<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>The third dimension of reading the sugar code by lectins: design of glycoclusters with cyclic scaffolds as tools with the aim to define correlations between spatial presentation and activity.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Murphy, Paul V.</td>
</tr>
<tr>
<td><strong>Publication Date</strong></td>
<td>2013-04-04</td>
</tr>
<tr>
<td><strong>Publication Information</strong></td>
<td>Murphy PV, André S, Gabius HJ (2013) &quot;The third dimension of reading the sugar code by lectins: design of glycoclusters with cyclic scaffolds as tools with the aim to define correlations between spatial presentation and activity&quot;. Molecules (Basel, Switzerland), 18 (4):4026-4053.</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>MDPI</td>
</tr>
<tr>
<td><strong>Link to publisher's version</strong></td>
<td><a href="http://dx.doi.org/10.3390/molecules18044026">http://dx.doi.org/10.3390/molecules18044026</a></td>
</tr>
<tr>
<td><strong>Item record</strong></td>
<td><a href="http://hdl.handle.net/10379/4326">http://hdl.handle.net/10379/4326</a></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
<td><a href="http://dx.doi.org/10.3390/molecules18044026">http://dx.doi.org/10.3390/molecules18044026</a></td>
</tr>
</tbody>
</table>
Review

The Third Dimension of Reading the Sugar Code by Lectins: Design of Glycoclusters with Cyclic Scaffolds as Tools with the Aim to Define Correlations between Spatial Presentation and Activity

Paul V. Murphy 1*, Sabine André 2 and Hans-Joachim Gabius 2

1 School of Chemistry, National University of Ireland Galway, University Road, Galway, Ireland
2 Faculty of Veterinary Medicine, Institute of Physiological Chemistry, Ludwig-Maximilians-University, Veterinärstr, 13, 80539 Munich, Germany; E-Mails: S.Andre@tiph.vetmed.uni-muenchen.de (S.A.); gabius@tiph.vetmed.uni-muenchen.de, gabius@lectins.de (H.-J.G.)

* Author to whom correspondence should be addressed; E-Mail: paul.v.murphy@nuigalway.ie; Tel.: +353-91-492460; Fax: +353-91-495576.

Received: 11 March 2013; in revised form: 22 March 2013 / Accepted: 1 April 2013 / Published: 4 April 2013

Abstract: Coding of biological information is not confined to nucleic acids and proteins. Endowed with the highest level of structural versatility among biomolecules, the glycan chains of cellular glycoconjuga tes are well-suited to generate molecular messages/signals in a minimum of space. The sequence and shape of oligosaccharides as well as spatial aspects of multivalent presentation are assumed to underlie the natural specificity/selectivity that cellular glycans have for endogenous lectins. In order to eventually unravel structure-activity profiles cyclic scaffolds have been used as platforms to produce glycoclusters and afford valuable tools. Using adhesion/growth-regulatory galectins and the pan-galectin ligand lactose as a model, emerging insights into the potential of cyclodextrins, cyclic peptides, calixarenes and glycophanes for this purpose are presented herein. The systematic testing of lectin panels with spatially defined ligand presentations can be considered as a biomimetic means to help clarify the mechanisms, which lead to the exquisite accuracy at which endogenous lectins select their physiological counterreceptors from the complexity of the cellular glycome.

Keywords: calixarene; cyclodextrin; galectin; glycome; glycopeptide; glycophane
1. The Concept of the Sugar Code

The ubiquitous presence of post-translational modifications, especially phosphorylation, teaches the lesson that it is more than its sequence that determines a protein’s activity profile. A covalent conjugation is able to convey new properties to the protein scaffold [1,2]. As a consequence, the range of protein functionality is likely broadened when processed. Sharing a frequent occurrence with phosphorylation, glycosylation, mostly on cell surface and extracellular proteins (from a monosaccharide to long and even highly branched chains), is known to be an integral part of this system of protein substitutions [3–10]. Clinically, the emerging insights into etiology of aberrations caused by congenital diseases of glycosylation and gain-of-glycosylation mutations are story-telling incidences to illustrate that glycans are not merely inert or readily interchangeable appendices for the protein [11–13]. Genetic engineering of animal models, too, underscores the essential nature of glycosylation. Serious defects up to embryonic lethality or neonatal death are caused after ablation of N-glycan synthesis (for recent review, please see [14]). In fact, these noted associations are already warranting the monitoring of glycan structure in detail, and this work, applying a combination of technically sophisticated methods, has revealed a level of structural complexity not reached by any other type of protein modification [15,16].

The intricacies of the underlying enzymatic machinery, with estimates that at least 1% of the genomic coding capacity is reserved for these enzymes [17], together with enormous versatility of regulation to dynamically shift the glycome profile by remodeling, are the means to let glycosylation become a highly refined process [5,6,18,19]. To give instructive examples, the introduction of certain sugars into glycan chains depends not just on one or a few enzymes. Instead, the fucosylation of mammalian glycans, a characteristic of branch ends and the N-glycan core, can be accomplished by thirteen transferases [20,21]. Twenty members belong to the family of sialyltransferases, which are also dedicated to generate elaborate glycan termini [22,23]. Next, the physiological fine-tuning depending on the availability of substrates comes into play to give the glycome its shape; the physiological impact of several types of glycoenzymes, for instance, was crucial to explain the role of glycan reprogramming on the pathway of how a tumor suppressor drives malignant cells into anoikis [24]. Nucleotide sugar transporters, too, deserve to be mentioned, harming their activity being another cause of diseases [25]. As a consequence, the glycophenotype, resulting from the interplay of all these components, can be considered as being as characteristic as a fingerprint for a cell, and carbohydrates can be viewed as a chemical platform to encode biological information.

