Carbohydrate Research 346 (2011) 939–948

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

Structural basis for ligand recognition in a mushroom lectin: solvent structure as specificity predictor

Diego F. Gauto [a], Santiago Di Lella [a], Darío A. Estrin [a], Hugo L. Monaco [c], Marcelo A. Martí [a,b,*]

[a] Departamento de Química Inorgánica, Analítica, y Química Física, INQUIMAE-CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón II, C1428EHA Ciudad de Buenos Aires, Argentina, Argentina
[b] Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón II, C1428EHA Ciudad de Buenos Aires, Argentina
[c] Biocrystallography Laboratory, Department of Biotechnology, University of Verona, Strada Le Grazie, 15, 37134 Verona, Italy

ARTICLE INFO

Article history:
Received 27 September 2010
Received in revised form 7 February 2011
Accepted 16 February 2011
Available online 23 February 2011

Keywords:
Lectin
Molecular dynamics
Solvent structure
Carbohydrate recognition domain
Affinity
Selectivity

ABSTRACT

Lectins are able to recognize specific carbohydrate structures through their carbohydrate recognition domain (CRD). The lectin from the mushroom *Agaricus bisporus* (ABL) has the remarkable ability of selectively recognizing the TF-antigen, composed of Galβ1-3GalNAc, Ser/Thr linked to proteins, specifically exposed in neoplastic tissues. Strikingly, the recently solved crystal structure of tetrameric ABL in the presence of TF-antigen and other carbohydrates showed that each monomer has two CRDs, each being able to bind specifically to different monosaccharides that differ only in the configuration of a single hydroxyl, like *N*-acetyl-D-galactosamine (GalNAc) and *N*-acetyl-D-glucosamine (GlcNAc). Understanding how lectin CRDs bind and discriminate mono and/or (poly)-saccharides is an important issue in glycobiology, with potential impact in the design of better and selective lectin inhibitors with potential therapeutic properties. In this work, and based on the unusual monosaccharide epimeric specificity of the ABL CRDs, we have performed molecular dynamics simulations of the natural (crystallographic) and inverted (changing GalNAc for GlcNAc and vice-versa) ABL–monosaccharide complexes in order to understand the selective ligand recognition properties of each CRD. We also performed a detailed analysis of the CRD local solvent structure, using previously developed methodology, and related it with the recognition mechanism. Our results provide a detailed picture of each ABL CRD specificity, allowing a better understanding of the carbohydrate selective recognition process in this particular lectin.

1. Introduction

Lectins are proteins able to recognize specific carbohydrate structures and as a consequence are involved in a wide variety of biological processes.[1–4] Found in most organisms, many plant, animal and fungal lectins have been structurally and functionally characterized.[4] The key elements of their biological role are the specificity and strength with which the carbohydrate recognition domain (CRD) binds a given saccharide.[1] The (lectin) agglutinin from the common edible mushroom *Agaricus bisporus* (ABL) is a member of the lectin family with the remarkable ability of being able to selectively recognize and bind with high affinity the Thomsen Friedenreich antigen (TF-antigen) or T-antigen (TA).[5] The TF-antigen is an important malignancy marker, Galβ1-3GalNAc, Ser/Thr linked to a protein, which is hidden in healthy cells and becomes exposed in neoplastic tissues.[3] In this context ABL has the property of inhibiting malignant cell line proliferation, without apparently affecting normal growing cells.[6] The biological mechanism of this inhibition is unclear, although it is believed to involve ABL binding to NeuAcα2-3Galβ1-3GalNAc in the oxygen regulated protein-150 (Orp 150).[7]

Interestingly, there are other fungal lectins capable of TF-antigen binding, which shows two CRDs with epimeric selectivity. For example, the *Scleroticum rolfsii* lectin has one CRD (CDRA) that also binds the TF-antigen in a similar manner as ABL, whereas the other binding site (CRDB) shows subtle differences between them, with SRL displaying an Asn at position 100 instead of the Ile which is present in ABL.[5,8]

The most striking fact of ABL, however, was revealed only recently when its crystal structure in the presence of TF-antigen, *N*-acetyl-D-galactosamine (GalNAc), and also *N*-acetyl-D-glucosamine (GlcNAc) was solved.[5] The results showed that, as expected,

Abbreviations: MD, molecular dynamics; CRD, carbohydrate recognition domain; WS, water site; WFP, water finding probability; HB, hydrogen bond; ABL, *Agaricus bisporus* lectin; GalNAc, *N*-acetyl-D-galactosamine; GlcNAc, *N*-acetyl-D-glucosamine; TA, T-antigen; TF-antigen, Thomsen Friedenreich antigen; VDW, Van der Waals; RDFS, radial distribution function.

* Corresponding author. Tel.: +54 11 4576 3380x124; fax: +54 11 5763341.
E-mail address: marcelo@qi.fcen.uba.ar (M.A. Martí).

0008-6215/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2011.02.016

each ABL monomer has a CRD (CRDA) with the ability to bind TA and also GalNAc. But strikingly, each monomer also presented a second CRD (CRDB) that specifically binds GlcNAc. The two CRDs, which are present in all monomers, seem completely independent and are able to distinguish two monosaccharides that differ only in the configuration of one hydroxyl (O4). GalNAc and GlcNAc are depicted in Figure 1.[5]

The way lectin CRDs bind and discriminate different mono and/or (poly)-saccharides is an important issue in glycobiology.[1,4,9]

Biological cells display a wide variety of potential glycoconjugate targets for lectins, and, therefore, understanding which proteins are able to specifically bind a specific target, and how ligand discrimination is achieved has important implications for the proper interpretation of the ongoing research in the field. Furthermore, understanding the molecular and structural basis of carbohydrate interaction may result in the design of better and selective lectin inhibitors with potential therapeutic properties.[10,11] Although the first approach for understanding protein–carbohydrate interaction relies on the experimental determination of protein–ligand complexes by X-ray or NMR, subtle regulation and discrimination may depend on protein and solvent dynamics which are not properly revealed by the static structure.[12] In this context Molecular Dynamics (MD) simulations have been shown to be an excellent complement for the experimental methods that allow a detailed and microscopic view on the protein–ligand interaction,[13–19] and particularly in the field of glycobiology.[20]

A secondary, but nonetheless important issue of protein–carbohydrate recognition consists in the role played by the solvent in relation to the binding process. During the ligand association process, significant solvent reorganization is produced along the contact surface. In some cases, water molecules strongly bound to the CRD must be displaced to allow proper contact between the structures, while some of them are retained bridging protein–ligand interactions.[15–17,21] The solvent role is particularly relevant for the highly hydrophilic surfaces of the CRDs, specially when compared to the hydrophobic patches typically found in the areas involved in protein–protein and protein–drug interactions.[22–26] Both the thermodynamics and kinetics of this solvent reorganization processes are complex and a clear understanding at the microscopic level has not been achieved so far. In this line of thought, many studies,[15–17,21,27] including our own[13,14] support the idea that displacement of tightly bound waters to CRDs surfaces should have a crucial effect on the binding free energy of the ligand and the study of these water thermodynamic properties yields valuable information on the carbohydrate binding structure and affinity. Specifically, using explicit water MD simulations combined with statistical mechanics analysis, we computed thermodynamic properties of selected water molecules located at several specific sites, called 'water sites' (WS), on the surface of several lectin CRDs and showed their relevance for carbohydrate recognition.[13,14] WS were defined as confined space regions close to the protein surface showing a high probability for harboring water molecules (water finding probability, WFP). Accordingly, the WS position is defined by the coordinates of the maximum probability point. Our work showed that those WS with higher WFP tend to occupy the same position of hydroxyl groups of the carbohydrate ligand in the protein–ligand complex. Therefore, by studying the properties of the WS on protein surfaces, carbohydrate binding configuration may be predicted.[13,14,28]

In this work, and based on the unusual monosaccharide epimeric specificity of the ABL CRDs we have performed MD simulations of the natural (crystallographic) and inverted (changing GalNAc for GlcNAc and vice-versa) ABL–saccharide complexes in order to understand the selective ligand recognition properties of each CRD. We also performed a detailed analysis of the CRD local solvent structure, using previously developed methodology,[13,14] and related it with the recognition mechanism. Our results provide a detailed and semi-quantitative picture of each CRD epimeric specificity allowing a better understanding of carbohydrate selective recognition process in lectins that includes solvent structural analysis.

