# Breaking the Sugar Code: Six Levels of Affinity Regulation in Glycan-lectin Interaction

**Harold Rüdiger[1] and Hans-Joachim Gabius[2]**

[1]Institut für Pharmazie und Lebensmittelchemie, Julius-Maximilians-Universität Würzburg, Am Hubland, 97074 Würzburg, Germany

[2]Lehrstuhl für Physiologische Chemie, Tierärztliche Fakultät, Ludwig-Maximilians-Universität München, Veterinärstraße 13, 80539 München, Germany

**E-Mail:** [1]ruediger@mail.uni-wuerzburg.de and [2]gabius@tiph.vetmed.uni-muenchen.de

*Received: 18th November 2011 / Published: 11th July 2012*

## Abstract

The glycan chains of cellular glycoconjugates harbour information essential for many physiological processes. It can be decoded by counter-receptors. Among them, lectins (carbohydrate-binding proteins except antibodies, enzymes working on the cognate glycans and sensor/transport proteins for free sugars) play a prominent role in translating the sugar code. The exquisite precision, with which distinct glycoconjugates and lectins form pairs, poses the question on the underlying molecular mechanism(s) that guide(s) lectins to only few target sites ('the needles in the haystack'), and this even with cell-type specificity. To provide a detailed concept and hereby give research a clear direction, structural characteristics on the side of the glycans are identified and systematically listed. As a result, a six-level scheme is devised, moving from sequence and shape to density of presentation, in glycoconjugates and membranes. It is flanked by an illustration of the range of modes for the topological display of binding sites in human lectins, via covalent and non-covalent clustering, to document the remarkable degree of sophistication reached within this recognition system. The discovery of orchestration of glycan display with presence of the matching effector in space and time and its functional consequences, e.g. to keep activated T effector cells under control preventing autoaggression or to drive pancreatic cancer cells into anoikis, sets

examples of clinical relevance on how this concept is realized. In summary, the combination of sequence/shape and topology (e. g. local density of cognate sites, together with cross-linking capacity of lectins to elicit efficient post-binding signalling) appears to be the key for the understanding of molecular specificity. At the same time, it makes intricate dynamic regulation and perspectives for rational drug design possible. Figuring out the details is a challenge for interdisciplinary research involving fields from computational chemistry to molecular medicine.

## Introduction

Cell sociology depends on molecular interactions with the environment. These recognition processes underlie the decision whether a cell can, for example, attach to a certain matrix or be responsive to potential effectors. The enormous advances in sequencing nucleic acids and proteins as well as in monitoring expression profiles by chip and proteomics technologies have led to profound insights into the versatility of the proteome as a platform for structure-activity considerations. However, a simple consideration of structural aspects with an eye on theoretical coding capacity in oligomers, as presented in Figure 1, reveals inherent limits of the linear structures of nucleic acids and proteins: linkage biochemistry with phosphodiester/peptide bonds mostly restricts the way monomers can be joined to a single mode. In other words, an oligomer of these two classes of biomolecules is entirely characterized by the sequence, except for rare cases of 2',5'-phosphodiester and isopeptide bonds. If at this level variability can yet commonly enter oligomer formation, then the coding capacity will automatically increase. The ensuing gain, however, comes at the (analytical) expense to make complete sequence determination much more demanding, and this is the case for carbohydrates.

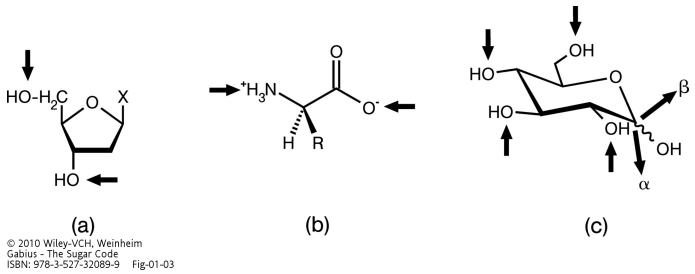


**Figure 1.** Illustration of the linkage points for oligomer formation in biomolecules by arrows. The phosphodiester bond in nucleic acid biosynthesis (a) and the peptide bond in protein biosynthesis (b) yield linear oligomers. In contrast, the glycosidic linkage in oligosaccharides can involve any hydroxyl group, opening the way to linear and also branched structures (c). Using D-glucose as an example, its active form UDP-Glc allows conjugation of this sugar to carbohydrate acceptors to any hydroxyl group,

as symbolized by arrows directed towards the hydroxy groups. The anomeric position in chain elongation can vary, as symbolized by two bold arrows pointing away from the molecule (from [5], with permission).

Simple visual inspection of the carbohydrate section in Figure 1, makes it readily apparent that sugars indeed offer new opportunities for structural variability. The chemistry of connecting monosaccharides to glycan chains allows to exploit both the anomeric nature of the glycosidic bond (α/β) and further diversity arising from changing the linkage position (Figure 1). The example of the 1,4-linked glucose polymers cellulose (β) and glycogen/starch (α) supports the significance of the anomer position, and an α(or β)1 – 4 linkage is just one of the five theoretically possible ways to build diglucosides by connecting the anomeric centre to a hydroxyl group. In nature, divergence in linkage position can in fact be nearly this complete. The sugar L-fucose (please see below for its detection as part of in AB0 blood-group epitopes by lectin application) can be added to glycan chains of glycoproteins in α1,2/3/4/6 positions (please note required sophistication on the level of the glycosyltransferases) to build AB0(H)/$Le^{a/b}$/$Le^{x/y}$ blood-group epitopes and the core fucosylation in complex-type *N*-glycans [1, 2]. These two examples highlight the special features of carbohydrates, often not realized when focusing on their roles as biochemical fuel or as cell-wall concrete. In view of their ubiquitous presence as integral part of glycoproteins and glycolipids [3], a much broader role is more than likely, especially when considering the complexity of the enzymatic machinery for glycan assembly indicated for fucosyltransferases. This notion is graphically expressed by referring to the building blocks ('letters') of cellular glycans as the third alphabet of life, owing to their special chemical properties, which enable high-density information coding: the basis of the concept of the sugar code [3, 4]. Using this term implies the question on the individual members of this alphabet, the ABC of the sugar code.