Toward this end, that is to serve as the third alphabet of life alongside nucleotides and amino acids, sugars have exceptional chemical properties. They actually enable the sugar ‘letters’ to reach an unsurpassed level of coding capacity. Although these features are basic in nature, they deserve to be mentioned in this context to raise awareness for how well sugars are adapted to building code words. Structural variability is possible not just by changing the sequence, as in proteins or nucleic acids. Beyond that, the following parameters can be independently altered when turning units (letters) into oligomers (words): the anomeric status, the linkage position between sugar units, the ring size, branching in oligosaccharides and introduction of site-specific substituents such as acetylation, phosphorylation or sulfation [26,27]. When the synthesis of the glycans is finally completed, their presentation by cell surface glycoproteins (and also by the class of glycolipids [28]) brings these
determinants into a strategic position for recognition events. In other words, owing to the generation of bioepitopes with a maximum of information in a minimum of space at readily accessible sites functional implications immediately arise.

In principle, the presence of the glycans can affect protein properties (acting in \textit{cis}) or the sugar epitope constitutes an entirely new site for recognition by respective receptors (acting in \textit{trans}). Intriguingly, a single sugar unit can act like a switch for the glycan’s three-dimensional structure, hereby building a bridge from the descriptive nature of glycophenotyping to what the respective sugar addition (e.g., core substitutions in \textit{N}-glycans) then triggers [29–32]. One effect is to cover surface regions of the protein so that the glycan’s shape modulates the potential for contacts to other proteins, in oligomerization and other types of inter-protein contacts. Also, stability and trafficking can depend on sugar signals, a wide field to be explored diligently [5,6,9,12,33]. Having herewith introduced the concept of the sugar code, \textit{i.e.} biological information storage by glycans and transfer into effects via different routes, we can proceed to looking at the mentioned translators of the sugar code, \textit{i.e.}, sugar receptors (lectins).

2. Glycans as Bioactive Ligands for Lectins

The specificity of carbohydrate recognition originates from the structural complementarity between the sugar and a protein. Hydrogen-bond networks (without or with involvement of water molecules) and C-H/\pi-interactions between a patch of suitably positioned C-H bonds (e.g., in \textit{d}-galactose) and a Trp residue cooperate along the way to give a snug fit [34–36]. In special cases, the presence of an anionic sugar part (e.g., in sialylated glycans) accounts for ionic bonds. Sensing the distinct mode of presentation of axial/equatorial hydroxyl groups of common “letters” of the sugar alphabet can also involve Ca$^{2+}$-ion(s) and coordination bonds [37]. These recognition modes combined, not onlyoligosaccharides but monosaccharides such as galactose vs mannose can readily be distinguished by lectins, as technically simple assays such as inhibition of lectin-mediated haemagglutination attest. To get a feeling for the extent of the physiological range of interactions via glycan recognition it is instructive to delineate the number of different protein folds with the capacity to bind sugars. A small number would indicate this type of recognition to be more a peculiarity than a frequently encountered mechanism. That would mean that the immense potential of the sugar code outlined above would not really be realized.

As the compilation in Table 1 documents, up to 14 different folds have proven capacity for glycan binding. In each case, examples for respective animal/human lectins are given together with information on glycan ligands. The proteins in the different families cover a wide range of activities, on the level of glycan routing and transport, cell adhesion and growth regulation as well as host defense, to give a few examples (for further information, please see [35,36]). Of note, the binding is remarkably specific to the cellular glycoconjugate, which is the target to ensure the correct flow of information. Despite a large number of theoretically possible contact sites, for example \textit{\beta}-galactosides, the lectins are indeed capable to home in on particular glycoproteins/glycolipids or glycosaminoglycan sequences, posing the challenge to identify the underlying molecular reasons. Fittingly, physiologic regulation works on both sides of the recognition system for optimal responsiveness, \textit{i.e.}, the presentation of the lectin and of its glycan counterreceptor(s), seen in distinct contexts of cell adhesion
or growth regulation [38–41]. In addition to making a lectin-reactive epitope available by dynamic remodeling or neosynthesis, topological factors appear to play a major role to guide the selection process. On the side of the glycans, six levels of affinity regulation have been identified, which underlie detectable preferences [36]. Spatial vicinity of ligands, as facilitated within microdomains so that perturbation of their integrity harms lectin reactivity [42], is an efficient means to build preferred contact regions, in terms of affinity and the nature of the counterreceptor. Depending on the cell type a particular glycoprotein (such as the $\alpha_5\beta_1$-integrin) or a ganglisside (GM1) can be the main binding partner for the same endogenous lectin, the association then setting in motion a post-binding signaling cascade, e.g., toward anergy, anoikis or growth arrest [39,40,43]. Of course, the in situ constellations operative in turning structure (at each of the six levels mentioned above) into distinct effects set attractive role models for the synthetic design of glycoclusters.

