Chapter 7
Adhesion/Growth-Regulatory Galectins: Insights into Their Ligand Selectivity Using Natural Glycoproteins and Glycotopes

Albert M. Wu, Tanuja Singh, Jia-Hau Liu, Sabine André, Martin Lensch, Hans-Christian Siebert, Mickael Krzeminski, Alexandre M.J.J. Bonvin, Herbert Kaltner, June H. Wu, and Hans-Joachim Gabius

Keywords Agglutinin • Docking • Glycosylation • Lectin phylogeny • Sugar code

"Biochemistry text books commonly make it appear that it is a foregone conclusion that the hardware of biological information storage and transfer is confined to nucleotides and amino acids, the letters of the genetic code. However, the remarkable talents of a third class of biomolecules are often overlooked" [1]. This statement from a recent review guides the readers to look at and fully appreciate the chemical/lectinochemical characteristics of carbohydrates that underlie the concept of the sugar code [2].

7.1 The Principles of the Sugar Code

Beyond simple sequence permutations, as in oligonucleotides and peptides, there are four parameters that account for the unsurpassed coding capacity of glycans: (1) variability of linkage points; (2) possibility for two types of interglycosidic linkage by anomer; (3) alteration of ring size (furanose or pyranose); and (4) introduction of branches, both in the central and in the peripheral regions of glycans [3]. To define a disaccharide, such as the epitope of the histo-blood group H determinant, writing down the sequence, i.e. Fuc-Gal, is therefore not at all sufficient, as it would be for an oligonucleotide or peptide. Instead, considering the four parameters listed above, the structure of this histo-blood group antigen is only unequivocally defined when writing α-L-Fuc*p*1-2-β-D-Gal*p* (first dimension of the sugar code).

As signified by the physiological importance of this structure, complex synthetic machinery is available to produce the indicated sequence diversity of glycans in nature, hereby realizing their theoretical coding capacity, e.g. facilitating α1-2, α1-3,

A.M. Wu (✉)
Glyco-Immunochemistry Research Laboratory, Institute of Molecular and Cellular Biology, College of Medicine, Chang-Gung University, Kwei-Shan, Tao-Yuan, 333, Taiwan
e-mail: amwu@mail.cgu.edu.tw

A.M. Wu (ed.), *The Molecular Immunology of Complex Carbohydrates-3*,
Advances in Experimental Medicine and Biology 705, DOI 10.1007/978-1-4419-7877-6_7,
© Springer Science+Business Media, LLC 2011

α1-4, and α1-6 linkages for fucose addition to glycan chains [4–8]. Each given linkage type forms a characteristic structural constellation as part of a branched glycan chain, often at terminal positions (the second dimension of the sugar code). Being spatially accessible at these sites is suggestive for being a target of proteins specific for distinct glycotopes. The way the structural assignment of the mentioned histo-blood group epitope had been initiated provides a telling example of the versatility of carbohydrates to engage in such intermolecular interactions. Their inherent specificity combined with hapten inhibition set the stage to track down the presence of the L-fucose moiety as part of the H-epitope. Hydrogen bonds, van der Waals interactions, and C–H/π-bonds with Trp as part of the contact sites are suitable means for yielding specific recognition and enabling, e.g. accurate discrimination against epimers [9–13]. When not tied to enzymatic processing, binding of a glycan defines a protein as a lectin (except for antibodies and transport proteins for free glycans), and the lectin of the European eel *Anguilla anguilla*, an agglutinin for erythrocytes of this type of blood group, was instrumental for identifying fucose as an essential part of the histo-blood group H epitope [14–17]. This molecular interplay, leading to hemagglutination, also underscores the mentioned spatial accessibility of glycans for intermolecular interactions. Thus, in the context of the cell surface, glycoproteins, such as integrins or laminin, can independently engage in protein–protein contacts and also in protein–carbohydrate interplay via their glycan chains.

At this stage, another particular factor comes into play that predisposes carbohydrates as ideal hardware for biological information transfer. Whereas peptides are highly flexible, the bulky rings of the sugars limit the flexibility around glycosidic linkages so that oligosaccharides often adopt only a few energetically privileged conformations, the third dimension of the sugar code [13, 18]. Since lectins accommodate such conformers in their contact sites frequently without major structural changes by mutual adaptation (please see below for flexible ligand docking), the recognition of cell surface glycans as signals or code words does not require arresting a flexible oligomer into a particular conformation in the context of a polymer. This, in contrast, is the case for bioactive conformations of peptide motifs as part of proteins. Overall, glycans thus adorn protein and lipid scaffolds with bioactive signals. They are decoded by respective receptor proteins. We will next turn to a distinct class of these translators of the sugar code with β-sandwich folding, i.e. the galectins.

7.2 Galectins: A Class of Potent Translators of the Sugar Code

A graphic example for this type of molecular complementarity, guiding decoding of glycan signals, is the illustration of the ribbon diagram (Fig. 7.1), which is emblematic for the fold of this family of adhesion/growth-regulatory lectins. Figure 7.2 shows the contact sites and profiles how the pentasaccharide of ganglioside GM1 makes contact to a galectin's binding site as a distinct, energetically privileged conformer, representing a "valley" in the conformational "landscape" of

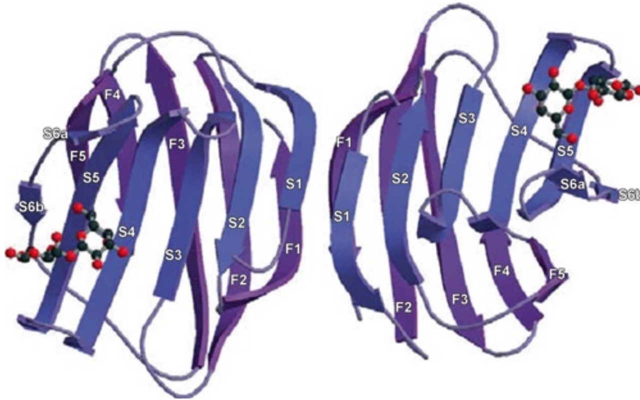

Fig. 7.1 Ribbon diagram of the homodimeric human galectin-1 [19]. The β-strands in the five-stranded (F1–F5) and six-stranded (S1–S6a/S6b) β-sheets are indicated by letter–number code, and the contact site for the carbohydrate ligand is shown by the insertion of a lactose molecule

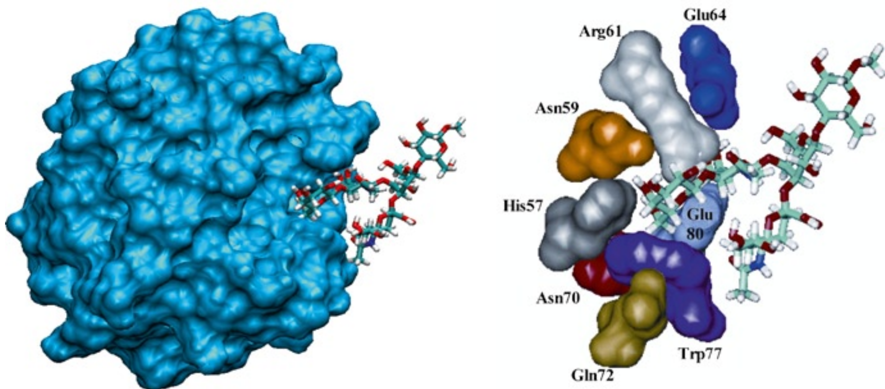

Fig. 7.2 *Left panel*: positioning of the contact site for the pentasaccharide chain of ganglioside GM1, shown for a galectin-1 subunit [19, 20]; *right panel*: the architecture of this site in a rat galectin-5 (rGal-5) [21]

the ligand [19–21]. Using the saturation transfer difference technique in nuclear magnetic resonance (NMR) spectroscopy, local vicinity between galectin-1 and the pentasaccharide have been mapped in detail [20]. Together with calculations of the binding energy of individual amino acids and sugar units, major contacts were ascribed to the Galβ1-3GalNAcβ-terminus (T_β antigen) and the α2-3-linked sialic acid of the branch [20]. Of note, this type of interaction between the lectin and the ganglioside's glycan has significant biological activity, triggering growth inhibition in neuroblastoma cells in vitro and autoimmune suppression by efficiently facilitating communication between regulatory and effector T cell populations [22–24]. Beyond galectins, ganglioside GM1 is also the docking site for cholera toxin.

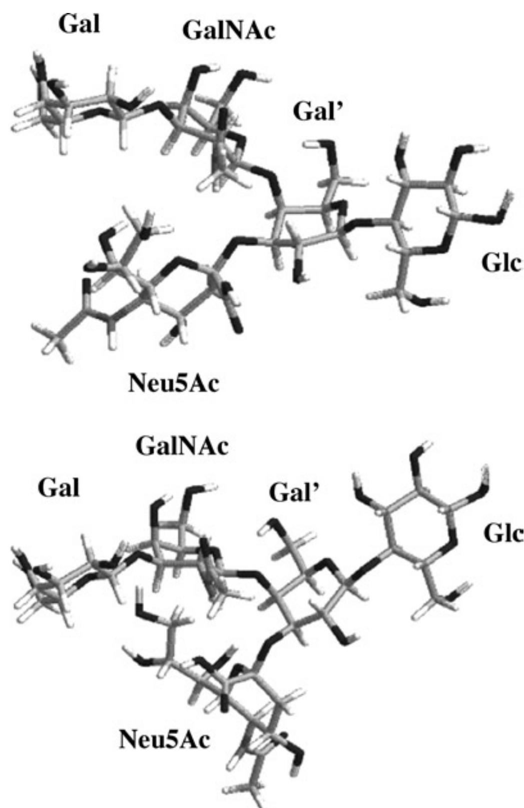

Fig. 7.3 Illustration of two bioactive conformers of the pentasaccharide of ganglioside GM1. Whereas the upper structure is the ligand for human galectin-1, a shape change is required to let the pentasaccharide acquire affinity to cholera toxin (*bottom*) [20]

