) arranged in the so-called bladed beta propeller fold, that is characteristic of many viral neuraminidases (**Figure 3b**).[14] At the center of the domain, it was located the dual binding/sialic acid cleavage site of the protein, comprising the conserved sialidase active site residues, namely Arg180, Glu407, Arg422, Arg512, Tyr540, Glu561, and Asp204.[15] Furthermore, the binding site comprised the Tyr369 and Val476, highly conserved among the MuV-HN genotypes, that revealed to be important determinants of Mumps recognition.[7,10,15]

As several studies highlighted the importance of the flexibility of specific loops in other viral sialidases,[4,16] the MuV-HN conformational behavior was explored by MD simulations. First, the MD simulation of the apo form of MuV-HN SBL-1 strain model was carried out using Amber18.[17] As result, it was noted that MuV-HN showed a slight conformational rearrangement along the simulated time, with a small increase of the backbone RMSD which converged to 1.5 Å (**Figure S2a**). The inter-strand loops RMSD was calculated (**Figure S2b**) showing that the fluctuations from the initial geometry affected mostly β5 and β6 loops. No relevant conformational changes were instead detected when calculating the β1, β2 and β3 inter-strand loops RMSD along the trajectory (**Figure S2b**). Furthermore, the atom-positional root-mean- square fluctuations (RMSFs) relative to the Cα-atoms was calculated along the trajectory, denoting smaller values in the secondary structure regions, further reinforcing the homology model quality (**Figure S2c**). Interestingly, large fluctuations occurred in the loop regions, especially at the positions corresponding to the β4 and β6 inter-strand loops. In particular, the β4 showed high flexibility from the RMSF plot (**Figure S2b**). Nonetheless, the highest flexibility belonged to the long loop connecting β4 and β5 sheets.

Our model was subsequently employed for docking calculation by Autodock 4.2[18] program in the presence of 3'SLN as ligand. Based on the predicted binding energy and cluster population, the best pose of 3'SLN/MuV-HN complex was selected (**Figures 4a,b**) and it was submitted to 100 ns MD run using AMBER 18 (**Figures S3a-e**). The stability of the complex has been assessed by the ligand RMSD calculated using the protein as reference; the results showed that the ligand remained stable around 1Å for almost the entire simulation, thus indicating it was anchored to the receptor along the simulated time. Moreover, the notable structural similarity between the most populated MD clusters calculated with respect to the ligand RMSD along the trajectory further confirmed that the sialotrisaccharide didn't exhibited conformational changes along the simulation (**Figures S3a,b**).

In accordance with NMR data, computational analysis showed that all three sugar moieties were in contact with the receptor surface, as reported in **Figure 4b**. The interaction network of the selected complex showed many similarities with those observed for MuV-HN from Hoshino strain.[4,7,15]

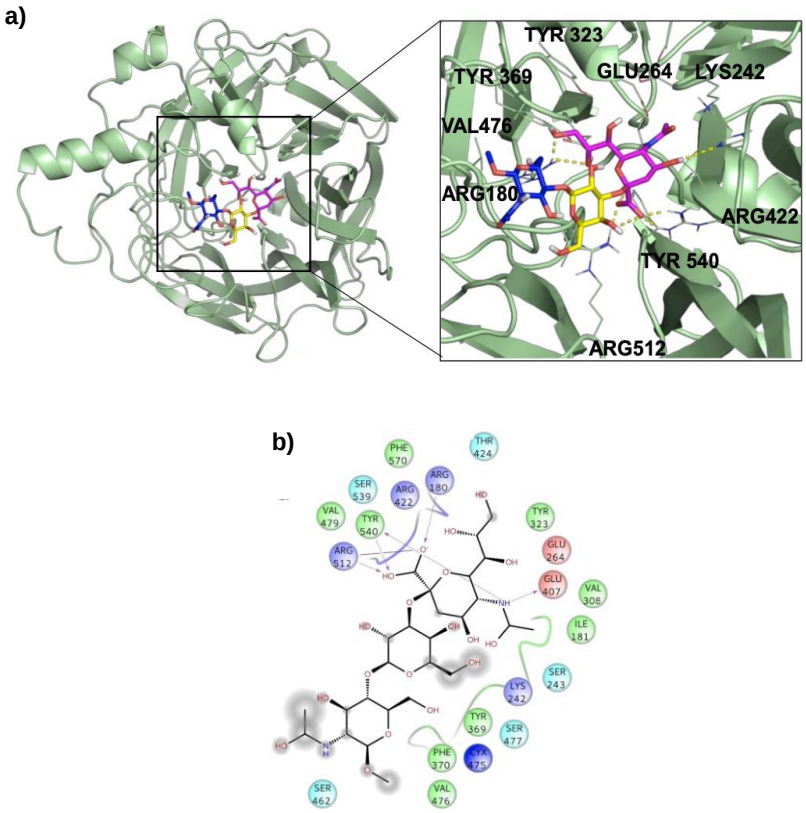


Figure 4. a) 3D model of MuV-HN from SBL-1 in complex with 3'SLN. Close up view of 3'SLN binding pose at the MuV-HN active site. The main amino acid residues involved in the binding are highlighted as lines. Sialic acid, Galactose and N-acetyl Glucosamine residues are colored in magenta, yellow and blue, respectively according to the SNFG nomenclature; b) Two-dimensional plot illustrating the interactions of 3'SLN with the MuV- HN binding site residues. Dotted arrows indicate hydrogen bonds with functional groups from side chains and solid arrows those involving the backbone functional groups. The residues shown, close to the ligand, are involved into hydrophobic and polar interactions.

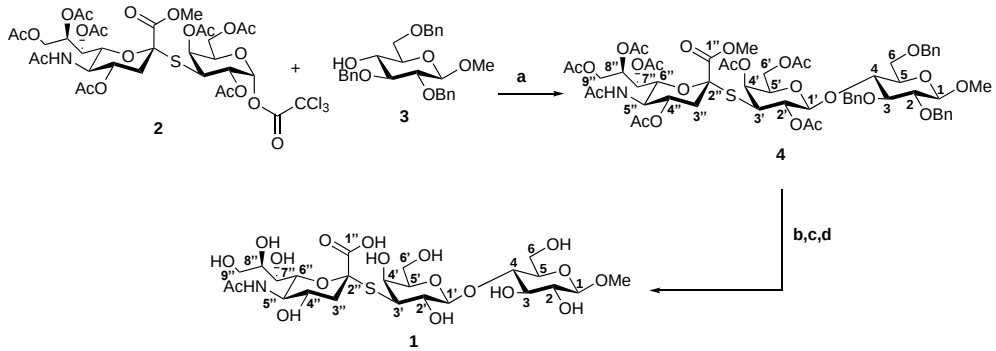
In detail, the sialic acid residue established strong interactions between its carboxylate moiety and three arginine residues, namely Arg180, Arg422 and Arg512. The *N*-Acetyl moiety formed the hydrogen bond with Glu407 and hydrophobic interactions with Ile181 and Val308. The GlcNAc ring was engaged in CH-π interactions with Tyr369. Gal residue was in proximity to the receptor surface, but no significant polar interactions were observed, as its hydroxyl groups were directed toward the solvent. Nonetheless, the Gal moiety was in proximity to the Val476 side chain thus entailing hydrophobic contacts with MuV-HN. Additionally, the analysis of the protein-ligand interactions showed that the main contacts were maintained along all the trajectory, including the electrostatic interactions of the Neu5Ac unit with the Arg triad, and the stacking interaction between Tyr369 and the GlcNAc moiety (**Figures S3c-e**). This further suggested the importance of such residues to the molecular recognition of sialoglycans by Mumps virus hemagglutinin-neuraminidase receptor.

In order to verify if a particular loop stabilization occurred in the MuV-HN bound form, the RMSF and loop RMSD analysis of the 3'SLN/MuV-HN complex was carried out and compared to the free state. As a result (**Figures S3a,b**), lower fluctuations upon binding of the inter-strands loops of β5 and β6 sheets were observed. A significant stabilization of the

inter-sheet loops, namely β2-β3, β4-β5 and β5-β6 was also noted. Interestingly, the lower atomic fluctuations displayed by the loops of β4, β5 and β6 sheets matched with the MuV-HN region binding to the 3'SLN ligand, therefore implying that the stabilizing interactions with the sialotrisaccharide influenced the involved loops conformational flexibility.