### Table 1. Overview of folds with capacity to bind sugars and of lectin classes.

<table>
<thead>
<tr>
<th>Type of fold</th>
<th>Example for lectin</th>
<th>Example for ligand</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$-sandwich (jelly-roll)</td>
<td>(a) galectins</td>
<td>$\beta$-galactosides</td>
</tr>
<tr>
<td></td>
<td>(b) calnexin, calreticulin</td>
<td>Glc, Man, GlcNAc</td>
</tr>
<tr>
<td></td>
<td>(c) ERGIC-53, VIP36, VIPL</td>
<td>Man, GlcNAc</td>
</tr>
<tr>
<td></td>
<td>(d) CRD* of Fbs1 in SCF E3 ubiquitin ligase and peptide-N-glycanase</td>
<td>Mann5GlcNAc2; mannanpentaose</td>
</tr>
<tr>
<td></td>
<td>(e) pentraxins</td>
<td>glycosaminoglycans, MOβDG, 3-sulfated Gal, GalNAc and GlcA, Man-6-phosphate</td>
</tr>
<tr>
<td></td>
<td>(f) G-domains of the LNS family (laminin, agrin)</td>
<td>heparin</td>
</tr>
<tr>
<td>C-type</td>
<td>asialoglycoprotein receptor, collectins, selectins</td>
<td>Fuc, Gal, GalNAc, Man, heparin tetrasaccharide</td>
</tr>
<tr>
<td>I-type (Ig fold)</td>
<td>N-CAM, TIM-3, siglec</td>
<td>Man5GlcNAc2, HNK-1 epitope, $\alpha_2,3/6$-sialylated glycans</td>
</tr>
<tr>
<td>P-type</td>
<td>mannose-6-phosphate receptors (MR) and proteins with MR homology domain (erlectin, OS-9)</td>
<td>Man-6-phosphate, Man5,8GlcNAc</td>
</tr>
<tr>
<td></td>
<td>(a) fibroblast growth factors</td>
<td>heparan sulfate</td>
</tr>
<tr>
<td></td>
<td>(b) cysteine-rich domain of C-type macrophage mannose receptor</td>
<td>GalNAc-4-sulfate in LacdiNAc</td>
</tr>
<tr>
<td>$\beta$-trefoil</td>
<td>(c) lectin domain in GalNAc-Ts$^s$ involved in mucin-type O-glycosylation</td>
<td>GalNAc</td>
</tr>
<tr>
<td></td>
<td>(d) hemolytic lectin CEL-III of sea cucumber and lectin EW29 of earthworm</td>
<td>Gal</td>
</tr>
<tr>
<td>$\beta$-propeller</td>
<td>(a) 4-bladed: tachylectin-3</td>
<td>S-type lipopolysaccharide</td>
</tr>
<tr>
<td></td>
<td>(b) 5-bladed: tachylectin-2</td>
<td>GlcNAc/GalNAc</td>
</tr>
<tr>
<td></td>
<td>(c) 6-bladed: tachylectin-1</td>
<td>KDO</td>
</tr>
<tr>
<td>$\beta$-propeller</td>
<td>(a) 4-bladed: tachylectin-3</td>
<td>S-type lipopolysaccharide</td>
</tr>
<tr>
<td></td>
<td>(b) 5-bladed: tachylectin-2</td>
<td>GlcNAc/GalNAc</td>
</tr>
<tr>
<td></td>
<td>(c) 6-bladed: tachylectin-1</td>
<td>KDO</td>
</tr>
</tbody>
</table>
Table 1. Cont.

<table>
<thead>
<tr>
<th>Type of fold</th>
<th>Example for lectin</th>
<th>Example for ligand</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-prism I</td>
<td>secretory proteins zg16p/b</td>
<td>not defined</td>
</tr>
<tr>
<td>β-prism II</td>
<td>pufferfish (fugu) lectin</td>
<td>Man</td>
</tr>
<tr>
<td>β-barrel with jelly-roll topology</td>
<td>tachylectin-4, eel (<em>Anguilla anguilla</em>)</td>
<td>Fuc</td>
</tr>
<tr>
<td>(a) ficolins</td>
<td>agglutinin, X-epilectin</td>
<td>GlcNAc</td>
</tr>
<tr>
<td>(b) intelectins (mammalian, <em>Xenopus</em>)</td>
<td>(c) tachylectin-5</td>
<td>Galf, pentoses</td>
</tr>
<tr>
<td>(d) slug (<em>Limax flavus</em>) lectin</td>
<td>CD44, TSG-6, LYVE-1, aggregating</td>
<td>N-acetylated sugars</td>
</tr>
<tr>
<td>proteoglycans</td>
<td>tachycytin and spider (*Selenocosmia</td>
<td>Sialic acid</td>
</tr>
<tr>
<td></td>
<td>huwena*) neurotoxin; cobra venom</td>
<td></td>
</tr>
<tr>
<td>link module</td>
<td>YKL-40 (human cartilage)</td>
<td>GalNAc; heparin-derived disaccharide</td>
</tr>
<tr>
<td>hevein-like domain</td>
<td>glycoprotein-39; chitinase-like lectin</td>
<td></td>
</tr>
<tr>
<td>(β/α)8 barrel</td>
<td>factor H (complement regulator)</td>
<td>glycosaminoglycans, sialic acid</td>
</tr>
<tr>
<td>(glycoside hydrolase family 18) short</td>
<td></td>
<td></td>
</tr>
<tr>
<td>consensus repeat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(complement control protein module)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* carbohydrate recognition domain, *b* N-acetylgalactosaminyltransferases; adapted from [44], with permission.

