Characterization of the Galectin-1 Carbohydrate Recognition Domain in Terms of Solvent Occupancy

Santiago Di Lella,[†,‡,§] Marcelo A. Martí,[‡] R. María S. Álvarez,[†,§] Darío A. Estrin,[*,‡] and Juan C. Díaz Ricci[*,†]

Instituto Superior de Investigaciones Biológicas (CONICET), Facultad de Bioquímica, Química y Farmacia, Universidad Nacional de Tucumán, Chacabuco 461, T4000ILI San Miguel de Tucumán, Tucumán, Argentina, Departamento de Química Inorgánica, Analítica y Química-Física (INQUIMAE-CONICET), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, C1428EHA Ciudad de Buenos Aires, Argentina, and Instituto de Química Física, Facultad de Bioquímica, Química y Farmacia, Universidad Nacional de Tucumán, San Lorenzo 456, T4000CAN San Miguel de Tucumán, Tucumán, Argentina

Received: December 28, 2006; In Final Form: March 16, 2007

Human galectin-1, a galactosil-terminal sugar binding soluble protein, is a potent multifunctional effector that participates in specific protein−carbohydrate and protein−protein interactions. Recent studies revealed that it plays a key role as a modulator of cellular differentiation and immunological response. In this work, we have investigated the solvation properties of the carbohydrate recognition domain of Gal-1 by means of molecular dynamics simulations. Water sites (ws) were identified in terms of radial and angular distribution functions, and properties such as water residence times, interaction energies, and free-energy contributions were evaluated for those sites. Our results allowed us to correlate the thermodynamic properties of the ws and their binding pattern with the *N*-acetilgalactoside ligand. These results let us further infer that the water molecules located at the ws, which exhibit much more favorable binding, are the ones replaced by −OH groups of the sugar.

1. Introduction

The binding between two biomolecules requires a significant reorganization of the solvent surrounding the contact surface. Rigidly associated water molecules to these surfaces should vacate their positions in order to allow proper binding. In both protein−protein and protein−ligand complexes, water molecules tend to occupy specific positions and orientations at the protein surface, or at ligand-binding sites, bridging sometimes the interaction between protein and ligand.[1,2] Special attention has been paid to the role of water molecules bound to contact surfaces between carbohydrates and proteins, particularly lectins. In this regard, the X-ray structures revealed the existence of water mediating ligand binding between carbohydrate−lectin complexes, as in the very well-studied Concanavalin A-trimannoside system.[3,4] The study of the role of these water molecules is not straight forward, and several approaches to this phenomenon have been reported. Weisner et al.[5] found that confined waters have residence times in the range of 1 ns to 10^6 ns, while for the more mobile waters, located on the protein surface, residence times are only 10−50 ps. Water molecules in large cavities of proteins, such as the 10−20 waters in the binding cavity of intestinal fatty acid binding protein, have been estimated to have residence times in the order of 1 ns.[5] It was shown that water molecules in binding sites have properties that may aid to ligand entry and binding, including fluctuations in its population, moderate disorder and mobility, reduced water hydrogen-bonding ability, and transient cavities.[6]

From a thermodynamic view point, the role played by water in protein−carbohydrate associations is still a controversial issue. Some authors reported that the displacement of ordered water molecules seems to lower the binding affinity.[7] The standard free energy of tying up a water molecule in the binding site of a protein has been studied by using the double-decoupling method, based on the underlying principle of statistical thermodynamics. This methodology allowed the calculation of the energies involved in transferring a water molecule from bulk water to gas phase and in transferring the water molecule from the binding pocket of the protein complex to the gas phase, by gradually turning off the electrostatic and van der Waals interactions of the water molecule from the rest of the system.[8] In this context, another study by Clarke et al.[9] that evaluated the effects of displacement of the well-ordered waters in Concanavalin A-trimannosides, also showed that rearrangement of these water molecules contributes favorably to the binding affinity. According to estimations carried out by Dunitz,[10] the entropic cost of tying up a water molecule into a cavity of a crystalline protein does not exceed 2.1 kcal mol^{-1} at 300 K. Li and Lazaridis[11] confirmed this view, by means of Molecular Dynamics (MD) simulations. Along with this observation, other authors also demonstrated that ligands designed to displace ordered water molecules exhibited enhanced affinities.[12−14]

All these studies strongly support the idea that water located in binding regions should drive the binding process. Therefore, the study of its properties may yield useful information to get a better understanding of the binding mechanism and energetics. Given the controversial findings mentioned above, it is clear

* To whom correspondence should be addressed. Phone: +54 11 4576 3378, ext. 105. Fax: +54 11 4576 3341. E-mail: dario@qi.fcen.uba.ar (D.A.E.); Phone: +54 381 4248921. Fax: +54 381 4248921. E-mail: juan@fbqf.unt.edu.ar (J.C.D.R.).

† Instituto Superior de Investigaciones Biológicas, Universidad Nacional de Tucumán.

‡ Universidad de Buenos Aires.

§ Instituto de Química Física, Universidad Nacional de Tucumán.