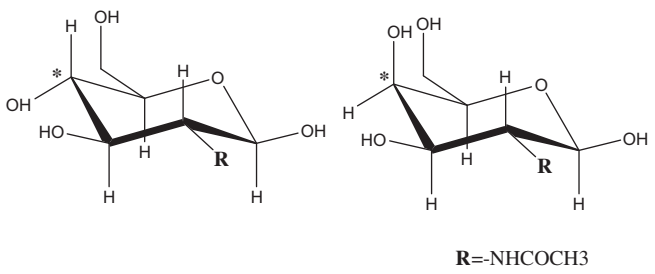

Figure 1. Schematic representation of (A) β-*N*-acetyl-D-glucosamine (β-GlcNAc) and (B) β-*N*-acetyl-D-galactosamine (β-GalNAc). The carbon atom holding the axial or equatorial O4 is depicted with an *.

2. Computational methods

2.1. Set up of the system

Starting from the crystallographic structure of GalNAc bound to ABL site A and GlcNAc bound to ABL site B (PDBID 1Y2X) -the corresponding ABL–*bi*-ligand complex was built and from now on will be referred to as ABL-GalNAc-A and ABL-GlcNAc-B state. In addition, both epimeric structures corresponding to GalNAc bound to site B (ABL-GalNAc-B) and GlcNAc bound to site A (ABL-GlcNAC-A) were constructed in silico by changing the position of the corresponding OH and H groups at position 4, yielding an inverted ABL–*bi*-ligand complex. As evidenced experimentally,[5] both CRDs are independent, meaning that binding of ligand to one of them does not affect the other. Therefore, production of two simulations with two ligands each is justified to reduce the computational cost. The site independency or oligomerization on the ligand structure and dynamics effect was further checked by simulating a dimer for one of the systems as described below.

Given that residue 72 in CRD A is a histidine, two tautomeric states corresponding to a histidine with a hydrogen either in the Nε (HIE tautomer) or the Nδ (HID tautomer) were built for the A site (ABL-GalNAc-A-HIE and ABL-GalNAc-A-HID, respectively). Finally, an ABL ligand free protein was built by removing the carbohydrate ligand from the GalNAc structure (ABL-free) with the residue 72 in the HIE tautomer. Each of these five systems was then subjected to the equilibration protocol described below and used for the production simulations.

2.2. Equilibration protocol and production simulation parameters

For each system, hydrogen atoms were added with the LEaP module of the Amber 10 package of programs.[29] Standard protonation states were assigned to titratable residues. In the case of histidine, protonation was assigned favoring the formation of hydrogen bonds in the crystal structure, except for cases in which it is expected that histidine protonation affects the CRD's characteristics; for these situations, we tested both tautomers (as described above). Each resulting construct was then set in a truncated octahedral box of TIP3P waters ($42 \times 40 \times 53$ Å). Each system was first optimized using a conjugate gradient algorithm for 2000 steps, followed by 200 ps-long constant volume MD thermalization during which the temperature of the system was slowly raised from 0 to 300 K. Heating was followed by a 200 ps-long constant temperature

and constant pressure MD simulation to equilibrate the system density. At the end of this equilibration, 20 ns long production MD simulations were performed for each case.

All amino acid parameters corresponded to the amber f99SB force field,[30] while Glycam-06 was used for the ligand parameters,[31] previously validated for carbohydrate binding.[32] Pressure and temperature were kept constant using the Berendsen barostat and thermostat,[33] respectively. All simulations were performed with periodic boundary conditions using the particle mesh Ewald (PME) summation method for long range electrostatic interactions. The SHAKE algorithm was applied to all hydrogen-containing bonds, allowing the use of a 2 fs time step.

2.3. Hydrogen bond analysis

In order to characterize the protein–ligand and/or water interactions, hydrogen bond analyses were performed. For each case, we computed the fraction of time during which hydrogen bonds, between ligand and protein atoms or water molecules (in the water sites) and protein atoms were established, as well as the mean distance between donor and acceptor atoms and the mean angles associated with each hydrogen bond. A hydrogen bond interaction was considered to be present whenever the distance r (donor–acceptor) was not larger than 3.0 Å, and angle donor-hydrogen-acceptor was not less than 120°.

2.4. Thermodynamic analysis using a continuous solvent approach

Thermodynamic parameters for all simulations were calculated using a combination of the force field molecular mechanical energies with a continuum solvent model approach as implemented in the AMBER 10 package.[34] Energetic contributions were computed corresponding to the electrostatic energy (ELE), Van der Waals contribution (VDW), and the internal energy arising from bond, angle and dihedral terms in the MM force field (INT), the sum of which yields to the total gas phase binding energy (GAS); solvation free energy (ΔG_{SV}) was estimated using the generalized-Born approximation,[33,35] which is based on the use of a cavitation and electrostatic energy components. The total solvation free energy contribution computed by the generalized Born model is also presented (GBTOT).

2.5. Water site structural and thermodynamic parameters

Water sites were defined as confined space regions close to the protein surface showing a high probability for harboring water molecules (water finding probability, WFP). The positions of the WS were defined by the coordinates of the maximum probability point using as a reference a surface residue of the protein which is able to interact favorably with the water molecules. The proposed scheme was already successfully used in our previous works[13,14] and will be only briefly described.

2.5.1. Water site identification

In order to identify the presence of water sites on the protein surface, the following methodology was used. First, we obtained radial distribution functions g(r) (RDF) for water molecules around selected potential hydrogen bonding donor/acceptor atoms in the recognition domain of each protein. These functions allow the identification of the region corresponding to the first solvation shell of the chosen atoms. Second, both translational angular g(θ) and dihedral g(φ) distribution functions were constructed taking the data of only those water molecules occupying the first solvation shell with respect to the reference atom, as defined by the peak in the g(r) functions. Visual analysis of the bidimensional plots allows clear identification of the WS, as the regions with high WFP. Using the above-mentioned analysis for each potential WS, the maximum WFP point is determined. For all subsequent calculations, this point defined the spatial center of the corresponding WS. A water molecule is defined as being inside the corresponding WS if its oxygen distance to the WS center is less than 0.6 Å, a value approximately corresponding to a volume of 1 Å^3 for the WS. In order to eliminate ambiguities in WS definition, angular bidimensional plot analysis was complemented by visual analysis of the three-dimensional grid clustering of the WFP performed by Visual Molecular Dynamics software[36] and analysis of the convergence of the computed parameters. Checking for convergence of the reported quantities (described above) was also performed using different fragments of the simulation or different references to define a given WS ensuring reproducibility in the WS definition and properties. This protocol has been successfully applied for WS identification in previous works from our group.[13,14]

