



# Combining glycocluster synthesis with protein engineering: an approach to probe into the significance of linker length in a tandem-repeat-type lectin (galectin-4)



Sabine André<sup>a</sup>, Guan-Nan Wang<sup>b</sup>, Hans-Joachim Gabius<sup>a</sup>, Paul V. Murphy<sup>b,\*</sup>

<sup>a</sup>Institute of Physiological Chemistry, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Veterinärstr. 13, 80539 Munich, Germany

<sup>b</sup>School of Chemistry, National University of Ireland Galway, University Road, Galway, Ireland

## ARTICLE INFO

### Article history:

Received 12 September 2013

Received in revised form 23 December 2013

Accepted 27 December 2013

Available online 11 January 2014

### Keywords:

Agglutinin

Glycoprotein

Lectin

Modelling

Terephthalamides

Triazoles

## ABSTRACT

Complementarity in lectin–glycan interactions in situ is assumed to involve spatial features in both the lectin and the glycan, giving a functional meaning to structural aspects of the lectin beyond its carbohydrate-binding site. In combining protein engineering with glycocluster synthesis, it is shown that the natural linker length of a tandem-repeat-type human lectin (galectin-4) determines binding properties in two binding assays (using surface-presented glycoprotein and cell surface assays). The types of glycocluster tested included bivalent lactosides based on tertiary amides of terephthalic, isophthalic, 2,6-naphthalic and oxalic acids as well as bivalent H(type 2) trisaccharides grafted on secondary/tertiary terephthalamides and two triazole-linker-containing cores. The presented data reveal a marked change in susceptibility to the test compounds when turning the tandem-repeat-type to a proto-type-like display. The testing of glycoclusters is suggested as a general strategy to help to delineate the significance of distinct structural features of lectins beyond their contact sites to the glycan.

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

Our understanding of the biological roles of complex carbohydrates is currently taking a quantum leap. Initially viewed to store energy and confer rigidity to cell walls, the sugars' unsurpassed versatility for generating structural variability of oligomers is being delineated as a platform for coding, with receptors (lectins) reading and translating sugar-encoded messages into cellular effects.<sup>1</sup> Fittingly, cases of functional orchestration of glycan/lectin expression, for example, in inflammation or by a tumour suppressor, are unravelled, stimulating project lines to define the diagnostic/prognostic potential of carbohydrate–protein interactions.<sup>2</sup> Moving on from work with plant lectins as models, the quest to explain the inherent specificity and selectivity of the interplay of tissue lectins with their (cognate) counterreceptors has inspired efforts to shape synthetic binding partners, in terms of the structure of headgroups and their spatial presentation.<sup>3</sup> The asialoglycoprotein receptor on hepatocytes has become a role model with its preference for trivalent binding partners, with its three contact sites separated by about 15 Å, 22 Å and 23 Å.<sup>3a</sup>

Our work in this area has thus far focused on bi- to tetravalent compounds, where the headgroups were grafted to

terephthalamide, *N,N'*-diglucosylterephthalamide, glycopane and triazole containing scaffolds.<sup>4</sup> Terephthalamide was found to be a scaffold suited for presenting carbohydrate ligands to plant and human lectins, as exemplified by the bioactivity of compound **1**.<sup>4a,b</sup> This glycocluster **1**, which is a constrained compound, features a distance of 7 to 8 Å between its two sugar headgroups (intersugar or interheadgroup distance).<sup>4a</sup> In order to explore structure–bioactivity relationships using such glycoclusters as tools, we prepared compounds **2–7** (Chart 1) as a set of analogues of **1**, with lactose as the headgroup. Beside the terephthalamide scaffold in **2**, we generated acceptors for ligands based on tertiary amides of oxalic acid (**3**), isophthalic acid (**6**) and 2,6-naphthalic acid (**5**). In these three cases, the interheadgroup distance is varied, with possibly the headgroup orientation also being varied, while maintaining the sugar and the acetylglycine residue bonded to the nitrogen atoms. In compounds **2** and **4**, a *t*-butyl residue was incorporated instead of the acetate residue present in compounds **3**, **5** and **6**; it was anticipated that this modification would give an indication as to whether the group attached to the anomeric nitrogen atom has an effect on the reactivity to a lectin. To assess the impact of the headgroup, lactose was extended in four substances to the trisaccharide fucosyl lactose, a positive regulator of reactivity for certain human lectins.<sup>4d,5</sup> Compound **10** is a direct analogue of fucosyl lactose containing **7**, which contains secondary amides rather than the tertiary amides found in **7** and the lactose deriva-

\* Corresponding author. Tel.: +353 91492460.

E-mail address: [paul.v.murphy@nuigalway.ie](mailto:paul.v.murphy@nuigalway.ie) (P.V. Murphy).

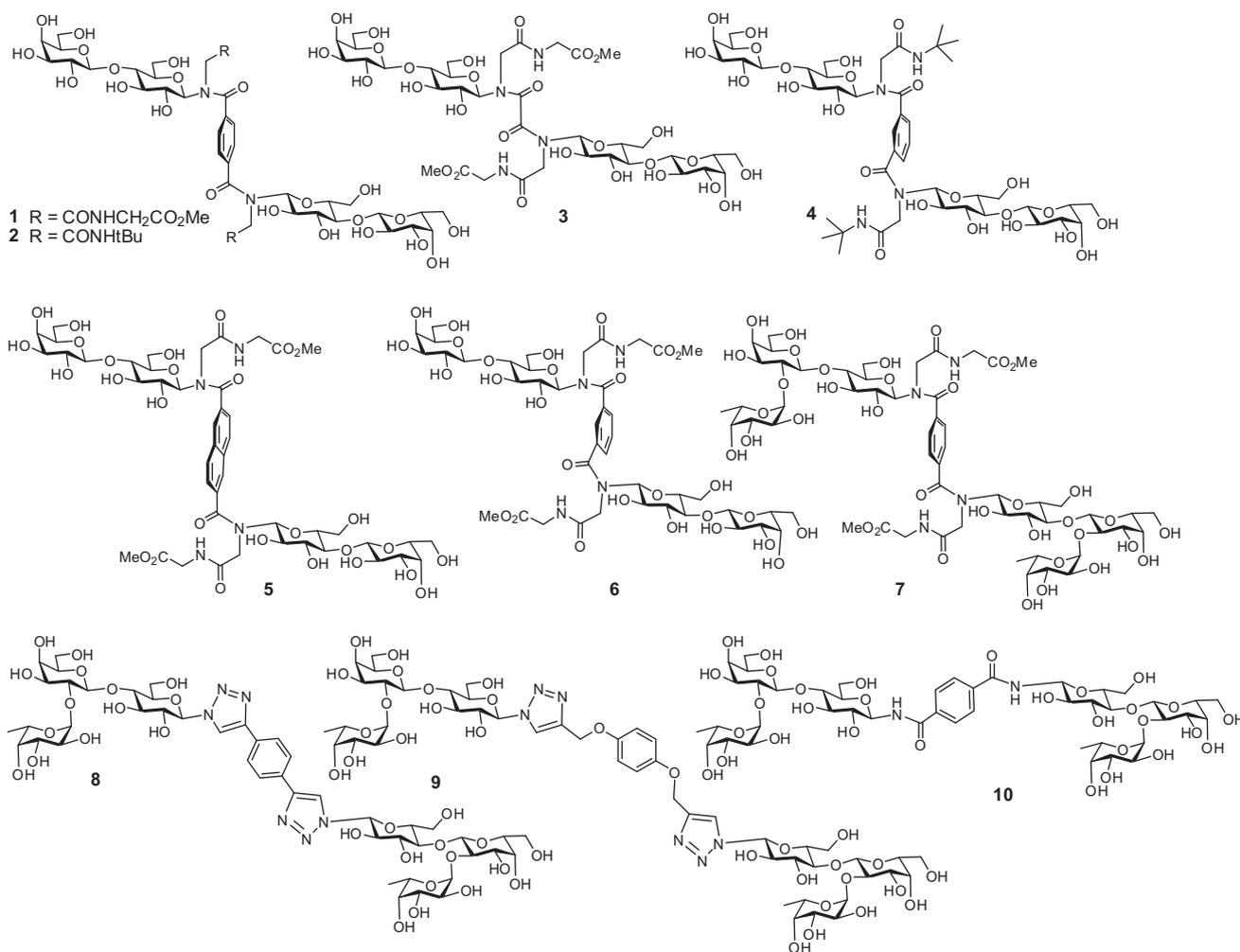
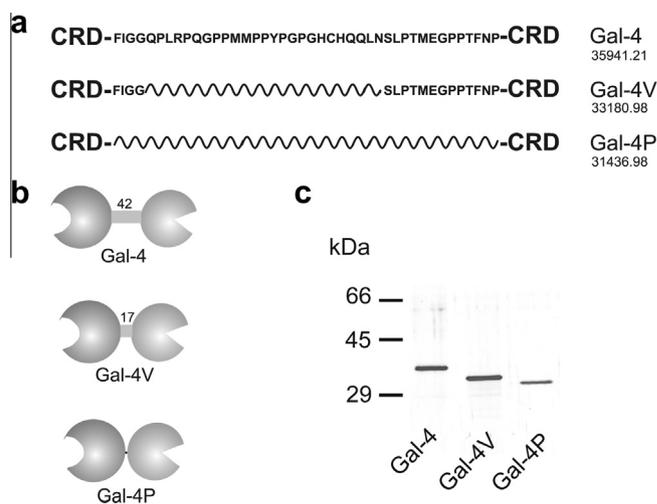


Chart 1. Structure of glycoclusters 1–10.

tives 2–6. Two additional bivalent fucosyl lactose-presenting substances, that is, **8** and **9**, were also included, here with sugar attached to triazole-containing linkers previously proven as being compatible for lectin reactivity.<sup>4c</sup> Since bioactivity of compounds of these classes has thus previously been documented for members of the adhesion/growth-regulatory family of human/animal galectins,<sup>4</sup> we can herein proceed to test these bivalent clusters as sensors to answer a distinct question on the structure–activity relationship of a human galectin.

Among the galectins, one of their three groups is established by the tandem-repeat design. Two different carbohydrate recognition domains (CRDs) are covalently connected by a linker (Fig. 1).<sup>6</sup> Human galectin-4 (Gal-4), a molecular transporter in glycoprotein routing and delivery in enterocytes and neurons,<sup>7</sup> has a single fixed length of this connecting peptide (Fig. 1). In assay settings which determine inhibition of lectin binding to glycoproteins, Gal-4 has proven to be rather sensitive to the presence of glycoclusters, secondary and tertiary terephthalamides among them.<sup>4a,8</sup> This capacity of glycoclusters to interfere with Gal-4 binding to a ligand-exposing surface (on a plastic surface or cell surface) enables addressing the issue of the relevance of linker length for binding capacity to Gal-4. To do so, two variants, that is, Gal-4V with length reduction and Gal-4P with complete linker removal mimicking a dimer of the proto-type group, were designed and produced (Fig. 1), then tested with the panel of compounds shown in Chart 1 in solid-phase/in vitro assays. A homodimeric



**Figure 1.** Illustration of the design of the two Gal-4 variants by comparatively presenting the sequence of the natural 42-amino-acid linker along with the positions of deletions between the N- and C-terminal CRDs and the resulting molecular masses (a) as well as the linker lengths (b) in Gal-4V and Gal-4P. The products of recombinant expression show purity and the expected positions in gel electrophoretic analysis (c).

(proto-type) galectin (CG-1A) with strong reactivity to N-glycans with LacNAc termini was used as a control,<sup>9</sup> and the sensor

character of the compounds was further tested in a second system of galectin/galectin variant. This was established by chimera-type galectin-3 (constituted by the CRD and a stalk of the N-terminal peptide with two sites for Ser phosphorylation and nine Pro/Gly-rich collagen-like tandem repeats; CG-3) and its proteolytically truncated form, that is, the CRD free of this collagenase-sensitive N-terminal tail.<sup>10</sup> The marked effects of protein engineering/truncation on the reactivity of lectins to certain glycoclusters underscore the structural significance of the natural linker length of Gal-4 and the presence of the collagen-like tail in CG-3.

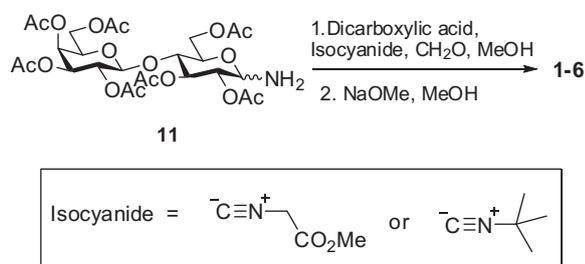
## 2. Results and discussion

### 2.1. Synthesis of the glycoclusters

The preparation of the bivalent glycoclusters **1–6** with lactose as headgroup started from the lactosamine **11** (Scheme 1) and involved the Ugi reaction. The reaction of **11** with formaldehyde, terephthalic acid and methyl isocyanacetate gave **1** after deacetylation, as described earlier.<sup>11</sup> When **11** was reacted with oxalic acid, 2,6-naphthalenedicarboxylic acid and isophthalic acid, respectively, in the presence of formaldehyde and methyl isocyanacetate, and the acetyl groups removed, then compounds **3**, **5** and **6** were obtained. When *t*-butyl isocyanide was reacted with lactosamine **1**, formaldehyde and terephthalic acid or isophthalic acid in methanol, the bivalent compounds **2** and **4** resulted after deacetylation.

The syntheses of bivalent glycoclusters presenting as headgroup the trisaccharide fucosyl lactose were next carried out (Schemes 2 and 3). Firstly, the fucosylated azide **12** was prepared as previously described.<sup>4d</sup> This azide was used in copper-catalysed azide alkyne cycloaddition reactions<sup>12,13</sup> with **14** and **15** to give, after deacetylation and subsequent HPLC-based purification, **8** and **9**, respectively. Although the azide **12** contained a small amount of the  $\alpha$ -anomer (<6%), it was possible to remove the cycloaddition products that resulted from the presence of this anomer by careful chromatography prior to the final deprotection step. Next, catalytic hydrogenation of **12** resulted in the 2'-*O*-fucosyl lactosamine **13** (Scheme 2), which when used in a coupling reaction with terephthaloyl chloride in the presence of DIPEA and followed by deacetylation was turned into **10**. Since the glycosyl amine **13**, formed by the reduction of the azide, contained a small amount of its  $\alpha$ -anomer (<10%), a mixture of diamides was obtained from the coupling reaction with the terephthaloyl chloride. Purification by HPLC was necessary in order to obtain **10** as a single product.