The structures of the individual 'letters' of the alphabet for the sugar language are illustrated in Figure 2, along with their acceptor position(s) in glycoconjugates [5]. As seen, sugars which are present within chains such as mannose use the full theoretical capacity for versatility. This type of linkage-point variability already mentioned above ensures that oligosaccharides surpass nucleic acids and proteins by orders of magnitude in the capacity to build isomer panels ('code words') [6]. Additionally, branching is rather common in glycoprotein glycans (only less than 20% of 3299 examined mammalian glycans are linear, more than 50% contain one or two branch sites, the rest even more [7]). Typically, sialic acid and fucose reside at branch ends (this explains their comparatively reduced acceptor range, given in Figure 2), along with some other moieties such as galactose (Gal), *N*-acetylgalactosamine (Gal*N*Ac) or *N*-acetylglucosamine (Glc*N*Ac), and there is a further mode to increase structural complexity. Following their synthesis, glycan chains can then be subjected to reversible additions. They are comparable in specificity and functionality with the site-specific post-translational modifications of distinct amino acid side chains in proteins. The introduction of specific substitutions into the mentioned (and other) units, e.g. sulfation in a branch-end Gal*N*Ac moiety [8] or in Glc*N*Ac (*O*- and *N*-sulfation possible) of glycosami-

noglycan chains [9], exemplifies this frequently taken route to further increase the coding capacity by sugars. Physiologically important cases involving phosphorylation and sulfation, here at least 35 enzymes are responsible for this modification in the human Golgi region, are presented in Figure 3 (please see legend for medical relevance).

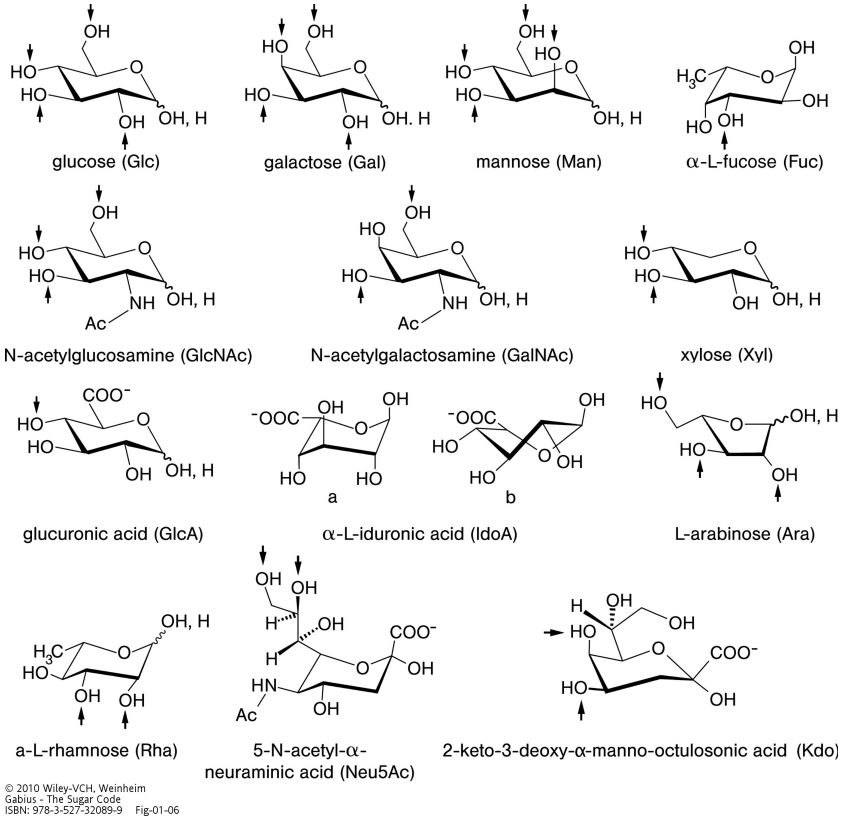


**Figure 2.** Illustration of the alphabet of the sugar language. Structural representation, name and symbol as well as the set of known acceptor positions (arrows) in glycoconjugates are given for each letter. Four sugars have L-configuration: fucose (6-deoxy-L-galactose), rhamnose (6-deoxy-L-mannose) and arabinose are introduced during chain elongation, whereas L-iduronic acid (IdoA) results from post-synthetic epimerization of glucuronic acid at C-5. The $^{1}C_{4}$ conformation of IdoA (a) is in equilibrium with the $^{2}S_{O}$ form (b) in glycosaminoglycan chains where this uronic acid can be 2-sulfated (please see Fig. 3 d). All other "letters" are D-sugars. Neu5Ac, one of the more than 50 sialic acids, often terminates sugar chains in animal glycoconjugates. Kdo is a constituent of lipopolysaccharides in the cell walls of Gram-negative bacteria and is also found in cell wall polysaccharides of green algae and higher plants. Foreign to mammalian glycobiochemistry, microbial polysaccharides contain the furanose ring form of D-galactose and also D-/L-arabinose indicated by an italic "*f*" derived from the heterocyclus furan. The α-anomer is prevalent for the pentose

arabinose, e.g. in mycobacterial cell wall arabinogalactan and lipoarabinomannan. β1 – 5/6-Linked galactofuranoside is present in the arabinogalactan and the β1 – 3/6 linkage in lipopolysaccharides (from [5], with permission).