2.5.2. Calculation of structural, dynamic and thermodynamic properties for the WS

In order to compute the potential energy associated with the interaction of water molecules in the WS with the protein and the rest of the solvent, for each snapshot along the whole simulation, the VDW and electrostatic interactions were computed. Contributions were calculated between the water located inside the WS and either the protein (E_p) or the other solvent molecules (E_w). Contributions were considered up to a distance of 8 Å to the WS. This cutoff has already been shown to yield reasonably convergent results.[13] For each WS, the mean interaction energies $\langle E_x \rangle$ were computed over the whole simulation. Total mean interaction energies $\langle E_t \rangle$ of a water molecule inside the WS were then computed. Using as a reference the interaction energy of a water molecule in bulk water ($\langle E_{wat\text{-}bulk} \rangle$), the differential mean interaction energy was computed as $\langle \Delta E \rangle = \langle E_t \rangle - \langle E_{wat\text{-}bulk} \rangle$. This difference in energy corresponds to the gain in potential energy of transferring a water molecule from the bulk solvent to the corresponding WS. For each WS, we also computed the WFP which is the probability of finding a water molecule inside the WS, and then normalized it with respect to that of the bulk water that corresponds to the water density at the corresponding temperature and pressure values. The values were computed using an arbitrary volume of 1 Å^3 for each WS. As a measure of the dispersion of each WS, we computed the radius of the volume that harbors 90% of the time a water molecule inside the corresponding WS. Therefore, the computed R90 results are a measurement of the size of the WFP in volume units for each WS. The volume of the WS also provides a measurement of its dispersion and it is consequently associated to the translational entropy, as shown by Lazaridis et al.[15] The greater the volume (larger R90) the more conformational freedom have the water molecules inside the WS and, therefore, they are expected to have larger entropy.

Finally, to analyze the correlation between the WS position with respect to the protein surface and the structure of the protein–ligand complex, we calculated the so-called R_{min} values. For this purpose, WS positions were superimposed on the protein–ligand complex crystallographic structure. R_{min} values were then computed as the distance of the WS position to the nearest heavy atom of the ligand in the superimposed structures. This procedure was performed with the help of the Visual Molecular Dynamics software of the University of Illinois,[36] and has been shown to provide significant results in ours previous works.[13,14]

2.6. Oligomerization effects analysis

In order to analyze possible effects of the oligomerization state on the presented results, we also built a dimer from the initial tetrameric structure (PDBid 1Y2X) complexed with GlcNAc and

GalNAc in the corresponding CRDs, as a test case and simulated it for 5 ns to MD in the same conditions described above for all monomeric systems. We then computed the RMSD of the ligand heavy atoms using as a reference the average structure from the 20 ns production simulation of the corresponding monomer. For all cases the resulting RMSD values were less than 0.6 Å (see SF6), clearly showing that no effects are observed in the ligand position or dynamics due to simulation artifacts. Moreover we also performed an H-bonds analysis (see Table S5A and B) for the dimer simulation and again the results are the same as those found for the monomer simulation presented in the results.

3. Results

The results are analyzed separately for each ABL CRD. For each case, a structural, hydrogen bond (HB), energetic and solvent structure analysis was performed and related to the binding specificity. The overall protein dynamics for all simulations is stable and proper equilibration is achieved in about 20 ns as evidenced by the RMSD versus time plot data shown in SM. All the computed values referred below correspond to the equilibrated segment of the simulation, unless otherwise stated.

3.1. Structural and dynamic analysis of ABL CRD A

The CRDA of ABL is mainly determined by residues Arg107, Tyr98, Tyr28, Gly49, Ser48, Tyr74, Asn73 and His72. MD simulations of the natural ligand GalNAc in the A site shows, as expected, that the ligand remains stable and bound to the protein during the whole simulation time (the RMSD of the average structure to the carbon atoms of the ligand to the initial X-ray derived structure is 1.2 Å), specially when His72 is simulated in the HIE tautomeric state (histidine protonated in the Nε). For the HID state (histidine protonated in the Nδ) the ligand seems to bind less tightly as evidenced by the higher mobility, and weaker interactions as shown below. A representative structure of GalNAc bound to ABL CRDA is presented in Figure 2A, and the corresponding HB analysis in Table 1.

In the HIE state, a very strong HB (>90% occupancy) is established between Ser48OH and *N*-acetyl carbonyl (GalNAcO7). Strong HBs are also present between the backbone NH of Asn73 and GalNAcO7, and between the backbone carbonyl of Gly49 and the axial GalNAc hydroxyl (GalNAcO4). Finally, a weak HB, presenting only less than 10% occupancy can be established between Asn73OD and GalNAc O1. When His72 is changed for the HID tautomer, the ligand moves and now the three main HBs are weaker. New transient HBs, however, are established between Ser48-O, Gly49-O, and GalNAcO3 and O4. These results confirm that stronger interactions are found when His72 is in the HIE tautomeric state for the natural GalNAc ligand. As expected, results (particularly for the HIE tautomer) are consistent with those observed in the X-ray structure in which the main predicted HBs are established between Asn73-N and Ser48-OG with GalNAcO7 and the Gly 49 carbonyl with the O3 and O4 from the GalNAc.

If the ligand is replaced by GlcNAc, as shown in Figure 2B and Table 1, some interactions are preserved while, as expected, several interactions, especially the equatorial O4-Gly49-O are significantly weakened. Visual inspection of the dynamics, interestingly shows that the GlcNAc ligand is bound only temporarily to the CRD A site, and is released during the simulation in about 10–15 ns, as evidently shown in Figure 3. The GlcNAc C atom RMSD versus time plot, using the initial structure as reference, clearly shows that ligand remains bound only about 5 ns, and starts dissociating between 5 and 10 ns after which it is completely released. The HB analysis, performed for the initial part of the simulation where the ligand is still in the CRD (during the first 10 ns) shows that for the HIE system, strong HBs between Ser48OH and GlcNAcO7, and Asn73NH and GlcNAcO7, are preserved with similar occupancies and structural parameters as for the natural ligand. However, the Gly49 interaction with the equatorial GlcNAcO4 is completely lost. When His72 is simulated in the HID state, strong interactions with Ser48 and Asn73 are completely lost, temporary HBs are established only between Gly49 and Ser48 backbone carbonyls and GlcNAcO3 and O4. Clearly the ABL CRD site A can only bind GalNAc and not its epimer GlcNAc.

To gain further insight into the ligand binding and discrimination mechanism we performed a thermodynamic analysis as mentioned in Section 2 for the binding of GalNAc to the CRDA site with His72 in both tautomeric states. Given that no stable CRDA site GlcNAc complex was obtained the same analysis could not be performed for the inverted ligand. The results, presented in Table 2, confirm the slight preference observed for His72 protonated in the ε position (HIE). Interestingly, the results show that GalNAc electrostatic and VDW binding energy contributions are of the same magnitude. Finally, using the continuum solvent method we computed the contribution of the OH to the binding energy. The results show that for GalNAc-HIE, the axial OH contributes with a favorable energy of 7.73 kcal/mol and for GalNAc-HID it contributes with a favorable energy of 6.41 kcal/mol. These values correspond to about one third of the total electrostatic contribution, showing the key role played by this moiety in modulating the protein–ligand interaction.