Since the Ugi reactions of **11** were successful, we also were interested to try to prepare the fucosyl lactose analogue of **1** from **13**. However, the Ugi reaction did not succeed for peracetylated compound **13**. To solve this problem the Ugi reaction of amine **18** was carried out instead. Compound **16** was treated with saturated solution of ammonium hydrogen carbonate to give **17**. Next, the amine group of **17** was temporarily protected by a treatment



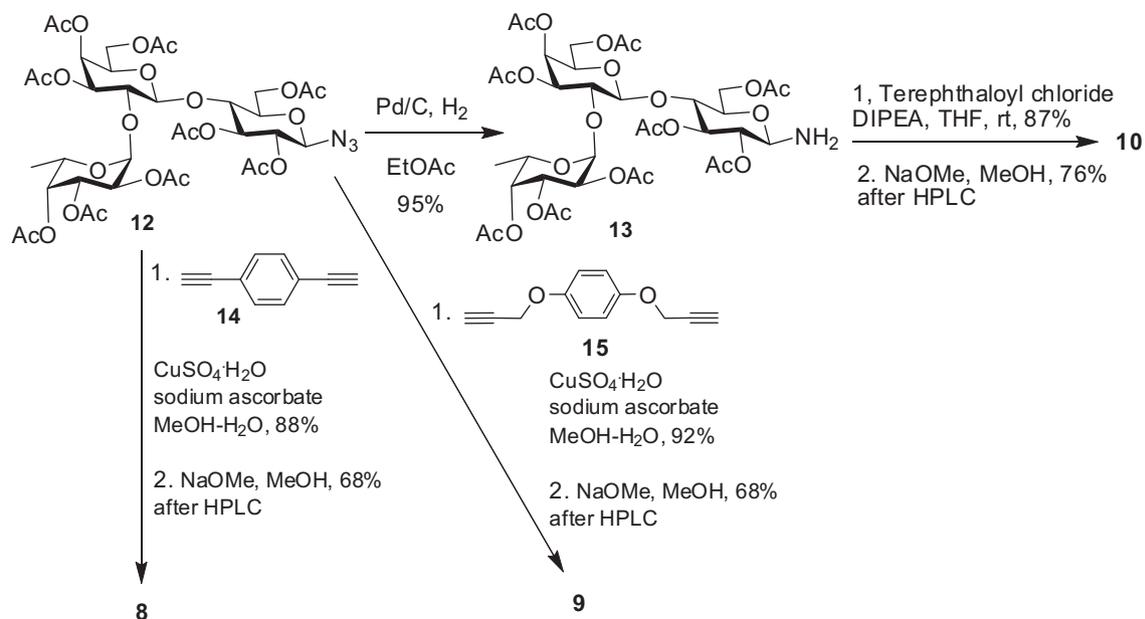
Scheme 1. Synthesis of bivalent glycoclusters **1–6**.

with *N*-(9-fluorenylmethoxycarbonyloxy)-succinimide (FmocOSu) in pyridine and the free hydroxyl groups were acetylated by adding acetic anhydride. The Fmoc was removed thereafter using morpholine in DMF to give the fucosyl lactosamine **18**. This was subsequently reacted with formaldehyde, methyl isocyanacetate and terephthalic acid to yield a diamide. The protecting groups were removed by Zemplén deacetylation and hydrogenolysis to give rise to **7** (Scheme 3).

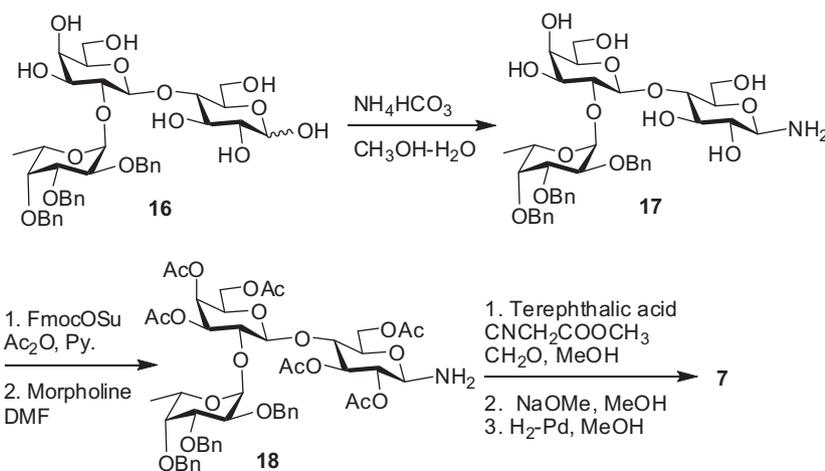
### 2.2. Structural analysis and molecular modelling

Compounds **1–7** and **10** are divalent glycosyl amides, structurally related to terephthalamides described previously.<sup>11</sup> In **1–7**, the amides are tertiary amides, whereas **10** harbours secondary amides. In terms of amide configuration, the amides in **10** are *trans* amides (or *Z* amides). This term *trans* is applied as in peptide chemistry. Compound **10** shows one set of signals in the NMR spectrum, and the chemical shift of the anomeric proton appears at  $\delta$  5.84 as a doublet with a coupling constant of  $J = 9.3$  Hz. This is consistent with **10** adopting a *trans–trans* (*ZZ*) structure shown in Figure 2, by comparison with related secondary amides.<sup>11</sup> However, and as shown previously for tertiary terephthalamides, the tertiary amides in **1–7** adopt configurations where the *cis* (*E*) amide is preferred. The NMR spectra show two signal sets for all these compounds. For example, for **2** (Fig. 3) the anomeric proton for the *cis* (or *E*) amide appears at  $\delta$  4.84 ( $J = 8.8$  Hz), whereas that for the *trans* amide was observed at  $\delta$  5.76 ( $J = 9.3$  Hz). By integration of suitable signals in the <sup>1</sup>H NMR spectra the ratio of *cis:trans* (*E:Z*) and consequently the *cis–cis* (*EE*) to the *cis–trans* (*EZ*) isomeric ratios are determined. Thus for compound **2** (Fig. 3) there are two well separated signals for the galactose anomeric protons in isomeric structures and these are observed at  $\delta$  4.50 (d,  $J = 7.8$  Hz, 2H, H-1' *Z*) and  $\delta$  4.44 (d,  $J = 7.8$  Hz, 2H, H-1' *E*); integration of these signals enabled establishment of the *EE:EZ* ratio in **2** to be 70:30. These ratios are summarized for compounds **1–7** in Table 1. All compounds **1–7** showed >2:1 preferences in solution for the *cis–cis* isomer. There are essentially two conformations that can be considered for the *cis–cis* isomers, and these are depicted in Figure 2. The carbohydrate headgroups could be stacked (stacked *cis–cis*) or presented in an extended conformation (extended *cis–cis*). The extended *cis–cis* structure is considered to be more relevant for binding to lectins, because headgroups might be too close for reactivity when stacked. It is also possible that lectins could bind to their ligand in the *EZ* or *cis–trans* structure, which is present to a lesser degree in solution. Dynamic transition to the *EZ* conformer is possible in the equilibrium as a result of amide bond rotation, and there is evidence that this exchange can take place for compounds related to **1–7**.<sup>11</sup> While the *trans–trans* (*ZZ*) isomer for compounds **1–7** (not shown) could also be adopted, it appears to be present at concentrations too low to be detected by NMR.

Based on the NMR data for compounds **1–7** we conclude that the major structural isomer is *cis–cis* or *EE*. Assuming that the extended *cis–cis* structure is more relevant than the stacked *cis–cis* structure and *cis–trans* structures for presenting the headgroup for binding to a lectin, we next carried out molecular dynamics simulations using Macromodel 8.0 (Schrodinger Inc., LLC, New York, NY, USA), to supplement our previous studies of divalent compounds.<sup>4a</sup> The main purpose of this work is to estimate distances between the carbohydrate headgroups and respective orientations. Models were built for the compounds **1–3** and **5–6**, and each compound was treated equally. Stochastic dynamics was applied to the selected structure at 300 K with an equilibration time of 1 ns and a time step of 1.5 fs using the OPLS-AA force field in the gas phase. A simulation was carried out in Macromodel for compound **2** using the GB/SA effective solvent model for water (not explicit water) and no major differences were observed



Scheme 2. Synthesis of 8–10.

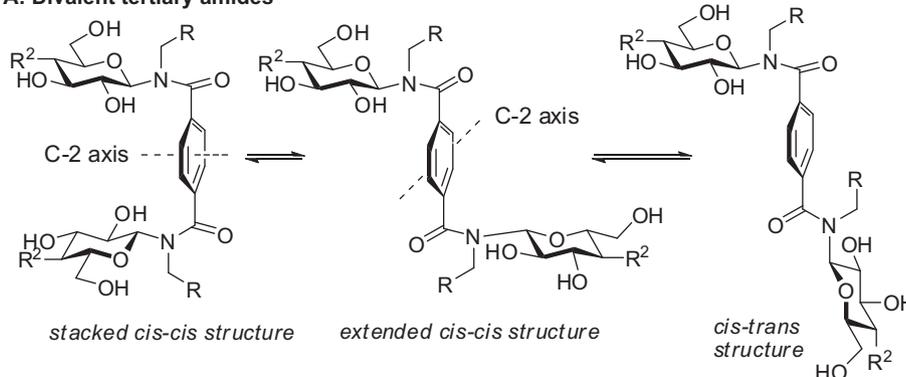
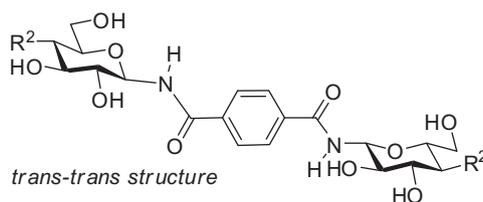
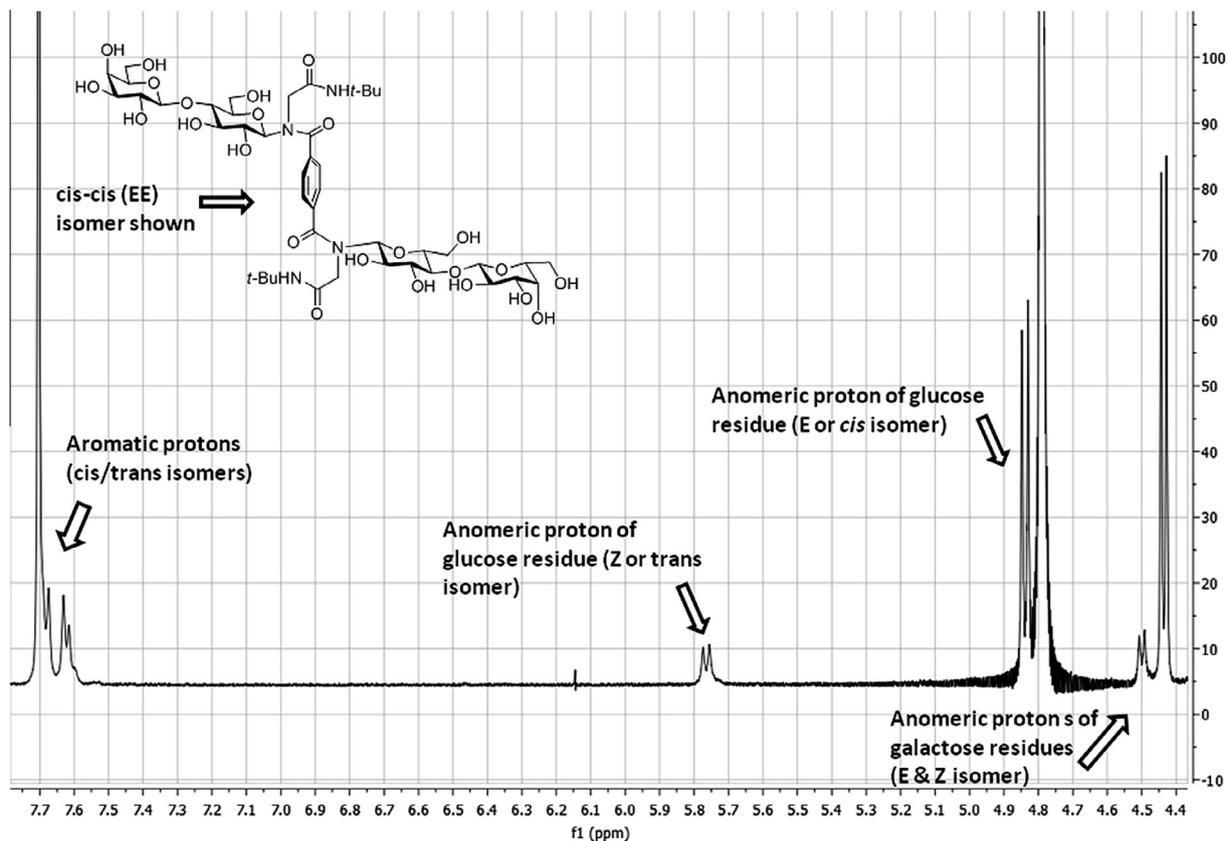


Scheme 3. Synthesis of 7.

between the gas phase simulation of **2** and estimating the GB/SA effective water model for **2**. During each of the subsequent simulations 100–200 structures were sampled and an internal coordinate system was used to define spatial parameters for these structures. This system firstly involves measuring the distances for the sampled structures between the anomeric carbon atoms of the glucose unit in the disaccharide (in Å) to give the interheadgroup distance. In addition, two dihedral angles were measured, a core dihedral and a galactose dihedral, for the sampled structures, providing parameters relevant to the interheadgroup orientation, and these are defined in the top part of Figure 4. Scatter plots were generated using the data obtained, and these are also shown in Figure 4. These plots can be considered representative of the spatial arrangements between the lactose headgroups accessible during the simulations, which were run at identical conditions.