The most telling example concerns the first mammalian lectin purified from rabbit liver, a hepatic receptor acting in clearance of glycoproteins from serum [45]. Testing its glycan reactivity, a geometrical increase of affinity was measured to arise from a numerical increase of valency in oligosaccharides when targeting this C-type lectin (please see also Table 1) [46]. Besides natural or synthetic N-glycans, cluster glycosides have been instrumental to trace the intriguing correlation of matching complementarity between ligand and receptor presentation [46–50]. These results led to the definition of the glycoside cluster effect, i.e., the affinity enhancement by multivalency over and beyond what is expected from the concentration increase [46,51]. In order to discern rules for the correlation between the topological aspects of ligand presentation and the lectin structure it is reasonable to focus on a certain family of lectins as test model system. For the scope of this review, we do so by dealing exclusively with adhesion/growth-regulatory galectins. Being deliberately placed in the top part of Table 1, these lectins share a β-sandwich fold and a sequence signature with a central Trp residue essential for the C-H/π-interaction with galactose noted above [36], a feature readily monitored using NMR and fluorescence spectroscopy [52–54].

3. Galectins: a Network of Bioeffectors

The common structural traits noted above concern the carbohydrate recognition domain (CRD) of galectins. Since the levels of affinity and specificity for cellular glycans, as well as functionality depend on more than monovalent binding, the active lectin is in general more than a single CRD. Examples of growth regulation by human galectin-1, given above [39–41,43], show that this lectin can
be expected to initiate signaling by cross-linking of counterreceptors, and, indeed, its structure is homodimeric [55]. Within the family of galectins, the relative spatial arrangement of CRDs divides these proteins into three groups: the homodimeric (proto-type) and the tandem-repeat-type family members, the latter with a linker peptide connecting two different CRDs, along with the chimera-type galectin-3 which harbors an N-terminal peptide with sites for Ser phosphorylation and collagen-like repeats enabling oligimerization in the presence of multivalent ligands [56–58] (Figure 1). Thus, the capacity of individual proteins to cross-link counterreceptors, as measured in precipitation analysis with a multivalent glycoprotein [59], and the stoichiometry of the complexes is expected to be different so that even functional competition can be predicted. Apparently, this is the case in blocking galectin-1’s growth-inhibitory activity on neuroblastoma and pancreatic cancer cells by the chimera-type galectin-3 [60,61]. Although different in spatial CRD display, the two lectins share specificity to the same glycoconjugate in these cells. Galectin-3 hereby precludes galectin-1 binding but fails to trigger the post-binding signaling leading to growth arrest in the tested cell systems. Because galectins physiologically form a complex network, with different proteins expressed for example in tumors [58,62–65], it is a pertinent issue to establish individual structure-activity profiles. Keeping the sugar structure constant by using the pan-galectin ligand lactose (or N-acetyllactosamine, LacNAc), glycoclusters are suited for this project line.

**Figure 1.** Schematic illustration of the three types of spatial CRD presentation in galectins, using the five chicken galectins as example. The ten Gly/Pro-rich repeats in CG-3 and the lengths of the linker peptide in CG-8 given in number of amino acids are indicated (from [66], with permission).

![Schematic illustration of the three types of spatial CRD presentation in galectins](image)

Given the different degrees of intra-group diversification in phylogensis, these experimental series can be run with the complex set of mammalian lectins, to connect the data to biomedical considerations, or, benefiting from organisms having a comparatively low number of galectin genes, with all proteins of such a restricted set of proteins. As shown in Figure 1, five galectins establish the entire panel of chicken galectins (CG), with representatives of each group included, i.e., three homodimeric, one chimera-type and one tandem-repeat-type protein expressed with two linker lengths
due to alternative splicing [58,67–72]. Regardless of the origin of the galectins the parameter measured in assays with the synthetic compounds is the inhibitory capacity of glyoclusters on the extent of galectin binding to a glycan-presenting matrix. Noting that galectins can be secreted and exert activity as lectins in auto- or paracrine manners on the level of the cell surface (e.g., in tumor growth regulation or in communication between effector and regulatory T cells [73]), the galectin is strictly kept in solution mimicking the physiological situation, while the binding partner is either a glycoprotein (e.g., asialofetuin) adsorbed to the surface of a well of a microtiter plate or a cell surface. The presence of an inhibitor will then reduce the read-out. Cell scanning is performed in cytofluorometric analysis and monitored in terms of percentage of positive cells and mean fluorescence intensity (Figure 2). With galectins as test proteins, four types of cyclic glyoclusters shown in Figure 3 have already been evaluated, with the review of the results herein starting with cyclodextrins.