Intriguingly, the bacterial protein accommodates a different conformer [20]. The pentasaccharide, by adopting the two low-energy "key-like" conformations shown in Fig. 7.3, is thus suited to fit into "locks" of widely separate architecture. Interestingly, selection of a conformer from an equilibrium is not confined to an oligomer. It can already take place on the level of disaccharides, obvious proof for the enormous versatility of carbohydrates as ligands. NMR spectroscopy, especially the determination of interproton distances by picking up nuclear Overhauser effects, was essential for structurally defining bound-state conformations [12, 13, 18]. The two conformers of the disaccharide Galβ1-2Gal, drawn in Fig. 7.4, together with the detectable interproton distances, represent the structures reactive either with a galectin or a plant agglutinin (toxin). The energy maps as part of Fig. 7.4 illustrate the meaning of a central low-energy valley with two further privileged positions. Overall, these results teach us the following lesson: animal/plant lectins are able to perform differential conformer selection, underscoring that conformers

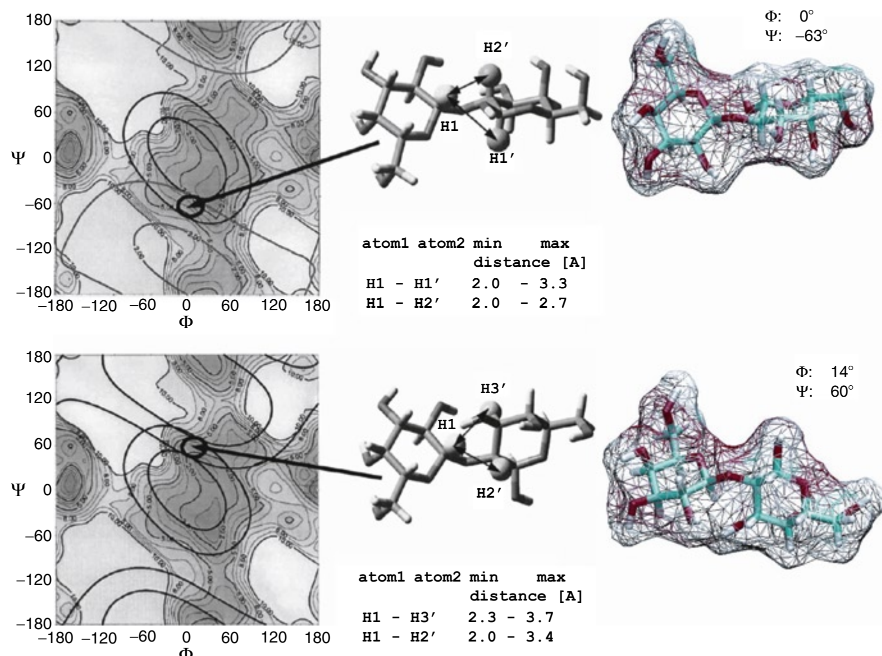

Fig. 7.4 Illustration of two bioactive conformers of the disaccharide Galβ1-2Gal, a potent ligand for galectins (*upper panel*) and the plant agglutinin viscumin (*bottom panel*). The two measured interproton contacts of the bound-state conformations of the disaccharide identify its position in the energy map defined by the dihedral angles *Ψ*/*Φ* of the glycosidic linkage, making the change in the *Ψ*-parameter obvious. A representation of the shape of the two conformers is given on the *right side*

can have distinct bioactivity profiles [8, 25, 26]. This finding of differential conformer selection has notable ramifications for drug design. It uncovers a chemical means to reduce cross-reactivity at the molecular level by designing conformationally restrained lectin ligands. This approach documents the medical perspectives for turning insights of the third dimension of the sugar code into innovative therapeutic agents targeting bacterial and plant toxins without reacting with human lectins. Their synthesis will give a clear direction if a lectin's fine specificity is characterized in detail, especially in comparison to related proteins.

As the term "galectin" implies, a key contact involves the β-galactoside part of the ligand. The galectins form a family of homologous proteins, and the current status of analysis indicates that they have potential to serve as specific regulators of a variety of cellular activities by virtue of homing in on certain physiological glycan ligands [1, 27, 28]. Beyond the interaction with ganglioside GM1, binding of the glycoprotein CD7 on activated T cells, the fibronectin receptor (α5β1-integrin) on tumor suppressor-positive carcinoma cells, and the tissue plasminogen activator in pancreatic cancer by galectin-1 sets distinct biosignaling in motion to elicit the

```
Gal-5  43 ....RFQINL.RCGGDIAFHLNPRFD...E.NAVVRNTQINNSWGPEERSLPGSMPFSRGQRFSVWILCEGHCFKV 109
 G4-N  45 RFHVNFAVGQ.DEGADIAFHFNPRFD...GWDKVVFNTMQSGQWGKEEKKK..SMPFQKGHHFELVFMVMSEHYKV 114
```

Fig. 7.5 Comparison of the sequences for prototype rat galectin-5 and the N-terminal domain of the rat tandem-repeat-type galectin-4 (G4-N or rGal-4N). Identical residues in both sequences are indicated as *white letters* on black background

particular cellular response [29–31]. Sequence diversification within this lectin family raises the question as to whether sequence divergence, which is apparent in the various family members (e.g. from the groups of prototype and tandem-repeat-type galectins, please see Fig. 7.5), will have an impact on ligand binding. In this respect, binding assays with free glycans are a valuable tool for comparative specificity analysis, e.g. in calorimetry [32], mapping the interaction between a ligand and the lectin in solution. Considering the importance of spatial parameters in natural glycoconjugates as regulators in lectin affinity [1], testing surface-immobilized glycoproteins constitutes a major refinement of specificity analysis. What is more, the selection of such test compounds will establish a panel that covers the range of common natural glycotopes (please see Table 7.1 for a listing) to enable fine-specificity determination as well as assessment of the impact of naturally clustered presentation [1, 33]. The next part of this chapter presents information on the experimental setup and answer the question as to whether the two mammalian lectins, whose sequences are presented in Fig. 7.5, have uniform or nonidentical binding properties to natural glycoproteins.

7.3 Natural Glycoproteins: Sensors to Map Functional Galectin Divergence?

A reliable and sensitive method for determining a lectin's binding specificity requiring minimum quantity of glycoprotein is the enzyme-linked lectin sorbent assay (ELLSA). A panel of glycoproteins well-characterized for the presentation of lectin-reactive glycotopes establishes the ligand side for the binding reaction. Their adsorption to the surface of a microtiter plate well leads to a ligand-presenting matrix. Keeping the lectin in solution precludes potentially harmful effects of adsorption on the receptor protein. The lectin will thus be confronted with different sets of glycotopes on the surface of the microtiter plate well, mimicking the interaction with cell surfaces. Quantitative monitoring of lectin binding requires labeling under activity-preserving conditions, conveniently using a fluorescent dye or by biotinylation, the extent of which is measured by two-dimensional gel electrophoresis and mass spectrometry [34, 35]. Extent of binding in the assay is routinely assessed to be dependent on carbohydrate and saturable. As shown in Fig. 7.6 for a galectin, the nature of the glycan chains, e.g. their status of sialylation, determines their reactivity. Typically for galectins, strong reactivity to terminal β-galactosides

Table 7.1 Codes and abbreviations of structural units in glycans of mammalian glycoproteins (gps) and glycosphingolipids [33]

	Codes[a]	Structural units	Sources
1	F	GalNAcα1-3GalNAc	Forssman pentasaccharide; animal tissue antigens mainly in glycosphingolipids
	F_{penta-}	GalNAcα1-3GalNAcβ1-3Galα1-4Galβ1-4Glc	
	F_{α}	GalNAcα1-3GalNAcα1-Ser/Thr	In mucin-type *O*-glycosylation of glycoproteins
	F_{β}	GalNAcα1-3GalNAcβ1-	Glycotope at the nonreducing end of F_{penta-}
2	A	GalNAcα1-3Gal	Human blood group A-active disaccharide
	A_h	GalNAcα1-3[LFucα1-2]Gal	Human blood group A-active trisaccharide
3	Tn	GalNAcα1-Ser/Thr	Tn antigen, only in mucin-type *O*-glycosylation of glycoproteins
4	T_{α}	Galβ1-3GalNAcα1-Ser/Thr	The mucin-type *O*-glycans sequence on the human erythrocyte membrane
	T_{β}	Galβ1-3GalNAcβ1-... ceramide	Brain glycoconjugates and gangliosides such as asialo-GM1 tetrasaccharide
5	I	Galβ1-3GlcNAc	Human blood group type I and II carbohydrate sequences. Branched or linear repeated II sequence is part of blood group I and i epitopes. I and II are precursors of ABH and Le^a, Le^b, Le^x, Le^y blood group-active antigens. Most of the lectins reactive with II are also reactive with I. Lectin Tri-II and mII determinants are present at the nonreducing end of carbohydrate chains in *N*- and *O*-glycans
6	II	Galβ1-4GlcNAc	
	II_{β}	Galβ1-4GlcNAcβ1-	
	Tri-II	Triantennary Galβ1-4GlcNAcβ1-	
	mII	Multivalent Galβ1-4GlcNAcβ1-	
7	B	Galα1-3Gal	Human blood group B-active disaccharide
	B_h	Galα1-3[LFucα1-2]Gal	Human blood group B-active trisaccharide
8	E	Galα1-4Gal	Blood group p^k and P_1 active disaccharide; sheep hydatid cyst glycoproteins, salivary glycoproteins of the Chinese swiftlet, glycosphingolipids in human erythrocytes and small intestine
9	L	Galβ1-4Glc	Constituent of mammalian milk
	L_{β}	Galβ1-4Glcβ1-	Lactosyl ceramides in brain and part of carbohydrate structures in gangliosides
10	P	GalNAcβ1-3Gal	Blood group P-related disaccharide; glycotope at the nonreducing end of globoside
	P_{α}	GalNAcβ1-3Galα1-	
11	S	GalNAcβ1-4Gal	Brain and asialo-GM2 disaccharide; human blood group Sd(a+)-related disaccharide in most human secretions, such as urine Tamm–Horsfall glycoprotein
	S_{β}	GalNAcβ1-4Galβ1-	