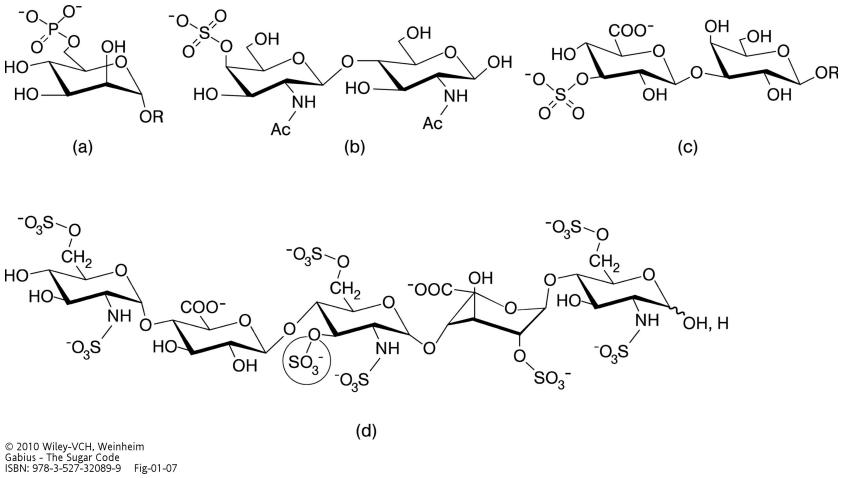


**Figure 3.** Illustration of phosphorylated (phosphated) and sulfated (sulfurylated) glycan "words". 6-Phosphorylation of a mannose moiety (in the context of a mannose-rich pentasaccharide) is the key section of a routing signal in lysosomal enzymes (a), 4-sulfation of the Gal*N*Acβ1 – 4Glc*N*Ac (Lacdi*N*Ac) epitope forms the "postal code" for clearance from circulation by hepatic endothelial cells of pituitary glycoprotein hormones labeled in such a way (b), the HNK-1 (human natural killer-1) epitope (3-sulfated GlcAβ1 – 3Galβ1 – 4Glc*N*Ac) is involved in cell adhesion/migration in the nervous system (c) and the encircled 3-*O*-sulfation in the pentasaccharide's center is essential for heparin's anti-coagulant activity (d). All sugars are in their pyranose form. Please note that the central glucosamine unit has *N*,*O*-trisulfation and that the 2-sulfated IdoA, given in the $^{1}C_{4}$ conformation, can also adopt the hinge-like $^{2}S_{O}$ skew-boat structure (please see Figure 2; about 60% or more for the $^{2}S_{O}$ form in equilibrium depending on the structural context) when present within glycosaminoglycan chains of the proteoglycan heparin. 2-Sulfation of IdoA serves two purposes: favoring the hinge-like $^{2}S_{O}$ conformation and precluding re-conversion to GlcA (from [5], with permission).

In summary, their chemical properties qualify carbohydrates to provide the most versatile toolbox to generate biochemical signals. In addition to enabling the exceptional structural diversity, deliberating their abilities to contribute to molecular recognition is rewarding. Both directional acceptor/donor hydrogen bonds (via hydroxyl groups) and van der Waals interactions/stacking (via patches of polarized C-H bonds) invite a rendezvous with proteins [3, 4]. What has become a burgeoning research field, i.e. work on structure and functions of the lectin class of glycan receptors [3, 4], started back in 1860 with experiments on snake venom [10].

## Translation of the Sugar Code by Proteins

Carbohydrate-binding proteins, which do not act as enzymes on the glycans (e. g. glycosyltransferases, glycosidases or sulfotransferases), are classified into three categories: sensor/transport proteins for free mono- to oligosaccharides, antibodies, and lectins. This separation is reflected in the current definition of the term 'lectin', which are binding partners for glycoconjugate glycans [10]. Occurrence of lectins is widespread, ranging from viruses and eubacteria, plants and invertebrates to mammals [11 – 13]. In support of the given notion for the physiological importance of the third alphabet of life, the structural diversity of glycans built from the set of units shown in Figure 2 and their substituted derivatives (see Figure 3) should be matched by the development of a wide array of recognition sites. Indeed, the establishment of carbohydrate reactivity in at least 15 folding patterns of proteins in animals, as compiled in Table 1, signifies that the structural platform for reading and translating sugar-encoded information is elaborate, definitely not a singular invention for one particular fold.

**Table 1.** Overview of protein folds with lectin activity

| type of fold | example for lectin | example for ligand |
|---|---|---|
| C-type | asialoglycoprotein receptor, collectins, selectins | Fuc, Gal, Gal*N*Ac, Man, heparin tetrasaccharide |
| I-type (Ig fold) | N-CAM, TIM-3, siglecs | $Man_6Glc$*N*$Ac_2$, HNK-1 epitope, α2,3/6-sialylated glycans |
| P-type | mannose-6-phosphate receptors (MR) and proteins with MR homology domain (erlectin, OS-9) | Man-6-phosphate, $Man_{5,8}Glc$*N*$Ac_2$ |
| β-sandwich (jelly-roll) | a) galectins | β-galactosides |
| | b) calnexin, calreticulin | $Glc_1Man_9Glc$*N*$Ac_2$ |
| | c) ERGIC-53 | $Man_xGlc$*N*$Ac_2$ |
| | d) CRD[a] of Fbs1 in SCF E3 ubiquitin ligase and peptide-*N*-glycanase | $Man_3Glc$*N*$Ac_2$; mannopentaose |
| | e) pentraxins | glycosaminoglycans, MOβDG, 3-sulfated Gal, Gal*N*Ac and GlcA, Man-6-phosphate |
| | f) G-domains of the LNS family (laminin, agrin) | heparin |
| β-trefoil | a) fibroblast growth factors | heparan sulfate |
| | b) cystein-rich domain of C-type macrophage mannose receptor | Gal*N*Ac-4-sulfate in Lacdi*N*Ac |
| | c) lectin domain in Gal*N*Ac-Ts[b] involved in mucin-type O-glycosylation | Gal*N*Ac |
| | d) hemolytic lectin CEL-III of sea cucumber and lectin EW29 of earthworm | Gal |
| β-propeller | a) 4-bladed: tachylectin-3 | S-type lipopolysaccharide Glc*N*Ac/Gal*N*Ac |
| | b) 5-bladed: tachylectin-2 | KDO |
| | c) 6-bladed: tachylectin-1 | |
| β-prism I | secretory proteins zg16 p/b | not defined |

**Table 1.** continued

| type of fold | example for lectin | example for ligand |
|---|---|---|
| β-prism II | pufferfish (fugu) lectin | Man |
| β-barrel with jelly-roll topology | tachylectin-4, eel (*Anguilla anguilla*) agglutinin, *Xenopus* | Fuc |
| fibrinogen-like domain | a) ficolins<br>b) intelectins (mammalian, *Xenopus*)<br>c) tachylectin-5<br>d) slug (*Limax flavus*) lectin | GlcNAc<br>Gal*f*, pentoses<br><br>*N*-acetylated sugars<br>sialic acid |
| link module | CD44, TSG-6, LYVE-1, aggregating proteoglycans | hyaluronic acid |
| hevein-like domain | tachycytin and spider (*Selenocosmia huwena*) neurotoxin; cobra venom cardiotoxin | GalNAc; heparin-derived disaccharide |
| $(\beta/\alpha)_8$ barrel (glycoside hydrolase family 18) | YKL-40 (human cartilage glycoprotein-39; chitinase-like lectin) | $(GlcNAc)_n$ |
| short consensus repeat (complement control protein module) | factor H (complement regulator) | glycosaminoglycans, sialic acid |

[a]carbohydrate recognition domain, [b]*N*-acetylgalactosaminyltransferases,
[c]please see Figure 2; adapted from [13] and extended, with permission

On the level of the glycans, the selection process between non-cognate and cognate epitopes for a lectin is expected to be of exquisite specificity which ensures the precision of lectin functionality. Despite the abundant presence of glycoconjugates on and around cells (the term "glycocalyx" is emblematic for this situation), lectins are capable to target distinct counter-receptors to fulfil specific assignments, for example to allow leukocytes to arrest at sites of inflammation (aptly termed selectins, a group of C-type lectins; see upper part of Table 1) or to regulate the growth of inflammatory and tumour cells [4, 14, 15]. In the latter process cascade, the proto-type homodimeric galectin-1, a cross-linking device with β-sandwich folding, elicits anoikis induction/growth inhibition in neuroblastoma/carcinoma cells or T effector cells (responsible for onset of autoimmune dys-regulation) by complex formation with the $\alpha_5\beta_1$-integrin or ganglioside GM1, respectively [16 – 19]. To explain the biochemical reasons why a certain glycoprotein/glycolipid serves as counter-receptor, depending on the cell type, is a central problem for functional glycomics. Strategic integration of various experimental approaches is required for this pressing issue to be resolved satisfactorily [4]. These efforts carry with them a promise for developing innovative medical applications. Toward this end, we first focus on the glycan side and systematically dissect the levels at which the glycan can regulate affinity in carbohydrate – protein (lectin) interactions. As a guideline for the following, Table 2 presents the overview of these parameters, which cooperate to turn a carbohydrate epitope into a privileged docking point for a lectin.