3.2. Structural and dynamic analysis of the ABL B site

The ABL B site, whose natural ligand is GlcNAc, is formed by residues Tyr114; Asp79; Ile80; Thr82; Asn83; Val81; Arg103; Ile102 and Tyr 97. The MD simulation with the natural GlcNAc ligand

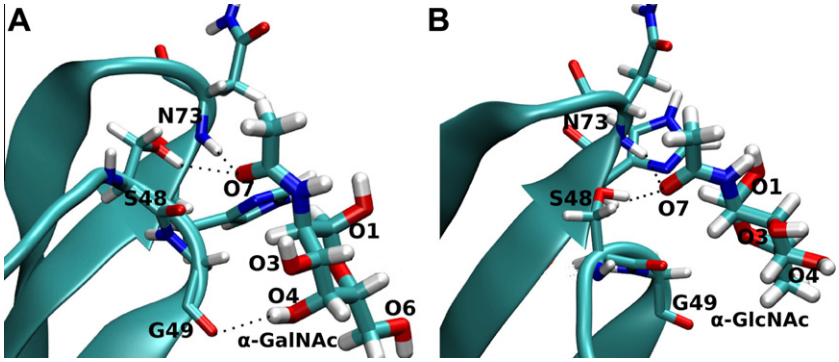

Figure 2. Representative structures of (A) α-*N*-acetyl-D-galactosamine (α-GalNAc) and B) α-*N*-acetyl-D-glucosamine (α-GlcNAc) bound to ABL HIE CRD A.

Table 1
H-bonds percent occupancies to CRDA

Acceptor	Donor	GalNAc-HIE	GalNAc-HID	GlcNAc-HIE[a]	GlcNAc-HID[a]
GalNAc/GlcNAc O7	Ser 48 HO	91.6	35.7	97.6	<1.0
GalNAc/GlcNAc O7	Asn 73 HN	50.0	35.4	35.0	<1.0
Gly 49 O	GalNAc HO4	42.7	13.1	<1,0	33.1
Asn 73 OD1	GalNAc/GlcNAc HO1	6.5	<1.0	18.2	<1.0
Ser 48 O	GalNAc HO4	<1.0	6.6	<1.0	6.8

GalNAc: *N*-acetyl-D-galactosamine; GlcNAc: *N*-acetyl-D-glucosamine; HIE: ε-Histidine; HID: δ-Histidine.
[a] For GlcNAc HB occupancies are computed during the initial 10 ns of the simulation when the ligand remains bound in the CRD.

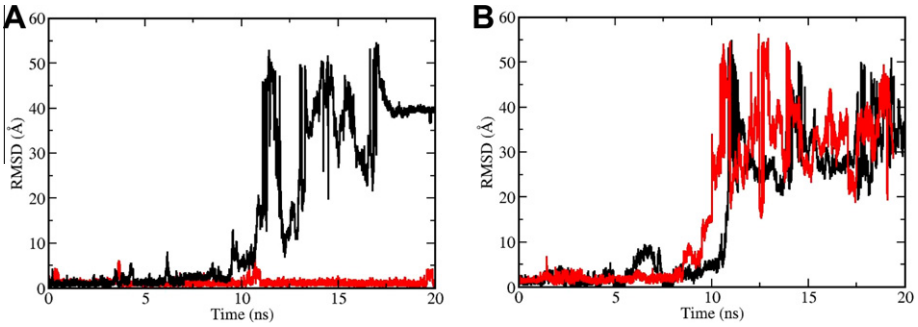

Figure 3. RMSD of the carbon atoms the carbohydrate (α-GalNAc in red and α-GlcNAc in black) in the CRD A site versus the MD simulation time. Different tautomeric states of HIS72 in the CRD A were tested: (A) HIE (Nε protonated) CRD A and (B) HID (Nδ protonated). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
MM-GBSA average values for the CRDA–GalNAc interaction

	CRDA: GalNAc-HIE	CRDA: GalNAc-HID
ELE	−19.1	−18.9
VDW	−18.3	−17.3
GAS	−37.4	−36.1
Δ*G* SV	26.9	27.0
GBTOT	−10.5	−9.5

GalNAc: *N*-acetyl-D-galactosamine.

Table 3
H-bonds percent occupancies to CRDB

Acceptor	Donor	GlcNAc	GalNAc
Asp 79 OD2	GlcNAc/GalNAc HO4	99.9	21.8
Asp 79 OD1	GalNAc/GalNAc HO4	<1.0	91.0
Asp 79 OD1	GlcNAc/GalNAc HO6	99.4	77.2
GlcNAc/GalNAcO7	Thr 82 NH	79.0	69.6
GlcNAc/GalNAcO3	Tyr 114 HO	72.5	93.6

GalNAc: *N*-acetyl-D-galactosamine; GlcNAc: *N*-acetyl-D-glucosamine.

shows that there are many H-bonds formed between protein and ligand that stabilize the ligand on CRDB. Moreover, the ligand position and orientation do not differ significantly from that of the crystal structure (the RMSD of the average structure to the carbon atoms of the ligand against the starting X-ray derived structure is 0.61 Å) suggesting that the interactions are very stable. The representative structure of the complex depicted in Figure 4A and HB analysis in Table 3, showed that GlcNAc binds to ABL CRD B, establishing several tight interactions. Asp79 forms two very strong (>98% occupancy) HBs with the equatorial GlcNAcO4 and GlcNAcO6. Thr82NH establish a strongly HB with acetyl carbonyl GlcNAcO7, while Tyr114 is strongly HB to GlcNAcO3, and interacts also transiently with the equatorial O4. When the ligand is changed to GalNAc, the interactions remain similar, but some HBs show lower occupancies, specially the Asp79 interactions with the GalNAc axial O4 and the O6 and Thr 82 NH with the O7 interaction. Also, noteworthy is the fact arising from the dynamics of Asp79. When the ligand is GlcNAc the Asp79 carboxylate is fixed and only

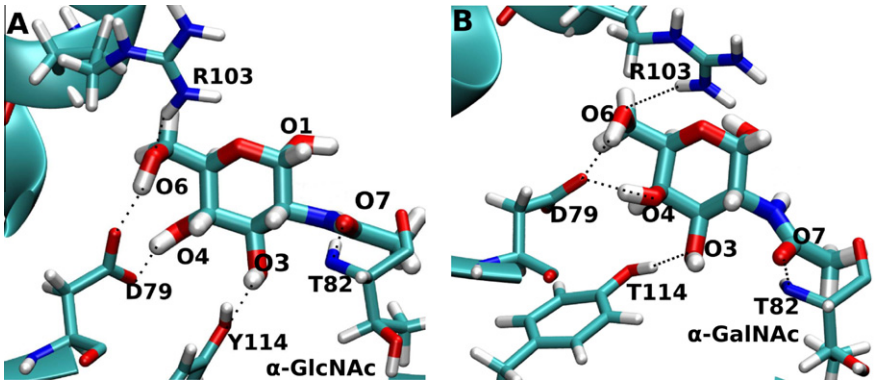

Figure 4. Representative structure of (A) α-*N*-acetyl-D-glucosamine (α-GlcNAc) and (B) α-*N*-acetyl-D-galactosamine (α-GalNAc) bound to ABL CRD B site.

one oxygen atom is HB to the OH4 during the whole simulation time. On the other hand, when GalNAc is the ligand the Asp79 carboxylate rotates alternating the HB between both carboxyl oxygens and the axial OH (GalNAcO4).

As for the CRD A site we computed the energetic contributions to ligand binding to the CRD B site for GalNAc and GlcNAc using the continuum solvent method as we can see in Table 4. The results show a clear preference for the natural ligand (about 8.0 kcal/mol). Analysis of the individual contributions shows that the preference lies in the electrostatic contribution, consistent with the analysis of the HB data. We also computed the contribution of the OH4 to the binding energy. The results show that for GlcNAc, the equatorial OH contributes with a favorable energy of 7.4 kcal/mol, while in GalNAc, the axial OH contributes only about that of 5.4 kcal/mol, showing, as expected, that the OH4 contributes to the specificity.