[a]α, β = anomer of sugars; m = multivalent

free of substitution is shared by the two mammalian family members (Table 7.2). Their sequences are listed in Fig. 7.5. Also, the common histo-blood group ABH substitutions are potent ligands [36–38]. This feature sets the galectins apart from the plant toxin ricin. The spatial architecture of this β-galactoside-specific lectin

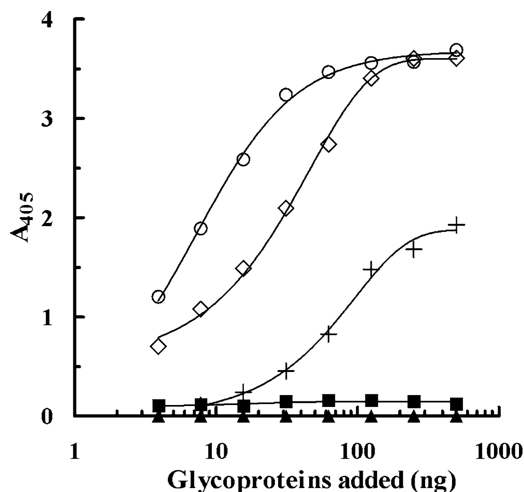

Fig. 7.6 Binding curves of rat galectin-5 to glycotopes in a solid-phase assay. The surface of 96-well microtiter plate wells was coated with glycoproteins dissolved in 0.05 M sodium carbonate buffer (0.05 M $NaHCO_3$/0.05 M Na_2CO_3, pH 9.6) overnight at 4°C (for details, please see [36]). Biotinylated rat galectin-5 (5 µg/ml) was used to determine and compare the extent of binding to the following glycoproteins (gps): cyst MSS 1st Smith degraded (*circle*), human asialo α_1-acid gp (*diamond*), human asialoglycophorin (+), human α_1-acid gp (*filled square*), and asialo ovine submaxillary mucin (*filled triangle*). Total volume of the assay was 50 µl. A_{405} was recorded after a period of 24 h

will obviously not tolerate these extensions to the core structure [39]. Thus, galectins and the β-galactoside-binding plant protein differ in fine specificity despite sharing binding capacity to β-galactosides. Next, even a conspicuous difference between the two galectins tested is revealed in Table 7.2: the binding properties to mucin-type *O*-glycans, especially the T_α-disaccharide, differ significantly. Thus, despite the close sequence similarity, the two galectins deviate from each other markedly in this parameter.

Our assay not only enables one to assess and compare the binding of lectins to surface-immobilized glycoproteins, but it also offers the attractive opportunity for fine-scale mapping by using free saccharides as inhibitors to glycotope binding. Basically, these inhibition assays are a control for specificity because they confirm the carbohydrate dependence of the binding [36–39]. Even more important, any difference in the relative potency identifies a disparity in the fine specificities between tested lectins, here especially the extension by α1-3-linked galactose and the effect of branching (Table 7.3). Moreover, the plant lectin proved relatively less sensitive to oligomer formation of the lactose core disaccharides (Table 7.3), as also summarized in Table 7.4. These results teach an important lesson that, despite sharing specificity to β-galactosides, the fine-specificity profiles can obviously be distinct between mammalian and plant lectins as well as among the members of this

Table 7.2 Comparison of the reactivity of two mammalian galectins (rGal-4N and rGal-5) for natural glycoproteins (gps) with that of a plant lectin (ricin)[a]

Glycoprotein (terminal epitope)[b]	Signal intensity for binding[c]		
	rGal-4N (ABH, I/II, T/Tn)	rGal-5 (A/B, I/II)	Ricin (II, T/Tn)
Terminal (I/II)-containing gps			
Cyst MSS 1st Smith degraded (I/II)	+++++	+++++	+++
Cyst Mcdon P-1 (I/II)	+++++	+++++	+++++
Cyst Tighe P-1 (I/II)	+++++	+++++	+++++
Human asialo α_1-acid gp (mII)	+++++	+++++	+++++
Asialofetuin (mII/I, T_α)	++++	++++	+++++
iII/Lac			
Pneumococcus type 14 polysaccharide	+++	+++	–
Blood group ABH, Le[a], Le[b], Le[x], and Le[y] *active gps*			
Cyst Mcdon (A_h > Le[b], Le[y])	+++++	++++	–
Cyst Beach phenol insoluble (B_h > Le[b], Le[y])	+++++	+++++	+
Cyst Tighe phenol insoluble (H, Le[a], Le[b], Le[x], Le[y])	++++	+	–
T_α/Tn-containing gps			
Asialo OSM (Tn, T_α, core 2 II)	+++++	–	+++++
Asialo PSM (Tn, T_α, A, A_h, H)	+++++	–	+++++
T_α-containing gps			
Human asialoglycophorin (T_α, Tn, mIIb/f)	+++++	++++	+++++
Antifreeze gp (T_α)	+++++	±	–
Crypto II, T_α/Tn-containing gps			
Human α_1-acid gp (α2-3/6 sialyl mII)	+	–	–
OSM (sialyl Tn, T_α, core 2 II)	–	–	–
PSM (sialyl Tn, T_α, A, A_h, H)	–	–	–

rGal-4N N-terminal domain of rat tandem-repeat-type galectin-4, *rGal-5* rat galectin-5, *ricin Ricinus communis* toxin

[a]For details, please see [36, 37, 39]

[b]The symbols in parentheses indicate the terminal epitopes and are listed in Table 7.1. iII/Lac = internal Galβ1-4Glc(NAc); m = multiantennary; mIIb/f = biantennary *N*-glycan with core fucosylation and bisecting GlcNAc

[c]The results were graded according to the spectrophotometric absorbance value at 405 nm (i.e. OD_{405}) after 2 h (ricin), 4 h (rGal-4N), and 24 h (rGal-5) incubation as follows: +++++, ($OD \geq 2.5$); ++++, ($2.5 > OD \geq 2.0$); +++, ($2.0 > OD \geq 1.5$); ++, ($1.5 > OD \geq 1.0$); +, ($1.0 > OD \geq 0.5$); ±, ($0.5 > OD \geq 0.2$); –, ($OD < 0.2$)

family of mammalian lectins. Such measurements thus open an intriguing route for further research, with clinical perspective and the fascinating possibility to be able to trace structure–activity relationships. To put this issue to a further test, we next examined the properties of two galectins from the same subgroup, i.e. the prototype galectins. Of note, both proteins tested, i.e. the chicken galectins (CGs) CG-1A/B

Table 7.3 Comparison of the relative potency of rGal-4N and rGal-5 for oligosaccharides with that of ricin[a]

Type of saccharide	Relative potency[b] rGal-4N (ABH, I/II, T/Tn)	rGal-5 (A/B, I/II)	Ricin (II, T/Tn)
Galβ1-3GlcNAcβ1-3Galβ1-4Glc (Type 1, Iβ1-3L)	7.7	10.0	0.9
Galβ1-4GlcNAcβ1-3Galβ1-4Glc (Type 2, IIβ1-3)	1.9	5.0	0.9
Galβ1-4GlcNAc |β1-6 Galβ1-4Glc |β1-3 Lacto-N-hexaose Galβ1-4GlcNAc [LNH; Di-II or II β1-3(IIβ1-6)L]	1.9	15.0	1.5
Galβ1-4GlcNAcβ1-2Man |α1-6 Galβ1-4GlcNAcβ1-2Manα1-3Manβ1-4GlcNAcβ1-4GlcNAc β1-N-Asn |β1-4 Galβ1-4GlcNAc Tri-antennary Galβ1-4GlcNAc (Tri-II)	>10.0 (11% inhibition)[c]	30.0	9.0
Galα1-3Galβ1-4GlcNAc (B active II)	1.2	50.0	0.4
Galβ1-4Glc (L)	1.0	1.0	1.0
Galα1-3Gal (B di)	0.2[d]	0.4	0.3
Galβ1-3GlcNAc (I)	0.2	3.3	0.8
Galβ1-4GlcNAc (II)	0.1	4.3	0.8

[a]Lectin abbreviations are listed in the footnote of Table 7.2. Inhibitory potency of oligosaccharides on binding of rGal-4N (50 ng/50 μl) to a Galβ1-containing gp, cyst MSS 1st Smith degraded gp (1 ng/50 μl) [37]; of rGal-5 (125 ng/50 μl) to a Galβ1-containing gp, cyst Beach P-1 (5 ng/50 μl) [36]; of ricin (2.5 ng/50 μl) to a Galβ1-containing gp, asialofetuin (25 ng/50 μl) [39]

[b]Relative potency of Galβ1-4Glc (L) is given as 1.0, leading to a normalization

[c]Could not reach 50% inhibition

[d]Extrapolation

Table 7.4 Comparison of the binding properties of rGal-4N and rGal-5 with ricin[a]