## Six Levels of Affinity Regulation

The first level is defined by the selection process of the "letters" shown in Figure 2. Prior to reaching the current classification of lectins based on sequence homology and folding pattern (for an overview, see Table 1), a technically simple and robust method delivered such data for a proper way to allocate proteins to categories. It was based on the capacity of at least bivalent proteins to agglutinate cells, technically most easily erythrocytes. As reflected in the given definition of the term 'lectin' with its emphasis on sugar binding, the inhibition of this haemagglutination (or polysaccharide precipitation) by mono- or disaccharides led to a common system to assign (phytohaem)agglutinins to groups, mannose- or galactose-specific lectins, to name some [10]. Testing a panel of glyco-compounds accordingly will thus reveal the sugar specificity expressed in terms of the most potent compound as a measure of affinity. Historically, a respective milestone was the delineation that human AB0(H) blood-group determinants are based on sugars. In detail, the agglutination of human erythrocytes of blood-group 0 status by the eel (*Anguilla anguilla*) lectin was inhibited most by α-methyl-L-fucopyranoside, 16fold less so by its β-anomer [20]. Other sugars such as D-fucose, arabinose, glucose, rhamnose or xylose were only very weakly active or not at all [20].

Of course, the simplicity of this assay spurred activity to detect lectins in diverse sources and with various specificities [10]. For example, the agglutination by the first member of the galectin family (from the electric organ tissue of the electric eel *Electrophorus electricus*) is precluded by the presence of a Gal-Glc(*N*Ac) disaccharide (the β1,4-linked lactose or *N*-acetyllactosamine but not α1,6-linked melibiose) [21]. With this information presented, the first level of affinity regulation is reached, in terms of mono- or disaccharides (Table 2). The agglutinin is characterized as lectin with specificity for such compounds. More closely looking into the family of galectins, the growing availability of various disaccharides solidified the notion for the importance of linkage points, as attested by affinity differences between β1,3/4-linked *N*-acetyllactosamine [22]. Even more important, gaining access to oligosaccharides, benefiting from the advances in synthetic carbohydrate chemistry [23], enabled to discern stringent structure-activity relationships and marked lectin affinity for distinct sugar epitopes, e.g. the blood group A determinant in the case of human galectin-3 [22]. Moving from mono- and disaccharides to more complex glycans takes affinity regulation to the second level (Table 2). Progress in preparative work on carbohydrates thus fuelled advances in our understanding of the size of contact area between a lectin and ligands. At the same time, the di- and oligosaccharides became objects of structural work themselves, by NMR spectroscopic analysis and by molecular modelling [24–26].

**Table 2.** Six levels of regulation of affinity for binding of a glycan to a lectin[a]

1. mono- and disaccharides (incl. anomeric position, linkage points and substitutions)
2. oligosaccharides (incl. branching and substitutions)
3. shape parameters of oligosaccharides
   a. shape of oligosacchride (*differential conformer selection*)
   b. differences in conformational flexibility between isomers
4. spatial parameters of glycans in natural glycoconjugates
   a. shape of glycan chain (example: modulation of conformational equilibrium by substitutions not acting as lectin ligand such as core fucosylation or introduction of bisecting Glc*N*Ac in *N*-glycans, influence of protein part)
   b. cluster effect with bi- to pentaantennary *N*-glycans or branched *O*-glycans (incl. modulation by substitutions, see a.)
5. cluster effect with neighboring glycan chains on the same glycoprotein (e.g. in mucins)
6. cluster effect with cellular glycoconjugates in spatial vicinity, e.g. in membranes with glycoproteins, glycolipids or complexes thereof, especially when presented in microdomains; modulation possible by dynamic remodeling of glycan chain, e.g. by enzymatic desialylation

[a]adapted from [13] and extended, with permission

These combined efforts revealed a remarkable property of many oligosaccharides: their limited flexibility that induces often the glycan to adopt only a few energetically privileged conformers, with implications for bio-recognition [4, 26]. A selection of preformed conformers for binding constitutes an entropic advantage in the thermodynamic balance sheet, when compared to a highly flexible ligand, and dynamic shifts in the shape profile have the potential to modulate affinity [4, 26]. The case study on galectin-1 and the pentameric lectin part of the cholera toxin ($AB_5$), which compete for binding to the pentasaccharide of ganglioside GM1, revealed selection of two different conformers, entailing the obvious perspective to exploit this disparity for the design of lectin-type-specific inhibitors [4]. A change of the shape, moving from the two dimensions of the structural formula to three dimensions in our concept, will affect the extent of complementarity with a protein in bio-recognition. In the given case, each lectin actually accommodates one from the three low-energy constellations how the sialic acid and the central galactose moiety are positioned. The α2,3-sialylgalactose can thus be considered to preferentially exist as three ‘keys’, one of them fitting into a ‘lock’ termed galectin-1, another into a B-subunit of the bacterial toxin. If this linkage were in α2,6-position, now involving the C5-C6 bond of galactose (an additional degree of rotational freedom), the binding properties will be completely different, in terms of direct contact (shape) and of flexibility. In the terminology of the sugar code, these two isomers, i.e. the α2,3/6-sialylgalactosides, which decorate the branch ends of glycan chains, are code words *sui generis*. In summary, the shape of the ligand and the extent of its flexibility factor into affinity regulation determine the level three (Table 2). To see these principles of lectin binding at work, the reader is referred to a movie illustrating the properties of the binding

partners and then the complex formation in the already mentioned case of galectin-3 and the blood-group A tetrasaccharide. Flexible docking on both sides is accomplished with the HADDOCK version 2.1 (see [27] for details on access information).

