Pure Appl. Chem., Vol. 84, No. 1, pp. 49–64, 2012.
http://dx.doi.org/10.1351/PAC-CON-11-10-01

Structural studies on the interaction of saccharides and glycomimetics with galectin-1: A 3D perspective using a combined molecular modeling and NMR approach*

Sonsoles Martín-Santamaría[1], Hans-Joachim Gabius[2], and Jesús Jiménez-Barbero[3]

[1]Department of Chemistry, Faculty of Pharmacy, Universidad San Pablo CEU, 28668-Boadilla del Monte, Madrid, Spain; [2]Institut für Physiologische Chemie, Tierärztliche Fakultät, Ludwig-Maximilians-Universität, München, Veterinärstrasse 13, 80539 München, Germany; [3]Department of Chemical and Physical Biology, CIB-CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain

Abstract: The interaction of a variety of saccharides and mimetics thereof with lectin receptors has been studied using a combination of molecular modeling protocols and NMR spectroscopy techniques. It is shown that both methods complement each other in a synergistic manner to provide a detailed perspective of the conformational and structural features of the recognition process.

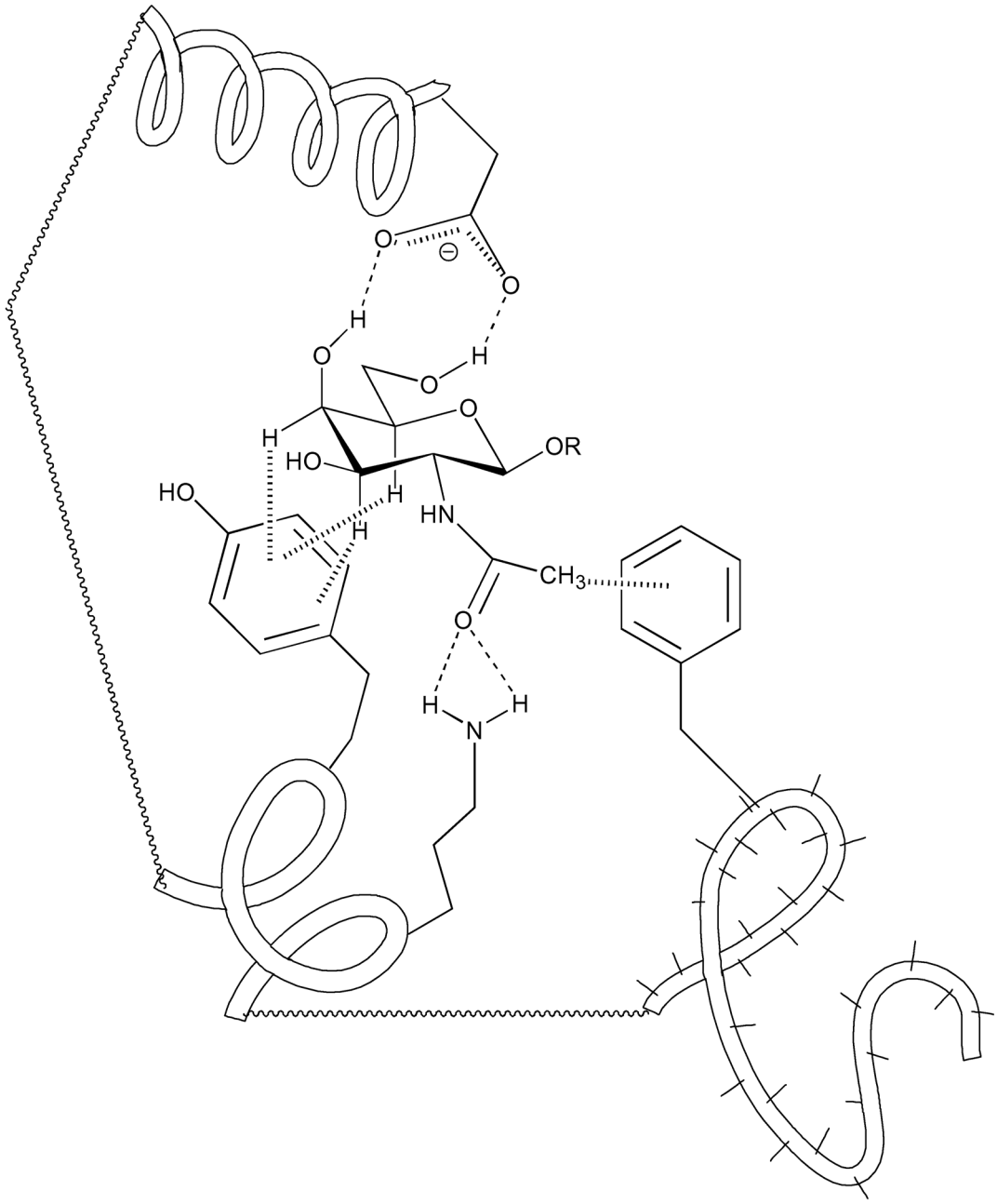
Keywords: bioactive molecules; carbohydrates; computer modeling; conformation; molecular recognition; NMR.

INTRODUCTION

In recent years, research on glycosciences has been much developed owing to the unequivocal demonstration that carbohydrates are involved in a variety of physiological processes, acting as signals for cellular recognition. Among these processes, immune and inflammatory responses, organogenesis, metastasis, and diverse infectious processes should be mentioned [1]. In this context, the elucidation of the mechanisms that govern how oligosaccharides are accommodated in the binding sites of lectins, antibodies, and enzymes is currently a topic of major interest because of its long-range potential for clinical applications [2]. Herein, we will focus on lectins, which are non-enzymatic and non-immunogenic carbohydrate-specific binding proteins that have been widely used as tools in different areas of chemical, biochemical, and biomedical investigations [3].

X-ray crystallography, NMR spectroscopy, and molecular modeling methods have been frequently employed to determine the 3D structures of diverse lectins, both in the free and in the carbohydrate-associated state. The analysis of the structures have permitted the demonstration that van der Waals and stacking interactions as well as H-bonds contribute, in a significant way (Scheme 1), to the modulation of the selectivity, stability, and affinity of the protein–carbohydrate complexes [4].

Pure Appl. Chem.* **84, 1–106 (2012). A collection of invited papers based on presentations at the 16th European Carbohydrate Symposium (Eurocarb-16), Sorrento, Italy, 3–7 July 2011.
‡Corresponding author



Scheme 1 Schematic view of the most frequent interactions involved in protein–carbohydrate interactions.

NMR methods are usually employed in combination with molecular modeling protocols to access to study protein–carbohydrate interactions at atomic resolution [5]. In fact, given the intrinsic flexibility of many saccharides, NMR can only access time-averaged information, and therefore the NMR-based parameters that contain structural and conformational information should be complemented by theoretical molecular modeling methods to get the key conformational features of the receptor-bound carbohydrate [6]. Nowadays, from a technical perspective, the boundaries of NMR for study-

ing large biomolecules, including lectins and their complexes, are continuously growing, owing to the development of new avenues in methodologies and to the access to magnets that operate at very high field [7]. In this context, the application of a multidisciplinary strategy combining molecular simulations and docking procedures with experimental data is probably the best approach to unravel the structural features that govern the recognition of saccharides by receptors [8]. Herein, we focus on some selected examples from our laboratories focused on the recognition of saccharides and glycomimetics with a model lectin, namely, galectin-1, which seems to be involved in inflammatory/autoimmune disorders and can be a potential therapeutic target in cancer and metastasis [9]. Galectins are β-galactoside-binding lectins defined by shared consensus amino acid sequences in their carbohydrate-recognition domain (CRD). These CRDs usually consist of approximately 130 amino acids [10]. At present, 15 members have been identified in mammals, although the number may increase in the future. In fact, there are 10 types of human galectins [11], and they are also present in invertebrates such as nematodes [12] and sponges [13]. These lectins are classified into three subgroups according to their topological structures: Galectin-1, -2, -5, -7, -10, -11, -13, -14, and -15 are non-covalent homodimers comprised of only one type of CRD (~15 kDa). In contrast, galectin-4, -6, -8, -9, and -12 show two different, although highly homologous, CRDs held together within a single polypeptide chain. Alternatively, galectin-3 contains one CRD attached to a non-lectin moiety constituted of proline- and glycine-rich short tandem repeats. This tail is covalently linked to the CRD and directly involved in its oligomerization process [14].

