Glycobiology vol. 17 no. 2 pp. 165–184, 2007
doi:10.1093/glycob/cwl062
Advance Access publication on October 23, 2006

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

Albert M. Wu[1,2], Tanuja Singh[2], Jia-Hau Liu[2], Mickael Krzeminski[3], Roland Russwurm[4], Hans-Christian Siebert[4], Alexandre M.J.J. Bonvin[3], Sabine André[4], and Hans-Joachim Gabius[4]

[2]Glyco-Immunochemistry Research Laboratory, Institute of Molecular and Cellular Biology, College of Medicine, Chang-Gung University, Kwei-san, Tao-yuan 333, Taiwan; [3]Department of NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands; and [4]Institute of Physiological Chemistry, Faculty of Veterinary Medicine, Ludwig Maximilians University, Veterinärstr. 13, 80539 Munich, Germany

Received on June 20, 2006; revised on September 18, 2006; accepted on October 15, 2006

Gene duplication and sequence divergence are driving forces toward establishing protein families. To examine how sequence changes affect carbohydrate specificity, the two closely related proto-type chicken galectins CG-14 and CG-16 were selected as models. Binding properties were analyzed using a highly sensitive solid-phase assay. We tested 56 free saccharides and 34 well-defined glycoproteins. The two galectins share preference for the II (Galβ1-4GlcNAc) versus I (Galβ1-3GlcNAc) version of β-galactosides. A pronounced difference is found owing to the reactivity of CG-14 with histo-blood group ABH active oligosaccharides and A/B active glycoproteins. These experimental results prompted to determine activity–structure correlations by modeling. Computational analysis included consideration of the flexibility of binding partners and the presence of water molecules. It provided a comparative description of complete carbohydrate recognition domains, which had so far not been characterized in animal galectins. The structural models assigned II, I selectivity to a region downstream of the central Trp moiety. Docking revealed that the tetrasaccharides can be accommodated in their free-state low-energy conformations. CG-14's preference for A versus B epitopes could be attributed to a contact between His124 and the *N*-acetyl group of GalNAc. Regarding intergalectin comparison, the Ala53/Cys51 exchange affects the interaction potential of His54/His52. Close inspection of simulated dynamic interplay revealed reorientation of His124 at the site of the His124/Glu123 substitution, with potential impact on ligand dissociation. In summary, this study identifies activity differences and provides information on their relation to structural divergence, epitomizing the value of this combined approach beyond galectins.

Keywords: blood group/docking/evolution/glycoprotein/lectin/mutation/protein–carbohydrate interaction

Introduction

Gene duplication and ensuing sequence diversification by mutations are key factors in the process to generate proteins with closely related structural properties (Thomas 1993). The introduction of amino acid substitutions can entail emergence of new and distinctive properties in members of a protein family. Such products are attractive study objects in the quest to delineate activity–structure relations. Of special interest for thorough study are those cases in which homologous proteins are differentially expressed, because they are then apparently assigned to nonidentical biological functions. Under these circumstances, detailed comparative analysis is expected to teach salient lessons on the versatility of a common domain acquired by amino acid substitutions. Toward this end, a corresponding research program starts with the definition of sequence divergence in the selected protein pair, followed by monitoring the presence of proteins in tissues. The next step then is biochemical determination of binding properties to delineate reactivity profiles with physiological ligands. These efforts are especially well invested when dealing with important regulators of cellular activities.

Owing to the emerging significance of protein (lectin)–carbohydrate recognition for a wide range of biological activities such as adhesion, growth control, migration, or glycan routing (Gabius 1997, 2001, 2006; Villalobo et al. 2006), we herein present a model study along this given line. We focus on a family of animal lectins exerting these listed functions, i.e. galectins. The intrafamily diversification of these cellular effectors has been carefully mapped at the level of sequences and phylogenetic dendrograms (Cooper 2002; Houzelstein et al. 2004). This information guides to the prioritization among members of this lectin family. As consequence, the case of two proto-type chicken galectins (CGs), termed CG-14 and CG-16 owing to their molecular weights as determined by gel electrophoresis, has attracted attention for a pilot study. A close inspection of this protein pair beyond structural relationship reveals further favorable features which give high priority to the study of their binding characteristics.

Structurally, the sequence alignment yielded 71 substitutions per 134 amino acid residues, and ensuing calculations on mutational rates suggested that the origin of the genes for

[1]To whom correspondence should be addressed; e-mail: amwu@mail.cgu.edu.tw

these two lectins dates back to a duplication event at around the divergence of birds from mammals approximately 3×10^8 years ago (Sakakura et al. 1990). Evidently, neither the common β-sandwich fold nor the position of the Trp moiety in the central part of the carbohydrate recognition domain (CRD) is impaired by these sequence changes (Siebert et al. 1997; Varela et al. 1999). Thus, the basic mode of interaction of these two galectins with the β-galactoside core of a glycan is maintained. But it is an open question whether the fine specificities for complex glycans will also be constant. The importance of answering the question on the CGs' binding properties to glycoproteins (gps) is underscored by the conspicuous differences in their developmental regulation. These were observed in dermal morphogenesis and organ-specific gene expression, which give reason to assume nonredundant functionalities (Sakakura et al. 1990; Akimoto et al. 1993, 1995). Experimentally, the two galectins also differ in hemagglutination and axon growth/guidance activities, CG-16 being a strong cross-linking agent and capable of blocking proliferation of neuroblastoma (SK-N-MC) cells (Schneller et al. 1995; Kopitz et al. 2004; André, Kaltner et al. 2005).

In order to distinguish sugar-encoded signals on the cell surface, galectins present extended binding sites to complex glycans of natural glycoconjugates (Leffler and Barondes, 1986; Kopitz et al. 2001; Hirabayashi et al. 2002; Siebert et al. 2003; Sturm et al. 2004; Fischer et al. 2005). It is therefore timely to resolve the issue whether and to what extent the fine specificities of these two homologous lectins will have been subject to divergence. In addition, spatial factors of cellular glycans imposed by clustering or substitutions are known to modulate lectin affinity significantly (André, Cejas et al. 1999; Andre et al. 2001; Andre et al. 2003; Andre, Unverzagt et al. 2004; Unverzagt et al. 2002; Dam et al. 2005). Consequently, we extended our panel of glycosubstances beyond commonly used free saccharides. In order to deliberately simulate the natural situation where both sequence and topological features contribute to lectin reactivity, 33 well-characterized gps and a bacterial polysaccharide were also tested. The molecular details of these probes' glycosylation, especially at the branch ends relevant as docking sites for galectins, and the binding profile of CG-16 had been described in our previous report (Wu et al. 2001). Here, we first document the results for CG-14. They were obtained under identical conditions to complete the comparative monitoring and to ensure its validity. Following the presentation of these experimental results, activity–structure relations derived from extensive in silico work on the two CRDs are reported. The computational docking analysis with state-of-the-art software provided detailed insights into the nature of the binding sites beyond the core β-galactoside. Ligand positioning was systematically graded according to the level of interaction energy, and contributions of individual amino acids to the contact to histo-blood group ABH tri- and tetrasaccharides were inferred. Hereby, effects of mutational substitutions in the CRDs could be determined for the first time in this lectin family. Equally important, the probing into the lectin–oligosaccharide interaction included simulation of flexibility for both binding partners. In contrast to crystallographic analysis, the potential for mutual adaptation could thus be inferred. In summary, the given results will have relevance beyond this model pair, because no information on structural aspects of binding of these determinants by mammalian lectins and the inherent dynamics is so far available. In more general terms, they established proof-of-principle for the combined strategy also suitable for the design of custom-made lectin variants and inhibitors.

Results

CG-14–glycan interaction

The first aim was to study the interaction profile of CG-14 by a highly sensitive microtiter plate enzyme-linked lectinosorbent assay (ELLSA) with a panel of natural gps. These test substances present a spectrum of distinct epitopes at their spatially accessible branch ends, simulating the physiological situation. Of particular note, their application in combination with free saccharides allows us to determine the effect of spatial parameters of glycan presentation on a protein backbone. Biotinylated CG-14 was used as probe. Labeling in the presence of ligand resulted in a loss of binding activity of less than 25%, as determined by hemagglutination and solid-phase assays (data not shown). The results obtained with different gps and a bacterial polysaccharide are summarized in Table I. Evidently, CG-14 selects distinct epitopes (Table I). It reacted strongly with three histo-blood group $\mathbf{A}_h$ active gps, namely Cyst MSM 10% ppt, Cyst Mcdon and Cyst MSS 10% 2× ppt (1–3), and one histo-blood group precursor equivalent gp (11). Binding to these gps reached absorbance values (A_{405}) of >3.0 within 4 h and required less than 14 ng gp for 1.5 units of absorbance. CG-14 also bound well to other histo-blood group $\mathbf{A}_h$ and $\mathbf{B}_h$ active gps (such as cyst Beach phenol insoluble, cyst 9 and 14 phenol insoluble 4–6) and **I**/**II**-containing precursor gps (cyst Beach P-1, cyst OG 10% 2×, and cyst MSS 1st Smith degraded 12–14). Not all β-galactosides were suitable binding partners. CG-14 reacted weakly or not at all with blood group **H** or $\mathbf{Le}^b$ active substances (cyst N-1 Le^a 20% 2×, cyst Tighe, and JS phenol insoluble 8–10), multi-**II** antennary gps (sialylated and asialo rat sublingual gp (RSL), human α_1-acid gp with/without sialylation and fetuin 15, 16, 20, 26–28, and also the *Pneumococcus* type 14 polysaccharide 17), **T**/**Tn**-containing salivary gps (sialylated and ovine, bovine and porcine salivary asialo gps 21, 29–32, active and inactive antifreeze gps 25 and 33, human glycophorin and its asialo derivative 22 and 23), and sheep hydatid cyst gp (34). Evidently, the topological way an epitope is presented matters significantly for ligand properties. As parameter for comparison, the amount of gp used for coating is listed in Table I. Because the actual amount of glycans adsorbed on the surface of the microtiter plate wells had not been determined in this assay setting, their relative binding properties were further examined in an inhibition assay. To mediate carbohydrate-dependent CG-14 binding to the plate surface, a gp presenting **I**, **II** determinants was selected for adsorption, as done previously in the study of CG-16 properties (Wu et al. 2001).

Inhibition of CG-14–glycan interaction by various polyvalent gps

The ability of various gps to inhibit the binding of CG-14 to a Galβ1–3/4GlcNAc containing gp (Beach P-1) purified from human ovarian cyst fluid was systematically analyzed by ELLSA. Five free saccharides were included for internal

Table I. Binding of CG-14 to human blood group A, B, H, P_1 Le^a, and Le^b active, sialo, and asialo gps by ELLSA*

Identification number	Glycoprotein (branch-end epitope)†	1.5 (A_{405}) unit (ng)	Maximum A_{405} absorbance	
			Absorbance reading‡	Binding intensity‡
ABH and Lewis blood group active glycoproteins				
1	Cyst MSM 10% ppt ($\mathbf{A_h}$ [$\mathbf{A_2}$])	1.0	3.0	++++ +
2	Cyst Mcdon ($\mathbf{A_h}$)	5.0	3.0	++++ +
3	Cyst MSS 10% 2× ($\mathbf{A_h}$)	6.0	3.0	++++ +
4	Cyst Beach phenol insoluble ($\mathbf{B_h}$)	26.0	2.6	++++ +
5	Cyst 14 phenol insoluble ($\mathbf{A_h}$)	28.0	2.6	++++ +
6	Cyst 9 phenol insoluble ($\mathbf{A_h}$)	90.0	2.5	++++ +
7	Cyst Tij 20% of 2nd 10% ($\mathbf{B_h}$ > **I/II**)	100.0	3.3	++++ +
8	Cyst N-1 Le^a 20% 2× (**Le^a**)	−	1.3	++
9	Cyst Tighe phenol insoluble (**H**, **Le^b**)	−	0.02	−
10	Cyst JS phenol insoluble (**H**)	−	0.0	−
Blood group precursor (equivalent) glycoproteins				
11	Cyst Mcdon P-1 (**I**, **II**, **T**, **Tn**)	14.0	3.8	++++ +
12	Cyst Beach P-1 (**I**, **II, T**, **Tn**)	50.0	3.7	++++ +
13	Cyst OG 10% 2× ppt (**I/II**)	60.0	2.6	++++ +
14	Cyst MSS first Smith degraded (**I**, **II**, **T**, **Tn**)	120.0	2.3	++++
N-linked and T/Tn-containing glycoproteins				
15	Asialo RSL (**II**)	−	0.8	+
16	Human asialo α_1-acid gp (m**II**)	−	0.6	+
17	*Pneumococcus* type 14 ps (**II**)	−	0.5	+
18	Asialo THGP Sd (a^+) W.M. (**II**, **S**)	−	0.4	±
19	THGP Sd (a^+) W.M. (sialyl **II**, **S**)	−	0.2	±
20	Asialofetuin (**II**, **T**)	−	0.1	−
21	Asialo BSM (**Tn**, GlcNAcβ1-3**Tn**)	−	0.1	−
22	Human glycophorin (sialyl $\mathbf{T_\alpha}$)	−	0.1	−
23	Human asialoglycophorin ($\mathbf{T_\alpha}$)	−	0.1	−
24	**Tn**-glycophorin (**Tn**)	−	0.08	−
25	Inactive antifreeze gp ($\mathbf{T_\alpha}$; M.W. $2.6–3.8 \times 10^3$)	−	0.04	−
26	Fetuin (sialyl **II, T**)	−	0.03	−
27	RSL (sialyl **II**)	−	0.02	−
28	Human α_1-acid gp (sialyl m**II**)	−	0.01	−
29	OSM (sialyl **Tn**)	−	0.0	−
30	Asialo OSM (**Tn**)	−	0.0	−
31	BSM (sialyl **Tn**, sialyl GlcNAcβ1-3**Tn**)	−	0.0	−
32	PSM (sialyl **Tn**, $\mathbf{T_\alpha}$, $\mathbf{A_h}$, **H**)	−	0.0	−
33	Active antifreeze gp ($\mathbf{T_\alpha}$; M.W. $1.0–2.1 \times 10^4$)	−	0.0	−
34	Sheep hydatid cyst gp ($\mathbf{E/P_1}$)	−	0.0	−