As can be seen, by comparing data for compounds **1** and **2**, the measured galactose and core dihedrals did not vary significantly, indicating that the relative orientations between the two potential contact sites for galectins are similar for these compounds and

apparently not greatly influenced by replacing the terminal acetate with the *t*-butyl group. Similarly, the distance between the two disaccharides did not change noticeably, averaging between 7–8 Å for both compounds during the simulations, as can also be expected for compound **7**. For **6**, which has the isophthalic acid, similar interlactose distances to compounds **1** and **2** were observed, but there was more flexibility in interheadgroup orientation, as can also be expected for the structurally similar compound **4**. With regard to compound **5**, which has naphthalene rather than benzene, the distance between the lactose residues is increased, as would be expected, and this varied between 9 and 10 Å during the simulation. Molecular dynamics simulations for compound **10** was not carried out but, on the basis of studies with a related lactose diamide,<sup>4a</sup> the distances between the residues would be expected to be 10 Å and similar to **5**. The interheadgroup orientation in **5** was similar to that found in compounds **1** and **2**. Coming to the ditriazoles, the distance between the two trisaccharides in **8** would be expected to approximate to 13 Å, whereas compound **9** is more flexible with distances of up to 16–17 Å being possible. At the other

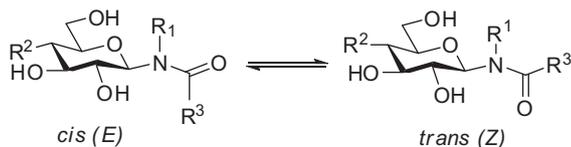
**A: Divalent tertiary amides****B: Divalent secondary amides****Figure 2.** Amide structures.**Figure 3.** Partial <sup>1</sup>H NMR spectrum of compound **2** (D<sub>2</sub>O, 500 MHz, δ 4.35–4.80 ppm).

side of the range of distances, for the oxalic acid derivative **3**, its lactose residues were separated by 3–4 Å and the interheadgroup dihedrals were also significantly different.

Broadly spoken, the distances between sugar headgroups in this range of compounds lie in the range of those of LacNAc termini in

biantennary complex-type N-glycans with backfolding (~5.9 Å), without backfolding (8.1 Å) and with anti-parallel configuration (22.1 Å), core substitutions regulating the position of the α1,3/6-arms, and 15–22 Å for the type 1 triantennary complex-type N-glycan.<sup>14</sup> In this sense, the synthetic compounds share spatial

**Table 1**  
Ratio of isomers of compounds 1–7



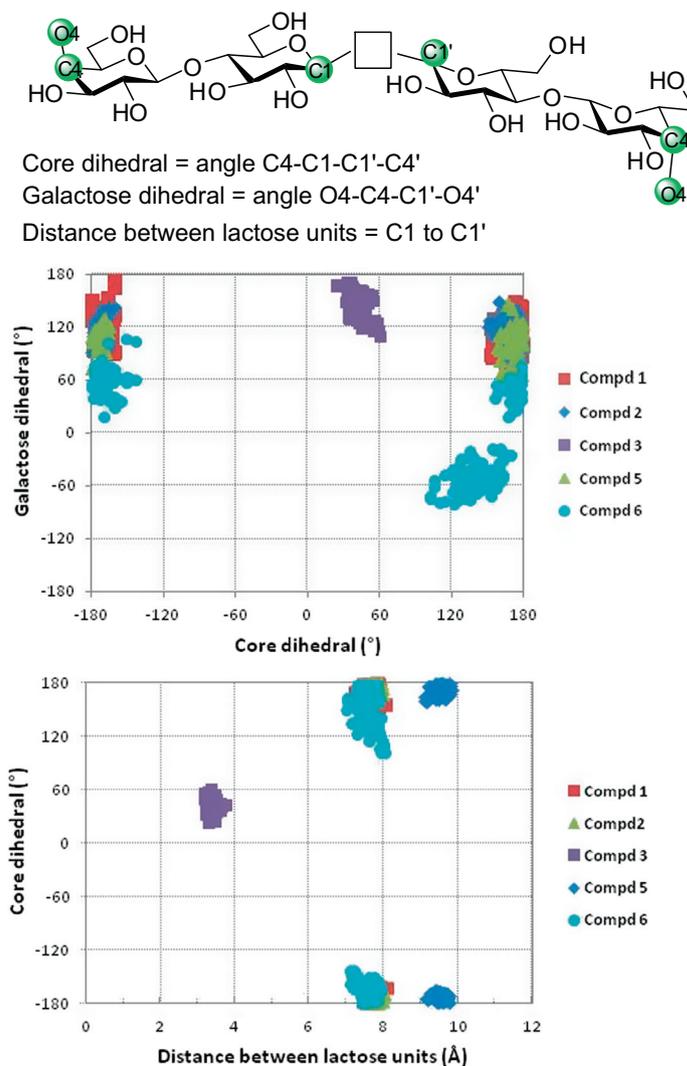
Compound	<i>cis:trans</i> ( <i>E:Z</i> ) ratio	<i>cis-cis:cis-trans</i> ratio ( <i>EE:EZ</i> , see Fig. 3)
1	85:15	70:30
2	85:15	70:30
3	85:15	70:30
4	85:15	70:30
5	92:8	84:16
6	92:8	84:16
7	88:12	76:24

properties with a natural constellation. Since Gal-4 is strongly reactive with N-glycans presenting LacNAc termini,<sup>15</sup> and every

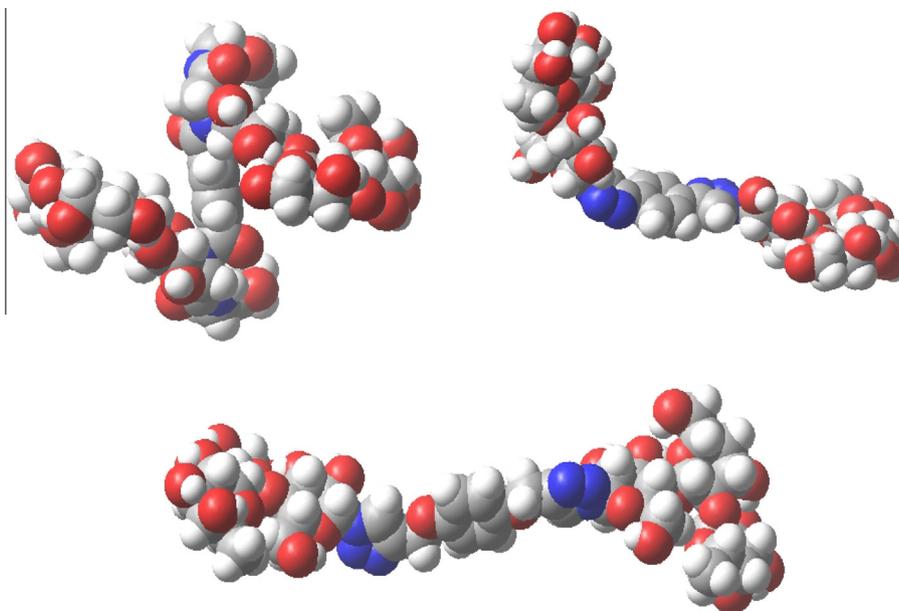
branch-end LacNAc can become a binding partner for a CRD of human galectins at saturation in solution, albeit with a gradient of decreasing affinity,<sup>9b,16</sup> the compounds of this panel may disclose information relevant for the discussion on the significance of the natural length of the linker. As in its function in glycoprotein routing, each Gal-4 is expected to react with headgroups of two synthetic molecules. Space-filling models have been shown for compounds 7–9 (Fig. 5).

### 2.3. Analysis of inhibitory potency

In the first type of assay, a glycoprotein (asialofetuin with up to nine LacNAc termini in the three triantennary complex-type N-glycans) was adsorbed to the plastic surface of microtiter plate wells, a matrix for carbohydrate-dependent binding of Gal-4. Binding was saturable, dependent on the presence of the terminal galactose moieties of the N-glycans and was blocked completely by an excess of hapten (lactose), for all galectin proteins tested (not shown). Titrations at an assay setting in the linear range of the OD-response with increasing concentrations of the test compounds determined the concentration to reduce extent of the signal by 50% ( $IC_{50}$ -value), a relative measure of the capacity to interfere with glycan–lectin interaction (for representative cases, please see

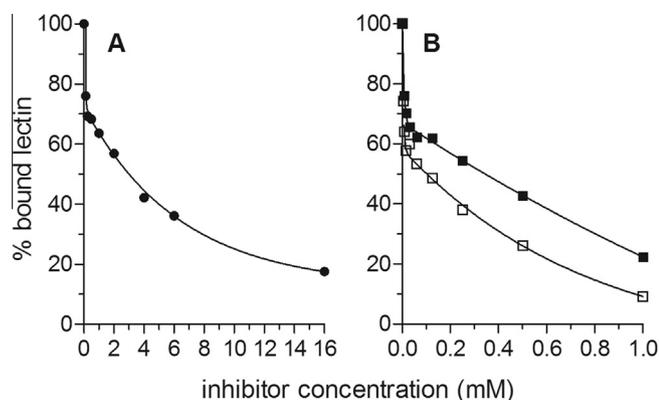


**Figure 4.** Spatial parameters for *EE* structures sampled during molecular dynamics simulations are represented using scatter plots. The definition for the dihedrals and distance between lactose units is included in the upper part of this figure.



**Figure 5.** Space filling models of fucosyl lactose derivatives **7** (top left), **8** (top right) and **9** (bottom).

**Fig. 6).** As given in **Table 2**, Gal-4 binding was rather effectively inhibited by the bivalent compounds, extending the previously collected evidence with a tertiary terephthalamide (compound **1**) and *N,N'*-diglucoylterephthalamides.<sup>4a</sup> The addition of the  $\alpha$ 1,2-linked fucose moiety to the lactose headgroup expectedly led to an increase in inhibitory potency, from 0.3 mM for lactose in the secondary terephthalamide<sup>4a</sup> to 0.05 mM (compound **10**). The introduction of the *t*-butyl group (for compounds **2** and **4**) can make a rather minor difference, and having oxalic acid in the backbone proved rather favourable, compared to isophthalic acid (**Table 2**). The two ditriazoles with the trisaccharide were the most potent inhibitors (**Table 2**). Overall, the data further substantiate the sensitivity of this lectin to bi- and trivalent glycoclusters, hereby establishing a solid basis to ask the question whether and how linker-length reduction in the tandem-repeat-type lectin will affect the reactivity towards the glycocluster. Two forms of engineered Gal-4 were therefore tested, by maintaining linker parts phylogenetically conserved in mammals (Gal-4V) and by turning Gal-4 into a linkerless, proto-type-like form (Gal-4P) (**Fig. 1**).



**Figure 6.** Courses of titrations illustrating the extent of inhibition of binding of labelled Gal-4V (5  $\mu$ g/mL) to surface-presented asialofetuin in the solid-phase assay upon increasing the concentration of lactose (A) as well as compound **9** ( $\square$ ) and compound **10** ( $\blacksquare$ ) (B).

The results obtained by testing these two variants under identical conditions clearly showed that: (i) reduction of linker length decreased the lectin's susceptibility to the test compounds; and (ii) this truncation of the linker and its complete removal led to similar results (**Table 2**). Most conspicuously, the two ditriazoles **8** and **9** dropped markedly in inhibitory capacity, a result definitely warranting confirmation by a different assay type. However, a special point precludes direct comparability for all data: the interpretation of the data for fucosyl lactose-presenting compounds **7–10** should take into account that the reactivity of Gal-4V to this epitope has been shown to decrease relative to that of the wild-type protein.<sup>5c</sup> An impact of the truncation both on affinity and on ligand selection has furthermore been revealed for surface binding to human neuroblastoma cells.<sup>17</sup> To document how a dimeric galectin will react with the glycoclusters we have added CG-1A. The range of  $IC_{50}$ -values of the two engineered Gal-4 proteins was rather similar to that of this control protein (**Table 2**). Its level of reactivity was in line with previous results for CG-1A and dilactosyl diamides/di-, tri- and tetravalent fucosyl lactosides.<sup>4c,d</sup>

In order to exclude the possibility that characteristics inherent to the solid-phase assay underlie the differences seen in **Table 2** we next performed cell-binding assays. As a test principle, binding parameters of the labelled galectin to the surface of cells in culture can be affected by the presence of the test compounds and this effect was quantified in terms of the percentage of positive cells and mean fluorescence intensity relative to mock-treated controls, which were always processed in parallel. This assay reflects the sensitivity of galectin binding to cell surface glycans *in vitro*, when the compounds are tested comparatively. To ensure identical conditions each series was performed with aliquots of the same cell suspension. Matching Gal-4 specificity, two lines with an abundance of LacNAc-presenting N-glycans were selected, that is, human pancreatic carcinoma cells expressing the tumour suppressor p16<sup>INK4a</sup> and the Lec2 mutant of the Chinese hamster ovary (CHO) cell panel.

In each case, titrations with lectin at  $2 \times 10^5$  cells per assay identified the linear range of the response, as in the solid-phase assays, to enable the work to be carried out with optimal sensitivity. Instead of the OD-value, this type of analysis provides information on number of positive cells and the mean fluorescence intensity.

**Table 2**  
IC<sub>50</sub>-values (in mM) of the test compounds for blocking binding of labelled galectins to the glycans of the surface-immobilized glycoprotein asialofetuin<sup>a</sup>

Compound	Gal-4 (5 µg/mL)	Gal-4V (5 µg/mL)	Gal-4P (2 µg/mL)	CG-1A (3 µg/mL)
<b>1</b> <sup>b</sup>	0.08	n.d.	n.d.	n.d.
<b>2</b>	0.05	0.5	0.4	0.1
<b>3</b>	0.08	0.6	0.6	0.2
<b>4</b>	0.12	0.4	0.5	0.14
<b>5</b>	n.d.	n.d.	n.d.	n.d.
<b>6</b>	0.1	0.4	0.4	0.12
<b>7</b>	n.d.	n.d.	n.d.	n.d.
<b>8</b>	0.008	0.2	0.3	0.2
<b>9</b>	0.01	0.1	0.3	0.3
<b>10</b>	0.05	0.3	0.4	0.4
Lac	1.6 (2.5 <sup>b</sup> )	3	6	0.3

n.d.: not determined.

<sup>a</sup> Given as concentration of sugar (not test compound).

<sup>b</sup> From Ref. 4a.

