Interfering with the sugar code: ten years later

Sara Sattin[a] and Anna Bernardi*[a]

Department of Chemistry, Università degli Studi di Milano, via C. Golgi, 19, 20133, Milan (Italy), e-mail: anna.bernardi@unimi.it

Dedicated to Jesus Jiménez Barbero on his 60th birthday

Abstract: This paper reviews the past ten years of our research in the field of glycomimetics synthesis and design. The structure-based design of mono- and polyvalent lectin ligands is one of the three main areas that we have explored in this period. We summarize here our work on ligands targeted against the dendritic cell receptor DC-SIGN. A second direction we have followed involves the discovery of lectin antagonists by screening of glycomimetic libraries. This approach has led to interesting hits against cholera toxin and related bacterial enterotoxins, as well as, in a different campaign, against the human C-type lectin dectin-2. Underlying both approaches, the development of robust synthetic methodologies for the rapid and selective synthesis of unnatural glycoconjugates has been our attempt to contribute to the diversification and enrichment of the chemical toolbox of glycoscience.

1. Introduction

The complexity of the glycome and the many biological functions of glycans that are continuously being elucidated have led to the concept of a "sugar-code", *i.e.* a set of sugar-mediated information that is "read" by specific sugar-binding proteins called lectins. Studying the sugar code and controlling its flow of information by interfering with it through molecular probes can not only advance fundamental understanding of biology, but also pave the way to new treatments. In 2008, in a Concept article in Chem Eur J, we discussed our early attempts to design and synthesize glycomimetic molecules targeted against lectins.[1] Here we review our progress in the field over the past 10 years.

The design of lectin ligands is especially challenging because of the intrinsic nature of lectins' binding sites, that are generally rather shallow and solvent-exposed. They have evolved to specifically recognize sugars, which have been described as "pre-organized water oligomer(s)",[2] with a generally low intrinsic affinity. Nature overcomes this problem by increasing affinity through polyvalency: multiple copies of glycans, in the form of glycoconjugates, are simultaneously presented to polyvalent lectins, thus increasing the local concentration of binding partners and profiting from avidity. Polyvalency has been embraced with success also for the synthesis of lectin antagonists.[3] Antagonists of lectin-sugar recognition have mostly been designed as glycomimetic structures, *i.e.* molecules that attempt to reproduce the three-dimensional structure and function of the native oligosaccharide ligand. Non-carbohydrate small molecules have also been reported to interact with lectins in the vicinity of the carbohydrate-binding site, but it has been shown that at least for C-type lectins they exert their action *via* allosteric mechanisms, rather than by straightforward competition with the sugar moiety.[4] Indeed, one could argue that, given the exquisite selectivity for (mono)saccharides tailored into lectins by evolution, a modified monosaccharide anchor may well be the most efficient way of generating antagonists. The additional elements that one can append to the anchor must serve multiple purposes, and in particular must increase the ligand affinity for the lectin while providing the molecule with improved pharmacokinetics and pharmacodynamics properties, in the context of drug design.[5]

Since we reviewed this topic about ten years ago, successful clinical trials for galectin modulator TD139 and selectin antagonists rivipansel (GMI‑1070) and uproleselan (GMI‑1271) have greatly increased the expectations of developing therapies based on carbohydrate-binding targets. From a synthetic point of view, recent remarkable developments of carbohydrate chemistry have unlocked a wide variety of structural modifications, leading to an assortment of different sugar mimics with the potential of performing as lectin antagonists.[6]

In this framework, our research over the past ten years has moved along three different lines that are covered in this Minireview: the structure-based design of new mono and polyvalent lectin ligands, the synthesis and screening of glycomimetic libraries (ligand discovery) and the development of synthetic methodologies for the fast and stereoselective synthesis of unnatural glycoconjugates.

2. Structure-based design of mono and polyvalent lectin ligands: the case of DC-SIGN

Our initial approach to glycomimetics design has largely relied on reproducing the three-dimensional structure of oligosaccharides' binding determinants using non-carbohydrate scaffolds (ligand-structure based design). In particular, over the past ten years, we have focused on mannose-based antagonists of the dendritic cell receptor DC-SIGN, which is implicated in the initial stages of many viral infections, including HIV[7] and Ebola.[8] In this context we have developed the pseudo-dimannoside **1** and pseudo-trimannoside **2** (Fig. 1) as glycomimetic antagonists with potential antiviral activity.[9] These molecules perform as structural and functional mimics of the corresponding di- and tri-mannosides (Manα1-2Man **3** and Manα1-2-Manα1-6-Manα **4**, Fig. 1) and bind to DC-SIGN in the carbohydrate recognition domain, as shown by X-ray crystallography.[10] Both ligands include a 1,2-dicarbomethoxy-cyclohexane scaffold (blue frame in Fig. 1), conformationally locked to mimic a mannose residue, which maintains the shape of the structure, while increasing the lipophilicity and reducing the susceptibility to enzymatic hydrolysis. In the course of these studies, various R substituents were adopted and tested in each molecule (including R= $-CH_2-CH_2-Cl$, $-CH_2-CH_2-N_3$, $-CH_2-CH_2-NH_2$, $-CH_2-CH_2$-triazole) and found to have minimal effect on the activity (if any). Indeed, the X-ray structure of the DC-SIGN complexes of both **1** and **2** shows the R group pointing towards the solvent.

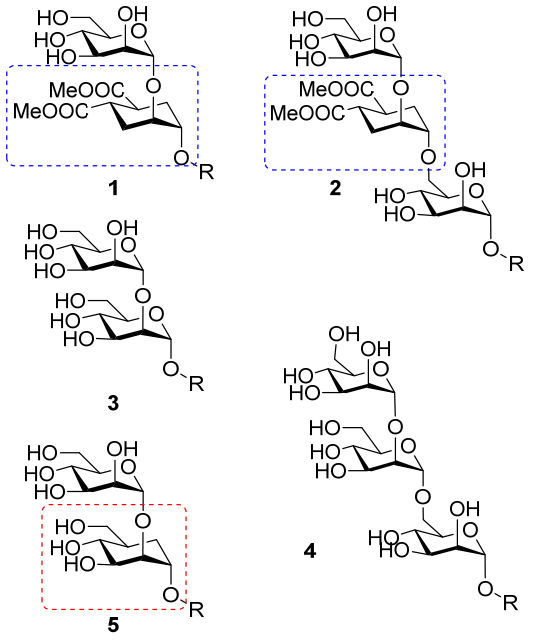


Figure 1. The Manα1-2Man disaccharide **3** and Manα1-2-Manα1-6-Manα trisaccharide **4** and their mimics **1,2,** and **5**

We showed that **1** binds to DC-SIGN with the same (millimolar) affinity as the parent disaccharide **3**, while being less cytotoxic than mannobiose.[11] For comparison, we also synthesized and tested the real carba analog of **3**, the pseudo-sugar **5** (Fig. 1), that includes a carbamannose residue at the reducing end (red frame). This new pseudo-disaccharide has a similar inhibitory potency against DC-SIGN as **1**, but its synthesis is based on the glycosylation with **6** of the appropriately protected carbamannose **7** (Fig. 2), which was obtained with ca.30% yields over three steps from tri-*O*-acetyl-D-glucal **8**.[12] In contrast, the synthesis of **1** entails as the mannosyl acceptor alcohol **9**, which is synthesized in good yields from **10** in two steps and a single chromatographic purification (Fig. 2). The affinity of these ligands, as well as of the natural sugar, for the DC-SIGN receptor is rather low, but since the lectin is tetravalent, the affinity can be boosted by polyvalent presentation (see below). Thus, the accessibility of the monovalent ligand is key to the further manipulation into polyvalent constructs, and the facile, high-yield synthesis of **1** made it the structure of choice to expand upon.

To optimize the ligand structure, we strived to exploit secondary interaction sites in the immediate vicinity of the sugar-binding site by introducing additional fragments/functional groups that could engage with them. This can be done either by random modifications, suggested by the chemotype of the ligand structure and loosely correlated to the characteristics of the target carbohydrate recognition domain, or, when structural data are available, by using structure-based design. It must be noted that structure-based design of lectins' ligands has seldom been successful. Lectins' binding sites, flat and open to solvent, are not the typical "pocket" which clearly defines excluded volumes and constrains binding modes. Indeed, more often than not, oligosaccharides adopt multiple binding orientations within a single lectin site. DC-SIGN is no exception: the Manα1-2Man disaccharide **3** can use two binding modes, depending on which of the mannose units coordinates the Ca^{2+} ion in the protein binding site, and, with larger oligomannosides, the number of orientations can be even higher.[13] As a consequence, the benefit of any ligand modification becomes hardly predictable. It is therefore a distinctive advantage of the pseudo-disaccharide **1** to display a single binding mode, that was revealed by X-ray crystallography[10a] and could be used as a template for structure-based ligand optimization.