Figure 2. Schematic illustration of the principle of the experimental read-out in cytofluorometric analysis of glyocluster activity. When a labeled lectin binds to cell surfaces, the signal (in percentage of positive cells and mean fluorescence intensity) describes the cells’ reactivity (black line). The presence of an inhibitor reduces staining (dotted line), shifting the binding profile into the direction of the background value (grey area).

4. Cyclodextrins, Cyclic Decapeptides and Calixarenes

Cyclodextrins are macrocycles of between six to eight α-D-glucose units produced by degradation of starch, which attract attention due to their high biocompatibility and solubility [74–76]. Persubstitution of the heptakis 6-deoxy-6-iodo-β-cyclodextrin core was performed with glycosides via their terminal sodium thiolate (for details, please see [77]). The chemical conjugation did not impair the reactivity of the sugar headgroup for lectins. Among the tested galectins, the relatively most sensitive protein was galectin-3 [78], monomeric in solution but capable to pentamerize in the presence of multivalent ligands [79].
Figure 3. Schematic illustration of the structures of four cyclic scaffolds tested for galectin reactivity, i.e., β-cyclodextrin, cyclic decapeptide, cone-type calixarene and glycophane.

The homodimeric galectins-1 and -7 were less responsive [78]. Such a grading had been noted before with a triiodobenzene-based trivalent cluster, to which 2-propynyllactosides had been conjugated [80]. The relative level of inhibition for each protein also depended on the nature of the matrix, here this refers to the type of glycoprotein used and its degree of N-glycan branching [78], a result later confirmed when testing other types of glycoclusters [81]. Equally noteworthy, the efficiency of the lactosylated cyclodextrin in interfering with lectin binding was enhanced when tested on galectin-1-presenting cells [78]. Whether the lectin is free in solution or associated to a cell affects the read-out,
precluding generalizations. This behaviour with single macrocycles was also observed with a pseudopolyrotaxane-based glycocluster, which had a beads-on-a-string arrangement with lactosylated cyclodextrin being “pearls” on a polyviologen “string” [82,83].

Similarly building on a natural scaffold, cyclic peptides have been decorated with sugar derivatives to generate neoglycopeptide clusters [84,85]. The decapeptides were tailored to have four attachment points for the derivatives via the side chain of the lysine moieties. A clear grading in susceptibility was seen when moving from galectin-1 to galectins-3 and -4, both in the solid-phase and in the cell assays [86]. Obviously, the presence of the linker in the tandem-repeat-type galectin-4 alters the reactivity to glyoclusters in a bivalent protein, when compared to homodimeric galectin-1. A major affinity difference to the glycoprotein used as matrix had been excluded by calorimetric titrations with human galectins [87]. Since reducing the linker length has consequences for the selection of cell surface ligands [88], the elucidation of the way the linker affects positioning of the two CRDs for cross-linking becomes a topic for further study. Clearly, its presence means more than an increase in inter-CRD distance.

This high level of sensitivity for galectin-4 was also seen with calixarenes [89]. In addition, the cone-like tetravalent presentation proved rather discriminatory between galectins-1 and -3. This calixarene display was later confirmed to be mostly inactive for galectin-1 when using an assay based on measuring surface plasmon resonance [90]. The differential reactivity between the two galectins could further be increased by an aromatic 3'-substitution at the galactose unit, yet unfavorably affecting solubility [91]. If it becomes an issue to preclude galectin-3 binding from cells, while maintaining surface binding of galectin-1, such a calixarene (please see Figure 3) with a substituted lactose becomes a possibility, although it is still cross-reactive with tandem-repeat-type members of the galectin family [91]. As a laboratory tool, an assumed functional divergence between galectins-1 and -3 can then be verified without having to resort to manipulations on the genetic level. As seen in Figure 3, a further matrix belongs to the set of macrocycles tested with galectins, i.e., glyco(cyclo)phanes and their acyclic forms. Since this compound class has only recently begun to be explored for lectin reactivity, we add information on reactivity to a plant lectin when mannose is conjugated.