BSM, bovine submandibular mucin; OSM, ovine salivary mucin; PSM, porcine salivary mucin; THGP, Tamm-Horsfall gp.
*250 ng of biotinylated lectin was reacted with various glycoproteins, ranging from 0.01 ng to 1 μg quantity.
†The human blood group activity and/or lectin determinants are given in bold in parentheses: **A** (GalNAcα1-3Gal); $\mathbf{A_h}$ (GalNAcα1-3[LFucα1-2]Gal); **H** (LFucα1-2Gal); **B** (Galα1-3Gal); $\mathbf{B_h}$ (Galα1-3[LFucα1-2]Gal); **T** (Galβ1-3GalNAc); **Tn** (GalNAcα1-Ser/Thr); **I/II** (Galβ1-3/4GlcNAc); **L** (Galβ1-4Glc).
‡The results were graded according to the spectrophotometric absorbance value at 405 nm (i.e. $O.D._{405}$) after 4-h incubation as follows: +++++ (O.D. > 2.5), ++++ (2.5 > O.D. ≥ 2.0),+++ (2.0 > O.D. ≥ 1.5), ++ (1.5 > O.D. ≥ 1.0), + (1.0 > O.D. ≥ 0.5), ± (0.5 > O.D. ≥ 0.2), and − (O.D. < 0.2).

calibration of the inhibitory potency of the gps. The results are compiled in Table II, and the inhibition curves are shown in Figure 1. Among the natural glycocompounds tested for the inhibition of this interaction, a gp presenting $\mathbf{A_h}$ determinants (1, cyst MSM 10% ppt) was the best. The CG-14–glycan interaction was also strongly inhibited by other gps with blood group $\mathbf{A_h}$ or $\mathbf{B_h}$ activity and their precursors (2–10). Their reactivities were 1.9×10^3, 3.5×10^3, and

Table II. Amount of various gps and free saccharides giving 50% inhibition of binding of CG-14 (125 ng/50 μL) to Beach P-1 (50 ng/50 μL)[a]

Order of activity	Curves in Figure 1	Inhibitor (branch-end epitope)[b]	Quantity giving 50% inhibition (ng)	Relative potency[c]
1	A	Cyst MSM 10% ppt ($\mathbf{A_h}$ [$\mathbf{A_2}$])	0.4	9.0×10^4
2	A	Cyst MSS 10% 2× ($\mathbf{A_h}$)	2.5	1.4×10^4
3	A	Cyst Mcdon ($\mathbf{A_h}$)	2.8	1.3×10^4
4	A	Cyst 14 phenol insoluble ($\mathbf{A_h}$)	10.0	3.6×10^3
5	A	Cyst Mcdon P-1 (**I**/**II**, **T**, **Tn**)	10.0	3.6×10^3
6	A	Cyst 9 phenol insoluble ($\mathbf{A_h}$)	10.0	3.6×10^3
7	B	Cyst Beach P-1 (**I**/**II**, **T**, **Tn**)	20.0	1.8×10^3
8	B	Cyst Tij 20% of second 10% ($\mathbf{B_h}$ > **I**/**II**)	27.0	1.3×10^3
9	B	Cyst Beach phenol insoluble ($\mathbf{B_h}$)	40.0	9.0×10^2
10	B	Cyst OG 10% 2× ppt (**I**/**II**)	65.0	5.5×10^2
11	B	Asialo BSM (**Tn**, GlcNAcβ1-3**Tn**)	70.0	5.1×10^2
12	A	Cyst MSS first Smith degraded (**I**, **II**, **T**, **Tn**)	1.7×10^2	2.1×10^2
13	A	Cyst Tighe P-1 (**I**/**II**)	1.8×10^2	2.0×10^2
14	B	BSM (sialyl **Tn**, sialyl GlcNAcβ1-3**Tn**)	2.0×10^2	1.8×10^2
15	B	*Pneumococcus* type 14 ps (**II**)	5.0×10^2	72.0
16	A, B	Galβ1-4GlcNAc (**II**)	7.7×10^2	46.7
17	A, B	Galβ1-4Glc (**L**)	1.4×10^3	25.7
18	A, B	Gaβ1 − 4GlcNAcβ1-2Man |α1-6 Galβ1-4GlcNAcβ1-2ManαI-3Manβ1-4GlcNAcβ1-4GlcNAcβ2-*N*-Asn |β1-4 Galβ1-4GlcNAc Tri-antennary Galβ1-4GlcNAc (Tri-II)	1.4×10^3	25.7
19	B	Asialo THGP Sd (a^+) W.M. (**II**, **S**)	2.8×10^3	12.9
20	A, B	GalNAc	2.9×10^4	1.2
21	A, B	Gal	3.6×10^4	1.0
22		PSM (sialyl **Tn**, $\mathbf{T_\alpha}$, $\mathbf{A_h}$, **H**)	$>2.8 \times 10^2$ (31.5%)[d]	—
23		Cyst N-1 Le^a 20% 2× ($\mathbf{Le^a}$)	$>2.8 \times 10^2$ (29.5%)	—
24		Asialo HSM (**Tn**)	$>1.4 \times 10^2$ (24.3%)	—
25		Cyst JS phenol insoluble (**H**)	$>2.8 \times 10^2$ (23.5%)	—
26		Asialo OSM (**Tn**)	$>2.8 \times 10^2$ (22.8%)	—
27		Active antifreeze gp ($\mathbf{T_\alpha}$; M.W. $1.0–2.1 \times 10^4$)	$>2.8 \times 10^2$ (19.9%)	—
28		HSM (sialyl **Tn**)	$>2.8 \times 10^2$ (19.8%)	—
29		RSL (sialyl **II**)	$>2.8 \times 10^2$ (18.4%)	—
30		Human asialo α_1-acid gp (m**II**)	$>2.8 \times 10^2$ (17.3%)	—
31		Human asialoglycophorin ($\mathbf{T_\alpha}$)	$>2.8 \times 10^2$ (16.4%)	—
32		OSM (sialyl **Tn**)	$>2.8 \times 10^2$ (13.4%)	—
33		Human glycophorin (sialyl $\mathbf{T_\alpha}$)	$>2.8 \times 10^2$ (12.8%)	—
34		Asialofetuin (**II**, **T**)	$>2.8 \times 10^2$ (12.4%)	—
35		Cyst Tighe phenol insoluble (**H**, $\mathbf{Le^b}$)	$>2.8 \times 10^2$ (11.1%)	—
36		**Tn**-glycophorin (**Tn**)	$>2.6 \times 10^2$ (7.7%)	—
37		Human α_1-acid gp (sialyl m**II**)	$>2.8 \times 10^2$ (7.1%)	—
38		Fetuin (sialyl **II**, **T**)	$>2.8 \times 10^2$ (6.6%)	—
39		THGP Sd (a^+) W.M. (sialyl **II**, **S**)	$>2.8 \times 10^2$ (1.5%)	—

HSM, hamster submacillary mucin.
[a]The inhibitory activity was estimated from the respective inhibition curve in Figure 1 and is expressed as the amount of inhibitor giving 50% inhibition of the control level of lectin binding; total assay volume was 50 μL.
[b]Listed in footnote of Table I.
[c]Relative potency = quantity of Gal (curve 21) required for 50% inhibition (taken as 1.0)/quantity of sample required for 50% inhibition.
[d]The inhibitory potency of inactive glycoproteins is expressed as the maximum amount of glycans tested that yield inhibition (in parentheses) below 50%.

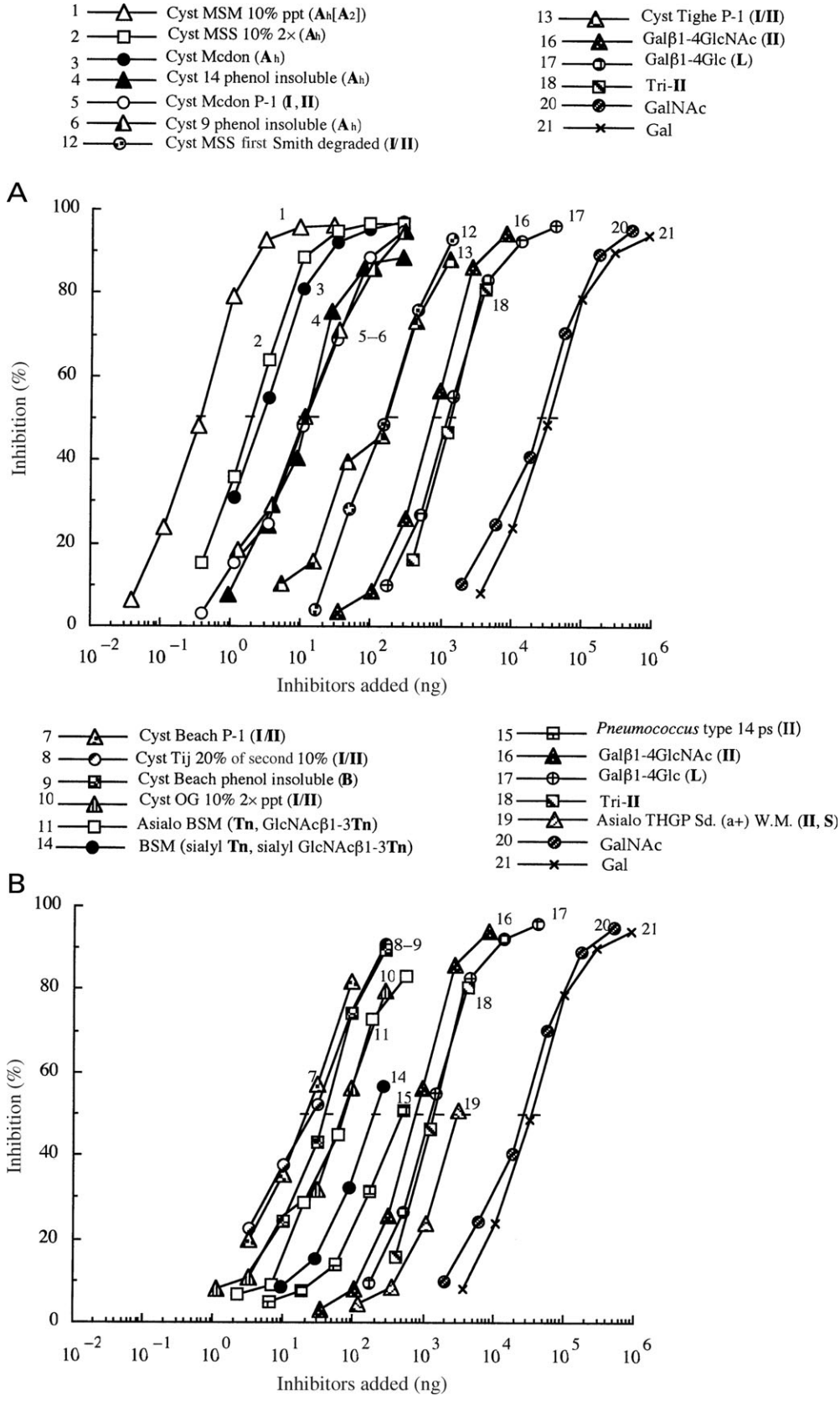

Fig. 1. Inhibition of CG-14 binding to a Galβ1-3/4GlcNAc-containing gp (Cyst Beach P-1) coated on the surface of ELLSA plate wells by a panel of gps presenting distinct branch-end epitopes. The quantity of gp in the coating solution was 50 ng per well. The quantity of lectin used for the inhibition assay was 125 ng per well. Total volume was 50 μL. A_{405} was recorded after a 4-h incubation period with substrate. The amount (ng) of gp required to reach 50% inhibition was determined.

9.0×10^4 times higher than that of monomeric Galβ1-4GlcNAc (**II**), Galβ1-4Glc (**L**), and Gal (16, 17, and 21), respectively. The inhibitory potencies of the gps, in general, are in accord with the maximum absorbances recorded in the binding assay (Tables I and II). However, a problem arises when carefully examining the quantitative data and applying the stoichiometric criterion that at least equal molarities of sugar unit and lectin should be reached to result in substantial inhibition. Evidently, the most potent gps reach a higher level of inhibitory capacity than expected on the basis of this criterion. Having herewith determined the reactivity of CG-14 to a series of gps in two assay settings, we next proceeded to define the binding specificity with free sugar compounds. These assays included epitopes known from glycolipids and galactose with aglyconic substituents.