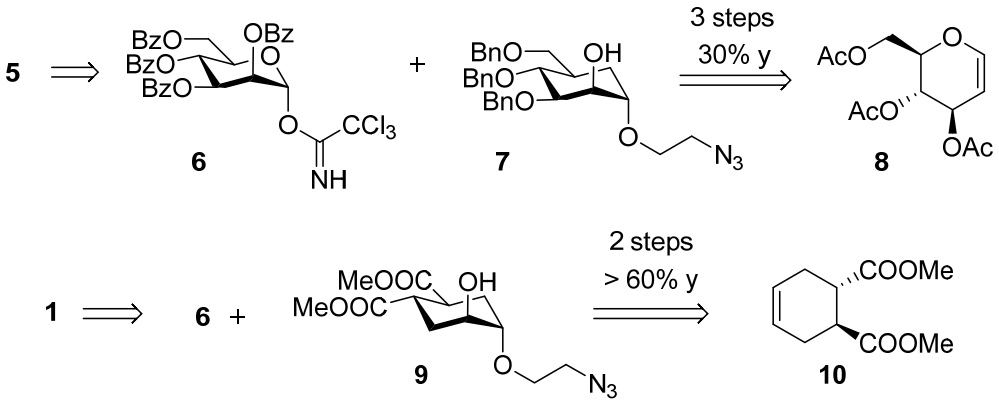


Figure 2. Retrosynthesis of the pseudo-mannobiosides **1** and **5**. The synthesis of the real carbamannose **5** is longer and more complex than the synthesis of **1**, but the two ligands perform equally as DC-SIGN antagonists.

The main issues to be confronted to optimize mannose-based DC-SIGN antagonists are affinity and selectivity against other similar mannose-binding lectins. Among the latter, a main competitor to be excluded is langerin, a lectin which is also mannose-selective, but has a protective effect against HIV infections. To optimize the structure of **1**, we partly relied on random modifications and library screenings detailed below, which resulted in selection of the bis-amide **11** (Fig. 3). The extended surface contact of this ligand within the DC-SIGN binding site leads to a slightly improved affinity, which can be magnified by polyvalent presentation, and to an unanticipated selectivity increase against langerin. Additional and more dramatic improvements were obtained by structure-based rational design, which was made possible after obtaining the X-ray

structure of the **1**/DC-SIGN complex[10a] and included comparative structural analysis of the two proteins.

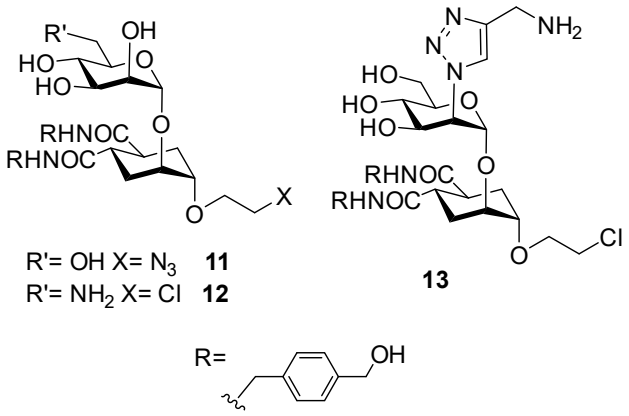


Figure 3. Structural optimization of **1**. Ligand **11** was identified by screening of a bis-amide library.[14] Ligands **12** and **13** were rationally designed by comparative structural analysis of DC-SIGN and langerin (**12**)[15] or by fragment-based drug discovery techniques (**13**).[16]

In particular, for the design of **12** we analyzed the structural and functional differences between the binding sites of DC-SIGN and langerin. Langerin was found to bind 6-sulfated sugars thanks to the presence of a Lys residue (K313) in its binding site. On the contrary, DC-SIGN features a mostly negative electrostatic potential in the area where Man-C6 of **1** finds itself upon binding. These characteristics were exploited to impair interaction of langerin with the pseudo-mannoside by replacing the hydroxyl group at mannose C6 with an amino group in **12**. Thus, in essence, we developed the structure of **12** to disfavor recognition by langerin and obtained a ligand that has an affinity for DC-SIGN similar to **11** (0.5 mM for **12**; 0.3 mM for **11**), but is much more selective against langerin than **11**. In fact, the interaction of **12** with langerin is practically no longer detectable by SPR inhibition assay. This use of comparative structural analysis of two receptors, which we called differential rational design, may be of general use in the development of selective lectin antagonists.[15]
Perhaps more spectacularly, in collaboration with the group of Sonsoles Martin-Santamaria, we were able to use fragment-based screening in the X-ray structure of the DC-SIGN complex of **1** (PDB 2XR5) to identify an ammonium-binding region near the mannose binding site and in proximity of mannose O2 (Fig. 4a). Functionalization of this position *via* a 2-azido-mannose intermediate led to the synthesis of **13**, modified at mannose C2 with a methylene-amino triazole (Fig. 3). This molecule provided one order of magnitude increase in affinity over **1** (K_d 52 ± 1 μM by ITC), while being fully selective against langerin. Gratifyingly, X-ray crystallography of the **13**/DC-SIGN complex revealed that the ligand interacts with the protein as anticipated by the calculations, and places its amino group in the predicted ammonium-binding region (Fig 4b, PDB: 6GHV).[16] One additional interesting feature of ligand **13** is that it no longer contains a mannose unit and therefore should not be a substrate of mannose hydrolases.
As mentioned above, to increase the activity of pseudo-mannoside ligands to useful levels we took advantage of the tetrameric structure of DC-SIGN, which presents four carbohydrate recognition domains, and adopted multivalent presentation. The binding modes that a multivalent glycoconjugate can exploit include chelation, defined as the

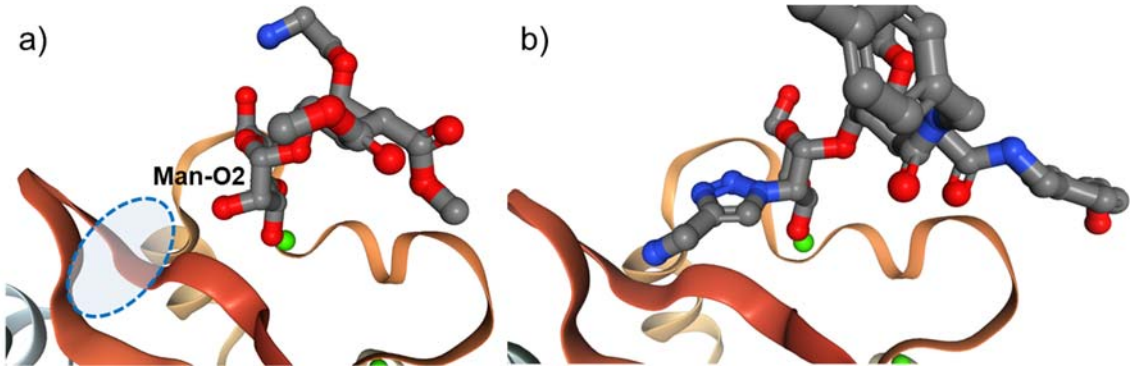


Figure 4. a) Structure of the DC-SIGN/**1** complex (PDB 2XR5). The ammonium-binding region is within the blue ellipse. b) Structure of the DC-SIGN/**13** complex (PDB 6GHV). The amino group is hosted within the ammonium binding region and participates in a H-bonding network which had been computationally predicted.

simultaneous binding of more than one binding site on a multimeric lectin, increased local ligand concentration (or statistical rebinding effect), as well as the ability of the polyvalent construct to cluster multiple copies of the protein receptor. While chelation must be carefully engineered to achieve optimal ligand presentation, the other effects are more easily attained and contribute to significant enhancements in ligand affinity. Indeed, early attempts based on polyester dendrimers and dendrons (Boltorn type) showed 1-2 orders of magnitude increase in affinity over the monovalent ligand **1** in SPR competition assays, depending on the valency of the structure. The ability of these constructs to block Ebola[9b] and HIV-1[9a] infections was established in cellular models and also in a human cervical explant model.[17]
Nonetheless, these polyvalent structures, exemplified by the tetravalent dendron **14** (Fig. 5), remained challenging to synthesize and purify, because the unhindered succinyl ester moieties used as linkers between the core scaffold and the ligand are easily hydrolysed, both in acidic and basic conditions. Replacement of the labile ester bond with a more robust amide functionality was attempted in dendron **15** (Fig. 5[18]), which could be efficiently synthesized from 2,2-bis(hydroxymethyl)propionic

acid **16** (Scheme 1)**.** The synthesis of the scaffold **21** involved the three building blocks **17**, **18** and **20** and was designed to minimize the number of protection/deprotection steps. Compared to **14**, the glycoconjugate construct **15** remained fully water soluble, but became stable to silica gel chromatography and to water solutions at physiological pH. Higher‐order constructs were easily assembled, using the azido-terminated linker at the dendron focal point, as demonstrated by the synthesis of a 16‐valent dendrimer. All these constructs were active against the serum protein Mannose Binding Lectin (MBL), a protein with a crucial role in the recognition of damage-associated molecular patterns, that can be targeted to reduce brain damage after brain ischemia.[19]