Carbohydrate specificity	Type of lectin: rGal-4N (ABH, I/II, T/Tn)	rGal-5 (A/B, I/II)	Ricin (II, T/Tn)
Monosaccharide specificity	β-Anomer of Gal and slightly enhanced by *N*-acetyl group in GalNAc	β-Anomer of Gal and enhanced by *N*-acetyl group in GalNAc	β-Anomer of Gal and reduced by *N*-acetyl group in GalNAc
Reactivity toward disaccharide structural units expressed in decreasing order (based on nanomoles comparison)	Galβ1-4Glc (L)>Galβ1-3GalNAc (T)>Galα1-3Gal (B)≥Galβ1-3GlcNAc (I)>Galβ1-4GlcNAc (II); Fucα1-2Gal (H) was inactive	Galβ1-4GlcNAc (II)>Galβ1-3GlcNAc (I)>Galβ1-3GalNAc (T)>Galβ1-4Glc (L)>GalNAcβ1-3Gal (P)>Galα1-3Gal (B)>GalNAcβ1-4Gal (S)≥GalNAcα1-3GalNAc (F)>GalNAcα1-3Gal (A); GalNAcα1-Ser/Thr (Tn) and Galα1-4Gal (E) were inactive	Galβ1-4Glc (L)>Galβ1-4GlcNAc (II)≥Galβ1-3GlcNAc (I)>Galβ1-3GalNAc (T)>Galα1-3Gal (B)≥Galα1-4Gal (E)≥GalNAcβ1-3Gal (P)>GalNAc α1-Ser/Thr (Tn)≥GalNAcβ1-4Gal (S)>GalNAcα1-3Gal (A); GalNAcα1-3 GalNAc (F) was inactive
The most active β-galactoside	[Fucα1-2]Galβ1-3GlcNAcβ1-3Galβ1-4Glc (H active Iβ1-3L) and Fucα1-2Galβ1-4Glc (H active L)	Galα1-3Galβ1-4GlcNAc (B active II) and Galα1-3Galβ1-4GlcNAcβ1-3Galβ1-4Glc (B active IIβ1-3L)	Galβ1-4GlcNAc core
Ratio of glycotope clusters (simple multivalent form)/ monomeric II	Triantennary glycopeptides with mostly type II termini were inactive	Triantennary glycopeptides with mostly type II termini were seven times more active than monomeric II	Triantennary glycopeptides with mostly type II termini were 12 times more active than monomeric II
Substituted branch-end glycans	Histo-blood group ABH precursor (equivalent) gps and enhanced strongly by blood group A, B, and H determinant sugar	Histo-blood group ABH precursor (equivalent) gps and enhanced strongly by blood group A, B determinant sugar	Histo-blood group precursor (equivalent) gps but hindered by ABH histo-blood group determinants
Ratio of complex polyvalent I/II glycotopes in natural glycoproteins/ monomeric II	8.8×10^5 times more active than monomeric II	3.2×10^2 times more active than monomeric II	4.6×10^2 times more active than monomeric II

[a]For detailed information on the binding properties of the lectins, please see [36, 37, 39]

(formerly referred to as C-16/C-14) closely related to mammalian galectin-1 [40], had been subjected to crystallographic analysis so that their crystal structures were available [41, 42]. In addition, the DNA sequence traits beyond the proteins' primary structures were known.

7.4 The Case Study of Two Prototype CGs

Owing to complete genome sequencing, the intended comparison can thus include the promoter region in addition to the coding sequence. Consequently, insights into divergence of regulatory mechanisms of gene expression can be inferred, too. In other words, the question can be answered as to whether the two genes have acquired their own characteristic expression profiles. The ensuing computational processing using algorithms to spot putative sites for binding of transcription factors revealed a notable degree of sequence divergence (Table 7.5). The arising assumption for disparate expression profiles was tested immunohistochemically. The preparation of noncross-reactive antibodies against these two lectins facilitated mapping their tissue presence [40]. The divergence of the two gene sequences, originally stemming from a duplication event at around the separation of birds from mammals in the phylogenetic tree approximately 3×10^8 years ago [43], indeed translated into nonidentical localization patterns (Table 7.6). These findings, together with sequence variations noted around key residues relevant for lectin activity, such as the Trp moiety at positions 68 (CG-1A) or 70 (CG-1B) or the His moiety at equivalent positions 52/54 (Fig. 7.7), nourish the suggestion that the interaction profiles with glycans will also be different. This question was answered by running the galectin–glycan binding assays under identical conditions [44, 45]. As summarized in Table 7.7, although both galectins showed pronounced preference for I/II-containing human blood group precursor equivalent glycoproteins, reactivity differed significantly for multiantennary *N*-glycans and the AB blood group epitopes. The inhibitory potency of di- and oligosaccharides was rather comparable (Table 7.8), leading us to the overall comparison given in Table 7.9. A prominent feature is that the polyvalent nature of glycotopes could be sensed discriminatively by these two closely related galectins (Table 7.9).

With sequences and crystal structures being available, this experimental analysis could be taken to the level of detailed structure–activity computations. As a means to move beyond common rigid-body docking, we implemented the high ambiguity-driven docking (HADDOCK) procedure for lectin interactions [45, 46]. It allowed lectins and ligands to maintain full flexibility, thereby yielding a detailed view on structural and energetic aspects. As stated above, in explaining the third dimension of the sugar code, the ligands were invariably accommodated by the lectin sites in their low-energy conformation, and a detailed calculation for energy of interaction disclosed the contribution of each amino acid to binding [45]. Files of movies showing the dynamic process of contact building are available at http://www.nmr.chem.uu.nl/haddock/movies/ [45]. A static view (snap-

Table 7.5 Compilation of putative binding sites for transcription factors in the proximal promoter regions of the genes for chicken galectins CG-1A/CG-1B

	CG-1A					CG-1B				
Motif for factor	Position (orientation)	Sequence	Core score	Matrix score	Program	Position (orientation)	Sequence	Core score	Matrix score	Program
AML-1a (Runx)	−1742 (−)	ACCAGa	1.0	1.0	PM	−1897 (+)	tGCGGT	1.0	1.0	PM
	−1518 (+)	tGTGGC	1.0	1.0	PM	−1761 (+)	aGTGGT	1.0	1.0	PM
	−1296 (+)	tGTGGC	1.0	1.0	PM	−1288 (+)	cGTGGT	1.0	1.0	PM
	−1226 (+)	tCTGGT	1.0	1.0	PM	−1280 (−)	ACCACt	1.0	1.0	PM
	−546 (−)	GCCACa	1.0	1.0	PM	−953 (−)	ATCACa	1.0	1.0	PM
	−95 (+)	tGTGAT	1.0	1.0	PM	−939 (+)	tGTGAT	1.0	1.0	PM
						−898 (−)	GCCACa	1.0	1.0	PM
						−511 (−)	ACCACa	1.0	1.0	M, PM
						−447 (−)	ACCACa	1.0	1.0	M, PM
						−263 (−)	GCCACa	1.0	1.0	PM
						−49 (+)	tGTGGT	1.0	1.0	M, PM
Barbie-Box		n.f.				−64 (+)	atttAAAGGcaaggg	1.0	1.0	PM
Cdx-1	−1892 (+)	CATTCtg	1.0	1.0	PM	−1549 (−)	gtCAAAT	1.0	1.0	PM
	−1630 (+)	aATAATg	0.984	0.981	M	−904 (+)	ATTGAag	1.0	1.0	PM
C/EBPα, β; CHOP-10	−1592 (−)	gggtttcGCAAAgc	0.984	0.973	M	−1658 (+)	gaTGTGCtaacat	0.952	0.953	M
	−1291 (+)	cccTTGCAcaaata/ccCTTGCacaaata	1.0/0.984	0.979/0.975	M	−1317 (−)	ccatgaGCAAAta	0.972	0.962	M
	−1006 (+)	agTTTGCtgaaggt	0.984	0.97	M	−813 (−)	ctttttCCAAAgg	1.0	0.973	M
	−929 (−)	cgctatCCACAga	0.979	0.96	M					
	−849 (−)	cattttTGCAAgca	1.0/0.996	0.973/0.99	M; PM					
	−773 (+)	acTTTGCtaagca	0.972	0.968	M					
COMP1	−1482 (+)	tggaatGGAAGcgactaaagcctc	0.964	0.892	PM	−1955 (−)	gcccttcccatgcCAATCctgtcc	1.0	0.84/0.936	M[a]; PM
	−556 (+)	catcagCATTGccacaaacagttg	0.914	0.812	M[a]					
	−134 (−)	cccttcttgcctgCCCCCcacacc	0.893	0.902	PM					
CP2		n.f.				−105 (−)	CTGGGgttggg	1.0	1.0	PM
c-Myc/Max	−1454 (−)	taacACGTGatt	1.0	0.995	M[a]		n.f.			
CUTL-1	−1446 (+)	gaTTGATccc	0.996	0.997	M[a]	−1139 (+)	cATTGAtggg	0.929	0.925	M[a]
						+132 (+)	caTCGATccc	1.0	0.999/1.0	M[a]; PM

(continued)

Table 7.5 (continued)

Motif for factor	CG-1A Position (orientation)	CG-1A Sequence	CG-1A Core score	CG-1A Matrix score	CG-1A Program	CG-1B Position (orientation)	CG-1B Sequence	CG-1B Core score	CG-1B Matrix score	CG-1B Program
E47 (Hand-1)	−673 (+)	tctagagTCTGGaatg	1.0	0.974	M[a]		n.f.			
Elk-1 (TCF-A)		n.f.				−1128 (+)	ccaactGGAAGtaccc	1.0	0.928	M[a]
En-1	−1671 (−)	acAACAC	1.0	1.0	PM	−1494 (+)	GTATTcc	1.0	1.0	PM
	−1207 (+)	GTTTTgc	1.0	1.0	PM	−986 (+)	GTTTTgc	1.0	1.0	PM
	−817 (−)	ccACTAC	0.996	0.992	M	−825 (+)	GTAATtg	1.0	1.0	M
						−825 (+)	GTAATtg	1.0	1.0	PM
						−612 (−)	ccATTTT	1.0	1.0	PM
						−333 (−)	ccATTTT	1.0	1.0	PM
						−302 (−)	agAATAC	1.0	1.0	PM
Evi-1	−1234 (−)	aTATTTcctctggta	0.861	0.904	PM	−1551 (−)	aGGTCAaatttcctt	0.842	0.903	PM
	−1137 (−)	aGCTCAgctttggtt	0.842	0.909	PM	−1456 (−)	agTCTCTaatc	0.988	0.981	PM[a]
						−1308 (−)	aaTATCTagaa	1.0	0.982	PM[a]
						+20(−)	gCATCAtgtcttgtg	0.921	0.927	PM
FoxD3	−262 (−)	aagcaAACAAtc	1.0	0.967	M[a]		n.f.			
GATA-3	−1760 (−)	caCTATCtg	1.0	0.995	M		n.f.			
GKLF	+78 (+)	gaagggaggAAGGG	1.0	0.944	M	−1850 (−)	CCTTCctttccttg	0.933	0.911	M
						−1489 (−)	CCCTCtcctgttgt	0.965	0.901	M
						−980 (−)	CCTCTtctatttct	0.949	0.901	M
						−861 (−)	CCTCTtctttcagt	0.949	0.906	M
						−694 (−)	CCCATccccttgtt	0.952	0.910	M
						−617 (−)	CCTCCccatttttc	0.914	0.9	M
						+40 (+)	agggaaggaATGGG	0.952	0.906	M
GR	−1248 (−)	agaGGACAcatgggatatt	0.989	0.97	M	−467 (+)	tcagtcctctgTGTACttg	0.984	0.974	M
HOX A3	−1703 (−)	aaatGTGCA	1.0	1.0	PM	−1990 (−)	aactGTCCT	1.0	0.995	PM
	−1619 (−)	agcaCTGAA	1.0	1.0	PM	−1902 (+)	TTCAGtgcg	1.0	1.0	PM
	−1284 (−)	acaaATATG	0.968	0.974	M	−1801 (−)	acctTTAGG	1.0	0.953/1.0	M; PM
	−1011 (+)	AGCACagtt	1.0	0.995	PM	−1575 (−)	caccAAAGG	1.0	0.995	PM
	−999 (+)	TGAAGgttg	1.0	1.0	PM	−1463 (+)	ATGAGttag	1.0	0.995	PM
	−938 (+)	CCTTCctgg	1.0	1.0	PM	−1213 (−)	gcccAAAGG	1.0	0.995	PM