Given that from NMR and computational analysis the entire trisaccharide seemed to be involved in the recognition and binding process, being the sialic acid the moiety anchoring the glycan to the binding pocket of the protein, we decided to synthesize a thioderivative of 3'SLN, namely thio-3'SL, as a potential HN inhibitor. The incorporation of a sulfur functional group on the anomeric position of the sialic acid should indeed affect the lability of the Neu5Ac-Gal glycosidic linkage without hampering the recognition of the sialoglycan by the enzyme.

The thio-3'SL (**1**) was prepared (**Scheme 1**) starting from the thrichloroacetimidate **2**,[19] and the methyl-2,3,6-tri-O-benzyl-β-D-glucopyranoside (**3**),[20] both prepared accordingly to literature methods. Glycosylation, catalyzed by BF_3•Et_2O, gives the protected trisaccharide **4** with a 50% yield. Removal of benzyl protective groups by hydrogenation and of the acetylic groups with MeONa in MeOH was followed by hydrolysis of the methyl ester with aqueous NaOH to give the thiotrisaccharide **1** with a 75% yield on three steps.



Scheme 1. Synthesis of thio-3'SL (compound **1**) with atoms labelling. Reagents and conditions: a) BF_3•Et_2O, 3Å molecular sieves, dry CH_2Cl_2, 0°C, 3 h; b) Pd/ C, H_2, MeOH, room temperature, 24 h; c) MeONa, MeOH, room temperature, 2 h; d) NaOH 1M in H_2O, room temperature, 2 h.

The hydrolytic stability of the thio-3'SL and its ability to interact with the MuV-HN protein was hence investigated by means of NMR spectroscopy and fluorescence analysis. As expected, the thio-3'SL was not hydrolyzed by MuV-HN, as it was observed by monitoring over time the NMR spectra of MuV- HN/thio-3'SL mixture (data not shown). However, it was recognized by MuV-HN protein as suggested by fluorescence data. The binding interaction of thio-3'SL with MuV-HN from SBL-1 strain was indeed analyzed by fluorescence spectroscopy exploiting the presence of aromatic amino acids in the protein binding pocket. In detail, changes in the fluorescence intensity of tryptophane residues of MuV-HN upon the addition of increased amount of thio-3'SL were followed (**Figure 5a**). The interaction was quantitatively evaluated by fitting the resulting binding curve achieved with the application of the quenching data elaboration described by Ribeiro *et al.*[21] As a result, a binding constant (K_d) value of 0.79 ± 0.06 was obtained. (**Figure 5b**). The molecular characterization of the complex has been then carried out by the combination of NMR and computational studies. The ligand interacting epitope has been indeed described by STD NMR experiments (**Figure 5c**). As expected, the strongest STD enhancements were observed for the sialic acid moiety and especially for its acetyl group and protons of Neu5Ac lateral chain (H7-H8-H9). Significant STD effects were further detected for protons H4-H5 and H6 of sialic acid ring.

Lower STD signals (below 50%) were instead observed for protons belonging to galactose and glucose residues indicating that they pointed farer from the protein surface.

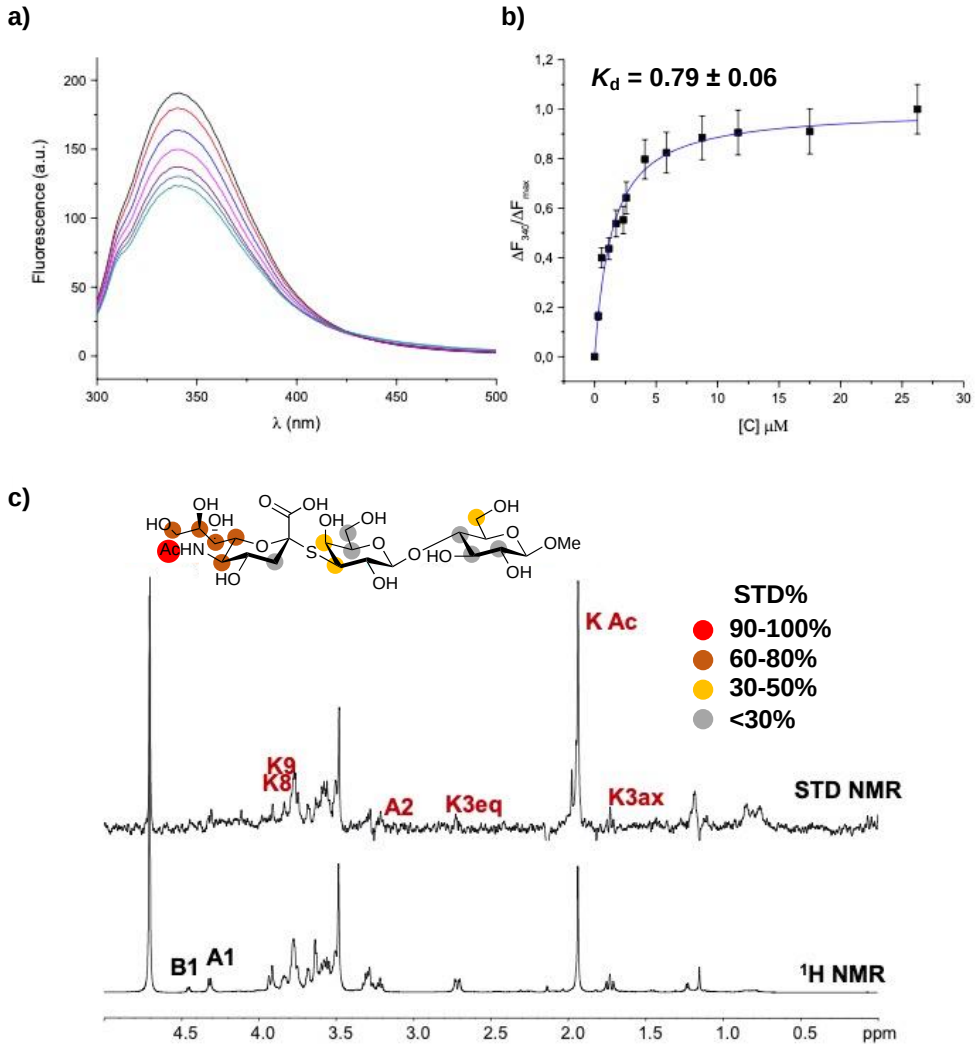


Figure 5. Analysis of MuV-HN from SBL-1 in complex with thio-3'SL: a) Fluorescence titration of MuV-HN (0.25 μM) upon the addition of increasing concentration of thio-3'SL from a stock solution of 700 μM. The emission spectra were recorded by using an excitation wavelength of 285 nm and a temperature of 10°C; b) relative binding isotherm, the value of the binding constant (Kb) is also reported. For each data point, 10% Y error bars are shown; c) STD NMR analysis of thio- 3'SL in the interaction with MuV-HN from SBL-1 strain. Interacting epitope map of thio 3'SL as derived by STD NMR data (top panel). The ^{1}H NMR spectrum and the STD NMR spectrum of MuV-SBL-1–thio-3'SL mixture with a molecular ratio of 1:50, at 283 K (bottom panel).

The thio-3'SL structure was then modelled into the binding site of MuV-HN SBL-1. The docking calculations allowed to describe the relevant receptor-ligand binding interactions, highlighting that the protein accommodated the synthetic analogue and the 3'SLN with a similar binding mode. Indeed, the superimposition of the two complexes (Figure 6) clearly showed a nearly identical orientation of the ligands into the receptor binding site. The interaction network was overall conserved; the Neu5Ac carboxylate engaged electrostatic interactions with the receptor Arg180, Arg422 and Arg512; Glu407, Glu264, and Tyr323

binding pocket residues were also interacting with sialic acid unit. Moreover, the Glc moiety of the trisaccharide formed hydrophobic interactions with Tyr369 and Val476 residues; this latter residue was also in close contact with Gal moiety.