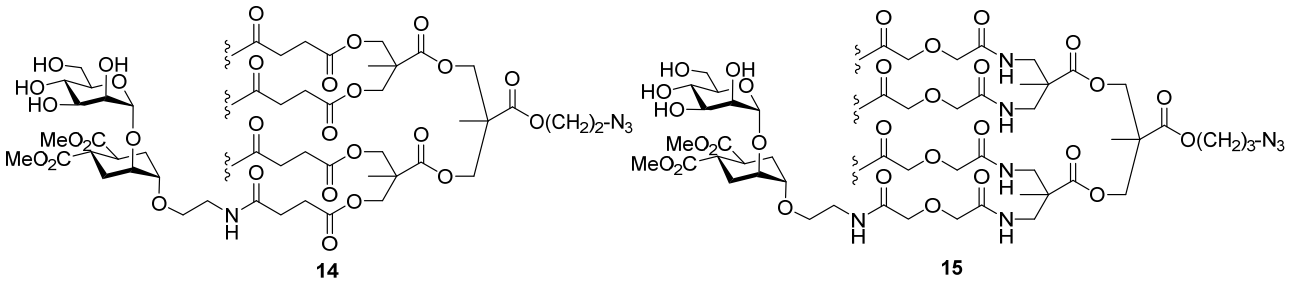
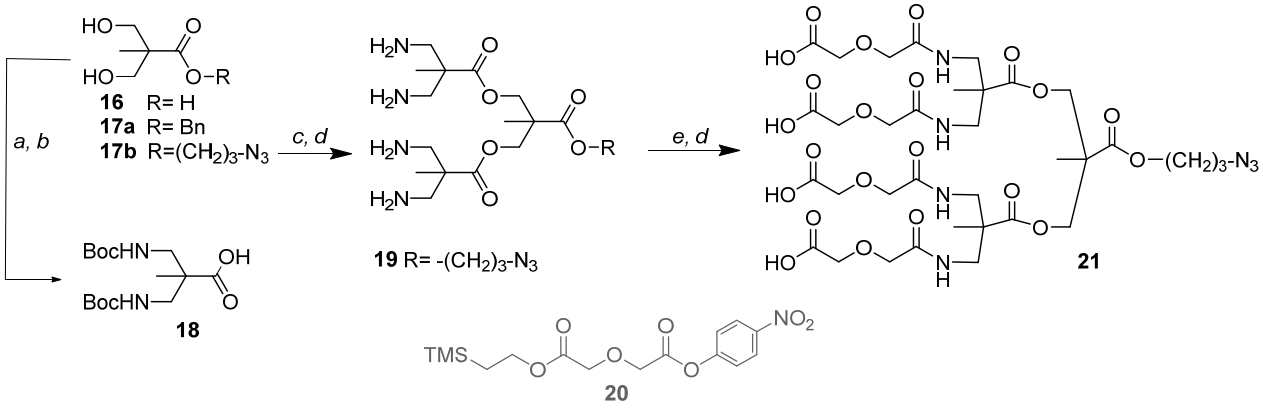


Figure 5. The tetravalent dendrons **14** and **15**.



Scheme 1. Synthesis of **21**: a) From **17a**: TsCl, pyridine, 60°C; then NaN_3, DMA, cat. Bu_4NI (85% over the two steps); b) H_2, Pd/C, Boc_2O, EtOH (**18**, 50%); c) **18**, DCC, DMAP, CH_2Cl_2, then **17b** (78%); d) TFA /CH_2Cl_2 quant. e) **20**, Et_3N, CH_2Cl_2 (76%).

In a different approach, in collaboration with the Rojo group, conjugation chemistry was switched from amide bond formation (as in **14** and **15**) to Cu-catalyzed azide-alkyne cycloaddition (CuAAC), using low-valency (2-6) polyalkyne scaffolds, which could be combined to build higher valency constructs (up to the 18-valent).[20] To this class belongs Polyman-19 (**22**, Fig. 6), a hexavalent scaffold, carrying six copies of **11**, which is highly effective against HIV infection both in cellular (IC_{50} 1 μM in *trans* infection studies) and in cervical explant models.[21]

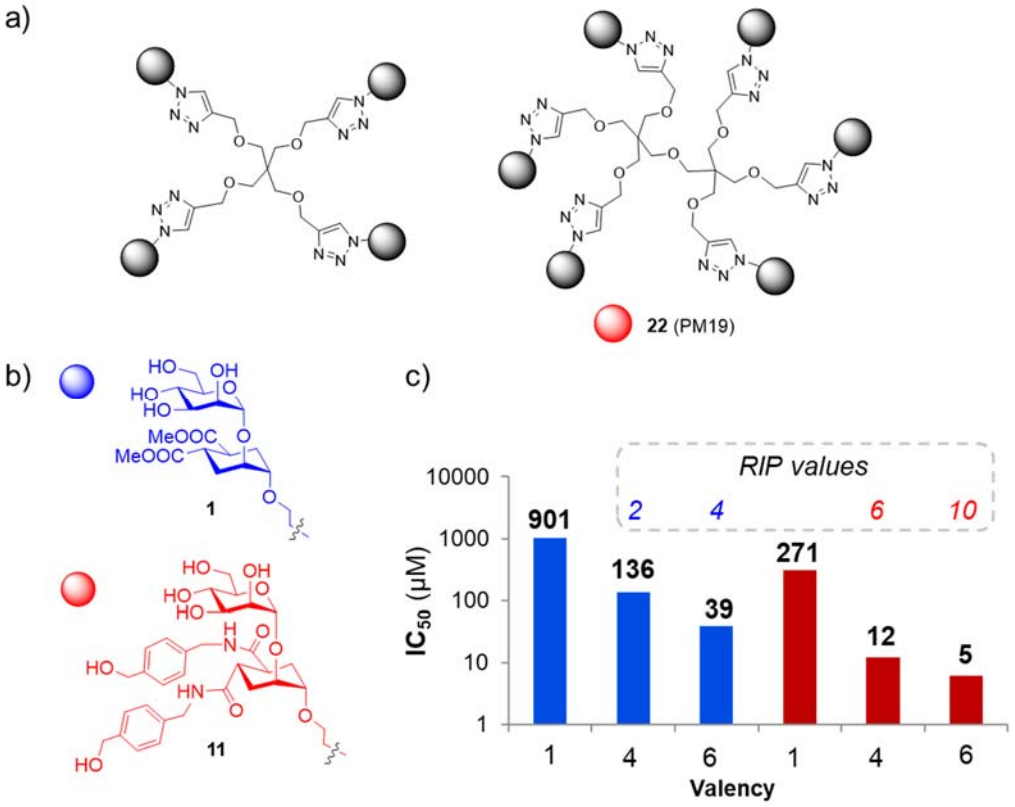


Figure 6. Multivalent presentation amplifies affinity differences. a) general structure of tetra- and hexa-valent dendron cores carrying b) ligand **1** or **11**; c) SPR inhibition studies on immobilized BSA-Man_3. Valency-corrected relative inhibitory potency (RIP) values of tetra- and hexa-valent dendrons carrying either ligand **1** (blue bars) or **11** (red bars) show a dependency on valency and on the nature of the monovalent ligand.

In the course of these studies, in collaboration with the group of Franck Fieschi, we were able to systematically compare the DC-SIGN affinity of a series of constructs through SPR inhibition assays, which afforded valency-corrected relative inhibitory potency (RIP) values.[20] The results showed a dependency both on valency and on the nature of the monovalent ligand (Fig. 6). In particular, affinity differences at the monovalent level were amplified by multivalent presentation: the monovalent ligand **11** is only three times more active than **1** (IC_{50} 0.3 and 0.9 mM, respectively), but its affinity grows faster with valency, so that the hexa-valent presentation is now 8 times more active (IC_{50} 0.005 mM *vs* 0.04 mM), with a RIP 10. Thus, these studies fully supported the hypothesis that optimizing the monovalent ligand structure is actually worth the effort.

The hexavalent presentation of **11**, **22** (PM19), has a low micromolar IC_{50} in the SPR inhibition assay, providing almost 2 orders of magnitude increase over monovalent **11**. Remarkably, this is achieved in the absence of chelating effects, since the estimated size of **22** at full extension (3.5 nm) is well below the distance between two contiguous sites in the DC-SIGN tetramer (3.8 – 4 nM, depending on the models). The chelation mechanism could be exploited by engineering the length and rigidity of the scaffold, which we did using water soluble phenylene-ethynylene rod-like elements as the dendrimer core. A systematic analysis of rod size, valency, length of the additional linker and type of monovalent ligand was made possible by the modularity of the synthesis (Fig. 7), which was based on CuAAC conjugation of preformed building blocks[22] (Fig. 6).