Up to this stage of affinity regulation, we have dealt with free saccharides, not glycan chains in natural glycoconjugates. Starting with the level four, these structural aspects will now be taken under scrutiny. They have a strong influence on the presentation of the carbohydrate ligand in its natural context, with increasing degree of complexity up to the dynamics of membrane microdomains. A crucial factor for affinity is the local density of ligand presentation. This topological parameter can be altered by several means. The first is the introduction of so-called core substitutions to *N*-glycans, i. e. the presence of core fucosylation (in α1,6-linkage in contrast to α1,2-linkage in blood-group AB0 epitopes) and the bisecting Glc*N*Ac moiety [4, 28, 29]. The additions to *N*-glycans occur non-randomly and have a marked influence on glycoprotein functionality, e. g. in binding antibody-antigen complexes to $F_c\gamma$ receptors (e. g. $F_c\gamma$RIII), where the core-fucose unit can impair the high affinity of antibodies seen without this modification [30]. Comprehensive molecular modeling for biantennary *N*-glycans has disclosed the enormous impact of these two substitutions on the conformational equilibrium of the respective *N*-glycan chain, which translates into affinity alterations for lectin binding [31]. Of potential clinical relevance, the rate of hepatic uptake of pharma-glycoproteins, a common route for their clearance from serum and thus governing their pharmacodynamics, can now be tailored rationally by selecting the cell system for recombinant production accordingly [31]. The substitutions thus work on the glycan chains like molecular switches, one aspect of level four (Table 2).

The second aspect on this level is attributed to altering the local density of lectin ligands by branching. The effectiveness of this structural parameter was first described by measuring the reactivity of mono-, bi-, and trivalent galactose-terminated glycans to the mammalian asialoglycoprotein receptor [32]. The numerical increase in valency resulted in a geometrical increase in affinity, an effect termed the glycoside cluster effect [33, 34]. Following this pioneering work, the natural role model has encouraged the respective design of glycoclusters with the aim to produce potent lectin inhibitors for medical applications [33 – 35]. The combination of the most suitable cluster design (in terms of affinity and discriminatory efficiency) with recruiting carbohydrate derivatives provides a perspective for rational optimization of inhibitor synthesis, an example for galectin-3 reported recently [36].

Reaching the same aim as by branching (level four b), the cluster effect can likewise be achieved by increasing the density of glycan chains on a scaffold, e. g. on a protein backbone of a mucin (level five). The resulting marked affinity enhancement for lectins can be explained by applying kinetic terms: the macroscopic off-rate of the lectin, when dissociating from a complex with a ligand, appears decreased at sufficient ligand density, likely coupled with an increase in the on-rate [37]. A lectin will thus switch places and oscillate between neighbouring sites instead of losing contact to the glycoconjugate. Spatial vicinity

can be generated not only by cluster formation on the same scaffold (level five). Microdomains in membranes are ideal platforms for bringing certain candidates for high-affinity binding together, an example for level six. Indeed, the ganglioside GM1 introduced above becomes a counter-receptor in this constellation, as shown by the detrimental effect of cholesterol depletion of membranes to harm microdomain integrity, and tight association with glycoproteins (e. g. with the $\alpha_5\beta_1$-integrin) can guide the post-binding signalling [38, 39]. The contact site itself, the pentasaccharide, can be made accessible on demand by dynamic glycan remodelling on the membrane.

Cell surface glyco-enzymes (glycosidases, glycosyltransferases), which are under strict control, can remove or establish contact sites. A graphic example is the up-regulation of a cell surface ganglioside sialidase on neuroblastoma cells, turning higher sialylated gangliosides into galectin-1-reactive ganglioside GM1, the signal to initiate cell growth inhibition [39]. In a broader context, reactivity to lectins can hereby be modulated differentially. Looking at siglecs it can be abolished by desialylation, while recognition features for certain C-type lectins or galectins are installed [14, 40]. This dynamic remodelling can be the prerequisite for cluster (microdomain) generation, therefore to high affinity, connecting the enzymatic machinery for performing structural alterations *in situ* with recognition by lectins (level six). After the completion of the description of the six levels listed in Table 2, it is mandatory to immediately mention that the intricate versatility of carbohydrate-protein interactions is also rooted in the structural design of lectins. As highlighted in Table 2, spatial parameters are salient, and it is not surprising that modular display of these sites matters, besides the folding pattern summarized in Table 1 and fine-structural aspects of the architecture of the contact site (see the movie for galectin-3 for details in a model case [27]). The family of C-type lectins, defined by a folding with two anti-parallel β-strands and two α-helices, is an illustrative example for the multitude of various types of display, from monomers to non-covalent/tandem-repeat-type oligovalent clusters [41].