6.3 Conclusions

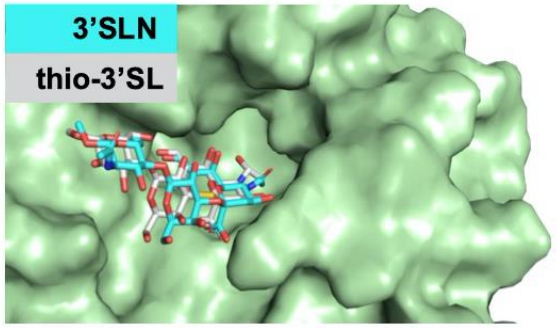
Hemagglutinin-Neuraminidase activity is essential for the infection and propagation of viruses belonging to the Paramyxoviridae family, including parainfluenza (hPIV), the Newcastle disease and mumps viruses. Over the past years, different series of inhibitors have been developed toward viral neuraminidases, especially targeting hPIVs. However, the design and development of novel inhibitors and high- affinity ligands could prove worthwhile for a better understanding of virus tropism and pathogenesis and may help in the fight against viral infections allowing the advancement of new licensed anti-viral drugs.

Here we focused our attention on mumps virus, the leading cause of the mumps disease that could affect the central nervous system, causing meningitis. Mumps virus possesses different surface glycoproteins and among them, the HN represents a multitasking protein mediating both the early and late stages of viral infection, including the host-cell sialoglycans recognition, the trigger of virus and host-cell membranes fusion and finally the release of progeny virions from infected cells. MuV-HN thus represents an attractive target for the structure-based design and development of novel anti-viral drugs since interfering with these processes may affect the viral pathogenicity and infectivity.

Thus, we investigated the structural features of MuV-HN from SBL-1 strain when bound to host-cell sialoglycans with the aim to exploit them as basis for the guided development of tailored inhibitors. The structural characterization of the HN protein in the interaction with the 3'SLN has been performed by means of NMR and computational studies. In detail, the kinetic of 3'SLN hydrolysis catalyzed by MuV-HN from SBL-1 strain was characterized by progress curve analysis derived by NMR assay. Notably, the kinetic parameters obtained by the elaboration through the Lambert W function were in the same range of the values obtained for the same ligand interacting with MuV-HN from Hoshino strain, as shown in our previous work.[7] STD NMR allowed to describe the ligand interacting epitope stressing the importance of the sialic acid in the recognition and interaction process, as expected, but also demonstrating that the binding pocket of the protein was pliable enough to accommodate the three composing sugars of the 3'SLN.

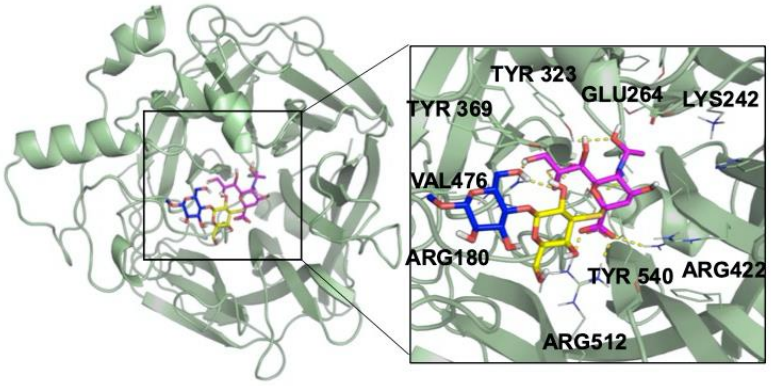
The homology modelling of the protein was also carried out with the aim to achieve a 3D view of the protein-ligand complex. The crystal structure of MuV-HN from the Hoshino strain was used for building the model; that revealed a huge similarity between the 3D structures. It is worth to note that, from the sequence analysis of the two proteins, few variable positions were observed. Few of the mutations affected the important functional regions, as the known B-cell epitopes; the most relevant is the 354 (P→Q) mutation at the 327-363 region, that modifies the physiochemical features of the surface exposed area. This variation was already suggested to be significant for altered antibody recognition and neutralization by MuV-HN.[10,22] Notably, the neuraminidase binding site was preserved, differently with respect to that observed when comparing the sequences of other MuV-HN genotypes.[10,22] The refined 3D model of the protein has been used for docking analysis and MD simulations of the MuV-HN/3'SLN complex.

a)



Docking pose	Cluster rank	n. of cluster conformations	Est. free energy of binding (kcal/mol)	Est. inhibition constant (mM)	RMSD from reference (Å)
3SLN	3	100	-1.46	85.58	2.371
Thio-3SL	1	92	-2.24	22.67	1.632

b)



c)

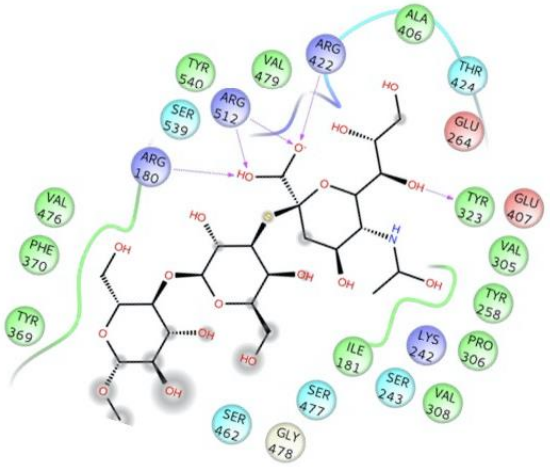


Figure 6. 3D model of MuV-HN from SBL-1 in complex with thio-3'SL. a) Docking of 3'SLN and thio-3'SLN into MuV-HN from SBL-1 strain structural model. The superimposition of the best selected binding mode of both ligands is shown. The parameters of the best models, calculated by Autodock, are also enlisted; b) Close up view of thio-3'SL binding pose at the MuV-HN active site. The main amino acid residues involved in the binding are indicated as lines. Sialic acid, Galactose and *N*-acetyl Glucosamine residues are colored in magenta, yellow and blue, respectively according to the SNFG nomenclature; c) Two-dimensional plot showing the interactions of thio-3'SL with the MuV-HN binding site residues. The residues shown are involved into hydrophobic and polar interactions with the ligand.

Our results demonstrated that the main interactions involving the three sugars of the glycan receptor observed when bound to MuV-HN from Hoshino strain were completely conserved in the binding with MuV- HN from SBL-1 strain. We indeed observed that the Neu5Ac residue was deeply located in the receptor pocket where it established strong polar interactions with the canonical neuraminidases sialic acid recognizing residues, including the Arginine triad (Arg180, Arg422, Arg512). Nonetheless, the previously described interactions involving Tyr369, which were reported to strongly influence the binding affinity of MuV-HN from Hoshino strain, were also present in the case of MuV-HN from SBL-1 strain.[15] Interestingly, the MD simulation studies performed on the apo and bound forms of MuV-HN revealed the stabilization of particular inter- strand- and inter-sheet loops upon the ligand binding, thus suggesting that the receptor loop flexibility may play a relevant role in the molecular recognition of sialoglycans by mumps virus hemagglutinin neuraminidase and mechanism of catalysis as well. Indeed, in the bound state, lower fluctuations were observed for loop regions in intimate contact with the ligand bound to Muv-HN, on which some of the binding site residues were located, as Val476 on β4-β5 loop and Arg512 on β5 loop. Previous studies on the conformational behavior of other neuraminidases[16,23,24] reported a shift from open to closed loop conformation upon sialoglycans binding, thus, we also inspected this possibility by measuring the distance across the loops delimiting the binding site (**Figure 7**).

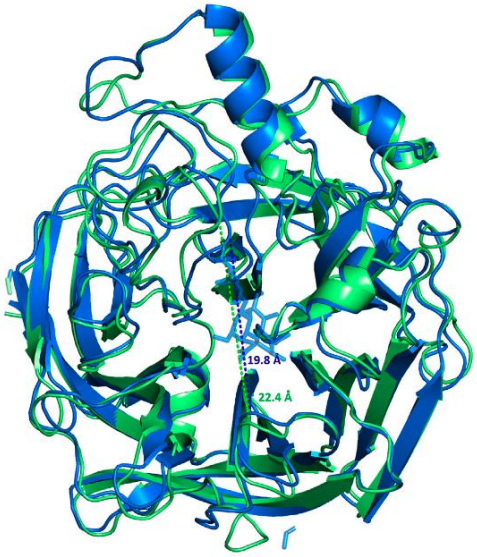


Figure 7. Superimposition of the most populated clusters of MuV-HN from SBL-1 strain in complex with 3'SLN ligand (blue) and the apo-MuV-HN (green), extracted from the MD simulations. The distances between the walls of the binding site are 19.8 Å (blue dashed line) and 22.4 Å (green dashed line) for the complexed and apo forms, respectively.