5. Glyco(cyclo)phanes

Similarly to cyclodextrins, this scaffold has received interest owing to its ability to accommodate guests such as sugars and hereby form inclusion complexes [92,93]. Also found naturally [94], the versatility in the degree of rigidity in the cyclophane scaffold has enabled the proposal of laboratory applications [95–97], including the testing of chiral variants of the scaffold for lectin ligand design. Presenting structural details beyond the model in Figure 3, two series of glyco(nano)phane-based clusters are depicted in Figure 4. They have either a phenylenediamine or a xylylenediamine within the macrocycle, their acyclic analogues also shown, with mannose (1–4) and lactose (5–8) grafted to the scaffold as the lectin ligand. The mannose-bearing compounds are included here to illustrate the bioactivity of conjugates beyond galectins. The compounds 1 and 3 are diastereoisomeric glyco(cyclo)phanes with α-mannose headgroups, ascertaining the versatility of solid-phase and cell assays as well as testing for differences in relative bioactivities. The saccharides in 1 are linked by a butanediol chain.
that has β-glycosidic linkages to the glucuronic acid, whereas the linkages to the butanediol chain are α-configured in 3. Compounds 2 and 4 are the corresponding acyclic analogues of 1 and 3. The substances 5–8 carry β-lactose units for the interaction with galectins, to extend their comparative analysis with cyclic scaffolds. The lactose moieties in 7 are attached via a triazole-containing linker to position 4 of the glucuronic acid embedded in the glycophane. In 5, the lactose residues are conjugated directly to the position 3 of the glucuronic acid residue. An overview of the synthetic routes to these products is given in Scheme 1 for compounds 3–6 (compounds 1 and 2 were prepared in an analogous manner from β-glucuronide 9) and Scheme 2 for compounds 7 and 8.

**Figure 4.** Structures of macrocyclic glycophane scaffolds (blue) with mannose (1,3) or with lactose (5,7) and their acyclic analogues (2,4,6,8).

The synthesis commenced from the α-glucuronide 10 [98]; its deacetylation followed by formation of an acetylated 6,3-lactone intermediate that was subsequently reacted with allyl alcohol, afforded the glycosyl acceptor 11. The α-mannose residue was attached to the glucuronic acid derivative by glycosidation with trichloroacetimidate 12, and subsequent Pd(0)-catalyzed removal of the allyl ester gave 13. Next, the Ugi reaction [99,100] was employed to produce 14; this was achieved in good yield (> 80%) in one pot by reacting 13 with phenylenediamine, formaldehyde and methyl isocyanacetate. Removal of the acetate-protecting groups from 14 gave acyclic compound 4. Alternatively, ring closure metathesis (RCM) followed by alkene reduction and acetate removal led to ring closure, thus to glycophane 3. The diastereoisomers 1 and 2 were prepared from β-glucuronide 9 by the same route of processing described for 3 and 4.
The acceptor 11 proved useful also in the synthesis of 5 and 6 [101]. Glycosidation with benzoylated lactose derivative 15 and subsequent allyl ester hydrolysis made 16 available. The coupling reaction of 16 with 1,4-xylylenediamine produced 17. Reduction and removal of the benzoyl groups generated 6, while RCM, alkene reduction and benzoyl group removal established the glycophane 5. As outlined in Scheme 2, the synthesis of compounds 7 and 8 was achieved from alkyne derivative 19, which was prepared according to literature procedures from the benzylidene 18 [66]. Alkyne 19 was converted to 20 in three steps, and then the copper-catalyzed azide alkyne cycloaddition (CuAAC) reaction [102] with azide 21 gave 22. Coupling using 1,4-xylylenediamine led to 23, which on deacetylation resulted in 8. When 23 was subjected to RCM [103,104] and the acetates were subsequently removed, this processing established the glycophane derivative 7. Overall, these glycophane based compounds are bivalent, without/with cyclization and it was postulated that they have different degrees of spatial flexibility for presenting the attached sugars. As a test case for this class of compounds, we here present information on modeling to sample a range of conformations and thus spatial headgroup constellations.

Conformational searching techniques based on the previously reported approach using Macromodel 8.0 to both compounds 1 and 2 [105,106] suggested that the lowest energy structures are conformers, where the two carbohydrate residues are stacked (Figure 5). It seemed unlikely due to the closeness of the headgroups that lectins would find access to such stacked sugars. Hence, the occurrence of extended conformers of both the glycophanes and their acyclic analogues in the trajectories were investigated by molecular dynamics simulations with Macromodel 8.0 (Schrodinger Inc., LLC, New York, NY, USA), to complement the earlier study [106].
Scheme 2. Synthesis of compounds 7 and 8.

Figure 5. Stacked (left) and extended (right) conformers that can potentially be accessed by the mannose-presenting macrocyclic (top) and acyclic (bottom) glyoclusters 1 and 2.
In general, stochastic dynamics was applied to a selected conformer at a temperature of 300 K with an equilibration time of 1 µs and a time step of 1.5 fs using the OPLS-AA force field in the gas phase. Before commencing the simulations for 1–4, the peptide side chains were also extended away from the carbohydrate to exclude forming hydrogen bonding with the carbohydrate residues. During each of the subsequent simulations 100–200 structures were sampled and an internal coordinate system established for determining three spatial parameters (Figure 6, top part). In detail, these were the distances between mannose anomeric carbon atoms (Å), a core dihedral, which is defined by atoms C-4 to C-1 to C-1 to C-4 of the mannose residues, and a glycosidic bond dihedral, which is defined by C1 to O1 to O1′ to C1′ of the mannose residues; the latter two parameters were obtained in order to reach a representation of the relative orientation of the mannose residues. Following these definitions of the coordinate system, the scatter plots which were generated are shown in Figure 6.