Gal-4 bound to the human cells in a carbohydrate-dependent manner (Fig. 7A). As exemplarily shown for compound **2** and also for compound **5**, lactose as a headgroup in glycoclusters was much more active than lactose free in solution (Fig. 7B). The direct comparison between the compounds with terephthalic acid versus the 2,6-naphthalic acid cores showed that there is no major difference between these scaffolds, also reflecting their rather similar inter-headgroup orientation described above. A grading of activities was seen for the trisaccharide-presenting clusters **7**, **9** and **10** (Fig. 7C). The tertiary terephthalamide **7** proved very active, as also seen for the ditriazole **9**, and, fully in line with a rather low potency of the lactose-presenting secondary terephthalamide in the solid-phase assay,<sup>4a</sup> compound **10** was found to be the least active. Its inhibitory capacity was in the range of that for more suited scaffolds **2** and **5** presenting lactose instead of the trisaccharide (Fig. 7B and C).

When testing labelled Gal-4V, inhibition by free sugars corroborated the comparatively reduced level of reactivity to the trisaccharide mentioned above (Fig. 7D). The same applied to the reactivity to the clusters **2** and **5** (Fig. 7E and F). Despite a 3-fold higher sugar concentration the extent of inhibition was less for Gal-4V than for Gal-4 binding (Fig. 7B and C versus Fig. 7E and F), as seen in the solid-phase assays (Table 2). As similarly seen for the second cell line (Lec2 CHO) and for labelled Gal-4P (not shown), results in both assay types consistently indicated that the linker has a bearing on the lectin's reactivity to bivalent glycoclusters and to the trisaccharide (2'-fucosyl lactose) headgroup. Evidently, this peptide portion of Gal-4 does more than simply connecting the two CRDs, and there is corroborating evidence for this assumption. Having measured the diffusion constant of the proteins, the (counterintuitive) decrease from  $1.02 \pm 0.01$  to  $0.92 \pm 0.01 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$  by truncation (reducing the molecular mass), along with acquisition of lactose dependence of this parameter,<sup>18</sup> argues in favour of an active role of the linker, above a certain length limit, in making Gal-4 structurally somehow compact. In this spatial form, probably with some flexibility to adapt in the presence of bi- to oligovalent ligands, the lectin is more sensitive to the test compounds than its variants. Headgroup distance in the tested range, which covers the distances in biantennary N-glycans, altered sensitivity for Gal-4 to a limited extent, while invariably being less active for Gal-4 after truncation.

Whether a different type of protein processing will also have a bearing on sensitivity was tested in a second system. Galectin-3 is special among galectins due to its N-terminal stalk with the two sites for Ser phosphorylation and the collagen-like repeats, this section connected as tail to the CRD. It is responsible for oligomerization in the presence of polyvalent ligands and high-affinity binding to clustered counterreceptors.<sup>19</sup> In this case, too, removal

of a protein part, here the tail, a natural proteolytic process, alters the sensitivity. As shown in Figure 8 for experiments with the wild-type CHO line, CG-3 was much more reactive with test compounds than the truncated version, extending previous observations with, for example, a bivalent glycocluster with triazole linkers that had a more than 2-fold higher capacity to interfere with CG-3 binding.<sup>4d</sup> The differences noted in the cytofluorimetric assays were in this range (Fig. 8B and D).

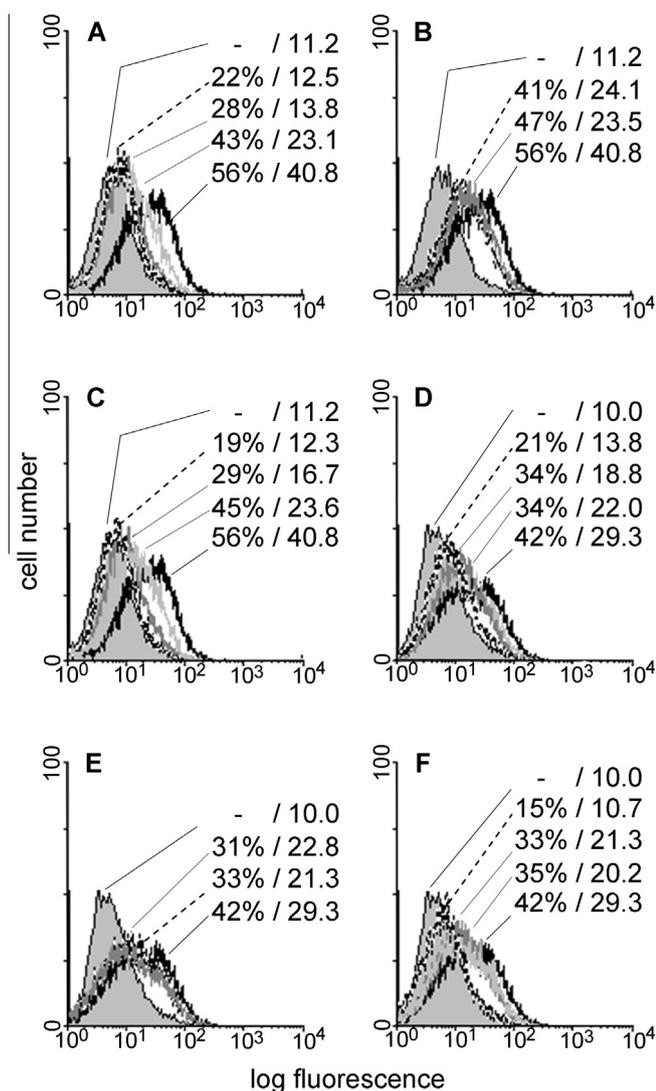
### 3. Conclusions

Two factors combine to establish the reactivity of a glycan to a lectin: the structure of the cognate epitope and its spatial presentation.<sup>1</sup> To attain the required high-level specificity and selectivity in counterreceptor selection in vivo the way sites for contact are positioned in lectins is assumed to matter markedly, too. If this is true, glycoclusters should sense engineered changes in lectins, serving as tools for structure–activity correlations. Our data using two engineered Gal-4 variants and the test panel of bivalent glycoclusters revealed special properties for the protein with the natural linker length. Its reduction decreased the susceptibility to glycoclusters in the two binding assays and the reactivity to the H-type trisaccharide, encouraging further testing, e.g. in agglutination assays using glycodendrimerosomes.<sup>20</sup> Also considering the results with CG-3, this approach, teaming up protein engineering with glycocluster synthesis and testing, is thus illustrated to be useful on the way to eventually resolve the question on the physiological significance of the length of sequence extension (Gal-3) or linker (Gal-4) in endogenous lectins, for galectins and other lectin families sharing these types of modular design.

### 4. Materials and methods

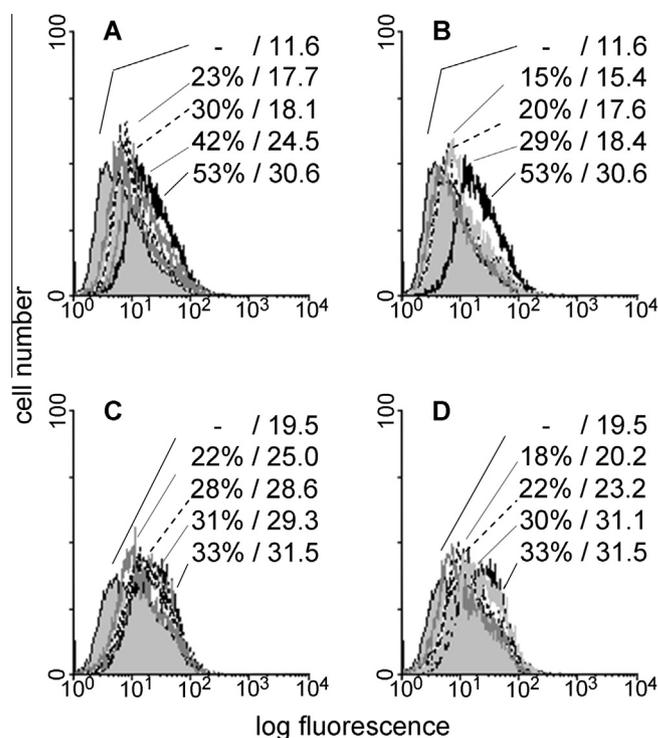
#### 4.1. General experimental

Unless otherwise noted, all commercially available compounds were used as provided without further purification. Solvents for chromatography were technical grade. Petroleum ether 40–60 °C was used for column chromatography and thin-layer chromatography (TLC). NMR spectra were recorded (25 °C). The frequency was 500 MHz for <sup>1</sup>H NMR and 125 MHz for <sup>13</sup>C NMR. Data are reported in the following order: chemical shift ( $\delta$ ) in ppm; multiplicities are indicated s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet); coupling constants (*J*) are given in Hertz (Hz). Chemical shifts are reported relative to internal Me<sub>4</sub>Si in CDCl<sub>3</sub> ( $\delta$  0.0) or HOD for D<sub>2</sub>O ( $\delta$  4.79, 25 °C) for <sup>1</sup>H and Me<sub>4</sub>Si in CDCl<sub>3</sub> ( $\delta$  0.0) or CDCl<sub>3</sub> ( $\delta$  77.0) for <sup>13</sup>C. <sup>1</sup>H NMR signals were assigned with the aid of COSY,



**Figure 7.** Semilogarithmic representation of fluorescent surface staining of human pancreatic carcinoma cells (Capan-1), reconstituted for the expression of the tumour suppressor p16<sup>INK4a</sup>, by labelled Gal-4 (10 μg/mL; A–C) and its Gal-4V variant (5 μg/mL; D–F). The control value (background) for the signal obtained by processing cells with the fluorescent indicator in the absence of the lectin is drawn as grey-shaded area of the respective scan data, the 100%-value (lectin-dependent staining in the absence of a test compound) as thick black line. The measured results on staining (percentage of positive cells/mean fluorescence intensity) are given in each panel in the order of listing the lactose concentration/type of compound (from top to bottom), all concentrations in sugar (not bivalent scaffold). In detail, the top number defines the control, the next pairs the test cases in the order of listing in the text and finally, at the bottom, the 100%-value. Inhibition of Gal-4-dependent staining by 1 mM/0.5 mM lactose and 50 μM fucosyl lactoside (A), by 40 μM sugar presented by compounds 2/5 (B) and by compounds 7, 9 and 10 (C) relative to the 100% control. Inhibition of Gal-4V-dependent staining by 1 mM/0.5 mM lactose and 120 μM fucosyl lactoside (D), by 120 μM sugar presented by compounds 5/2 (E) and by 120 μM compounds 7, 9 and 10 (F) relative to the 100% control.

<sup>13</sup>C NMR signals using DEPT, gHSQCAD and/or gHMBCAD. Low- and high-resolution mass spectra were in positive and/or negative mode, as indicated in each case. TLC was performed on aluminium sheets precoated with silica gel and spots visualized by UV and charring with H<sub>2</sub>SO<sub>4</sub>–EtOH (1:20), or cerium molybdate. Flash chromatography was carried out with silica gel 60 (0.040–0.630 mm) and using a stepwise solvent polarity gradient correlated with TLC mobility. CH<sub>2</sub>Cl<sub>2</sub>, MeOH, toluene and THF reaction solvents were used as obtained from a Pure Solv™ Solvent



**Figure 8.** Semilogarithmic representation of fluorescent surface staining of parental CHO cells by labelled CG-3 (5 μg/mL; A and B) and proteolytically truncated CG-3 (2 μg/mL; C and D); for further details, please see legend of Figure 7. Inhibition of CG-3-dependent staining by 10 mM lactose, 0.5 mM fucosyl lactoside and 0.5 mM lactose (A) and by 0.5 mM sugar presented by compounds 8, 9 and 7 (B). Inhibition of staining by truncated CG-3 by 10 mM lactose, 0.5 mM fucosyl lactoside and 2 mM lactose (C) and by 0.5 mM sugar presented by compounds 8, 9 and 7 (D).

Purification System. Anhydrous DMF, pyridine and EtOH were used as purchased from commercial suppliers.

## 4.2. Synthetic procedures

### 4.2.1. Typical procedure for the Zemplén deacetylation

The acetylated compound (0.02 mmol) was dissolved in methanol (5 mL), a catalytic amount of NaOMe (0.1 mL of a 0.2 M solution in MeOH) was added and the resulting mixture was stirred for 1 h at room temperature. Amberlite IR-120 (plus) was added, the mixture was neutralized and the resin was then removed by filtration and washed with water. Finally, the solvents were removed under diminished pressure and the residue subjected to subsequent chromatographic purification.

### 4.2.2. *N,N'*-Di(β-D-galactopyranosyl-(1→4)-β-D-glucopyranosyl)-*N,N'*-di[(*t*-butylcarbamoyl)-methyl]terephthalamide 2

Terephthalic acid (13.6 mg, 0.08 mmol), lactosyl amine 1 (100 mg, 0.16 mmol) and formaldehyde (16 μL of a 37% solution, 0.19 mmol) were suspended in MeOH (5 mL) and the mixture was stirred at room temperature for 1 h. *t*-Butyl isocyanide (19 μL, 0.16 mmol) was then added and the mixture was stirred at room temperature overnight. The reaction was heated to 45 °C for 10 h and then solvent was then removed under reduced pressure. Chromatography of the residue (CH<sub>2</sub>Cl<sub>2</sub>–CH<sub>3</sub>OH, gradient elution, 70:1 to 60:1 to 50:1) gave the protected intermediate as a white amorphous solid (98 mg, 77%). Zemplén deacetylation and subsequent purification using a BioGel P-2 gel column (eluent/H<sub>2</sub>O) and subsequent C18 reverse-phase column (gradient elution, H<sub>2</sub>O to H<sub>2</sub>O–CH<sub>3</sub>OH, 95:5) gave 2 as an interconverting mixture of

*EE* and *EZ* isomers (4:1);  $[\alpha]_D^{20} +28.0$  (c 0.1, D<sub>2</sub>O); <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O) data for *EE* isomer  $\delta$  7.70 (s, 4H), 4.84 (d,  $J = 8.8$  Hz, 2H, H-1), 4.44 (d,  $J = 7.8$  Hz, 2H, H-1'), 4.20 (ABq,  $J = 16.3$  Hz, 4H), 3.99 (d,  $J = 10.9$  Hz, 2H), 3.92 (d,  $J = 3.3$  Hz, 2H), 3.86–3.63 (m, 14H), 3.60–3.48 (m, 6H), 1.38 (s, 18H); selected <sup>1</sup>H NMR data for *EZ* isomer  $\delta$  7.68 (d,  $J = 7.8$  Hz, 2H), 7.62 (d,  $J = 7.8$  Hz, 2H), 5.76 (d,  $J = 9.3$  Hz, 2H, H-1), 4.50 (d,  $J = 7.8$  Hz, 2H, H-1'), 1.18 (s, 18H); <sup>13</sup>C NMR (125 MHz, D<sub>2</sub>O)  $\delta$  174.1 (C), 169.9 (C), 135.9 (C), 127.5, 102.8, 87.2, 77.5, 76.7, 75.3, 74.0, 72.4, 70.8, 69.7, 68.5 (each CH), 61.0 (CH<sub>2</sub>), 59.9 (CH<sub>2</sub>), 51.5 (C), 45.5 (CH<sub>2</sub>), 27.6 (CH<sub>3</sub>); HRMS-ESI: calcd for C<sub>44</sub>H<sub>70</sub>N<sub>4</sub>O<sub>24</sub>Na: 1061.4278; Found: 1061.4293.