Interestingly, the distance measured between the binding pocket walls showed an increased value for the apo-MuV-HN protein, thus suggesting that also for the mumps virus hemagglutinin neuraminidase the glycan receptor specificity may be tuned by loops flexibility.

Furthermore, we performed the synthesis of a thiotrisaccharide, thio-3'SL, as a potential inhibitor of the neuraminidase activity of the protein. The ability of thio-3'SL to interact with MuV-HN from SBL-1 strain was demonstrated by monitoring the intrinsic fluorescence of the tryptophane residues of the receptor and by NMR. Moreover, modelling analyses provided

a picture of the protein-ligand complex, showing that the thio-3'SL established contacts with MuV-HN receptor which were comparable to those made from 3'SLN, consistently to the similar bound ligands' topology. The main interactions mostly involved the Neu5Ac unit, that interacted with the characteristic receptor binding site amino acids Arg180, Arg422, Arg512, Glu407 and Tyr323. Accordingly, the third sugar exhibited stacking interactions with the Tyr369.

In conclusion, a combination of organic chemistry, NMR, fluorescence and computational studies could provide a 3D view of the MuV-HN from SBL-1 strain when interacting with the natural substrate, 3'SLN, and the synthetic thio-3'SL, showing a similar binding mode. Our analysis demonstrated that, in both cases, the main interactions involving the three sugars of the glycan receptor were also comparable to those observed when investigating MuV-HN from Hoshino strain.[15]

Our outcomes may help in the identification of new inhibitor scaffolds which could prove worthwhile in the fight against mumps virus.

6.4 Supporting information

6.4.1 Materials and methods

Protein expression and purification: Detailed method was described previously.[25] Briefly, the expression plasmid encoding the MuV-HN protein (amino acid positions 96–582)[15] of the SBL-1 strain was transfected into HEK293S cells lacking *N*-acetylglucosaminyltransferase I [293S GnTI(−) cells][26] using polyethyleneimine-MAX (Polysciences, Inc.). MuV-HN protein was expressed as a recombinant protein containing an *N*-terminal secretion signal sequence and a C-terminal His6 tag sequence using the expression vector pHLsec.[27,28] MuV-HN protein was first purified using a Ni^{2+}-nitrilotriacetic acid (NTA) affinity column (Cosmogel His-Accept; Nacalai Tesque). Then, the eluted MuV-HN protein was further purified using a size exclusion column (Superdex 200 Increase GL 10/300; GE Healthcare) in PBS without calcium and magnesium.

***Synthesis and Characterization of Chemical Materials*:** The 3'-sialyllactosamine (3'-SLN) was purchased from Tokyo Chemical Industry Co., Lt. A thioderivative of 3'SLN, namely thio-3'SL, has been synthesized as described below. ESI-MS analyses were performed in negative ion mode and were recorded on an LCQ-Fleet Ion Trap equipped with a standard Ionspray. ^{1}H NMR spectra were obtained at 500 MHz in $CDCl_3$ and D_2O. ^{13}C NMR spectra were obtained at 125 MHz in $CDCl_3$ and D_2O. Chemical shifts are reported in part per million (δ), using the residual solvent line as internal reference. Polarimetry analyses were performed on a Jasco DIP-370.

***Synthesis of compound 4*: 2** (25 mg, 0.027 mmol) and **3** (20 mg, 0.043 mmol) were solubilized in toluene and co-evaporated in vacuo for three times and then dried in vacuo overnight to remove traces of water. Under nitrogen atmosphere, anhydrous CH_2Cl_2 (1 mL) and freshly activated 3 Å molecular sieves (100 mg) were charged in the reaction flask and the mixture was stirred for 2h at room temperature. The mixture was cooled to 0°C and $BF_3{\cdot}Et_2O$ was added (3 µL, 0.024 mmol). The reaction was stirred at 0°C for 3h then the mixture was diluted with 15 mL of CH_2Cl_2 and filtered. The filtrate was washed with 10 mL of a saturated solution of $NaHCO_3$. The organic layer was collected, dried over anhydrous Na_2SO_4, filtered and the solvent evaporated to obtain 35 mg of crude product. Crude was purified by flash chromatography on silica gel (petroleum ether 5% in ethyl acetate; R_f = 0.35), to obtain pure 4 as a white foam (15 mg, 50%). ^{1}H NMR (500 MHz, $CDCl_3$, δ = 7.26): δ 7.37-7.22 (m, 15H, CH_2*Ph*); 5.64 (m, 1H, C*H*-8"); 5.29 (dd, J_1 = 10.2 Hz, J_2 = 2.2 Hz, 1H, C*H*-7"); 5.21 (d, *J* = 7.6 Hz, 1H, C*H*-1'); 5.10 (d, *J* = 7.5 Hz, 1H, N*H*); 5.01 (d, *J* = 11.0 Hz, 1H, CH_2-Ph); 4.86-4.79 (m, 5H, CH_2-Ph, C*H*-4", C*H*-2', CH_2-Ph, C*H*-3'); 4.63-4.61 (m, 3H, 3 x CH_2-Ph); 4.30 (d, *J* = 7.8 Hz, C*H*-1); 4.21 (dd, J_1 = 12.6 Hz, J_2 = 2.6 Hz, CH_2-9"); 4.09.4.02 (m, 2H, C*H*-5", CH-5); 3.96-3.84 (m, 4H, CH_2-9", CH_2-6, CH-4', CH_2-6'); 3.82 (s, 3H, COO*Me*); 3.72-3.62 (m, 5H, CH_2-6', C*H*-4, C*H*-6", C*H*-6, CH-3); 3.55 (s, 3H, O*Me*); 3.54-3.51 (m, 1H, C*H*-5'); 3.36 (dd, J_1 = 9.3 Hz, J_2 = 8.0 Hz, 1H, C*H*-5'); 2.63 (dd, J_1 = 12.8 Hz, J_2 = 4.6 Hz, 1H, C*H*-3_{eq}"); 2.17 (s, 3H, *Ac*); 2.12 (s, 3H, *Ac*); 2.03 (s, 6H, 2 x *Ac*); 2.01 (s, 3H, *Ac*); 1.96 (s, 3H, *Ac*); 1.87 (s, 4H, C*H*-3_{ax}", Ac); 1.82 (s, 3H, *Ac*); ^{13}C NMR (125 MHz, $CDCl_3$, δ = 77.16): ∂ 171.06; 170.57; 170.45; 170.42; 170.33; 170.30; 170.24; 169.83; 169.05; 139.37; 138.88; 138.63; 128.38; 128.35; 128.16; 127.66; 127.52; 127.45; 127.43; 127.19; 104.57 (*C*H-1); 101.49 (*C*H-1'); 83.16 (*C*H-3); 82.24 (*C*H-2); 81.14 (*C*-2"); 77.54 (*C*H-4'); 75.20 (*C*H-5'); 75.08 (*C*H_2Ph); 74.75 (*C*H_2Ph); 73.79 (*C*H-4); 73.56 (*C*H_2Ph); 72.46 (*C*H-5"); 69.86 (*C*H-3'); 69.59 (*C*H-6'); 69.32 (*C*H-2'); 69.11 (*C*H-4"); 67.19 (*C*H-7"); 67.15 (*C*H-8"); 62.48 (*C*H-9"); 62.16 (*C*H-6); 57.12 (O*Me*); 53.33 (COO*Me*); 49.28 (*C*H-

5); 45.94 (*C*H-6"); 37.21 (*C*H-3"); 23.30 (*Ac*); 21.56 (*Ac*); 21.07 (*Ac*); 20.93 (2x*Ac*); 20.82 (*Ac*); 20.74 (*Ac*); 20.60 (*Ac*); ESI-MS *m/z* (%): 1264.83 (100%) $[M+Na]^+$; $[\alpha]_D^{22°C}$ = +46 (0.5 mg/mL in MeOH).

