Different roles of the heterodimer architecture of galectin-4 in selective recognition of oligosaccharides and lipopolysaccharides having ABH antigens

Received for publication, March 15, 2024, and in revised form, June 18, 2024 Published, Papers in Press, July 15, 2024,
https://doi.org/10.1016/j.jbc.2024.107577

Jon I. Quintana[1], Mora Massaro[2,3], Alejandro J. Cagnoni[2,3], Reyes Nuñez-Franco[1], Sandra Delgado[1], Gonzalo Jiménez-Osés[1,4], Karina V. Mariño[3], Gabriel A. Rabinovich[2,5], Jesús Jiménez-Barbero[1,4,6,7,*], and Ana Ardá[1,4,*]

From the [1]CIC bioGUNE, Bizkaia Technology Park, Derio, Bizkaia, Spain; [2]Laboratorio de Glicomedicina, Instituto de Biología y Medicina Experimental (IBYME), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina; [3]Laboratorio de Glicómica Funcional y Molecular, Instituto de Biología y Medicina Experimental (IBYME), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina; [4]Ikerbasque, Basque Foundation for Science, Bilbao, Bizkaia, Spain; [5]Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Buenos Aires, Argentina; [6]Department of Organic Chemistry II, Faculty of Science and Technology, University of the Basque Country, Leioa, Spain; [7]Centro de investigación Biomédica En Red de Enfermedades Respiratorias, Madrid, Spain

Reviewed by members of the JBC Editorial Board. Edited by Robert Haltiwanger

The dimeric architecture of tandem-repeat type galectins, such as galectin-4 (Gal-4), modulates their biological activities, although the underlying molecular mechanisms have remained elusive. Emerging evidence show that tandem-repeat galectins play an important role in innate immunity by recognizing carbohydrate antigens present on the surface of certain pathogens, which very often mimic the structures of the human self-glycan antigens. Herein, we have analyzed the binding preferences of the C-domain of Gal-4 (Gal-4C) toward the ABH-carbohydrate histo-blood antigens with different core presentations and their recognition features have been rationalized by using a combined experimental approach including NMR, solid-phase and hemagglutination assays, and molecular modeling. The data show that Gal-4C prefers A over B antigens (two-fold in affinity), contrary to the N-domain (Gal-4N), although both domains share the same preference for the type-6 presentations. The behavior of the full-length Gal-4 (Gal-4FL) tandem-repeat form has been additionally scrutinized. Isothermal titration calorimetry and NMR data demonstrate that both domains within full-length Gal-4 bind to the histo-blood antigens independently of each other, with no communication between them. In this context, the heterodimeric architecture does not play any major role, apart from the complementary A and B antigen binding preferences. However, upon binding to a bacterial lipopolysaccharide containing a multivalent version of an H-antigen mimetic as O-antigen, the significance of the galectin architecture was revealed. Indeed, our data point to the linker peptide domain and the F-face of the C-domain as key elements that provide Gal-4 with the ability to cross-link multivalent ligands, beyond the glycan binding capacity of the dimer.

Galectins are a family of evolutionary conserved glycan-binding proteins that participate in a myriad of biological events, mostly related to development, angiogenesis, and immunity (1–3). Twelve members of this family have been identified so far in humans, which share the ability to bind to ubiquitous β-galactosides through a conserved carbohydrate recognition domain (CRD). Galectins are structurally classified into three subfamilies depending on the organization of the CRDs. Except for Galectin-3 (Gal-3), the only member of the chimera family, the other galectins possess two CRDs that can be either identical, as in prototype (monomers can dimerize under certain conditions) or different, as in tandem-repeat (heterodimers). This dimeric nature is known to strongly impact galectin-mediated biological effects (4–6). In prototype galectins (such as Gal-1, Gal-2, and Gal-7), the dimers are associated through noncovalent forces and thus, their stability is related to the respective dimerization equilibrium constants. For tandem-repeat galectins (such as Gal-4, Gal-8, Gal-9, and Gal-12), on the contrary, the two CRDs are covalently linked through short peptides of different lengths and sequences.

It has been proposed that the covalent dimer architecture of tandem-repeat galectins endows them with the ability to cross-link multivalent ligands (7). However, the molecular mechanisms behind these key cross-linking processes remain elusive. In particular, the role of the linker domain is ill-defined. It remains unclear, for instance whether or not this, unstructured and flexible peptide linker also allows for an interdomain communication, as described for the prototype dimeric galectin Gal-1 (8, 9). Indeed, Gal-1 binds Lac and LacNAc with a negative cooperativity effect through an allosteric communication mechanism across the noncovalent dimerization interface, in such a way that the binding to one domain weakens the binding to the second domain by *ca.* 15-fold. Moreover, it is also uncertain whether the linker domains of tandem-repeat galectins contribute to the recognition phenomena by involving clustering and/or aggregation.

* For correspondence: Jesús Jiménez-Barbero, jjbarbero@cicbiogune.es; Ana Ardá, aarda@cicbiogune.es.

Galectin-4 binding to ABH antigen containing molecules

The tandem-repeat galectins Gal-4, Gal-8, and Gal-9, as other galectins (10, 11), play a direct role in pathogen recognition and immune response. They have been shown to bind and kill bacteria that produce ABH-antigen-like glycan structures on their surface in the form of lipid-linked polysaccharides (*e.g.*, LPS) (12, 13). In this way, these galectins would fill a gap in our adaptive immune system (14), unmasking pathogens that use a molecular mimicry strategy, by camouflaging themselves with host-like carbohydrates. More recently (15), it has been shown that, not only extracellularly but also at the cytosol, Gal-4 is able to engage *Salmonella enterica* bacteria in a glycan-dependent manner, modulating the host immune response. As most galectins (16, 17), Gal-4 has also been correlated with cancer development, although no consensus on its role has been established so far (18–21). Its elevated levels in certain types of cancers (22, 23) led to propose that the tumor cells could directly interact with the red blood cells (RBCs) *via* Gal-4. The interaction, presumably *via* the ABH blood group antigens (BGAs) on the RBCs, could be a mechanism for the tumor cells dissemination and metastasis (24).

The histo-blood ABH-antigens are carbohydrate determinants defined by the trisaccharides Galα1-3(Fucα1–2)Galβ and GalNAcα1-3(Fucα1-2)Galβ (B and A, respectively) and the disaccharide Fucα1-2Galβ (H) (Table 1), which are widely distributed on the surface of RBCs, although not exclusively. They are, in fact, also expressed in a wide variety of human tissues and are present on most epithelial and endothelial cells. They are also displayed on other blood cells such as T-cells, B-cells, and platelets, and as soluble forms on the secretions of the so-called "secretor" individuals (25). Although the most frequent presentation of ABH antigens on erythrocytes is the so-called type-2 core structure (Galβ1–4GlcNAcβ), other antigen presentations indeed exist (Table 1), depending on the particular tissues and body fluids, which are known to influence their antigenicity (26, 27) and very likely, their lectin binding (13, 28).

Gal-4 has been shown to bind the A- and B-BGAs with higher affinity than lactose (Galβ1–4Glc) and N-acetyl-lactosamine (Galβ1–4GlcNAc, LacNAc) [(29), Cagnoni, Massaro *et al.* (69) co-submitted], which are weak and very weak binders, respectively (30, 31). Both domains of Gal-4 (Gal-4N and Gal-4C) contribute to the ABH antigens binding events. However, while some studies have proposed a substantially stronger interaction for the C- than the N-domain (32), others have reported rather small differences in affinity (33). Interestingly, even though Gal-4N showed substantial binding to blood group B-expressing *Escherichia coli* bacteria, this domain had no effect on bacteria viability, while Gal-4C was shown to both bind and kill the bacteria (12).

We have recently presented a systematic structural analysis on the interaction in solution between Gal-4N and the A and B group antigen oligosaccharides with diverse core presentations (28). Our results showed that Gal-4N always prefers (ca. twofold in affinity) the B-antigens over the A-antigens, and the type-6 (Galβ1-4Glc) presentation over the type-1 (Galβ1-3GlcNAc) and type-2 (Galβ1-4GlcNAc) analogues, which behaved similarly.

However, given the tandem-repeat dimer nature of human Gal-4, the study of the full-length Gal-4 (Gal-4FL) is essential to understand and eventually modulate its recognition properties. Thus, we herein present a systematic analysis of the interaction of Gal-4FL and its Gal-4C with different ligands that present the ABH-antigens. These include oligosaccharide antigens with different core presentations and a LPS in which the O-chain repeating is an H-antigen mimetic. Moreover, we have paid special attention to the role of the linker domain in the molecular recognition processes. Fittingly, the linker resulted essential for the binding to the multivalent presenting antigen ligand LPS, but not for oligosaccharide binding.

Table 1
ABH histo-blood group antigens and core disaccharide structures

Antigen	Terminal structure	Type	Core structure
A	GalNAcα1–3 (Fucα1–2)Galβ	Type-1	Galβ1–3GlcNAcβ
B	Galα1–3 (Fucα1–2)Galβ	Type-2	Galβ1–4GlcNAcβ
H	Fucα1–2Galβ	Type-3	Galβ1–3GalNAcα
		Type-4	Galβ1–3GalNAcβ
		Type-5	Galβ1–3Galβ
		Type-6	Galβ1–4Glcβ

Results

The single Gal-4C domain: binding to the A- and B-type histo-blood oligosaccharide antigens

The binding of Gal-4C to A- and B-antigen glycans with different core presentations was studied by using ligand- and receptor-based NMR strategies (34, 35), assisted by molecular modeling protocols.

A quantitative estimation of the binding affinities was obtained through $^{1}H,^{15}N$-heteronuclear single quantum coherence (HSQC) titration experiments, using ^{15}N-Gal-4C. The obtained K_D values (Table 2, Figs. S1, and S2) show that Gal-4C prefers the A over the B-antigens by *ca.* two-fold in binding affinity, irrespective of the core disaccharide structure. Regarding the core structure, by comparing the same antigen (A or B), for the tetrasaccharides, the type-6 and type-1 were the best binders, while the interaction with the type-2 analogues was twice weaker. The pentasaccharides displaying the type-4 core structure were systematically the worst binders.

For a structural rationalization of these affinity data, further NMR analysis and experiments were performed, providing binding information with atomic resolution both from the lectin's and the ligand's points of view (36).

First, the $^{1}H,^{15}N$-HSQC spectrum of Gal-4C was assigned by acquiring the standard set of 3D-NMR experiments (37) on

Table 2
K_D values as obtained from $^{1}H,^{15}N$-HSQC-NMR titration experiments, for the interaction of Gal-4C with AB antigen oligosaccharides

K_D (μM)	Type-1 (t)	Type-2 (t)	Type-6 (t)	Type-4 (p)
A-antigen	86 ± 8	170 ± 9	63 ± 8	185 ± 15
B-antigen	190 ± 10	300 ± 28	170 ± 10	458 ± 31

Tetrasaccharides (t) were used for types-1, 2, and 6, while pentasaccharides (p) were used for the type-4.

a double-labeled ^{13}C,^{15}N sample of Gal-4C and following well-established protocols, as implemented in the *ccpnmr* analysis software (https://ccpn.ac.uk/software/analysisassign/) (38), to successfully achieve the assignment of 86% of the NH backbone correlations (Fig. S3).