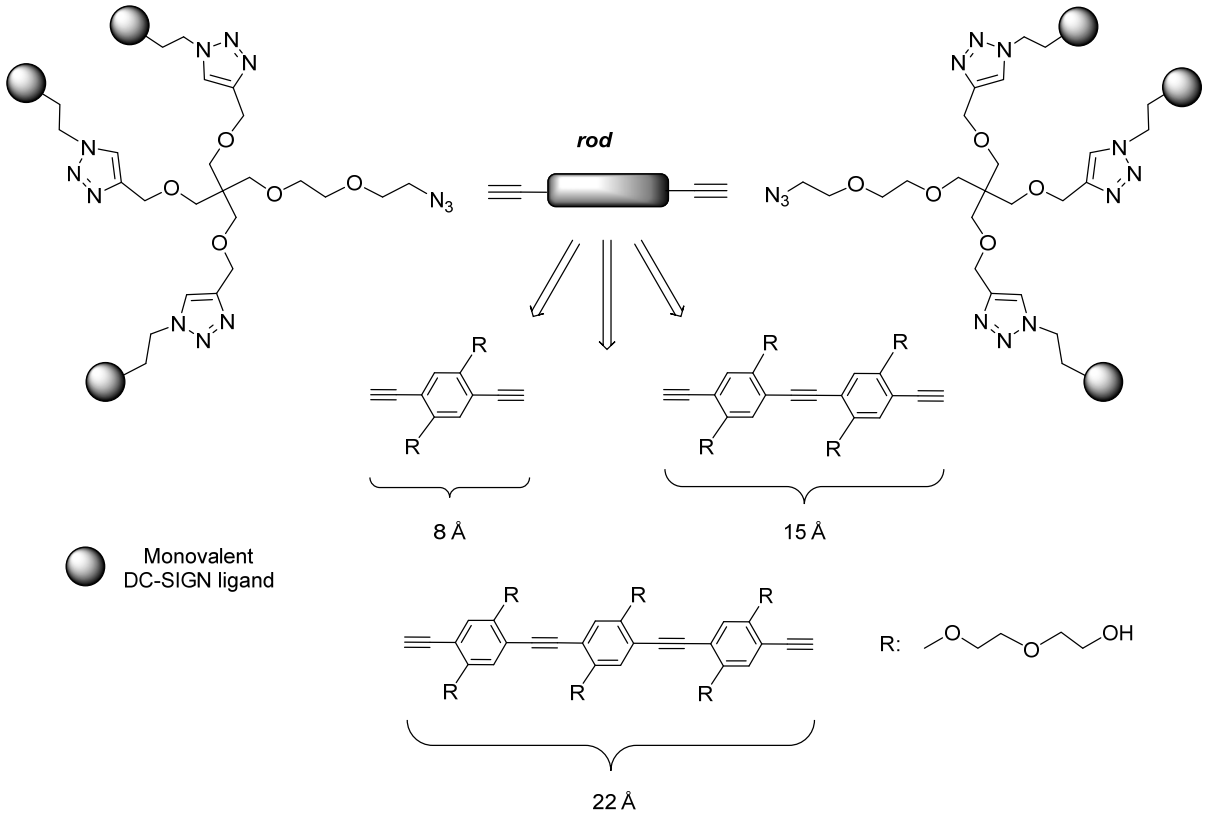


Figure 7. Modular synthesis of rod-based dendrimers. The full structures are assembled by CuAAC of preformed building blocks.

SPR inhibition studies yielded important information about the optimal design of the polyvalent scaffolds. The size and the rigidity of the system are both significant, and the best results are achieved exploiting the longest rod and minimizing the length of the flexible linker used to connect the monovalent ligand. This is apparent in the series of dimers **23-25** shown in Fig. 8. Their models docked in a dimer of DC-SIGN CRDs are shown in the left panel; the structures and IC_{50} values in DC-SIGN inhibition experiments are collected in the right panel. A combination of a short rod and a long flexible linker, as in **23**, allows chelation (as shown by the docked model in the left panel of Fig. 8), but it is less efficient, due to its high flexibility, than a long rod with a short linker (as in **25**), when the rod is long enough to attain chelation. Under these conditions, one order of magnitude activity is gained in the inhibition assay (IC_{50} of **25** 19 μM, IC_{50} of **11** 300 μM). Attaching a flexible linker to the long rod, as in **24**, reduces the activity of the ligand by a factor of 2 (IC_{50} of **24** 34 μM), because more rotational degrees of freedom are lost upon binding.

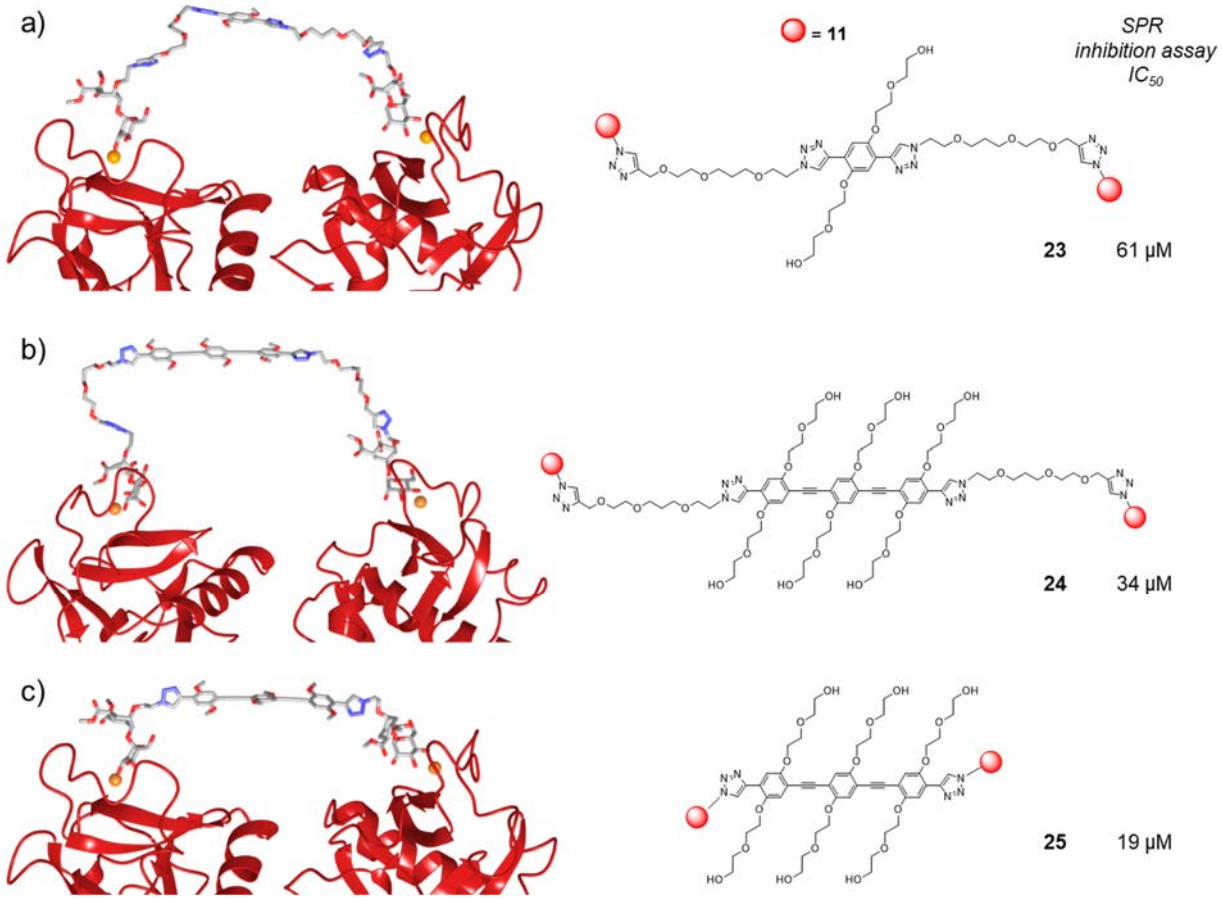


Figure 8. Rod-based divalent ligands. Left panel a-c: DC-SIGN complexes of the ligands, as obtained by docking and MD simulations. The ligand models are simplified for computational convenience: the monovalent pseudo-mannoside **1** replaces **11** and a methyl ether is used in place of the PEG chains on the aromatic rings. Right panel a-c: Ligand structures and DC-SIGN inhibition data. IC_{50} were measured by SPR as the concentration of the ligand required to inhibit 50% of DC-SIGN binding to immobilized mannosylated bovine serum albumin (Man-BSA). The IC_{50} of **11** in the same assay is 300 μM.

We used the long rod to build the hexavalent dendrimer **26** (Polyman26, or PM26, Fig. 9), which is essentially a stretched-out version of **22**, and achieved an impressive IC_{50} of 24 nM in *trans* HIV infection tests in a cellular model, a 2-orders-of-magnitude increase over **22** (IC_{50} 1 μM in the same assay).[22] .