Synthesis of compound 1: To a stirred solution of **4** (11 mg 8.85 µmol) in MeOH (1 mL), Pd/C (30 mg) was added, and the suspension was stirred under H_2 atmosphere for 24h. Pd/C was removed by filtration on a HPLC filter and solvent evaporated to obtain 8 mg of crude as a white solid. Crude was redissolved in a freshly prepared solution of MeONa in MeOH (0.5 mg of Na° in 1 mL of MeOH) and the solution was stirred for 2h, then solvent was evaporated, and crude dissolved in 1 mL of NaOH 1M. The solution was stirred for 2h, then diluted with 1 mL of water. The solution was treated with Amberlyst-15 till pH 4, then resin was removed by filtration and the aqueous solution lyophilized to obtain pure 1 as an amorphous white solid (6 mg, 75%). ^{1}H NMR (500 MHz, D_2O, δ = 4.79): δ 4.55 (d, *J* = 6.9 Hz, 1H, C*H*-1'); 4.42 (d, *J* = 7.9, 1H, C*H*-1); 4.03 (dd, J_1 = 12.3 Hz, J_2 = 2.2 Hz, 1H, CH_2-6); 3.96-3.92 (m, 1H, C*H*-8"); 3.90-3.87 (m, 4H, C*H*-4, C*H*-4', CH_2-6, CH_2-9"); 3.80-3.77 (m, 1H, C*H*-5'); 3.74-3.72 (m, 2H, 2x CH_2-6'); 3.71-3.64 (m, 4H, C*H*-4", C*H*-5", C*H*-6", C*H*-3); 3.63-3.59 (m, 3H, C*H*-5, C*H*-7", C*H*-9"); 3.58 (s, 3H, O*Me*); 3.43-3.46 (m, 2H, C*H*-3', C*H*-2'); 3.32 (t, *J* = 8.65 Hz, 1H, C*H*-2); 2.82 (dd, J_1 = 12.7 Hz, J_2 = 4.7 Hz, 1H, C*H*-3_{eq}"); 2.04 (s, 3H, *Ac*); 1.83 (t, *J* = 12.0 Hz, 1H, C*H*-3_{ax}"); ^{13}C NMR (125 MHz, D_2O): δ 174.97; 174.53: 104.12 (*C*H-1'); 103.09 (*C*H-1); 84.08 (*C*-2"); 78.14 (*C*H-6"); 77.76 (*C*H-5'); 74.98 (*C*H-4"), 74.83 (*C*H-5); 74.33 (*C*H-3); 72.82 (*C*H-2); 71.92 (*C*H-8"); 68.80 (*C*H-7"); 68.49 (*C*H-4); 68.44 (*C*H-2'); 68.11 (*C*H-4'); 62.54 (*C*H-9"); 61.24 (*C*H-6'); 60.05 (*C*H-6); 57.20 (O*Me*); 51.61 (*C*H-5"); 50.61 (*C*H-3'); 40.62 (*C*H-3"); 22.01 (*Ac*); ESI-MS *m/z* (%): 662.67 (100%) $[M-H]^-$; $[\alpha]_D^{21°C}$ = +37 (0.5 mg/mL in H_2O).

Nuclear magnetic resonance experiments: NMR spectra were collected by a Bruker 600-MHz Avance Neo instrument fitted with a cryo probe. NMR samples were dissolved in 50 mM deuterate phosphate buffer (NaCl 140 mM, Na_2HPO_4 10 mM, KCl 3 mM, pH 7.4) and the [D4](trimethylsilyl) propionic acid, sodium salt (TSP, 10 uM) was used as internal reference to calibrate all the spectra. Data acquisition and processing were analyzed using TOPSPIN 3.2 software. The chemical shifts of the glycan ligands were assigned by ^{1}H, COSY, TOCSY, NOESY and HSQC experiments (**Tables S1**, **S2**). ^{1}H NMR spectra were registered by using 16 k and 32 k data points. The homonuclear spectra were recorded with data sets of 4096x512 (t1 x t2) points and the data matrix processed with zero-filled in the F1 dimension up to 4096x2048 points. In order to improve the resolution, a cosine-bell function was used before Fourier transformation in both dimensions. Heteronuclear single quantum coherence (HSQC) experiments were carried out in the ^{1}H-detected mode by single quantum coherence with proton decoupling in the ^{13}C domain, setting data points of 2048x256. The method of States *et al.*[29] was employed for the experiments in the phase-sensitive mode.

Kinetic analysis: For the analysis of the enzyme kinetic of 3'SLN and thio-3'SL catalyzed by MuV-HN from SBL-1 strain, a 1:70 p/l molar ratio was used. The enzyme and the substrate were dissolved in 50 mM PBS/D2O buffer, pH 7.4. and T 298K. Both analyses were performed by using the following procedure: before the addition of MuV-HN of SBL-1 strain, a 1D proton spectrum with the application of composite pulses for water presaturation (zgppr) was acquired with 32 transients, thus obtaining the resonances at t 0. After the addition of the ligand, followed by a short shimming routine, ^{1}H NMR spectra using the above acquisition parameters were recorded at different time points for several hours. A timer was set to measure the delay between the addition of the protein and the collection of the first

NMR spectrum of the mixture. The delay time was incorporated into the kinetic analysis of 3'SLN. The well dispersed resonances of the substrate and the products were integrated at each time point; then, the corresponding concentrations were determined with respect to the integrals of H3eq resonances of the substrate. For the hydrolysis of 3'SLN by MuV-HN, the progress curve was fitted in Sigma Plot with the equation adapted from reference.[30]

Saturation transfer difference analysis: STD NMR spectra were acquired with a protein:ligand ratio of 1: 50, with the on-resonance pulse at 6.5ppm and the off-resonance at 40ppm. By using these conditions, no STD signals were observed in the control STD NMR spectrum of the ligand alone. A train of 50 ms (field strength of 21 Hz) Gaussian shaped pulse with an attenuation of 60 db has been used to saturate the protein. And zero-filled up to 64 k data points prior to processing. The epitope mapping of ligand 1 was achieved by the calculation of the ratio (I0–Isat)/I0, where (I0–Isat) is the intensity of the signal in the STD NMR spectrum and I0 is the peak intensity referred to the unsaturated reference spectrum (off-resonance).

Homology modelling and docking calculations: The sequence encoding for MuV-HN (UniProtKB: P19762-1) was obtained from Uniprot. For computational 3D structure calculation by homology modeling, the 3D coordinates of MuV-HN head domain (PDB ID: 5B2D, chain A) were considered as template. The sequence of the target was aligned to the template using BLAST[31] and the target- template alignment was submitted to SWISS-MODEL to generate the homology model.[11] Then, the obtained structure was optimized by means of Maestro suite of program.[12] For the optimization, missing hydrogen atoms were added and the protonation state of ionizable groups was computed by using Maestro Protein Preparation Wizard.[12] The structure was then submitted to 100,000 steps of steepest descent minimization with MacroModel and optimized with OPLS_2005 force field. The stability of the model was then evaluated by means of PROCHECK tool implemented within SAVES resource. The 3D coordinates of 3'-SLN and thio-3'SL were built with the help of glycan builder and Maestro. The bonds were parametrized and Kollman charges added by means of Maestro and the geometries of each ligand were optimized by 100,000 step of steepest descent minimization with OPLS2005 force field by using Macro Model.[32,33] Ligands were prepared for docking calculations using AutoDockTools;[18] docking calculations of all compounds were performed by using AutoDock 4.2.2,[18] and the analysis of the docking poses was also performed with AutoDockTools.[18]

The docking protocol was validated using the crystal structure of 3' sialyllactose ligand bound to MuV-HN (PDB entry: 5B2D) with the aim to assess if the binding pose of the crystal was correctly predicted by docking calculations. In detail, the structure of 3' sialyllactose was extracted from the 5B2D PDB file and docked into the binding pocket of MuV-HN from Hoshino strain. The docking parameters (grid spacing and coordinates, population size and number of energy evaluations) were optimized through this system and then employed for the docking of 3'SLN and thio-3'SL into MuV- HN SBL-1 strain model. In detail, the grid point spacing was set at 0.375 Å, and a hexahedral box was built with x, y, z dimensions: 58 Å, 56 Å, 60 Å centered in the centroid position among the active site residues Glu407, Arg422, Arg512, Tyr540 residues. A total of 200 runs using Lamarckian Genetic algorithm was performed, with a population size of 100 and 250,000 energy evaluations.