In the last few years, we have studied the interaction of different glycomimetics with a variety of galectins. Herein, we will pay attention to the interactions of these molecules with galectin-1, an endogenous carbohydrate-binding protein, which has been implicated in a variety of processes of paramount importance. For instance, it has been reported that galectin-1 promotes tumour aggressiveness by promoting angiogenesis and T-cell apoptosis [15], events with obvious broad implications in developing novel targeting strategies for galectin-1 in cancer. Moreover, this endogenous lectin is also widely expressed at inflammation sites. Therefore, it has also been postulated as an attractive immunosuppressive agent to restore immune cell tolerance and homeostasis in autoimmune and inflammatory settings and seems also to be involved in infection processes, with different putative roles. As an example, it has been described that galectin-1 displays an inhibitory role of galectin-1 in paramyxovirus infection [16]. On the other hand, galectin-1, which is rather abundant in organs that represent major reservoirs for HIV-1, such as the thymus and lymph nodes, promotes HIV-1 infection by facilitating virus attachment to the host cell surface glycans [17]. On this basis, it has been strongly suggested that blockade of galectin-1 might result in therapeutic benefits in disease [18].

In this context, the development of small ligands that are able to modulate the interaction of galectins with their natural ligands is an area of active research. A variety of molecules with different chemical nature have been tested as galectin inhibitors. Molecules different from sugars, as well as others that are based on sugar and sugar mimetics, have been designed and evaluated [19]. Different design methods have been proposed, from structure-based [20] to combinatorial approaches [21]. Strategies based on the presentation of multivalent ligands have also been reported [22].

From the structural perspective, insights into galectin binding can be very valuable for the understanding of the interaction mechanism of these compounds, and for further design of selective binders able to modulate galectins functions. It has been shown that intermolecular H-bonds as well as van der Waals and CH–π interactions are the key forces involved in the process [23]. From the structural perspective, galectin-1 residues directly involved in carbohydrate binding are part of a pocket formed by four adjacent β-sheets (S3, S4, S5, and S6), being the CRD located at opposite ends of the galectin-1 dimer (Fig. 1).

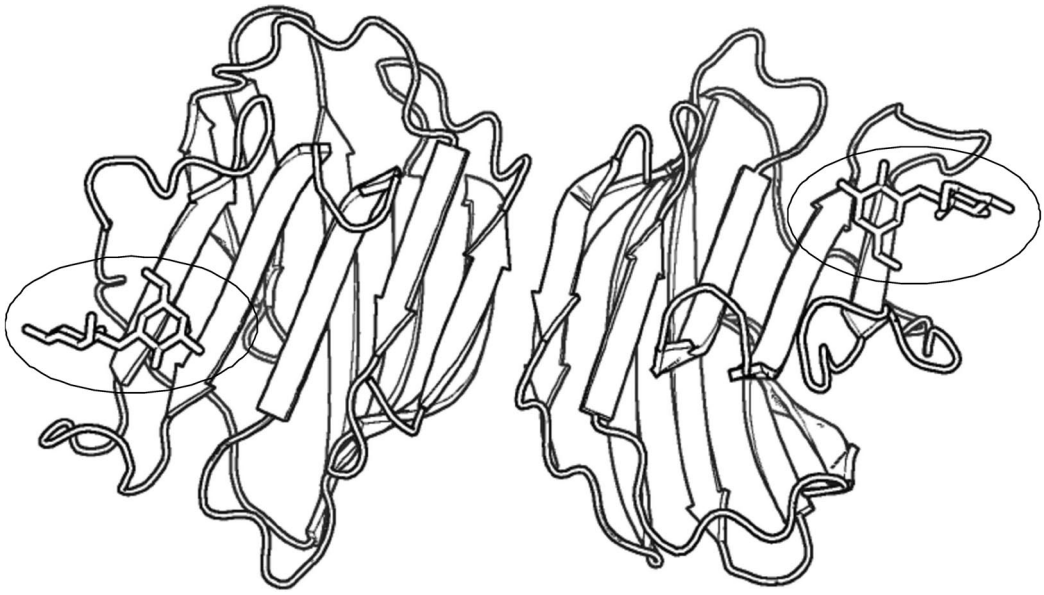


Fig. 1 Galectin-1 dimer (PDB1 gzw) in complex with lactose: the CRDs are located at opposite ends.

Analysis of the 3D structure X-ray crystallographic of the *h*gal1/lactose complex [24] reveals that galectin-1 binds to the C-4 and C-6 hydroxyls of galactose (involving H-bonds with His44, Asn46, and Asn61) and the C-3 hydroxyl of the second saccharide (H-bonds with Glu71) (Fig. 2). A closer look reveals the role of Asp54 (exclusive residue of galectin-1) in orienting Arg48 and Arg73 toward lactose. Moreover, Arg48 establishes a bifurcated H-bond with oxygens in both units of the disaccharide, and Trp68 and His52 participate in binding through characteristic CH–π interactions. Additionally, there is also a possible role of individual water molecules in the recognition process, as the one bridging the H-bond between OH3(Glc) and Asn47.

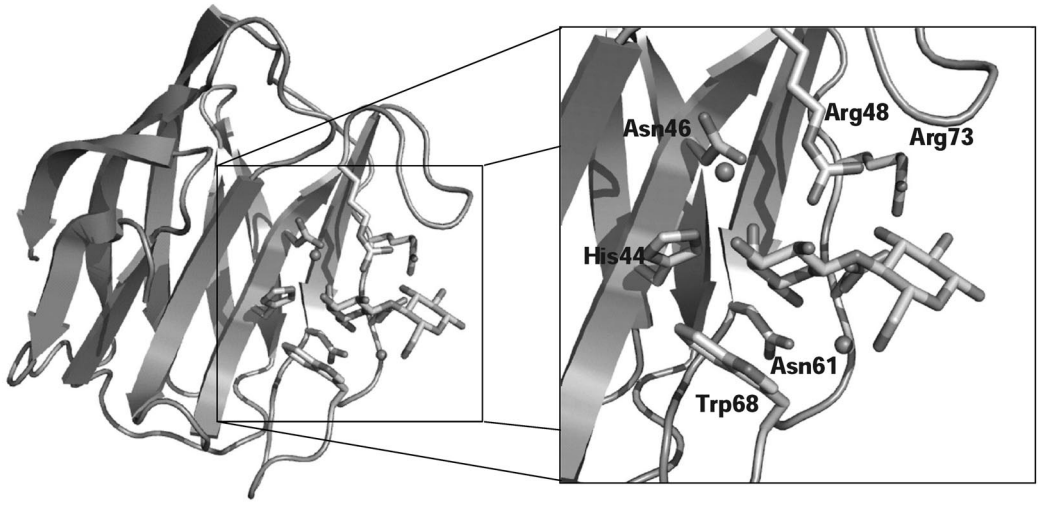
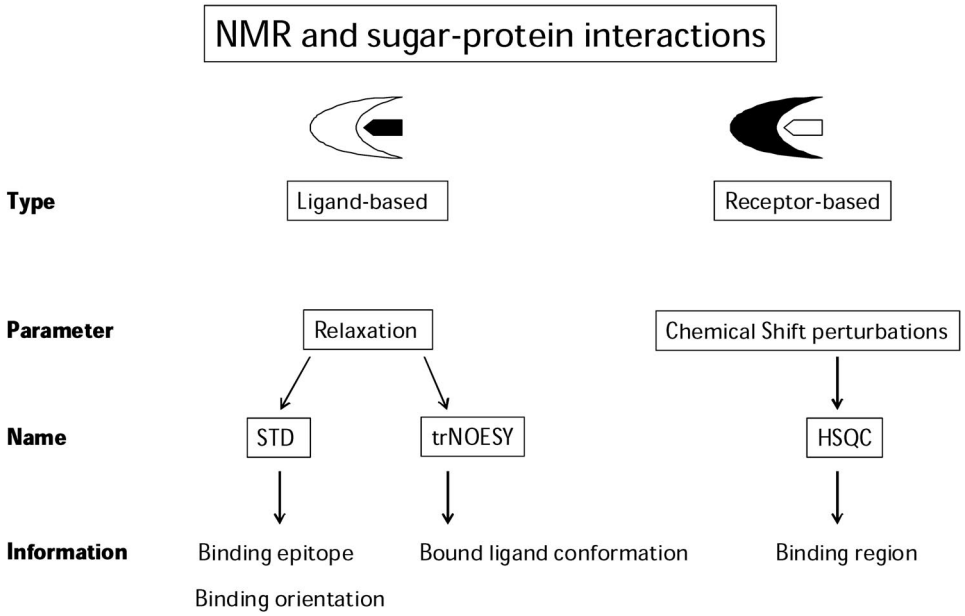


Fig. 2 The galectin-1 binding site (PDB1 gzw, galectin-1 in complex with lactose).