Inhibition of CG-14–glycan interaction by mono-, di-, and oligosaccharides

These measurements were performed under conditions identical to the assays with the panel of the gps. The complete results are presented in Table III and Figure 2 for substances reaching a level of 50% inhibition and in Table IV for substances which are less active. To enable direct comparison between the inhibitory potencies of gps and free saccharides, we had added five mono- and disaccharides at positions 16, 17, 18, 20, and 21 to Table II. Among the saccharides tested, Galβ1-4GlcNAcβ1-3Galβ1-4Glc (**II**β1–3**L**), [Fucα1-2]Galβ1-3GlcNAcβ1-3Galβ1-4Glc (**H** active **I**β1–3**L**), and the tri-antennary *N*-linked glycopeptides, with Gal(β1-4)GlcNAc at the nonreducing end (Tri-**II**), were the best, being 5.0, 5.0, 2.9 and 5.0×10^2, 5.0×10^2, 2.9×10^2 times more active than Galβ1-4GlcNAc (**II**) and Gal, respectively (1–3 compared with 9 and 7, Table III, Figure 2A). Apparently, the 2,4,2-branched tri-antennary *N*-glycan from asialofetuin is slightly better than a bi-antennary hexasaccharide, and the presence of three *N*-glycans in asialofetuin conveys ligand properties to the gp, which are inferior to those of *O*-glycan-bearing gps from human ovarian cyst fluid. **H** active **I**β1–3**L** pentasaccharide (2) was 3.8-fold more potent than **I**β1–3**L** (5), whereas **H** active **L** (6) was 2.6 times more potent than **L** (13). These results indicate that introduction of LFucα1-2 to parent **I**β1–3**L** or **L** significantly enhances the binding affinity. In line with the data on gp, binding the completion of **A-** and **B**-type headgroups yielded 3.7- or 3.8-fold better inhibitors compared with the fucose-free disaccharides (10 and 27 versus 21 and 34). In direct comparison, the **A**-type trisaccharide (10) clearly exceeded the **B**-type glycan (27). Bi-antennary hexasaccharides **II**β1–3(**II**β1–6)**L** and **I**β1–3(**II**β1–6)**L** were 2.2 and 6.5 times better than Galβ1-3/4GlcNAc (**I**/**II**) (4 versus 9 and 7 versus 24, Table III), indicating that the affinity of CG-14 for binding partners is significantly enhanced by the length of carbohydrate sequence and the number of antennae in free Galβ1-containing ligands. Introduction of an **I** epitope into bivalent sugar lowered the affinity, and a β1-6 extension of **II** did not enhance it (7 and 8, Table III). These two compounds were about 2.9 times less active than tri-antennary Galβ1-4GlcNAc (3, Table III), but still 6.5 times more active than Galβ1-3GlcNAc (**I**). Gal(β1-6)GlcNAc did not reach 50% inhibition, revealing the importance of the β1–4 linkage at the core galactose unit. These results show that CG-14 has a pronounced preference at this site in the order: Galβ1-4 > Galβ1-3 > Galβ1-6.

Testing further disaccharides revealed a clear preference pattern: GalNAcβ1-3Gal (**P**), GalNAcα1-3Gal (**A**), GalNAcα1-3GalNAc (**F**), and Galβ1-3GalNAc (**T**) were from 28.5, 18.2 to 16.7 times more active than Gal (14, 21–23 versus 37), but 30 times less active than **II**β1–3**L** (Type 2; no. 1) (Table III). These results indicate that the *N*-acetyl amino group of the GalNAc residue at the penultimate position (**T**) might interfere with optimal ligand accommodation, that a histo-blood-group-like α1-3 extension and its β-anomer are tolerated, and that the β1-4 linkage is better than the β1-3 linkage. GalNAc itself, irrespective of its anomer, is a very weak inhibitor. Glycotopes $\mathbf{P}_\alpha$ (GalNAcβ1-3Galα1-Me) and $\mathbf{B}_\alpha$ (Galα1-3Galα1-Me) were 1.4–15.4 times better than **P** and **B** without fixed anomeric position (12 versus 14 and 20 versus 34), respectively, illustrating that anomeric arrest in $\mathbf{P}_\alpha$ and $\mathbf{B}_\alpha$ epitopes can play a role in binding (Table III). Further testing of building blocks of natural glycans helped to dissect the contributions of distinct structures to the binding of CG-14 to a gp carrying several determinants. In line with the lack of activity of mucin with **Tn** and **T**α (Table III), these two epitopes as free saccharides are weakly inhibitory (Table IV). Activity of the histo-blood precursor gps in Table II is most likely attributable to the **I**, **II** structures. Having herewith completed the analysis of a series of 90 glycosubstances as inhibitors of CG-14 binding, it is possible to present a detailed comparison between the properties of CG-14/CG-16, taking the previously published data on CG-16 into account (Wu et al. 2001).

Comparison of binding properties

The fact that we deliberately kept all experimental parameters constant in the measurements on the two CGs ensured that the intended comparison of the data sets is valid. The results presented in Tables I–IV and Figures 1 and 2 on CG-14 as well as in our previous report on CG-16 (Wu et al. 2001) can thus be directly compared. To help the reader identify differences, the two data sets were scrutinized stepwise to spot any disparities. Table V documents the individual aspects of the carbohydrate specificities, arranged into eight categories. Obviously, the two homologous lectins share pronounced preference to **II**. The relative levels of discrimination between **II** and **I** were 20-fold for CG-16 and 6.5-fold for CG-14. Polyvalency is discriminatively sensed by CG-14. When further comparing the reactivity profiles of the tested glycotopes, the assays revealed differences in gp binding regarding substitutions which led to histo-blood group **AB** epitopes (Table V). They are reflected on the level of assays with free oligosaccharides. Whereas the **A-** and **B-**type trisaccharides are potent inhibitors in the case of CG-14, the interaction of CG-16 with first Smith degraded cyst MSS gp could hardly be reduced by their presence, only **H** active **L** (Fucα1-2Galβ1-4Glc) serving as inhibitor with about 10-fold diminished activity relative to lactose (data not shown). Evidently, the presence of the **A-** and **B**-type extensions in gps makes marked difference. These observations prompted a close inspection of the CRDs for activity–structure relations. The current lack of information on how animal galectins accommodate these types of oligosaccharides adds a further incentive to address this issue. As illustrated in Figure 3, the two

Table III. Amount of various saccharides giving 50% inhibition of binding of CG-14 (125 ng/50 μL) to Beach P-1 (50 ng/50 μL)[a]

Curves or identification number	Curves in Figure 2	Inhibitor	Quantity giving 50% inhibition (nmd)	Relative potency[b]
1	A	Galβ1-4GlcNAcβ1-3Galβ1-4Glc (Type 2, **II**β1-3**L**)	0.4	5.0×10^2
2	A	[Fucα1-2]Galβ1-3GlcNAcβ1-3Galβ1-4Glc (**H** active **I**β1-3**L**)	0.4	5.0×10^2
3	A	Tri-antennary Galβ1-4GlcNAc (Tri-**II**)	0.7	2.9×10^2
4	A	Galβ1-4GlcNAc |β1-6 Galβ1-4Glc |β1-3 Galβ1-4GlcNAc [LNH; Di-II or IIβ1-3(IIβ1-6)L]	0.9	2.2×10^2
5	A	Galβ1-3GlcNAcβ1-3Galβ1-4Glc (**I**β1-3**L**)	1.5	1.3×10^2
6	A	Fucα1-2Galβ1-4Glc (**H** active **L**)	1.5	1.3×10^2
7	A	Galβ1-4GlcNAc |β1-6 Galβ1-4Glc |β1-3 Galβ1-3GlcNAc [LNnH; Iβ1-3(IIβ1-6)L] Galβ1-3GlcNAc	2.0	1.0×10^2
8	A	Galβ1-4GlcNAcβ1-6Gal (**I** Ma)	2.0	1.0×10^2
9	A	Galβ1-4GlcNAc (**II**)	2.0	1.0×10^2
10	B	GalNAcα1-3[Fucα1-2]Gal ($\mathbf{A}_h$)	3.0	66.7
11	A	Galβ1-4Glcβ1-*O*-Me ($\mathbf{L}_\beta$)	3.5	57.1
12	B	GalNAcβ1-3Galα1-Me ($\mathbf{P}_\alpha$)	3.5	57.1
13	A	Galβ1-4Glc (**L**)	4.0	50.0
14	B	GalNAcβ1-3Gal (**P**)	5.0	40.0
15	B	Galβ1-3GalNAcα1-Me ($\mathbf{T}_\alpha$1-methyl)	7.0	28.5
16	B	Galβ1-3GalNAcβ1-Me ($\mathbf{T}_\beta$1-methyl)	8.0	25.0
17	C	Galβ1-3DAra	9.0	22.2
18	B	Galβ1-3GalNAcα1-benzyl ($\mathbf{T}_\alpha$1-benzyl)	9.0	22.2
19	B	Galβ1-3GalNAcβ1-4Galβ1-4Glc (**T**β1-4**L**)	9.5	21.1
20	B	Galα1-3Galα1-*O*-Me ($\mathbf{B}_\alpha$)	10.0	20.0
21	B	GalNAcα1-3Gal (**A**)	11.0	18.2
22	B	GalNAcα1-3GalNAc (**F**)	11.0	18.2
23	B	Galβ1-3GalNAc (**T**)	12.0	16.7
24	A	Galβ1-3GlcNAc (**I**)	13.0	15.4
25	C	*p*-NO_2phenyl αGal	19.0	10.5
26	C	*p*-NO_2phenyl αGalNAc	33.0	6.1
27	B	Galα1-3[Fucα1-2]Gal ($\mathbf{B}_h$)	40.0	5.0
28	C	Galα1-6Galα1-6Glcβ1-2Fruf (Stachyose)	60.0	3.3
29	C	Methyl βGal	70.0	2.9
30	B	**Tn**-containing glycopeptides (M.W.$< 3.0 \times 10^3$)	1.2×10^2	1.7
31	C	GalNAc	1.3×10^2	1.5
32	C	*p*-NO_2phenyl βGal	1.4×10^2	1.4
33	C	Methyl αGal	1.5×10^2	1.3
34	B	Galα1-3Gal (**B**)	1.5×10^2	1.3
35	C	Galα1-6Glc (Melibiose)	1.8×10^2	1.1
36	B	Fucα1-2Gal (**H**)	2.0×10^{2c}	1.0
37	A, B, C	Gal	2.0×10^2	1.0
38	C	Galα1-6Glcβ1-2Fruf (Raffinose)	2.8×10^2	0.7
39	B	Galα1-4Gal (**E**)	6.0×10^{2c}	0.3

Continued

Table III. *Continued*

Curves or identification number	Curves in Figure 2	Inhibitor	Quantity giving 50% inhibition (nmd)	Relative potency[b]
40	C	DAra	2.5×10^3	0.08
41	C	Glc	3.0×10^3	0.07
42	C	Methyl αGlc	3.0×10^3	0.07
43	C	Methyl βGlc	3.7×10^3	0.05
44	C	DFuc	4.0×10^3	0.05

LNH, lacto-*N*-hexaose; LNnH, lacto-*N*-neohexaose.
[a]The inhibitory activity was estimated from the respective inhibition curve in Figure 2 and is expressed as the amount of inhibitor giving 50% inhibition of the control lectin binding; total assay volume was 50 μL.
[b]Relative potency of sugars compared with Gal was taken as 1.0.
[c]Extrapolated value.

CGs are structurally closely related. Their CRDs are placed at equivalent positions in the jelly-roll-like folding pattern. Whereas the position of lactose hosted in each CRD can readily be depicted, as shown in the middle panel of Figure 4, it is yet an open question how the two homologous CGs interact with ABH type epitopes. To answer it, we performed a computational modeling and docking study. Of note, the binding partners maintained full flexibility in the interaction study, a new feature for docking studies with lectins.

Computational modeling and docking

The generation of the model for CG-14—together with crystallographic information on CG-16—was the first step in our modeling study. The ensuing comparison of the set of key amino acids in the regions of contact for the disaccharide of CG-14 and CG-16 revealed identity (Figure 4, upper panel). Sequence divergence concerned the immediate vicinity of these moieties. Notably, the orientation of the Trp residues will probably not be kept constant, at least in part due to the Thr67/Glu69 substitution following the common Gly to their side. As evidence presented in the last paragraph intimates, the sequence stretch downstream of the Trp moiety might be relevant for **II, I** selectivity. The models then served as platform to visualize contact sites for **ABH** epitopes. The sequence additions to turn **I, II** structures into **ABH** epitopes will most likely occupy the space, where two amino acid changes are apparent (Figure 4, bottom panel). In detail, Cys51/Ala53 and Glu123/His124 substitutions, among others, separate CG-16 from CG-14. In order to determine whether these differences may have a bearing on ligand binding, we proceeded from rigid-body docking to three stages of increasing level of model-building refinement. First, flexibility was allowed for the amino acid side chains and then for both protein backbone and the side chains, the ligand being treated as flexible throughout the entire procedure. Finally, calculations on complexes selected due to most favorable binding energy were run with explicit consideration of water molecules.

The first lesson taught by the in silico work was that the ligand's free-state low-energy conformation was not distorted by contact with the CRDs. In other words, structures corresponding to populated areas of Φ, Ψ, E-maps for the **ABH** epitopes (Figure 5) will fit into the respective section of the CRD of both CGs. They will not be subject to a major alteration from their free-state parameters. Such a conformation could thus be drawn from Figure 4 (bottom panel). The positions of three energetically preferred constellations of the ligand associated with the CGs in the simulations are marked in the Φ, Ψ, E-maps in Figure 5. The energy grading in these maps also epitomizes that only a portion of the conformational space is accessible for these oligosaccharides in the free state. This conformational preference turns into a valuable asset for affinity generation when conformers from this region are selected for binding. The actual numbers of average dihedral angles of the glycosidic bonds are listed in Table VI. In essence, they signify that the free-state low-energy topology is suited for complex formation. To answer the question whether it is common for these histo-blood group determinants to retain the free-state conformation, we scoured records on structures of their complexes with proteins. As a result, crystallographic information on three cases involving the α1-2/3 substitutions on the galactose moiety is given in Table VI for comparison. These data sets support the concept that the free-state low-energy conformation of these oligosaccharides behaves like a "key-like" structure selected for interaction. Nonetheless, to preclude missing a physiologically relevant bound-state conformation, full ligand flexibility was enforced without resulting in a viable alternative. Both CGs therefore select conformers from energetically preferred regions of the Φ, Ψ, E-maps for binding. Their low positional root mean square deviation (RMSD) values from the average structure (between 0.57 and 0.84 Å) signify that these ensembles are structurally well defined. Inspecting the contact sites closely, we find that there is no clash between an amino acid side chain and the ligand, which might preclude its access to CG-16's CRD. Having herewith identified new topological aspects of ligand accommodation for animal galectins shown in Figure 4 (bottom panel) and most probably excluded steric hindrance as explanation for the difference, we proceeded to characterize the energetic contributions of each amino acid/monosaccharide to the overall interaction energy, distinguishing between Coulomb and van der Waals energies.