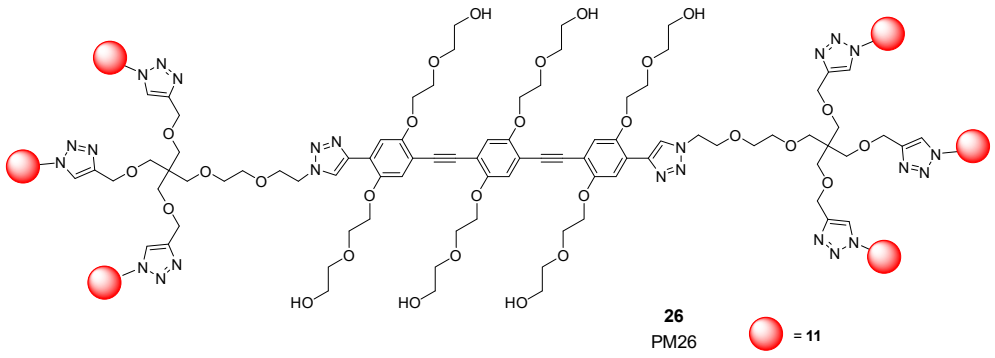


Figure 9. The rod-based hexavalent dendrimer **26** (Polyman26, PM26) shows antiviral activity in nM concentration in cellular models (IC_{50} 24 nM in HIV trans infection studies).

Confocal microscopy studies indicated that **26** is internalized in dendritic cells by DC-SIGN and routed to the lysosomes. The internalization promotes DC maturation and activates immune response in ways that suggest both a potential boost of antiviral activity *in vivo* and a possible use as immunomodulator and adjuvant in the context of vaccine production.[23] Thus, the

combination of chelation effects and higher local concentration of the ligand is beneficial to effectively block the infection and affords one of the most potent DC-SIGN antagonists described so far. It is worth noting that the design components adopted for the synthesis of **26**, due to their modularity, should be easily adapted to other multivalent targets with different binding geometry and distance requirements. We expect to be able to use the same elements to build polyvalent antagonists of lectins other than DC-SIGN.

3. Synthesis and screening of glycomimetic libraries (ligand discovery)

As mentioned above, initial attempts at optimizing the affinity of **1** for DC-SIGN tried to take advantage of hydrophobic interactions in the lectin binding site. To this end, a library of bis-amides was synthesized and evaluated by an in vitro assay that measures inhibition of DC-SIGN-mediated immature dendritic cell adhesion to mannan-coated plates (Fig 10[24]). A second round of screening was focused on bis-benzylamides and allowed to select compound **11** as a combination of increased activity, good solubility, synthetic accessibility, and selectivity against langerin.[14]

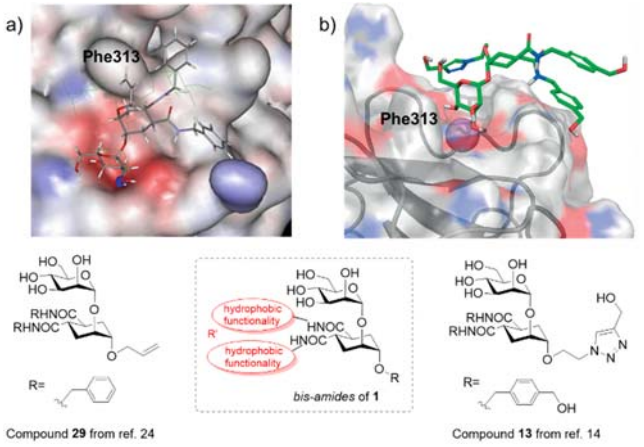
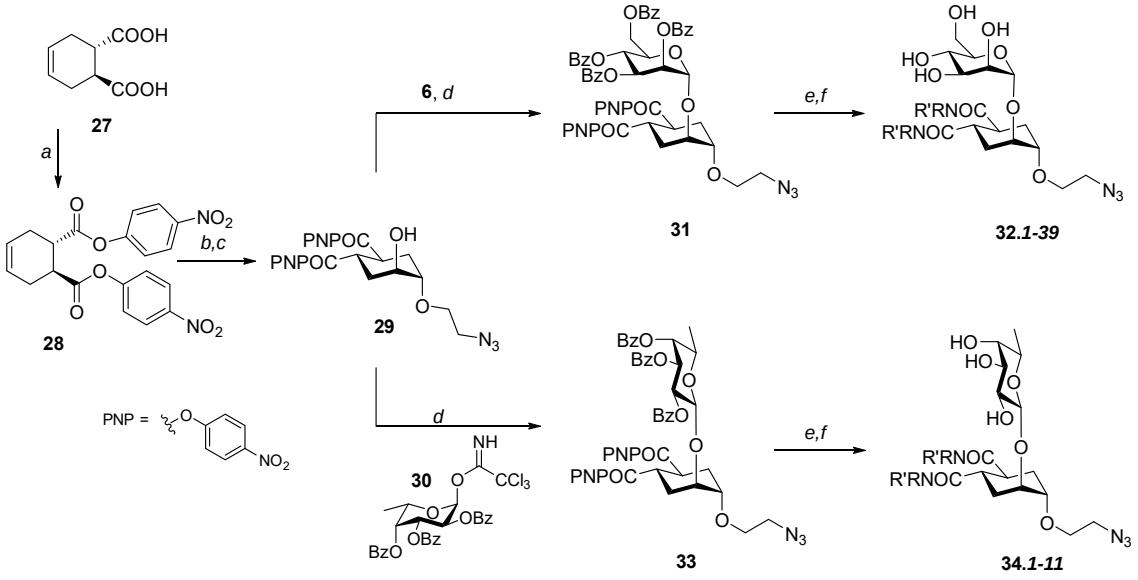


Figure 10. Synthesis and testing a library of bis-amide derivatives of **1** led to selection of ligand **11.** Modified from ref 24; b. Modified from ref 14.

A docking protocol was actually used to evaluate the possible binding mode of the pseudo-dimannoside bis-amides, but its predictivity turned out to be rather low. In our initial calculations,[24] the aglycone extends towards Phe313 (Figure 10a), a convincing pose at the time, since the predicted binding area overlapped with the region occupied by oligomannosides in the known X-ray structures. However, the X-ray structure of the DC-SIGN/**1** complex (PDB 2XR5, Figure 4a) later revealed a different orientation of the pseudo-dimannoside core. This



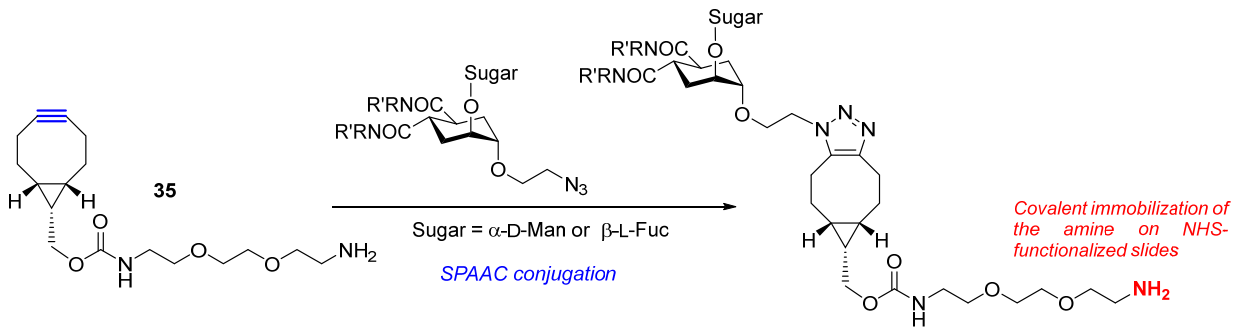
Scheme 2. Synthesis and general structure of mannose- (**32.*1-39***) and fucose-derived (**34.*1-11***) members of the glycomimetic library. a) 4-nitrophenyl trifluoroacetate, pyridine, DMF, 50°C, 73% b) MCPBA, 94% c) 2-azidoethanol, $Cu(OTf)_2$, 70% d) TMSOTf, -30°C, mannose 80%, fucose 74% e) RR'NH, THF/DMF f) NaOMe, MeOH.

information, together with NMR interaction data collected for **11**[14] allowed to select among the docked poses of **11** in DC-SIGN the structure shown in Figure 10b, where the cyclohexane scaffold contacts Val351 (as in PDB 2XR5) and one of the amide residues extends towards the external surface of the protein. This pose, which is basically rotated by 180° relative to Figure 10a, was finally confirmed when the X-ray structure of **13** in complex with DC-SIGN was obtained (PDB 6GHV, Figure 4b) and showed tha aglycone (which **13** shares with **11**) exactly in the position expected based on the NMR data.[16] This highlights the

difficulties involved in using docking protocols for low-affinity ligands in large and flat binding sites and emphasizes that virtual screening of fragments in the ligand/lectin complex may be the best design tool for the optimization of lectin antagonists.