Then, the comparison of the chemical shift perturbations (CSPs) produced on the protein signals (see Methods) by structurally related ligands of increasing complexity was essential to deduce the interacting lectin residues for each glycan analogue. For instance, 2′FucLac (H-6 trisaccharide) basically produced the same CSPs on Gal-4C as lactose (Fig. 1*A*). This fact shows, as expected (39), that the fucose (Fuc) moiety is not involved in contacts with the lectin. On the contrary, the presence of a terminal galactose (Gal) residue α1-3 linked to lactose (Galα1–3Galβ1–4Glc), which defines the B-antigens, produced important differences with respect to lactose binding (Fig. 1*B*). New amino acids at strand S3 (residues 220–222 and 226–227) were perturbed, revealing the αGal binding site (Fig. 1*D*). Additional CSPs at residues within the canonical lactose-binding site suggested a somehow subtle reorganization of intermolecular interactions involving the lactose moiety. The comparison between the A and B types yielded a conserved trend in the CSP profile (Figs. S4, S5, and S6) for the different core types. The A-antigen always produced larger perturbations than the corresponding B-antigen at amino acids K219, S220, F221, A222, and Q313 (Figs. 1*C* and S6), which gather along the S2-S3 strands and define a smooth hollow (Fig. 1*E*). Remarkably, the observed CSP of the side chain signal (NHε) of the key W256, which provides the key CH-π stacking interaction (40) with the central Galβ residue, was also different for the interactions with the A and B glycans. Finally, the distinct core types also yielded differences in the CSP plots (Fig. S7). For the types −1, −2, and −6, CSP were very similar, with the main differences just located at strand S5. In contrast, the type-4 antigens displayed significant CSP differences, especially at strands S4 and S5 (Fig. S7). This experimental evidence was used later for generating the 3D structures of the complexes (see below in Molecular dynamics (MD)).

To get information from the antigens' perspective, ^1H-saturation transfer difference (STD)-NMR spectra were acquired to get detailed information on the ligands' binding epitope. From a technical point of view, the on-resonance irradiation at the aromatic region produced stronger STD responses than that at the aliphatic one (Figs. 2 and S12–S14), highlighting the role of the aromatic residues in the recognition event. In any case, the STD-NMR spectra were very similar for all the ligands in terms of the observed intensities and profiles. For none of the ligands, STD effect was observed for the Fuc protons, neither for protons at the reducing-end residues, except for those belonging to the Ac group, whenever present. Basically, STD-NMR intensities were observed for the protons of the central Galβ residue and H1 and H2 of the terminal Gal/GalNAc residues (Fig. 2*A*). This set of intensities is essentially identical to that previously reported for the interaction with Gal-3 CRD (41). Remarkably, the comparison of the STD-NMR intensities at the acetyl region was highly informative (Fig. 2*B*). While the N-acetyl group of the terminal αGalNAc (in the A-analogues) always displayed a strong STD-NMR signal (both under aliphatic and aromatic irradiation), the N-acetyl group of the reducing end varied notably: for the type-1 (Galβ1-3GlcNAc) and type-4 (Galβ1-3GalNAc) antigens, a strong STD-NMR signal was observed upon aromatic irradiation, which nearly disappeared in the aliphatic irradiation spectrum. This difference between aromatic and aliphatic irradiation was less evident for the type-2 (Galβ1-4GlcNAc) antigens. These results anticipate that, for the Galβ1-3 linked antigens, the N-acetyl group at the reducing-end is closer to an aromatic group than that in the Galβ1-4 linked antigens.

Interestingly, the binding event for the interaction of the A-antigens with Gal-4C was in slow exchange in the chemical shift time scale between the sugar free and bound forms. This fact was evident since chemical exchange cross-peaks (same sign as the diagonal) were observed in the rotating frame Overhauser effect spectroscopy (ROESY) experiment recorded for the corresponding A-antigen in the presence of Gal-4C (Fig. S15). In contrast, this slow exchange regime was not observed for any of the B-antigens. This fact nicely correlates with the higher affinities measured (and presumably lower k_{off} and thus lower k_{ex}, thus achieving the $K_{ex} << \Delta\omega$ condition) for the A, but not for the B-antigens. As it has been previously shown (28, 41–43), this kinetic regime in binding processes can be highly informative, revealing the chemical shifts of the ligand in the bound form, which is not always easy to detect. In this case, since H4, H5, and H6 of the central β-Gal were in slow exchange, a rather large chemical shift difference (Δδ 1.5–2.5 ppm) was measured between the free and bound states, with upfield shifting for the bound form. This observation is in full agreement with the CH-π stacking of this β-Gal moiety with the conserved tryptophan at the galectins binding site (1, 40, 44). The profile of the observed chemical shift differences between the free and bound states for these protons was nearly identical for all A-type ligands, suggesting that the nature of this stacking interaction is fairly similar in all cases (Fig. S15). Nevertheless, some differences were found at the reducing end residues. While for the type-2 and -4 antigens, the weaker ligands, none of the protons at the reducing end was in slow chemical exchange, H4 in the type-1 and H3 and H5 in the type-6 of the corresponding reducing end residues displayed slow exchange. These observations strongly suggest that GlcNAc H4 faces the protein surface in the type-1 analogues, while H3 and H5 of the Glc moiety are facing the lectin in the type-6.

Thus, the NMR data show a clear trend for the binding preferences of Gal-4C for the oligosaccharide antigens. A antigens are always the best binders compared to the B and H analogues. Regarding the core structures, types-6, -1, and -2 bind similarly, with type-6 being always preferred closely followed by type-1, then type-2, and finally −4. Quantitatively (Table 2), the differences in affinity are however relatively small.

Thus, competitive solid-phase assays were carried out to further confirm that the type 6 blood group A (A-6) antigen is

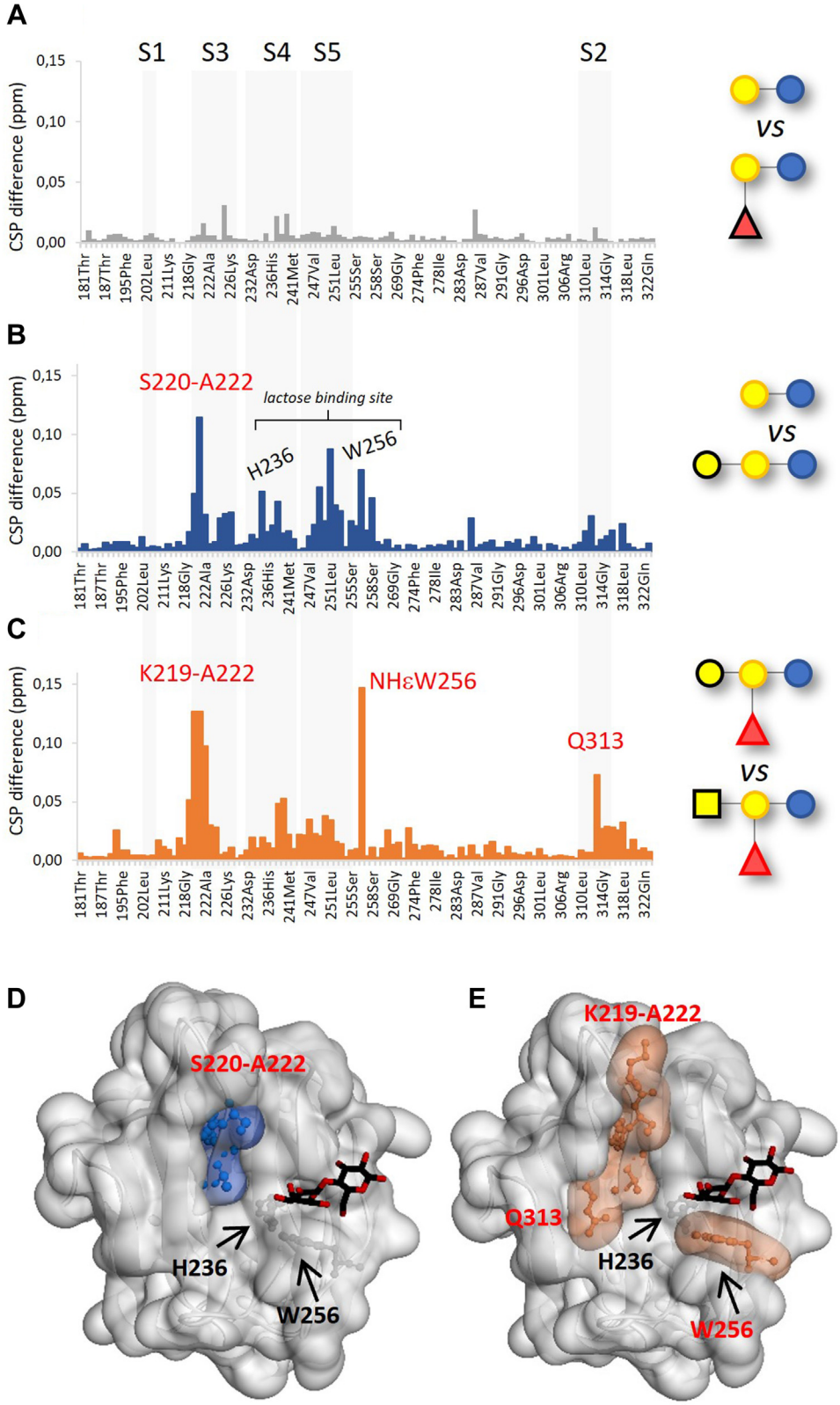

Figure 1. ^{1}H,^{15}N-HSQC chemical shift perturbation analysis of Gal-4C with different glycans related to A- and B-antigens. Plots for the average (^{1}H, ^{15}N) chemical shift difference of Gal-4C bound to (*A*) lactose and H-6 antigen, showing very little differences. *B*, lactose and the trisaccharide Galα1–3Galβ1–4Glc. Small differences imply the lactose-binding site, but the most important differences are found at residues S220-A222. *C*, B-6 and A-6 antigens. Important differences imply K219-A22, NHε W259, and Q313. Below (*D* and *E*), 3D structures for the complex of Gal-4C with lactose (in *black sticks*) (pdb 4YM3) highlighting key W256 and H236 residues: (*D*) residues perturbed according to plot in (*B*) are highlighted as *blue* surface; (*E*) residues perturbed according to plot in (*C*) are highlighted as *orange* surface. Gal-4C, C-domain of Gal-4; HSQC, heteronuclear single quantum coherence.