#### 4.2.3. *N,N'*-Di( $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)- $\beta$ -D-glucopyranosyl)-*N,N'*-di[(1-methoxycarbonyl)-methylamino-2-oxoethyl]oxalamide **3**

Oxalic acid (7 mg, 0.08 mmol), lactosyl amine **11** (100 mg, 0.16 mmol) and formaldehyde (16  $\mu$ L of a 37% solution, 0.19 mmol) were suspended in MeOH (5 mL) and the mixture was stirred at room temperature for 1 h. Methyl isocyanacetate (15  $\mu$ L, 0.16 mmol) was then added and the mixture was stirred at room temperature overnight. The reaction was then heated to 45 °C for 12 h, solvent was removed thereafter under reduced pressure. Chromatography of the residue (CH<sub>2</sub>Cl<sub>2</sub>–CH<sub>3</sub>OH, gradient elution, 80:1 to 60:1 to 50:1) gave the protected intermediate as a white amorphous solid (84 mg, 67%). Zemplén deacetylation and purification of the product using a BioGel P-2 gel column (eluant/H<sub>2</sub>O) followed by C18 reverse-phase column (elution, H<sub>2</sub>O) generated **3** as an interconverting mixture of *EE* and *EZ* isomers (3.1:1);  $[\alpha]_D^{20} +30.8$  (c 0.5, D<sub>2</sub>O); <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O) data for *EE* isomer  $\delta$  4.98 (d,  $J = 8.7$  Hz, 2H, H-1), 4.47 (d,  $J = 7.7$  Hz, 2H, H-1'), 4.30 (ABq,  $J = 16.7$  Hz, 4H), 4.11–4.07 (m, 4H), 3.98–3.90 (m, 4H), 3.89–3.62 (m, 24H), 3.56 (dd,  $J = 9.9$ , 7.7 Hz, 2H, H-2'); selected <sup>1</sup>H NMR data for *EZ* isomer  $\delta$  5.65 (d,  $J = 9.4$  Hz, 2H); <sup>13</sup>C NMR (125 MHz, D<sub>2</sub>O)  $\delta$  171.8 (C), 170.5 (C), 165.1 (C), 102.8, 86.4, 77.3, 76.8, 75.3, 74.1, 72.5, 70.9, 69.4, 68.5 (each CH), 61.0 (CH<sub>2</sub>), 59.8 (CH<sub>2</sub>), 52.8 (CH<sub>3</sub>), 43.7 (CH<sub>2</sub>), 41.2 (CH<sub>2</sub>); HRMS-ESI: calcd for C<sub>36</sub>H<sub>58</sub>N<sub>4</sub>O<sub>28</sub>Na: 1017.3135; Found: 1017.3108.

#### 4.2.4. *N,N'*-Di( $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)- $\beta$ -D-glucopyranosyl)-*N,N'*-di[(*t*-butylcarbamoyl)-methyl]isophthalamide **4**

Isophthalic acid (13.6 mg, 0.08 mmol), lactosyl amine **1** (100 mg, 0.16 mmol) and formaldehyde (16  $\mu$ L of a 37% solution, 0.19 mmol) were suspended in MeOH (5 mL) and the mixture was stirred at room temperature for 1 h. *t*-Butyl isocyanide (19  $\mu$ L, 0.16 mmol) was then added and the mixture was stirred at room temperature overnight. The reaction was then heated to 45 °C for 12 h, followed by removing solvent under reduced pressure. Chromatography of the residue (CH<sub>2</sub>Cl<sub>2</sub>–CH<sub>3</sub>OH, gradient elution, 80:1 to 60:1 to 50:1) gave the protected intermediate as a white amorphous solid (101 mg, 79%). Zemplén deacetylation and subsequent purification using a BioGel P-2 gel column (eluent/H<sub>2</sub>O) and C18 reverse-phase column (gradient elution, H<sub>2</sub>O to H<sub>2</sub>O–CH<sub>3</sub>OH, 95:5) produced **4** as an interconverting mixture of *EE* and *EZ* isomers (4:1);  $[\alpha]_D^{20} +22.3$  (c 0.27, D<sub>2</sub>O); <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O) data for *EE* isomer  $\delta$  7.80–7.68 (m, 4H), 4.79 (d, 2H, H-1, determined by 2D-NMR), 4.43 (d,  $J = 7.8$  Hz, 2H, H-1'), 4.20 (ABq,  $J = 16.4$  Hz, 4H), 4.04–3.94 (m, 2H), 3.92 (d,  $J = 3.2$  Hz, 2H), 3.91–3.59 (m, 14H), 3.59–3.44 (m, 6H), 1.38 (s, 18H); selected <sup>1</sup>H NMR data for *EZ* isomer  $\delta$  5.75 (d,  $J = 9.3$  Hz, 2H), 4.32 (d,  $J = 7.3$  Hz, 2H), 1.18 (s, 18H); <sup>13</sup>C NMR (125 MHz, D<sub>2</sub>O)  $\delta$  174.1 (C), 169.9 (C), 134.1 (C), 130.0, 129.5, 102.8, 87.3, 77.5, 76.7, 75.3, 74.1, 72.4, 70.8, 69.7, 68.5 (each CH), 61.0 (CH<sub>2</sub>), 59.9 (CH<sub>2</sub>), 51.5 (C), 45.7 (CH<sub>2</sub>), 27.6 (CH<sub>3</sub>); HRMS-ESI: calcd for C<sub>44</sub>H<sub>70</sub>N<sub>4</sub>O<sub>24</sub>Na: 1061.4278; Found: 1061.4257.

#### 4.2.5. [*N,N'*-Di( $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)- $\beta$ -D-glucopyranosyl)-*N,N'*-di[(1-methoxycarbonyl)-methylamino-2-oxoethyl]-naphthalene-2,6-dicarboxamide **5**

Naphthalene-2,6-dicarboxylic acid (18 mg, 0.08 mmol), lactosyl amine **1** (100 mg, 0.16 mmol) and formaldehyde (16  $\mu$ L of a 37% solution, 0.19 mmol) were suspended in MeOH (5 mL) and the mixture was stirred at room temperature for 1 h. Methyl isocyanacetate (15  $\mu$ L, 0.24 mmol) was then added and the mixture was stirred at room temperature overnight. The reaction was then heated to 45 °C for 12 h. The solvent was removed thereafter under reduced pressure. Chromatography of the residue (CH<sub>2</sub>Cl<sub>2</sub>–CH<sub>3</sub>OH, gradient elution, 70:1 to 60:1 to 50:1) led to the protected intermediate as a white amorphous solid (89 mg, 66%). Zemplén deacetylation and subsequent purification using a BioGel P-2 column (eluent: H<sub>2</sub>O) and then C18 reverse-phase column (gradient elution, H<sub>2</sub>O to H<sub>2</sub>O–CH<sub>3</sub>OH, 95:5) gave **5** as an interconverting mixture of *EE* and *EZ* isomers (4:1);  $[\alpha]_D^{20} +36$  (c 0.2, D<sub>2</sub>O); <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O) data for *EE* isomer  $\delta$  8.24 (s, 2H), 8.18 (d,  $J = 8.5$  Hz, 2H), 7.74 (d,  $J = 8.6$  Hz, 2H), 4.89 (d,  $J = 8.9$  Hz, 2H, H-1), 4.54–4.36 (m, 6H), 4.17 (s, 4H), 4.01 (d,  $J = 11.9$  Hz, 2H), 3.91 (d,  $J = 3.0$  Hz, 2H), 3.88–3.58 (m, 20H), 3.53–3.38 (m, 6H); selected <sup>1</sup>H NMR data for *EZ* isomer  $\delta$  5.83 (d,  $J = 8.6$  Hz, 2H); <sup>13</sup>C NMR (125 MHz, D<sub>2</sub>O)  $\delta$  175.0 (C), 171.93(C), 171.87 (C), 132.9 (C), 132.4 (C), 129.7, 127.0, 124.7, 102.7, 87.4, 77.5, 76.9, 75.3, 74.1, 72.4, 70.8, 69.7, 68.4 (each CH), 60.9 (CH<sub>2</sub>), 60.0 (CH<sub>2</sub>), 52.8 (CH<sub>3</sub>), 44.7(CH<sub>2</sub>), 41.3 (CH<sub>2</sub>); HRMS-ESI: calcd for C<sub>46</sub>H<sub>64</sub>N<sub>4</sub>O<sub>28</sub>Na: 1143.3605; Found: 1143.3601.

#### 4.2.6. *N,N'*-Di( $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4)- $\beta$ -D-glucopyranosyl)-*N,N'*-di[(1-methoxycarbonyl)-methylamino-2-oxoethyl]isophthalamide **6**

Isophthalic acid (20 mg, 0.12 mmol), lactosyl amine **1**<sup>21</sup> (152 mg, 0.24 mmol) and formaldehyde (23  $\mu$ L of a 37% solution, 0.29 mmol) were suspended in MeOH (5 mL) and the mixture was stirred at room temperature for 1 h. Methyl isocyanacetate (23  $\mu$ L, 0.24 mmol) was then added and the mixture was stirred at room temperature overnight. The reaction was then heated to 45 °C for 12 h, solvent removed thereafter under reduced pressure. Chromatography of the residue (CH<sub>2</sub>Cl<sub>2</sub>–CH<sub>3</sub>OH, gradient elution, 70:1 to 60:1 to 50:1) led to the protected intermediate as a white amorphous solid (129 mg, 65%). Zemplén deacetylation of this intermediate and purification by BioGel P-2 gel column (eluent/H<sub>2</sub>O) and then C18 reverse-phase chromatography (gradient elution, H<sub>2</sub>O to H<sub>2</sub>O–CH<sub>3</sub>OH, 95:5) generated **2** as an interconverting mixture of *EE* and *EZ* isomers (3.75:1);  $[\alpha]_D^{20} +27.3$  (c 0.6, D<sub>2</sub>O); <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O) data for *EE* isomer  $\delta$  7.79 (s, 1H), 7.78 (s, 2H), 7.75–7.71 (m, 1H), 4.83 (d,  $J = 8.8$  Hz, 2H, H-1), 4.43 (d,  $J = 7.6$  Hz, 2H, H-1'), 4.38 (d,  $J = 11.9$  Hz, 4H), 4.14 (d,  $J = 3.5$  Hz, 4H), 3.97 (d,  $J = 12.7$  Hz, 2H), 3.93 (d,  $J = 3.3$  Hz, 2H), 3.86–3.63 (m, 20H), 3.57–3.49 (m, 6H); selected <sup>1</sup>H NMR data for *EZ* isomer  $\delta$  5.77 (d,  $J = 9.5$  Hz, 2H), 4.42 (d,  $J = 7.2$  Hz, 2H); <sup>13</sup>C NMR (125 MHz, D<sub>2</sub>O)  $\delta$  174.0 (C), 171.8 (C), 134.0 (C), 130.0, 129.7, 102.8, 87.2, 77.5, 76.7, 75.3, 74.1, 72.4, 70.8, 69.7, 68.5 (each CH), 61.0 (CH<sub>2</sub>), 59.9 (CH<sub>2</sub>), 52.8(CH<sub>3</sub>), 44.8 (CH<sub>2</sub>), 41.3 (CH<sub>2</sub>); HRMS-ESI: calcd for C<sub>42</sub>H<sub>62</sub>N<sub>4</sub>O<sub>28</sub>Na: 1093.3448; Found: 1093.3466.

#### 4.2.7. *O*-(2,3,4-Tri-*O*-benzyl- $\alpha$ -L-fucopyranosyl)-(1 $\rightarrow$ 2)-*O*-(3,4,6-tri-*O*-acetyl- $\beta$ -D-galactopyranosyl)-(1 $\rightarrow$ 4)-1,2,3,6-tetra-*O*-acetyl- $\beta$ -D-glucopyranosyl amine **18**

Compound **16** (450 mg, 0.59 mmol) was dissolved in CH<sub>3</sub>OH–H<sub>2</sub>O (30 mL, 2:1) and treated with an excess of ammonium hydrogen carbonate for 7 days at 30 °C. The solution was concentrated to half its original volume and diluted with water. This procedure was repeated twice, followed by removing water. Toluene was then evaporated from the residue (25 mL  $\times$  3 times) and the residue containing **17** was suspended in pyridine (25 mL), Fmoc-OSu