The results obtained with DC-SIGN and langerin suggested that the structure of **1** could represent a general template for a library of lectin antagonists containing one natural monosaccharide as the lectin-targeting element and a tuning unit, which could provide affinity and selectivity by establishing interactions in the proximity of the primary binding site. This hypothesis was explored by synthesizing a library of mannose- and fucose-based glycomimetics which was screened in a microarray format, in collaboration with the group of Niels Reichardt at CIC biomaGUNE, against a set of C-type lectin receptors (CLRs) that included DC-SIGN, DC-SIGNR, langerin, and dectin-2.[25] The streamlined synthesis of the library, which included the bis(p-nitrophenylester) **29** (Scheme 2) as the key glycosylation acceptor intermediate, allowed to obtain over 50 different molecules which were immobilized on the screening chip using the hetero-bi-functional spacer **35** (Scheme 3).

The lectins screened responded differently to the amide substituents of the mimics, generating distinctive binding profiles. Langerin was bound weakly by most of the molecules examined; DC-SIGN and DC-SIGNR showed similar recognition profiles, tolerant of secondary amide substituents on the pseudo-mannobioside. A set of mannosylated tertiary amides selectively recognized dectin-2 over DC-SIGN and some of the fucosylated structures displayed a similar selectivity. Thus this screening campaign provided the first discovery of glycomimetic ligands for dectin-2 and gave useful indications for the design and optimization of dectin-2 selective antagonists.



Scheme 3. Covalent immobilization of the ligands using the hetero bifunctional spacer **35**. Strain Promoted Azide-Alkyne Cycloaddition (SPAAC) was used to conjugate the ligands to the linker. The amino-terminated tether allowed to immobilize the ligand on a glass slide functionalized with *N*-hydroxy succinimide (NHS).

A totally different approach was adopted for the synthesis of glycomimetic ligands for the cholera toxin (CT). CT is a soluble, pentavalent lectin secreted by *V. cholerae* that recognizes the GM1 ganglioside on host cells membrane in the intestinal tract. The binding determinants of GM1 are a galactose (Gal) and a sialic acid (NeuAc) residue, pre-organized for optimal binding by the conformation of the ganglioside.[1] Functional mimics of GM1 may be designed by tethering the two pharmacophoric sugar fragments, trying to preserve the relative orientation adopted in the natural oligosaccharide. Several trials at computer-aided design proved unsatisfactory: CT binding site is large and shallow in the area that should accomodate the linker, and does not provide enough of a constraint. Thus, we adopted a modular synthesis of bidentate ligands of general formula **36** (Fig. 11), which could be assembled by clicking sialyl azide to appropriate linker-armed *C*-or *N*-galactosides. As opposed to our previous GM1 mimics, based on ligand structure modeling,[1, 26] these molecules are not substrates for glycosyl hydrolases and, in principle, are well suited for development into low-cost prophylactic drugs against cholera.

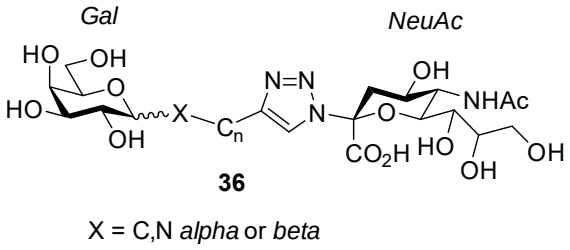


Figure 11. General structure of the bidentate ligands synthesized as cholera toxin antagonists.

Library members were ranked for CT affinity using weak affinity chromatography. Enhancements up to one or two orders of magnitude over the affinity of the individual sugar residues could be achieved.[27] This work proved that the connection of pharmacophoric sugar elements with non-hydrolizable linkers is a viable strategy for glycomimetics design. The X-ray structure of some of these ligands was obtained in complex with CT and a related bacterial enterotoxin. Both proteins are pentamers, with five identical binding sites and bind GM1 with similar affinity and binding mode. Pleasingly, the Gal and NeuAc residues of the library ligands were found to occupy their expected position in the lectin carbohydrate recognition domain (the complex of ligand **36a** is shown in Fig. 12a) while the substantially flexible linker is folded in a conformation that allows the pharmacophoric sugars to establish the canonical interactions seen with GM1. However, in two (out of five) of the binding sites the sialic acid moiety is stretched out and extends into the NeuAc binding site of another pentamer, thus linking two adjacent toxins in the crystal (Fig. 12b). This effect was confirmed to occur also in solution and provides an indication that protein aggregation can be produced even by relatively small ligands.[28] A similar effect was previously observed for the pseudo-trimannoside **2**, which promotes clustering of DC-SIGN tetramers in solution.[10b] For DC-SIGN, a cell membrane receptor, this effect most likely does not have a biological relevance and it is rather an artifact of the assay format. However, for CT and other soluble lectins, protein aggregation can -and should- be exploited as a relevant mechanism of inhibition.

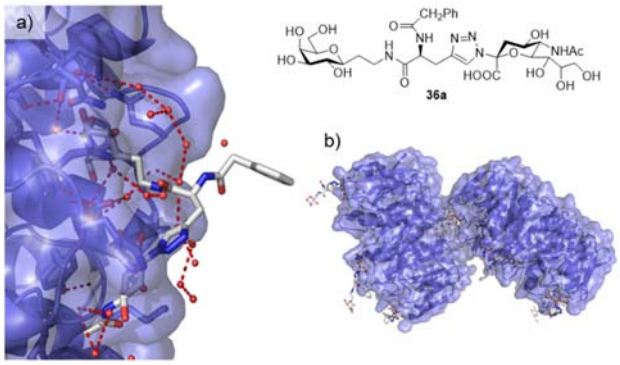


Figure 12. X-ray structure of bidentate ligand **36a** with a CT-like bacterial enterotoxin. a) The pharmacophoric residues Gal and NeuAc interact with the toxin as in the natural oligosaccharide ligand. b). Ligand **36a** promotes protein aggregation: two pentamers are shown.

4. Synthetic methodologies for the fast and stereoselective synthesis of unnatural glycoconjugates

4.1. Stereoselective synthesis of glycosylamides

Unnatural glycoconjugates are a large and diverse class of compounds that have been used as glycomimetics. Among them, we became particularly interested in α-linked glycosyl amides because natural glycopeptides and *N*-linked glycoconjugates are almost invariably β-linked and thus it is likely that the unnatural, α-linked isomers may go unnoticed by hydrolytic enzymes. Additionally, we were able to prove that α-glycosyl amides retain the normal pyranose conformation of the monosaccharide[29] and therefore represent true structural mimics. Indeed, α-fucosyl amides were efficient mimics of α-fucosides for DC-SIGN recognition[30] and targeting.[31]

Effective ways of synthesizing α-glycosylamides are missing: for the fucose derivatives mentioned above we relied on DeShong's method, which uses 2-acetoxy glycopyranosyl azides and Ph_3P, followed by an appropriate acylating agent.[32] With the same methodology we obtained a practical synthesis of α-linked FMOC-glycosyl-asparagine, both in the *gluco* and *galacto* series (*e.g.* **37**, Fig. 13), which we used for solid phase synthesis of unnatural glycopeptides.[33] Two model α-N-linked galactosyl-peptides (**38** and **39** in Fig. 13) were found to interact with Gal-

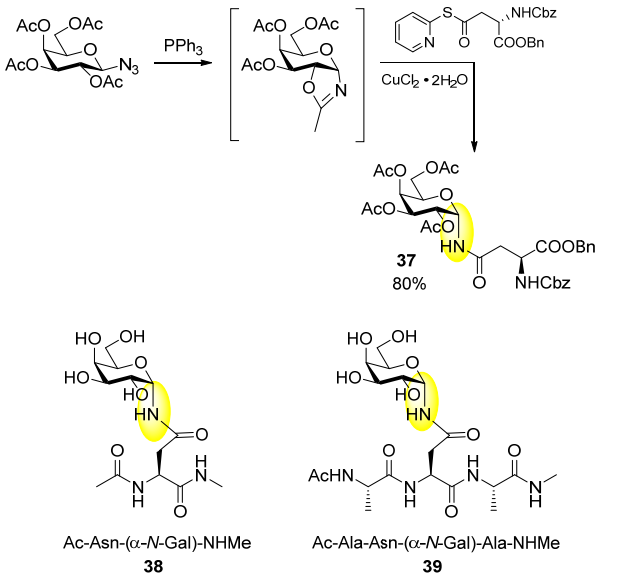


Figure 13. Synthesis of α-*N*-linked galactosylasparagine **37** and α-N-linked galactosyl-peptides **38** and **39**.

binding plant lectins (Viscum album agglutinin, VAA, and Erythrina cristagalli agglutinin, ECA, from coral tree)[34] establishing α-substituted neo-glycopeptides as potential tools in the design of glycomimetic structures.

As an alternative synthetic approach, we developed a traceless Staudinger ligation of anomeric glycosyl azides with functionalized phosphines **40** (Fig. 14),[29] which turned out to be particularly effective for the stereoselective synthesis of both α- and β-glycofuranosyl amides.[35] Indeed, the stereocontrol of furanoses does not depend on the anomeric configuration of the starting azide, but is determined by the C2 configuration and by the protection/deprotection state of the substrates. As exemplified for galactofuranose (Gal*f*) in Fig. 14, unprotected furanosyl azides of either anomeric configuration yield *1,2-cis* anomeric amides, while the corresponding *O*-acetyl derivatives afford the *1,2-trans* isomer (Fig. 14).