But we will focus here on the combination of NMR and molecular modeling methods. NMR is nowadays a standard method for 3D structure determination of small and large molecules, and also can be applied to monitor molecular recognition processes, with further applications in drug design and discovery [25]. NMR methods that are able to identify and characterize the binding of small molecules to a protein (lectin) fall into two main categories depending on the NMR signals of the detected species. Ligand-based methods monitor the process from the small molecule perspective. They most frequently rely on ^{1}H NMR spectroscopy, although ^{19}F NMR methods have also been developed [26]. They basically rely on changes in the relaxation properties of the ligand between its free and bound states. In contrast, receptor-based methods follow variations in the NMR signals of the lectin in the absence and presence of the ligand, and typically involve the production of isotopically ^{15}N-labeled protein. Most frequently, these methods require the acquisition of 2D heteronuclear correlation spectra to monitor the changes of the receptor ^{1}H and ^{15}N chemical shifts upon ligand binding [27].

From the ligand's perspective, we have employed transferred nuclear Overhauser effect spectroscopy (TRNOESY) [28] and saturation transfer difference (STD) [29] methods (Scheme 2).



Scheme 2 Experiments and information provided by NMR methods.

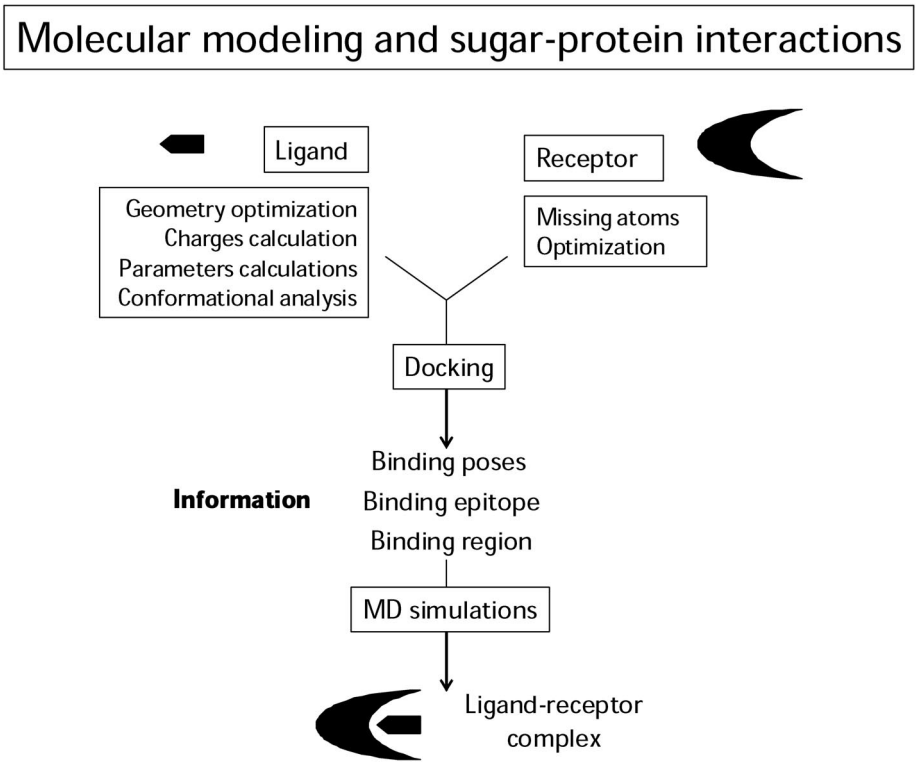
These techniques are particularly useful in the medium–low affinity range and, therefore, they have been adopted to detect ligand interactions, in different systems. In particular, for ligands that are not tightly bound to their receptors and thus exchange between the free and the receptor-bound states at reasonably fast rates, the TR-NOESY experiment provides adequate means to detect binding and to deduce the ligand conformation when bound to the lectin. Indeed, the TRNOESY experiment has found large applicability for monitoring protein–carbohydrate interactions [30]. There are several reasons for achieving this optimal situation. Protein–sugar interactions are rather weak, unless multivalency issues take place [31], and the proton–proton cross peaks that contain the key conformational information on inter-residual distances are accessible to TRNOESY analysis. In detail, it is well known that for large molecules (obviously including molecular complexes), proton–proton cross-relaxation rates are negative. Thus, for small molecules in their bound state, cross-relaxation rates in the presence of the protein (σ^{B}) are negative and opposite in sign to those of the free state (positive, σ^{F}). The value of σ^{B} for a par-

ticular proton pair will depend on the interproton distances, the spectrometer frequency, and on the correlation time of the complex. Provided that there is binding, negative NOEs will be observed. Therefore, the existence of a binding process may be easily visualized from the change of sign of cross peaks in the NOESY/TRNOESY experiments, when passing from the free to the bound state [32]. The STD method allows fast monitoring of ligand binding to lectins [33]. In favorable cases, it may supply information on the binding epitope of the ligand [34]. As for the TRNOESY method, the chemical exchange between the bound and free states of the ligand should be fast in the relaxation time scale. The combination of STD data with those obtained by docking of ligand-receptor complex is becoming a robust method to clarify binding modes with atomic resolution, as will be detailed below [35].

Methods based on the receptor are solidly based on the fact that NMR parameters are very sensitive to change to the chemical environment of the atoms under analysis, from chemical shifts to relaxation parameters [36]. Since chemical shifts are the easiest parameters to be obtained, they are mostly followed up for these protocols. Provided that the assignment of the resonances of the ^{15}N-labeled lectin are known [37], the ligand-induced changes in the chemical shifts of its ^{1}H-^{15}N signal resonances can be directly visualized on the primary sequence, or in the secondary or tertiary structure of the lectin [38]. Thus, these changes describe the binding process from receptor point of view [39]. Fittingly, and contrary to ligand-based NMR techniques, the observation of the interaction process may be accessible to a bigger range of affinities, from millimolar to nanomolar [40].

From the molecular modeling side, computational chemistry and biology are becoming routine methods to study many biological processes. Computational models may provide key information about the crucial interactions involved in the molecular recognition phenomena, and characterization of conformational, dynamics and energetic features governing these processes. In particular, to study the interaction of saccharides and glycomimetics with galectin-1, we have employed ligand–protein docking and molecular dynamics (MD) simulations (Scheme 3).

Ligand–protein docking methodologies have proven to be useful at different stages of the drug discovery process for predicting the binding mode of known active ligands or newly designed compounds, and for the selection/identification of new ligands from extense chemical libraries (virtual screening) [41]. The majority of the currently available methods work in two steps: (1) Sampling: search in the conformational space of the relative orientations of the small molecule interacting with the macromolecule, in order to generate tentative binding solutions or docking poses. All docking approaches introduce sampling algorithms able to perform a near complete search in the intramolecular conformational space of the ligand [42]. Today, most docking programs treat the ligand as flexible while the receptor is considered as rigid (or nearly rigid) structure. Handling efficiently the flexibility of the macromolecule is currently considered one of the major challenges in this field [43]. (2) Scoring: through the use of a force field, the strength of the interaction (binding affinity) is predicted, (un)favorable ligand–receptor interactions are evaluated, and selection (ranking) of the best-docked poses can be carried out. Scoring functions can be both physical or knowledge based and, although they have still to be improved, a number of large-scale assessments of scoring function performance are reported in the literature [44], which points out ligand–protein docking calculations as a sort of routine process in drug discovery and molecular recognition studies. There is a wide range of docking programs available, some representatives being Flex-X [45], Glide [46], DOCK [47], and AutoDock [48]. Based on the reported results in the literature, AutoDock4 was used for the obtaining of binding poses for the ligands under study. AutoDock is a free suite of automated docking tools consisting of two main programs: AutoDock, which performs the docking of the ligand to a set of grids describing the target protein; and AutoGrid, which precalculates these grids. The program allows flexibility in the side-chains of the macromolecule, and has a free-energy scoring function based on a linear regression analysis, the AMBER force field, and a large set of diverse protein–ligand complexes with known inhibition constants. Additionally, customized parameters may be used, which is a major advantage, not always available in current software for docking [49].