# Modular Display of Lectin Sites: Regulation of both Affinity and Selectivity

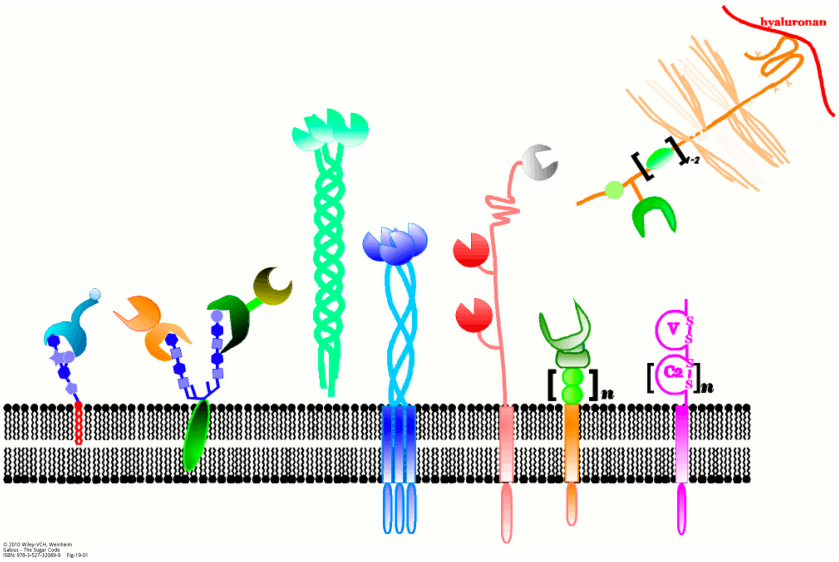


**Figure 4.** Illustration of the strategic ways how carbohydrate recognition domains (CRDs) in animal lectins are positioned to reach optimal ligand selection (for example to separate self from non-self glycan profiles in innate immunity) and topological complementarity. From left to right, the CRD display in the three subtypes within the galectin family (chimeric, proto-type and tandem-repeat-type arrangements binding to a ganglioside or a branched complex-type *N*-glycan without or with terminal α2,3-sialylation), the presentation of CRDs (C-type or fibrinogen-like domain) in serum and surfactant collectins or ficolins connected to their collagenous stalks and the non-covalent association of binding sites in transmembrane C-type lectins by α-helical coiled-coil stalks (for example asialoglycoprotein and Kupffer cell receptors, the scavenger receptor C-type lectin, CD 23, DC-SIGN or DC-SIGNR) are given. Similar to tandem-repeat-type galectins the C-type family of lectins also has a branch of members with this design, i. e. immulectins-1, -2 and 3. Next, the tandem-repeat display in the mannose-specific macrophage receptor (also found on dendritic cells, hepatic endothelial cells, kidney mesangial cells, retinal pigment epithelial cells and tracheal smooth muscle cells) and the related C-type lectin Endo180, an endocytic receptor for glycosylated collagen with eight domains, as well as in the cation-independent P-type lectin with 15 domains is presented. Capacity for sugar binding is confined to only few domains as depicted. The occurrence of lectin activity for Gal*N*Ac-4-$SO_4$-bearing pituitary glycoprotein hormones in the cysteine-rich domain, a member of the β-trefoil fold family with one $(QxW)_3$ domain in the N-terminal section of the macrophage mannose receptor (amino acids 8 – 128), which is linked via a fibronectin-type-II-repeat-containing module to the tandem-repeat section, is also included into the schematic drawing for these lectins with more than one type of CRD per protein chain. Moving further to the right side, the association of a distal CRD in selectins (attached to an epidermal-growth-factor (EGF)-like domain and two to nine

complement-binding consensus repeats) or in the siglec subfamily of I-type lectins using 1 – 16 C2-set immunoglobulin-like units as spacer equivalents to let the CRD reach out to contact ligands and to modulate capacity to serve in *cis*- or *trans*-interactions on the cell surface is shown. The force-dependent alterations of the topological arrangement of the two distal domains in selectins accounts for catch bonds of selectins, a canonical immunoreceptor tyrosine-based inhibitory motif (ITIM) together with a putative tyrosine-based motif is frequently present in the intracellular portion of siglecs. C2-set domains linked to fibronectin-type-III repeats establish the extracellular section of the I-type lectins L1 and neural cell adhesion molecule (N-CAM). In the matrix, the modular proteoglycans (hyalectans/lecticans: aggrecan, brevican, neurocan and versican) interact a.) with hyaluronan (and also link protein) via the link-protein-type modules of the N-terminal G1 domain (and an Ig-like module), b.) with receptors binding to the glycosaminoglycan chains in the central region and c.) with carbohydrates or proteins (fibulins-1 and -2 and tenascin-R) via the C-type lectin-like domain flanked by EGF-like and complement-binding consensus repeat modules (kindly provided by H. Kaltner; from [13], with permission)

In Figure 4 it can be seen an overview of ways how carbohydrate recognition domains are arranged to let them spot their targets. Local density can be varied, and, remarkably, association with other domains can implement properties conducive for functionality (for detailed explanations, see legend to Figure 4). Covalent connections between the modules facilitate the generation of tandem-repeat-type effectors with identical or different domains. Non-covalent association, as depicted for the i) asialo-glycoprotein receptor mentioned above as role model for the glycoside cluster effect, ii) collectins which distinguish the glycan signatures of infectious organisms such as yeast and malignant cells from normal self-glycans [42] and iii) cross-linking homodimeric galectins such as galectin-1, is the alternative for bringing recognition sites for glycans into spatial vicinity (see Figure 4 and its legend for further information on the mentioned lectins). Of note, intra-family diversity in this parameter brings about functional divergence. Galectins-1 and -3, shown in the left part of Figure 4, can compete for the same counter-receptor but are fundamentally different in triggering post-binding signaling. Consequently, galectin-3 can block anti-tumuoral activities of galectin-1, in neuroblastoma cells [16] and in the anoikis induction in pancreatic carcinoma cells upon tumor suppressor $p16^{INK4a}$ restoration, to list examples with clinical relevance [43]. This competition has prompted the design of galectin-3-specific glycoclusters to neutralize its antagonistic action on galectin-1 [36].

## Conclusions

In his thoughtful reflections on glycobiology [44], S. Roseman provides a simple answer to the question "why glycobiology has apparently lagged so far behind the other fields" (i.e. genomics or (functional) proteomics): "Glycoconjugates are much more complex, variegated, and difficult to study than proteins or nucleic acids". The combination of structural analysis and synthetic chemistry with molecular modelling and the experimental monitoring of interactions with proteins (lectins) teach instructive lessons about the factors that can regulate the translation of the sugar code on the side of the glycans. The exquisite selectivity

in physiological recognition and post-binding signalling, leading e. g. to regulation of growth and adhesion [3, 4], is attributed to six levels, from mono- or disaccharide binding to presentation of suited glycans as clusters in microdomains. The presented systematic dissection into these separate parameters has the purpose to give research a clear direction by sensitizing to relate e. g. arising structural information on glycan presentation by a glycoprotein to the potential impact of this constellation on lectin reactivity. Lectin-dependent cargo selection and routing in apical transport of glycoproteins substantiates this concept [45].The orchestration of changes in (glycoprotein) glycosylation and cell surface presence of a lectin effector by a tumour suppressor [17] sets a precedent for the intimate cooperation of both sides of glycan-protein interactions to achieve the required selectivity in terms of the binding partner (here for the $\alpha_5\beta_1$-integrin) and thereby post-binding signalling (here caspase-8-dependent anoikis induction). The complexity of glycan display, based on the chemical properties of sugars, therefore ensures to generate the physiologically required diversity of sequence/shape/density signatures for a wide array of specific recognition processes.

## References

[1] Becker, D.J. & Lowe, J.B. (2003) Fucose: biosynthesis and biological functions in mammals. *Glycobiology* **13:**41R-53R.
doi: http://dx.doi.org/10.1093/glycob/cwg054.

[2] Javaud, C., Dupuy, F., Maftah, A., Julien, R. & Petit, J.M. (2003) The fucosyltransferase gene family: an amazing summary of the underlying mechanisms of gene evolution. *Genetica* **118:**157 – 170.
doi: http://dx.doi.org/10.1023/A:1024101625214.

[3] Gabius, H.-J., (Ed.) (2009) *The Sugar Code. Fundamentals of glycosciences*. Wiley-VCH. Weinheim, Germany.

[4] Gabius, H.-J., André, S., Jiménez-Barbero, J., Romero, A. & Solís, D. (2011) From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem. Sci.* **36:**298 – 313.
doi: http://dx.doi.org/10.1016/j.tibs.2011.01.005.