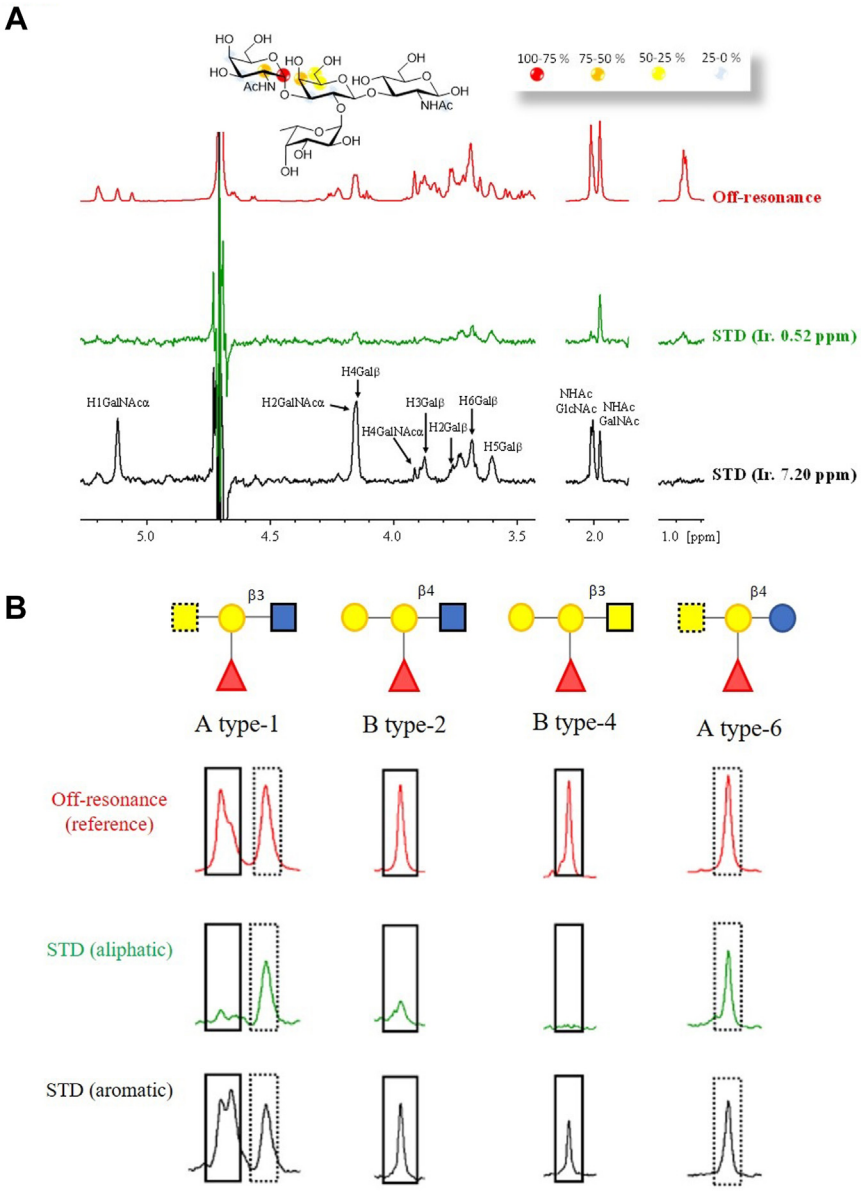

Figure 2. ^{1}H-STD-NMR results. For both (*A* and *B*): top (in *red*) the off-resonance (reference) spectrum; middle (in *green*) ^{1}H-STD spectrum with aliphatic on-resonance irradiation; bottom (in *black*) ^{1}H-STD spectrum with aromatic on-resonance irradiation. *A*, spectra for the interaction of Gal-4C with A-1 tetrasaccharide: (*B*) summary of the STD of the acetyl groups for the interaction of different *A* and *B* antigens with Gal-4C. Acetyls at the terminal positions are highlighted with *dotted line*, while those at the reducing end are highlighted with a *solid line*. Gal-4C, C-domain of Gal-4; STD, saturation transfer difference.

the preferential Gal-4C ligand. In particular, competition assays were performed to test whether these small binding differences observed in solution are translated into a competing binding setting. The capacity of A-1, A-2, and A-6 to neutralize Gal-4C interaction with immobilized asialofetuin (ASF) was evaluated through competitive solid-phase assays (45), using lactose as a reference (Fig. 3 and Table S2). As expected, all evaluated A-glycans were better binders than lactose (IC_{50} = 7.0 mM). Moreover, A-6 (IC_{50} = 0.8 mM) resulted to be the best inhibitor of Gal-4C, followed by A-1 (IC_{50} = 1.2 mM), and finally A-2 (IC_{50} = 1.8 mM). These results confirm the glycan-binding preferences of Gal-4C toward type-6 presentation of blood group A antigens, deduced by NMR.

BGA A-6 inhibits Gal-4C-induced hemagglutinating activity

The glycan-binding capacities of Gal-4C for the histo-blood oligosaccharide antigens were further explored using a cell-based assay. First, the hemagglutinating activity of Gal-4C with type A and type 0 erythrocytes, where BGAs are mostly presented on a type-2 core (46), was evaluated (Fig. 4*A*). Gal-4C was able to agglutinate the trypsin-fixed A-type erythrocytes at concentrations of 25 μg/ml or higher. In contrast, for 0 type erythrocytes, no hemagglutinating activity was observed for Gal-4C even at concentrations of 200 μg/ml. These results highlight the preferential recognition of Gal-4C toward blood group A antigens. Next, we sought to evaluate the inhibitory capacity of the A-6 oligosaccharide using a hemagglutination inhibition assay. Notably, the A-6 tetrasaccharide successfully

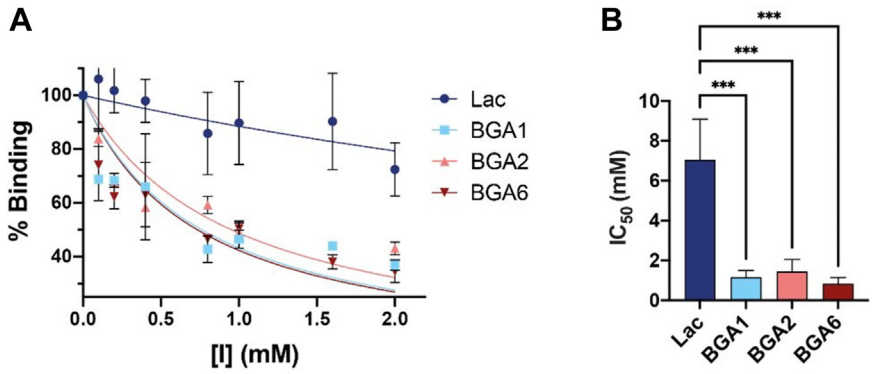

Figure 3. Gal-4C interactions with ASF is inhibited by blood group A antigens. *A*, binding curves for Gal-4C-ASF interactions in competition with Lactose (Lac). *B*, IC_{50} values for the interaction of human Gal-4C and Lac, A1, A2, and A6. One-way ANOVA followed by Tukey's posttest was used. Data are presented as means ± SEM and are from a representative of three independent experiments. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. ASF, asialofetuin; Gal-4C, C-domain of Gal-4.

inhibited Gal-4C-induced hemagglutinating activity using type A erythrocytes at concentrations higher than 0.6 mM (Fig. 4*B*).

Molecular modeling

Molecular dynamics (MD) simulations were performed to obtain 3D models of the different lectin-antigen complexes. The X-ray crystallographic structure (Protein Data Bank (PDB) code 4YM0 (39), Gal-4C bound to lacto-N-tetraose) was used as starting geometry for the lectin, while the global minimum conformations (41) of the ligands were docked in the lectin-binding site by superimposing the central βGal residue of the different antigens with the corresponding βGal moiety of the reference structure.

The analysis of the complexes (always assessing their structural features through the experimental NMR data) with the A-antigens showed that the terminal αGalNAc residue establishes key interactions with the protein, which are absent or much less persistent in the complexes with the B-antigens, having a terminal αGal residue instead (Fig. 5): (i) the methyl group of the NHAc perfectly fits in the cavity formed by amino acids S220-A222 at strand S3, establishing van der Waals contacts with these residues, (ii) the side chain of Q313 hydrogen bonds the CO of the NHAc group, and (iii) the OH-6 establishes H-bonding with NHε of W256. All these ligand and protein structural features are in perfect agreement with the STD and CSP data and account for the binding preference of the Gal-4C domain for the A over the B antigens. Interestingly, this preference is opposite to that observed for the N domain (28), which prefers the B over the A antigens. This switch is likely due to the presence of the larger F47 residue in the N-domain (instead of the smaller A222 at the equivalent

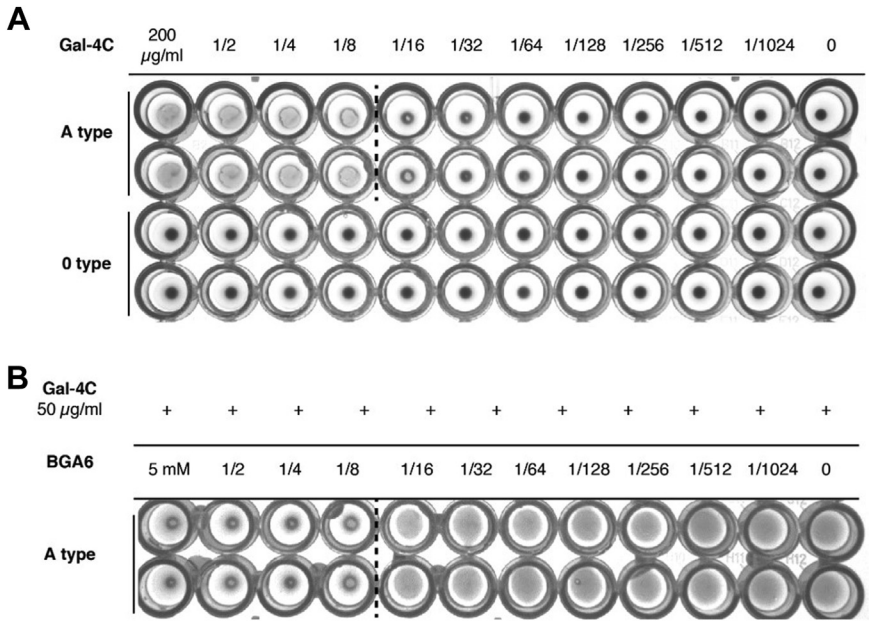

Figure 4. Hemagglutination assay and hemagglutination inhibition assay of Gal-4C. *A*, hemagglutinating activity of Gal-4C. Serial dilutions of recombinant Gal-4C were tested on trypsin-treated A and 0 type erythrocytes. *Dotted lines* indicate minimal agglutinating concentration. PBS buffer was used as negative control. *B*, inhibition of Gal-4C hemagglutinating activity. BGA A-6 was tested as Gal-4C competitive ligand in serial dilutions. *A* and *B*, data from a representative of two independent experiments. BGA, blood group antigen; Gal-4C, C-domain of Gal-4.

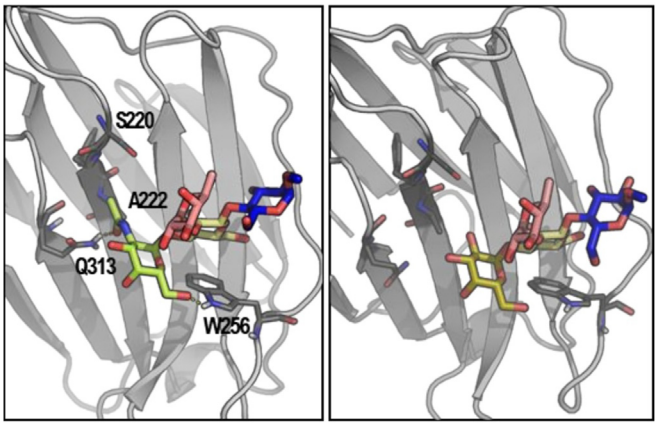

Figure 5. Complexes of Gal-4C with A-type-6 (left) and B-type-6 (right) tetrasaccharide antigens according to MD simulations. Selected snapshots highlighting key differences are shown. Side chains of residues Q313 and W256 establish persistent H-bonding (*yellow*) with the αGalNAc residue, while S220-A222 define a cavity hosting the Me group of the NHAc moiety. Color codes for monosaccharides: Glc *blue*, βGal *yellow*, Fuc *light red*, αGal *olive*, and αGalNAc *lime*. Gal-4C, C-domain of Gal-4; MD, molecular dynamics.

position in C-domain), which causes steric clash with the bulky N-Ac group of the A-antigens. This destabilizing interaction causes the decreased binding strength for the A-antigens respect to the B-antigens, which lack the N-Ac group (28).

With respect to the core disaccharide structure, MD simulations showed that, for the β1-4 linked antigens (types −2 and −6), positions two and three of the reducing end residue (GlcNAc or Glc, respectively) face the lectin (Fig. S16). In such a situation, for type-6 antigens, both positions (OH-2 and OH-3) display hydrogen-bonding interactions with E259, while for type-2, with an N-Acetyl group at position 2, one hydrogen bond is lost, while that involving OH-3 is weakened (less persistent along the MD simulation).

For the β1-3 linked antigens (type −1 and −4), positions four and six of the reducing end residues (GlcNAc or GalNAc, respectively) are now facing the lectin surface (Fig. S16). In the case of type-1 antigens, OH-4 has a similar orientation to that of OH-3 in the β1-4 linked analogues, resulting in similar interactions with E259 (Fig. S16), while OH-6 establishes hydrogen bonding with the K261 side chain. For the type-4 antigens, while the interaction of OH-6 with K261 is preserved, the axial orientation of the hydroxyl group at position four of the reducing end GalNAc precludes any interaction with the lectin. These results perfectly fit with the above described NMR experimental data.