	−580 (−)	ctaaCTTAA	1.0	0.995	PM	−927 (−)	acaaATATG	0.968	0.974	M
	−362 (−)	gccaTCTCG	1.0	1.0	PM	−274 (+)	TTGACttcg	1.0	0.995	PM
	−289 (−)	acagGAAGG	1.0	1.0	PM	−200 (−)	catcCTGGA	1.0	0.995	PM
	−186 (+)	CCTACttta	1.0	0.995	PM	−116 (+)	CCTATtggt	0.985	0.982	M
						+101 (+)	AGCAGagtg	1.0	0.995	PM
Lmo2-complex	−1760 (−)	cacTATCTg	1.0	1.0	PM		n.f.			
Msx-1	−1316 (−)	cAGTTAcct	0.988	0.927	M	−1708 (−)	cAGGCAatg	1.0	1.0	PM
	−1184 (−)	cAGTTAcag	0.988	0.991	M	−940 (+)	ttgTGATTt	1.0	1.0	PM
	−591 (−)	cACATAcag	0.928/1.0	0.95/1.0	M; PM	−827 (+)	tagTAATTg	1.0	0.935	M
	−544 (−)	cACAAAcag	1.0	1.0	PM	−827 (+)	tagTAATTg	1.0	1.0	PM
	−182 (+)	cttTAAATg	0.996	0.958	M	−768 (+)	cacAAAGTg	1.0	1.0	PM
	−78 (+)	aagTGATTt	1.0	1.0	PM					
Myogenin/NF-1	−338 (−)	catgggacgaaaGCCAAcatacactgctg	1.0	0.779	M[a]	−1802 (+)	aacctttaggttTTGGCatctctccatag	1.0	0.801	M[a]
MZF-1	−1427 (+)	atgctAGGGActg	1.0	0.989	PM	−1432 (+)	tcgGAGGA	1.0	1.0	PM
	−757 (+)	agtGGGGA	1.0	1.0	M, PM	−690 (−)	TCCCCttg	1.0	1.0	PM
						−619 (−)	ttcCTCCCcattt	1.0	0.996	PM
						−482 (−)	TCCCCaaa	1.0	1.0	PM
NF-E2	−50 (+)	cGCTGAgtcac	1.0	0.986/1.0	M[a]; PM[a]		n.f.			
NRF-2	−939 (−)	ccCTTCCtgg	1.0	1.0	PM[a]		n.f.			
Pax-2	−1458 (−)	tctctaacacGTGATtgat	0.902	0.862	M	−1947 (−)	catgccaatcCTGTCctga	0.972	0.963	PM
	−1048 (+)	tagtGTCAGgagttggagt	0.98	0.971	PM	−1879 (−)	agtgctgccaAGAACatgg	0.971	0.963	PM
	−1009 (+)	cacaGTTTGctgaaggttg	0.992	0.989	PM	−1819 (+)	taagGACATgtatcctcaa	0.992	0.973	PM
	−979 (−)	gcagggcagcGTGTCtcct	1.0	0.967	PM	−1592 (+)	aacaGACAAgcatgggcca	0.98	0.969	PM
	−48 (+)	ctgaGTCACgttggctgtg	0.992	0.982	PM	−1418 (−)	ttgggcatgcGTGACtccg	1.0/0.992	0.892/0.981	M; PM
	+35 (−)	ggggggtggcTTGTCttgt	0.98	0.968	PM	−1345 (+)	aggaGACATgatttttaaa	0.992	0.962	PM
						−1312 (−)	agcaaatatcTAGAAattt	0.98	0.965	PM
						−1215 (−)	ctgcccaaagGCTGGaaca	0.992	0.967	PM
						−1101 (−)	ggggttctgcATGACtggg	0.991/1.0	0.879/0.997	M; PM
						−919 (+)	gcaaGACAAggaaagattg	0.98	0.967	PM
						−653 (−)	gtgcaagagcCTGACactg	0.98	0.974	PM
						−601 (+)	agagGACAGgcttgctgtc	0.972	0.962	PM
						−564 (−)	atctacaagaGCTGGtggc	0.992	0.965	PM

(continued)

Table 7.5 (continued)

	CG-1A					CG-1B				
Motif for factor	Position (orientation)	Sequence	Core score	Matrix score	Program	Position (orientation)	Sequence	Core score	Matrix score	Program
						−276 (−)	cattgacttcGTGGCcaca	0.949	0.969	PM
						−128 (+)	cttaTTCTCgatcctattg	1.0	0.961	PM
						+15 (−)	cagcagcatcATGTCttgt	0.992	0.974	PM
						+23 (+)	tcatGTCTTgtgtgagtag	0.843	0.848	M
						+86 (−)	atgccttggcTTGACagca	0.989	0.987	PM
Pax-4a	−1420(−)	ggactggggtaagcttaaggtggcaTTTCT	0.95	0.786	M	−1420 (−)	ccttgggcatgCGTGActccg	0.977	0.875	M[a]
						−1076 (−)	gGGGTGggttca	1.0	0.985	M
						−893 (−)	aggcttggcctgctgcctgatatatTTTTT	1.0	0.813	M
						−213 (−)	gGGGTGggatta	1.0	0.995	M
Pax-5 (BSAP)	−325 (+)	ccaacataCACTGctgcgtagcagctaa	1.0	0.913	M[a]		n.f.			
Pax-6		n.f.				−665 (−)	tgcctagagtaGGTGCaagag	0.987	0.977	PM
Pbx-1	−1434(+)	taaCATCAt	1.0	0.989	PM	−1338 (−)	aTGATTttt	1.0	0.996	PM
	−275 (+)	tagGATCAt	0.986	0.982	PM	−323 (−)	aTGATTcct	1.0	0.995	PM
	−241 (+)	aaaAATCAt	1.0	0.996	PM					
POU2F1/2 (Oct-1)	−870 (+)	ctctatGCAAAgctc	1.0	0.901	M[a]	−1276 (+)	cttcatGCAAAggaa	1.0	0.929	M[a]
						−828 (+)	gtaGTAATtgggt	1.0	1.0	PM
POU3F1 (Oct-6, Tst-1)	−1686 (+)	gggGAATTggaatgg	1.0	0.935	M	−852 (−)	tcagtgtAATTAcaa	0.902	0.919	M
						−849 (+)	gtgTAATTacaattg	0.902	0.930	M
Sp1/GC box	−1057 (+)	ggGGCAGggt	0.972	0.978	M	−1218 (−)	aacCTGCCca	1.0	0.994	PM
	−721 (−)	accCAGCCct	0.962	0.954	M					
	−228 (−)	atcCTGCCca	1.0	0.994	PM					
SRY	−1570 (+)	AAACAaa	1.0	1.0	M; PM	−1995 (+)	AATCTaa	1.0	1.0	PM
	−1566 (+)	AAACAca	1.0	1.0	PM	−1976 (−)	tgTCCTT	1.0	1.0	PM
	−965 (−)	ctCCTTT	1.0	1.0	PM	−1818 (+)	AAGGAca	1.0	1.0	PM
	−685 (−)	tcTCTTT	1.0	1.0	PM	−1462 (−)	tgAGTTA	1.0	1.0	PM

	−509 (−)	gaTGTTT	1.0	1.0	PM	−995 (−)	tgTGTTT	1.0	1.0	PM
	−374 (+)	TAACAaa	1.0	1.0	PM					
	−217 (+)	AAAGAac	1.0	1.0	PM					
STAT1α/β, 2-6	−291 (−)	ttacAGGAA	1.0	1.0	PM[a]		n.f.			
TATA-Box	−1203 (−)	tgccctctTTTATat	1.0	0.985	M		n.f.			
USF	−1455 (+/−)	ctaaCACGTgattg/ctaacACGTGattg	1.0	0.995	M	−329 (+/−)	tttcCACGTgggat/ tttccACGTGggat	1.0	0.963	M
	−1096 (+/−)	atgcCACGTgaatg/atgccACGTGaatg	1.0	0.974	M					
	−1094 (−)	gccaCGTGAa	1.0	0.974	M[a]					
v-Maf		n.f.				−545 (−)	aaggtctaggTCAGCagct	1.0	0.896	M[a]
v-Myb		n.f.				−1927 (+)	aatAACGGgc	1.0	0.985	M[a]
YY1		n.f.				−617 (+)	cctcCCCATttttccaa	0.996	0.988	M
ZEB (AREB6)	−1569 (−)	aaCAAACac	1.0	1.0	PM		n.f.			
	−261 (−)	agCAAACaa	1.0	1.0	PM					