[5] Rüdiger, H. & Gabius, H.-J. (2009) The biochemical basis and coding capacity of the sugar code. In: Gabius, H.-J. (Ed.) *The Sugar Code. Fundamentals of glycosciences*. pp. 3 – 13, Wiley-VCH, Weinheim, Germany.

[6] Laine, R.A. (1997) The information-storing potential of the sugar code. In: Gabius, H.-J. & Gabius, S. (Eds.) *Glycosciences: Status and Perspectives*. pp. 1 – 14, Chapman & Hall, London – Weinheim.

[7] Seeberger, P.H. (2010) Bioinformatics: key to the future of chemical glycomics. In: Hicks, M.G. & Kettner, C. (Eds.) *Proceedings of the Beilstein Sympoisum on Glyco-Bioinformatics: Bits 'n' Bytes of Sugars*. pp. 25 – 36, Beilstein Institut zur Förderung der Chemischen Wissenschaften, Frankfurt. Logos-Verlag Berlin.

[8] Hooper, L.V., Manzella, S.M. & Baenziger, J.U. (1997) The biology of sulfated oligosaccharides. In: Gabius, H.-J. & Gabius, S. (Eds.) *Glycosciences: Status and Perspectives*. pp. 261 – 276, Chapman & Hall, London – Weinheim.

[9] Buddecke, E. (2009) Proteoglycans. In: Gabius, H.-J. (Ed.) *The Sugar Code. Fundamentals of glycosciences*. pp. 199 – 216, Wiley-VCH, Weinheim, Germany.

[10] Rüdiger, H. & Gabius, H.-J. (2009) The history of lectinology. In: Gabius, H.-J. (Ed.) *The Sugar Code. Fundamentals of glycosciences*. pp. 261 – 268, Wiley-VCH, Weinheim, Germany.

[11] Holgersson, S., Gustafsson, A. & Gaunitz, S. (2009) Bacterial and viral lectins. In: Gabius, H.-J. (Ed.) *The Sugar Code. Fundamentals of glycosciences*. pp. 279 – 300, Wiley-VCH, Weinheim, Germany.

[12] Rüdiger, H. & Gabius, H.-J. (2009) Plant lectins. In: Gabius, H.-J. (Ed.) *The Sugar Code. Fundamentals of glycosciences*. pp. 301 – 315, Wiley-VCH, Weinheim, Germany.

[13] Gabius, H.-J. (2009) Animal and human lectins. In: Gabius, H.-J. (Ed.) *The Sugar Code. Fundamentals of glycosciences*. pp. 317 – 328, Wiley-VCH, Weinheim, Germany.

[14] Schwartz-Albiez, R. (2009) Inflammation and glycosciences. In: Gabius, H.-J. (Ed.) *The Sugar Code. Fundamentals of glycosciences*. pp. 447 – 467, Wiley-VCH, Weinheim, Germany.

[15] Kaltner, H. & Gabius, H.-J. (2012) A toolbox of lectins for translating the sugar code: the galectin network in phylogenesis and tumors. *Histol. Histopathol.* **27**(4):397-416.

[16] Kopitz, J., von Reitzenstein, C., André, S., Kaltner, H., Uhl, J., Ehemann, V., Cantz, M. & Gabius, H.-J. (2001) Negative regulation of neuroblastoma cell growth by carbohydrate-dependent surface binding of galectin-1 and functional divergence from galectin-3. *J. Biol. Chem.* **276:**35917 – 35923.
doi: http://dx.doi.org/10.1074/jbc.M105135200.

[17] 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-Döberitz, M., Gress, T.M., Nishimura, S.-I., Rosewicz, S. & Gabius, H.-J. (2007) Tumor suppressor $p16^{INK4a}$: modulator of glycomic profile and galectin-1

expression to increase susceptibility to carbohydrate-dependent induction of anoikis in pancreatic carcinoma cells. *FEBS J.* **274:**3233 – 3256.
doi: http://dx.doi.org/10.1111/j.1742-4658.2007.05851.x.

[18] Sanchez-Ruderisch, H., Detjen, K.M., Welzel, M., André, S., Fischer, C., Gabius, H.-J. & Rosewicz, S. (2011) Galectin-1 sensitizes carcinoma cells to anoikis via the fibronectin receptor $\alpha_5\beta_1$-integrin. *Cell Death Differ.* **18:**806 – 816.
doi: http://dx.doi.org/10.1038/cdd.2010.148.

[19] Wu, G., Lu, Z.H., Gabius, H.-J., Ledeen, R.W. & Bleich, D. (2011) Ganglioside GM1 deficiency in effector T cells from NOD mice induces resistance to regulatory T cell suppression. *Diabetes* **60:**2341 – 2349.
doi: http://dx.doi.org/10.2337/db10-1309.

[20] Watkins, W.M. & Morgan, W.T.J. (1952) Neutralisation of the anti-H agglutinin in eel serum by simple sugars. *Nature* **169:**825 – 826.
doi: http://dx.doi.org/10.1038/169825a0.

[21] Teichberg, V.I., Silman, I., Beitsch, D.D. & Resheff, G. (1975) A β-D-galactoside binding protein from electric organ tissue of *Electrophorus electricus*. *Proc. Natl. Acad. Sci. U.S.A.* **72:**1383 – 1387.
doi: http://dx.doi.org/10.1073/pnas.72.4.1383.

[22] Sparrow, C., Leffler, H. & Barondes, S.H. (1987) Multiple soluble β-galactoside-binding lectins from human lung. *J. Biol. Chem.* **262:**7383 – 7390.

[23] Oscarson, S. (2009) The chemist's way to synthesize glycosides. In: Gabius, H.-J. (Ed.) *The Sugar Code. Fundamentals of glycosciences*. pp. 31 – 51, Wiley-VCH, Weinheim, Germany.

[24] Carver, J.P. (1993) Oligosaccharides: how can flexible molecules act as signals? *Pure Appl. Chem.* **65:**763 – 770.
doi: http://dx.doi.org/10.1351/pac199365040763.

[25] Imberty, A. (1997) Oligosaccharide structures: theory versus experiment. *Curr. Opin. Struct. Biol.* **7:**617 – 623.
doi: http://dx.doi.org/10.1016/S0959-440X(97)80069-3.

[26] von der Lieth, C.-W., Siebert, H.-C., Kozár, T., Burchert, M., Frank, M., Gilleron, M., Kaltner, H., Kayser, G., Tajkhorshid, E., Bovin, N.V., Vliegenthart, J.F.G. & Gabius, H.-J. (1998) Lectin ligands: new insights into their conformations and their dynamic behavior and the discovery of conformer selection by lectins. *Acta Anat.* **161:**91 – 109.
doi: http://dx.doi.org/10.1159/000046452.