(211 mg, 0.65 mmol) was added and the mixture was stirred overnight at ambient temperature. Acetic anhydride (12 mL) was then added and the mixture was stirred for another 12 h. After concentration and co-evaporation with toluene, the residue was purified by chromatography (EtOAc-PE, gradient elution, 5:1 to 2:1) to give a colourless oil. This oil was dissolved in DMF (5 mL) and morpholine (5 mL) was added. After 25 min the solution was diluted with toluene (10 mL) and concentrated. The residue was purified by chromatography (CH<sub>2</sub>Cl<sub>2</sub>-CH<sub>3</sub>OH, gradient elution, 80:1 to 70:1) to lead to **18** as a colourless oil (237 mg, 40% for three steps), *R*<sub>f</sub> 0.33 (CH<sub>2</sub>Cl<sub>2</sub>-CH<sub>3</sub>OH, 40:1). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 7.35–7.24 (m, 15H), 5.32 (d, *J* = 2.5 Hz, 1H), 5.22 (d, *J* = 3.4 Hz, 1H, H-1''), 5.13 (t, *J* = 9.5 Hz, 1H, H-3), 5.05 (dd, *J* = 10.0, 3.4 Hz, 1H), 4.97 (d, *J* = 11.6 Hz, 1H), 4.75 (t, *J* = 9.5 Hz, 1H, H-2), 4.72–4.63 (m, 5H), 4.49 (dd, *J* = 1.9, 11.5 Hz, 1H), 4.38 (d, *J* = 7.6 Hz, 1H, H-1'), 4.18–4.02 (m, 6H), 3.85–3.78 (m, 4H), 3.69 (d, *J* = 1.9 Hz, 1H), 3.55–3.51 (m, 1H), 2.08–2.07 (4s, 12H), 2.04 (s, 3H), 1.79 (s, 3H), 1.19 (d, *J* = 6.4 Hz, 3H); <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>): 170.4, 170.34, 170.30, 170.0, 169.9, 169.8, 138.7, 138.6, 138.4 (each C), 128.4, 128.3, 128.2, 127.7, 127.6, 127.54, 127.50, 127.3, 100.9, 97.9, 84.9, 79.4, 77.7, 76.3 (each CH), 74.8 (CH<sub>2</sub>), 74.0 (CH), 73.8 (CH), 73.4 (CH<sub>2</sub>), 73.1 (CH), 73.0 (CH), 72.9 (CH<sub>2</sub>), 72.6 (CH), 72.1 (CH), 70.5(CH), 67.22 (CH), 67.20 (CH), 62.5 (CH<sub>2</sub>), 61.0 (CH<sub>2</sub>), 20.91, 20.88, 20.8, 20.7, 20.6, 16.5 (each CH<sub>3</sub>). HRMS-ESI: calcd for C<sub>51</sub>H<sub>64</sub>NO<sub>20</sub>Na: 1032.3841; Found: 1032.3837

#### 4.2.8. *N,N*-Di[α-*L*-fucopyranosyl-(1→2)-β-*D*-galactopyranosyl-(1→4)-β-*D*-glucopyranosyl]-*N,N*-di[(1-methoxycarbonyl)methylamino-2-oxoethyl]terephthalamide **7**

Terephthalic acid (5 mg, 32.2 μmol), formaldehyde (37%, 6 μL, 77.3 μmol) and **18** (65 mg, 64.4 μmol) were suspended in anhydrous methanol. After 1 h, the methyl isocyanacetate (9 μL, 96.6 μmol) was added and the reaction was allowed to stirred at rt for 4 h and 45 °C for 24 h. Then solvent was removed under reduced pressure. Chromatography of the residue (CH<sub>2</sub>Cl<sub>2</sub>-CH<sub>3</sub>OH, gradient elution, 100:1 to 70:1 to 55:1) gave the protected intermediate as a white amorphous solid (19 mg, 24%). Zemplén deacetylation and subsequent C18 reverse-phase chromatography (H<sub>2</sub>O-CH<sub>3</sub>OH, gradient elution, 1:1 to 1:2 to 1:3 to 1:4) were carried out to produce the benzylated intermediate. Then the residue was dissolved in methanol, to which 10% Pd-C was added. The mixture was stirred under an atmosphere of hydrogen for 24 h at ambient temperature. When the reaction was completed, the mixture was filtered over celite and concentrated. Reverse-phase chromatography using a C-18 column (H<sub>2</sub>O-CH<sub>3</sub>OH, gradient elution, 1:0 to 98:2 to 97:3) gave **7** as an amorphous solid as an interconverting mixture of *EE* and *EZ* isomers (3:1); [α]<sub>D</sub><sup>20</sup> –28.0 (c 0.1, D<sub>2</sub>O); <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O) data for *EE* isomer δ 7.73 (s, 4H), 5.27 (d, *J* = 3.4 Hz, 2H, H-1''), 4.74 (d, *J* = 8.9 Hz, 2H, H-1), 4.51 (d, *J* = 7.8 Hz, 2H, H-1'), 4.39 (ABq, *J* = 16.6 Hz, 4H), 4.14 (m, 6H), 3.98 (d, *J* = 11.6 Hz, 2H, H-6a), 3.93–3.68 (m, 28H), 3.64 (*J* = 8.5 Hz, 2H), 3.47 (t, *J* = 9.1 Hz, 2H), 3.30–3.28 (m, 2H), 1.08 (d, *J* = 6.6 Hz, 6H); selected <sup>1</sup>H NMR data for *EZ* isomer δ 7.70 (d, *J* = 7.5 Hz, 2H), 7.63 (d, *J* = 7.5 Hz, 2H), 5.74 (d, *J* = 7.7 Hz, 2H, H-1), 5.35 (bs, 2H, H-1''), 4.57 (d, *J* = 7.5 Hz, 2H, H-1'); <sup>13</sup>C NMR (125 MHz, D<sub>2</sub>O) δ 174.0 (C), 171.8 (C), 171.7 (C), 136.0 (C), 127.5, 100.1, 99.5, 87.6, 77.5, 76.6, 75.2, 75.1, 74.2, 73.4, 71.6, 69.7, 69.5, 69.0, 68.2, 66.8 (each CH), 61.1(CH<sub>2</sub>), 60.1(CH<sub>2</sub>), 52.8 (CH<sub>3</sub>), 44.7 (CH<sub>2</sub>), 41.3 (CH<sub>2</sub>), 15.3 (CH<sub>3</sub>); HRMS-ESI: calcd for C<sub>54</sub>H<sub>82</sub>N<sub>4</sub>O<sub>36</sub>Na: 1385.4606; Found: 1385.4608.

#### 4.2.9. 1,4-Di[α-*L*-fucopyranosyl-(1→2)-β-*D*-galactopyranosyl-(1→4)-β-*D*-glucopyranosyl-1,2,3-triazol-4-yl]benzene **8**

Compound **12** (130 mg, 145.9 μmol) was dissolved in CH<sub>3</sub>OH-H<sub>2</sub>O (2:1, 12 mL), then 1,4-diethynyl benzene (9.6 mg, 73.0 μmol), sodium ascorbate (5.8 mg dissolved in 1 mL H<sub>2</sub>O, 29.2 mol) and

CuSO<sub>4</sub> (2.4 mg dissolved in 1 mL H<sub>2</sub>O, 14.6 μmol) were subsequently added and the mixture was stirred overnight, after which the solvent was removed and the residue was precipitated by CH<sub>2</sub>Cl<sub>2</sub> (50 mL) and water (15 mL). The organic phase was washed by water (15 mL × 2), dried by Na<sub>2</sub>SO<sub>4</sub> and concentrated. The crude residue was purified by flash chromatography (EtOAc-PE, gradient elution, 2:1 to 2.5:1) to give a white foam (122 mg, 88%); The protecting groups were removed from the peracetylated intermediate (41 mg, 0.021 mmol) by the Zemplén procedure to give **8** as a white amorphous solid (17 mg, 68%) after preparative reverse-phase HPLC (isocratic elution with water-CH<sub>3</sub>CN, 91:9, flow rate 10 mL/min) and lyophilization; [α]<sub>D</sub><sup>20</sup> –56.0 (c 0.1, D<sub>2</sub>O); <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O) δ 8.60 (s, 2H), 7.87 (s, 4H), 5.81 (d, *J* = 9.1 Hz, 2H, H-1), 5.35 (d, *J* = 1.9 Hz, 2H, H-1''), 4.61 (d, *J* = 7.6 Hz, 2H, H-1'), 4.27 (dd, *J* = 12.6, 6.0 Hz, 2H, H-5''), 4.13 (t, *J* = 9.2 Hz, 2H, H-2), 4.04–4.00 (m, 4H), 3.95–3.70 (m, 24H), 1.30 (d, *J* = 6.3 Hz, 6H); <sup>13</sup>C NMR (125 MHz, D<sub>2</sub>O) δ 147.1 (C), 129.4 (C), 126.3, 121.3, 100.3, 99.4, 87.4, 78.1, 76.4, 75.2, 74.9, 74.4, 73.5, 72.1, 71.6, 69.6, 69.1, 68.2, 66.9 (each CH), 61.1 (CH<sub>2</sub>), 59.8 (CH<sub>2</sub>), 15.3 (CH<sub>3</sub>); HRMS-ESI: calcd for C<sub>46</sub>H<sub>68</sub>N<sub>6</sub>O<sub>28</sub>Na: 1175.3979; Found: 1175.3937.

#### 4.2.10. 1,4-Di[α-*L*-fucopyranosyl-(1→2)-β-*D*-galactopyranosyl-(1→4)-β-*D*-glucopyranosyl]-1,2,3-triazol-4-ylmethylbenzene **9**

Compound **12** (90 mg, 0.10 mmol) was dissolved in CH<sub>3</sub>OH-H<sub>2</sub>O (2:1, 6 mL), then *p*-bispropargyloxybenzene<sup>22</sup> **15** (9.4 mg, 0.05 mmol), sodium ascorbate (4 mg dissolved in 1 mL H<sub>2</sub>O, 20 μmol) and CuSO<sub>4</sub> (1.6 mg dissolved in 1 mL H<sub>2</sub>O, 10 μmol) were subsequently added and the mixture was stirred overnight, after which the solvent was removed and the residue was participated by CH<sub>2</sub>Cl<sub>2</sub> (50 mL) and water (15 mL). The organic phase was washed by water (15 mL × 2), dried by Na<sub>2</sub>SO<sub>4</sub> and concentrated. Flash silica gel chromatography (EtOAc-PE, gradient elution, 2:1 to 2.5:1) gave the dimeric intermediate as a white foam (91 mg, 92%), *R*<sub>f</sub> 0.50 (PE-EtOAc, 1:4); [α]<sub>D</sub><sup>20</sup> –75.0 (c 1.0, CHCl<sub>3</sub>); <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 7.80 (s, 2H), 6.92 (s, 4H), 5.84 (d, *J* = 9.3 Hz, 2H, H-1), 5.45 (t, *J* = 9.5 Hz, 2H), 5.40 (d, *J* = 3.8 Hz, 2H, H-1''), 5.37–5.30 (m, 6H), 5.18–5.15 (m, 6H), 5.02–4.97 (m, 4H), 4.51 (d, *J* = 12.2 Hz, 2H), 4.46 (d, *J* = 7.6 Hz, 2H, H-1'), 4.41 (q, *J* = 6.5 Hz, 2H), 4.32 (dd, *J* = 12.2, 5.5 Hz, 2H), 4.17 (dd, *J* = 11.2, 6.5 Hz, 2H), 4.10 (dd, *J* = 11.2, 7.0 Hz, 2H), 4.03–3.97 (m, 4H), 3.92–3.84 (m, 4H), 2.17 (s, 6H), 2.13 (2s, 12H), 2.10 (s, 6H), 2.08 (s, 6H), 2.00 (s, 6H), 1.99 (s, 6H), 1.97 (s, 6H), 1.87 (s, 6H), 1.24 (d, *J* = 6.5 Hz, 6H); <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>): δ 170.7, 170.6, 170.5, 170.3, 170.1, 169.9, 169.8, 169.7, 168.9, 152.8, 145.0 (each C), 121.2, 115.9, 100.2, 95.6, 85.8, 76.3, 73.8, 73.4, 71.8, 71.5, 71.0, 70.9, 70.2, 68.0, 67.3, 67.0, 65.0 (each CH), 62.6, 62.1, 60.8 (each CH<sub>2</sub>), 20.8, 20.70, 20.68, 20.66, 20.65, 20.63, 20.60, 20.2, 15.6 (each CH<sub>3</sub>); LRMS (ESI) 1991.5 (M+Na<sup>+</sup>); HRMS-ESI: calcd for C<sub>84</sub>H<sub>108</sub>N<sub>6</sub>O<sub>48</sub> Na: 1991.6092; Found: 1991.6145. This intermediate (37 mg, 0.019 mmol) was dissolved in methanol (5 mL) to which a catalytic amount of NaOMe (0.1 mL of a 0.2 M solution in MeOH) was added and the resulting solution was stirred for 1 h at room temperature. Amberlite IR-120 (plus) was added to neutralize pH = 7, after which the resin was removed by filtration and washed with water. The solvent removed under diminished pressure to give **9** after lyophilization, as an amorphous solid (20 mg, 88%). [α]<sub>D</sub><sup>20</sup> –63.3 (c 0.12, D<sub>2</sub>O); <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O) δ 8.32 (s, 2H), 7.06 (s, 4H), 5.79 (d, *J* = 9.2 Hz, 2H, H-1), 5.35 (d, *J* = 3.2 Hz, 2H, H-1''), 5.27 (s, 4H), 4.60 (d, *J* = 7.7 Hz, 2H, H-1'), 4.26 (dd, *J* = 13.1, 6.5 Hz, 2H, H-5''), 4.08 (t, *J* = 9.3 Hz, 2H), 4.00–3.97 (m, 4H), 3.92–3.89 (m, 4H), 3.87–3.71 (m, 20H), 1.28 (d, *J* = 6.6 Hz, 6H); <sup>13</sup>C NMR (125 MHz, D<sub>2</sub>O) δ 152.2 (C), 143.7 (C), 124.4, 116.9, 100.2, 99.3, 87.3, 78.1, 76.3, 75.2, 74.9, 74.4, 73.5, 72.0, 71.6, 69.6, 69.1, 68.1, 66.9 (each

CH), 62.0, 61.1, 59.7 (each CH<sub>2</sub>), 15.3 (CH<sub>3</sub>); HRMS-ESI: calcd for C<sub>48</sub>H<sub>72</sub>N<sub>6</sub>O<sub>30</sub> Na: 1235.4191; Found: 1235.4204.