Binding of the Gal-4FL to the A- and B-type histo-blood oligosaccharide antigens: independent behavior of individual domains

The behavior of Gal-4FL was then analyzed. First, the $^{1}H,^{15}N$-HSQC spectrum of Gal-4FL was compared to those recorded for the isolated domains (Gal-4N (28) and Gal-4C) (Fig. 6, *A* and *B*). The superimposition of the spectrum of Gal-4FL with the sum of the spectra of the isolated domains (Gal-4N and Gal-4C) (Fig. S17) evidenced great similarities, suggesting that there are no major changes in the secondary structure of each of the domains when passing from the isolated domains to the full-length form. In fact, 51% of residues belonging to the N-domain and 43% belonging to the C-domain could be directly assigned on Gal-4FL based on the chemical shift similarity with the isolated domains (Fig. 6 and Table S3). The rest of the cross-peaks could not be assigned in Gal-4FL due to severe NMR signal crowding and/or large chemical shift changes. Unfortunately, the low solubility of Gal-4FL precluded the full assignment of the $^{1}H,^{15}N$-HSQC spectrum through triple resonance experiments. In any case, the number of identified residues was large enough to carry out a CSP analysis, as shown in Figure 6*A*, and to compare the data obtained for the single domains *versus* the full-length lectin. The most perturbed residues are mapped as red and pale orange surface into the corresponding 3D structures (Fig. 6, *B* and *C*). Residues not identified in Gal-4FL, but assigned in the isolated domains are highlighted in yellow. For the N-domain (Fig. 6*B*), CSP were in general rather low, while those peaks that were not identified in the Gal-4FL spectrum were sparsely distributed throughout the CRD. However, for the C-domain (Fig. 6*C*), important perturbations could be identified for several residues at the F-face, which cluster with residues that could not be assigned by comparison with the single domain. This fact suggests that several amino acids in this region have a different chemical environment when moving from the isolated domain to the full-length form, which could be a consequence of protein-protein interactions between the conformationally poorly defined and flexible linker and this hydrophobic region of the F-face.

With respect to the saccharide binding events, the isolated domains (Gal-4N (28) and Gal-4C) compete for the same BGA ligands, as discussed above. Thus, to discriminate the binding behavior of each domain individually, but within the full-length form, single point mutants were generated, with abrogate binding at either the N- or the C-domain. In particular, the mutant H63R (Gal-4FL-H63R), with the N-domain inactive for glycan binding, can only bind glycans through the C-domain, while the H236R mutant (Gal-4FL-H236R) can only interact through the N-domain. Indeed, the analysis of $^{1}H,^{15}N$-HSQC experiments confirmed that Gal-4FL-H63R and Gal-4FL-H236R were not able to bind the A-6 tetrasaccharide through their N- and C-domains, respectively, as key residues belonging to these domains were not affected upon ligand binding (Fig. 7: T78, G81 in Gal-4FL-H63R and S250, G228 in Gal-4FL-H236R). Additionally, the observed CSP profiles for residues at the N-domain upon glycan binding for Gal-4FL, Gal-4N, and Gal-4FLH-236R were basically identical (Fig. S18). Likewise, the CSPs for residues at the C-domain in Gal-4FL, Gal-4C, and Gal-4FL-H63R upon glycan binding were also the same (Fig. S18). Thus, it is experimentally demonstrated that binding abrogation at one of the domains does not affect glycan recognition at the other domain.

To confirm these results from a quantitative perspective, additional isothermal titration calorimetry (ITC) measurements were carried out (Table 3) [Cagnoni, Massaro *et al.* (69) co-submitted] The affinity of Gal-4FL-H236R for the A-6

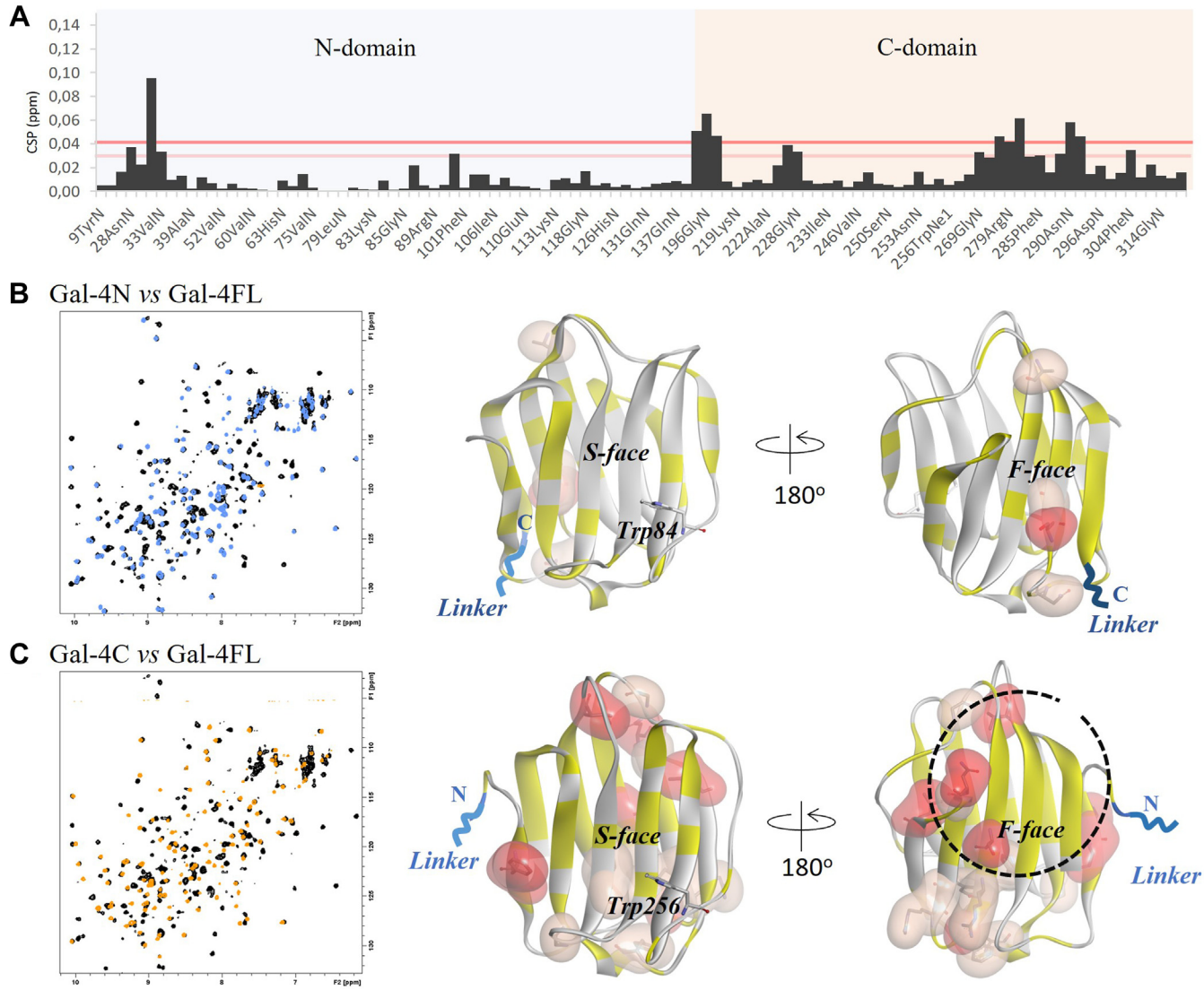

Figure 6. Comparison between Gal-4C and Gal-4N single domains *versus* full-length Gal-4. *A*, CSP analysis for identified crosspeaks based on chemical shift similarity. *B*, *left*: superimposition of the $^{1}H,^{15}N$-HSQC spectra of Gal-4FL (*black*) and Gal-4N (*blue*). *Right*, residues at the N-domain most perturbed when comparing Gal-4N and Gal-4FL according to the plot in (*A*). In *yellow*, nonidentified residues in Gal-4FL with respect to Gal-4N. *C*, left: superimposition of the $^{1}H,^{15}N$-HSQC spectra of Gal-4FL (*black*) and Gal-4C (*orange*). Right, residues at the C-domain most perturbed when comparing Gal-4C and Gal-4FL according to the plot in (*A*). In *yellow*, nonidentified residues in Gal-4FL with respect to Gal-4C. CSP, chemical shift perturbation; Gal-4FL, full-length Gal-4; Gal-4C, C-domain of Gal-4; HSQC, heteronuclear single quantum coherence.

tetrasaccharide (K_D 93 μM) was essentially the same to that obtained for Gal-4N (K_D 98 μM). Moreover, the thermodynamic profile was also basically unaltered. Likewise, Gal-4FL-H63R and Gal-4C recognized the A-6 antigen with affinities within the same order of magnitude (K_D 35 μM *versus* 26 μM), and with similar thermodynamic profiles. Thus, both domains of Gal-4FL bind the A-6 antigen, but independently of each other. Fittingly, the observed K_D for the Gal-4FL is in between those measured for the individual domains and their mutants.

Binding of Gal-4FL to a bacterial LPS composed by a H type-1 antigen analogue

LPS are composed of three structural elements: the hydrophobic Lipid A, a nonrepeating core oligosaccharide, and a polysaccharide O-antigen, which is its most chemically and structurally diverse component. The repeating unit of the O-antigen of the LPS of *E. Coli* O55 is composed of Gal, GalNAc, and L-colitose (Col, 3-deoxy-Fuc, Fig. S19). Interestingly, the structure of this O-chain is very similar to that of the H type-1 antigen (47), and thus can be regarded as its polymer mimetic version. It is well-known that LPS forms aggregates in water, resulting in very large entities that expose the O-antigen to the aqueous medium.

The interaction of the *E. Coli* O55 LPS with Gal-4 was thus studied by NMR. The protocol involved monitoring of the reduction of the signal intensities produced on the $^{1}H,^{15}N$-HSQC spectra of the different Gal-4 variants described above (single domains, FL, and mutants) upon addition of the LPS aggregates. The interaction of any protein with a very large molecular entity produces increased transverse NMR relaxation rates on the protein's NMR-active nuclei, with the

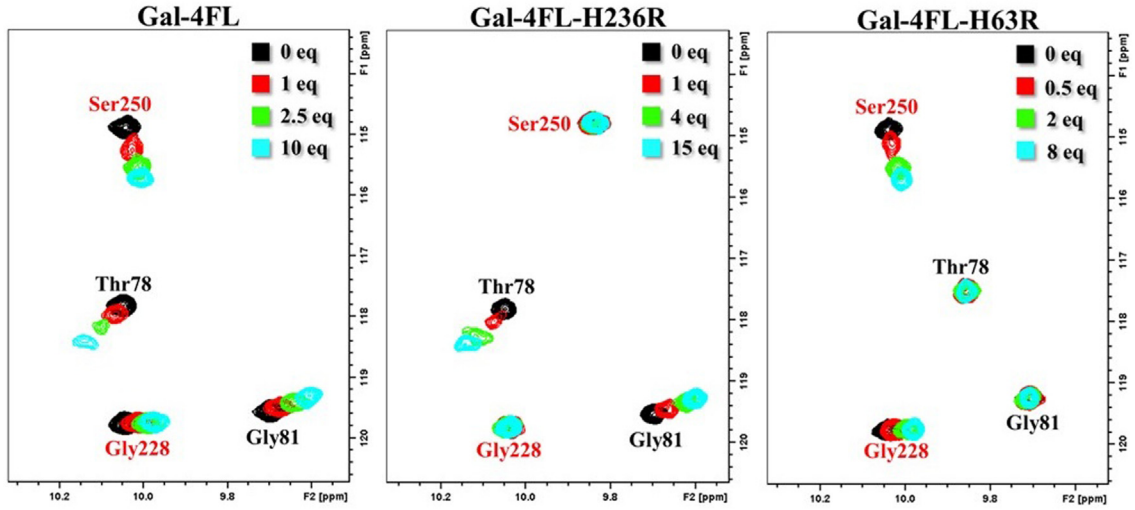

Figure 7. The behavior of Gal-4FL, Gal-4FL-H236R, and Gal-4FL-H63R and upon A-6 tetrasaccharide binding (^{1}H,^{15}N-HSQC titration data, number of ligand equivalents (eq) specified in each case). *Left*: residues at both domains are perturbed in Gal-4FL. *Middle*: residues only belonging to the N-domain (labels in *black*) are perturbed in Gal-4FL-H236R, exactly as in Gal-4FL. *Right*: residues only belonging to the C-domain (labels in *red*) are perturbed in Gal-4FL-H63R, exactly as in Gal-4FL. Gal-4FL, full-length Gal-4; HSQC, heteronuclear single quantum coherence.

concomitant decrease in their signal intensities, which may even disappear. The very large size of the generated complex and the presence of chemical exchange both contribute to the fast transverse relaxation rates.