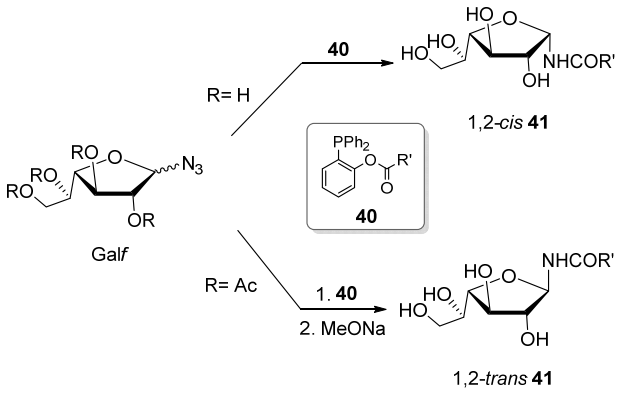
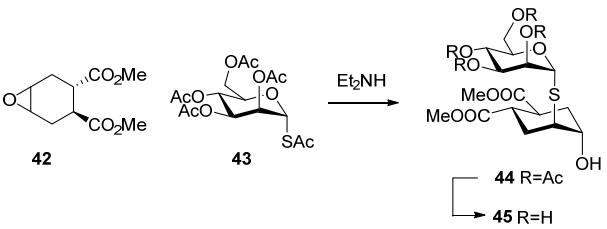


Figure 14. Reaction of galactofuranosyl azides with the functionalized phosphine **40**. Independent on the anomeric configuration of the azide, unprotected sugars afford *1,2-cis* anomeric amides and OAc derivatives the *1,2-trans* isomer.

Some of these compounds were tested in minimal inhibition concentration (MIC) assays against a panel of bacteria. One compound, the 1,2-*cis*-galactofuranosyl amide 1,2-*cis* **41** R'=-$(CH_2)_{14}CH_3$ displayed a modest activity against *Mycobacterium bovis* with a MIC value of 64 μg/mL.[36]

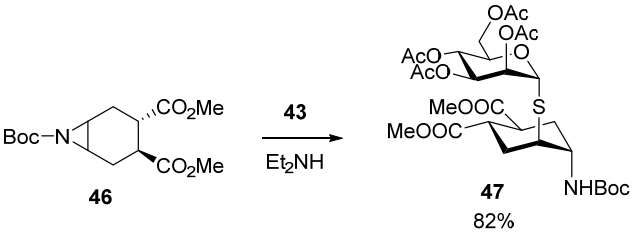
4.1. Thioglycosydes via one-pot reactions

Thiosugars are a second class of unnatural glycoconjugates that have attracted a great deal of attention because the *S*-glycosidic linkage is generally resistant towards both acid catalyzed and enzymatic hydrolysis. Additionally, C-S bonds are easier to synthesize than C-O bonds, thanks to the higher nucleophilicity of sulfur. We have recently exploited this feature in a streamlined approach to a thio analog of the pseudo-1,2-dimannoside structure **1**. The pseudo-thio-1,2-dimannoside **45** (Scheme 4) was obtained in a one-pot reaction between peracetylated mannosyl thiol (**43**) and epoxide **42**. In this reaction, after *in-situ* deacetylation of the anomeric thiol, a completely selective *trans*-diaxial opening of **42** affords **44** as a single isomer conformationally locked in the chair shown in Scheme 4.[37] Upon deacetylation, the pseudo-thio-1,2-dimannoside **45** was tested for DC-SIGN affinity and showed the same inhibitory activity of the parent pseudo-mannoside **1**. As opposed to **1**, **45** is stable even after 24 h treatment with high



Scheme 4. Synthesis of the pseudo-thio-1,2-dimannoside **44** by one pot opening reaction of epoxide **42**.

concentration of a mannosyl hydrolase (jack bean mannosidase).[38] Thus this method yields an agonist that is both more stable and easier to synthesize than the previous candidate. A similar transformation can be successfully achieved starting from *N*-acyl aziridines such as **46** (Scheme 5[38]), an approach that we are currently exploiting for the in-line synthesis of pseudo-glycopeptides.[39]



Scheme 5. One pot opening reaction of *N*-Boc aziridine **46**.

5. Conclusions

After a decade of research in the field, which lessons can we extract about designing glycomimetics as lectin antagonists? We can tentatively begin to distill some principles from the examples described above. First, some oligosaccharides have three-dimensional structures that are relatively well pre-organized (*i.e.*

not entirely flexible) and use a single monosaccharide as the primary binding element. Using this monosaccharide as an anchor and including it into a molecule that somehow mimics the native oligosaccharide shape has been successful. In doing so, there appears to be a distinctive advantage when removing hydroxyl groups that are not in direct contact with the protein.[2] The monosaccharide anchor can be modified or decorated with additional elements that can develop secondary interactions in the lectin binding site. Lipophilic elements can be used to this purpose in random screenings, taking advantage of the large and exposed nature of most lectin sites. However, the same approach was successfully adopted also in the context of virtual fragment screening, where the target becomes the monosaccharide complex of the lectin and additional binding regions are explored in the vicinity of the sugar. We expect this design tactic to be transferrable and to produce novel results for many lectin substrates.

Using modular synthetic plans that make use of "click" conjugation technologies and that include rigid elements we were also able to design polyvalent structures to fit size and shape of DC-SIGN, a tetravalent C-type lectin. The flexibility of this approach should allow to adapt the strategy to different targets, and to exploit all possible affinity amplification mechanisms against multivalent lectins.

The subject of polyvalency remains relevant because it is still rather difficult to get high affinity ligands with small, monovalent molecules. Similarly, the selectivity problem remains unsolved and, in fact, largely unexplored. In our experience, serendipity has been so far the most important factor, but both differential design (comparative analysis of lectins' site structures) and library screenings have also been successful.

A recurring observation in the course of these studies is that ligand-induced clustering of lectins occurs more often than previously imagined, even with relatively small molecules such as the pseudo-trisaccharide **2** or the CT ligand **36a**. Even more so, this effect is probably commonplace for polyvalent constructs, when the lectin is not immobilized. Although protein clustering is not usually taken into account as a design principle, it could and probably should be exploited for soluble lectins. On the contrary, when addressing membrane receptors, this effect can be observed *in vitro*, depending on the format of the assay, but, in the biological context, would depend on membrane clustering of the receptor, an effect that is exploited by living systems, but more complex to predict and trigger at will with artificial probes.

Finally, a rewarding take home message for a synthetic organic chemist is that facile access to the synthesis of the monovalent ligands is still of paramount importance for the progress of knowledge in this field. More generally, the improvement and diversification of synthetic methodologies is still much-needed to fully harness the power of chemistry in chemical biology projects.

Acknowledgments

The contribution of all the students and coworkers that cooperated to this decade of research and are mentioned in the references is gratefully acknowledged. The long-standing collaboration with Franck Fieschi and Javier Rojo in the DC-SIGN project has been a pleasure over the years and has made possible most of the results described in this paper. Financial support from the European Union is gratefully acknowledged.

Keywords: Carbohydrates • Glycomimetics • DC-SIGN • Cholera Toxin • multivalency

[1] A. Bernardi, P. Cheshev, *Chem Eur J* **2008**, *14*, 7434-7441.

[2] J. Cramer, C. P. Sager, B. Ernst, *J Med Chem* **2019**, *62*, 8915-8930.

[3] a) A. Bernardi, J. Jimenez-Barbero, A. Casnati, C. De Castro, T. Darbre, F. Fieschi, J. Finne, H. Funken, K. E. Jaeger, M. Lahmann, T. K. Lindhorst, M. Marradi, P. Messner, A. Molinaro, P. V. Murphy, C. Nativi, S. Oscarson, S. Penades, F. Peri, R. J. Pieters, O. Renaudet, J. L. Reymond, B. Richichi, J. Rojo, F. Sansone, C. Schaffer, W. B. Turnbull, T. Velasco-Torrijos, S. Vidal, S. Vincent, T. Wennekes, H. Zuilhof, A. Imberty, *Chem Soc Rev* **2013**, *42*, 4709-4727; b) S. Cecioni, A. Imberty, S. Vidal, *Chem. Rev.* **2015**, *115*, 525-561.

[4] B. G. Keller, C. Rademacher, *Curr Opin Struct Biol* **2019**, *62*, 31-38.

[5] R. Hevey, *Pharmaceuticals (Basel)* **2019**, *12*.

[6] A. Tamburrini, C. Colombo, A. Bernardi, *Med Res Rev* **2019**.