The overall distribution of energy contributions for the interaction between the tri- and tetrasaccharides and CRDs of

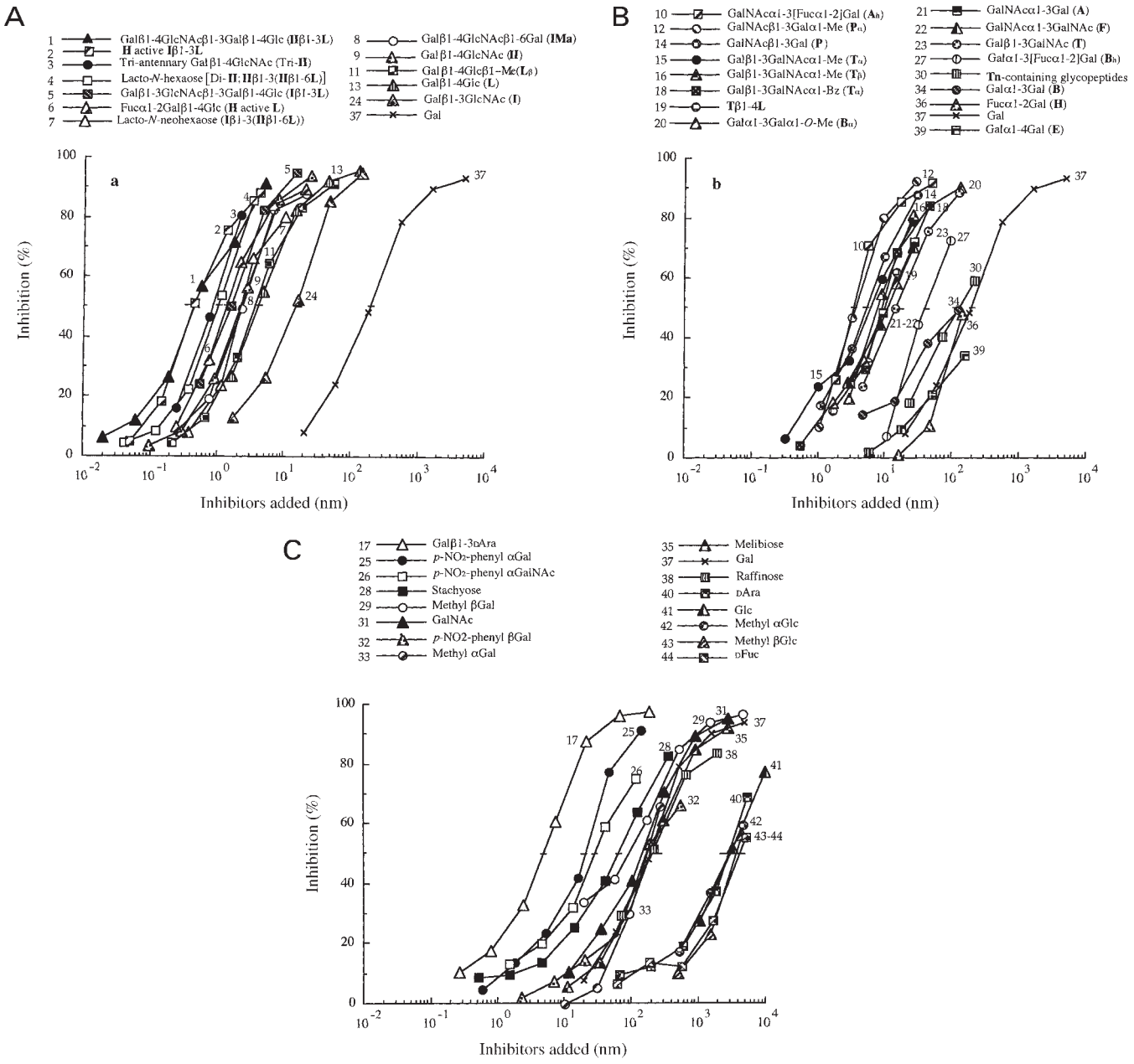

Fig. 2. Inhibition of CG-14 binding to a Galβ1-3/4GlcNAc-containing gp (Cyst Beach P-1) coated on the surface of ELLSA plate wells by various free saccharides. The quantity of gp in the coating solution was 50 ng per well. The lectin (250 ng per well) was preincubated with an equal volume of serially diluted inhibitor. The final CG-14 content was 125 ng per well. Total volume was 50 μL. A_{405} was recorded after a 4-h incubation period with substrate. The amount (nmol) of free saccharides required to reach 50% inhibition was determined.

CG-14 and CG-16 reflected a high degree of structural homology (Tables VII and VIII). These data characterize and grade in energetic terms the region responsible for hosting α1-2/3 substitutions in the two animal galectins. As a consequence, it is possible to offer a reasoning to rationalize the measured preference of CG-14 for histo-blood group **A** versus **B** epitopes, as seen in Tables I and II. Close inspection of patterns of interaction in silico revealed contact between the oxygen atom of the *N*-acetyl group of the α1-3-linked GalNAc moiety and the imidazole ring of His124 in CG-14, also shown in Figure 4 (bottom panel). Its loss in the histo-blood group **B**-tetrasaccharide might account for the reduced binding/inhibition capacity of gps with histo-blood group **B** in contrast to **A** epitopes. Regarding the differential reactivity of **AB** epitopes for CG-14 versus CG-16, as given in category 6 of Table V, the presented energetic considerations provide no immediate clue. What could prove helpful besides the estimation of energy terms, though, is thorough monitoring of the dynamic way the ligand–receptor pair is formed. The respective simulations indicated that rigid-body docking will inevitably miss rearrangements of the local environment as a response to the establishing contact. Indeed, the analysis of the routes to mutual structural adaptation offered a hint to pinpoint a potentially relevant difference between the CGs.

Table IV. Reduction of binding of CG-14 (125 ng/50 μL) to Beach P-1 (50 ng/50 μL) by saccharides acting as weak inhibitors

Inhibitor	Maximum amount of inhibitor used	Percentage of inhibition
Galβ1-6GlcNAc	1.6×10^2	46.3
Galβ1-3GalNAcα1-Ser (**T**$_\alpha$1-Ser)	23.6	46.1
Man	6.1×10^3	33.8
GalNAcβ1-4Gal (**S**)	28.9	33.5
Methyl βGalNAc	47.2	26.8
Methyl αGalNAc	47.2	24.2
GalNAcα1-Ser/Thr (**Tn**)	36.1	19.6
LFuc	5.2×10^3	2.1
GlcNAc	3.2×10^3	2.1
p-NO_2phenyl βGalNAc	1.1×10^2	14.4
LAra	3.6×10^3	−14.8
Chitin	1.2×10^2	−23.7

To allow the reader to follow the reasoning, files visualizing details of the simulated dynamics of the molecular interplay are instructive and are made accessible as Supplementary data.

The two substitutions in the first paragraph of this section were apparently not neutral in topological aspects. The Ala53/Cys51 exchange affected the orientation of the equivalent His54/His52 residues (Figure 4) and thereby the Coulomb energy term. This effect is clearly seen in Tables VII and VIII. These two tables list major energy contributions by individual amino acids. By itself, the noted substitution has no such direct influence on the energy balance sheet. Intriguingly, the second substitution, i.e. His124/Glu123, yet appears to be responsible for a notable difference in CRD–ligand interplay with potentially direct relevance to the experimental observation. The way the recognition process develops can best be followed by accessing the graphic files available at http://www.nmr.chem.uu.nl/haddock/movies/cg.html and looking at the movies. They document the different stages of the simulations of the interaction between the CGs and the histo-blood group **H**-trisaccharide/**A**-tetrasaccharide. Starting with the view on ligand

Table V. Comparison of the binding properties of the two proto-type CGs CG-14 and CG-16

Characterization number	Carbohydrate specificity	Galectin type	
		CG-14	CG-16 (Wu et al. 2001)
1	Monosaccharide specificity	β-Anomer of Gal, slightly enhanced by *N*-acetyl group in GalNAc	β-Anomer of Gal, reactivity reduced by *N*-acetyl group in GalNAc
2	Ratio between α and β anomers for *p*-nitrophenyl galactoside[a]	7.5 (α > β; curve 22 versus 28)	7.4 (α > β; curve 9 versus 12)
	Ratio between α and β anomers for methyl galactoside[a]	0.4 (β > α; curve 29 versus 25)	3.4 (α > β; curve 15 versus 20)
	Hydrophobicity for Galα-anomer[b]	Yes (8.1; ratio of curve 22/curve 29)	Yes (9.9; ratio of curve 9/curve 15)
	Hydrophobicity for Galβ-anomer[b]	No (0.5; ratio of curve 28/curve 25)	Yes (4.5; ratio of curve 12/curve 20)
3	Reactivity toward common branch-end epitopes expressed in decreasing order (based on nanomoles comparison)	Galβ1-4GlcNAc (**II**) > Galβ1-4Glc (**L**) > GalNAcβ1-3Gal (**P**) > GalNAcα1-3Gal (**A**) ≥ GalNAcα1-3GalNAc (**F**) ≥ Galβ1-3GalNAc (**T**) ≥ Galβ1-3GlcNAc (**I**) >> Galα1-3Gal (**B**) > Fucα1-2Gal (**H**) >> Galα1-4Gal (**E**) (Table III)	Galβ1-4GlcNAc (**II**) > Galβ1-4Glc (**L**) > Galβ1-3GlcNAc (**I**) > Galα1-3Gal (**B**); Galβ1-3GalNAc (**T**) and Galα1-4Gal (**E**) were inactive
4	The most active β-galactoside	Galβ1-4GlcNAcβ1-3Galβ1-4Glc (**II**β1-3**L**) and Galβ1-4GlcNAc (**II**) and its **H**-type derivative	Galβ1-4GlcNAc (**II**) mainly, extension to **ABH** epitopes reducing activity
5	Ratio of glycotope clusters (simple multivalent form)/monomeric **II**	Tri-antennary glycopeptides with mostly type **II** termini and 2,4,2-branching pattern from asialofetuin was three times more active than monomeric **II** (Table III)	Tri-antennary glycopeptides with mostly type **II** termini and 2,4,2-branching pattern from asialofetuin was three times more active than monomeric **II** (Table III)
6	Substituted branch-end glycans	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
7	Ratio of complex polyvalent glycotopes in macromolecules/monomeric **II**	77 times more active than monomeric **II** (Table II)	Only 5.5 times more active than monomeric **II**
8	The most complementary chain length	**H** active **I**β1-3**L** and Galβ1-4GlcNAcβ1-3Galβ1-4Glc (**II**β1-3**L**)	Galβ1-4GlcNAc (**II**) and Galβ1-3GlcNAcβ1-3Galβ1-4Glc (**I**β1-3**L**)

[a]From Table III.
[b]Based on *p*-nitrophenyl to be more active than the corresponding methyl glycosides as "Yes"; *p*-nitrophenyl to be less active than the corresponding methyl glycosides as "No".