#### 4.2.11. O-(2,3,4-Tri-O-acetyl- $\alpha$ -L-fucopyranosyl)-(1 $\rightarrow$ 2)-O-(3,4,6-tri-O-acetyl- $\beta$ -D-galactopyranosyl)-(1 $\rightarrow$ 4)-1,2,3,6-tetra-O-acetyl-D-glucopyranosyl amine 13

The azide **12** (1.2 g, 1.35 mmol) was dissolved in EtOAc to which was added 10% Pd-C (0.1 g). The reaction was left to stir overnight under H<sub>2</sub>. Then the reaction was diluted with EtOAc and filtered through Celite. Removal of the solvent gave **13** as a white foam (1.16 g, >95% yield, mixture of anomers,  $\beta$ : $\alpha$  = 10:1) which was used in the next step without further purification; R<sub>f</sub> 0.24, (EtOAc-CH<sub>2</sub>Cl<sub>2</sub>, 9:1); <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) data for the  $\beta$ -anomer:  $\delta$  5.38 (d, *J* = 3.9 Hz, 1H), 5.29 (dd, *J* = 10.3, 2.8 Hz, 2H), 5.18–5.13 (m, 2H), 5.00–4.96 (m, 2H), 4.78 (t, *J* = 9.4 Hz, 1H), 4.48 (dd, *J* = 11.9, 1.8 Hz, 1H), 4.45–4.40 (m, 2H), 4.24 (dd, *J* = 12.0, 6.1 Hz, 1H), 4.17–4.11 (m, 3H), 4.10–4.06 (m, 1H), 3.86–3.80 (m, 2H), 3.78 (t, *J* = 9.5 Hz, 1H), 3.64–3.61 (m, 1H), 2.16 (s, 3H), 2.15 (s, 3H), 2.12 (s, 3H), 2.08 (s, 3H), 2.07 (s, 3H), 2.06 (s, 3H), 2.00 (s, 3H), 1.98 (s, 3H), 1.97 (s, 3H), 1.22 (d, *J* = 6.5 Hz, 3H). Selected <sup>1</sup>H NMR data for the  $\alpha$  anomer:  $\delta$  5.35 (d, *J* = 3.0 Hz, 1H), 5.23 (t, *J* = 9.4 Hz, 1H), 4.73 (t, *J* = 9.4 Hz, 1H), 4.28 (dd, *J* = 12.0, 5.5 Hz, 1H); <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) for the  $\beta$ -anomer:  $\delta$  170.74, 170.70, 170.6, 170.34, 170.30, 170.1, 170.03, 170.01, 169.67 (each C), 100.2, 95.5, 85.0, 74.7, 74.0, 73.5, 72.4, 72.0, 71.4, 71.1, 70.7, 68.1, 67.5, 67.1, 64.8 (each CH), 62.7 (CH<sub>2</sub>), 60.9 (CH<sub>2</sub>), 20.92, 20.86, 20.8, 20.67, 20.66, 20.65, 20.63, 20.60, 20.59, 15.53 (each CH<sub>3</sub>); LRMS (ESI) 866.2 [M+H]<sup>+</sup>; HRMS-ESI: calcd for C<sub>36</sub>H<sub>53</sub>N<sub>1</sub>O<sub>23</sub>: 866.2921; Found: 866.2930.

#### 4.2.12. N,N-Di( $\alpha$ -L-fucopyranosyl-(1 $\rightarrow$ 2))- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 4))- $\beta$ -D-glucopyranosyl terephthalamide 10

The lactosyl amine **13** (112 mg, 0.13 mmol) and DIPEA (0.16 mL, 0.19 mmol) in dry THF (10 mL) were added dropwise at room temperature into freshly recrystallized terephthaloyl chloride (26 mg, 0.065 mmol) in dry THF. The reaction was stirred overnight and then the solvent was removed. Chromatography of the residue (EtOAc-PE, gradient elution, 1:1 to 2:1) gave the intermediate as a white amorphous solid (105 mg, 87% including trace of  $\alpha$ -anomer); R<sub>f</sub> 0.75 (PE-EtOAc, 1:4); [ $\alpha$ ]<sub>D</sub><sup>20</sup> –77.2 (c 1.1, CHCl<sub>3</sub>); <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)  $\delta$  7.82 (s, 4H), 7.03 (d, *J* = 9.0 Hz, 2H), 5.40–5.34 (m, 6H), 5.34 (d, *J* = 2.6 Hz, 2H), 5.30 (d, *J* = 3.6 Hz, 2H), 5.19 (dd, *J* = 11.0, 3.2 Hz, 2H), 5.01–4.98 (m, 6H), 4.51–4.41 (m, 6H), 4.32–4.29 (m, 2H), 4.18 (dd, *J* = 11.1, 6.6 Hz, 2H), 4.08 (dd, *J* = 11.1, 6.8 Hz, 2H), 3.89–3.84 (m, 8H), 2.17 (s, 6H), 2.14 (s, 6H), 2.13 (s, 6H), 2.11 (s, 6H), 2.10 (s, 6H), 2.05 (s, 6H), 2.00 (s, 6H), 1.984 (s, 6H), 1.976 (s, 6H), 1.25 (d, *J* = 6.7 Hz, 6H); <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>)  $\delta$  171.7, 170.7, 170.6, 170.5, 170.3, 170.2, 170.0, 169.7, 169.6, 165.9, 136.2 (each C), 127.7, 99.9, 95.7, 79.0, 74.9, 74.1, 73.4, 71.7, 71.4, 71.1, 71.0, 70.8, 68.1, 67.4, 67.0, 65.0 (each CH), 62.3 (CH<sub>2</sub>), 61.0 (CH<sub>2</sub>), 20.9, 20.78, 20.76, 20.69, 20.67, 20.66, 20.64, 20.60, 15.6 (each CH<sub>3</sub>); LRMS (ESI) 1883.4 [M+Na]<sup>+</sup>; HRMS-ESI: calcd for C<sub>80</sub>H<sub>104</sub>N<sub>2</sub>O<sub>48</sub> Na: 1883.5656; Found: 1883.5609. This intermediate (40 mg, 21.5  $\mu$ mol, including the minor  $\alpha$ -anomer) was dissolved in methanol (5 mL) to which a catalytic amount of NaOMe (0.1 mL of a 0.2 M solution in MeOH) was added and the resulting solution was stirred for 1 h at room temperature. Amberlite IR-120 (plus) was added to neutralize pH = 7, after which the resin was removed by filtration and washed with water. The solvent removed on a rotary evaporator to generate **10** as a yellow solid (23 mg, 97%). The compound was further purified by reverse-phase semi-preparative HPLC (isocratic elution with water-CH<sub>3</sub>CN, 97:3, flow rate 10 mL/min, *t*<sub>R</sub> = 22 min) and the product was obtained after lyophilization as a white solid; [ $\alpha$ ]<sub>D</sub><sup>20</sup> –41.5 (c 0.325, D<sub>2</sub>O); <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O)  $\delta$  7.96 (s, 4H), 5.35 (d, *J* = 2.7 Hz, 2H, H-1''), 5.23 (d, *J* = 9.1 Hz, 2H, H-1),

4.58 (d, *J* = 7.7 Hz, 2H, H-1'), 4.28 (dd, *J* = 13.0, 6.4 Hz, 2H), 3.99 (d, *J* = 11.5 Hz, 2H), 3.91–3.63 (m 28H), 1.29 (d, *J* = 6.5 Hz, 6H); <sup>13</sup>C NMR (125 MHz, D<sub>2</sub>O)  $\delta$  171.0 (C), 136.4 (C), 127.9, 100.2, 99.3, 79.8, 77.0, 76.2, 75.2, 75.1, 73.5, 71.6, 71.5, 69.6, 69.1, 68.1, 66.9 (each CH), 61.1 (CH<sub>2</sub>), 59.9 (CH<sub>2</sub>), 15.2 (CH<sub>3</sub>); HRMS-ESI: calcd for C<sub>44</sub>H<sub>68</sub>N<sub>2</sub>O<sub>30</sub> Na: 1127.3755; Found: 1127.3728.

### 4.3. Protein engineering, production and labelling

Starting with full-length cDNA for human Gal-4 cloned from mRNA preparation of human leukaemic KG-1 cells, the BamHI restriction site (position 769 to 774) was silenced without affecting coding using the sense primer 5'-CTGAATGGCTCGTGGGGCTCAGAGGAGAAGAAGATCACC-3' (nucleotide exchanges underlined; melting temperature >78 °C) and the antisense primer 5'-GGTGATCTTCTCTCCTCIGAGCCCCACGACCATTGAG-3' (nucleotide exchanges underlined; melting temperature >78 °C) in a modified QuikChange<sup>®</sup> site-directed mutagenesis procedure (Agilent Technologies, Böblingen, Germany). The extension reaction was performed in two steps. Briefly, two 50  $\mu$ L reaction mixtures were prepared in separate tubes containing 20 pmol either of the sense or the antisense primer, 200 ng template plasmid and 1 U PfuTurbo<sup>®</sup> DNA polymerase (Agilent Technologies). After an initial preheating step at 95 °C for 30 s, three cycles (denaturation at 95 °C for 30 s, annealing at 55 °C for 1 min, extension at 68 °C for 8 min) were run. To complete the primer-directed sequence extension 25  $\mu$ L of each tube were transferred to one tube and 1 U PfuTurbo<sup>®</sup> DNA polymerase was added. Subsequently, thermal cycling which consisted of preheating at 94 °C for 30 s and 18 cycles (denaturation at 94 °C for 30 s, annealing at 55 °C for 1 min, extension at 68 °C for 5 min) was carried out. After incubation in the presence of DpnI (10 U) at 37 °C for 2 h to digest the methylated parental DNA template, 5  $\mu$ L of the reaction mixture were used to transform XL-1-Blue electrocompetent cells. Plasmids were isolated from kanamycin-resistant colonies grown on LB agar plates and treated with BamHI to ascertain the sequence conversion. Then, the cDNAs of the N-domain extended by a nucleotide sequence encoding for a tetrapeptide stretch at its C-terminus and of the C-domain with a 13mer-peptide stretch of the linker were first amplified separately. Creating a new joining hinge of the shortened cDNA using two BamHI restriction sites has the advantage that the wild-type codons, at these positions encoding Gly and Ser, were preserved. In detail, the following primer pairs were used: the sense primer 5'-CATATGGCCTATGTCCCCGACCG-3' with an internal NdeI restriction site (underlined) and the antisense primer 5'-GGATCCTCCGATGAAGTTGATTGATTGAAGTTG-3' with an internal BamHI restriction site (underlined) for the N-domain with the nucleotide sequence for the tetrapeptide stretch of the linker and the sense primer 5'-GGATCCTGCCACC ATGGAAGGA-3' with an internal BamHI restriction site (underlined) and the antisense primer 5'-GTCGACTTAGATCTGGACAT AGG-3' with an internal SalI restriction site (underlined) in the cDNA encoding for the 13mer-peptide stretch of the linker followed by the sequence of the C-domain. The cDNAs were then propagated in the pET-Blue-1 AccepTor vector (Novagen, Bad Soden, Germany), digestion with the restriction enzymes and gel extraction led to two vector-released cDNAs, which were ligated into the pET24a expression vector (Novagen) yielding a 897 bp insert encoding the Gal-4V sequence. cDNA for Gal-4P was prepared by first amplifying the cDNAs of N- and C-domains separately and reverting the artificial ClaI restriction site at the new joining hinge to the wild-type codons at these positions. In detail, we used the following primer pairs: the sense primer 5'-CATATGGCCTATGTCCCCGACCG-3' with an internal NdeI restriction site (underlined) and the antisense primer 5'-ATCGAT GATTGATTGAAGTTGAGATCCCC-3' with an internal ClaI restriction

site (underlined) for the N-domain as well as the sense primer 5'-ATCGATGTCCATATTTCCGGAGG-3' with an internal ClaI restriction site (underlined) and the antisense primer 5'-GTCCACTTAGATCTGGACATAGG-3' with an internal Sall restriction site (underlined) for the C-domain. The cDNAs were then propagated as described above, digestion with restriction enzymes (NdeI/ClaI, ClaI/Sall) followed by elution of the 846 bp insert encoding Gal-4P from the gel. The inserted ClaI restriction site between the N- and C-domains encoding for Ile and Asp finally needed to be reversed to the wild-type codons, at these positions encoding for Asn and Pro, in a modified QuikChange® site-directed mutagenesis procedure (Agilent Technologies) using a sense primer 5'-CTGCAACTTCAATC AATCAACCCTGTGCCATATTTCCGGAGGCTG-3' (nucleotide exchanges underlined) and the antisense primer 5'-CAGGGTCCGAAATAT GGCACAGGGTGATTGATTGAAGTTGCAG-3' (nucleotide exchanges underlined). Plasmids were isolated from kanamycin-resistant colonies grown on LB agar plates and sequencing ascertained absence of any deviations. Proteins were produced in the BL21(DE3) pLysS *E. coli* strain with TB medium (Roth, Karlsruhe, Germany) at 22 °C with optimal yields of 9 mg/L for Gal-4V and 3 mg/L for Gal-4P reached with a final IPTG concentration of 75 µM. Recombinant production of the two CG-3 proteins followed the previously described protocol.<sup>10a</sup> Quality controls by one- and two-dimensional electrophoresis, gel filtration and haemagglutination to ascertain homogeneity, quaternary structure and activity as well as biotinylation under activity-preserving conditions followed by product analysis to quantify label incorporation and activity assessment were performed as described.<sup>23</sup>

#### 4.4. Solid-phase/cell assays

The plastic surface of microtiter plate wells was coated with the glycoprotein asialofetuin (0.5 µg per well in 50 µL phosphate-buffered saline) overnight at 4 °C, and further processing followed the protocol given in detail previously.<sup>4</sup> Assays were routinely done in triplicates with up to six independent series. Cell binding was measured for human pancreatic carcinoma (Capan-1) cells expressing the tumour suppressor p16<sup>INK4a</sup> (kindly provided by K. M. Detjen, Berlin, Germany)<sup>2b</sup> as well as with the wild-type (parental)/Lec2 mutant lines of the CHO system (kindly provided by P. Stanley, Bronx, USA).<sup>24</sup> Process steps and controls were performed as described previously.<sup>2b,4,25</sup> Aliquots of cell suspensions at the same passage were routinely processed at least in duplicates, with at least three independent series and standard deviations not exceeding 13.6% after normalization of data based on internal controls and background values.

#### Acknowledgements

The generous financial support by the Science Foundation Ireland (08/SRC/B1393), the EC research/ITN programs GlycoHIT (contract no. 260600), COST (CM1102) and GLYCOPHARM (contract no. 217297) and the Verein zur Förderung des biologisch-technologischen Fortschritts in der Medizin e.V. is gratefully acknowledged. We are thankful to Drs. B. Friday and B. Tab for helpful discussions.