Finally, in order to have an estimation of the entropy change due to the ligand binding process, we computed the protein conformational entropy in the unbound and bound states for the natural and inverted ligands. The results show that the presence of both natural ligands significantly reduces protein mobility, which results in an about 78.0 kcal/mol loss in entropy, at 300 K. This is consistent with previous data from our group which show that carbohydrate ligands usually reduce mobility of the protein for several Galectins.[37,38] On the other hand, the change in entropy when both the inverted ligands are bound is close to zero, consistent with the fact that they are only loosely bound to the protein, and, therefore, have a minor impact in protein dynamics.

3.3. Solvent structure and WS of ABL CRD A

As mentioned in the introduction, previous works from our group show that analysis of the solvent structure in the free protein can be a good predictor of protein–ligand complex structure characteristics.[13,14] The analysis is based on the definition and thermodynamic characterization of high WFP regions close to the protein surface denoted as WS. Given the subtle epimeric preferences of the ABL CRD A and B, it is interesting to see whether solvent occupancy could predict or help to explain the observed differences. To analyze the solvent structure we defined the possible WS of both CRDs and computed the WS structural and thermodynamic properties as in our previous works.[13,14] We were able to identify 5 WS for site A and 4 WS for site B (see below).

As expected, the coordinates of the WS described in the CRD and the position of water molecules in the original X-ray structure are in good agreement. For example, in CRD A, WS1 is very close to CW2 and WS2 is very close to CW1 (see SF 7A). In the case of CRD B, WS3 almost superimposes to CW1 (see SF 7B). The other identified WS are not present in the x-ray structure.

We start the analysis by looking at the solvent structure of ABL CRD A. Figure 5A shows the radial probability distribution function (RDF) of water oxygen atoms with respect to Gly49-O, while Figure 5B shows the corresponding bidimensional angular plot and Figure 5C the identified WS location on the CRD A site around the epimeric zone. These results show that Gly49O has a well defined first solvation shell and analysis of the water finding probability densities allows defining three WS. It should be noted that although WS2 in the bidimensional plot seems poorly defined and located between WS1 and WS3 it is extremely relevant as will be shown below. The resulting structural and thermodynamic parameters for the corresponding WS are shown in Table 5.

Table 4
MM-GBSA average values for the CRDB–GlcNAc/GalNAc interaction

	ABLB GlcNAc	ABLB GalNAc
ELE	−75.3	−66.3
VDW	−15.2	−16.7
GAS	−90.5	−83.0
ΔG SV	61.2	61.9
GBTOT	−29.1	−21.1

GalNAc: *N*-acetyl-D-galactosamine; GlcNAc: *N*-acetyl-D-glucosamine.

The results show that although WS1 is far from any saccharide atom in the complex, both WS2 and WS3 are close to the axial O4 position. WS3 is closest to axial O4 in the natural ligand, forming the same HB pattern and is displaced when GalNAc binds. For the inverted ligand the equatorial O4 position is further from the WS3 (and also WS2) predicted location, suggesting that the WS position analysis reflects epimeric preference. WS2 is not displaced by any of the ligands but as will be shown below is extremely important for epimeric specificity. WS4 and WS5 are very close, and possibly replaced by atoms N7 and O7 of the ligand. This is in agreement with the fact that the *N*-acetyl group is the main anchoring point for the carbohydrate ligand and with the results from our previous work which show that the WS are able to predict the location of the ligand hydrophilic groups.[8,13,14]

To further relate the solvent (WS) structure and the epimeric preference we computed the radial and bidimensional plots using Gly49 as a reference for the corresponding HO4 atom in both the ABL GalNAc, and GlcNAc complexes with its histidine on CRD protonated in the Nε (for GlcNAc the first part of the simulation was analyzed before the ligand detached from the protein). The corresponding RDFs are shown in Supplementary data—Figure SF 4—while the bidimensional plots are shown in Figure 6. The data show that for GalNAc the RDF shows the first solvation shell peak at 3 Å the same distance as the water solvation shell, on the other hand for GlcNAc the RDF shows the peak at 5 Å (and is very small), clearly showing that the equatorial OH cannot come close enough to interact strongly with the carbonyl. Furthermore comparison of the bidimensional plots of Figures 6A and B with 5B clearly shows that the axial OH4 of GalNAc overlaps well with WS3, but equatorial OH4 of GlcNAc does not, again demonstrating the predictive role of the WS analysis.

As mentioned above, WS2 is not displaced by the ligands. Interestingly, analysis of the ligand bound (GalNAc and GlcNAc) simulations shows the presence of a strong water bridge between the protein and the ligand. The corresponding bridging water, perfectly corresponds with WS2, and is HB to His72-Nδ and to the axial O4 in the natural ligand. To comparatively analyze the structure of this WS in the free, GalNAc and GlcNAc cases we computed the RDF (Fig. SF 6 and bidimensional using His72-Nδ as the reference atom, as shown in Fig. 7). The results clearly show an RDF peak and high WFP zone in the unbound CRD, which becomes more localized when GalNAc is bound, but almost completely disappears for the GlcNAc bound system. The increased stability of water inside WS2 when GalNAc is bound is also evidenced in the WFP that rises three times to a value of 12.0. Finally, in order to determine the energetic contribution of the bridging water to ligand binding we computed the GalNAc interaction energy with the generalized Born model approach but including water as part of the analyzed system. The results presented in Table 6 show that the bridging water molecule contribution raises the GalNAc electrostatic interaction energy in 6.5 kcal/mol and the total predicted binding energy including solvation by 2.5 kcal/mol which further increases the differences observed in Table 2.

3.4. Solvent structure and WS of ABL CRD B

Similarly, we can analyze the solvent structure around the CRD B site using as reference the Asp79 carboxylate group that interacts with the OH4 and the Thr 82 N atom that interacts with the *N*-acetyl group from the carbohydrate ligand. Figure 8A shows the RDF for

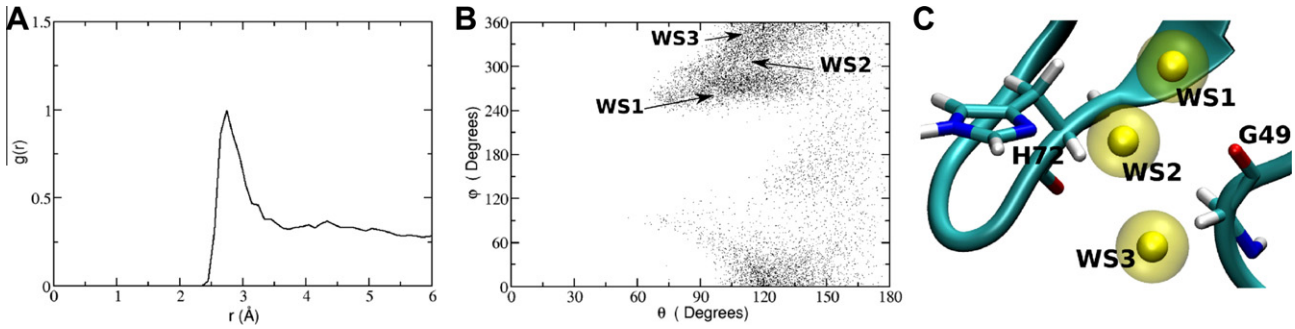

Figure 5. (A) g(*r*) plot between Gly49O and solvent O. (B) Bidimensional plot distribution of WFP for O solvent around Gly49O. (C) WS on CRD A site around the zone determining specificity.