10.1021/jp068989k CCC: $37.00
Published on Web 05/25/2007

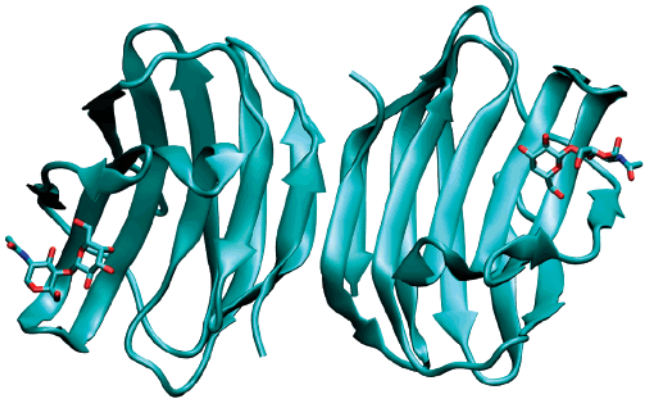

Figure 1. Representation of galectin-1 homodimer, with *N*-acetyllactosamine bound in both CRDs. PDB: 1W6P.

that the role of water in protein−ligand binding is complex and that further study is required. A good model system for this type of studies is galectin-1 (Gal-1).[15] This lectin is a galactosil-terminal sugar binding soluble protein, expressed as a homodimer of 14,5 Kda subunits; the integrity of the dimer is mainly maintained by interactions held between the monomers interface and by well-conserved hydrophobic residues that are part of a hydrophobic core; there are also many hydrogen bonds between the two faces in contact. Gal-1 is found in several tissue types and appears to play an important role in the regulation of cellular growth and differentiation.[16] Recently, important immunological implications have been attributed to galectin-1.[17−20] Natural ligands of this lectin include laminin,[21] fibronectin,[22] CD45 LAMP I and II,[23] α-β integrin,[24] and the glycolipid GM1.[25] Each monomer is a β-barrel consisting in two antiparallel β-sheets, shown in Figure 1. Several galactosil terminated carbohydrate chains bind to the Gal-1 carbohydrate recognition domain (CRD) located at amino acids 30−90. The Gal-1 CRD consists of an antiparallel β-sandwich homologue to other galectins CRDs, usually codified by only one exon. The main amino acids involved in the protein−carbohydrate interaction are His44, Asn46, Arg48, His52, Asn61, Trp68, Glu71, and Arg73.

Comparison of the ligand-bound and ligand-free crystallographic structures of human galectin-1 (h-Gal-1)[26] showed that several crystallographic water molecules are interacting with specific groups of the amino acids in the CRD. Some of these water molecules are replaced by oxygen atoms of the carbohydrate ligand while others are not.

A very promising approach to study the role of water molecules interacting with protein surfaces are MD simulations in explicit solvent. MD simulations provide a detailed description of the structural and dynamical properties of the surface-bound water molecules. X-ray structures of Gal-1 have provided significant information about the structural requirements for ligand binding. However, high-resolution structures cannot yield dynamical details regarding the binding process, whereas computer simulation techniques can provide valuable information not available experimentally. By simulating the protein in explicit solvent, an atomistic picture of the dynamics and the structural mechanism of water molecules association with the protein may be obtained. Moreover, statistical analysis allows computing the thermodynamic contribution of particular molecules or atom groups that are inaccessible through experimental means. Particularly promising along this line is the inhomogeneous fluid solvation theory developed by Lazaridis et al.,[27] which allows computing the thermodynamic contributions of tightly localized individual waters.

In this work we have performed MD simulations of solvated Gal-1. From the MD simulations we were able to precisely define tightly bound water binding sites and to establish a relationship between the thermodynamic properties of the waters that interact with the CRD and the role they play in the binding process.

2. Computational Methods

Molecular Dynamics Simulations. Galectin-1 (human, C2S mutant) coordinates were retrieved from the Protein Data Bank, code 1W6N (X-ray, 1.65 Å resolution).[26] The C2S mutant (wild type cystein replaced by a serine) was used due to the availability of crystallographic data of the protein as well as many protein complexes with different ligands. For the simulation, a single monomer of the dimeric protein was prepared using the LEaP module of AMBER 8,[28] which allowed the addition of hydrogen atoms. The protein was solvated with 4555 three-site point charge (TIP3P) water molecules in a octahedral box of 26 Å of radius, with the addition of four Na^+ ions in order to neutralize the system. No crystallographic water molecule was retained in the initial configuration of the galectin-1. The 20 ns MD simulations were performed using the AMBER force field with the PARM99 set of parameters. In order to remove initial unfavorable contacts, the system was minimized in a 500 cycles run, followed by a 100 ps annealing simulation, during which the temperature was raised from 0 to 300 K. Pressure was then equilibrated at 1 atm over 200 ps. The amino acids protonation states were assumed to correspond to physiological pH. Since histidine residues may exhibit tautomerism, we analyzed more carefully the protonation states of His44 and His52. Both residues were protonated at the N-ϵ epsilon atom. In the case of His44, this is justified by the hydrogen-bond network suggested by Ford et al. for the experimental protein structure, where the HE2 should be involved in a hydrogen bond with O4 of the galactosil residue. His52 protonation state is not evident from the X-ray structure. However, this residue is very exposed to the solvent, and thus it may be assumed that it will be protonated at the ϵ position. This is justified because free histidine forms a mixture with approximately a 4:1 preference for N-ϵ protonation since NE and ND atoms have distinct microscopic pK_a, the NE pK_a being slightly more basic than the ND pK_a.[29]

The simulation used the periodic boundary condition approximation and the particle mesh Ewald (PME) summation method, with an 8 Å cutoff. The SHAKE algorithm was applied to all hydrogen-containing bonds. The MD simulations were performed at 300 K, using a 2 fs time-step. Snapshots of the coordinates were saved every 2 ps, over a total of 20 ns trajectory. The resulting 10 000 instantaneous configurations saved on disk were analyzed in order to obtain thermodynamic parameters.

In separate experiments, simulations were also performed for five isolated amino acids, corresponding to the five amino acids identified in the CRD having a key role in sugar binding. MD simulations were performed using the same computational setup, except for the octahedral solvation boxes, which, in this case, were of 15 Å of radius.

Finally, a MD simulation was performed of an octahedral box of water, in order to evaluate the thermodynamic parameters of water-in-water as reference, with the same force field.