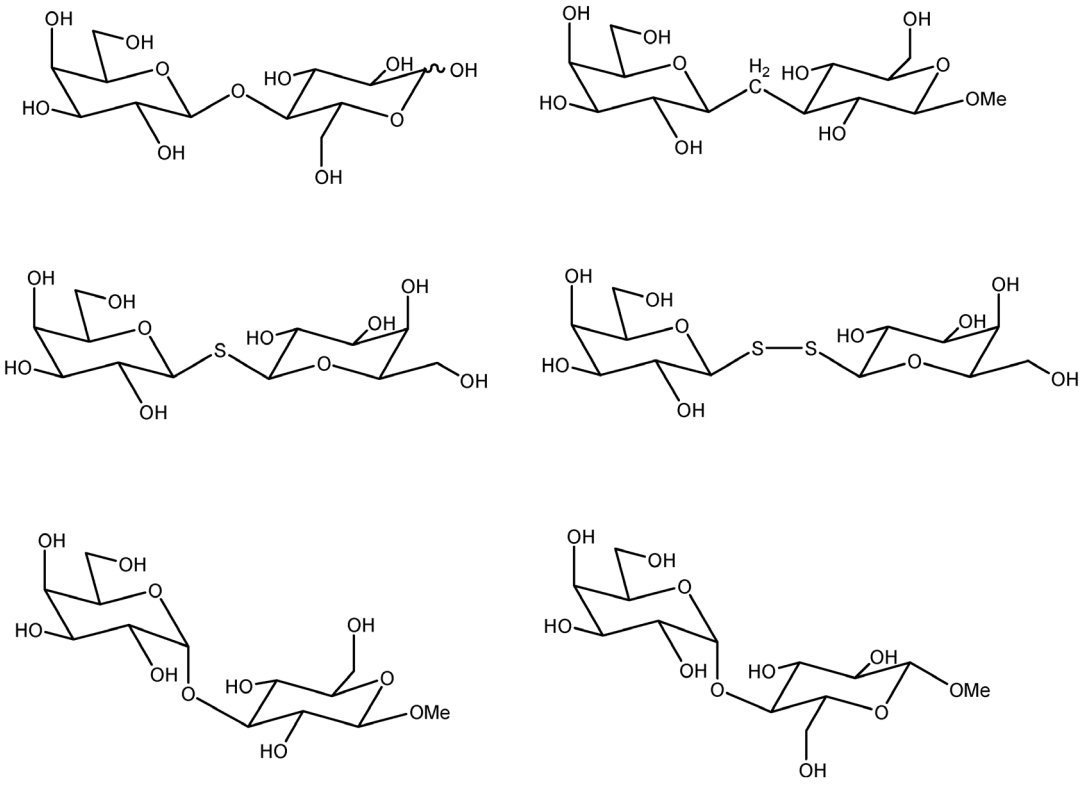


Scheme 3 Experiments and information provided by molecular modeling methods.

Predicted ligand–receptor complexes through docking protocols usually require further refinement and stability analysis. In this sense, MD simulations are indispensable as a technique to optimize the geometry, and to study the dynamics and energetics governing the stability of the predicted binding mode. MD simulations are one of the main tools in studying the structure, dynamics, and thermodynamics of biological macromolecules. In the case of proteins, their dynamics character, or personality [50], drive their functions, and MD simulations have shown to be a potent tool to investigate such processes [51]. The system is described as a collection of classical particles which interact via potentials with mainly pairwise components. Integrating Newton's equations of motion for each atom, new coordinates can be generated in a time sequence fashion. Thus, MD simulations are widely applied in the investigation of a wide range of dynamic properties and processes in the "fast" time scale, including ps–ns dynamics of side-chains and ns–μs loop and local hinge motions. In the examples herein reported, the AMBER force field, as implemented in Amber9 program [52] was selected to carry out the MD simulations. Amber is a suite of programs, extensively validated and widely used for proteins and DNA. When nonstandard residues or atoms are present, it is necessary to provide external parameters. This means that new units for the new atoms or residues must be calculated (by means of quantum chemistry methods), including charges, bonds, angles, and dihedrals. In the case of carbohydrates, parameters have been implemented in the Glycam force field [53], which has been tested widely and used successfully in simulations of the type described herein.

RESULTS AND DISCUSSION ON SELECTED EXAMPLES

The conformational behavior of different glycomimetics of lactose (Scheme 4), a well-described ligand of galectin-1 that has been studied, when bound to galectin-1 has been analyzed employing a combined NMR/molecular modeling protocol.



Scheme 4 Schematic view of the molecules mentioned in the text. From left to right and top to bottom: lactose, the *C*-glycosyl analogue of Galβ(1 → 3)Glc, GalSGal, GalSSGal, Galα(1 → 3)Gal, and Galα(1 → 4)Gal.

First, a *C*-glycosyl analogue was employed using a combined molecular modeling/ligand-based NMR approach. *C*-glycosyl mimics of natural glycosides have been extensively used as molecular probes [54]. In these analogues, the anomeric oxygen has been replaced by a methylene group, thus providing stability toward acid conditions [55]. However, it has to be considered that there are striking geometry differences between natural saccharides and their *C*-glycosyl analogues, owing to the differences in bond lengths and angles, as well as in the lack of anomeric effects for the synthetic molecules [56]. There are very few examples of the recognition mode of these molecules by lectins [57]. The existing examples have mainly focused on the lactose [58] and mannobiose and mannotriose families [59]. Thus, the knowledge of the binding features of a variety of *C*-glycosyl molecules and its comparison with that of respective *O*-glycosides is of paramount relevance to define the potential of these analogues as molecular probes from a global perspective [60]. Thus, we have focused on the molecular recognition of Galβ-*C*-(1 → 3)-Glcβ-*O*Me by human galectin-1. Previously, the conformational behavior of Galβ-*C*-(1 → 3)-Glcβ-*O*Me and its O-linked parent compound in water solution has been deduced by using a combination of NMR and molecular mechanics calculations [61]. The employed protocol involved the calculation of the potential energy maps for both compounds by using the MM3* force

field, followed by with the determination of the expected vicinal proton–proton couplings (J) and NOEs for the calculated population distribution [62]. The analysis of the experimental/NMR data for Galβ-*O*-(1 → 3)-Glcβ-*O*Me predicted the presence of a very major conformation in solution defined by the syn-Φ/syn-Ψ global minimum (A). In contrast, for the Galβ-*O*-(1 → 3)-Glcβ-*O*Me analogue, a conformational distribution of ca. 40:30:30 among the anti-Ψ minimum (B), the syn-Φ/syn-$\Psi_{(+)}$ minimum (A), and the syn-Φ/syn-$\Psi_{(-)}$ geometry (D) was deduced. Thus, A and D represent conformers with a *syn exo*-anomeric orientation for Φ, but differ in the orientation for Ψ, being either positive (*syn*-$\Psi_{(+)}$) or negative (*syn*-$\Psi_{(-)}$), respectively (Fig. 3). Thus, clear conformational differences in water solution were found between the glycoside and the glycomimetic [63].