Putative transcription factor binding sites found in the proximal promoter regions (from −2,000 bp upstream of the transcription (in the case of CG-1A: translation) start site to +150 bp downstream) of chicken galectins CG-1A and CG-1B. *Position (orientation)*: Position (relative to the transcription start site) and orientation of the putative binding site ((+): sense strand; (−): antisense strand). *Sequence*: DNA sequence of the putative binding site; core sequence is set in uppercase. *Core score*: score for the matrix core sequence; *matrix score*: score for the entire matrix sequence. *Program*: the algorithm yielding the displayed hit – M: Match™; PM: P-Match™Names of factors: *AML1a* acute myeloid leukemia gene 1a (Runx/runt homology domain), *Barbie-Box* barbiturate-inducible element, *Cdx-1* chicken homeobox gene cdxA product, *C/EBP* CCAAT/enhancer binding protein, *CHOP-10* C/EBP homologous protein 10, *COMP1* cooperates with myogenic proteins 1, *CP2* CCAAT-binding protein 2, *c-Myc/Max* c-Myc (cellular counterpart of viral myelocytomatosis oncogene): Max heterodimer, *CUTL-1* cut-like homeodomain protein 1, *E47 (Hand-1)* E-box binding protein (heart- and neural crest derivatives-expressed 1), *Elk-1 (TCF-A) ets*-like factor 1 (transcription factor A), *En-1* engrailed 1, *Evi-1* ectopic viral integration site 1-encoded factor, *FoxD3* fork head box D3, *GATA-3* GATA-box binding factor 3, *GKLF* gut-enriched Krueppel-like factor, *GR* glucocorticoid receptor, *HOX A3* homeobox cluster protein A3, *Lmo2* LIM-only protein 2, *Msx-1* msh-like homeobox protein 1, *myogenin/NF-1* myogenin/nuclear factor 1, *MZF-1* myeloid zinc finger protein 1, *NF-E2* nuclear factor erythroid 2, *NRF-2* nuclear respiratory factor 2, *Pax* paired box gene product, *Pbx-1* pre B-cell leukemia transcription factor 1, *Sp1* stimulatory protein 1, *SRY* sex-determining region Y gene product, *STAT* signal transducer and activator of transcription 1, p91, *USF* upstream stimulatory factor, *v-Maf* avian musculoaponeurotic fibrosarcoma virus AS42 nuclear oncoprotein, *v-Myb* viral myoblastoma oncoprotein, *YY1* Yin and Yang 1 (nuclear factor E1, delta factor), *ZEB* zinc finger E-box-binding protein, Atp1a1 regulatory element binding protein 6 (AREB6), *n.f.* no significant sequence hit found; these differences are highlighted in *gray*

[a]Hits detected by "high-quality" matrices

It is possible to obtain more than one hit for the same factor at the same or almost the same position. This is due to the use of two independent search algorithms and more than one factor-specific weight matrix. For the listed cases of such "multiple hits" at the same site, their scores and/or orientation differ significantly

Table 7.6 Immunohistochemical profiling for the presence of the prototype chicken galectins CG-1A and CG-1B in various organs of adult animals[a]

Type of organ	Staining intensity[b]	
	CG-1A	CG-1B
Larynx		
Respiratory epithelium	++[c]	–
Lamina propria mucosae	–	+
Trachea		
Respiratory epithelium	++[c]	–
Lamina propria mucosae	–	++
Lung		
Respiratory epithelium	–	+++[c]
Connective tissue	–	+
Esophagus		
Lamina propria mucosae	–	++
Gut		
Epithelial lining of villi and intestinal glands	–	–
Lamina propria mucosae	–	+
Liver		
Hepatocytes (parenchyma)	+++[c]	–
Kidney		
Epithelium		
Collecting ducts (medulla)	–	–
Proximal/distal tubules (MTN I, MTN II, RTN)	+++[c]	–
Skin		
Epidermis		
Stratum corneum	–	–
Stratum intermedium	–	+++[d]
Stratum basalis	–	+[c]
Dermis	–	++
Subcutis	–	+

MNT-I mammalian-type nephron I (juxtamedullar), *MNT-II* mammalian-type nephron II (mid-cortical), *RTN* reptilian-type nephron (superficial)
[a]For details, please see [40]
[b]The intensity of staining is grouped into categories: – no staining, + weak staining, ++ medium staining, +++ strong staining
[c]Only cytoplasmic
[d]Cytoplasmic and nuclear

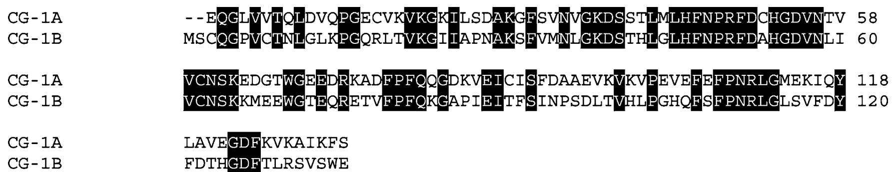

Fig. 7.7 Comparison of the sequences for chicken galectins CG-1A (C-16) and CG-1B (C-14). Identical residues in both sequences are indicated as *white letters* on black background. Please note the sequence differences in the vicinity of residues indispensable for sugar binding, such as His 52/54 or Trp 68/70

Table 7.7 Comparison of reactivity of two prototype chicken galectins for natural glycoproteins (gps)[a]

Glycoprotein (terminal epitope)[b]	Signal intensity for binding[c]	
	CG-1A (I/II)	CG-1B (AB, I/II)
Terminal (I/II)-containing gps		
Cyst Beach P-1 (I/II)	+++++	+++++
Cyst Mcdon P-1 (I/II)	+++++	+++++
Cyst Tighe P-1 (I/II)	+++++	ND
Human asialo α_1-acid gp (mII)	++++	+
Asialofetuin (mII/I, T_α)	+++	–
iII/Lac		
Pneumococcus type 14 polysaccharide	++++	+
Blood group ABH, Le[a], Le[b], Le[x], and Le[y] active gps		
Cyst Mcdon (A_h>Le[b], Le[y])	+	+++++
Cyst Beach phenol insoluble (B_h>Le[b], Le[y])	+	+++++
Cyst Tighe phenol insoluble (H, Le[a], Le[b], Le[x], Le[y])	–	–
T_α/Tn-containing gps		
Asialo OSM (Tn, $T_{\alpha,}$ core 2 II)	–	–
Asialo PSM (Tn, T_α, A, A_h, H)	–	ND
T_α-containing gps		
Human asialoglycophorin (T_α, Tn, mIIb/f)	+	–
Antifreeze gp (T_α)	–	–
Crypto II, T_α/Tn-containing gps		
Human α_1-acid gp (α2-3/6 sialyl mII)	–	–
OSM (sialyl Tn, T_α, core 2 II)	–	–
PSM (sialyl Tn, T_α, A, A_h, H)	–	–

CG-1A/B chicken galectin-1A/B, *iII/Lac* internal Galβ1-4Glc(NAc), *m* multiantennary, *mIIb/f* biantennary *N*-glycan with core fucosylation and bisecting GlcNAc, *ND* not determined

[a]For details, please see [44, 45]

[b]The symbols in parentheses indicate the terminal epitopes and are listed in Table 7.1

[c]The results were graded according to the spectrophotometric absorbance value at 405 nm (i.e. OD_{405}) after 4 h incubation as follows: +++++, (OD≥2.5); ++++, (2.5>OD≥2.0); +++, (2.0>OD≥1.5); ++, (1.5>OD≥1.0); +, (1.0>OD≥0.5); ±, (0.5>OD≥0.2); –, (OD<0.2)

Table 7.8 Comparison of relative potency of CG-1A and CG-1B for oligosaccharides[a]

Type of saccharide	Relative potency[b]	
	CG-1A (I/II)	CG-1B (AB, I/II)
Galβ1-4GlcNAc |β1-6 Galβ1-4Glc |β1-3 Galβ1-4GlcNAc Lacto-N-hexaose [LNH; Di-II or II β1-3(IIβ1-6)L]	3.5	4.4
Galβ1-4GlcNAcβ1-2Man |α1-6 Galβ1-4GlcNAc β1-2Manα1-3Man β1-4GlcNAc β1-4GlcNAc β1-N-Asn |β1-4 Gal β1-4GlcNAc Triantennary Galβ1-4GlcNAc (Tri-II)	6.0	5.7

(continued)

Table 7.8 (continued)

Type of saccharide	Relative potency[b]	
	CG-1A (I/II)	CG-1B (AB, I/II)
Galβ1-4GlcNAc (II)	2.0	2.0
Galβ1-4Glc (L)	1.0	1.0
Galα1-3Gal (B di-)	0.01	0.02
Galβ1-3GlcNAc (I)	0.1	0.3

[a]Galectin abbreviations are listed in the footnote of Table 7.7. Inhibitory potency of oligosaccharides on binding of CG-1A (125 ng/50 μl) to a Galβ1-containing gp, cyst MSS 1st Smith degraded gp (50 ng/50 μl) [44], and of CG-1B (125 ng/50 μl) to a Galβ1-containing gp, cyst Beach P-1 (50 ng/50 μl) [45]

[b]Relative potency of Galβ1-4Glc (L) is given as 1.0 for normalization

Table 7.9 Comparison of the binding properties of two prototype chicken galectins, CG-1A and CG-1B[a]