Molecular dynamics simulations: MD simulations of apo-MuV-HN and MuV-HN/3'SLN complex were carried out using AMBER 18 suite of programs.[17] Prior to MD simulation, the complex structure was refined upon the addition of missing hydrogen atoms and computing of protonation state of ionisable groups by means of Maestro Protein Preparation Wizard.[12]

Capping residues at C- and N- termini were also added using Maestro. By using the Leap module, atom types and charges were assigned according to AMBER ff14SB force field for the protein and GLYCAM-06j-1 force field to represent the ligand. The complex was hydrated with an octahedral box containing explicit TIP3P water molecules buffered at 10Å, also, Na^+ counter ions were added to neutralize the system by using the Leap module. The system minimization was performed using Sander and MD simulations were carried out using the CUDA, which is distributed within the AMBER 18 package. The smooth particle mesh Ewald method was used to represent the long-range electrostatic interactions in the system while each simulation was under periodic boundary conditions, and the grid spacing was set to 1Å. The minimization, heating and equilibration procedures prior to MD simulation were divided in several steps. The system was first minimized by applying a restriction to the protein or complex which was gradually released in the following steps. The minimization consisted of 1,000 steps of the steepest descent algorithm followed by 8,000 steps of conjugate gradient algorithm; with the application of 100 kcal·mol−1 A−2 harmonic potential as constraint that was progressively lowered before the final production. Then slow system thermalization from 0 K to 100 K was carried out applying a solute restraint of 10 kcal/mol using the Langevin thermostat in the canonical ensemble (NVT). Then, the system was heated from 100 K to 300 K in an isothermal-isobaric ensemble under the same restraint using Langevin thermostat for isothermal simulation and Berendsen barostat for constant pressure simulations (NPT). Thereafter, temperature was kept constant at 300 K during 100 ps without harmonic restraint. The systems then advanced in an isothermal-isobaric ensemble along the production run which was performed using the Langevin thermostat and Berendsen barostat, with a 2 fs time step and lasting 100 ns. The resulting trajectory was processed and analyzed (RMSD, RMSF, clustering and hydrogen bonds monitored) by means of the cpptraj module included in Amber Tools.[34] The trajectory clustering was carried out with respect to the ligand RMSD and using the K-means algorithm[35] to acquire a total of 5 representative structures of the MD simulation.

Fluorescence titration: Steady-state fluorescence spectra have been collected on a Fluoromax-4 spectrofluorometer (Horiba, Edison,NJ, United States) at the fixed temperature of 10°C. Emission spectra were recorded in the emission range of 300–500 nm upon excitation at 285 nm. The slit widths were fixed at 4 nm for the excitation and 4 nm for the emission wavelength. A quartz cuvette with a path length of 1 cm and 1.5 ml volume was employed. For each experiment, MuV-HN solution at fixed concentration of 0.25 μM in PBS buffer (pH 7.4) was titrated by adding small aliquots (1–50 μl of a ligand stock solution of 700 μM) of 3'-SLN and thio-3'SL respectively. The binding curve was obtained by plotting ΔF/ΔFmax values versus ligand concentration as described by Ribeiro *et al*.[21]

6.4.2 Supporting tables

Unit	1	2	3	4	5	6	7	8	9
GlcNAc A									
^{1}H	4.52	3.77	3.73	3.73	3.61	4.02/3.86			
^{13}C	101.18	55.09	72.16	78.28	74.78	60.00			
CH_3									
^{1}H		1.96							
^{13}C		26.76							
Gal B									
^{1}H	4.56	3.58	4.12	3.97	3.72	3.74/3.73			
^{13}C	102.54	69.41	75.53	67.46	75.26	61.01			
Neu5Ac K									
^{1}H			$2.77_{eq}/1.81_{ax}$	3.70	3.85	3.65	3.61	3.90	3.88/3.66
^{13}C			39.81	68.46	51.66	72.93	68.18	71.84	62.69
CH_3									
^{1}H					1.95				
^{13}C					26.76				

Table S1. ^{1}H and ^{13}C chemical shifts (ppm) of 3'SLN (Neu5Ac-α-(2,3)-Gal-β-(1,4)-GlcNAc-β-$O(CH_2)_2NH_2$).

Unit	1	2	3	4	5	6	7	8	9
Glc A									
^{1}H	4.42	3.32	3.52	3.60	3.56	4.03/3.86			
^{13}C	103.16	72.74	74.84	78.04	74.36	60.00			
Gal B									
^{1}H	4.55	3.40	3.38	3.88	3.69	3.74|3.72			
^{13}C	104.13	68.22	50.58	68.74	77.75	61.01			
Neu5Ac K									
^{1}H			$2.82_{eq}/1.83_{ax}$	3.68	3.87	3.59	3.59	3.94	3.88/3.65
^{13}C			40.38	68.46	51.49	74.88	68.06	71.95	62.69
CH_3									
^{1}H					2.04				
^{13}C					21.90				

Table S2. ^{1}H and ^{13}C chemical shifts (ppm) of thio-3'SL (Neu5Ac-α-(2,3)-Gal-β-(1,4)-Glc-β-OCH_3).

6.4.3 Supporting figures

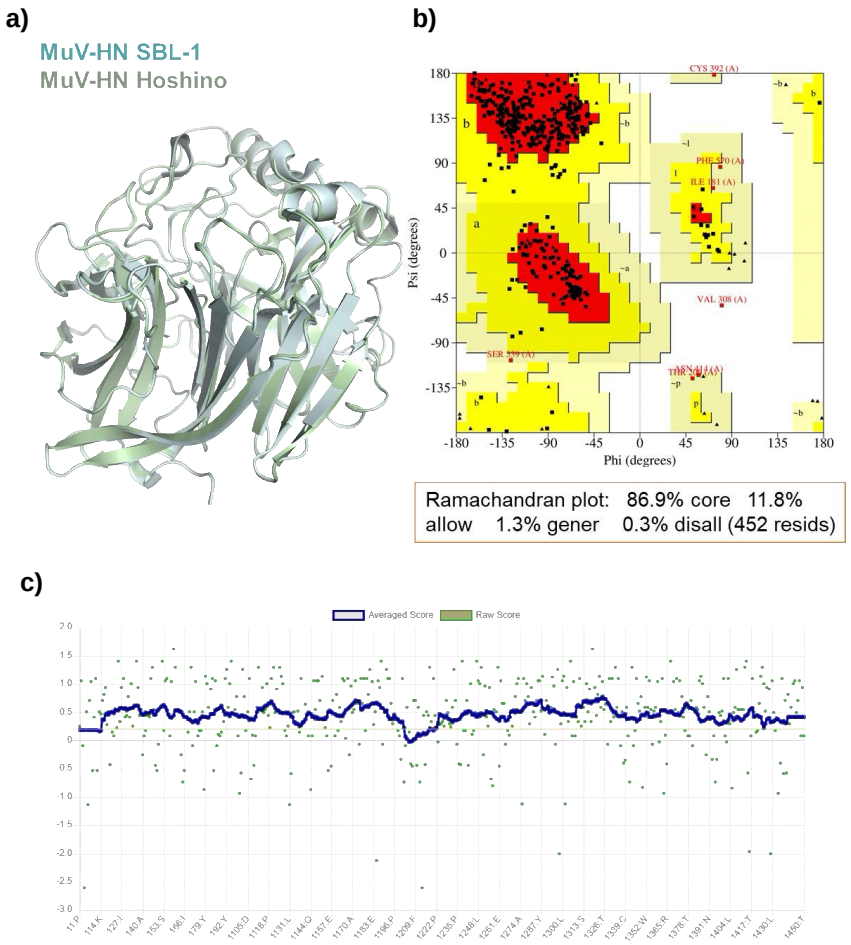


Figure S1. Validation of the homology model of MuVHN-SBL-1 a) Superimposition of the head domains of MuV-HN SBL-1 model (green), and MuV-HN Hoshino strain crystal structure, PDB entry 5B2D (blue); b) PROCHECK analysis (Ramachandran plot); c) Verify3D plot of the homology model.