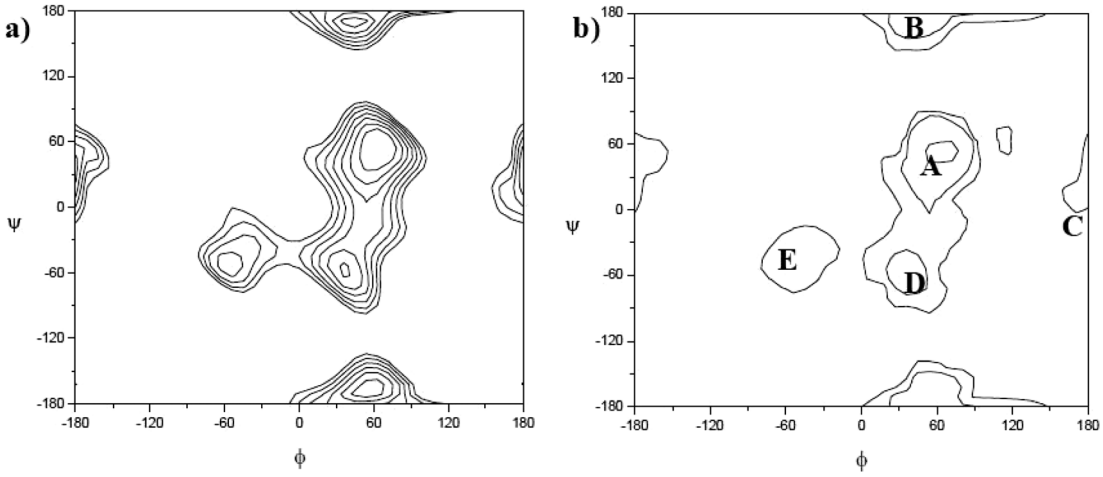


Fig. 3 Adiabatic (a) and population distribution (b) maps for the *C*-glycosyl analogue of Galβ(1 → 3)Glc. Energy contours are given every 0.5 kcal/mol. Distribution contours are given at 10, 1, and 0.1 % of the population.

For the bound state, TRNOESY experiments were employed. The obtained spectra for the free and galectin-1 bound states spectra were very similar except for the change in sign of the cross-peaks for the complex (negative) in comparison to those of the free molecular (positive), as expected for the existence of a binding event [64]. From the conformational perspective, the presence of the Gal H1'-Glc H3 cross-peak permitted the deduction that the *syn*-ΦΨ conformer is the major geometry present in the galectin-bound state. Therefore, for the *O*-glycoside, the major *syn* conformation present in solution [65] is that bound for human galectin-1. In contrast, when the analogous analysis was performed for the *C*-glycosyl mimetic, important differences with respect to the spectra obtained in solution in the absence of galectin-1 were found. The key observation involved the change in relative intensities of the H7proS-Gal H2 and H7proS-(Glc H2 + Glc H4) cross-peaks when passing from the free to the galectin-1 bound condition. The intensity ratio between these cross-peaks was very different between both states, indicating a change in the existing conformation when passing from the free to the galectin-1 bound situations. In fact, the large intensity observed for the H7proS-(Glc H2 + Glc H4) cross-peak in the bound state strongly suggested that this galectin exclusively selects the *syn*-$\psi_{(+)}$ in the bound state. Therefore, the recognition of the *C*-glycosyl mimetic by galectin-1 takes place with a conformational selection process. Only one of the three conformations coexisting in water solution is indeed selected by the lectin. The global minimum of the *C*-glycosyl mimetic, which displays an *anti*-ψ orientation, is not bound by human galectin-1. Then, in order to find out the structural requirements for the conformational selection process in these molecules, a molecular modeling approach, involving docking and MD simulations, was adopted.

First, the *O*-glycoside was docked into the galectin binding site. The best pose proposed by the docking protocol was indeed the *syn*-$\psi_{(+)}$ geometry, in agreement with the TRNOESY analysis described above. The obtained binding mode is similar to that reported by lactose. Indeed, the calculated rmsd for the *O*-glycoside with respect to the position of lactose in crystallographic structure only amounts to 0.7 Å (Fig. 2). The indol ring of Trp68 provides stacking interactions with the bottom face of the Gal ring, while His44, Asn46, and Arg48 establish H-bonds with Gal O3 and Gal O4 at the upper face. For this βGal-(1 → 3) linked disaccharide, Arg48 and Glu71 make H-bonding with Glc O4, which occupies the place of Glc O3 in lactose. Only in this *syn*-$\psi_{(+)}$ orientation is it possible to provide these stabilizing contacts. Finally, His52 also shows interactions with Gal O2 Gal and Glc O6 Glc.

Docking studies were also performed for the *C*-glycosyl analogue, to assess the observed conformational selection process. Thus, the three existing conformations in free solution (*syn*-$\psi_{(+)}$, *anti*-ψ, and *syn*-$\psi_{(-)}$) were employed as starting geometries for the docking protocol. The three conformations provided good poses in the docking solutions. Again, although the predicted binding energy values should only be considered in a first approximation, the binding energies for the *syn*-$\psi_{(+)}$ and *syn*-$\psi_{(-)}$ were very similar (within 0.5 kJ mol^{-1}) and better than that predicted for *anti*-ψ orientation, which was destabilized in ca. 3 kJ mol^{-1}. The orientation of the *syn*-$\psi_{(+)}$ conformer of the *C*-glycosyl analogue within the galectin-1 binding site is basically identical to that described above for the *O*-glycoside and lactose (Fig. 2), with a similar set of key intermolecular distances. The best docking pose obtained for the major conformation in free solution, the *anti*-ψ geometry, permitted the verification that, although the interactions with the Gal moiety are analogous to those already described, a significant part of the Glc ring is immersed in the bulk water, displaying less intermolecular interactions with galectin-1. Arg48 is now far away, precluding its interaction with the Glc moiety. Therefore, Autodock provided key structural explanations to account for the selective recognition of the *syn*-ψ+$_{(+)}$ in relation to the other two geometries. Nevertheless, MD simulations were then performed to provide a more quantitative perspective of the conformational selection process with atomic resolution. These data were validated through detailed MD simulations in water solution. For the *O*-glycoside and its *C*-glycosyl analogue in the experimentally observed *syn*-Ψ+ conformation, the MD showed a very stable structure for the protein backbone in the complex, close to 1.5 Å. Also, the Φ_H/Ψ_H glycosidic torsion angles stayed close to the starting values, with minor fluctuations around the starting values (ca. 50/20º) and the stacking interaction between the Gal ring and the indol moiety of Trp68 were maintained during the MD run. Analysis of the intermolecular interactions revealed that the complexes between *h*gal and these molecules in the *syn*-Ψ+ conformation were stable, owing to the existence of similar H-bonds and stacking interactions to those described for lactose. In contrast, more fluctuations in the protein backbone and much less protein–ligand intermolecular contacts were observed between the *C*-glycosyl analogue and galectin-1 when the MD simulation started with the *anti*-Ψ geometry. Thus, the combination of NMR and modeling procedures permitted to determine the binding mode of a glycoside and its *C*-glycosyl glycomimetic and to provide a fairly detailed explanation in structural terms for the existence of a conformational selection process in the case of the glycomimetic.

Also, two thioanalogues (thiodigalactoside, TDG, and dithiodigalactoside, DTDG) of lactose have been recently studied and their properties to bind galectin-1 have been analyzed by means of a combined NMR/molecular modeling protocol. Thioglycosides offer the advantage to be resistant to hydrolysis, so they have been proposed as chemical platforms for galectin-1 inhibitor design.

Binding assays have shown that both TDG and DTDG are able to bind galectin-1: TDG with similar binding affinity to lactose, and in the case of DTDG, with much poorer affinity. These results were corroborated from experimental STD NMR data, which confirmed a rather weak or no activity for DTDG. Guiding model building to explain the negative impact of the disulfide linkage, in contrast, the STD-based epitope mapping for TDG showed intense signals for the H4, H5, and H6 protons, a characteristic signature for galactose/galectin-1 contacts. This result, together with the available crystal structures of galectins complexes with TDG [66,67], provided input for the independent 3D model building of the TDG/*h*gal1 complex. The obtained results were also compared with the recently pub-

lished X-ray structure of this complex [67], which was not available when our modeling results were completed (PDB code 3OYW). The results permitted to assess that TDG can establish a series of interactions with the lectin using both galactose units. The binding site proposed by docking studies was also in close agreement with that experimentally found in the X-ray structure of the galectin-1/lactose complex [23]. Again, the key amino acids are Trp68, His44, Asn46, Arg48, Glu71, His52, and Arg73. This finding is important, since the experimental STD-based protocol also demonstrated, in a nonambiguous manner, that these S-containing molecules compete with lactose for the same galectin binding site. From the structural viewpoint, the complex showed a net of van der Waals and H-bond interactions, very similar to those found in the interaction of galectin-1 with lactose, with special mention to the sugar-aromatic CH–π stacking interaction between the hydrophobic face of the galactose unit and the indol moiety of Trp68. Thus, the coordinates resulting from the best docking pose were employed as starting geometry for MD simulations with explicit solvent.