Carbohydrate specificity	Type of galectin	
	CG-1A	CG-1B
Monosaccharide specificity	β-Anomer of Gal, reactivity reduced by *N*-acetyl group in GalNAc	β-Anomer of Gal and slightly enhanced by *N*-acetyl group in GalNAc
Reactivity toward disaccharide structural units expressed in decreasing order (based on nanomoles comparison)	Galβ1-4GlcNAc (II) > Galβ1-4Glc (L) > Galβ1-3GlcNAc (I) > Galα1-3Gal (B); Galβ1-3GalNAc (T) and Galα1-4Gal (E) were inactive	Galβ1-4GlcNAc (II) > Galβ1-4Glc (L) > GalNAcβ1-3Gal (P) > GalNAcα1-3Gal (A) ≥ GalNAcα1-3GalNAc (F) ≥ Galβ1-3GalNAc (T) ≥ Galβ1-3GlcNAc (I) Ð Galα1-3Gal (B) > Fucα1-2Gal (H) Ð Galα1-4Gal (E)
The most active β-galactoside	Galβ1-4GlcNAc (II) mainly, extension to ABH epitopes reducing activity	Galβ1-4GlcNAcβ1-3Galβ1-4Glc (IIβ1-3L) and Galβ1-4GlcNAc (II) and its H-type derivative
Ratio of glycotope clusters (simple multivalent form)/monomeric II	Triantennary glycopeptides with mostly type II termini and 2,4,2-branching pattern from asialofetuin was three times more active than monomeric II	Triantennary glycopeptides with mostly type II termini and 2,4,2-branching pattern from asialofetuin was three times more active than monomeric II
Substituted branch-end glycans	Histo-blood group precursor (equivalent) gps but hindered by ABH histo-blood group determinants	Histo-blood group ABH precursor (equivalent) gps and enhanced strongly by blood group A, B determinant sugar
Ratio of complex polyvalent glycotopes in natural glycoproteins/monomeric II	Only 5.5 times more active than monomeric II	77 Times more active than monomeric II
The most complementary chain length	Galβ1-4GlcNAc (II) and Galβ1-3GlcNAcβ1-3Galβ1-4Glc (Iβ1-3L)	H active Iβ1-3L and Galβ1-4GlcNAcβ1-3Galβ1-4Glc (IIβ1-3L)

[a]For detailed information on the binding properties of the galectins, please see [45]

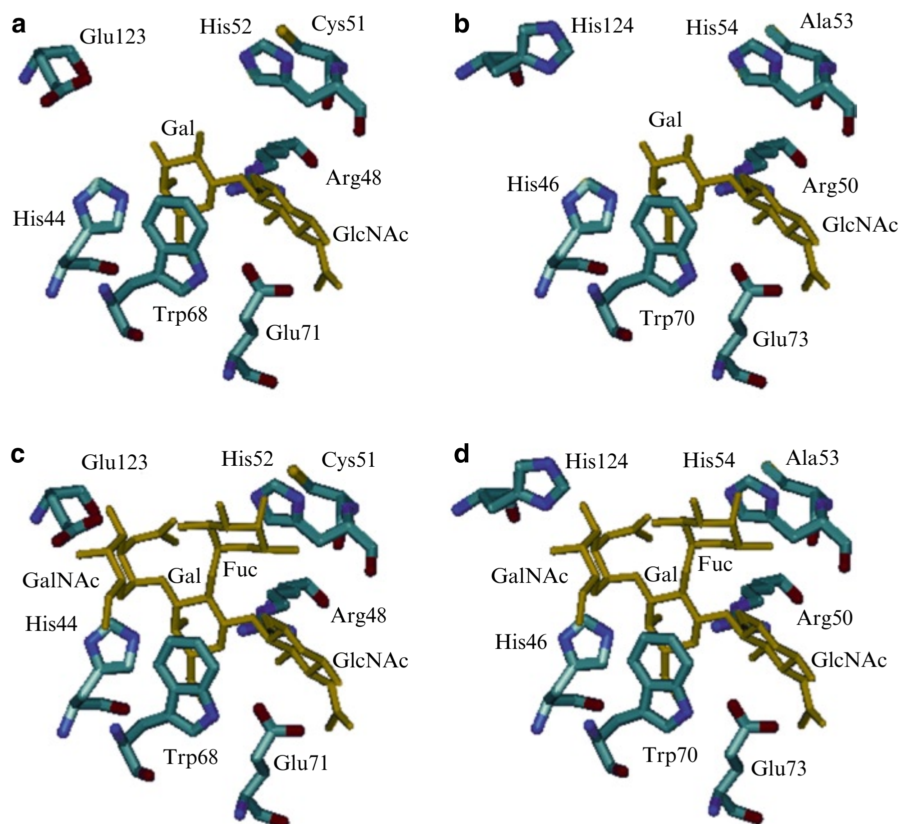

Fig. 7.8 Graphical illustration of the carbohydrate recognition domains (CRDs) of CG-1A (C-16) (*left panel*) and CG-1B (C-14) (*right panel*) pinpointing the contact sites for Galβ1-4GlcNAc (*upper panel*) and for histo-blood group A tetrasaccharide. The files for seeing the molecules in motion are available at http://www.nmr.chem.uu.nl/haddock/movies/. The positioning of the key contact sites was deliberately kept constant for direct comparison. Both sugars remained in low-energy conformations, and the comparison identified the regions of each CRD responsible for contact to the α1-2/3 substitutions on the core galactose moiety [45]

shot) of the binding-site architecture and the major contacts is illustrated in Fig. 7.8. With this information, it is possible to gage the impact of amino acid substitutions between the proteins on ligand binding, e.g. that of the Ala53/Cys51 exchange (see Fig. 7.8) on the orientation of equivalent His54/His52 residues and the resulting consequences on energetic terms [45]. Clearly, the outlined data attest that even closely related members of the galectin family should not be considered as redundant modules. They prompt further analysis to tie the presence of amino acid substitutions to functional divergence. The strategic combination of the binding assays with the flexible ligand docking in silico and calculation of binding-energy terms is expected to be of pivotal importance for understanding the initial step of translating the sugar code in molecular terms [47, 48]. Also, with the emergence of insights into lectin involvement as endogenous effectors in

disease manifestation and progression, drug design on human galectins becomes a therapeutical perspective [49, 50], the concept and data presented herein serving as a proof-of-principle case study.

7.5 Summary

The glycan part of cellular glycoconjugates harbors bioactive signals encoded in glycotopes. They are decoded and translated into signaling and cellular responses by lectins such as those of the family of galectins sharing the β-sandwich folding. As a physiologically relevant means for evaluating the binding properties of these medically important lectins to carbohydrate ligands in their natural structural context, we have introduced a panel of glycoproteins with a presentation of structurally well-defined glycotopes. Their application in a solid-phase assay reveals fine-specificity differences between closely related lectins, even from the same subgroup, and also among mammalian and plant lectins sharing specificity to β-galactosides. These results support the notion for nonredundant assignments of different lectins from the same family in vivo, convincingly backed by promoter analysis and immunohistochemical expression profiling of two prototype CGs. By implementing flexible ligand docking to this research area in a proof-of-principle case study, we began to discern detailed structure–activity relationships, a promising approach for drug design when applied to human lectins.

Acknowledgments This work was supported by grants from the Chang-Gung Medical Research Project (CMRP No. 33025 and 170441; Kwei-san, Tao-yuan, Taiwan), the National Science Council (NSC 97-2320-B-182-020-MY3 and 97-2628-B-182-002-MY3; Taipei, Taiwan), the Mizutani Foundation for Glycoscience (Tokyo, Japan), the research initiative LMU*excellent* (Munich, Germany), the Verein zur Förderung des biologisch-technologischen Fortschritts in der Medizin e. V. (Heidelberg, Germany), and a European Community Marie Curie Research Training Network grant (contract no. MRTN-CT-2005-019561).

References

1. Gabius HJ, Wu AM (2006) The emerging functionality of endogenous lectins: a primer to the concept and a case study on galectins including medical implications. Chang Gung Med J 29:37–62
2. Gabius HJ (ed) (2009) The sugar code. Fundamentals of glycosciences. Wiley-VCH, Weinheim
3. Laine RA (1997) The information-storing potential of the sugar code. In: Gabius HJ, Gabius S (eds) Glycosciences: status and perspectives. Chapman & Hall, London, pp 1–14
4. Brockhausen I, Schachter H (1997) Glycosyltransferases involved in N- and O-glycan biosynthesis. In: Gabius HJ, Gabius S (eds) Glycosciences: status and perspectives. Chapman & Hall, London, pp 79–113
5. Reuter G, Gabius HJ (1999) Eukaryotic glycosylation: whim of nature or multipurpose tool? Cell Mol Life Sci 55:368–422