Water Sites Definition. The CRD of Gal-1 has already been described,[26,30] and eight important amino acids were found to be specially relevant in determining binding: histidine 44, asparagine 46, arginine 48, histidine 52, asparagine 61, tryp-

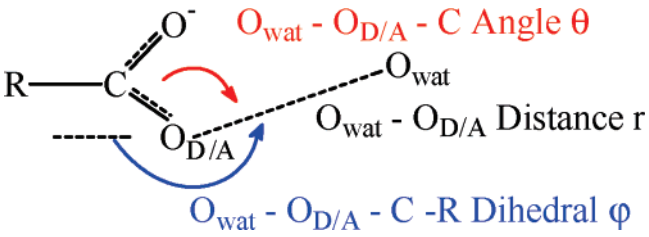

Figure 2. Definition of the θ angle and φ dihedral angle. O_{wat}: water oxygen atom. D/A: donor/acceptor atom.

tophane 68, glutamic acid 71, arginine 73. Each one of these amino acids present one, two, or three atoms which could act as donor or acceptor of hydrogen bonds with water; therefore, they could be used in order to describe the solvent structure around them. Accordingly, 15 sites were described taking as reference the selected donor/acceptor atoms in the above-mentioned amino acids.

Water translational radial distribution functions $g(r)^{tr}$ were calculated for the above listed atoms, by counting the presence of water molecules within a range dr (of about 0.1 Å) around a distance r of the atom. From this data the $g(r)^{tr}$ was normalized considering the accessible volume for the surrounding waters. The accessible volume was determined from both translational angular $g(\theta)^{tr}$ and dihedral $g(\varphi)^{tr}$ distributions. These distribution functions were obtained by counting the number of waters found within dθ around a θ angle and dφ around a φ dihedral angle. For both angular functions only water molecules up to the first solvation shell (as determined from $g(r)^{tr}$) were considered. The θ angle and φ dihedral are defined as follows (see Figure 2): θ is the angle formed by atoms O_{wat}, the donor/acceptor atom of the amino acid, and a reference atom; φ is the dihedral defined by atoms O_{wat}, the donor/acceptor atom of the amino acid, and two reference atoms. The two reference atoms are an atom bound to the donor/acceptor atom of the amino acid, and another atom bound to that first reference, in the amino acid structure.

From the $g(\theta)^{tr}$ and $g(\varphi)^{tr}$ values for each water molecule, bidimensional scatter plots (θ vs φ) were also obtained. This kind of approximation allowed us to describe the spatial distribution of the water molecules along the simulation. We have defined a water site as the volume around a point showing maximum water occupancy probability, which is occupied by a single water molecule at each snapshot. The 15 sites found taking as reference the D/A atoms allowed the description of the water sites. As the definition requires the presence of only one water molecule in the water site (ws), and, in many cases, this water molecule is associated to different D/A atoms, many regions were defined taking as reference more than one D/A atom. All those regions (ws) were then thermodynamically characterized by computing the average thermodynamic parameters over the simulation time.

Hydrogen bond analyses were also performed in order to further characterize the reference atoms that describe the water sites. The fraction of time in which hydrogen bonds between water molecules in water sites and reference atoms are established, as well as the mean distance between donor and acceptor atoms and the mean angles associated with the hydrogen bonds. A water molecule occupying the water site was considered to form a hydrogen bond when distance r (donor−acceptor) was not larger than 3.5 Å, and angle donor−hydrogen−acceptor was not less than 140°.

Calculation of Thermodynamical Parameters. Since the interaction energy between any given water molecule (located in a particular defined ws) and its surroundings fluctuates significantly, the mean value for all water molecules visiting the site, over the whole simulation, is taken as the relevant quantity that contains the information of the thermodynamic properties of the particular ws. Once water sites were identified, for each snapshot along the simulation, van der Waals and electrostatic potential contributions were computed for the water molecule visiting each water site. Both contributions were computed up to a distance of 8 Å. In order to validate this choice, we have performed calculations for larger values of the cutoff radius. The fluctuations in the computed values were less than 0.3 kcal/mol, indicating that the 8 Å cutoff yielded reasonably converged results. Water and protein atoms contributions were evaluated separately in order to discern between the interaction energy of the site with the protein and with the solvent separately. The convergence of the mean interaction energy values was determined by plotting the mean value against the simulation time. (See Supporting Information).

The definition of a ws in the binding region involves the presence of a region where water molecules are more probably to be found than in bulk. First, we defined the probability of finding water molecules in this determined zone, $p(\upsilon)$. The $p(\upsilon)$ function is given by the water finding probability in a volume υ normalized with respect to the bulk solvent density. Although this parameter is able to identify and characterize the ws, it turns out to be convenient also to quantify this magnitude in energetic terms, by defining J as

$$J = -RT \ln p(\upsilon) \tag{1}$$

Specifically, we report results for $p(\upsilon)$ and J computed at a given volume υ centered on the highest probability density found for each water site. In order to make this parameter comparable within different water sites, the values reported in Tables 2 and 3 were estimated at an arbitrarily chosen volume υ of 1 Å^3. The latter is approximately the volume defined by the radial and angular distributions functions and will be the volume assigned to ws.