The dynamic behavior of the ligand and its interactions with the receptor were scrutinized by means of MD simulations. Stable key interactions in the Gal-S-Gal/*h*gal1 complex were identified, perfectly fitting the experimental STD data. The major interactions involved one of the Gal rings that made key contacts with Trp68 and His44, and was fairly embedded in the binding pocket. Nevertheless, the second Gal moiety, although more exposed, also displayed additional contacts with polar side-chains of galectin-1. Despite the different chemical nature of Gal-S-Gal and its distinct glycosidic linkage with respect to lactose, the observed interactions resemble those present in the crystallographic structure of lactose bound to galectin-1.

The possibility of formation of intermolecular H-bonds was also checked, and different atoms were found to be involved in persistent H-bonding as acceptors (NE2 His44, OD1 Asn61, OE2 Glu71, OE1 Glu71, OD1 Asn46, NE2 His52 of the protein and most of the hydroxyl oxygens of the thiodisaccharide), and donors (HD1 His44, HD21 Asn61, HD22 Asn61, HH21 Arg48, HH22 Arg48, HH11 Arg48, HH12 Arg48, HH21 Arg73, HH22 Arg73, HH11 Arg73, HH12 Arg73, HD21 Asn46, HD22 Asn46, HD1 His52, as well as most of the hydroxyl hydrogens of the sugar). Indeed, these are basically the same that those involved in H-bonding in the crystallographic structure of the lactose/*h*gal1 complex [68]. There are some bifurcated H-bonds (i.e., both carboxylic oxygens of Glu71 with HO2, one of them also interacting with HO6'), as well as cooperative H-bonding networks (i.e., the above-mentioned bifurcated interaction of Glu71 is enhanced by the H-bond interaction of OE1 Glu71 with the guanidinium moiety of Arg73). Also, a H-bond network formed by Arg48 as donor to HO2 Gal, which in turn donates to Glu71 was also found. Analogous interactions have also been found in the recently reported crystallographic TDG/*h*gal1 complex [67], as well as in the crystallographic lactose/*h*gal1 complex [23] (Fig. 2), with HO3 Glc of lactose playing the role of HO2 Gal in TDG. These results further validate the computational methodology employed. Very tiny changes were observed between both complexes with slight variations in the orientation of the side-chains and, therefore, in the corresponding intermolecular donor–acceptor distances. In particular, the orientation of the sequence His44-Arg48, which interacts with the embedded Gal unit, is basically identical in the crystal and in the modeled structure, while there are slight differences in the residues that interact with the outer Gal unit of Gal-S-Gal (or Glc of lactose). It seems that the orientation of the side-chains of Glu71 and Arg73 fits better with the thioglycoside in the MD-based complex than within the X-ray data. In any case, the key H-bonds and CH–π interactions are indeed analogous. A better affinity of *h*gal1 for Gal-S-Gal than for lactose has been experimentally deduced (ca. 1.7-fold better), in satisfactory agreement with the associated relative free energy. Nevertheless, the assignment of this small energy difference to a particular interaction would be highly speculative.

As for TDG, the binding site for docked DTDG corresponded to that for the lactose/*h*gal1 complex, and the complex showed a net of van der Waals, H-bonds, and CH–π stacking interactions. However, whereas both galactose moieties of TDG were accommodated in the binding site, no such situation was encountered in the case of DTDG. One moiety can enter the binding site, with arrest of the

S–S bond at the θ angle of about –90º, while the second will then be completely exposed to the solvent. The lack of contact building of the second moiety of the symmetric DTDG perfectly explains the limits to the affinity. Evidently, the disulfide bonding, itself being subject to a conformational arrest, impairs the potential for interactions with galectins. According to the average structure found in the 6-ns MD simulation, there are also interactions between the ligand and the receptor, but significantly less numerous than for the Gal-S-Gal analogue. In any case, these remained fairly stable during the MD run. The intermolecular contacts only involved one of the Gal rings that make key interactions with Trp68 and His44. In this case, the second Gal moiety was exposed to the solvent and did not display additional contacts with the polar side-chains of galectin-1 (see below). The analysis of the MD simulation allowed identification of one particular water molecule mediating the interaction between Arg48 and Glu71 of galectin-1. In fact, this water molecule is occupying the position of OH2 Gal in the Gal-S-Gal/*h*gal1 complex in the H-bond network mentioned above for the Gal-S-Gal complex. The presence of the dithio moiety places the second Gal unit away from the amino acid side-chains, and the equivalent OH2 hydroxyl group cannot establish any stable intermolecular interaction with the lectin [68].

Although galectins bind to β-linked galactosides, their binding ability to α-linked galactose derivatives has also recently being evaluated [69]. In this case, in close collaboration with Prof. K. H. Mayo, the investigation was performed by using a combination of receptor-based NMR methods (^{15}N-^{1}H HSQC chemical shift perturbation analysis) and a ligand-based STD NMR epitope mapping study, combined with docking analysis and MD simulations. Since the assignment of the HSQC spectrum of galectin-1 has been achieved [37], the analysis of ^{1}H-^{15}N HSQC data in the absence and presence of different α1 → 3, α1 → 4, and α1 → 6 linked galactosides permitted to assess the location of the binding site for the different ligands. In all cases, major chemical shift perturbations were found for the signals of the amino acids at the regular lactose binding site or at its surroundings. Additionally, and from the ligand's perspective, the binding epitope of the different ligands when bound to galectin-1 was monitored by STD NMR experiments. Thus, information on the degree of involvement of the two different sugar residues of each disaccharide was deduced. In all instances, the non-reducing Gal moiety of each disaccharide was the major epitope for recognition by galectin-1 (Fig. 4).

For the α1 → 6 linked disaccharide, the STD spectra show that only the non-reducing end was recognized, since clear STD signals were only observed for this residue. For the Galα(1 → 3)Gal and Galα(1 → 4)Gal disaccharides, besides the major STD effects observed for the non-reducing end, the interaction also involved the ligands' reducing ends. These data furnished experimental input for the computational simulations, which afforded 3D pictures of the bound state. In this case, we also performed docking studies followed by MD simulations on the galectin-1 monomer bound to the three α-linked disaccharides. In all cases, the docking protocol was focused on the carbohydrate recognition site, as deduced by the HSQC experiments. The most populated clusters from the docking procedure were selected for the MD simulations, since they were also consistent with the experimental STD data. Then, MD simulations for the complexes were run for 3 ns, following an equilibration period of 100 ps. During all simulations, protein structures were stable, and the disaccharides remained at the binding site without diffusing into the solvent. Different behavior was observed for the different ligands.

For instance, for Galα(1 → 3)Gal, interconversion between conformers with different Ψ values was observed. According to the MD, these fluctuations appeared to be primarily related to transient intermolecular interactions with residues of galectin-1. Also, during the course of the simulation, it appears that, apart from the typical H-bond between His44 and Gal O4' (which defines the recognition of galactosides by galectins, as mentioned above), only one additional H-bond persisted, i.e., the one between Asn61 and Gal O6', with several other H-bonds forming transiently between both Gal residues and galectin-1. Obviously, in this and the other cases, the α-glycosidic linkage produced a drastic change of H-bond pattern at the reducing end, in comparison to that observed for lactose. For instance, as opposed to lactose Arg48 remained far from the α-linked galactose residues. Regarding CH–π interactions, those formed between the apolar face of the non-reducing galactose and the indole ring of Trp68, their frequencies of occurrence appeared to be smaller than for lactose. Similar situations took