In our case, the addition of LPS to a 1:1 mixture of isolated Gal-4 domains (Gal-4N and Gal-4C) produced a marked cross-peak intensity reduction for all residues (Fig. S20). The total intensity loss was much more pronounced for the C-domain than for the N-domain (84% *versus* 46%) (Fig. 8), which correlates well with the higher affinity of the C-domain for type core-1 structures (Galβ1-3GlcNAc) compared to the N-domain (Table S1).

For Gal-4FL, the intensity loss was basically homogeneous among the residues, irrespective of the domain and position within the 3D-domain structure (Fig. S20). Importantly, and in contrast to the observations for the isolated domains, precipitation was clearly observed in the NMR tube. It is very likely that this effect is a consequence of the exclusive ability of the full-length protein to cross-link different LPS aggregates due to its divalent glycan binding nature: in this way, networks of cross-linked species are generated (Fig. 8*C*). The two Gal-4FL-H63R and Gal-4FL-H236R mutants, which have lost the capacity to bind the sugar antigens through one of their domains, were then evaluated. Surprisingly, although they have also lost the glycan bivalent binding capacity, their behaviors were fairly similar to that of the WT form (Fig. 8). The signal intensity loss was slightly lower than for the WT protein, but again the decrease was homogeneous for all residues at both domains (despite one of the domains being inactive) (Fig. S21). Moreover, precipitation was observed in both cases.

This unexpected result for the mutants strongly suggests that these Gal-4 mutants are still able to display multivalent ligand binding capacity. It is tempting to propose that this phenomenon occurs through a lectin oligomerization phenomenon that is independent of sugar binding. Thus, as a next step, the ability of Gal-4FL to oligomerize was explored. ^{1}H-NMR DOSY spectra (48, 49) were recorded at different protein concentrations. Fittingly, the results (Fig. 9) showed that, indeed, as protein concentration increased, the diffusion coefficient became smaller. This fact unequivocally proves that species with larger hydrodynamic radius are being formed. Thus, it can be concluded that Gal-4FL displays a certain tendency to aggregate in solution. This tendency can obviously be enhanced in the presence of interactions with certain ligands, especially those showing multivalent presentation of the binding epitope.

Since the existence of this aggregation tendency is independent of sugar binding, it is very likely that the linker promotes this oligomerization, since it is present in the full-length and in the two mutants, but not in the single domains. Interestingly, the linker in Gal-4 has a unique length, unlike other tandem-repeat galectins, for which isoforms with different linker lengths exit. Gal-4 linker sequence is actually very different to that of Gal-8 and Gal-9 but shares high similarity with the nonlectin N-terminal region of Gal-3. Remarkably, this intrinsically disordered peptide has been shown to mediate aggregation of Gal-3 to form liquid-liquid phase separation (50–52). An additional structural element contributing to Gal-4FL oligomerization could be the F-face of the CRDs, as proposed for Gal-9 (53, 54). While, according to

Table 3
ITC data for Gal-4FL, binding-compromised mutants, and its individual domains binding to A-6 antigen

K_D (μM)	K_D (μM)	ΔH[a]	-TΔS[a]
Gal-4FL	71 ± 3	−7.1 ± 0.4	1.5
Gal-4FL-H236R	93 ± 4	−8.3 ± 0.5	2.8
Gal-4N	98 ± 5	−8.9 ± 0.1	3.4
Gal-4FL-H63R	35 ± 2	−7.9 ± 0.1	1.8
Gal-4C	26[b] ± 2	−8.3 ± 1	2.0

[a] kcal/mol.
[b] the discrepancy with the value reported in Table 1 is attributed to the fact that they were obtained through different techniques, as previously observed (28).

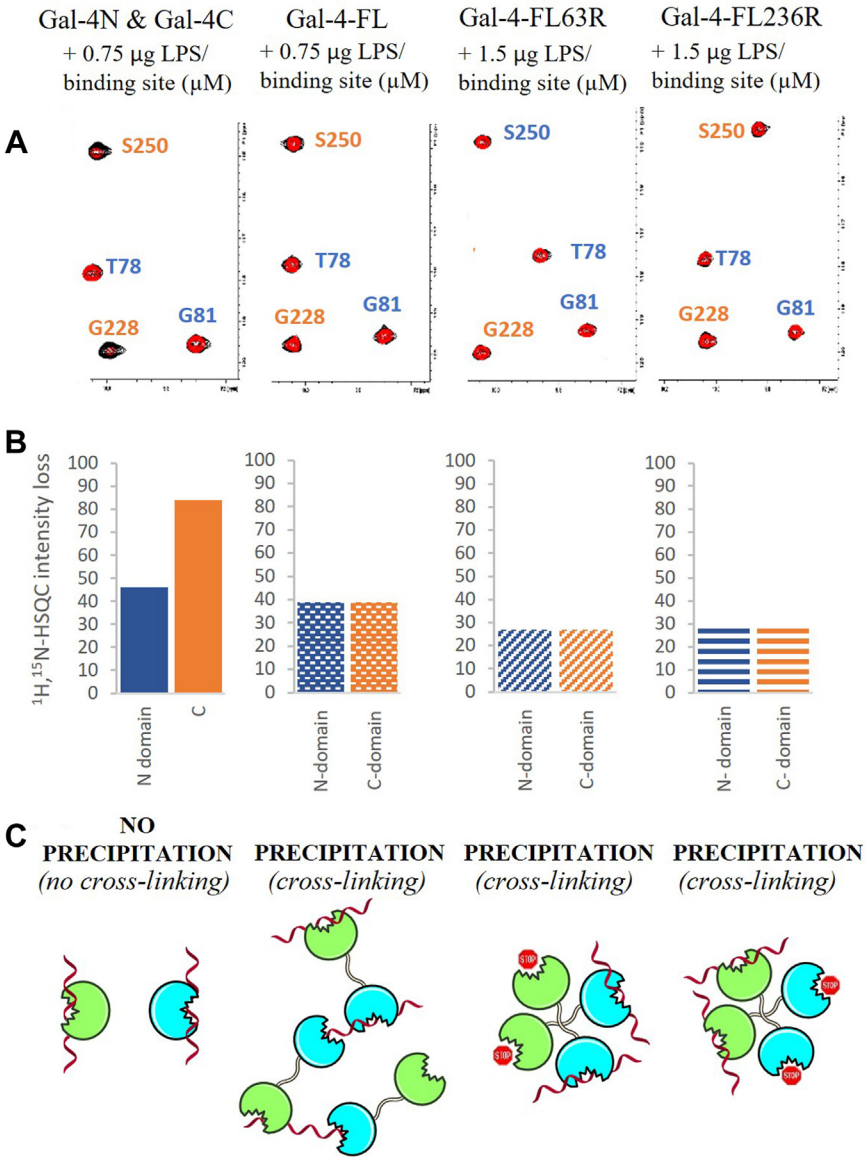

Figure 8. Interaction of different Gal-4 variants with *Escherichia coli* O55 LPS analyzed through NMR. *A*, region of the $^{1}H,^{15}N$-HSQC spectra showing crosspeaks corresponding to both the N and C-domains. In *black*, the protein/s alone. In *red*, in the presence of the LPS. *B*, quantification of the total signal intensity loss for the N and C domains upon LPS addition. *C*, cartoon models for the Gal-4/LPS complexes and the putative cross-linking networks. HSQC, heteronuclear single quantum coherence; LPS, lipopolysaccharide.

our NMR data (see above), when moving from the single domain to the full-length form, residues at the N-domain do not experience important changes in their chemical environments, several amino acids at the F-face of the C-domain are importantly affected. Interestingly, these amino acids define a hydrophobic surface region (Fig. S22) that could mediate self-association, as reported for the C-domain of Gal-9 (54).

Gal-4FL oligomerization (formation of dimers) was thus explored using AlphaFold multimer (https://github.com/google-deepmind/alphafold) (55, 56) followed by refinement with MD simulations. Interestingly, a dimer formed through the interaction of the hydrophobic F-face of the Gal-4C domains and involving a small region of the linker (Fig. 10), stays stable along MD simulations, representing a plausible model for Gal-4 dimerization. Nonetheless, it is likely that Gal-4 oligomerization is a complex process, even resulting in heterogeneous species, which deserves further investigations.

Discussion

Herein, the binding preferences of the C-domain of Gal-4 toward the blood group oligosaccharide antigens have been scrutinized and structurally rationalized by using an experimental approach, including NMR studies, competitive solid-phase assays and hemagglutination assays, as well as molecular modeling simulations. As previously shown (39), and as in the case of Gal-3 (41), the Fuc residue in ABH-antigens does not provide intermolecular stabilizing interactions with the protein, while the terminal Gal/GalNAc residues defining the B-antigens and A-antigens, respectively, are intimately

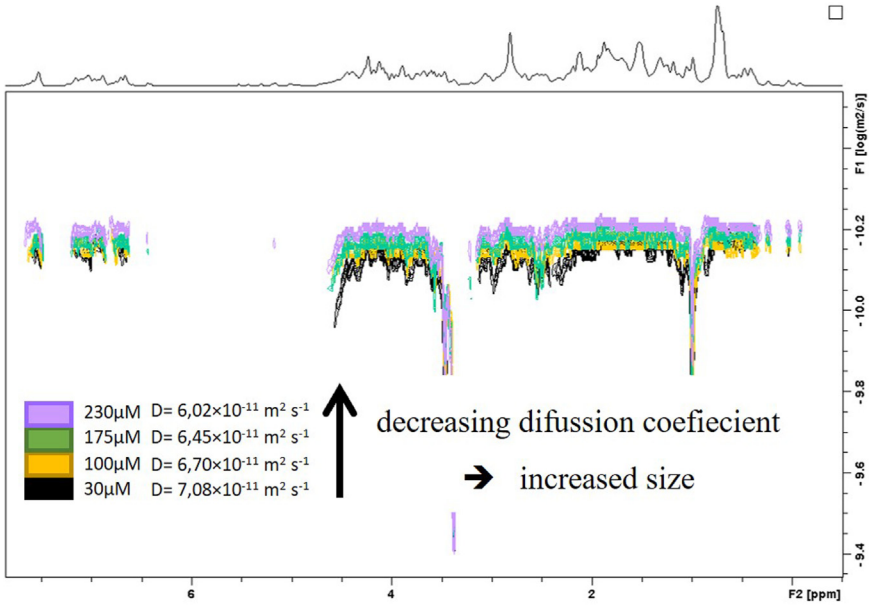

Figure 9. ^{1}H-DOSY spectra of Gal-4FL at different protein concentrations. As protein concentration increases, the diffusion coefficient decreases, highlighting the increase in size of the formed entities. Gal-4FL, full-length Gal-4.

engaged with the lectin. Gal-4C prefers A-antigens over B-antigens by *ca.* two-fold in affinity, irrespective of the core disaccharide structure. This NMR-based conclusion agrees with very recent data based on ELISA assays (57). The detailed analysis at atomic resolution accounts for specific protein-glycan interactions as the origin of these preferences: the side chains of Q313 and W256 establish H-bonds with the terminal αGalNAc residue, and the S220-A222 triad defines a cavity for the Me group of the NHAc moiety. This is the opposite trend observed for the N-domain, which always prefers B antigens. However, both domains share a preference for type-6 core structures, which for the C-domain is due to stabilizing H-bonds between OH-2 and OH-3 of the reducing Glc residue with E259. The type-2 antigen, which is the most abundant presentation of BGAs in erythrocytes is recognized twice as weak as the type-6 analogue.