Table 5
Calculated values for the interaction energies, WFP and R_{min} for the WS previously defined in Section 2

Reference	WS	E_p	E_w	E_t	ΔE	WFP	GalNAc		GlcNAc	
CRDA							R_{min}	Atom	R_{min}	Atom
Gly 49 O	WS1	−17.0	−3.3	−20.3	−2.9	8.2	—	—	—	—
Gly 49 O	WS2	−15.8	−4.8	−20.6	−3.2	4.5	2.35	O4	>5.0	O4
Gly 49 O	WS3	−15.7	−4.5	−20.2	−2.8	5.5	1.9	O4	3.16	O4
							1.3	O3	1.9	O3
ASN 73 N	WS4	−7.1	−10.4	−17.5	−0.1	5.5	1.4	N7	1.4	N7
SER48 OG	WS5	−6.4	−13.0	−19.4	−2.0	5.3	0.7	O7	0.7	N7
CRDB							R_{min}	Atom	R_{min}	Atom
Asp 79 CG	WS1	−19.4	−3.4	−22.8	−5.4	11.0	0.65	C4	0.70	C4
Asp 79 CG	WS2	−14.0	−4.4	−18.4	−1.0	4.8	3.6	O4	1.77	O4
Asp 79 CG	WS3	−12.4	−2.1	−14.5	2.9	4.9	—	—	—	—
Thr 82 N	WS4	−6.9	−11.6	−18.5	−1.1	11.0	0.2	O7	0.7	O7

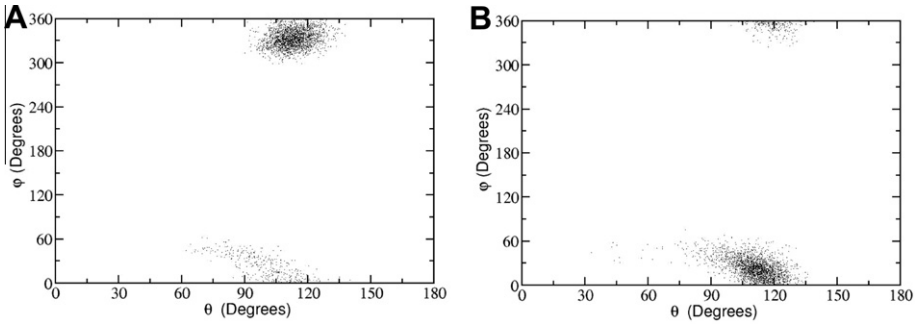

Figure 6. (A) Bidimensional plot of the axial oxygen of α-*N*-acetyl-D-galactosamine (α-GalNAc) against Gly 49 O. (B) Bidimensional plot of the equatorial of α-*N*-acetyl-D-glucosamine (α-GlcNAc) against Gly49O.

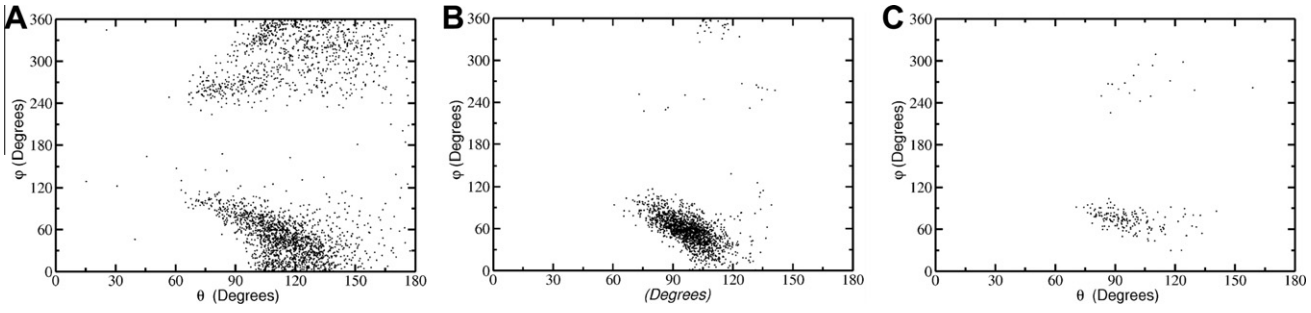

Figure 7. Bidimensional plots of angular distribution for oxygen of the water molecules around HIS72 Nδ. (A) First solvation shell structure of the free protein. (B) First solvation shell structure of normal complex. (C) First solvation shell structure of inverted complex.

the AspCG, the corresponding bidimensional plot and the position of the three identified WS. The thermodynamic and structural parameters for the corresponding WS are shown in Table 5. The results show that WS1 is close and will be displaced by C4 on both substrates. WS3 is far from any ligand atom and is not related to the complex. WS2 is the best epimeric specificity predictor since

Table 6
MM-GBSA average values considering the interaction of the specific water molecule bridging ligand bound to the CRD

Energy	With H_2O (kcal/mol)	Without H_2O (kcal/mol)	ΔE (kcal/mol)
Elect	−25.5	−19.1	−6.5
VDW	−17.8	−18.3	0.5
Gas	−43.3	−37.4	−5.9
ΔG SV	29.7	26.9	2.8
GBtotal	−13.0	−10.5	−2.5

it is displaced by the natural GlcNAc ligand, but does not match the OH4 in the inverted case. Finally, WS4 is perfectly replaced by the O7 from both ligands (GlcNac and GalNac) confirming again the predictive power of the WS.

Again to further compare the solvent structure with the ligand bound structure we computed the RDFs and bidimensional plots for the OH4 in the GlcNAc and GalNAc complexes. The results presented in SF5 and Figure 8 show that the peak position of the OH4 is for both ligands around 3 Å, similar to the distance of the first solvation shell, consistent with the fact that the Asp79 carboxylate is able to interact with O4 in both epimers. The bidimensional plots depicted in Figure 9, however, show significant differences. While the mean angular position of the natural ligand GlcNAc O4 is in between the positions of WS1 and WS2 (closer to WS2—see Fig. 8), the position of GalNAcO4 (inverted ligand) is only slightly superimposed to the border of WS1. In summary, analysis of the WS shows consistent with our previous work that solvent structure is a good predictor of the ligand bound structure, moreover the analysis also reveals a slight preference for GlcNAc over GalNAc.

4. Discussion

The two ABL CRDs offer the unique possibility of studying how each of them is able to distinguish two very similar monosaccharides that differ only in the configuration of hydroxyl group 4 GalNAc and GlcNAc. Using MD simulations of the natural and inverted (where the ligand is changed for its epimer in silico) complexes we have determined the structural reasons that define epimeric specificity and have measured in an approximate but quantitative way the epimeric preference. Our simulations provide an extension of our previous crystallographic work[5] allowing a detailed analysis of (a) Each particular interaction and their relative contribution to the binding energy that determine the epimeric specificity, (b) a semi-quantitative prediction of the degree of epimeric specificity of each site, and (c) an analysis of the solvent structure, in the uncomplexed protein, showing how solvent structure is related to protein–ligand complex and could, therefore, be used as a specificity predictor. A discussion of the results in the general context of protein–carbohydrate binding is presented below.

4.1. Preference of the ABL CRD A site for GalNAc

The ABL CRD A site binds preferentially GalNAc over GlcNAc, in fact our MD data show that it is completely unable to bind GlcNAc. The relevant interactions for holding the saccharide in place are formed between the *N*-acetyl carbonyl of the ligand and Ser48 and Asn73, plus the specific interactions with the HO4 which is HB to the Gly49 carbonyl and a water bridge to His72. The water bridge is crucial for determining the His72 tautomeric state to HIE type. Since only three to four strong HB interactions are established between protein and carbohydrate changing the position of O4 as in GlcNAc results in the loss of 1/2 to 2/3 of the HB interactions which forbids binding of GlcNAc, determining epimeric specificity. This is consistent with the experimental evidence which shows that only the GalNAc electron density is present at this site, while galactose, or GlcNAc is not. The lack of binding to galactose clearly highlights the importance of the *N*-acetyl HB.