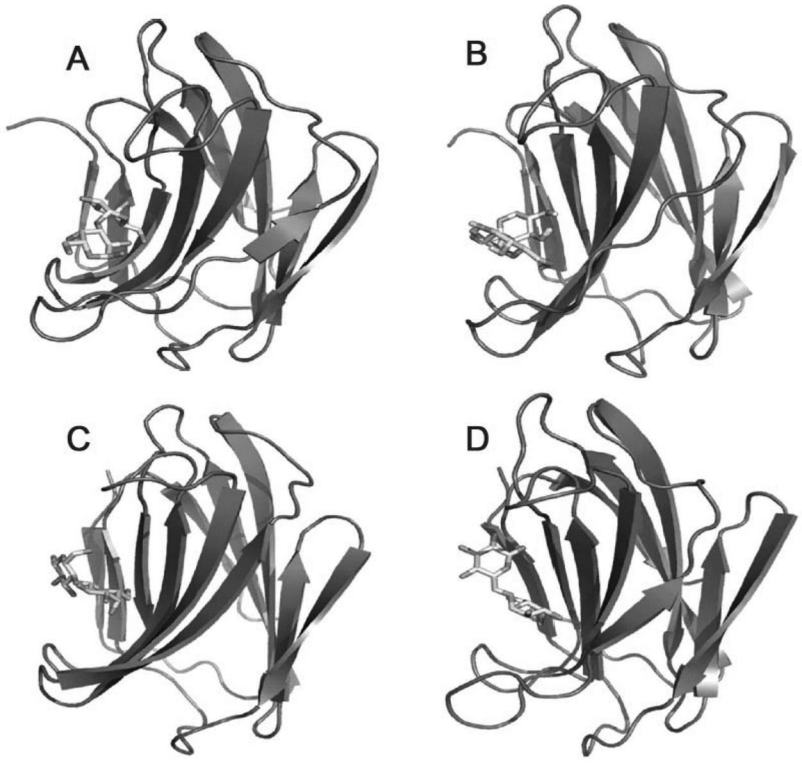


Fig. 4 Structural models of the complexes of galectin-1 with Galα(1 → 3)Gal (A), Galα(1 → 4)Gal (B), Galβ(1 → 4)Glc (C), and Gal α(1 → 6)Glc (D), based on docking analysis and the experimental input from STD NMR spectroscopy.

place for the other disaccharides. Taken together, the MD results, in accordance with the epitope mapping, define a structural model for binding of α-galactosides, also being consistent with the observed weaker binding of these disaccharides towards galectin-1, especially for Galα(1 → 4)Gal, determined experimentally, between one and two orders of magnitude, in competitive binding assays.

In summary, the combination of NMR and molecular modeling protocols provide a perfect complementary to access to detailed information, at atomic resolution, of conformational events and molecular recognition processes. It is expected that the continuous advances in both methodologies will provide scientists with more and more robust possibilities to address essential chemistry-related processes in a variety of scientific fields.

ACKNOWLEDGMENTS

We thank MICINN for financial support. Projects CTQ2009-08536 to JJB and CTQ2011-22724 to SMS. JJB and HJG also thank EU for supporting this type of research through the GlycoHit project (HEALTH-F4-2010-260600). We warmly thank illuminating discussions with Prof. K. H. Mayo (Minnesotta), Prof. Cañada, and Drs. J. F. Espinosa and P. Vidal (Madrid).

REFERENCES

1. H.-J. Gabius (Ed.). *The Sugar Code. Fundamentals of Glycosciences*, Wiley-VCH, Weinheim (2009).
2. H.-J. Gabius, S. André, J. Jiménez-Barbero, A. Romero, D. Solís. *Trends Biochem. Sci*. **36**, 298 (2011).
3. H.-J. Gabius, H. C. Siebert, S. André, J. Jiménez-Barbero, J. H. Rüdiger. *Chembiochem* **5**, 740 (2004).
4. H. Rüdiger, H.-C. Siebert, D. Solís, J. Jiménez-Barbero, A. Romero, C.-W. von der Lieth, T. Díaz-Mauriño, H.-J. Gabius. *Curr. Med. Chem*. **7**, 389 (2000).
5. V. Roldós, F. J. Cañada, J. Jiménez-Barbero. *Chembiochem.* **12**, 990 (2011).
6. H. Kogelberg, D. Solís, J. Jiménez-Barbero. *Curr. Opin. Struct. Biol.* **13**, 646 (2003).
7. V. Slynko, M. Schubert, S. Numao, M. Kowarik, M. Aebi, F. H. T. Allain. *J. Am. Chem. Soc*. **131**, 1274 (2009).
8. M. R. Wormald, A. Petrescu, Y.-L. Pao, A. Glithero, T. Elliott, R. A. Dwek. *Chem. Rev.* **102**, 371 (2002).
9. M. Salatino, D. O. Croci, G. A. Bianco, J. M. Ilarregui, M. A. Toscano, G. A. Rabinovich. *Exp. Opin. Biol. Ther*. **8**, 45 (2008).
10. D. Solís, A. Romero, M. Menéndez, J. Jiménez-Barbero. In *The Sugar Code. Fundamentals of Glycosciences*, H.-J. Gabius (Ed.), pp. 233–245, Wiley-VCH, Weinheim (2009).
11. A. A. Klyosov, Z. J. Witczak, D. Platt (Eds.). *Galectins*, John Wiley, Hoboken (2008).
12. Y. Arata, J. Hirabayashi, K. Kasai. *J. Biol. Chem*. **272**, 26669 (1997).
13. F. G. Hanisch, S. E. Baldus, T. A. Kümmel. *Glycobiology* **6**, 321 (1996).
14. M. Krzeminski, T. Singh, S. André, M. Lensch, A. M. Wu, A. M. J. J. Bonvin, H.-J. Gabius. *Biochim. Biophys. Acta* **1810**, 150 (2011).
15. L. A. Earl, S. Bi, L. G. Baum. *J. Biol. Chem*. **285**, 2232 (2010).
16. E. L. Levroney, H. C. Aguilar, J. A. Fulcher, L. Kohatsu, K. E. Pace, M. Pang, K. B. Gurney, L. G. Baum, B. Lee. *J. Immunol*. **175**, 413 (2005).
17. M. Ouellet, S. Mercier, I. Pelletier, S. Bounou, J. Roy, J. Hirabayashi, S. Sato, M. J. Tremblay. *J. Immunol*. **174**, 4120 (2005).
18. C. E. Maljaars, S. André, K. M. Halkes, H.-J. Gabius, J. P. Kamerling. *Anal. Biochem*. **378**, 190 (2008).
19. J. Jiménez-Barbero, E. Dragoni, C. Venturi, F. Nannucci, A. Ardá, M. Fontanella, S. André, F. J. Cañada, H.-J. Gabius, C. Nativi. *Chem.—Eur. J.* **15**, 10423 (2009).
20. C. M. Henry. *Chem. Eng. News* **79**, 69 (2001).
21. S. André, C. E. Maljaars, K. M. Halkes, H. J. Gabius, J. P. Kamerling. *Bioorg. Med. Chem. Lett*. **17**, 793 (2007).
22. S. André, B. Liu, H.-J. Gabius, R. Roy. *Org. Biomol. Chem.* **1**, 3909 (2003).
23. M. C. Fernandez-Alonso, F. J. Cañada, J. Jiménez-Barbero, G. Cuevas. *J. Am. Chem. Soc.* **127**, 7379 (2005).
24. M. F. López-Lucendo, D. Solís, S. André, J. Hirabayashi, K.-i. Kasai, H. Kaltner, H.-J. Gabius, A. Romero. *J. Mol. Biol.* **343**, 957 (2004).
25. M. Pellecchia, D. S. Sem, K. Wüthrich. *Nat. Rev. Drug Discov*. **1**, 211 (2002).
26. T. Diercks, J. P. Ribeiro, F. J. Cañada, S. André, J. Jiménez-Barbero, H.-J. Gabius. *Chem.—Eur. J.* **15**, 5666 (2009).
27. P. Rajagopal, E. B. Waygood, J. Reizer, M. H. Saier, R. E. Klevit. *Protein Sci.* **6**, 2624 (1997).
28. A. A. Bothner-By, J. H. Noggle. *J. Am. Chem. Soc.* **101**, 5152 (1979).
29. T. Peters, B. Meyer. *Angew. Chem., Int. Ed.* **42**, 890 (2003).
30. A. Poveda, J. Jiménez-Barbero. *Chem. Rev.* **27**, 133 (1998).