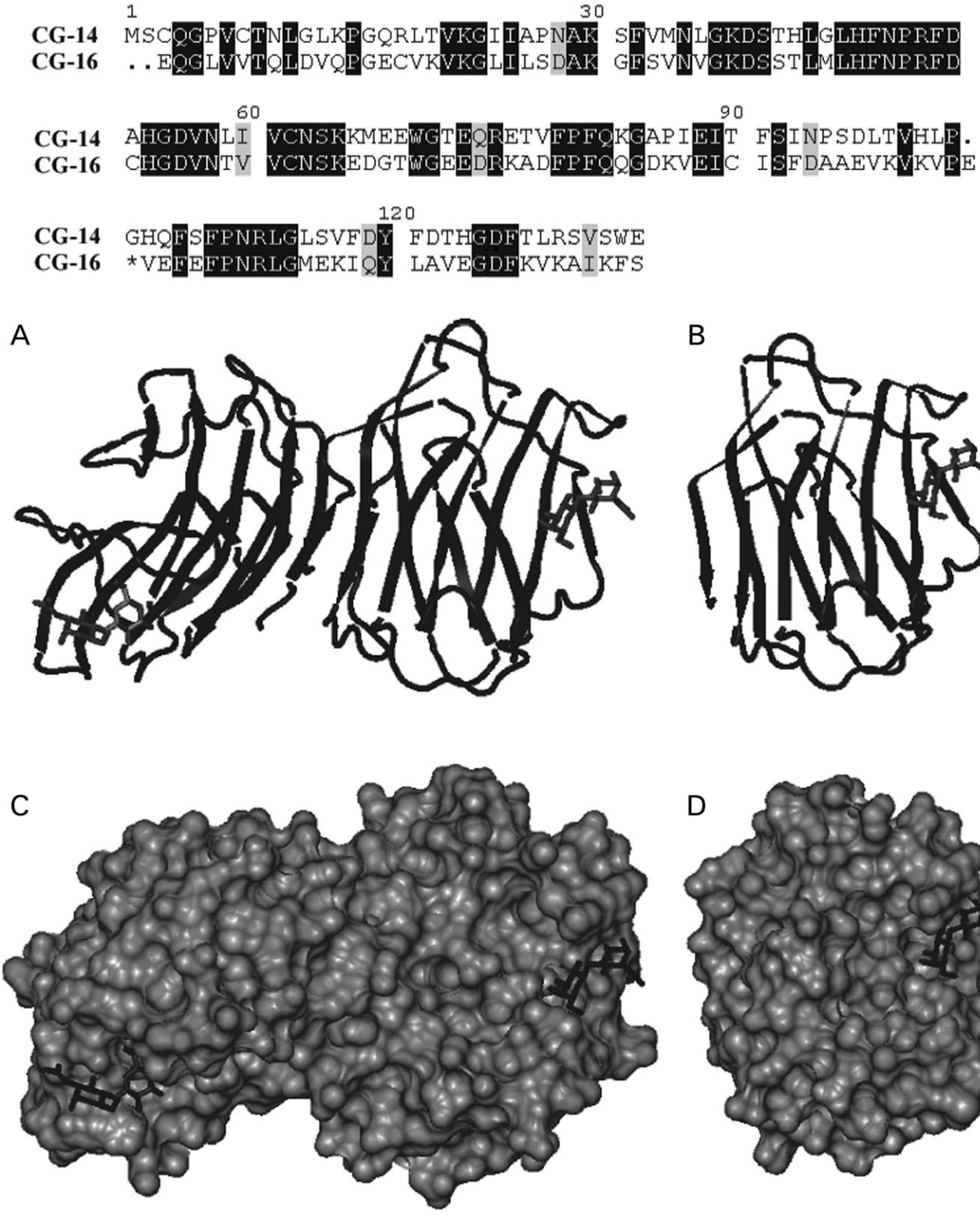

Fig. 3. Comparison of structural aspects between CG-14 and CG-16. Their sequences were set into relation using the program CLUSTAL W (1.60) for multiple sequence alignment (http://bioportal.bic.nus.edu.sg/clustalw/clustalw.html). Identical residues in both sequences are indicated in white letters on black background, similarities in black letters on gray background (top part). The middle and bottom parts present schematic representations of the crystal structure of homodimeric CG-16 (PDB 1QMJ) [(**A**) and (**C**)] and the structure of CG-14 protomer [(**B**) and (**D**)] derived from homology-based calculations (André, Kaltner et al. 2005) in two versions depicting either the protein backbone with the set of antiparallel β-sheets [(A) and (B)] or the contour surface of the two proteins [(C) and (D)]. The latter version is also used in the real-time documentation of the dynamics in the interaction process (http://www.nmr.chem.uu.nl/haddock/movies/cg.html). The contact site for Galβ1-4GlcNAc (**II**) in the CRDs is highlighted by appropriately positioning the disaccharide. Details on the individual amino acids in the vicinity of the ligands of increasing degree of substitution for the two CRDs are given in Figure 4.

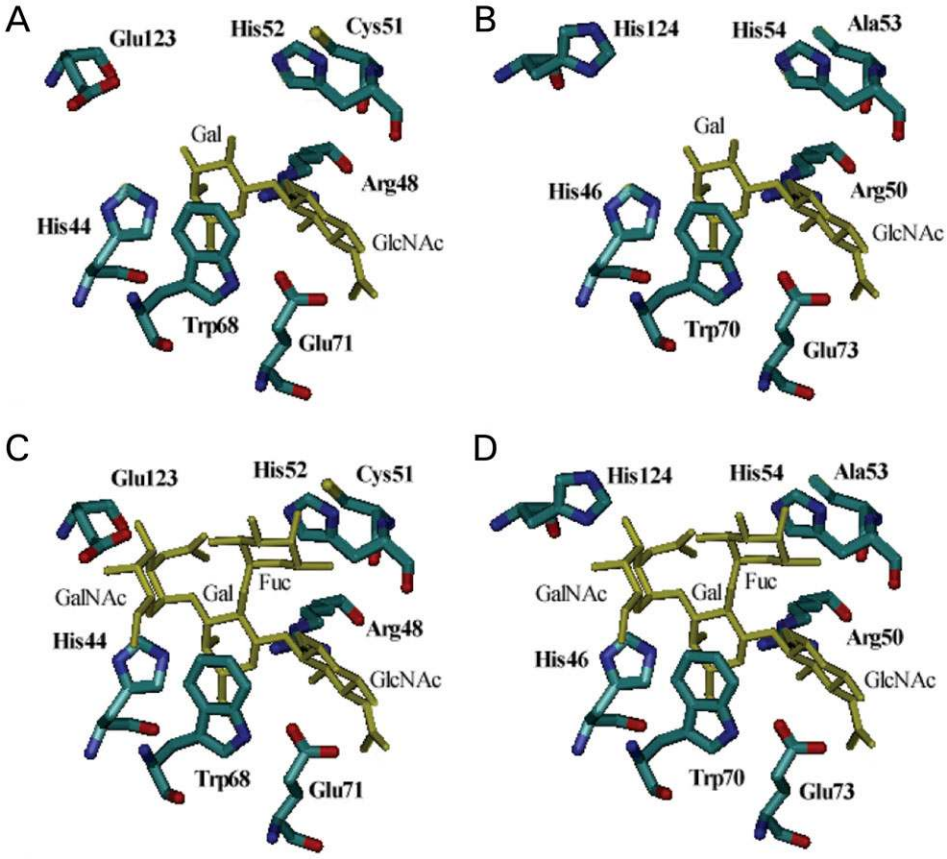

Fig. 4. Graphical illustration of the CRDs of CG-16 (left panel) and CG-14 (right panel) pinpointing the contact sites for Galβ1-4GlcNAc (upper panel) and histo-blood group A tetrasaccharide. Positioning of key contact sites was deliberately kept constant for direct comparison. Both sugars are drawn in low-energy conformations, and the comparison identifies the regions of each CRD responsible for contact to the α1-2/3 substitutions on the core galactose moiety.

flexibility, the four files visualize increasing degree of flexibility also on the level of the protein. Thereby, the impression of a fine-tuned mutual adaptation process reaching a snug fit is conveyed, especially in the case of CG-14 and the tetrasaccharide. Crucial amino acids are marked in the visual presentations for assignment of surface profile changes. By doing so, the re-orientation of the imidazole ring of His124 of CG-14 as consequence of tetrasaccharide movement into the CRD of CG-14 can be followed. This particular movement might impede ligand dissociation, singling the respective substitution out for further analysis.

With respect to the different levels of selectivity for **I** and **II** structures, the graphic files and Figure 4 draw attention to the Glu73/Glu71 moieties (and Arg75/Arg73) which come close to the GlcNAc ring and their environment. The relatively strong interaction of the GlcNAc residue to the Glu73 moiety in CG-14 is a major factor to restrict internal flexibility to a single Φ–Ψ combination (Table VI). Next, sequence divergence starting from the conserved Trp-Gly motif (Thr72-Glu73-Gln74 in CG-14, Glu70-Glu71-Asp72 in CG-16) is found in this region. This difference is conducive for disparate ligand specificity. Conversely, a stretch of sequence identity from the central Trp moiety downstream up to Arg75-Glu76 is spotted between CG-14 and human galectin-1. Fittingly, this lectin reacts 2.5-fold more potently with **II** versus **I** disaccharides in an inhibition assay using asialofetuin as ligand (Sparrow et al. 1987). Equally fitting in terms of an activity–structure correlation, the N-terminal domain of rat galectin-4 with its preference to **I** and also to **T** deviates from this shared sequence markedly with Lys-Glu-Glu-Lys-Lys, and rat galectin-5 with similar **I** and **II** reactivity also shows divergence with Pro-Glu-Glu-Arg-Ser.

Discussion

The first aim of this study was to define the binding profile of CG-14. The specificity to glycotopes was determined in a highly sensitive solid-phase assay using free saccharides as inhibitors. It is yet becoming clear that not only the sequence but also the topology of ligand presentation contributes significantly to the affinity regulation of lectin–carbohydrate interactions. These factors evidently cooperate to endow cell surface glycans with their characteristic binding preferences toward lectins. In order to adequately take this emerging factor into account besides the glycotope specificity, we performed binding assays with a panel of well-characterized gps. In comparison with previous reports on CG-14, using frontal affinity chromatography or binding to neoglycoproteins, our data are in accord with its weak reactivity against sialylated *N*-glycans, an increase in affinity upon chain elongation of LacNAc and the binding to the complex-type tri-antennary glycans irrespective of the branching mode (Hirabayashi et al. 2002; André

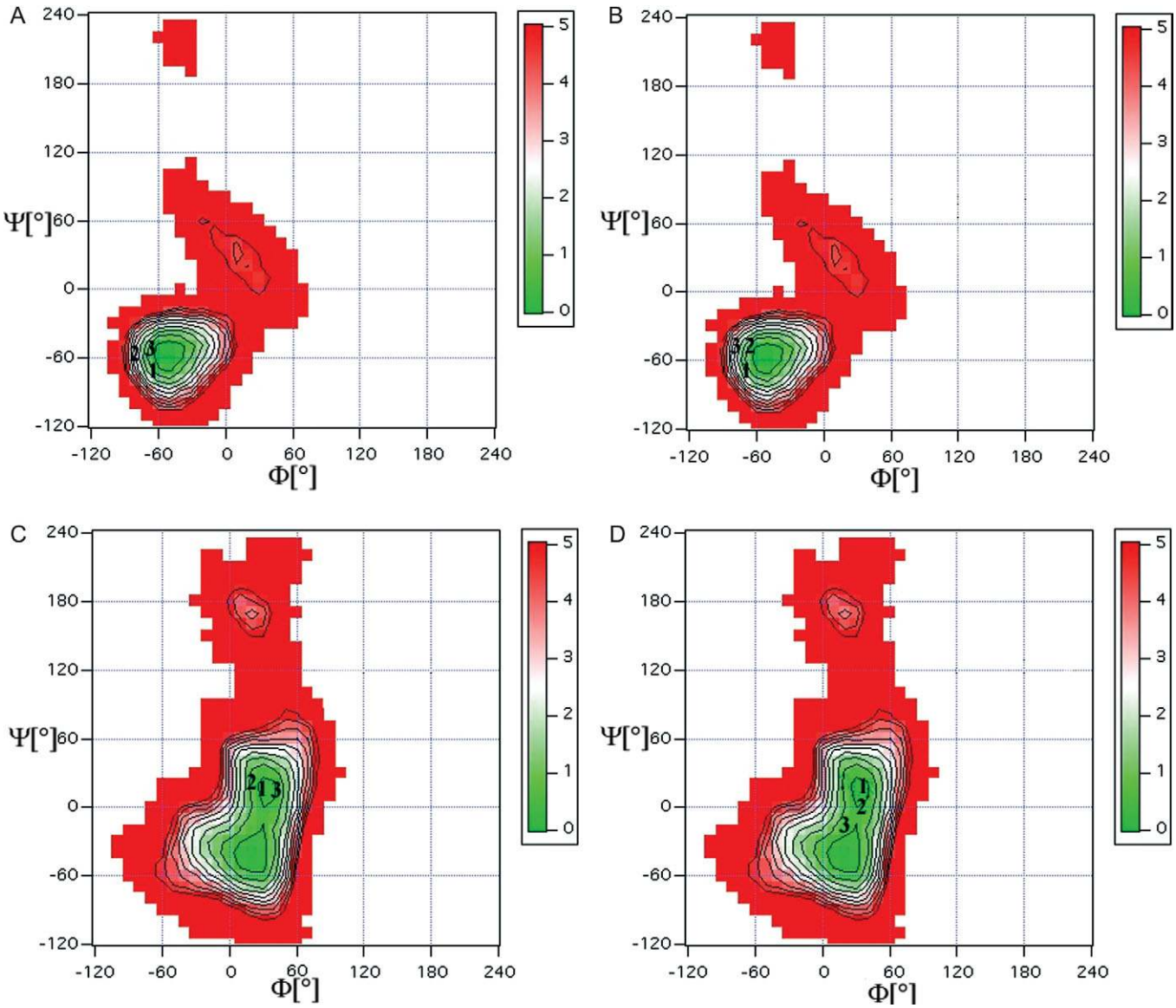

Fig. 5. Illustration of Φ, Ψ, E-plots of the glycosidic linkages in GalNAcα1-3Gal [(**A**) and (**B**)] and in Fucα1-2Gal [(**C**) and (**D**)] of the histo-blood group A-tetrasaccharide, with the internal energy calibration from 0 to 5 kcal/mol. The positions of three low-energy conformations of the sugar ligand in complex with CG-16 [(A) and (C)] or CG-14 [(C) and (D)] are defined by the numbers 1, 2, and 3 in the maps.

et al. 2006). These results thus served as positive quality controls. Moving on, especially the series with gps prepared from human ovarian cyst fluid revealed a pronounced responsiveness of CG-14 to clustered multivalency. This finding underscores the importance of combining measurements on glycotope specificity with topology of glycan presentation.

Addressing the second issue of our study, i.e. systematic comparison of the presented data sets with the published evidence on CG-16 (Wu et al. 2001), differences were detected. CG-16 exhibits a rather strong preference for **II** versus **I** relative to CG-14. In line with previous solid-phase assay data (Wu et al. 2001), the free enthalpy change measured by titration calorimetry reached 0.6 kJ/mol between **II**/**I** as ligands (Bharadwaj et al. 1999). The way CG-16 separates **II**/**I** structures is favorable to localize such glycotopes in cells and tissues, as well as to help fractionate complex glycan mixtures with respect to these types of galactosylation. Thus, this avian lectin can be added as tool to the array of plant lectins commonly used in histochemistry and lectin affinity chromatography (Osawa and Tsuji, 1987; Peumans and van Damme, 1998; Rüdiger and Gabius, 2001; Gabius et al. 2004; Manning et al. 2004). Of note, the previously analyzed rat galectins-4 (N-domain) and -5 either predominantly bind **I** or show no significant capacity for discrimination (Wu et al. 2002, 2004, 2006). Interestingly, the considerable reactivity of galectin-4 to **T** or **Tn** glycotopes not shared by CGs attests to the occurrence of further marked affinity changes within the galectin family (Wu et al. 2002, 2004). They can only thoroughly be mapped by systematic application of this approach. A further distinct difference in ligand selection between CG-14 and CG-16 concerns gps with the histo-blood group **AB** epitopes.