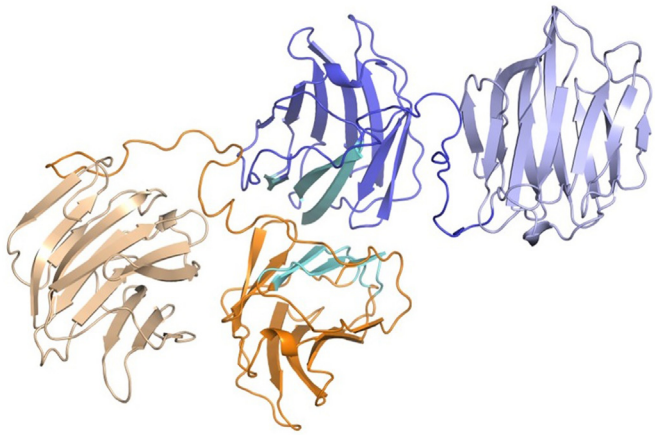

Figure 10. Molecular model for Gal-4 dimerization. The model was generated with AlphaFold multimer and refined by MD simulations (one representative snapshot shown). One Gal-4 unit is colored in different shades of *blue* (N-domain in *light blue*, and C-domain and linker in *dark blue*), and the other Gal-4 unit is colored in different shades of *orange* (N-domain in *light orange*, and C-domain and linker in *dark orange*). The b-sheets of the F-Face of the C-domain of each Gal-4 are colored in *cyan*. MD, molecular dynamics.

The behavior of the full-length form of Gal-4 was then examined. By using single mutants at either CRD binding site, which abolish glycan binding at the corresponding domain, ITC and NMR experiments agree on the fact that both domains within the full-length form bind the A type-6 tetrasaccharide in an independent manner, with no cooperativity or communication between the domains. Thus, for simple oligosaccharide binding events, the heterodimeric architecture of the lectin has no major significance.

However, the interaction with an LPS containing a multivalent presentation of an H-antigen mimetic as the O-antigen, highlighted the relevance of the heterodimer architecture of Gal-4, which goes beyond its mere bivalent glycan binding capacity. Indeed, the Gal-4FL mutants, which display only one active glycan binding site, also aggregate in the presence of the LPS, behaving as the bivalent Gal-4FL, very different from the monovalent single domains. This observation suggests that these two mutants exhibit a multivalent binding capacity, most likely through protein oligomerization. ^{1}H-DOSY NMR experiments corroborated that Gal-4FL shows a tendency to self-associate. Interestingly, in contrast to other galectin heterodimers, the linker domain of Gal-4 has a fixed length. Indeed, its sequence is highly similar to that of the non-lectin N-domain of Gal-3, which has been demonstrated to mediate Gal-3 oligomerization (50–52). Our data suggest that, besides the linker, an additional element for oligomerization could imply the F-face of the C-domain, as has also been postulated for Gal-9C (54).

This study highlights the architecture of Gal-4 as a fundamental element for the recognition of multivalent presented

glycan ligands, such as LPS. Both CRD domains behave independently in terms of oligosaccharide binding, but an intrinsic trend of the full-length form to oligomerize, provided by the linker and C-domain, allows for an enhanced capacity to cross-link these multivalent ligands, which is beyond Gal-4 bivalent binding nature.

Experimental procedures

Proteins expression and purification

Plasmids

The constructs for Gal-4FL and Gal-4C were synthesized by GenScript. For Gal-4FL, the DNA fragment coding from amino acids 1 to 323, including an additional N-terminal SUMO fusion protein and His-tag, was inserted into the pET28a-SUMO expression vector (Novagen) (58). For Gal-4FL-H63R and Gal-4FL- H236R, point mutations (H63R and H236R) were inserted into two different pET28a-SUMO expression vectors. The plasmid for the SUMO ULP1 protease, used to cleave the His6-tagged SUMO, was a gift from Hideo Iwai (Addgene plasmid # 31122). For Gal-4C, amino acid sequence coding for amino acids 179 to 323, including a C-terminus His-tag and a thrombin-cleavage site and was inserted into a pET29b(+) expression vector (59).

Expression of unlabeled proteins

The plasmid for the Gal-4C was transformed into *E. coli* BL21 cells, while plasmids for the Gal-4FL variants were transformed into *E. coli* Rosetta (DE3) cells. In both cases, a single colony was inoculated into 200 ml LB medium containing 50 μg mL^{-1} kanamycin and cultured at 37 °C overnight. A precise quantity of the cultured was then added to 2l of fresh LB containing kanamycin to achieve an initial absorbance at 600 nm (A_{600}) of 0.1. Cells were grown at 37 °C until A_{600} reached 0.6 to 0.8, subsequently, protein expression was induced by addition of 1 mM of IPTG, and the culture was allowed to grow overnight at 20 °C (58, 59).

Expression of ^{15}N- and ^{13}C, ^{15}N-labeled proteins

The transformed cells were inoculated into 5 ml of LB containing the antibiotic and were grown 6 h at 37 °C. The small preculture was centrifuged and resuspended in 1 ml of M9 minimal medium, transferred to 200 ml of M9 medium and then incubated overnight at 37 °C. A precise quantity of the culture was then added to obtain a starting A_{600} of 0.1 in 2l of fresh M9 medium containing 1 g/L $^{15}NH_4Cl$ for the ^{15}N-labeled samples, or 1 g/L $^{15}NH_4Cl$ and 20% w/v U–^{13}C-glucose for the ^{13}C,^{15}N-labeled samples. Upon reaching an A_{600} of 0.6 protein expression was induced by addition of 1 mM of IPTG. *E. coli* BL21 cells were harvest after incubation overnight at 20 °C. *E. coli Rosetta* (DE3) cells were harvest after 24 h at 25 °C.

Protein purification

The obtained culture pellets were resuspended in lysis buffer (50 mM sodium phosphate pH 8, 600 mM NaCl, 1 mM PMSF, and 1 mM DTT for the full-length proteins) and sonicated at 4 °C. The crude extract was clarified by centrifugation at 35,000 rpm for 1 h at 4 °C. The soluble fraction was purified by Ni-NTA affinity chromatography and further purified by size-exclusion chromatography in a HiLoad 26/600 Superdex 75 column. For the full-length proteins, the SUMO tag was removed by incubation 1 h at room temperature (RT) using SUMO protease (10 units of SUMO protease per mg of protein). Afterward, the protein was loaded onto a 5 ml Ni-NTA column, and the desired fragment collected in the wash and loaded in a HiLoad 26/600 Superdex 75 column. Protein purity was checked by 4 to 12% SDS-PAGE and by LC-MS.

Ligands

Lactose was purchased from Sigma-Aldrich. Glycans H type-6, B type-6 trisaccharide, A type –1, –2 and –6 tetrasaccharides, A type-4 pentasaccharide, B type –1, –2 and –6 tetrasaccharides, and B type-4 pentasaccharide were purchased from Elicityl (references GLY031–3, GLY074–2, GLY035–1, GLY035–2, GLY035–3, GLY128, GLY038–1, GLY038–2, GLY038–3, and GLY129). The LPS from *E. Coli* O55 was purchased from Sigma-Aldrich.

Nuclear magnetic resonance

General

The total volume for the NMR samples was 500 μl. The proteins were dissolved in phosphate-buffered saline (50 mM sodium phosphate, 150 mM NaCl, pH 7.4, with 1 mM DTT for the full-length proteins), either in D_2O or 90:10 H_2O:D_2O depending on the NMR experiment. The pH was adjusted with the required amount of NaOH and HCl or NaOD and DCl. Ligands' ^1H and ^{13}C resonance assignment is provided in Figs. S8–S11, based on the analysis of standard 2D-TOCSY, NOESY, and HSQC spectra.

^1H,^{15}N-HSQC based titrations

^1H,^{15}N-HSQC spectra were recorded either on an 800 MHz (with a cryoprobe) or 600 MHz Bruker spectrometers. The samples were prepared using 50 to 100 μM of the ^{15}N labeled lectin in the corresponding buffer in 90:10 H_2O:D_2O. The experiments were acquired at 298 K. Plots for the CSP of the protein-backbone NH groups with each ligand (Figs. S4 and S5) was analyzed as implemented in CcpNmr Analysis 2.4.2 (https://ccpn.ac.uk/software/version-2) and calculated using the formula: CSP (ppm) = $[(\Delta\delta H^2 + (0.14{\cdot}\Delta\delta N)^2)/2]^{½}$). Plots of the average (^1H, ^{15}N) chemical shift difference between Gal-4C bound to two related oligosaccharide ligands (Figs. 1, S6, and S7) was calculated as $\sqrt{\frac{1}{2}[\Delta\delta(1H)^2+(0.14\,x\Delta\delta(15N)^2\,]}$, where $\Delta\delta(1H)$ = δ ^1H (Gal 4C bound to Ligand1) - δ ^1H (Gal-4C bound to Ligand2), and $\Delta\delta(15N)$ = δ ^{15}N (Gal-4C bound to Ligand1) - δ ^{15}N (Gal-4C bound to Ligand2). For K_D estimations, six to nine titration points were recorded for each glycan, and K_D values were calculated as implemented in CcpNmr Analysis 2.4.2. The error is expressed as standard deviation from a minimum of five and a maximum of ten amino acids, depending on the system.

DOSY experiments

^{1}H-DOSY NMR experiments were acquired by using a stimulated echo sequence using bipolar gradients, at different concentrations of Gal-4FL (30 μM, 100 μM, 175 μM, and 230 μM), at 288K, on an 800 MHz (cryo) spectrometer. In total, 32 1D ^{1}HNMR experiments were acquired with 520 scans, with a linear gradient stepped between 2% and 95%. Big delta and little delta were 350 ms and 3.5 ms, respectively.

^{1}H-STD NMR

All ^{1}H-STD NMR experiments were acquired on an 800 MHz (with a cryoprobe). The samples were prepared in the corresponding deuterated buffer. ^{1}H-STD NMR spectra were acquired with 1024 scans, 2 s of saturation time using a train of 50 ms Gaussian-shaped pulses, and 3 s of relaxation delay. The spin-lock filter applied to remove the residual signals of the lectin was set at 40 ms. The protein:ligand ratio and the temperature were optimized for each lectin. For Gal-4C, the temperature was 298 K. Subsequently, 50 μM of the lectin with 50 equivalents of the ligand were used. The on-resonance frequencies were set at 7.19 ppm (aromatic region) and at 0.52 ppm (aliphatic region), while the off-resonance frequency was set at 100 ppm. The experiments with Gal-4FL, Gal-4FL-H63R, and Gal-4FL-H236R were acquired at 298 K. The on-resonance frequency was set at 7.03 ppm, and the off-resonance at 100 ppm. The experiments were performed using 25 μM of Gal-4FL and 50 μM of Gal-4FL-H63R and Gal-4FL-H236R, all with 50 equivalents of ligand.

ROESY spectra

ROESY NMR spectra for the glycans were acquired, in the 800 MHz Bruker spectrometer, in the presence of 50 μM Gal-4C using 1:10 protein:ligand ratios in the corresponding deuterated phosphate-buffered saline buffer, at 298 K.