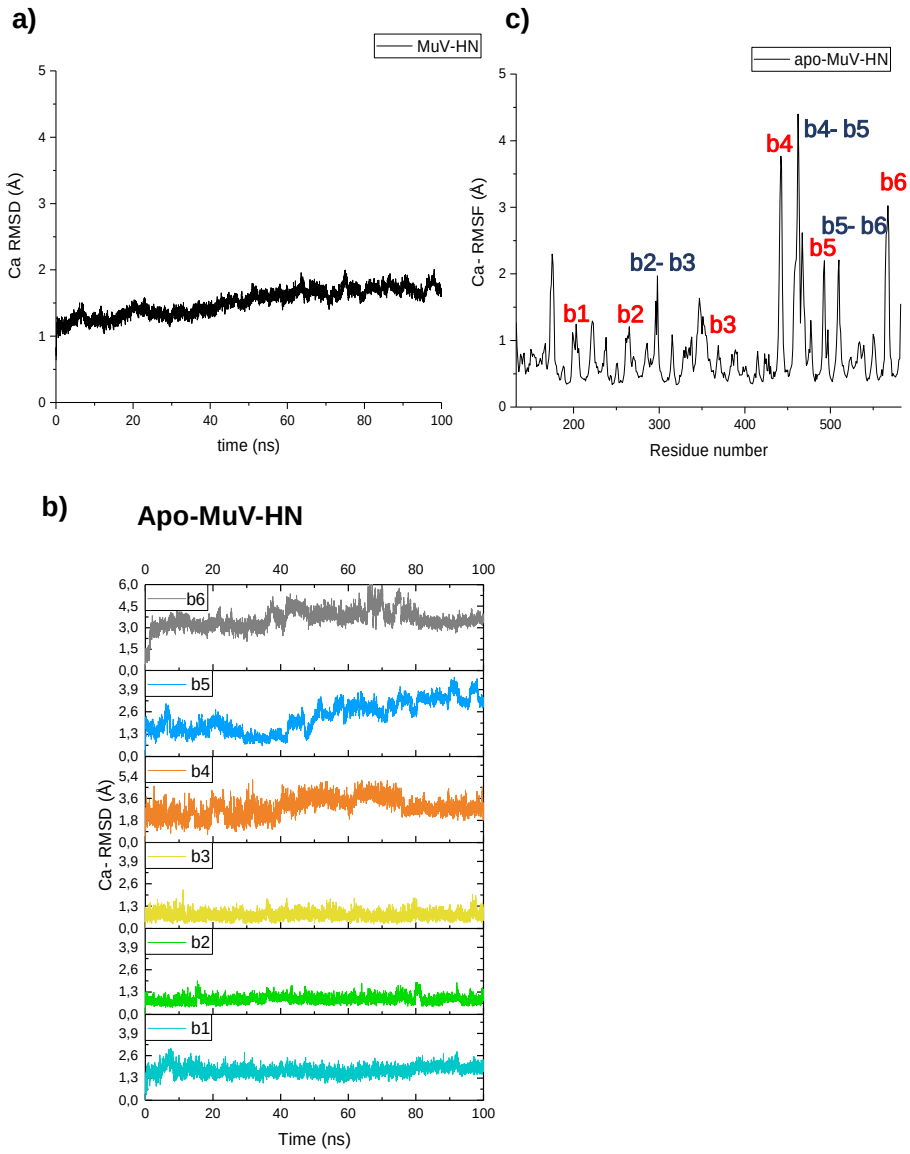


Figure S2. MD simulation analysis of MuV-HN homology model. a) Backbone RMSD of the protein along the trajectory; b) backbone RMSD of the inter-strand loops, namely β1, β2, β3, β4, β5, β6; c) per residue atomic fluctuation of the structure, calculated using the protein C□ atoms. The peaks in the RMSF plot corresponded to the mobile loops connecting the six ß-sheets and the b strands in each of the b-sheets. Inter-strand loops: β1 (Asn199-Asn209), β2 (Val259-Gln276), β3 (Tyr365-Val375), β4 (Thr438-Ser446), β5 (Phe501-Ile515), β6 (Glu561-Phe570) Inter-sheet loops: β2- β3 (Ile291-Asn300), β4- β5 (Ile452-Ser477), β5- β6 (Phe530- Pro535).

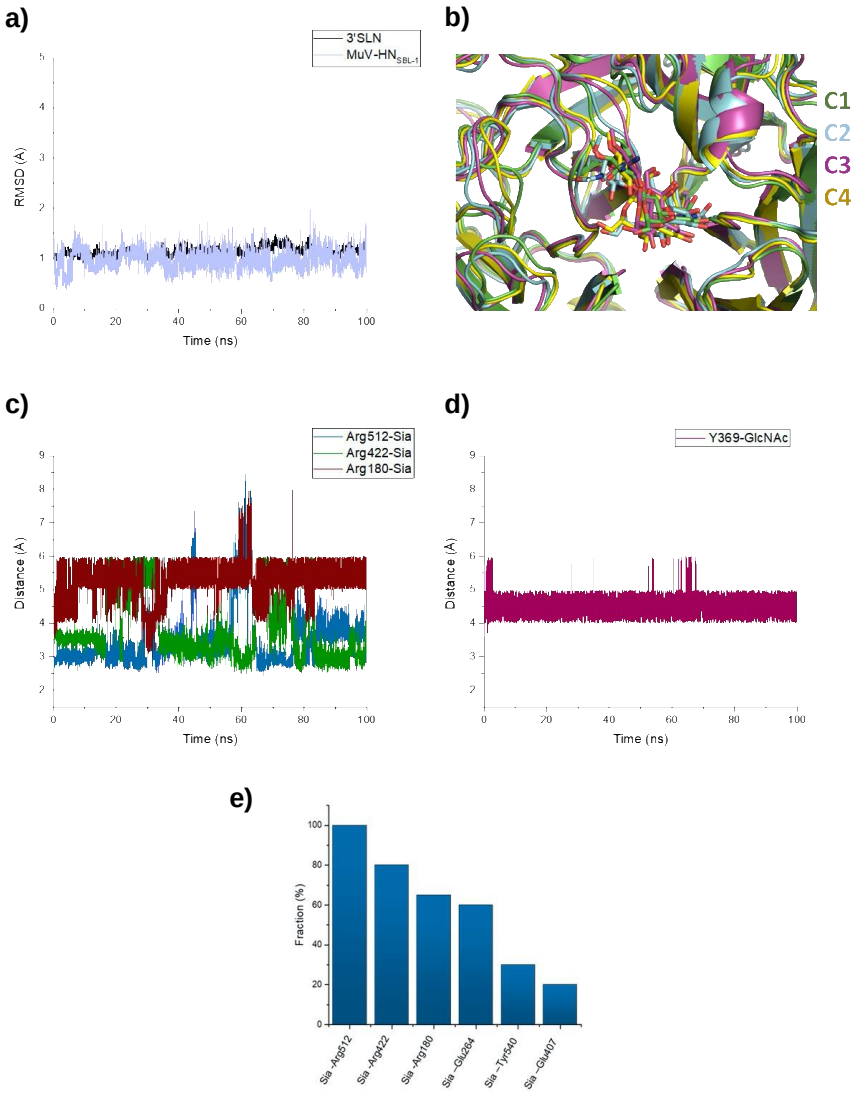


Figure S3. MD simulation analysis of MuV-HN/3'SLN complex. a) Protein and ligand backbone RMSD variation along the MD; b) superimposition of the most populated clusters (C1-C4) derived according to k-mean algorithm estimation; c) distance between MuV-HN triarginyl cluster: Arg180, Arg422 and Arg512 and Sia carboxylate. The average between the heavy atoms directly involved in the interaction has been considered; d) average distance between Tyr369 aromatic ring protons and H1-H3-H5 atoms of GlcNAc unit; e) frequency most representative protein/ligand distances.