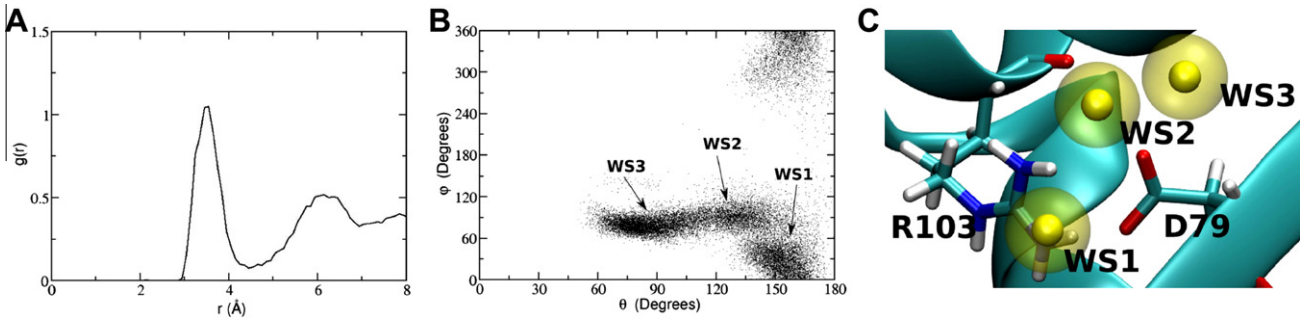

Figure 8. (A) g(*r*) plot between Asp 79 CG and O by waters. (B) Bidimensional plot distribution of WFP for water molecules around Asp 79 CG as reference. (C) WS on CRD B site around zone determining specificity.

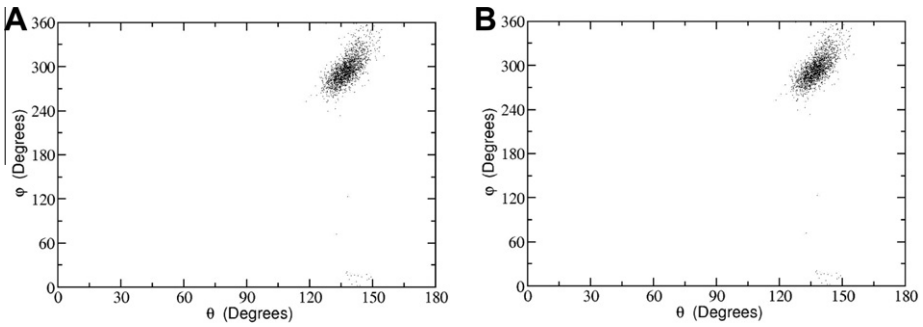

Figure 9. (A) Bidimensional plot of the equatorial oxygen O4 of α-*N*-acetyl-D-glucosamine (α-GlcNAc) atom with Asp 79 CG as reference. (B) Bidimensional plot of the axial O4 of α-*N*-acetyl-D-galactosamine (α-GalNAc) atom with Asp 79 CG as reference.

4.2. Preference of the ABL CRD B site for GlcNAc

Compared to CRD A, the epimeric preference in CRD B seems much less pronounced. In CRD B stable MD simulations were obtained with both the natural GlcNAc, and also the inverted GalNAc ligand. Overall HB analysis shows that for the natural ligand 4 very strong HBs are present that involve monosaccharide hydroxyls O3, O4, O6 and carbonyl O from the *N*-acetyl group. The change from GlcNAc to GalNAc results in a shift in the saccharide relative position mainly due to the loss of the Asp79 carboxyl with the OH4 interaction that concomitantly weakens the other interactions. The predicted difference in binding free energy is 8 kcal/mol corresponding to about 1/4 of the natural ligand interaction energy. The results suggest that CRD B may bind GlcNAc but also GalNAc with lower affinity. The X-ray data, however, showed no electron density in the CRD B when the crystals were soaked with GalNAc. This could be partially explained by the fact that in these ABL crystals only the CRD A site is freely accessible, the CRD B being partially blocked by crystal packing as evidenced by crystal breakdown when soaking is performed with the CRDB preferred ligand GlcNAc. In fact, the crystals of the complex ABL-GlcNAc were obtained later and belong to a different crystal form in which the CRD B site is not blocked. To test whether CRD B is able to bind GalNAc and further assess its epimeric specificity, future experiments are required. An interesting possibility suggested by the present analysis consists in mutating the CRD A site key residues (His72 and Ser48) to inhibit GalNAc binding to CRD A and then evaluate GalNAc binding to these mutants.

4.3. Comparison between CRD A and CRD B

The main difference between CRD A and CRD B is the presence of the acidic group of Asp79 in CRD B. The negatively charged group makes two very strong HBs which results in higher electrostatic interaction energy. This high interaction energy although partially compensated by the solvation free energy results in higher predicted interaction energy in CRD B compared to CRD A. To confirm this, experimental measurements of the ligand dissociation constants or binding free energy are needed. The VDW contribution to the binding energy is similar and accounts for about −15 kcal/mol in both cases. The epimeric selectivity difference is higher for CRD A for two reasons; first in CRD A the HO4 makes two HB contacts in the natural ligand which are lost for the other epimer. In CRD B only one HB is present between HO4 and the protein. Second, the total number of HBs is smaller for CRD A, and, therefore, losing one interaction has a bigger relative effect. It should be also noted that although in many lectins, like Gal-1,[38] the non polar face of the carbohydrates interacts with aromatic residues of the protein,[39] establishing a π-stacking interaction, no such interaction is present in any of the two ABL binding sites.

4.4. Relation of solvent structure with epimeric preference and carbohydrate binding to lectins

As described in the introduction previous data from our group showed that solvent structure in the ligand free protein mimics some properties of the protein–carbohydrate complexes, since high WFP regions in the free protein tend to be occupied by carbohydrate hydroxyl groups.[13,14] Epimeric discrimination relies on subtle structural differences and, therefore, represents a difficult test case to examine how well solvent structure may predict protein–carbohydrate complex conformation. Our results show that analysis of solvent structures as characterized by the presence of WS may to some extent predict or at least explain the epimeric preference. For CRD A WS with about five times more WFP than bulk solvent (WS3) is positioned (as evidenced by the bidimensional plot) where the OH4 of GalNAc but not GlcNAc is located in the complex. Furthermore, the WS2 bridge as the GalNAc OH4, significantly contributing to the binding but is lost when GlcNAc is forced to bind. For CRD B the solvent contribution is less clear, since the highest WFP WS is positioned closer to the C4 of the saccharide, however, analysis of the WS2 position shows that it will be displaced by the ligand GlcNAc equatorial OH4, and is far from that of GalNAc, and, therefore, careful analysis of the structure reveals epimeric selectivity. In summary, our results support our hypothesis that solvent structure analysis allows to predict characteristics of protein–carbohydrate complexes, and may even be able to determine possible epimeric selectivity.

4.5. General Implication for carbohydrate selectivity and recognition

From a general viewpoint, the results presented here show how difficult and subtle saccharide selective recognition is. The results show that only for CRD A where two of four of the natural ligand HB interactions are lost, the protein completely discriminates the inverted ligand that is not able to remain bound. This is not the case for CRD B where only about one out of four HBs is lost, and both GlcNAc and GalNAc remain bound. Interestingly, X-ray data (PDBid 1Y2U) shows that although inversion of the ligands in the CRDs does not occur and that GlcNAc does not bind to CRD A, a disaccharide containing GlcNAc and a Galβ1 (Lacto-N-Biose) binds to an adjacent site and forms a stable complex. The second GalNAc adds between three and four HBs and, therefore, compensates for the losses described above. This shows that single epimeric selectivity can only be achieved by a monosaccharide and requires, as expected, fine tuning of the HO4 protein interactions and a weak binding, since tight binding by multiple interactions as in CRD B or by a disaccharide to CRD A results in the lack of selectivity. Although examination of more cases is needed, the present analysis suggests that carbohydrate selectivity may be more difficult to achieve than previously thought, suggesting the use of carefully chosen non natural configuration glycomimetic compounds as potential drug candidates to be developed and tested.