6. Patsos G, Corfield A (2009) *O*-Glycosylation: structural diversity and functions. In: Gabius HJ (ed) The sugar code. Fundamentals of glycosciences. Wiley-VCH, Weinheim, pp 111–137
7. Wilson IBH, Paschinger K, Rendić D (2009) Glycosylation of model and 'lower' organisms. In: Gabius HJ (ed) The sugar code. Fundamentals of glycosciences. Wiley-VCH, Weinheim, pp 139–154
8. Zuber C, Roth J (2009) *N*-Glycosylation. In: Gabius HJ (ed) The sugar code. Fundamentals of glycosciences. Wiley-VCH, Weinheim, pp 87–110
9. Lemieux RU (1989) The origin of the specificity in the recognition of oligosaccharides by proteins. Chem Soc Rev 18:347–374
10. Lemieux RU (1996) How water provides the impetus for molecular recognition in aqueous solution. Acc Chem Res 29:373–380
11. Gabius HJ (1998) The how and why of protein–carbohydrate interaction: a primer to the theoretical concept and a guide to application in drug design. Pharm Res 15:23–30
12. Gabius HJ, Siebert HC, André S, Jiménez-Barbero J, Rüdiger H (2004) Chemical biology of the sugar code. Chembiochem 5:740–764
13. Solís D, Romero A, Menéndez M, Jiménez-Barbero J (2009) Protein–carbohydrate interactions: basic concepts and methods for analysis. In: Gabius HJ (ed) The sugar code. Fundamentals of glycosciences. Wiley-VCH, Weinheim, pp 233–245
14. Watkins WM, Morgan WTJ (1952) Neutralisation of the anti-H agglutinin in eel serum by simple sugars. Nature 169:825–826
15. Kilpatrick DC, Green C (1992) Lectins as blood typing reagents. Adv Lectin Res 5:51–94
16. Watkins WM (1999) A half century of blood-group antigen research: some personal recollections. Trends Glycosci Glycotechnol 11:391–411
17. Rüdiger H, Gabius HJ (2009) The history of lectinology. In: Gabius HJ (ed) The sugar code. Fundamentals of glycosciences. Wiley-VCH, Weinheim, pp 261–268
18. von der Lieth CW, Siebert HC, Kožár T, Burchert M, Frank M, Gilleron M, Kaltner H, Kayser G, Tajkhorshid E, Bovin NV, Vliegenthart JFG, Gabius HJ (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
19. López-Lucendo MF, Solís D, André S, Hirabayashi J, Kasai KI, Kaltner H, Gabius HJ, Romero A (2004) Growth-regulatory human galectin-1: crystallographic characterisation of structural changes induced by single-site mutations and their impact on the thermodynamics of ligand binding. J Mol Biol 343:957–970
20. Siebert HC, André S, Lu SY, Frank M, Kaltner H, van Kuik JA, Korchagina EY, Bovin NV, Tajkhorshid E, Kaptein R, Vliegenthart JFG, von der Lieth CW, Jiménez-Barbero J, Kopitz J, Gabius HJ (2003) Unique conformer selection of human growth-regulatory lectin galectin-1 for ganglioside GM1 versus bacterial toxins. Biochemistry 42:14762–14773
21. André S, Kaltner H, Lensch M, Russwurm R, Siebert HC, Fallsehr C, Tajkhorshid E, Heck AJR, von Knebel DM, Gabius HJ, Kopitz J (2005) Determination of structural and functional overlap/divergence of five proto-type galectins by analysis of the growth-regulatory interaction with ganglioside GM1 *in silico* and *in vitro* on human neuroblastoma cells. Int J Cancer 114:46–57
22. Kopitz J, von Reitzenstein C, Burchert M, Cantz M, Gabius HJ (1998) Galectin-1 is a major receptor for ganglioside GM1, a product of the growth-controlling activity of a cell surface ganglioside sialidase, on human neuroblastoma cells in culture. J Biol Chem 273:11205–11211
23. Kopitz J, von Reitzenstein C, André S, Kaltner H, Uhl J, Ehemann V, Cantz M, Gabius HJ (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
24. Wang J, Lu ZH, Gabius HJ, Rohowsky-Kochan C, Ledeen RW, Wu G (2009) Cross-linking of GM1 ganglioside by galectin-1 mediates regulatory T cell activity involving TRPC5 channel activation: possible role in suppressing experimental autoimmune encephalomyelitis. J Immunol 182:4036–4045

25. Siebert HC, Gilleron M, Kaltner H, von der Lieth CW, Kožár T, Bovin NV, Korchagina EY, Vliegenthart JFG, Gabius HJ (1996) NMR-based, molecular dynamics- and random walk molecular mechanics-supported study of conformational aspects of a carbohydrate ligand (Galβ1-2Galβ1-R) for an animal galectin in the free and in the bound state. Biochem Biophys Res Commun 219:205–212
26. Gilleron M, Siebert HC, Kaltner H, von der Lieth CW, Kožár T, Halkes KM, Korchagina EY, Bovin NV, Gabius HJ, Vliegenthart JFG (1998) Conformer selection and differential restriction of ligand mobility by a plant lectin. Conformational behavior of Galβ1-3GlcNAcβ1-R, Galβ1-3GalNAcβ1-R and Galβ1-2Galβ1-R' in the free state and complexed with mistletoe lectin as revealed by random walk and conformational clustering molecular mechanics calculations, molecular dynamics simulations and nuclear Overhauser experiments. Eur J Biochem 252:416–427
27. Kaltner H, Gabius HJ (2001) Animal lectins: from initial description to elaborated structural and functional classification. In: Wu AM (ed) The molecular immunology of complex carbohydrates. Adv Exp Med Biol, vol 491. Plenum Press, New York, pp 79–93
28. Cooper DNW (2002) Galectinomics: finding themes in complexity. Biochim Biophys Acta 1572:209–231
29. Rappl G, Abken H, Muche JM, Sterry W, Tilgen W, André S, Kaltner H, Ugurel S, Gabius HJ, Reinhold U (2002) $CD4^{+}CD7^{-}$ leukemic T cells from patients with Sézary syndrome are protected from galectin-1-triggered T cell death. Leukemia 16:840–845
30. André S, Sanchez-Ruderisch H, Nakagawa H, Buchholz M, Kopitz J, Forberich P, Kemmner W, Böck C, Deguchi K, Detjen KM, Wiedenmann B, von Knebel DM, Gress TM, Nishimura SI, Rosewicz S, Gabius HJ (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
31. Roda O, Ortiz-Zapater E, Martínez-Bosch N, Gutiérrez-Gallego R, Vila-Perelló M, Ampurdanés C, Gabius HJ, André S, Andreu D, Real FX, Navarro P (2009) Galectin-1 is a novel functional receptor for tissue plasminogen activator in pancreatic cancer. Gastroenterology 136:1379–1390
32. Ahmad N, Gabius HJ, Kaltner H, André S, Kuwabara I, Liu FT, Oscarson S, Norberg T, Brewer CF (2002) Thermodynamic binding studies of cell surface carbohydrate epitopes to galectins-1, -3, and -7: evidence for differential binding specificities. Can J Chem 80:1096–1104
33. Wu AM (2003) Carbohydrate structural units in glycoproteins and polysaccharides as important ligands for Gal and GalNAc reactive lectins. J Biomed Sci 10:676–688
34. Purkrábková T, Smetana K Jr, Dvo ánková B, Holíková Z, Böck C, Lensch M, André S, Pytlík R, Liu FT, Klíma J, Smetana K, Motlík J, Gabius HJ (2003) New aspects of galectin functionality in nuclei of cultured bone marrow stromal and epidermal cells: biotinylated galectins as tool to detect specific binding sites. Biol Cell 95:535–545
35. Kübler D, Hung CW, Dam TK, Kopitz J, André S, Kaltner H, Lohr M, Manning JC, He L, Wang H, Middelberg A, Brewer CF, Reed J, Lehmann WD, Gabius HJ (2008) Phosphorylated human galectin-3: facile large-scale preparation of active lectin and detection of structural changes by CD spectroscopy. Biochim Biophys Acta 1780:716–722
36. Wu AM, Singh T, Wu JH, Lensch M, André S, Gabius HJ (2006) Interaction profile of galectin-5 with free saccharides and mammalian glycoproteins: probing its fine specificity and the effect of naturally clustered ligand presentation. Glycobiology 16:524–537
37. Wu AM, Wu JH, Liu JH, Singh T, André S, Kaltner H, Gabius HJ (2004) Effects of polyvalency of glycotopes and natural modifications of human blood group ABH/Lewis sugars at the Galβ1-terminated core saccharides on the binding of domain-I of recombinant tandem-repeat-type galectin-4 from rat gastrointestinal tract (G4-N). Biochimie 86:317–326
38. Wu AM, Wu JH, Tsai MS, Liu JH, André S, Wasano K, Kaltner H, Gabius HJ (2002) Fine specificity of domain-I of recombinant tandem-repeat-type galectin-4 from rat gastrointestinal tract (G4-N). Biochem J 367:653–664

39. Wu JH, Singh T, Herp A, Wu AM (2006) Carbohydrate recognition factors of the lectin domains present in the *Ricinus communis* toxic protein (ricin). Biochimie 88:201–217
40. Kaltner H, Solís D, Kopitz J, Lensch M, Lohr M, Manning JC, Mürnseer M, Schnölzer M, André S, Sáiz JL, Gabius HJ (2008) Prototype chicken galectins revisited: characterization of a third protein with distinctive hydrodynamic behaviour and expression pattern in organs of adult animals. Biochem J 409:591–599
41. Varela PF, Solís D, Díaz-Mauriño T, Kaltner H, Gabius HJ, Romero A (1999) The 2.15 Å crystal structure of CG-16, the developmentally regulated homodimeric chicken galectin. J Mol Biol 294:537–549
42. López-Lucendo MF, Solís D, Sáiz JL, Kaltner H, Russwurm R, André S, Gabius HJ, Romero A (2009) Homodimeric chicken galectin CG-1B (C-14): crystal structure and detection of unique redox-dependent shape changes involving inter-and intrasubunit disulfide bridges by gel filtration, ultracentrifugation, site-directed mutagenesis and peptide mass fingerprinting. J Mol Biol 386:366–378
43. Sakakura Y, Hirabayashi J, Oda Y, Ohyama Y, Kasai KI (1990) Structure of chicken 16-kDa β-galactoside-binding lectin: complete amino acid sequence, cloning of cDNA and production of recombinant lectin. J Biol Chem 265:21573–21579
44. Wu AM, Wu JH, Tsai MS, Kaltner H, Gabius HJ (2001) Carbohydrate specificity of a galectin from chicken liver (CG-16). Biochem J 358:529–538
45. Wu AM, Singh T, Liu JH, Krzeminski M, Russwurm R, Siebert HC, Bonvin AMJJ, André S, Gabius HJ (2007) Activity–structure correlations in divergent lectin evolution: fine specificity of chicken galectin CG-14 and computational analysis of flexible ligand docking for CG-14 and the closely related CG-16. Glycobiology 17:165–184
46. Dominguez C, Boelens R, Bonvin AMJJ (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125:1731–1737
47. Villalobo A, Nogales-González A, Gabius HJ (2006) A guide to signaling pathways connecting protein–glycan interaction with the emerging versatile effector functionality of mammalian lectins. Trends Glycosci Glycotechnol 18:1–37
48. Gabius HJ (2008) Glycans: bioactive signals decoded by lectin. Biochem Soc Trans 36:1491–1496
49. Rüdiger H, Siebert HC, Solís D, Jiménez-Barbero J, Romero A, von der Lieth CW, Díaz-Mauriño T, Gabius HJ (2000) Medicinal chemistry based on the sugar code: fundamentals of lectinology and experimental strategies with lectins as targets. Curr Med Chem 7:389–416
50. Yamazaki N, Kojima S, Bovin NV, André S, Gabius S, Gabius HJ (2000) Endogenous lectins as targets for drug delivery. Adv Drug Deliv Rev 43:225–244