Water Residence Times Calculation. A wide variation in the definition of residence time can be found in the literature, most of them based on the ideas proposed by Impey et al.,[31] as used here. To obtain a precise definition of residence time for water molecules, a $P_i(t,t_n;t^*)$ function is introduced. This function depicts the coordination state of the water molecule i and is equal either to 0 or 1. It takes the value 1 if the molecule i lies in a defined region of the space (the defined ws) at both time steps t_n and $t + t_n$ and remains in there for a continuous period of t^*. Otherwise, it takes the value of 0. In our case, we have considered the value of t^* equal to the simulation frame time-step used, e.g., 2 ps. Evaluating this function allows us to determine the probability to find a water molecule associated to a determinate atom of the protein during a discrete period of time. The resulting correlation function can be fitted to either one or a sum of exponentials. Although the process of solvent diffusion is complex and may involve several relaxation times, the use of a double exponential model results in an accurate fit in all the evaluated cases. This is in agreement with previous calculations by Makarov et al.[32] The two exponentials would represent two diffusion components observed during the simulations. One fast, represented by water molecules that just enter and leave the first coordination shell; and a slow one, involving water molecules that stay in the site for a prolonged period of time. Thus, $P(t)$ can be expressed as

$$P(t) = W_o[(1 - w)e^{-t/\tau_1} + w\, e^{-t/\tau_2}] \tag{2}$$

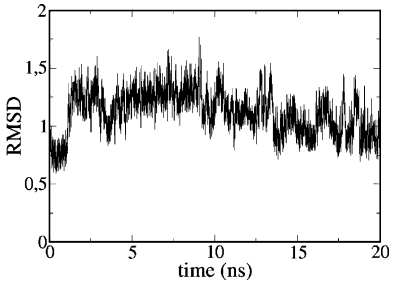

Figure 3. Root-mean-square deviation for the 20 ns MD simulation of Gal-1 under periodic boundary conditions.

where W_0 is a constant, $(1 - w)$ and w are the weights for the fast and slow components, respectively, and τ_1 and τ_2 are the two characteristic residence times.

3. Results and Discussion

Molecular Dynamics of Gal-1. A 20 ns MD simulation of a monomer of Gal-1 was performed. The C-α root-mean-square deviation (rmsd) is shown in Figure 3. We have chosen to perform the MD simulations for only one subunit of the homodimeric Gal-1. However, the results are not affected by this assumption, since the rmsd is smaller than 1.7 throughout almost the whole time scale of our simulation.

Determination of the Water Binding Sites. As already mentioned in the Introduction, the crystallographic structure of h-Gal-1 showed several structural water molecules. Particularly interesting are, for example, the three water molecules that, according to the experimental information, interact strongly with the CRD: water molecules labeled A, B, and C, which interact with the NE of histidine 44 and a NH of the arginine 48, with the ND2 of the asparagine 61 and the OE of glutamate 71, and with both OE of glutamate 71 and both NH of arginine 48, respectively.

During the time scale of a MD simulation, however, many different water molecules interact with particular residues at the protein surface. Also, a specific water molecule may diffuse during the simulation over several places of the protein surface or to the bulk of the solvent. This dynamical behavior makes it difficult to define the interaction of a water molecule with a particular region of the protein surface, as the ones observed in the static X-ray structure. Therefore, in order to characterize the contribution of specific waters located in the CRD of Gal-1, we had first to unambiguously define a water site to compute its thermodynamic properties.

To identify relevant ws in the Gal-1 CRD region, we have computed the $g(r)^{tr}$ for several selected amino acids; then we computed the $g(\theta)^{tr}$ and $g(\varphi)^{tr}$ for all the water molecules corresponding to the first peak in the $g(r)^{tr}$. From the bidimensional θ vs φ scatter plot we identified the maximum water occupancy probability regions that define a ws. For example, Figure 4 shows the $g(r)^{tr}$ using the NH2 atom of arginine 48 as a reference.

By inspection of Figure 4, it can be seen that water is highly ordered at about 2.9 Å; the peak integrates for about three water molecules.

In Figure 5 the scatter plot θ vs φ plot shows that there are three maximum water finding probability positions at about (θ, φ) values of (92°, 170°), (135°, 170°), and (135°, 350°), respectively. The identified ws sites were named ws48NH2a, ws48NH2b, and ws48NH2c, respectively. It is important to note here that these three arbitrary defined water sites present a different degree of localization. That is to say, a water molecule in ws48NH2a may be more rigidly localized, which may be attributed to a configuration of lower translational entropy than the others, taking into account the spatial dispersion of the angular distribution. The three ws explicited above are depicted in Figure 6.

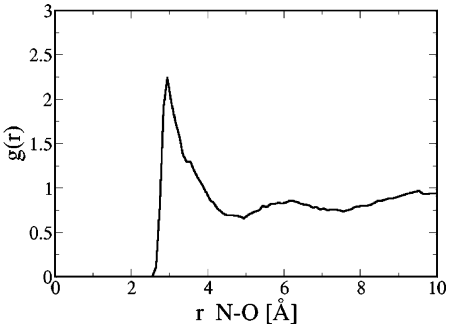

Figure 4. NH2−O (water) radial distribution function for arginine 48. The peak, located at 2.9 Å, integrates for about three water molecules.