Table VI. Average dihedral Φ and Ψ angles of the glycosidic linkages in histo-blood group H-trisaccharide and the A-tetrasaccharide in the free state, in complex with CG-14/CG-16 on the basis of calculations or in complex with other receptor types in crystals

Disaccharide/protein	GalNAcα1-3Gal[a]	Fucα1-2Gal[a]	Galβ1-4GlcNAc/Glc[a]
Free state	−50°/−60°	50°/20°, 10°/20°, 0°/−30°	50°/20°
CG-14 (1)[b]		19°/20°	54°/−1°
CG-14 (2)[b]		31°/9°	54°/−1°
CG-14 (3)[b]		16°/12°	56°/−4°
CG-16 (1)[b]		19°/8°	48°/6°
CG-16 (2)[b]		13°/8°	15°/29°
CG-16 (3)[b]		54°/−4°	53°/4°
CG-14 (1)[b]	−67°/−77°	33°/20°	49°/5°
CG-14 (2)[b]	−59°/−93°	21°/26°	49°/−9°
CG-14 (3)[b]	−74°/−57°	40°/19°	49°/ 2°
CG-16 (1)[b]	−72°/−65°	38°/17°	53°/0°
CG-16 (2)[b]	−65°/−48°	36°/6°	81°/43°
CG-16 (3)[b]	−74°/−50°	16°/−19°	123°/17°
Coprinopsis cinera agglutinin (1ULF)[c]	−64°/−51°	61°/26°	45°/18°
	−65°/−48°	52°/21°	43°/18°
Coprinopsis cinera agglutinin (1ULD)[c]		54°/20°, 55°/26°, 49°/28°, 58°/16°	49°/18°, 48°/18°, 47°/22°, 53°/17°
Cholera toxin B-subunit (1TLO)[c]	−63°/− 51°,−72°/− 46°, −66°/−53°	49°/24°, 51°/20°, 52°/20°	
Ralstonia solanacearum agglutinin (1BS5)[c]		−5°/−13°, −6°/−13°	
Griffonia simplicifolia agglutinin IV (1GSL)[c]		50°/18°	

[a]Glycosidic angles are defined as follows: Φ = H1-C1-O1-C'X and Ψ = C1-O1-C'X-H'X.
[b]Data for the CGs were derived from docking analysis selecting the energy-minimum position.
[c]The Protein Data Bank codes are given as sources for the listed information.

These two detected activity changes prompted us to initiate computer modeling. It was performed with the aim to discover activity–structure relations which might explain emergence of substitutions at distinct sites in functional terms. In view of the clinical relevance of mammalian galectins in tumor progression, the development of reliable test protocols for computational simulations could be helpful in drug designs. Its aim will be to devise potent blocking reagents, e.g. to block galectin-1-dependent glioma invasion (Camby et al. 2002). The recent cases of the C-type lectins DC-SIGN/DC-SIGNR and their Val351/Ser363 substitution (it eliminates van der Waals interactions of DC-SIGNR to the 2-hydroxyl group of L-fucose in Le^x epitopes and restricts its binding to high-mannose-type *N*-glycans) (Feinberg et al. 2001; Guo et al. 2004; van Liempt et al. 2004) and the extended network of interactions in E- versus P-selectins facilitated by a rotation of the Asn83 residue and a movement of the sialic acid moiety in the sLe^x glycotope (Somers et al. 2000) exemplify ways how sequence divergence can entail diversity in ligand selection. Furthermore, the four haplotypes of the human mannan-binding lectin (wild-type, Gly34Asp, Gly37Gln, Arg32Cys) illustrate the relevance of allelic variance for a lectin's functionality regulated by the quaternary structure (Dean et al. 2006). These advances have been mainly expedited by crystallography. Computer modeling faces at least two serious challenges in the case of galectins. First, long-range effects of substitutions may be difficult to predict. As determined in crystals of Cys2Ser/Arg111His mutants of human galectin-1, inversion by 180° of the side chain of Asp123 and shifts in orientations of His52/Trp68 were caused by sequence changes distant to the CRD (López-Lucendo et al. 2004). Knowledge-based modeling is thus prone to not fully cover such effects. Next, ligand accommodation can involve a dynamic series of mutual steps of structural adaptation. Arguing against a rigid lock-and-key recognition in solution, small-angle neutron scattering was instrumental in determining a significant decrease in the gyration radius of human galectin-1 by ligand binding (He et al. 2003). This challenge can be met by introducing flexibility to the simulations for both binding partners. Therefore, rigid-body docking was deliberately avoided in our analysis, which strived for examining aspects of the dynamics of the molecular interplay. This objective engenders the illustration of the recognition process by motion pictures. They are accessible to the reader in addition to common figures.

With focus on the ligand, the modeling indicates that the binding of the histo-blood group tri- and tetrasaccharides is not accompanied by a major distortion of the free-state low-energy conformations. They were previously described by molecular mechanics calculations and NMR spectroscopy (Yan and Bush, 1990; Imberty et al. 1995; Koca et al. 1995; Casset et al. 1996; Otter et al. 1999; Azurmendi and Bush,

Table VII. Compilation of values for energy of interaction[a] between CG-14 and the histo-blood group H-trisaccharide and A-tetrasaccharide determinants (for spatial positioning of the listed amino acid residues, please see Figure 4)

CG-14-H_{tri}/A_{tetra}	E_{tot} (kcal/mol)	E_{vdW}	E_{Coul}
His46	−9.6/−6.6	−2.1/−3.3	−7.5/−3.3
Arg50	−28.2/−30.0	−1.2/−2.3	−27.0/−27.7
His54	−20.5/−13.7	−5.5/−5.4	−15.0/−8.3
Asp56	−11.6/+7.9	−1.3/−0.4	−10.3/+8.3
Asn63	−8.1/−7.0	−0.2/−0.2	−7.9/−6.8
Lys65	−/−6.6	−/−0.4	−/−6.2
Glu69	−/−8.2	−/−2.5	−/−5.7
Trp70	−7.4/−13.9	−7.0/−9.0	−0.4/−4.9
Glu73	−27.9/−43.3	−3.3/+0.4	−24.6/−43.7
Arg75	−11.7/−6.0	−1.4/−1.3	−10.5/−4.7
His124	−/−24.0	−/−2.0	−/−22.0
Σ	−125.0/−151.4	−22.0/−26.4	−103.0/−125.0
Gal	−50.8/−52.7	−15.0/−13.7	−35.8/−39.0
Fuc	−10.8/−10.2	−5.2/−4.8	−5.6/−5.4
GalNAc	−/−32.5	−/−12.7	−/−19.8
Σ	−61.6/−95.4	−20.2/−31.2	−41.4/−64.2

[a]The total energy (E_{tot}) is subdivided into the van der Waals (vdW) and Coulomb (E_{Coul}) terms; the calculations of intermolecular energies use an 8.5 Å cut-off with OPLS nonbonded parameters (Jorgenson and Tirado-Rives 1988). The energies are averages calculated over the final ensembles of 10 conformers. They are reported for the interaction of CG-14 with H_{tri} and A_{tetra} (E_{Htri}/E_{Atetra}).

Table VIII. Compilation of values for energy of interaction[a] between CG-16 and the histo-blood group H-trisaccharide and A-tetrasaccharide determinants (for spatial positioning of the listed amino acid residues, please see Figure 4)

CG-16-H_{tri}/A_{tetra}	E_{tot}	E_{vdW}	E_C
His44	−9.6/−1.0	−2.0/−1.9	−7.6/+0.9
Arg48	−30.1/−33.0	−1.1/−1.8	−29.0/−31.2
His52	−13.7/−21.3	−4.9/−6.2	−8.8/−15.1
Asp54	−4.3/+7.9	−0.9/−0.4	−3.4/+8.3
Asn61	−7.5/−6.5	−0.3/−0.6	−7.2/−5.9
Lys63	−/−15.5	−/−0.5	−/−15.0
Thr67	−/−7.7	−/−1.8	−/−5.9
Trp68	−14.9/−14.0	−10.2/−9.0	−4.7/−5.0
Glu71	−30.0/−40.4	−2.7/+0.8	−27.3/−41.2
Arg73	−6.2/−5.3	−2.2/−1.2	−4.0/−4.1
Glu123	−/−42.3	−/−1.8	−/−40.5
Σ	−116.3/−179.1	−24.3/−24.4	−92.0/−154.7
Gal	−47.5/−44.5	−15.0/−12.4	−32.5/−32.1
Fuc	−6.7/−6.8	−4.8/−4.9	−1.9/−1.9
GalNAc	−/−69.2	−/−10.9	−/−58.3
Σ	−54.2/−120.5	−19.8/−28.2	−34.4/−92.3

[a]The total energy (E_{tot}) is subdivided into the van der Waals (vdW) and Coulomb (E_C) terms; the calculations of intermolecular energies use an 8.5 Å cut-off with OPLS nonbonded parameters (Jorgenson and Tirado-Rives 1988). The energies are averages calculated over the final ensembles of 10 conformers. They are reported for the interaction of CG-16 with H_{tri} and A_{tetra} (E_{Htri}/E_{Atetra}).

2002). These rather rigid ligands (please see Figure 5 for accessible conformational space) appear to home into the CRDs without a major entropic penalty in the thermodynamic balance sheet owing to the loss of conformational freedom. The implied selection of a low-energy conformer present in solution has already been described for CG-16 and the disaccharide Galβ1-2Gal (Siebert et al. 1996). The computed topology of the interactions between the tri- and tetrasaccharides and the CRDs provide a first view on the extended binding site of these animal lectins for α1-2/α1-3-substituted β-galactosides. So far, respective information was only available from a protein of the basidiomycete *Coprinopsis cinerea* (Walser et al. 2004). The presented interaction analysis details the individual contributions at the level of amino acids and sugar units. A substantial portion originates from the GalNAc unit, exceeding Fuc as source of binding energy significantly. That this region of the CRD is most likely relevant for physiological ligands is epitomized by the 2.75 and 8.44 kJ/mol gains in free enthalpy when testing Galβ1-4Glc in comparison with the histo-blood group **H**-trisaccharide and **A**-tetrasaccharide (at 279.6–280.4 K) and human galectin-3 (Bachhawat-Sikder et al. 2001). Moreover, placement of bulky substituents at the 3′-position of galactose such as a 4-methoxy-2,3,5,6-tetrafluorobenzamido moiety which stacks against Arg144 is documented for human galectin-3 (Sörme et al. 2005).

With focus on the protein, the modeling data presented make it possible to relate distinct sequence substitutions to alterations in direct contacts and/or orientations of crucial side chains in the model of the two CGs. It makes sense to avoid a mutational loss of an indispensable contact site but to modulate its functionality by altering its immediate vicinity, as evidenced for the region downstream of the central Trp residue and also the His124/Glu123 substitution. In addition to **I**/**II** preference, the vicinity of the Trp residue comes into contact with the carboxyl group of sialyllactose in human galectin-1 (Ala in position 67) (Siebert et al. 2003). It thus regulates the activity of sialylated glycans, too. By making precise calculation-based predictions on the spatial vicinity between distinct constituents of ligand and lectin, this work gives research a clear direction to map the spatial aspects of the interaction experimentally. For this purpose, monitoring saturation transfer in NMR-spectroscopic experiments can prove valuable (Siebert et al. 2003). Besides addressing questions on the impact of sequence changes, the computational analysis will also find application in the design of inhibitors, as noted at the end of the *Introduction*. Equally important, rational design of new lectin variants will become feasible. Explicitly, this kind of in silico analysis may help gauge the effect of mutagenesis when aiming at producing galectin variants with custom-made properties as laboratory tools. Finally, the full consideration of flexibility for both binding partners visualizes that an intricate structural interplay can allow for initial contact formation. The insights into the routes toward an optimal fit which the simulations provided can help narrow the existing gap in our understanding between the static view of ligand binding obtained by crystallography and not-yet-rigorously defined dynamics in solution.

Materials and methods

Free saccharides and glycopeptides

Mono-, di-, and oligosaccharides were purchased from or prepared by Dextra (Reading, Berkshire, UK) and Sigma (Munich, Germany). Tri-antennary glycopeptides with Gal(β1-4)GlcNAc termini and a small percentage of Gal(β1-3)GlcNAc termini in the α1-3 branch were obtained from asialofetuin after pronase digestion and repeated fractionation by BioGel P-4 400 mesh column chromatography (Wu et al. 1998). The **Tn-**presenting glycopeptides (**Tn** clusters) used for this study were mixtures obtained in the filterable fraction (molecular mass cut-off <3000 Da) (Wu et al. 1997).

Gps and polysaccharides

Sources and further processing of these compounds have been given in detail in the previous report on CG-16 (Wu et al. 2001). The predominant galectin-reactive sugar determinant of each test substance is listed in Tables I and II, for example, defining gps from human ovarian cyst fluid (i.e. Cyst Beach, MSS, Tighe, Mcdon, and JS phenol insoluble) to harbor **I**/**II** and histo-blood group **ABH** determinants.