[27] Krzeminski, M., Singh, T., André, S., Lensch, M., Wu, A.M., Bonvin, A.M.J.J. & Gabius, H.-J. (2011) Human galectin-3 (Mac-2 antigen): defining molecular switches of affinity to natural glycoproteins, structural and dynamic aspects of glycan binding by flexible ligand docking and putative regulatory sequences in the proximal promoter region. *Biochim. Biophys. Acta.* **1810:**150 – 161.
doi: http://dx.doi.org/10.1016/j.bbagen.2010.11.001.

[28] Wilson, I.B.H., Paschinger, H. & Rendic, D. (2009) Glycosylation of model and 'lower' organisms. In: Gabius, H.-J. (Ed.) *The Sugar Code. Fundamentals of glycosciences*. pp. 139 – 154, Wiley-VCH, Weinheim, Germany.

[29] Zuber, C. & Roth, J. (2009) *N*-Glycosylation. In: Gabius, H.-J. (Ed.) *The Sugar Code. Fundamentals of glycosciences*. pp. 87 – 110, Wiley-VCH, Weinheim, Germany.

[30] Ferrara, C., Grau, S., Jager, C., Sondermann, P., Brunker, P., Waldhauer, I., Hennig, M., Ruf, A., Rufer, A.C., Stihle, M., Umana, P. & Benz, J. (2011) Unique carbohydrate-carbohydrate interactions are required for high affinity binding between $F_c\beta$RIII and antibodies lacking core fucose. *Proc. Natl. Acad. Sci. U.S.A.* **108:**12669 – 12674.
doi: http://dx.doi.org/10.1073/pnas.1108455108.

[31] André, S., Kozár, T., Kojima, S., Unverzagt, C. & Gabius, H.-J. (2009) From structural to functional glycomics: core substitutions as molecular switches for shape and lectin affinity of *N*-glycans. *Biol. Chem.* **390:**557 – 565.
doi: http://dx.doi.org/10.1515/BC.2009.072.

[32] Lee, Y.C., Townsend, R.R., Hardy, M.R., Lönngren, J., Arnarp, J., Haraldsson, M. & Lönn, H. (1983) Binding of synthetic oligosaccharides to the hepatic Gal/GalNAc lectin. *J. Biol. Chem.* **258:**199 – 202.

[33] Lee, Y.C. (2010) Warfare between pathogens and hosts: the trickery of sugars. *Trends Glycosci. Glycotechnol.* **22:**95 – 106.

[34] Lee, R.T. & Lee, Y.C. (1997) Neoglycoconjugates. In: Gabius, H.-J. & Gabius, S. (Eds.) *Glycosciences: Status and Perspectives*. pp. 55 – 77, Chapman & Hall, London – Weinheim.

[35] Chabre, Y.M. & Roy, R. (2009) The chemist's way to prepare multivalency. In: Gabius, H.-J. (Ed.) *The Sugar Code. Fundamentals of glycosciences*. pp. 53 – 70, Wiley-VCH, Weinheim, Germany.

[36] André, S., Grandjean, C., Gautier, F.M., Bernardi, S., Sansone, F., Gabius, H.J. & Ungaro, R. (2011) Combining carbohydrate substitutions at bioinspired positions with multivalent presentation towards optimising lectin inhibitors: case study with calixarenes. *Chem Commun (Camb)* **47:**6126 – 6128.
doi: http://dx.doi.org/10.1039/c1cc11163a.

[37] Dam, T.K. & Brewer, C.F. (2010) Multivalent lectin-carbohydrate interactions energetics and mechanisms of binding. *Adv. Carbohydr. Chem. Biochem.* **63:**139 – 164. doi: http://dx.doi.org/10.1016/S0065-2318(10)63005-3.

[38] Wang, J., Lu, Z.H., Gabius, H.-J., Rohowsky-Kochan, C., Ledeen, R.W. & Wu, G. (2009) Cross-linking of GM1 ganglioside by galectin-1 mediates regulatory T cell activity involving TRPC 5 channel activation: possible role in suppressing experimental autoimmune encephalomyelitis. *J. Immunol.* **182:**4036 – 4045. doi: http://dx.doi.org/10.4049/jimmunol.0802981.

[39] Kopitz, J., Bergmann, M. & Gabius, H.-J. (2010) How adhesion/growth-regulatory galectins-1 and -3 attain cell specificity: case study defining their target on neuroblastoma cells (SK-N-MC) and marked affinity regulation by affecting microdomain organization of the membrane. *IUBMB Life* **62:**624 – 628. doi: http://dx.doi.org/10.1002/iub.358.

[40] Lopez, P.H. & Schnaar, R.L. (2009) Gangliosides in cell recognition and membrane protein regulation. *Curr. Opin. Struct. Biol.* 19: 549 – 557. doi: http://dx.doi.org/10.1016/j.sbi.2009.06.001.

[41] Gready, J.N. & Zelensky, A.N. (2009) Routes in lectin evolution: case study on the C-type lectin-like domains. In: Gabius, H.-J. (Ed.) *The Sugar Code. Fundamentals of glycosciences*. pp. 329 – 346, Wiley-VCH, Weinheim, Germany.

[42] Kawasaki, N. & Kawasaki, T. (2010) Recognition of endogenous ligands by C-type lectins: interaction of serum mannan-binding protein with tumor-associated oligosaccharide epitopes. *Trends Glycosci. Glycotechnol.* **22:**141 – 151. doi: http://dx.doi.org/10.4052/tigg.22.141.

[43] Sanchez-Ruderisch, H., Fischer, C., Detjen, K.M., Welzel, M., Wimmel, A., Manning, J.C., André, S. & Gabius, H.-J. (2010) Tumor suppressor $p16^{INK4a}$: down-regulation of galectin-3, an endogenous competitor of the pro-anoikis effector galectin-1, in a pancreatic carcinoma model. *FEBS J.* **277:**3552 – 3563. doi: http://dx.doi.org/10.1111/j.1742-4658.2010.07764.x.

[44] Roseman, S. (2001) Reflections on glycobiology. *J. Biol. Chem.* **276:**41527 – 41542. doi: http://dx.doi.org/10.1074/jbc.R100053200.

[45] Morelle, W., Stechly, L., André, S., van Seuningen, I., Porchet, N., Gabius, H.-J., Michalski, J.C. & Huet, G. (2009) Glycosylation pattern of brush border-associated glycoproteins in enterocyte-like cells: involvement of complex-type *N*-glycans in apical trafficking. *Biol. Chem.* **390:**529 – 544. doi: http://dx.doi.org/10.1515/BC.2009.075.