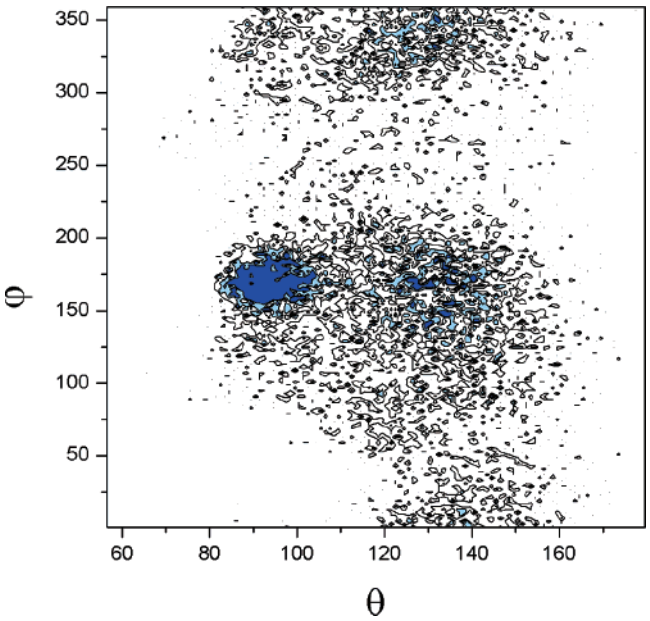

Figure 5. Contour plot of the bidimensional histogram of the angular distribution of water molecules around NH2 of arginine 48 depicting different probabilities. Dark blue corresponds to high probability regions. Angle θ (O_{wat}−$NH2_{Arginine48}$−$CZ_{Arginine48}$) and dihedral φ (O_{wat}−$NH2_{Arginine48}$−$CZ_{Arginine48}$−$NE_{Arginine48}$). CZ: carbon-ζ. NE: nitrogen ϵ. (See Figure 6.)

As can be seen from Figure 6, the same site ws48NH2b should be also identified when using atom NE2 from histidine 44 as a reference. Its correspondent $g(r)^{tr}$ and θ vs φ scatter plots (Figure S1A,B) plots also allowed the identification of this site. In the same way, we can characterize ws48NH2c taking OE2 of glutamic 71 as reference.

By using this analysis with atoms of all the amino acids participating in the binding with the ligand, we were able to identify eight ws in or near the CRD region of Gal-1, presented in Table 1. The corresponding identified ws are shown in Figure 7, with its main reference position listed in Table 1. ($g(r)^{tr}$ and contour plots of the bidimensional histogram of the angular distribution (θ, φ) of water molecules used to identify this site are shown in Figures S1−14, Supporting Information.)

The hydrogen bond analysis performed for these sites and shown in Table 1 also confirmed the definition of an existing water site.

Thermodynamic Properties of the Identified ws. For each site identified we have computed the following properties as

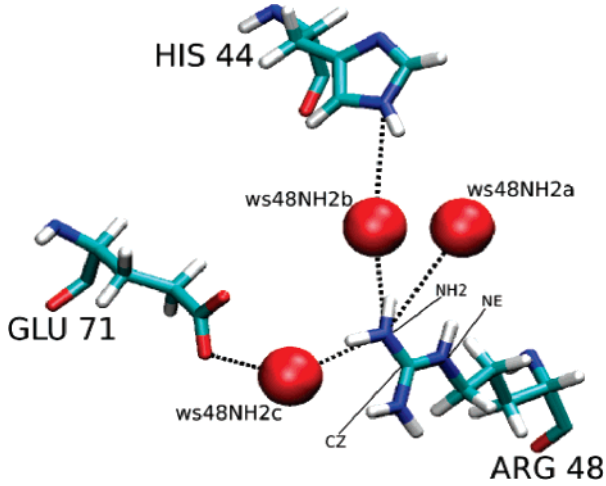

Figure 6. Snapshot of the MD simulation, representing a part of the CRD of Gal-1, showing the three ws, named ws48NH2a, b, and c. Note that ws48NH2b is shared with the one referenced by the NE2 of HIS44, and ws48NH2c, with OE2 of GLU71.

TABLE 1: H Bond Parameters of Water Sites of the CRD of Gal-1[a]

ws name	ref amino acid	ref atom	t_{HB}	$\langle r \rangle$	$\langle \alpha \rangle$
ws1	arginine 48	NH2	14.0	3.018	155.6
	histidine 44	NE2	31.8	3.040	157.0
	asparagine 46	OD1	16.4	2.887	158.4
ws2	asparagine 61	ND2	57.4	3.001	161.6
	glutamic 71	OE1	50.9	2.746	161.3
ws3	arginine 48	NH1	43.5	2.959	157.5
	arginine 48	NH2	12.3	3.010	152.6
	glutamic 71	OE1	17.3	2.926	157.6
	glutamic 71	OE2	46.0	2.733	161.4
ws4	asparagine 46	ND2	15.4	3.104	156.1
	histidine 52	NE2	13.4	3.085	156.0
ws5	histidine 52	ND1	20.6	3.046	157.9
ws6	histidine 44	ND1	21.8	3.042	157.7
ws7	tryptophan 68	NE1	23.4	3.044	155.9
ws8	arginine 73	NH2	37.6	3.010	158.1

[a] t_{HB}, $\langle r \rangle$, and $\langle \alpha \rangle$ represent the percentage of time in which H bonds are established, the mean values of H bond distance and angles, respectively, with respect to the reference amino acids.

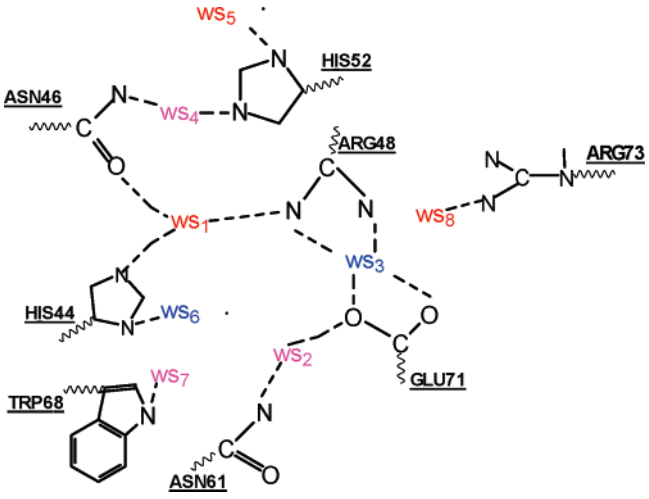

Figure 7. Schematic representation of the CRD of human galectin-1, showing the defined eight ws.

defined in the Computational Methods section: (*i*) the mean interaction energy between the water located at the water site with the solvent, $\Delta E_{w-waters}$, (*ii*) the protein, ΔE_{w-prot}, (*iii*) the whole system, ΔE_{w-all}, and (*iv*) the energy *J*. Table 2 shows the corresponding values for each site.