31. P. I. Kitov, J. M. Sadowska, G. Mulvey, G. D. Armstrong, H. Ling, N. Pannu, R. J. Read, D. R. Bundle. *Nature* **403**, 669 (2000).
32. V. L. Bevilacqua, D. S. Thomson, J. H. Prestegard. *Biochemistry* **29**, 5529 (1990).
33. J. P. Ribeiro, S. André, F. J. Cañada, H.-J. Gabius, A. P. Butera, R. J. Alves, J. Jiménez-Barbero. *ChemMedChem* **5**, 415 (2010).
34. M. Mayer, B. Meyer. *J. Am. Chem. Soc*. **123**, 6108 (2001).
35. K. Martinez-Mayorga, J. L. Medina-Franco, S. Mari, F. J. Canada, E. Rodriguez-Garcia, P. Vogel, H. Li, Y. Blériot, P. Sinaÿ, J. Jiménez-Barbero. *Eur. J. Org. Chem*. **20**, 4119 (2004).
36. L. P. Calle, F. J. Cañada, J. Jiménez-Barbero. *Nat. Prod. Rep.* **18**, 1118 (2011).
37. I. V. Nesmelova, M. Pang, L. G. Baum, K. H. Mayo. *J. NMR Assign.* **2**, 203 (2008).
38. I. V. Nesmelova, E. Ermakova, V. A. Daragan, M. Pang, M. Menéndez, L. Lagartera, D. Solís, L. G. Baum, K. H. Mayo. *J. Mol. Biol.* **397**, 1209 (2010).
39. A. Canales, B. López-Méndes, J. Angulo, R. Ojeda, M. Bruix, R. Lozano, G. Giménez-Gallego, M. Martín-Lomas, P. M. Nieto, J. Jiménez-Barbero. *FEBS J*. **273**, 4716 (2006).
40. A. Canales, J. Angulo, R. Ojeda, M. Bruix, R. Fayos, R. Lozano, G. Giménez-Gallego, M. Martín-Lomas, P. M. Nieto, J. Jiménez-Barbero. *J. Am. Chem. Soc.* **127**, 5778 (2005).
41. (a) A. R. Leach, B. K. Shoichet, C. E. Peishoff. *J. Med. Chem.* **49**, 5851 (2006); (b) L. Tan, J. Batista, J. Bajorath. *Chem. Biol. Drug Des.* **76**, 191 (2010); (c) C. L. Pierri, G. Parisi, V. Porcelli. *Biochim. Biophys. Acta* **1804**, 1695 (2010).
42. R. Dias, W. F. de Azevedo Jr. *Curr. Drug Targets* **9**, 1040 (2008).
43. C. B. Rao, J. Subramanian, S. D. Sharma. *Drug. Discov. Today* **14**, 394 (2009).
44. (a) T. Tuccinardi. *Comb. Chem. High Throughput Screen.* **12**, 303 (2009); (b) C. Sotriffer, R. Mannhold, H. Kubinyi, G. Folkers (Eds.). *Virtual Screening: Principles, Challenges, and Practical Guidelines*, Wiley-VCH (2011); (c) D. Plewczynski, M. Łaźniewski, R. Augustyniak, K. Ginalski. *J. Comput. Chem*. **32**, 742 (2011).
45. *BioSolveIt*, <www.biosolveit.de/FlexX/>.
46. Schrödinger. *Glide*, <www.schrodinger.com/products/14/5/>.
47. University of California, San Francisco. *Dock 6.5*, <http://dock.compbio.ucsf.edu/>.
48. Scripps Research Institute. *AutoDock*, <http://autodock.scripps.edu/>.
49. R. Huey, G. M. Morris, A. J. Olson, D. S. Goodsell. *J. Comput. Chem.* **28**, 1145 (2007).
50. (a) K. Henzler-Wildman, D. Kern. *Nature* **450**, 964 (2007); (b) D. D. Boehr, R. Nussinov, P. E. Wright. *Nat. Chem. Biol*. **5**, 789 (2009).
51. (a) E. Fadda, R. J. Woods. *Drug Discov. Today* **15**, 596 (2010); (b) H. V. Goodson, I. V. Gregoretti. *Methods Cell Biol.* **95**, 175 (2010); (c) C. Bissantz, B. Kuhn, M. Stahl. *J. Med. Chem*. **22**, 5061 (2010).
52. D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross, P. A. Kollman. AMBER 9, University of California, San Francisco (2006).
53. K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L Foley, R. J. Woods. *J. Comput. Chem*. **29**, 622 (2008).
54. J. F. Espinosa, F. J. Cañada, J. L. Asensio, H. Dietrich, M. Martin-Lomas, R. R. Schmidt, J. Jimenez-Barbero. *Angew. Chem., Int. Ed. Engl.* **35**, 303 (1996).
55. J. Jiménez-Barbero, J. F. Espinosa, J. L. Asensio, F. J. Cañada, A. Poveda. *Adv. Carbohydr. Chem. Biochem*. **56**, 235 (2000).
56. J. L. Asensio, F. J. Cañada, X. Cheng, N. Khan, D. R. Mootoo, J. Jiménez-Barbero. *Chem.—Eur. J.* **6**, 1035 (2000).
57. J. F. Espinosa, F. J. Cañada, J. L. Asensio, M. Martín-Pastor, H. Dietrich, R. R. Schmidt, M. Martin-Lomas, J. Jiménez-Barbero. *J. Am. Chem. Soc.* **118**, 10862 (1996).

58. See, for instance: A. García-Herrero, E. Montero, J. L. Muñoz, J. F. Espinosa, A. Vián, J. L. García, J. L. Asensio, F. J. Cañada, J. Jiménez-Barbero. *J. Am. Chem. Soc.* **124**, 4804 (2002).
59. L. M. Mikkelsen, M. J. Hernáiz, M. Martín-Pastor, T. Skrydstrup, J. Jiménez-Barbero. *J. Am. Chem. Soc*. **124**, 14940 (2002).
60. V. García-Aparicio, M. Sollogoub, Y. Blériot, V. Colliou, S. André, J. L. Asensio, F. J. Cañada, H.-J. Gabius, P. Sinaÿ, J. Jiménez-Barbero. *Carbohydr. Res.* **342**, 1918 (2007).
61. P. Vidal, B. Vauzeilles, Y. Blériot, M. Sollogoub, P. Sinaÿ, J. Jiménez-Barbero, J. F. Espinosa. *Carbohydr. Res.* **342**, 1910 (2007).
62. M. Martín-Pastor, J. F. Espinosa, J. L. Asensio, J. Jiménez-Barbero. *Carbohydr. Res.* **298**, 15 (1997).
63. J. L. Asensio, J. F. Espinosa, H. Dietrich, F. J. Cañada, R. R. Schmidt, M. Martin-Lomas, S. André, H. J. Gabius, J. Jiménez-Barbero. *J. Am. Chem. Soc.* **121**, 8995 (1999).
64. A. Bernardi, D. Arosio, D. Potenza, I. Sánchez-Medina, S. Mari, F. J. Cañada, J. Jiménez-Barbero. *Chem.—Eur. J.* **10**, 4395 (2004).
65. J. L. Asensio, F. J. Cañada, J. Jiménez-Barbero. *Eur. J. Biochem.* **233**, 618 (1995).
66. M. A. Bianchet, H. Ahmed, G. R. Vasta, L. M. Amzel. *Proteins* **40**, 378 (2000).
67. K. A. Stannard, P. M. Collins, K. Ito, E. M. Sullivan, S. A. Scott, E. Gabutero, I. Darren Grice, P. Low, U. J. Nilsson, H. Leffler, H. Blanchard, S. J. Ralph. *Cancer Lett*. **299**, 95 (2010).
68. S. Martín-Santamaría, S. André, E. Buzamet, R. Caraballo, G. Fernández-Cureses, M. Morando, J. P. Ribeiro, K. Ramírez-Gualito, B. de Pascual-Teresa, F. J. Cañada, M. Menéndez, O. Ramström, J. Jiménez-Barbero, D. Solís, H. J. Gabius. *Org. Biomol. Chem.* **9**, 5445 (2011).
69. M. C. Miller, J. P. Ribeiro, V. Roldós, S. Martín-Santamaría, F. J. Cañada, I. Nesmelova, S. André, M. Pang, A. Klyosov, L. G. Baum, J. Jiménez-Barbero, H. J. Gabius, K. H. Mayo. *Glycobiology* **21**, 1627 (2011).