#### References

- (a) Sharon, N. *Complex Carbohydrates. Their Chemistry, Biosynthesis, and Functions*; Addison-Wesley: London, 1975; (b) Rauvala, H. *Trends Biochem. Sci.* **1983**, *8*, 323–325; (c) *The Sugar Code. Fundamentals of glycosciences*; Gabius, H.-J., Ed.; Wiley-VCH: Weinheim, 2009; (d) Gabius, H.-J.; André, S.; Jiménez-Barbero, J.; Romero, A.; Solís, D. *Trends Biochem. Sci.* **2011**, *36*, 298–313.
- (a) Sperandio, M. *FEBS J.* **2006**, *273*, 4377–4389; (b) André, S.; Sanchez-Ruderisch, H.; Nakagawa, H.; Buchholz, M.; Kopitz, J.; Forberich, P.; Kemmner, W.; Böck, C.; Deguchi, K.; Detjen, K. M.; Wiedenmann, B.; von Knebel Doeberitz,

- M.; Gress, T. M.; Nishimura, S.-I.; Rosewicz, S.; Gabius, H.-J. *FEBS J.* **2007**, *274*, 3233–3256; (c) Sanchez-Ruderisch, H.; Fischer, C.; Detjen, K. M.; Welzel, M.; Wimmel, A.; Manning, J. C.; André, S.; Gabius, H.-J. *FEBS J.* **2010**, *277*, 3552–3563; (d) Amano, M.; Eriksson, H.; Manning, J. C.; Detjen, K. M.; André, S.; Nishimura, S.-I.; Lehtiö, J.; Gabius, H.-J. *FEBS J.* **2012**, *279*, 4062–4080; (e) Gebert, J.; Kloor, M.; Lee, J.; Lohr, M.; André, S.; Wagner, R.; Kopitz, J.; Gabius, H.-J. *Histochem. Cell Biol.* **2012**, *138*, 339–350; (f) Dawson, H.; André, S.; Karamitopoulou, E.; Zlobec, I.; Gabius, H.-J. *Anticancer Res.* **2013**, *33*, 3053–3059; (g) Scott, D. W.; Patel, R. P. *Glycobiology* **2013**, *23*, 622–633.
- (a) *Neoglycoconjugates. Preparation and Applications*; Lee, Y. C., Lee, R. T., Eds.; Academic Press: San Diego, 1994; (b) Roy, R. *Trends Glycosci. Glycotechnol.* **2003**, *15*, 291–310; (c) Chabre, Y. M.; Roy, R. *The Chemist's Way to Prepare Multivalency. In The Sugar Code. Fundamentals of glycosciences*; Gabius, H.-J., Ed.; Wiley-VCH: Weinheim, 2009; pp 53–70; (d) Oscarson, S. *The Chemist's Way to Synthesize Glycosides. In The Sugar Code. Fundamentals of glycosciences*; Gabius, H.-J., Ed.; Wiley-VCH: Weinheim, 2009; pp 31–51; (e) Murphy, P. V.; André, S.; Gabius, H.-J. *Molecules* **2013**, *18*, 4026–4053.
- (a) Leyden, R.; Velasco-Torrijos, T.; André, S.; Gouin, S.; Gabius, H.-J.; Murphy, P. V. *J. Org. Chem.* **2009**, *74*, 9010–9026; (b) André, S.; Velasco-Torrijos, T.; Leyden, R.; Gouin, S.; Tosin, M.; Murphy, P. V.; Gabius, H.-J. *Org. Biomol. Chem.* **2009**, *7*, 4715–4725; (c) André, S.; Jarikote, D. V.; Yan, D.; Vincenz, L.; Wang, G.-N.; Kaltner, H.; Murphy, P. V.; Gabius, H.-J. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 313–318; (d) Wang, G.-N.; André, S.; Gabius, H.-J.; Murphy, P. V. *Org. Biomol. Chem.* **2012**, *10*, 6893–6907.
- (a) Wu, A. M.; Wu, J. H.; Liu, J.-H.; Singh, T.; André, S.; Kaltner, H.; Gabius, H.-J. *Biochimie* **2004**, *86*, 317–326; (b) Krzeminski, M.; Singh, T.; André, S.; Lensch, M.; Wu, A. M.; Bonvin, A. M. J. J.; Gabius, H.-J. *Biochim. Biophys. Acta* **1810**, *2011*, 150–161; (c) Vokhmyanina, O. A.; Rapoport, E. M.; André, S.; Severov, V. V.; Ryzhov, I.; Pazynina, G. V.; Korchagina, E.; Gabius, H.-J.; Bovin, N. V. *Glycobiology* **2012**, *22*, 1207–1217.
- (a) Kasai, K.-i.; Hirabayashi, J. *J. Biochem.* **1996**, *119*, 1–8; (b) Cooper, D. N. W. *Biochim. Biophys. Acta* **2002**, *1572*, 209–231; (c) Kaltner, H.; Gabius, H.-J. *Histol. Histopathol.* **2012**, *27*, 397–416.
- (a) Delacour, D.; Gouyer, V.; Zanetta, J.-P.; Drobecq, H.; Leteurtre, E.; Grard, G.; Moreau-Hannedouche, O.; Maes, E.; Pons, A.; André, S.; Le Bivic, A.; Gabius, H.-J.; Manninen, A.; Simons, K.; Huet, G. *J. Cell Biol.* **2005**, *169*, 491–501; (b) Stechly, L.; Morelle, W.; Dessein, A. F.; André, S.; Grard, G.; Trinel, D.; Dejonghe, M. J.; Leteurtre, E.; Drobecq, H.; Trugnan, G.; Gabius, H.-J.; Huet, G. *Traffic* **2009**, *10*, 438–450; (c) Stancic, M.; Slijepcevic, D.; Nomden, A.; Vos, M. J.; de Jonge, J. C.; Sikkema, A. H.; Gabius, H.-J.; Hoekstra, D.; Baron, W. *Glia* **2012**, *60*, 919–935; (d) Velasco, S.; Díez-Revuelta, N.; Hernández-Iglesias, T.; Kaltner, H.; André, S.; Gabius, H.-J.; Abad-Rodríguez, J. *J. Neurochem.* **2013**, *125*, 49–62.
- (a) André, S.; Specker, D.; Bovin, N. V.; Lensch, M.; Kaltner, H.; Gabius, H.-J.; Wittmann, V. *Bioconjug. Chem.* **2009**, *20*, 1716–1728; (b) André, S.; Lahmann, M.; Gabius, H.-J.; Oscarson, S. *Mol. Pharm.* **2010**, *7*, 2270–2279; (c) André, S.; Grandjean, C.; Gautier, F. M.; Bernardi, S.; Sansone, F.; Gabius, H.-J.; Ungaro, R. *Chem. Commun.* **2011**, 6126–6128; (d) André, S.; Renaudet, O.; Bossu, I.; Dumy, P.; Gabius, H.-J. *J. Pept. Sci.* **2011**, *17*, 427–437.
- (a) Beyer, E. C.; Zweig, S. E.; Barondes, S. H. *J. Biol. Chem.* **1980**, *255*, 4236–4239; (b) Gupta, D.; Kaltner, H.; Dong, X.; Gabius, H.-J.; Brewer, C. F. *Glycobiology* **1996**, *6*, 843–849; (c) Solís, D.; Romero, A.; Kaltner, H.; Gabius, H.-J.; Díaz-Mauriño, T. *J. Biol. Chem.* **1996**, *271*, 12744–12748; (d) Varela, P. H.; Solís, D.; Díaz-Mauriño, T.; Kaltner, H.; Gabius, H.-J.; Romero, A. *J. Mol. Biol.* **1999**, *294*, 537–549; (e) Wu, A. M.; Wu, J. H.; Tsai, M.-S.; Kaltner, H.; Gabius, H.-J. *Biochem. J.* **2001**, *358*, 529–538; (f) André, S.; Kaltner, H.; Lensch, M.; Russwurm, R.; Siebert, H.-C.; Fallsehr, C.; Tajkhorshid, E.; Heck, A. J. R.; von Knebel Doeberitz, M.; Gabius, H.-J.; Kopitz, J. *Int. J. Cancer* **2005**, *114*, 46–57; (g) Göhler, A.; Büchner, C.; André, S.; Doose, S. H.; Kaltner, H.; Gabius, H.-J. *Biochimie* **2012**, *94*, 2649–2655.
- (a) Kaltner, H.; Kübler, D.; López-Merino, L.; Lohr, M.; Manning, J. C.; Lensch, M.; Seidler, J.; Lehmann, W. D.; André, S.; Solís, D.; Gabius, H.-J. *Anat. Rec.* **2011**, *294*, 427–444; (b) Yongye, A. B.; Calle, L.; Ardá, A.; Jiménez-Barbero, J.; André, S.; Gabius, H.-J.; Martínez-Mayorga, K.; Cudic, M. *Biochemistry* **2012**, *51*, 7278–7289.
- (a) Bradley, H.; Fitzpatrick, G.; Glass, W. K.; Kunz, H.; Murphy, P. V. *Org. Lett.* **2001**, *3*, 2629–2632; (b) Murphy, P. V.; Bradley, H.; Tosin, M.; Pitt, N.; Fitzpatrick, G. M.; Glass, W. K. *J. Org. Chem.* **2003**, *68*, 5693–5704.
- (a) Campo, V. L.; Carvalho, I.; Da Silva, C. H. T. P.; Schenkman, S.; Hill, L.; Nepogodiev, S. A.; Field, R. A. *Chem. Sci.* **2010**, *1*, 507–514; (b) Meldal, M.; Tornøe, C. W. *Chem. Rev.* **2008**, *108*, 2952–3015; (c) Morvan, F.; Meyer, A.; Jochum, A.; Sabin, C.; Chevolut, Y.; Imbert, A.; Praly, J. P.; Vasseur, J. J.; Souteyrand, E.; Vidal, S. *Bioconjugate Chem.* **2007**, *18*, 1637–1643; (d) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2002**, *41*, 2596–2599.
- Tornøe, C. W.; Christensen, C.; Meldal, M. *J. Org. Chem.* **2002**, *67*, 3057–3064.
- (a) Lee, Y. C. *J. Biochem.* **1997**, *121*, 818–825; (b) André, S.; Kožár, T.; Schuberth, R.; Unverzagt, C.; Kojima, S.; Gabius, H.-J. *Biochemistry* **2007**, *46*, 6984–6995; (c) André, S.; Kožár, T.; Kojima, S.; Unverzagt, C.; Gabius, H.-J. *Biol. Chem.* **2009**, *390*, 557–565.
- (a) Ideo, H.; Seko, A.; Yamashita, K. *J. Biol. Chem.* **2005**, *280*, 4730–4737; (b) Morelle, W.; Stechly, L.; André, S.; van Seuningen, I.; Porchet, N.; Gabius, H.-J.; Michalski, J.-C.; Huet, G. *Biol. Chem.* **2009**, *390*, 529–544.
- Dam, T. K.; Gabius, H.-J.; André, S.; Kaltner, H.; Lensch, M.; Brewer, C. F. *Biochemistry* **2005**, *44*, 12564–12571.
- Kopitz, J.; Ballikaya, S.; André, S.; Gabius, H.-J. *Neurochem. Res.* **2012**, *37*, 1267–1276.

18. Göhler, A.; André, S.; Kaltner, H.; Sauer, M.; Gabius, H.-J.; Doose, S. *Biophys. J.* **2010**, *98*, 3044–3053.
19. (a) Kopitz, J.; von Reitzenstein, C.; André, S.; Kaltner, H.; Uhl, J.; Ehemann, V.; Cantz, M.; Gabius, H.-J. *J. Biol. Chem.* **2001**, *276*, 35917–35923; (b) André, S.; Liu, B.; Gabius, H.-J.; Roy, R. *Org. Biomol. Chem.* **2003**, *1*, 3909–3916; (c) Ahmad, N.; Gabius, H.-J.; André, S.; Kaltner, H.; Sabesan, S.; Roy, R.; Liu, B.; Macaluso, F.; Brewer, C. F. *J. Biol. Chem.* **2004**, *279*, 10841–10847; (d) Kopitz, J.; Bergmann, M.; Gabius, H.-J. *IUBMB Life* **2010**, *62*, 624–628.
20. Percec, V.; Leowanawat, P.; Sun, H.-J.; Kulikov, O.; Nusbaum, C. D.; Tran, T. M.; Bertin, A.; Wilson, D. A.; Peterca, M.; Zhang, S.; Kamat, N. P.; Vargo, K.; Moock, D.; Johnston, E. D.; Hammer, D. A.; Pochan, D. J.; Chen, Y.; Chabre, Y. M.; Shiao, T. C.; Bergeron-Brlak, M.; André, S.; Roy, R.; Gabius, H.-J.; Heiney, P. A. *J. Am. Chem. Soc.* **2013**, *135*, 9055–9077.
21. Aravind, S.; Park, W. K. C.; Brochu, S.; Roy, R. *Tetrahedron Lett.* **1994**, *35*, 7739–7742.
22. Srinivasan, M.; Sankararaman, S.; Hopf, H.; Dix, I.; Jones, P. G. *J. Org. Chem.* **2001**, *66*, 4299–4303.
23. (a) Gabius, H.-J.; Engelhardt, R.; Cramer, F.; Bätge, R.; Nagel, G. A. *Cancer Res.* **1985**, *45*, 253–257; (b) Gabius, H.-J.; Wosgien, B.; Hendry, M.; Bardosi, A. *Histochemistry* **1991**, *95*, 269–277; (c) Purkrábková, T.; Smetana, K., Jr.; Dvoránková, B.; Holíková, Z.; Böck, C.; Lensch, M.; André, S.; Pytlík, R.; Liu, F.-T.; Klíma, J.; Smetana, K.; Motlík, J.; Gabius, H.-J. *Biol. Cell* **2003**, *95*, 535–545; (d) Habermann, F. A.; André, S.; Kaltner, H.; Kübler, D.; Sinowatz, F.; Gabius, H.-J. *Histochem. Cell Biol.* **2011**, *135*, 539–552; (e) Martín-Santamaría, S.; André, S.; Buzamet, E.; Caraballo, R.; Fernández-Cureses, G.; Morando, M.; Ribeiro, J. P.; Ramírez-Gualito, K.; de Pascual-Teresa, B.; Cañada, F. J.; Menéndez, M.; Ramström, O.; Jiménez-Barbero, J.; Solís, D.; Gabius, H.-J. *Org. Biomol. Chem.* **2011**, *9*, 5445–5455; (f) Giguère, D.; André, S.; Bonin, M.-A.; Bellefleur, M.-A.; Provencal, A.; Cloutier, P.; Pucci, B.; Roy, R.; Gabius, H.-J. *Bioorg. Med. Chem.* **2011**, *19*, 3280–3287.
24. Patnaik, S. K.; Stanley, P. *Methods Enzymol.* **2006**, *416*, 159–182.
25. Kopitz, J.; Fik, Z.; André, S.; Smetana, K., Jr.; Gabius, H.-J. *Mol. Pharm.* **2013**, *10*, 2054–2061.