These plots illustrate spatial arrangements accessed by the divalent mannosides 1–4. The panels in Figure 6 report that structures, in which the internannose distance is <8 Å, are often stacked conformers, whereas those with a distance >10 Å can be considered to represent extended conformations. While some overlapping features are observed when comparing the scatter data for the different molecules, each has its own profile.

**Figure 6.** Scatter plots of data for selected conformers of compounds 1–4 as generated from molecular dynamics simulations using Macromodel 8.0. The definitions of distance between mannose units, glycoside bond dihedral and core dihedral are shown at the top.
Figure 6. Cont.

For example, macrocycle 1 and its acyclic counterpart 2 can access conformations, where the distance between mannose residues is ~15 Å and the glycoside bond dihedral angle varies from −50° to +50°. However, the core dihedral can change from 50°–150° for compound 1 and from 120° to −160° for compound 2. Thus, macrocyclization appears to induce a different core orientation between the two mannose residues, even though the distance between these residues and the glycoside bond dihedral may be similar in both the macrocyclic and the acyclic structures. When comparing both macrocycles 1 and 3, it is clear that the distance between mannose residues differs. The distances were found to vary from 6–14 Å for 3 but more restrained for 1 between 14–16 Å during the simulations; the increased rigidity of 1 compared to 3 can be viewed in respective movies of the dynamics simulations (for access, please go to http://youtu.be/yzZxCGNQ6j0 for compound 1 and http://youtu.be/RZObaf6MD24 for compound 3).

In terms of bioactivity, the glyoclusters based on mannose were shown to be active as ligands for two leguminous lectins, with cyclization accounting for a trend toward enhanced activity [106]. As testing cells with shifts in the glycome revealed, the specific nature of the glycan display has a marked bearing on relative levels of inhibition [106]. Such an impact of structural aspects of glycosylation had been noted before when examining properties of different glycoproteins with complex-type N-glycosylation in the solid-phase assay on galectins [78,81]. These observations preclude extrapolations and require the establishment of a broad experimental basis. They also attest reactivity of the presented mannose moieties to the plant lectins, encouraging work on the lactosides and galectins.

Running the same protocols for the lactose-presenting compounds 5–8 containing 1,4-xylylendiamine as opposed to the phenylenediamine unit in compounds 1–4 also led to stacked conformers. To interact with galectins extended conformations (Figures 7 and 8) may be more relevant. In this case of lactose, the core dihedral was defined by atoms Gal C-4 to Glc C-1 to Glc C-1′ to Gal C-4′ of the lactose residues, the galactose dihedral is the dihedral angle defined by Gal O-4 to Gal C-4 to Gal C-4′ to Gal O-4′ and the distance is that measured between the two anomic carbon atoms of glucose (Figure 9). The scatter plots illustrate spatial arrangements accessible to the divalent lactosides 5–8, excluding stacked conformers. At inter-lactose distances of >10 Å extended
conformations will be reached. As noted above, each compound has its own conformational profile. Both macrocyclic compounds showed more rigidity in terms of distance between the lactose residues when compared with the acyclic analogues; the distance between lactose residues in compound 5 varied from 14–16 Å, increasing to 19–23 Å for 7, what reflects the impact of linker characteristics. The extended conformers for the acyclic analogues can apparently reach inter-lactose distances ranging from 10 Å to 25 Å.

**Figure 7.** Examples of stacked (left) and extended (right) conformers of the lactose-presenting macrocyclic glyocluster 5.

**Figure 8.** Models of lactose-bearing compounds 6 (top), 7 (middle) and 8 (bottom) in extended conformations.
Figure 9. Spatial parameters for extended conformers of compounds 5–8. The definition for the galactose dihedral, core structure and distance between lactose units is shown on top.

In terms of bioactivity, the glycophane 5 was markedly more active against the CRD of human galectin-3, a product of proteolytic truncation, than the full-length protein [101]. The spatial characteristics of the cyclic form 5 facilitated selectivity between galectin-3 and its truncated version at a discriminatory level of an about 5fold difference [101]. The disparity between the cyclic and acyclic compounds 5 and 6 for this protein is more than 10fold, the IC_{50}-values at 0.3 mM and 5 mM, respectively. The cyclic form 7 was found to be more active for CG-8, the chicken tandem-repeat-type galectin, than its acyclic analogue 8, the increase of relative inhibitory capacity to lactose being 4.5 fold [66]. The triazole at the anomeric carbon appears to be a favorable linker part, as corroborated with other types of glyoclusters [107]. Overall, these data and the examples from the previous paragraphs document the activity of cyclic glyoclusters, giving reason to add comments on ways to enhance their inhibitory capacity/selectivity and the potential for applications.
6. How to Optimize Inhibition, What to Expect