6.5 Bibliography

[1]Willocks L.J., Guerendiain D., Austin H.I., Morrison K.E., Cameron R.L., Templeton K.E., *et al*., *Epidemiol. Infect.*, **2017**, 145(*15*), 3219-3225

[2]Donahue M., Hendrickson B., Julian D., Hill N., Rother J., Koirala S., *et al.*, *MMWR Morb. Mortal. Wkly. Rep*., **2020**, 69(*22*), 666-669

[3]Lamb R.A., Kolakofsky D., *Fields Virology 4th ed.,* **2001**

[4]Kubota M., Matsuoka R., Suzuki T., Yonekura K., Yanagi Y., Hashiguchi T., *J. Virol*., **2019**, 93(*15*), e00344-19

[5]Chan J., Watson J.N., Lu A., Cerda V.C., Borgford T.J., Bennet A.J., *Biochemistry*, **2012**, 51 (*1*), 433-441

[6]Goudar C.T., Harris S.K., McInerney M.J., Suflita J.M., *J. Microbiol. Methods.,* **2004**, 59(*3*), 317-326

[7]Forgione R. E., Di Carluccio C., Kubota M., Manabe Y., Fukase K., Molinaro A., *et al.*, *Sci. Rep.*, **2020**, 10(*1*), 1589

[8]Di Carluccio C., Forgione R.E., Martini S., Berti F., Molinaro A., Marchetti R., *et al.*, *Carbohydr. Res*. **2021**, 503, 108313

[9]Palacios G., Jabado O., Cisterna D., de Ory F., Renwick N., Echevarria J.E., *et al.*, *J. Clin. Microbiol*., **2005**, 43(*4*), 1869-1878

[10]Gouma S., Vermeire T., Van Gucht S., Martens L., Hutse V., Cremer J., *et al.*, *Sci. Rep.*, **2020**, 10(*1*), 5456

[11]Waterhouse A., Bertoni M., Bienert S., Studer G., Tauriello G., Gumienny R., *et al.*, *Nucleic Acids Res.* **2018**, 46(*W1*), W296-W303

[12]Schrödinger Suite, **2019**, Schrödinger Suite 2019-2 Protein Preparation Wizard; Epik, Impact, Prime. New York, NY: Schrödinger, LLC

[13]Laskowski R.A., MacArthur M.W., Moss D.S., Thornton, J.M., *J. Appl. Cryst*., **1993**, 26, 283-291

[14]Colman P.M., Hoyne P.A., Lawrence M.C., *J. Virol*., **1993**, 67 (*6*), 2972-2980

[15]Kubota M., Takeuchi K., Watanabe S., Ohno S., Matsuoka R., Kohda D., *et al.*, *Proc. Natl. Acad. Sci. USA*, **2016**, 113, 11579-11584

[16]Winger M., von ltzstein M., *J. Am. Chem. Soc.,* **2012**,134(*44*), 18447-18452

[17]Case D.A., Cerutti D.S., Cheatham T., Darden T., Duke R.E., Gieses T.J., *et al., AMBER 2018*, **2018**, San Francisco, CA: University of California

[18]Morris G.M., Huey R., Lindstrom W., Sanner M.F., Belew R.K., Goodsell D.S., *et al*., *J. Comput. Chem.*, **2009**, 30, 2785-2791

[19]Turnbull W.B., Field R.A., *Perkin Trans.*, **2000**, 1, 1859-1866

[20]Yoneda Y., Kawada T., Rosenau T., and Kosma P., *Carbohydrate Research*, **2005**, 340(*15*), 2428-2435

[21]Ribeiro M.M.B., Franquelim H.G., Castanho, M.A.R.B., Veiga A.S., *J. Pept. Sci.*, **2008**, 14(*4*), 401-406

[22]Gouma S., Vermeire T., Van Gucht S., Martens L., Hutse V., Cremer J., *et al.*, *Sci. Rep.,* **2018**, 8(*1*), 13337

[23]Amaro R.E., Cheng X., Ivanov I., Xu D., McCammon J.A., *J. Am. Chem. Soc*., **2009**, 131 (*13*), 4702-4709

[24]Amaro R.E., Swift R.V., Votapka L., Li W.W., Walker R.C., and Bush R.M., *Nat. Commun.*, **2011**, 2, 388

[25]Kubota M., Hashiguchi T., *Methods Mol. Biol*., **2020**, 2132, 641-652

[26]Reeves P.J., Callewaert N., Contreras R., Khorana H.G., *Proc. Natl. Acad. Sci.*, **2002**, 99(*21*), 13419-13424

[27]Aricescu A.R., Assenberg R., Bill R.M., Busso D., Chang V.T., Davis S.J., *et al*., *Acta Crystallogr. D. Biol. Cryst.* **2006***,* 62, 1114-1124

[28]Hashiguchi T., Kajikawa M., Maita N., Takeda M., Kuroki K., Sasaki K., *et al.*, *Proc. Natl. Acad. Sci.* **2007**, 104, 19535-19540

[29]States D.J., Haberkorn R.A., Ruben, D.J., *J. Magn. Reson*. **1982**, 1969(*48*), 286-292

[30]Her C., Alonzo A.P., Vang J.Y., Torres E., Krishnan V.V., *J. Chem. Educ.*, **2015**, 92(*11*), 1943-1948

[31]Altschul S. F., Gish W., Miller W., Myers E.W., Lipman D.J., *J. Mol. Biol*., **1990**, 215 (*3*), 403-410

[32]Schrodinger, **2012**, Epik Version 2.3, Impact version5.8, Prime Version 3.1. New York, NY: Schrodinger, LLC

[33]Schrödinger Release, **2021**, Schrödinger Release 2021-1: MacroModel. New York, NY: Schrödinger, LLC

[34]Case D.A., Aktulga H.M., Belfon K., Ben-Shalom I.Y., Brozell S.R., Cerutti D.S., *et al., AMBER 2021*, **2021**, San Francisco, CA: University of California

[35]Jain A.K., Murty M.N., Flynn P.J., *ACM Computing Surveys*, **1999**, 31, 264-323

List of Publications & Author Contributions

1. A Simple Biomimetic Receptor Selectively Recognizing the GlcNAc₂ Disaccharide in Water (Chapter 2).
Francesconi O., Milanesi F.,[1,2,3,4] Nativi, C., Roelens S., *Angew. Chem. Int. Ed.,* 2021, 60, 11168-11176.

2. Molecular Recognition of Disaccharides in Water: Preorganized Macrocyclic or Adaptive Acyclic? (Chapter 3).
Francesconi O., Milanesi F.,[1,2,3,4] Nativi C., Roelens S., *Chem. Eur. J*., 2021, 27, 10456-10460.

3. Characterization of natural and synthetic sialoglycans targeting the Hemagglutinin-Neuraminidase of mumps virus (Chapter 6).
Marchetti R., Forgione R.E., Di Carluccio C., Milanesi F.,[1,4] Kubota M., Nieto F.F., Molinaro A., Hashiguchi T., Francesconi O., Silipo A., *Front. Chem.,* 2021, *9*, article 711346.

4. Conformationally Constrained Sialyl Analogues as New Potential Binders of h-CD22 (Chapter 5).
Forgione R.E., Nieto F.F., Di Carluccio C., Milanesi F.,[1,4] Fruscella M., Papi F., Nativi C., Molinaro A., Palladino P., Scarano S., Minunni M., Montefiori M., Civera M., Sattin S., Francesconi O., Marchetti R., Silipo A., *ChemBioChem.,* 2022, *23*, e202200076.

5. *A biomimetic receptor for glycans. Selective recognition of the core GlcNAc₂ disaccharide of the sialylglycopeptide SGP* (Chapter 4).
Milanesi F.,[1,2,3,4] Unione L., Arda A., Nativi C., Roelens S., Jiménez-Barbero J., Francesconi O., *Manuscript in preparation*, 2022.

Notes:
[1]Synthesis and characterization of chemical materials.
[2]^{1}H NMR, ITC and CD titrations.
[3]NMR studies.
[4]Manuscript preparation.