Protein backbone resonance assignment

The experiments required for achieving the backbone resonance assignment of Gal-4C were carried out at 25 °C on a 800 MHz Bruker spectrometer equipped with a cryoprobe. 3D HNCO, HN(CA)CO, HN(CO)CACB, HNCACB, HN(CO)CA, and HNCA experiments were acquired for the free Gal-4C containing the His-Tag. The entire peak analysis provided the unambiguous identification of 86% of the expected NH signals for Gal-4C. The spectra were processed with Bruker TopSpin 3.5.2 (https://www.bruker.com/es/products-and-solutions/mr/nmr-software/topspin.html) and analyzed *via* CARA NMR 1.9.1.4 (http://cara.nmr.ch/doku.php/home).

Competitive solid-phase assays—The assay was adapted from Rapoport *et al.* (45) Briefly, 96-well plates (flat-bottom) were coated with 10 μg/ml ASF (Sigma-Aldrich) in sodium carbonate buffer (pH 9.6) and incubated overnight at 4 °C. Then, wells were washed three times with 100 μl/well PBS-Tween 0.05% (w/v) and blocked with 100 μl/well PBS-bovine serum albumin (BSA) 2% (1 h, RT) in a humid chamber. Meanwhile, equal volumes of human Gal-4C (60 μg/ml, final concentration) in PBS-BSA 0.3% buffer were preincubated with the corresponding inhibitors in serial dilutions (2 h, 37 °C). Glycans were evaluated as potential galectin competing ligands in the presence of immobilized ASF (2 h, 37 °C). After washing, Gal-4C was detected using a biotinylated anti-Gal-4 detection antibody (R&D, BAF1227) in PBS-BSA 0.3% w/v (1 h, RT). Then, plates were rinsed three times before adding 0.33 μg/ml horseradish peroxidase-labeled streptavidin (Sigma-Aldrich) for 30 min. After washing, 100 μl TMB solution (0.1 mg/ml tetramethylbenzidine and 0.06% H_2O_2 in citrate-phosphate buffer, pH 5.0) was added to plates. The reaction was stopped by adding a 2N H_2SO_4 solution. Absorbances were determined at 450 nm in a Multiskan MS microplate reader (Thermo Fisher Scientific).

Hemagglutination assay and hemagglutination inhibition assays

Hemagglutination assays were performed as previously described (60). Briefly, human A and 0 type RBCs were trypsinized and fixed with glutaraldehyde 1% w/v and stored at 4 °C until use (61). The assay was performed in 96-well microplates (U-bottom) with sequential dilutions of human recombinant Gal-4C in PBS buffer. For hemagglutination inhibition assays, serial dilutions of BGA6 glycan were incubated in the presence of 50 μg/ml Gal-4C. Trypsinized RBCs (3.7 x10^4/ml, final concentration) in PBS buffer were added, gently mixed by pipetting, and incubated 1 to 2 h at RT.

Isothermal titration calorimetry

ITC experiments were performed using MicroCal PEAQ-ITC calorimeter. Samples containing 100 to 200 μM the lectins in PBS buffer (50 mM sodium phosphate pH 7.4, 150 mM NaCl) were titrated with stocks of 3 to 10 mM in PBS of the BGAs. During the automated experiment, small aliquots (2–3 μl) of the sugar stocks were added to the cell containing the target lectin. Curve fitting to a single binding site model was performed with MicroCal Origin 7 software (https://www.originlab.com/). The error is expressed as the standard deviation of a minimum of three and maximum of five measurements for each interacting system.

Generation of all-atom Gal-4FL dimer models

First, AlphaFold multimer (55, 56) was used to generate 30 structures for each of the five deep-neural network models, yielding a total of 150 structures. At this stage, only the C-terminal domains and the last 13 residues of the flexible linker (CHQQLNSLPTMEG) attached to such domains were used in the prediction. After visual inspection, three out of the 150 generated structures (one from the AlphaFold model two set and two from the AlphaFold model three set) in which the F-faces of the C-terminal domains appeared to be interacting in the dimer were selected. Then, the rest of Gal-4FL was modeled as a monomer using AlpaFold (model 1, one single run), and the resulting structure was manually attached to each of the C-terminal domains at the corresponding position of the flexible linker using PyMol (https://www.pymol.org/)

(62). The three full-length dimers thereby constructed were refined by MD simulations (see below), and one of them was selected for discussion based on the conservation of the F-face dimerization interface.

Molecular dynamics

Simulations were run with Amber 20 (https://ambermd.org/) (63), using the ff19SB (64) force field for the protein. Protein complexes were immersed in a box with 10 Å buffer of OPC3 (65) water molecules and neutralized by adding explicit Na+ or Cl- counterions. A two-stage optimization approach was performed. The first stage minimizes only the positions of solvent molecules and ions, and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems were then heated by incrementing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions. Harmonic restraints of 10 kcal mol-1 were applied to the solute, and the Andersen temperature coupling scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Water molecules were treated with the SHAKE algorithm (66) such that the angle between the hydrogen atoms is kept fixed through the simulations. Long-range electrostatic effects were modeled using the particle mesh Ewald method (67). An 8 Å cutoff was applied to Lennard-Jones interactions. Each system was equilibrated for 2 ns with 2 fs time step at a constant volume and temperature of 300 K. Three independent production trajectories were then run for an additional 600 ns under the same simulation conditions, leading to accumulated simulation times of 1.8 μs for each modeled full-length h-Gal-4 dimer. Snapshots from the MD simulations were extracted using the cpptraj module (68). in Amber 20 and rendered with PyMol.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding authors on reasonable request.

Supporting information—This material is available free of charge *via* the Internet at http://pubs.acs.org.

Acknowledgments—We thank generous funding by the European Research Council (RECGLYCANMR, Advanced Grant No. 788143), the Agencia Estatal de Investigación (Spain) for grants PDI2021 to 1237810B-C21, PID2021 to 125946OB-I00 and the Severo Ochoa Excellence Accreditation CEX2021 to 001136-S, and CIBERES, an initiative of Instituto de Salud Carlos III (ISCIII), Madrid, Spain. A. J. C., K. V. M., and G. A. R. are researchers at the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina. MM was supported by a doctoral fellowship from CONICET.

Author contributions—J. I. Q., M. M., R. N.-F., and S. D. investigation; J. I. Q., M. M., R. N.-F., and S. D. validation; J. I. Q., M. M., R. N.-F., and S. D. visualization; J. I. Q., M. M., R. N.-F., S. D., and A. A. writing–original draft; J. I. Q., M. M., R. N.-F., and S. D. formal analysis; J. I. Q., A. J. C., G. J.-O., K. V. M., G. A. R., J. J.-B., and A. A. project administration; A. J. C., G. J.-O., K. V. M., G. A. R., J. J.-B., and A. A. methodology; A. J. C., G. A. R., K. V. M., J. J.-B., and A. A. writing–review and editing; A. J. C., K. V. M., G. A. R., G. J.-O., J. J.-B., and A. A. resources.

Funding and additional information—Agencia Estatal de Investigacion of Spain: PDI2021 to 1237810B-C21 and CEX2021 to 001136-S from MCIN/AEI/10.13039/501100011033. European Research Council (RECGLYCANMR, Advanced Grant no. 788143 to J. J. B.). This work was also supported by grants from the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT, PICT 2015–0564 to K. V. M., PICT V 2014–3687 to GAR, PICT 2018–00955 to A. J. C.) and Florencio Fiorini Foundation (to A. J. C.); as well as grants from Sales, Bunge & Born, René Baron Foundations (Argentina) and The Richard Lounsbery Foundation (USA) (to G. A. R.). We also thank the Ferioli, Ostry, and Carballo families for generous donations.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: ASF, asialofetuin; BGA, blood group antigen; BSA, bovine serum albumin; CRD, carbohydrate recognition domain; CSP, chemical shift perturbation; Gal-4C, C-domain of Gal-4; Gal-4FL, full-length Gal-4; HSQC, heteronuclear single quantum coherence; ITC, isothermal titration calorimetry; LPS, lipopolysaccharide; MD, molecular dynamics; RBC, red blood cell; ROESY, rotating frame Overhauser effect spectroscopy; RT, room temperature; STD, saturation transfer difference.

References

1. Cummings, R. D., Liu, F.-T., Rabinovich, G. A., Stowell, S. R., and Vasta, G. R. (2022) *Galectins*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY)
2. Mariño, K. V., Cagnoni, A. J., Croci, D. O., and Rabinovich, G. A. (2023) Targeting galectin-driven regulatory circuits in cancer and fibrosis. *Nat. Rev. Drug Discov.* **22**, 295–316
3. Troncoso, M. F., Elola, M. T., Blidner, A. G., Sarrias, L., Espelt, M. V., and Rabinovich, G. A. (2023) The universe of galectin-binding partners and their functions in health and disease. *J. Biol. Chem.* **299**, 105400
4. Rabinovich, G. A., Toscano, M. A., Jackson, S. S., and Vasta, G. R. (2007) Functions of cell surface galectin-glycoprotein lattices. *Curr. Opin. Struct. Biol.* **17**, 513–520
5. Bättig, P., Saudan, P., Gunde, T., and Bachman, M. F. (2004) Enhanced apoptotic activity of a structurally optimized form of galectin-1. *Mol. Immunol.* **41**, 9–18
6. Earl, L. A., Bi, S., and Baum, L. G. (2011) Galectin multimerization and lattice formation are regulated by linker region structure. *Glycobiology* **21**, 6–12
7. Nabi, I. R., Shankar, J., and Dennis, J. W. (2015) The galectin lattice at a glance. *J. Cell Sci.* **128**, 2213–2219
8. Bertuzzi, S., Gimeno, A., Núñez-Franco, R., Bernardo-Seisdedos, G., Delgado, S., Jiménez-Osés, G., *et al.* (2020) Unravelling the time scale of conformational plasticity and allostery in glycan recognition by human galectin-1. *Chemistry* **26**, 15643–15653
9. Nesmelova, I. V., Ermakova, E., Daragan, V. A., Pang, M., Menéndez, M., Lagartera, L., *et al.* (2010) Lactose binding to galectin-1 modulates structural dynamics, increases conformational entropy, and occurs with apparent negative cooperativity. *J. Mol. Biol.* **397**, 1209–1230
10. Vasta, G. R. (2009) Roles of galectins in infection. *Nat. Rev. Microbiol.* **7**, 424–438