TABLE 2: Calculated Values for the Interaction Energies, Probabilities $p(v)$, and *J* Energies for the ws Defined Previously[a]

ws name	$\Delta E_{w-waters}$	ΔE_{w-prot}	ΔE_{w-all}	$\Delta\Delta E$	$p(v)$	J
ws1	−12.11	−6.62	−18.73	−1.31	8.166	−1.26
ws2	−5.43	−15.19	−20.62	−3.20	15.130	−1.63
ws3	−3.63	−20.09	−23.72	−6.30	14.154	−1.59
ws4	−13.94	−3.54	−17.48	−0.06	3.669	−0.78
ws5	−11.26	−3.95	−15.21	2.21	2.082	−0.44
ws6	−8.38	−5.93	−14.31	3.11	1.350	−0.18
ws7	−12.34	−7.02	−19.36	−1.94	2.263	−0.49
ws8	−2.01	−22.20	−24.21	−6.79	13.241	−1.55

[a] $\Delta\Delta E$ values are estimated as the difference between ΔE_{w-all} and the energy calculated for a water molecule in bulk water (−17.42 kcal/mol). Values are given in kcal/mol.

TABLE 3: Calculated Values for the Interaction Energies and *J* Energies for the ws Associated to the Listed Reference Atoms of Solvated Amino Acids[a]

amino acid	ref atom	site	$\Delta E_{w-waters}$	ΔE_{w-prot}	ΔE_{w-all}	$\Delta\Delta E$	$p(v)$	J
histidine	ND1	a	−12.58	−6.34	−18.92	−1.50	6.467	−1.21
	NE2	a	−13.35	−7.52	−20.87	−3.45	11.397	−1.46
arginine	NH (1 or 2)	a	−11.57	−12.01	−23.58	−6.16	8.443	−1.28
	NH (1 or 2)	b	−12.30	−10.38	−22.68	−5.26	15.384	−1.64
asparagine	OD1	a	−14.82	−4.03	−18.85	−1.43	7.147	−1.18
	ND2	b	−12.44	−6.23	−18.67	−1.25	4.263	−0.87
glutamic	OE (1 or 2)	a	−7.96	−13.92	−21.88	−4.46	5.659	−1.04
tryptophan	NE1	a	−14.45	−3.85	−18.30	−0.88	9.025	−1.32

[a] Site a or b indicates the water site name when there is more than one defined for the same atom. $\Delta\Delta E$ values are the difference between ΔE_{w-all} and the energy calculated for a water in bulk water (−17.42 kcal/mol). Values are given in kcal/mol.

As observed in Table 2, the water sites named 1, 2, 3, and 8 showed significantly higher $p(v)$ values than the ones presented for water sites 4, 5, 6, and 7. The corresponding differences expressed in energy (*J*) values are less pronounced but still significant. However, no such distinction is shown in the values calculated for the interaction energy, for which the values fluctuate without apparent correlation, although they are of the same magnitude as *J* values. In order to gain some insight on the nature of the values computed in Table 2 we calculated the same parameters for isolated amino acids in a 15 Å solvation sphere. The results are listed in Table 3.

A detailed examination of the MD simulation and the comparison between Tables 2 and 3 can be used to give an interpretation of the observed energy contributions. When carefully analyzing the neighborhoods of ws1 along the simulation, we notice a very localized water site due to three major hydrogen bonds acting either as donor or acceptor. Since the water is rigidly retained here, other water molecules can firmly interact with it in its first solvation shell and therefore increase the ΔE_{w-w} measured. These interactions are shown in Figure 8. The corresponding radial distribution function for this site, with respect to the NE2 of histidine 44, is shown in Figure 9 showing a tightly localized peak. It is also interesting to remark the fact that not one of the three reference residues (Figure 8) alone is able to explain the ws properties, since the thermodynamic values for ARG, ASN, and HIS in Table 3 are clearly different than the values for this ws.

Water sites 2 and 3 are highly localized in their correspondent positions. They show the lowest *J* energy values (−1.63 and

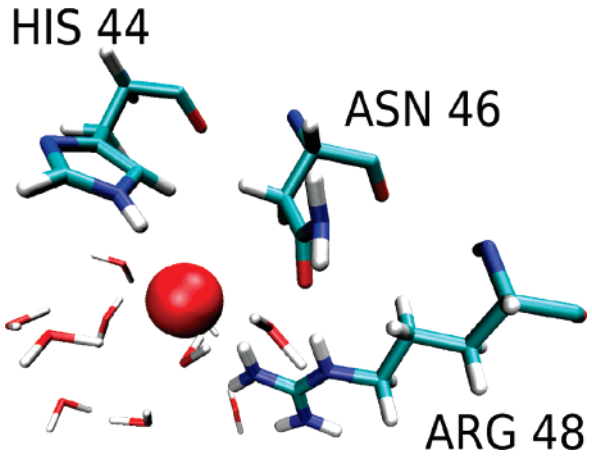

Figure 8. Selected snapshot, showing the interactions between the ws1 (red ball), the amino acids nearby, and the water molecules of the first coordination shell.