5. Conclusion

In this work, we present an explanation based on MD simulations, free energy calculations, and solvation structure analyses of ABL lectin carbohydrate binding and epimeric specificity. Our results show that high epimeric specificity of ABL CRDA for GalNAc is mainly determined by His72 and a bridging water molecule. For CRDB epimeric specificity is predicted to be lower and overall higher affinity for GlcNAc is predicted compared to CRDA. Hereby, solvent structure in the unbound CRDs has been proved to be a reasonable tool to obtain critical information regarding ligand structure, allowing the prediction of subtle epimeric specificity in the studied system.

Acknowledgments

Computer power was partially provided by CECAR at FCEN-UBA. This work was supported by Grants from the Argentinean National Agency for the Promotion of Science and Technology PICT 2007-01650 to M.A.M. and PICT 2007-0157 to D.A.E., from the University of Buenos Aires UBA08-X625 to M.A.M., and CONICET-PIP2009 and UBA to D.A.E. D.F.G. and S.D.L. are grateful to CONICET for a Doctoral and post-doctoral fellowships, respectively. M.A.M. and D.A.E. are staff members of CONICET.

Supplementary data

RMSD plots for the simulated systems, radial distribution functions for the definition of binding positions, illustration of the WS positions and hydrogen bond analysis results are provided. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.carres.2011.02.016.

References

1. Varki, A.; Cummings, R.; Esko, J.; Hart, G.; Freeze, H.; Marth, J. In *Essentials of Glycobiology*, 2nd ed.; Cold Spring Harbor Laboratory Press, 1999.
2. Springer, G. F. *Science* **1984**, *224*, 1198–1206.
3. Springer, G. F. *J. Mol. Med.* **1997**, *75*, 594–602.
4. Singh, R. S.; Bhari, R.; Kaur, H. P. *Crit. Rev. Biotechnol.* **2010**, *30*, 99–126.
5. Carrizo, M. E.; Capaldi, S.; Perduca, M.; Irazoqui, F. J.; Nores, G. A.; Monaco, H. L. *J. Biol. Chem.* **2005**, *280*, 10614–10623.
6. Yu, L.; Fernig, D. G.; Smith, J. A.; Milton, J. D.; Rhodes, J. M. *Cancer Res.* **1993**, *53*, 4627–4632.
7. Yu, L. G.; Andrews, N.; Weldon, M.; Gerasimenko, O. V.; Campbell, B. J.; Singh, R.; Grierson, I.; Petersen, O. H.; Rhodes, J. M. *J. Biol. Chem.* **2002**, *277*, 24538–24545.
8. Leonidas, D. D.; Swamy, B. M.; Hatzopoulos, G. N.; Gonchigar, S. J.; Chachadi, V. B.; Inamdar, S. R.; Zographos, S. E.; Oikonomakos, N. G. *J. Mol. Biol.* **2007**, *368*, 1145–1161.
9. DeMarco, M. L.; Woods, R. J. *Glycobiology* **2008**, *18*, 426–440.
10. Bies, C.; Lehr, C. M.; Woodley, J. F. *Adv. Drug Delivery Rev.* **2004**, *56*, 425–435.
11. Pathiaseril, A.; Woods, R. J. *J. Am. Chem. Soc.* **2000**, *122*, 331–338.
12. Agostino, M.; Jene, C.; Boyle, T.; Ramsland, P. A.; Yuriev, E. *J. Chem. Inf. Model.* **2009**, *49*, 2749–2760.
13. Di Lella, S.; Marti, M. A.; Alvarez, R. M. S.; Estrin, D. A.; Díaz Ricci, J. C. *J. Phys. Chem. B.* **2007**, *111*, 7360–7366.
14. Gauto, D. F.; Di Lella, S.; Guardia, C. M.; Estrin, D. A.; Marti, M. A. *J. Phys. Chem. B.* **2009**, *113*, 8717–8724.
15. Li, Z.; Lazaridis, T. *J. Phys. Chem. B.* **2005**, *109*, 662–670.
16. Li, Z.; Lazaridis, T. *J. Phys. Chem. B.* **2006**, *110*, 1464–1475.
17. Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G. *J. Am. Chem. Soc.* **2001**, *123*, 12238–12247.
18. Fadda, E.; Woods, R. J. *Drug Discovery Today* **2010**, *15*, 596–609.
19. Ford, M. G.; Weimar, T.; Kölhi, T.; Woods, R. J. *Proteins: Struct., Funct., Genet.* **2003**, *53*, 229–240.
20. Frank, M.; Schloissnig, S. *Cell. Mol. Life Sci.* **2010**, *67*, 2749–2772.
21. Kadirvelraj, R.; Foley, B. L.; Dyekjaer, J. D.; Woods, R. J. *J. Am. Chem. Soc.* **2008**, *130*, 16933–16942.
22. Tame, J. R. H.; Murshudov, G. N.; Dodson, E. J.; Neil, T. K.; Wilkinson, A. J. *Science* **1994**, *264*, 1578–1581.
23. Sleigh, S. H.; Tame, J. R. H.; Dodson, E. J.; Wilkinson, A. J. *Biochemistry* **1997**, *36*, 9747–9758.
24. Naisimith, J. H.; Field, R. A. *J. Biol. Chem.* **1996**, *271*, 972–976.
25. Loris, R.; Maes, D.; Poortmans, F.; Wyns, L.; Bouckaert, J. *J. Biol. Chem.* **1996**, *271*, 30614–30618.
26. Weisner, S.; Kurian, E.; Prendergast, F. G.; Halle, B. *J. Mol. Biol.* **1999**, *286*, 233–246.
27. Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2008**, *130*, 2817–2831.
28. Robinson, D. D.; Sherman, W.; Farid, R. *ChemMedChem* **2010**, *5*, 618–627.
29. Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.; The Amber biomolecular simulation programs *J. Computat. Chem.* **2005**, *26*, 1668–1688.
30. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712–725.
31. Case, D. A.; Cheatham, T. E.; Darden, T. A.; Gohlke, H.; Luo, R.; Merz, K. M. J.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
32. Corzana, F.; Motawia, M. S.; Herve du Penhoat, C.; van den Berg, F.; Blennow, A.; Perez, S.; Engelsen, S. B. *J. Am. Chem. Soc.* **2004**, *126*, 13144–13155.
33. Constanciel, R.; Contreras, R. *Theor. Chim. Acta* **1984**, *65*, 1–11.
34. Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S.R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P.A. AMBER 10, **2008**.
35. Still, W. C.; Tempczyrk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
36. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
37. Guardia, C.M.A.; Gauto, D.F.; Di Lella, S.; Rabinovichm, G.A.; Marti, M.A.; Estrin, D. A. Submitted for publication.
38. Di Lella, S.; Ma, L.; Rabinovich, G. A.; Díaz Ricci, J. C.; Asher, S.; Alvarez, R. M. S. *Biochemistry* **2009**, *48*, 786–791.
39. Díaz, M. D.; Fernández-Alonso, M. D. C.; Cuevas, G.; Cañada, F. J.; Jiménez-Barbero, J. *Pure Appl. Chem.* **2008**, *80*, 1827–1835.