11. Vasta, G. R. (2020) Lectin in host defense against microbial infections. In: Hsieh, S.-L., ed. *Advances in Experimental Medicine and Biology*, Springer, Singapore: 169–196
12. Stowell, S. R., Arthur, C. M., Dias-Baruffi, M., Rodrigues, L. C., Gourdine, J. P., Heimburg-Molinaro, J., *et al.* (2010) Innate immune lectins kill bacteria expressing blood group antigen. *Nat. Med.* **16**, 295–301
13. Blenda, A. V., Kamili, N. A., Wu, S.-C., Abel, W. F., Ayona, D., Gerner-Smidt, C., *et al.* (2022) Galectin-9 recognizes and exhibits antimicrobial activity toward microbes expressing blood group-like antigens. *J. Biol. Chem.* **298**, 101704
14. Arthur, C. M., Patel, S. R., Mener, A., Kamili, N. A., Fasano, R. M., Meyer, E., *et al.* (2015) Innate immunity against molecular mimicry: examining galectin-mediated antimicrobial activity. *BioEssays* **37**, 1327–1337
15. Li, C.-S., Lo, T.-H., Tu, T.-J., Chueh, D.-Y., Yao, C.-I., Lin, C.-H., *et al.* (2023) Cytosolic galectin-4 enchains bacteria, restricts their motility, and promotes inflammasome activation in intestinal epithelial cells. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2207091120
16. Girotti, M. R., Salatino, M., Dalotto-Moreno, T., and Rabinovich, G. A. (2020) Sweetening the hallmarks of cancer: galectins as multifunctional mediators of tumor progression. *J. Exp. Med.* **217**, e20182041
17. Rabinovich, G. A., and Conejo-Garcia, J. R. (2016) Shaping the immune landscape in cancer by galectin-driven regulatory pathways. *J. Mol. Biol.* **428**, 3266–3281
18. Cao, Z. Q., and Guo, X. L. (2016) The role of galectin-4 in physiology and diseases. *Protein Cell* **7**, 314–324
19. Rao, U. S., and Rao, P. S. (2017) Surface-bound galectin-4 regulates gene transcription and secretion of chemokines in human colorectal cancer cell lines. *Tumor Biol.* **39**. https://doi.org/10.1177/1010428317691687
20. Kaur, S., Singh, J., and Kaur, M. (2023) Multifaceted role of galectin-4 in cancer: a systematic review. *Eur. J. Clin. Invest.* **53**, e13987
21. Tsai, C. H., Tzeng, S. F., Chao, T. K., Tsai, C. Y., Yang, Y. C., Lee, M. T., *et al.* (2016) Metastatic progression of prostate cancer is mediated by autonomous binding of galectin-4-O-glycan to cancer cells. *Cancer Res.* **76**, 5756–5767
22. Barrow, H., Rhodes, J. M., and Yu, L.-G. (2013) Simultaneous determination of serum galectin-3 and -4 levels detects metastases in colorectal cancer patients. *Cell Oncol.* **36**, 9–13
23. Watanabe, M., Takemasa, I., Kaneko, N., Yokoyama, Y., Matsuo, E., Iwasa, S., *et al.* (2011) Clinical significance of circulating galectins as colorectal cancer markers. *Oncol. Rep.* **25**, 1217–1226
24. Helwa, R., Heller, A., Knappskog, S., and Bauer, A. S. (2017) Tumor cells interact with red blood cells *via* galectin-4 - a short report. *Cell Oncol.* **40**, 401–409
25. Dean, L. (2005) *Blood Groups and Red Cell Antigens*, National Center for Biotechnology Information (US), Bethesda (MD)
26. Clausen, H., and Hakomori, S. (1989) ABH and related histo-blood group antigens; immunochemical differences in carrier isotypes and their distribution. *Vox Sang.* **56**, 1–20
27. Tanaka, A., Kimura, A., Yamamoto, Y., Uede, K., and Furukawa, F. (2008) Expression of histo-blood group A type 1, 2 and 3 antigens in normal skin and extramammary paget's disease. *Acta Histochem. Cytochem.* **41**, 165–171
28. Quintana, J. I., Delgado, S., Núñez-Franco, R., Cañada, F. J., Jiménez-Osés, G., Jiménez-Barbero, J., *et al.* (2021) Galectin-4 N-terminal domain: binding preferences toward A and B antigens with different peripheral core presentations. *Front. Chem.* **9**, 664097
29. Horlacher, T., Oberli, M. A., Werz, D. B., Kröck, L., Bufali, S., Mishra, R., *et al.* (2010) Determination of carbohydrate-binding preferences of human galectins with carbohydrate microarrays. *ChemBioChem* **11**, 1563–1573
30. Bum-Erdene, K., Leffler, H., Nilsson, U. J., and Blanchard, H. (2016) Structural characterization of human galectin-4 N-terminal carbohydrate recognition domain in complex with glycerol, lactose, 3′-sulfo-lactose and 2′-fucosyllactose. *Sci. Rep.* **6**, 20289
31. Sindrewicz, P., Li, X., Yates, E. A., Turnbull, J. E., Lian, L.-Y., and Yu, L.-G. (2019) Intrinsic tryptophan fluorescence spectroscopy reliably determines galectin-ligand interactions. *Sci. Rep.* **9**, 11851
32. Vokhmyanina, O. A., Rapoport, E. M., André, S., Severov, V. V., Ryzhov, I., Pazynina, G. V., *et al.* (2012) Comparative study of the glycan specificities of cell-bound human tandem-repeat-type galectin-4, -8 and -9. *Glycobiology* **22**, 1207–1217
33. Ideo, H., Seko, A., and Yamashita, K. (2005) Recognition mechanism of galectin-4 for cholesterol 3-sulfate. *J. Biol. Chem.* **280**, 4730–4737
34. Gimeno, A., Valverde, P., Ardá, A., and Jiménez-Barbero, J. (2020) Glycan structures and their interactions with proteins. A NMR view. *Curr. Opin. Struct. Biol.* **62**, 22–30
35. Valverde, P., Quintana, J. I., Santos, J. I., Ardá, A., and Jiménez-Barbero, J. (2019) Novel NMR avenues to explore the conformation and interactions of glycans. *ACS Omega* **4**, 13618–13630
36. Ardá, A., and Jiménez-Barbero, J. (2018) The recognition of glycans by protein receptors. Insights from NMR spectroscopy. *Chem. Commun.* **54**, 4761–4769
37. Inagaki, F. (2013) *Encyclopedia of Biophysics*. Springer Berlin Heidelberg, Berlin, Heidelberg: 2033–2037
38. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., *et al.* (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696
39. Bum-Erdene, K., Leffler, H., Nilsson, U. J., and Blanchard, H. (2015) Structural characterization of human galectin-4 C-terminal domain: elucidating the molecular basis for recognition of glycosphingolipids, sulfated saccharides and blood group antigens. *FEBS J.* **282**, 3348–3367
40. Asensio, J. L., Ardá, A., Cañada, F. J., and Jiménez-Barbero, J. (2013) Carbohydrate–aromatic interactions. *Acc. Chem. Res.* **46**, 946–954
41. Gimeno, A., Delgado, S., Valverde, P., Bertuzzi, S., Berbis, M. A., Echavarren, J., *et al.* (2019) Minimizing the entropy penalty for ligand binding: lessons from the molecular recognition of the histo blood-group antigens by human galectin-3. *Angew. Chem. Int. Ed. Engl.* **58**, 7268–7272
42. [preprint] Platzer, G., Mayer, M., Böttcher, J., Geist, L., Fuchs, J. E., McConnell, D. B., *et al.* (2021) Inverse PI by NMR: analysis of ligand ^{1}H-chemical shifts in the protein-bound state. *bioRxiv*. https://doi.org/10.1002/cphc.202300636
43. Moure, M. J., Gimeno, A., Delgado, S., Diercks, T., Boons, G. J., Jiménez-Barbero, J., *et al.* (2021) Selective ^{13}C-labels on repeating glycan oligomers to reveal protein binding epitopes through NMR: polylactosamine binding to galectins. *Angew. Chem. Int. Ed. Engl.* **60**, 18777–18782
44. Modenutti, C. P., Capurro, J. I. B., Di Lella, S., and Martí, M. A. (2019) The structural biology of galectin-ligand recognition: current advances in modeling tools, protein engineering, and inhibitor design. *Front. Chem.* **7**, 823
45. Rapoport, E. M., Pochechueva, T. V., Kurmyshkina, O. V., Pazynina, G. V., Severov, V. V., Gordeeva, E. A., *et al.* (2010) Solid-phase assays for study of carbohydrate specificity of galectins. *Biochemistry (Moscow)* **75**, 310–319
46. Oriol, R., Le Pendu, J., and Mollicone, R. (1986) Genetics of ABO, H, Lewis, X and related antigens. *Vox Sang.* **51**, 161–171
47. Stenutz, R., Weintraub, A., and Widmalm, G. (2006) The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol. Rev.* **30**, 382–403
48. Groves, P., Rasmussen, M. O., Molero, M. D., Samain, E., Cañada, F. J., Driguez, H., *et al.* (2004) Diffusion ordered spectroscopy as a complement to size exclusion chromatography in oligosaccharide analysis. *Glycobiology* **14**, 451–456
49. Kandiyal, P. S., Kim, J. Y., Fortunati, D. L., and Mok, K. H. (2019) Size determination of protein oligomers/aggregates using diffusion NMR spectroscopy. *Methods Mol. Biol.* **2039**, 173–183
50. Chiu, Y.-P., Sun, Y.-C., Qiu, D.-C., Lin, Y.-H., Chen, Y.-Q., Kuo, J.-C., *et al.* (2020) Liquid-liquid phase separation and extracellular multivalent interactions in the tale of galectin-3. *Nat. Commun.* **11**, 1229
51. Zhao, Z., Xu, X., Cheng, H., Miller, M. C., He, Z., Gu, H., *et al.* (2021) Galectin-3 N-terminal tail prolines modulate cell activity and glycan-mediated oligomerization/phase separation. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2021074118
52. Sato, S. (2023) Why does galectin-3 have a unique intrinsically disordered region? "Raison d'être" for the disordered structure and liquid–liquid

phase separation –Part 1–. *Glycoforum*. https://doi.org/10.32285/glycoforum.26A1
53. Nagae, M., Nishi, N., Murata, T., Usui, T., Nakamura, T., Wakatsuki, S., *et al.* (2006) Crystal structure of the galectin-9 N-terminal carbohydrate recognition domain from Mus musculus reveals the basic mechanism of carbohydrate recognition. *J. Biol. Chem.* **281**, 35884–35893
54. Nonaka, Y., Ogawa, T., Oomizu, S., Nakakita, S., Nishi, N., Kamitori, S., *et al.* (2013) Self-association of the galectin-9 C-terminal domain *via* the opposite surface of the sugar-binding site. *J. Biochem.* **153**, 463–471
55. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589
56. [preprint] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., *et al.* (2021) Protein complex prediction with AlphaFold-Multimer. *bioRxiv* **10**. https://doi.org/10.1101/2021.10.04.463034
57. Slámová, K., Červený, J., Mészáros, Z., Friede, T., Vrbata, D., Křen, V., *et al.* (2023) Oligosaccharide ligands of galectin-4 and its subunits: multivalency scores highly. *Molecules* **28**, 4039
58. Rustiguel, J. K., Soares, R. O. S., Meisburger, S. P., Davis, K. M., Malzbender, K. L., Ando, N., *et al.* (2016) Full-length model of the human galectin-4 and insights into dynamics of inter-domain communication. *Sci. Rep.* **6**, 33633
59. Rustiguel, J. K., Kumagai, P. S., Dias-Baruffi, M., Costa-Filho, A. J., and Nonato, M. C. (2016) *Protein Expr. Purif.* **118**, 39–48
60. Maller, S. M., Cagnoni, A. J., Bannoud, N., Sigaut, L., Pérez Sáez, J. M., Pietrasanta, L. I., *et al.* (2020) An adipose tissue galectin controls endothelial cell function *via* preferential recognition of 3-fucosylated glycans. *FASEB J.* **34**, 735–753
61. Nowak, T. P., and Barondes, S. H. (1975) Agglutinin from Limulus polyphemus: purification with formalinized horse erythrocytes as the affinity adsorbent. *Biochim. Biophys. Acta, Prot. Struct. Mol.* **393**, 115–123
62. [software]. *The PyMOL Molecular Graphics System, Version 2.4* (2020). Schrödinger LLC, New York, NY
63. Case, D. A., Belfon, K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, I. I. I.,T. E., *et al.* (2020) *AMBER 2020*, University of California, San Francisco
64. Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Migues, A. N., *et al.* (2020) ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theor. Comput.* **16**, 528–552
65. Izadi, S., and Onufriev, A. V. (2016) Accuracy limit of rigid 3-point water models. *J. Chem. Phys.* **145**, 074501
66. Miyamoto, S., and Kollman, P. A. (1992) Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962
67. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092
68. Roe, D. R., and Cheatham, T. E., 3rd (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theor. Comput.* **9**, 3084–3095
69. Cagnoni, A. J., Massaro, M., Cutine, A. M., Gimeno, A., Pérez-Sáez, J. M., Manselle Cocco, M. N., *et al.* (2024) Exploring galectin interactions with human milk oligosaccharides and blood group antigens identifies BGA6 as a functional galectin-4 ligand. *J. Biol. Chem.* https://doi.org/10.1016/j.jbc.2024.107573