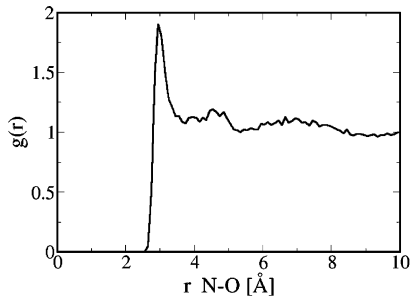

Figure 9. Radial distribution function for the NE2 of histidine 44. The second peak, while not strictly localized, is clearly defined.

−1.59 kcal/mol, respectively). This situation can be explained in terms of the hydrogen bond interactions held by water molecules occupying these sites. In the case of water site 2, water molecules can establish hydrogen bonds with two protein atoms: ASN61ND2 and GLU71OE2, whereas in case of ws 3, with four protein atoms: ARG48NH1, ARG48NH2, GLU71OE1, and GLU71OE2. In both cases, an oxygen atom of a negative-charged glutamate is present, favoring a larger electrostatic interaction, and thus, considerably increasing the observed energy interaction values. Comparing Tables 2 and 3, it can be seen that the trends in the interaction energy and *J* energy contributions are the same in the isolated free amino acids and those in the protein, particularly when looking at the charged residues. Although this observation may not be directly correlated, due to the presence of more than one amino acid reference in the water site, it is clear that the presence of those charged residues is critical in determining the solvation structure.

The results presented for water sites 4, 5, and 6 show an almost zero, or slightly positive $\Delta\Delta E$, meaning that interaction energy in the site is close to that in bulk water. Consistently, the water finding probability is low compared with those favorably interacting with the protein. The situation observed for ws 7 is different. Although it presents a significant negative value for the energy contribution, the resulting *J* energy is just slightly negative.

In ws 8, the very high interaction energy value can be explained in view of the presence of the charged arginine 71 nitrogen. Since, as in ws 2 and 3 the trends in the values are similar to those for the isolated amino acid, here again the presence of the charged center determines the energetic profile of the site.

The hydrogen bond analysis shown in Table 1 is in strict correspondence with these results. Hydrogen bonds were established for significant times in water sites 1, 2, 3, and 8. On the other hand, hydrogen bonds were established less frequently for water sites 4, 5, 6, and 7.

TABLE 4: Water Residence Time for the Selected Water Sites[a]

water site	τ_1 (ps)	τ_2 (ps)
ws1	67.2	454.5
ws2	35.6	207.6
ws3	67.5	268.1
ws4	71.6	362.4
ws5	34.4	188.6
ws6	55.3	322.7
ws7	34.7	203.4
ws8	36.2	204.6

[a] τ_1 and τ_2 represent the fast and slow characteristic times for water diffusion, respectively

Water Sites Residence Times. In Table 4, the water residence times calculated for each water site defined in the CRD of Gal-1 are reported.

The data presented in Table 4 shows that all ws present at least two characteristic residence times, the larger one being between 5 and 10 times the shorter one. The quality of the fit (indicated by a value of r^2 always greater than 0.998) justified the employment of two-exponential fitting. This means that two different processes seem to be contributing to the dynamics of the system. As already stated, the short residence time arises from solvent molecules that enter the site and immediately leave, whereas the longer corresponds to molecules staying in the site for longer periods of time. No clear correlation between any of these values and the thermodynamic properties associated to those sites, or the role they played in ligand binding, is observed. The lack of correlation between residence time values and the surface solvent accesibility[33] and between the spatial (density) and temporal (residence time) order[32] has already been reported for azurin and myoglobin, respectively.

Relationship between Solvation and Ligand Binding. In the first place, we will provide a detailed comparison of water molecules in the experimental X-ray structure and the results of the MD simulations. Water molecules present in the crystal of the unbound protein do correspond to water sites 1, 2, and 3 found in this work. This situation could be expected in view of the favorable interaction energy and *J* values associated with these sites. Nevertheless, ws8, which has also shown a particularly favorable energetic and dynamical profile, does not show a correspondence with any crystallographic water molecule. However, Ford et al. have shown that the crystal structure of Gal-1 bound to N-acetyl-lactosamine exhibits an hydrogen bond between the carbonyl O of GlcNAc-C═O and the NH2 of arginine 73, a position equivalent to ws8 found in the present work.

The most striking fact comes from a comparison of the crystallographic structure of the Gal-1 CRD complexed with its ligand *N*-acetilactosamine and the areas representing the ws in the MD of the Gal-1. The most strongly defined water sites ($p(\upsilon)$ values of more than 8, and *J* values lower than −1.25 kcal/mol, see Table 2) allow to a univocal determination of the replacements occurring during the sugar binding process. In other words, those ws in the CRD of the Gal-1 that are substituted by defined atoms of the carbohydrate are those with higher water probability density. More specifically, since the system has been analyzed in terms of its dynamical behavior, it is not the substitution of a particular water by the carbohydrate that is considered, but the position of a ws later occupied by an

oxygen of the sugar. In this context, it is possible to assign the following substitutions:

- ws1 by Gal-O4
- ws2 by Gal-O6
- ws3 by GlcNAc-O3
- ws4, 5, 6, 7 are not replaced by any ligand atom
- ws8 by GlcNAc-C═O

Clearly the relative *J* interaction energies associated with the ws that are replaced by an oxygen atom of the carbohydrate (ws 1, 2, 3, and 8) are significantly larger than those corresponding to ws not undergoing substitution during ligand binding. Also interesting is the fact that water site 1, which shows the highest residence time for the long process, has been shown by Ford et al. to provide the binding specificity for galactose due to the interactions between Arg-48 and His-44 (here described as water site 1) and the 4-hydroxyl group of the galactosyl residue.[30]

4. Conclusions

The studies summarized here clearly indicate that the analysis of the CRD solvation properties provides significant information about the nature of the ligand binding. Our results show that there is a correlation between the thermodynamics of water molecules, in particular sites of the protein, and the role the ws play in the binding activity. In this context, our results show that it may be possible to predict the sites were oxygen atoms of the ligand will be held in the ligand-bound structure, which is information that may not be easily obtained by experimental tools. Moreover, by inspecting the relative positions of these sites, a prediction of the bound ligand structure may be suggested for carbohydrate binding proteins.