Cloning, expression, and purification of CG-14

Total RNA from embryonic chicken kidney (14th day of incubation) was isolated using an RNAeasy kit (Qiagen, Hilden, Germany) according to the manufacturer's protocols; 1 μg of the resulting total RNA was used as template for first-strand synthesis of cDNA. Reverse transcription was performed for 50 min at 42 °C in a 20 μL reaction volume containing first-strand buffer (50 mM Tris–HCl, pH 8.3, 75 mM KCl, 15 mM $MgCl_2$), dNTPs (1 mM each), oligo$(dT)_{12-18}$ primer (0.5 μg), 200 units Superscript™ II RNase H^- reverse transcriptase (Invitrogen/Life Technologies, Karlsruhe, Germany), and diethylpyrocarbonate-treated water. The reaction was terminated by incubation at 70 °C for 15 min, and 2 μL of this mixture was transferred into a new tube for PCR amplification, with the sense primer 5′-CGTA CGGGATCCATCTCTTGTCAGGGACC-3′ and the antisense primer 5′-CGCTAGGGATCCTTACCCCAGCTGACAG-3′ harboring *Bam*HI restriction sites (underlined). The reaction mixture was prepared as recommended by the manufacturer of the *Taq* polymerase (Qiagen). After an initial denaturation step at 94 °C for 4 min, 35 cycles (denaturation at 94 °C for 45 s, annealing at 63 °C for 45 s, extension at 72 °C for 1 min) were run. The final extension step was performed at 72 °C for 5 min. The produced amplicon was ligated into the *Eco*RV-linearized pET-Blue-1 AccepTor™ vector (Novagen®, Merck Biosciences, Darmstadt, Germany) having single 3′dU overhangs and propagated in the commercially available *Escherichia coli* strain NovaBlue. DNA sequencing ascertained complete agreement to the cDNA sequence as published in GenBank (accession no. M11674). To facilitate cloning into an efficient expression system, further sequence modification and subcloning were necessary. Therefore, using the pET-Blue-1/CG-14 plasmid as template in a megaprimer PCR, the *Hin*dIII restriction site was impaired by a single nucleotide exchange from C to T at position 93, which caused no amino acid substitution. The amplicon obtained in this mutagenesis procedure was used for a further standard PCR, with the sense primer 5′-CGCTAGCCATGGCTTGT CAGGGACC-3′ containing an internal *Noc*I restriction site (underlined) and the antisense primer 5′-CGT ACGAAGCTTTTACTCCCAGCTGACAG-3′ with an internal *Hin*dIII restriction site (underlined). The steps of the PCR protocol were the same as outlined above in this paragraph. Following digestion of the product exploiting the restriction sites, the amplicon was cloned into the *Nco*I/*Hin*dIII sites of the prokaryotic expression vector pQE60 (Qiagen) used for transformation. Bacteria were grown in terrific broth medium (Roth, Karlsruhe, Germany) at 37 °C, and recombinant production was performed under conditions defined to be optimal by systematic testing, applying isopropyl-β-D-thiogalactoside at a final concentration of 0.5 mM to yield approximately 6 mg CG-14/L. Purification by affinity chromatography on lactosylated Sepharose 4B, obtained after resin activation by divinyl sulfone, as crucial step, quality and purity controls by one- and two-dimensional SDS–PAGE/silver staining and mass spectrometry, biotinylation under activity-preserving conditions and controls for activity by hemagglutination, and solid-phase assays with immunological detection of extent of bound lectin with/without label were performed as described in Gabius et al. (1984), Gabius (1990), André, Kaltner et al. (2004), and André, Kojima et al. (2005).

Lectin-enzyme binding assay

The assay was performed according to the procedures and conditions described previously (Duk et al. 1994). The volume of each glycosubstance-containing solution applied to the plate for coating was 50 μL/well, and all incubations, except for the coating step performed overnight at 4 °C, were carried out at room temperature (20–25 °C). Tris-buffered saline (TBS; 0.05 M Tris–HCl, 0.15 M NaCl, pH 7.35) containing 0.05% Tween 20 (TBS-T) was used to prepare the solutions and for washing plate wells between incubation steps.

The 96-well microtiter plates (F96 Maxisorp, Nunc-Immuno plate, Kamstrupvej, Denmark) were coated with gps in 0.05 M carbonate buffer (0.05 M $NaHCO_3$/0.05 M Na_2CO_3), pH 9.6, overnight at 4 °C. After washing the plate three or four times with 200 μL TBS-T, solution with biotinylated CG-14 (250 ng/well) was added and the plate was incubated for 30 min. The plates were then carefully washed three or four times to remove unabsorbed lectin, and the ExtrAvidin/Alkaline Phosphatase Conjugate (diluted 1 : 10 000, Sigma) was then added to detect the specifically bound probes. After a 1 h incubation step, the plates were again washed at least four times with TBS-T and then incubated with *p*-nitrophenyl phosphate (Sigma phosphatase substrate 5 mg tablets) in 0.05 M carbonate buffer, pH 9.6, containing 1 mM $MgCl_2$ (one tablet/5 mL). The resulting absorbance was read at 405 nm in a microtiter plate reader after 2–4-h incubation at room temperature in the dark with the substrate-containing solution.

For inhibition studies, serially diluted inhibitor samples were mixed with an equal volume of lectin solution containing a fixed amount of CG-14 (250 ng/well). The control lectin sample was diluted 2-fold with TBS-T. After incubation at room temperature for 1 h, the binding of these samples (lectin–inhibitor mixture) to cyst Beach P-1-coated plate was determined as described in the paragraph above.

The inhibitory activity of each compound was delineated from the inhibition curves. It is presented in the tables as the amount of inhibitor (ng or nmol per well) resulting in 50% inhibition of the control level of lectin binding in the absence of any inhibitor.

All experiments were performed in duplicate or triplicate, and data are presented as the mean values of the combined results. The standard deviation did not exceed 10% and in most experiments was even less than 5% of the mean value. Blank wells, where either the coating step or that of incubation with biotinylated lectin was omitted from the standard protocol, gave low absorbance values (below 0.1) after a 2–4-h period of incubation with the chromogenic substrate at room temperature. This showed that blocking the wells before lectin addition was not necessary to achieve low background values, when Tween 20 was present in the buffer.

Computational analysis

The crystal structure of CG-16 (PDB code 1QMJ) and a homology-based structural model of CG-14 (André, Kaltner et al. 2005) were used for further in silico processing in the modeling analysis, together with low-energy conformations of the histo-blood group **ABH** tri- and tetrasaccharides derived from crystal structures (PDB codes 1ULD and 1ULF) and molecular mechanics calculations yielding detailed Φ, Ψ, E-plots to define the conformational space of the ligands (Siebert et al. 2000). Systematic investigations on the way these ligands can be accommodated in the CRDs of both galectins were performed by the information-driven docking program HADDOCK (High Ambiguity Driven DOCKing) (version 1.3) (Dominguez et al. 2003). The parallhdg.5.3 data set (Linge et al. 2003), which includes the optimized potentials for liquid simulations (OPLS) parameters for nonbonded interactions (Jorgenson and Tirado-Rives, 1988), was applied to the protein part, and respective sets for the carbohydrates were adapted from crystallography and NMR system (Brunger et al. 1998). Coulomb/van der Waals energy terms characterizing the interaction were calculated using a 8.5 Å cut-off value. The dielectric constant ε was set to 10 in the first stage of the calculations, when no solvent molecules were present. When water molecules were explicitly added for the second part of the simulations, ε was set to 1. In order to figure out contributions of individual amino acid residues and monosaccharides to the energy balance of the complete interaction process, as carried out in detail previously for a carbohydrate-binding 15-mer peptide (Siebert et al. 2006), the two energetic terms were separately computed for each contact between an amino acid of the lectin and a sugar moiety below the cut-off limit.

Modeling initial contact formation between ligand and lectin was guided by the evidence that the CGs maintain the characteristic way a ligand's central galactose unit resides in the vicinity of the Trp moiety of the CRD (Siebert et al. 1997; André, Kaltner et al. 2005). In detail, the distance restraints for defining the relative positions of the binding partners suitable to drive the simulations were derived from inspection of the mode of interaction between human galectin-1 and galactose as part of lactose in the crystal state (PDB code 1W6M), using the LigPlot program (Wallace et al. 1995). Five distance restraints (with distance values between 2 and 4 Å) involving the same type of amino acid at equivalent positions were accordingly obtained in the cases of both CGs. They involve atom pairs of the sugar and the respective contacting amino acid residue, namely the oxygen atom O1 of galactose and the hydrogen atom NHε2 of His44 (CG-16) or His46 (CG-14), the oxygen atoms O4 and the O5 atoms of the sugar and the hydrogen atom NH2η of Arg48 (CG-16) or Arg50 (CG-14), the oxygen atom O6 of the sugar and the hydrogen atom NHδ2 of Asn61 (CG-16) or Asn63 (CG-14), as well as the HO6 hydrogen atom of the ligand and the oxygen atom Oε of Glu71(CG-16) or Glu73 (CG-14). Having defined this set of knowledge-based interatom distances, the ensuing calculations were divided into separate stages of increasing level of refinement in order to describe the conformational space of both binding partners in the complex comprehensively. Protein–ligand docking allowing full flexibility of the oligosaccharide was performed during the simulated annealing stage of HADDOCK. In other words, rigid-body docking was avoided. The hardware for these computations was a Linux Beowulf cluster consisting of 16 dual processor nodes (Opteron 246, 2 GHz). To gauge the incurred CPU time costs, one complete docking run took approximately 8 h on 10 processors. The first stage encompassed a conformational search at high temperature (2000 K), followed by three annealing stages (please see Dominguez et al. 2003 for program details). The total number of integration steps was 16 000. The ligand maintained full flexibility throughout the protocol to let it adopt its bound-state conformation should complex formation be accompanied by a distortion of the free-state low-energy structure. Regarding the lectins, the degree of flexibility was increased stepwise, first only for side chains of the amino acids and later for the CGs' backbones as well. As a result, a total of 300 models of the complex between each galectin and ligand were generated, which were then systematically ranked according to their interaction energy. The top 250 structures of each complex were subjected to a final refinement in the presence of explicit water molecules. The resulting models were grouped into clusters on the basis of similarity of their binding modes. Clustering was performed using a 1 Å cut-off value on the basis of the pairwise RMSD matrix for the carbon atoms of each ligand after fitting on the backbone atoms of the protein. The cluster harboring the most favorable intermolecular energy terms was selected as representative model of the complex, and this structure was closely examined regarding energetic aspects of the intermolecular recognition on the basis of amino acids/monosaccharides. The motion pictures visualizing the docking process were generated with Pymol (http://www.pymol.org).

Supplementary data

Supplementary data are available at Gycobiology online (http://glycob.oxfordjournals.org/).

Acknowledgments

This study was generously supported by funding from the Chang-Gung Medical Research Project (CMRPD no. 33022, Kwei-san, Tao-yuan, Taiwan), the National Science Council,

Taiwan (NSC 94-2320-B-182-044, NSC 94-2320-B-182-053), the Mizutani Foundation for Glycoscience (Tokyo, Japan), the Verein zur Förderung des biologisch-technologischen Fortschritts in der Medizin e. V. (Heidelberg, Germany), and an EC Marie Curie research training network (contract no. MRTN-CT-2005–019561). A personal travel grant (to S.A.) from the Deutscher Akademischer Austauschdienst (DAAD; Bonn, Germany) and the National Science Council (Taipei, Taiwan) to support the cooperation between Taiwan and Germany and support from the Dutch Organisation for Scientific Research (NWO) through the Molecule to Cell program (to M.K.) are also gratefully acknowledged. Moreover, we are indebted to Drs B. Friday, G. Hase, and S. Namirha for helpful comments.

Conflict of interest statement

None declared.

Abbreviations

BSM, bovine submandibular mucin; CGs, chicken galectins; CRD, carbohydrate recognition domain; ELLSA, enzyme-linked lectinosorbent assay; gps, glycoproteins; HADDOCK, High Ambiguity Driven DOCKing; HSM, hamster submacillary mucin; OPLS, optimized potentials for liquid simulations; OSM, ovine salivary mucin; RMSD, root mean square deviation; RSL, rat sublingual gp; TBS, Tris-buffered saline; TBS-T, Tris-buffered saline containing Tween 20; THGP, Tamm-Horsfall gp

References

Akimoto Y, Obinata A, Hirabayashi J, Sakakura Y, Endo H, Kasai K-I, Hirano H. 1993. Secretion of endogenous 16-kDa β-galactoside-binding lectin from vitamin A-pretreated chick embryonic cultured skin. Exp Cell Res. 205:251–260.

Akimoto Y, Obinata A, Hirabayashi J, Sakakura Y, Endo H, Kasai K-I, Hirano H. 1995. Changes in expression of two endogenous β-galactoside-binding isolectins in the dermis of chick embryonic skin during development in vivo and in vitro. Cell Tissue Res. 279:3–12.

André S, Cejas Ortega PJ, Alamino Perez M, Roy R, Gabius H-J. 1999. Lactose-containing starburst dendrimers: influence of dendrimer generation and binding-site orientation of receptors (plant/animal lectins and immunoglobulins) on binding properties. Glycobiology. 9:1253–1261.

André S, Kaltner H, Furuike T, Nishimura S, Gabius H-J. 2004. Persubstituted cyclodextrin-based glycoclusters as inhibitors of protein–carbohydrate recognition using purified plant and mammalian lectins and wild-type and lectin-gene-transfected tumor cells as targets. Bioconjugate Chem. 15:87–98.

André S, Kaltner H, Lensch M, Russwurm R, Siebert H-C, Fallsehr C, Tajkhorshid E, Heck AJR, von Knebel-Döberitz M, Gabius H-J, et al. 2005. Determination of structural and functional overlap/divergence of five proto-type galectins by analysis of the growth-regulatory interaction with ganglioside GM_1 in silico and in vitro on human neuroblastoma cells. Int J Cancer. 114:46–57.

André S, Kojima S, Gundel G, Russwurm R, Schratt X, Unverzagt C, Gabius H-J. 2006. Branching mode in complex-type triantennary *N*-glycans as regulatory element of their ligand properties. Biochim Biophys Acta. 1760: 768–782.

André S, Kojima S, Prahl I, Lensch M, Unverzagt C, Gabius H-J. 2005. Introduction of extended LEC14-type branching into core-fucosylated biantennary *N*-glycan. FEBS J. 272:1986–1998.

André S, Liu B, Gabius H-J, Roy R. 2003. First demonstration of differential inhibition of lectin binding to synthetic tri- and tetravalent glycoclusters from cross-coupling to rigidified 2-propynyl lactoside. Org Biomol Chem. 1:3909–3916.

André S, Pieters RJ, Vrasidas I, Kaltner H, Kuwabara I, Liu F-T, Liskamp RMJ, Gabius H-J. 2001. Wedgelike glycodendrimers as inhibitors of binding of mammalian galectins to various glycoproteins, lactose maxiclusters and cell surface glycoconjugates. ChemBioChem. 2:822–830.