Two main parameter changes are possible in order to increase the inhibitory capacity of a glycocluster on the lectin-glycan interaction, beyond the spatial features discussed so far in this review. The first structural region for tailoring is the aglyconic anomeric extension. It can strengthen binding of the sugar headgroup and provides the functional group for attachment to the scaffold. Experience with galectins, starting with detecting the slight enhancing effect of reactivity of p-aminophenyllactoside relative to lactose for galectin-1 [108,109] and further exploring other aglyconic substituents, revealed quantitative effects not above a several-fold increase [110,111]. More strongly, the nature of the sugar headgroup can exert an effect on affinity and selectivity, the second parameter change. Respective systematic profiling, for example by binding and inhibition assays using glycoproteins/glycan derivatives and free or resin-immobilized galectins, even as isotopically labeled probes in NMR spectroscopy, by docking protocols, titration calorimetry or glycan arrays documented the significance of this factor [112–121]. Exploiting this parameter holds the promise to accomplish improvements. Already a rather minor biomimetic adaptation such as taking advantage of the physiologic target specificity of galectin-4 for the sulfatide headgroup in apical delivery of glycoproteins [122,123] can likely be very helpful. This teamed up with dendrimeric display, taking into account this lectin’s sensitivity for high ligand density [86,89,124], offers a tempting perspective. Combining the power and creativity of synthetic carbohydrate chemistry [125] and the well-studied chemical routes toward multivalent display [126], for example in starburst dendrimers tailored to galectin-1 [127], with comprehensive case-by-case testing of galectin panels is expected to track down any discriminatory ability that glycoclusters have for these (and other) lectins. At the same time, the affinity increase by cluster design can be conducive to let it act as postal address for cargo delivery by bitopic conjugates, e.g., directing a phototoxin (hematoporphyrin) or inhibitors of matrix metalloproteinases to sites of high galectin density [128,129]. In addition to the cytofluorometric assays cyto- and histochemical determination of staining by labeled galectins [130,131] in the absence/presence of inhibitors will also define their inhibitory potential on lectin binding to physiological counterreceptors. Of note, here changes in signal intensity can be monitored at the membrane or the extracellular space, and also in the cytoplasm and nucleus, providing a sensor for the influence of glycan on intracellular lectin interactions.

Finding optimal combinations of scaffold, linker and sugar headgroup together with its display and favorable dynamics will be helpful to relate spatial aspects of functionality. Considering a perspective for lectin blocking, such parameters will be necessary to minimize cross-reactivity among lectins of a family and also between different lectin families sharing carbohydrate specificity, a major caveat not to be neglected before claiming any application of lectin-targeted drug design [132]. Moreover, galectins are known as multifunctional proteins, the exact nature of the effect depending on the context. Taking galectin-1 as example, it can be growth inhibitory on carcinoma cells [39,43] but favor tumor progression and invasion not only in glioblastoma but also pancreatic carcinoma [133,134], at the same time for example serving as versatile immune regulator [73]. As discussed in detail previously, caution needs to be exercised and very detailed insights into the multifaceted galectin functionality at the different sites of localization gained before deliberating to interfere with binding of a certain galectin, within the natural network and also in the context of other lectins [132].
7. Conclusions

The growing insights into the way lectin-glycan recognition contributes to cell physiology have inspired chemists to design biomimetic glycoclusters. Tested primarily with plant lectins such as concanavalin A or peanut agglutinin as models, covalent conjugation of sugar derivatives to diverse scaffolds has been shown to retain the bioactivity of the sugar headgroup. The enhancement of avidity by cluster formation in neoglycoconjugates has been exploited for different applications, among them directing cargo to cells or lectin localization in tissues and cells [51,135–137]. Generally, different modes of CRD presentation are found among lectins, within a protein a tandem-repeat display or even widely separated domains and in non-covalent aggregates, as illustrated in Figure 1 for galectins or for C-type lectins in [138]. This spatial parameter, together with a matching display on the glycan side, is assumed to guide complex formation. The cross-linking is the prerequisite to start signaling for growth control, two examples for the biochemical details of the intracellular cascade presented in [139,140]. Evidently, by making the design of a range of glycan displays possible, glycoclusters become highly welcome tools for delineating structure-activity profiles. In molecular detail, spatial aspects can then be examined in assays of increasing biorelevance.

Presented data for lactose-bearing glycoclusters revealing respective differences between galectin-1 vs. galectins-3 and -4 substantiated that the mode of spatial presentation can markedly matter. Whether this line of research can be viewed to have a therapeutic perspective critically depends on collecting a wealth of information not just on one or few proteins but on the complexity of (a) the natural lectin network, (b) the inherent multifunctionality of its individual members and (c) the glycome, on the mentioned six levels of affinity regulation. Undoubtedly, the synthetic compounds will have their merit in laboratory experiments to relate spatial presentation to reactivity, a key source of specificity/selectivity in translating the sugar code.

Acknowledgments

Our work has been generously supported by the European Commission through Marie Curie Intra-European Fellowships (500748, 514958, 220948), Marie Curie Initial Training Network GLYCOPHARM (PITN-GA-2012-317297), the GlycoHIT program (grant agreement 260600), the Programme for Research in Third-Level Institutions (PRTLI), administered by the Higher Education Authority, the Verein zur Förderung des biologisch-technologischen Fortschritts in der Medizin e.V. (Heidelberg, Germany) and the Irish Research Council, Enterprise Ireland and Science Foundation Ireland (04/BR/C0192, 06/RFP/CHO032, 12/IA/1398). Inspiring discussions with Bernd Friday are gratefully acknowledged, as is the valuable input by the reviewers.

References


© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).