[7] a) T. B. H. Geijtenbeek, R. Torensma, S. J. van Vliet, G. C. F. van Duijnhoven, G. J. Adema, Y. van Kooyk, C. G. Figdor, *Cell* **2000**, *100*, 575-585; b) T. B. H. Geijtenbeek, D. S. Kwon, R. Torensma, S. J. van Vliet, G. C. F. van Duijnhoven, J. Middel, I. L. M. H. A. Cornelissen, H. S. L. M. Nottet, V. N. KewalRamani, D. R. Littman, C. G. Figdor, Y. van Kooyk, *Cell* **2000**, *100*, 587-597.

[8] C. P. Alvarez, F. Lasala, J. Carrillo, O. Muniz, A. L. Corbi, R. Delgado, *J. Virol.* **2002**, *76*, 6841-6844.

[9] a) S. Sattin, A. Daghetti, M. Thepaut, A. Berzi, M. Sanchez-Navarro, G. Tabarani, J. Rojo, F. Fieschi, M. Clerici, A. Bernardi, *ACS Chem Biol* **2010**, *5*, 301-312; b) J. Luczkowiak, S. Sattin, I. Sutkeviciute, J. J. Reina, M. Sanchez-Navarro, M. Thepaut, L. Martinez-Prats, A. Daghetti, F. Fieschi, R. Delgado, A. Bernardi, J. Rojo, *Bioconjug Chem* **2011**, *22*, 1354-1365.

[10] a) M. Thépaut, C. Guzzi, I. Sutkeviciute, S. Sattin, R. Ribeiro-Viana, N. Varga, E. Chabrol, J. Rojo, A. Bernardi, J. Angulo, P. M. Nieto, F. Fieschi, *J. Am. Chem. Soc.* **2013**, *135*, 2518-2529; b) I. Sutkeviciute, M. Thepaut, S. Sattin, A. Berzi, J. McGeagh, S. Grudinin, J. Weiser, A. Le Roy, J. J. Reina, J. Rojo, M. Clerici, A. Bernardi, C. Ebel, F. Fieschi, *ACS Chem Biol* **2014**, *9*, 1377-1385.

[11] J. J. Reina, S. Sattin, D. Invernizzi, S. Mari, L. Martinez-Prats, G. Tabarani, F. Fieschi, R. Delgado, P. M. Nieto, J. Rojo, A. Bernardi, *ChemMedChem* **2007**, *2*, 1030-1036.

[12] V. Bordoni, V. Porkolab, S. Sattin, M. Thépaut, I. Frau, L. Favero, P. Crotti, A. Bernardi, F. Fieschi, V. Di Bussolo, *RSC Adv.* **2016**, *6*, 89578-89584.

[13] a) H. Feinberg, R. Castelli, K. Drickamer, P. H. Seeberger, W. I. Weis, *J Biol Chem* **2007**, *282*, 4202-4209; b) J. D. Martinez, P. Valverde, S. Delgado, C. Romano, B. Linclau, N. C. Reichardt, S. Oscarson, A. Arda, J. Jimenez-Barbero, F. J. Canada, *Molecules* **2019**, *24*.

[14] N. Varga, I. Sutkeviciute, C. Guzzi, J. McGeagh, I. Petit-Haertlein, S. Gugliotta, J. Weiser, J. Angulo, F. Fieschi, A. Bernardi, *Chem Eur J* **2013**, *19*, 4786-4797.

[15] V. Porkolab, E. Chabrol, N. Varga, S. Ordanini, I. Sutkevičiūte, M. Thepaut, M. J. Garcia-Jimenez, E. Girard, P. M. Nieto, A. Bernardi, F. Fieschi, *ACS Chem Biol* **2018**, *13*, 600-608.

[16] L. Medve, S. Achilli, J. Guzman-Caldentey, M. Thepaut, L. Senaldi, A. Le Roy, S. Sattin, C. Ebel, C. Vives, S. Martin-Santamaria, A. Bernardi, F. Fieschi, *Chem Eur J* **2019**, *25*, 14659-14668.

[17] A. Berzi, J. J. Reina, R. Ottria, I. Sutkeviciute, P. Antonazzo, M. Sanchez-Navarro, E. Chabrol, M. Biasin, D. Trabattoni, I. Cetin, J. Rojo, F. Fieschi, A. Bernardi, M. Clerici, *Aids* **2012**, *26*, 127-137.

[18] G. Goti, A. Palmioli, M. Stravalaci, S. Sattin, M. G. De Simoni, M. Gobbi, A. Bernardi, *Chem Eur J* **2016**, *22*, 3686-3691.

[19] F. Orsini, P. Villa, S. Parrella, R. Zangari, E. R. Zanier, R. Gesuete, M. Stravalaci, S. Fumagalli, R. Ottria, J. J. Reina, A. Paladini, E. Micotti, R. Ribeiro-Viana, J. Rojo, V. I. Pavlov, G. L. Stahl, A. Bernardi, M. Gobbi, M. G. De Simoni, *Circulation* **2012**, *126*, 1484-1494.

[20] N. Varga, I. Sutkeviciute, R. Ribeiro-Viana, A. Berzi, R. Ramdasi, A. Daghetti, G. Vettoretti, A. Amara, M. Clerici, J. Rojo, F. Fieschi, A. Bernardi, *Biomaterials* **2014**, *35*, 4175-4184.

[21] A. Berzi, N. Varga, S. Sattin, P. Antonazzo, M. Biasin, I. Cetin, D. Trabattoni, A. Bernardi, M. Clerici, *Viruses* **2014**, *6*, 391-403.

[22] S. Ordanini, N. Varga, V. Porkolab, M. Thepaut, L. Belvisi, A. Bertaglia, A. Palmioli, A. Berzi, D. Trabattoni, M. Clerici, F. Fieschi, A. Bernardi, *Chem Commun* **2015**, *51*, 3816-3819.
[23] A. Berzi, S. Ordanini, B. Joosten, D. Trabattoni, A. Cambi, A. Bernardi, M. Clerici, *Sci. Rep.* **2016**, *6*, 35373.
[24] N. Obermajer, S. Sattin, C. Colombo, M. Bruno, U. Svajger, M. Anderluh, A. Bernardi, *Mol Divers* **2011**, *15*, 347-360.
[25] L. Medve, S. Achilli, S. Serna, F. Zuccotto, N. Varga, M. Thépaut, M. Civera, C. Vivès, F. Fieschi, N. Reichardt, A. Bernardi, *Chem Eur J* **2018**, *24*, 14448-14460.
[26] A. Bernardi, A. Checchia, P. Brocca, S. Sonnino, F. Zuccotto, *J. Am. Chem. Soc.* **1999**, *121*, 2032-2036.
[27] P. Cheshev, L. Morelli, M. Marchesi, C. Podlipnik, M. Bergstrom, A. Bernardi, *Chem Eur J* **2010**, *16*, 1951-1967.
[28] J. E. Heggelund, A. Mackenzie, T. Martinsen, J. B. Heim, P. Cheshev, A. Bernardi, U. Krengel, *Sci. Rep.* **2017**, *7*, 2326.
[29] A. Bianchi, A. Bernardi, *J.Org. Chem.* **2006**, *71*, 4565-4577.
[30] a) M. Andreini, D. Doknic, I. Sutkeviciute, J. J. Reina, J. Duan, E. Chabrol, M. Thépaut, E. Moroni, F. Doro, L. Belvisi, J. Weiser, J. Rojo, F. Fieschi, A. Bernardi, *Org. Biomol. Chem.* **2011**, *9*, 5778-5786; b) C. Guzzi, J. Angulo, F. Doro, J. J. Reina, M. Thepaut, F. Fieschi, A. Bernardi, J. Rojo, P. M. Nieto, *Org Biomol Chem* **2011**, *9*, 7705-7712.
[31] D. Arosio, F. Chiodo, J. J. Reina, M. Marelli, S. Penades, Y. van Kooyk, J. J. Garcia-Vallejo, A. Bernardi, *Bioconjug Chem* **2014**, *25*, 2244-2251.
[32] F. Damkaci, P. DeShong, *J. Am. Chem. Soc.* **2003**, *125*, 4408-4409.
[33] C. Colombo, A. Bernardi, *Eur. J. Org. Chem.* **2011**, *2011*, 3911-3919.
[34] F. Marcelo, F. J. Canada, S. Andre, C. Colombo, F. Doro, H. J. Gabius, A. Bernardi, J. Jimenez-Barbero, *Org Biomol Chem* **2012**, *10*, 5916-5923.
[35] F. Nisic, G. Speciale, A. Bernardi, *Chem Eur J* **2012**, *18*, 6895-6906.
[36] F. Nisic, PhD thesis, Univeristà degli Studi di Milano **2010**.
[37] A. Tamburrini, S. Achilli, F. Vasile, S. Sattin, C. Vives, C. Colombo, F. Fieschi, A. Bernardi, *Bioorg. Med. Chem.* **2017**, *25*, 5142-5147.
[38] A. Tamburrini, PhD thesis, Università degli Studi di Milano **2019**.
[39] N. Hribernik, A. Tamburrini, E. Falletta, A. Bernardi, *manuscript in preparation*, **2020**.