André S, Unverzagt C, Kojima S, Frank M, Seifert J, Fink C, Kayser K, von der Lieth C-W, Gabius H-J. 2004. Determination of modulation of ligand properties of synthetic complex-type biantennary *N*-glycans by introduction of bisecting GlcNAc in silico, in vitro and in vivo. Eur J Biochem. 271:118–134.

Azurmendi HF, Bush LA. 2002. Conformational studies of blood group A and blood group B oligosaccharides using NMR residual dipolar couplings. Carbohydr Res. 337:905–915.

Bachhawat-Sikder K, Thomas CJ, Surolia A. 2001. Thermodynamic analyis of the binding of galactose and poly-*N*-acetyllactosamine derivatives to human galectin-3. FEBS Lett. 500:75–79.

Bharadwaj S, Kaltner H, Korchagina EY, Bovin NV, Gabius H-J, Surolia A. 1999. Microcalorimetric indications for ligand binding as a function of the protein for galactoside-specific plant and avian lectins. Biochim Biophys Acta. 1472:191–196.

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al. 1998. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D54:905–921.

Camby I, Belot N, Lefranc F, Sadeghi N, de Launoit Y, Kaltner H, Musette S, Darro F, Danguy A, Salmon I, et al. 2002. Galectin-1 modulates human glioblastoma cell migration into the brain through modifications to the actin cytoskeleton and levels of expression of small GTPases. J Neuropathol Exp. Neurol. 61:585–596.

Casset F, Peters T, Etzler M, Korchagina EY, Nifant'ev N, Pérez S, Imberty A. 1996. Conformational analysis of blood group A trisaccharide in solution and in the binding site of *Dolichos biflorus* lectin using transient and transferred nuclear Overhauser enhancement (NOE) and rotating-frame NOE experiments. Eur J Biochem. 239:710–719.

Cooper DNW. 2002. Galectinomics: finding themes in complexity. Biochim Biophys Acta. 1572:209–231.

Dam TK, Gabius H-J, André S, Kaltner H, Lensch M, Brewer CF. 2005. Galectins bind to the multivalent glycoprotein asialofetuin with enhanced affinities and a gradient of decreasing binding constants. Biochemistry. 44:12564–12571.

Dean MM, Heatley S, Minchinton RM. 2006. Heteroligomeric forms of codon 54 mannose binding lectin (MBL) in circulation demonstrate reduced in vitro function. Mol Immunol. 43:950–961.

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.

Duk M, Lisowska E, Wu JH, Wu AM. 1994. The biotin/avidin-mediated microtiter plate lectin assay with the use of chemically modified glycoprotein ligand. Anal Biochem. 221:266–272.

Feinberg H, Mitchell DA, Drickamer K, Weis WI. 2001. Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. Science. 294:2163–2166.

Fischer C, Sanchez-Ruderisch H, Welzel M, Wiedenmann B, Sakai T, André S, Gabius H-J, Khachigian L, Detjen KM, Rosewicz S. 2005. Galectin-1 interacts with the α5β1 fibronectin receptor to restrict carcinoma cell growth via induction of p21 and p27. J Biol Chem. 280:37266–37277.

Gabius H-J. 1990. Influence of type of linkage and spacer on the interaction of β-galactoside-binding proteins with immobilized affinity ligands. Anal Biochem. 189:91–94.

Gabius H-J. 1997. Animal lectins. Eur J Biochem. 243:543–576.

Gabius H-J. 2001. Glycohistochemistry: the why and how of detection and localization of endogenous lectins. Anat Histol Embryol. 30:3–31.

Gabius H-J. 2006. Cell surface glycans: the why and how of their functionality as biochemical signals in lectin-mediated information transfer. Crit Rev Immunol. 26:43–80.

Gabius H-J, Engelhardt R, Rehm S, Cramer F. 1984. Biochemical characterization of endogenous carbohydrate-binding proteins from spontaneous murine rhabdomysarcoma, mammary adenocarcinoma and ovarian teratoma. J Natl Cancer Inst. 73:1349–1357.

Gabius H-J, Siebert H-C, André S, Jiménez-Barbero J, Rüdiger H. 2004. Chemial biology of the sugar code. ChemBioChem. 5:740–764.

Guo Y, Feinberg H, Conroy E, Mitchell DA, Alvarez R, Blixt O, Taylor ME, Weis WI, Drickamer K. 2004. Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. Nat Struct Mol Biol. 11:591–598.

He L, André S, Siebert H-C, Helmholz H, Niemeyer B, Gabius H-J. 2003. Detection of ligand- and solvent-induced shape alterations of cell-growth-regulatory human lectin galectin-1 in solution by small angle neutron and X-ray scattering. Biophys J. 85:511–524.

Hirabayashi J, Hashidate T, Arata Y, Nishi N, Nakamura T, Hirashima M, Urashima T, Oka T, Futai M, Müller WEG, et al. 2002. Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. Biochim Biophys Acta. 1572:232–254.

Houzelstein D, Gonçalves IR, Fadden AJ, Sidhu SS, Cooper DNW, Drickamer K, Leffler H, Poirier F. 2004. Phylogenetic analysis of the vertebrate galectin family. Mol Biol Evol. 21:1177–1187.

Imberty A, Mikros E, Koca J, Mollicone R, Oriol R, Pérez S. 1995. Computer simulation of histo-blood group oligosaccharides: energy maps of all constituting disaccharides and potential energy surfaces of 14 ABH and Lewis carbohydrate antigens. Glycoconjugate J. 12:331–349.

Jorgenson WL, Tirado-Rives J. 1988. The OPLS (optimized potentials for liquid simulations) potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc. 110: 1657–1666.

Koca J, Pérez S, Imberty A. 1995. Conformational analysis and flexibility of carbohydrates using the CICADA approach with MM3. J Comput Chem. 16:296–310.

Kopitz J, Russwurm R, Kaltner H, André S, Dotti CG, Gabius H-J, Abad-Rodríguez J. 2004. Hippocampal neurons and recombinant galectins as tools for systematic carbohydrate structure-function studies in neuronal differentiation. Dev Brain Res. 153:1043–1060.

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.

Leffler H, Barondes SH. 1986. Specificity of binding of three soluble rat lung lectins to substituted and unsubstituted β-galactosides. J Biol Chem. 261: 10119–10126.

Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M. 2003. Refinement of protein structures in explicit solvent. Proteins. 50: 496–506.

López-Lucendo MF, Solís D, André S, Hirabayashi J, Kasai K-I, Kaltner H, Gabius H-J, Romero A. 2004. Growth-regulatory human galectin-1: crystallographic characterisation of the structural changes induced by single-site mutations and their impact on the thermodynamics of ligand binding. J Mol Biol. 343:957–970.

Manning JC, Seyrek K, Kaltner H, André S, Sinowatz F, Gabius H-J. 2004. Glycomic profiling of developmental changes in bovine testis by lectin histochemistry and further analysis of the most prominent alteration on the level of the glycoproteome by lectin blotting and lectin affinity chromatography. Histol Histopathol. 19:1043–1060.

Osawa T, Tsuji T. 1987. Fractionation and structural assessment of oligosaccharides and glycopeptides by use of immobilized lectins. Annu Rev Biochem. 56:21–42.

Otter A, Lemieux RU, Ball RG, Venot AP, Hindsgaul O, Bundle DR. 1999. Crystal state and solution conformation of the B blood group trisaccharide α-l-Fuc*p*-(1–2)-[α-d-Gal*p*]-(1–3)-β-d-Gal*p*-OCH_3. Eur J Biochem. 259: 295–303.

Peumans WJ, van Damme EJM. 1998. Plant lectins: specific tools for the identification, isolation and characterization of *O*-linked glycans. Crit Rev Biochem Mol Biol. 33:209–259.

Rüdiger H, Gabius H-J. 2001. Plant lectins: occurrence, biochemistry, functions and applications. Glycoconjugate J. 18:589–613.

Sakakura Y, Hirabayashi J, Oda Y, Ohyama Y, Kasai K-I. 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.

Schneller M, André S, Cihak J, Kaltner H, Merkle H, Rademaker GJ, Haverkamp J, Thomas-Oates J, Lösch U, Gabius H-J. 1995. Differential binding of two chicken β-galactoside-specific lectins to homologous lymphocyte subpopulations and evidence for inhibitory activity of the dimeric lectin on stimulated T cells. Cell Immunol. 166:35–43.

Siebert H-C, Adar R, Arango R, Burchert M, Kaltner H, Kayser G, Tajkhorshid E, von der Lieth C-W, Kaptein R, Sharon N, et al. 1997. Involvement of laser photo CIDNP (chemically induced dynamic nuclear polarization)-reactive amino acid side chains in ligand binding by galactoside-specific lectins in solution. Eur J Biochem. 249:27–38.

Siebert H-C, André S, Asensio JL, Cañada FJ, Dong X, Espinosa JF, Frank M, Gilleron M, Kaltner H, Kozár T, et al. 2000. A new combined computational and NMR-spectroscopical strategy for the identification of additional conformational constraints of the bound ligand in an aprotic solvent. ChemBioChem. 1:181–195.

Siebert H-C, André S, Lu S-Y, Frank M, Kaltner H, van Kuik JA, Korchagina EY, Bovin NV, Tajkhorshid E, Kaptein R, et al. 2003. Unique conformer selection of human growth-regulatory lectin galectin-1 for ganglioside GM_1 versus bacterial toxins. Biochemistry. 42: 14762–14773.

Siebert H-C, Born K, André S, Frank M, Kaltner H, von der Lieth C-W, Heck AJR, Jiménez-Barbero J, Kopitz J, Gabius H-J. 2006. Carbohydrate chain of ganglioside GM_1 as a ligand: identification of the binding strategies of three 15 mer peptides and their divergence from the binding modes of growth-regulatory galectin-1 and cholera toxin. Chem Eur J. 12:388–402.

Siebert H-C, Gilleron M, Kaltner H, von der Lieth C-W, Kozár T, Bovin NV, Korchagina EY, Vliegenthart JFG, Gabius H-J. 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.

Somers WS, Tang J, Shaw GD, Camphausen RT. 2000. Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to sLe^x and PSGL-1. Cell. 103:467–479.

Sörme P, Arnoux P, Kahl-Knutsson B, Leffler H, Rini JM, Nilsson UJ. 2005. Structural and thermodynamic studies on cation-π interactions in lectin–ligand complexes: high-affinity galectin-3 inhibitors through fine-tuning of an arginine–arene interaction. J Am Chem Soc. 127:1737–1743.

Sparrow CP, Leffler H, Barondes SH. 1987. Multiple soluble β-galactoside-binding lectins from human lung. J Biol Chem. 262:7383–7390.

Sturm A, Lensch M, André S, Kaltner H, Wiedenmann B, Rosewicz S, Dignass AU, Gabius H-J. 2004. Human galectin-2: novel inducer of T cell apoptosis with distinct profile of caspase activation. J Immunol. 173: 3825–3837.

Thomas JH. 1993. Thinking about genetic redundancy. Trends Genet. 9: 395–399.

Unverzagt C, André S, Seifert J, Kojima S, Fink C, Srikrishna G, Freeze H, Kayser K, Gabius H-J. 2002. Structure-activity profiles of complex biantennary glycans with core fucosylation and with/without additional α2,3/α2,6-sialylation: synthesis of neoglycoproteins and their properties in lectin assays, cell binding, and organ uptake. J Med Chem. 45: 478–491.

van Liempt E, Imberty A, Bank CMC, van Vliet SJ, van Kooyk Y, Geijtenbeek TBH, van Die I. 2004. Molecular basis of the differences in binding properties of the highly related C-type lectins DC-SIGN and L-SIGN to $Lewis^X$ trisaccharide and *Schistosoma mansoni* egg antigens. J Biol Chem. 279:33161–33167.

Varela PF, Solís D, Díaz-Mauriño T, Kaltner H, Gabius H-J, Romero A. 1999. The 2.15 Å crystal structure of CG-16, the developmentally regulated homodimeric chicken galectin. J Mol Biol. 294:537–549.

Villalobo A, Nogales-González A, Gabius H-J. 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.

Wallace AC, Laskowski RA, Thornton JM. 1995. Ligplot—a program to generate schematic diagrams of protein ligand interactions. Protein Eng. 8: 127–134.

Walser PJ, Haebel PW, Künzler M, Sargent D, Kües U, Aebi M, Ban N. 2004. Structure and functional analysis of the fungal galectin CGL2. Structure. 12:689–702.

Wu AM, Singh T, Wu JH, Lensch M, André S, Gabius H-J. 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.

Wu AM, Song SC, Chang SC, Wu JH, Chang KSS, Kabat EA. 1997. Further characterization of the binding properties of a GalNAc specific lectin

from *Codium fragile* subspecies tomentosoids. Glycobiology. 7: 1061–1066.

Wu JH, Song SC, Chen YY, Tsai MC, Kabat EA, Wu AM. 1998. Multi-antennary Galβ1-4GlcNAc and Galβ1-3GalNAc clusters as important ligands for a lectin isolated from the sponge *Geodia cydonium*. FEBS Lett. 427:134–138.

Wu AM, Wu JH, Liu J-H, Singh T, André S, Kaltner H, Gabius H-J. 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.

Wu AM, Wu JH, Tsai M-S, Kaltner H, Gabius H-J. 2001. Carbohydrate specificity of a galectin from chicken liver (CG-16). Biochem J. 358: 529–538.

Wu AM, Wu JH, Tsai M-S, Liu J-H, André S, Wasano K, Kaltner H, Gabius H-J. 2002. Fine specificity of domain-I of recombinant tandem-repeat-type galectin-4 from rat gastrointestinal tract (G4-N). Biochem J. 367:653–664.

Yan Z-Y, Bush A. 1990. Molecular dynamics simulations and the conformational mobiliy of blood group oligosaccharides. Biopolymers. 29: 799–811.