Acknowledgment. This work was partially supported by CONICET, National University of Tucumán, and University of Buenos Aires. S.D.L. thanks D. Bikiel and V. Molinero for revealing discussions.

Supporting Information Available: Figures showing $g(r)$ plots, φ versus θ contour plots, and an interaction energy versus simulation time plot. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Tame, J. R. H.; Murshudov, G. N.; Dodson, E. J.; Neil, T. K.; Wilkinson, A. J. *Science* **1994**, *264*, 1578−1581.

(2) Sleigh, S. H.; Tame, J. R. H.; Dodson, E. J.; Wilkinson, A. J. *Biochemistry* **1997**, *36*, 9747−9758.

(3) Naismith, J. H.; Field, R. A. *J. Biol. Chem.* **1996**, *271*, 972−976.

(4) Loris, R.; Maes, D.; Poortmans, F.; Wyns, L.; Bouckaert, J. *J. Biol. Chem.* **1996**, *271*, 30614−30618.

(5) Weisner, S.; Kurian, E.; Prendergast, F. G.; Halle, B. *J. Mol. Biol.* **1999**, *286*, 233−246.

(6) Henchman, R. H.; McCammon, J. A. *Protein Sci.* **2002**, *11*, 2080−2090.

(7) Mikol, V.; Papageorgiou, C.; Borer, X. *J. Med. Chem.* **1995**, *38*, 3361−3367.

(8) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2004**, *126*, 7683−7689.

(9) Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G. *J. Am. Chem. Soc.* **2001**, *123*, 12238−12247.

(10) Dunitz, J. D. *Science* **1994**, *264*, 670−670.

(11) Li, Z.; Lazaridis, T. *J. Phys. Chem.* B **2005**, *109*, 662−670.

(12) Watson, K. A.; Mitchell, E. P.; Johnson, L. N.; Son, J. C.; Bichard, C. J. F.; Orchard, M. G.; Fleet, G. W. J.; Oikonomakos, N.; Leonidas, D. D.; Kontou, M.; Papageorgioui, A. *Biochemistry* **1994**, *33*, 5745−5758.

(13) Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C.-H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Erickson-Viitanen, S. *Science* **1994**, *263*, 380−384.

(14) Conelly, P. R.; Aldape, R. A.; Wilson, K. P. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1964−1968.

(15) Cooper, D. N. W.; Barondes, S. H. *Glycobiology* **1999**, *9*, 979−984.

(16) Leffler, H. *Trends Glycosci. Glycotechnol.* **1997**, *9*, 9−19.

(17) Toscano, M. A.; Ilarregui, J. M.; Bianco, G. A.; Rubinstein, N.; Rabinovich, G. A. *Medicina (B Aires*) **2006**, *66*, 357−362.

(18) Rubinstein, N.; Ilarregui, J. M.; Toscano, M. A.; Rabinovich, G. A. *Tissue Antigens* **2004**, *64*, 1−12.

(19) Rabinovich, G. A. *Br. J. Cancer* **2005**, *92*, 1188−1192.

(20) Liu, F. T.; Rabinovich, G. A. *Nat. Rev. Cancer* **2005**, *5*, 29−41.

(21) Zhou, Q.; Cummings, R. D. *Arch. Biochem. Biophys.* **1993**, *300*, 6−17.

(22) Andre, S.; Kojima, S.; Yamazaki, N.; Fink, C.; Kaltner, H.; Kayser, K.; Gabius, H. J. *J Cancer Res. Clin. Oncol.* **1999**, *125*, 461−474.

(23) Walzel, H.; Schulz, U.; Neels, P.; Brock, J. *Immunol. Lett.* **1999**, *67*, 193−202.

(24) Gu, M.; Wang, W.; Song, W. K.; Cooper, D. N.; Kaufman, S. J. *J. Cell Sci.* **1994**, *107*, 175−181.

(25) Kopitz, J.; von Reitzenstein, C.; Burchert, M.; Cantz, M.; Gabius, H. J. *J. Biol. Chem.* **1998**, *273*, 11205−11211.

(26) López-Lucendo, M. F.; Solís, D.; André, S.; Hirabayashi, J.; Kasai, K.; Hebert, K.; Gabius, H.; Romero, A. *J. Mol. Biol.* **2004**, *343*, 957−970.

(27) Li, Z.; Lazaridis, T. *J. Phys. Chem. B* **2006**, *110*, 1464−1475.

(28) Case, D. A.; Darden, T. A.; Cheatman, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, D. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. LEaP module of *AMBER 8*; University of California: San Francisco, 2004.

(29) Tanokura, M. *Biochim. Biophys. Acta* **1983**, *742*, 576−585.

(30) Ford, M. G.; Weimar, T.; Kölhi, T.; Woods, R. J. *Proteins: Stuct., Funct., Genet.* **2003**, *53*, 229−240.

(31) Impey, R. W.; Madden, P. A.; McDonald, I. R. *J. Phys. Chem.* **1983**, *87*, 5071−5083.

(32) Makarov, V. A.; Andrews, B. K.; Smith, P. E.; Pettitt, B. M. *Biophys. J.* **2000**, *79*, 2966−2974.

(33) Luise, A.; Falconi, M.; Desideri, A. *Proteins: Stuct., Funct., Genet.* **2000**, *39*, 56−67.